

UNIVERSIDADE DE SÃO PAULO

Escola Superior de Agricultura “Luiz de Queiroz”

Departamento de Ciência Florestais

*Biometria Florestal segundo
o Axioma da Verossimilhança*

Com Aplicações em Mensuração Florestal

Prof. Dr. JOÃO LUÍS F. BATISTA

Quantitative Resource Management/Forest Biometrics, PhD

Tese apresentada à Escola Superior de Agricultura “Luiz de Queiroz” da Universidade de São Paulo (ESALQ-USP) para obtenção do título de Livre-Docente junto ao Departamento de Ciências Florestais.

Piracicaba
2014

Sumário

Motivação	1
I Considerações Históricas	5
1 Considerações sobre a História da Mensuração Florestal	7
1.1 Introdução	7
1.2 O Nascimento da Mensuração Florestal	8
O Contexto Histórico	8
O Surgimento da Ciência Florestal	10
Desenvolvimento da Mensuração Florestal	11
A Mensuração Florestal como Disciplina no Século XIX	15
A Mensuração Florestal no Início do Século XX	17
1.3 As Transformações no Início do Século XX	18
Determinação do Volume de Árvores em Pé	19
Árvore Média	21
De Estimação Ocular a Predição Ocular ¹	23
Determinação da Produção de Povoamentos Florestais	26
Determinação do Crescimento e Produção da Floresta	33
2 Considerações sobre a História da Inferência Estatística	39
2.1 Introdução	39
2.2 Inferência Estatística até 1900	40
Bernoulli e a Lei Fraca dos Grandes Números	40
De Moivre e a Aproximação da Binomial	41
Simpson e o Conceito de Erro	42
Laplace: Probabilidade Inversa e Teorema Central do Limite	42
Bayes e seu Teorema	43
Legendre e os Quadrados Mínimos	43
Gauss, Laplace e Quadrados Mínimos	44
Galton, Pearson e Yule	45
2.3 Biometria e Estatística	46

	Sociedade e Instituto	49
	Ronald Fisher e a sua Definição de Biometria	49
	Biometria Florestal	50
2.4	Inferência Estatística Clássica	51
	Quadrados Mínimos, Regressão e Modelos Lineares	51
	Testes de Significância	52
	Teoria de Neyman-Pearson	53
	Fisher e o Valor-p	56
	Teoria da Amostragem	58
	Intervalo de Confiança	59
	Estatística Matemática	59
2.5	Problemas da Inferência Clássica	61
	Teoria de Neyman-Pearson	62
	Valor-p	63
	Espaço Amostral	64
	Distribuições Amostrais	66
	Probabilidade como Medida de Evidência	69
3	Inferência Clássica e Mensuração Florestal	71
3.1	Introdução	71
3.2	A Estatística nos Textos de Mensuração Florestal	72
3.3	A Influência da Inferência Clássica	76
3.4	Novas Disciplinas	79
	Inventário Florestal	79
	Biometria Florestal	80
3.5	Inferência Clássica na Biometria Florestal: Problemas	83
	Escolha de Modelos para Distribuição de Diâmetros	84
	Estimação e Predição	86
	Especificação e Seleção de Modelos	87
II	Inferência Segundo o Axioma da Verossimilhança	89
4	Inferência e Biometria Florestal	91
4.1	Inferência para Biometria Florestal	91
4.2	Biometria Florestal e Mensuração Florestal	92
4.3	Roteiro de Desenvolvimento	93
5	Cenário Estocástico e Modelagem	95
5.1	Cenário Estocástico	96
	Processo Estocástico	98

5.2	Modelos e Modelagem	99
	Modelo Estocástico	101
	Modelo Estocástico Empírico	102
5.3	Predição	103
	Interpolação e Extrapolação	104
6	Cálculo de Probabilidades	107
6.1	Conceito de Probabilidade	107
	Conceito A Priori	107
	Conceito A Posteriori	108
	Conceito Subjetivo	109
	A Articulação Problemática dos Conceitos de Probabilidade . .	110
6.2	Desenvolvimento Axiomático	111
	Espaço de Probabilidade	111
	Cálculo de Probabilidades	117
6.3	Probabilidade Condicional e Independência	119
	Teorema de Bayes	123
	Eventos Independentes	125
	Independência, Cenário Estocástico e Processo Estocástico . .	128
7	Distribuições Estocásticas	131
7.1	Variável Aleatória e Distribuição Estocástica	132
7.2	Função de Distribuição e Função de Densidade	138
	Distribuição Estocástica Discreta	138
	Distribuição Estocástica Contínua	140
	Família de Distribuições	141
7.3	Famílias de Distribuições Discretas	142
	Bernoulli	142
	Uniforme Discreta	143
	Binomial	144
	Poisson	146
	Geométrica	148
	Binomial Negativa	149
7.4	Famílias de Distribuições Contínuas	152
	Retangular ou Uniforme	152
	Exponencial	155
	Weibull	157
	Gama	160
	Gaussiana	163
	Log-normal	168
7.5	Propriedades das Distribuições Estocásticas	171

Quantis	171
Simetria	174
Transformação de Escala	176
Esperança Matemática	177
7.6 Relações entre Famílias de Distribuições	184
Lei Forte dos Grandes Números	184
Teorema Central do Limite	184
8 Estimação por Verossimilhança	191
8.1 Critério de Estimação	192
Propriedades de um Critério de Estimação	192
8.2 Conceito de Verossimilhança	194
Interpretação da Verossimilhança	194
Definição de Verossimilhança	195
Razão de Verossimilhança	196
Força de Evidência e de Plausibilidade	200
8.3 Função de Log-verossimilhança	202
Axioma da Verossimilhança	206
8.4 Estimativa de Máxima Verossimilhança	208
Estimação por Máxima Verossimilhança	209
Curva de Log-verossimilhança e Incerteza associada à MLE	214
Intervalo de Log-Verossimilhança	222
Propriedades das MLE	224
8.5 Espaço Paramétrico Multidimensional	227
Superfície de Verossimilhança	227
Região de Log-Verossimilhança	229
Parâmetros Inconvenientes	239
Parâmetros Ortogonais	241
Verossimilhança Estimada	244
Verossimilhança Perfilhada	249
Multidimensionalidade e Parâmetros Inconvenientes na Prática	256
9 Especificação e Seleção de Modelos	259
9.1 Introdução	259
9.2 Especificação de Modelos	260
Aspecto Prático	261
Aspecto Teórico	262
Componente Estocástico e Componente Determinístico	264
9.3 Seleção de Modelos	266
Discrepâncias	266
Critérios de Seleção de Modelos: Propriedades Esperadas	273

9.4	Critério de Informação de Akaike	275
	Discrepância de Kullback-Leibler	275
	Discrepância de Kullback-Leibler e Verossimilhança	278
	Variantes do AIC	281
	Delta de AIC e Pesos de Akaike	283
	Inferência por Multimodelos	285
	Critério de Informação de Akaike e Verossimilhança	287
9.5	Validação Cruzada	292
	Critérios de Comparação de Modelos na Validação Cruzada	294
	Análise Gráfica e Validação Cruzada nos Modelos de Predição	308
9.6	Avaliação de Modelos Florestais	313
9.7	Conclusão	314
III	Aplicações da Inferência Segundo a Verossimilhança	315
10	Modelagem da Distribuição de Diâmetros	317
10.1	Introdução	317
10.2	Modelagem da Distribuição de DAP	317
10.3	Problemas na Modelagem de Distribuições	324
	Critérios para Comparação de Modelos e Métodos de Ajuste	325
	Problemas dos Testes de Aderência	328
10.4	Aplicação: Estrutura de Tamanho de Floresta Nativa	332
	A Situação de Aplicação	333
	Dados do Levantamento Florestal	333
	Modelos Candidatos	334
	Estrutura dos Dados e Modelos Candidatos	335
	Seleção dos Modelos	336
	Resultados dos Ajustes dos Modelos	337
	Estrutura da Floresta como Função de Atributos Quantitativos	348
	Conclusão	354
	Apêndices	355
A	Lista de Figuras	357
B	Lista de Tabelas	365
C	Lista de Definições e Teoremas	367
	Referências Bibliográficas	371

Motivação

“There comes a time in the life of a scientist when he must convince himself either that his subject is so robust from a statistical point of view that the finer points of statistical inference are irrelevant, or that the precise mode of inference he adopts is satisfactory. Most will be able to settle for the former, and they are perhaps fortunate in being able to conserve their intellectual energy for their main interests; but some will be forced, by the paucity of their data or the complexity of their inferences, to examine the finer points of their own arguments, and in so doing they are likely to become lost in the quicksands of the debate on statistical inference.

. . .

I find myself, in this debate, like a man who wants to build a house, but sees nothing but sand all around. He consults several geological maps (provided by professional statisticians), only to discover that they are in marked disagreement with each other. But being more interested in a roof over his head than in geology, he determines to build somewhere, choosing the site with a mixture of intuition and hope, being well aware that experience may force him to build again elsewhere, or that one day the geologists will provide a definitive map. In the meantime, if his house prove habitable, he may feel old enough to invite other to build nearby.”

(A.W.E. Edwards, “Likelihood”, Prefácio da Edição de 1972)

Estas citações do prefácio da primeira edição do livro “*Likelihood*”, por A.W.E. Edwards, expressam de forma concisa e, até certo ponto, poética a situação que levou a elaboração deste documento. Assim como Edwards, o autor deste documento também se viu na necessidade de se convencer de que a forma de inferência estatística utilizada na Biometria e Mensuração Florestal era satisfatória.

Mas, sendo a Mensuração Florestal uma forma de inferência em si mesma baseada em vários conceitos quantitativos e que possui uma longa história anterior ao advento da inferência estatística, foi necessário empreender primeiramente

um estudo dos conceitos e procedimentos de Mensuração Florestal que foram estabelecidos antes do século XX. Combinando este estudo com o estudo do desenvolvimento da inferência estatística clássica e seguindo-se o desenvolvimento dos conceitos e procedimentos que surgiram na Mensuração Florestal no século XX a partir da década de 1930, foi possível obter uma panorâmica do processo de desenvolvimento da Biometria e Mensuração Florestal como vem sendo praticada desde as décadas finais do século XX.

Tais estudos compreende a primeira parte deste documento. Embora não sejam estudos históricos rigorosos, eles permitem discriminar com clareza os conceitos e procedimentos da Mensuração Florestal que antecedem o surgimento da Inferência Estatística Clássica daqueles que sofreram intensa influência dos conceitos quantitativos desenvolvidos por ela. A prática da Biometria e Mensuração Florestal atuais não pode ser considerada exatamente como seguidora da inferência clássica, mas estes estudos mostram porque o estado atual de confusão ou conflito entre as diferentes formas de inferência que existe na esfera da Estatística se reflete numa boa medida na esfera da Biometria e Mensuração Florestal. Estas duas esferas são indissociáveis.

A segunda citação do prefácio do livro de Edwards é uma indicação da segunda parte deste documento. A confusão e incoerência reinantes na aplicação da inferência estatística clássica na Biometria e Mensuração Florestal convenceram o autor deste documento de que o problema não está na “construção da casa”, mas sim nos “mapa geológicos” que se contradizem ao indicarem onde os alicerces da casa devem ser assentados. Os argumentos apresentados por Royall (1997) e, na medida da compreensão do autor, por Hacking (1965) são contundentes e não deixam margem de dúvida de que Edwards (1992) foi sábio na “construção da sua casa”. Mas, a contribuição de uma série de estatísticos e pesquisadores atesta a sabedoria do caminho proposto por Edwards, como Birnbaum (1962), Berger e Wolpert (1988), Hilborn e Mangel (1997), Anderson *et al.* (2000), Burnham e Anderson (2001, 2002), Anderson e Burnham (2002), Sober (2002, 2004), Johnson e Omland (2004) e Hobbs e Hilborn (2006). A segunda parte deste documento é, portanto, a tentativa de “construção de uma casa” vizinha à de Edwards (1992), mas sem ignorar os conselhos e indicações de Hacking (1965) e de Royall (1997), buscando “peneirar” da teoria da inferência estatística clássica os “materiais” que se mostraram apropriados.

O objetivo da segunda parte deste documento é apresentar o desenvolvimento completo da inferência baseada em modelos, segundo a abordagem da verossimilhança, para a Biometria Florestal aplicada à Mensuração Florestal. Este desenvolvimento é iniciado com a apresentação de definições básicas a respeito de modelos estocásticos, modelagem, inferência e predição. A partir destas definições, apresenta-se a teoria básica sobre o cálculo de probabilidades, passando-se à descrição de alguns modelos estocásticos (distribuições probabilísticas) importantes

para a Biometria Florestal, apresentando-se também algumas propriedades básicas destes modelos. Segue-se um capítulo sobre estimação por verossimilhança em que o problema da estimação e da inferência sobre os parâmetros dos modelos é tratado unicamente segundo a abordagem da verossimilhança. Este capítulo é o “*cerne conceitual*” da inferência baseada em modelo por verossimilhança, sendo um aspecto particularmente importante a apresentação do Axioma da Verossimilhança. O último capítulo desta parte trata da especificação e seleção de modelos, também privilegiando a abordagem da verossimilhança. Todo o desenvolvimento é realizado de modo um tanto matemático, apresentando-se “definições” dos conceitos essenciais ao desenvolvimento da inferência. O desenvolvimento teórico é ilustrado com exemplos práticos que, hora abordam questões diretamente ligadas à análise de dados, hora abordam problemas teóricos associados a certos tipos de dados ou de modelos em particular.

O objetivo da terceira e última parte deste desenvolvimento era apresentar a aplicação da inferência baseada em modelos, segundo o Axioma da Verossimilhança, a alguns problemas particulares da Biometria voltada à Mensuração Florestal. Por limitações de tempo, somente um capítulo é apresentado, o qual aborda a modelagem da distribuição do diâmetro das árvores com o objetivo de se representar a estrutura dos arvoredos e florestas.

Este documento é antes de tudo uma proposta para o tratamento da Biometria voltada aos problemas da Mensuração Florestal. Ainda que ambicioso, não se trata de um desenvolvimento completo e exaustivo, pois há ainda muitos pontos a desenvolver, assim como existe uma infinidade de problemas de Mensuração Florestal que aguardam serem tratados segundo esta perspectiva. Muitos outros problemas e questões da Mensuração Florestal seriam beneficiados pela revisão de conceitos básicos da mensuração e a reorientação da inferência estatística segundo o Axioma da Verossimilhança. É esperança do autor que isto possa ser realizada nos próximos anos, em colaboração com outros pesquisadores que se sintam motivados pelo tema. Este documento alcançará plenamente seus objetivos se alguns pesquisadores e estudantes se sentirem motivados a aprofundarem seus estudos e pesquisas nesta abordagem, resultando no maior aprimoramento do seu desenvolvimento e no maior número de suas aplicações.

Parte I

Considerações Históricas sobre a Mensuração Florestal, sobre a Inferência Estatística e sobre a influência desta sobre aquela

Capítulo 1

Considerações sobre a História da Mensuração Florestal: Origem e Desenvolvimento

1.1 Introdução

Atualmente, a Biometria Florestal é considerada uma disciplina bem definida e completamente estabelecida como dentro da Engenharia Florestal. Em termos histórico, entretanto, esta concepção é relativamente recente. A Biometria Florestal surgiu com a aplicação dos conceitos e métodos estatísticos na Engenharia Florestal e na Mensuração Florestal, em particular, sendo que a expressão “Biometria Florestal” passou a ser de uso corrente somente na segunda metade do século XX. Já a Mensuração Florestal é bastante antiga, datando do início da Engenharia Florestal como profissão no século XVIII.

A perspectiva histórica é fundamental para a compreensão da realidade. Ela não explica a estrutura das concepções e realizações atuais, mas ela ajuda a entender como as formas de conceber e de realizar vieram a ser, a partir de formas de concepção e realização anteriores. A perspectiva histórica nos ajuda a captar a temporalidade dos conhecimentos e a transitoriedade das práticas humanas. Ela poderá nos auxiliar a visualizar as virtudes e vícios que a Biometria Florestal atual herdou, abrindo a possibilidade de se vislumbrar novos caminhos para a sua renovação constante.

Assim, iniciaremos esta reflexão sobre a Biometria Florestal com algumas considerações históricas sobre o seu surgimento. O objetivo não é uma descrição factual detalhada, nem uma análise histórica rigorosa, mas apresentar considerações que clarifique a origem e o desenvolvimento dos conceitos e concepções dominantes na Biometria Florestal atual.

Neste capítulo, caracterizamos primeiramente de forma bastante breve o surgimento da Mensuração Florestal como disciplina nos estados germânicos do século XVIII e seu desenvolvimento no século XIX. Num segundo passo, analisamos as transformações mais relevantes que ocorreram na Mensuração Florestal na primeira metade do século XX na América do Norte, antes da introdução dos métodos estatísticos. Nos capítulos seguintes, consideramos os aspectos históricos do desenvolvimento da inferência estatística e suas influências sobre a Mensuração Florestal.

1.2 O Nascimento da Mensuração Florestal

O Contexto Histórico

No século XVIII, a “Alemanha” era constituída pelo Sacro Império Romano, não sendo ainda uma nação, como a França, a Espanha ou a Inglaterra. O “Império” era composto por mais de trezentos estados praticamente independentes, cada qual liderado por um príncipe soberano com sua corte e gozando de liberdade política, militar, monetária e religiosa. A união entre os principados era antes linguística, cultura e artística que propriamente política. Dentre os principados, 63 eram eclesiásticos, governados por arcebispos, bispos e abades, enquanto que 51 eram “cidades livres”, sujeitas a uma vassalagem bastante frouxa, semelhante àquela que os príncipes deviam ao imperador que liderava o “Império”. Com exceção da Saxônia e da Bavária, a maior parte do território germânico era cultivada por servos da gleba, isto é, servos legalmente ligados às terras que cultivavam, sujeitos a quase todos antigos deveres feudais (Durant, 1965).

Ao contrário da França, Espanha ou Inglaterra, onde os reis tiveram sucesso em tornar a nobreza subserviente a um governo central, nos estados germânicos, os eleitores, príncipes, duques, condes, bispos e abades privaram o imperador de qualquer poder real sobre os seus domínios e atraíram a baixa nobreza para compor as suas cortes principescas. Como consequência, a riqueza dos proprietários da terra era direcionada a eles, que a gastavam em suntuosidades arquitetônicas e de vestimenta, mas também em arte, cultura, bibliotecas, museus e em empreendimentos científicos. Assim, dentre os estados europeus do século XVIII, os principados germânicos se tornaram líderes em educação. Um exemplo: em 1717, Frederico Guilherme I da Prússia tornou a educação primária obrigatória em seu principado, construindo 1700 escolas durante os 20 anos seguintes (Durant, 1965).

O século XVIII também foi um período de grandes transformações em toda a Europa. A partir do meio do século XVII, a civilização européia medieval, cujo

cerne se assentava na fé, na tradição e na autoridade eclesiástica e monárquica, foi radicalmente questionada com base na razão filosófica a partir de conceitos gerados pela Nova Filosofia e pela Revolução Científica, ambas ligadas entre si e resultantes de um movimento cultural maior chamado de Iluminismo. Este processo generalizado de racionalização e secularização se estabeleceu rapidamente e derrubou a antiga hegemonia da Teologia na condução do estudo do mundo. Gradativamente a crença no sobrenatural foi sendo erradicada da cultura intelectual européia e a filosofia escolástico-Aristotélica foi substituída pela “filosofia mecânica” ou pela “visão mecanicista do mundo”, como fruto do domínio das idéias de René Descartes (Cartesianismo) no ambiente intelectual (Israel, 2001).

O rompimento com o “mundo aristotélico” e, portanto, com a visão de mundo medieval, estabeleceu a visão mecanicista como base para o conhecimento do mundo e desencadeou a Revolução Científica que, embora tenha iniciado na Astronomia, envolveu todas as ciências físicas. A Matemática se transformou na ferramenta essencial: os resultados deveriam ser expressos em números, os argumentos qualitativos eram rejeitados. A Matemática também trouxe a exigência de investigações rigorosas e precisas, o que requer melhores instrumentos, de modo que a fabricação de instrumentos teve um grande desenvolvimento. Estabeleceu-se neste período as ciências como nós as conhecemos (Ronan, 1983).

A partir aproximadamente da década de 1760, a Matemática passou a ter emprego generalizado não só nas ciências, mas também nos mais variados campos da atividade humana, da linguagem ao manejo florestal. O espírito quantificador (*l'esprit géométrique*) incluía a paixão pela ordem, a sistematização, a mensuração e o cálculo. As aplicações do “método matemático”, que era predominantemente geométrico nos séculos XVI e XVII, ampliaram-se com a estabilização da Álgebra e com o desenvolvimento da Análise Matemática como forma de trabalho com o Cálculo Diferencial-Integral. Acima de tudo, a Matemática como instrumento científico de estudo do mundo era o elemento-chave do espírito quantificador (Heilbron, 1990).

Mas a Matemática como instrumento científico requer, além do “software” da teoria, um “hardware” que permita observar quantitativamente a natureza. A segunda metade do século XVIII presenciou um dramático desenvolvimento na fabricação de instrumentos e no aumento da sua precisão. Alguns exemplos demonstram esta dramaticidade. O aumento da precisão na medição de arcos (ângulos), que sofrera um aumento de três vezes no século XVII, teve um aumento de 200 vezes no século XVIII, permitindo aos astrônomos observar fenômenos até então completamente desconhecidos, como a mudança do eixo da Terra em relação ao seu plano de translação. Os relógios e cronômetros tiveram sua precisão aumentada em duas ordens de magnitude, assim o erro na determinação da longitude durante as navegações passou de aproximadamente 100 milhas em 1750, para apenas uma ou duas milhas em 1800. Os termômetros alcançaram a precisão de um centésimo

de grau, enquanto que os barômetros tiveram sua precisão aumentada em 10 vezes em apenas 50 anos. A avalanche de medidas quantitativas, que passaram a ser produzidas pela disponibilidade de instrumentos mais precisos, desafiava a capacidade humana de lidar com a abundância de informações quantitativas. Surgiu então, no último terço do século, a apresentação tabular dos números, como forma de organização racional da informação quantitativa, e as primeiras representações gráficas, como forma de sumarização da informação (Heilbron, 1990).

O Iluminismo, em sua ânsia de racionalização das atividades humanas, produziu uma abordagem instrumentalista no estudo da sociedade e na administração do estado. A proliferação de fatos e figuras sobre a economia, associada à racionalização da administração do estado, geraram a burocratização do aparato financeiro-administrativo dos estados. Este aparato exigia funcionários públicos com domínio numérico e treinados na administração “científica” dos negócios públicos. Nos principados germânicos, este movimento resultou no surgimento das “ciências camerais” (*Kameralwissenschaften*) voltadas a todos os assuntos necessários à administração racional do estado. Já em 1725, elas começaram a figurar no currículo das universidades de Halle e de Frankfurt-an-der-Order (Lowood, 1990).

O Surgimento da Ciência Florestal¹

Até o início do uso do carvão mineral como fonte de energia, que ocorreu no meio do século XIX, a madeira figurou como recurso essencial à vida cotidiana e à atividade econômica. As florestas constituíam uma porção considerável dos principados germânicos e era o maior setor da economia estatal na Europa Central. No início século XVIII, os estados germânicos ainda se recuperavam da devastadora Guerra dos 30 Anos (1618-1648), mas foram novamente arrasados pela Guerra dos Sete Anos (1756-1763) no meio do século. A situação florestal permaneceu problemática e inspirou grande preocupação nos administradores públicos. O temor por uma escassez geral de madeira foi levantado inúmeras vezes ao longo de todo século. Ao contrário das demais nações européias, os estados germânicos permaneceram desunidos e não estabeleceram colônias nas Américas, na África ou na Ásia que pudessem suprir suas necessidades de recursos florestais.

¹A literatura sobre o surgimento da “Ciência Florestal” é bastante escassa, sendo a maior parte dela publicada em alemão. Trabalhos realizados por historiadores, que contextualizem apropriadamente o início das atividades profissionais e científicas florestais em relação à época histórica e aos condicionantes culturais, sociais e políticos são ainda mais raros. O trabalho de Lowood (1990) é uma agradável surpresa, pois se apoia sobretudo em fontes primárias e apresenta o assunto de forma harmoniosa com outras pesquisas históricas do período. Na presente seção e na próxima (Desenvolvimento da Mensuração Florestal) apresentamos uma compilação das idéias deste trabalho, ordenando-as segundo nossos objetivos, mas sem qualquer análise crítica, por isto, as citações sucessivas ao trabalho de Lowood (1990) serão omitidas.

A “Ciência Florestal” (*Forstwissenschaft*) surge assim como uma ciência cameral imperativa para administração pública em vários principados germânicos. O gerenciamento das florestas tornou-se um aspecto da organização do estado, sujeito a organização científica racional que buscava organizar pedaços espalhados de conhecimento em um sistema científico e transformar todos os tipos de atividades previamente guiadas pelos costumes numa ciência. O elemento de união das partes do sistema era a racionalidade econômica.

Ao longo da segunda metade do século XVIII e início do século XIX, as práticas tradicionais dos florestais-caçadores (*holzgerechte Jäger*), sem formação acadêmica e mais compromissados com a condução da caça do que com o gerenciamento da floresta, foram sendo substituídas por técnicas e métodos quantitativos de medição da floresta e por procedimentos técnicos de definição das áreas de colheita e da regeneração natural ou artificial das florestas, segundo a *Forstwissenschaft* nascente. Multiplicaram-se as publicações técnico-científicas que tratavam especificamente do gerenciamento das florestas para funcionários cameralistas pelas mãos dos autores germânicos clássicos como Georg Hartig (1764-1834), Johann Heinrich Cotta (1763-1844) e Johann Wilhelm Hossfeld (1767-1837). Ao final do século surgiram as primeiras escolas específicas para a formação dos profissionais florestais, fundadas por Hans Dietrich von Zanthier (1717-1778) em Wernigerode em 1764, por Cotta na Saxônia em 1785 e por Hartig na Prússia em 1789.

Desenvolvimento da Mensuração Florestal

A Mensuração Florestal se desenvolveu simultaneamente ao surgimento da ciência florestal. A administração racional das florestas almejada pelos funcionários públicos cameralistas, exigia o uso de informações quantitativas sobre as florestas e de procedimentos matemáticos de gerenciamento da floresta. As práticas tradicionais dos florestais-caçadores (*holzgerechte Jäger*) de estimação visual da produção e do gerenciamento da floresta pela sua subdivisão em um número de lotes de mesma área, conforme o número de anos do ciclo de corte ficavam muito aquém das expectativas quantitativas dos cameralistas.

O racionalismo econômico da administração do estado levava os cameralistas a exigir o conhecimento preciso da produção de madeira das florestas a fim de convertê-la em unidades monetárias, isto é, eles buscavam “valorar a floresta” (*Forsttaxation*). Também era necessário um sistema que permitisse a previsão da produção futura da floresta, um sistema flexível o suficiente para considerar as variações climáticas ao longo das décadas e a variação da intensidade de colheita de madeira ao longo do tempo, permitindo a definição de procedimentos de “regulação da floresta” (*Forsteinrichtung*).

O primeiro a tentar um sistema de manejo nestes termos foi um florestal-

caçador da antiga geração que, preocupado com o suprimento de madeira na Saxônia, deu prioridade à madeira sobre a caça. Johann Gottlieb Beckmann (1700-1770) era inspetor florestal e concebeu um sistema de medição da quantidade de madeira em pé que era bastante simples. Uma equipe de auxiliares fazia uma varredura do lote florestal por meio do caminhamento lado-a-lado onde cada auxiliar observava uma faixa de algumas jardas. Cada árvore avistada recebia um prego cuja cor indicava a classe de tamanho dela. Ao final da varredura, contava-se os pregos restantes e se subtraía do número inicial de pregos para se determinar o número de árvores em cada classe de tamanho. A partir do conhecimento da quantidade média de madeira que as árvores de cada classe produziam, se determinava a quantidade de madeira em pé no lote.

Ao redor de 1760, Beckmann começou a divulgar o seu método que logo se tornou conhecido em toda a Alemanha, abrindo caminho para abordagens mais matemáticas. Carl Christoph Oettelt (1727-1802) foi um dos primeiros “florestais matemáticos” (*Forstgeometer*) que, levado pela confiança inabalável na Matemática, típica da Revolução Científica, procurou desenvolver a *Forstwissenschaft* para obter a valoração da floresta com base em princípios gerais e em técnicas baseadas neles. Ele escreveu o primeiro trabalho de mensuração científica de árvores — *Prova prática de que a Matemática executa serviços indispensáveis à Ciência Florestal*² — no qual evoca a geometria: “uma árvore é igual a um cone de base circular”. Os florestais-caçadores tradicionais reagiram afirmando que a geometria florestal era desnecessária, porque “não se mede madeira como se mede ouro”. Mas os *Forstgeometers* corrigiram e melhoraram seus cálculos geométricos numa série de publicações que culminaram no trabalho definitivo de estereometria florestal por Hossfeld — *Estereometria Prática*³ — publicado em 1812.

Em relação à determinação do volume ou massa de madeira de árvores individuais, Cotta adotou uma abordagem diferente da linha geométrica iniciada por Oettelt. Cotta era cético com relação aos cálculos geométricos, mas não da necessidade de quantificação. Ele acreditava que o único meio seguro de determinar o volume ou a massa de uma árvore é cortá-la em pedaços pequenos e medí-los (método destrutivo). Assim, ele construiu as *Erfahrungstabellen* — literalmente “tabelas de experiência”, mais apropriadamente “tabelas empíricas” — onde se tabulava o volume ou massa de árvores individuais em função da altura da árvore, da circunferência do tronco e de classes de idade e qualidade do solo. As tabelas empíricas faziam a ponte entre o formato cônico teórico das árvores e o seu vo-

²Karl Christoph Oettelt, *Practischer Beweis, dass die Mathesis bey dem Forstwesen unentbehrliche Dienste thue*, Eisenach, 1765; também 1786, 1798, 1803.

³Johann Wilhelm Hossfeld, *Niedere und höhere praktische Stereometrie; oder kurze und leichte Messung und Berechnung aller regel- und unregelmässigen Körper und selbst der Bäume in Walde*, Leipzig: Weidmann, 1812

lume medido, permitindo aos florestais o treinamento prático na estimativa ocular do volume da árvore em pé.

Ao final do século XVIII e início do século XIX, os cientistas florestais germânicos haviam definidos os passos para determinar, prever e controlar a madeira das florestas. Em 1804, Cotta publicou a exposição mais clara e mais amplamente lida — *Instruções sistemáticas para a valoração de florestas*⁴ — onde estes passos são definidos como: (1) o “levantamento geométrico”, isto é, topográfico, da extensão da floresta; (2) o cálculo do volume ou massa de madeira de árvores individuais; (3) o cálculo para o povoamento; (4) o cálculo para a floresta como um todo. As taxas de crescimento eram computadas para cada nível de organização: árvore, povoamento e floresta. Estes passos forneciam as informações necessárias ao gerenciamento da floresta, pois para Cotta o problema fundamental era determinar o valor da floresta em pé, dados os custos de manutenção e dada a colheita em ciclos de corte de 100 a 150 anos. Procedimentos semelhantes também foram apresentados no trabalho de Georg Hartig — *Novas Instruções para o Real-Prussiano florestal-matemático e florestal-avaliador*⁵ — publicado em 1819. Cotta e Hartig se tornaram os autores clássicos (*Forstklassiker*) e as principais referências para o Manejo Florestal no século XIX.

Para estes *Forstklassiker*, a Mensuração e o Manejo Florestal eram essencialmente a mesma ciência, dedicada à definição dos procedimentos de “regulação florestal” (*Forsteinrichtung*) que produziria a floresta geradora da máxima produção sob um sistema econômico perfeito. Três princípios, ou *regulae silvarum*, ligavam o ideal da “floresta regulada” e os métodos centrados na mensuração e no cálculo: “diversidade mínima”, “balancete contábil” e “produção sustentada”.

Diversidade Mínima.

As informações necessárias à regulação da floresta estavam escondidas sob a multiplicidade de formas e sob a diversidade de árvores na floresta real. Assim os florestais criaram novas categorias que permitissem a mensuração e o cálculo das informações desejadas: a “árvore-padrão” (*Normalbaum*), a classe de tamanho (*Stärkeklasse*), a parcela amostral (*Probemorgen*) e a classe de idade (*Periode, Altersklasse*).

Hossfeld foi o típico *Forstmathematiker* que buscava obter as informações da floresta por meio do cálculo matemático. Ele definia os seus métodos com base na economia de esforços, argumentando que o produto de algumas medidas baseadas em uma ou duas características criteriosamente observadas era tão boa quanto

⁴Heinrich Cotta, *Systematische Anleitung zur Taxation der Waldungen*, 2 vols., Berlin, 1804.

⁵Georg Ludwig Hartig, *Neue Instructionen für die Königlich-Preussischen Forst-Geometer und Forst-Taxatoren, durch Beispiele erklärt*, Berlin: In Commission bey der Kummerischen Buchhandlung, 1819.

somas de grande quantidade de medidas que demandavam um grande tempo de trabalho no campo.

Cotta acreditava que o florestal treinado para identificar a árvore-padrão no campo e para memorizar as tabelas empíricas, podia definir rapidamente o volume ou a massa de madeira de um talhão da floresta. Hartig defendia a estrita aderência aos resultados de algumas poucas parcelas e a “métodos confiáveis”. Cotta concordava com Hartig na necessidade de ignorar detalhes discrepantes e se concentrar nos números derivados de uma parcela, argumentado que a mensuração seletiva gerava resultados aceitáveis para produção e crescimento da floresta e que não é necessário se preocupar com o acúmulo de erros, pois as diferenças individuais se cancelam no agregado.

Balancete contábil.

Os cameralistas e os cientistas florestais tinha ambos fé nos números, mas divergiam sobre a premissa subjacente a esta fé. Oettelt, Hossfeld e Cotta viam o gerenciamento da floresta como dependente da Matemática e da natureza, não o contrário. Já para os cameralistas, o papel da quantificação da Ciência Florestal era descritivo e não prescritivo. O ponto de acordo entre ambos era a conversão da quantidade de madeira em valor da floresta, mas a partir daí cameralistas e florestais se separavam. O cameralista fazia o orçamento monetário do estado (*Geld-Etat*), enquanto que o florestal fazia um orçamento que comparava a produção com o que a floresta realmente possui ao longo do tempo (*Forst-Etat*). Vários autores (Friedrich von Burgsdorf, Hartig, Cotta) descrevem o Manejo Florestal como operações análogas a um contador mantendo o balancete contábil. Hartig descreve a tarefa de construção do *Forst-Etat* como a busca de um equilíbrio: “onde estiver faltando um balancete seguro, baseado na Matemática e na Filosofia Natural, sempre haverá super ou subutilização da madeira”.

Produção Sustentada.

A produção sustentada (*Nachhaltigkeit*) figura com conceito chave do Manejo Florestal em praticamente todos autores germânicos que trataram do assunto. A produção, ao contrário da massa de madeira de um lote florestal, não é uma quantidade determinada pela natureza, é necessário definir procedimentos adequados para obter uma produção que seja permanente, certa e relativamente equitativa ao longo do tempo. A produção sustentada requer um sistema de manejo baseado na mensuração, predição e no planejamento. Mesmo face o desafio da predição do comportamento da floresta, os cientistas florestais germânicos não se intimidaram e propuseram várias abordagens para definir os procedimentos preditivos e prescritivos para a regulação da floresta que, baseados na *Forstmathematike*, estabeleciam uma estratégia de longo prazo para regeneração e corte da floresta que

resultaram na “floresta normal”. Ao final do século XIX, o resultado da aplicação destas abordagens pode ser observado na grande uniformidade da estrutura das florestas e na extrema regularidade do arranjo espacial dos lotes florestais na paisagem germânica.

A Mensuração Florestal como Disciplina no Século XIX

Até o meio do século XIX, a literatura florestal européia foi totalmente dominada pelos trabalhos germânicos. A esta altura a *Forstwissenschaft* já estava consolidada e serviu de inspiração e modelo para os reformadores das práticas florestais em outras nações, como França, Inglaterra e Estados Unidos da América, que durante o século XIX sentiram a necessidade da conservação e do manejo das florestas baseada em princípios científicos e executada florestais profissionalmente treinados (Fernow, 1907; Lowood, 1990).

No caso dos Estados Unidos, a Ciência Florestal foi difundida tanto por norte-americanos com treinamento na Alemanha, como Gifford Pinchot (1865-1946), quanto por alemães que emigraram para a América do Norte, como Carl Alwin Schenck (1868-1955) e Bernhard Eduard Fernow (1851-1923) (Lowood, 1990). Pinchot foi o primeiro chefe do Serviço Florestal estado-unidense⁶ (National Wildlife Federation, 2012). Schenck partiu da Alemanha para assumir a administração de uma propriedade florestal — “*George W. Vanderbilt’s Biltmore Estate*” — onde fundou em 1898 a primeira escola para treinamento profissional florestal dos Estados Unidos: a “*Biltmore Forest School*” (Forest History Society, 2008; Forest History Society Library and Archives, 2008). Fernow precedeu Pinchot no que era a Divisão de Florestas do Departamento de Agricultura dos Estados Unidos (*USDA Division of Forestry*). Em 1898 deixou esta posição, tornando-se o primeiro diretor do “*New York State College of Forestry*”, na Universidade de Cornell, a faculdade dos Estados Unidos com um programa universitário de 4 anos de formação profissional florestal (Forest History Society, 2009).

A Ciência Florestal trazida por estes e outros profissionais já havia se estabelecido dentro de um paradigma de princípios gerais e disciplinas bem definidas. Esta Ciência Florestal teve forte influência sobre o movimento de conservação das florestas e na administração dos recursos florestais nos Estados Unidos. Influência que foi duradoura porque as novas gerações de profissionais foram formadas com base nela. Fernow, numa série de palestras apresentadas na *Yale Forest School* em 1907, sobre a história da prática florestal, apresenta uma coletânea detalhada do progresso alcançado pelos florestais germânicos no século XIX (Fernow, 1907), afirmando que a Mensuração Florestal já se tornara uma disciplina estabelecida e bem desenvolvida com seus métodos e instrumentos. Resumidamente, os elemen-

⁶US Forest Service.

tos principais que Fernow atribui à Mensuração Florestal são:

Mensuração do diâmetro e altura de árvores. Instrumentos de medição de diâmetro já era utilizados desde o início do século XIX, segundo descrições realizadas por Cotta, em 1804, e por Hartig, em 1813. A melhoria das sutas até a eficiência conhecida no início do século XX, tempo de Fernow, foi desenvolvida por Carl Gustav Heyer (1797-1856) a partir de 1840. Instrumentos satisfatórios para medição da altura já era descritos por Hossfeld, em 1800, seguindo-se um período de grande desenvolvimento que alcançou o hipsômetro de espelho de Martin Faustmann (1822-1876). Instrumentos para medição de diâmetros a diferentes alturas do tronco tinham sido desenvolvidos por Max Robert Pressler (1815-1886) e Karl Breymann (1807-1870). A técnica de obter a altura da árvore a partir do diâmetro é atribuída a Gottlieb Koenig (1779-1849).

Mensuração do volume de madeira em árvores e pilhas. O uso de fórmulas de cubagem de toras estão associadas aos nomes de Oettelt, Heinrich Ludwig Smalian (1785-1848), Franz Xaver Huber (1769-1842), Hossfeld e Pressler. O método do deslocamento de água já era conhecido desde 1782 devido aos trabalhos de Karl Wilhelm Hennert (1739-1800) e a medição do volume de madeira e de casca numa pilha já era ensinado por Oettelt, no século XVIII.

Determinação do volume de árvores em pé. O método de dedução por fórmulas de sólidos geométricos, iniciado por Oettelt, foi desenvolvido por Hossfeld, Pressler e Adalbert Schiffel (1851-1914). As “*tabelas de volume*”, que já tinham um certo uso no século XVIII, passam a ser generalizados com as tabelas publicadas por Cotta, em 1804 e 1817, e por Koenig, em 1813 e 1840. O fator de forma foi inicialmente utilizado por Huber para classificação das árvores em classes de forma. Hossfeld é creditado com a primeira fórmula para o cálculo de fator de forma em 1812, enquanto que Smalian, em 1834, introduziu o conceito de fator de forma “normal” ou verdadeiro. O fator de forma foi amplamente utilizado na construção de várias tabelas de volume. Já o nome Schiffel é associado ao quociente de forma como método de caracterização da forma do tronco das árvores.

Determinação do Crescimento da Floresta. A idéia do uso de parcelas permanentes para o acompanhamento do crescimento das árvores é atribuída a Hossfeld, em 1823, que também propôs um sistema de interpolação das medidas das parcelas no caso em que o número de medições era limitado. Huber publicou em 1824 a primeira “tabela de produção normal” baseada na “árvore média” (*Normalbaum*) como índice do povoamento. As primeiras “tabelas de incremento relativo” (*increment per cent tables*), como forma de estimar o crescimento futuro com base

na performance passada de árvores individuais, foram publicadas por Hartig, em 1795, e por Cotta, em 1804. Em 1860, Pressler ampliou as possibilidades de estudo do crescimento com a criação da sua sonda ou trado, que permite estudar os anéis de crescimento da árvore sem destruí-la.

A Mensuração Florestal no Início do Século XX

No início do século XX, os princípios, métodos e técnicas desenvolvidas na Alemanha já faziam parte da base educacional dos florestais na América do Norte, como atestam os livros texto de Schenck (1905), Graves (1906), Chapman (1912), Recknagel e Bentley (1919), Winkenwerder e Clark (1922), Chapman (1924), Belyea (1931) e Chapman e Demeritt (1932). Estes livros eram destinados à formação universitária de florestais e já apresentavam em grande parte a Mensuração Florestal na forma como é conhecida hoje. Sua análise permite identificar os procedimentos de Mensuração Florestal que no início do século XX eram a base do exercício profissional na América do Norte.

Determinação do Volume de Árvores Abatidas. Utilizava-se os métodos e fórmulas de cubagem aplicado a toras ou porções do tronco, inclusive para a determinação do volume de casca. Algumas regras de cubagem já eram utilizadas e se destacava a importância do estudo da forma do tronco por meio de fatores de forma e de quocientes de forma.

Determinação do Volume de Árvores em Pé. Os instrumentos e seus princípios básicos de funcionamento são claramente apresentados no caso da medição do diâmetro e da altura das árvores. Distingue-se claramente a aplicação de fatores de forma para determinação do volume de árvores em pé. Detalha-se a construção e aplicação dos três tipos básicos de tabelas de volume: tabelas locais, tabelas de dupla entrada e tabelas de classe de forma. Também se apresenta a construção e aplicação de *tabelas de forma*, análogas às atuais equações de forma ou de afilamento. Discute-se a construção e aplicação da relação hipsométrica como forma de se determinar a altura das árvores a partir da medição do diâmetro.

Levantamento em Povoamentos Florestais. Descreve-se o uso de parcelas retangulares ou circulares (“*plots*”) e parcelas em faixa (“*strips*”) como meio de captar tanto o padrão dominante quanto a variabilidade natural do povoamento. Mostra-se claro o conhecimento das consequências da intensidade amostral (“*sampling intensity*”) sobre a acurácia das estimativas. Indica-se o uso do delineamento sistemático na amostragem por parcelas e parcelas em faixa, baseado na praticidade e na cobertura espacial das parcelas sob neste delineamento. A discussão

dos métodos de levantamento mostra já se tinha conhecimento dos principais fatores que causam a variabilidade natural da produção de madeira em povoamentos e como as parcelas e faixas devem ser locadas no campo de modo a amostrar adequadamente tais fatores. A aplicação da estratificação na amostragem dos povoamentos (*amostragem estratificada*) é apresentada como meio de se levantar povoamentos com grande variabilidade interna.

Estudo do Crescimento de Árvores e Povoamentos Florestais e sua Aplicação no Manejo Florestal . Indica-se a utilização de parcelas permanentes para medir o incremento de árvores e florestas por meio da remedição das árvores individualmente. Discute-se a necessidade da observação dos componentes do crescimento das florestas: crescimento das árvores sobreviventes, mortalidade, ingresso e colheita. Apresentam-se aplicações e limitações da análise de tronco para estudo do crescimento e incremento de povoamentos. São apresentados conceitos de incremento — incremento periódico, incremento corrente anual e incremento médio anual — e a forma de sua aplicação no manejo dos povoamentos. Apresenta-se a forma atual de determinação da idade técnica de corte.

* * *

Esta lista, apesar de não ser exaustiva, é suficiente para mostrar que o formato básico da Mensuração Florestal, como é conhecida hoje, já estava solidamente estabelecida no início do século XX na formação dos florestais na América do Norte. Havia, no entanto, alguns procedimentos que eram totalmente dependentes da experiência do mensuracionista, pois se apoiavam no seu julgamento pessoal, enquanto que uma profusão de conceitos, muitas vezes contraditórios, geravam um certo estado de confusão teórica. Mas no início do século XX, a Mensuração Florestal passou por transformações que a depuraram da maioria destes problemas.

1.3 As Transformações no Início do Século XX

A Mensuração Florestal não permaneceu inalterada no século XX. Principalmente na segunda metade do século, o desenvolvimento tecnológico permitiu o desenvolvimento de uma grande gama de instrumentos e ferramentas mais precisos e práticos. Sutas, clinômetros, bússolas e teodolitos sofreram grandes transformações. Mas o destaque tecnológico principal talvez deva ser atribuído à tecnologia digital em geral e ao computador eletrônico em particular. Esta última ferramenta revolucionou as concepções de coleta, armazenamento e processamento das medidas obtidas em campo. A tecnologia digital introduziu a concepção atual de

“*dado*”, permitindo a manipulação de uma quantidade de medidas e informações obtidas em campo, com a qual os florestais do início de século XX jamais sonhariam ser possível trabalhar.

Mas antes que o impacto tecnológico pudesse ser claramente sentido, o que ocorreu, principalmente depois da Segunda Guerra Mundial, a intensificação e disseminação da prática profissional florestal gerou na América do Norte mudanças ainda no primeiro terço do século XX. A análise mais minuciosa dos mesmos livros-texto discutidos acima atesta tais transformações. Na explanação que se segue, não se visa de produzir uma lista exaustiva das mudanças, mas analisar como alguns conceitos e métodos da tradição germânica foram profundamente transformados ou mesmo abandonados ainda no início do século XX.

Determinação do Volume de Árvores em Pé

A *tabela de volume* é talvez o exemplo mais claro de princípios e métodos desenvolvidos no século XIX que estão em uso na Mensuração Florestal nos dias atuais. Segundo Spurr (1952), o crédito para concepção atual de tabelas de volume é dado a Cotta que em 1804 publicou uma tabela de volume para faia e 1817 publicou um conjunto de tabelas padronizadas. Cotta inicia seu trabalho estabelecendo um postulado que ainda é a base fundamental para o desenvolvimento de equações de volume. Spurr apresenta este postulado da seguinte forma:

“Tree volume is dependent upon diameter, height and form. When the correct volume of a tree has been determined, it is valid for all other trees of the same diameter, height and form.”

(Spurr, 1952, p.56)

As tabelas de volume de Cotta forneciam o volume de um cone para os diversos diâmetros e alturas dos troncos. Uma segunda tabela fornecia fatores para conversão do volume cônico em volume sólido das árvores. Esta segunda tabela estava organizada por espécie, tendo para cada espécie 5 classes diferentes de forma. Na linguagem atual, as tabelas de Cotta informavam o volume sólido dos troncos em função de um “*fator de forma cônico*” (Spurr, 1952).

Atualmente as tabelas de volume foram substituídas pelas *equações de volume*, o que dispensa o uso de conversão de uma ou várias tabelas. Entretanto, o postula de Cotta ainda é a base fundamental para a determinação do volume de árvores em pé, tendo sido generalizado inclusive para a determinação da biomassa, com o acréscimo do conceito de densidade básica da madeira. As tabelas de volume de Cotta também representam uma abordagem essencialmente empírica para a questão. Duzentos anos após os trabalhos de Cotta, a abordagem totalmente empírica permanece atual.

Mas o predomínio das tabelas de volume ainda não existia no início do século XX. As tabelas de volume ainda conviviam com outras formas mais simples de se predizer o volume de árvores em pé, apesar de, em 1896, Pinchot e Graves (1896) terem publicado tabelas de volume e produção para *Pinus strobus* (*white pine*) na região da Pennsylvania e Nova Iorque, talvez as primeiras tabelas na América do Norte. Além da abordagem de tabelas empíricas, os florestais lançavam mão de várias formas de determinação teórica do volume, segundo propostas de florestais alemães do século XVIII e XIX, em que o volume do tronco da árvore como um todo é aproximado pelo volume de um sólido, geralmente o cone ou o parabolóide.

As fórmulas mais citadas na literatura Norte Americana do início do século XX são as aproximações de Hossfeld, Schiffel e Pressler:

$$\text{Hossfeld:} \quad v = \frac{h}{4} (3 g_{h/3} + g_0)$$

$$\text{Schiffel:} \quad v = \frac{h}{6} (4 g_{h/2} + g_0)$$

$$\text{Pressler:} \quad v = \left(\frac{2}{3}\right) g_0 h_{d/2}$$

onde

g_0 — área transversal à base da árvore,

$g_{h/2}$ — área transversal do tronco à metade da altura total,

$g_{h/3}$ — área transversal do tronco a um terço da altura total,

h — altura total da árvore, e

$h_{d/2}$ — altura onde o diâmetro do tronco é metade do DAP.

Note que nas expressões acima o volume de madeira é matematicamente deduzido das medidas de diâmetro do tronco e altura da árvore.

Schenck (1905) apresenta três métodos de determinação do volume de árvores em pé: (i) estimacão ocular, (ii) mensuração do diâmetro e da altura e (iii) o método do fator de forma, que combinaria estimacão e mensuração. Ele designa o método da mensuração do diâmetro e da altura como sendo “científico”, que consiste na aplicação das fórmulas acima com as medidas de diâmetro e altura. Curiosamente, ele conclui o capítulo que trata da determinação do volume de árvores em pé listando uma série de publicações de tabelas de volume desenvolvidas para espécies da América do Norte. Mas ele não enumera formalmente a abordagem de tabela de volume como um método.

Graves (1906), além da estimativa ocular, comenta a predição por tabelas de volume e fatores de forma, mas apresenta as fórmulas acima como o *método de fórmulas*, embora dedique um capítulo à discussão de tabelas de volume. Recknagel e Bentley (1919) apresentam cinco métodos para determinação do volume de árvores em pé, onde o uso de tabelas de volume e fatores de forma é listado juntamente com as fórmulas de Schiffel e Pressler, bem como um método de estimativa ocular.

Já em Chapman (1924), a fórmula de Schiffel aparece numa apresentação histórica do desenvolvimento do conceito de quociente de forma. Em Belyea (1931), as fórmulas são apresentadas como nota de rodapé também na discussão de quociente de forma.

O método de tabelas de volume aparece como método preferencial para determinação do volume de árvore em pé no livro “*Elements of Forest Mensuration*” de Chapman e Demeritt (1932). Neste livro, cinco capítulos completos são dedicados ao tema, que é desenvolvido da seguinte forma: (i) construção de tabelas de volume locais, (ii) construção de tabelas de volume de dupla entrada (“*standard volume tables*”), (iii) construção de gráficos de alinhamento (“*alignment charts*”) para tabelas de dupla entrada, (iv) tabelas de dupla entrada para volume de pranchas (“*board feet*”), e (v) estimativa do volume de tábuas por meio de tabelas de volume.

Os livros mostram que houve uma transformação na concepção da melhor abordagem para a determinação do volume de árvores em pé. Apesar de Cotta ter publicado a primeira tabela de volume em 1804 e Schenck (1905) listar uma série de publicações de tabelas de volume na América do Norte, a dedução teórica pelas fórmulas de Hossfeldt, Schiffel e Pressler foi inicialmente considerada como um método de determinação do volume tão bom quanto o método empírico da tabela de volume. Mas estas fórmulas foram gradativamente substituídas pelas tabelas de volume, e a abordagem de predição empírica se transformou no método preferencial para determinação do volume de árvores em pé.

Árvore Média

A *árvore média* (“*average tree*”) é um conceito dominante na Mensuração Florestal no início do século XX, talvez ainda sobre a influência do conceito da *árvore padrão* (“*Normalbaum*”) dos florestais alemães do século XVIII. Para compor uma estimativa da produção num povoamento, os florestais de então frequentemente tomava a *situação média* do povoamento com a árvore média, ou a árvore de *volume médio*. A árvore média também é o método mais natural em levantamentos onde a produção é obtida por meio da contagem das árvores apropriadas para colheita ou venda.

O uso da árvore média estava frequentemente associado à estimativa ocular.

Schenck (1905), identificava três métodos para estimar a produção de um povoamento: (1) estimacão (ocular), (2) mensuracão e cálculo exato, e (3) combinacão de estimacão e mensuracão, e acrescenta que a mensuracão só é possível em florestas com poucas árvores com defeitos. Este autor detalha o método de estimacão da seguinte forma:

“The volume of a wood is ascertained by cruisers’ estimates in the following ways:

- a. By estimating the number of trees and the volume of the average tree with due allowance for defects.
- b. By counting the trees and estimating the volume of average trees with allowance for defects.
- c. By estimating the volume of each tree separately, sounding it with an axe, when necessary, and judging its soundness from all sides.

The above methods (a, b, c) are applied either to sample plots or to sample strips or to the entire area.”

(Schenk, 1905, p.26)

Graves (1906) mostra que o conceito de árvore média e a forma de obtê-la eram bem mais flexíveis no início do século XX:

“ There are several ways of estimating the volume of the average trees required in these methods of rough sample areas. Some estimate by the eye the average yield per tree. Some estimate the average number of logs per tree, and knowing from the experience at the local sawmills the average contents of the logs, determines the average yield of the standing trees. Another way is to select several trees of average size, estimate their volume, and use the average of these as the required average yield per tree in the forest.”

(Graves, 1906, p.193)

Chapman (1924) detalha as “seis classes de média empregadas na estimacão da produçã”, mostrando que a árvore média ainda era um conceito chave na determinacão da produçã de povoamentos e florestas. Já Recknagel e Bentley (1919), após apresentarem vários métodos baseado em parcelas, apresentam o método da *árvore amostral média*, mas enfatiza a necessidade de que este método seja aplicado para cada espécie separadamente.

Note que a palavra “estimacão” (*estimating*) é utilizada em profusão por estes autores. Mas a conotacão desta palavra para eles é mais próxima a conotacão

coloquial atual do que do termo técnico biométrico corrente. Para eles, estimar representa uma forma se obter um valor numérico por um processo não determinado, o que deve ser entendido como “estimação ocular”. A árvore média era essencialmente um resultado de diversas técnicas baseadas na estimação ocular, seja via estimação direta da árvore média, seja via estimação de componentes que geram a árvore média.

Belyea (1931) representa uma mudança nesta concepção, pois ele coloca forte ênfase na determinação das áreas dos povoamentos (“*forest survey*”) e distingue o processo de determinação da produção em áreas pequenas e grandes. Em áreas pequenas, o uso da árvore média é considerado como uma das abordagens possíveis, já no caso das áreas grandes, os exemplos enfatizam a necessidade de se obter um *total* para as unidades amostrais e, pelo conhecimento da área do povoamento, se obter a produção do povoamento. É interessante o detalhamento que Belyea faz dos diferentes tipos de unidades amostrais e sua instalação no campo, discutindo tamanho, forma e número de unidades amostrais para se gerar uma boa determinação da produção florestal. Nesta perspectiva, a importância da árvore média e da estimação ocular é fortemente reduzida.

Assim como a estimação ocular, a importância da árvore média na determinação da produção de povoamentos e florestas foi gradativamente reduzida. A determinação do volume por meio do total nas unidades amostrais e da área do povoamento ou floresta tornou-se o procedimento padrão. O conceito de árvore média não é encontrado nos livros de Mensuração Florestal do final do século XX, como por exemplo os livros de Husch, Millers e Beers (1982) e de Avery e Brukhart (1983).

*De Estimação Ocular a Predição Ocular*¹

A “estimação ocular” (“*ocular estimation*”) ou “juízo pessoal” (“*personal judgement*”), que consiste no julgamento subjetivo do profissional florestal para gerar uma estimativa da produção do povoamento, é talvez o exemplo mais flagrante de um procedimento profissional considerado padrão e que foi abandonado.

Nos livros que datam das primeiras décadas (Schenck, 1905; Graves, 1906; Chapman, 1912), a referência à estimação ocular é bastante clara, denotando a aceitação geral do método. Ainda no livro de Chapman de 1924, a estimação ocular direta é colocada como uma das abordagens válidas para estimar a produção

¹Os termos *estimação* e *predição* são frequentemente foco de confusão na Engenharia Florestal em geral e no Brasil em particular. Neste trabalho, o termo *estimação* será utilizado para designar o processo de obtenção de uma *estimativa* para um parâmetro populacional por meio de modelagem estocástica em geral, o que na maioria dos casos implica na amostragem. Já o termo *predição* será usado para designar o processo de obtenção do valor para uma variável aleatória por meio de um modelo estocástico que substitua a mensuração, como por exemplo uma relação hipsométrica.

num povoamento, com muitas reservas.

Graves (1906) ao descrever o método de “*timber cruising*” de determinação do volume de povoamentos afirma:

“It is absolutely impossible to learn from books how to estimate timber, for it is not a matter of method, but of judgment, which can be acquired only through experience and practice in the woods. A cruiser is able to judge by the eye the merchantable contents of a tree because he has seen trees of the same character and size cut and used, and he knows what they produced. In the same way a cruiser can estimate the contents of a stand by comparing it to similar stands whose actual product he knows.”

(Graves, 1906, p.191)

Recknagel e Bentley (1919) afirmam que seja pelo julgamento ocular, seja por medições da área do povoamento e da dimensão das árvores, o resultado é sempre uma estimativa, pois o valor exato da produção de um povoamento só pode ser conhecido depois que o povoamento é colhido. Mas eles afirmam que os compradores de madeira esperavam que a estimativa apresentada não diferisse em mais de 10% da produção efetivamente obtida, e argumentam na direção da eliminação da estimativa ocular:

This means that more refined methods of work must be employed. Actual measurement of land surfaces must take the place of the guess of the old-time cruiser, or even the maps which were formerly thought to be good enough; the diameters and heights of trees must be measured with instruments, instead of being judged by the eye; the trees must be counted and tallied instead of merely guessing at the average number per acre; calculations based on sound mathematical principles must take the place of rules of thumb. In other words, every possible source of error must be eliminated, and the work conducted with all the exactness and precision that time and money will permit. At the best, it will still be an estimate; therefore measurements, where they can be made, should be made accurately, and so far as possible the element of personal choice should be gotten rid of, in order that the results may be perfectly fair and unbiased.

(Recknagel e Bentley, 1919, p.65-66)

Chapman (1924) apresenta a estimativa ocular como método, mas tece várias críticas ao método. Ao discutir os métodos de treinamento de medidores florestais ele afirma:

“In actual training of timber cruisers it has been found that ability to secure accurate estimates is greatest in men who have best developed their mental faculties by education, close observation, memory and systematic coordination. This same combination of qualities is desirable for success in any line. Many cruisers lack this ability and remain permanently inefficient to a marked degree. The only reason that such individuals have in the past continued to practice timber cruising as profession is the almost complete absence of a reliable check on their results for years at a stretch, and the comparative indifference of purchasers to the accuracy of estimates due to a rising market and a plentiful lumber supply.”

(Chapman, 1924, p.303-304)

No livro de Belyea de 1931, a simples estimação ocular direta da produção não é formalmente apresentada como um “método” de estimação da produção, mas alguns métodos citados utilizam a avaliação ocular das dimensões de árvores dentro das unidades amostrais, como a parcela circular. Mas o autor ressalta que a mensuração de algumas árvores “adiciona confiança à estimativa final”. Nestes métodos, o objetivo era fazer uma boa e rápida estimativa da “árvore média” (“*average tree*”), para a partir dela obter a produção do povoamento.

Mesmo Chapman no seu livro de 1932, relativiza a relevância da estimação ocular direta, ressaltando que a qualidade das estimativas nesta abordagem é fortemente dependente das habilidades da pessoa realizando o trabalho e, portanto, é antes uma arte que uma ciência.

Em 1965, Grosenbaugh propôs a amostragem 3P, cujo nome significa amostragem com *probabilidade proporcional* à *predição*, como método eficiente para estimar o volume de madeira de um conjunto de árvores colocado à venda (*timber sale*). A predição neste caso é a avaliação ocular do volume de todas as árvores à venda, sendo que algumas delas são amostradas para serem efetivamente medidas (Shiver e Borders, 1996). Atualmente, este método de amostragem é de uso corrente na América do Norte e a habilidade de avaliação ocular do diâmetro, altura e volume de *árvores individuais* é considerada muito útil para os técnicos de medição de campo (Schreuder, Gregoire e Wood, 1993).

A estimação ocular direta da *produção de povoamentos* utilizada no início do século foi totalmente abandonada. Ela gerava uma informação que não podia ser diretamente mensurada ou objetivamente determinada, fazendo o processo de determinação da produção se assentar sobre uma base subjetiva. Atualmente, o que se realiza é uma *predição ocular*, restrita a *atributos mensuráveis* de árvores individuais, que podem ser verificados e permitem a calibração da avaliação ocular. Assim, a predição ocular desempenha, nos métodos atuais de determinação da produção, o papel de um expediente que aumenta a velocidade do trabalho de

campo, aumentando a eficiência do processo sem comprometer a qualidade da informação gerada.

Determinação da Produção de Povoamentos Florestais

A determinação do volume de árvores em pé já representava um desafio para o florestal no início do século XX, mas vários dos métodos utilizados permanecem válidos até o presente. Já a determinação da produção de um povoamento, é um assunto de complexidade bem maior⁷.

Graves (1906) sabiamente afirma que existem duas classes de problemas distintos no que tange à determinação da produção de povoamentos. Primeiramente, há a questão de determinar a produção visando a valoração da terra para compra e venda ou visando o planejamento da colheita. Uma segunda questão, que exige maior exatidão na mensuração, é o estudo do incremento e a determinação da produção futura. Ele então discute uma série de métodos em prática na época, destacando que estes métodos diferem no grau de exatidão dos resultados que produzem.

O livro de Graves representa um marco para Mensuração Florestal na América do Norte, e sua apresentação do tema reflete com fidelidade os conhecimentos e procedimentos de mensuração da época. Vale a pena, portanto, uma discussão detalhada de como Graves apresenta os métodos de determinação da produção de povoamentos.

Primeiramente, ele cita dois métodos baseados na estimativa ocular:

Timber cruising (ou timber estimating): determinação direta da produção por peritos em estimativa ocular com grande experiência de campo, os *curisers*. Esta determinação direta não envolve a medição das árvores ou a delimitação de áreas para a observação.

Estimativa ocular (estimate by eye): também se baseia na estimativa ocular, mas requer a contagem de árvores em faixas ou parcelas de área conhecida, com posterior extrapolação das observações para o povoamento como um todo.

A **enumeração completa (ou censo) das árvores comerciais do povoamento** é apresentada como um método viável em várias situações. A determinação do volume das árvores individuais podia ser realizada por estimativa ocular ou por medição do DAP e aplicação de tabelas de volume.

⁷A discussão deste tema envolve necessariamente a questão da estimativa ocular, já discutida acima. O desenvolvimento dos métodos de determinação da produção de povoamentos tem aspectos conceituais que, por um lado se inter-relacionam com a questão da estimativa ocular, mas que por outro lado a transcendem. Para uma clara apresentação do tema, portanto, uma certa redundância no texto se fez necessária.

Em seguida, Graves enumera um série de métodos para delimitação de áreas de observação dentro dos povoamentos. Estes métodos são:

Método de Michigan: Graves argumenta que em regiões planas é mais difícil a contagem de árvores e a orientação dentro da floresta, então ele descreve um método de caminhamento com contagem dos passos no qual se acompanha a área de floresta onde as árvores estão sendo enumeradas. A área total acompanhada é de 40 acres.

Estimação com base em pequenas parcelas quadradas contíguas: as parcelas vão sendo instaladas no campo até que uma área total de 40 acres seja observada.

Método da faixa de 40 rods: consiste em observar a floresta em faixas de meia milha de comprimento por 40 rods⁸ de largura, o que equivale a aproximadamente 200 m. A área da faixa totaliza os mesmos 40 acres.

Método de Erickson: consiste em cruzar a floresta com faixas de 4 rods (aproximadamente 20 m) distanciadas entre si em um quarto de milha (aproximadamente 400 m). Nestas faixas, se toma nota das estimativas oculares do DAP e do diâmetro superior sem casca das toras comerciais de cada árvore.

A **estimação por tabelas de povoamento** (*stand tables*) é apresentada como mais um método de determinação da produção de povoamentos. Graves argumenta que tabelas de povoamentos fornecem a produção por acre para povoamentos de diferentes características. Para povoamentos equiâneos e completamente estocados (*fully stocked even-aged stands*), a tabela de povoamento é uma tabela normal de produção (*normal yield table*), mas quando a tabela de povoamento apresenta a produção média que de fato foi observada em povoamentos equiâneos numa dada localidade ou região, ela é uma tabela de produção empírica. Graves afirma que a maioria das tabelas de povoamento na América do Norte eram empírica, e que a estimativa da produção de um povoamento pode ser obtida comparando-se a estimativa ocular de campo com os valores apresentados por tabelas empíricas.

Finalmente, Graves cita a estimação por meio de “**levantamentos de valoração**” (*valuation surveys*) como uma abordagem mais acurada de determinação da produção. Os diâmetros das árvores eram realmente medidos numa área de tamanho conhecido. “Estudos detalhados” para a determinação do volume das árvores eram realizados, o que geralmente consistia na aplicação de tabelas de volume ou no abate e cubagem de árvores amostrais, mas também poderia incluir a construção de relações hipsométricas e a aplicação de tabelas de volume de dupla entrada.

⁸1 rod = 16.5 pés = 5.03 m.

A área onde a medição e os estudos eram realizados era designada por “área de valoração” (*valuation area*), “área amostral” (*sampling area*) ou “parcela amostral” (*sampling plot*). Graves afirma que tais levantamentos são utilizados não só para verificar as estimativas oculares, mas também como uma “base matemática completa para computação do volume numa dada localidade” (*complete mathematical basis for the computation of the volume of a given tract*).

Graves (1906) reconhece dois tipos principais destes levantamentos.

Levantamento em Faixas (*strip surveys*): se o levantamento era realizado em faixas estreitas, ele era chamado de “levantamento em faixas”. Graves reconhece duas formas de alocar as faixas no campo: sistematicamente (*parallel and equidistant*) ou como áreas amostrais isoladas. Este segundo caso se aplicava a regiões montanhosas ou onde a madeira comercial se concentra em algumas partes do povoamento, justificando a alocação das faixas de forma mais ou menos irregular. Era muito comum que nesta situação a determinação da produção fosse realizada por tipos florestais distintos, de modo que a elaboração de um mapa da floresta, contendo os tipos florestais, era frequentemente produzido juntamente com a estimativa da produção. Graves afirma que geralmente as faixas amostrais representavam de 5 a 10% da área do povoamento, podendo chegar a 20 ou 30% em povoamentos menores.

Levantamentos por Parcelas Exatas (*accurate plot survey*): se o levantamento era realizado em área com outra forma além da faixa, ele era designado por “levantamento por parcelas” (*plot survey*), mas, neste caso, maior cuidado era tomado na delimitação da parcela e na determinação da sua área, sendo necessária a utilização de bússola, trenas, correntes de agrimensor, espelhos em ângulo (*angle mirror*) e outros equipamentos para instalação exata das parcelas. Os levantamentos em parcela eram utilizados principalmente para estudos de crescimento e construção de tabelas de produção.

Sobre a locação das parcelas, Graves argumenta que isto depende do objetivo do levantamento e da forma, idade, composição, e grau de uniformidade da floresta. Ele argumenta que alguns princípios devem ser seguidos, de forma que é antes uma questão de julgamento do que de regras.

Um aspecto que Graves (1906) apresenta em grande detalhe no método do levantamento por parcelas é como realizar o abate de árvores amostrais para determinar o volume pela cubagem e, a partir destes dados, como determinar o volume do povoamento. Primeiramente ele apresenta o “método da árvore amostral média” (*mean sample-tree method*), que consiste em medir o diâmetro das árvores no povoamento, encontrar o “diâmetro médio”, abater árvores com diâmetro igual a

ele e determinar os seus volumes. O “diâmetro médio” no caso é o diâmetro da árvore de área transversal média (diâmetro médio quadrático), o que pressupõe que a variação de altura e forma das árvores é irrelevante. O volume do povoamento é calculado assumindo a razão do volume total do povoamento pelo volume da árvore média é igual à razão da área basal do povoamento pela área transversal da árvore média, o que pressupõe igualmente que a variação na altura e na forma das árvores é irrelevante.

Graves também discute três formas de se selecionar as árvores a serem abatidas: (i) por proximidade com o diâmetro médio, (ii) por grupos arbitrários em função da heterogeneidade do povoamento, (iii) por classes de diâmetro arbitrariamente definidas. Uma discussão extensa é apresentada sobre os métodos de Draudt, de Urich, de Hartig, de Block, do Fator de Forma da Floresta e de Metzger, que consistem em diferentes formas de se estabelecer a relação entre o volume das árvores abatidas e o volume total do povoamento.

Schenck (1905), cujo livro não é tão detalhado como o de Graves (1906), apresenta essencialmente os mesmos métodos de determinação da produção de povoamentos. Ele também faz distinção entre os métodos baseados em estimativa ocular e os métodos de mensuração (*exact mensuration*) utilizados nos levantamentos de valoração (*valuation surveys*). Ele também apresenta os métodos de seleção de árvores amostrais e de obtenção do volume da floresta a partir do volume das árvores amostrais. Sua classificação destes métodos, entretanto, é ligeiramente distinta: método de Draudt-Urich, método de Hartig e método da árvore amostral média. Um aspecto interessante na apresentação de Schenck é que ele declara explicitamente a pressuposição básica destes métodos: que a forma-altura das árvores amostrais é igual a forma-altura da floresta. Ele comenta que esta premissa é mais segura que a pressuposição de que a razão do volume da floresta está para o volume das árvores amostrais assim como o número de árvores da floresta está para o número de árvores amostrais. Este autor completa afirmando que uma pressuposição *mais insegura* (sic) é a que assume que a relação entre os volumes da floresta e das árvores amostrais é a mesma que a relação da área da floresta e da área amostrada. Nas suas próprias palavras:

“More unsafe is the assumption that the volumes of forest and sample trees bear the ratio of the acreage occupied by the forest on the one hand and by the sample trees on the other hand.”

(Schenck, 1905, p.27)

Chapman (1915) analisando os métodos de valoração da produção florestal, afirma que o método mais utilizado para determinação da produção de um povoamento (na época) era o levantamento em faixas, em geral com faixas sistematicamente locadas no campo. Mas, nesta designação, ele incluía tanto os métodos

em que as árvores eram medidas como os métodos de estimação ocular. Ele argumenta que se as faixas fossem consideradas como uma “verdadeira média da floresta”, a produção do povoamento poderia ser determinada com base na produção por unidade de área, mas afirma que este nunca é o caso, pois os povoamentos são sempre espacialmente heterogêneos. Em seguida ele apresenta uma série de considerações sobre a locação das faixas no campo e sobre a subdivisão dos povoamentos em áreas mais homogêneas para que a determinação da produção com base na área fosse um método mais acurado.

O livro de Recknagel e Bentley (1919) é voltado para o Manejo Florestal, o que torna estes autores muito críticos dos métodos baseados em estimativas oculares. Eles afirmam que a simples estimativa da produção de madeira comercial tem validade somente para a compra e venda. O gerenciamento de uma floresta requer, além desta informação, informações sobre o “estoque imaturo” (*immature stock*), ou seja, as árvores que não atingiram tamanho comercial, e uma predição da produção futura da floresta, o que só pode ser obtido por uma estimativa “moderna”, suplementada por estudos de crescimento da floresta. Assim, estes autores reconhece três classes de métodos de determinação da produção de povoamentos:

- (i) a estimativa ocular direta: geralmente realizada com base numa pequena porção do povoamento: 1 a 2%;
- (ii) “*métodos aproximados*”: as distâncias são medidas por passos e os diâmetro e altura das árvores por julgamento ocular, mas com verificação por instrumentos, e as estimativas são obtidas por inspeção de até 25% do povoamento;
- (iii) “*métodos precisos*”: as distâncias são medidas com instrumentos, as alturas com hipsômetros e os diâmetros com sutas ou vara de Biltmore; as estimativas são baseadas em 5, 10 ou até 20% do povoamento.

Os “levantamentos florestais”, definidos por Graves (1906), são enquadrados nos métodos “aproximados” e “precisos”, mas Recknagel e Bentley (1919) enfatizam a importância da medição topográfica (*land survey*) da área dos povoamentos, mesmo nos “métodos aproximados”. Inovando em relação aos livros que os antecederam, eles propõe o uso de pequenas parcelas circulares nos levantamentos, enfatizando como vantagem a sua instalação no campo. Ao enfatizarem a importância da medição da área das parcelas, este autores sugerem a possibilidade, em florestas de maior uniformidade, de se obter boas estimativas com base em área:

“If the several stands in the timber to be estimated are very uniform and regular in character, and present no special difficulty to the problem of selecting a few sample acres that shall be in all respects representative and average, then the method of accurately

measured plots may be used. As the name implies the plots are measured carefully; and the calculation of the volume of timber on the plots is done with as much care as the measurement of the area of the plot.”

(Recknagel e Bentley, 1919, p.84)

Mas o problema da determinação do volume do povoamento a partir do volume das árvores persiste. Recknagel e Bentley também são bastante críticos em relação a este ponto e afirmam:

“If the plots are established for the additional purpose of volume determination and growth studies, then one of the several methods of volume calculation, described in standard texts on Forest Mensuration may be applied. There are several of these methods, but since most of them are of purely academic interest to the American forester, only two of them will be described here. The first one, called the Mean Sample Tree method, is applicable to stands of even-age and great regularity; the second one, called the Volume Curve Method, can be used to good advantage where the conditions are not so uniform.”

(Recknagel e Bentley, 1919, p.85-86)

Os métodos de árvore amostral média (*Mean Sample Tree*) e o da curva de volume (*Volume Curve*) são descritos de modo semelhante ao de Graves (1906). Estes autores ainda discutem dois aspectos da determinação do volume de povoamentos que não aparecem de forma tão clara nos livros que os antecederam: os erros que afetam a qualidade das estimativas e os custos de se obter estimativas de produção.

A apresentação de Chapman (1924) sobre a determinação da produção de povoamentos não traz elementos essencialmente novos ao tema, mas sua abordagem denota uma clara evolução do tema. Contrariando a terminologia de Graves (1906), ele assume o termo “estimação da produção” (*timber estimating*) como termo genérico que engloba todos os métodos, o que é feito em função do reconhecimento de que todos os métodos de determinação produzem “estimativas”:

“Standing timber cannot be ‘measured’. There is always a residual error in the closest work. Hence the use of the term ‘estimates’.”

(Chapman, 1924, p.304)

Por outro lado, ele estabelece uma sequência de conceitos que implicam numa complexidade crescente:

“estimação da produção” (*timber estimating*): obter uma estimativa da produção de madeira numa área florestada;

“levantamento florestal” (*forest survey*): além da estimativa da produção, inclui informações adicionais de carácter espacial como mapa topográfico da área, mapa da cobertura florestal, com indicação de tipos de florestas, qualidade de madeira, áreas queimadas, etc., e mapa de solos; em relação à produção, o levantamento florestal implica num “inventário” da quantidade e da qualidade da madeira; e

“valoração da produção” (*timber appraisal*): consiste numa operação separada, que não deve ser entendida como parte da Mensuração Florestal, que acrescenta ao levantamento florestal informações sobre as condições e custos de colheita, os custos de intervenções silviculturais, as condições econômicas e de mercado para comercialização da madeira e custos de transporte e processamento da madeira.

Um outro elemento que difere na apresentação de Chapman (1924) é a forte ênfase no levantamento topográfico do povoamento ou área florestada como base fundamental para uma estimativa de produção acurada. Boa parte da apresentação se refere a locação e medição de parcelas e faixas, sempre enfatizada a medição precisa das distâncias e áreas.

Com relação à determinação do volume de árvores em pé, ele assume como regra o uso de tabelas de volume, mas sugere a cubagem de árvores abatidas na ausência de tabelas de volume apropriadas. Os métodos tradicionais de combinação do volume das árvores amostradas para determinação do volume do povoamento não são citados. A ênfase é dada na estimação com base na área amostrada e na área do povoamento, discutindo em detalhes a questão do número de unidades amostrais (faixas e parcelas) e sua locação no campo em relação ao tamanho da área florestada e a sua uniformidade, sempre tecendo considerações sobre a acuidade das estimativas geradas.

Belyea (1931) aborda o assunto de forma muito semelhante a Chapman (1924), embora de forma mais concisa, pois um único capítulo é dedicado ao tema. A discussão se concentra no tamanho, número, forma e locação no campo das unidades amostrais. A estimação visual direta da produção do povoamento não é discutida, mas a estimação visual de atributos de árvores individuais (volume, número de toras, diâmetro e altura) é descrita como parte dos métodos, em geral com verificação periódica por medidas obtidas por instrumentos ou tabelas de volume. Belyea distingue a determinação do volume em áreas florestadas pequenas e grandes. Para as áreas pequenas, ele sugere o censo. No caso das grandes áreas, o volume da floresta é determinado sempre com base na área amostrada em relação à área da floresta.

Esta longa exposição dos métodos de determinação da produção de florestas mostra que o desenvolvimento dos métodos ocorreu na direção da sua simplificação. No início, a combinação da produção de árvores individuais para obtenção da

produção da floresta era um tema complexo e um tanto confuso. Existiam várias abordagens teóricas que partiam de premissas sobre a uniformidade das árvores da floresta em termos de altura e forma, e a distinção da amostragem por áreas e por árvores não era totalmente clara. A heterogeneidade espacial da floresta era considerada um obstáculo à uma estimação acurada da produção da floresta com base em área.

Com o passar do tempo, a discussão de como combinar a produção de árvores amostrais individuais segundo os diferentes métodos e premissas da variação natural de diâmetro, área basal, altura e forma das árvores foi substituída pela discussão do tamanho, forma e localização de unidades amostrais. A estimação da produção da floresta com base na área amostrada, considerada inicialmente problemática, se estabeleceu com sendo a norma, talvez em boa parte devido ao acréscimo de informações sobre a floresta que a sua utilização gera. Com efeito, o estabelecimento como método padrão da estimação com base na área amostral foi precedido tanto pela aceitação da importância dos métodos de medição topográfica da floresta e das unidades amostrais, quanto pelo reconhecimento da necessidade de confecção de mapas detalhando a heterogeneidade da floresta (*forest surveying*).

Determinação do Crescimento e Produção da Floresta

A mensuração do crescimento visando a predição da produção futura da floresta já está presente nos textos de mensuração norte-americanos do começo do século XX. Alguns aspectos são de impressionante atualidade. Por exemplo, as definições dos conceitos de incrementos como incremento corrente anual, incremento médio anual, incremento corrente periódico e incremento médio periódico já são apresentadas em Schenck (1905) e em Graves (1906). Mas, por outro lado, Schenck (1905) discute vários conceitos que caíram em desuso como por exemplo, incremento total (*total increment*), incremento do passado e do futuro (*increment of the past*), incremento absoluto e incremento relativo (*absolut increment, relative increment*). Já Graves (1906), se restringe aos conceitos ainda em uso atualmente, acrescentado apenas o conceito de crescimento percentual (*growth percent*) que atualmente não é muito utilizado. O comportamento e relacionamento de incremento corrente e incremento médio é discutido muito bem por Schenck (1905), sem a necessidade do uso de conceitos avançados de matemática, como cálculo diferencial.

Schenck (1905) apresenta os padrões do crescimento de árvores individuais em diâmetro, altura e forma de maneira impressionantemente atual, mas não faz uso de conceitos fisiológicos e ecológicos. Sua apresentação se baseia na medição dos anéis de crescimento de árvores abatidas por meio da análise do tronco, tópico que é discutido em grande detalhe. Uma dificuldade desta abordagem é a determinação do volume do tronco para cada anel de crescimento para um grande número

de árvores. Assim uma série de métodos aproximados são apresentados, como por exemplo, método gráfico de Schenck (o próprio autor), método de Wagener, métodos de Pressler e método de Breyman. Também discute a proposta de Pressler para determinar o crescimento de árvores em pé pela amostra do lenho retirada por sonda (trado), o qual relaciona crescimento em volume com crescimento em diâmetro medido na amostra.

Abordagem semelhante é encontrada no livro de Graves (1906). A análise de tronco utilizando amostras do lenho também é a base para sua discussão da mensuração de crescimento de árvores individuais e de grupos de árvores. Graves apresenta detalhadamente o crescimento em diâmetro, área transversal, altura e volume, sempre discutindo a determinação do crescimento médio, máximo e mínimo para um grupo de árvores. Se o estudo do crescimento em diâmetro resulta naturalmente do estudo dos anéis de crescimento, o estudo do crescimento em altura já se faz bem mais complexo com este tipo de dado. No caso dos estudos de crescimento em volume, grande parte da discussão se concentra na questão da determinação do crescimento volumétrico em função do crescimento da área transversal, assumindo-se que o crescimento em altura e a mudança da forma da árvore podem ser negligenciados em curtos períodos de tempo.

Schenck (1905) já considerava a utilização das tabelas de volume como o método preferencial para se estudar o crescimento das árvores. Ele apresenta dois métodos para construção das tabelas: o método matemático, que consiste em agrupar as árvores segundo algumas categorias, a idade sendo uma delas, e calcular a média das classes; e o método gráfico que alcança o mesmo objetivo, mas utilizando curvas construídas em gráficos onde os dados são grafados.

No caso do crescimento de povoamentos e florestas ambos os autores, Schenck (1905) e Graves (1906), consideram que a solução é o desenvolvimento de tabelas de produção (*yield tables*) e ambos discutem sua construção e aplicação, mas Graves o faz com grande detalhamento, dedicando um capítulo inteiro ao tema. A definição de Graves (1906) de tabelas de produção pode ser considerada a base conceitual para a construção dos modelos de crescimento e produção atuais:

“A yield table is a tabular statement of the yield per acre, at different periods, of a forest of a specific character growing on a specific class of soil and treated under a specific method of management.”

(Graves, 1906, p.316)

Graves (1906) apresenta as tabelas “normais” (*normal yield tables*), conceito aprendido dos mestres florestais europeus, que são tabelas de produção que apresentam o crescimento “máximo médio” (*average maximum growth*) de povoamentos “completamente povoados” (*fully-stocked stands*). Mas ele acrescenta o conceito de tabelas “empíricas” (*empirical yield tables*), que apresentam o cresci-

mento médio da floresta como um todo incluindo as áreas boas e as áreas de baixa densidade de árvores (*poorly stocked*).

Os dois tipos de tabelas são discutidos no caso de diferentes tipos de florestas: equiâneas puras, inequiâneas e mistas. São detalhados tanto os tipos de dados quanto os métodos de construção das tabelas para diferentes tipos de floresta. Graves (1906) conclui sua apresentação de tabelas de produção afirmando a importância da remedição de parcelas amostrais permanentes e descreve os principais tipos de parcelas em uso nos E.U.A na época. Mais uma vez, os métodos atuais de coleta de dados em parcelas permanentes diferem muito pouco do que Graves apresenta.

Chapman (1912, 1915) restringe a discussão do crescimento a curtos períodos de tempo, afirmando que neste caso o crescimento do futuro pode ser estimado pelo crescimento do passado (anéis de crescimento), mas restringe a abordagem a períodos de no máximo dez anos. Para períodos maiores, ele também se apoia nas tabelas de produção.

Recknagel e Bentley (1919) também apresentam e discutem o crescimento de árvores individualmente pela análise de tronco. Eles também reconhecem que a aplicação da análise de tronco para um grande número de árvores representa uma grande quantidade de trabalho de campo e de cálculos, principalmente no que se refere ao cálculo do volume das árvores nas diferentes idades. Assim, eles sugerem seguir o método de Graves, de trabalhar com uma árvore média por grupo de árvores, sendo que o crescimento volumétrico é obtido pelas fórmulas de Pressler ou Schneider. Com relação ao crescimento de povoamentos, estes autores expressam bem o aumento de complexidade da questão:

“The study of increment in individual trees presents a number of problems, as has been shown in the preceding chapter; increment in whole stands is even more complex, for to the difficulties already encountered in the problem for single trees there must be added the everchanging character of the stand, especially as to the number and the sizes of the trees composing it.”

(Recknagel e Bentley, 1919, p.112)

Eles reconhecem um grau crescente de complexidade à medida que a estrutura da floresta se torna mais complexa na sequência de povoamentos puros equiâneos (*pure even-aged stands*), povoamentos puros inequiâneos (*pure many-aged stands*) e povoamentos mistos manejados sob corte seletivo (*mixed stands of selection form*). A solução para acompanhar o crescimento em tais povoamentos é a mesma discutida por Graves e Schenck: tabelas de produção.

Chapman (1924) representa uma grande mudança em relação aos trabalhos que o antecederam. Sua apresentação do tema se inicia com um capítulo onde os

“princípios do crescimento” de árvores e florestas são discutidos, seguidos de capítulos de apresentam as questões de (i) determinação da idade de povoamentos, (ii) crescimento de árvores em diâmetro, (iii) crescimento de árvores em altura, (iv) crescimento de árvores em volume, (v) fatores que afetam o crescimento de povoamentos, (vi) tabelas de produção normais para povoamentos equiâneos, (vii) aplicação de tabelas de produção para predição de povoamentos equiâneos, (viii) determinação do crescimento percentual, (ix) métodos para medição e predição do crescimento corrente ou periódico de povoamentos, e (x) coordenação de levantamentos florestais com a determinação do crescimento de florestas.

O aspecto mais marcante do texto de Chapman (1924) é que ele coloca claramente que a unidade de estudo do crescimento da floresta deve ser o povoamento e não as árvores individuais. Consequentemente, ele minimiza a importância da análise de tronco de árvores individuais:

“Volume-growth analysis of individual trees although apparently the most accurate and scientific basis of growth, is in reality the least important and most inefficient when expense is compared with results. ”

(Chapman, 1924, p.328)

Ele reconhece a importância da análise de tronco para o conhecimento dos padrões e das leis que regem o crescimento das árvores individualmente, mas mostra que a determinação do crescimento volumétrico médio do povoamento por este método é na prática muito menos acurado do que se assume. Além disto, ele reconhece que o conhecimento do crescimento de árvores individuais é apenas parte do problema do estudo do crescimento da floresta. Ele argumenta que o crescimento em tamanho das árvores implica na redução do número de árvores vivas ao longo da vida de um povoamento e, como a mortalidade ocorre majoritariamente nas árvores de menor diâmetro, o crescimento do diâmetro médio, e consequentemente do volume médio, da floresta não é função apenas do crescimento das árvores individuais. Ele demonstra assim que não é possível uma representação acurada do crescimento da floresta com base *apenas* no crescimento das árvores individualmente, é necessário acompanhar o processo de sobrevivência/mortalidade.

Nos capítulos em que Chapman (1924) discute o crescimento das árvores em diâmetro, altura e volume, o tema não é dominado por detalhes do processo de análise de tronco. Antes, ele discute os padrões de crescimento de árvores de diferentes tamanhos e idade, mostrando a relação do crescimento com a estrutura e padrões de crescimento dos povoamentos. Ele também afirma que a melhor abordagem para a predição de crescimento é aquela baseada nas tabelas de produção, mas seu argumento não se restringe à questão da dificuldade de se determinar o “crescimento volumétrico da árvore média”, pois reconhece que as tabelas de produção apresentam o padrão do crescimento do povoamento ao mostrar o número

de árvores vivas em diferentes idades. Chapman deixa claro que para se obter predições de alta acurácia do crescimento de florestas é necessário empreender estudos de crescimento com base em parcelas permanentes que são periodicamente remedidas:

“The best possible basis for growth predictions is the actual records of the growth in successive periods of specific forest stands whose history is known and whose conditions of management is fixed. The establishment of sample areas which are measured successively by ten-year periods will give a firm basis for growth predictions superior, either to the method of comparison, based on the growth of older stands, or to the effort to predict growth of stands from that of the trees which they contain.”

(Chapman, 1924, p.328-329)

Se os trabalhos de Schenck (1905) e Graves (1906) podem ser considerados atuais no que tange ao crescimento das árvores individualmente, o trabalho de Chapman (1924) é atual tanto na discussão do crescimento de árvores quanto nos princípios do crescimento de florestas e povoamentos. Sua apresentação do crescimento das árvores e da sua relação com o funcionamento dinâmico dos povoamentos pode ser considerada uma base conceitual para o desenvolvimento dos atuais modelos de crescimento e produção. Infelizmente, sua explanação sobre tema está marcadamente restrita às florestas equiâneas puras.

A apresentação de Belyea (1931) sobre o tema não apresenta desenvolvimento conceitual muito além do trabalho de Chapman (1924). Sua análise do crescimento das árvores individuais é menos detalhado do que os autores que o precederam e ele chega até mesmo a retomar algumas abordagens do crescimento percentual (*growth per cent*) presentes nos trabalhos de Schenck (1905) e Graves (1906), mas não apresentadas por Chapman (1924), como as fórmulas de Pressler e Schneider que buscam associar o crescimento relativo em volume com o crescimento relativo em diâmetro.

Contudo, o enfoque do trabalho de Belyea (1931) no estudo do crescimento de povoamentos e florestas é a construção e aplicação das tabelas de produção. Na sua apresentação sobre o tema, já se encontra a terminologia e os detalhes técnicos utilizados nos modelos de crescimento e produção atuais. Suas etapas para a construção de tabelas seguem a determinação do “índice de sítio” (*site index*) dos povoamentos com base no crescimento em altura das árvores dominantes, o estudo do crescimento da área basal em função da idade por classes de índice de sítio e, por fim, a análise da produção volumétrica do povoamento em função da idade com base no crescimento em área basal e no índice de sítio. Embora apresentada na forma de tabela de produção, sua abordagem na construção das tabelas é equivalente a abordagem apresentada por Clutter e colaboradores (1983)

para construção de modelos de crescimento e produção no nível de povoamento para florestas equiâneas puras.

* * *

O formato básico da Mensuração Florestal já estava dado no início do século XX, mas as transformações ocorridas até a década de 1930 depuraram muito este formato básica, resultando, do ponto de vista mensuracional e florestal, na concepção de Mensuração Florestal que ainda hoje é adotada.

Talvez a contribuição mais relevante, original e tipicamente florestal à Mensuração Florestal, após a década de 1930, foi o método da enumeração angular para amostrar povoamentos florestais, proposto por Walter Bitterlich. Embora o desenvolvimento do método tenha acontecido antes da Segunda Grande Guerra, ele só foi apresentado publicamente na tese de doutoramento de Bitterlich, que foi aceita em 1949 (Bitterlich, 1984). Nas décadas seguintes, diversos autores realizaram diversos desenvolvimentos a partir da ideia original, tornando esta área da Mensuração Florestal bastante rica em métodos e técnicas.

Excluindo o método de Bitterlich, as grandes alterações e melhorias que ocorreram no restante do século XX resultaram da introdução de elementos externos, não só à Mensuração Florestal, mas até às Ciências Florestais. Tais elementos foram os conceitos quantitativos que surgiram com o desenvolvimento da inferência estatística.

Capítulo 2

Considerações sobre a História da Inferência Estatística

Surgimento da Inferência Clássica

2.1 Introdução

Assim como as Ciências Florestais (*Forstweissenschaft*), a Inferência Estatística teve seu início com o “espírito quantificador” do século XVIII, mas foi necessário um processo mais longo para que, no século XX, ela se estabelecesse como parte fundamental dos métodos de pesquisa científica. Neste capítulo, tecemos algumas considerações sobre o desenvolvimento histórico da inferência estatística até o ponto em que a “*Inferência Clássica*”, também chamada de “*Inferência Frequentista*”, se estabeleceu na primeira metade do século XX. Esta forma de inferência é o paradigma dominante nos dias atuais, pois os cursos e livros introdutórios de Estatística voltados à formação universitária nas mais diferentes áreas de conhecimento ensinam esta forma de inferência.

Iniciamos com uma breve exposição do desenvolvimento da inferência estatística até o final do século XIX. Em seguida, apresentamos alguns aspectos relativos ao surgimento das palavras “*Biometria*” e “*Estatística*”, como nomes para a ciência nascente na passagem do século XIX para o século XX. Como terceiro passo, discutimos os principais elementos metodológicos da inferência clássica. Em cada um destes elementos, tecemos breves considerações históricas sob o seu desenvolvimento no século XX até o ponto de sua contribuição para o paradigma clássico.

2.2 Inferência Estatística até 1900¹

Bernoulli e a Lei Fraca dos Grandes Números

Talvez o marco do início da inferência estatística seja o trabalho de Jacob Bernoulli (1654-1705) publicado postumamente em 1713: “*Ars Conjectandi*”². Este livro é frequentemente tomado como o início da teoria matemática da probabilidade e da emergência do conceito de probabilidade. Bernoulli primeiro apresenta o conceito de probabilidade “*a priori*”, derivado dos jogos de azar, onde a probabilidade é a razão entre o número de casos favoráveis ao evento desejado pelo número total de casos possíveis. A premissa de que os casos são igualmente prováveis está implícita neste raciocínio.

Mas Bernoulli pergunta como a probabilidade poderia ser calculada em situações envolvendo problemas como ocorrência de doenças, as variações climáticas ou jogos de habilidade. Nestes problemas é impossível enumerar antecipadamente os casos igualmente prováveis. Então, ele propõe determinar a probabilidade “*a posteriori*”, isto é, após se observar um certo número de casos, se calcula a razão entre o número de casos favoráveis pelo número total de casos observados.

Esta forma empírica de se determinar probabilidade não era nova no tempo de Bernoulli, mas o seu tratamento formal deste conceito de probabilidade consistiu uma inovação. Bernoulli fornece então uma prova formal da ideia intuitiva de que à medida que o número de casos observados cresce, a incerteza sobre o valor da probabilidade calculada a posteriori decresce. Esta prova é conhecida hoje como a “*Lei Fraca dos Grandes Números*”, e pode ser estabelecida da seguinte forma: seja X o número de casos favoráveis de um evento que tem probabilidade p de acontecer, observados num conjunto de N observações. Para um número positivo pequeno ϵ e um número positivo grande c , é possível determinar N de tal forma que:

$$P\left(\left|\frac{X}{N} - p\right| \leq \epsilon\right) = c P\left(\left|\frac{X}{N} - p\right| > \epsilon\right)$$

Logo, é possível definir um N para qual a incerteza sobre a probabilidade a posteriori seja arbitrariamente pequena:

$$P\left(\left|\frac{X}{N} - p\right| > \epsilon\right) = \frac{1}{c + 1}.$$

¹O interesse central deste capítulo é a compreensão das formas de inferência estatística que surgiram no século XX. Mas para melhor compreensão histórica, apresentamos brevemente o desenvolvimento da Inferência Estatística até 1900, seguindo o livro de Stigler (1986). Assim, as citações sucessivas a esta obra foram omitidas.

²Arte da Inferência

Bernoulli abriu assim um caminho através do qual os eventos incertos, além dos jogos de azar, pudessem receber um tratamento probabilístico formal. Mas do ponto de vista quantitativo, o resultado que alcançou estava longe de ser prático. Para uma probabilidade verdadeira de $p = 30/50$, seria necessário observar $N = 25550$ casos para que a incerteza da probabilidade empírica (X/N) ficasse em $\epsilon = 1/50$.

De Moivre e a Aproximação da Binomial

Abraham De Moivre (1667-1754) foi um matemático francês radicado na Inglaterra que procurou dar sequência ao trabalho de Jacob Bernoulli. Ele intuiu que se o caminho que Bernoulli seguiu não tinha dado o resultado esperado, seria mais sábio seguir um outro caminho. Assim, ele partiu do cálculo de probabilidade binomial, cuja abordagem e coeficientes tinham sido desenvolvidos pelo próprio Bernoulli. O resultado que ele alcançou é conhecido hoje como “*Aproximação Normal para Distribuição Binomial*”. Embora De Moivre tenha entendido perfeitamente que seu resultado gerava uma curva, o conceito de distribuição de probabilidade ainda não estava totalmente estabelecido e, portanto, ele não concebia esta curva em termos da distribuição Normal, mas apenas como uma aproximação conveniente ao cálculo de probabilidade binomial quando o número de casos era grande.

De Moivre, como Bernoulli, tentou encontrar limites razoáveis para determinação da probabilidade de um evento. Durante a sua vida ele alcançou resultados progressivamente melhores, chegando ao que foi apresentado na edição de 1738 no seu livro “*Doctrine of Chances*”. Partindo de um experimento binomial realizado N vezes, com probabilidade de sucesso p , ele demonstrou que a probabilidade do número de eventos favoráveis X , quando N é grande, pode ser aproximada por uma série numérica que hoje conhecemos como distribuição Normal. Implícito nos resultados de De Moivre também está o fato que

$$\mathbf{E} \left\{ \left| \frac{X}{N} - p \right| \right\} \cong \sqrt{\frac{2p(1-p)}{\pi N}},$$

isto é, que o “*valor mais provável*”, hoje chamado de valor esperado ou esperança matemática ($\mathbf{E}\{\cdot\}$), para a diferença em módulo entre a probabilidade empírica (X/N) e a probabilidade verdadeira (p) é proporcional ao inverso da raiz quadrada do número de casos observados ($1/\sqrt{N}$). Infelizmente, De Moivre se ateu à sua aproximação para binomial e não explorou esta possibilidade, onde a incerteza seria quantificada por um tipo de grandeza diferente da medida da probabilidade.

Simpson e o Conceito de Erro

No século XVIII, uma questão muito discutida entre matemáticos e astrônomos era a precisão das medidas astronômicas da posição e trajetória dos astros. Thomas Simpson (1710-1761), um talentoso matemático autodidata inglês, deu duas contribuições que auxiliaram no desenvolvimento da inferência estatística. A primeira delas foi de ordem conceitual. Ele foi o primeiro a tratar não as medidas diretamente, mas o “erro” delas, deslocando o foco matemático para o “*erro médio*” e para “*distribuição dos erros*”. Sua segunda contribuição foi de ordem mais técnica, mas também muito importante. Simpson ao tratar o problema do erro das medidas assumiu uma distribuição de frequência conhecida para os erros, a partir da qual realizou o seu desenvolvimento matemático. A distribuição frequência que ele assumiu era uma distribuição triangular e, assim, ele foi o primeiro a abordar a questão através de uma distribuição *contínua* para os erros de medidas.

Uma dos resultados alcançados por Simpson, hoje tomado como trivial, foi a demonstração de que a média aritmética é a melhor combinação entre medidas repetidas como forma de gerar o “*melhor tipo possível de erro*”.

Laplace: Probabilidade Inversa e Teorema Central do Limite

Pierre Simon Laplace (1749-1827) foi um grande matemático francês, com muitas contribuições importantes para o desenvolvimento da matemática em geral, e da teoria de probabilidades em particular. Numa publicação de 1774, Laplace estabeleceu o “*Princípio da Probabilidade Inversa*” que permitiu levar adiante os desenvolvimentos sobre os erros de medidas.

Na época havia apenas dois conceitos de probabilidade: a probabilidade *a priori* e a probabilidade *a posteriori*. Na visão da ciência mecanicista, difundida pelo movimento Iluminista e compartilhada por Laplace, a aleatoriedade de um fenômeno poderia estar nos seus resultados, mas não nas suas causas, pois o funcionamento do mundo natural era visto como totalmente determinístico. Logo, era incoerente raciocinar em termos da *probabilidade de uma causa* para um evento observado. Laplace usou, então, a expressão “*probabilidade inversa*” para designar o uso de probabilidade para realizar a inferência sobre a causa de um fenômeno com resultados incertos.

Aplicando a ideia de curva de erro, ele concluiu que no caso que se tenha o valor verdadeiro de uma medida V , da qual se tem três observações a, b e c , para uma dada “curva de erro” ϕ se tem que

$$\phi(V|a, b, c) = \frac{\phi(a - V)\phi(b - V)\phi(c - V)}{\int \phi(a - U)\phi(b - U)\phi(c - U)dU}$$

A “curva de erro” é o que hoje se chama de “função de densidade probabilística”. Laplace provou que para qualquer curva de erro, a média aritmética e a

soma das observações pode ser aproximada pela mesma curva de De Moivre, à medida que o número de observações cresce. Note que a aplicação de Laplace estabelece a probabilidade, isto é, a densidade probabilística, para o valor *verdadeiro* de uma medida, que não pode ser fenômeno aleatório, por isto, a expressão probabilidade inversa.

Laplace desenvolveu o Teorema Central do Limite numa série de publicações no final do século XVIII, mas em seu livro “*Théorie Analytique des Probabilités*”, publicado em 1812, o teorema é apresentado em sua forma plena com uma dedução rigorosa. Esta obra de Laplace é tomada como uma prova da solidez da teoria da probabilidade no início do século XIX, mas um longo caminho ainda teria que ser percorrido para a inferência estatística.

Bayes e seu Teorema

O Princípio da Probabilidade Inversa é hoje conhecido por “*Teorema de Bayes*”, pois o reverendo Thomas Bayes (1701-1761) desenvolveu os mesmos resultados, deduzindo-os com grande rigor, mas não com muita clareza, sendo o seu trabalho lido postumamente perante a Royal Society em 1763.

O Teorema de Bayes estabelece que dado uma série de eventos mutuamente exclusivos A_i ($i = 1, 2, \dots, n$) e um evento em particular E , no qual o interesse é focalizado, a probabilidade condicional de um dos eventos A_k , dado que o evento E ocorreu, pode ser calculada por:

$$P(A_k|E) = \frac{P(E|A_k)P(A_k)}{\sum_{i=1}^n P(E|A_i)P(A_i)}$$

Uma forma de interpretar o Teorema de Bayes é considerar os eventos A_i como possíveis causas do evento E . O teorema mostra como calcular a probabilidade de uma das possíveis causas (A_k) ser a causa efetiva, dado que E aconteceu. Em princípio, Laplace assumiu que todas as causas eram igualmente prováveis, isto é, $P(A_i) = 1/n$.

Legendre e os Quadrados Mínimos

Outro matemático que se dedicou à questão dos erros das medidas astronômicas foi Adrien Marie Legendre (1752-1833) que publicou em 1805 seu trabalho “*Nouvelles méthodes pour la détermination des orbites des comètes*”. Ele apresenta como forma de deduzir os resultados mais acurados possíveis a partir de medidas observadas, o que leva a um sistema de equações em que o erro de cada medida é uma função linear de um certo número de variáveis. Se o número de equações no sistema é igual ao número de incógnitas, a solução era simples, mas Legendre

considera a possibilidade do número de equações (observações) ser maior que o número de incógnitas.

Legendre coloca o problema em termos de “distribuir” os erros entre as equações (observações). Ele sugere que entre os princípios possíveis, o mais geral, mais exato e mais fácil de aplicar é tornar mínima a soma dos quadrados dos erros. Legendre argumenta que o método evita que os erros extremos se tornem dominantes e apresenta várias vantagens: (i) se existe um ajuste perfeito o método o encontrará; (ii) se subsequentemente, uma equação for descartada, a solução é facilmente reencontrada; (iii) a média aritmética é um caso particular do método; (iv) o centro de gravidade de massas também é um caso particular.

A linguagem direta e prática de Legendre teve forte influência nas áreas aplicadas. Dez anos após a publicação do trabalho de Legendre, o método dos quadrados mínimos já era considerado o método padrão em Astronomia e em Geodésia na França, Itália e Prússia. Mas o trabalho de Legendre não apresentava nenhuma consideração probabilística para o tratamento dos erros de medição.

Gauss, Laplace e Quadrados Mínimos

Coube a Carl Friederich Gauss (1777-1855), matemático e astrônomo alemão, a união entre o método dos quadrados mínimos e a teoria de probabilidade. Gauss demonstrou que se as incógnitas no sistema de equações lineares fossem assumidas como igualmente prováveis e tomando uma função para “curva dos erros”, os “valores mais prováveis para incógnitas seria aqueles que maximizassem o produto da função da curva dos erros para os n erros. O problema se resumiria a solução de um sistema de equações simultâneas que equivale à solução pelo método dos quadrados mínimos.

O que Gauss demonstrou em seu trabalho de 1809, foi que o método de quadrados mínimos e o método da máxima verossimilhança (maximizar o valor mais provável) são equivalentes quando se assume como curva de erros a distribuição Normal. Em seu trabalho, entretanto, ele não conseguiu justificar bem porque ele utilizou aquela curva de erros específica (a distribuição normal). Gauss elencou as propriedades esperada para uma curva de erros: (i) seu valor seja máximo quando o erro é zero, (ii) seja uma curva simétrica ao máximo e (iii) assumo valor zero quando o erro está fora das amplitudes possíveis. Depois ele mostrou que quando uma mesma incógnita é observada sob várias condições a curva de erro que maximiza a média aritmética é a curva normal. Então, na situação mais geral, com várias incógnitas, a curva normal leva ao método dos quadrados mínimos.

Laplace leu o trabalho de Gauss e, em 1811, publicou um longo trabalho onde discutiu os sistemas lineares, mostrando que quando se assume como distribuição de erros a distribuição Normal, a estimativas de quadrados mínimos tem o menor erro esperado. Mas utilizando o Teorema Central do Limite, ele argumentou

em favor da distribuição Normal como melhor aproximação quando o número de equações no sistema é grande.

Doze anos depois, Gauss publicou um novo trabalho em que comenta a análise de Laplace e mostra que como a análise envolve apenas segundos momentos, se a medida de exatidão for o erro quadrado esperado o resultado de Laplace vale independentemente do número de equações no sistema. Este resultado é hoje conhecido como Teorema de Gauss-Markov, sendo o principal fundamento teórico para a aplicação dos modelos lineares.

Galton, Pearson e Yule

Durante o século XIX, as aplicações da teoria de probabilidades às questões demográficas, sociais e de história natural (biologia) se intensificaram e se tornaram o elemento preponderante para o desenvolvimento da inferência estatística. Dentre as inúmeras pessoas que contribuíram para este desenvolvimento destacam-se três ingleses: Francis Galton (1822-1911), Karl Pearson (1857-1936) e George Udny Yule (1871-1961), que contribuíram para o desenvolvimento dos conceitos de regressão linear e correlação.

Francis Galton foi um nobre inglês com grande curiosidade e inteligência intuitiva que procurou aplicar os conceitos probabilístico para vários problemas. Através de uma série de exemplos empíricos, ele mostrou que o Teorema Central do Limite podia ter uma aplicação mais ampla do que as condições colocadas por Laplace, que era por demais restritivas para serem aplicadas a problemas biológicos e sociais. Ele criou uma “escala estatística” para converter qualquer medida ou categoria para uma escala que seguia a distribuição Normal padronizada, isto é, de média zero e desvio padrão unitário. No seu estudo de hereditariedade da estatura de pessoas, ele estabeleceu uma relação linear (uma reta) entre a altura dos pais e a altura dos filhos. Ele notou que a inclinação da reta era tal que, quando a altura dos pais era superior à média populacional, os filhos tendiam a ser mais baixos que os pais, mas, quando a altura dos pais era inferior à média populacional, os filhos tendiam a ser mais altos que os pais. Galton denominou tal efeito de “*tendência de regredir para média*”, cunhando a expressão “*regressão linear*” utilizada até hoje.

Galton também percebeu que quando as duas variáveis utilizadas na regressão eram transformadas para distribuição normal padronizada, a regressão da média da altura dos pais pela altura dos filhos e a regressão da altura dos filhos pela média da altura dos pais produziam a mesma reta. Ele sugeriu então, que a inclinação da reta podia ser tomada como uma expressão da proximidade de “*co-relação*” entre as duas alturas. Esta expressão tornar-se o “*coeficiente de correlação*”.

Karl Pearson foi o primeiro cientista a liderar um departamento de estatística. No início final do século XIX e início do século XX, seu departamento

era o único que oferecia formação profissional em estatística. Cabe a Pearson o desenvolvimento formal do conceito de curvas assimétricas para distribuição de probabilidades. A partir de equações diferenciais ligadas à distribuição binomial, ele desenvolveu uma “família” de curvas que classificou em cinco tipos.

Tipo I: distribuição beta assimétrica,

Tipo II: distribuição beta simétrica,

Tipo III: distribuição gama,

Tipo IV: distribuição assimétrica baseada na função tangente e

Tipo V: distribuição normal.

Pearson desenvolveu um modo para estimar o coeficiente de correlação da distribuição normal bivariada, hoje conhecido como “*coeficiente de correlação de Pearson*” e soube disseminar a sua aplicação através de uma série de exemplos, principalmente com dados de hereditariedade. Pearson buscou aplicar os métodos estatísticos a muitos conjuntos de dados biológicos e sociais e talvez tenha sido o primeiro a se preocupar com a divulgação dos métodos desenvolvidos para aplicação por cientistas não matemáticos.

George Udny Yule (1871-1951) foi um engenheiro que trabalhou no início da sua carreira como assistente de Karl Pearson no seu laboratório. Ele ampliou as investigações de Pearson sobre o coeficiente de correlação numa série de elementos que hoje é chamado de teoria da regressão linear. Ele demonstrou que as melhores estimativas para os coeficientes da expressão de regressão lineares eram as estimativas de quadrados mínimos e deduziu, no contexto da regressão linear, o princípio do desdobramento da soma de quadrados que se tornaria o fundamento da análise de variância desenvolvida por Ronald Fisher.

Yule, como Pearson, também foi importante para divulgação e propagação dos métodos estatísticos para diversas ciências. O seu livro “*An introduction to the Theory of Statistics*”, publicado em primeira edição em 1919 e em segunda em 1927, é apresentada como a fonte básica de teoria estatística pelos vários pesquisadores e professores de Mensuração Florestal como Belyea (1931), Schumacher e Hall (1933) e Bruce e Schumacher (1935, 1942).

2.3 *Biometria e Estatística*

O surgimento da Biometria (“*biometry*” ou “*biometrics*”) é de difícil definição. Stigler (2000) remonta à idade antiga, em alusões ao velho testamento da bíblia e a

histórias do mundo greco-romano, para tentar demonstrar que sempre existiu uma continuidade nos problemas que a Biometria trata atualmente, embora o enfoque e a ênfase nas soluções destes problemas tenha diferido muito ao longo da história. Já Kendall (1960) coloca o início da Biometria como ciência na segunda metade do século XVII. Segundo ele, surgiram neste período uma série de publicações sobre demografia e estatísticas de organização do Estado que demonstram que os autores tratavam os dados quantitativos de forma semelhante a que se utiliza atualmente na Biometria.

Embora os argumentos de Stigler (2000) tenham seus méritos, é difícil aceitar o surgimento de uma disciplina científica como um evento histórico anterior ao estabelecimento do que hoje entendemos por ciência. Os argumentos de Kendall (1960) também têm sua validade, mas o período do século XVII e XVIII é justamente o período do surgimento da concepção de ciência como a temos hoje, e o “espírito quantificador” é um elemento essencial no surgimento desta concepção (Heilbron, 1990), de modo que o tratamento de dados quantitativos não pode ser tomado como o marco de uma ciência em particular.

Na história recente da pesquisa científica, um elemento marcante para a definição de uma disciplina científica é a publicação de uma obra que a defina como uma área do conhecimento científico. No caso da Biometria, o marco foi a publicação do periódico científico “*Biometrika*” em 1901 (Stigler, 2000).

Segundo Stigler (2000) a publicação da “*Biometrika*” seria resultado principalmente do trabalho de Francis Galton e Karl Pearson. Mas segundo Egon Pearson (Pearson, 1965), a publicação foi fruto da intensa colaboração científica entre Karl Pearson e Walter Frank Raphael Weldon (1860-1906) iniciada em 1894, na qual Galton atuava mais como um mediador. Já Cox (2001) afirma categoricamente, com base em carta enviada por Weldon para Karl Pearson, que a iniciativa partiu de Weldon e que, embora Galton tenha sido um dos editores no início, o periódico foi iniciado e mantido majoritariamente pelos esforços de Weldon e Karl Pearson. Na época, Karl Pearson e Weldon trabalhavam no “University College London”, Karl Pearson na cadeira de Matemática Aplicada e Weldon na de Zoologia. Galton já era um ancião de 64 anos com a carreira bem estabelecida e atuou como um mediador no debate entre a abordagem matemática teórica de Karl Pearson e a abordagem biológica eminentemente aplicada de Weldon (Pearson, 1965).

Segundo Cox (2001), o nome “*Biometrika*” foi sugerido por Karl Pearson como nome de uma publicação voltada aos estudos em “*Biometry*”, ou seja, um veículo científico voltado para a divulgação de estudos quantitativos da Natureza. O nome seria a sua resposta à indagação de Weldon em sua carta em 16 de novembro de 1900 após uma reunião na “Royal Society”:

“The contention ‘that numbers mean nothing and do not exist in Nature’ is a very serious thing, which will have to be fought. Most

other people have got beyond it but most biologist have not.

Do you think it would be too hopelessly expensive to start a journal of some kind?"

(Cox, 2001)

O elemento importante nesta história, é que a publicação da “*Biometrika*”, e consequentemente o termo Biometria, surgiu como uma resposta à resistência do uso de métodos quantitativos nas Ciências da Vida. A intensa colaboração entre Karl Pearson and Weldon estava centrada na representação da variabilidade das observações biológicas na forma de distribuições de probabilidades. A abordagem inicial, marcada pela aplicação da distribuição Normal, evoluiu para a utilização de distribuições assimétricas que resultaram na família de distribuições desenvolvida por Pearson. No decorrer dos debates, a questão da aderência de observações empíricas às distribuições teóricas esteve sempre presente e contribuiu para o desenvolvimento do conceito da estatística Qui-quadrado (Pearson, 1965). O desenvolvimento teórico da Biometria foi estimulado por questões de cunho aplicado, e as questões aplicadas e teóricas dos métodos quantitativos em desenvolvimento andavam juntas.

Ao grupo que trabalhava com Karl Pearson veio se juntar outra pessoa que posteriormente teve grande destaque: William Sealy Gosset (1876-1937), que ficou conhecido pelo pseudônimo de “Student” com o qual publicou seus trabalhos estatísticos. Químico e matemático por formação, “Student” foi contratado em 1899 pela Messrs Guinness Son & Co. como “*brewer*” (cervejeiro). A Messrs Guinness havia iniciado uma política de contratação de jovens com formação científica em Oxford ou Cambridge (Pearson, 1939) e o método científico e as determinações laboratoriais estavam começando a ser seriamente tratados no processo de produção de cerveja (McMullen, 1939). As condições deste processo de produção, como a variabilidade dos materiais, a susceptibilidade a mudanças de temperatura e a necessidade de uma série de experimentos curtos, se mostravam inadequadas à aplicação das teorias estatísticas da época que eram baseadas em grandes amostras. Apesar das diferenças, “Student” passou um ano estudando com Karl Pearson no University College London e criou uma colaboração científica que perdurou por toda a sua vida. Desta colaboração surgiram vários desenvolvimentos voltados a pequenas amostras, dos quais o mais famoso é a distribuição do desvio padrão amostral, que resultou no que hoje é conhecido como “teste *t*” de Student (McMullen, 1939).

O interesse de “Student” era eminentemente prático. Embora conhecido como “estatístico”, a estatística não era a sua atividade profissional principal, nem tomava a maior parte do seu tempo de trabalho (McMullen, 1939; Pearson, 1939). Suas investigações estatísticas foram realizadas somente quando ele e seus colegas enfrentavam problemas que requeriam uma solução de carácter estatístico

(Pearson, 1939).

Quando Karl Pearson se aposentou do University College London como “*Galton Professor of Eugenics*” em 1933, seu departamento de Genética, na época chamado de “*Eugenics*”, foi dividido em dois departamentos: o departamento de Genética e o departamento de Estatística (“*Statistics*”). Para liderar o departamento de Genética foi convidado Ronald Aylmer Fisher (1890-1962), enquanto que para o departamento de Estatística foi convidado Egon Sharpe Pearson (1895-1980), filho de Karl Pearson (Stigler, 2007).

Sociedade e Instituto

Um aspecto curioso é que o termo “Estatística” precede o termo “Biometria” na sua utilização para designar a nova disciplina científica que nascia no final do século XIX. Por exemplo, em 1834 foi fundada a “*Statistical Society of London*”, que em 1887 se tornaria a “*Royal Statistical Society*”, cujo objetivo era publicar “fatos calculados sobre as condições da sociedade” (Hill, 1984). No nome “*Statistical Society*”, portanto, “Estatística” designava originalmente a coleção, organização e publicação de informações quantitativas (estatísticas) sobre a sociedade. De fato, Hill (1984) mostra que houve muita discussão dentro da “*Statistical Society*” até que a Estatística fosse tomada como uma ciência no conceito que temos atualmente dela, isto é, como uma ciências de análise de dados.

Outro exemplo é o “*International Statistical Institute*” (ISI), que foi fundado em 1885 e até hoje promove congressos internacionais atraindo pessoas de instituições governamentais e da academia. No final do século XIX, a palavra “estatística” presente no nome do ISI se referia antes aos dados coletados para se obter informações sobre um país, que ao conjunto de teorias e métodos voltados para a análise de dados.

Ronald Fisher e a sua Definição de Biometria

Ronaldo Fisher é hoje considerado o pai da Estatística moderna, sendo difícil encontrar um tema ou método estatístico que ele não tenha trabalhado ou sugerido alguma linha de pesquisa (Rao, 1992).

Dentre as suas contribuições destaca-se, a introdução do conceito de “graus de liberdade”, o conceito de verossimilhança, o método da máxima verossimilhança para estimativa de parâmetros, a organização dos princípios do delineamento experimental e a análise de variância de experimentos, inclusive com o teste F . Uma apresentação detalhada de suas contribuições pode ser encontrada em Rao (1992).

Curiosamente, as posições que Fisher ocupou em universidades não eram da área de estatística, embora ele se considerasse um estatístico bem conhecido (*well-known statistician*) e um geneticista relativamente bem conhecido (*fairly well-*

known geneticist) (Edwards, 1990). Sua atuação como pesquisador começou em 1919 quando assumiu uma posição em Rothamsted onde iniciou o desenvolvimento dos métodos estatísticos de análise de experimentos. Em 1933 assumiu a cadeira de Genética no University College of London e em 1943 assumiu a cadeira de Genética em Cambridge (Kendall, 1963).

Apesar a sua contribuição marcante estabelecendo os fundamentos da teoria estatística, tudo o que fazia era fortemente motivado por aplicações práticas, e seus trabalhos publicados diferem muito do que hoje se vê nos periódicos de estatística (Rao, 1992). Tendo sido convidado a ser “Presidente de Fundação” da International Biometric Society (IBS), fundada em 1947, publicou um curto artigo onde apresenta a sua visão da Biometria como ciência (Fisher, 1948), onde apresentou uma definição que até hoje a IBS utiliza como definição “oficial” (vide página oficial da IBS em <http://www.tibs.org/>):

“Biometry, active pursuit of biological knowledge by quantitative methods”.
(Fisher, 1948)

Posteriormente à publicação da “*Biometrika*”, o uso dos termos “Estatística” e “Biometria” parece ter seguido *pari passu*, sendo que gradativamente eles passaram a ser utilizados como sinônimos. Embora o termo “*Biometry*” tenha sido utilizado por alguns autores antes da publicação da “*Biometrika*”, o seu estabelecimento como nome de disciplina científica e seu uso generalizado entre os cientistas deve ser creditado ao sucesso do periódico científico (Stigler, 2000).

Biometria Florestal

A expressão “*Biometria Florestal*” é utilizada neste trabalho no sentido definido por Fisher, isto é, como a atividade que busca o desenvolvimento do conhecimento nas Ciências Florestais por meio de métodos quantitativos. Já o termo “Estatística”, é utilizado para designar os conceitos e métodos de análise de dados, independentemente da área de aplicação.

Esta distinção é até certo ponto artificial. Primeiramente, porque o desenvolvimento dos princípios e suas contribuições teóricas e metodológicas iniciais estão todos ligados a busca de soluções de problemas práticos e não ao desenvolvimento puramente teórico. Em segundo lugar, o aspecto puramente teórico da Estatística é essencialmente matemático, sendo muito difícil distinguí-lo da Matemática como ciência. Já os aspectos da aplicação prática da Estatística estão intimamente ligados à área do conhecimento onde a teoria estatística é aplicada, sendo, conseqüentemente, difícil separá-los totalmente da área de conhecimento particular onde a aplicação ocorre.

2.4 *Inferência Estatística Clássica*

No início do século XX, vários métodos de inferência estatística tinham sido propostos e sua aplicação em problemas práticos começava a ser amplamente utilizada em várias áreas do conhecimento. A visão da inferência estatística como a forma mais apropriada de inferência quantitativa para lidar com as incertezas dos resultados científicos se estabelecia na comunidade científica e acadêmica. O estabelecimento da revista científica “*Biometrika*” contribuiu para que o conjunto de métodos de inferência se afirmasse como uma área de conhecimento científico.

Foi neste período que a Estatística adquiriu solidez teórica e se firmou como um paradigma de inferência quantitativa. Esta forma de inferência ficou conhecido como a “*Inferência Clássica*” ou “*Inferência Frequentista*”. Segue-se uma breve exposição dos aspectos do seu desenvolvimento que são mais relevantes para a Biometria Florestal aplicada à Mensuração Florestal.

Quadrados Mínimos, Regressão e Modelos Lineares

Durante o século XIX, o método de quadrados mínimos teve grande desenvolvimento na pesquisa aplicada, principalmente em Astronomia, mas foi pouco estudado por pesquisadores e matemáticos aplicados de outras áreas do conhecimento (Seal, 1967). Com os trabalhos de Galton, os pesquisadores de métodos estatísticos passaram a ter grande interesse nos métodos de regressão linear, mas até o início do século XX, quadrados mínimos e regressão linear eram considerados temas totalmente independentes (Seal, 1967). Poucos matemáticos e estatísticos estudavam a “teoria dos erros” que vinha sendo desenvolvida desde Laplace e Gauss na Astronomia (Aldrich, 2005).

Yule utilizava o método de quadrados mínimos para ajustar os modelos sem reservas, mas ignorava totalmente a teoria de distribuição desenvolvida por Laplace e Gauss (Aldrich, 2005). Coube a Fisher a união entre a abordagem de regressão linear com a abordagem de quadrados mínimos. Ele introduziu a ideia de se testar a nulidade dos coeficientes de regressão utilizando a teoria para pequenas amostras de “Student” através da distribuição t de Student. Também foi Fisher que introduziu a ideia de que as variáveis preditivas poderiam ser variáveis qualitativas, o que culminariam na sua análise de variância como método para análise de dados experimentais, através do que veio a ser conhecido como o teste F . Fisher estendeu tais método a uma classe muito mais ampla de problemas e aplicações (Seal, 1967).

Os trabalhos de Fisher, principalmente na década de 1920, talvez tenham despertado o interesse na união das duas abordagens sendo seguidos por uma série de desenvolvimentos. Um aspecto marcante, é que somente em 1935 o método de quadrados mínimos e de regressão foi apresentado na notação matricial que é uti-

lizada até hoje (Seal, 1967). Mas talvez seja o livro de Draper e Smith, “*Applied Regression Analysis*”, cuja primeira versão foi publicada 1966, que se tornou a referência básica para a teoria e aplicação da regressão nas diversas áreas de conhecimento no final do século XX (veja Drape e Smtih, 1981).

Um aspecto a ser notado, é que enquanto o método de quadrados mínimos, como proposto na época de Gauss e utilizado na Astronomia, Geodésia e outras áreas aplicadas, tinha um objetivo prático de predição de medidas, a abordagem de regressão, desenvolvida na linha de continuidade do pensamento de Galton, Pearson, Yule e Fisher, tinha como objetivo predominantemente uma técnica explicativa auxiliar no teste de hipóteses científicas. Mas a denominação deste método por “regressão linear” tornou-se dominante, tanto nas publicações estatísticas teórica quanto nas publicações práticas de aplicação do método nas diferentes áreas do conhecimento. Assim, o desenvolvimento teórico das “técnicas de regressão linear” foi, e ainda é, fortemente guiado por questões de importância explicativa, como por exemplo, teste estatístico dos coeficientes de regressão, seleção e construção do “melhor” modelo e verificação rigorosa das pressuposições.

Testes de Significância

Desde o século XIX, a inferência estatística consistia em associar o “*erro provável*” (“probable error”) à estimativa de uma estatística. Embora De Moivre tenha se referido a um conceito que é baseado no erro provável, foi provavelmente Friederich Wilheml Besse (1784-1846), matemático e astrônomo alemão, quem o utilizou pela primeira vez em 1818. Posteriormente, Gauss adotou-o e desenvolveu várias métodos para calculá-lo. Para os pesquisadores que utilizavam a “teoria dos erros” baseada na distribuição Normal, era comum comparar uma dada estimativa a três vezes o seu erro provável para determinar a sua “*significância*”. Com a aplicação da distribuição Normal a variáveis biológicas, Galton levantou sérias objeções a expressão por causa da palavra “erro”, mas foi Karl Pearson que estabeleceu o conceito de “*desvio padrão*” (Cowles e Davis, 1982).

Os estudos de Karl Pearson no final do século XIX sobre o ajuste de dados à distribuições resultaram no chamado “*teste de aderência*” (“*goodness-of-fit test*”), onde a estatística utilizada, chamada de Qui-quadrado tinha distribuição conhecida. No exame de vários exemplos, Pearson calculou a probabilidade de ocorrência dos valores de Qui-quadrado calculados, para dados com distribuição Normal, acrescentando comentários sobre a qualidade do ajuste dos dados à distribuição. A partir daí surge a referência de um teste estatístico não em termos da variabilidade da estimativa, como o erro provável ou o desvio padrão, mas como medida de probabilidade da discrepância observada (Cowles e Davis, 1982).

No início do século XX, “Student” introduziu a ideia de se realizar testes estatísticos com base em pequenas amostras, tendo deduzido a distribuição t de Stu-

dent para o caso de médias da distribuição Normal, quando o desvio padrão não é conhecido. Mas foi Fisher que expandiu a ideia do cálculo de estatísticas baseadas em pequenas amostras, deduzindo as suas distribuições exatas e aplicando a abordagem à diferentes problemas práticos, de modo a generalizar a ideia de teste estatístico. Também foi Fisher quem cunhou a expressão “*nível de significância*” para se referir a um valor fixo de probabilidade na distribuição de uma estatística, a partir do qual a discrepância da estatística calculada devesse ser considerada “*significativa*” (Lehmann, 2011). Em 1925, Fisher publicou o livro “*Statistical Methods for Research Workers*”, que teve grande influência nos meios estatísticos e científicos, em que divulga a sua abordagem de inferência estatística baseada nos testes de significância (Lehmann, 2011). Embora o nível arbitrário de 5% parece ter sido utilizado por vários estatísticos antes desta época, foi provavelmente este livro que propagou a utilização deste limite (Cowles e Davis, 1982). Assim, a ideia de “teste de significância” ficou fortemente associada à pessoa de Fisher.

Teoria de Neyman-Pearson

Egon Pearson, filho de Karl Pearson e seu sucessor na área de estatística no “University College of London”, e Jerzy Neyman (1894-1981), matemático polonês, se tornaram famosos pelo trabalho em colaboração que desenvolveram juntos entre os anos de 1928 e 1936. Desta parceria resultou o desenvolvimento de uma nova teoria para a realização dos testes de significância, ou dos testes estatísticos de hipótese, com base na teoria da decisão. Alguns autores denominam esta teoria por “*paradigma de Neyman-Pearson*”.

Nesta teoria, uma determinada hipótese científica deve ser formulada matematicamente em termos de duas hipóteses: a hipótese nula e a hipótese alternativa. Em princípio, as hipóteses devem ser complementares, de modo que a rejeição de uma implica na aceitação da outra e vice-versa. Na sua forma mais simples, a hipótese nula estabelece um valor fixo conhecido para um parâmetro populacional, mas ela também pode estabelecer a igualdade entre dois ou mais parâmetros populacionais. O aspecto essencial da hipótese nula é que, tomando-a como premissa, torna-se possível deduzir o comportamento probabilístico de uma dada grandeza calculada a partir dos dados de uma amostra. Esta grandeza é denominada genericamente de “*estatística*”. Em decorrência da simplicidade da hipótese nula, a hipótese alternativa é sempre complexa, de modo que é impossível se saber o comportamento probabilístico da “*estatística*” quando se toma a hipótese alternativa como premissa.

Neste ponto, a teoria de Neyman-Pearson traz a teoria da decisão para o julgamento das hipóteses. O teste das hipóteses é definido, então, como uma decisão de se rejeitar ou não a hipótese nula, pois somente sob a hipótese nula é que o comportamento da “*estatística*” é conhecido. Surgem então duas possibilidades

de erro na decisão. Existe a possibilidade de se rejeitar uma hipótese nula que é verdadeira, que é denominada por “*erro tipo I*”, e existe a possibilidade de não se rejeitar uma hipótese nula falsa, que é o “*erro tipo II*”. Para uma melhor compreensão, a situação pode ser ilustrada pelo seguinte quadro, que confronta a decisão com o dado da realidade:

DECISÃO EM RELAÇÃO À HIPÓTESE NULA	HIPÓTESE VERDADEIRA	
	Hipótese Nula	Hipótese Alternativa
Rejeitar	Erro Tipo I	Decisão correta
Não Rejeitar	Decisão correta	Erro Tipo II

Como o dado da realidade é sempre desconhecido, é impossível se evitar os erros. Por outro lado, o comportamento da “estatística” só pode ser conhecido sob as condições da hipótese nula, logo, só é possível se controlar a possibilidade do erro tipo I. Por exemplo, considere que se deseja testar a hipótese nula de que uma dada média populacional tem valor nulo (zero), contra a hipótese alternativa de que o seu valor é diferente de zero. Uma “estatística” apropriada para testar a hipótese nula é a “estatística” t de Student, que é calculada pela expressão

$$t = \frac{\bar{x} - 0}{\sqrt{s^2/n}},$$

onde \bar{x} é a média amostral, s^2 é a variância amostral e n é o tamanho da amostra. Se a hipótese nula for verdadeira, a “estatística” se comportará conforme o gráfico da figura 2.1.

Na figura, estão marcadas duas áreas para as quais o valor da “estatística” t é maior que dois em módulo ($t \geq |2,00|$). Estas duas áreas formam uma *região de rejeição* que equivale à probabilidade de 10%. Pode-se, então, definir a seguinte *regra de decisão*: rejeitar a hipótese nula sempre que o valor observado da “estatística” t (com 5 graus de liberdade) for maior ou igual ao *valor crítico* de $t = |2,00|$. Ou seja, rejeita-se a hipótese nula sempre que o valor observado da “estatística” estiver na região de rejeição. A aplicação desta regra de decisão no longo prazo implica na situação que, de todas as vezes que a hipótese nula for verdadeira, 10% das vezes a hipótese nula será equivocadamente rejeitada. Ou seja, se terá, no longo prazo, uma margem de erro tipo I de 10%.

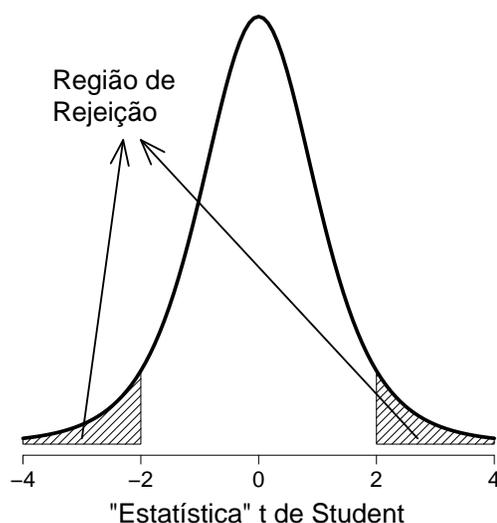


Figura 2.1: Distribuição probabilística da “estatística” t de Student com 5 graus de liberdade, mostrando a região de rejeição equivalente à probabilidade de 10%, a qual é definida pelo valor crítico de $t = |2,00|$.

Na teoria de Neyman-Pearson, se estabelece convencionalmente uma margem aceitável do erro tipo I, que é chamada de “*nível de probabilidade*”. O nível de probabilidade convencionalmente aceito na maioria das áreas de conhecimento científico, inclusive nas Ciências Florestais, e que foi insistentemente utilizado por Fisher nos seus exemplos é de 5%. A partir do nível de probabilidade, se encontra, em cada situação particular, o valor crítico da “estatística” sendo utilizada para testar a hipótese nula, na figura 2.1 o valor é $t = |2,00|$. Se o valor da “estatística” efetivamente observado na situação particular for maior que o valor crítico, se rejeita a hipótese nula. Caso contrário, a hipótese nula não é rejeitada. Esta forma de testar a hipótese nula, baseada na regra de decisão, garante que, a longo prazo, o número de vezes que se encontre num erro tipo I é no máximo igual ao nível de probabilidade convencionalmente estabelecido, ou seja, 5%.

É importante perceber que a regra de decisão também implica numa margem de erro tipo II, mas é impossível se conhecer esta margem, pois seria necessário se conhecer a hipótese alternativa verdadeira. No caso do exemplo acima, a hipótese alternativa estabelece que a média populacional é *diferente* de zero, logo, seria necessário conhecer o valor da média populacional.

A partir da concepção básica apresentada, a teoria sofreu um grande desenvolvimento e surgiram vários conceitos utilizados atualmente no teste estatístico de

hipótese, como poder do teste, teste da razão de verossimilhança e teste uniformemente mais poderoso. Historicamente, o aspecto importante é que Neyman e Pearson desenvolveram a sua teoria de teste estatístico em bases bastante diferentes da concepção de Fisher. Mas, a organização teórica de Neyman e Pearson produziu uma justificativa mais rigorosa para a teoria de testes baseados na distribuição Normal, que contrastava com desenvolvimento em bases intuitivas utilizado por Fisher (Lehmann, 2011).

Em 1938, Neyman assumiu a posição de professor na Universidade da Califórnia, em Berkeley, tornando-se uma figura muito influente na formação da nova geração de estatísticos norte-americanos, após a Segunda Grande Guerra (Lehmann, 2011).

Fisher e o Valor-p

Ronaldo Fisher parece nunca ter aceitado totalmente a proposta de Neyman e Pearson. Embora sua posição tenha variado bastante ao longo do tempo, suas divergências eram tanto de carácter teórico quanto prático. Frequentemente ele sugeria testes muito diferentes dos propostos pela teoria de Neyman-Pearson para a mesma situação e os mesmos dados (Lehmann, 2011). Assim, no período de 1915 a 1933 houve um intenso debate envolvendo Fisher, Neyman e Pearson sobre estas duas visões concorrentes de teste estatístico de hipótese e a ferocidade do debate chegou a gerar uma forte divisão entre os profissionais da nascente ciência da Estatística (Lehmann, 1993). Passado este período, o debate arrefeceu, mas sem que os contendores chegassem a um acordo.

É curioso que ambas visões partem dos mesmos fundamentos da Inferência Clássica, mas afirmaram existir grandes diferenças filosóficas. Ainda mais curioso, é que, apesar das diferenças filosóficas, atualmente, os estatísticos clássicos acreditam que as duas teorias são complementares na prática, sendo possível combinar as melhores características de ambas (Lehmann, 1993).

Dentre as diferenças entre a teoria de Neyman-Pearson e a abordagem de Fisher, o conceito de “valor-p” (*p value*) talvez seja o aspecto que mais influencia a aplicação prática dos testes estatísticos de hipótese. Ao invés de partir do nível de probabilidade, isto é, de uma margem convencionalmente estabelecida para o erro tipo I, a abordagem do valor-p parte do valor da “estatística” observado em cada situação particular.

Retomando o exemplo do teste para verificar se a média populacional é nula, considere que o valor da “estatística” t observado foi de 2,75. O valor-p consiste na probabilidade de se observar este valor efetivamente observado ($t = 2,75$) ou um valor mais extremo. O valor mais extremo é definido pela hipótese alternativa. Como a hipótese alternativa é de que a média é diferente de zero, os valores mais extremos são $t \geq |2,75|$. A figura 2.2 ilustra a região na distribuição da “esta-

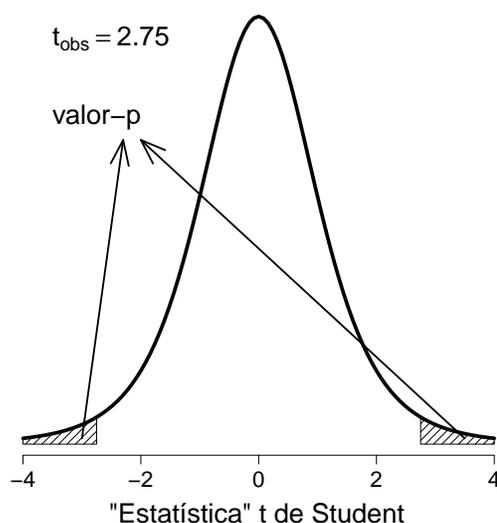


Figura 2.2: Distribuição probabilística da “estatística” t de Student com 5 graus de liberdade, mostrando a região correspondente ao valor-p para o valor observado de $t_{\text{obs}} = 2,75$ (valor-p = 4%).

tística” t correspondente ao valor-p associado ao $t_{\text{obs}} = 2,75$. A hipótese nula é rejeitada quando o valor-p observado for menor que o nível de probabilidade estabelecido, ou seja, 5% na maioria das ciências. Quando isto acontece, diz-se que o “valor-p é *significativo*” ou que o “teste estatístico se mostrou *significativo*”.

Para um mesmo nível de probabilidade, a utilização do valor-p ou do valor crítico da teoria de Neyman-Pearson geram exatamente a mesma conclusão prática, uma vez que as duas abordagens são logicamente equivalentes. Contudo, o valor-p se tornou muito mais popular nas aplicações práticas dos testes de hipóteses devido uma afirmação de Fisher na 13ª edição do seu livro “*Statistical Methods for Research Workers*” de 1958 (Lehmann, 2011):

“The actual value of P obtainable from the table by interpolation indicates the strength of the evidence against the hypothesis.”

Ou seja, Fisher afirma que o valor-p pode ser tomado como medida da força da evidência presente nos dados contra a hipótese nula. Esta forma de interpretar é muito semelhante à maneira como os pesquisadores testam suas hipóteses científicas e se tornou a forma preferencial de se interpretar os testes estatísticos de hipóteses. A adesão a esta interpretação foi tão dramática e ficou tão enraizada na comunidade científica que, no dias atuais, encontrar um valor-p significativo

se tornou um “imperativo estatístico” para a aceitação de uma hipótese científica. Na maioria das áreas de conhecimento científico, os editores de periódicos científicos tomam tal interpretação como um dogma da inferência estatística. Não seria exagero dizer que, atualmente, a antiga frase referente à carreira de pesquisador que dizia “*publish or perish*” pode ser adaptada para “*find significant p values or perish*”.

Teoria da Amostragem

O desenvolvimento da teoria da amostragem se deu de uma forma aparentemente mais lenta que outros métodos estatísticos. Durante o final do século XIX e início século XX, houve muita discussão nos encontros do ISI (“International Statistical Institute”), sobre as vantagens da utilização da amostragem para substituir os censos de coleta de dados sociais e econômicos. A maioria dos membros do ISI eram defensores do censo, mas a intensa defesa da “*amostragem representativa*” foi gradativamente gerando maior espaço para esta abordagem. Os defensores da amostragem, entretanto, não tinham a mesma ideia do que era uma amostra “representativa”. A maioria tinha um conceito de representatividade que levava a um tipo de amostragem hoje denominada de “*amostragem intencional*” ou “*amostragem seletiva*”, pois eles acreditavam que a representatividade da amostra era gerada pela seleção cuidadosa por parte de uma pessoa conhecedora da população. A amostragem intencional gerava com frequência resultados extremamente enviesados (Kruskal e Mosteller, 1980).

Em 1934, Neyman apresentou um trabalho diante da Royal Statistical Society que hoje é tomado como o marco do início da teoria da amostragem. O trabalho era intitulado “*On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection*”³, e demonstrava os problemas e falhas da seleção intencional de amostras. Este trabalho marca a associação da “representatividade” com a necessidade de aleatorização na seleção da amostra e a importância da estratificação (Kruskal e Mosteller, 1980). Apesar do grande impacto deste trabalho, ele não é totalmente original, pois vários pesquisadores estatísticos já vinham desenvolvendo trabalhos semelhantes. Fisher, por exemplo, já advogava a necessidade da aleatorização na condução de experimentos, enquanto que vários dos resultados ligados a amostragem estratificada já tinham sido antecipados por A.A. Tchouprof num trabalho publicado na revista *Metron* em 1923 (Fienberg e Tanur, 1996).

As contribuições de Neyman, no entanto, foram fundamentais para o desenvolvimento da teoria da amostragem e o trabalho de 1934 é considerado como

³Sobre os dois aspectos diferentes do método representativo: o método da amostragem estratificada e o método da seleção intencional.

o ponto de início da amostragem probabilística, pois atribuiu um claro conceito matemático à ideia de representatividade.

Intervalo de Confiança

A ideia de intervalo de confiança também está fortemente associado à Neyman, que no seu trabalho de 1934, publicou um apêndice onde apresenta o conceito e, em 1937, publicou um trabalho inteiramente ao desenvolvimento rigoroso do conceito. Neyman afirmou que desenvolveu o conceito a partir da ideia de “probabilidade fiducial” que Fisher publicara em 1930 (Edwards, 1976; Zabell, 1992; Lehmann, 2011). Fisher chegou a ideia de “probabilidade fiducial” a partir da ideia de “probabilidade inversa” como método de inferência estatística que foi amplamente utilizada por Laplace, mas com a qual ele não concordava (Zabell, 1992; Lehmann, 2011). Entretanto, Fisher nunca desenvolveu plenamente a sua abordagem (Zabell, 1992), enquanto que a proposta de Neyman passou a ser considerada como a solução completa para o problema (Edwards, 1976; Lehmann, 2011).

Na verdade, Neyman não foi o primeiro a utilizar a ideia de intervalo de confiança. Em 1929, Working e Hotelling já haviam publicado um trabalho onde deduziram “bandas de confiança” para regressão linear, e Neyman cita que antes de Fisher e Hotelling, Wilson e Milot já tinham trabalhado com conceito semelhante para a probabilidade de sucesso numa distribuição binomial. A ideia de intervalo de confiança já “estava no ar” no meio estatístico e o mérito de Neyman foi elevar a discussão das aplicações particulares para uma teoria geral (Lehmann, 2011).

A partir da apresentação de Neyman em 1934 e 1937, o conceito de intervalo de confiança passou a ser amplamente utilizado em aplicação práticas e no desenvolvimento da teoria estatística. O resultado do desenvolvimento teórico é o que atualmente se chama “estimação de intervalo”, isto é, obtenção de um intervalo para uma estimativa que seja uma medida da sua precisão, em contraposição à “estimação pontual” que se refere simplesmente à obtenção da estimativa (Lehmann, 2011).

Estatística Matemática

“*Estatística Matemática*” (*Mathematical Statistics*) é a expressão, em geral utilizada por estatísticos profissionais, para designar o que também poderia ser chamado de “*Teoria Estatística*”, isto é, um conjunto de conhecimentos estatísticos gerais e, portanto, aplicáveis a qualquer problema prático de análise de dados. A palavra “matemática” na expressão tem carácter adjetivo, como a expressão em inglês deixa bem claro, indicando que é uma estatística desenvolvida puramente

no âmbito matemático. Logo, a teoria estatística é, na sua essência, matemática. De fato, a Estatística Matemática se sobrepõe parcialmente à “Teoria de Probabilidades”, que é certamente uma ramo da Matemática.

O que distingue a Estatística Matemática da Teoria de Probabilidades é o fato de que na segunda o tratamento matemático está estritamente ligado a modelos probabilísticos. Já na primeira, os modelos probabilísticos são tomados como representando populações hipotéticas das quais se dispõe de informação apenas através de amostras. A Teoria de Probabilidade realiza a inferência a partir do modelo probabilístico assumindo o conhecimento pleno deste modelo. A Estatística Matemática realiza a inferência a partir de uma amostra gerada pelo modelo probabilístico, considerando que o modelo não é conhecido em si, mas somente os dados da amostra. Alguns procedimentos estatísticos, como a estatística não paramétrica, chegam ao ponto de tornar dispensável qualquer conhecimento a respeito do modelo probabilístico subjacente à amostra.

Apesar de elementos fundamentais da teoria estatística, como o Teorema Central do Limite e o Teorema de Gauss-Markov, tenham sido demonstrados no século XIX, foi somente na primeira metade do século XX que ela se estabeleceu como tal, isto é, como estatística matemática. O trabalho de Fisher “*On the Mathematical Foundations of Theoretical Statistics*”, publicado em 1922, é considerado a semente do seu desenvolvimento, sendo frequentemente definido como o trabalho individualmente mais influente do desenvolvimento da Estatística (Lehmann, 2011).

Fisher inicia o trabalho apresentando uma série de definições e desenvolve vários conceitos que se tornaram o fundamento da estatística matemática, como por exemplo os conceitos de suficiência, de eficiência, de estimador não viciado e o conceito técnico de verossimilhança como algo distinto de probabilidade. Ele também propôs o método da máxima verossimilhança com forma de estimação de parâmetros e demonstrou que este método era mais eficiente que o método dos momentos proposto por Karl Pearson.

Mas três aspectos de carácter mais filosófico são importantes e tiveram grande influência não só na teoria estatística mas na própria concepção da estatística. O primeiro é a definição de probabilidade:

“It is a parameter which specifies a simple dichotomy in an infinite hypothetical population, and it represents neither more nor less than the frequency ratio which we imagine such a population to exhibit.”

(Fisher, 1922, p.312)

O segundo aspecto é que Fisher define o propósito da Estatística como sendo a substituição dos dados por número relativamente pequeno de quantidades (as estimativas dos parâmetros) que representam adequadamente o todo. O terceiro é

a definição dos problemas fundamentais da estatística. Fisher apresenta três deles: (i) ESPECIFICAÇÃO, ou a definição do modelo probabilístico apropriado para a população infinita hipotética; (ii) ESTIMAÇÃO, isto é, discussão dos métodos de determinação dos valores dos parâmetros do modelo com base em amostras; e (iii) DISTRIBUIÇÃO, que inclui a discussão da distribuição, ou comportamento aleatório, das estimativas dos parâmetros ou de funções delas.

Estes três aspectos filosóficos do trabalho de Fisher se tornaram os princípios fundamentais sobre o qual se desenvolveu, e ainda se desenvolve, a inferência clássica ou inferência clássica. O desenvolvimento da estatística matemática foi rápido. Em 1930, os “*Annals of Mathematical Statistics*” são fundados sob a égide da “American Statistical Association” (ASA). Mas em 1933, quando o apoio da ASA mingou, seus fundadores criaram o o “*Institute of Mathematical Statistics*” como forma de apoio aos “*Annals*” (Stigler, 1996). Foi também em 1933 que Neyman e Pearson publicaram o trabalho que consolidou a sua proposta de testes estatísticos (Neyman e Pearson, 1933).

Durante a década de 1930 também ocorreu um importante debate sobre o livro de Horace Secrist, professor de estatística na Universidade Northwestern (EUA), que diagnosticava as causas da Grande Depressão econômica e propunha soluções. Secrist, utilizando uma grande quantidade de dados, concluiu através de análise de regressão que tanto os custos, quando os lucros das empresas, tendiam à média (mediocridade), de forma análoga à tendência da regressão da altura de pais e filhos revelada por Galton. Mas Harold Hotteling, estatístico matemático, demonstrou que o que Secrist havia encontrado era um resultado criado pela forma de analisar os dados. Na verdade, custo e lucro oscilavam aleatoriamente e, se outra forma de analisar fosse realizada, era possível demonstrar exatamente o contrário do que Secrist havia encontrado. A disputa envolveu estatísticos práticos do lado de Secrist e estatísticos matemáticos do lado de Hotteling, e acabou por estabelecer a proeminência da estatística matemática sobre os métodos exclusivamente práticos. A década de 1930 talvez deva ser considerada a época do estabelecimento da Estatística Matemática (Stigler, 1996).

2.5 *Problemas da Inferência Clássica*

No que foi apresentado acima, procurou-se levantar alguns dos aspectos históricos mais relevantes do desenvolvimento e do estabelecimento da Estatística como ciência, do século XVIII até o século XX. Muitos contribuíram para seu desenvolvimento, embora a apresentação tenha se restringido a alguns poucos, dentre os principais contribuidores da formação da nova ciência. Contudo, problemas sempre existiram e ainda existem, por isto, é importante apresentar os principais

aspectos problemáticos da Inferência Clássica. Nesta apresentação, segue-se principalmente Royall (1997), dada a sua clareza e contudência na crítica da Inferência Clássica, mas também Edwards (1992) e Hacking (1965).

Teoria de Neyman-Pearson

A proposta de Neyman e Pearson de como realizar o teste de hipóteses, baseada na teoria da decisão, pode ser considerada uma abordagem logicamente coerente e conceitualmente bem estruturada. Contudo, a base da qual ela parte é que o objetivo do teste de hipóteses é orientar o pesquisador numa tomada de decisão que consiste na escolha em duas possibilidades de ação. Portanto, rejeitar a hipótese nula, na teoria de Neyman-Pearson, significa escolher um linha de ação, enquanto que não rejeitá-la implica numa outra linha de ação. O teste de hipóteses se resume à escolha entre duas linhas de ação. Os erros tipos I e II são as margens de erro na escolha entre estas linhas de ação. O melhor teste de hipóteses é aquele que minimiza um dos dois tipos de erro, geralmente o erro tipo I, uma vez que não é possível minimizar ambos simultaneamente.

Neyman, no seu livro “*First Course in Probability and Statistics*”, publicado em 1950, enfatiza que a interpretação apropriada do teste de hipótese deve ficar restrita à questão da tomada de decisão. Ele usa a expressão “*comportamento indutivo*” (*inductive behavior*) para esta forma de testar hipóteses e descarta, como inapropriada, qualquer consideração ao respeito da falsidade da hipótese nula ou da hipótese alternativa. Mas, neste livro, ele vai mais além e estabelece o comportamento indutivo como o fundamento filosófico para a Estatística, afirmando que a Estatística Matemática lida como problemas relacionados às características da performance de regras de comportamento indutivo baseadas em experimentos aleatórios (Royall, 1997).

A teoria da decisão, que é o elemento que dá coerência e solidez à abordagem de Neyman-Pearson para o teste de hipóteses, é também o seu principal problema. Na pesquisa científica, raramente os pesquisadores pensam suas hipóteses com formas de direcionar a sua ação e, menos ainda, o teste de hipóteses como uma regra decisão sobre que linha de ação deve ser seguida. Na ciência, os dados são vistos como portadores da “evidência empírica” que é utilizada para se rejeitar ou aceitar hipóteses, que são formulações teóricas. Por vezes, se busca a evidência empírica presente nos dados como o juiz na escolha entre hipóteses candidatas a explicar o fenômeno estudado. A elaboração do teste de hipóteses como uma forma de tomada de decisão, ainda que logicamente coerente, é muito artificial e estranha à prática científica e à maneira como que os cientistas pensam as suas pesquisas.

Na maioria das áreas de conhecimento, as aplicações da Estatística na pesquisa científica não são bem representadas pela teoria de Neyman-Pearson. Muitos esta-

tísticos rejeitam esta abordagem e interpretam os métodos estatísticos como formas de revelar a evidência contida nos dados a respeito das hipóteses a serem testadas. Contudo, nesta operação, eles utilizam as mesmas ferramentas, métodos e conceitos estatísticos da teoria de Neyman-Pearson. Mais ainda, os mesmos tipos de problemas gerais são identificados — teste de hipótese, estimação e intervalo de confiança — e os procedimentos são avaliados, como na teoria de Neyman-Pearson, em termos de suas propriedades probabilísticas — tamanho do teste, poder do teste, viés, variância, etc (Royall, 1997).

Nas aplicações práticas da estatística, ao se relatar o resultado de um teste estatístico, por exemplo, que a hipótese nula foi rejeitada, o procedimento padrão requer que também se relate o nível de significância do teste e, às vezes, o poder do teste. No caso de se relatar o valor da estimativa de um parâmetro, por exemplo, a produção média de uma floresta, também se informa que o procedimento de estimação (estimador) não é viciado e também o valor do erro padrão da estimativa. Caso seja informado o intervalo de confiança da estimativa, também se informa o coeficiente de confiança utilizado, ou seja, a probabilidade de cobertura do procedimento que gerou o intervalo de confiança informado. Nos relatórios técnicos e científicos, o padrão é informar não apenas o resultado do procedimento estatístico, mas também as propriedades probabilísticas do procedimento utilizado. Isto mostra que o procedimento estatístico não está sendo utilizado como uma forma de escolha entre linhas de ação, mas como um modo de indicar “o que os dados dizem”, isto é, como um modo de interpretar os dados como evidência (Royall, 1997).

Infelizmente, ao se interpretar os procedimentos da teoria de Neyman-Pearson como uma forma de avaliar a evidência que os dados contêm a respeito de hipóteses ou estimativas, o resultado é a perda da coerência lógica da estrutura robusta que a teoria de Neyman-Pearson conferem aos procedimentos estatísticos. Felizmente, na maioria das situações, a incoerência lógica não resulta em interpretações totalmente erradas ou equivocadas. Contudo, em algumas circunstâncias, esta incoerência pode afetar seriamente a validade das conclusões científicas. Royall (1997) apresenta várias situações na área das ciências médicas.

Valor-p

Como na teoria de Neyman-Pearson, existem vários aspectos problemáticos com a abordagem Fisheriana do teste de significância. Mas, dada as diferentes interpretações e formas de aplicação que foram desenvolvidas, a discussão se torna muito complexa, uma vez que exigiria a referência ao trabalho de um grande número de estatísticos. Há, contudo, um aspecto claramente problemático na abordagem Fisheriana e a sua relevância é imensa, devido à universalidade da sua aplicação na pesquisa científica. Trata-se da interpretação do valor-p com medida da força

de evidência contra a hipótese nula.

É quase universal a concepção que quanto menor o valor-p maior a força de evidência que os dados observados apresentam contra a hipótese nula. Mas, como Royall (1997) demonstra claramente, esta concepção é simplesmente errada, ainda que o próprio Fisher a tenha afirmado (veja a citação na página 57).

Tomando-se a hipótese nula como premissa, o valor-p é a probabilidade de se observar o valor efetivamente observado da “estatística” utilizada no teste ou um valor mais extremo. Ignore-se o problema do que significaria a expressão “ou um valor mais extremo”, cuja única justificativa é que ela exprime a situação em que a hipótese alternativa se mostra mais realista que a hipótese nula, embora este aspecto não tenha nenhuma relação com a evidência contida nos dados. Um valor-p pequeno indica que, sob a hipótese nula, o que foi observado é um evento raro, logo o que se observou é uma evidência contra a hipótese nula. Mais ainda, quanto menor o valor-p, isto é, quanto mais raro a hipótese nula diz ser o que foi observado, mais forte a evidência do observado contra a hipótese nula.

Em essência, a interpretação do valor-p como evidência contra a hipótese nula afirma que situações que geram eventos raros não acontecem e que a raridade de um evento é evidência contra a existência da situação que o gerou, o que, obviamente é um absurdo. Royall (1997) chama esta interpretação de “*Lei da Improbabilidade*” e discute detalhadamente o seus problemas, avaliando vários enunciados possíveis para a “lei”, mas descarta todos como incoerentes ou problemáticos.

Embora a interpretação seja errada, a interpretação do valor-p dentro do contexto das regras de decisão da teoria de Neyman-Pearson é perfeitamente coerente. Por isto, rejeitar a hipótese nula todas as vezes que o valor-p é menor que o nível de probabilidade de 5%, garante que se terá no longo prazo uma margem de erro tipo I menor ou igual a 5%. Mas, o valor-p não informa absolutamente nada sobre o que de fato está acontecendo em situação particular em que a hipótese nula é testada. É a teoria de Neyman-Pearson que explica porque a interpretação errada do valor-p como evidência não resulta em grandes prejuízos na maioria das situações particulares da pesquisa científica. A popularidade desta interpretação, contudo, se tornou um grande obstáculo para que comunidade científica adote no julgamento de hipóteses abordagens apropriadas e coerentes para a interpretação da evidência contida nos dados.

Espaço Amostral

Um outro aspecto problemático da inferência clássica é a interferência do “*espaço amostral*” no resultado da inferência. O espaço amostral é o conjunto matemático de valores possíveis de serem observados como resultado de um estudo, seja ele um levantamento amostral, seja um experimento. Logo, o espaço amostral depende da concepção teórica que se tem do fenômeno sendo observado e do *pro-*

cedimento de observação adotado, mas não tem nenhuma ligação como os dados efetivamente observados.

Um exemplo simples torna o conceito de espaço amostral mais claro. Um dado laboratório, chamado de Lab-A, decidiu testar a toxicidade de um produto aplicando-o em 20 cobaias de laboratório. Após a aplicação, observou-se que seis delas morreram. Para testar a hipótese nula de que a toxicidade do produto é de 50%, o Lab-A utilizará uma distribuição binomial e observará que o valor-p é igual a 0,05765915, logo ele é ligeiramente maior que o nível de probabilidade de 5%, o que levaria o Lab-A a não rejeitar a hipótese nula. Um outro laboratório, chamado de Lab-B, na falta de cobaias suficientes, decidiu aplicar o mesmo produto nas cobaias à medida que elas se tornassem disponíveis, mas decidiu também que o experimento seria concluído quando a sexta cobaia morresse. Ao seguir o procedimento, o Lab-B observou que a sexta cobaia a morrer foi a vigésima cobaia a receber o produto. Para testar a hipótese nula de que a toxicidade do produto é de 50%, este laboratório terá que usar a distribuição binomial negativa e verificará que o valor-p é 0,02069473, o que levaria o Lab-B a rejeitar a hipótese nula com grande certeza.

Ambos laboratórios têm exatamente os mesmos dados para testar a hipótese, ou seja, 6 cobaias morreram dentre as 20 que receberam o produto, mas, ao utilizarem o teste de hipótese da inferência clássica, eles chegam a conclusões opostas. Isto acontece porque o procedimento utilizado para observar o fenômeno foi diferentes nos dois laboratórios. A diferença de procedimento resulta que os dados efetivamente observados estão associados a distribuições diferentes: dist. binomial no Lab-A e dist binomial negativa no Lab-B. Assim, o cálculo do valor-p depende não só do que foi efetivamente observado (6 cobaias mortas em 20), mas também do que se poderia observar, o que depende da distribuição em questão. No Lab-A, as observações “mais extremas que a observada” poderiam ser qualquer número de cobaias mortas entre 7 e 20, uma vez que apenas 20 receberam o produto. Já no Lab-B, o número de cobaias mortas é fixo (6), mas o número de cobaias que receberam o produto é variável. Logo, as observações “mais extremas” poderiam ser de 13 a 0 cobaias sobreviventes, que, com as seis cobaias mortas, dariam um total de 19 a 6 cobaias que receberiam o produto.

O problema se repete se for utilizada a abordagem do cálculo da região de rejeição que é mais característica da teoria de Neyman-Pearson, pois a região de rejeição depende da distribuição utilizada. O problema da interferência do espaço amostral no resultado da inferência é, portanto, uma implicação da inferência clássica, independentemente das divergências que possam existir entre as diferentes abordagens dentro dela.

Vários estatísticos clássico procuram justificar esta interferência como apropriada com base no princípio de que os métodos de observação são elementos fundamentais na pesquisa científica e, conseqüentemente, deveriam ser levados

em conta na inferência estatística. O fato é que a interferência do espaço amostral implica em levar em consideração na inferência observações que não foram feitas e, portanto, são meras possibilidades. Esta forma de inferência é equivalente à pergunta: “*O que aconteceria, se tivesse acontecido o que não aconteceu, mas que acreditamos que poderia ter acontecido?*”. Esta pergunta é um modo de se formular conjecturas, mas certamente não é a maneira adequada de se realizar a inferência científica.

Distribuições Amostrais

Distribuições amostrais são distribuições que descrevem o comportamento de estimativas obtidas com base em amostras. Por exemplo, o Teorema Central do Limite é um teorema da Teoria de Probabilidades que determina que a distribuição normal, também chamada de distribuição Gaussiana, é a distribuição limite para a média amostral, à medida que o tamanho da amostra tende ao infinito. A distribuição Gaussiana pode ser tomada, portanto, como a distribuição da média amostral no caso de grandes amostras.

As distribuições amostrais, como resultados de teoremas da Teoria de Probabilidades, são conceitos bem estabelecidos. A inferência clássica, contudo, é dominada pela idéia de que a única forma de se analisar a incerteza associada a qualquer estimativa ou “estatística” é por meio de uma distribuição amostral e, conseqüentemente, somente elas permitem uma inferência estatística *rigorosa*.

A dominância do conceito de distribuição amostral parece derivar do fato de que os estatísticos clássicos concebem as observações de um fenômeno natural como sendo “*uma amostra tomada de uma população hipotética de tamanho infinito que poderia gerar infinitas amostras*”. Conseqüentemente, as estimativas, ou qualquer quantidade calculada a partir dos dados, devem necessariamente possuir uma distribuição amostral, pois, em tese, poderiam ser observados “infinitas vezes”.

Esta concepção se deve provavelmente à influência de Ronald Fisher que, em seu influente trabalho de 1922, descreve a maneira de se alcançar o objetivo da inferência estatística:

“This object is accomplished by constructing a hypothetical infinite population, of which the actual data are regarded as constituting a random sample. The law of distribution of this hypothetical population is specified by relatively few parameters, which are sufficient to describe it exhaustively in respect of all qualities under discussion.”

(Fisher, 1922)

Ao apresentar o conceito da distribuição de probabilidades (*frequency curve*), ele

reforça esta concepção:

“Consequently, it should be clear that the concept of a frequency curve includes that of a hypothetical infinite population, distributed according to a mathematical law, represented by the curve.”

(Fisher, 1922)

Em outra passagem, ele enfatiza a idéia das observação como uma amostra de uma população infinita, independentemente da natureza da população, mas nesta passagem ele omite o adjetivo “hipotética”:

“It should be noted that there is no falsehood in interpreting any set of independent measurements as a random sample from an infinite population; for any such set of numbers are a random sample from the totality of numbers produced by the same matrix of causal conditions: the hypothetical population which we are studying is an aspect of the totality of the effects of these conditions, of whatever nature they may be. The postulate of randomness thus resolves itself into the question, “Of what population is this a random sample ?” which must frequently be asked by every practical statistician.”

(Fisher, 1922)

No mesmo trabalho, ao definir os três tipos fundamentais de problemas da Estatística ele coloca como terceiro tipo os problemas de definição das distribuições amostrais:

“(3) Problems of Distribution. These include discussions of the distribution of statistics derived from samples, or in general any functions of quantities whose distribution is known.”

(Fisher, 1922)

Os estatísticos clássicos deram maior importância aos problemas de determinação da distribuição amostral que a outro tipo de problema que Fisher também descreveu com fundamental, mas para os quais apresentou poucos desenvolvimentos. Trata-se do problema de especificação do modelo ou da distribuição probabilística da população infinita hipotética, que Fisher colocou como o primeiro tipo de problema:

“(1) Problems of Specification. These arise in the choice of the mathematical form of the population.”

(Fisher, 1922)

Nos inúmeros exemplos desenvolvidos por Fisher, o problema de especificação é tratado como uma questão de desenvolver uma construção racional razoável que justifique a escolha de um dado modelo matemático para a população infinita hipotética. Ele falou pouco sobre este problema que hoje é designado como especificação e seleção de modelos, mas é importante notar que, majoritariamente, os métodos desenvolvidos por Fisher tomavam a distribuição Gaussiana como premissa, ou seja, como o modelo matemático da população infinita hipotética.

Talvez por causa disto, os estatísticos clássicos tendem a tratar a questão da seleção de modelos na perspectiva da obtenção da “estimativa de uma medida de comparação de modelos”, com sua respectiva distribuição amostral. Por exemplo, o Critério de Informação de Akaike (*Akaike Information Criterion* — *AIC*) é frequentemente utilizado como uma medida apropriada para comparação de modelos. Mas, na inferência clássica, sempre se procura definir as suas propriedades probabilísticas e, se possível, convertê-lo num teste com distribuição amostral conhecida, como a distribuição de Qui-quadrado (veja por exemplo Claeskens e Hjort, 2008).

O estatístico clássico não parte dos dados, mas dos modelos. Ele deduz as propriedades amostrais do AIC para uma “classe de modelos”, como por exemplo os modelos lineares homocedásticos e estabelece que a maneira apropriada para se utilizar o AIC é através de tais propriedades amostrais. Mas, frequentemente, a tal “classe de modelos” é uma única distribuição probabilística, como no caso dos modelos lineares, que seguem todos eles a distribuição Gaussiana, diferindo apenas na função linear que descreve a média da distribuição. Um procedimento de comparação de modelos inteiramente condicional aos dados efetivamente observados e que, conseqüentemente, não busca definir uma distribuição amostral, deduzir propriedades probabilísticas ou calcular um valor-p será considerando um procedimento *pouco rigoroso* de inferência.

Hacking (1965) argumenta que a concepção de uma população hipotética infinita, utilizada com grande criatividade por Fisher, tem uma relação direta com a definição que von Mises oferecia para “probabilidade”. Por sua vez, a definição de von Mises guarda estreita semelhança com a abordagem tentada por Jacob Bernoulli no século XVIII. Mas, como Hacking (1965) deixa claro, a concepção de população hipotética infinita não ajuda a elucidar de modo definitivo o conceito de “probabilidade”, de modo que ele sugere o conceito de “cenário estocástico” (*chance set-up*) como uma definição logicamente mais apropriada para fundamentar o desenvolvimento da inferência estatística. Partindo-se do conceito de “cenário estocástico”, a necessidade imperativa das distribuições amostrais se desvanece, abrindo a possibilidade de outras abordagens para a inferência estatística.

Probabilidade como Medida de Evidência

O problema da definição de probabilidade é tratado com grande perspicácia por Hacking (1965) que demonstra a inadequação das diferentes definições para o desenvolvimento de uma forma de inferência estatística que seja apropriada à pesquisa científica. Mas, além deste problema, e talvez diretamente derivado dele, existe o problema da probabilidade ser tomada como a medida apropriada de evidência que as observações fornecem a respeito das hipóteses científicas.

Edwards (1992) argumenta que Ramsey, filósofo da matemática, demonstrou com sucesso que graus *absolutos* de crença em proposições devem seguir as leis de probabilidade para serem coerentes, mas que isto não implica que a probabilidade seja a medida apropriada de evidência que deve ser utilizada no teste de hipóteses científicas. As diversas formas de inferência estatística propostas — não só a inferência clássica, mas também a probabilidade inversa e a inferência bayesiana — procuram utilizar a probabilidade como uma medida de evidência absoluta no teste de hipóteses. Este é um problema central nestas formas de inferência e a teoria de Neyman-Pearson evita o problema realizando a inferência com base na teoria da decisão, evitando qualquer consideração sobre evidência.

Qualquer forma de cálculo de probabilidade tem como premissa um certo “*modelo*” sobre o funcionamento da realidade para poder justificar que os valores de probabilidade calculados sejam realistas. Como qualquer hipótese científica é em si num “*modelo*” da realidade, o uso da probabilidade como medida absoluta de evidência no teste de uma dada hipótese científica requer, portanto, a aceitação de um “*meta-modelo*” da realidade que englobe a hipótese sendo testada.

Para desenvolver sua teoria de inferência, Edwards (1992) recorre aos argumentos de Hacking (1965) e Birnbaum (1972) que afirmam que a verossimilhança é a medida apropriada sobre a qual se deve desenvolver o conceito de inferência estatística. Mas, Royall (1997) talvez seja o autor que mais enfatise que a verossimilhança é uma medida *relativa* de evidência, pois só permite a comparação de uma hipótese contra outra hipótese. A inferência estatística deve ser, portanto, uma forma de inferência baseada na evidência relativa contida nos dados a respeito de hipóteses concorrentes. Logo, ela não permite o teste de *uma hipótese*, mas apenas a comparação de uma hipótese *contra* outra hipótese.

Capítulo 3

Inferência Clássica e Mensuração Florestal

Surge a Biometria Florestal

3.1 Introdução

Pelos os textos apresentados no capítulo 1, fica claro que as expressões “Biometria Florestal” e “Inventário Florestal” não eram de uso corrente na América do Norte até a década de 1930. Os textos também revelam que a literatura de formação universitária em Mensuração Florestal não incluía nenhum treinamento em Estatística e, em essência, as inferências quantitativas eram guiadas pelo bom senso prático dos mensuracionistas.

Neste capítulo, apresenta-se algumas evidências de como a Inferência Clássica influenciou as transformações que ocorreram na Mensuração Florestal a partir da década de 1930. A expressão Biometria Florestal surge na literatura florestal depois que uma série de conceitos estatísticos foram assimilados nas práticas da Mensuração Florestal e nas análises dos experimentos na Ciências Florestais. É importante ressaltar que esta influência se deu inicialmente somente na direção da Estatística para a Biometria Florestal, uma vez que os criadores da Inferência Clássica não estiveram diretamente ligados aos problemas práticos florestais.

Inicialmente, mostra-se o aumento crescente de conceitos estatísticos nos textos de Mensuração Florestal na América do Norte, a partir da década de 1930. Depois, apresenta-se uma apreciação da presença dos livros que iniciaram a divulgação da Inferência Clássica como referências básicas nos textos de Mensuração Florestal. Segue-se uma discussão sobre o surgimento do Inventário Florestal e da Biometria Florestal como duas novas disciplinas pelo desdobramento do conceito de Mensuração Florestal. O capítulo é concluído com a apresentação de

alguns problemas da aplicação da Inferência Clássica na Biometria e Mensuração Florestal.

3.2 *A Estatística nos Textos de Mensuração Florestal*

O livro de Belyea (1931) talvez seja o primeiro a reconhecer a importância das técnicas estatísticas para inferência quantitativa na Mensuração Florestal, advogando a necessidade de formação estatística nos cursos de graduação. Embora o livro siga os mesmos conteúdos e a mesma forma de apresentação de livros anteriores, ele inclui um capítulo sobre a aplicação de métodos estatísticos. Neste capítulo, Belyea trata do uso da teoria da amostragem para determinação do erro amostral de levantamentos florestais:

“... an observation to an adequate number of samples will reveal conclusions, the stability of which permits us to apply to the whole forest generalities derived from the sample. This is known as the application of the theory of simple sampling.”

(Belyea, 1931, p.103)

Entretanto, Belyea vislumbrava uma importância muito maior para os métodos estatístico na Mensuração Florestal. No prefácio do livro, ele defende a necessidade da formação universitária em Engenharia Florestal incluir uma disciplina em métodos estatísticos:

“Some criticism, perhaps more or less justifiable, may be offered at the inclusion of a chapter on elementary statistics in a book of this type. It should be understood that there has never been the slightest attempt to cover the whole subject of statistics within the limits of a single chapter. Furthermore, the discussion has purposely been centered around the development, application and further treatment of a single phase — the standard deviation, a most important one in gaging the reliability of any series of biometric data. Considering the more recent advances and the present status of general work in the field of forest measurement, it is difficult to see how some considerations of this subject could be overlooked or omitted. It is the believe of the writer that the time is rapidly approaching when it will be necessary for all forestry students of collegiate grade to take a complete course in statistical methods. This, at the present, is a requisite in at least one forest school in the country and could well be imitated by others.”

(Belyea, 1931, p.iii-iv)

Este prefácio nos dá uma indicação clara de que no início da década de 1930 a formação em métodos estatísticos já era considerada um requisito para uma sólida formação em Mensuração Florestal, ao menos em uma escola Norte-americana.

Um segundo texto de Mensuração Florestal que é provavelmente um dos primeiros a fazer uso intensivo de métodos estatísticos é a artigo de Francis Schumacher e Francisco do Santos Hall publicado no “*Journal of Agricultura Research*” em 1933 (Schumacher e Hall, 1933). O artigo trata do uso da relação entre DAP, altura e volume sólido de árvores na escala logarítmica para a construção de tabelas de volume. Este artigo pode ser considerado um trabalho clássico de mensuração florestal, pois apresenta a racionalização básica para construção de tabelas de volume e propõe um modelo que desde então vem sendo utilizado em muitas situações práticas, sendo denominado de modelo Schumacher-Hall.

Deixando de lado o aspecto do desenvolvimento conceitual, o artigo de Schumacher-Hall chama a atenção pelo uso intensivo dos métodos de regressão linear vigentes na época, os quais diferem bastante da abordagem atual. Os autores iniciam o artigo afirmando que o método geral de construção de tabelas de volume consistem em *correlacionar* o volume de árvores individuais com o diâmetro e altura. Em seguida eles criticam os métodos em uso comum, pois eles implicam numa ignorância sobre a relação matemática entre as medidas das árvores, uma vez que eles se apoiam no processo de construção de curvas *à mão livre*¹, de modo total ou parcial.

“All the methods in common use imply ignorance as to the mathematical relationship of volume to dimensions; for they rely upon the free-hand curve process, either entirely or partially.”

(Schumacher e Hall, 1933, p.719)

Como exemplo de procedimento baseado inteiramente em curvas à mão livre ele cita o livro de Chapman e Demeritt (1932). Como exemplo do procedimento baseado parcialmente em curvas à mão livre ele cita o trabalho de Reineke e Bruce, publicado como boletim técnico pelo U.S. Department of Agriculture². Schumacher e Hall, ao contrário, depois de apresentarem a teoria que leva ao modelo linear do volume em função do DAP e da altura em escala logarítmica, ajustam o modelo para nove espécies arbóreas diferentes utilizando regressão linear.

¹Os métodos de curva à mão livre consistiam em fazer um gráfico de dispersão do volume médio das árvores por classe de DAP-altura e determinar *à sentimento* a melhor curva que passava entre as observações ou as médias que oscilavam num gráfico do volume em função do DAP, com várias curvas para as classes de altura e num gráfico do volume em função da altura com várias curvas para as classes de DAP.

²Reineke, L.H. e Bruce, D. An Alinement-Chart Method for preapering forest tree-volume tables. U.S. Dept. Agr. Tech. Bul. 304, 28p., 1932.

Além da estimativa dos coeficiente das equações de volume, os autores ainda fazem uso dos conceitos de regressão linear e estimação estatística para comparar as equações de volume de duas espécies, buscando responder a questão de se a tabela de volume de uma espécie pode ser utilizada para outra. Embora os procedimentos estatísticos utilizados sejam diferentes dos utilizados atualmente, eles são totalmente coerente com a forma com que a regressão linear era tratada na época, sendo estatisticamente rigorosos.

Um aspecto comum às análises apresentada tanto por Belyea quanto por Schumacher e Hall é a abordagem utilizada para fazer a inferência quantitativa. Os autores utilizam o erro padrão da estimativa como medida de incerteza para avaliar as estimativas obtidas. Belyea faz isto no contexto de valiar a qualidade de uma estimativa da produção volumétrica de uma floresta. Schumacher e Hall usam o erro padrão da estimativa para avaliar as estimativas dos coeficientes de regressão da equação de volume e das próprias previsões geradas pela equação. Estes dois autores também tomam os erros padrão das estimativas de equações de duas espécies para julgar se uma mesma equação poderia ser aplicada à previsão do volume das árvores de ambas as espécies, sem perda significativa de exatidão.

É importante lembrar que o livro de Belyea é dedicado ao ensino universitário da Engenharia Florestal e o artigo de Schumacher e Hall é um texto técnico-científico que propõe um forma mais objetiva de se construir e avaliar as tabelas de volume. Portanto, estes dois trabalhos podem ser tomados como um exemplo da maneira como a inferência estatística foi introduzida na Mensuração Florestal como forma de medir quantitativamente a incerteza associada a estimativas e previsões.

Mas as técnicas estatísticas foram rapidamente assimiladas na Mensuração Florestal. Em 1935, Bruce e Schumacher publicam o seu livro “*Forest Mensuration*” rompendo com o formato tradicional dos livros de Mensuração Florestal até então. No prefácio, eles fazem uma dura crítica ao ensino da Mensuração Florestal a partir de uma analogia com a carpintaria. Na carpintaria, o ensino é orientado primeiramente para o aprendizado e domínio de ferramentas e técnicas básicas, e depois para o desenvolvimento de projetos progressivamente mais complexos. Mas o ensino de mensuração é focalizado numa sequência de operações que em si não permitem resolver qualquer problema prático:

“Forest mensuration, in the past, has been taught pretty much as carpentry has not been taught. The student was successively shown how to make a log rule, a local volume table, a regional volume table, a current growth study, a yield table. If, once out of school, he had occasion to do any of these things, he probably fount that, owing to some peculiar local conditions the procedure he had so carefully learned was inapplicable. ... Had he been taught

primarily to use his tools, he would have devised the necessary procedure more easily and confidently and without the embarrassing feeling that he was being a daring innovator, doing violence to the tradition of his profession.”

(Bruce e Schumacher, 1935, p.v)

Bruce e Schumacher afirmam que houve um período de grande desenvolvimento destas ferramentas, das quais algumas são gráficas, mas a maior contribuição veio da “moderna” ciência da Estatística.

“Moreover, the past few years have been a period of great progress in devising and adapting new tools. Some of these are graphic, such as the alignment chart and anamorphosis, but the largest contribution is from the modern science of statistics. The standard deviation, the frequency curve, the correlation coefficient, and the regression equation have been found admirably adapted to forestry work by the still rather small group of investigators who have learned their use.”

(Bruce e Schumacher, 1935, p.v-vi)

Assim, os autores elaboraram um livro de Mensuração Florestal que inova pelos conteúdos tratados, pela forma de apresentação do conteúdo e pelos termos utilizados. O livro é dividido em quatro partes, que em si já demonstram a nova concepção de mensuração. A parte I (*Direct Measurement*) trata da medição de árvores: diâmetro, altura, volume e idade. A parte II (*Direct Estimate by Sampling*) é totalmente nova, fornecendo a formação básica em estatística: medidas de posição (média, mediana e moda), medidas de dispersão (desvio padrão), a distribuição Normal (denominada curva normal de erros), o erro padrão, distribuições assimétricas e aplicações da distribuição Normal. Esta parte é concluída com um capítulo inteiramente dedicado à amostragem. Na parte III (*Indirect Estimate Based on One or More Independent Variables*), é tratada a questão dos modelos de predição, essencialmente a relação hipsométrica e as tabelas de volume. Forte ênfase é dada aos métodos gráficos, que incluem o uso da curva “a mão livre”, e a aplicação do método dos quadrados mínimos para o caso de uma e de duas variáveis independentes. O livro é concluído com a parte IV (*Prediction of Growth and Yield*) que aborda a predição de crescimento e produção de forma bastante convencional. Curiosamente, os métodos estatísticos não são discutidos nesta parte do livro, nem mesmo na forma de exemplos de aplicação.

Sete anos depois, Bruce e Schumacher publicaram a segunda edição do livro, no prefácio da qual eles reconhecem o desenvolvimento e aplicação de novas ferramentas estatísticas, mas enfatizam a eficiência e adaptabilidade dos métodos apresentados no livro.

“The seven years that have elapsed since this book was first written have been a period of rapid evolution in the procedures of forest mensuration but the fundamental tools described here have merely proved their efficiency and adaptability. There have been new tools devised in the science of statistics and adapted to mensuration, but their use is primarily in the field of advanced research and it has seemed unwise to describe them in this elementary text.”

(Bruce e Schumacher, 1942, p.v)

Eles reconhecem o rápido desenvolvimento da aplicação da teoria da amostragem e a necessidade de refazer a parte do livro relativa à ela.

“Much attention has been given in late years to the theory and technique of sampling, and the chapter thereon has been almost completely rewritten.”

(Bruce e Schumacher, 1942, p.v)

De fato o capítulo de amostragem foi quase que completamente reelaborado, apresentando de forma mais detalhada e didática a seleção de amostras aleatórias e introduzindo a amostragem sistemática. A apresentação também demonstra as vantagens do processo de estratificação e faz uma comparação entre amostragem aleatória e sistemática. A amostragem sistemática como delineamento amostral ainda não se mostra clara, o que resulta em certa confusão na comparação de amostragem aleatória e sistemática.

O livro de Bruce e Schumacher, apesar do mesmo título de livros anteriores “*Forest Mensuration*”, parece inaugurar um novo paradigma não só do ensino da Mensuração Florestal, mas principalmente do conteúdo e prática desta disciplina. No novo paradigma, o delineamento amostral e o ajuste de modelos passam a ser ferramentas fundamentais para a prática da mensuração. Ambas ferramentas ainda se mostram, em 1942, de uma forma relativamente incipiente, pois muito progresso teórico e prático estava por ocorrer. Mas, um novo caminho para Mensuração Florestal foi definitivamente apresentado neste livro.

3.3 *A Influência da Inferência Clássica*

Talvez a área de ação humana que sofreu maior influência da Estatística durante o século XX foi a da pesquisa científica. O desenvolvimento do conceito de teste estatístico de hipóteses e das concepções de amostragem, bem como o seu fortalecimento teórico, fez com que a Estatística exercesse enorme influência sobre as

concepções teóricas de várias ciências e disciplinas³, através da transformação de seus métodos e de suas técnicas de pesquisa. Foi exatamente assim que ocorreu o surgimento da Biometria Florestal a partir da Mensuração Florestal. No início do século XX, os principais conceitos teóricos da Estatística se estabeleceram de maneira mais firme e sua aplicação se ampliou para inúmeras áreas de conhecimento.

Talvez a influência tenha sido exercida mais fortemente através de livros que eram dedicados aos pesquisadores em geral e ao mesmo tempo organizaram e sistematizaram a Estatística como área de conhecimento. Um destes primeiros livros foi o de Yule “*An Introduction to the Theory of Statistics*” (Yule, 1919), publicado pela primeira vez em 1919, mas que em 1927 já estava na sua oitava edição. Belyea no seu livro de Mensuração Florestal de 1931 (Belyea, 1931) indica a primeira edição do livro de Yule como a referência básica em estatística. Schumacher e Hall, no seu trabalho sobre equações de volume de 1933 (Schumacher e Hall, 1933), tomaram a oitava edição como referência para os métodos de regressão que utilizaram. Bruce e Schumacher, no seu livro de 1935 (Bruce e Schumacher, 1935), fazem duas referências à oitava edição do livro de Yule como a fonte para métodos mais acurados que os métodos por eles utilizados na determinação da mediana e da moda.

Mas talvez o livro de estatística de maior influência em todo o século XX tenha sido o livro de Fisher “*Statistical Methods for Research Workers*” (Fisher, 1925a), no qual Fisher apresenta a *sua maneira* de aplicar a estatística, o que viria a constituir o cerne da inferência clássica. A recepção do livro não foi muito positiva por parte dos especialistas da área, sendo a queixa mais frequente a ausência de provas matemáticas dos métodos utilizados. Mas o livro foi um sucesso como livro técnico. A primeira edição de 1925 se esgotou em três anos, sendo necessária uma segunda edição que novamente foi um sucesso, de modo que a cada dois ou três anos uma nova edição se fazia necessária, embora as tiragens crescessem a cada edição. A última edição, a décima quarta, foi publicada postumamente em 1970 com as notas que Fisher havia preparado antes da sua morte em 1962 (Lehmann, 2011).

O impacto do livro de Fisher se fez sentir tanto no campo da estatística, através dos inúmeros desenvolvimentos da teoria estatística que inspirou e influenciou, quanto na disseminação da aplicação dos métodos estatísticos na diversas áreas do conhecimento. O livro promoveu uma mudança completa nos métodos estatís-

³A partir deste ponto torna-se necessário a distinção clara entre ciência e disciplina. A palavra ciência será utilizada para designar uma área de “*conhecimento teórico*”, isto é, o conhecimento que busca “*explicar como as coisas são*” e, portanto, aumentar a compreensão do mundo. A palavra disciplina será usada para designar uma área de “*conhecimento prático*”, isto é, o conhecimento que busca “*definir com as coisas devem ser*” e, conseqüentemente, dirigir a ação sobre o mundo.

ticos, sendo que suas edições representam a transformação histórica dos métodos, pois a cada nova edição Fisher fazia melhorias na apresentação de alguns conteúdos e incluía novos assuntos (Lehmann, 2011).

Mas a grande disseminação da abordagem Fisheriana de inferência clássica não foi devida apenas ao livro de Fisher. Em 1937, George Snedecor (1881-1974), professor do departamento de estatística da Iowa State University, publicou o livro “*Statistical Methods*”, com o subtítulo “*Applied to Experiments in Agriculture and Biology*”. Escrito na forma de uma conversa entre autor e leitor, o livro tem um estilo bem mais amigável e de leitura bem mais fácil que o livro de Fisher. O livro atingiu uma audiência muito mais ampla, tendo vendido mais de 200.000 cópias nas suas várias edições. Durante alguns anos, foi o livro mais citado no Science Citation Index, que hoje faz parte do site Web of Science (Lehmann, 2011).

Os livros de Fisher e Snedecor tiveram um forte impacto na Engenharia Florestal, particularmente na Mensuração Florestal. O livro de Chapman e Meyer (1949), no capítulo sobre técnicas estatísticas fundamentais, apresenta ambos os livros como referências estatísticas. O livro de Snedecor figura entre as referências bibliográficas nos livros de inventário florestal de Spurr (1952) e de mensuração de Husch (1963). Já o livro de Biometria Florestal de Prodan (1968) é rico em referências a livros de estatística e os livros de Fisher e Snedecor figuram entre elas.

Dois publicações sobre estatística que tiveram influência na formação de Engenheiros Florestais na América do Norte são os manuais (“handbooks”) publicados pelo Serviço Florestal Norte-Americano e de autoria de Frank Freese (1922-1996) “*Elementary Forest Sampling*” (1962) e “*Elementary Statistical Methods for Foresters*” (1967). Os livros de Fisher e Snedecor figuram entre as referências para leitura adicional e, no prefácio de ambos manuais, Freese agradece aos autores pela permissão para a reimpressão de tabelas estatísticas. Os manuais de Freese, por sua vez, se tornaram referência estatística básica para os livros de Husch, Miller e Beers (1982) e de Avery e Brukhart (1983), que são livros texto mais recentes voltados ao ensino da Mensuração Florestal de nível universitário.

À medida que a inferência clássica foi sendo aplicada na Mensuração Florestal, a concepção de mensuração foi gradativamente transformada resultando na sua divisão em duas novas disciplinas. A Biometria Florestal e o Inventário Florestal tiveram a sua origem na abordagem clássica de inferência quantitativa, seguindo as linhas de pensamento proposta por Fisher e Neyman.

3.4 *Novas Disciplinas*

A partir da sua introdução na Mensuração Florestal nas década de 1930 e 1940, a Inferência Clássica foi adquirindo maior aplicação e importância nos procedimentos de mensuração. Aparentemente, isto aconteceu em duas linhas paralelas. A primeira delas na introdução da teoria da amostragem nos procedimentos de estimação da produção de povoamentos e florestas. A outra linha foi a aplicação inicialmente do método dos quadrados mínimos e, posteriormente, da teoria de modelos lineares, regressão e análise de variância, na construção de tabelas de volume, tabelas de produção e de crescimento, e na análise de resultados experimentais.

Inventário Florestal

Já em 1942, surge um livro inteiramente dedicado à aplicação da teoria e dos métodos de amostragem na área florestal: “*Sampling Methods in Forestry and Range Management*” (Schumacher e Chapman, 1942). Mas um texto que marca a maturidade da aplicação da teoria da amostragem nos levantamentos florestais é o livro “*Forest Inventory*” de Stephen H. Spurr de 1952. Ao contrário dos livros de mensuração mencionados até agora, que eram voltados ao ensino universitário da Mensuração Florestal, Spurr coloca “*Forest Inventory*” como um relatório abrangente do estado das técnicas de mensuração da época. Spurr admite que sua abordagem é basicamente estatística e rompe explicitamente com o paradigma do século XVIII que considerava possível descrever matematicamente a forma das árvores:

“ The approach is largely a statistical one, tempered by logic and the knowledge that the forest tree is not simply a geometrical solid, but a biological unit which is not definable in purely mathematical terms.”

(Spurr, 1952, p.iii-iv)

Embora o livro seja voltado para leitores praticantes e interessados em Mensuração Florestal, Spurr menciona que além do conhecimento em mensuração o livro assume que o leitor conheça as técnicas estatísticas mais simples usadas na pesquisa em agricultura.

O livro é dividido em quatro partes: medição de árvores e parcelas, estimação do volume, estimação do crescimento, delineamento amostral e do inventário. Assim, uma parte, composta de sete capítulos (69 páginas), é inteiramente dedicada à questão de amostragem e organização do inventário. Um capítulo é inteiramente dedicado à teoria da amostragem, onde são tratados delineamentos amostrais que

ainda hoje são considerados básicos para compreensão das técnicas de levantamentos florestais: amostragem aleatória simples, amostragem estratificada, subamostragem, amostragem dupla e amostragem sistemática. Todo o processo de estimação é fundamentado no conhecimento da área dos povoamentos e da floresta e uma outra inovação do livro é a grande ênfase na aplicação de fotogrametria na determinação da área e na classificação dos povoamentos florestais.

A fotogrametria representa o início das técnicas modernas de sensoriamento remoto, logo, o livro de Spurr apresenta o inventário florestal nos moldes em que é realizado até hoje. Mais ainda, ele consolida o uso da palavra “*inventário*” em expressões como “*inventário florestal*” e “*delineamento de inventário*”. Assim, o livro de Spurr é aparentemente um marco do surgimento da nova disciplina de “*Inventário Florestal*” dentro, ou a partir, da Mensuração Florestal.

Vários livros foram posteriormente publicados na América do Norte com o título Mensuração Florestal, como por exemplo os livros de Husch, Miller e Beers (1982) e o de Avery e Burkhardt (1983). No entanto, a disciplina “*Inventário Florestal*” está aparentemente consagrada àquilo que originalmente foi a aplicação da teoria e dos métodos de amostragem dentro da Mensuração Florestal. Na década de 1980, a aplicação da teoria de amostragem como método de inventário florestal já está madura e com alto grau de sofisticação, como mostra o livro de Pieter G. de Vries “*Sampling Theory for Forest Inventory*”, com o subtítulo “*A Teach-Yourself Course*” (de Vries, 1986). Dois exemplos mais recentes são o livro de Schreuder, Gregoire e Wood (1993) “*Sampling Methods for Multiresource Forest Inventory*” e o livro editado por Kangas e Maltamo (2006) “*Forest Inventory: Methodology and Applications*”. Estes três últimos livros demonstram a complexidade e sofisticação com que a teoria da amostragem é hoje desenvolvida e aplicada na Mensuração Florestal.

O Brasil é também um exemplo da força da introdução da teoria da amostragem na Engenharia Florestal. Como a profissão foi introduzida no país na década de 1960 e foi consolidada na década de 1970, a expressão “*mensuração florestal*” foi tradicionalmente muito pouco utilizada. A expressão consagrada para designar a área é a de “*Dendrometria e Inventário Florestal*”. A dendrometria é relativa à medição das árvores individualmente, enquanto que o inventário florestal refere-se à quantificação da produção na floresta como um todo.

Biometria Florestal

A segunda linha de aplicação da estatística, relativa ao método de quadrados mínimos e à teoria de regressão linear, aparentemente teve um desenvolvimento mais lento. Schumacher e Hall, no seu artigo de 1933, afirmam categoricamente a superioridade do método dos quadrados mínimos em comparação com os métodos de curva a mão livre na construção de tabelas de volume. No seu livro paradigmático

de 1935, Bruce e Schumacher apresentam o método de quadrados mínimos como uma maneira mais eficaz de se estabelecer uma relação quantitativa entre variáveis florestais. No entanto, a grande ênfase do livro na construção de tabelas de volume e produção é nos métodos gráficos que se apoiam total ou parcialmente na curva a mão livre. Bruce e Schumacher justificam a maior aplicação dos métodos gráficos pelo fato do método de quadrados mínimos serem computacionalmente muito laboriosos e necessitarem a apresentação matematicamente explícita da forma de relação entre as variáveis.

“In general, the advantage of the method of least squares is that it eliminates all errors of personal judgement and gives assurance that the result has the highest probability of correctness possible with the data available. The disadvantages are that the computations are laborious and that the method is applicable only to cases where the form of the equation is known or may reasonably be used as a hypothesis. . . . In general, therefore, since most of the relations with which forestry is concerned have no known equation form, this method is relatively little used in practical work.”

(Bruce e Schumacher, 1935, p.186)

Estes dois aspectos, a laboriosidade computacional e a necessidade do conhecimento prévio da forma da equação, parecem ter se mantido por algumas décadas como o grande obstáculo para aplicação do método de quadrados mínimos e da regressão linear na Mensuração Florestal. Husch, no seu livro “*Forest Mensuration and Statistics*” de 1963, ainda apresenta, 28 anos depois, os mesmos argumentos.

“The construction of volume tables using regression analysis is based on the premise that volume is related to the chosen independent variables according to a definite mathematical function which will reveal itself from a series of samples. With graphic techniques, this mathematical function is not necessarily explicitly defined, but it is implicit in the method. . . . When least squares fitting is used, the form of the equation expressing the relation of volume to size measurements must be decided upon, a priori, and the constants calculated which will give the best fit for the chosen equation.”

(Husch, 1963, p.114)

Apesar de um capítulo inteiramente dedicado à análise de regressão e correlação, Husch utilizam os métodos gráficos nos exemplos práticos de construção de relações hipsométricas e tabela de volume.

O livro de 1963 de Husch é de certa forma também um marco na Mensuração Florestal. O livro é dividido em duas partes. A primeira parte, composta de 14 capítulos, apresenta em 267 páginas todo o conteúdo de Mensuração Florestal, incluindo Inventário Florestal, delineamentos amostrais e um capítulo inteiramente dedicado à amostragem com probabilidade proporcional ao tamanho, onde detalha o método de Bitterlich.

A segunda parte (“Introductory Statistics in Forestry”) é inteiramente dedicada a conteúdos de estatística e sua aplicação na Engenharia Florestal. Ela se inicia com uma apresentação introdutória de probabilidade seguida da apresentação de métodos gráficos e estatísticas descritivas de sumarização para uma variável: histogramas, medidas de posição e medidas de dispersão. Um capítulo é dedicado à apresentação de distribuições estatísticas: Normal, binomial e Poisson. O capítulo dedicado à amostragem comprova o grau de maturidade que o assunto já havia alcançado. Além dos delineamentos básicos de amostragem aleatória, sistemática e estratificada, são tratados delineamentos mais complexos amostragem com probabilidade de seleção variável, amostragem em múltiplos estágios e amostragem por conglomerados. O conceito e aplicação de intervalo de confiança já é claramente apresentado.

Curiosamente, um capítulo é dedicado à questão da computação por máquinas (*Machine Computation*), onde Husch discute o uso de calculadoras de mesa e o uso de sistemas eletrônicos com uso de cartões perfurados, isto é, os primeiros computadores eletrônicos. A tecnologia do início da década de 1960 se não permitia a implementação rotineira da regressão linear, já preparava uma tecnologia que o permitiria. Husch parece ter percebido a importância desta tecnologia.

Nos últimos capítulos da segunda parte, Husch apresenta a abordagem da inferência clássica. Um capítulo sobre testes de significância (*Tests of Significance*) onde apresenta o teste t de Student para amostras pareadas e para amostras com variância igual ou diferente, o teste F para igualdade de variâncias, e o teste de Qui-quadrado para tabelas de contingência e para aderência de distribuições. Segue um capítulo sobre análise de regressão e correlação (*Regression and Correlation Analysis*) que segue a abordagem hoje considerada convencional. O tratamento da estatística é concluído com um capítulo sobre análise de variância (*Analysis of Variance*) que apresenta a análise para experimentos inteiramente ao acaso e blocos ao acaso, bem como para arranjos fatoriais, e que discute ainda a questão da homogeneidade de variâncias.

Mas o livro que provavelmente inaugurou a disciplina de Biometria Florestal foi o livro “*Forstliche Biometrie*” de Michail Prodan de 1961, cuja a tradução para o inglês foi publicada em 1968 com o título “*Forest Biometrics*”. Trata-se de um livro de estatística aplicada à Engenharia Florestal que ultrapassa em amplitude e profundidade as questões práticas da Mensuração Florestal. Prodan produziu um apresentação completíssima da teoria e métodos estatísticos, ainda

que muito resumidamente, aplicando-os e ilustrando-os com questões florestais. O livro começa com dois capítulos conceituais que tratam de prerrequisitos matemáticos (*Mathematical Aids*) e levantamentos estatísticos (*Statistical Surveys*). Segue um capítulo onde o conceito de distribuição probabilística é apresentado (*Frequency Distributions. Concept and Examples*) e dois capítulos sobre medidas descritivas: medidas de tendência central (*Measures of Central Tendency*) e medidas de dispersão (*Measures of Dispersion and Variation*). Prodan discute a questão do agrupamento dos dados (*Errors Caused by the Grouping of Observations into Classes*) e conceitos básicos de probabilidade (*Calculation of Probabilities*). Numa sequência de sete capítulos, Prodan expõem a teoria das distribuições probabilísticas até as distribuições amostrais. Seguem-se capítulos voltados à uso da estatística para inferência: um capítulo sobre análise de variância (*Analysis of Variance*), um capítulo sobre amostragem (*Sample Surveys*), um capítulo sobre planejamento de experimentos (*The Planning of Experiments*), quatro capítulos sobre regressão e correlação (*The Calculation of Correlation and Regression, Curve Fitting by Orthogonal Polynomials, Fitting Regression Subject to Periodic Variation, Multiple Regression and Correlation*), um capítulo sobre funções de crescimento (*Growth Functions*), um capítulo sobre econometria (*Methods and Terminology of Economic Statistics*), e um capítulo sobre séries temporais (*Time Series*). Prodan conclui o livro com um breve capítulo sobre programação linear (*Brief Introduction to Linear Programming*).

A abrangência e profundidade do livro são impressionantes e admiráveis. Pela estrutura do livro podemos concluir que o conceito de Biometria Florestal de Prodan era simplesmente de “*estatística aplicada à Engenharia Florestal*”. Mas, a estatística que Prodan tinha em mente era a da inferência clássica.

3.5 Inferência Clássica na Biometria Florestal: Problemas

Mais que a aplicação de métodos e técnicas estatísticas, a incorporação da inferência clássica à Mensuração Florestal resultou em profundas transformações conceituais. Talvez a transformação mais profunda seja aquela que gerou as disciplinas de Biometria Florestal e de Inventário Florestal, que são duas partes de uma mesma questão prática que originalmente era tratada por uma única disciplina florestal: a Mensuração Florestal. Esta divisão só faz sentido na medida que se aceite uma distinção dos métodos e das técnicas estatísticas de construção de modelos e análise de experimentos, daqueles de estimação com base em amostras.

As questões práticas tiveram grande influência no desenvolvimento da inferên-

cia clássica, pois a elaboração dos métodos e técnicas se deu a partir de problemas práticos e a constituição da teoria estatística se deu pela generalização das aplicações práticas. Tais questões práticas, entretanto, não eram de origem florestal ou mensuracional. Assim, a aplicação da Inferência Clássica nas Ciências Florestais em geral, e na Mensuração Florestal em particular, resultou em alguns problemas e incoerências que permaneceram ao longo do tempo. Segue-se a apresentação de três problemas que são relevantes para a prática da Mensuração Florestal.

Escolha de Modelos para Distribuição de Diâmetros

Desde há muito tempo a “*estrutura da floresta*” foi identificada com a estrutura de tamanho das árvores, a qual, por sua vez, é geralmente representada pela frequência das árvores nas diferentes classes de DAP. Já em 1898, de Liocourt apresentou um tratamento matemático para a representação da frequência por classes de DAP, estabelecendo um padrão que ficou conhecido como “*Lei de de Liocourt*” (de Liocourt, 2001). O tratamento matemático do problema é, portanto, muito anterior ao advento da Inferência Clássica.

Este tipo de problema era provavelmente muito frequente nas ciências naturais pois um dos primeiros problemas práticos abordados por Karl Pearson no final do século XIX e início do século XX foi como verificar se a frequência relativa de uma dada medida segue um modelo teórico proposto. Ele nomeou este problema como o problema do “*test of goodness-of-fit*”, atualmente traduzido como problema do teste de aderência, e sugeriu como solução o famoso teste de Qui-Quadrado (Stigler, 1986).

Com o estabelecimento do conceito de teste de significância e da teoria de Neyman-Pearson, o teste de aderência passou a ser aplicado aos problemas de representação da distribuição do DAP, utilizando-se como nível de probabilidade, ou nível de significância, o valor limite de 5%. Inicialmente, aplicando-se o teste de Qui-quadrado e, posteriormente, o teste de Komolgorov-Smirnov, generalizou-se que o problema da representação da distribuição de DAP pode ser colocado nos seguintes termos: “*A estrutura de diâmetro de uma floresta pode ser adequadamente representada pela distribuição W ?*”. Tome-se W como uma distribuição julgada razoável para floresta em questão, seja ela a distribuição normal, a exponencial, a Weibull, a gama, ou qualquer uma outra. Se o teste implementado se mostra não significativo no nível de 5%, a distribuição testada é julgada como apropriada.

O primeiro problema que surge é quando se testa mais de uma distribuição e duas ou mais delas se mostram não significativas. Nenhum dos testes implementados pode dar qualquer orientação sobre qual das distribuições que se mostraram não significativas é mais apropriada como modelo da distribuição de DAP. Os testes comparam a frequência relativa esperada pelo modelo com a frequência re-

lativa observada, logo, não fornecem qualquer informação comparativa entre os modelos candidatos. Alguns pesquisadores podem lançar mão do conceito do valor-p como medida de evidência contra a hipótese nula, e assim selecionar a distribuição com o *maior* valor-p, ainda que todos eles sejam maior que 5%. Mas, como foi apresentado, o valor-p não é medida de evidência contra a hipótese nula e lançar mão deste expediente seria imaginar que é possível resolver um problema por meio de um erro de lógica.

Um segundo problema surge mesmo no caso que se faça uma aderência total a teoria de Neyman-Pearson e se descarte a idéia de valor-p. Se uma única distribuição se mostrar não significativa, a decisão tomada será a de não se rejeitar a hipótese nula. No teste de aderência, a hipótese nula estabelece que a frequência observada não difere da frequência esperada pela distribuição sendo testada. Se ela não for rejeitada, pode se encorcer num erro tipo II, isto é, não rejeitar uma hipótese nula falsa. Mas, a decisão foi tomada com base no nível de probabilidade de 5%, que é a margem de *erro tipo I aceitável* e, conseqüentemente, toma-se a decisão sem qualquer conhecimento da margem de erro tipo II, que é o erro realmente relevante no caso da escolha de uma distribuição para representar o comportamento dos dados.

Além desses dois problemas de carácter geral, existe um problema de carácter específico à modelagem da distribuição de DAP. Num levantamento florestal, as árvores são observadas segundo parcelas ou outro tipo qualquer de unidade amostral. A agregação de todas as árvores medidas num levantamento, ignorando-se o fato delas formarem conglomerados nas parcelas, leva a se testar uma distribuição de DAP que não existe. Com efeito, a distribuição de DAP de uma floresta, como qualquer outro atributo dela, varia de local para local e, portanto, varia de parcela para parcela. O agregado de todas as árvores do levantamento é uma distribuição de lugar nenhum da floresta. Pode-se argumentar que o agregado representa a “*distribuição de DAP média*” da floresta, mas, neste caso, é importante lembrar que o DAP ou altura médios são o DAP e a altura de árvore nenhuma, pois eles são informações abstratas da floresta como uma população. Se a “*distribuição de DAP média*” for tomada e interpretada como uma informação direta da floresta, o procedimento é seguro.

Contudo, frequentemente, a distribuição de DAP é tomada como ponto de partida para confecção de tabelas de produção dos povoamentos para fins de manejo florestal. Neste caso, só faz sentido se falar de uma “*distribuição de DAP média*” para o povoamento, que é a unidade de manejo, e não para floresta como um todo. Se, por outro lado, a distribuição de DAP for parte de um sistema de simulação de crescimento da floresta, ela deverá ser representada para cada uma das unidades de simulação que, geralmente, são as parcelas ou unidades amostrais. A situação prática da modelagem da estrutura de tamanho da floresta implica que os dois problemas estatísticos apresentados acima serão multiplicados pelo nú-

mero de povoamentos a serem manejados ou pelo número parcelas presentes no levantamento florestal.

Colocar o problema da representação da distribuição de DAP nos termos: “A estrutura de diâmetro de uma floresta pode ser adequadamente representada pela distribuição W ?” é um equívoco. Esta forma de colocar o problema é provavelmente resultado da facinação que os testes de aderência geraram nos primeiros profissionais florestais que procuraram tratar o problema de uma forma mais rigorosa utilizando a inferência estatística. A colocação é muito semelhante à forma como Karl Pearson olhava o seus problemas, que não eram florestais.

Na perspectiva da Mensuração Florestal, uma forma mais adequada de se definir o problema é a seguinte: “Qual é o melhor modelo (distribuição probabilística) para representar a estrutura de DAP da floresta em cada unidade (povoamento ou parcela) utilizada para o manejo ou a simulação?”. Nesta perspectiva, os testes de aderência são absolutamente inúteis, pois eles não podem auxiliar na busca de uma resposta para esta pergunta e, ao se evitar tais testes, evita-se todos os seus problemas de incoerência e inadequação.

Estimação e Predição

Estimação e predição são dois conceitos frequentemente confundidos nas literatura técnica florestal, particularmente na literatura brasileira. Deve se entender por estimação o procedimento de determinação do valor de um parâmetro de um modelo ou população como, por exemplo, se estimar o DAP médio de uma floresta ou os coeficientes de regressão de uma equação de volume. Já a predição é o procedimento de se obter o valor de uma nova observação com base num modelo estatístico.

No caso da Mensuração Florestal, a predição de uma nova observação significa utilizar um modelo para determinar o atributo de uma árvore ou arvoredo que geralmente é de medição trabalhosa ou só pode ser medido por meio destrutivo. Por exemplo, se utiliza a relação hipsométrica para se *predizer* a altura das árvores de uma parcela com base no DAP medido. Outro exemplo é o uso de equações de biomassa para se *predizer* a biomassa de árvores lenhosas sem necessidade de abatê-las. Também se pode utilizar modelos de crescimento e produção para se *predizer* a área basal de uma arvoredo particular, que não foi medido aos 3 anos de idade, com base nas medidas tomadas no mesmo arvoredo quando ele tem 7 anos.

Juntamente com a medição, a estimação e a predição constituem as três operações básicas realizadas na Mensuração Florestal, mas a apresentação detalhada dos conceitos destas operações foge ao objetivo do presente capítulos.

No desenvolvimento da Inferência Clássica, o primeiro problema reconhecido foi o problema de estimação. Somente mais tarde é que os estatísticos se volta-

ram especificamente para o problema da predição, uma vez que este problema só pode ser colocado depois que o problema de estimação foi razoavelmente resolvido. Na verdade, a discriminação destes dois conceitos deve ser tomada como um dos capítulos na história do desenvolvimento da inferência estatística. Fisher, em seu trabalho clássico de 1922, enumera o problema de estimação, mas não cita o problema de predição. Somente na segunda metade do século XX é que surgem trabalhos explicitamente voltados para o problema de predição, como por exemplo Allen (1971), Hanisch e Stoyan (1983), e Akaike (1985).

Inicialmente, a aplicação da inferência estatística aos problemas de mensuração também foram dominados pela questão de estimação e de inferência sobre os parâmetros dos modelos. Até mesmo a comparação e seleção de modelos foi vista predominantemente através da questão da inferência sobre as variáveis preditoras, ou os parâmetros a elas relacionados, que deveriam estar presente nos modelos. Na década de 1980, surgiram em que a análise e comparação dos modelos começaram a adquirir um carácter mais voltado à questão de predição, como os trabalhos de Cao *et al.* (1980), Burkhart *et al.* (1981), Bragg e Henary (1985), Burkhart (1987), Hann *et al.* (1987), Lappi e Bailey (1988), Bjornstad (1990) e Lappi (1991).

Atualmente, existe grande clareza de que os problemas práticos da utilização de modelos na Mensuração Florestal são em essência problemas de predição. Mas, a formação teórica em estatística, ainda baseada na Inferência Clássica, exige de muitos biometristas florestais que uma boa dose da sua atenção seja destinada a questões teóricas de pouca relevância para aplicação da Biometria à Mensuração Florestal.

Especificação e Seleção de Modelos

Fisher, em seu clássico trabalho sobre os fundamentos matemáticos da estatística teórica (Fisher, 1922), cita os problemas de especificação e seleção de modelos como formando um dos três tipos fundamentais de problemas aos quais a estatística se dedica. Mas, Fisher nunca dedicou um esforço muito grande a estes problemas. Na maioria dos seus exemplo práticos, ele parte de uma construção racional para especificação do modelo e dedica a maior parte de seu esforço aos problemas de estimação e distribuição amostral. Deve ser levado em conta que, na maioria das aplicações às quais Fisher se dedicou, o modelo especificado era a distribuição normal. Mas, o reconhecimento do problema de especificação de modelo demonstra que ele tinha uma perspectiva bem ampla para as possíveis aplicações da Estatística.

Muito do desenvolvimento da Inferência Clássica está centrada em modelos baseados na distribuição normal, logo, a expressão “seleção de modelos” é numa boa medida entendida como a seleção de modelos dentro de uma das *classes de*

modelos definidas pelo desenvolvimento da teoria da Inferência Clássica. Exemplos de tais classes são: os modelos lineares e os modelos não-lineares, sejam os modelos clássicos de efeitos fixos, sejam os modelos mais recentes de efeitos mistos, todos eles baseados na distribuição normal; os modelos lineares generalizados que, como o nome diz, generalizam a abordagem tradicional dos modelos lineares baseados na distribuição normal; os modelos de séries temporais, também baseados em grande medida na distribuição normal; modelos geoestatísticos, baseados na distribuição normal multivariada; os modelos de sobrevivência, estes sim utilizam outras distribuições de probabilidades que não a distribuição normal.

O desenvolvimento da teoria da Inferência Clássica foi em grande medida definida e restrita a essas classes de modelo. Os modelos lineares clássicos, por exemplo, tiveram um grande desenvolvimento teórico e muito dos elementos teóricos dos modelos não-lineares foi paralelo à abordagem seguida nos modelos lineares. Contudo, alguns procedimentos de inferência bem estabelecidos nos modelos lineares são duvidosos ou questionáveis nos modelos não-lineares, como por exemplo, os testes de hipóteses sobre os coeficientes de regressão e a interpretação do coeficiente de determinação. Outro exemplo é toda a teoria sobre as “*deviance*” dos modelos lineares generalizados que segue de forma análoga à análise de variância nos modelos lineares clássicos. Os modelos lineares clássicos serviram como “balizas” para o desenvolvimento teórico da Inferência Clássica nas outras classes de modelo. Ainda hoje, muito do debate a respeito da seleção de modelos e teste de hipóteses nos modelos lineares de efeitos mistos reflete a perspectiva dos procedimentos aplicados aos modelos lineares clássicos.

A teoria estatística na Inferência Clássica é, contudo, muito pobre no que se refere a comparação e seleção de modelos *entre* as classes de modelo. Talvez os trabalhos de Akaike (1973a, 1973b, 1974, 1981, 1985, 1987) sejam os primeiros trabalhos a abrirem um caminho para uma teoria de comparação de modelos entre classes, embora os primeiros trabalhos ainda estivessem restritos à comparação dentro de cada classe.

Na Biometria Florestal, a visão dos modelos voltados aos problemas de Mensuração Florestal seguiu as classes de modelos desenvolvidos pela teoria da Inferência Clássica, tanto no seu ensino quanto nas suas formas de aplicação. No início da aplicação da inferência este procedimento era natural. Mas, é até certo ponto curioso que a Biometria Florestal que desenvolveu uma “taxonomia” própria de modelos para aplicação a problemas práticos de mensuração (relações hipsométricas, equações de volume, equações de produção, curvas de sítio, modelos de crescimento, etc.), nunca teve a preocupação de desenvolver uma forma de inferência e comparação de modelos adaptada a cada uma das suas próprias classes de modelo.

Parte II

Inferência Baseada em Modelos Segundo o Axioma da Verossimilhança

Capítulo 4

Inferência e Biometria Florestal

Inferência Preditiva Baseada em Modelos

4.1 Inferência para Biometria Florestal

Na primeira parte deste documento, mostrou-se que a utilização da inferência estatística na Mensuração Florestal foi iniciada na década de 1930, quando a Inferência Clássica já tinha sido estabelecida como um corpo razoavelmente estável de conceitos e técnicas. Embora a prática profissional em Mensuração Florestal e Biometria Florestal esteja fortemente assentada nesta forma de inferência, a abordagem dominante de inferência nos trabalhos de pesquisa em Biometria e Mensuração Florestal já não coincide mais exatamente com a abordagem da Inferência Clássica. Por exemplo, verifica-se na última década um clara redução da importância de se realizar testes estatísticos de hipóteses e reportar os respectivos valores-p nos trabalhos de Biometria Florestal publicados nos periódicos científicos internacionais. Por outro lado, percebe-se o uso crescente de critérios de comparação de modelos como o AIC (*Akaike Information Criterion*). A literatura internacional em Biometria Florestal parece sugerir que a Inferência Clássica já não é totalmente apropriada, mas nenhuma forma de inferência alternativa foi claramente proposta.

O objetivo da segunda parte deste documento é apresentar uma forma de inferência estatística para Biometria Florestal que é radicalmente distinta da Inferência Clássica, mas que é mais apropriada para auxiliar a solução dos problemas práticos florestais. Esta forma de inferência é baseada na abordagem da verossimilhança, que foi brilhantemente desenvolvida por Edwards (1992), maravilhosamente aprimorada por Royall (1997) e inspiradoramente exemplificada na área de recursos naturais por Hilborn e Mangel (1997), ainda que estes dois últimos autores não cite os dois primeiros.

Esta forma de inferência, que pode ser brevemente definida como “*inferência baseada em modelos segundo o axioma da verossimilhança*”, será apresentada a partir de seus princípios fundamentais e, portanto, se inicia com o estabelecimento do próprio conceito de fenômeno aleatório ou estocástico. Contudo, a apresentação estará focalizada na Biometria Florestal enquanto disciplina voltada para solução de alguns problemas práticos da Mensuração Florestal. Portanto, é necessário se delimitar com maior precisão o que se entende por “*Biometria Florestal*” neste documento.

4.2 *Biometria Florestal e Mensuração Florestal*

A Biometria Florestal discutida neste documento tem escopo e amplitude bem menor que aqueles utilizados por Husch e por Prodan em seus livros da década de 1960. Neste escopo mais restrito, entendemos Biometria Florestal como a aplicação de métodos de inferência quantitativa para construção de modelos voltados à solução de problemas práticos de Mensuração Florestal. Isto reduz a disciplina a um conjunto de problemas de características bem definidas. Primeiramente, são problemas essencialmente preditivos, isto é, os modelos de interesse na Mensuração Florestal visam prever medidas a respeito de árvores, como altura e volume, ou de arvoredos¹, como área basal e produção madeireira. O caráter preditivo é uma marca dos modelos em mensuração, logo, não trataremos de questões de modelagem voltadas à explicação de fenômenos florestais, como por exemplo os modelos de análise de experimentos.

Uma segunda característica deste conceito de Biometria Florestal é que ela é essencialmente empírica, ou seja, os modelos preditivos são desenvolvidos a partir de dados coletados em campo. Os componentes teóricos dos modelos são de importância secundária, uma vez que o objetivo é produzir modelos que gerem informações confiáveis para situações particulares de árvores ou arvoredos. O caráter empírico deste conceito implica na exclusão da busca por modelos gerais ou universais, apoiados numa racionalidade teórica sobre os fenômenos estudados. Logo, estão excluídos de nossas reflexões os “*modelos de processo*” ou ainda os modelos que procuram combinar a modelagem de processo com a modelagem empírica, segundo a conceituação apresentada por Dixon *et al.* (1990). A pala-

¹Utilizaremos a palavra “arvoredado” para designar genericamente um “conglomerado de árvores”, que é formado ou definido por uma unidade amostral: parcela, ponto de Bitterlich, ou qualquer outro tipo. Consequentemente, neste documento, “arvoredado” tem um sentido específico que difere de “povoamento florestal”, para o qual frequentemente é utilizado como sinônimo. A expressão “povoamento florestal” será utilizada como uma unidade administrativa da floresta, podendo ser um talhão ou quadra, como no manejo de florestas plantadas, ou um compartimento, como no manejo de florestas nativas.

vra “processo” será utilizada com um sentido completamente diferente, que será apresenado mais adiante.

A terceira característica é que ela é fundamentalmente prática, isto é, seu objetivo é disponibilizar métodos e técnicas que permitam a solução de problemas práticos nas situações particulares. O caráter prático confere privilégio às soluções praticamente eficientes sobre aquelas com elaboração teoricamente sofisticadas.

Ou seja, a Biometria Florestal aqui discutida é uma ferramenta que auxilia na buscar de soluções para os problemas de mensuração de árvores e arvoredos e que tem quatro características básicas. Primeiramente, seu objetivo é solucionar problemas práticos de mensuração, por isto, ela é *prática nos seus objetivos*. Em segundo lugar, ela sempre parte de observações e dados sobre as árvores e os arvoredos, logo, ela é *empírica em seu fundamento*. Em terceiro lugar, ela se utiliza de modelos para solucionar os problemas e, por isto, ela é *baseada em modelos na sua forma*. Por fim, a função dos modelos é gerar previsões dos atributos das árvores e arvoredos, o que faz a Biometria Florestal ser *preditiva na sua função*.

4.3 Roteiro de Desenvolvimento

O desenvolvimento que se segue nos próximos capítulos procura organizar os conceitos fundamentais de inferência estatística com o objetivo de gerar uma forma de inferência mais próxima à Biometria Florestal, como foi definida pelas quatro características apresentadas acima. A inferência desenvolvida estará totalmente apoiada na abordagem da verossimilhança, ou seja, ela será baseada *única e exclusivamente* em dois fundamentos estatísticos: a *Lei da Verossimilhança* e o *Princípio da Verossimilhança*, os quais juntos formam o que Edwards (1992) chama de “*Axioma da Verossimilhança*”.

Mas, a forma de pensamento da Inferência Clássica está fortemente impregnada no pensamento quantitativo dos biometristas florestais, uma vez que praticamente toda a sua formação estatística é baseada neste paradigma. Por isto, o desenvolvimento não é iniciado com o “Axioma da Verossimilhança”, mas de algumas definições essenciais para se estabelecer a maneira apropriada de se pensar a Biometria Florestal na perspectiva da abordagem da verossimilhança. Estas definições se referem ao fenômeno da estocasticidade (aleatoriedade), ao processo de construção de modelos (modelagem) e à predição baseada em modelos. Dentre as definições apresentadas, a definição de “*cenário estocástico*” é muito importante, pois coloca a observação de fenômenos estocásticos em bases muito diferentes da concepção de uma “*população hipotética infinita que pode ser amostrada indefinidamente*”.

Segue-se um capítulo sobre a teoria do cálculo de probabilidades, que inicia com a apresentação das definições de probabilidade e seus problemas. Para evitar os problemas que os conceitos específicos de probabilidade podem gerar no desenvolvimento matemático do cálculo de probabilidades, segue-se o desenvolvimento axiomático. O desenvolvimento axiomático é puramente matemático e possui a vantagem de que a concepção particular do fenômeno estocástico, ou mesmo de probabilidade, se torna absolutamente irrelevante para o desenvolvimento. Embora o primeiro desenvolvimento axiomático de probabilidade tenha sido proposto por Kolmogorov (Hacking, 1965; Stigler, 1986), segue-se a apresentação axiomática segundo Mood, Graybill e Boes (1974).

O capítulo seguinte trata das distribuições estocásticas, ou distribuições de probabilidade como são mais frequentemente chamadas. As distribuições estocásticas constituem os modelos básicos para modelagem na inferência estatística e além de se apresentar o seu conceito, são apresentadas algumas distribuições relevantes para Biometria Florestal. São apresentadas também algumas propriedades importantes das distribuições estocásticas e dois resultados básicos de relação entre distribuições. Este capítulo segue em grande medida a apresentação de Mood, Graybill e Boes (1974).

O capítulo sobre a estimação por verossimilhança é o “*cerne conceitual*” do desenvolvimento da inferência apresentado. Primeiramente se discute algumas propriedades desejáveis nos critérios de estimação de modelos e se mostra que a verossimilhança satisfaz todos estes critérios. Apresenta-se, então, a Lei e o Princípio de Verossimilhança e, a partir deles, se desenvolve uma teoria de estimação tanto *pontual* quanto *intervalar*. O desenvolvimento deste capítulo segue Edwards (1992) e Royall (1997).

O último capítulo, que conclui o desenvolvimento da inferência por verossimilhança, trata da especificação e seleção de modelos. Sobre a especificação de modelos, reconhece-se dois componentes: o prático e o teórico. Já sobre a seleção dos modelos, discute-se alguns critérios básicos para seleção e se introduz o conceito de discrepâncias. O critério de informação, que representa a introdução da Teoria Matemática da Informação na Estatística, é apresentado com o detalhamento do Critério de Informação de Akaike como a melhor forma de implementar o critério da informação na seleção de modelos estocásticos. A validação cruzada também é discutida como forma de aprimorar a seleção de modelos.

O capítulo sobre especificação e seleção de modelos conclui o desenvolvimento da inferência segundo a verossimilhança. Todo o desenvolvimento é realizado na forma de *definições* para que os conceitos-chave da inferência sejam claramente apresentados e, para melhorar o entendimento, o desenvolvimento teórico é ilustrado com exemplos práticos que, hora abordam questões diretamente ligadas à análise de dados, hora abordam problemas teóricos associados a certos tipos de dados ou de modelos em particular.

Capítulo 5

Cenário Estocástico e Modelagem

Fundamentos da Inferência Preditiva

O primeiro passo na construção de uma estrutura conceitual para inferência preditiva é definir o contexto em que as operações de inferência e predição se darão. Este contexto tem que ser tal que permita a construção de uma ligação entre os problemas práticos da Biometria Florestal e a estrutura teórica da inferência estatística, que se inicia com o conceito de probabilidade.

Tradicionalmente, este contexto parte da concepção de “*experimento aleatório*”, a qual raramente é desenvolvida e acaba por recorrer a ideia de que as observações são uma amostra de uma população infinita de observações possíveis, sugerindo que o processo de observação pode ser repetido indefinidamente. Esta concepção, embora possa ter sido e ainda ser útil para o desenvolvimento de métodos estatísticos aplicados, está demasiadamente distantes da realidade prática da maioria das áreas de conhecimento das ciências aplicadas e do desenvolvimento tecnológico, particularmente da Biometria Florestal. Na prática, não é possível se tomar infinitas amostras em nenhuma situação real. Esta concepção está mais próxima da “criatividade matemática” do que da “criatividade pragmática”. Não é de se espantar, portanto, que ela tenha sido utilizada de forma muito natural e depois tenha sido popularizada por um homem com grande senso prático, mas também com uma imensa capacidade de pensamento matemático: Ronald Fisher (veja Hacking, 1965).

Partindo-se da concepção do experimento aleatório como gerador de infinitas amostras de uma população infinita, o conceito de “distribuição amostral” surge como um “*conceito natural*”, o que conduz naturalmente à abordagem da Inferência Clássica. O conceito de distribuição amostral era muito caro para Ronald Fisher que o elencou como um dos problemas fundamentais da teoria estatística (Fisher, 1922). Contudo, ele é excessivamente restritivo para uma inferência preditiva, pois o foco da inferência permanece na questão da estimação. É necessário

que a inferência preditiva seja fundamentada numa idéia que seja simultaneamente mais simples e mais ampla, que o conceito tradicional de experimento aleatório.

5.1 *Cenário Estocástico*

Hacking (1965) argumentou que o conceito de experimento aleatório, como gerador de infinitas amostras de uma população infinita, não é necessário ao desenvolvimento lógico-matemático rigoroso da inferência estatística. Ele propôs substituí-lo pelo conceito de “*cenário estocástico*” como o mais simples conceito necessário ao desenvolvimento teórico da inferência estatística. Devido a sua coerência e robustez, os argumentos de Hacking (1965) são contundentes, por isto, o desenvolvimento que se segue parte do conceito de cenário estocástico como o primeiro alicerce para inferência.

Definição 5.1 (Cenário Estocástico). *Cenário estocástico é um dispositivo ou uma parte do mundo no qual pode ser conduzido um ou mais ensaios (experimentos ou observações). Cada ensaio gera um único resultado, dentre um conjunto de resultados possíveis, que tem uma certa chance de ocorrer.*

Nesta definição, a palavra “*ensaio*” é utilizada no seu sentido mais amplo, significando qualquer forma de obter informação sobre um dispositivo natural ou sobre uma parte do mundo. Logo, tanto experimentos quanto observações não experimentais podem ser tomados como ensaios. Nesta obra, a palavra “ensaio” será sempre utilizada nesta acepção mais ampla.

A informação gerada por um ensaio é chamada de “*resultado*”. Os “*resultados*” podem ser experimentais ou observacionais, quantitativos ou qualitativos, sempre dependendo do ensaio em questão. A ideia fundamental é que os “*resultados*” são *dados* pelo cenário estocástico, sendo que a sua qualidade técnica e científica depende unicamente da condução dos ensaios. Portanto, após a condução de um certo número de ensaios no contexto de um dado cenário estocástico, os próprios ensaios conduzidos não podem ser objeto de análise nem de aprimoramento pela inferência estatística, salvo no caso do planejamento de futuros experimentos ou observações.

Outro aspecto importante da definição de cenário estocástico é que os resultados são estocásticos e não determinísticos. Cada resultado possível tem “uma certa chance de ocorrer” e, conseqüentemente, não é possível se determinar o resultado de um ensaio antes que ele seja conduzido. Só é possível se obter as probabilidades para os diferentes resultados possíveis.

Esta definição de cenário estocástico é suficientemente ampla para abranger todos os problemas práticos em que a utilização da inferência estatística se faz

necessária. Alguns exemplos tornarão mais clara a sua generalidade.



Exemplo 5-1: Cenário Estocástico: Exemplos Clássicos

Os exemplos clássicos de cenários estocásticos são os jogos de azar presentes em todos livros de probabilidade, desde o lançamento de moedas e dados, até jogos mais complexos envolvendo baralhos e regras sofisticadas de pareamento de cartas.

Por exemplo, ao se lançar uma moeda, há dois resultados possíveis: cara e coroa. Cada lançamento é um “ensaio” que gerará um destes dois resultados. Se a moeda for “justa”, espera-se que a chance de cada resultado será de 50%. O “dispositivo” que envolve todos os aspectos do lançamento da moeda é o “cenário estocástico”.



Exemplo 5-2: Cenário Estocástico: Exemplos Florestais

No caso de exemplos florestais, o interesse maior desta obra está na Mensuração Florestal. Qualquer operação de medição de atributo de árvores ou de arvoredos pode ser vista como um cenário estocástico.

Por exemplo, a medição do DAP das árvores na floresta é um cenário estocástico, pois os valores da medida serão números reais que estarão dentro de uma amplitude de valores que depende da estrutura de tamanho da floresta. A medição de uma árvore é um “ensaio”, pois gera apenas um valor dentre os inúmeros valores possíveis. A estrutura de tamanho das árvores da floresta, juntamente com o procedimento de medição do DAP, compreende o “dispositivo” que é o cenário estocástico. Cada árvore medida é a “repetição” do “ensaio” no cenário estocástico.

De forma análoga, a medição de parcelas de inventário florestal e os cálculos necessários para se obter o volume de madeira também pode ser visto como um cenário estocástico. Neste caso, o “dispositivo” é a estrutura de produção da floresta, juntamente com (i) os procedimentos de locação das parcelas, que inclui o tamanho e forma de parcela e a forma de localização espacial da parcela (delineamento amostral), (ii) os procedimentos de medição das árvores individuais, e (iii)

os procedimentos de cálculo, que envolvem os modelos de predição necessários para se obter o volume, isto é, relação hipsométrica e equação de volume. Cada parcela é um “ensaio” e o volume de madeira obtido é o “resultado”.



Processo Estocástico

No conceito de experimento aleatório da inferência clássica, a incorporação das escalas espacial e temporal ao modelo requer uma redefinição do que seria a população finita, na qual são realizadas infinitas amostras. Esta população seria fixa no espaço e no tempo? Se fosse variável, como esta população varia no espaço e no tempo? Como deduzir a distribuição amostral de uma estatística se a população amostrada é variável?

Neste aspecto, o conceito de cenário estocástico tem uma grande vantagem sobre o conceito de experimento aleatório. Ele permite a incorporação explícita das escalas espacial e temporal ao conceito básico a partir do qual se desenvolve a inferência, sem que haja um aumento exacerbado de complexidade.

Definição 5.2 (Processo Estocástico). *Processo estocástico é um cenário estocástico cujos ensaios são conduzidos segundo uma escala explícita de tempo e/ou de espaço.*

Os processos estocásticos são, portanto, cenários estocásticos em que a condução dos ensaios é realizada de modo a considerar explicitamente a escala espacial, temporal ou ambas. O cenário estocástico em que a localização espacial dos dados é considerada explicitamente e se avalia as relações espaciais entre eles é um *processo estocástico espacial*. Já nos casos em que os dados são observados explicitamente ao longo do tempo e a informação temporal é parte explícita do cenário estocástico observado, a designação apropriada é *processo estocástico temporal*. Se ambas escalas são parte do cenário estudado, então o processo deve ser designado por *processo estocástico espaço-temporal*.



Exemplo 5-3: Processo Estocástico: Exemplos Florestais

Os modelos de Biometria Florestal e os procedimentos de Mensuração Florestal são com frequência processos estocásticos quando a informação de tempo e espaço é considerada explicitamente.

Medidas repetidas do DAP ou da altura no tempo são tomadas para o estudo do crescimento das árvores individuais. Neste caso, cada árvore é um “ensaio”, sendo o valor de DAP ou altura em cada medição um “resultado”. Contudo, tais resultados podem ser organizados numa escala explícita de tempo, sendo que para cada medida de DAP ou altura se terá um valor de tempo, mais especificamente de idade da árvore. O crescimento em DAP ou altura das árvores individuais de uma floresta pode ser visto, portanto, como um processo estocástico temporal.

No caso de um levantamento ou inventário florestal, se obtém dados de um conjunto de parcelas. Como apresentado acima, a produção de madeira obtida em cada parcela pode ser vista como o “resultado” de um “ensaio” de um cenário estocástico. Contudo, se a localização da parcela é obtida na forma de coordenadas num plano e *se esta informação de localização* for utilizada na análise dos dados do levantamento, como por exemplo numa análise geo-estatística, então as parcelas de inventário devem ser vista com “ensaios” de um processo estocástico espacial, uma vez que a escala espacial será explicitamente utilizada na inferência.

Num inventário florestal contínuo, em que se realiza a observação em parcelas permanentes georreferenciadas, os dados gerados estarão explícitos tanto na escala temporal quanto na espacial. Logo, as parcelas podem ser vistas como ensaios de um processo estocástico espaço-temporal.



5.2 Modelos e Modelagem

Fisher, no seu trabalho “*On the Mathematical Foundations of Theoretical Statistics*” (1922), definiu o propósito da estatística como sendo o de “reduzir” os dados a um pequeno número de quantidades que representam o todo, pois contêm tanta informação relevante quanto possível. Em seguida, identificou os três tipos básicos de problemas na estatística: (i) os problemas de especificação, isto é, escolha do modelo matemático; (ii) os problemas de estimação ou os métodos para o cálculo dos valores para os parâmetros do modelo; e (iii) problemas de distribuição que envolvem a discussão da distribuição das estimativas dos parâmetros.

Um usuário experiente da estatística clássica perguntaria sobre a inferência estatística por meio dos testes de hipóteses. Certamente, este aspecto da inferência estatística foi incorporado posteriormente por Fisher nesta concepção inicial, a qual teve um grande desenvolvimento nas décadas de 1920 e 1930.

A concepção inicial de 1922 coloca os problemas da estatística com grande clareza e concisão, mas indica um aspecto da inferência clássica que provavelmente é derivado em grande parte da concepção de experimento aleatório. Trata-se dos problemas das distribuições que envolvem a estimativas dos parâmetros ou, em poucas palavras, dos problemas das distribuições amostrais.

Na inferência estatística baseada em verossimilhança, os modelos estatísticos devem ser considerados explicitamente, por isto, ela pode ser apropriadamente chamada de “inferência baseada em modelos” ou, em inglês, “*model based inference*”. Como será demonstrado nesta obra, a questão das distribuições amostrais perdem muito da sua importância na inferência baseada em verossimilhança, particularmente no caso da inferência preditiva.

Contudo, os dois primeiros problemas apontados por Fisher — o problema da especificação do modelo e o problema da estimação dos parâmetros do modelo — permanecem sendo problemas fundamentais, pois eles se referem diretamente aos modelos. Logo, tais problemas devem ser entendidos como parte de um problema mais amplo, que é o problema da “*construção de modelos*” ou o problema da “*modelagem estocástica*”. Deve-se, no entanto, acrescentar a estes dois problemas o da “*especificação dos dados*”, isto é, o problema da obtenção das observações para modelagem. A modelagem estocástica pode, portanto, ser definida a partir destes três problemas fundamentais:

Definição 5.3 (Modelagem Estocástica). *Modelagem estocástica é o procedimento de construção de um modelo estocástico que represente adequadamente um determinado cenário ou processo estocástico. Ela envolve três tipos de problemas:*

Especificação de Dados: *problemas relacionadas a como conduzir os ensaios do cenário ou processo estocástico que geram os resultados, isto é, a como realizar as observações que geram os dados para a construção do modelo.*

Especificação de Modelo: *problemas relacionadas à determinação do melhor modelo estocástico para representar adequadamente o cenário ou processo estocástico em estudo;*

Estimação: *problemas relacionadas à determinação do valor dos parâmetros do modelo estocástico no contexto do cenário ou processo estocástico em estudo.*

Esta definição mostra que o propósito da modelagem é fornecer um modelo que represente adequadamente o cenário ou processo estocástico de interesse. Na Biometria Florestal, busca-se uma inferência preditiva, portanto a *representação adequada* significa a produção de predições com a qualidade desejada. Logo, os problemas de predição e de modelagem não são independentes. O procedimento

“ótimo” de modelagem depende do contexto de predição e do que se espera obter com a predição.

Da mesma forma, os três tipos de problemas na modelagem, não são independentes entre si. Há um aspecto de *condicionalidade* da especificação do modelo e da estimação em relação aos dados, isto é, a especificação do modelo mais apropriado a um cenário estocástico e a estimação dos seus parâmetros *estará sempre condicionada* aos dados disponíveis. Assim, a especificação de modelo e a estimação serão sempre realizadas no contexto definido pelos dados, ou seja, uma vez obtido os dados busca-se o melhor modelo e a melhor estimação.

Por outro lado, ao se *planejar* um estudo empírico, seja um estudo *observacional* (levantamento, inventário), seja um estudo *experimental*, o procedimento apropriado para a coleta dos dados será definido no contexto dos modelos estocásticos potenciais para o problema prático em questão. Os procedimentos de especificação de modelo e de estimação de parâmetros podem ser estudados de modo a indicar qual o tipo de dado mais apropriado para se chegar a uma conclusão sólida a respeito de modelos e estimativas. Tal tipo ideal de conjunto de dados indicará qual abordagem ou qual método deve ser seguido para obtê-lo, tanto para problemas que envolvem delineamentos amostrais (estudos observacionais) quanto para aquele que envolvem delineamentos experimentais (estudos experimentais).

Dos três tipos de problemas da inferência, o problema de especificação dos dados é o mais dependentes do contexto do problema prático em questão. Ele envolve não só os aspectos estatísticos, mas também os aspectos operacionais e de custo, que podem superar as questões de inferência e predição em termos de relevância prática. Dada a particularidade e a característica quase idiossincrática do problema de especificação de dados, é muito difícil traçar linhas gerais para resolução deste problema. Assim, o desenvolvimento que se segue focaliza apenas nos problemas de especificação de modelos e de estimação.

Modelo Estocástico

O desenvolvimento de uma forma de inferência baseada em modelos, como a inferência por verossimilhança, demanda a apresentação precisa do que é o modelo estocástico. Uma definição detalhada do modelo estocástico requer conceitos matemáticos que serão apresentados mais adiante neste desenvolvimento de inferência preditiva. Mas, para maior clareza, uma definição geral e preliminar se faz necessária.

Definição 5.4 (Modelo Estocástico). *Modelo estocástico é um dispositivo matemática, cuja estrutura combina elementos estocásticos e elementos determinísticos, que simula resultados possíveis dos ensaios de um cenário ou processo estocástico.*

Um aspecto importante nesta definição é que o modelo estocástico *simula resultados possíveis* de um cenário ou processo estocástico. Se o objetivo do modelo for “*explicar*” um fenômeno natural, espera-se que os resultados possíveis sejam gerados por um modelo seja uma descrição ou explicação apropriada do fenômeno natural subjacente ao cenário ou processo estocástico. Este seria o objetivo de uma “*inferência explicativa*”.

No contexto da Biometria Florestal aplicada à Mensuração Florestal, os “*resultados*” devem ser entendidos como as observações obtidas num cenário estocástico, que podem ser tanto medidas ou atributos observáveis, como altura de árvore ou biomassa de arvoredos, quanto atributos abstratos, como a diversidade de espécies de arvoredos. Como se trata de uma inferência preditiva, não é necessário que o modelo descreva ou explique o fenômeno natural subjacente ao cenário ou processo estocástico.

Outro aspecto importante a ressaltar na definição, é que o modelo estocástico é um dispositivo matemático que combina elementos estocásticos e elementos determinísticos. Assim como os resultados dos ensaios de um dado cenário ou processo estocástico são incertos, os resultados simulados pelo modelo estocástico também serão incertos, uma vez que o seu objetivo é simular resultados possíveis. Não se trata, no entanto, de estabelecer desde o início que, num dado modelo, o elemento determinístico é o aspecto explicado pelo modelo, enquanto que o elemento estocástico representa a incerteza ou o que não é explicado pelo modelo. O elemento estocástico também é parte do modelo e, portanto, também é “*explicado pelo modelo*”. A interpretação do elemento determinístico e do elemento estocásticos do modelo deve ser definida a partir do contexto da inferência que se realiza. No contexto da inferência preditiva, o elemento estocástico é frequentemente tomado como sendo o aspecto incontrolável da predição e, portanto, como indicador da incerteza da predição.

Modelo Estocástico Empírico

Como foi apresentado acima, o presente desenvolvimento de inferência quantitativa adota uma abordagem essencialmente empírica. Para enfatizar o caráter empírico dos modelos estocásticos tratados, apresenta-se uma definição específica de modelo empírico.

Definição 5.5 (Modelo Estocástico Empírico). *Modelo estocástico empírico é um modelo matemático, cuja estrutura pode ser arbitrária ou biologicamente fundamentada, mas é construído com base em resultados previamente obtidos do cenário ou processo estocástico sendo modelado.*

O comportamento do modelo estocástico empírico, isto é, as predições por ele geradas, é fortemente condicionado pelos resultados utilizados na sua cons-

trução.

Esta definição mostra que a estrutura teórica dos modelos empíricos, que é sempre matemática, pode ser biologicamente fundamentada ou pode ser arbitrária, no sentido de não ser uma explicação, mas uma aproximação matemática conveniente do fenômeno natural subjacente ao cenário ou processo estocástico. Seja qual for o caso, as observações previamente obtidas do fenômeno são o aspecto *mais importante* na construção do modelo empírico e são determinantes no seu comportamento.

5.3 Predição

O propósito da inferência estatística na Biometria Florestal aplicada à Mensuração Florestal é auxiliar na determinação de medidas e atributos de árvores e arvoredos que não podem ser medidos, isto é, não podem ser determinados por procedimentos de medição diretos ou indiretos. A impossibilidade da determinação por medição pode ser de ordem prática, como no caso de medidas e atributos que só podem ser medidos de forma destrutiva ou de ordem pragmática como nos procedimentos de medição excessivamente onerosos. Mas a impossibilidade pode ser também de ordem operacional, isto é, o atributo é na verdade um atributo abstrato que não existe no momento, como a altura de uma árvore ou a produção de um arvoredo num dado momento no futuro ou no passado.

O propósito é, portanto, obter a *predição* de medidas e atributos de árvores e de arvoredos. Mas, uma definição geral de predição se faz necessária.

Definição 5.6 (Predição). *Predição é o procedimento de determinação de certos resultados possíveis dos ensaios de um dado cenário ou processo estocástico com base num modelo estocástico, podendo utilizar outros resultados do mesmo cenário ou processo estocástico, que não foram utilizados na construção do modelo.*

Retoma-se o conceito de cenário estocástico pois ele estabelece o contexto em que as observações empíricas são obtidas, isto é, os dados são coletados. O cenário estocástico também estabelece o contexto em que a discussão do modelo estocástico mais apropriado é realizada. Por estabelecer um contexto, o cenário estocástico é de especificação totalmente qualitativa, sendo um requisito para a realização da predição. No caso da Biometria Florestal, os cenários estocásticos são definidos pela especificação dos problemas de Mensuração Florestal.

A definição também mostra que a predição pode ser realizada exclusivamente com base no modelo estocástico construído, ou pode utilizar também novas observações do fenômeno estudado, que não estavam disponíveis quando o modelo

foi construído. As novas observações podem ser utilizadas tanto para “calibrar” o modelo para situação particular de predição, quanto como elemento a ser combinado com os resultados gerados pelo modelo para se gerar a predição.

Interpolação e Extrapolação

Ao se tratar de predição, é natural que surja a questão da interpolação e da extrapolação na aplicação dos modelos estocásticos, pois tais termos possuem diferentes interpretações em função da área de conhecimento. Nas Engenharias, interpolação e extrapolação tem um significado puramente matemático relacionado às operações realizadas com as funções matemáticas. Já na Estatística, o uso destes termos implica numa interpretação relativa à origem dos dados utilizados no modelo.

A definição apresentada se assemelha à interpretação estatística, pois relaciona interpolação e extrapolação com os cenários ou processos estocásticos de construção do modelo e de aplicação do modelo.

Definição 5.7 (Predição Interpolativa). *Predição Interpolativa é aquela realizada quando um modelo estocástico é aplicado ao mesmo cenário ou processo estocástico a partir do qual ele foi construído, isto é, os resultados utilizados na construção do modelo são provenientes de ensaios do mesmo cenário ou processo estocástico em que o modelo é aplicado.*

Definição 5.8 (Predição Extrapolativa). *Predição Extrapolativa é aquela realizada quando um modelo estocástico é aplicado a um cenário ou processo estocástico análogo ao cenário ou processo estocástico a partir do qual ele foi construído, isto é, os resultados utilizados na construção do modelo não são provenientes do mesmo cenário ou processo estocástico em que o modelo é aplicado.*



Exemplo 5-4: Exemplos Predição Interpolativa

Os exemplos mais tradicionais de predição interpolativa na medição de árvores são as relações hipsométricas utilizadas para prever a altura das árvores de uma parcela, quando apenas um subconjunto delas teve a altura medida. Também o uso de equações de volume, equações de biomassa e equações de forma (equações de afilamento) são exemplo de predição interpolativa quando a sua construção é realizada a partir de uma amostra destrutiva de árvores da floresta em que a equação será utilizada.

Também a aplicação de modelos de crescimento e produção podem ser considerada predição interpolativa quando os dados utilizados para construção dos mode-

los podem ser considerados uma amostra representativa das situações referentes à mesma floresta sendo modelada e manejada. Estas situações envolvem principalmente aspectos como espécie ou material genético (clone), tratamentos silviculturais prescrições de manejo, sítios e solos, condições climáticas e classes de idades.



Exemplo 5-5: Exemplos Predição Extrapolativa

O exemplo convencional de predição extrapolativa é a utilização de fatores ou modelos biométricos publicados na literatura técnica para aplicação em situações análogas de espécie florestal, região e classe de idade. A utilização de fatores de forma, fatores de empilhamento, equações de volume e equações de biomassa publicadas são provavelmente os exemplos mais frequentes de predição extrapolativa.



Como os modelos estocástico aqui tratados são empíricos, é importante se ter em mente que a situação ideal de predição é a situação de *predição interpolativa*, pois, neste caso, pode-se contar com toda a robustez da fundamentação da Teoria da Amostragem nas predições realizadas. Já no caso da *predição extrapolativa*, a qualidade das predições dependerá majoritariamente da qualidade do *raciocínio analógico* que estabelece as semelhanças e as diferenças entre o cenário ou processo estocástico de construção do modelo e o cenário ou processo de aplicação do mesmo. Qualquer ponderação de ordem geral é impossível nesta situação e, para o bem da qualidade das predições em Biometria Florestal, é cauteloso que a predição extrapolativa seja evitada.

* * *

O desenvolvimento que se segue focaliza apenas nos problemas de especificação de modelos e de estimação. Em benefício da completude da exposição, serão apresentados dois capítulos que tratam da teoria de probabilidades, iniciando no próprio conceito de probabilidade e indo até o conceito de distribuição estocástica.

Os conceitos de teoria de probabilidade são fundamentais para uma exposição clara da abordagem da inferência preditiva por meio da verossimilhança.

Capítulo 6

Cálculo de Probabilidades

Descrição Matemática da Estocasticidade

Para realizar um desenvolvimento completo da inferência estatística por verossimilhança, iniciamos a exposição com a apresentação de alguns resultados fundamentais da teoria de probabilidades. Primeiramente, discutimos os três conceitos de probabilidade atualmente utilizados na pesquisa científica e apresentamos o tratamento axiomática de probabilidade, que é como a Matemática atual lida com a probabilidade, pois ele independente de qual conceito prático é aceito. Em seguida, desenvolvemos as propriedades básicas e o conceito de probabilidade até o Teorema de Bayes.

Para a elaboração deste capítulo, baseamos-nos principalmente no texto de Mood, Graybill e Boes (1974), mas também em outros textos tradicionais de probabilidade e de estatística matemática: Cramer (1955), Hoel, Port e Stone (1971), Ross (1989) e Arnold (1990). Como o desenvolvimento deste tema é essencialmente lógico-dedutivo, não apresentaremos citações a estas obras, exceto quando elas refletirem uma interpretação particular do autor citado.

6.1 *Conceito de Probabilidade*

Conceito A Priori

Considera-se como o início da Teoria de Probabilidades as correspondências entre Blaise Pascal (1623-1662), matemático, físico e escritor francês, e Pierre Fermat (1601-1665), advogado e matemático amador francês, a respeito de cálculos matemáticos envolvendo jogos de azar. Assim, o primeiro conceito de probabilidade foi desenvolvido partir dos jogos de azar nos quais as possibilidades do resultado

de um jogo podem ser deduzidas antes da realização do jogo.

Um jogo de azar pode ser considerado um “*ensaio de um cenário estocástico*”, isto é, um ensaio cujo o resultado final é incerto. Se o ensaio for realizado várias vezes, o resultado nunca será o mesmo em todas as vezes, mas cada um dos resultados é igualmente provável. Num jogo de azar, os resultados igualmente prováveis podem ser antecipadamente enumerados, de modo que o conceito matemático de probabilidade *a priori* foi definido pelo quociente entre o número de resultados favoráveis e o número total de resultados possíveis. Considerando um evento A num ensaio, a probabilidade deste evento, portanto, será

$$P(A) = \frac{\text{Número de resultados favoráveis a } A}{\text{Número total de resultados possíveis do ensaio}}$$



Exemplo 6-1: Lançamento de um Dado

No lançamento de um dado justo qual a probabilidade do resultado ser um número primo? Um dado justo tem seis resultados igualmente prováveis, destes resultados, quatro são números primos: 1, 2, 3 e 5. Logo a probabilidade é $4/6 = 2/3$.



Exemplo 6-2: Sorteio de Baralho

Num sorteio completamente aleatório de uma carta de um baralho, qual a probabilidade da carta ser uma dama de copas? Os baralhos tem 52 cartas, sendo que apenas uma delas é dama de copas. Logo a probabilidade é $1/52$.



Conceito A Posteriori

Uma séria limitação do conceito de probabilidade *a priori* é que ele só é aplicável a um número muito pequeno de fenômenos, que são os jogos de azar. No mundo natural, existe uma infinidade de fenômenos, mas não é possível enumerar antecipadamente todos os resultados que estes fenômenos podem gerar. A possibilidade

de aplicar o cálculo de probabilidades a estes fenômenos começou com Jacob Bernoulli no século XVIII que utilizou o conceito de probabilidade *a posteriori*.

No conceito *a posteriori*, o fenômeno natural é observado um grande número de vezes. Assume-se que o fenômeno natural se comporta de modo uniforme ao longo das observações, de modo que se possa aplicar a ele o mesmo conceito de ensaio de um cenário estocástico, sendo que cada resultado observado é igualmente provável. Após as observações de vários resultados do ensaio, a probabilidade de um evento será a “*frequência relativa*” deste evento entre todos os resultados observados. Assim, a probabilidade *a posteriori* do evento A é calculada como sendo:

$$P(A) = \frac{\text{Número de resultados em que o evento } A \text{ foi observado}}{\text{Número total de resultados observados}}$$

Este conceito de probabilidade existia desde antes de Jacob Bernoulli. Mas foi Bernoulli quem primeiro demonstrou matematicamente que com um número de observações arbitrariamente grande, era possível tornar a frequência relativa muito próxima da probabilidade verdadeira do fenômeno, conferindo certo rigor matemático ao conceito *a posteriori* (Stigler, 1986).

Conceito Subjetivo

A probabilidade *a priori* pode ser obtida pelo exame racional do fenômeno estudado, utilizando-se deduções lógicas, de modo que ele também pode ser chamado de conceito *teórico* de probabilidade. Já a probabilidade *a posteriori* só é possível com base na observação do fenômeno e, portanto, ela pode ser chamada de conceito *empírico* de probabilidade. Ambos conceitos são objetivos no sentido que independem do julgamento pessoal do pesquisador que estuda o fenômeno. A qualidade da probabilidade *a priori* depende do rigor lógico-dedutivo com que o fenômeno é estudado. Já a qualidade da probabilidade *a posteriori* está assentada no rigor metodológico na condução do experimento ou na observação do fenômeno.

Mas a probabilidade pode ser vista como uma “*percepção*” da incerteza de um fenômeno expressa de forma pessoal, isto é, subjetiva. A probabilidade seria uma “*medida*” da incerteza ou do desconhecimento do pesquisador sobre o fenômeno sendo estudado. Alguns chegam a afirmar que a probabilidade seria uma “*medida do grau da crença*” do pesquisador num resultado observado ou concluído. Assim, quando um pesquisador afirma estar 70% certo de que suas conclusões estão válidas, ele estaria expressando esta “*crença*”. O mesmo acontece quando alguém olha o céu e afirma que há uma probabilidade de 50% de chover no fim da tarde.

O conceito subjetivo de probabilidade tem duas limitações sérias. Em primeiro lugar, não é possível avaliar a qualidade com que uma dada probabilidade é determinada, pois ela surge de um julgamento subjetivo. Em segundo lugar, não parece

haver comprovação empírica de que incertezas ou crenças subjetivas expressas na forma de porcentagem tenha validade além da pura validade qualitativa e possam ser tratadas como quantidades numéricas. A validade matemática do conceito subjetivo de probabilidade é provavelmente o fundamento mais problemático da Inferência Estatística Bayesiana.

A Articulação Problemática dos Conceitos de Probabilidade

Conforme o que foi apresentado, é possível perceber que a articulação dos três conceitos de probabilidade é problemática. O conceito subjetivo é com certeza de articulação mais discutível e problemática. Se a sua articulação com o conceito teórico de probabilidade é aparentemente impossível, a sua relação com o conceito empírico também é altamente duvidosa. Será que qualquer julgamento subjetivo de probabilidade é empiricamente coerente, isto é, pode ser demonstrado coerente pelas observações? Certamente, algumas pessoas que atuam em áreas específicas do conhecimento podem se mostrar com um alto grau de coerência empírica na sua área de conhecimento. Mas, esta coerência pode ser uniformemente assumida para qualquer tipo de julgamento e independentemente da pessoa que realiza o juízo? Esta é com certeza uma premissa muito pouco razoável, e argumentos teóricos sobre o conhecimento humano e a formação do juízo na mente humana são de muito pouca validade quando a questão é essencialmente de ordem prática e empírica.

Mas a articulação entre os conceitos teórico e empírico também é problemática. Jacob Bernoulli demonstrou no século XVIII que à medida que o número de ensaios observados tende ao infinito, a frequência relativa, que é a medida empírica de probabilidade, tende a um valor constante que pode se assumir como medida teórica de probabilidade do fenômeno observado. Mas o conceito de um “*número infinito*” de ensaios, se é teoricamente compreensível é empiricamente impossível. O que leva diretamente à questão prática de quantos ensaios são necessários observar para se ter uma incerteza desprezível sobre a probabilidade calculada pela frequência relativa. Em essência, este é um dos problemas centrais da modelagem matemática de fenômenos estocásticos que só podem ser resolvidos de modo empírico para situações particulares. Assim, a articulação entre os conceitos teórico e empírico de probabilidade, que está na raiz da modelagem matemática dos fenômenos naturais, já traz em seu bojo uma tensão teoricamente insolúvel que é central na modelagem estocástica (teórica) das observações científicas.

6.2 *Desenvolvimento Axiomático*

A definição axiomática de probabilidade é a definição matemática que, sendo completamente abstrata, não considera nenhum dos conceitos de probabilidade discutidos acima e, conseqüentemente, não trata de qualquer aspecto da aplicação prática da probabilidade e do seu cálculo. Mas, ela fundamenta todo o desenvolvimento matemático da Teoria de Probabilidades

Tradicionalmente, a definição de probabilidade se inicia com a expressão “*experimento aleatório*”, embora raramente os livros de matemático apresentem uma definição desta expressão. Mas, o ponto de partida do desenvolvimento aqui apresentado é do conceito de *cenário estocástico*, como proposto por Hacking (1965) e apresentado no capítulo anterior. Num desenvolvimento axiomático de probabilidade, os conceitos se tornam progressivamente mais matemáticos, de forma que para manter a coerência matemática será tomado como referência o desenvolvimento apresentado por Mood, Graybill e Boes (1974).

Espaço de Probabilidade

Todo o desenvolvimento axiomático se fundamenta na teoria matemática de conjuntos. Assume-se um conhecimento básico de teoria de conjuntos, de modo que não se apresenta as definições e as operações básicas desenvolvidas nesta teoria.

A primeira definição é a de “*espaço amostral*”:

Definição 6.1 (Espaço Amostral). *Espaço amostral* (Ω) é o conjunto cujos elementos tem uma relação de um-para-um com os resultados possíveis dos ensaios de um cenário estocástico.

O espaço amostral pode ser um conjunto numérico ou um conjunto de “*códigos*” que representem os resultados do cenário estocástico. Ele também pode ter tamanho finito ou infinito. No caso de espaços amostrais infinitos, disse-se que ele tem tamanho “*infinito contável*” se for possível fazer uma associação de um-para-um de cada elemento dele com os elementos do conjunto dos números naturais. Se isto não for possível, disse-se que o espaço amostral tem tamanho “*infinito incomensurável*”.



Exemplo 6-3: Sexo de Árvores do Pinheiro-do-Paraná

O pinheiro-do-paraná (*Araucaria angustifolia*, *Araucariaceae*) é uma espécie arbórea conífera (gimnospermas) que é dióica, isto é, cada planta tem um único sexo. Considerando o cenário estocástico como sendo a determinação do sexo de uma árvore selecionada aleatoriamente numa população de pinheiro-do-paraná, o

espaço amostral é finito com a seguinte forma:

$$\Omega = \{M, F\},$$

sendo que M indica “sexo masculino” e F indica “sexo feminino”.

Por outro lado, se o ensaio do cenário estocástico consistir na seleção aleatória de duas árvores, o espaço amostral é

$$\Omega = \{MM, MF, FM, FF\}.$$



Exemplo 6-4: Ninhada de Cachorro-do-Mato

O cachorro-do-mato (*Cerdocyon thous*, *Canidae*) é um canídeo, também conhecido como raposa-da-floresta, que vive na América do Sul da Colômbia e sul da Venezuela até o Paraguai, Uruguai e norte da Argentina. Ele é um animal de hábito noturno e vive em casais monogâmicos, cuja fêmea dá à luz a uma ou duas ninhadas por ano, sendo que cada ninhada tem de três a seis filhotes. Assumindo como cenário estocástico o número de filhotes numa ninhada de cachorro-do-mato, o espaço amostral associado a este cenário é finito com a forma:

$$\Omega = \{3, 4, 5, 6\}.$$

Por outro lado, se o cenário estocástico em estudo for o número de filhotes que o casal de cachorros-do-mato pode ter em um ano, é necessário considerar que o casal pode ter uma ou duas ninhadas por ano. Neste caso, o espaço amostral se torna

$$\Omega = \{3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$



Exemplo 6-5: Parcela de Regeneração Natural

O número de plântulas numa parcela de regeneração natural pode ser considerado um ensaio de um cenário estocástico com espaço amostral infinito contável:

$$\Omega = \{0, 1, 2, \dots\} \Leftrightarrow \Omega = \mathbb{N},$$

isto é, o espaço amostral é o conjunto dos números naturais. Note que não se espera que seja possível efetivamente contar um número infinito de plântulas numa parcela, mas a impossibilidade de se determinar a priori um limite superior para o número de plântulas na parcela implica em assumir que o espaço amostral é infinito.



Exemplo 6-6: Diâmetro de uma Árvore

O resultado da medida do diâmetro de uma árvore selecionada aleatoriamente numa florestal pode ser considerado um resultado de um ensaio de um cenário estocástico com espaço amostral infinito incomensurável:

$$\Omega = \mathbb{R}^+ = \{x \in \mathbb{R} : x > 0\},$$

ou seja, o espaço amostral é o conjunto dos números reais positivos.

Também nesta situação, não se espera observar uma árvore com diâmetro infinito, nem que o diâmetro possa ser determinado com precisão infinita, que é uma implicação do número ser real. Contudo, a impossibilidade de se determinar a priori o diâmetro máximo possível de ser observado, bem como a exatidão infinita na medição do diâmetro, resulta que o melhor “espaço amostral” para se lidar com este fenômeno é o dos números reais positivos.



O cálculo de probabilidades se faz com base na ocorrência de “eventos”, sendo, portanto, necessário definir este conceito.

Definição 6.2 (Evento). *Evento é um subconjunto do espaço amostral.*



Exemplo 6-7: Ninhada de Cachorro-do-Mato II

Considerando o exemplo do cenário estocástico do número de filhotes que um casal de cachorros-do-mato pode ter em um ano, é possível se observar os seguintes eventos:

- o resultado é um número ímpar: $A = \{3, 5, 7, 9, 11\}$,
- o resultado é um número par: $B = \{4, 6, 8, 10, 12\}$,
- o resultado é um número primo: $C = \{3, 5, 7, 11\}$.



Exemplo 6-8: Parcela de Regeneração Natural II

Na observação do número de plântulas numa parcela, eventos possíveis são

- o número de plântula é menor que 5: $A = \{0, 1, 2, 3, 4\}$
- não foram observadas plântulas na parcela: $B = \{0\}$



Fazendo uma ligação entre a linguagem prática do uso de probabilidade e a linguagem matemática, disse-se que “o evento A ocorreu” quando o resultado do ensaio num cenário estocástico for um elemento de A . A aplicação da teoria de conjuntos gera algumas implicações sobre os eventos de um espaço amostral, que também acarretam uma definição matemática de expressões ligadas ao uso de probabilidade.

- O evento pode ser constituído de um único elemento do espaço amostral, sendo chamado de *evento elementar*.
- O conjunto vazio (\emptyset) também é um evento, pois $\emptyset \subset \Omega$.
- O espaço amostral é também um evento, pois $\Omega \subset \Omega$. Ele é chamado de *evento certo*.

- Associado ao evento A existe o evento *complementar* A^c , que consiste de todos elementos em Ω que não estão em A : $A^c = \{x \in \Omega : x \notin A\}$.
- Uma série de eventos A_1, A_2, \dots é dita uma série de eventos *disjuntos* ou eventos *mutuamente exclusivos* se $A_i \cap A_j = \emptyset; i \neq j; i = j = 1, 2, \dots$
- Seja $\{A_k\}$ uma coleção de eventos de Ω indexados por K , então

$$(i) \left[\bigcup_{k \in K} A_k \right]^c = \bigcap_{k \in K} A_k^c$$

$$(ii) \left[\bigcap_{k \in K} A_k \right]^c = \bigcup_{k \in K} A_k^c$$

O cálculo de probabilidades não fica restrito somente a eventos mas também a combinações de eventos, sendo, portanto, necessária mais uma definição.

Definição 6.3 (Espaço de Eventos). *Espaço de eventos \mathcal{A} é a classe de todos eventos associados a um cenário estocástico.*



Exemplo 6-9: Ninhada de Cachorro-do-Mato III

O espaço amostral do cenário estocástico do número de filhotes que um casal de cachorros-do-mato produz num ano é $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. Cada elemento do espaço amostral configura um evento elementar. Considerando todas as combinações possíveis entre estes eventos, verifica-se que o espaço de eventos \mathcal{A} tem $2^{10} = 1024$ eventos, que incluem os 10 eventos elementares, o conjunto vazio (\emptyset) e o evento certo (Ω).



Exemplo 6-10: Sexo de Árvores do Pinheiro-do-Paraná II

Voltando ao exemplo do cenário estocástico do sexo de duas árvores de pinheiro-do-paraná, nota-se que espaço amostral é formado por quatro eventos elementares: $\Omega = \{MM, MF, FM, FF\}$. Assim, o espaço de eventos \mathcal{A} deste cenário estocástico é composto de $2^4 = 16$ eventos.



Mas é necessário definir algumas propriedades do espaço de eventos para que possamos calcular probabilidades de combinações de eventos em espaços amostrais infinitos.

Definição 6.4 (Sigma Algebra). *O espaço de eventos \mathcal{A} é chamado de uma sigma álgebra se possuir as seguintes propriedades:*

(i) $\Omega \in \mathcal{A}$.

(ii) Se $A \in \mathcal{A}$, então $A^c \in \mathcal{A}$.

(iii) Se A_1, A_2, \dots é uma série de eventos então

$$\left[\bigcup_{i=1}^{\infty} A_i \right] \in \mathcal{A} \quad e \quad \left[\bigcap_{i=1}^{\infty} A_i \right] \in \mathcal{A}$$

Com esta última definição podemos então apresentar a definição de probabilidade:

Definição 6.5 (Probabilidade). *A probabilidade $P(\cdot)$ é a função com domínio em \mathcal{A} (uma sigma álgebra de eventos) e contradomínio no intervalo $[0, 1]$ que satisfaz os seguintes axiomas:*

(i) $P(A) \geq 0$ para todo $A \in \mathcal{A}$.

(ii) $P(\Omega) = 1$.

(iii) Se A_1, A_2, \dots é uma série de eventos mutuamente exclusivos em \mathcal{A} , então

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

A junção dos conceitos de espaço amostral, espaço de eventos (sigma álgebra) e probabilidade forma um trio sobre o qual se pode desenvolver as regras de cálculo de probabilidades.

Definição 6.6 (Espaço de Probabilidade). *Espaço de probabilidade é o trio $[\Omega, \mathcal{A}, P(\cdot)]$, sendo Ω o espaço amostral, \mathcal{A} o espaço de eventos (assumido como uma sigma álgebra) e $P(\cdot)$ a função de probabilidade com domínio em \mathcal{A} .*

Esta definição completa a definição axiomática de probabilidade. Doravante, sempre que se referir a um “evento”, se estará referindo a um evento definido no espaço de probabilidade $[\Omega, \mathcal{A}, P(\cdot)]$, por isto, serão evitadas referências redundantes a ele. Mas toda vez que se ler “seja um evento A ”, entenda-se “seja um evento A no espaço de probabilidade $[\Omega, \mathcal{A}, P(\cdot)]$ ”.

Cálculo de Probabilidades

A partir da definição axiomática de probabilidade é possível se deduzir uma série de resultados que nos permitem realizar o cálculo de probabilidades. Os principais resultados são:

1. $P(\emptyset) = 0$.
2. Se A é um evento em \mathcal{A} , então: $P(A^c) = 1 - P(A)$.
3. Utilizando a notação de intersecção de eventos $(A \cap B) = (A, B)$, para dois eventos A e $B \in \mathcal{A}$, então

$$\begin{aligned} P(A) &= P(A, B) + P(A, B^c) \quad e \\ P(A - B) &= P(A, B^c) = P(A) - P(A, B) \end{aligned}$$

4. Se A e $B \in \mathcal{A}$ e $A \subset B$, então

$$\begin{aligned} P(A) &\leq P(B), \\ P(A) &= P(A, B), \quad e \\ P(B) &= P(A) + P(A^c, B) \end{aligned}$$

5. A união de qualquer dois eventos A e $B \in \mathcal{A}$:

$$P(A \cup B) = P(A) + P(B) - P(A, B)$$

6. A união de qualquer dois eventos *disjuntos* A e $B \in \mathcal{A}$:

$$P(A \cup B) = P(A) + P(B).$$

7. A união de uma série de eventos $A_1, A_2, \dots, A_n \in \mathcal{A}$:

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \\ &\quad - \sum \sum_{i < j} P(A_i, A_j) + \sum \sum \sum_{i < j < k} P(A_i, A_j, A_k) + \\ &\quad + \dots + (-1)^{n+1} P(A_1, A_2, \dots, A_n) \end{aligned}$$

8. **(Desigualdade de Boole)** Se $A_1, A_2, \dots, A_n \in \mathcal{A}$, então:

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$$

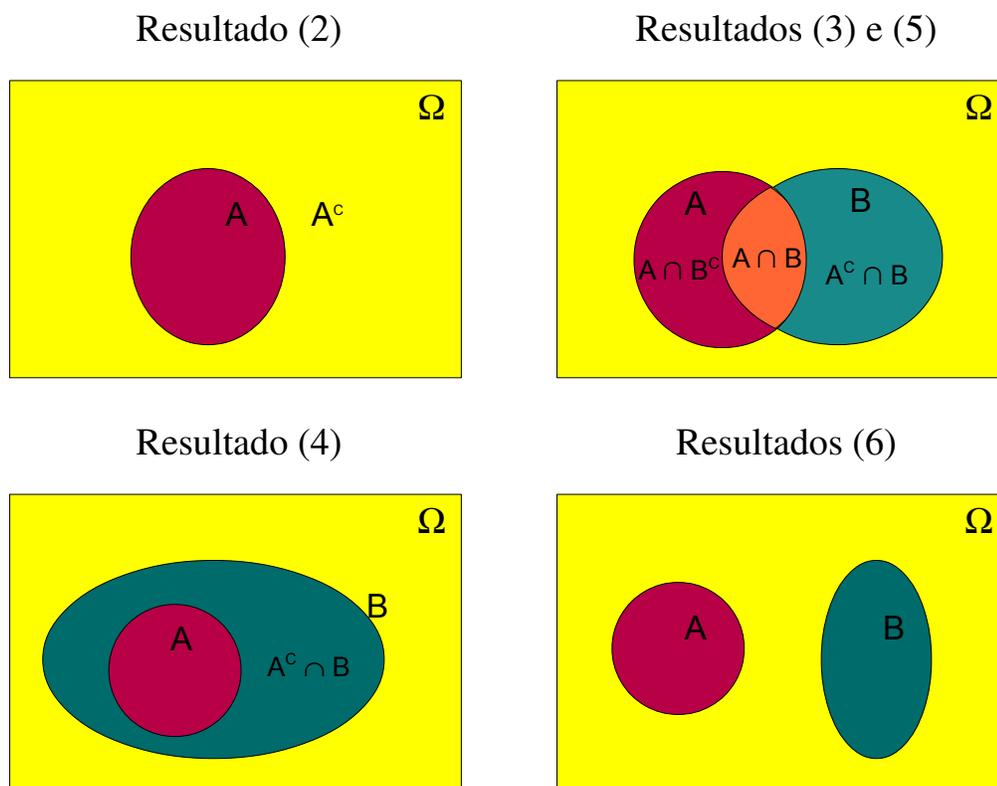


Figura 6.1: Figuras ilustrativas dos resultados do cálculo de probabilidade utilizando diagramas de Venn.

A figura 6.1 ilustra alguns destes resultados utilizando diagramas de Venn.



Exemplo 6-11: Ninhada de Cachorro-do-Mato IV

Tomando o exemplo do cenário estocástico do número de filhotes que um casal de cachorros-do-mato gera em um ano, considere que a probabilidade de uma ninhada não ocorrer (0 filhotes na ninhada) é equiprovável a ela ocorrer e ter qualquer número de filhotes entre três e seis. Neste caso, o espaço amostral, formado por duas ninhadas, é composto pelos seguintes eventos elementares equiprováveis:

veis:

$$\Omega = \left\{ \begin{array}{cccccc} (0, 0), & (0, 3), & (0, 4), & (0, 5), & (0, 6), \\ (3, 0), & (3, 3), & (3, 4), & (3, 5), & (3, 6), \\ (4, 0), & (4, 3), & (4, 4), & (4, 5), & (4, 6), \\ (5, 0), & (5, 3), & (5, 4), & (5, 5), & (5, 6), \\ (6, 0), & (6, 3), & (6, 4), & (6, 5), & (6, 6) \end{array} \right\}$$

Os seguintes eventos podem, então, ser analisados:

A — o número de filhotes no ano é 9: $A = \{(3, 6), (4, 5), (5, 4), (6, 3)\}$

B — o número de filhotes em pelo menos uma das ninhadas é 6:

$$B = \{(0, 6), (3, 6), (4, 6), (5, 6), (6, 6), (6, 0), (6, 3), (6, 4), (6, 5)\}$$

(A, B) — o número de filhotes no ano é 9 e pelo menos uma das ninhadas foi de 6 filhotes: $(A, B) = \{(3, 6), (6, 3)\}$.

A probabilidade do evento A não ocorrer é

$$P(A^c) = 1 - P(A) = 1 - 4/25 = 21/25.$$

A probabilidade de ocorrer o evento A ou o evento B é:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A, B) \\ &= 4/25 + 9/25 - 2/25 = 11/25. \end{aligned}$$



6.3 Probabilidade Condicional e Independência

A probabilidade condicional surge da necessidade de se considerar a probabilidade de ocorrência de um evento uma vez que sabemos que um outro evento aconteceu. Por exemplo, dentre as árvores com diâmetro menor que 10 *cm*, qual a probabilidade da árvore ser uma árvore dominada?

Definição 6.7 (Probabilidade Condicional). *Sejam dois eventos A e $B \in \mathcal{A}$, a probabilidade condicional do evento A , dado o evento B é*

$$P(A|B) = \frac{P(A, B)}{P(B)} \quad \text{se } P(B) \neq 0,$$

permanecendo indefinida se $P(B) = 0$.

Assumindo $P(B) > 0$, desta definição podem ser deduzidos uma série de resultados:

1. Esta definição implica que há dois caminhos para se obter a probabilidade da intersecção de dois eventos:

$$P(A|B) = \frac{P(A, B)}{P(B)} \Rightarrow P(A, B) = P(A|B)P(B)$$

$$P(B|A) = \frac{P(A, B)}{P(A)} \Rightarrow P(A, B) = P(B|A)P(A)$$

2. $P(\emptyset|B) = 0$
3. Se A é um evento em \mathcal{A} , então $P(A^c|B) = 1 - P(A|B)$.
4. Para qualquer dois eventos A_1 e A_2 em \mathcal{A}

$$P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1, A_2|B)$$

5. Se A_1, \dots, A_n são eventos mutuamente exclusivos em \mathcal{A} , então

$$P(A_1 \cup \dots \cup A_n|B) = \sum_{i=1}^n P(A_i|B)$$

6. Se A_1, \dots, A_n são eventos em \mathcal{A} , então

$$P\left(\bigcup_{i=1}^n A_i|B\right) \leq \sum_{i=1}^n P(A_i|B)$$

7. **(Regra da Multiplicação)** Se A_1, \dots, A_n são eventos em \mathcal{A} , tais quais $P(A_1, A_2, \dots, A_n) > 0$, então

$$P(A_1, A_2, \dots, A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2) \dots P(A_n|A_1, A_2, \dots, A_{n-1})$$

8. **(Regra da Probabilidade Total)** Seja B_1, \dots, B_n uma coleção de eventos disjuntos em \mathcal{A} , que satisfaz as condições

$$\Omega = \bigcup_{i=1}^n B_i \quad \text{e}$$

$$P(B_i) > 0 \quad \text{para } i = 1, \dots, n.$$

Então, para qualquer evento A em \mathcal{A} , temos

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i).$$

9. **(Corolário da Probabilidade Total)** Seja A e B dois eventos em \mathcal{A} , tal que $P(B) > 0$, então:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$

A probabilidade condicional facilita o cálculo de probabilidades quando se trabalha com ensaios que envolvem várias etapas ou estágios. Com base no acontecido num estágio, pode se calcular a probabilidade de ocorrência num estágio posterior, ou vice-versa. O cálculo de probabilidade de eventos que envolvem todos os estágio é frequentemente facilitado utilizando-se probabilidade condicional, particularmente a regra de multiplicação e a regra da probabilidade total.



Exemplo 6-12: Ninhada de Cachorro-do-Mato V

Voltando novamente ao exemplo da ninhada de cachorro-do-mato, considere um cenário estocástico com uma estrutura diferente das anteriores. As fêmeas do cachorro-do-mata pode ter até duas ninhadas no ano, uma no verão e outra no inverno, cada uma com três a seis filhotes. Assume-se que a probabilidade de ocorrer duas ninhadas no ano é $2/3$, a de ocorrer somente a ninhada de verão é $2/9$, a de ocorrer somente a ninhada de inverno é $1/18$ e a de não ocorrer nenhuma ninhada é $1/18$. Já o número de filhotes em cada ninhada é equiprovável.

Define-se como A o evento de ocorrer 3 filhotes no ano e com B_i o evento de ocorrência das ninhadas, com as seguintes possibilidades: B_1 – duas ninhadas, B_2 – somente ninhada de verão, B_3 – somente ninhada de inverno e B_4 – nenhuma ninhada.

Pela regra da probabilidade total, a probabilidade de ocorrer 3 filhotes no ano é

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \\ + P(A|B_3)P(B_3) + P(A|B_4)P(B_4)$$

$$P(A) = (0)(2/3) + (1/4)(2/9) + (1/4)(1/18) + (0)(1/18) = 5/72$$



Exemplo 6-13: Exame de Sangue

Um dado tipo de exame de sangue resulta positivo, caso a pessoa tenha a doença, com probabilidade de 99,9%. Por outro lado, em pessoas saudáveis, o exame resulta positivo em 10% dos casos. Sabendo que incidência da doença na população é de 2%, qual a probabilidade de um exame qualquer resultar positivo?

O evento E^+ é o resultado positivo no exame, enquanto o evento D^+ é a pessoa examinada ser possuidora da doença, sendo seu complemento D^- a pessoa não possuir a doença. A regra da probabilidade total indica que

$$P(E^+) = P(E^+|D^+)P(D^+) + P(E^+|D^-)P(D^-) \\ = (0,999)(0,02) + (0,001)(0,98) = 0,02096 \approx 2\%$$



Exemplo 6-14: Reprodução de Árvores

Considere o processo de regeneração natural de uma população de paineira (*Ceiba speciosa*, *Malvaceae*) em floresta estacional semidecidual. Assume-se que a probabilidade de uma semente dispersa pelo vento germinar com sucesso é de 1%. Das sementes que germinam, somente 5% se estabelecem com arboretas no sub-bosque da floresta, enquanto, das arboretas do sub-bosque, somente 10% chega a se estabelecer no dossel da mata. Das árvores que atingem o dossel da mata, todas elas atingem a fase reprodutiva e produzem sementes anualmente. Qual a

probabilidade de uma dada semente dispersada no vento chegar a produzir outras sementes?

Seja A_i o evento da passagem com sucesso pela fase de desenvolvimento i ($i = 1, 2, 3, 4$), então

$$\begin{aligned} P(A_1, A_2, A_3, A_4) &= P(A_1) P(A_2|A_1) P(A_3|A_1, A_2) P(A_4|A_1, A_2, A_3) \\ &= (0,01)(0,05)(0,10)(1) = 5 \times 10^{-5} = 5 \times 10^{-3}\% \end{aligned}$$



Teorema de Bayes

O Teorema de Bayes é um resultado da teoria de probabilidades deduzido por Thomas Bayes e publicado postumamente em 1763. O teorema permite encontrar a probabilidade de um dado evento que faz parte de uma série de eventos dado que um evento externo à série ocorreu. No século XX, com a aceitação do conceito subjetivo de probabilidade, o teorema passou a ser a base de uma forma de inferência estatística: a Inferência Bayesiana.

Definição 6.8 (Teorema de Bayes). *Seja B_1, \dots, B_n uma coleção de eventos disjuntos em \mathcal{A} , que satisfaz as condições*

$$\begin{aligned} \Omega &= \bigcup_{i=1}^n B_i \quad e \\ P(B_i) &> 0 \quad \text{para } i = 1, \dots, n. \end{aligned}$$

Então, para qualquer evento A em \mathcal{A} , temos

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^n P(A|B_i)P(B_i)}.$$

O teorema de Bayes parte do conceito de probabilidade condicional

$$P(B_k|A) = \frac{P(B_k, A)}{P(A)},$$

mas considera que há dois caminhos para se obter a probabilidade da intersecção de dois eventos

$$P(B_k, A) = P(A|B_k)P(B_k) \Rightarrow P(B_k|A) = \frac{P(A|B_k)P(B_k)}{P(A)}$$

e aplica a regra da probabilidade total

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i) \Rightarrow P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^n P(A|B_i)P(B_i)}.$$

Assim, o teorema de Bayes permite calcular probabilidades de questões bem mais complexas.



Exemplo 6-15: Ninhada de Cachorro-do-Mato VI

Retomando o exemplo da ninhada de cachorro-do-mato segundo o centário anterior (exemplo V), considere A como sendo o evento da ocorrência de 6 filhotes no ano. Segundo a ocorrência das ninhadas no ano, este evento pode ocorrer das seguintes maneiras:

- duas ninhadas no ano, cada uma com três filhotes:
 $P(A|B_1) = (1/4)(1/4) = 1/16$;
- somente ninhada de verão com seis filhotes: $P(A|B_2) = (1/4)$;
- somente ninhada de inverno com seis filhotes: $P(A|B_3) = (1/4)$; e
- nenhuma ninhada no ano: $P(A|B_4) = 0$.

Dado que foram observados seis filhotes, qual a probabilidade deles serem resultados de duas ninhadas no ano? A solução é encontrada aplicando-se o Teorema de Bayes:

$$\begin{aligned} P(B_1|A) &= \frac{P(A|B_1)P(B_1)}{\sum_{i=1}^4 P(A|B_i)P(B_i)} \\ &= \frac{(1/16)(2/3)}{(1/16)(2/3) + (1/4)(2/9) + (1/4)(1/18) + (0)(1/18)} = 3/8. \end{aligned}$$



Exemplo 6-16: Exame de Sangue II

Um dado tipo de exame de sangue resulta positivo, caso a pessoa tenha a doença, com probabilidade de 99,9%. Por outro lado, em pessoas saudáveis, o exame resulta positivo em 10% dos casos. Sabendo que incidência da doença na população é de 2%, qual a probabilidade de um exame qualquer resultar num “falso positivo”, isto é, a pessoa ser saudável embora o exame tenha sido positivo?

O evento E^+ é o resultado positivo no exame, enquanto o evento D^+ é a pessoa examinada ser possuidora da doença, sendo seu complemento D^- a pessoa não possuir a doença. Aplica-se o teorema de Bayes

$$\begin{aligned} P(D^-|E^+) &= \frac{P(E^+|D^-)P(D^-)}{P(E^+|D^+)P(D^+) + P(E^+|D^-)P(D^-)} \\ &= \frac{(0,001)(0,98)}{(0,999)(0,02) + (0,001)(0,98)} = 0,04675573 \approx 5\%. \end{aligned}$$

Este resultado parece paradoxal, pois de acordo com o exemplo 3-13, $P(E^+) \approx 2\%$, logo a frequência de diagnóstico positivo equivocado é maior que a frequência de resultados positivos na população. Isto acontece porque a probabilidade de resultado positivo em pessoas saudáveis ainda é muito grande ($P(E^+|D^-) = 0,001$) para a raridade da doença ($P(D^+) = 0,02$). Se a probabilidade de resultado positivo em pessoa saudável for reduzido em uma ordem de grandeza, a probabilidade de falso positivo se torna igualmente uma ordem de grandeza menor $P(D^-|E^+) = 0,004876592 \approx 0,5\%$.



Eventos Independentes

Quando a probabilidade condicional de um evento não depende do evento condicionante, isto é, $P(A|B)$ não depende de B , diz-se que os eventos são independentes. O cálculo de probabilidade envolvendo eventos independentes se torna bastante simplificado.

Definição 6.9 (Eventos Independentes). *Dois evento A e B em \mathcal{A} são independentes se, e somente se, qualquer uma das seguintes condições é satisfeita:*

- (i) $P(A, B) = P(A)P(B)$
- (ii) $P(A|B) = P(A)$ se $P(B) > 0$

(ii) $P(B|A) = P(B)$ se $P(A) > 0$

Definição 6.10 (Múltiplos Eventos Independentes). *Seja A_1, A_2, \dots, A_n uma coleção de eventos em \mathcal{A} , os eventos são independentes se, e somente se,*

$$P(A_1, A_2, \dots, A_n) = P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$$

A independência de eventos implica em alguns resultados no cálculo de probabilidades:

1. Se A e B são dois eventos em \mathcal{A} , tal que $P(A) \neq 0$ e $P(B) \neq 0$, então
 - Se A e B são independentes, eles **não** são mutuamente exclusivos.
 - Se A e B são mutuamente exclusivos, eles **não** são independentes.
2. Se dois eventos A e B em \mathcal{A} são independentes, então:
 - A^c e B são independentes,
 - A e B^c são independentes, e
 - A^c e B^c são independentes.
3. Numa coleção de eventos A_1, A_2, \dots, A_n , a independência *dois-a-dois* não implica em independência entre os n eventos.



Exemplo 6-17: Ninhada de Cachorro-do-Mato V

Revisitando o exemplo da ninhada de cachorro-do-mato, considere-se o cenário estocástico onde as duas ninhadas do ano sempre ocorrem e o número de filhotes nas ninhadas é equiprovável. Neste caso, o espaço amostral é

$$\Omega = \left\{ \begin{array}{cccc} (3, 3) & (3, 4) & (3, 5) & (3, 6) \\ (4, 3) & (4, 4) & (4, 5) & (4, 6) \\ (5, 3) & (5, 4) & (5, 5) & (5, 6) \\ (6, 3) & (6, 4) & (6, 5) & (6, 6) \end{array} \right\}.$$

Considere-se então os seguintes eventos:

- A_1 o número de filhotes na primeira ninhada é ímpar:
 $A_1 = \{(3, i), (5, i)\} (i = 3, 4, 5, 6) \Rightarrow P(A_1) = 8/16 = 1/2.$

- A_2 o número de filhotes na segunda ninhada é ímpar:
 $A_2 = \{(i, 3), (i, 5)\} (i = 3, 4, 5, 6) \Rightarrow P(A_2) = 8/16 = 1/2.$
- A_3 o número de filhotes no ano é ímpar:
 $A_3 = \{(3, j), (5, j), (j, 3), (j, 5)\} (j = 4, 6) \Rightarrow P(A_3) = 8/16 = 1/2.$

Os eventos A_1 , A_2 e A_3 são independentes dois-a-dois

$$\begin{aligned} P(A_1, A_2) &= P(A_2|A_1)P(A_1) \\ &= (4/8)(1/2) = (1/2)(1/2) = P(A_2)P(A_1) \end{aligned}$$

$$\begin{aligned} P(A_1, A_3) &= P(A_3|A_1)P(A_1) \\ &= (4/8)(1/2) = (1/2)(1/2) = P(A_3)P(A_1) \end{aligned}$$

$$\begin{aligned} P(A_2, A_3) &= P(A_3|A_2)P(A_2) \\ &= (4/8)(1/2) = (1/2)(1/2) = P(A_3)P(A_2). \end{aligned}$$

Mas os três eventos não são independentes porque

$$P(A_1, A_2, A_3) = 0 \neq P(A_1)P(A_2)P(A_3) = 1/8.$$

Logo, independência dois-a-dois não garante a completa independência numa coleção de eventos.



Exemplo 6-18: Aniversariantes na Festa

Numa festa há 25 pessoas. Assume-se que os dias do ano são equiprováveis para nascimentos e que o dia de aniversários das pessoas são eventos independentes. Qual é a probabilidade de pelo menos duas pessoas na festa fazerem aniversário no mesmo dia?

Considere o evento A como a ocorrência de pelo menos duas pessoas com aniversário no mesmo dia. É muito complicado calcular a probabilidade de A , pois este evento implica não só em duas pessoas tendo coincidência de aniversários, três pessoas, quatro pessoas, etc., mas também em várias coincidências duas-a-duas, três-a-três, quatro-a-quatro, etc.

Contudo, o evento A^c é bem mais simples, pois ele corresponde a nenhuma coincidência de aniversário, sendo mais fácil calcular a probabilidade. Considere D_i o evento em que a $i^{\text{ésima}}$ pessoa na festa ($i = 1, 2, \dots, 25$) tomar um dos dias do ano como seu aniversário numa amostragem sem reposição. Se as datas de aniversário são independentes, como foi assumido, a probabilidade de nenhuma coincidência de data dado as datas de aniversário das 25 pessoas é igual ao produto de uma série de probabilidades condicionais:

$$P(A^c) = P(A^c|D_1) P(A^c|D_1, D_2) \dots P(A^c|D_1, D_2, \dots, D_{25}).$$

Para que não haja coincidência de data, a primeira pessoa terá disponíveis os 365 dias do ano, já a segunda pessoa terá 365-1 dias no ano, enquanto que a terceira terá 365-2, e assim sucessivamente até a vigésima quinta pessoa, que terá 365-24 dias disponíveis. Assim, as probabilidades condicionais resultam em

$$P(A^c) = \left(\frac{365}{365}\right) \left(\frac{365-1}{365}\right) \left(\frac{365-2}{365}\right) \dots \left(\frac{365-24}{365}\right) = \frac{365!/(365-25)!}{365^{25}}$$

Assim, a probabilidade de pelo menos uma coincidência de aniversário é

$$P(A) = 1 - P(A^c) = 1 - \frac{365!/(365-25)!}{365^{25}} = 0.5686997 \approx 57\%$$

Este exemplo mostra que podem existir mais coincidências de eventos aleatórios independentes na vida cotidiana que geralmente as pessoas imaginam.



Independência, Cenário Estocástico e Processo Estocástico

A premissa fundamental das Ciências Naturais é que a independência entre fenômenos naturais é relativamente rara, do contrário como se poderia construir explicações com base em causas e efeitos. Contudo, a *independência probabilística* apresentada acima não corresponde necessariamente à ausência da relação de causa e efeito nos fenômenos naturais, embora ela possa ser tomada pela inferência quantitativa no contexto de uma dada pesquisa científica como evidência contra a existência de relações causais.

No caso da Biometria Florestal, a independência probabilística é o aspecto que diferencia cenários e processos estocásticos. Embora não esteja explícito na definição de cenário estocástico, a imprevisibilidade dos resultados dos ensaios

dos cenários estocásticos implica na independência entre eles. Assim, um determinado resultado particular num dado cenário estocástico não traz nenhuma informação sobre os resultados seguintes a serem observados neste mesmo cenário. Portanto, os cenários estocásticos são modelados por modelos que têm como uma de suas premissas a independência entre as observações.

Por outro lado, os processos estocásticos consideram explicitamente as escalas espacial, temporal ou ambas. Por isto, os modelos voltados aos processos estocásticos devem necessariamente incorporar uma relação de dependência espacial, temporal ou espaço-temporal explícita entre as observações. Logo, nestes modelos a premissa é a ausência de independência e, conseqüentemente, uma parcela significativa do esforço da sua construção é voltada à representação adequada da relação de dependência.

É importante ressaltar que a distinção entre cenário estocástico e processo estocástico não é um dado do fenômeno natural sendo estudado. Esta distinção é antes de tudo uma questão de abordagem na modelagem. Considere-se, por exemplo, a questão do inventário florestal. A abordagem clássica para realização da inferência num inventário florestal é a aplicação da Teoria da Amostragem, na qual, por efeito do delineamento amostral utilizado, observações independentes são geradas. Logo, a abordagem clássica modela os dados como resultados de um cenário estocástico. A abordagem geoestatística, por outro lado, utiliza a informação da localização geográfica de cada unidade amostral e modela as observações a partir da existência de uma relação de dependência espacial entre elas. Ambas abordagens são igualmente válidas, a escolha entre elas é fundamentalmente uma questão de utilidade e praticidade.

* * *

Neste capítulo, foram apresentados os conceitos básicos do cálculo de probabilidades. Mas, para se avançar em direção aos modelos estocásticos utilizados na Biometria Florestal, é necessário se apresentar o conceito de distribuição estocástica, que é o conceito-chave para modelar medidas e atributos de árvores e florestas. Mas, o desenvolvimento do conceito de distribuição estocástica será apresentado somente no contexto dos cenários estocásticos, ou seja, no contexto de que os resultados dos ensaios destes cenários são independentes. O contexto dos processos estocásticos exige que as distribuições estocásticas incluam a relação de dependência entre as observações, o que só pode ser realizado nas distribuições estocásticas multivariadas. Mas estas distribuições não serão abordadas.

Capítulo 7

Distribuições Estocásticas Modelos Básicos da Inferência

Definidas as regras do cálculo de probabilidades, a teoria de probabilidades avança para um conceito chave para a modelagem estocástica: o conceito de distribuição estocástica. Este conceito se faz necessário porque o cálculo de probabilidades, como apresentado no capítulo anterior, é baseado em eventos cuja a codificação pode ter uma variedade de formas, não sendo apenas numérica. Mas, em grande espaços amostrais, ou em espaços amostrais infinitos, a codificação dos eventos se torna demasiadamente complexa ou inviável para permitir uma abordagem probabilística. Considere, no exemplo da ninhada do cachorro-do-mato, como ficaria o espaço amostral se fossem estudadas dez fêmeas e não apenas uma.

No cálculo de probabilidades, também se considera os eventos elementares como sendo equiprováveis. Por exemplo, na observação dos DAP das árvores de uma floresta, o DAP de uma dada árvore é equiprovável ao DAP de todas as demais, caso as árvores sejam selecionadas de modo completamente aleatório. Mas, os “valores numéricos” dos DAP não são equiprováveis. Na prática do manejo e conservação das florestas, o que importa não são os DAP das árvores individualmente, mas a frequência com que eles assumem os diversos valores numéricos possíveis.

É necessário, portanto, ir além do espaço de probabilidade, que permite o cálculo de probabilidades de eventos, para uma situação onde as medidas quantitativas das árvores, dos arvoredos e das florestas possam ser tratadas probabilisticamente. O conceito que permite este avanço é o conceito de distribuição estocástica, sendo que o seu desenvolvimento é apresentado neste capítulo, o qual segue, em linhas gerais, a abordagem utilizada por Mood, Graybill e Boes (1974).

7.1 Variável Aleatória e Distribuição Estocástica

A expressão “variável aleatória” é uma expressão consagrada tanto na teoria de probabilidade quanto na estatística. Contudo, apesar de sua popularidade, ela é uma expressão enganadora e causadora de confusão. Primeiramente, uma variável aleatória não é uma variável, mas uma função. Em segundo lugar, uma variável que varia “aleatoriamente”, sugere a ideia de algo que varia de modo completamente imprevisível, mas uma variável aleatória está longe de ser algo imprevisível. Também Hacking (1965) e Mood, Graybill e Boes (1974) expressam desconforto com esta expressão, mas enquanto o primeiro a evita totalmente, os últimos aceitam a sua popularidade e a adotam.

A definição de variável aleatória será utilizada para estabelecer uma ponte entre o conceito de espaço de probabilidade e o conceito de distribuição estocástica. Em geral, uma variável aleatória é considerada equivalente a uma distribuição estocástica, embora haja uma pequena diferença sutil. De qualquer forma, o conceito de distribuição é claro e não dá margens a confusões. Neste documento, a expressão “variável aleatória” será abandonada e evitada assim que o conceito de distribuição estocástica esteja estabelecido.

Definição 7.1 (Variável Aleatória). *Dado um espaço de probabilidade $[\Omega, \mathcal{A}, P(\cdot)]$, uma variável aleatória, denotada por X ou $X(\cdot)$, é uma função com domínio no espaço amostral Ω e contradomínio no conjunto dos números reais (\mathbb{R}). A função $X(\cdot)$ deve ser tal que o conjunto A_r , definido como $A_r = \{\omega : X(\omega) \leq r\}$, pertence ao espaço de eventos \mathcal{A} para todo número real r .*

O espaço amostral de um cenário estocástico não precisa ser necessariamente definido por um conjunto numérico, podendo ser um conjunto de elementos “codificados”. Assim, variável aleatória é uma “função” que estabelece uma relação entre os elementos de um espaço amostral e os números reais. Portanto, variável aleatória é o conceito que permite transformar qualquer espaço amostral num conjunto numérico.



Exemplo 7-1: Mortalidade de Árvores

Considere o cenário estocástico em que uma árvore é selecionada de uma floresta e se observa se ela está viva ou morta. O espaço amostral deste cenário é $\Omega = \{\text{VIVA}, \text{MORTA}\}$. Assim, é possível definir a variável aleatória X tal que $X(\omega) = 0$, se $\omega = \text{VIVA}$, e $X(\omega) = 1$, se $\omega = \text{MORTA}$.

Esta definição de X satisfaz a condição de variável aleatória, pois o espaço de eventos \mathcal{A} consiste de quatro subconjuntos: \emptyset , $\{\text{VIVA}\}$, $\{\text{MORTA}\}$, Ω . Logo,

verifica-se que a condição é alcançada, pois

$$\begin{aligned} \text{se } r < 0, & \quad \{\omega : X(\omega) \leq r\} = \emptyset, \\ \text{se } 0 \leq r < 1, & \quad \{\omega : X(\omega) \leq r\} = \{\text{MORTA}\}, \\ \text{e se } r \geq 1, & \quad \{\omega : X(\omega) \leq r\} = \Omega = \{\text{MORTA, VIVA}\}. \end{aligned}$$



Exemplo 7-2: Árvores para Serraria

Considere o cenário estocástico em que foi observado um conglomerado de 2 árvores numa floresta e se registrou quais árvores eram aptas para serraria (S) ou não (N). O espaço amostral deste cenário estocástico é

$$\Omega = \{(S, S), (S, N), (N, S), (N, N)\}.$$

Defina-se a variável aleatória X como o número de árvores aptas para serraria no conglomerado, então

$$\begin{aligned} X(\omega) &= 2, \text{ se } \omega = (S, S), \\ X(\omega) &= 1, \text{ se } \omega = (N, S), \\ X(\omega) &= 1, \text{ se } \omega = (S, N), \text{ e} \\ X(\omega) &= 0, \text{ se } \omega = (N, N). \end{aligned}$$

Também neste caso, a definição satisfaz a condição de variável aleatória:

$$\begin{aligned} \text{se } r < 0, & \quad \{\omega : X(\omega) \leq r\} = \emptyset, \\ \text{se } 0 \leq r < 1, & \quad \{\omega : X(\omega) \leq r\} = \{(N, N)\}, \\ \text{se } 1 \leq r < 2, & \quad \{\omega : X(\omega) \leq r\} = \{(N, N), (S, N), (N, S)\}, \\ \text{e se } r \geq 2, & \quad \{\omega : X(\omega) \leq r\} = \{(N, N), (S, N), (N, S), (S, S)\} = \Omega. \end{aligned}$$



 *Exemplo 7-3: Diâmetro das Árvores na Floresta*

Uma floresta onde as árvores são observadas, cada qual tendo uma certa “espessura” do tronco, pode ser considerada um cenário estocástico. A espessura do tronco de uma dada árvore é observada por meio de uma suta e o resultado deste ensaio é o número observado na suta. Assim, a medição do diâmetro do tronco das árvores é uma variável aleatória, pois atribui número reais (os valores numéricos observados na suta) à espessura dos troncos das árvores. A condição imposta à criação da variável aleatória é satisfeita uma vez que os próprios resultados são números reais.

Esta situação pode ser igualmente aplicada a qualquer outro atributo tanto das árvores, como altura, forma, volume ou biomassa, quanto dos arvoredos, como área basal, densidade de estande, altura média das árvores dominantes ou produção de madeira.



Como apresentado, uma variável aleatória é sempre definida em termos dos resultados dos ensaios de um cenário estocástico, mas, na maioria das aplicações, ela será utilizada com base numa função matemática. Assim, se faz necessário defini-la em termos de “forma funcional” .

Definição 7.2 (Função de Distribuição). *A função de distribuição de uma variável aleatória X , denotada por $F(\cdot)$, é uma função com domínio no conjunto dos números reais (\mathbb{R}) e contradomínio no intervalo $[0, 1]$ e que satisfaz a condição:*

$$F(x) = P(X \leq x) = P(\{\omega : X(\omega) \leq x\})$$

para qualquer número real x . Cada função de distribuição define uma distribuição estocástica.

A função de distribuição, também chamada de “função de distribuição acumulada”, é definida de forma única para cada variável aleatória, isto é, uma dada variável aleatória terá uma, e apenas uma, função de distribuição. Se a função de distribuição de uma variável aleatória for conhecida, ela pode ser utilizada para calcular probabilidades associadas à variável aleatória sem qualquer referência direta ao espaço de probabilidade definido a partir do cenário estocástico. Logo, o conceito de função de distribuição simplifica grandemente o cálculo de probabilidades em problemas mais complexos. Cada função de distribuição determina um modelo estocástico específico que pode ser utilizado para o cálculo de probabilidades, sendo que estes modelos são designados por *distribuições estocásticas*, mas

também são chamados de *distribuições estatísticas* ou *distribuições de probabilidade*.



Exemplo 7-4: Mortalidade de Árvores II

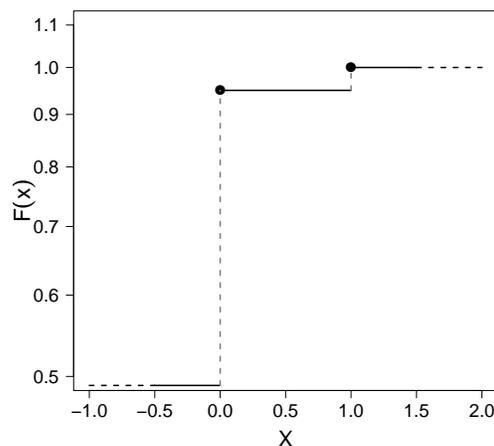
Considere o cenário estocástico da observação de árvores vivas e mortas numa floresta apresentado acima. Se X denota o número de árvores mortas, então pode-se ter a seguinte função de distribuição para X :

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ \frac{95}{100} & \text{se } 0 \leq x < 1 \\ 1 & \text{se } 1 \leq x \end{cases}$$

Esta função de distribuição implica que a probabilidade de se observar uma árvore morta na floresta é de 5%, mas ela também pode ser apresentada numa forma funcional mais tradicional:

$$F(x) = \sum_{i=0}^x \left(\frac{5}{100}\right)^i \left(1 - \frac{5}{100}\right)^{1-i}, \quad x = 0, 1$$

O gráfico desta distribuição estocástica é



Exemplo 7-5: Árvores para Serraria II

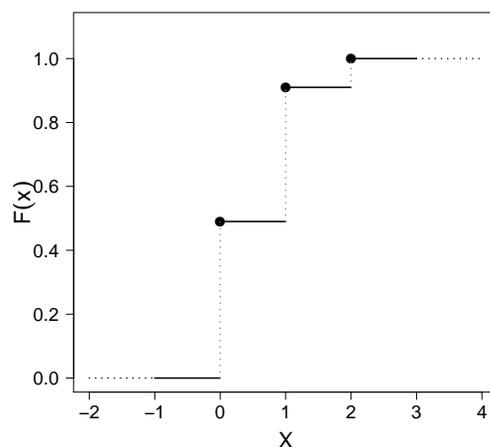
Considere o cenário estocástico das árvores para serraria apresentado acima. Se X é o número de árvores aptas para serraria no conglomerado de duas árvores, uma função de distribuição possível para X é

$$F(x) = \begin{cases} 0 & \text{se } x < 0 \\ 49/100 & \text{se } 0 \leq x < 1 \\ 91/100 & \text{se } 1 \leq x < 2 \\ 1 & \text{se } 2 \leq x \end{cases}$$

que também pode ser expressa numa expressão algébrica mais compacta

$$F(x) = \sum_{i=0}^x \binom{x}{i} \left(\frac{30}{100}\right)^i \left(1 - \frac{30}{100}\right)^{2-i}, \quad x = 0, 1, 2.$$

O gráfico desta distribuição é

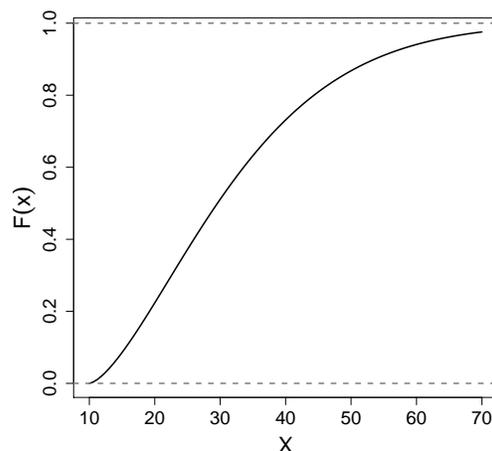


Exemplo 7-6: Diâmetro das Árvores na Floresta II

Considere o cenário estocástico da observação dos diâmetros das árvores de uma floresta, apresentado acima. Uma função de distribuição possível para o diâmetro das árvores é

$$F(x) = 1 - \exp \left[- \left(\frac{x - 10}{25} \right)^{1.5} \right], \quad x \geq 0,$$

cujo gráfico é



Os exemplos acima mostram que a função de distribuição não precisa ser totalmente contínua, mas existem três propriedades que ela necessariamente deve ter:

Propriedades da Função de Distribuição:

- (i) O contradomínio de $F(\cdot)$ é o intervalo $[0, 1]$. Como o seu domínio é o conjunto dos números reais, tem-se que $F(-\infty) \equiv \lim_{x \rightarrow -\infty} F(x) = 0$ e $F(\infty) \equiv \lim_{x \rightarrow \infty} F(x) = 1$.
- (ii) $F(\cdot)$ é uma função monotônica não decrescente, isto é, $F(a) \leq F(b)$ para $a \leq b$.
- (iii) $F(\cdot)$ é uma função contínua à direita, ou seja

$$\lim_{0 < h \rightarrow 0} F(x + h) = F(x).$$

Na verdade, estas três propriedades podem ser tomadas como propriedades definidoras da função de distribuição.

Definição 7.3 (Função de Distribuição II). *Qualquer função com domínio em \mathbb{R} e contradomínio no intervalo $[0, 1]$ e que satisfaz as três propriedades apresentadas acima é uma função de distribuição.*

Esta definição implica na existência de um número infinito de distribuições estocásticas (funções de distribuição) possíveis, mas, na prática, trabalha-se com um conjunto relativamente pequeno delas.

7.2 Função de Distribuição e Função de Densidade

O conceito de função de distribuição, ao definir a ideia de “*distribuições estocásticas*”, permite que se possa tratar de problemas complexos de cálculo de probabilidades sem referência direta aos conceitos de espaço de probabilidade e de variável aleatória. Mas para que o conceito possa ser aplicado de modo adequado aos problemas práticos, é necessário ainda distinguir dois tipos de funções de distribuição: as *discretas* e as *contínuas*.

As funções discretas são modelos estocásticos apropriados para dados de contagem, como número de árvores mortas, número de plântulas ou número de árvores defeituosas ou doentes. Já as funções contínuas são modelos apropriadas ao tratamento de dados resultantes de medidas, como DAP, altura e biomassa das árvores, ou como o índice de sítio ou a produção dos arvoredos.

Distribuição Estocástica Discreta

Primeiramente, define-se o que são funções de distribuição discretas, utilizando o conceito de variável aleatória.

Definição 7.4 (Função de Distribuição Discreta). *Uma variável aleatória X é definida como discreta se o conjunto imagem de X é contável. Se uma variável aleatória X é discreta, a sua função de distribuição correspondente $F(\cdot)$ é definida como discreta.*

A expressão “*o conjunto imagem de X é contável*” significa que existe um subconjunto dos números reais que é finito ou infinito contável $(x_1, x_2, \dots, x_n, \dots)$, tal que X só assume valores neste subconjunto. O que é equivalente a afirmar que o domínio da função de distribuição correspondente $F(\cdot)$ é contável.

Toda distribuição estocástica definida por uma função de distribuição tem associada a ela uma outra função, que no caso de funções discretas, torna o cálculo de probabilidades mais expedito. Esta função é chamada de “*função de densidade discreta*” e pode ser definida a partir da função de distribuição.

Definição 7.5 (Função de Densidade Discreta). *Função de densidade discreta $f_X(\cdot)$ associada a uma distribuição estocástica é a função definida por*

$$f_X(x) = F(x_i) - \lim_{0 < h \rightarrow 0} F(x_i - h),$$

sendo $F(\cdot)$ uma função de distribuição discreta, cujo o domínio contável é $\{x_1, x_2, \dots, x_n, \dots\}$ e i um índice que indica os elementos neste domínio ($i = 1, 2, \dots, n, \dots$). Contudo, como

$$\lim_{0 < h \rightarrow 0} F(x_i - h) = F(x_{i-1})$$

a definição de função de densidade discreta resulta

$$\begin{aligned} f_X(x) &= F(x_i) - F(x_{i-1}) = P(X \leq x_i) - P(X \leq x_{i-1}) \\ f_X(x) &= P(X = x), \quad x = x_i, \quad i = 1, 2, \dots, n, \dots \end{aligned}$$

Esta definição mostra que, no caso de distribuições discretas, a função de densidade é a própria probabilidade dos valores que a variável aleatória pode assumir. O uso de função de densidade é, portanto, a maneira mais direta de se calcular probabilidades nas distribuições discretas.

Como foi visto no caso da função de distribuição, as funções de densidade discretas também podem ser definidas em termos de três propriedades fundamentais.

Definição 7.6 (Função de Densidade Discreta II). *Qualquer função $f(\cdot)$ com domínio em \mathbb{R} e contradomínio no intervalo $[0, 1]$ é definida como função de densidade discreta se para um conjunto contável $x_1, x_2, \dots, x_n, \dots$*

(i) $f(x_i) > 0$, para $i = 1, 2, \dots, x_n, \dots$;

(ii) $f(x) = 0$, para $x \neq x_i$; $i = 1, 2, \dots, x_n, \dots$;

(iii) $\sum f(x) = 1$, sendo a somatória aplicada sobre todos os pontos do conjunto contável $(x_1, x_2, \dots, x_n, \dots)$

Para uma dada distribuição estocástica, a função de densidade e a função de distribuição tem uma correspondência direta. Assim, é possível obter a função de distribuição discreta a partir da função de densidade discreta. Considere-se um conjunto contável em \mathbb{R} , definido por uma série de pontos $x_1, x_2, \dots, x_n, \dots$, então

$$F(x) = P(X \leq x) = \sum_{i: x_i \leq x} f_X(x_i).$$

Dada a função de distribuição podemos obter a função de densidade, e vice-versa. Portanto, a função de distribuição e a função de densidade são duas maneiras matemáticas diferentes de se definir a mesma distribuição estocástica.

Distribuição Estocástica Contínua

As distribuições estocástica contínuas podem ser definidas a partir da propriedade de continuidade da função de distribuição.

Definição 7.7 (Função de Distribuição Contínua). *Uma variável aleatória X é dita contínua, se a sua função de distribuição $F(\cdot)$ for uma função absolutamente contínua, isto é, se*

$$F(x) = \lim_{0 < h \rightarrow 0} F(x + h) = \lim_{0 < h \rightarrow 0} F(x - h)$$

Como no caso das funções discretas, a função de densidade contínua pode ser definida em termos da função de distribuição:

Definição 7.8 (Função de Densidade Contínua). *Dada uma função de distribuição contínua $F(\cdot)$, a função de densidade contínua $f_X(\cdot)$ é aquela, para qual*

$$F(x) = \int_{-\infty}^x f_X(u) du.$$

Como no caso das distribuições discretas, também nas distribuições contínua é possível obter a função de densidade e a função de distribuição, e vice-versa. Por exemplo, se for dada uma função de distribuição contínua $F(x)$, a função de densidade é

$$f_X(x) = \frac{dF(x)}{dx}$$

Como se trata de funções *absolutamente contínuas*, a derivação permitem a obtenção da função de densidade a partir da função de distribuição, enquanto que a integração permite o reverso.

Nulidade de Probabilidades Pontuais. É importante notar que a propriedade de continuidade absoluta na distribuição estocástica, permite o cálculo da probabilidade de intervalos de números reais, seja pela diferença da função de distribuição

$$\begin{aligned} P(x_1 \leq X \leq x_2) &= P(X \leq x_2) - P(X \leq x_1) \\ &= F(x_2) - F(x_1), \end{aligned}$$

seja pela integração da função de densidade

$$\begin{aligned} P(x_1 \leq X \leq x_2) &= \int_{-\infty}^{x_2} f_X(u) du - \int_{-\infty}^{x_1} f_X(u) du \\ &= \int_{x_1}^{x_2} f_X(u) du. \end{aligned}$$

Por outro lado, a probabilidade de um único número real se torna nula:

$$P(X = x) = P(x \leq X \leq x) = F(x) - F(x) = \int_x^x f_X(u)du = 0$$

Mas a variação relativa da probabilidade na vizinhança infinitesimal dos números reais é descrita pela função de densidade, pois ela é a derivada da função de distribuição, a qual descreve a probabilidade acumulada:

$$\lim_{\Delta x \rightarrow 0} \frac{P(X \leq x + \Delta x) - P(X \leq x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = f_X(x).$$

A função de densidade contínua também pode ser definida a partir de duas propriedades fundamentais.

Definição 7.9 (Função de Densidade Contínua II). *Qualquer função $f(x)$ absolutamente contínua com domínio em \mathbb{R} é uma função de densidade contínua se*

- (i) $f(x) \geq 0$, para todo $x \in \mathbb{R}$ (contradomínio no intervalo $[0, \infty)$); e
- (ii) $\int_{-\infty}^{\infty} f(x) dx = 1$.

Diferenças entre Funções de Densidade Discretas e Contínuas. Há duas diferenças principais. A primeira é que a função discreta tem contradomínio no intervalo $[0, 1]$, enquanto que a contínua tem no intervalo $[0, \infty)$. A segunda diferença é que a função discreta gera a probabilidade para de cada ponto num subconjunto contável de números reais ($\{x_1, x_2, \dots, x_n, \dots\}$). A função contínua pode ser integrada para se obter a probabilidade num intervalo de números reais ($[x_1, x_2]$), mas o seu valor para cada ponto não tem interpretação absoluta. A interpretação dos valores pontuais da função contínua é sempre relativa: o aumento, ou diminuição, dos valores pontuais da função indica o aumento, ou diminuição, da probabilidade.

Família de Distribuições

Para maior clareza na apresentação das distribuições estocásticas e da sua aplicação aos problemas de mensuração é necessário fazer a distinção entre a distribuição estocástica enquanto modelo geral, com uma dada forma funcional, e a distribuição estocástica específica, como modelo de um cenários estocástico particular. A expressão “distribuição estocástica” será utilizada desde ponto em diante como se referindo a um modelo de um cenário estocástico particular, pois ela pode ser representada graficamente por uma única curva. A distribuição estocástica como modelo geral, que possui uma dada forma funcional, será designada por

“*família de distribuições estocásticas*”. Neste caso, a representação gráfica pode ser apenas esquemática, apresentando-se várias curvas num mesmo gráfico, pois uma família de distribuições é composta por um número infinito de distribuições específicas.

Definição 7.10 (Família de Distribuições e Parâmetros). *Uma família de distribuições é definida por uma forma funcional específica da função de densidade e da função de distribuição.*

As constantes presentes na forma funcional são chamadas de parâmetros. Os valores numéricos atribuídos aos parâmetros determinam uma distribuição específica dentro da família de distribuições com gráfico particular para a função de densidade e para função de distribuição.

7.3 Famílias de Distribuições Discretas

A caracterização das famílias de distribuições estocásticas discretas é mais fácil quando se utiliza a função de densidade, pois, neste caso, a função apresenta diretamente as probabilidades, sendo possível apresentá-la graficamente também. Apresentam-se algumas famílias discretas de aplicação frequente nas Ciências Florestais e em Ecologia.

Bernoulli

A família de distribuições Bernoulli é o modelo para os cenários estocásticos que envolvem apenas dois resultados possíveis, sendo o espaço amostral $\Omega = \{0, 1\}$. O resultado 1 codifica um “*sucesso*”, isto é, o evento de interesse foi observado. Já o resultado 0 codifica um “*fracasso*”, ou seja, o evento de interesse não foi observado. A função de densidade desta família tem a forma

$$f(x) = f(x; p) = \begin{cases} p^x (1 - p)^{1-x}, & x = 0 \text{ ou } 1; \\ 0 & \text{nos demais casos;} \end{cases} \quad 0 \leq p \leq 1. \quad (7.1)$$

sendo que a constante p é o “*parâmetro*” da distribuição e indica a probabilidade da ocorrência de “*sucesso*”. A figura 7.1 apresenta o gráfico da função de densidade.



Exemplo 7-7: Árvores com Cancro

Numa floresta de eucalipto, a taxa de árvores com cancro é de 3,5%. Portanto, a probabilidade de uma árvore selecionada aleatoriamente nesta floresta ter cancro é

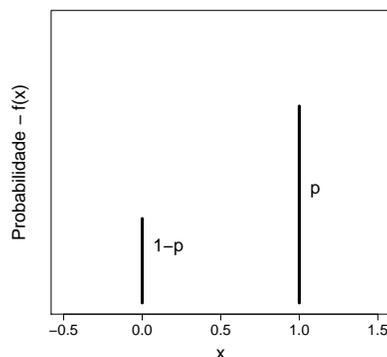


Figura 7.1. Gráfico da função densidade da distribuição Bernoulli.

$p = 0,035$, enquanto que a probabilidade dela não ter cancro $1 - p = 1 - 0,035 = 0,965$.



Uniforme Discreta

A família de distribuições uniformes discretas é a família das distribuições em que todos os valores possíveis são equiprováveis. Sua função de densidade é

$$f(x) = \begin{cases} 1/N, & x = 0, 1, \dots, N; \\ 0, & \text{nos demais caso;} \end{cases} \quad (7.2)$$

Note que o único parâmetro da distribuição é o número de elementos no conjunto domínio da função (N).

A expressão que indica que a função de densidade é nula para os casos fora do seu domínio, sendo uma ênfase da própria definição de função de densidade discreta. Ressalte-se que o domínio de toda função de densidade discreta é um subconjunto *contável* dos números reais, fora do qual seu valor é nulo. Este detalhe das funções de densidade discretas não será mais apresentado nas demais famílias discreta, devendo ficar subentendido.



Exemplo 7-8: Cubagem de Árvores

Para construção de equações de volume, foi tomada uma amostra destrutiva de 150

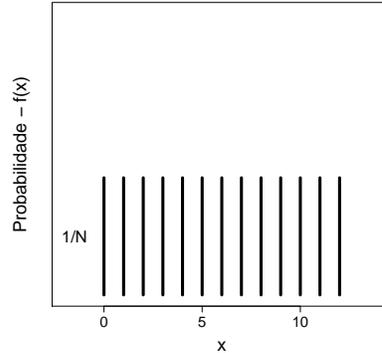


Figura 7.2: Gráfico da função densidade de uma distribuição uniforme discreta ($N = 12$).

árvores numa floresta, na qual as árvores foram abatidas, cubadas, e tiveram seus volumes determinados. Para seleção das árvores para amostra, foram definidas 10 classes de DAP (classes de 1 a 10), com 2 *cm* de amplitude, sendo os limites das classes: 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 e 30 *cm*, respectivamente. Para que a amostra destrutiva não privilegiasse nenhuma das classes de DAP, tomou-se o mesmo número de árvores em cada classe (15 árvores), gerando uma amostra com distribuição retangular, com a probabilidade de uma árvore pertencer a cada classe sendo sempre de 1/10.



Binomial

Um cenário estocástico composto de N realizações independentes de uma distribuição Bernoulli, com a mesma probabilidade de se observar um sucesso (parâmetro p), resultará na distribuição binomial. Portanto, a família de distribuições binomial define o número de sucessos em N realizações independentes de um evento estocástico com resultado binário (sucesso ou fracasso). A forma da sua função de densidade é

$$f(x) = f(x; N, p) = \binom{N}{x} p^x (1-p)^{N-x}, \quad (7.3)$$

$$N \in \mathbb{N}, 0 \leq p \leq 1; \quad x = 0, 1, \dots, N.$$

A distribuição tem dois parâmetros: o parâmetro N é o número de realizações independentes, sendo, portanto, um número natural, e o parâmetro p que é a probabilidade de sucesso. Distribuições da família binomial são apresentadas na figura 7.3.

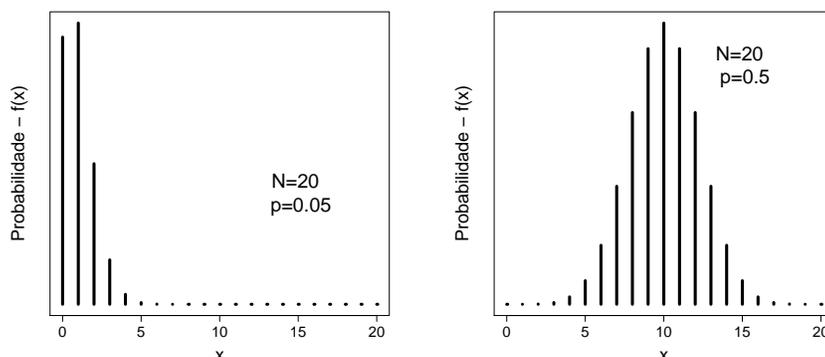


Figura 7.3. Gráficos de duas distribuições da família binomial.



Exemplo 7-9: Mortalidade de Árvores III

Considere o cenário estocástico em que a mortalidade das árvores é observada numa floresta cuja a taxa de árvores mortas é de 5%. Contudo, as árvores não são selecionadas individualmente, mas em parcelas contendo 100 árvores. O número de árvores mortas por parcela pode ser adequadamente modelado por uma distribuição binomial com densidade

$$f(x) = \binom{100}{x} (0,05)^x (1 - 0,05)^{100-x}, \quad x = 0, 1, \dots, 100,$$

A probabilidade de se observar 7 árvores mortas numa parcela é

$$f(7) = \binom{100}{7} (0,05)^7 (1 - 0,05)^{100-7} = 0,1060255 \approx 11\%.$$

Já a probabilidade de se observar uma parcela sem nenhuma árvore morta é

$$f(0) = \binom{100}{0} (0,05)^0 (1 - 0,05)^{100-0} = 0,005920529 \approx 0,6\%.$$



 *Exemplo 7-10: O Curioso Jogo das Moedas*

Dois amigos, utilizando uma moeda justa, criaram um jogo que consiste em cada um escolher um lado da moeda, cara (C) ou coroa (K), e manter fixa a escolha. A cada lançamento da moeda, o perdedor para R\$ 1,00 ao vencedor. Mas eles se surpreenderam ao perceber que raramente o jogo terminava empatado após uma sequência de 10 lançamentos. Qual a probabilidade deste jogo terminar empatado após 10 lançamentos? O modelo adequado para este jogo é a distribuição binomial e a probabilidade é

$$f(5) = \binom{10}{5} (0,5)^5 (1 - 0,5)^{10-5} = 0,2460938 \approx 25\%.$$

Logo a probabilidade do jogo *não* terminar empatado é

$$P(X \neq 5) = 1 - P(X = 5) = 1 - f(5) = 1 - 0,2460938 \approx 75\%.$$

Mas será que a probabilidade de empate cresce com o aumento do número de lançamentos? A tabela abaixo mostra um resultado curioso.

Número de Lançamentos	Probabilidade (%)	
	Empate	Não-empate
10^2	8	92
10^3	3	97
10^4	1	99
10^6	0,1	99,9



Poisson

A família de distribuições Poisson está associada a cenários estocásticos em que são enumeradas as ocorrências de um dado evento num certo intervalo de tempo ou numa certa região do espaço, sendo que a ocorrência dos eventos é completamente independente. A forma funcional da densidade para este caso é

$$f(x) = f(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad \lambda > 0; \quad x = 0, 1, \dots \quad (7.4)$$

Note que o domínio da distribuição é um subconjunto *infinito contável*, isto é, não se estabelece um limite para o máximo de ocorrências que podem ser enumeradas.

A distribuição tem apenas um parâmetro (λ), que está associado à frequência média com que o evento ocorre. A figura 7.4 apresenta duas distribuições da família Poisson.

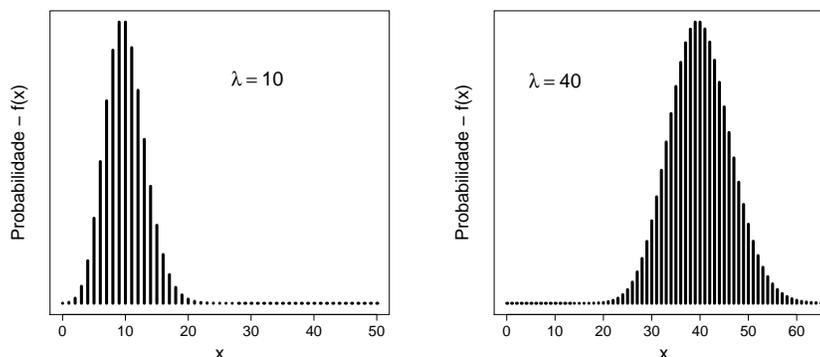


Figura 7.4. Gráficos da função densidade de duas distribuições Poisson.



Exemplo 7-11: Árvores de Jatobá na Mata

Numa mata, a densidade de árvores de jatobá (*Hymenea courbaril*) é de 0,5 árvores por hectare. Qual a probabilidade de se encontrar duas árvores num hectare? Assumindo que as árvores de jatobá tem uma ocorrência espacialmente completamente aleatória, a distribuição Poisson é um bom modelo para este cenário estocástico. Assim, a probabilidade de se encontrar duas árvores de jatobá em um hectare é

$$f(2) = \frac{e^{-0,5} 0,5^2}{2!} = 0,07581633 \approx 8\%.$$

Qual a probabilidade de se encontrar pelo menos uma árvore de jatobá em um hectare de mata?

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) = 1 - f(0) \\ &= 1 - \frac{e^{-0,5} 0,5^0}{0!} = 1 - 0,6065307 \approx 39\%. \end{aligned}$$



 *Exemplo 7-12: Plântulas de Regeneração Natural*

Numa floresta, a densidade de plântulas é de $5,5 \text{ m}^{-2}$. Assumindo que a ocorrência das plântulas na floresta é completamente aleatória no espaço, qual a probabilidade de numa parcela de 1 m^2 se observar 10 plântulas? Tomando a distribuição Poisson como modelo para este cenário estocástico, a probabilidade é

$$f(10) = \frac{e^{-5,5} 5,5^{10}}{10!} = 0,02852622 \approx 3\%.$$

Já a probabilidade de um parcela de 1 m^2 não ter nenhuma plântula é

$$f(0) = \frac{e^{-5,5} 5,5^0}{0!} = 0,004086771 \approx 0,4\%.$$



Geométrica

Como a binomial, a família de distribuições geométricas pode ser entendida como modelo de um cenário estocástico que corresponde a uma série de observações independentes, onde o parâmetro p indica a probabilidade do evento esperado ocorrer em cada observação. Mas no caso da família geométrica, o interesse está no número de observações necessárias até o primeiro sucesso, isto é, até a primeira ocorrência do evento esperado. Sua função de densidade é

$$f(x) = f(x; p) = p(1-p)^x, \quad 0 \leq p \leq 1; \quad x = 0, 1, \dots \quad (7.5)$$

A figura 7.5 apresenta dois exemplos de distribuição geométrica.

 *Exemplo 7-13: Cobaias em Laboratório*

Dois laboratórios decidiram testar a toxidez de um produto cuja taxa de mortalidade é de 50%. O laboratório A tomou um grupo de 20 cobaias e aplicou o produto. Qual a probabilidade deste laboratório observar 6 mortes? Este é um exemplo da distribuição binomial com parâmetros $n = 20$ e $p = 0,5$, logo a probabilidade é

$$f(6) = \binom{6}{20} (0,5)^6 (1-0,5)^{20-6} = 0.0369443 \approx 3,6\%.$$

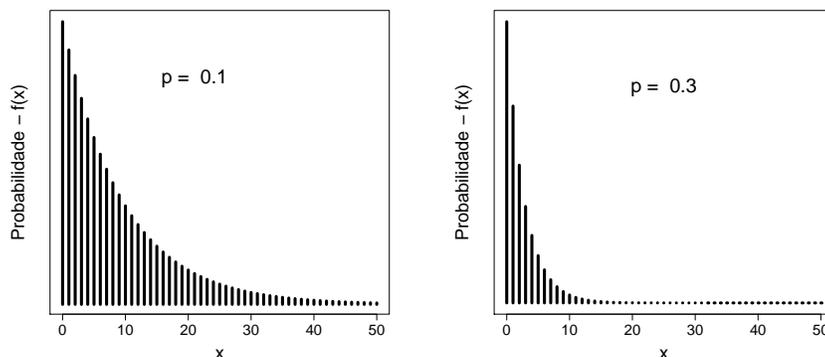


Figura 7.5. Gráficos da função densidade de duas distribuições geométricas.

Já o laboratório B, com carência de cobaias, decidiu aplicar o produto à medida que as cobaias ficasse disponíveis para o teste. Qual a probabilidade do laboratório B aplicar o produto em 6 cobaias, até que apareça a primeira cobaia morta? Neste caso, o modelo estocástico apropriado é a distribuição geométrica com parâmetro $p = 0,5$, assim a probabilidade é

$$f(6) = (0,5)(1 - 0,5)^6 = 0,0078125 \approx 0,8\%.$$



Binomial Negativa

A família de distribuições binomiais negativas pode ser vista como uma generalização da geométrica. O cenário estocástico também o composto de uma série de observações independentes, em que a probabilidade de sucesso de cada observação é dada pelo parâmetro p . Mas na binomial negativa o interesse está no número de observações que *falharão*, isto é, o número de *insucessos*, até que se observe r sucessos. Ou seja, a distribuição geométrica é uma distribuição binomial negativa com parâmetro $r = 1$.

A função da densidade da família binomial negativa é

$$f(x) = f(x; r, p) = \binom{r+x-1}{x} p^r (1-p)^x, \quad (7.6)$$

$$0 \leq p \leq 1, r = 1, 2, \dots; \quad x = 0, 1, \dots$$

Tomando-se $r = 1$ a densidade da família binomial negativa é reduzida à densidade da família geométrica.

A figura 7.6 apresenta dois exemplos de distribuição binomial negativa.

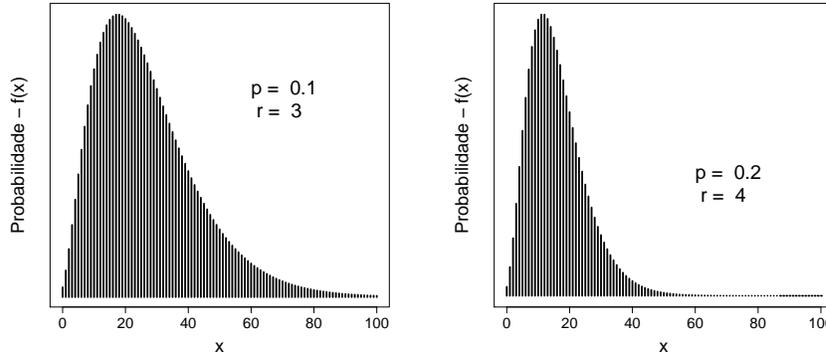


Figura 7.6. Gráficos da densidade de duas distribuições binomiais negativas.

O termo binomial da função de densidade pode ser expresso de uma outra forma, como a razão entre duas funções gama. A função gama é definida pela seguinte integral

$$\Gamma(c) = \int_0^{\infty} \exp(-u) u^{c-1} du, \quad c \in \mathbb{R}.$$

Quando a constante c for um número inteiro, então

$$\Gamma(c) = (c - 1) \Gamma(c - 1).$$

Logo, $\Gamma(c) = (c - 1)!$ e, conseqüentemente, a função gama generaliza o conceito de *fatorial* para números reais não inteiros. Assim, substituindo-se o termo binomial, a função de densidade pode ser apresentada da seguinte forma

$$f(x) = f(x; r, p) = \frac{\Gamma(r + x)}{\Gamma(r) x!} p^r (1 - p)^x.$$

Mas a distribuição binomial negativa aceita transformações mais radicais. É possível redefinir os seus parâmetros e apresentar sua densidade numa forma diferente. As diferentes formas de apresentar a função de densidade de uma família de distribuições, com parâmetros diferentes, são chamadas de “*parametrizações*”. Uma parametrização alternativa para binomial negativa é a seguinte

$$f(x) = f(x; k, \mu) = \frac{\Gamma(k + x)}{\Gamma(k) x!} \left(\frac{k}{k + \mu} \right)^k \left(\frac{\mu}{k + \mu} \right)^x \quad (7.7)$$

$$k > 0, \mu > 0; \quad x = 0, 1, \dots$$

Nesta parametrização, os parâmetros k e μ não são necessariamente números inteiros como o parâmetro r da primeira parametrização.

Por se tratar da mesma função de densidade, sempre existe uma relação matemática entre as duas parametrizações. Assim, os parâmetros de uma podem ser expresso como função dos parâmetros da outra, e vice-versa:

$$k = r \quad \text{e} \quad \mu = \frac{r(1-p)}{p} \quad \iff \quad r = k \quad \text{e} \quad p = \frac{\mu + k}{\mu}.$$

A segunda forma de parametrização é bastante utilizada em estudos ecológicos em que a distribuição binomial negativa é utilizada para modelar a enumeração de eventos que não ocorrem de forma uniforme, seja no tempo, seja no espaço, resultando numa certa agregação das observações. Neste caso, o parâmetro μ é interpretado como o número médio da contagem (das enumerações), enquanto que k é como parâmetro de “dispersão”.



Exemplo 7-14: Cobaias em Laboratório II

Dois laboratórios decidiram testar a toxidez de um produto cuja taxa de mortalidade é de 50%. O laboratório A tomou um grupo de 20 cobaias e aplicou o produto. Qual a probabilidade deste laboratório observar 6 mortes? Como visto acima, este é um exemplo da distribuição binomial com parâmetros $n = 20$ e $p = 0,5$ e a probabilidade é 3,6%.

O laboratório C, com carência de cobaias, decidiu aplicar o produto à medida que as cobaias ficasse disponíveis para o teste, e decidiu que o teste seria suspenso quando a sexta cobaia morresse. Qual a probabilidade do laboratório C aplicar o produto em 20 cobaias? Neste caso, o modelo estocástico apropriado é a distribuição binomial negativa com parâmetros $r = 6$ e $p = 0,5$. Aplicar o produto em 20 cobaias, havendo 6 mortes, implica a sobrevivência de 14 cobaias, assim a probabilidade é

$$f(14) = \binom{6+14-1}{14} p^6 (1-p)^{14} = 0,01108932 \approx 1,1\%.$$



 *Exemplo 7-15: Árvores de Jatobá na Mata II*

Numa dada mata, a densidade de árvores de jatobá (*Hymenea courbaril*) é de 0,5 árvores por hectare (μ), mas a distribuição das árvores na mata é agregada com parâmetro de dispersão (k) igual a 4. Qual a probabilidade de se encontrar duas árvores num hectare? Neste caso, utiliza-se a segunda parametrização da família binomial negativa:

$$f(2) = = \frac{\Gamma(4+2)}{\Gamma(4) 2!} \left(\frac{4}{4+0,5}\right)^4 \left(\frac{0,5}{4+0,5}\right)^2 = 0,07707347 \approx 7,7\%.$$

Qual a probabilidade de se encontrar pelo menos uma árvore de jatobá em um hectare de mata?

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) = 1 - f(0) \\ &= 1 - \frac{\Gamma(4+0)}{\Gamma(4) 0!} \left(\frac{4}{4+0,5}\right)^4 \left(\frac{0,5}{4+0,5}\right)^0 \\ &= 1 - 0,6242951 = 0,3757049 \approx 36\%. \end{aligned}$$



7.4 Famílias de Distribuições Contínuas

No caso das distribuições contínuas, a probabilidade é calculada para intervalos de números reais pela integração da função da densidade no intervalo, ou pela diferença da função de distribuição para os extremos do intervalo. Entretanto, a função de densidade ainda é mais apropriada para visualizar o comportamento da distribuição, uma vez que ela permite uma análise comparativa para os diversos valores que a distribuição pode assumir. Com o objetivo de exemplificar melhor as famílias de distribuições contínuas, serão apresentadas as funções de densidade e de distribuição para algumas famílias de uso frequente nas Ciências Florestais e em Ecologia.

Retangular ou Uniforme

A família das distribuições retangulares são aquelas em que todos os valores numéricos num intervalo dos números reais são equiprováveis. Suas funções de

densidade e de distribuição são

$$f(x) = f(x; a, b) = \frac{1}{a - b} \quad (7.8)$$

$$F(x) = F(x; a, b) = \frac{x - a}{a - b}. \quad (7.9)$$

Ambas funções definidas no intervalo $a \leq x \leq b$, sendo que os únicos parâmetros da família são os limites inferior (a) e superior (b) deste intervalo. Os gráficos de uma distribuição retangular são apresentados na figura 7.7.

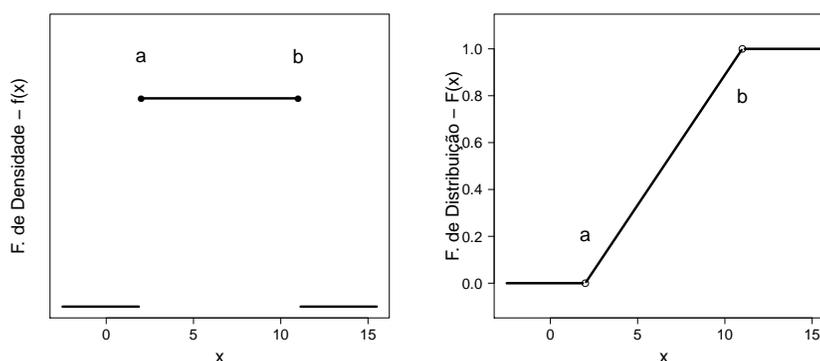


Figura 7.7: Função de densidade ($f(x)$) e função de distribuição ($F(x)$) de uma distribuição retangular com parâmetros $a = 2$ e $b = 11$.

Uma distribuição específica da família da distribuição retangular, que é chamada de distribuição *uniforme*, é quando as observações são definidas no intervalo $[0, 1]$. Neste caso, as funções de densidade e de distribuição se tornam muito simplificadas

$$f(x) = 1 \quad \text{e} \quad F(x) = x. \quad (7.10)$$

Para se obter uma amostra de observações independentes da distribuição uniforme, basta selecionar valores do intervalo $[0, 1]$ de modo equiprovável. A partir de uma amostra da distribuição uniforme é possível se gerar uma amostra aleatória de qualquer outra distribuição.

 *Exemplo 7-16: Localização Aleatória das Árvores*

Em algumas florestas, como na Floresta Estacional Semi-decidual, a localização das árvores na floresta é aleatória, de forma que o padrão espacial é completamente aleatório. Neste caso, num parcela quadrada de 100 m de lado (1 hectare) as coordenadas de localização das árvores seguem a distribuição retangular no intervalo $[0, 100]$. A figura 7.8 exemplifica este fenômeno.

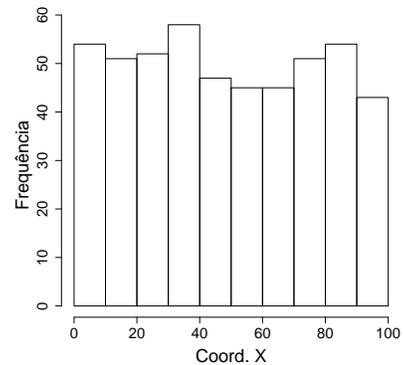
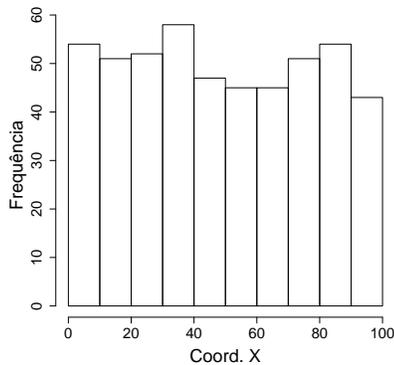
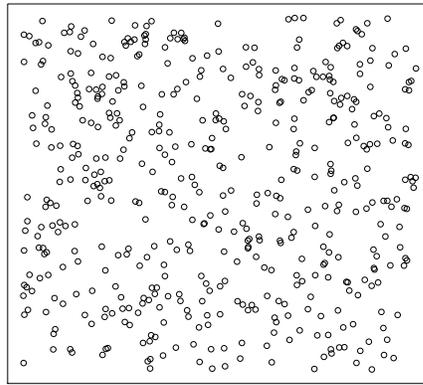


Figura 7.8: Exemplo de padrão espacial completamente aleatório e histogramas das coordenadas cartesianas (x, y) da localização das árvores na parcela quadrada de 100 m de lado.



Exponencial

A distribuição exponencial é um modelo para um cenário estocástico com observação de variáveis contínuas análogo ao modelo da distribuição Poisson para variáveis discretas. Suas funções de densidade e distribuição são

$$f(x) = f(x; \lambda) = \lambda \exp(-\lambda x) \quad (7.11)$$

$$F(x) = F(x; \lambda) = 1 - \exp(-\lambda x) \quad (7.12)$$

sendo $\lambda > 0$ e $x \geq 0$. Os gráficos da distribuição exponencial são apresentados na figura 7.9

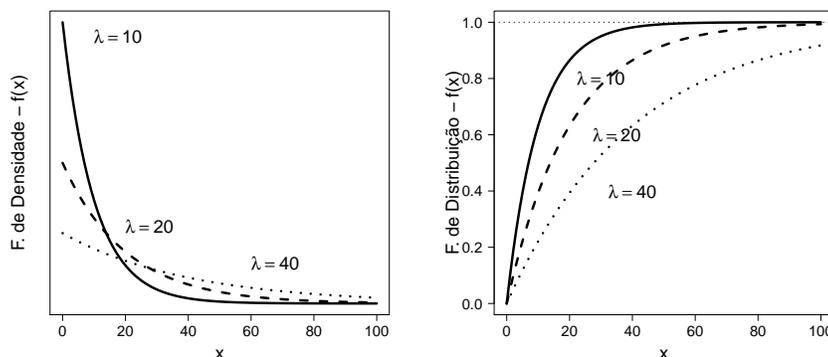


Figura 7.9: Gráficos da função densidade ($f(x)$) e da função de distribuição ($F(x)$) de três distribuições exponenciais.

Na família exponencial, a probabilidade de se observar os resultados decresce com o aumento do valor do resultado, por isto, o seu parâmetro λ é frequentemente designado por “taxa”, pois ele define a taxa de redução da probabilidade. Logo, a redução da probabilidade é constante ao longo de toda curva, sendo esta uma propriedade básica da família exponencial.

A família exponencial também pode ser parametrizada transformando o parâmetro de taxa (λ) num parâmetro de escala (β): $\beta = 1/\lambda$, de modo que a sua função de densidade se torna:

$$f(x) = f(x; \beta) = \frac{1}{\beta} \exp(-x/\beta). \quad (7.13)$$

Esta parametrização torna a família exponencial mais semelhante à parametrização tradicional da família Weibull, da qual a exponencial é um caso particular.

 *Exemplo 7-17: “Lei” de De Liocourt*

A “lei” de De Liocourt (de Liocourt, 2001; Rubin, Manion e Faber-Langendoen, 2006) estabelece que uma floresta inequidiana está *balanceada* se há uma razão constante entre o número de árvores em classes de diâmetro de igual amplitude sucessivas, ou seja,

$$\frac{N_1}{N_2} = \frac{N_2}{N_3} = \dots = \frac{N_{k-1}}{N_k} = \dots = \frac{N_{n-1}}{N_n} = q$$

send N_k o número de árvores na $k^{\text{ésima}}$ classe e q a razão constante. A implicação da razão constante é que o número de árvores por classe de diâmetro segue uma série geométrica decrescente:

$$N_k = N_1 q^{-(k-1)}.$$

Considerando que as classes de diâmetro tem amplitude constante w , cujos limites superiores são $(l_1, l_2, \dots, l_k, \dots, l_n)$, então $l_k = k w$. Se a distribuição de diâmetros da floresta pode ser modelada pela função de distribuição $F(\cdot)$, a frequência relativa do número de árvores numa dada classe de diâmetros é dada por

$$\frac{N_k}{N} = F(l_k) - F(l_{k-1}) = F(k w) - F((k-1) w),$$

para $N = N_1 + N_2 + \dots + N_n$, que é o número total de árvores no arvoredo. Assumindo que a distribuição é exponencial, a frequência relativa da $k^{\text{ésima}}$ classe é

$$\begin{aligned} \frac{N_k}{N} &= (1 - \exp[-\lambda k w]) - (1 - \exp[-\lambda (k-1) w]) \\ &= (1 - \exp[-\lambda w]) \exp[-\lambda (k-1) w]. \end{aligned}$$

Na expressão acima, o termo entre parênteses é a frequência relativa da primeira classe de diâmetro, logo, o número de árvores na $k^{\text{ésima}}$ classe é dado por

$$\begin{aligned} N_k &= N (1 - \exp[-\lambda w]) \exp[-\lambda (k-1) w] \\ &= N_1 \exp[-\lambda (k-1) w]. \end{aligned}$$

Então, a razão q da frequências de classes sucessivas pode ser expressa em função do parâmetro da exponencial

$$\begin{aligned} q &= \frac{N_{k-1}}{N_k} = \frac{N_1 \exp[-\lambda (k-2) w]}{N_1 \exp[-\lambda (k-1) w]} = \exp[\lambda w] \\ q &= \exp[\lambda w] \iff \lambda = \frac{\ln(q)}{w}. \end{aligned}$$

Assim, a distribuição balanceada de “De Liocourt” é uma reparametrização da família exponencial definida em termos da razão de classes sucessivas q e da amplitude de classe w . Logo, a função de distribuição da pode ser apresentada na forma

$$F(x) = 1 - \exp[-\lambda x] = 1 - q^{-x/w}, \quad x \geq 0, \quad q > 1, w > 0.$$



Weibull

A família Weibull pode ser considerada uma generalização da família exponencial. Suas funções de densidade e distribuição são

$$f(x) = f(x; \alpha, \beta, \gamma) = \left(\frac{\gamma}{\beta}\right) \left(\frac{x - \alpha}{\beta}\right)^{\gamma-1} \exp\left[-\left(\frac{x - \alpha}{\beta}\right)^\gamma\right] \quad (7.14)$$

$$F(x) = F(x; \alpha, \beta, \gamma) = 1 - \exp\left[-\left(\frac{x - \alpha}{\beta}\right)^\gamma\right] \quad (7.15)$$

sendo $\alpha, \beta, \gamma > 0$ e $x \geq \alpha$. A família Weibull com parâmetros $\alpha = 0$ e $\gamma = 1$ fica reduzida à família exponencial com parâmetro $\lambda = 1/\beta$.

Em famílias de distribuições com vários parâmetros, como a Weibull, é comum se designar cada parâmetro por sua influência sobre o comportamento da distribuição. Assim, o parâmetro α é chamado de parâmetro de “localização”, pois ele influencia a posição a partir da qual a distribuição se inicia. O parâmetro β é dito parâmetro de “escala”, pois influencia o “espalhamento” ou dispersão da distribuição, enquanto que o parâmetro γ é chamado de parâmetro de “forma”, uma vez que controla a forma da curva de densidade. A distribuição com parâmetros de localização, de escala e de forma é normalmente chamada como *Weibull três parâmetros*. Uma forma mais simples de trabalho com a Weibull é aquela com dois parâmetros, em que o parâmetro de localização é dispensado ($\alpha = 0$).

Famílias de distribuições como a Weibull que possuem um parâmetro de forma apresentam grande flexibilidade para representar diferentes cenários estocásticos. É importante notar que a influência dos parâmetros sobre a distribuição não é “independente”, isto é, há uma interação entre os parâmetros de modo que curvas de distribuições bem próximas podem resultar de combinações de parâmetro com valores bastante distintos. Os gráficos de algumas distribuições da família Weibull são apresentados na figura 7.10

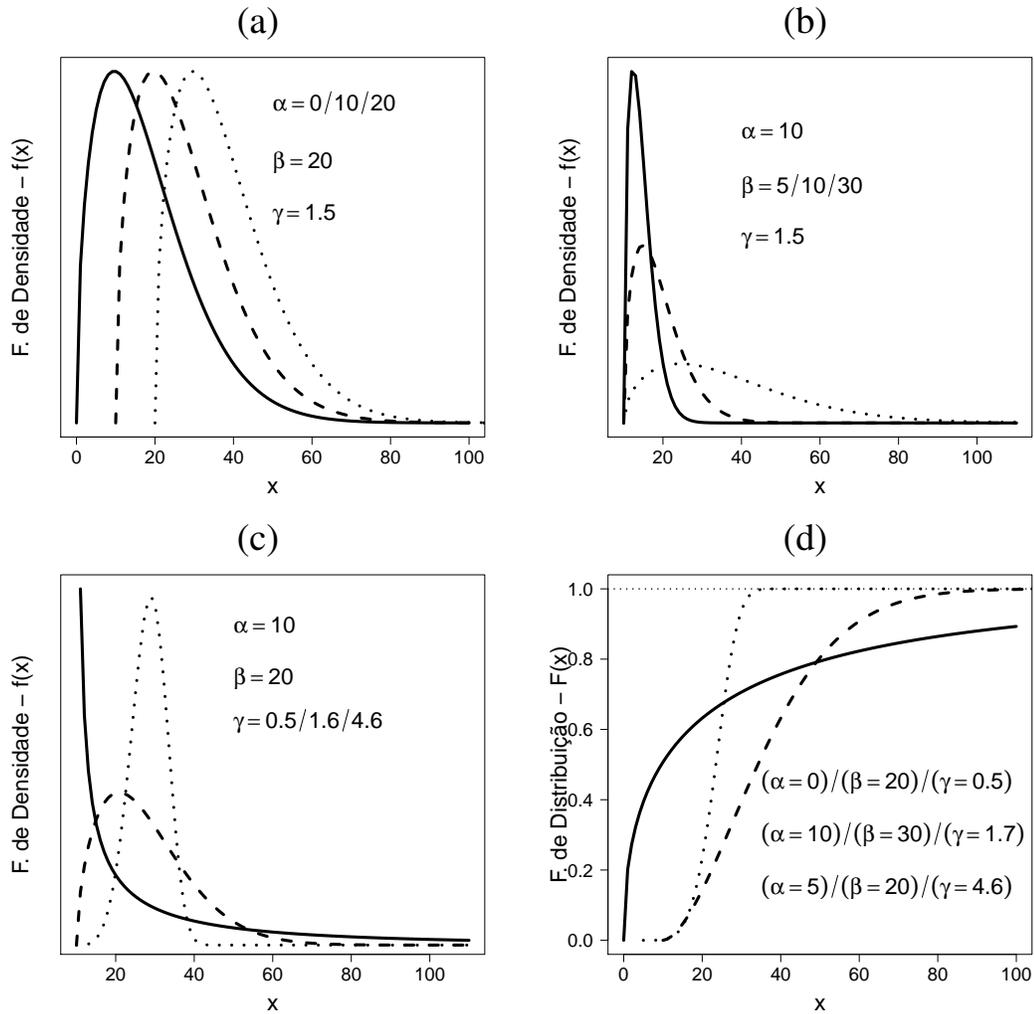


Figura 7.10: Gráficos da função densidade (a-c) e da função de distribuição (d) de distribuições da família Weibull mostrando a influência dos parâmetros sobre o comportamento das distribuições.

Weibull (1951) propôs esta família de distribuições como uma aplicação na área de resistência de materiais. Por exemplo, a Weibull é um bom modelo para distribuição do tempo de falha de cada elo de uma corrente, sabendo-se que a probabilidade da corrente falhar num dado momento depende da probabilidade de qualquer um dos elos falhar. Mas esse autor argumentou e exemplificou que a família Weibull pode ser aplicada a uma ampla classe de problemas em que a ocorrência de um evento, em qualquer parte de um objeto ou sistema, pode ser dita como tendo ocorrido no objeto ou sistema como um todo, como no exemplo da corrente.

Nas aplicações de resistência de materiais, a família Weibull surge da família exponencial, quando o tempo de falha é elevado a uma dada potência, isto é, o tempo sofre uma transformação de escala (Shapiro e Gross, 1981). Se t é o tempo de falha de um material e segue uma distribuição exponencial com parâmetro λ , então, $z = t^\gamma$ seguirá uma distribuição Weibull com parâmetros $\beta = 1/\lambda$ e γ .

Pinder e colaboradores (1978) propôs a aplicação da família Weibull no estudo de curva de sobrevivência de populações biológicas, mostrando que o parâmetros de forma permite controlar a taxa de mortalidade específica da idade e, portanto, a forma da curva de sobrevivência. Na década de 1970, a família Weibull começou a ser aplicada na modelagem da distribuição de diâmetros de florestas de coníferas (Bailey e Dell, 1973; Dell, 1973; Schreuder e Swank, 1974; Ek, Issos e Bailey, 1975). Na década de 1980, ela foi aplicada a florestas plantadas de *Pinus* no Brasil (Couto, 1980) e a espécies arbóreas de florestas tropicais (Batista, 1989).



Exemplo 7-18: Estrutura de Tamanho de Espécies Arbóreas

Batista (1989) estudando a estrutura de tamanho de populações de espécies arbóreas em floresta nativa no município de Bom Jardim, Maranhão, utilizou a família Weibull para modelar as distribuições do DAP de 60 espécies.

O DAP mínimo de medição foi de 14 *cm*, de modo que o parâmetro de localização para todas as espécies foi considerado conhecido. Já os parâmetros de escala e forma variaram entre as espécies. Mas foi possível associar os valores dos parâmetros com a classificação das espécies arbóreas em três grupos da sucessão ecológica: início, meio e final de sucessão (figura 7.11).

Um quarto grupo, composto de apenas duas espécies, apresentou uma estrutura de tamanho que sugere que a regeneração seja intermitente.

Os gráficos mostram que se o modelo de De Liocourt (família exponencial) pode ser um modelo adequado para as florestas nativas em geral, ele não é adequado para as espécies arbóreas que compõe a floresta tropical nativa, pois poucas espécies tiveram o parâmetro da forma (γ) próximo da unidade.



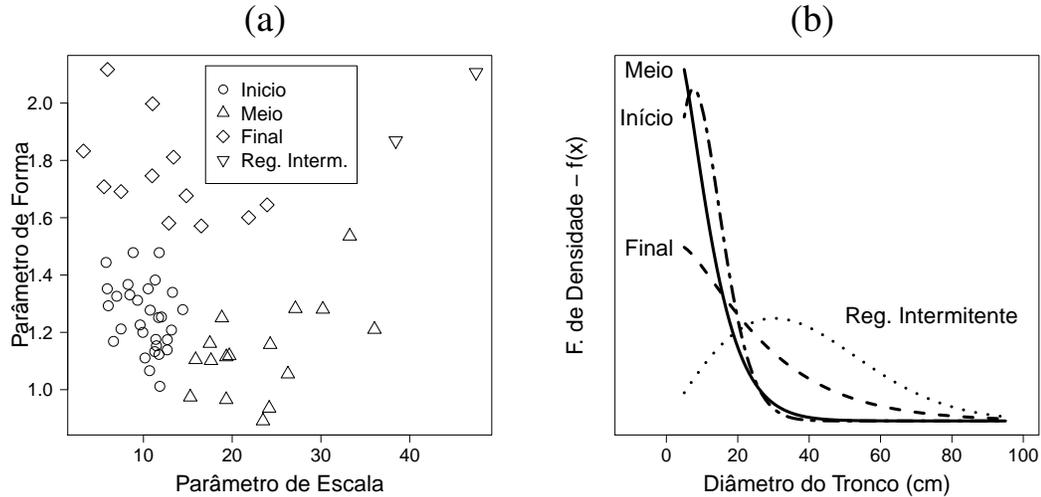


Figura 7.11: (a) Variação dos parâmetros de escala e forma para as 60 espécies arbóreas indicando os grupos da sucessão ecológica. (b) Curvas da função de densidade para os valores médios de cada grupo.

Gama

A família gama também generaliza a família exponencial, mas de uma forma distinta da família Weibull. Sua função de densidade é

$$f(x) = f(x; \beta, \gamma) = \frac{(x/\beta)^{\gamma-1} \exp[-(x/\beta)]}{\beta \Gamma(\gamma)} \quad (7.16)$$

sendo $\beta, \gamma > 0$ e $x \geq 0$. A família gama não possui uma apresentação em “*forma fechada*” para sua função de distribuição, isto é, ela só pode ser apresentada na forma de uma integral da função de densidade

$$F(x) = \frac{1}{\beta \Gamma(\gamma)} \int_0^x (t/\beta)^{\gamma-1} \exp[-(t/\beta)] dt$$

Somente quando o parâmetro da forma (γ) é definido como número inteiro é que a apresentação em forma fechada se torna possível. Exemplos de distribuição da família gama são apresentados na figura 7.12

A família exponencial é um caso particular da família gama quando o parâmetro da forma é unitário ($\gamma = 1$). Logo, assim como a família Weibull, a família gama também pode ser explicada a partir da família exponencial em termos de resistência de materiais. Se um mecanismo é composto de diversos componentes, sendo que o tempo de falha de cada componente segue uma distribuição exponencial, então o tempo de falha do mecanismo seguirá uma distribuição gama. De

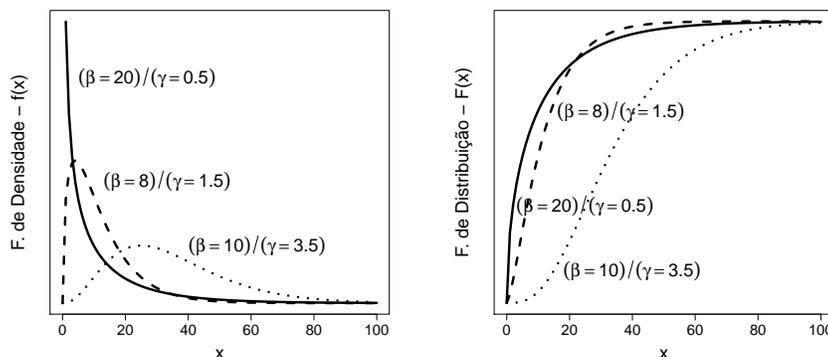


Figura 7.12: Gráficos da função densidade ($f(x)$) e da função de distribuição ($F(x)$) de algumas distribuições gama.

modo análogo à resistência de materiais, a família gama também é utilizada na área biomédica para modelar a falha de sistemas multi-orgãos, como o pulmão e os rins (Shapiro e Gross, 1981). Embora de utilização menos extensiva que a família Weibull, a família gama também tem sido utilizada para modelar a distribuição dos diâmetros das árvores de arvoredos (Bailey, 1980).

Uma parametrização que torna a família gama mais semelhante à parametrização tradicional da família exponencial é a transformação do parâmetro de escala (β) em parâmetro de taxa (λ): $\lambda = 1/\beta$. Nesta parametrização, a função de densidade se torna

$$f(x) = f(x; \lambda, \gamma) = \frac{\lambda^\gamma x^{\gamma-1} \exp[-(\lambda x)]}{\Gamma(\gamma)}. \quad (7.17)$$

Caso o parâmetro da forma for unitário ($\gamma = 1$), a densidade da família gamma (equação 7.17) é reduzida à densidade da família exponencial (equação 7.11).



Exemplo 7-19: Estrutura de Tamanho de Floresta Plantada

A forma flexível da família gama a torna um modelo apropriado para modelar a estrutura de tamanho de florestas plantadas com diferentes idades. À medida que uma floresta plantada se desenvolve, a sua estrutura de tamanho muda gradativamente de uma distribuição composta de árvores com pequenos diâmetros para uma floresta com uma ampla variação de tamanho. Neste processo, ocorre uma alteração na assimetria da distribuição.

A figura 7.13 apresenta a variação na estrutura de tamanho das árvore de uma

floresta plantada de *Eucalyptus grandis* em diferentes idades. À medida que a floresta se torna mais velha, a distribuição do DAP se torna mais simétrica. Nessa figura, a estrutura de tamanho em todas idades é modelada por distribuições gama.

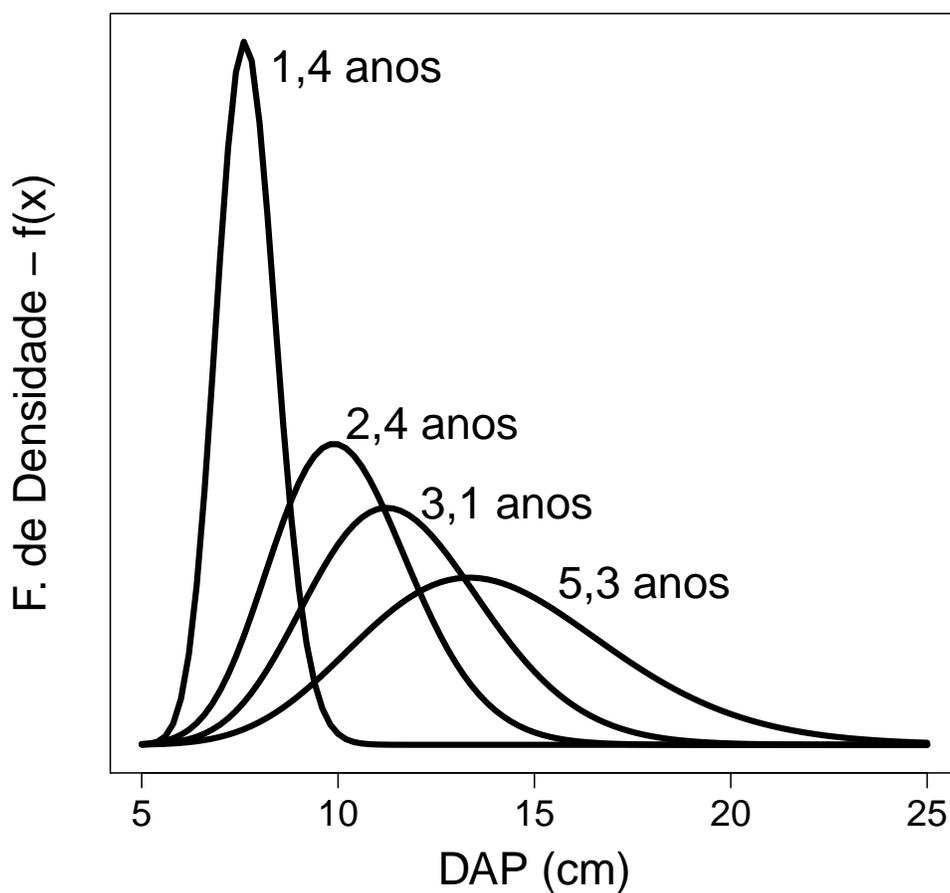


Figura 7.13: Distribuições gama como modelo das distribuições de diâmetros de árvores de uma mesma floresta plantada de *Eucalyptus grandis* em diferentes idades.



Gaussiana¹

A família de distribuições Gaussianas tem a seguinte função de densidade

$$f(x) = f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right] \quad (7.18)$$

$$-\infty \leq \mu \leq \infty, \sigma > 0; \quad -\infty \leq x \leq \infty.$$

O parâmetro μ é chamado de *média*, sendo um parâmetro de localização, no caso a localização central da distribuição. Já o parâmetro σ é chamado de *desvio padrão*, sendo o parâmetro de escala, isto é, controla a dispersão dos valores ao redor da média μ . Também é comum designar o parâmetro de escala por σ^2 , chamado de *variância*.

A família Gaussiana não tem um parâmetro que controle a forma, logo, as distribuições gaussianas têm sempre a mesma forma fixa, tradicionalmente designada como “*forma de sino*” (*bell shape*). Um outro aspecto particular da família Gaussiana é que os seus parâmetros, média e desvio padrão, atuam de modo independente sobre o comportamento das distribuições (figura 7.14).

Gaussiana Padronizada

Assim como a família gama, da família Gaussiana não possui uma forma fechada para a função de distribuição. Isto implica que, no cálculo de probabilidades, a integração da função de densidade nunca resulta numa expressão matemática explícita, sendo necessário métodos numéricos² para obter as probabilidades.

Para facilitar este processo, costumava-se tabelar os valores de probabilidade para uma única distribuição da família Gaussiana chamada de *Distribuição Gaussiana Padronizada*. Com o desenvolvimento dos computadores eletrônicos, este procedimento se tornou desnecessário, mas o procedimento para se obter a distribuição padronizada continuou a ser designado por *padronização* de uma variável.

¹A distribuição Normal ou distribuição Gaussiana é uma distribuição central para a inferência estatística. Seu nome consagrado é “distribuição normal”, mas o “normal” significa que é uma distribuição que indica a “norma” de comportamento. Esta concepção surgiu com a dedução da própria distribuição como modelo para o comportamento dos erros de medições astronômicas e geodésicas. Embora muitos autores tenham contribuído na sua definição, a designação “Gaussiana” é uma homenagem a Carl Friederich Gauss (1777-1855) que desenvolveu a teoria de probabilidades associada a ela e também fez grande uso dela em aplicações práticas. Considerando que várias pessoas contribuíram para o seu estabelecimento, é questionável homenagear apenas uma pessoa. Mas utilizaremos o nome “família Gaussiana” porque a aplicaremos de modo análogo às aplicações prática desenvolvidas por Gauss.

²Métodos numéricos são formas de se calcular de modo aproximado o valor de funções. Para efeitos práticos, o cálculo numérico permite obter o valor de uma função com erro negligenciável.

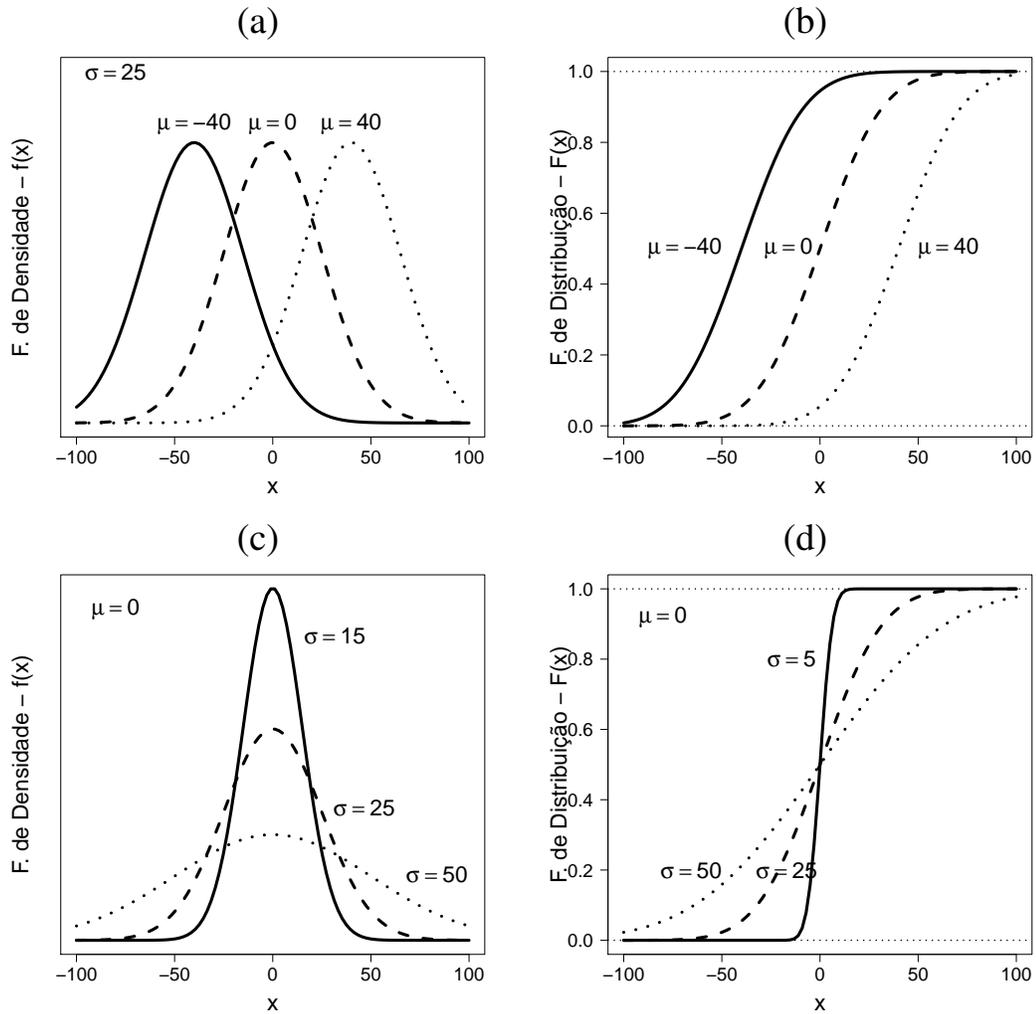


Figura 7.14: Exemplos de distribuições da família Gaussiana, mostrando-se os gráficos das funções de densidade (a,c) e das respectivas funções de distribuição (b,d). As influências da média (a,b) e do desvio padrão (c,d) se mostram totalmente independentes.

O procedimento consiste em subtrair dos valores da variável (X) a sua média (μ_X), dividindo o resultado pelo desvio padrão (σ_X):

$$Z = \frac{X - \mu_X}{\sigma_X}. \quad (7.19)$$

Se a variável pertencer à família Gaussiana, a variável resultante (Z) terá distribuição Gaussiana padronizada com média nula ($\mu_Z = 0$) e desvio padrão unitário ($\sigma_Z = 1$).

Este procedimento simplifica bastante a função de densidade da distribuição padronizada, tornando-a

$$f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}, \quad -\infty \leq x \leq \infty. \quad (7.20)$$

Dada a importância da distribuição Gaussiana padronizada nos problemas práticos de Biometria Florestal, sua função de distribuição, embora sem forma fechada, será referida como $\Phi(\cdot)$

$$\Phi(z) = \int_{-\infty}^z \frac{e^{-u^2/2}}{\sqrt{2\pi}} du.$$

O fato da média ser nula e o desvio padrão unitário implica nos valores da distribuição Gaussiana padronizada poderem ser interpretados em termos de “*distância da média em unidades de desvio padrão*”. Por exemplo, o valor $z = 2$ indica uma posição de dois desvios padrão acima da média, enquanto que o valor $z = -1,5$ indica uma posição de um desvio padrão e meio abaixo da média.

Assim, a distribuição padronizada ilustra uma propriedade importante da família Gaussiana associada à sua forma de sino. Os intervalos centrados na média, cuja largura é determinada por uma proporção do desvio padrão, definem uma probabilidade fixa. Por exemplo, no intervalo centrado na média com largura de dois desvios padrão, isto é, o intervalo que vai de $\mu - \sigma$ até $\mu + \sigma$, define a probabilidade de aproximadamente 68% (figura 7.15). O intervalo de quatro desvios padrão ($[\mu - 2\sigma; \mu + 2\sigma]$) define a probabilidade de 95%, enquanto o intervalo de seis desvios padrão ($[\mu - 3\sigma; \mu + 3\sigma]$) define a probabilidade de 99,7%.

Ressalta-se que os valores nas “*caldas*” da curva de densidade Gaussiana, isto é, os valores que distanciam mais de três desvios padrão da média, ou para além de -3 e 3 na distribuição padronizada, tem baixíssima probabilidade de ocorrência.

O procedimento de padronização pode, e frequentemente é, aplicado a qualquer variável observada ou medida. Quando a padronização é aplicada a mais de uma variável, o resultado é um conjunto de variáveis padronizadas, todas elas como média nula e desvio padrão unitário, mas não necessariamente com distribuição Gaussiana padronizada. O procedimento de padronização uniformiza a escala das variáveis, eliminando as diferenças entre as médias e transformando a escala de dispersão para uma escala medida em unidades de desvio padrão.

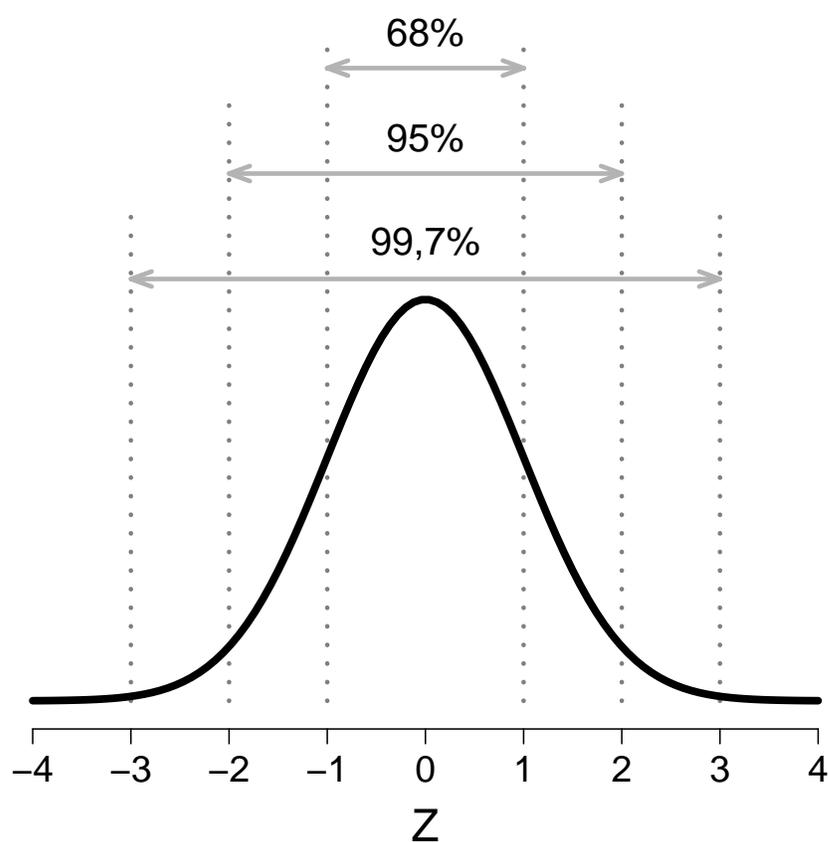


Figura 7.15: Distribuição Gaussiana Padronizada ilustrando a propriedade da família Gaussiana de probabilidades fixas para intervalos centrados na média com largura definida em proporções do desvio padrão.



Exemplo 7-20: Área Basal em Floresta Tropical

Em geral, o comportamento de medidas dos atributos das árvores não segue a família Gaussiana, mas quando as medidas das árvores individuais são agregadas para o conjunto de árvores num arvoredo, a medida resultante da agregação se aproxima do comportamento das distribuições Gaussianas. Um exemplo deste comportamento é a área basal de arvedos de floresta tropical do Maranhão (figura 7.16). A área basal do arvoredo é produzida pela soma das áreas transversais das árvores individuais e, por isto, ela tende a uma distribuição simétrica que pode ser modelada pela família Gaussiana.

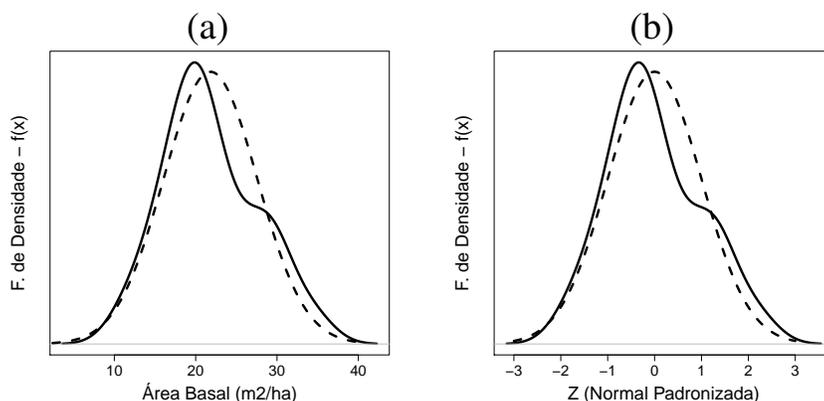


Figura 7.16: Distribuição da área basal ($m^2 ha^{-1}$) de arvedos de floresta tropical do Município de Bom Jardim, Maranhão, na escala original (a) e na forma de variável padronizada (b). A curva irregular (linha contínua) é a densidade estimada empiricamente por procedimento não paramétrico, enquanto a curva simétrica (linha tracejada) é o modelo Gaussiano.



Exemplo 7-21: Proporção de Árvores Dominadas

Mesmo os observações arborimétricas qualitativas, das quais se enumera os casos e se calcula proporções, podem se aproximar de uma distribuição Gaussiana, quando um número suficientemente grande de árvores é agregado para o cálculo da proporção. Figura 7.17 apresenta um exemplo da proporção de árvores dominadas em arvedos de floresta plantada de eucalipto *urograndis*, na região central do estado de São Paulo.

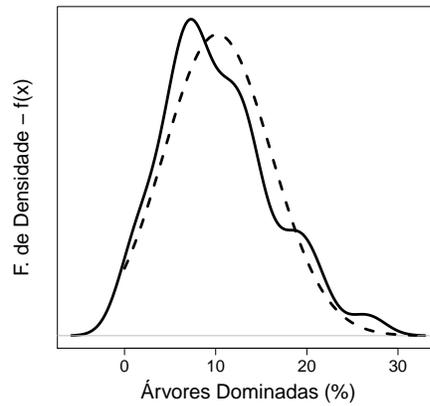


Figura 7.17: Proporção de árvores dominadas em arvoredos de floresta plantada de eucalipto *urograndis*, na região central do estado de São Paulo. A curva irregular (linha sólida) é a densidade estimada empiricamente por procedimento não paramétrico, enquanto que a curva simétrica (linha tracejada) é o modelo Gaussiano.



Log-normal

A família log-normal é a família análoga à família Gaussiana para variáveis em escala logarítmica. Ou seja, se a distribuição de probabilidades de uma dada medida (X) segue uma distribuição da família log-normal, então o logaritmo natural desta medida ($\ln(X)$) segue uma distribuição Gaussiana.

Sua função de densidade é

$$f(x) = f(x; \mu, \sigma) = \frac{1}{x \sigma \sqrt{2\pi}} \exp \left\{ -\frac{[\ln(x) - \mu]^2}{2\sigma^2} \right\} \quad (7.21)$$

$$\mu > 0, \sigma > 0; \quad x > 0.$$

e os parâmetros μ e σ correspondem à média e ao desvio padrão na escala logarítmica ($\ln(x)$).

Na família log-normal o parâmetro μ não é o parâmetro de localização, pois ele influencia a dispersão dos dados, sendo, portanto, o parâmetro de escala. Por outro lado, o parâmetro σ torna-se o parâmetro de forma, pois ao contrário da

família Gaussiana, na família log-normal as distribuições podem variar de forma. (figura 7.18).

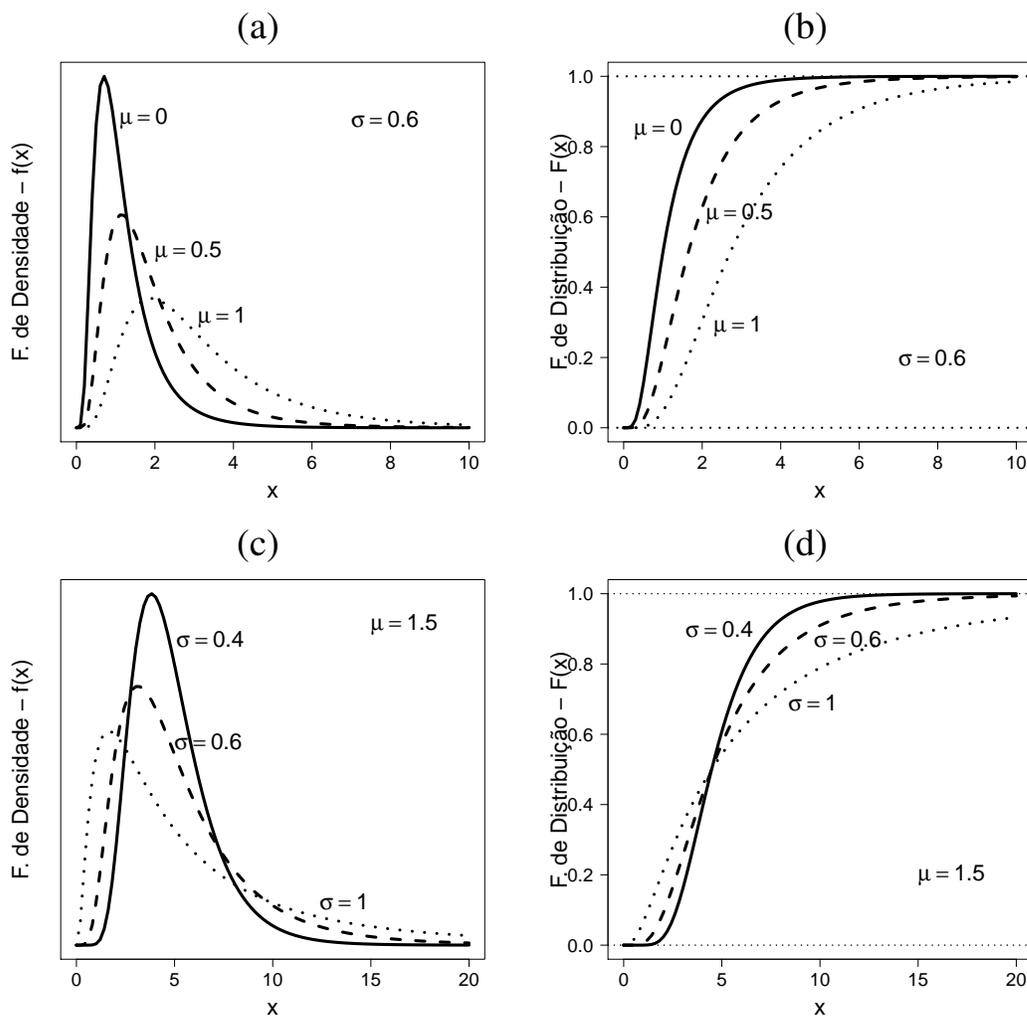


Figura 7.18: Gráficos da função densidade (a,c) e a função de distribuição (b,d) de algumas distribuições log-normais. Os gráficos (a) e (b) mostram que o parâmetro μ influencia a dispersão dos valores, enquanto os gráficos (c) e (d) mostram a influência do parâmetro σ sobre a forma.

 *Exemplo 7-22: Abundância de Espécies Arbóreas Tropicais*

A abundância de espécies arbóreas numa floresta tropical pode ser modelada pela família log-normal. A figura 7.19 apresenta a densidade das abundâncias de espécies arbóreas na forma de número de árvores e na forma do logaritmo do número de árvores. A família log-normal se apresenta como uma boa aproximação para os resultados observados.

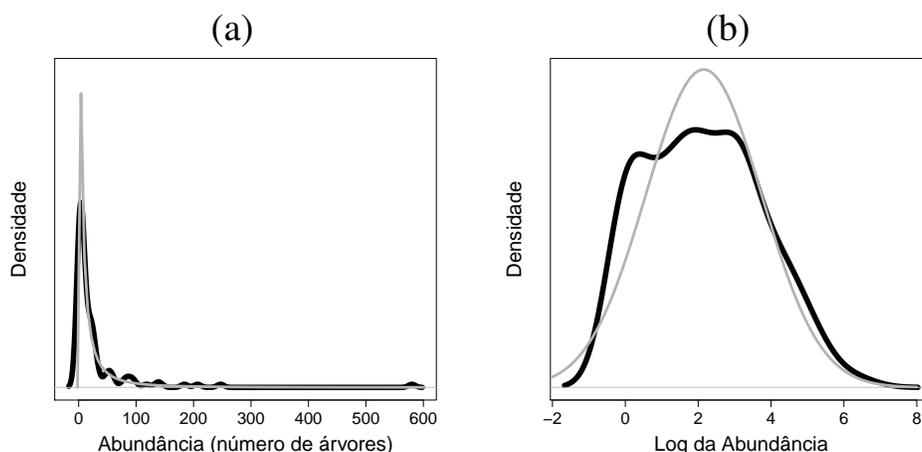


Figura 7.19: Densidade da abundância de espécies arbóreas numa floresta tropical na escala de número de árvores (a) e na escala de logaritmo do número de árvores (b). A curva irregular é a densidade estimada empiricamente por procedimento não paramétrico, enquanto que a curva regular é o modelo log-normal, o qual se mostra simétrico na escala logarítmica (gráfico (b)).



 *Exemplo 7-23: Volume de Árvores Individuais*

O volume de madeira de árvores individuais é uma variável que frequentemente apresenta uma distribuição fortemente assimétrica que pode ser adequadamente modelada pela família log-normal. A figura 7.20 apresenta o volume de árvores de *Eucalyptus saligna* de florestas plantadas na região central do estado de São Paulo. A família log-normal se mostra muito próxima ao comportamento do volume das árvores individuais.

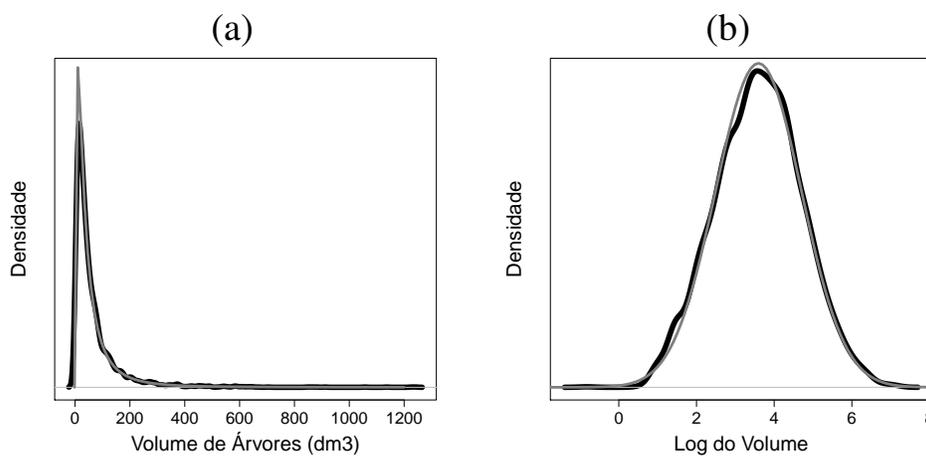


Figura 7.20: Densidade do volume de madeira de árvores individuais de *Eucalyptus saligna* de florestas plantadas na região central do estado de São Paulo. Em (a) o volume é apresentado na sua escala original (dm^3) e em (b) na escala logarítmica. A curva irregular é a densidade estimada empiricamente por procedimento não paramétrico, enquanto que a curva regular (simétrica no gráfico (b)) é o modelo log-normal.



7.5 Propriedades das Distribuições Estocásticas

Quantis

Os quantis das distribuições estocásticas são úteis na modelagem de dados em geral e na Mensuração Florestal em particular. Eles permitem a análise gráfica do comportamento de medidas observadas em campo, de modo a se busca por análise visual a distribuição que é mais apropriada para a sua modelagem.

Definição 7.11 (Quantil). O $q^{\text{ésimo}}$ quantil de uma distribuição estocástica com função de distribuição $F(\cdot)$, denotado por ξ_q , é o menor número ξ que satisfaz a condição

$$F(\xi_q) \geq q.$$

Se a distribuição for contínua, o $q^{\text{ésimo}}$ quantil satisfaz a condição $F(\xi_q) = q$.

Alguns quantis recebem uma denominação particular. Os quantis relativos às

porcentagens da distribuição são designados por *percentis*. Por exemplo, o quantil para $q = 0,1$ é chamado de percentil 10%. O quantil central da distribuição, isto é, o quantil referente a $q = 0,5$, é chamado de *mediana*. Os quantis que dividem uma distribuição em quartos são chamados de primeiro *quartil* ($q = 0,25$), segundo *quartil* ($q = 0,5$ — mediana) e terceiro *quartil* ($q = 0,75$).

Uma forma de medir a dispersão de uma distribuição é a *distância interquartil*, sendo definida como a diferença entre o terceiro e o primeiro quartil

$$\Delta_q = \xi_{0,75} - \xi_{0,25}.$$

Note que é possível obter algebricamente qualquer quantil nas distribuições em que a função inversa da função de distribuição é conhecida, o que acontece com várias distribuições contínuas. Neste caso, sabendo que $0 \leq q \leq 1$, o quantil ξ_q é dado por

$$\xi_q = F^{-1}(q).$$



Exemplo 7-24: Gráfico Quantil-Quantil

O gráfico Quantil-Quantil, ou gráfico QQ, é uma forma de confrontar os quantis teóricos segundo uma dada distribuição estocástica com os quantis empíricos de uma medida observada em campo (Chambers, Cleveland, Kleiner e Tukey, 1983). Esta forma de confronto é muito útil para a prática da análise de dados na Biometria Florestal, pois ela permite avaliar em quais aspectos a distribuição observada se distingue da distribuição estocástica teórica candidata a modelar o fenômeno estudado.

Os quantis empíricos (x_i ; $i = 1, 2, \dots, n$) são obtidos ordenando-se de modo crescente os valores da medida observada em campo e atribuindo a cada valor ordenado i a probabilidade

$$q_i = \frac{i - 0,5}{n}.$$

Os quantis teóricos (ξ_i) são calculados aplicando a função inversa da função de distribuição às mesmas probabilidades q_i obtidas para os quantis empíricos:

$$\xi_i = F^{-1}(q_i).$$

O gráfico QQ resulta da plotagem dos pares (x_i, ξ_i) para todos os valores observados. Caso a variável tenha a distribuição estocástica teorizada, os pontos no

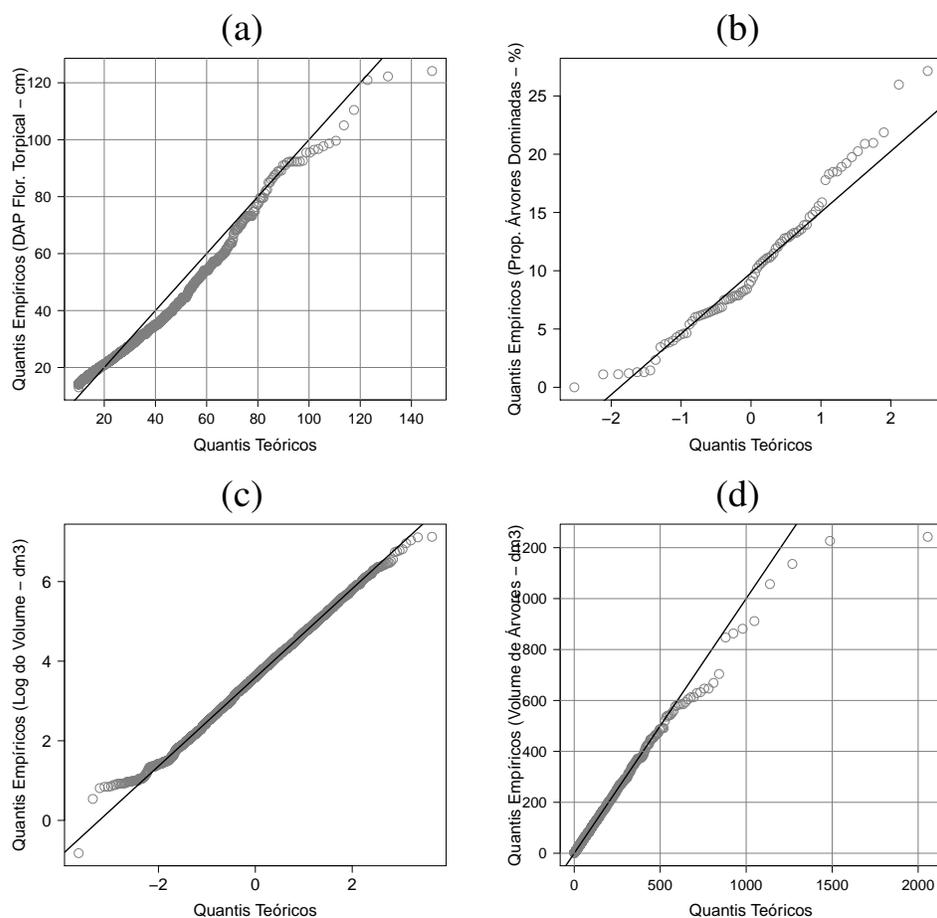


Figura 7.21: Gráficos Quantil-Quantil (gráfico QQ) para verificar a distribuição de variáveis observadas. (a) Distribuição Exponencial: DAP (cm) de árvores de floresta tropical do município de Bom Jardim, Maranhão. (b) Distribuição Gaussiana: proporção (%) de árvores dominadas em povoamentos de floresta plantada de eucalipto *urograndis*, na região central do estado de São Paulo. (c-d) Distribuição Log-normal: volume de madeira de árvores de floresta plantada de *Eucalyptus saligna* na região central do estado de São Paulo. Em (c) o volume foi transformado para escala logarítmica e comparado com os quantis teóricos da distribuição Gaussiana padronizada. Em (d) o volume é comparado com os quantis teóricos da distribuição log-normal.

gráfico se alinharão ao longo de uma reta. A figura 7.21 apresenta exemplos de gráficos QQ para as distribuições exponencial, Gaussiana e log-normal.



Simetria

A simetria é uma propriedade importante em várias situações de inferência estatística na Mensuração Florestal, pois os modelos mensuracionais geralmente têm esta propriedade como premissa.

Definição 7.12 (Distribuição Simétrica). *Uma distribuição estocástica com densidade $f(\cdot)$ é definida simétrica se*

$$f(-x) = f(x).$$

Quando a simetria da distribuição não se dá em relação à origem, mas em relação a um ponto κ na reta dos números reais, a distribuição é dita “simétrica em relação a κ ” se

$$f(\kappa - x) = f(\kappa + x).$$

A família Gaussiana tem a simetria como uma de suas propriedades fundamentais, como a sua forma de sino demonstra. A distribuição Gaussiana *padronizada*, assim como qualquer distribuição Gaussiana com média nula ($\mu = 0$), é simétrica, enquanto as demais distribuições Gaussianas, isto é, aquelas de média não nula, são simétricas em relação à média μ .

Simetria com base em Quantis

A simetria também pode ser definida em termos dos quantis da distribuição (Chambers, Cleveland, Kleiner e Tukey, 1983). Se a mediana da distribuição é $\xi_{0,5}$, então, a distribuição será simétrica se

$$\xi_{0,5} - \xi_q = \xi_{1-q} - \xi_{0,5} \quad \text{para todo } q, 0 < q < 0,5.$$

Ou seja, a simetria implica que a diferença entre os quantis individuais e a mediana é a mesma para os quantis equidistantes da mediana em termos de suas probabilidades (q).

Por exemplo, numa distribuição simétrica, a distância entre a mediana e o primeiro quartil é igual à distância entre o terceiro quartil e a mediana

$$\xi_{0,5} - \xi_{0,25} = \xi_{0,75} - \xi_{0,5}.$$

Logo, a distância interquartil é o dobro desta distância

$$\Delta_q = \xi_{0,75} - \xi_{0,25} = 2(\xi_{0,5} - \xi_{0,25}) = 2(\xi_{0,75} - \xi_{0,5}).$$



Exemplo 7-25: Gráfico de Simetria

Utilizando a simetria definida em termos de quantis é possível se construir um gráfico para avaliar a simetria de variáveis observadas em experimentos ou levantamentos. Neste caso utiliza-se a diferença entre a mediana (\tilde{x}) e as observações ordenadas de modo crescente ($x_{(i)}$)

$$\tilde{x} - x_{(i)} = x_{(n+1-i)} - \tilde{x},$$

para $i = 1$ a $n/2$, se n for par, ou para $i = 1$ a $(n+1)/2$, se n for ímpar (Chambers, Cleveland, Kleiner e Tukey, 1983). A figura 7.22 apresenta os gráficos de simetria para altura total e volume de madeira de árvores de *Eucalyptus saligna*, mostrando que na escala original estas variáveis não são simétricas, mas a transformação para a escala logarítmica as torna simétricas.

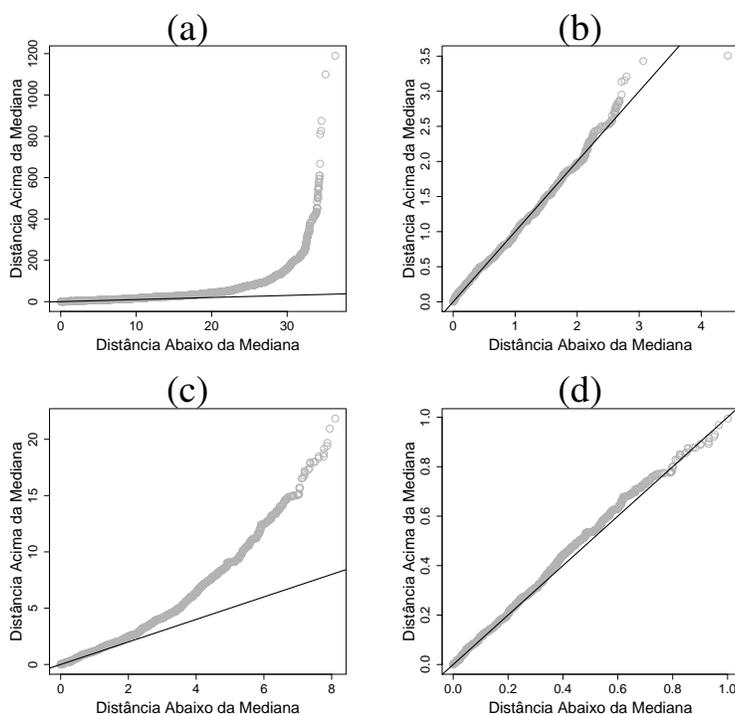


Figura 7.22: Gráficos de simetria para medidas de árvores de florestas plantadas de *Eucalyptus saligna* da região central do estado de São Paulo: volume de madeira (a) na escala original e (b) na escala logarítmica; altura total (c) na escala original e (d) na escala logarítmica.



Transformação de Escala

As medidas de árvores e arvoredos mais frequentemente utilizadas na Biometria Florestal, como DAP, altura, biomassa, área basal e produção de madeira, são geralmente variáveis matemáticas contínuas. Por isto, se utiliza modelos de distribuições estocásticas contínuas para representá-las. Contudo, frequentemente estas medidas sofrem transformações de escala, sendo provavelmente a transformação mais comum a transformação para escala logarítmica.

É necessário, portanto, considerar as alterações que ocorrem nas distribuições estocásticas quando as medidas que elas visam modelar sofrem transformações de escala. Este é um aspecto importante, pois, na prática, a transformação de escala implica na transformação de uma distribuição em outra distribuição. O teorema abaixo, que segue a apresentação de Hoel *et al.* (1971), descreve com obter a distribuição estocástica da medida transformada.

Teorema 7.1 (Transformação de Distribuição). *Considere uma função $g(\cdot)$ estritamente monotônica³ e diferenciável no intervalo I , de modo que $g(I)$ denote o seu contradomínio e $g^{-1}(\cdot)$ a sua função inversa.*

Seja X uma variável com distribuição estocástica contínua de densidade $f_X(\cdot)$, tal que $f_X(x) = 0$ para $x \notin I$. Então, a transformação de escala $Y = g(X)$ terá função de densidade

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \quad \text{para } y \in g(I) \quad (7.22)$$

e $f_Y(y) = 0$ para $y \notin g^{-1}(I)$.

De modo equivalente, pode-se afirmar que a função de densidade de Y será tal que

$$f_Y(y)|dy| = f_X(x)|dx|, \quad y \in g(I) \quad \text{e} \quad x \in g^{-1}(I).$$



Exemplo 7-26: Transformação da Distribuição Uniforme

Considere uma variável X com distribuição uniforme e a seguinte transformação de escala $Y = g(X) = -\lambda^{-1} \ln(1 - X)$. Qual a distribuição da variável transformada Y ?

³Considere o intervalo I e dois pontos x_1 e x_2 quaisquer, nele contidos ($x_1, x_2 \in I$), tal que $x_1 < x_2$. A função contínua $g(\cdot)$ é *estritamente crescente*, no intervalo I , se $g(x_1) < g(x_2)$ e *estritamente decrescente* se $g(x_1) > g(x_2)$. Dizer que uma função é *estritamente monotônica* é dizer que ela é *estritamente crescente* ou *estritamente decrescente*.

Primeiramente, verifica-se os intervalos das variáveis. Como X é uma variável uniforme, ela está restrita ao intervalo $[0, 1]$ e possui densidade $f_X(x) = 1$. A variável transformada Y , por outro lado, estará no intervalo $[0, \infty)$.

Em segundo lugar, elabora-se a função inversa da função de transformação e sua primeira derivada:

$$\begin{aligned} Y = g(X) = -\lambda^{-1} \ln(1 - X) &\implies X = g^{-1}(Y) = 1 - \exp[-\lambda Y] \\ &\implies \frac{d}{dy} g^{-1}(y) = \lambda \exp[-\lambda y] \end{aligned}$$

Por fim, aplica-se o teorema para se obter a função de densidade transformada

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| = (1) |\lambda \exp[-\lambda y]| \\ &= \lambda \exp[-\lambda y]. \end{aligned}$$

Ou seja, a variável transformada Y segue a família exponencial com parâmetro λ .

É interessante notar um aspecto especial desta transformação de escala:

$$\begin{aligned} g^{-1}(y) &= 1 - \exp[-\lambda y] = F_Y(y) \\ g(x) &= -\lambda^{-1} \ln(1 - x) = F_Y^{-1}(x). \end{aligned}$$

A função inversa da transformação ($g^{-1}(y)$) é a própria função de distribuição da família exponencial e, conseqüentemente, a transformação ($g(x)$) é a função inversa da função de distribuição da exponencial. Pode-se concluir, portanto, que é possível se obter qualquer distribuição estocástica contínua a partir da distribuição uniforme. Para isto, basta utilizar como transformação a função inversa da função de distribuição desejada.



Esperança Matemática

Um conceito muito útil na inferência estatística é o conceito de *esperança matemática* ou *valor esperado*. A esperança matemática pode ser entendida como uma operação que é realizada sobre uma distribuição resultando numa contante correspondente ao “valor médio” da distribuição, ou, numa interpretação física, ao “centro de gravidade” da distribuição.

A esperança é particularmente útil quando o modelo estocástico de interesse envolve a distribuição Gaussiana, mas ela se aplica a todas distribuições estocásticas, discretas ou contínuas.

Definição 7.13 (Esperança Matemática). *A esperança matemática de uma distribuição estocástica com função de densidade discreta $f(\cdot)$, definida num subconjunto contável de números reais $(x_1, x_2, \dots, x_n, \dots)$, é*

$$\mathbf{E}\{X\} = \sum x_i f(x_i).$$

Para distribuições contínuas com função de densidade $f(\cdot)$, a esperança matemática se torna

$$\mathbf{E}\{X\} = \int_{-\infty}^{\infty} x f(x) dx.$$

O conceito de esperança matemática pode ser ampliado para tratar transformações que se apliquem sobre as distribuições estocásticas.

Definição 7.14 (Esperança de Variável Transformada). *Seja X uma variável com densidade $f(x)$ sobre a qual se aplica uma transformação por meio da função monotônica $g(\cdot)$. A esperança matemática da variável transformada $g(X)$ é definida como*

$$\mathbf{E}\{g(X)\} = \sum g(x_i) f(x_i).$$

se X tem distribuição discreta definida num subconjunto contável de números reais $(x_1, x_2, \dots, x_n, \dots)$, e

$$\mathbf{E}\{g(X)\} = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

se X tem distribuição contínua.

Esta definição permite que se obtenha a esperança matemática numa transformação de escala sem a necessidade de se deduzir a distribuição estocástica que resulta da transformação. Ressalte-se que, no caso da esperança matemática, não é necessário que a transformação seja *estritamente* monotônica, basta que seja monotônica.

Propriedades da Esperança Matemática. Sendo uma operação que se aplica a distribuições estocásticas, a esperança matemática tem as seguintes propriedades:

- (i) $\mathbf{E}\{c\} = c$, para uma constante c .
- (ii) $\mathbf{E}\{c g(X)\} = c \mathbf{E}\{g(X)\}$, para uma constante c .
- (iii) $\mathbf{E}\{c_1 g_1(X) + c_2 g_2(X)\} = c_1 \mathbf{E}\{g_1(X)\} + c_2 \mathbf{E}\{g_2(X)\}$.
- (iv) Se $g_1(x) \leq g_2(x)$, então $\mathbf{E}\{g_1(X)\} \leq \mathbf{E}\{g_2(X)\}$.

Variância

Uma esperança matemática especial é a esperança do quadrado dos desvios em relação ao valor esperado, pois ela indica um “desvio médio quadrático” em relação ao “valor médio” da distribuição. Esta esperança recebe o nome de *variância*, pois é uma operação análoga à variância na distribuição Gaussiana.

Definição 7.15 (Variância de uma Distribuição). *Seja X uma variável com densidade $f(x)$, discreta ou contínua, sua variância é definida como*

$$\mathbf{Var}\{X\} = \mathbf{E}\{(X - \mathbf{E}\{X\})^2\}.$$

A aplicação das propriedades da esperança matemática implica que a variância de X também pode ser obtida por

$$\mathbf{Var}\{X\} = \mathbf{E}\{X^2\} - (\mathbf{E}\{X\})^2.$$

A variância, como uma forma especial de esperança matemática, também tem propriedades particulares:

- (i) $\mathbf{Var}\{c\} = 0$, para uma constante c .
- (ii) $\mathbf{Var}\{c g(X)\} = c^2 \mathbf{Var}\{g(X)\}$, para uma constante c .
- (iii) Se X_1, X_2, \dots, X_n é uma sequência de resultados independentes de uma distribuição estocástica⁴, isto é, $P(X_i = x_i | X_j = x_j) = P(X_i = x_i)$, para $i \neq j$, sendo $i, j = 1, 2, \dots, n$, então

$$\mathbf{Var}\{X_1 + X_2 + \dots + X_n\} = \mathbf{Var}\{X_1\} + \mathbf{Var}\{X_2\} + \dots + \mathbf{Var}\{X_n\}.$$

Talvez a maneira mais prática de se entender a esperança matemática e a variância é como “operadores matemáticos” que se aplica sobre as distribuições estocásticas para se gerar informações a respeito delas na forma de constantes matemáticas. A tabela 7.1 apresenta a esperança matemática e a variância das famílias de distribuição estocásticas tratadas anteriormente.

Ressalte-se que tanto a esperança quanto a variância são sempre funções dos parâmetros das distribuições, mas no caso particular da família Gaussiana elas são os próprios parâmetros da distribuição.

⁴Neste capítulo o conceito de distribuição estocástica *multivariada* não é tratado, portanto, não se aborda o conceito de independência estocástica no contexto de distribuições marginais e distribuições condicionais.

Tabela 7.1. Esperança matemática e variância das distribuições apresentadas.

DISTRIBUIÇÃO	PARÂMETROS	ESPERANÇA	VARIÂNCIA
<i>Distribuições Discretas</i>			
Bernoulli	p	p	$p(1 - p)$
Uniforme	N	$1/N$	$(N + 1)(N - 1)/12$
Binomial	N, p	Np	$Np(1 - p)$
Poisson	λ	λ	λ
Geométrica	p	$(1 - p)/p$	$(1 - p)/p^2$
Binomial Negativa	r, p μ, k	$r(1 - p)/p$ μ	$r(1 - p)/p^2$ $\mu + \mu^2/k$
<i>Distribuições Contínuas</i>			
Retangular	a, b	$(a + b)/2$	$(b - a)^2/12$
Uniforme	$0, 1$	$1/2$	$1/12$
Exponencial	λ β	$1/\lambda$ β	$1/\lambda^2$ β^2
Weibull	α, β, γ	$\alpha + \beta \Gamma^*(1)$	$\beta (\Gamma^*(2) - [\Gamma^*(1)]^2)$
Gama	β, γ λ, γ	$\beta \gamma$ γ/λ	$\beta^2 \gamma$ γ/λ^2
Gaussiana	μ, σ	μ	σ^2
Log-normal	μ, σ	$\exp[\mu + \sigma^2/2]$	$\exp[2\mu + \sigma^2][\exp(\sigma^2) - 1]$

$\Gamma^*(k) = \Gamma[(\gamma + k)/\gamma]$, sendo $\Gamma(\cdot)$ a função gama e γ o parâmetro de forma.



Exemplo 7-27: Distribuição Poisson e Índice de Dispersão

As propriedades da esperança matemática e da variância da distribuição de Poisson são tradicionalmente utilizada para se calcular um índice de dispersão de organismos que permite inferir se o padrão espacial dos organismos é completamente aleatório. O índice de dispersão utiliza a contagem dos organismos em unidades amostrais de área fixa (parcelas), como o apresentado na figura 7.23.

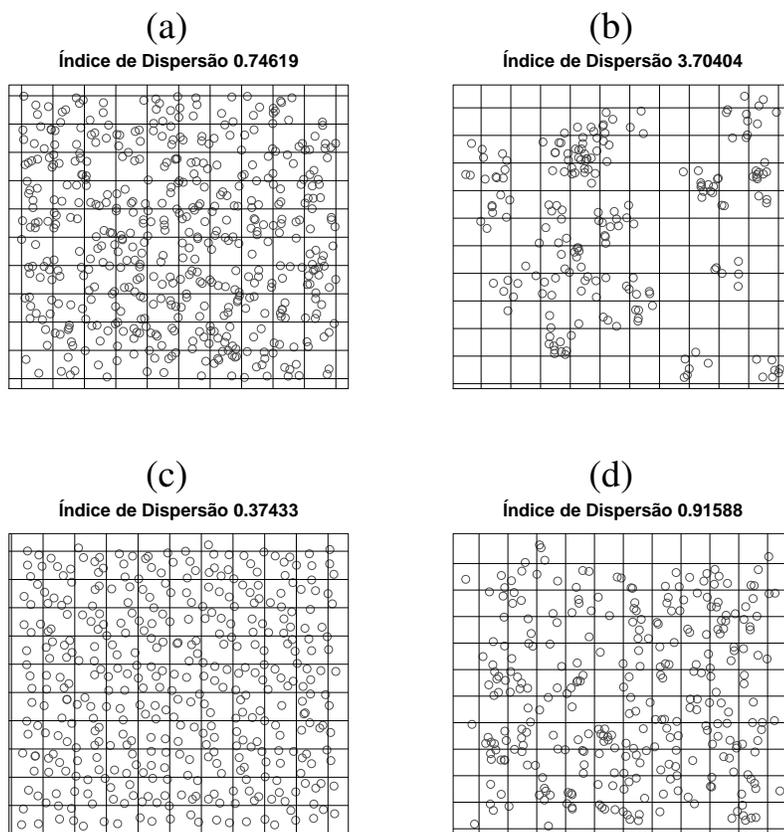


Figura 7.23: Padrões espaciais: (a) completamente aleatório, (b) agregado, (c) regular, e (d) agregado cujo índice de dispersão não consegue identificar.

É possível demonstrar que se o padrão espacial dos organismos for completamente aleatório, o número de organismos por parcela seguirá a distribuição Poisson, que tem a esperança matemática e variância iguais ao parâmetro da distribuição. ($E\{X\} = \text{Var}\{X\} = \lambda$). Se a contagem dos organismos nas parcelas segue de fato uma distribuição Poisson, média amostral (\bar{x}) e variância amostral (s_x^2)

estarão bastante próximas. Logo, a sua razão pode ser tomada como um índice de dispersão (s_x^2/\bar{x}), o qual será aproximadamente um (1) para organismos com padrão espacial aleatório.

Índices de dispersão maiores que um implicam na variância do número de organismos por parcela ser maior que a esperança, indicando que o padrão espacial se desvia do padrão aleatório em direção à agregação. Por outro lado, índices de dispersão menores que um implicam na variância ser menor que a esperança e, conseqüentemente, que o padrão espacial se desvia da aleatoriedade em direção à regularidade. A figura 7.23 apresenta quatro padrões espaciais com os respectivos índices de dispersão obtidos pelas enumeração dos pontos nas parcelas definidas pelo reticulado do gráfico.

Este índice, contudo, tem uma forte limitação na detecção dos desvios do padrão aleatório. A sua capacidade de identificar os padrões que se desviam da aleatoriedade está diretamente ligada ao tamanho da parcela. O índice falha quando a agregação ocorre numa escala igual ou menor ao tamanho da parcela, como no padrão apresentado na figura 7.23 (d).



Exemplo 7-28: Viés Logarítmico e Distribuição Log-normal

Dentre os vários modelos lineares utilizados nos estudos de Biometria Florestal, são comuns os modelos em que a relação linear se dá na escala logarítmica. Por exemplo, o modelo de equação de volume proposto por Schumacher e Hall (1933) é um modelo linear múltiplo em que a relação entre o volume (v_i), o DAP (d_i) e a altura (h_i) das árvores é tomada na escala logarítmica:

$$\ln(v_i) = \beta_0 + \beta_1 \ln(d_i) + \beta_2 \ln(h_i) + \varepsilon_i,$$

sendo $\ln(\cdot)$ o logaritmo natural e ε_i o termo estocástico do modelo que, por premissa, segue uma distribuição Gaussiana com média nula ($\mu = 0$) e variância constante (σ^2).

O componente determinístico nos modelos lineares determina a esperança matemática da variável resposta, dados os valores das variáveis predictoras. Neste modelo em particular, tem-se

$$\mathbf{E}\{\ln(v_i)\} = \beta_0 + \beta_1 \ln(d_i) + \beta_2 \ln(h_i).$$

Assim, o logaritmo do volume ($\ln(v_i)$) segue uma distribuição Gaussiana com média igual à esta esperança matemática e variância constante (σ^2).

Ao se ajustar este modelo para uma situação particular, as estimativas de quadrados mínimos dos parâmetros ($\widehat{\beta}_0$, $\widehat{\beta}_1$ e $\widehat{\beta}_2$) permitem se obter uma predição do *logaritmo do volume* da árvore:

$$\widehat{\ln(v_i)} = \widehat{\beta}_0 + \widehat{\beta}_1 \ln(d_i) + \widehat{\beta}_2 \ln(h_i).$$

Para se obter uma predição do volume da árvore, esta predição é matematicamente transformada para escala original:

$$\exp[\widehat{\ln(v_i)}] = \widehat{v}_i = \exp[\widehat{\beta}_0] d_i^{\widehat{\beta}_1} h_i^{\widehat{\beta}_2}.$$

Mas, na escala original, esta predição do volume possui um viés conhecido como *viés logaritmo*.

O viés logaritmo ocorre porque se o logaritmo do volume ($\ln(v_i)$) segue uma distribuição Gaussian, o volume na escala original (v_i) segue uma distribuição log-normal, cuja esperança matemática é

$$\begin{aligned} \mathbf{E}\{v_i\} &= \exp[\mu_i + \sigma^2/2] = \exp[\beta_0 + \beta_1 \ln(d_i) + \beta_2 \ln(h_i) + \sigma^2/2] \\ &= \exp[\beta_0] d_i^{\beta_1} h_i^{\beta_2} \exp[\sigma^2/2]. \end{aligned}$$

Logo, a predição adequada do volume na escala original é

$$\widehat{v}_i = \exp[\widehat{\beta}_0] d_i^{\widehat{\beta}_1} h_i^{\widehat{\beta}_2} \exp[\widehat{\sigma}^2/2],$$

sendo que a estimativa apropriada para variância do componente estocástico ($\widehat{\sigma}^2$) é o quadrado médio do resíduo:

$$\widehat{\sigma}^2 = \frac{\sum_{i=1}^n \left(\ln(v_i) - \widehat{\ln(v_i)} \right)^2}{n - 3}.$$

O termo $\exp[\widehat{\sigma}^2/2]$ da predição do volume é conhecido como *correção do viés logarítmico*. Flewelling e Pienaar (1981) discutiram em detalhes a correção do viés logarítmico em modelos florestais.



7.6 Relações entre Famílias de Distribuições

As relações entre as famílias de distribuições estocásticas podem ser de grande utilidade na inferência estatística. Certas relações facilitam o cálculo de probabilidades, outras permite simplificar os modelos de inferência. Segue-se algumas relações de importância na Mensuração Florestal.

Lei Forte dos Grandes Números

A lei forte dos grandes números estabelece um princípio de grande aplicação prática nas ciências e engenharias em geral, e que fundamenta várias aplicações práticas de Mensuração Florestal. Ela estabelece que, com uma amostra suficientemente grande, pode-se ter certeza que a média amostral estará muito próxima à esperança matemática da variável observada, independentemente da distribuição estocástica envolvida.

Teorema 7.2 (Lei Forte dos Grandes Números). *Seja (X_1, X_2, \dots, X_n) uma sequência de resultados independentes de uma mesma distribuição estocástica, cuja esperança matemática é $\mathbf{E}\{X_i\} = \mu$. Então, com probabilidade 1 (evento certo), a média amostral \bar{x}_n se aproxima da esperança matemática*

$$\frac{X_1 + X_2 + \dots + X_n}{n} = \frac{S_n}{n} = \bar{x}_n \rightarrow \mu,$$

à medida que $n \rightarrow \infty$ (Ross, 1989).

A lei forte dos grandes números garante que no caso de grandes amostras, a média amostral pode ser tomada como uma boa estimativa da esperança matemática de uma medida de árvores, arvoredos ou floresta, independentemente da distribuição estocástica dos valores individuais da medida sendo estudada. Obviamente, a questão central da aplicação desta lei é quando uma amostra tem tamanho suficientemente grande para a aplicação da lei.

Na utilização prática desta lei, o tamanho de amostra n adequado para se poder contar com a aproximação depende da família de distribuições envolvida e, dentro de cada família, da distribuição específica em questão, isto é, dos valores dos parâmetros da distribuição em cada caso particular. Geralmente, em famílias contínuas, será preciso um n menor que em famílias discretas. Numa mesma família, quanto mais simétrica os valores dos parâmetros que tornem a distribuição, menor o tamanho de amostra necessário para aplicação da lei.

Teorema Central do Limite

O teorema central do limite expande o resultado da lei forte dos grandes números, ampliando as potencialidades de utilização prática.

Teorema 7.3 (Teorema Central do Limite). *Seja (X_1, X_2, \dots, X_n) uma sequência de resultados independentes de uma mesma distribuição estocástica, cuja esperança matemática é $\mathbf{E}\{X_i\} = \mu$ e a variância é $\mathbf{Var}\{X_i\} = \sigma^2$. Então, a distribuição de*

$$\frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}},$$

tende à distribuição Gaussiana padronizada à medida que $n \rightarrow \infty$. Ou seja, tem-se que para qualquer número real z ($-\infty < z < +\infty$)

$$\lim_{n \rightarrow \infty} P \left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq z \right] = \Phi(z),$$

onde $\Phi(\cdot)$ é a função de distribuição da Gaussiana padronizada (Ross, 1989).

O teorema central do limite estabelece um limite que não converge para uma constante, mas para uma função de distribuição, por isto a convergência é chamada de *convergência em distribuição*.

Em termos de distribuição estocástica, a sua aplicação é irrestrita, pois ele se aplica a qualquer distribuição. Mas, vale aqui as mesmas considerações feita sobre a lei forte dos grandes números. O aspecto fundamental para aplicação é o tamanho da amostra (n) suficientemente grande para garantir a aplicação com segurança do teorema. Distribuições contínuas geralmente exigirão um tamanho menor que distribuições discretas, enquanto que as distribuições com menor assimetria necessitarão de amostras menores que aquelas com maior assimetria.



Exemplo 7-29: Média Amostral

O Teorema Central do Limite é de grande utilidade quando se está interessado no comportamento da média amostral. A figura 7.24 mostra como, à medida que o tamanho da amostra cresce, a média amostral se aproxima rapidamente da distribuição Gaussiana, tanto para variáveis contínuas quanto discretas.

Um aspecto importante, é que o tamanho de amostra que consiste numa aproximação razoável da distribuição Gaussiana não depende somente da família da distribuição da variável observada, mas da distribuição específica da variável em cada situação particular, isto é, ela depende do valor que os parâmetros assumem. A figura 7.24 utiliza a família da distribuição Weibull para mostrar que em alguns casos uma amostra de tamanho 5 já gera uma distribuição da média amostral bastante simétrica e amostras de tamanho 20 já estariam bem próximas da distribuição Gaussiana. Para outras distribuições, no entanto, somente com amostras de tamanho próximo a 100 resultam em médias amostrais com distribuição próxima

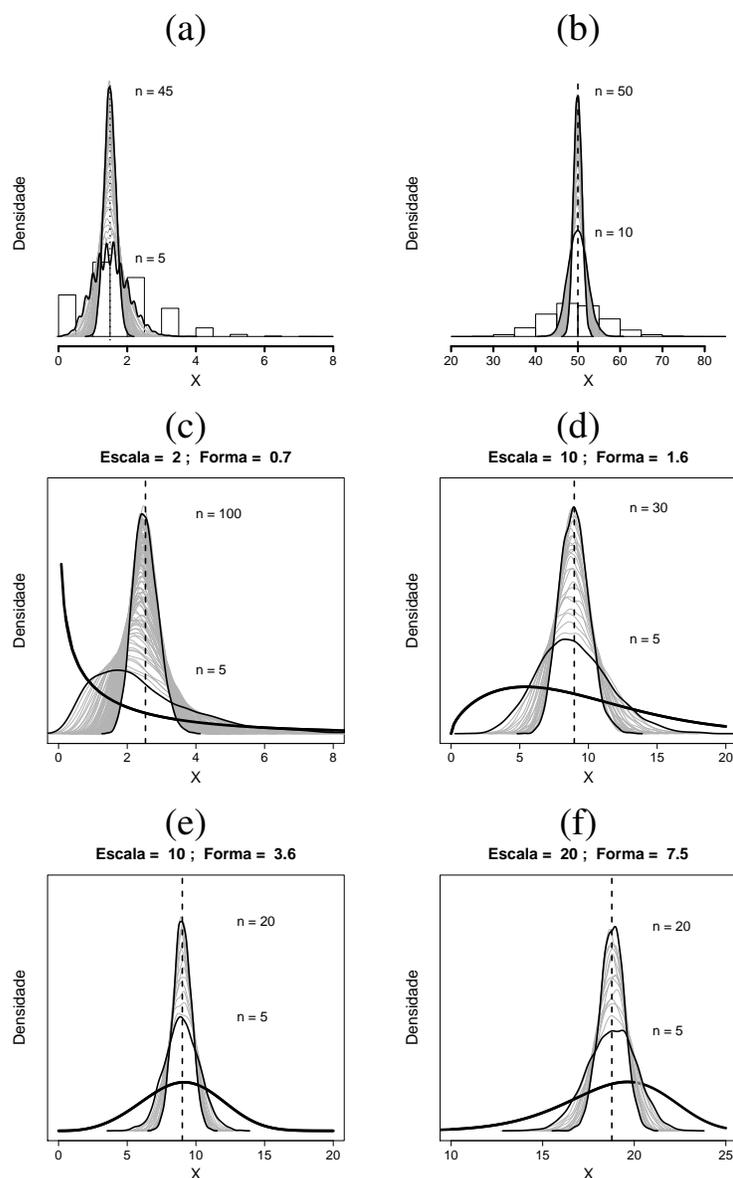


Figura 7.24: Densidades empíricas para a média amostral de variáveis com diferentes distribuição estocásticas (X): (a) binomial ($n = 10, p = 0,15$), (b) Poisson ($\lambda = 50$), (c-f) Weibull com diferentes valores dos parâmetros. As curvas em cinza resultam de diferentes tamanhos de amostra, o menor e o maior tamanho de amostra são indicados no gráfico. As curvas com linhas de maior grossura indicam a distribuição estocástica dos dados.

à Gaussiana. Neste caso particular da família Weibull, a assimetria da distribuição da variável é um aspecto determinante.



Aproximação Gaussiana para Binomial

Uma das aplicações do teorema central do limite é que, para uma grande amostra n , a distribuição binomial pode ser aproximada pela distribuição Gaussiana padronizada. Ou seja, se X tem distribuição binomial com parâmetros n e p , então para n suficientemente grande

$$\begin{aligned} P \left[z_1 \leq \frac{X - \mathbf{E}\{X\}}{\sqrt{\mathbf{Var}\{X\}}} \leq z_2 \right] &= P \left[z_1 \leq \frac{X - np}{\sqrt{np(1-p)}} \leq z_2 \right] = \\ &= P \left[z_1 \sqrt{np(1-p)} + np \leq X \leq z_2 \sqrt{np(1-p)} + np \right] = \Phi(z_2) - \Phi(z_1). \end{aligned}$$

Aproximação Gaussiana para a Média Amostral

A aplicação mais ampla do teorema central do limite é no cálculo de probabilidades envolvendo a média amostral. Para uma amostra aleatória de n observações independentes de uma variável X com distribuição estocástica, cuja esperança matemática é $\mathbf{E}\{X\} = \mu$ e a variância é $\mathbf{Var}\{X\} = \sigma^2$, então, para n suficientemente grande, as probabilidades envolvendo a média amostral $\bar{X}_n = \sum X_i/n$ podem ser obtidas por

$$\begin{aligned} P \left[z_1 \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z_2 \right] &= P \left[z_1 \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq z_2 \right] = \\ &= P \left[z_1 \frac{\sigma}{\sqrt{n}} + \mu \leq \bar{X}_n \leq z_2 \frac{\sigma}{\sqrt{n}} + \mu \right] = \Phi(z_2) - \Phi(z_1). \end{aligned}$$

Este resultado nos mostra que, para uma probabilidade constante $(\Phi(z_2) - \Phi(z_1))$, o intervalo ao redor da média amostral (\bar{X}_n) diminui com a raiz quadrada do tamanho da amostra (\sqrt{n}) , isto é, a exatidão da média amostral aumenta com \sqrt{n} . Logo, em teoria, é possível se obter uma média amostral arbitrariamente precisa.



Exemplo 7-30: Aproximações Gaussianas

A figura 7.25 mostra a aproximação Gaussiana em dois casos particulares onde o intervalo de 95% de probabilidade ao redor da média pode ser calculado, para

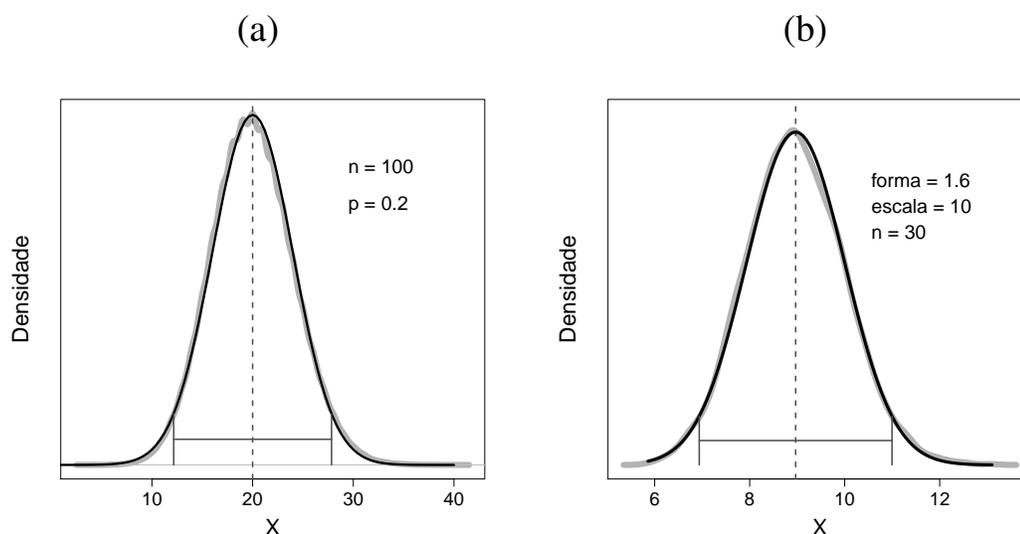


Figura 7.25: Intervalo de 95% de probabilidade resultante da aproximação Gaussiana para (a) variável binomial e (b) média amostral resultante de uma variável com distribuição Weibull. A linha com maior irregularidade é a distribuição original, enquanto que alinha perfeitamente regular e simétrica é a aproximação Gaussiana.

efeitos práticos, a partir da distribuição Gaussiana. O primeiro caso é de uma variável binomial com parâmetros $n = 100$ e $p = 0,2$. O segundo caso é da média amostral resultante de amostras de tamanho 30 calculadas de uma variável com distribuição Weibull com parâmetro de forma de 1,6 e de escala de 10. Note que, para efeitos práticos, não há diferença entre o intervalo de probabilidade calculado utilizando a distribuição original e aquele utilizando a aproximação Gaussiana.



* * *

Neste capítulo foi apresentado o conceito de distribuição estocástica que é o conceito-chave para inferência estatística. Também foram apresentadas aquelas famílias de distribuição de maior relevância para Biometria e Mensuração Florestal, e concluiu-se a apresentação com algumas propriedades e relações importantes entre distribuições. De forma muito sumária, este capítulo e o anterior resumem

os aspectos mais abstratos da teoria de probabilidades necessários à inferência estatística.

Para se continuar o desenvolvimento da inferência preditiva baseada em verossimilhança é necessário trazer para a discussão os dados gerados pelas observações do mundo natural. É necessário construir uma “*ponte conceitual*” entre os modelos teóricos das distribuições estocásticas e os resultados empíricos provenientes da observação das árvores, dos arvoredos e das florestas que são os objetos de estudo da Mensuração Florestal. A *verossimilhança* é esta “ponte conceitual”.

Capítulo 8

Estimação por Verossimilhança

Ligação entre Observações e Modelos

As distribuições estocásticas são o componente básico dos modelos estocásticos utilizados na estatística em geral e na Biometria Florestal em particular. A sua aplicação em situações particulares requer que os dados resultantes de observações empíricas possam ser incorporados ao modelo. Este processo é geralmente chamado de “*ajuste do modelo*” e se traduz na determinação de valores numéricos para os parâmetros da distribuição, isto é, na “*estimação*” dos parâmetros. Depois da obtenção dos dados e da escolha dos modelos candidatos a representar o cenário estocástico em estudo, a estimação é o próximo passo no processo de modelagem estocástica.

Os diferentes procedimentos para estimação dos parâmetros são frequentemente chamados de “*estimadores*”. Existe uma extensa literatura estatística que discute as vantagens e desvantagens dos vários tipos de estimadores quando aplicados a diferentes distribuições e a diferentes tipos de dados. A abordagem da verossimilhança implica no seu tipo específico de estimador e em formas particulares de avaliá-lo. Adotar a abordagem da verossimilhança significa, portanto, se restringir a um subconjunto das técnicas estatísticas propostas para a análise de dados. Contudo, o “preço” desta restrição é “pago com juros e dividendos” pela adoção de uma forma de inferência que é mais coerente, mais flexível e, na perspectiva de sua aplicação, mais simples em sua teoria.

Este capítulo utiliza o conceito de verossimilhança e suas implicações para a estimação dos parâmetros de distribuições estocásticas e desenvolve técnicas de inferência visando a avaliação das estimativas obtidas. A ênfase é em modelos voltados para inferência preditiva, mas o desenvolvimento segue a abordagem proposta por Edwards (1992) e Royall (1997).

8.1 Critério de Estimação

Há várias maneiras de se determinar o valor numérico dos parâmetros de um modelo estocástico. Talvez, o mais antigo seja o método de quadrados mínimos, que consistem em encontrar os valores que minimizam a soma do quadrado dos desvios em relação ao parâmetro a ser estimado. Para distribuições simétricas e centradas no parâmetro, este método é conveniente e eficiente. Karl Pearson propôs estimar os parâmetros de uma distribuição estocástica com base nos *momentos* da distribuição, isto é, com base na média (primeiro momento), variância (segundo momento centrado na média), assimetria (terceiro momento centrado na média), etc (Stigler, 1986; Lehmann, 2011). Uma outra abordagem se baseia no uso dos quantis da distribuição estocástica: sempre que os parâmetros possam ser expressos como função dos quantis, também é possível se estimar os parâmetros com base em quantis amostrais.

Quadrados mínimos, momentos e quantis são propriedades das distribuições estocásticas e, portanto, possuem ligação direta com os parâmetros. Mas estes diferentes métodos não são igualmente eficientes para as diferentes distribuições. Na distribuição Gaussiana, o método dos quadrados mínimos pode ser visto como um método não só eficiente como também natural, uma vez que a própria densidade é expressa em termos dos desvios ao quadrado. Em algumas distribuições, os parâmetros podem ser expressos como funções relativamente simples dos quantis, tornando o método dos quantis fácil de implementar. Mas os diferentes métodos atendem a critérios distintos de eficiência.

Um aspecto importante a se ressaltar é que os métodos dos momentos e o método dos quantis realizam a estimação dos parâmetros com base numa redução substancial da quantidade de informação presente nos dados. O primeiro porque os valores individuais das observações são reduzidos aos momentos, o segundo porque, uma vez ordenadas as observações, somente as observações correspondentemente alguns poucos quantis são efetivamente utilizadas na estimação. Já o método dos quadrados mínimos, utiliza *quase* todas as observações individuais. O “*quase*” acontece, porque ao buscar minimizar a soma dos quadrados dos desvios, ele torna nula a soma dos desvios das n observações e, conseqüentemente, o valor de uma das observações pode ser aritmeticamente obtido pelo valor da soma das outras $n - 1$ observações. Para se calcular a verossimilhança, contudo, são necessários os valores individuais de todas as observações.

Propriedades de um Critério de Estimação

A verossimilhança, como método de estimação de parâmetros, foi proposta por Fisher que, em seu trabalho sobre os fundamentos da matemática estatística de 1922, apresentou o uso do *método da máxima verossimilhança*. Ao longo dos

anos, vários autores desenvolveram este método, deduzindo as suas propriedades e demonstrando as suas vantagens. Hacking (1965) chegou a afirmar que a verossimilhança é, ou deveria ser, a base da inferência estatística no sentido que ela confere e garante o único arcabouço lógico consistente para a inferência estatística. Royall (1997) reforçou que a abordagem da verossimilhança supera as limitações e problemas das duas escolas tradicionais de inferência estatística: a clássica e a Bayesiana.

Mas Edwards (1992), no seu livro “Likelihood”, talvez seja o autor que melhor delineou uma forma de inferência totalmente baseada na verossimilhança. No início da sua proposta, Edwards listou as propriedades que se espera de uma medida matemática apropriada como critério de inferência estatística. Segue uma apresentação adaptada desta lista:

Transitividade: se a medida aponta a hipótese H_1 como superior à hipótese H_2 , e H_2 como superior à H_3 , então, H_1 deve ser superior à H_3 .

Aditividade: deve ser possível combinar, preferencialmente de modo aditivo, a medida relativa a duas hipóteses obtida a partir de um conjunto de dados, com a medida relativa às mesmas duas hipóteses obtida de outros conjuntos de dados independentes.

Invariância sob a transformação dos dados: a medida não deve ser afetada por uma transformação um-para-um dos dados.

Invariância sob a transformação dos parâmetros: no caso de parâmetros contínuos, a medida deve ser independente da forma particular de parametrização do modelo estocástico.

Relevância: a medida deve se mostrar intuitivamente aceitável nas aplicações práticas. Ao aplicá-la, deve ser intuitivamente natural se reagir de modo equivalente a medidas de igual valor e a medida não deve ser afetada por informação julgada irrelevante.

Consistência: no caso em que uma hipótese seja “a hipótese verdadeira”, esta hipótese deve atrair para si o valor máximo da medida, quando o tamanho da amostra tende para o infinito.

Compatibilidade: a medida deve possuir uma relação simples com a maneira com que os resultados do cenário estocástico são incorporados ao Teorema de Bayes, naqueles casos em que existam probabilidades a priori válidas.

Dentre os diversos métodos de estimação de parâmetros, a única medida que é capaz de satisfazer todas estas propriedades em qualquer modelo estocástico é a verossimilhança.

8.2 *Conceito de Verossimilhança*

Uma definição de dicionário¹ para palavra “verossímil” é:

1. que parece verdade; 2. que é possível ou provável por não contrariar a verdade; plausível

e para “verossimilhança”:

1. qualidade do que é verossímil ou verossimilhante.

No caso da inferência estatística, verossimilhança tem uma definição técnica, mas que possui uma ligação direta com a definição de “semelhante à verdade” ou “plausível” apresentada pelo dicionário.

A palavra é uma tradução direta do termo da língua inglesa “*likelihood*”. Seu conceito estatístico foi originalmente apresentado por Fisher para diferenciá-lo do conceito de probabilidade, uma confusão que foi introduzida na matemática pela ideia de probabilidade inversa. No seu trabalho “*On the Mathematical Foundations of Theoretical Statistics*” (Fisher, 1922), ele afirma:

“We must return to the actual fact that one value of p , of the frequency of which we know nothing, would yield the observed result three times as frequently as would another value of p . If we need a word to characterise this relative property of different values of p , I suggest that we may speak without confusion of the *likelihood* of one value of p being thrice the likelihood of another, bearing always in mind that likelihood is not here used loosely as a synonym of probability, but simple to express the relative frequencies with which such values of the hypothetical quantity p would in fact yield the observed sample.”

(Fisher, 1922, p.326)

O argumento de Fisher é que não faz sentido se falar na *probabilidade* de um parâmetro assumir certos valores. Este é o problema fundamental da ideia de probabilidade inversa.

Interpretação da Verossimilhança

A definição de Fisher deixa claro que a verossimilhança de um certo valor do parâmetro é proporcional à frequência relativa, isto é, à probabilidade deste valor produzir os resultados observados. Assim, o valor de um parâmetro que gera

¹Dicionário Houaiss da Língua Portuguesa, 1^a edição. Rio de Janeiro: Objetiva, 2009.

uma probabilidade maior de se produzir os resultados observados é um valor mais verossímil. Logo, a verossimilhança só tem sentido comparativo, pois é uma *medida relativa*. Mas uma medida relativa do quê? Há duas formas de interpretação possível.

Evidência Relativa. A primeira interpretação é que a verossimilhança é uma medida de evidência relativa. Como a verossimilhança compara o valor para um parâmetro de um modelo estocástico *dado que certos resultados foram efetivamente observados*, ela indica qual valor do parâmetro é favorecido com base na evidência contida nos dados observados. A verossimilhança é uma medida da evidência relativa com que os diferentes valores possíveis do parâmetro são favorecidos pelos dados.

Plausibilidade Relativa. A verossimilhança também pode ser interpretada através do conceito de plausibilidade. Como o estudo dos valores do parâmetro de um modelo estocástico é sempre *condicionado* pelos resultados observados, a verossimilhança é uma medida da *plausibilidade relativa* dos valores que o parâmetro pode assumir. Os valores que geram uma maior probabilidade para os resultados observados são mais plausíveis. Também nesta interpretação, a verossimilhança é uma medida *relativa*, isto é, compara a plausibilidade de um valor *vis-à-vis* outro valor.

Definição de Verossimilhança

Seguindo-se o desenvolvimento formal apresentado por Edwards (1992), a verossimilhança pode ser definida da seguinte maneira:

Definição 8.1 (Verossimilhança). *A verossimilhança $\mathcal{L}\{H|R\}$ de uma hipótese H , dado o resultado R e um modelo estocástico M , é proporcional à probabilidade $P_M(R|H)$, sendo a constante de proporcionalidade arbitrária.*

A hipótese H é qualquer hipótese a respeito do modelo estocástico M . No caso de estimação, a hipótese de maior interesse se refere aos valores que o parâmetro do modelo estocástico M pode assumir. O resultado R é um conjunto de dados observados no contexto do cenário estocástico em questão. A probabilidade do resultado R , dada a hipótese H , é calculada pelo modelo estocástico M .

A definição faz uma distinção fundamental entre probabilidade e verossimilhança. Na probabilidade, o resultado é condicionado pelo parâmetro, na verossimilhança o valor do parâmetro é condicionado pelo resultado observado. Na probabilidade, R é variável e H é constante, enquanto que na verossimilhança, H é variável e R é constante.

A constante arbitrária faz com que a definição seja aplicável tanto a distribuições discretas quanto contínuas. A definição de verossimilhança pode, em verdade, ser estabelecida em termos da função de densidade do modelo estocástico:

Definição 8.2 (Verossimilhança II). *A verossimilhança $\mathcal{L}\{H|R\}$ de uma hipótese H , dado o resultado R e um modelo estocástico M , é proporcional à densidade $f(R|H)$ do modelo estocástico M , sendo a constante de proporcionalidade arbitrária.*

O fato da verossimilhança ser *proporcional* à densidade, não implica que aquela tenha as mesmas propriedades desta. De fato, a verossimilhança não tem as mesmas propriedades da densidade. No caso das distribuições discretas, os valores de densidade (probabilidade) estão restritos ao intervalo $[0, 1]$ e seu somatório no espaço amostral é igual a um. No caso de distribuições contínuas, a função de densidade integrada em todo espaço amostral resulta no valor unitário. Nenhuma destas propriedades pode ser atribuída à verossimilhança. Mesmo em distribuições discretas, a verossimilhança será contínua, caso a hipótese em questão se refira a um parâmetro definido como número real. O único aspecto que se pode afirmar sobre a verossimilhança é que o seus valores, sendo iguais aos da função de densidade, são números reais não-negativos, isto é, estão no intervalo $[0, \infty)$.

Razão de Verossimilhança

A determinação do valor da constante arbitrária é irrelevante, pois o uso que se faz da verossimilhança é sempre um uso *relativo* ou *comparativo* entre hipóteses. A verossimilhança aplicada a uma situação particular, consiste em comparar diversas hipóteses duas-a-duas, dado o mesmo conjunto de resultados observados. A aplicação da verossimilhança está sempre *condicionada* a um conjunto de resultados específicos obtidos num determinado estudo empírico.

Assim, na prática se trabalha sempre com a *razão de verossimilhança* entre duas hipóteses:

Definição 8.3 (Razão de Verossimilhança). *A razão de verossimilhança de duas hipóteses H_1 e H_2 , para um certo conjunto de dados R e um modelo estocástico M , é a razão das suas verossimilhanças para as hipóteses neste conjunto de dados:*

$$\mathcal{L}_R\{H_1, H_2|R\} = \frac{\mathcal{L}\{H_1|R\}}{\mathcal{L}\{H_2|R\}} = \frac{f(R|H_1)}{f(R|H_2)}.$$



Exemplo 8-1: Regeneração Natural

Num estudo de regeneração natural em floresta, deseja-se determinar o número médio de plântulas e, para isto, uma parcela circular de 3 m de raio ($28,3 \text{ m}^2$) foi locada no campo. Duas hipóteses foram formuladas:

- Hipótese 1: o número médio de plântulas na parcela é 16 (5700 ha^{-1});
- Hipótese 2: o número médio de plântulas na parcela é 35 (12500 ha^{-1});

Na parcela em campo, observaram-se 24 plântulas (8470 ha^{-1}).

Assumindo que a distribuição espacial das plântulas na floresta é completamente aleatória, o modelo apropriado para este cenário estocástico é a distribuição Poisson. Denotando por x o número de plântulas por parcela a função de densidade é

$$f(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 1, 2, \dots; \quad \lambda > 0$$

em que λ é o parâmetro, mas que corresponde ao “valor esperado” (esperança matemática) ou número médio de plântulas por parcela na floresta.

As probabilidades de se observar $x = 24$ de acordo com as hipóteses são:

$$\text{Hipótese } H_1 : \lambda = 16 \Rightarrow f(24|\lambda = 16) = \frac{e^{-16} 16^{24}}{24!} = 0,01437018$$

$$\text{Hipótese } H_2 : \lambda = 35 \Rightarrow f(24|\lambda = 35) = \frac{e^{-35} 35^{24}}{24!} = 0,01160434$$

Logo, a observação $x = 24$ favorece a hipótese H_1 sobre a hipótese H_2 . A razão de verossimilhança é:

$$\mathcal{L}_R\{H_1, H_2|x = 24\} = \frac{f_x(24)}{f_x(24)} = \frac{0,01437018}{0,01160434} = 1,238345.$$

Ou seja, a observação $x = 24$ aponta que a hipótese H_1 é 1,2 vezes mais verossímil que a hipótese H_2 .

Pode se dizer que a hipótese H_1 é 1,2 vezes mais favorecida que a hipótese H_2 pela evidência da observação $x = 24$. Também pode se dizer que, dado que se observou $x = 24$, a hipótese H_1 é 1,2 vezes mais plausível que a hipótese H_2 .



A definição de razão de verossimilhança implica numa propriedades bastante conveniente da verossimilhança no que se refere à combinação de resultados envolvendo dois conjuntos de dados independentes.

Teorema 8.1 (Combinação de Razões de Verossimilhança). *As razões de verossimilhança de duas hipóteses H_1 e H_2 para dois conjuntos de dados independentes R_1 e R_2 , podem ser multiplicadas para se gerar a razão de verossimilhança das duas hipóteses na combinação destes dois conjunto de dados:*

$$\mathcal{L}_R\{H_1, H_2|R_1, R_2\} = \mathcal{L}_R\{H_1, H_2|R_1\} \mathcal{L}_R\{H_1, H_2|R_2\}$$

Esta propriedade é uma implicação da própria definição de verossimilhança:

$$\begin{aligned} \mathcal{L}_R\{H_1, H_2|R_1, R_2\} &= \frac{f(R_1, R_2|H_1)}{f(R_1, R_2|H_2)} = \frac{f(R_1, |H_1) f(R_2|H_1)}{f(R_1, |H_2) f(R_2|H_2)} \\ &= \frac{\mathcal{L}\{H_1|R_1\}}{\mathcal{L}\{H_2|R_1\}} \frac{\mathcal{L}\{H_1|R_2\}}{\mathcal{L}\{H_2|R_2\}} \\ &= \mathcal{L}_R\{H_1, H_2|R_1\} \mathcal{L}_R\{H_1, H_2|R_2\}. \end{aligned}$$

Esta propriedade pode ser ampliada para a um número arbitrário (k) de conjuntos de dados independentes:

$$\begin{aligned} \mathcal{L}_R\{H_1, H_2|R_1, R_2, \dots, R_k\} &= \frac{f(R_1, R_2, \dots, R_k|H_1)}{f(R_1, R_2, \dots, R_k|H_2)} \\ &= \frac{f(R_1|H_1) f(R_2|H_1) \dots f(R_k|H_1)}{f(R_1|H_2) f(R_2|H_2) \dots f(R_k|H_2)} \\ &= \frac{f(R_1|H_1)}{f(R_1|H_2)} \frac{f(R_2|H_1)}{f(R_2|H_2)} \dots \frac{f(R_k|H_1)}{f(R_k|H_2)} \\ &= \mathcal{L}_R\{H_1, H_2|R_1\} \mathcal{L}_R\{H_1, H_2|R_2\} \dots \mathcal{L}_R\{H_1, H_2|R_k\} \\ &= \prod_{i=1}^k \mathcal{L}_R\{H_1, H_2|R_i\}. \end{aligned}$$

Mas note também que

$$\begin{aligned} \mathcal{L}_R\{H_1, H_2|R_1, R_2, \dots, R_k\} &= \frac{f(R_1, R_2, \dots, R_k|H_1)}{f(R_1, R_2, \dots, R_k|H_2)} \\ &= \frac{f(R_1|H_1) f(R_2|H_1) \dots f(R_k|H_1)}{f(R_1|H_2) f(R_2|H_2) \dots f(R_k|H_2)} \\ &= \frac{\prod_{i=1}^k \mathcal{L}\{H_1|R_i\}}{\prod_{i=1}^k \mathcal{L}\{H_2|R_i\}}. \end{aligned}$$

Logo, a razão de verossimilhança entre duas hipóteses dado uma série de k observações *independentes* pode ser entendida tanto como o produto das k razões de verossimilhança, quanto como a razão entre os produtos das k verossimilhanças de cada hipótese.



Exemplo 8-2: Regeneração Natural II

Num estudo de regeneração natural em floresta, deseja-se determinar o número médio de plântulas e, para isto, foram locadas no campo dez parcela circular de 3 m de raio ($28,3 \text{ m}^2$). Duas hipóteses foram formuladas:

- Hipótese 1: o número médio de plântulas na parcela é $\lambda_1 = 16$ (5700 ha^{-1});
- Hipótese 2: o número médio de plântulas na parcela é $\lambda_2 = 35$ (12500 ha^{-1});

Assume-se que a distribuição espacial das plântulas na floresta é completamente aleatória, portanto, o modelo apropriado para este cenário estocástico é a distribuição Poisson. A verossimilhança de cada hipótese, dada as observações é:

$$\mathcal{L}\{H_i|x_j\} = \frac{e^{-\lambda_i} \lambda_i^{x_j}}{x_j!} \Rightarrow \prod_{j=1}^{10} \mathcal{L}\{H_i|x_j\} = e^{-10 \lambda_i} \prod_{j=1}^{10} \frac{\lambda_i^{x_j}}{x_j!}$$

sendo que o índice i denota a hipótese ($i = 1, 2$) e o índice j denota a observação ($j = 1, 2, \dots, 10$). A tabela abaixo apresenta a verossimilhança das hipóteses segundo cada observação:

Parcela	Número de Plântulas	Verossimilhança	
		Hipótese 1	Hipótese 2
1	26	0,0056596	0,0218697
2	29	0,0010574	0,0427689
3	25	0,0091969	0,0162461
4	25	0,0091969	0,0162461
5	18	0,0830055	0,0006118
6	24	0,0143702	0,0116043
7	29	0,0010574	0,0427689
8	23	0,0215553	0,0079573
9	25	0,0091969	0,0162461
10	24	0,0143702	0,0116043
Produto		$1,818688 \times 10^{-21}$	$1,124413 \times 10^{-19}$
Razão		61,82605	

Nota-se que a hipótese 2 se mostrou mais de 61 vezes mais plausível que a hipótese 1.



Força de Evidência e de Plausibilidade

Mas como interpretar quantitativamente a razão de verossimilhança? Qual a força de evidência contida numa razão de verossimilhança de 2? Como interpretar em termos de plausibilidade uma razão de verossimilhança de 4?

Nas aplicações práticas, a razão de verossimilhança será sempre uma medida contínua. Para se poder interpretá-la é necessário se ter *valores de referência* que serão sempre arbitrários, isto é, são definidos por circunstâncias externas e independentes à situação de aplicação prática. Royall (1997) sugere que há duas maneiras fáceis de se desenvolver uma compreensão quantitativa da razão de verossimilhança. Uma delas é considerar um exemplo não complicado, onde a intuição é forte, e examinar a razão de verossimilhança de várias observações. A outra é caracterizar a razão de verossimilhança em termos do seu impacto sobre probabilidades a priori. Segue-se estas duas maneiras como apresentadas por Royall (1997).

Experimento Canônico. Considere-se o experimento em que se tem duas urnas, uma contendo todas as bolas brancas e outra onde metade das bolas são brancas. O experimento consiste em selecionar uma das urnas e se realizar sucessivas retiradas de bolas da urna, sempre retornando a bola para a urna após a retirada, aleatorizando-se novamente as bolas dentro da urna. Há duas hipóteses sobre o conteúdo da urna selecionada: “todas as bolas são brancas” (H_1) e “metade das bolas são brancas” (H_2), enquanto que as observações são a evidência.

Suponha que a primeira bola retirada é branca ($x_1 = b$) e, em sequência a segunda bola também é branca ($x_2 = b$) e ainda a terceira bola sai branca novamente ($x_3 = b$). Muitos considerariam estas três observações como uma evidência bastante forte a favor da hipótese H_1 (“todas as bolas são brancas”) contra a hipótese

H_2 (“metade das bolas são brancas”). A razão de verossimilhança neste caso é

$$\begin{aligned} \mathcal{L}_R\{H_1, H_2|x_1 = b, x_2 = b, x_3 = b\} &= \\ &= \frac{\mathcal{L}\{H_1|x_1 = b\}}{\mathcal{L}\{H_2|x_1 = b\}} \frac{\mathcal{L}\{H_1|x_2 = b\}}{\mathcal{L}\{H_2|x_2 = b\}} \frac{\mathcal{L}\{H_1|x_3 = b\}}{\mathcal{L}\{H_2|x_3 = b\}} \\ &= \frac{P(x_1 = b|H_1)}{P(x_1 = b|H_2)} \frac{P(x_2 = b|H_1)}{P(x_2 = b|H_2)} \frac{P(x_3 = b|H_1)}{P(x_3 = b|H_2)} \\ &= \frac{1}{1/2} \frac{1}{1/2} \frac{1}{1/2} = 2^3 = 8. \end{aligned}$$

Ou seja, três bolas brancas sucessivas são uma evidência oito vezes mais forte em favor da hipótese H_1 vis-à-vis a hipótese H_2 . Dito de outra forma, três bolas brancas sucessivas tornam a hipótese “todas as bolas são brancas” oito vezes mais plausível que a hipótese “metade das bolas são brancas”.

De fato, há uma relação simples entre o número de bolas brancas sucessivas (n_b) observadas e a razão de verossimilhança: $\mathcal{L}_R\{H_1, H_2|Y = n_b\} = 2^{n_b}$. É possível traduzir a razão de verossimilhança em termos de número de bolas brancas observadas neste experimento canônico (tabela 8.1). Note que este experimento canônico implica que se o número de bolas cresce numa progressão aritmética, a razão de verossimilhança crescerá numa progressão geométrica.

Tabela 8.1: Número de bolas brancas sucessivas observadas (n_b) no experimento canônico em função da razão de verossimilhança. Reproduzido de Royall (1997).

Razão	10	20	50	100	1000	10.000	100.000	1.000.000
n_b	3,3	4,3	5,6	6,6	10,0	13,3	16,6	19,9

Se muitas pessoas consideram 3 bolas brancas como uma evidência “*bastante forte*”, é comum que as pessoas concordem que 5 bolas brancas sucessivas é uma evidência “*muito forte*” em favor de H_1 vis-à-vis H_2 . Cinco bolas correspondem a uma razão de verossimilhança de $2^5 = 32$. Assim, convencionou-se utilizar a razão de verossimilhança de 8 como uma força de evidência, ou de plausibilidade, mínima para se distinguir duas hipóteses, e a razão de 32 como uma força de evidência, ou de plausibilidade, praticamente indubitável para favorecer uma hipótese vis-à-vis a outra.

Impacto da Razões de Verossimilhança. Algumas pessoas podem considerar o experimento canônico como algo dúbio. Por que uma razão de verossimilhança de 8 teria o mesmo significado em todas as situações de aplicação prática?

Esta pergunta pode ser respondida, voltando-se à inferência por “probabilidade inversa”, pelo Teorema de Bayes. Tome-se como exemplo o experimento canônico descrito acima e aplique-se a probabilidade inversa. A “probabilidade” da hipótese H_1 é

$$P(H_1|Y = n_b) = \frac{P(Y = n_b|H_1) P(H_1)}{P(Y = n_b|H_1) P(H_1) + P(Y = n_b|H_2) P(H_2)}$$

em que n_b é o número de bolas brancas sucessivas observadas. Da mesma forma a “probabilidade” da hipótese H_2 é

$$P(H_2|Y = n_b) = \frac{P(Y = n_b|H_2) P(H_2)}{P(Y = n_b|H_1) P(H_1) + P(Y = n_b|H_2) P(H_2)}.$$

A razão destas duas “probabilidades” é

$$\begin{aligned} \frac{P(H_1|Y = n_b)}{P(H_2|Y = n_b)} &= \frac{P(Y = n_b|H_1) P(H_1)}{P(Y = n_b|H_2) P(H_2)} = \frac{P(Y = n_b|H_1)}{P(Y = n_b|H_2)} \frac{P(H_1)}{P(H_2)} \\ &= \mathcal{L}_R\{H_1, H_2|Y = n_b\} \frac{P(H_1)}{P(H_2)} = 2^{n_b} \frac{P(H_1)}{P(H_2)}. \end{aligned}$$

Logo, a razão entre as “probabilidades” de duas hipóteses, dado que certo resultado foi observado, é igual ao produto da razão de verossimilhança e a razão das “probabilidades” destas hipóteses a priori. No jargão do Teorema de Bayes, a razão das “probabilidades” *a posteriori* de duas hipóteses é o produto da razão de verossimilhança pela a razão *a priori* das hipóteses.

A razão de verossimilhança é portanto o fator que modula ou transforma a razão das “probabilidades” de duas hipóteses *antes* de se observar os dados (razão a priori) pela evidência contida nos dados a respeito das duas hipóteses. Um dado valor numérico da razão de verossimilhança tem o mesmo significado de alterar a razão das “probabilidades” a priori independentemente da situação de aplicação prática. Se a razão de verossimilhança for 8, isto implica que os resultados observados aumentam em oito vezes a “probabilidade” da hipótese H_1 sobre a hipótese H_2 .

O mais importante é que este significado permanece válido mesmo que não se disponha ou que não exista a razão das “probabilidades” a priori. Logo, a interpretação da razão de verossimilhança não necessita da aplicação do Teorema de Bayes e, portanto, a inferência por “probabilidade inversa” pode ser dispensada.

8.3 Função de Log-verossimilhança

Que tipo de medida matemática é a verossimilhança? O que a verossimilhança mede exatamente? Edwards (1992) usa a expressão “*medida de suporte*” no sen-

tido que ela indica quanto suporte os resultados observados dão para uma certa hipótese. Já Royall (1997) prefere falar em “*medida relativa de evidência*”, isto é, os resultados observados são uma evidência que favorece melhor qual de duas hipóteses. Já Lele (2004) apresenta o conceito de “*função de evidência*”, mostrando que é possível se generalizar o conceito, utilizando outras funções de evidência além daquela baseada em verossimilhança, mas, em certas situações, a função de evidência baseada em verossimilhança é ótima, isto é, a melhor dentre todas funções de evidência.

A verossimilhança será designada como “*medida relativa de evidência*”, pois esta expressão enfatiza o aspecto empírico da medida, isto é, que a medida é condicionada pelos resultados observados, pelos dados. Mas a verossimilhança também pode ser entendida como uma “*medida relativa de plausibilidade*”, onde a plausibilidade é definida pelos dados.

Com o objetivo de satisfazer os requisitos esperados de uma medida matemática para inferência estatística (Edwards, 1992), será utilizada como medida de evidência o logaritmo natural da verossimilhança. A transformação logarítmica faz com que a combinação da verossimilhança por multiplicação seja transformada na combinação por adição, satisfazendo o requisito de aditividade.

Definição 8.4 (Função de Log-Verossimilhança). *No contexto de um modelo estocástico M , a medida relativa de evidência de uma dada hipótese H fornecida pela observação do resultado R , é dada pelo logaritmo natural da verossimilhança desta hipótese:*

$$\mathbf{L}\{H|R\} = \ln(\mathcal{L}\{H|R\})$$

No caso do resultado observado R consistir numa série de observações independentes dos resultados de um cenário estocástico (x_1, x_2, \dots, x_n) , a função de log-verossimilhança é o logaritmo natural do produto da função de densidade do modelo M , o que resulta na soma do logaritmo natural desta densidade:

$$\mathbf{L}\{H|x_1, x_2, \dots, x_n\} = \ln \left[\prod_{i=1}^n (f(x_i|H)) \right] = \sum_{i=1}^n \ln (f(x_i|H)) .$$

Para simplificar a notação matemática, doravante a expressão “resultado observado R ” será tomada sempre como uma série de observações independentes dos resultados de um cenário estocástico, ou seja, $R = (x_1, x_2, \dots, x_n)$.



Exemplo 8-3: Regeneração Natural III

Voltando ao exemplo da regeneração natural em floresta, as duas hipóteses em estudo

- Hipótese 1: o número médio de plântulas na parcela é $\lambda_1 = 16$ (5700 ha^{-1});
- Hipótese 2: o número médio de plântulas na parcela é $\lambda_2 = 35$ (12500 ha^{-1});

são referentes ao modelo Poisson e podem ser trabalhadas através da função de log-verossimilhança:

$$\begin{aligned} \mathbf{L}\{\lambda_i|x_1, \dots, x_{10}\} &= \sum_{i=1}^{10} \ln\left(\frac{e^{-\lambda_i} \lambda_i^{x_j}}{x_j!}\right) \\ &= \sum_{i=1}^{10} [-\lambda_i + \lambda_i x_j - \ln(x_j!)] \\ &= -10 \lambda_i + \lambda_i \sum_{j=1}^{10} x_j - \sum_{j=1}^{10} \ln(x_j!) \end{aligned}$$

A tabela abaixo apresenta a log-verossimilhança das hipóteses segundo cada observação:

Parcela	Número de Plântulas	Log-verossimilhança	
		Hipótese 1	Hipótese 2
1	26	-5,1744	-3,8227
2	29	-6,8520	-3,1519
3	25	-4,6889	-4,1199
4	25	-4,6889	-4,1199
5	18	-2,4888	-7,3992
6	24	-4,2426	-4,4564
7	29	-6,8520	-3,1519
8	23	-3,8371	-4,8337
9	25	-4,6889	-4,1199
10	24	-4,2426	-4,4564
Soma		-47,75617	-43,63186

Um aspecto importante no uso da função de log-verossimilhança é que os valores numéricos permanecem numa escala de tratamento bem mais prático e fácil que os valores da função de verossimilhança, principalmente no caso de grandes amostras.



A função de log-verossimilhança em si não é interpretável, uma vez que, num sistema de inferência estatística, a evidência fornecida pelos dados a respeito de uma hipótese só pode ser julgada em comparação com uma outra hipótese, isto é, somente em termos de *evidência relativa*. Na prática do uso da verossimilhança como forma de inferência, a função de log-verossimilhança só será utilizada com base na razão de verossimilhança, isto é, como uma função de log-verossimilhança relativa.

Definição 8.5 (Função de Log-Verossimilhança Relativa). *No contexto de um modelo estocástico M , a medida de evidência relativa para duas hipóteses H_1 e H_2 fornecida pelo resultado observado R é o logaritmo natural da razão de verossimilhança das hipóteses, sendo o mesmo que a diferença da log-verossimilhança das duas hipóteses.*

$$\begin{aligned} \mathbf{L}_R\{H_1, H_2|R\} &= \ln(\mathcal{L}_R\{H_1, H_2|R\}) \\ &= \ln\left(\frac{\mathcal{L}\{H_1|R\}}{\mathcal{L}\{H_2|R\}}\right) = \mathbf{L}\{H_1|R\} - \mathbf{L}\{H_2|R\}. \end{aligned}$$



Exemplo 8-4: Regeneração Natural IV

Relembrando as duas hipóteses em estudo do exemplo da regeneração natural em floresta referentes ao modelos Poisson:

- Hipótese 1: o número médio de plântulas na parcela é $\lambda_1 = 16$ (5700 ha^{-1});
- Hipótese 2: o número médio de plântulas na parcela é $\lambda_2 = 35$ (12500 ha^{-1}),

A função de log-verossimilhança relativa neste caso é:

$$\begin{aligned} \mathbf{L}_R\{\lambda_2, \lambda_1|R\} &= \mathbf{L}\{\lambda_2|x_j\} - \mathbf{L}\{\lambda_1|x_j\} \\ &= \sum_{j=1}^{10} [-\lambda_2 + \lambda_2 x_j - \ln(x_j!)] - \sum_{j=1}^{10} [-\lambda_1 + \lambda_1 x_j - \ln(x_j!)] \\ &= \sum_{j=1}^{10} [-\lambda_2 + \lambda_2 x_j - \ln(x_j!)] - [-\lambda_1 + \lambda_1 x_j - \ln(x_j!)] \\ &= (\lambda_2 - \lambda_1) \left[\sum_{j=1}^{10} x_j - 10 \right] \end{aligned}$$

A tabela abaixo apresenta a log-verossimilhança das hipóteses e a log-verossimilhança relativa segundo cada observação:

Parcela	Número de Plântulas	Log-verossimilhança		
		Hipótese 1	Hipótese 2	Relativa
1	26	-5,1744	-3,8227	1,3517
2	29	-6,8520	-3,1519	3,7000
3	25	-4,6889	-4,1199	0,5690
4	25	-4,6889	-4,1199	0,5690
5	18	-2,4888	-7,3992	-4,9103
6	24	-4,2426	-4,4564	-0,2138
7	29	-6,8520	-3,1519	3,7000
8	23	-3,8371	-4,8337	-0,9965
9	25	-4,6889	-4,1199	0,5690
10	24	-4,2426	-4,4564	-0,2138
Soma		-47,75617	-43,63186	4,124316

A log-verossimilhança relativa favorece a hipótese H_2 , sobre a H_1 , por uma diferença maior que quatro. Em termos de verossimilhança, esta diferença implica numa razão de $\exp(4,124316) = 61,82551 \approx 62$. A tabela mostra que a contribuição das observações individuais para a log-verossimilhança relativa não é a mesma. Algumas, observações favorecem a hipótese H_1 , enquanto outras favorecem a hipótese H_2 . A observação 5 favorece de sobre madeira a hipótese H_1 , mas as duas observações de 29 plântulas (observações 2 e 7) são as que fazem o conjunto favorecer a hipótese H_2 sobre a H_1 .



A função de log-verossimilhança relativa será utilizada para se obter uma estimativa e para se realizar inferências sobre esta estimativa, ou seja, ela também será utilizada para se avaliar a qualidade da estimativa obtida. Para se obter uma *função relativa*, no entanto, é necessário se tomar uma das hipóteses como valor de referência para o cálculo da função. Alguns fundamentos são necessários para se entender qual é esta hipótese de referência.

Axioma da Verossimilhança

Os dois fundamentos para o sistema de estimação baseado em verossimilhança são a Lei da Verossimilhança e o Princípio da Verossimilhança, juntos eles formam, na

expressão de Edwards (1992), o *Axioma da Verossimilhança*. Estes fundamentos podem ser estabelecidos em termos da função de log-verossimilhança relativa.

Definição 8.6 (Lei da Verossimilhança). *No contexto de um modelo estocástico M , o fato do resultado R ter sido observado é uma evidência que favorece a hipótese H_1 vis-à-vis a hipótese H_2 , se a função de log-verossimilhança relativa destas hipóteses for maior que zero:*

$$L_R\{H_1, H_2|R\} > 0.$$

O valor da função de log-verossimilhança relativa mede o grau de evidência com que o resultado R favorece a hipótese H_1 vis-à-vis a hipótese H_2 .

Tomando as hipóteses como referentes aos valores do parâmetro do modelo estocástico, a Lei da Verossimilhança estabelece a função de log-verossimilhança relativa como a medida com que os resultados empíricos (R) dirigem o processo de estimação do parâmetro. Mas é necessário um outro fundamento para estabelecer a função de log-verossimilhança relativa como a função mais apropriada para estimação.

Definição 8.7 (Princípio da Verossimilhança). *No contexto de um modelo estocástico e face um certo resultado observado, a função de log-verossimilhança relativa de duas hipóteses quaisquer contém toda evidência contida no resultado observado referente às duas hipóteses.*

O Princípio da Verossimilhança estabelece que no contexto de um modelo estocástico e condicionado a um certo conjunto de dados, a comparação entre duas hipóteses pode ser realizada exclusivamente com base na função de log-verossimilhança, pois ela contém toda informação a respeito das duas hipóteses que os dados podem fornecer. Nenhuma outra informação ou dispositivo é necessário para a comparação das hipóteses.

O Princípio da Verossimilhança afirma que, embora a função de log-verossimilhança relativa seja uma medida *relativa* a duas hipóteses, ela pode ser interpretada de modo *absoluto*. Em todas as situações de modelos estocásticos e de observações empíricas, um certo valor da função de log-verossimilhança relativa é sempre a mesma medida relativa, isto é, ele indica o *mesmo grau de evidência* com que os dados favorecem uma hipótese sobre outra. Para enfatizar esta implicação, o Princípio de verossimilhança pode ser apresentado de um modo alternativo, conforme Royall (1997).

Definição 8.8 (Princípio da Verossimilhança II). *No contexto de um modelo X_1 , duas hipóteses H_A e H_B são comparadas com base no resultado observado R_1 . No contexto de um outro modelo X_2 , duas hipóteses H_C e H_D são comparadas*

com base no resultado observado R_2 . Se a função de log-verossimilhança relativa das hipóteses H_A e H_B for igual à função das hipóteses H_C e H_D , isto é,

$$\mathbf{L}_R\{H_A, H_B|R_1\} = \mathbf{L}_R\{H_C, H_D|R_2\} = K$$

então, o resultado R_1 , no contexto do modelo X_1 , favorece a hipótese H_A sobre a H_B com o mesmo grau de evidência K com que o resultado R_2 favorece a hipótese H_C sobre a H_D no contexto do modelo X_2 .

O Princípio da Verossimilhança afirma que a função de log-verossimilhança relativa é a medida *exclusiva* e *universal* para comparação de duas hipóteses no contexto de qualquer modelo estocástico.

Embora a Lei da Verossimilhança seja de aceitação geral entre os estatísticos, muitos profissionais da área não aceitam o Princípio da Verossimilhança. Contudo, Berger e Wolpert (1988) escreveram uma monografia onde fazem uma profunda e extensa defesa do Princípio da Verossimilhança, rebatendo as críticas contrárias a ele. Mais ainda, Hacking (1965) e Birnbaum (1962, 1972) demonstraram que tanto a lei quanto o princípio são implicações lógicas de dois princípios de inferência ainda mais fundamentais, que são o Princípio da Condicionalidade e o Princípio da Suficiência. Curiosamente, estes dois princípios mais básicos raramente são questionados pelos estatísticos.

8.4 Estimativa de Máxima Verossimilhança

A Lei e o Princípio da Verossimilhança estabelecem os dois fundamentos para a estimação dos parâmetros dos modelos estocásticos. Destes dois fundamentos decorre que, no contexto de um modelo estocástico e considerando exclusivamente os resultados observados, a melhor estimativa para o parâmetro do modelo é aquela que maximiza a função de log-verossimilhança. Esta estimativa é chamada de “*estimativa de máxima verossimilhança*” ou, no inglês, “*Maximum Likelihood Estimate*” (MLE).

Definição 8.9 (Estimativa de Máxima Verossimilhança (MLE)). *No contexto de um modelo estocástico M , com função de densidade $f(\cdot; \theta)$, foi observado um certo resultado R . A estimativa de máxima verossimilhança (Maximum Likelihood Estimate - MLE) do parâmetro θ do modelo é aquela que maximiza a função de log-verossimilhança:*

$$\hat{\theta} = \max_{\arg(\theta)} \mathbf{L}\{\theta|R\}$$

A MLE pode ser pensada como uma estimativa com máxima evidência relativa ou de máxima plausibilidade relativa. Ela é o valor do parâmetro de um modelo cuja evidência contida nos resultados observados favorece ao máximo em comparação com todos os valores possíveis.

Estimação por Máxima Verossimilhança

O conjunto de todos os valores que o parâmetro do modelo estocástico pode assumir é chamado de “*espaço paramétrico*” do modelo. Numa linguagem mais matemática, a MLE pode ser definida como o valor do parâmetro que maximiza a função de log-verossimilhança no espaço paramétrico.

Definição 8.10 (Estimativa de Máxima Verossimilhança (MLE) II). *Num modelo estocástico M , com função de densidade $f(\cdot; \theta)$, a estimativa de máxima verossimilhança (Maximum Likelihood Estimate - MLE) é valor do parâmetro θ que, dado o resultado R , maximiza a função de log-verossimilhança no espaço paramétrico Θ .*

O problema de estimação por máxima verossimilhança consiste, portanto, no clássico problema matemático de encontrar o ponto de máximo de uma função. A função a ser estudada é a função de log-verossimilhança, que é o logaritmo natural da função densidade do modelo estocástico, mas com os resultados (dados) fixos e o parâmetro variável. Doravante, para fins de simplificação e clareza, serão omitidas as referências aos resultados observados, uma vez que eles são um elemento constante na função a ser maximizada.

Se a função de log-verossimilhança for contínua no espaço paramétrico, a MLE é obtida encontrando o valor do parâmetro que torna a primeira derivada nula:

$$\hat{\theta} \Rightarrow \frac{dL\{\theta\}}{d\theta} = 0$$

Edwards (1992) denomina esta expressão de “*equação de suporte*”. Assim, pode-se dizer que a MLE é a solução da equação de suporte.

Quando o resultado observado consiste numa série de observações independentes de um cenário estocástico ($x_i; i = 1, 2, \dots, n$), a primeira derivada da função de log-verossimilhança se transforma na soma das primeiras derivadas do logaritmo natural da função de densidade das observações e a equação de suporte toma a forma de um somatório:

$$\begin{aligned} \frac{dL\{\theta\}}{d\theta} = 0 &\Rightarrow \frac{d}{d\theta} \sum_{i=1}^n \ln(f(x_i; \theta)) = \sum_{i=1}^n \frac{d}{d\theta} \ln(f(x_i; \theta)) = 0 \\ &\Rightarrow \sum_{i=1}^n \frac{1}{f(x_i; \theta)} \frac{df(x_i; \theta)}{d\theta} = 0 \end{aligned}$$

Tabela 8.2: Estimadores de máxima verossimilhança (MLE) para algumas distribuições estocásticas apresentadas anteriormente, com base numa amostra de n observações independentes.

DISTRIBUIÇÃO	PARÂMETROS	ESTIMADOR DE MÁXIMA VEROSSIMILHANÇA	
<i>Distribuições Discretas</i>			
Binomial	N, p	N conhecido	$\hat{p} = \sum_{i=1}^m x_i / (N n)$
Poisson	λ	$\hat{\lambda} = \sum_{i=1}^n x_i / n$	
Geométrica	p	$\hat{p} = n / (\sum_{i=1}^n x_i + n)$	
<i>Distribuições Contínuas</i>			
Exponencial	λ	$\hat{\lambda} = n / \sum_{i=1}^n x_i$	

Os estimadores de máxima verossimilhança (MLE) são facilmente obtidos para as algumas das distribuições estocásticas de uso comum na Biometria Florestal. A tabela 8.2 apresenta os estimadores de algumas distribuições apresentadas no capítulo anterior.

Métodos Numéricos

Um aspecto importante é que nem sempre a equação de suporte tem uma solução analítica para θ , isto é, não existe uma expressão algébrica deduzida da equação de suporte em que o parâmetro θ esteja isolado dos demais termos da equação. Nestes casos, a única forma de se obter a MLE é pelos métodos numéricos de solução de equações. Mais ainda, existem modelos estocásticos tão complexos, que a função de log-verossimilhança é demasiadamente complicada para se deduzir a equação de suporte. Para tais modelos, a única solução é utilizar métodos de otimização que maximizem a própria função de log-verossimilhança.

Do ponto de vista do Biometrista, a maior complexidade dos problemas implica não só na necessidade de um certo embasamento teórico em métodos numéricos e em métodos de otimização, mas também no domínio de softwares que permitam a implementação das rotinas de otimização dos modelos de interesse. Este “investimento” num maior conhecimento matemático, no entanto, gera um

retorno na forma de maior realismo dos modelos e de maior flexibilidade na solução de problemas práticos importantes.



Exemplo 8-5: Distribuição de Abundância de Espécies

A distribuição da série logarítmica ou distribuição log-série foi proposta por Ronald Fisher como modelo estocástico para a distribuição de abundância de espécies. Fisher a desenvolveu para ajustar dois conjuntos de dados referentes à abundância de espécies de mariposas capturadas em armadilhas luminosas e à abundância de uma coleção de espécies de borboleta tropicais. A distribuição da série logarítmica apresentada por Fisher relaciona o número de espécies observados na amostra com a distribuição do número de indivíduos (abundância) por espécie (Fisher, Corbet e Williams, 1943; Kempton e Taylor, 1974; Kempton, 1975). Partindo da distribuição Poisson e da Binomial Negativa, como uma generalização da Poisson, Fisher desenvolveu esta distribuição como uma série de números inteiros na forma

$$\alpha \theta, \frac{\alpha \theta^2}{2}, \frac{\alpha \theta^3}{3}, \dots, \frac{\alpha \theta^n}{n},$$

em que o primeiro termo é o *número esperado* de espécies cuja abundância é um indivíduo, o segundo termo é o número esperado de espécies com dois indivíduos, e assim sucessivamente (Fisher, Corbet e Williams, 1943). Posteriormente, o parâmetro α passou a ser identificado como um índice de diversidade das espécies da comunidade (Magurran, 1988)

Fisher apresenta duas relações que associam o índice α com duas informações que podem ser obtidas numa amostra, isto é, numa série de abundâncias observadas:

(i) o número de espécies (S) $\Rightarrow \alpha = -S / \ln(1 - \theta)$, para $0 < \theta < 1$;

(ii) e o número de indivíduos na amostra (N) $\Rightarrow N = \alpha \theta / (1 - \theta)$.

Unindo estas duas relações, Fisher deduz a equação

$$S = \alpha \ln \left(1 + \frac{N}{\alpha} \right),$$

cujas solução numérica em α , a partir de S e N conhecidos, produz a estimativa do índice de diversidade $\hat{\alpha}$.

MLE da Distribuição Log-série

Partindo do conceito de distribuição estocástica, uma caminho diferente e mais geral pode ser encontrado para se obter a estimativa de α . Tomando o número de espécies numa comunidade como S^* e considerando que a série logarítmica é uma *distribuição estocástica discreta*, o número esperado de espécies com uma dada abundância x pode ser apresentado por

$$S^* P(X = x) = S^* f(x) = \alpha \frac{\theta^x}{x} \quad \Rightarrow \quad f(x) = \frac{\alpha}{S^*} \frac{\theta^x}{x}.$$

Aplicando a relação entre o índice α e o número de espécies na expressão acima, a função de densidade se torna:

$$\alpha = -S/\ln(1 - \theta) \quad \Rightarrow \quad f(x) = -\frac{\theta^x}{x \ln(1 - \theta)}.$$

A série logarítmica pode ser apresentada de forma teórica, isto é, sem menção a uma aplicação em particular, como uma distribuição estocástica discreta com um único parâmetro, cuja a função densidade é dada por

$$P(X = x) = f(x; \theta) = -\frac{\theta^x}{x \ln(1 - \theta)}; \quad x = 1, 2, \dots; \quad 0 < \theta < 1.$$

(Weisstein, sd; Krishnamoorthy, 2006).

Qual seria a estimativa de máxima verossimilhança (MLE) do parâmetro θ ? Partindo da função de densidade, a MLE é encontrada pela solução da primeira derivada da função de log-verossimilhança

$$\frac{d\mathbf{L}\{\theta\}}{d\theta} = 0 \quad \Rightarrow \quad \sum_{i=1}^n \frac{1}{f(x_i; \theta)} \frac{df(x_i; \theta)}{d\theta} = 0.$$

No caso da distribuição da série logarítmica, se tem

$$f(x; \theta) = -\frac{\theta^x}{x \ln(1 - \theta)} \quad \Rightarrow \quad \frac{df(x_i; \theta)}{d\theta} = -\frac{\theta^{x-1}}{\ln(1 - \theta)} - \frac{\theta^x}{x [\ln(1 - \theta)]^2 (1 - \theta)},$$

e a primeira derivada da função de log-verossimilhança se torna

$$\begin{aligned} \frac{d\mathbf{L}\{\theta\}}{d\theta} &= \sum_{i=1}^n \left(-\frac{x_i \ln(1-\theta)}{\theta^x} \right) \left(-\frac{\theta^{x_i-1}}{\ln(1-\theta)} - \frac{\theta^x}{x_i [\ln(1-\theta)]^2 (1-\theta)} \right) \\ &= \sum_{i=1}^n \left(\frac{x_i}{\theta} + \frac{1}{\ln(1-\theta)(1-\theta)} \right) = \frac{\sum_{i=1}^n x_i}{\theta} + \frac{n}{\ln(1-\theta)(1-\theta)} \end{aligned}$$

$$\frac{d\mathbf{L}\{\theta\}}{d\theta} = 0 \Rightarrow \frac{\sum_{i=1}^n x_i}{n} + \frac{\theta}{\ln(1-\theta)(1-\theta)} = 0.$$

A solução numérica desta expressão gera a MLE do parâmetro θ .

Contudo, se a parametrização de Fisher é utilizada, isto é, $N = \sum x_i$ e $S = n$, juntamente com a definição do índice α , obtém-se

$$\begin{aligned} \frac{\sum_{i=1}^n x_i}{n} = \frac{N}{S} = -\frac{\theta}{\ln(1-\theta)(1-\theta)} &\Rightarrow N = \alpha \theta / (1-\theta) \\ &\Rightarrow S = \alpha \ln \left(1 + \frac{N}{\alpha} \right). \end{aligned}$$

Portanto, embora Fisher não mencione em seu trabalho de 1943, a estimativa do índice de diversidade α é uma estimativa de máxima verossimilhança, obtida pela solução da relação entre o número de indivíduos e o número de espécies na amostra por métodos numéricos.

Estimador pelo Método dos Momentos

Curiosamente a MLE é também a mesma estimativa obtida pelo método dos momentos. A esperança matemática da distribuição, ou seja, o primeiro momento da distribuição, é

$$\begin{aligned} E\{X\} &= \sum_{i=1}^{\infty} x_i \left(-\frac{\theta^x}{x_i \ln(1-\theta)} \right) = -\frac{1}{\ln(1-\theta)} \sum_{i=1}^{\infty} \theta^x \\ E\{X\} &= -\frac{\theta}{(1-\theta) \ln(1-\theta)}. \end{aligned}$$

Numa amostra, a abundância média é a razão do número de indivíduos na amostra (N) pelo número de espécies na amostra S , ou seja, $\bar{x} = N/S$. Logo, tem-se a

seguinte relação empírica

$$\frac{N}{S} = - \frac{\theta}{(1-\theta) \ln(1-\theta)} \Rightarrow S = \alpha \ln \left(1 + \frac{N}{\alpha} \right).$$

A qual resulta na mesma equação proposta por Fisher para estimação do índice α .



Curva de Log-verossimilhança e Incerteza associada à MLE

A representação gráfica é sempre um auxílio poderoso à compreensão de conceitos matemáticos, por isto, o gráfico da função de log-verossimilhança será chamada de “*curva de log-verossimilhança*”. Ao construir uma representação gráfica da função de log-verossimilhança é importante considerar que os valores de verossimilhança e log-verossimilhança não tem interpretação absoluta, mas apenas relativa. Assim a curva de log-verossimilhança toma com ponto de referência a estimativa de máxima verossimilhança (MLE) do parâmetro e apresenta a função de log-verossimilhança *relativa*, de modo que o ponto de máximo da curva seja o MLE.

Definição 8.11 (Curva de Log-verossimilhança). *A curva de log-verossimilhança de um modelo estocástico com parâmetro θ é o gráfico da função de log-verossimilhança relativa (ordenadas) em função dos valores do parâmetro (abscissas): $\mathbf{L}_R\{\theta, \hat{\theta}\} \times \theta$. A função de log-verossimilhança relativa é definida como*

$$\begin{aligned} \mathbf{L}_R\{\theta, \hat{\theta}\} &= \ln(\mathcal{L}_R\{\theta, \hat{\theta}\}) = \ln \left(\frac{\mathcal{L}\{\theta\}}{\mathcal{L}\{\hat{\theta}\}} \right) \\ &= \mathbf{L}\{\theta\} - \mathbf{L}\{\hat{\theta}\} = \mathbf{L}\{\theta\} - \max[\mathbf{L}\{\theta\}], \end{aligned}$$

onde $\hat{\theta}$ é a estimativa de máxima verossimilhança (MLE).

A forma da curva de log-verossimilhança depende da definição funcional do modelo estocástico em questão, mas o *grau de curvatura* na vizinhança da estimativa de máxima verossimilhança é uma indicação da incerteza associada à estimativa. Curvas de log-verossimilhança suaves, isto é, com pequena curvatura, indicam uma incerteza maior, pois os valores do parâmetro poderá mudar sem muita alteração no valor da função de log-verossimilhança. Ao contrário, quando o grau de curvatura é elevado, pequenas mudanças no valor do parâmetro causam grandes mudanças no valor da função de log-verossimilhança, o que indica uma menor incerteza a respeito da MLE. A curvatura da função de log-verossimilhança negativa é controlada pelo tamanho da amostra.

Aproximação Quadrática

A expansão de Taylor na vizinhança da MLE ($\hat{\theta}$)

$$\mathbf{L}\{\theta\} = \mathbf{L}\{\hat{\theta}\} + (\theta - \hat{\theta}) \frac{d\mathbf{L}\{\hat{\theta}\}}{d\theta} + \frac{1}{2} (\theta - \hat{\theta})^2 \frac{d^2\mathbf{L}\{\hat{\theta}\}}{d\theta^2} + \frac{1}{3} (\theta - \hat{\theta})^3 \frac{d^3\mathbf{L}\{\hat{\theta}\}}{d\theta^3} + \dots$$

fornece uma aproximação polinomial para a curva de log-verossimilhança. No ponto da estimativa de máxima verossimilhança ($\hat{\theta}$) a primeira derivada é nula e, portanto, o termo de primeiro grau da expansão de Taylor também é nulo.

Como a expansão é na vizinhança da MLE, as diferenças dos valores do parâmetro para a MLE vão rapidamente para zero à medida que a potência cresce. Em muitos modelos estocásticos, esta tendência é bem mais rápida nas derivadas de ordem de derivação maior, de modo que os termos com ordem superior ao quadráticos se tornam negligenciáveis:

$$\lim_{r \rightarrow \infty} (\theta - \hat{\theta})^r \frac{d^r \mathbf{L}\{\hat{\theta}\}}{d^r \theta} = 0.$$

Este efeito tende a ser ampliado com o tamanho da amostra, logo, para amostras *razoavelmente grandes*, a função de log-verossimilhança pode ser aproximada pela expansão de Taylor até o termo de segunda ordem, ou seja, a curva de log-verossimilhança pode ser aproximada por uma curva parabólica.

Definição 8.12 (Aproximação Quadrática da Curva de Log-Veros.). *Para amostras razoavelmente grandes, a função de log-verossimilhança pode ser aproximada na vizinhança da estimativa de máxima verossimilhança (MLE — $\hat{\theta}$) pela expansão de Taylor até o termo de segunda ordem:*

$$\mathbf{L}\{\theta\} \approx \mathbf{L}\{\hat{\theta}\} + \frac{1}{2} (\theta - \hat{\theta})^2 \frac{d^2 \mathbf{L}\{\hat{\theta}\}}{d\theta^2}.$$

Assim, a aproximação quadrática para curva de log-verossimilhança é dada por

$$\mathbf{L}_R\{\theta, \hat{\theta}\} \approx \frac{1}{2} (\theta - \hat{\theta})^2 \frac{d^2 \mathbf{L}\{\hat{\theta}\}}{d\theta^2},$$

que descreve uma curva parabólica com vértice no ponto $(\hat{\theta}, 0)$.



Exemplo 8-6: Estrutura de Tamanho de Floresta Tropical

A distribuição exponencial foi estudada como modelo estocástico para a distribuição do DAP, como representando a estrutura de tamanho, de uma floresta tropical

do município de Bom Jardim, Maranhão (veja figura 7.21 (a), referente ao exemplo 7.5, na página 172). O modelo exponencial é um modelo com um único parâmetro cuja estimativa de máxima verossimilhança (MLE) é apresentada na tabela 8.2. A partir da função de densidade

$$f(x; \lambda) = \lambda \exp(-\lambda x); \quad x \geq 0; \quad \lambda > 0;$$

tem-se a função de log-verossimilhança

$$\mathbf{L}\{\lambda\} = \sum_{i=1}^n \ln[\lambda \exp(-\lambda x_i)] = n \ln(\lambda) - \lambda \sum_{i=1}^n x_i,$$

que define o gráfico da curva de log-verossimilhança. A partir da sua primeira derivada se obtém a MLE

$$\frac{d \mathbf{L}\{\lambda\}}{d \lambda} = \frac{n}{\lambda} - \sum_{i=1}^n x_i = 0 \quad \Rightarrow \quad \hat{\lambda} = \frac{n}{\sum_{i=1}^n x_i}.$$

No caso da distribuição de DAP da floresta tropical de Bom Jardim, foram medidas 11 986 árvores ($n = 11\,986$), mas o diâmetro mínimo para medição foi de 15 cm, de modo que a variável para distribuição exponencial é $x_i = d_i - 15$. A relação da MLE com o DAP médio aritmético se torna:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n (d_i - 15)} \quad \Rightarrow \quad \bar{d} = \frac{1}{\hat{\lambda}} + 15 = 25,41186 \text{ cm}.$$

A segunda derivada da função de log-verossimilhança é

$$\frac{d^2 \mathbf{L}\{\lambda\}}{d \lambda^2} = -\frac{n}{\lambda^2},$$

que é necessariamente negativa, provando que a MLE é o ponto de máximo da função de log-verossimilhança. Nota-se que a derivada de ordem r será dada pela expressão

$$\frac{d^r \mathbf{L}\{\lambda\}}{d \lambda^r} = (-1)^{r-1} \frac{n^{r-1}}{\lambda^r}.$$

Assim, no ponto da MLE ($\hat{\lambda}$) as derivadas da função de log-verossimilhança se tornam:

$$\frac{d^r \mathbf{L}\{\hat{\lambda}\}}{d \lambda^r} = (-1)^{r-1} \frac{n^{r-1}}{\hat{\lambda}^r}$$

e os diferentes termos da expansão de Taylor são:

$$(\lambda - \hat{\lambda})^r \frac{d^r \mathbf{L}\{\hat{\lambda}\}}{d\lambda^r} = (\lambda - \hat{\lambda})^r (-1)^{r-1} \frac{n^{r-1}}{\hat{\lambda}^r}.$$

No caso da floresta tropical de Bom Jardim, os termos da expansão de Taylor para uma vizinhança de 5% da MLE ($\lambda - \hat{\lambda} = 0,05 \hat{\lambda}$) resultam nos seguintes valores:

Ordem (r)	Valor do Termo
2	$2,996500 \times 10^{01}$
3	$-1,498250 \times 10^{00}$
4	$7,491250 \times 10^{-02}$
5	$-3,745625 \times 10^{-03}$
6	$1,872813 \times 10^{-04}$
7	$-9,364063 \times 10^{-06}$
8	$4,682031 \times 10^{-07}$
9	$-2,341016 \times 10^{-08}$
10	$1,170508 \times 10^{-09}$

Note-se que, neste caso, os termos de quarta ordem ou maiores podem ser considerados negligenciáveis, enquanto que o termos de terceira ordem, embora uma ordem de grandeza menor que o termo quadrático, ainda tem certa relevância.

A aproximação quadrática, na vizinhança da MLE, para curva de log-verossimilhança da distribuição exponencial é dada por

$$\begin{aligned} \mathbf{L}\{\lambda\} &\approx \mathbf{L}\{\hat{\lambda}\} + \frac{1}{2} (\lambda - \hat{\lambda})^2 \frac{d^2 \mathbf{L}\{\hat{\lambda}\}}{d\lambda^2} \\ &\approx n \left[\ln(\hat{\lambda}) - 1 \right] - \frac{1}{2} (\lambda - \hat{\lambda})^2 \left(\frac{n}{\hat{\lambda}^2} \right). \end{aligned}$$

A figura 8.1 apresenta vários gráficos da curva de verossimilhança para os dados da estrutura de tamanho da floresta tropical de Bom Jardim, MA. Os gráficos exemplificam a aproximação quadrática da curva de log-verossimilhança, mostrando que, à medida que o gráfico se restringe a uma vizinhança gradativamente menor ao redor da MLE, a aproximação quadrática se mostra progressivamente mais apropriada.



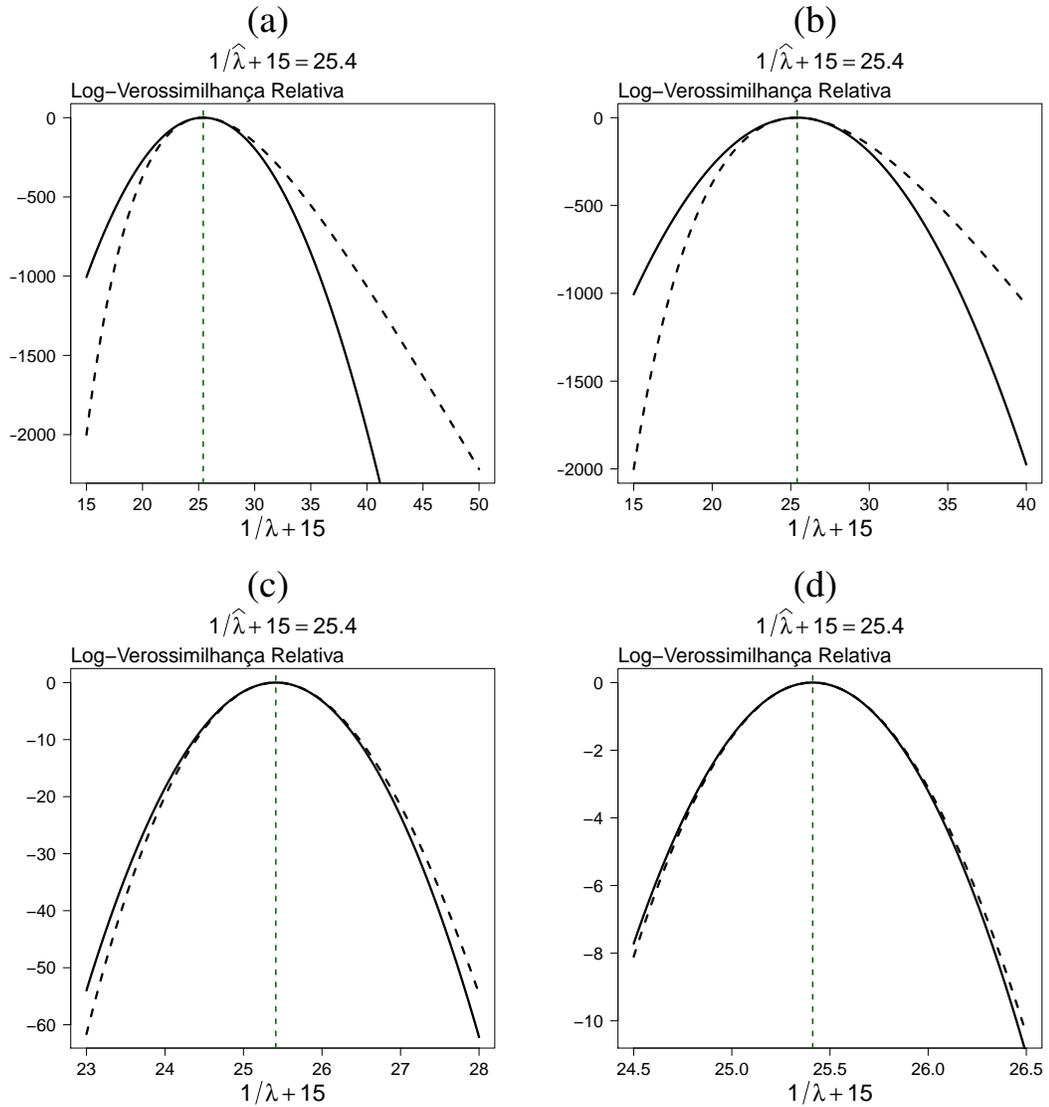


Figura 8.1: Distribuição exponencial como modelo estocástico para distribuição de DAP de floresta tropical no município de Bom Jardim, Maranhão. Os gráficos mostram a curva de log-verossimilhança (linha tracejada) para a estimativa de máxima verossimilhança do parâmetro da distribuição exponencial ($\hat{\lambda}$) e sua aproximação quadrática (linha sólida). À medida que o gráfico retrata uma vizinhança mais próxima de $\hat{\lambda}$ (gráficos de (a) para (d)), a aproximação quadrática se mostra mais apropriada.

Informação Observada

Como a MLE é o ponto de *máximo* da curva de log-verossimilhança, a sua segunda derivada em relação ao parâmetro é negativa, o que implica que a aproximação é dada por uma curva parabólica côncava, cujo *grau de curvatura* é a segunda derivada avaliada na MLE. Quando o resultado observado consiste numa série de observações independentes do modelo estocástico $M(x_i; i = 1, 2, \dots, n)$ com densidade $f(\cdot; \theta)$, a segunda derivada da função de log-verossimilhança no ponto da MLE é dada por

$$\begin{aligned} \frac{d^2 \mathbf{L}\{\hat{\theta}\}}{d\theta^2} &= \frac{d^2}{d\theta^2} \sum_{i=1}^n \ln(f(x_i; \hat{\theta})) = \sum_{i=1}^n \frac{d^2}{d\theta^2} \ln(f(x_i; \hat{\theta})) \\ &= \sum_{i=1}^n \left[\frac{1}{f(x_i; \hat{\theta})} \frac{d^2 f(x_i; \hat{\theta})}{d\theta^2} - \frac{1}{[f(x_i; \hat{\theta})]^2} \frac{d f(x_i; \hat{\theta})}{d\theta} \right]. \end{aligned}$$

Como se trata de uma soma de n termos, o valor absoluto tenderá a crescer à medida que o número de observações independentes (n) cresce. A implicação prática desta propriedade matemática é que à medida que o tamanho da amostra (n) cresce, a curvatura da curva de log-verossimilhança aumenta na vizinhança da MLE, isto é, a curva se “fecha” nas proximidades da MLE, reduzindo a incerteza associada à estimativa. Edwards (1992) designa o valor negativo da segunda derivada da função de log-verossimilhança por “*informação*”.

Definição 8.13 (Informação da Curva de Log-Verossimilhança). A “*informação*” a respeito de um dado valor possível para o parâmetro de um modelo estocástico (θ) é dada pela curvatura da curva de log-verossimilhança no ponto correspondente a este valor dentro do espaço paramétrico ($\theta \in \Theta$), sendo portanto, obtida pela segunda derivada da função de log-verossimilhança negativa:

$$I(\theta) = - \frac{d^2 \mathbf{L}\{\theta\}}{d\theta^2}.$$

Se o ponto em questão corresponde à estimativa de máxima verossimilhança (MLE— $\hat{\theta}$), então a curvatura é chamada de “*informação observada*” ($I(\hat{\theta})$).

Esta designação sugere que o grau de curvatura da curva de log-verossimilhança é uma medida da quantidade de informação que se tem a respeito do valor do parâmetro do modelo estocástico. No caso da MLE, a “*informação observada*” é uma medida da informação que se tem a respeito da estimativa, quanto maior a curvatura da curva de log-verossimilhança na vizinhança da MLE, mais informação observada se tem e, conseqüentemente, menor incerteza a respeito da estimativa.

 *Exemplo 8-7: Distribuição Binomial e Tamanho de Amostra*

A distribuição binomial pode ser utilizada para exemplificar a influência do tamanho da amostra sob a curvatura da curva de log-verossimilhança do seu parâmetro p (probabilidade de sucesso). A função de densidade da distribuição binomial é

$$f(x; N, p) = \binom{N}{x} p^x (1-p)^{N-x},$$

$$N \in \mathbb{N}, 0 \leq p \leq 1; \quad x = 0, 1, \dots, N.$$

Dado que o parâmetro N é conhecido, a função de log-verossimilhança para uma amostra com n observações independentes tem a seguinte forma

$$\mathbf{L}\{p\} = \sum_{i=1}^n \ln \left[\binom{N}{x_i} \right] + \ln(p) \sum_{i=1}^n x_i + n N \ln(1-p).$$

A primeira derivada da função de log-verossimilhança e respectiva MLE são

$$\frac{d\mathbf{L}\{p\}}{dp} = \frac{\sum_{i=1}^n x_i}{p} - \frac{nN - \sum_{i=1}^n x_i}{1-p} = 0 \quad \Rightarrow \quad \hat{p} = \frac{\sum_{i=1}^n x_i}{Nn}.$$

Já a segunda derivada tem a forma

$$\frac{d^2\mathbf{L}\{p\}}{dp^2} = - \left[\frac{\sum_{i=1}^n x_i}{p^2} + \frac{nN - \sum_{i=1}^n x_i}{(1-p)^2} \right],$$

que resulta na informação observada igual a

$$I(\hat{p}) = \frac{d^2\mathbf{L}\{\hat{p}\}}{dp^2} = n \left(\frac{N}{\hat{p}(1-\hat{p})} \right).$$

Assim, a curvatura da função de log-verossimilhança é diretamente proporcional ao tamanho da amostra (n). A curvatura também é inversamente proporcional ao produto $\hat{p}(1-\hat{p})$ e, portanto, à variância da distribuição binomial dividida pelo número de ensaios (N), que representa o número de realizações independentes de um evento binário. A figura 8.2 ilustra graficamente o aumento da curvatura da curva de log-verossimilhança da binomial com o aumento do tamanho da amostra.



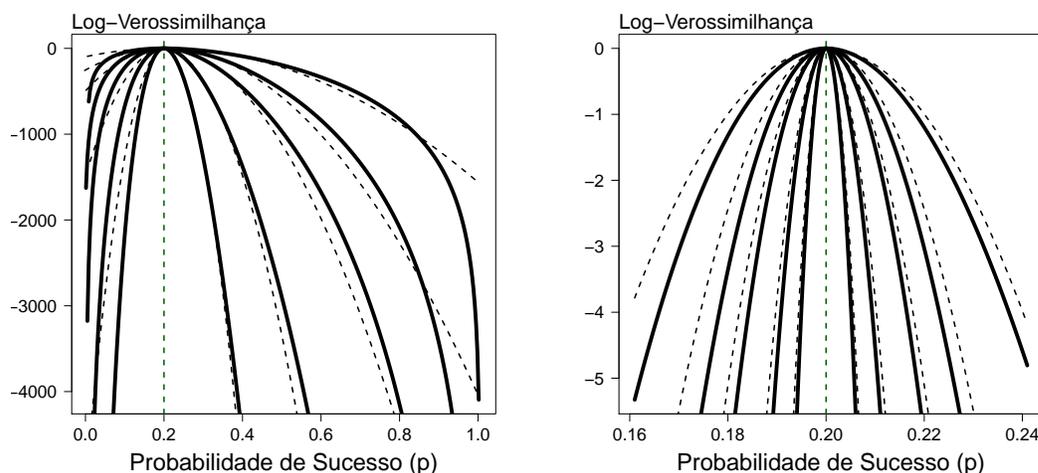


Figura 8.2: Exemplo ilustrativo do aumento da curvatura da curva de log-verossimilhança com o aumento do tamanho da amostra para distribuição binomial ($N = 50$ e $p = 0,20$). As curvas seguem, da menor curvatura para de maior curvatura, tamanhos de amostra de 20, 50, 100, 300 e 1000 observações independentes. São mostradas a curva de log-verossimilhança (linha sólida) e sua aproximação quadrática (linha tracejada). Os dois gráficos com vizinhanças de tamanho diferente mostram que à medida que o tamanho da amostra cresce a curvatura da curva de log-verossimilhança aumenta e a aproximação quadrática se torna mais apropriada.



Exemplo 8-8: Estrutura de Tamanho de Floresta Tropical II

No exemplo da estrutura de tamanho de uma floresta tropical no município de Bom Jardim, Maranhão, a distribuição exponencial foi tomada como um modelo estocástico candidato a representar as observações. A segunda derivada da função de log-verossimilhança do modelo exponencial é

$$\frac{d^2 \mathbf{L}\{\hat{\lambda}\}}{d\lambda^2} = -\frac{n}{\lambda^2},$$

logo, a informação da curva de log-verossimilhança é dada por

$$I(\lambda) = -\frac{d^2 \mathbf{L}\{\hat{\lambda}\}}{d\lambda^2} = \frac{n}{\lambda^2}.$$

Assim, a curvatura da curva de log-verossimilhança é diretamente proporcional ao tamanho da amostra (n), mas também é diretamente proporcional à variância da distribuição exponencial ($1/\lambda^2$).

Na aplicação do modelo exponencial à distribuição de DAP de uma floresta, como a de Bom Jardim, a estrutura da floresta é obtida pela observação dos DAP das árvores: d_i ($i = 1, 2, \dots, n$). Neste caso, a informação observada é

$$I(\hat{\lambda}) = \frac{n}{\hat{\lambda}^2} = \frac{n}{[n/\sum_{i=1}^n (d_i - d_{\text{MIN}})]^2} = n (\bar{d} - d_{\text{MIN}})^2,$$

sendo d_{MIN} o diâmetro mínimo para a medição do DAP.

A curvatura da aproximação quadrática do modelo exponencial aplicado à distribuição de DAP de uma floresta é proporcional não só ao tamanho da amostra (n), mas também ao quadrado da diferença entre o DAP médio e o DAP mínimo de medição. Logo, para um mesmo DAP mínimo de medição e um mesmo tamanho de amostra, a aproximação quadrática será mais apropriadas nas florestas com grandes árvores (maior DAP médio). Por outro lado, a redução do DAP mínimo de medição, por exemplo de 15 para 5 *cm*, terá um efeito duplo sobre a melhoria da aproximação quadrática do modelo exponencial, pois, num levantamento florestal com parcelas de área fixa, um maior número de árvores serão medidas (maior n) e a diferença entre o DAP médio e o DAP mínimo de medição será maior.



Intervalo de Log-Verossimilhança

A medida da informação fornecida pela curva de log-verossimilhança não é diretamente interpretável, uma vez que a própria log-verossimilhança não é interpretável. Por outro lado, tornou-se convencional associar a uma estimativa estatística um *intervalo* que caracterize a incerteza associada à estimativa. Na inferência clássica se utiliza o “intervalo de confiança”, enquanto que na inferência Bayesiana é comum se referir ao “intervalo de credibilidade” ou “intervalo de confiança Bayesiano”, ou ainda ao “intervalo de máxima densidade posterior” (*highest posterior density interval*).

Na inferência baseada na verossimilhança, a definição de um intervalo consiste na delimitação de uma vizinhança ao redor da MLE na qual os valores do parâmetro podem ser considerados igualmente plausíveis, ou seja, os valores são igualmente favorecidos pela evidência contida nas observações realizadas. A delimitação desta vizinhança necessita de um limite para a razão de verossimilhança abaixo do qual os valores da razão sejam considerados, para efeitos práticos, equivalentes em termos plausibilidade e evidência.

Estabeleceu-se uma convenção em que o valor utilizado é o valor de 8, seguindo os argumentos apresentados por Royall (1997) no seu exemplo do experimento canônico. A razão de 32 pode ser utilizado quando se deseja uma atitude mais conservadora na discriminação entre os valores do parâmetro de um modelo. O limite de 8, portanto, define o “*intervalo de log-verossimilhança*”.

Definição 8.14 (Intervalo de Log-Verossimilhança). *O intervalo de log-verossimilhança de uma estimativa de máxima verossimilhança (MLE) $\hat{\theta}$ é definido como a vizinhança de $\hat{\theta}$ na qual a razão de verossimilhança é igual ou maior a um oitavo:*

$$\mathcal{L}_R\{\theta, \hat{\theta}\} = \frac{\mathcal{L}\{\theta\}}{\mathcal{L}\{\hat{\theta}\}} \geq \frac{1}{8},$$

ou na qual a diferença na curva de log-verossimilhança é igual ou maior ao logaritmo natural negativo de oito:

$$\mathbf{L}_R\{\theta, \hat{\theta}\} = \mathbf{L}\{\theta\} - \mathbf{L}\{\hat{\theta}\} \geq -\ln(8).$$

A interpretação do intervalo de log-verossimilhança é direta e intuitiva. O intervalo de log-verossimilhança define uma vizinhança no entorno da MLE em que os valores do parâmetro são, para efeitos práticos, tão plausíveis quanto a MLE, isto é, são valores igualmente favorecidos quando comparados à MLE pela evidência contida nos dados observados. A maioria das pessoas que utiliza a inferência clássica confere esta interpretação ao intervalo de confiança, mas, do ponto de vista lógico, a interpretação do intervalo de confiança é totalmente diferente desta.

Se a curva parabólica for uma boa aproximação para a curva de log-verossimilhança na vizinhança da MLE, o intervalo de log-verossimilhança pode ser delimitado por

$$-\frac{1}{2}(\theta - \hat{\theta})^2 \frac{d^2 \mathbf{L}\{\hat{\theta}\}}{d\theta^2} = \frac{1}{2}(\theta - \hat{\theta})^2 I(\hat{\theta}) = \ln(8) \Rightarrow \theta = \hat{\theta} \pm \sqrt{\frac{2 \ln(8)}{I(\hat{\theta})}} \quad (8.1)$$

Assim, a aproximação quadrática implica num intervalo simétrico em relação à MLE, o que é aceitável somente quando a aproximação quadrática é apropriada, isto é, no caso de amostras suficientemente grandes.



Exemplo 8-9: Distribuição Binomial e Tamanho de Amostra II

Voltando ao exemplo acima da distribuição binomial, a influência da tamanho da amostra sobre a curvatura da curva de log-verossimilhança afeta diretamente a

aproximação quadrática e a qualidade do intervalo de log-verossimilhança obtido por ela. A figura 8.3 ilustra como para pequenos tamanhos de amostra a curva de log-verossimilhança pode ser bastante assimétrica, mas para tamanhos de amostra suficientemente grandes a curva se torna simétrica podendo o intervalo de log-verossimilhança ser calculado pela aproximação quadrática.

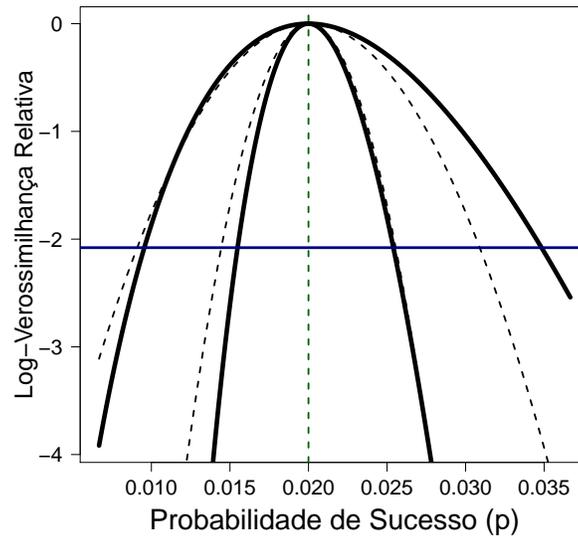


Figura 8.3: Influência do tamanho da amostra sobre a qualidade do intervalo de log-verossimilhança gerado pela aproximação quadrática em função da curvatura da curva de log-verossimilhança. As curvas são referentes à distribuição binomial ($N = 30$ e $p = 0,02$), sendo que a de menor curvatura tem tamanho de amostra 20 e a de maior curvatura tamanho de amostra 100. São mostradas a curva de log-verossimilhança (linha tracejada) e sua aproximação quadrática (linha sólida). A linha horizontal representa a diferença de log-verossimilhança igual a $-\ln(8)$.



Propriedades das MLE

Na inferência baseada em verossimilhança, a aceitação do Princípio da Verossimilhança implica na aceitação de que as MLE são as melhores estimativas possíveis e

que o seu estudo se reduz ao estudo da curva de log-verossimilhança. No entanto, as MLE também são tomadas como boas estimativas nas abordagens clássica e Bayesiana, devido às suas propriedades no cenário de “probabilidade de longo prazo”, isto é, sob condições de amostragem repetidas indefinidamente.

Apresenta-se estas propriedades como forma de ilustrar que a superioridades das MLE, no caso de grandes amostras, é um aspecto de consenso entre os diferentes tipos de inferência estatística. Segue-se aqui a apresentação de Greene (2000), que apresenta as propriedades lembrando que certas condições de regularidade devem ser satisfeitas para que as propriedades sejam válidas².

Invariância sob transformação do parâmetro. Esta propriedade foi um dos requisitos apresentados para uma medida aceitável para inferência estatística que indicou a verossimilhança como tal medida. Uma implicação prática desta propriedade é que as funções monotônicas (transformação um-para-um) das MLE também são estimativas de máxima verossimilhança.

Consistência. Uma estimativa é dita *consistente* quando o seu valor converge para o “valor verdadeiro do parâmetro” quando o tamanho da amostra cresce ($n \rightarrow \infty$). A implicação prática é que, no caso de grandes amostras, não há que se preocupar com vícios ou vieses nas MLE.

Na inferência baseada em verossimilhança, a função de log-verossimilhança mede o grau de evidência com que as observações apoiam os possíveis valores do parâmetro. A propriedade de consistência implica que à medida que o número de observações aumenta, o valor com o grau máximo de evidência (MLE) converge para o “valor verdadeiro” do parâmetro.

Eficiência Assintótica. O Teorema do Limite Inferior de Cramer-Rao afirma que o um limite inferior para variância das estimativas não-viciadas de qualquer parâmetro é dada pelo inverso da “*informação de Fisher*” (Mood, Graybill e Boes, 1974). As MLE converge para este limite à medida que o tamanho da amostra cresce ($n \rightarrow \infty$).

No contexto da inferência por verossimilhança, a informação de Fisher é equivalente a medida de informação definida acima. Logo, a eficiência assintótica

²Segundo Greene (2000), as condições são: (1) a função de verossimilhança do modelo estocástico tem suas três primeiras derivadas com respeito ao parâmetro definida em todo espaço paramétrico e em todo espaço amostral; (2) são satisfeitas as condições necessárias para existência da esperança matemática da primeira e segunda derivadas da função de log-verossimilhança; (3) Em todo espaço paramétrico, o valor absoluto da terceira derivada da função de log-verossimilhança é menor que uma função que tenha esperança matemática finita. Estas condições garantem a existência da esperança matemática da primeira e da segunda derivada da função de log-verossimilhança e que ela possa ser aproximada pela expansão de Taylor até o termo quadrático.

simplesmente implica que o grau de curvatura da curva da log-verossimilhança atinge um limite à medida que o número de observações cresce. Este limite é a incerteza mínima possível, ou variância mínima possível, para qualquer tipo de estimativa.

Normalidade Assintótica. As MLE convergem em distribuição para distribuição Gaussiana à medida que o tamanho da amostra cresce.

No contexto da inferência por verossimilhança é bastante estranho se falar da “distribuição” de estimativas, o que só faz sentido quando se considera um cenário em que amostragens ou experimentos são repetidos indefinidamente. Segundo Edwards (1992), o que esta propriedade afirma é que a verossimilhança de θ e de uma transformação monotônica de θ ambas tendem para a forma da distribuição Gaussiana. Ou seja, para uma amostra grande o suficiente a amplitude de incerteza ao redor do máximo é tão pequena que a transformação nesta amplitude é praticamente linear, e dentro desta amplitude a função de log-verossimilhança é bem aproximada pela expansão de Taylor até o termo quadrático.

Edwards (1992) demonstra esta tendência da verossimilhança aproximar a forma da distribuição Gaussiana. Primeiramente, ele define o “*span*” da curva de log-verossimilhança (w) como a raiz quadrada do inverso da “informação observada”:

$$w^2 = 1 / I(\hat{\theta}) = -1 / \frac{d^2 \mathbf{L}\{\hat{\theta}\}}{d\theta^2}.$$

O span é uma medida da largura da curva de log-verossimilhança na vizinhança da MLE, dada pela aproximação da curva pela expansão de Taylor até o termo quadrático. Se o tamanho de amostra é grande o suficiente para que a curva de log-verossimilhança na vizinhança da MLE coincida com a aproximação, então a log-verossimilhança toma a forma parabólica

$$\mathbf{L}\{\theta\} = \mathbf{L}\{\hat{\theta}\} - (\theta - \hat{\theta})^2 / (2w^2).$$

e a verossimilhança deve ter a forma

$$\exp(\mathbf{L}\{\theta\}) = k \exp\left(-(\theta - \hat{\theta})^2 / (2w^2)\right)$$

que é a forma funcional da densidade da distribuição Gaussiana, tendo a MLE ($\hat{\theta}$) com média e o span (w) como desvio padrão.

Esta demonstração não tem nenhuma implicação de que, para grandes amostras, as “MLE têm distribuição Normal”, pois uma afirmação que toma uma MLE como variável aleatória não tem sentido na inferência baseada em verossimilhança. No entanto, a demonstração implica que para grandes amostras a curva

de log-verossimilhança tenderá, na vizinhança da MLE, a uma forma parabólica centrada na MLE e, portanto, o intervalo de log-verossimilhança tenderá a ser simétrico em relação à MLE.

8.5 Espaço Paramétrico Multidimensional

A estimação por verossimilhança foi apresentada até este ponto com respeito a modelos estocásticos com um único parâmetro, isto é, modelos onde o espaço paramétrico é unidimensional. Muitos modelos estocásticos, no entanto, tem dois ou mais parâmetros de modo que o espaço paramétrico se torna *multidimensional*. Os resultados teóricos apresentados são válidos para estes modelos, pois a unidimensionalidade não é um requisito nestes resultados.

Por exemplo, os resultados a respeito da razão de verossimilhança entre duas hipóteses permanecem válidos, sejam as hipóteses referentes ao valor de um parâmetro, sejam referentes aos valores de vários parâmetros. Da mesma forma, a MLE de um parâmetro é obtida pela solução da equação de suporte, mas no caso de um modelo estocástico com parâmetro multidimensional as MLE são obtidas pela solução de um *sistema* de equações de suporte. As propriedades das MLE apresentadas também permanecem válidas.

Nos modelos com espaço paramétrico unidimensional, toda a informação sobre o modelo podia ser estudada pela inspeção da curva de log-verossimilhança. Nos modelos com espaço paramétrico bidimensional este estudo pode ser realizado com gráficos de contorno, mas a partir de três dimensões a inspeção direta da função de log-verossimilhança se torna impraticável e novas técnicas de inspeção e estudo se fazem necessárias.

Superfície de Verossimilhança

Quando o modelo estocástico tem espaço multidimensional, pode se dizer que ele possui um parâmetro multidimensional que corresponde a um *vetor* de parâmetros unidimensionais $\theta = (\theta_1, \theta_2, \dots, \theta_p)$. A função de log-verossimilhança se torna uma função multivariada representada geometricamente por uma superfície multidimensional: a *superfície de verossimilhança*.

Geometricamente, o ponto de máximo desta superfície é o ponto onde o hiperplano tangente à superfície tem inclinação nula. Assim, as MLE são encontradas

solucionando o sistema de equações de suporte:

$$\frac{d\mathbf{L}\{\boldsymbol{\theta}\}}{d\boldsymbol{\theta}} = 0 \quad \Rightarrow \quad \begin{cases} \partial\mathbf{L}\{\theta_1\} / \partial\theta_1 = 0 \\ \partial\mathbf{L}\{\theta_2\} / \partial\theta_2 = 0 \\ \vdots \\ \partial\mathbf{L}\{\theta_p\} / \partial\theta_p = 0 \end{cases} \quad \Rightarrow \quad \hat{\boldsymbol{\theta}} = \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \vdots \\ \hat{\theta}_p \end{pmatrix},$$

sendo $\hat{\boldsymbol{\theta}}$ o vetor das estimativas de máxima verossimilhança dos parâmetros (MLE).

Definição 8.15 (Superfície de Log-verossimilhança). *A superfície de log-verossimilhança de um modelo estocástico M , com parâmetro multidimensional $\boldsymbol{\theta}$ (p parâmetros), é a figura formada no espaço com $p + 1$ dimensões que descreve a função de log-verossimilhança relativa ($\mathbf{L}_R\{\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\}$). A função de log-verossimilhança relativa é definida como*

$$\begin{aligned} \mathbf{L}_R\{\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\} &= \ln(\mathcal{L}_R\{\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\}) = \ln\left(\frac{\mathcal{L}\{\boldsymbol{\theta}\}}{\mathcal{L}\{\hat{\boldsymbol{\theta}}\}}\right) \\ &= \mathbf{L}\{\boldsymbol{\theta}\} - \mathbf{L}\{\hat{\boldsymbol{\theta}}\} = \mathbf{L}\{\boldsymbol{\theta}\} - \max[\mathbf{L}\{\boldsymbol{\theta}\}], \end{aligned}$$

sendo $\hat{\boldsymbol{\theta}}$ o vetor com as p estimativas de máxima verossimilhança (MLE) relativas aos p parâmetros do modelo.

Edwards (1992) mostra que, como a curva de verossimilhança, a superfície de verossimilhança pode ser aproximada por uma superfície parabólica utilizando-se a expansão de Taylor no formato matricial até o termo quadrático.

Definição 8.16 (Aproximação Quadrática da Superfície de Log-veros.). *A aproximação quadrática da função de log-verossimilhança na vizinhança das MLE é a expansão de Taylor até o termo quadrático:*

$$\mathbf{L}\{\boldsymbol{\theta}\} \approx \mathbf{L}\{\hat{\boldsymbol{\theta}}\} + \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}),$$

em que $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$ é a matriz das segundas derivadas da função de log-verossimilhança em relação aos parâmetros unidimensionais no ponto das MLE ($\hat{\boldsymbol{\theta}}$):

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = \begin{pmatrix} \partial^2 \mathbf{L}\{\hat{\boldsymbol{\theta}}\} / \partial\theta_1^2 & \partial \mathbf{L}\{\hat{\boldsymbol{\theta}}\} / \partial\theta_1 \partial\theta_2 & \dots & \partial \mathbf{L}\{\hat{\boldsymbol{\theta}}\} / \partial\theta_1 \partial\theta_p \\ \partial \mathbf{L}\{\hat{\boldsymbol{\theta}}\} / \partial\theta_2 \partial\theta_1 & \partial^2 \mathbf{L}\{\hat{\boldsymbol{\theta}}\} / \partial\theta_2^2 & \dots & \partial \mathbf{L}\{\hat{\boldsymbol{\theta}}\} / \partial\theta_2 \partial\theta_p \\ \vdots & \vdots & \ddots & \vdots \\ \partial \mathbf{L}\{\hat{\boldsymbol{\theta}}\} / \partial\theta_p \partial\theta_1 & \partial \mathbf{L}\{\hat{\boldsymbol{\theta}}\} / \partial\theta_p \partial\theta_2 & \dots & \partial^2 \mathbf{L}\{\hat{\boldsymbol{\theta}}\} / \partial\theta_p^2 \end{pmatrix}.$$

Assim, a aproximação quadrática para superfície de log-verossimilhança será

$$\mathbf{L}_R\{\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\} \approx \frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}),$$

que descreve uma superfície hiper-parabólica com vértice no ponto $(\hat{\boldsymbol{\theta}}, 0)$.

Se a aproximação for apropriada, como no caso de grandes amostras, a função de verossimilhança será

$$\mathcal{L}\{\boldsymbol{\theta}\} = \exp(\mathbf{L}\{\boldsymbol{\theta}\}) = k \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right)$$

que tem a forma funcional da distribuição Gaussiana *multivariada* com vetor de médias $\hat{\boldsymbol{\theta}}$ e matriz de variância-covariância igual a $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{-1}$.

O termo quadrático da superfície hiper-parabólica pode ser tomado como uma medida aproximada da curvatura da superfície de verossimilhança e, consequentemente, uma medida da quantidade de informação contida nas estimativas de máxima verossimilhança. Logo, a matriz das segundas derivadas da função de log-verossimilhança é definida como a “matriz de informação”.

Definição 8.17 (Matriz de Informação Observada). *A matriz das segundas derivadas da função de log-verossimilhança é definida como matriz de informação do parâmetro multidimensional ($\boldsymbol{\theta}$):*

$$I(\boldsymbol{\theta}) = \boldsymbol{\Sigma}.$$

No ponto das MLE ($\hat{\boldsymbol{\theta}}$) a matriz das segundas derivadas é definida como a matriz de informação *observada*:

$$I(\hat{\boldsymbol{\theta}}) = \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}.$$

Região de Log-Verossimilhança

De forma análoga ao intervalo de verossimilhança, nos modelos com parâmetro multidimensional será definida uma *região de log-verossimilhança* delimitada pela mesma constante de razão de verossimilhança

$$\mathbf{L}_R\{\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}\} = \mathbf{L}\{\boldsymbol{\theta}\} - \mathbf{L}\{\hat{\boldsymbol{\theta}}\} \geq -\ln(8).$$

No caso da aproximação hiper-parabólica ser apropriada na vizinhança da MLE, esta região pode ser aproximadamente delimitada por

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = 2 \ln(8),$$

onde $\Sigma_{\hat{\theta}}$ é a *matriz de informação observada*.

Assim como a inspeção direta da superfície de log-verossimilhança é impraticável, também o estudo direto de regiões de verossimilhança não é viável. O aspecto multidimensional da superfície e da região de log-verossimilhança requer o desenvolvimento de métodos especiais que permitam a inferência sobre os parâmetros num espaço com dimensionalidade reduzida. Tai métodos devem gerar inferências cuja interpretação nas situações práticas sejam simples e claras.



Exemplo 8-10: Estimação na Distribuição Gaussiana

A distribuição Gaussiana tem densidade

$$f(x) = f(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right]$$

$$-\infty \leq \mu \leq \infty, \sigma > 0; \quad -\infty \leq x \leq \infty,$$

que para uma amostra de n observações independentes, resulta na função de log-verossimilhança

$$\mathbf{L}\{\mu, \sigma\} = -n \ln(\sigma) - \frac{n}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Como a distribuição tem dois parâmetros, o sistema de equações de suporte necessária à obtenção das estimativas de máxima verossimilhança é composto por um sistema com duas equações:

$$\frac{\partial \mathbf{L}\{\mu, \sigma\}}{\partial \mu} = \frac{1}{\sigma^2} \left[\sum_{i=1}^n x_i - n\mu \right] = 0$$

$$\frac{\partial \mathbf{L}\{\mu, \sigma\}}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{\sigma^3} = 0,$$

cuja solução produz as MLE:

$$\hat{\mu} = \frac{\sum_{i=1}^n x_i}{n} \quad \hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}}.$$

A MLE da média ($\hat{\mu}$) é a média amostral, sendo portanto, equivalente à estimativa de quadrados mínimos, isto é, a estimativa que minimiza os desvios em relação à média ao quadrado. Já a MLE do desvio padrão ($\hat{\sigma}$) não é a mesma estimativa tradicionalmente utilizada na inferência clássica, pois o denominador na expressão

não são os graus de liberdade ($n - 1$). Contudo, no caso de grandes amostras, esta diferença se torna irrelevante.

A função de log-verossimilhança no ponto das MLE é

$$\begin{aligned} \mathbf{L}\{\hat{\mu}, \hat{\sigma}\} &= -n \ln(\hat{\sigma}) - \frac{n}{2} \ln(2\pi) - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (x_i - \hat{\mu})^2 \\ &= -n \ln(\hat{\sigma}) - \frac{n}{2} \ln(2\pi) - \frac{n}{2}, \end{aligned}$$

de modo que a função de log-verossimilhança relativa se torna

$$\begin{aligned} \mathbf{L}_R\{\mu, \hat{\mu}, \sigma, \hat{\sigma}\} &= -n \ln(\sigma) - \frac{n}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 - \\ &\quad - \left[-n \ln(\hat{\sigma}) - \frac{n}{2} \ln(2\pi) - \frac{n}{2} \right] \\ &= n \left[\ln\left(\frac{\hat{\sigma}}{\sigma}\right) + \frac{1}{2} \left(1 - \frac{\sum_{i=1}^n (x_i - \mu)^2 / n}{\sigma^2} \right) \right] \end{aligned}$$

A matriz de informação da distribuição Gaussiana é composta pelas seguintes segundas derivadas:

$$\begin{aligned} \frac{\partial \mathbf{L}^2\{\mu, \sigma\}}{\partial \mu^2} &= -\frac{n}{\sigma^2} & \frac{\partial \mathbf{L}^2\{\mu, \sigma\}}{\partial \sigma^2} &= \frac{n}{\sigma^2} - \frac{3 \sum_{i=1}^n (x_i - \mu)^2}{\sigma^4} \\ \frac{\partial \mathbf{L}^2\{\mu, \sigma\}}{\partial \mu \partial \sigma} &= -\frac{2 \sum_{i=1}^n (x_i - \mu)}{\sigma^3}. \end{aligned}$$

Estas derivadas apresentam muitas informações sobre a superfície de log-verossimilhança Gaussiana. Note-se a terceira derivada em relação à média ($\partial \mathbf{L}^3 / \partial \mu^3$) é nula, enquanto que as derivadas de maior ordem envolvendo a média também se tornaram nulas ($\partial \mathbf{L}^3 / \partial \mu^2 \partial \sigma$ e $\partial \mathbf{L}^4 / \partial \sigma^2 \partial \mu^2$). Quanto às derivadas de maior ordem envolvendo o desvio padrão, estas serão inversamente proporcionais a potências crescentes do desvio padrão e, portanto, aproximam-se rapidamente de zero à medida que a ordem das derivadas cresce. Isto indica que a aproximação quadrática pela série de Taylor será apropriada mesmo para pequenos tamanhos de amostra, uma vez que as derivadas de terceira ou maior ordem podem ser negligenciadas.

No ponto da MLE, as segundas derivadas se tornam:

$$\begin{aligned}\frac{\partial \mathbf{L}^2\{\hat{\mu}, \hat{\sigma}\}}{\partial \mu^2} &= -\frac{n}{\hat{\sigma}^2} \\ \frac{\partial \mathbf{L}^2\{\hat{\mu}, \hat{\sigma}\}}{\partial \sigma^2} &= \frac{n}{\hat{\sigma}^2} - \frac{3 \sum_{i=1}^n (x_i - \hat{\mu})^2}{\hat{\sigma}^4} = \frac{n}{\hat{\sigma}^2} - \frac{3n\hat{\sigma}^2}{\hat{\sigma}^4} = -2\frac{n}{\hat{\sigma}^2} \\ \frac{\partial \mathbf{L}^2\{\hat{\mu}, \hat{\sigma}\}}{\partial \mu \partial \sigma} &= -\frac{2 \sum_{i=1}^n (x_i - \hat{\mu})}{\hat{\sigma}^3} = 0.\end{aligned}$$

Assim, a matriz de informação observada da distribuição Gaussiana é

$$\Sigma_{\hat{\theta}} = \frac{-n}{\hat{\sigma}^2} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Logo, a aproximação quadrática da superfície de log-verossimilhança se torna

$$\begin{aligned}\mathbf{L}_R\{\mu, \hat{\mu}; \sigma, \hat{\sigma}\} &\approx \frac{1}{2} \begin{pmatrix} -n \\ \frac{-n}{\hat{\sigma}^2} \end{pmatrix} \begin{pmatrix} \mu - \hat{\mu} \\ \sigma - \hat{\sigma} \end{pmatrix}' \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \mu - \hat{\mu} \\ \sigma - \hat{\sigma} \end{pmatrix} \\ &\approx -\frac{n}{2\hat{\sigma}^2} [(\mu - \hat{\mu})^2 + 2(\sigma - \hat{\sigma})^2].\end{aligned}$$

Assim, a curvatura da superfície de log-verossimilhança na distribuição Gaussiana é diretamente proporcional ao tamanho da amostra (n) e inversamente proporcional à estimativa da variância ($\hat{\sigma}^2$). A figura 8.4 apresenta uma exemplo de superfície de log-verossimilhança da distribuição Gaussiana, onde é possível observar que a superfície tem a forma bastante próxima a um paraboloide, isto é, a aproximação quadrática se mostra bastante apropriada.



Exemplo 8-11: Modelo de Regressão Linear Simples

O modelo de regressão linear simples é o modelo linear clássico de maior simplicidade, mas pode ser exemplificado como a multidimensionalidade da superfície de log-verossimilhança rapidamente se impões na modelagem quantitativa. Este modelo pode ser apresentado como um modelo Gaussiano em que o parâmetro

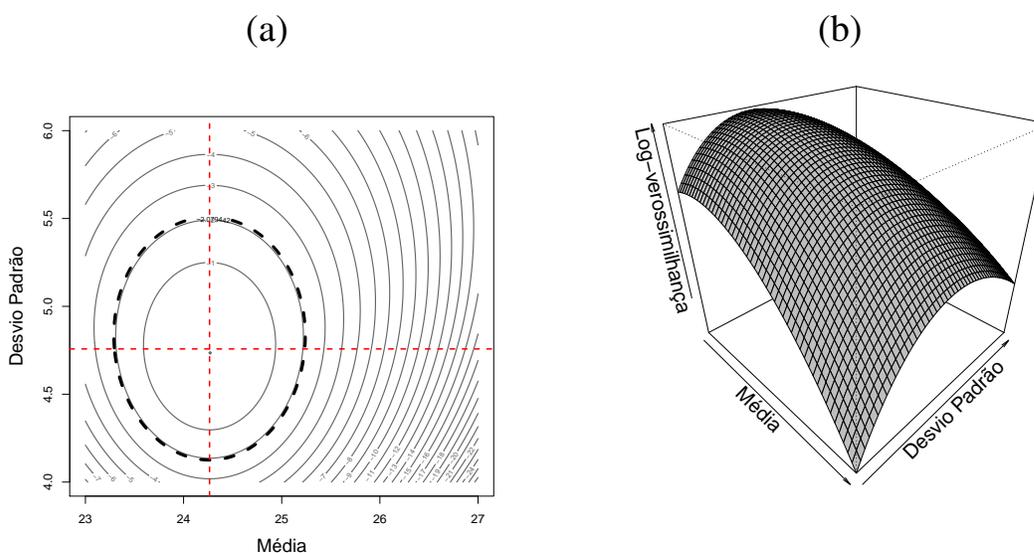


Figura 8.4: Superfície de log-verossimilhança da distribuição Gaussiana, para um tamanho de amostra $n = 100$. O gráfico (a) é um gráfico de contorno que apresenta as isolinhas de log-verossimilhança, a linha tracejada delimita a região de log-verossimilhança maior ou igual a $-\ln(8)$. O gráfico (b) é um gráfico de perspectiva que mostra uma representação tridimensional da superfície.

média é substituído por uma expressão linear em função de uma variável preditora. Neste caso, a função de densidade da distribuição Gaussiana se torna

$$f(x|z; \beta_0, \beta_1, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x - \beta_0 - \beta_1 z)^2}{2\sigma^2}\right]$$

$$-\infty \leq \beta_0, \beta_1 \leq \infty; \sigma > 0; -\infty \leq x, z \leq \infty;$$

send x a variável resposta e z é a variável preditora. Já β_0 e β_1 são chamados de coeficientes de regressão e são os parâmetros que definem a relação linear entre a média e a variável preditora z .

A função de log-verossimilhança neste modelo é

$$\mathbf{L}\{\beta_0, \beta_1, \sigma\} = -n \ln(\sigma) - \frac{n}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \beta_0 - \beta_1 z_i)^2,$$

cujas primeiras derivadas geram o seguinte sistema de equações de suporte

$$\begin{aligned}\frac{\partial \mathbf{L}\{\beta_0, \beta_1, \sigma\}}{\partial \beta_0} &= -\frac{1}{\sigma^2} \left[n\beta_0 + \beta_1 \sum_{i=1}^n z_i - \sum_{i=1}^n x_i \right] = 0 \\ \frac{\partial \mathbf{L}\{\beta_0, \beta_1, \sigma\}}{\partial \beta_1} &= -\frac{1}{\sigma^2} \left[\beta_0 \sum_{i=1}^n z_i + \beta_1 \sum_{i=1}^n z_i^2 - \sum_{i=1}^n x_i z_i \right] = 0 \\ \frac{\partial \mathbf{L}\{\beta_0, \beta_1, \sigma\}}{\partial \sigma} &= -\frac{1}{\sigma} \left[\frac{\sum_{i=1}^n (x_i - \beta_0 - \beta_1 z_i)^2}{\sigma^2} - n \right] = 0.\end{aligned}$$

As duas primeiras equações deste sistema, reduzidas aos termos dentro dos colchetes igualados a zero, formam o *Sistema de Equações Normais* da regressão linear clássica:

$$\begin{cases} \beta_0 n + \beta_1 \sum_{i=1}^n z_i = \sum_{i=1}^n x_i \\ \beta_0 \sum_{i=1}^n z_i + \beta_1 \sum_{i=1}^n z_i^2 = \sum_{i=1}^n x_i z_i \end{cases}$$

A solução analítica deste sistema existe e resulta nas estimativas de máxima verossimilhança (MLE) para os coeficientes de regressão:

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^n x_i z_i - \sum_{i=1}^n x_i \sum_{i=1}^n z_i / n}{\sum_{i=1}^n z_i^2 - (\sum_{i=1}^n z_i)^2 / n} \\ \hat{\beta}_0 &= \frac{\sum_{i=1}^n x_i}{n} - \hat{\beta}_1 \frac{\sum_{i=1}^n z_i}{n} = \bar{x} - \hat{\beta}_1 \bar{z}.\end{aligned}$$

Estas estimativas são iguais às estimativas de quadrados mínimos, logo, no modelo Gaussiano linear simples, as MLE também minimizam a soma dos quadrados dos desvios em relação à média, que, no caso, é a função linear em z . Mas, também há a MLE do desvio padrão (σ), que é obtida em função das MLE dos coeficientes de regressão:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (x_i - \hat{\beta}_0 - \hat{\beta}_1 z_i)^2}{n}}.$$

A superfície de log-verossimilhança é definida pela função de log-verossimilhança

relativa:

$$\begin{aligned} \mathbf{L}_R\{\beta_0, \hat{\beta}_0, \beta_1, \hat{\beta}_1, \sigma, \hat{\sigma}\} &= \mathbf{L}\{\beta_0, \beta_1, \sigma\} - \mathbf{L}\{\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}\} \\ &= -n \ln(\sigma) - \frac{n}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \beta_0 - \beta_1 z_i)^2 \\ &\quad - \left[-n \ln(\hat{\sigma}) - \frac{n}{2} \ln(2\pi) - \frac{n}{2} \right] \\ &= n \ln\left(\frac{\hat{\sigma}}{\sigma}\right) + \frac{1}{2} \left[n - \frac{\sum_{i=1}^n (x_i - \beta_0 - \beta_1 z_i)^2}{\sigma^2} \right]. \end{aligned}$$

Esta superfície de log-verossimilhança se apresenta num espaço quadridimensional, de modo que sua representação gráfica não é possível. O estudo desta superfície pelas segundas derivadas, no entanto, ainda é possível:

$$\begin{aligned} \frac{\partial \mathbf{L}^2\{\beta_0, \beta_1, \sigma\}}{\partial \beta_0^2} &= -\frac{n}{\sigma^2} & \frac{\partial \mathbf{L}^2\{\beta_0, \beta_1, \sigma\}}{\partial \beta_1^2} &= -\frac{1}{\sigma^2} \sum_{i=1}^n z_i^2 \\ \frac{\partial \mathbf{L}^2\{\beta_0, \beta_1, \sigma\}}{\partial \sigma^2} &= \frac{1}{\sigma^2} \left[\frac{3}{\sigma^2} \sum_{i=1}^n (x_i - \beta_0 - \beta_1 z_i)^2 - n \right] \\ \frac{\partial \mathbf{L}^2\{\beta_0, \beta_1, \sigma\}}{\partial \beta_0 \partial \beta_1} &= -\frac{1}{\sigma^2} \sum_{i=1}^n z_i \\ \frac{\partial \mathbf{L}^2\{\beta_0, \beta_1, \sigma\}}{\partial \beta_0 \partial \sigma} &= \frac{2}{\sigma^3} \left[n \beta_0 - \beta_1 \sum_{i=1}^n z_i - \sum_{i=1}^n x_i \right] \\ \frac{\partial \mathbf{L}^2\{\beta_0, \beta_1, \sigma\}}{\partial \beta_1 \partial \sigma} &= \frac{2}{\sigma^3} \left[\beta_0 \sum_{i=1}^n z_i + \beta_1 \sum_{i=1}^n z_i^2 - \sum_{i=1}^n x_i z_i \right]. \end{aligned}$$

A aproximação quadrática da superfície de log-verossimilhança será formada pela matriz de informação observada que é composta pelas segundas derivadas no ponto das MLE. Nota-se que nas segundas derivadas em relação aos coeficientes de regressão (β_0 e β_1) e ao desvio padrão (σ), que são as duas últimas expressões acima, o conteúdo dos colchetes forma o Sistema de Equações Normais. Assim, as derivadas no ponto das MLE, são nulas, pois as MLE dos coeficientes de regressão são obtidas tornando o Sistema de Equações Normais igual a zero. Logo, a matriz de informação observada resultante é

$$\Sigma_{\hat{\theta}} = -\frac{1}{\hat{\sigma}^2} \begin{pmatrix} n & -\sum z_i & 0 \\ -\sum z_i & \sum z_i^2 & 0 \\ 0 & 0 & 2n \end{pmatrix}.$$

A aproximação quadrática da superfície de log-verossimilhança fica

$$\begin{aligned} \mathbf{L}_R\{\beta_0, \hat{\beta}_0; \beta_1, \hat{\beta}_1; \sigma, \hat{\sigma}\} &\approx \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \\ &\approx -\frac{1}{2\hat{\sigma}^2} \begin{pmatrix} \beta_0 - \hat{\beta}_0 \\ \beta_1 - \hat{\beta}_1 \\ \sigma - \hat{\sigma} \end{pmatrix}' \begin{pmatrix} n & -\sum z_i & 0 \\ -\sum z_i & \sum z_i^2 & 0 \\ 0 & 0 & 2n \end{pmatrix} \begin{pmatrix} \beta_0 - \hat{\beta}_0 \\ \beta_1 - \hat{\beta}_1 \\ \sigma - \hat{\sigma} \end{pmatrix} \\ &\approx \frac{1}{2\hat{\sigma}^2} \left[n(\beta_0 - \hat{\beta}_0)^2 - 2\left(\sum z_i\right)(\beta_0 - \hat{\beta}_0)(\beta_1 - \hat{\beta}_1) + \right. \\ &\quad \left. + \left(\sum z_i^2\right)(\beta_1 - \hat{\beta}_1)^2 + 2n(\sigma - \hat{\sigma})^2 \right]. \end{aligned}$$

A região de log-verossimilhança é definida igualando-se a função de log-verossimilhança relativa ao valor limite de $-\ln(8)$:

$$\begin{aligned} \mathbf{L}_R\{\beta_0, \hat{\beta}_0, \beta_1, \hat{\beta}_1, \sigma, \hat{\sigma}\} &= -\ln(8) \\ n \ln\left(\frac{\hat{\sigma}}{\sigma}\right) + \frac{1}{2} \left[n - \frac{\sum_{i=1}^n (x_i - \beta_0 - \beta_1 z_i)^2}{\sigma^2} \right] &= -\ln(8). \end{aligned}$$

No caso da aproximação quadrática, a região de log-verossimilhança é definida por

$$\begin{aligned} \mathbf{L}_R\{\beta_0, \hat{\beta}_0; \beta_1, \hat{\beta}_1; \sigma, \hat{\sigma}\} &\approx -\ln(8) \\ \Rightarrow \frac{1}{2\hat{\sigma}^2} \left[n(\beta_0 - \hat{\beta}_0)^2 - 2\left(\sum z_i\right)(\beta_0 - \hat{\beta}_0)(\beta_1 - \hat{\beta}_1) + \right. \\ &\quad \left. + \left(\sum z_i^2\right)(\beta_1 - \hat{\beta}_1)^2 + 2n(\sigma - \hat{\sigma})^2 \right] = -\ln(8). \end{aligned}$$

No caso do modelo linear simples, em que a superfície de log-verossimilhança é definida numa espaço quadridimensional, a aproximação quadrática é algebricamente *ligeiramente mais tratável* que a função de log-verossimilhança relativa. Mas, é fácil perceber que, à medida que aumentar o número de variáveis preditoras e, consequentemente, o número de dimensões do espaço paramétrico, a aproximação quadrática se tornará progressivamente mais complexa, de modo que a solução será mantê-la na expressão matricial. Conclui-se, portanto, que a aproximação quadrática não se constitui uma solução para o problema do espaço paramétrico multidimensional e outras técnicas de inspeção do espaço paramétrico se fazem necessárias.





Exemplo 8-12: Estimação na Distribuição Weibull

A distribuição Weibull também pode ser apresentada na forma de dois parâmetros com densidade

$$f(x; \beta, \gamma) = \left(\frac{\gamma}{\beta}\right) \left(\frac{x}{\beta}\right)^{\gamma-1} \exp\left[-\left(\frac{x}{\beta}\right)^\gamma\right]; \quad \beta, \gamma > 0; \quad x \geq 0;$$

na qual β é o parâmetro de escala e γ é o parâmetro de forma. Sua função de log-verossimilhança, com base numa amostra de n observações independentes, é

$$\mathbf{L}\{\beta, \gamma\} = n \ln(\gamma) - n \gamma \ln(\beta) + (\gamma - 1) \sum_{i=1}^n \ln(x_i) - \frac{1}{\beta^\gamma} \sum_{i=1}^n x_i^\gamma$$

sendo o sistema de equações de suporte

$$\frac{\partial \mathbf{L}\{\beta, \gamma\}}{\partial \beta} = -\frac{n \gamma}{\beta} + \frac{\gamma}{\beta^{\gamma+1}} \sum_{i=1}^n x_i^\gamma = 0 \quad \Rightarrow \quad \hat{\beta} = \left[\frac{\sum_{i=1}^n x_i^\gamma}{n} \right]^{1/\gamma}$$

$$\frac{\partial \mathbf{L}\{\beta, \gamma\}}{\partial \gamma} = \frac{n}{\gamma} + \sum_{i=1}^n \ln(x_i) - \frac{1}{\beta^\gamma} \sum_{i=1}^n x_i^\gamma \ln(x_i) = 0.$$

A partir da primeira equação do sistema de equações de suporte, a estimativa de máxima verossimilhança (MLE) do parâmetro da escala (β) é obtida em função da MLE do parâmetro da forma (γ). A segunda equação do sistema de equações de suporte, contudo, não possui solução analítica, logo, a sua solução requer métodos numéricos.

A superfície de log-verossimilhança para distribuição Weibull é definida por

$$\begin{aligned} \mathbf{L}_R\{\beta, \hat{\beta}, \gamma, \hat{\gamma}\} &= \mathbf{L}\{\beta, \gamma\} - \mathbf{L}\{\hat{\beta}, \hat{\gamma}\} \\ &= n \ln(\gamma) - n \gamma \ln(\beta) + (\gamma - 1) \sum_{i=1}^n \ln(x_i) - \frac{1}{\beta^\gamma} \sum_{i=1}^n x_i^\gamma + \\ &\quad - \left[n \ln(\hat{\gamma}) - n \hat{\gamma} \ln(\hat{\beta}) + (\hat{\gamma} - 1) \sum_{i=1}^n \ln(x_i) - n \right] \\ \mathbf{L}_R\{\beta, \hat{\beta}, \gamma, \hat{\gamma}\} &= n \left[-\ln\left(\frac{\hat{\gamma}}{\gamma}\right) + \hat{\gamma} \ln(\hat{\beta}) - \gamma \ln(\beta) + 1 \right] + (\gamma - \hat{\gamma}) \sum_{i=1}^n \ln(x_i) + \\ &\quad - \frac{1}{\beta^\gamma} \sum_{i=1}^n x_i^\gamma. \end{aligned}$$

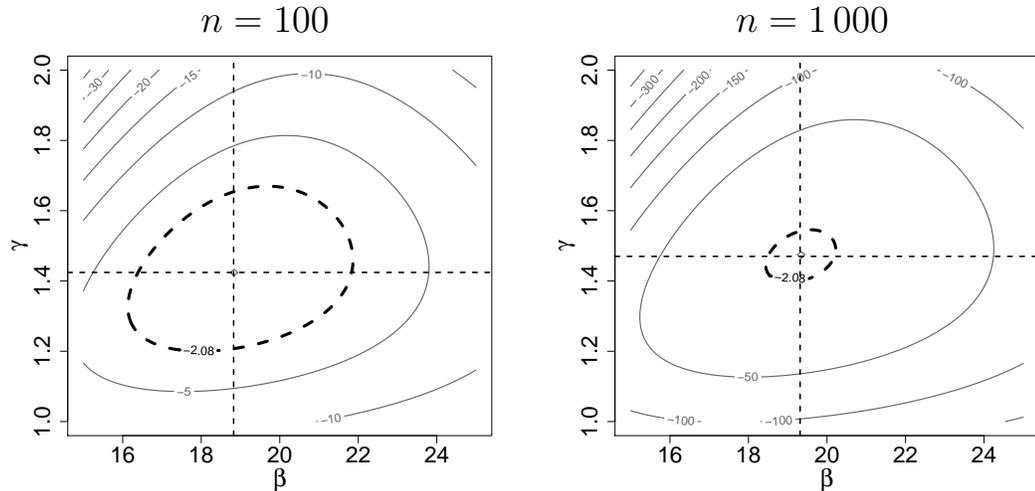


Figura 8.5: Gráficos de contorno da superfície de log-verossimilhança da distribuição Weibull para dois tamanhos de amostra. A linha tracejada indica a região de log-verossimilhança correspondente ao limite de $-\ln(8)$.

A complexidade desta superfície pode ser observada na figura 8.5 que apresenta gráficos de contorno para superfícies geradas por dois tamanhos de amostra ($n = 100$ e $n = 1000$). A figura também ilustra o efeito do tamanho da amostra (n) sobre a curvatura da superfície de log-verossimilhança.

As segundas derivadas da função de log-verossimilhança são

$$\begin{aligned} \frac{\partial^2 \mathbf{L}\{\beta, \gamma\}}{\partial \beta^2} &= \frac{n\gamma}{\beta^2} - \frac{\gamma(\gamma+1)}{\beta^{\gamma+2}} \sum_{i=1}^n x_i^\gamma \\ \frac{\partial^2 \mathbf{L}\{\beta, \gamma\}}{\partial \gamma^2} &= -\frac{n}{\gamma} - \frac{[\ln(\beta)]^2}{\beta^\gamma} \sum_{i=1}^n x_i^\gamma + \\ &\quad + 2 \frac{\ln(\beta)}{\beta^\gamma} \sum_{i=1}^n x_i^\gamma \ln(x_i) - \frac{1}{\beta^\gamma} \sum_{i=1}^n x_i^\gamma [\ln(x_i)]^2 \\ \frac{\partial^2 \mathbf{L}\{\beta, \gamma\}}{\partial \beta \partial \gamma} &= -\frac{n}{\beta} + \frac{1}{\beta^{\gamma+1}} \left[(1 - \gamma \ln(\beta)) \sum_{i=1}^n x_i^\gamma + \gamma \sum_{i=1}^n x_i^\gamma \ln(x_i) \right] \end{aligned}$$

Estas derivadas mostram que a curvatura da superfície de log-verossimilhança de uma distribuição Weibull é bem complexa e, conseqüentemente, a aproximação quadrática é problemática. As derivadas mostram a complexa interação entre os parâmetros de forma (γ) e de escala (β) e a forte influência do parâmetro da forma

sobre a superfície, uma vez que os valores das observações (x_i) estão sempre elevados à potência deste parâmetro. Logo, a forma da distribuição Weibull é uma indicação da complexidade da superfície de log-verossimilhança.

Contudo, a partir das MLE dos parâmetros é possível calcular numericamente as segundas derivadas e, conseqüentemente, obter a matrix de informação observada ($\Sigma_{\hat{\theta}}$) e o vetor dos parâmetro ($(\theta - \hat{\theta})$), gerando a aproximação quadrática para superfície de log-verossimilhança da Weibull:

$$\frac{1}{2} \begin{pmatrix} \beta - \hat{\beta} \\ \gamma - \hat{\gamma} \end{pmatrix}' \begin{pmatrix} \partial^2 \mathbf{L}\{\hat{\beta}, \hat{\gamma}\} / \partial \beta^2 & \partial^2 \mathbf{L}\{\hat{\beta}, \hat{\gamma}\} / \partial \beta \partial \gamma \\ \partial^2 \mathbf{L}\{\hat{\beta}, \hat{\gamma}\} / \partial \beta \partial \gamma & \partial^2 \mathbf{L}\{\hat{\beta}, \hat{\gamma}\} / \partial \gamma^2 \end{pmatrix} \begin{pmatrix} \beta - \hat{\beta} \\ \gamma - \hat{\gamma} \end{pmatrix}.$$

A figura 8.6 ilustra esta situação mostrando os gráfico de densidade de duas distribuições Weibull e os respectivos gráficos de contorno para as superfícies de log-verossimilhança, construídas com base em 1000 observações. Nota-se que a aproximação quadrática é muito problemática para a distribuição assimétrica ($\gamma = 0, 5$), mas já se mostra razoável quando para a distribuição com forma aproximadamente simétrica ($\gamma = 3, 6$). É importante ressaltar que, em circunstâncias práticas, 1000 observações constituem uma amostra extremamente grande.



Parâmetros Inconvenientes

Como visto, mesmo quando o interesse de conhecimento sobre um modelo estocástico com parâmetro multidimensional recaia em todos os parâmetros do vetor, frequentemente não é praticável avaliar a superfície de verossimilhança ou a região de verossimilhança. Há situações, no entanto, em que o interesse da investigação não está focado em todos os parâmetros do vetor paramétrico, mas apenas em alguns deles, mas os parâmetros sem interesse não podem ser simplesmente descartados, pois isto implicaria na completa transformação do modelo. Ou seja, a análise dos parâmetros de interesse não pode ser realizada sem a presença daqueles parâmetros sem interesse.

Os parâmetros sem interesse que não podem ser descartados são chamados de “parâmetros inconvenientes” (*nuisance parameters*), pois embora não haja interesse neles a análise do modelo não pode prescindir deles. Assim, soma-se ao problema do estudo de parâmetro multidimensionais, o problema da presença de parâmetros inconvenientes no vetor paramétrico. Para deixar a apresentação

mais clara, utiliza-se a notação proposta por Royall (1997) em que o parâmetro multidimensional é composto por dois sub-vetores (θ, γ) , onde θ é parâmetro multidimensional de interesse e γ é o parâmetro multidimensional inconveniente.

O problema da multidimensionalidade, como colocado por Royall (1997), é o problema de que se deseja representar, interpretar e relatar a evidência a respeito

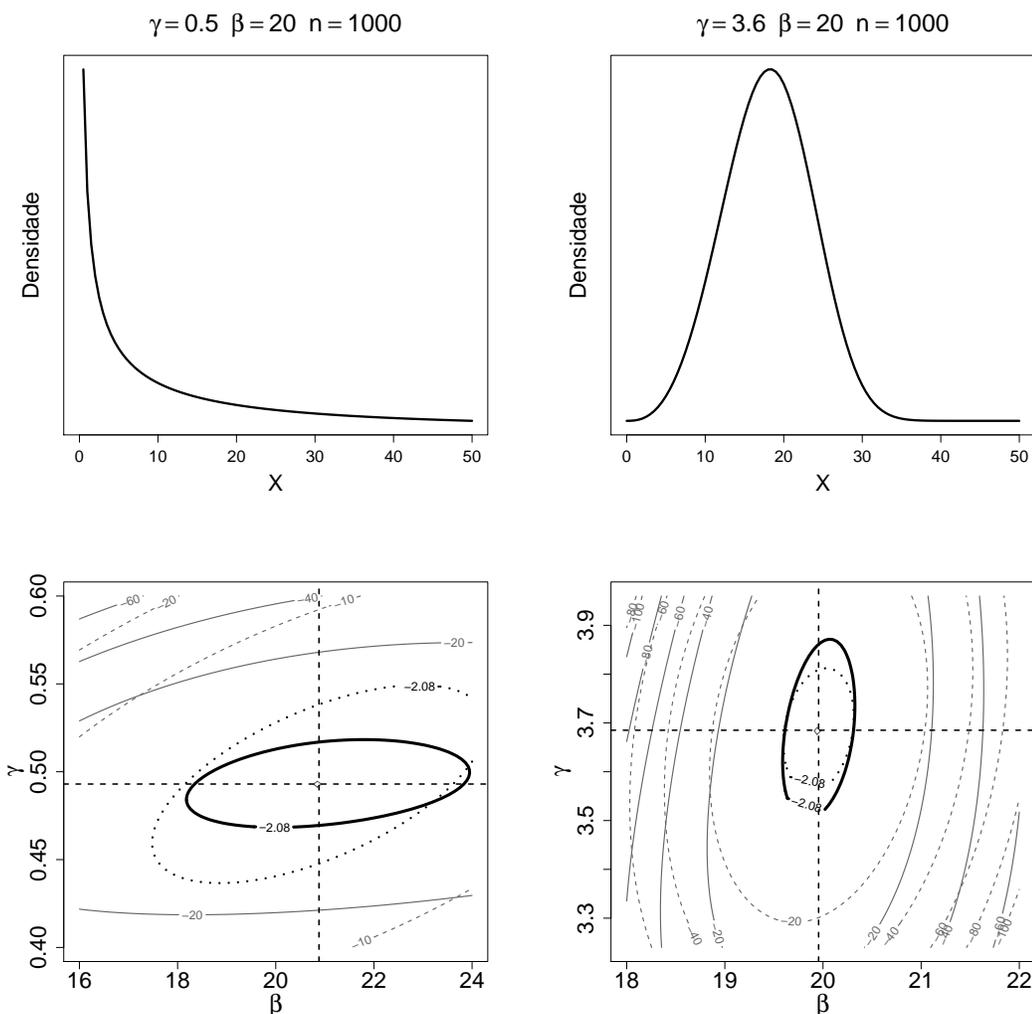


Figura 8.6: Gráficos da Weibull para dois valores diferentes do parâmetro da forma (γ) e mesmo valor do parâmetro da escala (β). Os gráficos no alto mostram a curva da função de densidade para as duas distribuições. Os gráficos de contorno respectivos mostram as isolinhas da superfície de log-verossimilhança (linhas contínuas), construída com uma amostra de 1000 observações ($n = 1000$), e as isolinhas da aproximação quadrática (linhas tracejadas).

(ou plausibilidade) do parâmetro θ somente. Mas a verossimilhança $\mathcal{L}\{\theta, \gamma\}$ depende de ambos, θ e γ . A razão de verossimilhança $\mathcal{L}\{\theta_1, \gamma\}/\mathcal{L}\{\theta_2, \gamma\}$ mede a evidência relativa (ou a plausibilidade relativa) dos valores (θ_1, γ) em relação aos valores (θ_2, γ) . Para um valor fixo de γ , esta razão é uma medida relativa de θ_1 contra θ_2 , mas ela depende do valor de γ . Portanto, a razão de verossimilhança é a comparação de θ_1 vis-à-vis θ_2 condicionada ao valor de γ . Deseja-se realizar inferências a respeito de θ somente, mas geralmente não é possível descartar o parâmetro inconveniente γ .

A impossibilidade de descarte do parâmetro inconveniente se deve ao fato de que os parâmetros (θ, γ) , e não somente θ , determinam a distribuição estocástica dos dados observados. Um valor fixo de θ não determina uma única distribuição estocástica, mas sim uma família de distribuições à medida de γ varia. A razão de verossimilhança compara duas distribuições estocásticas com valores de parâmetro fixos, mas não duas famílias de distribuição.

O exemplo mais comum de parâmetro inconveniente é o desvio padrão nos modelos baseados na distribuição Gaussiana. Seja nos modelos lineares seja nos modelos não lineares, o foco da inferência estatística são os coeficientes de regressão, mas a inferência é diretamente influenciada pelo desvio padrão do modelo. Nas situações em que a estimativa do desvio padrão é problemática, a relação entre os coeficientes de regressão e o desvio padrão pode tornar a inferência a respeito dos coeficientes inviável. Este problema ocorre em amostras com tamanho pequeno e, principalmente, no caso de multicolinearidade, isto é, nas situações em que as diversas variáveis preditoras são correlacionadas entre si. Nos modelos Gaussianos em que se assume a heteroscedasticidade, isto é, o desvio padrão não é tomado como uma constante, mas é ele mesmo modelado em função de variáveis preditoras, qualquer inferência restrita aos coeficientes de regressão não é somente problemática, mas de interpretação muito complexa.

Não existe uma solução geral para este problema, pois não há uma solução teórica geral que permita descartar o parâmetro inconveniente. Entretanto, existe uma série de soluções possíveis para situações particulares de aplicação.

Parâmetros Ortogonais

Há distribuições em que a função de verossimilhança pode ser reduzida ao produto de duas funções, cada uma dependendo de componentes diferentes do parâmetro multidimensional. Neste caso, a função de log-verossimilhança se torna a adição de duas funções e os parâmetros são ditos *ortogonais*.

Definição 8.18 (Parâmetros Ortogonais). *Num modelo estocástico com espaço parâmetro multidimensional (θ, γ) , os parâmetros θ e γ são ditos ortogonais se a função de log-verossimilhança pode ser apresentada como a adição de duas*

funções, cada qual dependendo apenas de um dos parâmetros:

$$\mathbf{L}\{\boldsymbol{\theta}, \gamma\} = \mathbf{L}_1\{\boldsymbol{\theta}\} + \mathbf{L}_2\{\gamma\}.$$

Nos modelos estocásticos com parâmetros ortogonais, a razão de verossimilhança eliminará automaticamente o parâmetro inconveniente:

$$\mathbf{L}_R\{(\boldsymbol{\theta}_1, \gamma), (\boldsymbol{\theta}_2, \gamma)\} = \mathbf{L}\{\boldsymbol{\theta}_1, \gamma\} - \mathbf{L}\{\boldsymbol{\theta}_2, \gamma\} = \mathbf{L}_1\{\boldsymbol{\theta}_1\} - \mathbf{L}_1\{\boldsymbol{\theta}_2\}.$$

Há situações em que os parâmetros não são ortogonais, mas a ortogonalidade pode ser obtida pela reparametrização do modelo.



Exemplo 8-13: Distribuição Poisson da Regeneração Natural

Royall (1997) dá um exemplo de parâmetros ortogonais que pode ser aplicado no estudo da regeneração natural de florestas. Suponha que se deseje saber a razão do número médio de indivíduos regenerantes de duas fases sucessivas da regeneração. Por exemplo, num levantamento do estoque de palmiteiro (*Euterpe edulis* Martius), na região do Vale do Ribeira, estado de São Paulo, foram encontrados 701,86 mudas ha^{-1} (altura entre 10 e 50 cm) e 2002,92 plântulas ha^{-1} (até 10 cm de altura) (Batista, Vettorazzi e Couto, 2000). Qual a MLE para razão entre estes números médios e a incerteza associada a ela?

A distribuição Poisson é um modelo candidato para este tipo de problema, mas neste caso a função de log-verossimilhança é

$$\begin{aligned} \mathbf{L}\{\lambda_1, \lambda_2\} &= \ln(\lambda_1) \sum_{i=1}^n x_{1i} - n\lambda_1 - \sum_{i=1}^n \ln(x_{1i}!) + \\ &\quad + \ln(\lambda_2) \sum_{i=1}^n x_{2i} - n\lambda_2 - \sum_{i=1}^n \ln(x_{2i}!), \end{aligned}$$

onde λ_1 e λ_2 se referem a número médio de mudas e de plântulas, respectivamente, enquanto que x_{1i} e x_{2i} são os números de mudas e plântulas, respectivamente, na i ésima parcela ($i = 1, 2, \dots, n$). Embora este modelo seja ortogonal, ele não fornece a informação desejada que é a razão entre os números médios

$$\theta = \frac{\lambda_1}{\lambda_2}.$$

Se for definido um parâmetro inconveniente, e neste caso sem sentido prático, como sendo

$$\gamma = \lambda_1 + \lambda_2,$$

tem-se a reparametrização do modelo original na seguinte forma

$$\lambda_1 = \frac{\gamma \theta}{1 + \theta} \quad e \quad \lambda_2 = \frac{\gamma}{1 + \theta},$$

resultando na função de log-verossimilhança

$$\begin{aligned} \mathbf{L}\{\theta, \gamma\} &= \ln\left(\frac{\gamma \theta}{1 + \theta}\right) \sum_{i=1}^n x_{1i} - n\left(\frac{\gamma \theta}{1 + \theta}\right) + \sum_{i=1}^n \ln(x_{1i}!) + \\ &+ \ln\left(\frac{\gamma}{1 + \theta}\right) \sum_{i=1}^n x_{2i} - n\left(\frac{\gamma}{1 + \theta}\right) + \sum_{i=1}^n \ln(x_{2i}!), \end{aligned}$$

que pode ser reescrita de modo a enfatizar seu carácter ortogonal:

$$\begin{aligned} \mathbf{L}\{\theta, \gamma\} &= \ln(\theta) \sum_{i=1}^n x_{1i} - \ln(1 + \theta) \left(\sum_{i=1}^n x_{1i} + \sum_{i=1}^n x_{2i} \right) + \\ &+ \ln(\gamma) \left(\sum_{i=1}^n x_{1i} + \sum_{i=1}^n x_{2i} \right) - n\gamma + \\ &+ \sum_{i=1}^n \ln(x_{1i}!) + \sum_{i=1}^n \ln(x_{2i}!). \end{aligned}$$

A função de log-verossimilhança no que se refere ao parâmetro θ se resume a

$$\mathbf{L}\{\theta\} = \ln(\theta) \sum_{i=1}^n x_{1i} - \ln(1 + \theta) \left(\sum_{i=1}^n x_{1i} + \sum_{i=1}^n x_{2i} \right),$$

pois os demais termos da função de log-verossimilhança são ortogonais à θ ou constantes. Assim, a sua MLE é obviamente dada pela razão dos somatórios

$$\begin{aligned} \frac{\partial \mathbf{L}\{\theta, \gamma\}}{\partial \theta} &= \frac{\sum_{i=1}^n x_{1i}}{\theta} - \frac{\sum_{i=1}^n x_{1i} + \sum_{i=1}^n x_{2i}}{1 + \theta} = 0 \Rightarrow \\ \Rightarrow \hat{\theta} &= \frac{\sum_{i=1}^n x_{1i}}{\sum_{i=1}^n x_{2i}}. \end{aligned}$$

Os dados do Vale do Ribeira indicam uma MLE para a razão de mudas e plântulas igual a

$$\hat{\theta} = \frac{701,86}{2002,92} = 0,3504184.$$

A curva de log-verossimilhança é dada pela expressão:

$$\mathbf{L}_R\{\theta\} = -\ln\left(\frac{\hat{\theta}}{\theta}\right) \sum_{i=1}^n x_{1i} + \ln\left(\frac{1+\hat{\theta}}{1+\theta}\right) \left(\sum_{i=1}^n x_{1i} + \sum_{i=1}^n x_{2i}\right).$$

A segunda derivada da função de log-verossimilhança é

$$\frac{\partial^2 \mathbf{L}\{\theta\}}{\partial \theta^2} = -\frac{\sum_{i=1}^n x_{1i}}{\theta^2} + \frac{\sum_{i=1}^n x_{1i} + \sum_{i=1}^n x_{2i}}{(1+\theta)^2}.$$

Logo, a informação observada é

$$I(\hat{\theta}) = -\frac{\partial^2 \mathbf{L}\{\hat{\theta}\}}{\partial \theta^2} = \frac{(\sum_{i=1}^n x_{2i})^3}{\sum_{i=1}^n x_{1i} (\sum_{i=1}^n x_{1i} + \sum_{i=1}^n x_{2i})} = 4\,232,611.$$

e a aproximação quadrática da curva de log-verossimilhança é

$$-\frac{1}{2}(\theta - \hat{\theta})^2 I(\hat{\theta}) = -2\,116,306(\theta - \hat{\theta})^2.$$

A figura 8.7 apresenta a curva de log-verossimilhança para razão entre o número de mudas e de plântulas (θ) observada no Vale do Ribeira. Observa-se que, na vizinhança da MLE, a aproximação quadrática se mostra adequada para representar o intervalo de log-verossimilhança. Assim, o intervalo de log-verossimilhança pode ser calculado pela aproximação quadrática:

$$\begin{aligned} \theta = \hat{\theta} \pm \sqrt{\frac{2 \ln(8)}{I(\hat{\theta})}} &\Rightarrow \theta = 0,3504184 \pm \sqrt{\frac{2 \ln(8)}{4\,232,611}} \\ &\Rightarrow \theta = 0,3504184 \pm 0,03134615 \\ &\Rightarrow \theta = 0,35 \pm 0,03 \Rightarrow \theta \in [0,32 \pm 0,38]. \end{aligned}$$



Verossimilhança Estimada

A verossimilhança estimada consiste em obter a verossimilhança fixando o valor do parâmetro inconveniente. O melhor valor para fixá-lo é a estimativa de máxima verossimilhança (MLE) do parâmetro. Suponha-se um dado modelo com

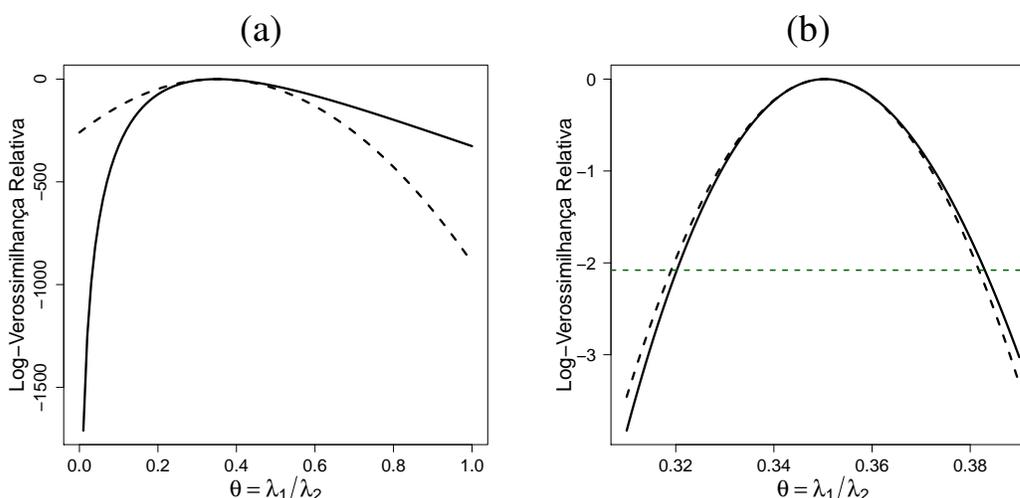


Figura 8.7: Curva de log-verossimilhança para razão entre o número de mudas e de plântulas do palmeiteiro (*Euterpe edulis* Martius) na região do Vale do Ribeira. Os gráficos apresentam a curva de log-verossimilhança (linha sólida), a aproximação quadrática (linha traço-ponto) e a linha que define a razão de verossimilhança de 8, em todo espaço paramétrico de θ (a) e na vizinhança da MLE (b).

espaço paramétrico multidimensional (θ, γ) com p parâmetros, sendo que o vetor de parâmetros de interesse (θ) têm dimensão r e o vetor de parâmetros inconvenientes (γ) têm dimensão q , tal que $p = r + q$. A verossimilhança estimada é obtida fixando-se o parâmetro inconveniente na MLE ($\hat{\gamma}$) e, conseqüentemente, a inspeção da superfície de log-verossimilhança é reduzida do espaço com dimensão $p + 1$ para o espaço com dimensão $r + 1$. Se o interesse estiver em um único parâmetro ($r = 1$), então a verossimilhança estimada reduz a superfície para a curva de log-verossimilhança.

Definição 8.19 (Verossimilhança Estimada). *Num modelo estocástico com espaço paramétrico multidimensional (θ, γ) , sendo γ o vetor dos parâmetros inconvenientes, a função de log-verossimilhança estimada do vetor de parâmetros de interesse θ é a função de log-verossimilhança relativa dado que o vetor de parâmetros inconvenientes permaneça fixo na MLE, ou seja*

$$\mathbf{L}_E\{\theta\} = \mathbf{L}_R\{\theta|\hat{\gamma}\},$$

sendo $\hat{\gamma}$ a MLE do vetor de parâmetros inconvenientes.

Assim como na aproximação quadrática, a verossimilhança estimada é empregada como uma aproximação da superfície de log-verossimilhança na vizinhança

da MLE do parâmetro de interesse. Mas, um aspecto importante a ser considerado é que a verossimilhança estimada é apenas um *expediente* para se contornar o parâmetro inconveniente. Ela não é *verdadeiramente* uma verossimilhança, pois ela é calculada como se o valor do parâmetro inconveniente fosse conhecido, o que nunca é o caso. Uma forte limitação da verossimilhança estimada é que ela tende a produzir uma superfície de verossimilhança com maior curvatura que aquela que existe de fato na superfície de log-verossimilhança. Ou seja, ela tende a superestimar a precisão da MLE do parâmetro de interesse.



Exemplo 8-14: Estimação na Distribuição Gaussiana II

No exemplo da estimação na distribuição Gaussiana foi apresentada a estimativa de máxima verossimilhança (MLE) e o estudo da curvatura da superfície de log-verossimilhança na vizinhança da MLE. Dos dois parâmetros da Gaussiana, normalmente o interesse está focalizado na média (μ), de modo que o desvio padrão (σ) pode ser considerado um parâmetro inconveniente. O exemplo 8.5 apresenta a função de log-verossimilhança relativa:

$$\mathbf{L}_R\{\mu, \hat{\mu}, \sigma, \hat{\sigma}\} = n \left[\ln \left(\frac{\hat{\sigma}}{\sigma} \right) + \frac{1}{2} \left(1 - \frac{\sum_{i=1}^n (x_i - \mu)^2 / n}{\sigma^2} \right) \right],$$

assim, a log-verossimilhança estimada para média é obtida pela função

$$\begin{aligned} \mathbf{L}_E\{\mu\} &= \mathbf{L}_R\{\mu|\hat{\sigma}\} = \frac{n}{2} \left(1 - \frac{1}{n\hat{\sigma}^2} \sum_{i=1}^n (x_i - \mu)^2 \right) \\ &= \frac{n}{2} \left(1 - \left[1 + \frac{1}{\hat{\sigma}^2} (\mu - \hat{\mu})^2 \right] \right) \\ \mathbf{L}_E\{\mu\} &= -\frac{n}{2\hat{\sigma}^2} (\mu - \hat{\mu})^2 \end{aligned}$$

na qual o único termo variável é o parâmetro da média (μ). Logo, a curva da log-verossimilhança estimada neste caso é em si uma função quadrática no parâmetro de interesse (μ).

Também pelo exemplo 8.5, a aproximação quadrática da superfície de log-verossimilhança na distribuição Gaussiana é

$$\mathbf{L}_R\{\mu, \hat{\mu}; \sigma, \hat{\sigma}\} \approx -\frac{n}{2\hat{\sigma}^2} [(\mu - \hat{\mu})^2 + 2(\sigma - \hat{\sigma})^2],$$

logo, a aproximação quadrática para verossimilhança estimada da média é

$$\mathbf{L}_E\{\mu\} = \mathbf{L}_R\{\mu|\hat{\sigma}\} \approx -\frac{n}{2\hat{\sigma}^2} (\mu - \hat{\mu})^2.$$

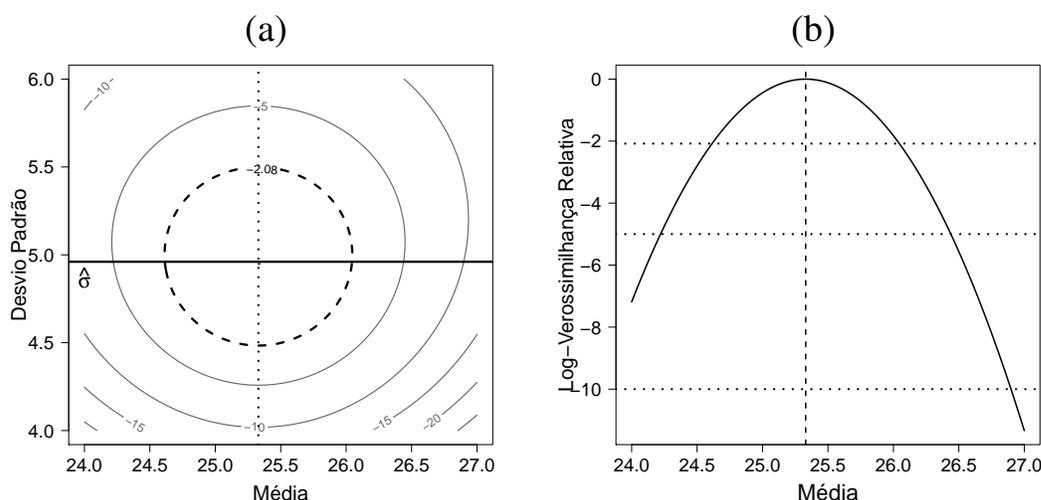


Figura 8.8: Gráfico de contorno (a) mostrando a superfície de log-verossimilhança para uma distribuição Gaussiana com base numa amostra de 200 observações. A linha sólida corresponde à direção da MLE do desvio padrão ($\hat{\sigma}$). Gráfico da log-verossimilhança estimada da média (b) em que as linhas horizontais pontilhadas correspondentem às isolinhas da log-verossimilhança relativa do gráfico de contorno.

Ou seja, na distribuição Gaussiana, a verossimilhança estimada da média é equivalente à aproximação quadrática quando se fixa o valor do desvio padrão na sua MLE ($\sigma = \hat{\sigma}^2$).

Como na distribuição Gaussiana, a superfície de log-verossimilhança é tridimensional, geometricamente, a verossimilhança estimada da média corresponde a um “corte” na superfície na direção da “linha” da MLE do desvio padrão. A figura 8.8 ilustra esta representação geométrica mostrando a superfície de log-verossimilhança de uma distribuição Gaussiana gerada com base numa amostra de 200 observações. A log-verossimilhança estimada da média é indicada no gráfico de contorno pela linha horizontal sólida da MLE do desvio padrão. A curva da log-verossimilhança estimada é apresentada no gráfico ao lado indicando as isolinhas da superfície de log-verossimilhança na forma de linhas horizontais pontilhadas. O gráfico mostra a “topografia” da superfície de log-verossimilhança na linha de “corte” correspondente à MLE do desvio padrão.



 *Exemplo 8-15: Modelo de Regressão Linear Simples II*

Nos modelos lineares, geralmente, os parâmetros de interesse são os coeficientes de regressão, enquanto o desvio padrão pode ser entendido como um parâmetro inconveniente. O exemplo 8.5 apresentou a função de log-verossimilhança relativa para o modelo linear simples:

$$\begin{aligned} \mathbf{L}_R\{\beta_0, \hat{\beta}_0, \beta_1, \hat{\beta}_1, \sigma, \hat{\sigma}\} &= \mathbf{L}\{\beta_0, \beta_1, \sigma\} - \mathbf{L}\{\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}\} \\ &= -n \ln(\sigma) - \frac{n}{2} \ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \beta_0 - \beta_1 z_i)^2 \\ &\quad - \left[-n \ln(\hat{\sigma}) - \frac{n}{2} \ln(2\pi) - \frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (x_i - \hat{\beta}_0 - \hat{\beta}_1 z_i)^2 \right] \\ &= n \ln\left(\frac{\hat{\sigma}}{\sigma}\right) - \frac{1}{2} \left[\frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \beta_0 - \beta_1 z_i)^2 - \frac{1}{\hat{\sigma}^2} \sum_{i=1}^n (x_i - \hat{\beta}_0 - \hat{\beta}_1 z_i)^2 \right]. \end{aligned}$$

Ao invés de desenvolver mais a álgebra desta expressão para simplificá-la, como foi realizado no exemplo 8.5, interessa aqui transformá-la na verossimilhança estimada dos coeficientes de regressão. Para isto, toma-se o valor do desvio padrão como sendo igual à sua MLE ($\sigma = \hat{\sigma}$):

$$\mathbf{L}_E\{\beta_0, \beta_1 | \hat{\sigma}\} = -\frac{1}{2\hat{\sigma}^2} \left[\sum_{i=1}^n (x_i - \beta_0 - \beta_1 z_i)^2 - \sum_{i=1}^n (x_i - \hat{\beta}_0 - \hat{\beta}_1 z_i)^2 \right].$$

Por meio de uma série de operações algébricas, é possível demonstrar que esta expressão é equivalente a

$$\begin{aligned} \mathbf{L}_E\{\beta_0, \beta_1 | \hat{\sigma}\} &= \\ &= -\frac{1}{2\hat{\sigma}^2} \left[n(\beta_0 - \hat{\beta}_0)^2 - 2\left(\sum z_i\right)(\beta_0 - \hat{\beta}_0)(\beta_1 - \hat{\beta}_1) + \left(\sum z_i^2\right)(\beta_1 - \hat{\beta}_1)^2 \right], \end{aligned}$$

que é a aproximação quadrática da superfície de logverossimilhança tomando o desvio padrão igual à sua MLE ($\sigma = \hat{\sigma}$). Portanto, a verossimilhança estimada para os coeficientes de regressão do modelo linear simples é equivalente a aproximação quadrática para estes coeficientes. Logo, a verossimilhança estimada para cada coeficiente de regressão individualmente pode ser obtida por:

$$\mathbf{L}_E\{\beta_0 | \hat{\beta}_1, \hat{\sigma}\} = -\frac{n}{2\hat{\sigma}^2} (\beta_0 - \hat{\beta}_0)^2 \quad \text{e} \quad \mathbf{L}_E\{\beta_1 | \hat{\beta}_0, \hat{\sigma}\} = -\frac{\sum z_i^2}{2\hat{\sigma}^2} (\beta_1 - \hat{\beta}_1)^2.$$



Verossimilhança Perfilhada

Outro “expediente” para se obter a verossimilhança contornando o parâmetro inconveniente γ é encontrar a MLE de γ para cada valor de θ . Ou seja, para cada valor fixo do parâmetro de interesse, se encontra o valor do parâmetro inconveniente que maximiza a função de log-verossimilhança.

Definição 8.20 (Verossimilhança Perfilhada). *Num modelo estocástico com espaço paramétrico multidimensional (θ, γ) , sendo γ o vetor dos parâmetros inconvenientes, a função de log-verossimilhança perfilhada do vetor de parâmetros de interesse θ é dada por*

$$\mathbf{L}_P\{\theta\} = \mathbf{L}_R\{\theta, \hat{\gamma}(\theta)\} = \max_{\gamma} (\mathbf{L}\{\theta, \gamma\}) ,$$

onde $\mathbf{L}_R\{\cdot\}$ é a função de log-verossimilhança relativa e $\hat{\gamma}(\theta)$ é a MLE dos parâmetros inconvenientes para cada valor dado dos parâmetros de interesse θ .

Sendo um dispositivo de cálculo da verossimilhança, a verossimilhança perfilhada também não é uma verossimilhança verdadeira. Mas ela representa a melhor aproximação numérica que se pode obter de “perfis cortados” na superfície de verossimilhança. Sua curvatura é sempre maior que a da verossimilhança estimada pois ela reflete a incerteza a respeito do parâmetro inconveniente para cada valor do parâmetro de interesse.



Exemplo 8-16: Estimação na Distribuição Gaussiana III

Na distribuição Gaussiana, o parâmetro da média é normalmente tomado como parâmetro de interesse, enquanto o desvio padrão é tomado como parâmetro inconveniente. Assim, na verossimilhança perfilhada da média, o parâmetro desvio padrão passa ser uma função maximiza a verossimilhança em função do valor do parâmetro média: $\hat{\sigma}(\mu)$. Logo, se na função log-verossimilhança relativa

$$\mathbf{L}_R\{\mu, \hat{\mu}, \sigma, \hat{\sigma}\} = n \left[\ln \left(\frac{\hat{\sigma}}{\sigma} \right) + \frac{1}{2} \left(1 - \frac{\sum_{i=1}^n (x_i - \mu)^2 / n}{\sigma^2} \right) \right] ,$$

o desvio padrão (σ) for substituído pela função $\hat{\sigma}(\mu)$, obtem-se a log-verossimilhança perfilhada:

$$\mathbf{L}_P\{\mu\} = n \left[\ln \left(\frac{\hat{\sigma}}{\hat{\sigma}(\mu)} \right) + \frac{1}{2} \left(1 - \frac{\sum_{i=1}^n (x_i - \mu)^2 / n}{[\hat{\sigma}(\mu)]^2} \right) \right] ,$$

sendo $\hat{\sigma}$ a MLE do desvio padrão.

A função $\hat{\sigma}(\mu)$ deve maximizar a log-verossimilhança perfilhada para cada valor de μ , por isto, para encontrá-la basta resolver a equação que iguála a primeira derivada da log-verossimilhança perfilhada a zero:

$$\begin{aligned} \frac{\partial \mathbf{L}_P\{\mu\}}{\partial \hat{\sigma}(\mu)} &= -\frac{n}{\hat{\sigma}(\mu)} + \frac{1}{[\hat{\sigma}(\mu)]^3} \sum_{i=1}^n (x_i - \mu)^2 = 0 \\ \Rightarrow \hat{\sigma}(\mu) &= \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}. \end{aligned}$$

A função de log-verossimilhança perfilhada se torna bem mais simples se inserir esta função na sua expressão original:

$$\mathbf{L}_P\{\mu\} = \frac{n}{2} \ln \left(\frac{\hat{\sigma}}{\hat{\sigma}(\mu)} \right) = \frac{n}{2} \ln \left(\frac{\sum_{i=1}^n (x_i - \hat{\mu})^2}{\sum_{i=1}^n (x_i - \mu)^2} \right).$$

A figura 8.9 compara as curvas de log-verossimilhança perfilhada e estimada para uma amostra de tamanho $n = 50$. Ressalte-se que, para este tamanho de amostra, numa vizinhança suficientemente próxima à MLE, como o intervalo de log-verossimilhança, as curvas são praticamente idênticas.

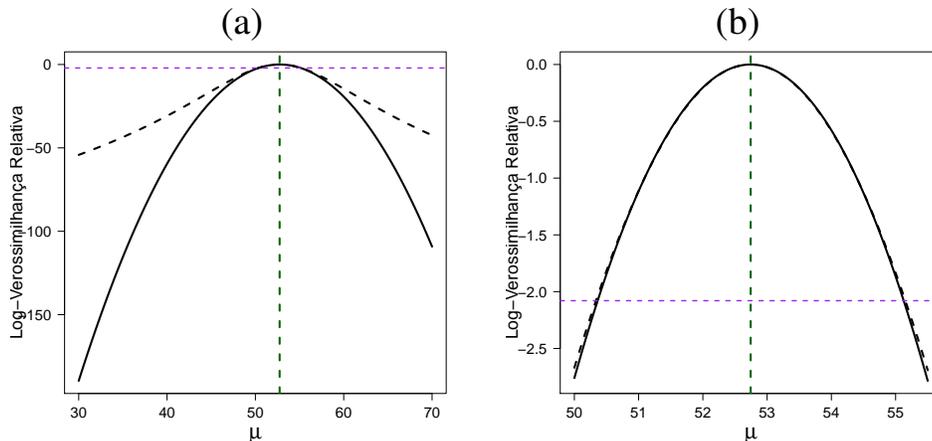


Figura 8.9: Curvas de log-verossimilhança estimada (linha sólida) e de log-verossimilhança perfilhada (linha tracejada) para a média da distribuição Gaussiana em duas vizinhanças da estimativa de máxima verossimilhança (MLE) numa amostra de tamanho $n = 50$. A linha tracejada vertical indica a posição da MLE da média ($\hat{\mu}$), enquanto a linha tracejada horizontal indica a diferença de log-verossimilhança de $-\ln(8)$.

A figura 8.10 apresenta o gráfico de contorno da superfície de log-verossimilhança de uma amostra de tamanho $n = 100$ da distribuição Gaussiana, mostrando o “trajeto” das curvas de log-verossimilhança estimada e perfilhada para a média na superfície. A curva de log-verossimilhança estimada possui um “trajeto reto” (linha horizontal), pois nela o valor do desvio padrão é fixado na MLE. Já a curva de log-verossimilhança perfilhada tem um “trajeto curvo”, pois, para cada valor da média, ela procura o valor do desvio padrão que maximiza a log-verossimilhança, isto é, em cada direção da média (direção vertical no gráfico), ela procura o ponto mais alto na superfície.

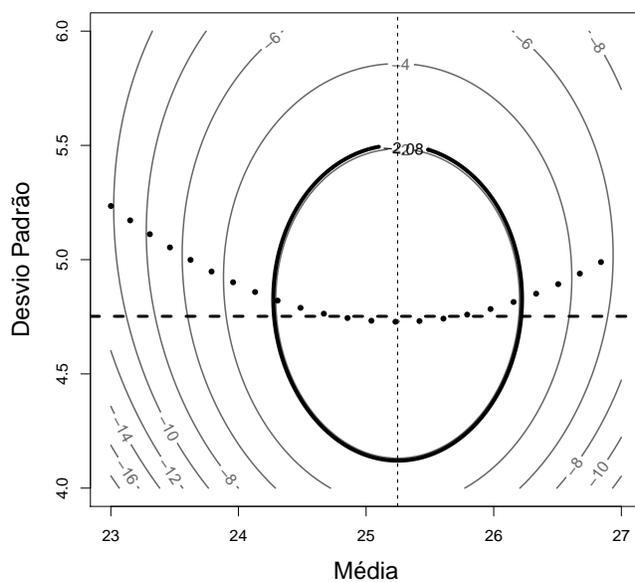


Figura 8.10: Gráfico de contorno da superfície de log-verossimilhança para a uma amostra de tamanho $n = 100$ da distribuição Gaussiana, mostrando as curvas da log-verossimilhança estimada (linha tracejada) e de log-verossimilhança perfilhada (linha pontilhada). A linha tracejada vertical indica a posição da MLE da média ($\hat{\mu}$), enquanto a região delimitada por linha sólida espessa corresponde à diferença de log-verossimilhança de $-\ln(8)$.





Exemplo 8-17: Modelo de Regressão Linear Simples III

Voltando ao exemplo da regressão linear simples, é interessante se comparar as curvas de log-verossimilhança estimada e perfilhada. A curva estimada foi apresentada no exemplo anterior (exemplo 8.5). Para se deduzir da curva perfilhada, parte-se novamente da função de log-verossimilhança relativa:

$$\mathbf{L}_R\{\beta_0, \hat{\beta}_0, \beta_1, \hat{\beta}_1, \sigma, \hat{\sigma}\} = n \ln\left(\frac{\hat{\sigma}}{\sigma}\right) - \frac{1}{2} \left[\frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \beta_0 - \beta_1 z_i)^2 - n \right].$$

Inicialmente, encontra-se a função do desvio padrão em função dos coeficientes de regressão ($\hat{\sigma}(\beta_0, \beta_1)$), pela solução da primeira derivada da função de log-verossimilhança relativa em relação ao desvio padrão:

$$\begin{aligned} \frac{\partial \mathbf{L}_R\{\beta, \beta_1, \sigma\}}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \beta_0 - \beta_1 z_i)^2 = 0 \\ \Rightarrow \hat{\sigma}(\beta_0, \beta_1) &= \left[\frac{\sum_{i=1}^n (x_i - \beta_0 - \beta_1 z_i)^2}{n} \right]^{1/2}. \end{aligned}$$

Introduzindo a função $\hat{\sigma}(\beta_0, \beta_1)$ n função de log-verossimilhança, obtem-se a função da *superfície perfilhada* dos coeficientes de regressão:

$$\mathbf{L}_P\{\beta_0, \beta_1\} = \mathbf{L}_R\{\beta_0, \beta_1 | \hat{\sigma}(\beta_0, \beta_1)\} = n \ln(\hat{\sigma}) - \frac{n}{2} \ln\left(\frac{1}{n} \sum_{i=1}^n (x_i - \beta_0 - \beta_1 z_i)^2\right).$$

Esta expressão é função apenas dos coeficientes de regressão e, portanto, a partir dela é possível se obter as funções que permitem obter a log-verossimilhança perfilhada de cada coeficiente individualmente. Novamente, isto é obtido igualando-se a zero a primeira derivada desta função em relação aos coeficientes de regressão:

$$\begin{aligned} \frac{\partial \mathbf{L}_R\{\beta_0, \beta_1, \sigma\}}{\partial \beta_0} &= n \frac{\sum_{i=1}^n (x_i - \beta_0 - \beta_1 z_i)}{\sum_{i=1}^n (x_i - \beta_0 - \beta_1 z_i)^2} = 0 \\ \Rightarrow \hat{\beta}_0(\beta_1) &= \frac{\sum_{i=1}^n x_i}{n} - \beta_1 \frac{\sum_{i=1}^n z_i}{n} = \bar{x} - \beta_1 \bar{z} \\ \frac{\partial \mathbf{L}_R\{\beta_0, \beta_1, \sigma\}}{\partial \beta_1} &= n \frac{\sum_{i=1}^n [(x_i - \beta_0 - \beta_1 z_i) z_i]}{\sum_{i=1}^n (x_i - \beta_0 - \beta_1 z_i)^2} = 0 \\ \Rightarrow \hat{\beta}_1(\beta_0) &= \frac{1}{\sum_{i=1}^n z_i^2} \left[\sum_{i=1}^n x_i z_i - \beta_0 \sum_{i=1}^n z_i \right]. \end{aligned}$$

Inserindo-se estas funções na função da superfície perfilhada se obtém a curva de log-verossimilhança perfilhada de cada coeficiente de regressão individualmente:

$$\begin{aligned} \mathbf{L}_P\{\beta_0\} &= \mathbf{L}_R\{\beta_0|\hat{\sigma}(\beta_0, \beta_1), \hat{\beta}_1(\beta_0)\} \\ &= n \ln(\hat{\sigma}) - \frac{n}{2} \ln \left\{ \frac{1}{n} \sum_{i=1}^n [x_i - \beta_0 - \hat{\beta}_1(\beta_0)z_i]^2 \right\} \\ \mathbf{L}_P\{\beta_1\} &= \mathbf{L}_R\{\beta_1|\hat{\sigma}(\beta_0, \beta_1), \hat{\beta}_0(\beta_1)\} \\ &= n \ln(\hat{\sigma}) - \frac{n}{2} \ln \left\{ \frac{1}{n} \sum_{i=1}^n [x_i - \hat{\beta}_0(\beta_1) - \beta_1 z_i]^2 \right\} \end{aligned}$$

A figura 8.11 apresenta as curvas de log-verossimilhança estimadas e perfilhadas para os coeficientes de regressão do modelo linear simples, com base em amostras simuladas de dois tamanhos ($n = 50$ e $n = 200$). Nota-se que as curvas perfilhadas tem curvatura sempre menor que as curvas estimadas, pois elas incorporam a incerteza associada a estimação dos demais parâmetros do modelo para cada valor do parâmetro sendo perfilhado.

O aumento do tamanho da amostra reduz a vizinhança ao redor das MLE relativa ao intervalo de log-verossimilhança de $-\ln(8)$, mas a diferença entre as curvas estimadas e perfilhada permanece. Esta é uma forte indicação da inadequação da verossimilhança estimada na avaliação dos coeficientes de regressão dos modelos lineares clássicos.

Um procedimento comum na análise dos modelos de regressão linear é o teste da hipótese de que o valor do coeficiente de inclinação (β_1) é nulo. Observando-se os gráficos da figura 8.11 referentes ao coeficiente de inclinação, verifica-se que, segundo a curva *perfilhada*, o intervalo de log-verossimilhança para diferença de $-\ln(8)$ não inclui o valor zero. Conclui-se, portanto, que é *pouco plausível* que o coeficiente de regressão seja nulo.

É importante enfatizar que a interpretação do intervalo de log-verossimilhança não deve ser confundida como a interpretação dos testes estatísticos de hipóteses. Por isto, não é apropriado dizer que o parâmetro do coeficiente de regressão “*difere significativamente*” de zero.



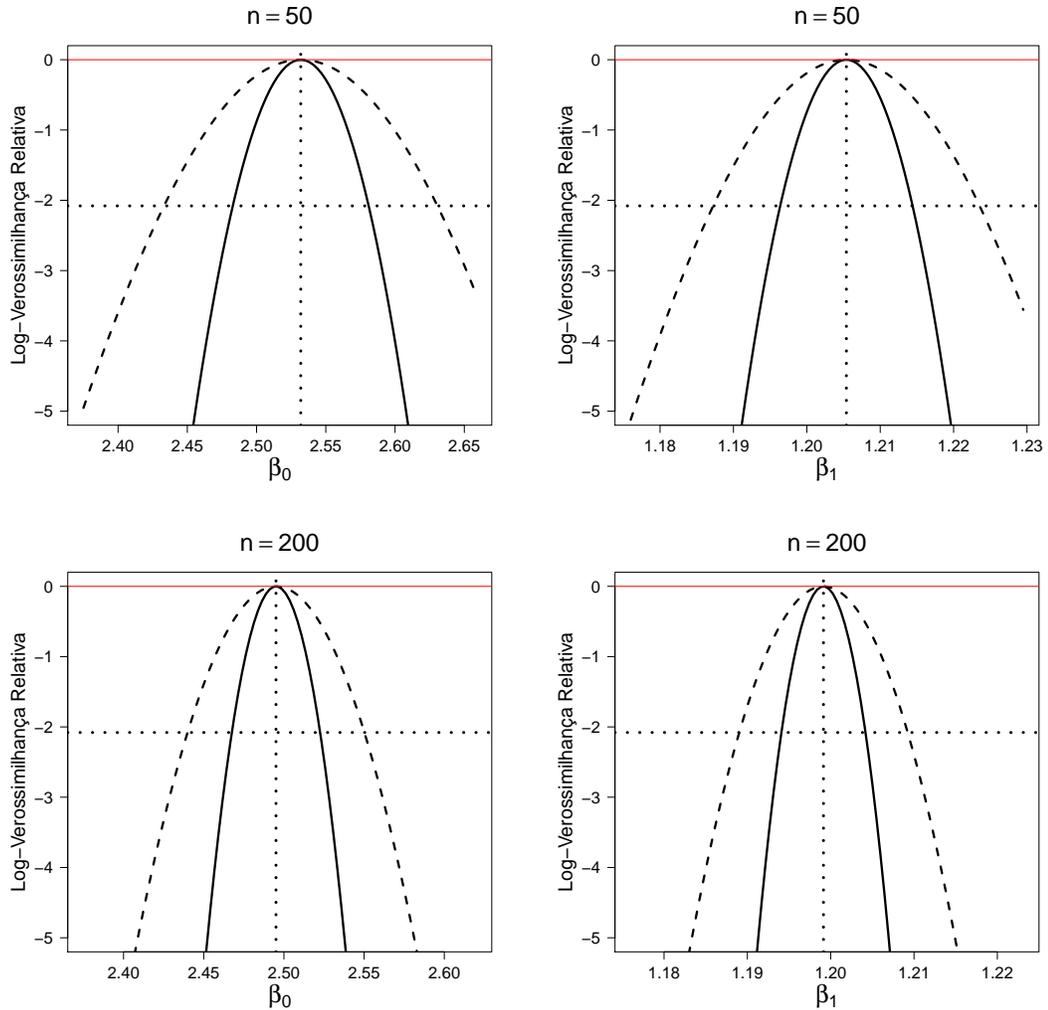


Figura 8.11: Curvas de log-verossimilhança estimada (linha sólida) e de log-verossimilhança perfilhada (linha tracejada) dos coeficientes de regressão (β_0 e β_1) do modelo linear simples clássico, com base em dados simulados com dois tamanhos de amostra: $n = 50$ e $n = 200$ observações. A linha pontilhada vertical indica a posição da MLE, enquanto que a linha pontilhada horizontal indica a diferença de log-verossimilhança de $-\ln(8)$.



Exemplo 8-18: Estimação na Distribuição Weibull II

No exemplo 8.5 foi estudada a superfície de log-verossimilhança da distribuição Weibull com dois parâmetros, mostrando a sua complexidade e a dificuldade de realizar inferências a partir dela. Assim as curvas de log-verossimilhança esti-

mada e perfilhada podem ser importantes auxílios no estudo da distribuição Weibull. A função de log-verossimilhança relativa que define a superfície de log-verossimilhança é

$$\mathbf{L}_R\{\beta, \hat{\beta}, \gamma, \hat{\gamma}\} = n \left[-\ln\left(\frac{\hat{\gamma}}{\gamma}\right) + \hat{\gamma} \ln(\hat{\beta}) - \gamma \ln(\beta) + 1 \right] + (\gamma - \hat{\gamma}) \sum_{i=1}^n \ln(x_i) - \frac{1}{\beta^\gamma} \sum_{i=1}^n x_i^\gamma.$$

As curvas de log-verossimilhança estimada são diretamente obtidas substituindo-se o parâmetro inconveniente pela uma estimativa de máxima verossimilhança:

$$\begin{aligned} \mathbf{L}_E\{\beta\} &= n \left[\hat{\gamma} \ln\left(\frac{\hat{\beta}}{\beta}\right) + 1 \right] - \frac{1}{\beta^{\hat{\gamma}}} \sum_{i=1}^n x_i^{\hat{\gamma}} \\ \mathbf{L}_E\{\gamma\} &= n \left[-\ln\left(\frac{\hat{\gamma}}{\gamma}\right) + (\hat{\gamma} - \gamma) \ln(\hat{\beta}) + 1 \right] + (\hat{\gamma} - \gamma) \sum_{i=1}^n \ln(x_i) - \frac{1}{\hat{\beta}^\gamma} \sum_{i=1}^n x_i^\gamma, \end{aligned}$$

onde $\hat{\beta}$ e $\hat{\gamma}$ são as MLE dos parâmetros.

A curva de log-verossimilhança perfilhada do parâmetro da escala (β) é muito diferente da curva perfilhada do parâmetro da forma (γ). Como a estimativa de máxima verossimilhança do parâmetro de escala (β) é uma função do parâmetro da forma (γ)

$$\hat{\beta}(\gamma) = \left[\frac{\sum_{i=1}^n x_i^\gamma}{n} \right]^{1/\gamma},$$

a curva perfilhada do parâmetro da forma (γ) pode ser obtida substituindo-se a expressão acima na função de log-verossimilhança relativa, o que resulta em

$$\mathbf{L}_P\{\gamma\} = n \left[-\ln\left(\frac{\hat{\gamma}}{\gamma}\right) + \hat{\gamma} \ln(\hat{\beta}) - \ln\left(\sum_{i=1}^n x_i^\gamma\right) + \ln(n) \right] - (\hat{\gamma} - \gamma) \sum_{i=1}^n \ln(x_i)$$

Já a curva perfilhada do parâmetro de escala (β) requer uma função $\hat{\gamma}(\beta)$ que retorne o valor do parâmetro de forma (γ) que maximiza a log-verossimilhança para um dado valor fixo do parâmetro de escala (β). Contudo esta função só pode ser obtida encontrando-se o valor de γ que, para um valor fixo de β , seja a raiz da primeira derivada da função de log-verossimilhança em relação à γ . Portanto, a

curva perfilhada de β é dada por

$$\mathbf{L}_P\{\beta\} = n \left[-\ln \left(\frac{\hat{\gamma}}{\hat{\gamma}(\beta)} \right) + \hat{\gamma} \ln(\hat{\beta}) - \hat{\gamma}(\beta) \ln(\beta) + 1 \right] + \\ - [\hat{\gamma} - \hat{\gamma}(\beta)] \sum_{i=1}^n \ln(x_i) - \frac{1}{\beta^{\hat{\gamma}(\beta)}} \sum_{i=1}^n x_i^{\hat{\gamma}(\beta)},$$

onde $\hat{\gamma}(\beta)$ é a raiz da expressão:

$$\frac{\partial \mathbf{L}\{\beta, \gamma\}}{\partial \gamma} = \frac{n}{\gamma} + \sum_{i=1}^n \ln(x_i) - \frac{1}{\beta^\gamma} \sum_{i=1}^n x_i^\gamma \ln(x_i) = 0.$$

Logo, a curva perfilhada do parâmetro de escala (β) só pode ser construída por métodos números.

A figura 8.12 apresenta a superfície de log-verossimilhança e as curvas estimadas e perfilhadas para os parâmetros da distribuição Weibull ajustada a dados de DAP de árvores de uma floresta nativa do município de Bom Jardim, MA. Os gráficos da superfície mostram que, à medida que a inspeção da superfície se restringe à vizinhança mais imediata das estimativas de máxima verossimilhança, a forma da superfície se torna mais próxima da forma quadrática. Os gráficos das curvas de log-verossimilhança mostram que, no intervalo de log-verossimilhança ($-\ln(8)$), a curva perfilhada se apresenta com *maior curvatura* (mais “fechada”) que a curva estimada, no caso da parâmetro de escala (β), e praticamente coincidente com esta, no caso do parâmetro de forma (γ).



Multidimensionalidade e Parâmetros Inconvenientes na Prática

Royall (1997) ainda apresenta outras formas de lidar com os parâmetro inconvenientes e a multidimensionalidade, como a verossimilhança marginal, a verossimilhança condicional e a verossimilhança condicional sintética. Estes métodos, como o método dos parâmetros ortogonais resultam em verossimilhanças verdadeiras, ao contrário da verossimilhança estimada e perfilhada que são apenas aproximações voltadas à redução da dimensionalidade da superfície de verossimilhança. A aplicabilidade destes métodos, no entanto, está ligada a problemas e interesses específicos que geralmente não estão focados diretamente no valor dos parâmetros do modelo estocástico, mas em *relações* (soma, produto, razão) entre os parâmetros ou entre os modelos.

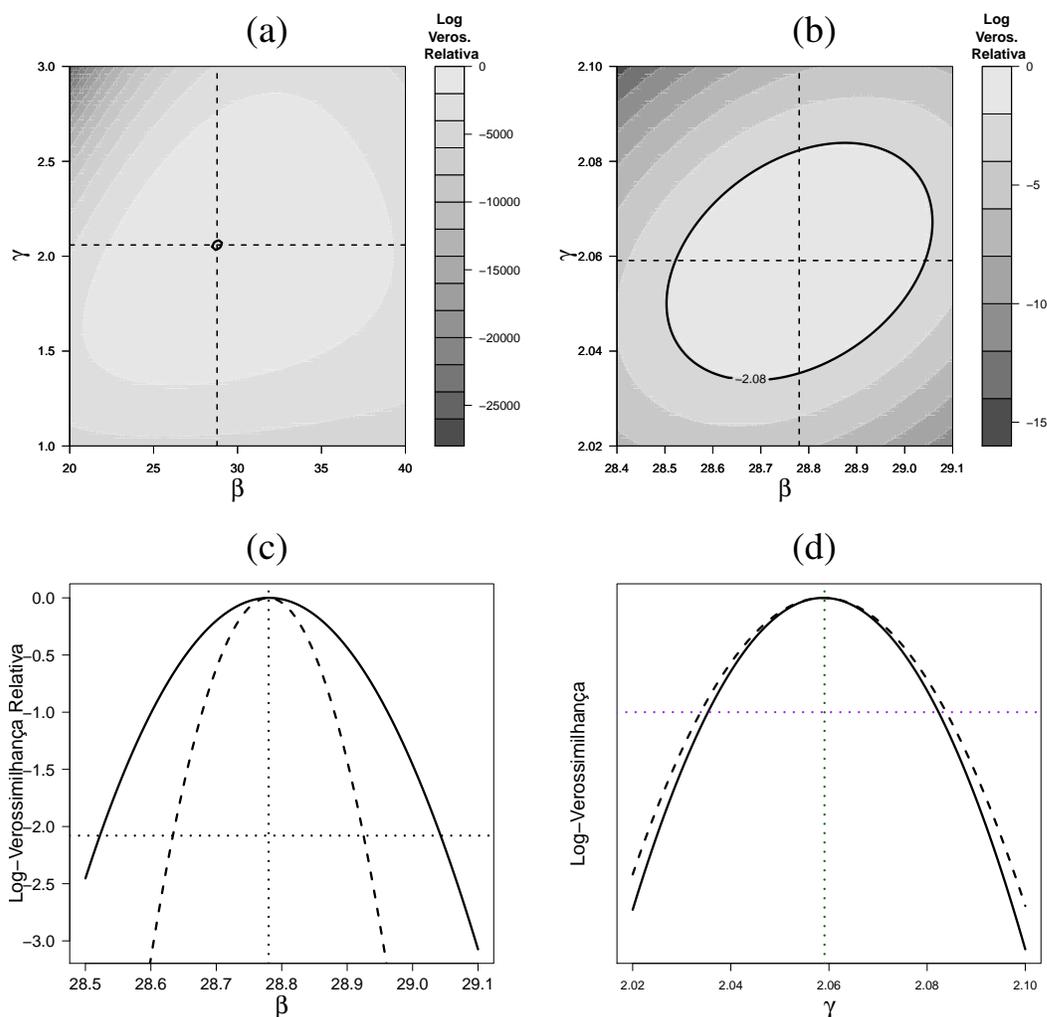


Figura 8.12: Superfície de log-verossimilhança da distribuição Weibull para dados de DAP de árvores de floresta nativa do município de Bom Jardim, MA, apresentando duas vizinhanças das MLEs ((a) e (b)); as linhas tracejadas indicam a posição das MLEs. Curvas de log-verossimilhança estimada (linha tracejada) e perfilhada (linha sólida) para os parâmetros de escala (β) e forma (γ) nos mesmos dados ((c) e (d)). A linha pontilhada vertical indica a posição da MLE, enquanto que a linha pontilhada horizontal indica a diferença de log-verossimilhança de $-\ln(8)$.

A solução de problemas práticos frequentemente requer modelos bastante complexos, de forma que o estudo teórico da superfície de log-verossimilhança pode se tornar muito trabalhoso. Considere, por exemplo, um problema de regressão linear com 2 variáveis preditoras. Este modelo, que é baseado na família Gaus-

siana, terá como parâmetros três coeficientes de regressão e o desvio padrão, de modo, que a superfície de log-verossimilhança será definida num espaço com 5 dimensões. Nesta situação, o estudo de regiões da superfície se mostra inviável e a inferência por verossimilhança se restringe à análise das curvas de log-verossimilhança perfilhada para cada parâmetro individualmente.

A curva perfilhada é geralmente utilizada, em preferência à curva estimada, por se tratar de uma abordagem mais conservadora na maioria das situações. Outra razão para seu uso mais frequente é que ela pode ser obtida por métodos numéricos a partir da função de log-verossimilhança, sem a necessidade da análise teórica do modelo em estudo. Esta operação é frequentemente chamada de “*perfilhamento da verossimilhança*” (*likelihood profiling*). Assim, na prática, tanto a obtenção das estimativas de máxima verossimilhança, quanto o estudo da sua incerteza são realizados por métodos numéricos utilizando softwares estatísticos especialmente desenvolvidos para esta finalidade.

* * *

Neste capítulo, apresentou-se a inferência estatística em termos da estimação dos parâmetros de modelos estocásticos e da avaliação da incerteza associada a estas estimativas. Em tese, após a estimação, um dado modelo já pode ser aplicado na resolução do problema prático para o qual ele foi desenvolvido. Entretanto, sempre existe mais de um modelo potencialmente útil para resolução de um dado problema prático particular. Surge, então, a questão de como se especificar e selecionar o modelo estocástico mais apropriado para cada problema. Trata-se agora de um outro aspecto da inferência estatística, que é o da especificação do modelo estocástico. A questão fundamental não é mais a comparação de valores para os parâmetros de um mesmo modelo, mas sim a comparação e o confronto entre modelos estocásticos candidatos à resolução do problema. É importante considerar que os modelos podem ser radicalmente distintos nas suas estruturas funcionais e também possuir número diferente de parâmetros a serem estimados. Portanto, a especificação e seleção de modelos leva a um contexto teórico distinto do contexto da estimação.

Capítulo 9

Especificação e Seleção de Modelos “O Ponto de Arquimedes” da Modelagem

9.1 Introdução

Estimar os parâmetros de um modelo estocástico é um aspecto importante da inferência estatística, mas, como foi visto, todo o processo de estimação está condicionado a um dado modelo estocástico. Igualmente importante é determinar qual o modelo estocástico mais apropriado para um determinado problema prático em questão.

Para Fisher (1922), a especificação é um dos problemas fundamentais da Estatística, consistindo na determinação da forma matemática da distribuição da população hipotética da qual a amostra observada seria proveniente. A especificação seria um problema essencialmente prático e de carácter qualitativo.

“As regards problems of specification, these are entirely a matter for the practical statistician, for those cases where the qualitative nature of the hypothetical population is known do not involve any problems of this type. In other cases we may know by experience what forms are likely to be suitable, and the adequacy of our choice may be tested *a posteriori*. We must confine ourselves to those forms which we know how to handle, or for which any tables which may be necessary have been constructed. More or less elaborate forms will be suitable according to the volume of the data. Evidently these are considerations the nature of which may change greatly during the work of a single generation.”

(Fisher, 1922a, p. 314p.)

Fisher enfatiza o aspecto empírico da especificação, ressaltando a importância da verificação com base no conceito de teste de aderência:

“... of even greater importance is the introduction of an objective criterion of goodness of fit. For empirical as the specification of the hypothetical population may be, this empiricism is cleared of its dangers if we can apply a rigorous and objective test of the adequacy with which the proposed population represents the whole of the available facts.”

(Fisher, 1922a, p. 314p.)

Seguindo a abordagem de Fisher, a implementação da verificação empírica na especificação de modelo permaneceu focalizada nos testes de aderência por muitas décadas. Em comparação ao desenvolvimento da teoria de estimação, a teoria sobre especificação de modelos se desenvolveu de forma mais lenta. Soma-se ao desenvolvimento teórico a impossibilidade de se implementar uma abordagem mais empírica, que se baseasse no ajuste de vários modelos e na comparação de performance.

Somente na década de 1970, com a utilização mais generalizada do computador eletrônico é que se tornou possível a abordagem empírica baseada na performance dos modelos. O impacto da tecnologia computacional pode ser apreciado na própria mudança da terminologia. Deixou-se de falar em “especificação” e passou-se a falar em “*seleção*” de modelos. A argumentação de aspectos teóricos de adequação dos modelos foi quase totalmente abandonada e substituída pela avaliação empírica da performance.

Entretanto, a maior parte dos métodos de seleção de modelos focalizava na seleção de modelos “*aninhados*”, isto é, na redução da dimensionalidade do modelo face a existência de um “modelo completo”. As técnicas de uso mais comum, como o coeficiente de determinação, teste F e C_p de Mallow, foram desenvolvidos sempre no contexto da aplicação do conceito de parcimônia para determinação do menor modelo de igual performance ao modelo completo. Mas nas últimas décadas do século XX, uma teoria voltada a seleção de modelos, no seu aspecto mais amplo, começou a ser desenvolvida com base na aplicação da teoria da informação na estatística.

Neste capítulo, será tratada a questão de especificação e seleção de modelos. Procura-se mostrar que a seleção de modelos não substitui totalmente o problema da especificação, uma vez que em qualquer aplicação prática a adequação de um modelo não depende apenas da sua performance relativa a outros modelos.

9.2 *Especificação de Modelos*

O problema da especificação de modelos tem dois aspectos. O primeiro é o “*aspecto prático*” relativo ao problema ou situação em que o modelo será utilizado

para predição. A situação ou problema em questão propõe objetivos ou estabelece restrições de ordem prática que devem ser satisfeitas para que as predições sejam as mais realistas possíveis. O segundo aspecto da especificação é o “*aspecto teórico*” referente à estrutura teórica dos modelos em estudo. Cada modelo estocástico, sendo uma formulação matemática, tem a sua estrutura própria, que requer que certas premissas, ou pressuposições, a respeito do fenômeno estudado sejam assumidas como válidas.

Aspecto Prático

Em cada problema prático ou situação de aplicação, os modelos estocásticos devem: (i) responder aos objetivos que se pretende alcançar, (ii) satisfazer as exigências essenciais do problema ou situação, e (iii) considerar as características e limitações impostas pelo problema, situação prática ou dados coletados. Como a aplicação de modelos estocásticos é um problema essencialmente prático, é ilusório acreditar que existam métodos padrão de especificação de modelos que possam ser aplicados uniformemente a todos os problemas ou situações. Uma aplicação verdadeiramente “prática” não pode prescindir das considerações práticas que são particulares a cada problema. Assim, é fundamental que os objetivos, limitações e exigências específicos de cada questão sejam objetivamente estabelecidos para que possam ser ponderados no processo de especificação de modelos. Em resumo, o conhecimento e a familiaridade com o problema prático é o elemento fundamental na especificação de modelos.



Exemplo 9-1: Predição da Altura \times Predição do Volume

Compare-se a questão da predição da altura contra a predição do volume de árvores individuais num levantamento cujo objetivo é a determinação da produção de uma floresta. Em ambas predições, o objetivo é determinar valores para atributos das árvores individualmente que seriam de medição custosa (altura) ou de medição destrutiva (volume).

As exigências impostas a essas predições, contudo, tem carácter diferente. As predições de altura serão utilizados como informação na predição do volume das árvores individualmente e, portanto, são predições que influenciam predições. Enquanto as predições de volume são agregados para se determinar os volumes das parcelas, ou unidades amostrais, que, por sua vez, são utilizados num procedimento de estimação da produção média por unidade de área ou na produção total da floresta. O problema da propagação da incerteza de predição tem pouca relevância na predição do volume, pois o processo de agregação tende a compensar as

incertezas, aumentando a precisão relativa no agregado. Já na predição da altura, a propagação da incerteza de predição é muito importante, uma vez que a incerteza na predição da altura contamina a predição do volume.

Os dados utilizados nos dois casos têm características e limitações bastante diferentes também. Os dados para construção de modelos para predição de volume são obtidos por meio de procedimento específico para este fim, que é a amostragem destrutiva das árvores na floresta onde se realiza o levantamento. Já os dados para construção dos modelos para predição da altura são obtidos durante o procedimento de medição das parcelas ou unidades amostrais durante a execução do levantamento. Espera-se, portanto, que os dados relativos às alturas das árvores sejam mais susceptíveis a erros de medição que os dados relativos ao volume, que resultam da cubagem rigorosa de árvores abatidas. Por outro lado, os dados de altura permitem uma modelagem mais detalhada, pois eles são identificados para cada parcela ou unidade amostral do levantamento, enquanto que os dados de volume são geralmente identificados no nível de talhão ou comportamento da floresta.

Tais considerações são raramente encontradas nos trabalhos científicos de Biometria Florestal, pois elas estão ligadas às particularidades de cada situação prática florestal. Contudo, elas são indispensáveis para se produzir modelos de qualidade para Mensuração Florestal.



Aspecto Teórico

O aspecto teórico do procedimento de especificação de modelo deve oferecer respostas ao aspecto prático do problema ou situação em estudo por meio da explicitação das *premissas do modelo*. Tais premissas podem ser de três tipos: (i) premissas referentes à estrutura do fenômeno estudado, (ii) premissas relativas ao procedimento de observação e estudo do fenômeno, isto é, premissas referentes à coleta de dados, e (iii) premissas impostas artificialmente pela estrutura determinística e estocástica do modelo. É fundamental que se possa distinguir entre estes três tipos de premissas uma vez que a relevância prática delas é muito diferente. Essa distinção se faz particularmente importante no momento da verificação e comparação empírica dos modelos candidatos.

Em Mensuração Florestal, é comum que o primeiro tipo de premissa seja ignorado, uma vez que os modelos são tomados como meras “*aproximações*” do

fenômeno estudado. Entretanto, parte do julgamento da qualidade da aproximação depende da estrutura do fenômeno estudado. A situação ótima ocorre quando é possível verificar empiricamente as questões estruturais do fenômeno em relação aos modelos.

Em teoria, a incompatibilidade entre as premissas referentes à estrutura do fenômeno e aquelas impostas pela estrutura matemática do modelo são consideradas muito importantes. Mas, essa incompatibilidade não tem necessariamente um grande impacto sobre a qualidade das previsões geradas pelo modelo. Neste caso, as considerações sobre as técnicas matemáticas e estatísticas para adaptação do modelo, como transformação da variável resposta, podem ser de pouca relevância prática.



Exemplo 9-2: Predição da Altura \times Predição do Volume II

Os modelos de predição de altura e de volume têm premissas semelhantes quanto à estrutura do fenômeno estudado. Em ambos os casos, são modelos destinados à predição de *medidas* de atributos de árvores individuais (*medidas arborimétricas*) e, conseqüentemente, os erros de predição se assemelham aos erros de medição.

Nestes modelos de predição, espera-se que os mesmos produzam erros simétricos em relação ao erro médio, o qual, por sua vez, espera-se que seja nulo. Logo, as distribuições estocásticas candidatas para modelar este fenômeno são aquelas simétrica à média, cuja estimativa espera-se que seja nula, e que possa cobrir valores positivos e negativos. A distribuição “natural” para este tipo de problema é a distribuição Gaussiana, que foi desenvolvida para solucionar o problema do erro de medição de medidas astronômicas. Por isto, é até certo ponto “natural” que as técnicas de regressão linear e não linear sejam de uso tão intensivo na Biometria Florestal.

Em relação às premissas impostas pelos procedimentos de coleta de dados, espera-se nos modelos para erros de medição, as observações sejam independentes, pois não se justifica que a medição do atributo de uma árvore tenha qualquer influência sobre a mensuração do mesmo atributo em outra árvore. Contudo, como as árvores são medidas em conglomerados, definidos por parcelas ou outras unidades amostrais, é possível que haja certo grau de dependência entre as medidas obtidas, embora o procedimento de medição de cada árvore individualmente possa ser tomado como relativamente independente. Modelos mais sofisticados, como os modelos de efeitos mistos, podem incorporar este aspecto de dependência espacial, sem a eliminação do erro aleatório simétrico de média nula, que é o componente estocástico apropriado para modelagem do erro de predição.

Uma premissa que é frequentemente imposta pelos modelos de regressão linear e não linear em muitos problemas de Biometria Florestal é a homogeneidade da variância (*homocedasticidade*). É possível que na predição da altura a homocedasticidade seja uma premissa razoável, mas na predição do volume das árvores individualmente ela é certamente inapropriada. A “*Lei de Taylor*” ou “*Lei da Potência*” (Taylor, 1961a; Taylor, 1961b; Taylor, 1984; Taylor, Perry, Woiwod e Taylor, 1988) nos garante que a norma na natureza é que exista uma relação de potência entre a variância e a média dos atributos de indivíduos, de modo que a homocedasticidade deve ser antes a exceção do que a regra. A transformação da variável resposta, com o único objetivo de satisfazer a premissa da homocedasticidade, representa, portanto, uma concepção à uma condição imposta artificialmente pelos modelos de regressão clássicos. Uma solução mais realista, requer o uso de modelos mais sofisticados, onde a heterocedasticidade possa ser incorporada ao modelo.



Componente Estocástico e Componente Determinístico

O aspecto prático da especificação de modelos pede por um procedimento que parte do problema em direção aos modelos candidatos, isto é, em direção ao aspecto teórico. Logo, o aspecto prático se concretiza na especificação de uma *classe de modelos* que atendam às particularidades do problema em estudo. Mas, dois *componentes* devem ser levados em consideração no processo de especificação dessa classe de modelo: o componente estocástico e o componente determinístico.

O componente estocástico é se refere às distribuições estocásticas candidatas à modelagem do problema. Como foi mostrado anteriormente (capítulo 7), cada distribuição estocástica representa um padrão próprio de estocasticidade definida pelas funções de densidade e de distribuição. Portanto, o primeiro passo na especificação de uma classe de modelos é a especificação da distribuição estocástica, ou das distribuições estocásticas, que se mostram apropriadas para a modelagem do problema. Em alguns problemas, a especificação pode chegar a definir uma única distribuição, mas em outros problemas, a especificação teórica do componente estocástica é muito difícil ou impraticável, de modo que apenas um grupo de distribuições candidatas pode ser definido, apoiando-se nas evidências empíricas (dados) para a definição da distribuição estocástica mais apropriada.

O componente determinístico ligado a especificação de uma classe de modelos estocástico está associado a determinação dos parâmetros da distribuição estocás-

tico candidata, ou das distribuições estocásticas candidatas. Os parâmetros do modelo podem ser tomados como constantes ou como funções de medidas preditoras, sendo que tais funções podem ter a estrutura linear ou não-linear. Para um mesmo problema e distribuição estocástica, diversas funções, ou formas funcionais, podem ser candidatas para os parâmetros do modelo. Boa parte da literatura em Biometria Florestal consiste em trabalhos que comparam e discutem o desempenho dessas funções para as várias situações práticas da Mensuração Florestal.



Exemplo 9-3: Modelos de Predição e de Distribuição

Um bom exemplo das diferenças do componente estocástico na especificação de modelos na Biometria Florestal é o contraste entre modelos arborimétricos¹, isto é, modelos para predição de medidas de árvores individuais, e modelos para representar a distribuição de DAP das árvores de um povoamento florestal.

Como foi apresentado acima (exemplo 9.2), os modelos arborimétricos de predição são especificados utilizando a distribuição Gaussiana como tendo a estrutura estocástica mais apropriada para a modelagem das medidas e os erros de predição. Nestes modelos, a especificação de modelos candidatos diferentes normalmente se restringe à especificação do componente determinístico. Os diversos modelos de relação hipsométrica ou de equação de volume encontrados na literatura florestal se referem essencialmente à forma funcional do parâmetro média, que na distribuição Gaussiana é idêntica ao valor esperado. Assim, a questão essencial na especificação e seleção dos modelos arborimétricos de predição é a definição da forma funcional apropriada para a média.

Nos modelos de distribuição de DAP, por outro lado, a questão fundamental é exatamente a determinação da estrutura estocástica apropriada para representar os dados, uma vez que as diferenças na distribuição do DAP refletem diferenças na distribuição estocástica. Nestes modelos, toma-se o componente determinístico como sendo definido por parâmetros constantes em cada situação particular de povoamento florestal. Logo, o componente determinístico é apenas uma questão de estimação, e não especificação ou seleção de modelos.

No caso dos problemas de predição arborimétrica, a especificação de modelos

¹Neste trabalho, utiliza-se a terminologia utilizada por Batista, do Couto e da Silva Filho (2014). Medidas e modelos que são referentes às árvores individualmente são designados por *arborimétricos*, enquanto que os referentes a agregados de árvores, como parcelas ou unidades amostrais (*arvoredos*) são chamados de *arbustimétricos*. Reserva-se a designação *silvimétricos* apenas para medidas e modelos que integram dados observados nos vários arvoredos (parcelas ou unidades amostrais) que compõem a observação da floresta com um todo.

consiste na definição de uma classe de modelos Gaussianos com diferentes formas funcionais para a média. A definição dessas formas funcionais implica não só na escolha entre formas lineares e não-lineares, mas também na seleção de quais medidas ou variáveis preditoras serão utilizadas. Por outro lado, no caso dos problemas de distribuição de DAP, a especificação de uma classe de modelos implica na determinação de classes genéricas quanto à *forma* da distribuição estocástica, como distribuições simétricas, assimétricas (à direita ou à esquerda), distribuições em J-invertido, distribuições com valores extremos, etc.



9.3 Seleção de Modelos

O problema da especificação dos modelos é naturalmente enfrentado antes do problema de seleção. Definida uma classe de modelos apropriada ao problema e tendo-se a disposição um conjunto de dados, surge a questão de qual, ou quais, dos modelos é mais apropriado para o problema em questão. Os procedimentos de seleção de modelos devem atender a essa demanda, mas a Estatística oferece inúmeros critérios e métodos que se pode utilizar na seleção de modelos. Forster (2000) apresenta uma revisão dos principais métodos e procura construir um arcabouço conceitual geral, contudo, fica claro que a literatura técnica está longe de um consenso sobre seleção de modelos e que uma abordagem geral consistente e amplamente aceita para o problema ainda não foi alcançada.

Contudo, dois conceitos parecem ser importantes para se delinear uma abordagem consistente sobre a seleção de modelos. O primeiro deles é o conceito teórico de *discrepâncias* apresentado por Linhart e Zucchini (1986). O outro, é o conceito de *critério de seleção*, que é o instrumento quantitativo para seleção dos modelos na prática.

Discrepâncias

Linhart e Zucchini (1986) apresentaram o conceito de discrepância como fundamento teórico para seleção de modelos e o aplicaram a uma série de problemas estatísticos clássicos. Zucchini (2000) fez uma apresentação mais concisa e didática do mesmo assunto. Embora o conceito de discrepância não tenha se tornado de uso generalizado nas aplicações estatísticas, ele fornece um instrumental de racionalização do problema de seleção de modelos bem fundamentado e suficien-

temente geral para ser aplicado nas mais variadas situações práticas.

Modelo e Família Operante

Inicialmente os autores definem o conceito de *modelo operante* (*operating model*) que seria o “modelo verdadeiro” que gerou os dados observados. Na definição apresentada neste trabalho, o modelo operante é o modelo estocástico subjacente ao cenário estocástico em estudo. Este modelo nunca é totalmente conhecido, podendo ser com frequência completamente desconhecido. Há entretanto, situações nas quais algumas características deste modelo podem ser determinadas, e uma *família operante*, isto é, numa classe de modelos, pode ser especificada.

Definição 9.1 (Modelo e Família Operante). *Modelo Operante* é o modelo estocástico subjacente ao cenário estocástico em estudo, ou seja, é o modelo que gera os dados observados.

Família Operante é a classe de modelos estocásticos à qual o modelo operante pertence.



Exemplo 9-4: Modelos de Predição e de Distribuição II

A distinção entre modelos de predição e de distribuição de DAP também se torna clara com o conceito de família e modelo operante. No caso dos modelos aborimétricos de predição, assume-se que a família operante, em termos de componente estocástico, é a *Família Gaussiana*. Em termos do componente determinístico, não é possível, contudo, se especificar se é uma Família Gaussiana com média linear ou não linear, ou com desvio padrão homocedástico ou heterocedástico. Já nos problemas de distribuição de DAP, a família operante não pode ser especificada, pois essa é exatamente a questão central da modelagem.



Família e Modelo de Aproximação

Os modelos candidatos à explicar o cenário estocástico observado são chamados de *modelos de aproximação* (*approximating models*). Neste caso, também é muito raro que se saiba exatamente os valores dos parâmetros dos modelos candidatos. Logo é mais apropriado se falar de *famílias de aproximação*, que seriam candidatas a representar o modelo operante, a expressão “família” é usada neste termo no

sentido de “família de distribuições estocásticas” com seus componentes estocásticos e determinísticos definidos.

Definição 9.2 (Modelos e Famílias de Aproximação). *Famílias de Aproximação são classes de modelos estocásticos, candidatas a representar o cenário estocástico, ou seja, são candidatas a representar o modelo operante. As famílias de aproximação são famílias de distribuições estocásticas com os componentes estocásticos e determinísticos definidos.*

Modelo de Aproximação é um modelo de uma dada família de aproximação cujos valores dos parâmetros foram determinados.



Exemplo 9-5: Modelos de Predição e de Distribuição III

Tomando-se como exemplo de modelo de predição os modelos volumétricos, os modelos de aproximação são na sua maioria da Família Gaussiana. Mas, ao se definir diferentes componentes determinísticos se define diferentes *subfamílias de aproximação*. Por exemplo, considere os modelos volumétricos clássicos: o modelo de Spurr ou da variável combinada, o modelo Log-Spurr, o modelo Schumacher-Hall, o modelo Meyer, o modelo Stoaate, etc.. Em essência, todos estes modelos são da família Gaussiana, assumindo-se que o desvio padrão é constante, isto é, são modelos Gaussianos homocedásticos. A diferença que existe entre eles está no componente determinístico que define o parâmetro *média*, ou valor esperado da distribuição, como uma função linear de medidas predictoras: DAP e altura. Mesmo o modelo Schumacher-Hall na forma não-linear, também é um modelo Gaussiano homocedástico cuja média é definida por uma função não-linear do DAP e da altura. Nos modelos de predição, portanto, as discrepâncias de aproximação se devem geralmente às diferenças no componente determinístico dos modelos, pois, num mesmo problema, assume-se geralmente o mesmo componente estocástico para todos modelos candidatos.

Já no caso do modelos de distribuição de DAP, as famílias de aproximação são definidas pelo componente estocástico, isto é, elas são as próprias famílias estocásticas, como por exemplo, as famílias exponencial, gama, Weibull, etc. Neste caso, as discrepâncias de aproximação são devidas ao componente estocástico de cada modelo candidato, sendo o componente determinístico de importância secundária.



O exemplo (9.3) sugere que a definição de *subfamília de aproximação* será útil no estudo dos modelos de predição tipicamente utilizados na Biometria Florestal.

Definição 9.3 (Subfamílias de Aproximação). *Subfamílias de Aproximação são classes de modelos estocásticos candidatas a representar o modelo operante, que pertencem a uma mesma família de distribuição estocástica. Logo, as subfamílias de aproximação têm o mesmo componente estocástico da família de aproximação a que pertencem, diferindo entre si apenas em termos do componente determinístico.*

Dentro de uma dada família de aproximação, podem ser definidas tantas subfamílias quantas especificações matemáticas diferentes forem definidas para o componente determinístico.



Exemplo 9-6: Predição do Volume de Árvores

Os “modelos” volumétricos utilizados na Biometria Florestal são subfamílias da família Gaussiana em que, tradicionalmente, assume-se o desvio padrão constante. Exemplos dessas subfamílias, em que a média é tomada como uma função linear do DAP e da altura, são:

$$\begin{array}{ll}
 \text{Spurr:} & v_i = \beta_0 + \beta_1 (d_i^2 h_i) + \varepsilon_i \\
 \text{Stoate:} & v_i = \beta_0 + \beta_1 d_i^2 + \beta_2 (d_i^2 h_i) + \beta_3 h_i + \varepsilon_i \\
 \text{Meyer:} & v_i = \beta_0 + \beta_1 d_i + \beta_2 d_i^2 + \beta_2 (d_i h_i) + \\
 & \quad + \beta_4 (d_i^2 h_i) + \beta_5 h_i + \varepsilon_i \\
 \text{Meyer Modif.:} & v_i = \beta_0 + \beta_1 d_i + \beta_2 d_i^2 + \beta_2 (d_i h_i) + \beta_4 (d_i^2 h_i) + \varepsilon_i \\
 \text{Näslund:} & v_i = \beta_0 + \beta_1 d_i^2 + \beta_2 (d_i^2 h_i) + \beta_4 (d_i h_i^2) + \beta_5 h_i^2 + \varepsilon_i \\
 \text{Log-Spurr:} & \ln(v_i) = \beta_0 + \beta_1 \ln(d_i^2 h_i) + \varepsilon_i \\
 \text{Schumacher-Hall (linear):} & \ln(v_i) = \beta_0 + \beta_1 \ln(d_i) + \beta_2 \ln(h_i) + \varepsilon_i \\
 \text{Baden-Würt (linear):} & \ln(v_i) = \beta_0 + \beta_1 \ln(d_i) + \beta_2 \ln^2(d_i) + \\
 & \quad + \beta_3 \ln(h_i) + \beta_4 \ln^2(h_i) + \varepsilon_i
 \end{array}$$

nas quais: i é o índice que identifica cada árvore na amostra ($i = 1, 2, \dots, n$), v_i é o volume lenhoso da árvore, d_i é o DAP, h_i é a altura, $\ln(\cdot)$ é o logaritmo natural, β_j ($j = 1, 2, \dots, 5$) são os coeficientes de regressão e ε_i é o componente estocástico do modelo que segue a distribuição Gaussiana com média nula ($\mu = 0$) e desvio padrão constante (σ).

Como o componente estocástico é o mesmo para todas essas subfamílias, todas elas pertencem à família Gaussiana homocedástica. A única diferença entre essas famílias é o componente determinístico que define a média, que é dado pela soma dos coeficientes de regressão multiplicados pelas respectivas variáveis preditoras. Por exemplo, a subfamília Spurr segue a distribuição Gaussiana com desvio padrão constante e média dada por:

$$\mu_v = \mathbf{E}\{v_i|d_i, h_i\} = \beta_0 + \beta_1 (d_i^2 h_i).$$

Assim, nessa subfamília, o volume de cada árvore individualmente é dado por

$$v_i = \mu_v + \varepsilon_i = \beta_0 + \beta_1 (d_i^2 h_i) + \varepsilon_i.$$

Ao se estimar os parâmetros dessas subfamílias para um conjunto de 175 árvores de *Eucalyptus grandis* na região de Itatinga, são obtidos os seguintes *modelos de aproximação*:

Spurr: $\hat{\mu}_v = -0.001471 + 0.00003718 (d_i^2 h_i)$
 $\hat{\sigma} = 0.01581$

Stoate: $\hat{\mu}_v = 0.04141 - 0.003347 d_i^2 + 0.000049 (d_i^2 h_i) +$
 $-0.001419 h_i$
 $\hat{\sigma} = 0.01577$

Meyer: $\hat{\mu}_v = -0.00689 + 0.01802 d_i - 0.001297 d_i^2 +$
 $-0.0001826 (d_i h_i) + 0.000070 (d_i^2 h_i) +$
 $-0.003669 h_i$
 $\hat{\sigma} = 0.0158$

Meyer Modif.: $\hat{\mu}_v = -0.09781 + 0.03028 d_i - 0.00171 d_i^2 +$
 $-0.000677 (d_i h_i) + 0.0000869 (d_i^2 h_i)$
 $\hat{\sigma} = 0.01576$

Näslund: $\hat{\mu}_v = 0.05339 - 0.000229 d_i^2 + 0.0000306 (d_i^2 h_i) +$
 $+0.000018 (d_i h_i^2) - 0.000219 h_i^2$
 $\hat{\sigma} = 0.01577$

$$\begin{aligned} \text{Log-Spurr: } \hat{\mu}_{\ln(v)} &= -10.23165 + 1.00262 \ln(d_i^2 h_i) \\ \hat{\sigma} &= 0.06571 \end{aligned}$$

$$\begin{aligned} \text{Schumacher-Hall (linear): } \hat{\mu}_{\ln(v)} &= -10.3182 + 1.974641 \ln(d_i) + \\ &\quad + 1.0555 \ln(h_i) \\ \hat{\sigma} &= 0.06583 \end{aligned}$$

$$\begin{aligned} \text{Baden-Würt (linear): } \hat{\mu}_{\ln(v)} &= 1.93777 + 2.34655 \ln(d_i) + \\ &\quad - 0.06972 \ln^2(d_i) + -7.00025 \ln(h_i) + \\ &\quad + 1.27100 \ln^2(h_i) \\ \hat{\sigma} &= 0.06537 \end{aligned}$$

Todos estes modelos diferem apenas nos seu componente determinístico, uma vez que eles diferem apenas em termos da expressão da média, e todos têm o mesmo componente estocástico referente à família Gaussiana.



Discrepâncias de Aproximação e de Estimação

Segue-se que, entre o modelo operante e um dado modelo de aproximação, há dois tipos de discrepâncias. A primeira é relativa à qualidade da aproximação da família de aproximação e, portanto, é chamada de *discrepância de aproximação*. Ela resulta da diferença entre a forma funcional da família operante e a forma funcional da família de aproximação. Entenda-se “forma funcional” como contendo os componentes estocástico e determinístico.

Forster (2000) utiliza a expressão “viés do modelo” (*model bias*) para designar essa discrepância. Como no caso da estimação de parâmetros, a palavra “viés” possui um conceito tecnicamente bem definido, o seu uso na seleção de modelos parece emprestar um rigor técnico e uma segurança que são falsos. Por isto, a palavra “discrepância” se mostra mais adequada.

O segundo tipo de discrepância é aquela que resulta da incerteza a respeito dos valores mais apropriados para os parâmetros no modelo de aproximação. Os vários conjuntos de valores possíveis para os parâmetros resultam em vários mode-

los de aproximação distintos dentro de uma mesma família de aproximação. Estes vários modelos tem a *mesma* discrepância de aproximação, pois estão na mesma família, mas terão valores de discrepância distintos devido aos valores dos parâmetros. Como os valores dos parâmetros são estimados com base numa amostra, isto é, num conjunto de observações do cenário estocástico, este segundo tipo de discrepância é chamada de *discrepância de estimação*. Forster (2000) demonstra que essa discrepância é equivalente à “variância da estimação” (*estimation variance*)

A soma da discrepância de aproximação e da discrepância de estimação resulta na *discrepância global* (*overall discrepancy*), que é a discrepância total entre o modelo operante e um dado modelo de aproximação, já com o valor dos parâmetros determinados. A discrepância global é a única discrepância que geralmente podemos observar, pois é a diferença entre as observações, geradas pelo modelo operante, e a predições geradas pelo modelo de aproximação já estimado.

Definição 9.4 (Discrepâncias). *Discrepância de Aproximação é a discrepância entre a família operante e uma dada família de aproximação. Ela é devida à diferença da forma funcional (componentes estocástico e determinístico) entre as famílias.*

Discrepância de Estimação é a discrepância devida a diferentes valores de parâmetros entre os modelos de uma mesma família de aproximação. Ela é resultado da estimação dos parâmetros do modelo.

Discrepância Global é a soma da discrepância de aproximação e da discrepância de estimação, sendo a única discrepância observável, pois é a diferença entre as observações e as predições.

Discrepâncias e Seleção de Modelos

O conceito de discrepância deixa claro que o problema de seleção de modelos é, na sua essência, um *problema de aproximação*, isto é, de escolha da forma funcional mais apropriada para um modelo estocástico que represente as observações. Entretanto, é impossível separá-lo do problema de estimação, uma vez que para se poder calcular qualquer medida de adequação de um modelo é necessário determinar o valor de seus parâmetros, ou seja, de estimar os parâmetros.

Mais ainda, o conceito de discrepância nos mostra que qualquer medida de adequação de um modelo ou de comparação entre modelos é uma medida de discrepância global. Logo, nessas medidas há um “confundimento”, isto é, não é possível distinguir, das discrepâncias de aproximação e de estimação.

É importante ter em mente que, enquanto a discrepância de aproximação é devida *exclusivamente* à forma funcional da família candidata, a discrepância de estimação dependerá do procedimento de estimação (estimador) e do tamanho da amostra. A situação ideal seria comparar os modelos usando uma medida

de discrepância onde a discrepância de estimação foi *minimizada* para todas as famílias candidatas.

Contudo, os diversos procedimentos de estimação não tem o mesmo desempenho nos diferentes tamanhos de amostra, isto é, os procedimentos mais eficientes em grandes amostras podem não ser os melhores em amostras pequenas. A disponibilidade de sistemas computacionais eficientes tem privilegiado o método de máxima verossimilhança sobre os demais, salvo problemas mais complexos nos quais o método da máxima verossimilhança *restrita* ou outros métodos particulares se fazem mais vantajosos. No capítulo anterior, foram apresentados os argumentos em favor do método da máxima verossimilhança. Contudo, sabe-se que, no ajuste de um modelo, a influência do tamanho da amostra e do estimador *não* são independentes do modelo. Modelos mais simples necessitam de menores tamanhos de amostras e de estimadores mais simples que modelos mais complexos. Assim, manter constante o tamanho de amostra e utilizar um mesmo estimador não é necessariamente a forma mais “justa” de comparar famílias de aproximação, principalmente quando existe uma grande amplitude de complexidade entre as famílias candidatas.

Crítérios de Seleção de Modelos: Propriedades Esperadas

Para que a discrepância global possa ser utilizada na seleção do modelo mais apropriado a partir de um conjunto de modelos candidatos é necessário que ela seja convertida numa medida quantitativa. Essa medida é geralmente designada por “*critério*” de seleção de modelos. Alguns exemplos de critérios de seleção tradicionalmente utilizados na Biometria Florestal são o coeficiente de determinação (R^2), o erro padrão da estimativa ($s_{\hat{y}}$) e o critério de Furnival².

Mas, antes de se discutir os critérios de seleção é importante apresentar as propriedades que se espera de um critério de seleção ideal, conforme a concepção de que o critério é a medida quantitativa que expressa a discrepância global dos modelos candidatos. Seguem-se algumas propriedades diretamente ligadas às características e necessidades da Biometria Florestal que um critério de seleção ideal deveria satisfazer.

Capacidade Preditiva. Na Biometria Florestal, a performance dos modelos deve ser julgada em termos da sua capacidade preditiva, uma vez que a função dos modelos biométricos em Mensuração Florestal é essencialmente preditiva.

²A fim de se padronizar a terminologia, se utiliza a expressão *critério de Furnival*, embora, tradicionalmente, este critério tenha sido chamado de *índice* pelo próprio George M. Furnival (Furnival, 1961).

Independência do Modelo. O critério para comparação da performance não deve depender ou ser condicionado pelo modelo ou modelos a serem testados.

Critérios condicionados ao modelo só fazem sentido quando se tem como premissa o conhecimento da *classe* do modelo operante, seja essa classe uma família ou uma subfamília estocástica. Por exemplo, nos modelos de predição pode se tomar como premissa que o modelo operante é pertence à família Gaussiana homocedástica. Neste caso, o erro padrão da estimativa e o critério de Furnival são critérios adequados para seleção dos modelos candidatos. Caso se considere a possibilidade do modelo operante ser heterocedástico, a validade destes critérios passa a ser questionável. A determinação da classe do modelo operante, quando possível e apropriada, é a parte fundamental do procedimento de *especificação* dos modelos e, portanto, deve preceder a fase de enumeração e seleção dos modelos candidatos.

Condicionamento apenas aos Dados. A comparação da performance dos modelos deve estar condicionada apenas aos dados observados e disponíveis para a análise, pois as observações são o critério de evidência empírica e o juiz da plausibilidade dos modelos. A forma como as observações foram obtidas, que define o espaço amostral do modelo, é irrelevante, pois questões envolvendo a qualidade da coleta de dados são prévias e externas à seleção de modelos. Só faz sentido utilizar um certo conjunto de dados para selecionar modelos *depois* que a confiabilidade dos dados foi estabelecida.

Para efeitos práticos, também é importante que a comparação dos modelos esteja diretamente ligada à situação prática observada e que não seja associada ou dependente de um juízo a respeito do *procedimento* de construção do modelo. Critérios fundamentados nas *propriedades estatísticas sob amostragem repetida de uma população hipotética infinita* dizem mais sobre o procedimento do que sobre a situação prática em questão. Tais critérios tem relevância prática somente no que tange ao planejamento de um estudo ou levantamento, mas perdem a relevância depois que os dados foram obtidos. Em essência essa exigência significa a completa aceitação e aderência ao Princípio de Verossimilhança.

Partição Aditiva. É conveniente que uma medida quantitativa da performance dos modelos possa ser particionada segundo subconjuntos do conjunto de dados original, de modo que a medida global seja resultado da adição das medidas nos subconjuntos. Por exemplo, na comparação de modelos volumétricos é conveniente que a medida global possa ser obtida pela soma das medidas por classes de DAP, de forma que o desempenho dos modelos possa ser coerentemente comparado tanto no todo quanto nas classes de DAP.

Como essas propriedades se referem a um critério *ideal*, é possível que critérios úteis para determinadas classes de modelo não satisfaçam todas elas. Voltando ao exemplo da família Gaussiana homocedástica, o erro padrão da estimativa e o critério de Furnival não satisfazem o critério da partição aditiva, mas são muito úteis nessa classe de modelo. Espera-se, contudo, que todas essas propriedades sejam satisfeitas por um critério que se proponha como *geral* e *universal*.

9.4 Critério de Informação de Akaike

O Critério de Informação de Akaike (*Akaike Information Criterion* — AIC) é o critério geral que satisfaz as propriedades esperadas de um critério ideal de seleção de modelos. Ele possui a vantagem adicional de ser um critério que minimiza a discrepância de estimação, tornando a discrepância global mais dependente da discrepância de aproximação.

Para se entender a generalidade e a universalidade do AIC na comparação de modelos estocásticos, é necessário seguir a sua fundamentação teórica a partir do conceito da discrepância de Kullback-Leibler.

Discrepância de Kullback-Leibler

A teoria da informação é considerada um ramo da teoria matemática de probabilidade e da estatística matemática que se desenvolveu a partir do conceito técnico de informação, tendo sido aplicado com sucesso em várias áreas do conhecimento científico (Kullback, 1959). Um dos primeiros trabalhos que usa o termo “informação” num sentido matemático foi o trabalho de Fisher (1925b), que definiu informação como a acurácia intrínseca de uma estatística. Shannon (1948) desenvolveu uma medida logarítmica de informação para aplicação no estudo matemático da comunicação. A medida proposta por Shannon foi aplicada a uma grande variedade de áreas de conhecimento, incluindo a Ecologia, na qual se popularizou como índice de diversidade de Shannon.

Kullback e Leibler utilizaram a medida de Shannon na aplicação da teoria da informação no estudo de métodos estatísticos, visando o aprimoramento das técnicas de discriminação de populações e hipóteses (Kullback e Leibler, 1951; Kullback, 1959). Mas os resultados alcançados por estes autores foram de carácter eminentemente teórico. Mais recentemente, a abordagem da teoria de informação se popularizou nas aplicações estatísticas, como demonstraram Burnham e Anderson (2002).

Seguindo Shannon (1948), Kullback e Leibler (Kullback e Leibler, 1951; Kullback, 1959) definiram o logaritmo da razão entre duas funções de densida-

des como a informação contida nos dados para discriminar duas populações ou hipóteses. Ou seja

$$\log \frac{f_1(x)}{f_2(x)}$$

é a informação contida em x para discriminar duas populações ou hipóteses representadas pelas funções de densidade $f_1(\cdot)$ e $f_2(\cdot)$. A discrepância³ entre essas densidades seria dada pela “informação média” relativa à observação de um evento pertencente ao espaço amostral ($E \in \Omega$):

$$D(1, 2) = \int_E \log \frac{f_1(x)}{f_2(x)} f_1(x) dx .$$

Note que a expressão “informação média” tem o sentido da esperança matemática da informação com base na densidade $f_1(\cdot)$. Consequentemente, a discrepância da densidade 1 para a 2 ($D(1, 2)$) é distinta da discrepância de densidade 2 para a 1 ($D(2, 1)$), que é

$$D(2, 1) = \int_E \log \frac{f_1(x)}{f_2(x)} f_2(x) dx .$$

Por isto, trata-se de uma discrepância não de uma “distância”. Rao propôs em 1945 uma medida de distância entre distribuições estocásticas, mas essa medida só se aplica a duas distribuições da mesma família de distribuições e, portanto, só se aplica para distinção de populações e amostras no contexto de um mesmo modelo estocástico (Atkinson e Mitchell, 1981).

A discrepância proposta por Kullback e Leibler é chamada de *discrepância de Kullback-Leibler* em homenagem aos seus autores, que demonstraram que ela tem propriedades bastante convenientes (Kullback e Leibler, 1951; Kullback, 1959). Primeiramente ela é sempre não negativa, isto é, $D(1, 2) \geq 0$ com igualdade somente se $f_1(x) = f_2(x)$. Em segundo lugar, ela é aditiva para eventos aleatórios independentes. Seja x e y dois eventos independentes, então

$$D_{xy}(1, 2) = D_x(1, 2) + D_y(1, 2)$$

Por fim, ela é invariante para uma ampla classe de transformações do espaço de probabilidade (Kullback e Leibler, 1951).

Contudo, a discrepância de Kullback-Leibler é essencialmente uma medida teórica, uma vez que assume o completo conhecimento das densidades das populações ou hipóteses sendo comparadas. As aplicações apresentadas por Kullback (1959) são essencialmente em questões teóricas de estatística matemática.

³Na literatura estatística e matemática, as palavras *discrepância*, e *divergência* são geralmente utilizadas como sinônimos no caso da discriminação de densidades probabilística, enquanto que a palavra *distância* é reservada às discrepâncias que tem uma propriedade específica. Assim, a discrepância Kullback-Leibler é também chamada de divergência Kullback-Leibler.

Discrepância de Kullback-Leibler e Discrepância de Aproximação. A discrepância de Kullback-Leibler é uma medida da discrepância entre duas distribuições estocásticas e, por isto, ela se oferece como uma medida natural para a discrepância de aproximação. Considere que uma das densidades é o modelo operante ($\varphi(\cdot)$), enquanto que a outra se refere a uma dada família de aproximação ($f(\cdot)$). Neste caso, a discrepância de Kullback-Leibler se torna

$$\begin{aligned} \int_E \log \frac{\varphi(x)}{f(x)} \varphi(x) dx &= \int_E \log[\varphi(x)] \varphi(x) dx - \int_E \log[f(x)] \varphi(x) dx \\ &= C - \int_E \log[f(x)] \varphi(x) dx. \end{aligned}$$

A discrepância se desdobra na diferença entre duas integrais. A primeira integral depende unicamente do modelo operante, que depende exclusivamente do cenário estocástico em estudo, sendo, portanto, constante para todas as famílias de aproximação candidatas.

Essa expressão mostra que, para um dado cenário estocástico, a discrepância de Kullback-Leibler indicará como melhor família de aproximação aquela que maximiza a esperança matemática do logaritmo de sua densidade, com respeito a um dado evento observado (E).

Definição 9.5 (Discrepância Relativa de Aproximação). *Dado um certo cenário estocástico, com modelo operante de densidade $\varphi(\cdot)$, a discrepância de aproximação para uma dada família de aproximação com densidade $f(\cdot)$ é dada pela discrepância de Kullback-Leibler:*

$$\begin{aligned} \int_E \log \frac{\varphi(x)}{f(x)} \varphi(x) dx &= \int_E \log[\varphi(x)] \varphi(x) dx - \int_E \log[f(x)] \varphi(x) dx \\ &= C - \int_E \log[f(x)] \varphi(x) dx, \end{aligned}$$

sendo E um evento observado. A discrepância relativa de aproximação é dada pela esperança matemática da densidade da família de aproximação com respeito ao evento observado:

$$- \int_E \log[f(x)] \varphi(x) dx.$$

Essa definição de discrepância relativa permanece inteiramente conceitual, pois para calculá-la é necessário o conhecimento do modelo operante (φ). Para que ela possa ser utilizada em situações práticas é necessário que um modo de contornar a necessidade do conhecimento do modelo operante seja encontrada. Para

isto, é importante notar que essa expressão matemática corresponde ao conceito da *esperança matemática* ou *valor esperado* do logaritmo da função de densidade do modelo de aproximação condicionada ao evento observado E

$$- \int_E \log[f(x)] \varphi(x) dx = -\mathbf{E}_\varphi \{ \log[f(x)] | E \} ,$$

sendo a esperança matemática definida em relação ao modelo operante (φ).

Discrepância de Kullback-Leibler e Verossimilhança

Hirotoyu Akaike foi o primeiro a associar a discrepância de Kullback-Leibler com o princípio da *máxima verossimilhança* (Akaike, 1981). Num primeiro trabalho em 1973, ele ligou a abordagem da teoria da informação com o princípio da máxima verossimilhança de estimação de modelos estocásticos (Akaike, 1973a). Numa sequência de trabalhos posteriores (Akaike, 1973b; Akaike, 1974; Akaike, 1985) ele desenvolveu mais precisamente o seu critério de informação. Inicialmente, o critério de informação foi utilizado para problemas tradicionais de análise estatística como análise de fatores (Akaike, 1973a; Akaike, 1987), componentes principais (Akaike, 1973a), modelos de séries temporais (Akaike, 1973b; Akaike, 1974), análise de variância e regressão (Akaike, 1973a; Sakamoto, Ishiguro e Kitagawa, 1986) e tabelas de contingência (Sakamoto, Ishiguro e Kitagawa, 1986). Mas o critério foi colocado essencialmente como um índice da qualidade de ajuste para o problema de especificação de modelos (Akaike, 1974).

Burnham e Anderson (2002) apresentam duas deduções detalhadas do Critério de Informação de Akaike. Ambas deduções se assentam no contexto de amostragens repetidas indefinidamente de uma população infinita, que é o princípio fundamental da inferência clássica. Nenhuma referência é feita ao Princípio de Verossimilhança e, aparentemente, o desenvolvimento do próprio Akaike se baseia nessa abordagem⁴. Curiosamente, Burnham e Anderson (2002) fazem menção a um “resultado intrigante” (*puzzling result*) que permite uma dedução mais sucinta. Embora o resultado seja matematicamente válido, eles não o consideram conceitualmente válido. Entretanto, essa consideração parece resultar do fato destes autores utilizarem o princípio da inferência clássica e não se apoiarem no Princípio de Verossimilhança. Segue-se uma apresentação do critério de informação de Akaike seguindo este “resultado intrigante” e evocando o Princípio de Verossimilhança.

Critério de Informação de Akaike e Princípio de Verossimilhança. Inicialmente é necessário considerar que para uma discrepância ser calculada é neces-

⁴É importante distinguir o Princípio de Verossimilhança do princípio de máxima verossimilhança utilizado por Akaike.

sário trabalhar não com uma família de aproximação, mas com um modelo de aproximação específico. Logo é necessário considerar a existência dos parâmetros do modelo. Neste caso, a discrepância de Kullback-Leibler se torna:

$$\begin{aligned} \int_E \log \frac{\varphi(x)}{f(x|\boldsymbol{\theta}_0)} \varphi(x) dx &= \int_E \log[\varphi(x)] \varphi(x) dx - \int_E \log[f(x|\boldsymbol{\theta}_0)] \varphi(x) dx \\ &= C - \int_E \log[f(x|\boldsymbol{\theta}_0)] \varphi(x) dx, \end{aligned}$$

sendo $\boldsymbol{\theta}_0$ o “valor verdadeiro” dos parâmetros do modelo de aproximação candidato $f(\cdot)$. A discrepância relativa deste modelo de aproximação é

$$- \int_E \log[f(x|\boldsymbol{\theta}_0)] \varphi(x) dx = -\mathbf{E}_x \{ \log[f(x|\boldsymbol{\theta}_0)] \} .$$

Note-se que a integração se faz em relação ao evento observado (E) que pertence ao espaço amostral do cenário estocástico ($E \in \Omega$). Como o modelo operante ($\varphi(\cdot)$) não é conhecido, a melhor ação possível é buscar a esperança matemática em relação aos dados observados ($\mathbf{E}_x \{ \cdot \}$), sem necessidade de se conhecer o valor dos parâmetros ($\boldsymbol{\theta}_0$).

Para que a discrepância do modelo de aproximação candidato possa ser determinada numericamente é necessário estimar os parâmetros. Essa discrepância relativa pode, então, ser desenvolvida como uma expansão da série de Taylor em relação ao valor das estimativas ($\hat{\boldsymbol{\theta}}$)

$$-\mathbf{E}_x \{ \log[f(x|\boldsymbol{\theta}_0)] \} = - \left[\mathbf{E}_x \left\{ \log[f(x|\hat{\boldsymbol{\theta}})] \right\} - w(\boldsymbol{\theta}_0) \right] ,$$

sendo $w(\boldsymbol{\theta}_0)$ uma função do valor verdadeiro dos parâmetros, que independe de dados (Burnham e Anderson, 2002).

O processo de estimação implica que à discrepância de aproximação, representada pela discrepância relativa original, é adicionada a discrepância de estimação. Portanto, a medida apresentada acima, resultante da expansão da série de Taylor, é uma discrepância global na qual os componentes de aproximação e de estimação são indistinguíveis. Para que essa medida seja um bom critério de comparação da qualidade da aproximação dos modelos candidatos é necessário que o componente de estimação seja minimizado.

Os modelos de aproximação candidatos serão comparados com base num conjunto de dados, isto é, a comparação da performance dos modelos será *condicional* às observações obtidas. A minimização da discrepância de estimação implica na maximização da esperança matemática do logaritmo da densidade do modelo candidato, que é a própria função de log-verossimilhança. Logo, as estimativas que

minimizam a discrepância de estimação são as estimativas de máxima verossimilhança (MLE) e a esperança matemática da medida de discrepância é o valor da máxima verossimilhança do modelo candidato. O Princípio de Verossimilhança assegura que *toda informação* contida nos dados a respeito de um modelo está contida na função de log-verossimilhança e, portanto, ele nos assegura que ao maximizarmos a função de log-verossimilhança estamos maximizando a informação contida nos dados a respeito da qualidade da aproximação do modelo candidato. Assim, a medida de discrepância relativa se torna

$$-\mathbf{E}_x \{ \log[f(x|\boldsymbol{\theta}_0)] \} = - \left[\mathbf{L}_f \{ \hat{\boldsymbol{\theta}} \} - w(\boldsymbol{\theta}_0) \right] .$$

onde $\mathbf{L}_f \{ \hat{\boldsymbol{\theta}} \}$ é a máxima log-verossimilhança do modelo candidato (f) dado um conjunto de observações e $\hat{\boldsymbol{\theta}}$ são as MLEs.

Por questões de conveniência, Akaike sugeriu que a função $w(\boldsymbol{\theta}_0)$ fosse tomada como o número de parâmetros livres do modelo (K) e, por “motivos históricos”, a expressão toda foi multiplicada por dois (Akaike, 1985), resultando no Critério de Informação de Akaike (AIC).

Definição 9.6 (Critério de Informação de Akaike). *O Critério de Informação de Akaike (AIC) é uma medida da discrepância global entre o modelo operante e um dado modelo de aproximação, sendo obtido pela expressão*

$$AIC = -2 \mathbf{L}_f \{ \hat{\boldsymbol{\theta}} \} + 2 K ,$$

onde $\hat{\boldsymbol{\theta}}$ são as estimativas de máxima verossimilhança do modelo (MLE) referentes ao valor máximo da função de log-verossimilhança ($\mathbf{L}_f \{ \hat{\boldsymbol{\theta}} \}$) e K o número de parâmetros livres do modelo.

O AIC é uma medida de discrepância relativa entre o modelo operante e o modelo de aproximação. Condicionalmente, a um conjunto de observações, o modelo candidato que melhor aproxima o modelo operante é aquele de **menor** AIC.

Aos dois componentes do AIC pode se atribuir interpretações ligeiramente diferentes. O valor negativo da máxima log-verossimilhança ($-\mathbf{L}_f \{ \hat{\boldsymbol{\theta}} \}$) deve ser interpretada como a discrepância relativa entre o modelo operante e o modelo candidato. Essa discrepância relativa é uma medida da discrepância global, em que estão fundidas a discrepância de aproximação (da família do modelo candidato) e a discrepância de estimação (do modelo candidato ajustado). Mas, utilizando-se as estimativas de máxima verossimilhança para os parâmetros (MLE) a discrepância de estimação é minimizada. Já o termo que envolve o número de parâmetros (K) deve ser interpretado como uma penalização à complexidade da *família do*

modelo candidato, uma vez que essa penalização independe do ajuste do modelo (Forster e Sober, 1994).

Outro aspecto relevante do AIC é a relação entre os dois componentes. À medida que o tamanho do conjunto de dados aumenta, a importância relativa do termo que penaliza a complexidade do modelo se torna menos importante. Para grandes amostras, o valor em módulo do máximo da log-verossimilhança estará numa escala muito maior que o número de parâmetros, tornando o componente de penalização da complexidade do modelo irrelevante. Já nas situações de amostras pequenas, a diferença do número de parâmetros nas famílias candidatas pode ser um fator decisivo em favor das famílias de modelos mais simples (Forster e Sober, 1994). A complexidade das famílias de aproximação candidatas também se torna relevante quando os modelos têm desempenho muito próximo. Nestes casos, o AIC favorecerá a família mais simples em prejuízo da mais complexa, seguindo o critério da parcimônia.

Variantes do AIC

Uma certa controvérsia envolve o termo K na expressão do AIC. Alguns autores (Sakamoto, Ishiguro e Kitagawa, 1986; Kuha, 2004) afirmam que ele implica na premissa de que o modelo operante esteja dentre os modelos candidatos. Outros autores (Burnham e Anderson, 2002) negam que essa premissa seja necessária para dedução do AIC. Parece existir concordância que para modelos candidatos bem próximos ao modelo operante o termo K é uma boa aproximação para função $w(\theta_0)$. No caso de modelos que não são uma boa aproximação, uma forma alternativa de cálculo do critério foi apresentada por em 1976 Takeushi (Burnham e Anderson, 2002)

$$\text{TIC} = -2 \mathbf{L}_f\{\hat{\boldsymbol{\theta}}\} + 2 \text{tr}[\mathbf{J}(\boldsymbol{\theta}_0) \mathbf{I}(\boldsymbol{\theta}_0)^{-1}],$$

onde $\text{tr}[\cdot]$ é o traço do produto da matriz \mathbf{J} e da inversa de \mathbf{I} . Ambas matrizes são dependentes do valor verdadeiro dos parâmetros ($\boldsymbol{\theta}_0$), mas não dos dados. Em princípio o Critério de Informação de Takeushi (TIC) seria mais geral que o AIC, mas sua grande limitação é a estimação das matrizes \mathbf{J} e \mathbf{I} . A estimação dessas matrizes pode trazer grande instabilidade ao processo de seleção dos modelos (Burnham e Anderson, 2002), o que torna o ganho de generalidade fornecido pelo TIC duvidoso.

O critério de uso mais amplo tem sido o AIC, pois, nas situações em que os modelos candidatos são boas aproximações, o uso do número de parâmetros (K) é apropriado.

A expressão matemática do AIC é geralmente interpretada como contendo uma medida relativa da capacidade preditiva ou explicativa do modelo, que é o valor da máxima log-verossimilhança, e um peso para onerar os modelos com

muitos parâmetros, que é o termo envolvendo K . Entre dois modelos com igual poder preditivo, isto é, mesmo valor da máxima log-verossimilhança, o AIC indicaria como melhor o modelo mais simples, seguindo o princípio da parcimônia. A partir dessa interpretação, surgiram propostas de outras formas de penalizar os modelos em função da sua complexidade, como por exemplo os critérios CAIC e CAICF (Bozdogan, 1987), ou ICOMP (Bozdogan, 2000).

Uma variante do AIC, que encontrou grande aplicação prática, é aquela que apresenta uma correção no caso de pequenas amostras. Na perspectiva de inferência clássica, o AIC para pequenas amostras possui um vício ou viés que favorece os modelos com muitos parâmetros. Logo, para uma comparação mais equilibrada entre os modelos se faria necessária uma penalização em função da relação entre o número de parâmetros (K) e o tamanho da amostra (n). Burnham e Anderson (2004) afirmam que para amostras em que a razão entre o tamanho de amostra e o número de parâmetros do modelo for menor que 40 ($n/K < 40$) a utilização do AIC com correção para pequenas amostras (AICc) é necessária.

Definição 9.7 (Correção do AIC). *O AIC com correção para pequenas amostras — AICc — é calculado pela expressão*

$$AICc = -2 \mathbf{L}_f\{\hat{\boldsymbol{\theta}}\} + 2K + \frac{2K(K+1)}{n-K-1},$$

na qual $\mathbf{L}_f\{\hat{\boldsymbol{\theta}}\}$ é o máximo da função de log-verossimilhança, n é o tamanho da amostra e K o número de parâmetros do modelo.

Outra variante do AIC, proposta por Schwarz (1978), é o BIC (*Bayesian Information Criterion*) que é obtido pela expressão

$$BIC = -2 \mathbf{L}_f\{\hat{\boldsymbol{\theta}}\} + K \log(n).$$

O elemento distinto do BIC é que a penalização dos modelos com maior número de parâmetros (K) cresce com o logaritmo natural do tamanho da amostra (n). A expressão “critério de informação” neste caso é inadequada, uma vez que Schwarz não utilizou a teoria da informação para construir essa medida, mas a construiu num contexto de inferência Bayesiana (Burnham e Anderson, 2004). Mas o BIC não deve ser considerado uma forma completa de comparação Bayesiana de modelos, pois ele não inclui uma determinação explícita e criteriosa das distribuições *a priori* dos modelos candidatos (Kuha, 2004). Contudo, ele assume implicitamente que a distribuição *a priori* dos parâmetros é uma distribuição multivariada Gaussiana (multivariada normal) (Ward, 2008).

Embora o BIC tenha sido construído num contexto Bayesiano, a sua utilização em substituição ao AIC não é necessariamente uma questão de escolher entre inferência Bayesiana e clássica (Burnham e Anderson, 2004). O argumento utilizado

em favor do BIC contra AIC é que o BIC consegue detectar o *modelo verdadeiro* com probabilidade tendendo a um à medida que o tamanho da amostra cresce (Claeskens e Hjort, 2008).

Contudo, a necessidade do *modelo verdadeiro* estar presente no conjunto de modelos testados não é um requisitos essencial para a dedução do BIC (Burnham e Anderson, 2002). Por outro lado, a convergência em probabilidade do BIC para um *modelo alvo* não implica logicamente que exista um *modelo verdadeiro* ou que o modelo alvo seja o modelo verdadeiro (Burnham e Anderson, 2004). Schwarz (1978) deduziu o BIC no contexto da estimação da dimensão do modelo, o que corresponde a situações como a definição da ordem do modelo numa regressão polinomial ou em modelos de séries temporais, ou a definição das variáveis preditoras numa situação de regressão linear com um grande número de variáveis preditoras candidatas (*stepwise regression*). Mais ainda, uma das premissas nas quais Schwarz (1978) se apoia é que os dados foram gerados por um modelo da família Koopman-Darmois, cuja densidade tem a forma

$$f(x, \theta) = \exp(\theta y(x) - b(\theta)) ,$$

na qual θ é o vetor de parâmetros do qual se deseja estimar a dimensão apropriada, $y(\cdot)$ é uma função vetorial que depende unicamente dos dados e $b(\cdot)$ é uma função única dos parâmetros. Portanto, o contexto de desenvolvimento do BIC é bem mais restrito e particular que o contexto geral de seleção de modelos.

Embora o BIC possa ser utilizado fora de um contexto Bayesiano de comparação de modelos (Burnham e Anderson, 2002), o objetivo do contexto Bayesiano em que ele foi deduzido é a identificação do modelo com a maior probabilidade de ser o modelo verdadeiro (Kuha, 2004). Já o desenvolvimento do AIC, nega a necessidade de existência de um modelo verdadeiro e utiliza a predição esperada de dados futuros como critério de comparação dos modelos (Kuha, 2004), isto é, o AIC busca identificar o modelo com a maior *acurácia preditiva estimada* (*estimated predictive accuracy*) (Forster e Sober, 1994).

Num estudo empírico baseado em simulação de modelos de abundância ecológica, Ward (2008) comparou a performance do AIC, AICc e BIC, além de outras medidas de seleção de modelos. Ele concluiu que o BIC se comportou de modo marcadamente parecido ao AICc, tendendo a selecionar modelos de menor complexidade. Ele atribuiu este resultado ao fato de que os tamanhos de amostra simulados foram relativamente pequenos.

Delta de AIC e Pesos de Akaike

Burnham e Anderson propuseram em vários trabalhos (Burnham e Anderson, 2001; Burnham e Anderson, 2002; Burnham e Anderson, 2004) o re-escalamento do AIC (ou AICc) para que o modelo com o menor AIC tivesse um valor

nulo como medida de performance. Num conjunto com R modelos, o valor reescalado de AIC (AICc) é

$$\Delta_i = \text{AIC}_i - \min_{j=1}^R (\text{AIC}_j) ,$$

sendo i ($i = 1, 2, \dots, R$) o índice do modelo. A organização dos modelos candidatos na ordem crescente de Δ_i resulta num “ranqueamento” dos modelos segundo a qualidade da sua performance, dos melhores para os piores, sendo uma forma de interpretação direta e fácil do AIC.

Utilizando a mesma regra canônica apresentada no capítulo de estimação, tem-se a regra de seleção de modelos na qual aqueles modelos com $\Delta_i > \log(8)$ são considerados marcadamente inferiores ao melhor modelo, enquanto que os modelos com $\Delta_i \leq \log(8)$ são igualmente plausíveis ao melhor modelo.

A partir dessa grandeza, esses autores sugerem que a “verossimilhança do modelo dadas as observações” é proporcional a

$$\mathcal{L}\{f_i|\mathbf{x}\} \propto \exp(-\Delta_i/2) ,$$

sendo uma medida da “força de evidência” em favor do modelo i . Essas expressões são, contudo, equivocadas e geram uma certa confusão de conceitos. O exponencial de delta é na verdade proporcional à *razão de verossimilhança* entre o modelo i e o melhor modelo, entendido como o modelo de menor AIC, condicionalmente aos dados observados:

$$\exp(-\Delta_i/2) \propto \frac{\mathcal{L}\{f_i|\mathbf{x}\}}{\mathcal{L}\{\text{melhor } f|\mathbf{x}\}} .$$

Burnham e Anderson (Burnham e Anderson, 2001; Burnham e Anderson, 2002; Burnham e Anderson, 2004) prosseguem e sugerem a normalização dos Δ_i , gerando os “pesos de Akaike” (*Akaike weights*):

$$w_i = \frac{\exp(-\Delta_i/2)}{\sum_{j=1}^R \exp(-\Delta_j/2)} ,$$

cujo total é unitário ($\sum w_i = 1$). Os pesos de Akaike são, logicamente, dependentes do conjunto dos modelos candidatos testados e são interpretados, segundo estes autores, como “o peso de evidência em favor do modelo i ser o verdadeiro melhor modelo K-L [Kullback-Leibler] para situação em estudo dado que um dos R modelos deve ser o melhor K-L modelo do conjunto de R modelos” (Burnham e Anderson, 2002, p.75).

A utilização do Δ_i é uma mera questão de conveniência, pois eles não trazem nenhuma informação que já não esteja no conjunto dos valores de AIC a partir do qual foram calculados. Já na expressão exponencial dos Δ_i , uma interpretação

equivocada é introduzida. Quando o melhor modelo é comparado com ele mesmo obtém-se $\Delta_i = 0$ e $\exp(-\Delta_i/2) = 1$, o que implica em

$$\exp(-\Delta_i/2) \leq 1.$$

Logo, não faz sentido afirmar que essa expressão exponencial representa a força de evidência relativa de cada modelo i . Essa expressão é, na verdade, proporcional ao *inverso* da força de evidência *em favor* do melhor modelo vis-à-vis do modelo i :

$$\exp(-\Delta_i/2) \propto \left[\frac{\mathcal{L}\{\text{melhor } f|\mathbf{x}\}}{\mathcal{L}\{f_i|\mathbf{x}\}} \right]^{-1}.$$

A normalização dos Δ_i gerando os pesos de Akaike não altera essa situação de interpretação equivocada. Os pesos de Akaike não podem ser um “peso de evidência em favor do modelo i ”, pois quando todos modelos são comparados com o melhor modelo, os dados favorecem sempre o melhor modelo sendo, portanto, evidência contra os demais. Talvez seja mais apropriado interpretar os pesos de Akaike como uma medida de *relevância prática* dos modelos quando comparados ao melhor modelo. O melhor modelo tem a maior relevância prática, sendo que os demais tem relevância progressivamente menor à medida que se desce no ordenamento produzido pelos Δ_i . É importante lembrar que os pesos de Akaike não estão condicionados apenas aos dados observados, mas também ao melhor modelo do conjunto de modelos testados, que é um conjunto arbitrariamente definido.

Inferência por Multimodelos

A seleção de modelos em aplicações práticas resulta frequentemente em situações em que dois ou mais modelos se mostram igualmente plausíveis. Nessas situações a incerteza a respeito do melhor modelo para predição ou estimação se torna máxima, pois, embora possa haver outras razões não estatísticas para se optar por um dos modelos, um certo grau de arbitrariedade estará presente na escolha final do melhor modelo. Um outro aspecto, levantado por alguns autores é que, mesmo com a utilização de critérios estatísticos para seleção de modelos, a escolha de um único modelo pode resultar na prática num “vício de especificação equivocada” (*misspecification bias*) na predição ou estimação, caso o modelo selecionado não seja o modelo “verdadeiro” (Buckland, Burnham e Augustin, 1997; Burnham e Anderson, 2004).

A proposta para contornar estes problemas é a combinação de vários modelos na realização da predição ou estimação, o que é normalmente chamado de “inferência por multimodelos” (*multimodel inference*) ou “média de modelos” (*model*

averaging). O Critério de Informação de Akaike é utilizado então para estabelecer o peso da predição ou estimação gerada por cada modelo candidato e uma média ponderada é calculada resultando na predição ou estimação final (Buckland, Burnham e Augustin, 1997; Burnham e Anderson, 2001; Burnham e Anderson, 2002; Burnham e Anderson, 2004). A inferência por multimodelos tornaria o vício de seleção de modelo em um componente da variância da predição ou estimação, incorporando assim, a incerteza de seleção de modelos no processo de predição e estimação (Buckland, Burnham e Augustin, 1997; Burnham e Anderson, 2004).

Embora a possibilidade de combinação da predição ou estimação gerada por dois ou mais modelos que se mostraram igualmente plausíveis possa ser de utilidade em situações de aplicação prática, os princípios que fundamentam a inferência por multimodelos são problemáticos. A concepção de que a seleção de um único “melhor modelo” pode resultar num “vício de especificação equivocada” só faz sentido quando o processo de seleção de modelos é visto única e absolutamente na perspectiva da inferência clássica. Se o Princípio de Verossimilhança for considerado, a ideia de “vício de especificação equivocada” de modelo é totalmente abstrata sem qualquer validade prática, uma vez que ela implica que algo, além dos dados observados, pode nos informar a respeito do comportamento dos modelos sendo estudados. Este “algo” é fundamentalmente uma concepção hipotética sobre o comportamento do modelo sob condições de uma população infinita re-amostrada repetidamente, que é o fundamento de toda inferência clássica.

Os proponentes da inferência por multimodelos parecem não perceber que a implicação lógica da ideia de “vício de especificação equivocada” é a completa invalidação de *qualquer* procedimento de seleção de modelos. Para se evitar este vício, a seleção de modelos deveria ser substituída pelo procedimento de inferência por multimodelos que, uma vez automatizado, tornaria dispensável qualquer julgamento humano no processo de predição ou estimação. Por outro lado, se o “vício de especificação equivocada” for considerado seriamente é necessário justificar porque a “especificação equivocada” influencia somente o “melhor” modelo e não o conjunto de modelos testados. Se o “melhor” modelo foi especificado equivocadamente, como garantir que os demais modelos testados também não o foram? Como garantir que o próprio *conjunto* de modelos testados não foi especificado equivocadamente?

Outra grande limitação da inferência por multimodelos é que a incerteza só pode ser incorporada à variância da predição ou estimação caso se aceite a premissa de que os modelos candidatos estudados representam uma *amostra aleatória de todos os modelos aplicáveis* à situação estudada (Buckland, Burnham e Augustin, 1997). É difícil imaginar qualquer situação de aplicação prática em que essa premissa possa ser considerada razoável ou mesmo desejável.

Do ponto de vista empírico, Richards (Richards, 2005) levantou limitações sobre a capacidade da inferência por multimodelos ser de fato superior a aborda-

gem tradicional do melhor modelo. Em dois estudos de simulação baseados em modelos mecanístico do processo de foragem de pólen por abelhas ele verificou que enquanto em um dos estudos a abordagem multimodelos produzia melhores estimativas, no outro ela produzia sempre estimativas piores ao melhor modelo. Ele concluiu que a superioridade da abordagem multimodelos ainda é incerta.

Critério de Informação de Akaike e Verossimilhança

O Critério de Informação de Akaike (AIC) é uma forma empírica de se computar uma medida de discrepância relativa entre os modelos de aproximação e o modelo operante (discrepância de Kullback-Leibler) que é calculado a partir da função de log-verossimilhança. Dessa forma, ele está fundamentado na Lei e no Princípio de Verossimilhança e, portanto, reforça a concepção de que toda comparação de modelos deve estar condicionada a um conjunto de observações. Por outro lado, por se fundamentar na verossimilhança, o AIC tem as propriedades desejadas num critério de seleção de modelos: invariância à parametrização e aditividade para observações independentes. Assim, o AIC estabelece uma relação direta entre a abordagem da teoria da informação aplicada à inferência estatística e o paradigma da inferência estatística baseada na verossimilhança.

Mas o AIC também pode ser entendido como uma medida da capacidade preditiva dos modelos ligada ao erro de predição em condições de validação cruzada (Akaike, 1985; Davies, Neath e Cavanaugh, 2005). Mas a validação é em si um aspecto muito importante do processo de especificação e seleção de modelos na Biometria Florestal.



Exemplo 9-7: Estrutura de Tamanho de Floresta Tropical

Nos exemplos 8.5 e 8.5 dados de diâmetro de uma floresta tropical no município de Bom Jardim, MA, foram utilizados para exemplificar o ajuste da distribuição Weibull pela abordagem da verossimilhança. Cabe questionar se a distribuição Weibull seria de fato o modelo estocástico mais apropriado para modelar a estrutura de tamanho dessa floresta tropical. A distribuição gama é uma distribuição igualmente flexível para representar variáveis não-negativas. Por outro lado, a Lei de De Liocourt sugere que a distribuição exponencial seria um modelo apropriado para qualquer floresta que tenha uma estrutura de tamanho “balanceada” (veja exemplo 7.4). Mas a distribuição exponencial pode ser tomada como um caso particular das distribuições Weibull e Gama quando o parâmetro da forma (γ) é unitário. Assim, essas três famílias de distribuição podem ser consideradas como *modelos de aproximação* candidatos para representar os dados de DAP dessa floresta tropical, sendo que uma família, a exponencial, pode ser considerada como

um caso particular das outras duas.

Numa primeira abordagem ao problema, pode se buscar qual o melhor modelo para representar a estrutura de tamanho no nível das parcelas individuais do levantamento da floresta. A modelagem no nível das parcelas traz como premissa a ideia de que a estrutura de tamanho da floresta não é homogênea, mas varia de local para local. Essa abordagem também tem a premissa de que as árvores observadas em cada parcela são uma amostra representativa da estrutura de tamanho local. Um estudo mais aprofundado nessa abordagem exigiria uma reflexão sobre o tamanho de parcela adequado para representar a estrutura de tamanho local e sobre o arranjo espacial das parcelas na floresta para representar a variação espacial da estrutura de tamanho. Tais reflexões, embora importantes, são raramente realizadas nos levantamentos florestais

Nos dados da floresta nativa de Bom Jardim, o levantamento foi realizado por meio de 87 parcelas com tamanhos de 1250, 2500 e 5000 m^2 , logo o número de árvores em cada parcela é variável, não só devido à variação de densidade da floresta, mas também devido ao tamanho das parcelas de inventário. Os três modelos foram ajustados a cada parcela, sendo o AIC também computado para cada parcela individualmente. A figura 9.1 apresenta as diferenças do AIC obtidos no ajuste dos três modelos (parcela-a-parcela) em função do número de árvores e do DAP médio. A distribuição exponencial se mostrou claramente inferior às outras duas distribuições, sendo que o seu desempenho não tem relação com o número de árvores na parcela (figuras 9.1 a-b), mas parece ligeiramente dependente do DAP médio (figuras 9.1 d-e). Essa figura mostra, no entanto, que em muitas parcelas a distribuição exponencial é um modelo tão bom quanto a Weibull ou gama.

As distribuições Weibull e gama se mostraram muito próximas no seu desempenho, não havendo relação entre a diferença de desempenho destes modelos e o número árvores ou o DAP médio na parcela (figuras 9.1 c,f). A distribuição gama se mostrou superior a Weibull em quatro parcelas, enquanto que o inverso ocorreu em duas parcelas, o que dá uma ligeira vantagem à distribuição gama sobre a Weibull. Na maioria das parcelas, contudo, as duas distribuições se mostram equivalentes.

O desempenho diferenciado entre os modelos nas diferentes parcelas sugere que as estimativas dos parâmetros podem estar associadas a alguns atributos das parcelas. No caso das distribuições Weibull e gama, o parâmetro da forma não ter nenhuma relação com o número de árvores na parcela (figura 9.2 a), mas parece estar inversamente relacionado ao DAP médio (figura 9.2 c). Já o parâmetro de

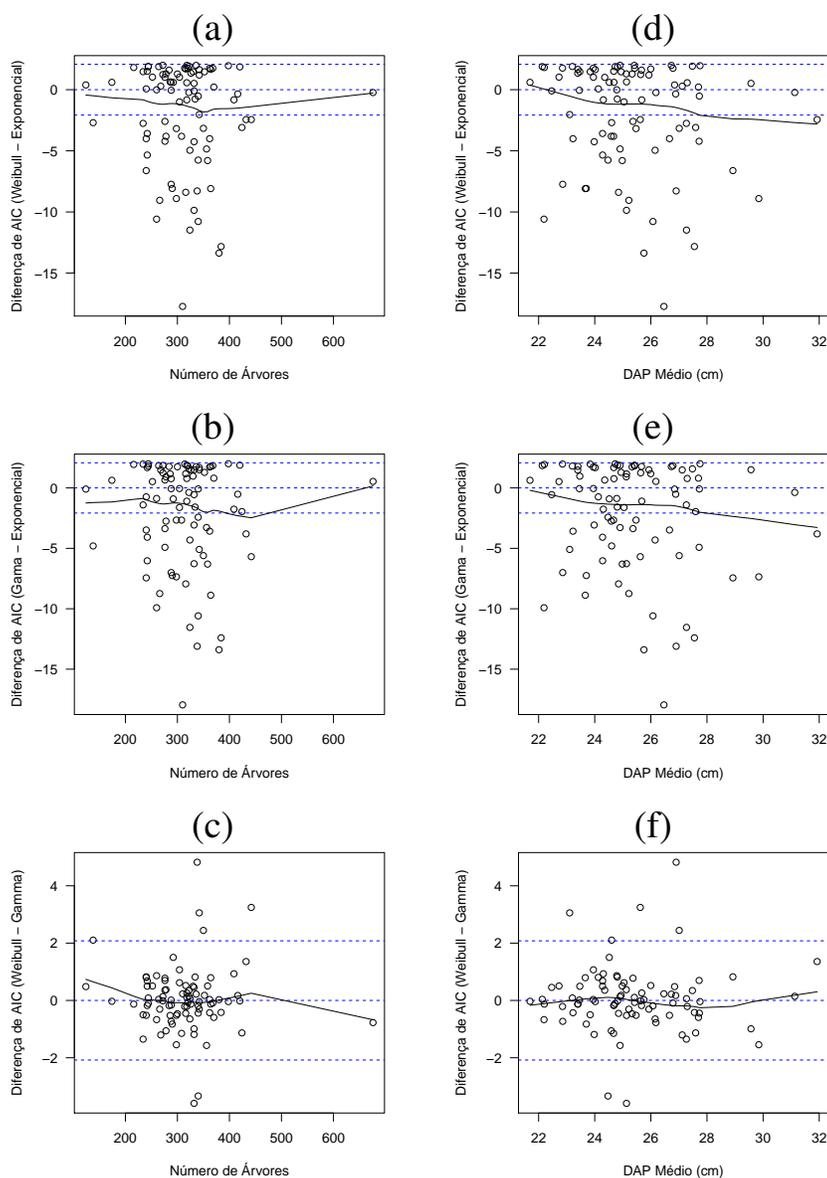


Figura 9.1: Diferenças do AIC grafada contra o número de árvores por parcela (a, b, c) e o DAP médio (d, e, f), para as distribuições Weibull e exponencial (a, d), gama e exponencial (b, e) e Weibull e gama (c, f). As linhas horizontais dos gráficos indicam as posições de diferença de AIC iguais a $\ln(8)$, zero e $-\ln(8)$. As curvas em todos os gráficos são linhas de tendência de relação entre as variáveis encontradas por suavização.

escala, ele também não têm relação com o número de árvores (figura 9.2 b), mas possui uma forte relação com o DAP médio (figura 9.2 d). Essa forte relação é esperada, pois na distribuição exponencial o parâmetro de escala é o inverso da média.

O delineamento amostral do levantamento foi a amostragem estratificada, onde os estratos representaram situações topográficas distintas. Logo, as 87 parcelas do levantamento podem ser agrupadas segundo as condições topográficas para

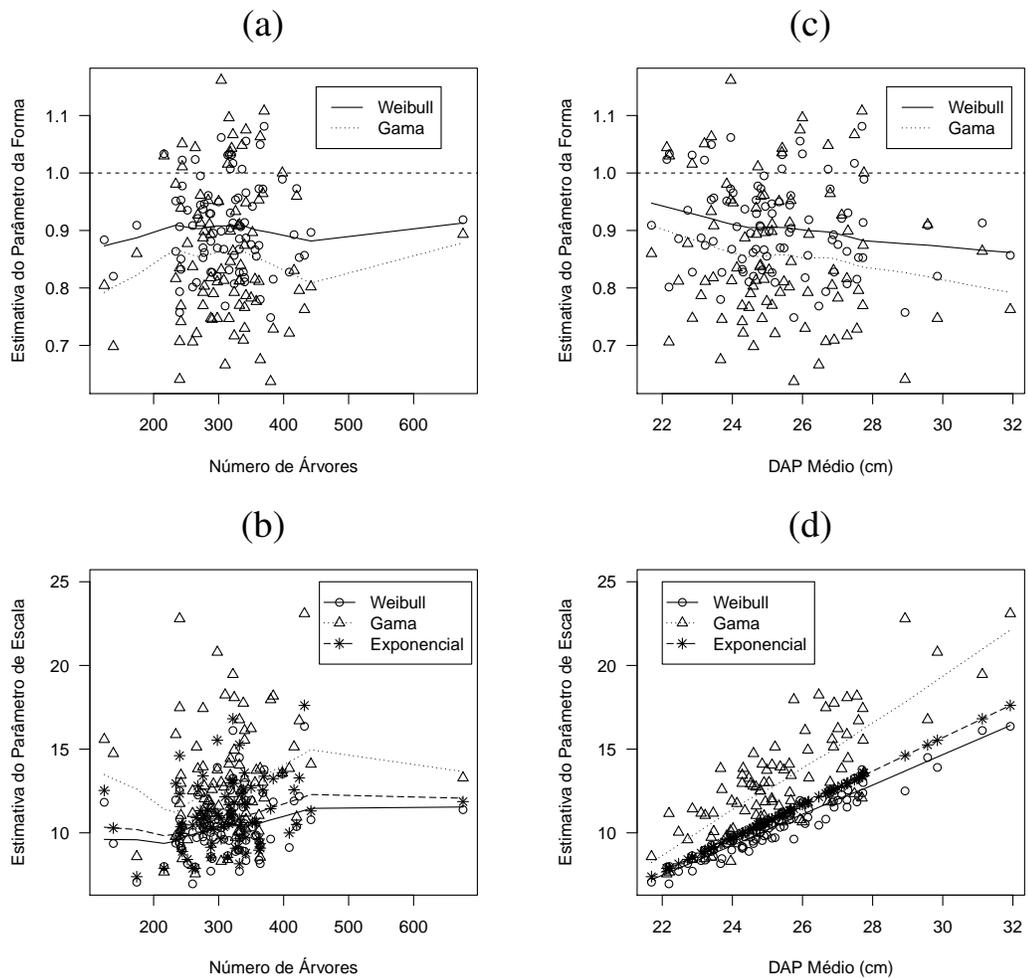


Figura 9.2: Estimativas do parâmetro da forma (a, c) e do parâmetro de escala (b, d) em função do número de árvores na parcela (a, b) e do DAP médio da parcela (c, d). As curvas em todos os gráficos são linhas de tendência de relação entre as variáveis encontradas por suavização.

se investigar se o desempenho dos modelos independe dessa condição. Como o AIC tem propriedade aditiva, os AIC das parcelas podem ser somados para se obter um AIC para cada condição topográfica. Por outro lado, todas as árvores do levantamento podem ser agregadas num único conjunto, para o qual as três distribuições podem ser ajustadas. Surgem então, além da questão do melhor modelo de aproximação, a questão da melhor escala para modelar a estrutura de tamanho da floresta. A tabela 9.1 apresenta os valores de AIC observados para as três distribuições nessas duas escalas de modelagem.

Tabela 9.1: Critério de Informação de Akaike (AIC) para os modelos exponencial (E), Weibull (W) e gama (G), ajustados para o agregado de todas as árvores do levantamento (“Ajuste Geral”) e por parcela (“Ajuste por Parcela”). No ajuste geral, os AIC foram obtidos a partir da soma da log-verossimilhança por situação topográfica e no total, já os AIC do ajuste por parcela foram obtidos somando-se os AIC de cada parcela.

Situação Topográfica	Número de Parc.	Árv.	AIC - Critério de Informação de Akaike					
			Ajuste Geral			Ajuste por Parcela		
			E	W	G	E	W	G
Encosta	22	3360	23055	23037	23042	23015	23006	23009
Platô Grande	34	4728	32309	32197	32197	32313	33234	32227
Platô Pequeno	31	3884	26247	26099	26110	26121	26027	26026
Total	87	11972	81607	81325	81341	81450	81267	81262

E – exponencial; W – Weibull; G – gama

Como esperado, o ajuste das três distribuições se mostra melhor quando a distribuição é ajustada parcela-a-parcela. Como o AIC penaliza o ajuste em função do número de parâmetros, o ajuste parcela-a-parcela se mostra melhor mesmo considerando o número muito maior de parâmetros que essa forma de ajuste implica. Enquanto no ajuste geral, a distribuição exponencial tem um parâmetro e as distribuições Weibull e gama têm dois, no ajuste parcela-a-parcela, o número de parâmetros do modelo exponencial é igual ao número de parcelas (um parâmetro de escala por parcela), enquanto que nos modelos Weibull e gama o número de parâmetros é duas vezes o número de parcelas (um parâmetro de forma e um parâmetro de escala por parcela).

Nota-se que a distribuição gama se mostra ligeiramente superior à Weibull no ajuste parcela-a-parcela (diferença de 5 no AIC total), mas no ajuste geral é a distribuição Weibull que se mostra superior (diferença de 16 no AIC total). A tabela 9.1 também ressalta que o desempenho relativo destes modelos varia de

uma situação topográfica para outra. Este é um bom exemplo de como a seleção de modelos não é independente da estrutura dos dados que se assume para ajustar os modelos candidatos. O resultado da seleção de modelos pode ser muito diferente dependendo da escala de ajuste em que os modelos são comparados.



9.5 Validação Cruzada

A apresentação explícita da ideia da validação cruzada como técnica para verificar a “validade” de uma análise estatística é atribuída à literatura psicométrica da década de 1930 em ligação com estudos de predição (Linhart e Zucchini, 1986; Arlot e Celisse, 2010). Stone (1974) apresenta uma breve história da validação cruzada e a desenvolve na perspectiva da escolha do melhor método para análise estatística, apresentando-a, portanto, no contexto de seleção de modelos.

Na construção dos modelos florestais em geral, e dos modelos empíricos de Biometria Florestal em particular, a expressão “*validação*” tem significado bastante variável, sendo frequentemente empregada para denotar o processo de determinação da “validade” de um dado modelo para uma situação prática particular. Alguns autores, entretanto, sugerem o uso da expressão “*avaliação de modelos florestais*”, uma vez que tecnicamente é duvidoso que se possa determinar a validade de um modelo em termos da sua “veracidade” (Vanclay e Skovsgaard, 1997). Neste trabalho, a expressão *validação cruzada* é utilizada no mesmo contexto de Stone (1974), isto é, como uma forma de *avaliação comparativa* entre modelos de aproximação candidatos a representar um dado cenário estocástico.

A validação cruzada consiste em subdividir a amostra de um estudo, alocando as observações disponíveis em duas subamostras. Numa delas, as observações são utilizadas para construção dos modelos, isto é, para estimação dos parâmetros, sendo essa subamostra chamada de subamostra de “*estimação*”, de “*ajuste*”, ou ainda de “*treinamento*” (Carmack, Spence e Schucany, 2012). As observações da outra subamostra são utilizadas para verificar a validade dos modelos por meio das predições das observações dessa subamostra, com base nos modelos ajustados na subamostra de estimação. Essa subamostra é chamada de subamostra de “*validação*” ou de “*teste*” (Carmack, Spence e Schucany, 2012).

Definição 9.8 (Validação Cruzada). *A validação cruzada consiste em subdividir uma amostra em duas subamostras com o objetivo de verificar a capacidade preditiva de modelos candidatos. A subamostra de estimação é utilizada para ajustar*

os modelos candidatos, isto é, para estimar os seus parâmetros. A subamostra de validação é utilizada para verificar a validade dos modelos candidatos através da predição das observações dessa amostra.

Quando a subdivisão é realizada uma única vez, a validação é chamada de *validação simples* (*hold-out validation*), mas quando a subdivisão é repetida diversas vezes, a validação é dita *validação cruzada* (Arlot e Celisse, 2010). Uma forma importante de validação cruzada é a formação das subamostras pela exclusão das observações “*uma-a-uma*” (*leave-one-out* ou *one-item-out*), que consiste em formar a subamostra de estimação excluindo-se uma única observação, enquanto que a subamostra de validação é composta apenas por essa observação excluída. Se a amostra tem n observações, a subdivisão pode ser repetida n vezes, obtendo-se a predição para todas as observações na amostra sem que elas tenham sido utilizadas no ajuste dos modelos que as predisseram.

Definição 9.9 (Validação Cruzada uma-a-uma). *Na validação cruzada uma-a-uma, a subamostra de estimação é formada pela exclusão de apenas uma das observações, enquanto que a subamostra de validação consiste nessa única observação excluída. Este procedimento é repetido para todas as observações na amostra, de modo a que todas as observações sejam preditas por modelos nos quais elas não participaram da estimação.*

A validação cruzada por exclusão também pode ser generalizada na forma da exclusão de k observações ($k = 1, 2, \dots$), formando subamostras de validação de tamanho k (Linhart e Zucchini, 1986; Yanagihara e Fujisawa, 2012). Essas k observações podem ser escolhidas aleatoriamente na amostra de n observações ou podem consistir de um agregado particular de observações. Quando a subdivisão é realizada aleatoriamente, a validação cruzada pela exclusão de observações “*k-a-k*” (*leave-k-out*) é raramente exaustiva, pois mesmo para valores pequenos de k é impraticável realizar todas as $\binom{n}{k}$ combinações possíveis de subdivisão da amostra (Arlot e Celisse, 2010).

Quando validação cruzada *k-a-k* utiliza agregados particulares de observações, a identificação dos agregados é dada pela própria estrutura dos dados. Por exemplo, num levantamento florestal, a amostra pode possuir uma estrutura hierárquica de observações, como as árvores agregadas em parcelas, e as parcelas agregadas em talhões. Na construção de equações volumétricas, a validação cruzada poderia ser realizada pela exclusão tanto das árvores individualmente, quanto das parcelas ou dos talhões. No primeiro caso, tem-se a validação cruzada uma-a-uma, enquanto no segundo tem-se a validação *k-a-k*, onde o número de observações k por agregado (parcela ou talhão) é variável.

Definição 9.10 (Validação Cruzada k -a- k). *Na validação cruzada k -a- k , a subamostra de estimação é formada pela exclusão de k observações, as quais formam a subamostra de validação. A seleção das k observações pode ser realizada aleatoriamente ou seguir agregados ou grupos naturais da estrutura da amostra disponível. O procedimento é repetido um número arbitrário de vezes.*

Um aspecto frequentemente enfatizado sobre a “validade” dos modelos verificada na validação cruzada é a detecção do problema de “sobre-ajuste” dos modelos (*model overfitting*). O sobre-ajuste acontece quando o modelo estocástico se ajusta excessivamente à amostra, refletindo antes atributos particularidades dela do que as características essenciais da população da qual ela foi tomada. Essa detecção do sobre-ajuste resulta da subamostra de estimação ser independente da subamostra de validação, de modo que o desempenho do modelo sobre-ajustado é marcadamente inferior na subamostra de validação quando comparado com seu desempenho na subamostra de estimação.

Critérios de Comparação de Modelos na Validação Cruzada

A detecção do problema de sobre-ajuste e a comparação do desempenho dos modelos requer a definição de uma “função de perda” (*loss function*) para avaliação dos modelos. Essa função de perda é definida por Arlot e Celisse (2010) como a esperança matemática, com respeito à amostra observada, de uma “função de contraste” (*contrast function*). É possível se utilizar funções de contraste diferentes, dependendo do tipo de modelo e de problema prático sendo estudado. Apresenta-se aqui dois tipos de modelos e de problemas que são de interesse frequente na Biometria Florestal.

Critério de Validação para Modelos de Distribuição

Um problema comum na Biometria Florestal é a modelagem do comportamento estocástico de uma medida no arvoredo ou na floresta, como por exemplo o DAP. Neste caso, o objetivo é que o modelo represente a *distribuição estocástica* dos valores da medida. O foco do problema não é a predição de valores individuais da medida, mas sim o comportamento estocástico do conjunto de valores da medida. No caso do exemplo do DAP, o modelo da distribuição dos diâmetros das árvores representa a estrutura de tamanho das árvores de um arvoredo ou floresta, sendo uma informação importante para sua caracterização ecológica e para o seu manejo florestal.

Estatisticamente, este problema é definido como a estimação de modelos de distribuição estocástica, ou mais especificamente de funções de densidade probabilística de modelos estocásticos. Arlot e Celisse (2010) mostram que neste caso a função de contraste deve ser a função logarítmica negativa da função de densidade

do modelo. Sendo $X = \{x_1, x_2, \dots, x_n\}$ uma amostra observada num cenário estocástico, cujo modelo operante é a densidade $\varphi(\cdot)$, e o modelo de aproximação candidato é $f(\cdot)$, então a função de perda é

$$\mathbf{E}_X \left\{ \ln \left(\frac{\varphi(x)}{f(x)} \right) \right\} = \int \ln \left(\frac{\varphi}{f} \right) \varphi d\mu,$$

que é a discrepância de Kullback-Leibler (Arlot e Celisse, 2010). Logo, o “critério natural” para se determinar o valor da função de perda numa amostra, através da validação cruzada uma-a-uma, é o Critério de Informação de Akaike (AIC) na seguinte forma:

$$\text{AIC}_{-1}(f) = -2 \sum_{i=1}^n \mathbf{L}_f \{ \hat{\theta}_{[-i]} \} + 2 K_f,$$

sendo $\mathbf{L}_f \{ \hat{\theta}_{[-i]} \}$ a log-verossimilhança do modelo f para a i ésima observação com base na estimativa dos parâmetros com a exclusão dessa observação ($\hat{\theta}_{[-i]}$), e K_f é o número de parâmetros do modelo.

Stone (1977) demonstrou que a seleção de modelos de densidade pela validação cruzada, na qual a função de contraste é o logaritmo da função de densidade, é assintoticamente equivalente ao Critério de Informação de Akaike quando todos os modelos candidatos são estimados por máxima verossimilhança. Logo, espera-se que para amostras suficientemente grandes

$$\text{AIC}_{-1}(f) \approx \text{AIC}(f),$$

de modo que o AIC pode ser utilizado na sua forma tradicional, sem a validação cruzada.

Contudo, é importante lembrar que, no caso da validação cruzada, as observações utilizadas para predição não são utilizadas para a estimação dos parâmetros dos modelos candidatos. Assim, o termo que penaliza os modelos em função do número de parâmetros se torna teoricamente dispensável de forma que a soma da log-verossimilhança dos valores preditos (*LL — Log-Likelihood*)

$$\text{LL}(f) = \sum_{i=1}^n \mathbf{L}_f \{ \hat{\theta}_{[-i]} \}$$

seria um critério suficiente para discriminar os modelos candidatos.

Critério de Validação para Modelos de Predição

Outro problema muito frequente na Biometria Florestal é o problema de predição, em que o objetivo é prever observações individuais de uma dada medida.

Isto ocorre tanto nos modelos arborimétricos, cujo objetivo é prever medidas de árvores individuais (altura, volume, biomassa), quanto nos modelos arbustimétricos, cuja meta é prever medidas de arvoredos como nas equações de produção (volume ou biomassa de madeira) e nos modelos de crescimento e produção (produção em função do tempo). Neste caso, a função de contraste “natural” é a função de quadrados mínimos (*least squares function*) (Arlot e Celisse, 2010).

Parte-se de um cenário estocástico no qual foi tomada uma amostra de observações pareadas $(\mathbf{X}, Y) = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$, sendo \mathbf{X} um conjunto de variáveis preditoras e Y a variável a ser predita, sendo i o índice das observações individuais ($i = 1, 2, \dots, n$). O modelo de predição tem forma

$$Y_i = f(\mathbf{X}_i) + \varepsilon_i,$$

onde a $f(\cdot)$ é função das variáveis preditoras que pode ser linear ou não-linear em seus parâmetros e o termo do erro (ε) tem distribuição estocástica simétrica em relação à sua esperança condicional que é nula: $\mathbf{E}\{\varepsilon_i|\mathbf{X}_i\} = 0$, com variância constante. Na prática, assume-se que o modelo estocástico operante pertence à subfamília Gaussiana homocedástica.

A função de perda, baseada na função de quadrados mínimos, é

$$\mathbf{E}_{(\mathbf{X}, Y)} \{(Y - f(\mathbf{X}))^2\}.$$

Note que se o interesse no problema não for a predição de observações individuais, mas a *estimação* da curva de predição, isto é, da “curva média” $\mu(Y) = \mathbf{E}\{Y|\mathbf{X} = \mathbf{x}\}$, então a função de perda se torna

$$\mathbf{E}_{(\mathbf{X}, Y)} \{(\mu(Y) - f(\mathbf{X}))^2\}.$$

Contudo, é importante lembrar que a predição das observações é realizada pelos valores da “curva média”, pois as observações individuais, para fins de predição, podem ser definidas como

$$Y_i = \mu(Y_i|\mathbf{X} = \mathbf{x}_i) + \varepsilon_i,$$

sendo ε_i é o erro aleatório que independe do modelo candidato $f(\cdot)$. Assim, também nos problemas de predição, a avaliação de modelos por meio da validação cruzada deve ser orientada pela discrepância (“distância”) entre a “curva média” e o modelo candidato, de modo que o critério de validação para predição e para estimação são equivalentes (Arlot e Celisse, 2010).

Critérios para Modelos de Predição no Contexto Geral da Amostra

A função de perda apresentada acima implica num critério de validação para o contexto global da amostra, onde os erros de predição são avaliados para amostra

como um todo. Neste contexto, o critério apropriado é “*erro quadrático médio de predição*” (*prediction mean squared error - PMSE*)⁵

$$PMSE = \mathbf{E}_{(\mathbf{X}, Y)} \{ (Y - f(\mathbf{X}))^2 \} .$$

O *PMSE* está diretamente ligado ao conceito da soma de quadrado de predição (*prediction sum of squares - PRESS*) utilizado no diagnóstico de medidas influentes em regressão linear, pelos resíduos deletados (Neter, Wasserman e Kutner, 1990). Mas o conceito de *PRESS* foi desenvolvido inicialmente no contexto da seleção de variáveis e modelos na regressão linear (Mosteller e Tukey, 1977; Draper e Smith, 1981).

Também Allen (1971) utilizou o *PMSE* como critério no contexto da seleção de variáveis na regressão linear, mas ele mostra que o *PMSE* é constituído de três elementos:

$$PMSE = \mathbf{Var}\{Y\} + \mathbf{Var}\{\hat{Y}\} + [\mathbf{Bias}\{\hat{Y}\}]^2 .$$

O primeiro termo ($\mathbf{Var}\{Y\}$) é a variância estrutural, isto é, a variância das observações devida à estrutura estocástica dos dados. O segundo termo ($\mathbf{Var}\{\hat{Y}\}$) é a variância das predições, enquanto que o terceiro termo ($[\mathbf{Bias}\{\hat{Y}\}]^2$) é o quadrado do viés de predição.

A variância estrutural depende apenas das observações presentes na amostra, sendo, portanto, independente do modelo candidato sob avaliação. Já a variância das predições e o viés das predições são medidas da qualidade preditiva dos modelos candidatos em termos da acurácia ou exatidão das predições, que reflete a discrepância global entre os modelos candidatos e o modelo operante.

Embora somente a variância das predições e o viés das predições caracterizem de fato o desempenho de um modelo candidato, a incerteza associada a uma predição é dada pelo *PMSE*, pois a incerteza da predição de uma observação individual inclui a variância do erro aleatório do modelo (ε). É provável que, na maioria das situações práticas de estudo, a variância estrutural tenha magnitude muito maior que a magnitude dos outros dois termos, de modo que ela obscureça as diferenças entre os modelos candidatos, quando estes são comparados através do *PMSE*. A variância estrutural é um componente “natural” do erro de predição, isto é, o componente que não pode ser alterado ou melhorado pela escolha do modelo candidato, mas que terá efeito marcante na incerteza da predição medida pela amplitude do *intervalo de predição*. No *PMSE*, a variância estrutural é o elemento que define a incerteza do processo de *predição*, enquanto que a variância da predição e o viés da predição estão associados à incerteza dos processos de *aproximação* e de *estimação*.

⁵Utiliza-se aqui o nome proposto por Allen (1971) e Hocking (1976), embora não no contexto de seleção modelos lineares pela exclusão de variáveis individuais, mas no contexto da seleção de modelos pela exclusão das *observações* uma-a-uma.

Na prática de comparação e seleção de modelos, é vantajoso se utilizar um critério cuja escala esteja na mesma escala da variável resposta, sendo predita. Por isto, no lugar do *PSME*, utiliza-se com mais frequência como critério que representa a variação nas predições, o erro padrão da estimativa (*SEE — standard error of estimate*):

$$SEE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{[i]})^2},$$

sendo y_i a i ésima observação e $\hat{y}_{[i]} = f_{[-i]}(\mathbf{x}_i)$ o seu valor predito na validação cruzada. Também na regressão linear existe um erro padrão da estimativa, calculado como a raiz quadrada do quadrado médio do resíduo, que é análogo a este, mas é obtido com base nos resíduos e não nos erros de predição.

Outro critério importante é o viés de predição (*PB - prediction bias*)

$$PB = \widehat{\mathbf{Bias}}\{\widehat{Y}\} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{[i]}) = \bar{y} - \bar{\hat{y}}_{[.]},$$

no qual \bar{y} é a média amostral e $\bar{\hat{y}}_{[.]}$ é a média das predições. Enquanto que o *SEE* é uma medida da exatidão das predições, o *PB* indica a magnitude de um erro sistemático nas predições.

Critérios para Modelos de Predição no Contexto Condicional

Nos problemas de predição, a predição de uma observação da variável predita (Y_i) é realizada condicionalmente aos valores das variáveis preditoras (\mathbf{X}_i), de forma que o modelo de predição pode ser definido não somente pelo pareamento das observações segundo o índice i , mas também na forma condicional:

$$(Y_i | \mathbf{X} = \mathbf{x}_i) = f(\mathbf{x}_i) + \varepsilon_i.$$

Logo, o que o modelo de predição faz é *estimar* a esperança matemática da variável predita condicionada às variáveis preditoras:

$$f(\mathbf{x}_i) = \mathbf{E}\{Y_i | \mathbf{X} = \mathbf{x}_i\}.$$

A avaliação da função de perda pode ser mais informativa quando realizada não no contexto global da amostra, mas no contexto condicional das variáveis preditoras:

$$\mathbf{E}_{(Y|\mathbf{X})} \{(Y - f(\mathbf{X}))^2\}.$$

Uma forma de implementar essa avaliação é utilizar os mesmos critérios apresentados acima (*SEE* e *PB*) para *classes* das variáveis preditoras.

Essa abordagem é frequentemente utilizada nos trabalhos de Biometria Florestal. No caso de modelos arborimétricos, as relações hipsométricas, as equações volumétricas e as equações de biomassa são comumente avaliadas para diferentes classes de DAP, enquanto que as equações de afilamento são avaliadas para diferentes posições ao longo do tronco. No caso de modelos arbustimétricos, equações de produção são avaliados para diferentes classes de sítio e área basal, enquanto que modelos de crescimento, além de serem avaliados para sítio e área basal, também são avaliados para diferentes idades da floresta.

Contudo a avaliação condicional por classes das variáveis preditoras possui várias limitações. Em modelos com diversas variáveis preditoras é impraticável identificar classes que envolvam todas as variáveis, o que enfatiza a noção de que as classes são essencialmente arbitrárias e dependentes da pessoa que realiza o estudo. Por outro lado, para que a avaliação por classe seja satisfatória é necessário que o número de observações por classe permita obter valores estáveis dos critérios de avaliação.

Uma abordagem alternativa é a avaliação com base nas observações individualmente por meio de métodos gráficos. Na avaliação gráfica, a arbitrariedade das classes das variáveis preditoras é eliminada e vários gráficos similares podem ser construídos para as principais variáveis preditoras do modelo.

Critérios em Modelos de Predição: AIC e SEE

Como apresentado acima, os modelos de predição na Biometria Florestal geralmente têm como classe de modelo básica a subfamília Gaussiana homocedástica. Nessa família, a log-verossimilhança é função da diferença entre os valores observados (y_i) e valores preditos (\hat{y}_i)

$$L\{\hat{\theta}\} = \ln \left[\sum_{i=1}^n (y_i - \hat{\mu}(y_i))^2 \right] = \ln \left[\sum_{i=1}^n (y_i - \hat{y}_i)^2 \right],$$

Logo, no caso da validação cruzada, a log-verossimilhança se torna uma função da soma de quadrado de predição (PRESS) e, portanto, do erro padrão da estimativa (*SEE*)

$$LL = \sum_{i=1}^n L\{\hat{\theta}_{[-i]}\} = \ln [\text{PRESS}] = \ln [n (\text{SEE})^2].$$

Assim, AIC e *SEE* em uma relação direta

$$AIC_{-1} = -2 \ln [n (\text{SEE})^2] + 2K.$$

Portanto, nos modelos Gaussianos, estes critérios são equivalentes à comparação dos modelos segundo a validação cruzada. Contudo, o AIC e a log-verossimilhança ($\mathbf{L}\{\hat{\boldsymbol{\theta}}_{[-i]}\}$) permitem apenas uma interpretação comparativa, isto é, somente compara os modelos indicando o melhor. Já o *SEE* e o *PB*, permitem uma interpretação na escada da variável predita e, conseqüentemente, podem ser utilizados para avaliar a qualidade preditiva de um dado modelo em termos da sua exatidão preditiva.

Cr terios em Modelos de Predi o: Crit rio de Furnival

O crit rio de Furnival (Furnival, 1961) procura contornar o problema de que, quando a vari vel resposta   transformada, o *SEE* se apresenta numa escala transformada e, conseqüentemente, o *SEE* de modelos com vari vel transformada e n o transformada n o s o compar veis. O crit rio de Furnival   em ess ncia a re-transforma o do *SEE* para escala original da vari vel resposta. Ele pode ser definido da seguinte maneira:

$$FC = [\tau'(y)]^{-1} SEE ,$$

na qual $\tau'(y)$   a primeira derivada da fun o de transforma o da vari vel resposta y , os colchetes ($[\cdot]$) representam a *m dia geom trica* dos valores e o erro padr o da estimativa (*SEE*) est  na escala transformada.

A transforma o mais comum nos modelos de predi o da Biometria Florestal   a transforma o por meio de logaritmos naturais. Neste caso, a primeira derivada  

$$\tau(y) = \ln(y) \Rightarrow \tau'(y) = 1/y ,$$

conseqüentemente, o crit rio de Furnival toma a seguinte forma:

$$FC = \left(\prod_{i=1}^n 1/y_i \right)^{-1/n} SEE ,$$

onde o *SEE* est  na escala do logaritmo natural. Para se facilitar o c lculo da n - sima raiz da produt ria de $1/y_i$, aplica-se a fun o logar mica seguida da fun o exponencial, pois essa   fun o inversa daquela:

$$\begin{aligned} FC &= \exp \left[\ln \left(\prod_{i=1}^n y_i^{-1} \right)^{-1/n} \right] SEE , \\ &= \exp \left[-\frac{1}{n} \sum_{i=1}^n -\ln(y_i) \right] SEE , \end{aligned}$$

A forma simplificada do critério de Furnival para transformação logarítmica se torna:

$$FC = \exp \left[\frac{1}{n} \sum_{i=1}^n \ln(y_i) \right] SEE .$$

Portanto, na comparação de modelos de predição em que a variável resposta foi transformada em alguns, mas não em outros, o *SEE* deve ser substituído pelo critério de Furnival. É importante, lembrar que o critério de Furnival é um critério apenas para o contexto geral da amostra.

Critérios em Modelos de Predição: AIC e o Coeficiente de Determinação

Um critério de uso geral no caso dos modelos de regressão linear, isto é, os modelos pertencentes à subfamília Gaussiana homocedástica com função linear para a média, é o *coeficiente de determinação*:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} .$$

Como critério de seleção de modelos, o coeficiente de determinação tem uma interpretação absoluta, pois ele expressa a proporção da variação ao redor da média que é explicada pelo modelo candidato. Na expressão acima, a razão dos somatórios é define a proporção da variância *não explicada* pelo modelo (numerador) em relação à variação ao redor da média (denominador). Como essa razão é subtraída da unidade, segue a interpretação tradicional do coeficiente. A variância não explicada pelo modelo é tradicionalmente a soma de quadrado de resíduos, mas no contexto da validação cruzada seria a PRESS, de modo que o coeficiente de determinação na validação cruzada uma-a-uma é

$$R_{-1}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{[-i]})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} .$$

Se na expressão do coeficiente de determinação, a PRESS for isolada obtém-se:

$$\sum_{i=1}^n (y_i - \hat{y}_{[-i]})^2 = (1 - R_{-1}^2) \sum_{i=1}^n (y_i - \bar{y})^2 .$$

Logo, a função de log-verossimilhança pode ser definida em termos do coeficiente de determinação:

$$LL = \sum_{i=1}^n \mathbf{L}\{\hat{\theta}_{[-i]}\} = \ln \left[(1 - R_{-1}^2) \sum_{i=1}^n (y_i - \bar{y})^2 \right]$$

e também o Critério de Informação de Akaike:

$$AIC_{-1} = -2 \ln \left[(1 - R_{-1}^2) \sum_{i=1}^n (y_i - \bar{y})^2 \right] + 2K.$$

Critérios de Seleção de Modelos na Subfamília Gaussiana Homocedástica

Nos modelos de predição pertencentes à subfamília Gaussiana homocedástica, a função de perda “natural” é a soma de quadrados da diferença entre os valores observados e os preditos. Deste fato resulta que, na validação cruzada, os critérios de seleção de modelos para avaliar a capacidade preditiva dos modelos são sempre uma função do PRESS e, conseqüentemente, da função de log-verossimilhança. Portanto, os critérios de seleção dos modelos preditivos, como o SEE e o coeficiente de determinação, são sempre variações do Critério de Informação de Akaike (AIC). Contudo, enquanto o AIC permanece um critério de seleção relativo, essas variações possibilitam, em contextos particulares dos modelos de predição, uma interpretação em termos absolutos, sendo, portanto, muito úteis para a seleção de modelos na prática da Biometria Florestal.



Exemplo 9-8: Equação de Volume para um Povoamento

O Problema. A construção de uma equação volumétrica para um único povoamento florestal é um caso em que os dados têm uma estrutura bastante simples e homogênea. Trata-se de um povoamento de *Eucalyptus grandis* na Estação Experimental de Ciências Florestais de Itatinga (ESALQ-USP), em que foram instaladas 3 parcelas de 360 m². Após a medição do DAP e altura de todas as árvores da parcelas, as árvores foram abatidas e cubadas rigorosamente, resultando num conjunto de 175 árvores. O objetivo é construir uma equação volumétrica para a predição do volume de madeira das árvores individualmente para gerar, numa segunda etapa, a estimativa do volume de madeira do povoamento.

Modelos Candidatos. Na construção da equação volumétrica, foram considerados quatro modelos logarítmicos:

$$\text{Husch: } \ln(v_i) = \beta_0 + \beta_1 \ln(d_i) + \varepsilon_i$$

$$\text{Spurr: } \ln(v_i) = \beta_0 + \beta_1 \ln(d_i^2 h_i) + \varepsilon_i$$

$$\text{Schumacher-Hall: } \ln(v_i) = \beta_0 + \beta_1 \ln(d_i) + \beta_2 \ln(h_i) + \varepsilon_i$$

$$\text{Baden-Würt: } \ln(v_i) = \beta_0 + \beta_1 \ln(d_i) + \beta_2 \ln^2(d_i) + \\ + \beta_3 \ln(h_i) + \beta_4 \ln^2(h_i) + \varepsilon_i.$$

O modelo de Husch é um modelo tradicionalmente utilizado para equações volumétricas locais, no qual o volume de madeira é função do DAP na escala logarítmica. O nome Husch é utilizado por Finger (1992). Os modelos de Spurr e Schumacher-Hall são modelos tradicionais para equações volumétricas padrão (dupla entrada), sendo amplamente reconhecidos como de grande aplicabilidade. Em relação ao modelo de Baden-Würt (Finger, 1992), espera-se problemas de ajuste devido à multicolinearidade devido certa redundância em sua estrutura, pois apresenta as variáveis DAP e altura em logaritmo e quadrado do logaritmo.

Ajuste dos Modelos. Os modelos foram ajustados pelo método de quadrados mínimos ordinários, que é equivalente ao ajuste por máxima verossimilhança neste caso, e as pressuposições fundamentais dos modelos lineares foram alcançadas na escala logarítmica, incluindo a distribuição dos resíduos bem próxima à distribuição Gaussiana. Os critérios de validação cruzada foram calculadas na escala original do volume das árvores (m^3). Como todos os modelos foram ajustados na escala logarítmica, a correção do viés logarítmico foi realizada.

Ajuste Obtido. A tabela 9.2 apresenta os critérios de ajuste dos modelos candidatos aos dados das 175 árvores. Como esperado, somente o modelo de equação local (modelo de Husch) se mostra claramente inferior aos demais. Considerando o limite para distinguir a plausibilidades dos modelos como sendo $\ln(8)$, nota-se que o Critério de Informação de Akaike (AIC) não faz distinção marcante entre os modelos de dupla entrada, nem na escala em que estes foram ajustados (escala logarítmica), nem na escala de predição do volume (m^3). O problema de multicolinearidade do modelo Baden-Würt não pode ser percebido por nenhum dos critérios de ajuste apresentados, mas é detectado pelo elevado erro padrão para as estimativas dos coeficientes de regressão, ao ponto de que três deles não sejam significativos.

Tabela 9.2: Critérios de *ajuste* dos modelos de equação de volume para *Eucalyptus grandis* na Estação Experimental de Ciências Florestais de Itatinga (ESALQ-USP). Estes critérios são baseados nos *resíduos* obtidos no ajuste dos modelos com todas as observações. O erro padrão da estimativa ($s_{\hat{y}}$) e o coeficiente de determinação ajustado (R_a^2) foram computados na escala de ajuste dos modelos ($\ln(m^3)$), enquanto que o critério de informação de Akaike (AIC) foi computado na escala de ajuste ($\ln(m^3)$) e de predição (m^3).

MODELO	$s_{\hat{y}}$	R_a^2	COEFICIENTES DE REGRESSÃO		
			NÃO SIGNIF.	AIC	
	$\ln(m^3)$			$\ln(m^3)$	m^3
Husch	0,0902	0,9442	—	-339,40	-874,11
Spurr	0,0657	0,9704	—	-449,63	-986,65
Schumacher-Hall	0,0658	0,9705	—	-448,03	-985,15
Baden-Würt	0,0654	0,9712	$\beta_0, \beta_2, \beta_3$	-448,52	-985,84

Validação Cruzada. Os critérios da validação cruzada (tabela 9.3), tanto da validação árvore-a-árvore (uma-a-uma) quanto da validação parcela-a-parcela (k - a - k) acrescentam pouco ao que a análise dos resíduos mostra. A distinção entre o modelo de equação local e os modelos de dupla entrada é clara, mas as diferenças entre os modelos de dupla entrada é igualmente tão pequena na validação cruzada quanto na análise dos resíduos.

Observa-se que os critérios aplicados aos resíduos mostram um desempenho melhor em todos os modelos que os critérios aplicados à validação cruzada árvore-a-árvore e estes, por sua vez, são melhores que os critérios aplicados à validação parcela-a-parcela. Contudo, o comportamento relativo dos modelos é muito semelhante nos três casos. A validação cruzada, no entanto, indica o modelo Spurr como superior aos demais e o modelo Baden-Würt com o de pior desempenho dentre os modelos de dupla entrada.

A tabela 9.3 também ilustra que o erro padrão de estimativa (SEE) dos modelos é bastante bom, ao redor dos 7%, e permite discriminar os modelos. Contudo, o erro de predição (\sqrt{PSME}) que de fato ocorrerá fica entre 30 e 40%, dependendo do tipo de validação cruzada realizada, e praticamente não difere entre os modelos candidatos.

Validação Cruzada por Classes de DAP. A validação cruzada condicional por classe de DAP (tabela 9.4) mostra que a análise do desempenho dos modelos por meio da validação cruzada não diferem marcadamente da análise dos resíduos. O

Tabela 9.3: Critérios da validação cruzada dos modelos de equação volumétrica de um arvoredo de *Eucalyptus grandis* na Estação Experimental de Ciências Florestais de Itatinga (ESALQ-USP). A *validação cruzada uma-a-uma* foi realizada árvore-a-árvore, enquanto que a *validação cruzada k-a-k* foi implementada parcela-a-parcela. Todos os critérios foram calculados na escala de predição do volume das árvores, isto é, na escala de metros cúbicos.

MODELO	ERRO PADRÃO DA ESTIMATIVA (<i>SEE</i>)		VIÉS DE DA PREDIÇÃO (<i>PB</i>)		ERRO DE DE PREDIÇÃO (\sqrt{PSME})		LOG- VEROS. (LL)	AIC
	(m^3)	(%)	($10^{-4} m^3$)	(%)	(m^3)	(%)		
VALIDAÇÃO CRUZADA ÁRVORE-A-ÁRVORE								
Spurr	0,01588	6,9	0,5424	0,02	0,0814	35,2	493,4	-980,84
Schumacher-Hall	0,01591	6,9	0,6115	0,03	0,0814	35,2	492,6	-977,36
Baden-Würt	0,01599	6,9	-0,4624	-0,02	0,0814	35,2	493,2	-974,43
Husch	0,02139	9,3	-3,2485	-0,14	0,0826	35,7	437,4	-868,82
VALIDAÇÃO CRUZADA PARCELA-A-PARCELA								
<i>Geral</i>								
Spurr	0,01611	7,0	-2,0951	-0,09	0,08142	35,2	489,3	-972,55
Schumacher-Hall	0,01620	7,0	-2,0828	-0,09	0,08135	35,2	488,6	-969,23
Baden-Würt	0,01640	7,0	-6,4918	-0,28	0,08148	35,2	487,9	-963,71
Husch	0,02196	9,5	-6,6842	-0,29	0,08277	35,8	432,7	-859,34
<i>Parcela 1 (56 árvores)</i>								
Spurr	0,01821	7,3	-11,479	-0,46	0,08174	32,7	148,2	-290,33
Schumacher-Hall	0,01813	7,3	-11,124	-0,45	0,08173	32,7	148,4	-288,78
Baden-Würt	0,01897	7,6	-18,735	-0,75	0,08193	32,8	146,9	-281,83
Husch	0,02558	10,2	-18,886	-0,76	0,08371	33,5	133,6	-261,24
<i>Parcela 2 (59 árvores)</i>								
Spurr	0,01419	6,3	6,2013	0,27	0,08906	39,4	174,2	-342,56
Schumacher-Hall	0,01452	6,4	5,9544	0,26	0,08112	39,4	173,6	-339,10
Baden-Würt	0,01448	6,4	-1,1571	-0,05	0,08907	39,4	173,2	-334,46
Husch	0,01713	7,5	5,8951	0,26	0,08957	39,6	154,5	-303,02
<i>Parcela 3 (59 árvores)</i>								
Spurr	0,01579	7,2	-1,4846	-0,07	0,07051	32,3	166,9	-327,77
Schumacher-Hall	0,01587	7,2	-1,5385	-0,07	0,07053	32,3	166,7	-325,35
Baden-Würt	0,01565	7,2	-2,0553	-0,09	0,07048	32,3	167,7	-323,42
Husch	0,02254	10,3	-7,6824	-0,35	0,07233	33,1	144,6	-283,18

modelo de Husch (equação local) se mostra claramente inferior aos modelos de dupla entrada, com marcante tendência de viés, principalmente na última classe

de DAP. Dentre os modelos de dupla entrada, o modelo Baden-Würt apresenta um desempenho ligeiramente superior tanto em termos do erro padrão de predição (SEE) quanto em termos do viés de predição (PB), uma vez que estes se mostram inferiores e mais constantes nas diferentes classes de DAP.

Crítério de Informação de Akaike — AIC. O único critério que discrimina marcadamente os modelos é o AIC. Considerando que as diferenças revelantes de desempenho são aquelas em que a diferença no AIC são maiores em valor absoluto que logaritmo natural de oito ($\ln(8) = 2,079$), o AIC indica o modelo de Spurr como superior aos demais modelos de dupla entrada. Essa superioridade é vista na validação cruzada (tabela 9.3), mas não na análise dos resíduos (tabela 9.2). Contudo, a diferença de desempenho apontada pelo AIC na validação cruzada não é observada na log-verossimilhança, o que mostra que o elemento no AIC que está distinguindo os modelos é de fato a penalização pelo número de parâmetros (tabela 9.3).

Análise de Resíduos e Validação Cruzada. A comparação do AIC entre a análise dos resíduos e a validação cruzada (tabela 9.5) mostra que há uma redução no desempenho preditivo dos modelos. Curiosamente, a equação local (Husch) teve uma piora no desempenho tão pequena quanto o melhor modelo de dupla entrada (Spurr). Essas mudanças nos valores do AIC foram o único critério capaz de discriminar claramente os modelos de dupla entrada. Nessa discriminação, o modelo Schumacher-Hall se mostrou um pouco inferior ao modelo de Spurr, enquanto que o modelo Baden-Würt se mostrou claramente inferior. Novamente, essa discriminação pode refletir antes uma diferença de complexidade entre os modelos que uma verdadeira diferença na capacidade preditiva.

Embora a diferença de AIC entre a análise de resíduos e as validações cruzadas elimine algebricamente a diferença no número de parâmetros do modelo, é curioso notar que a diferença de AIC é aproximadamente duas vezes o número de parâmetros para comparação da validação árvore-a-árvore e algo entre três e quatro vezes para comparação da validação parcela-a-parcela. Já a diferença no AIC entre a validação árvore-a-árvore e a parcela-a-parcela parece não estar associada ao número de parâmetros no modelo.



Tabela 9.4: Critérios da validação cruzada por classe de DAP aplicados tanto para os resíduos do ajuste quanto para os erros de predição dos modelos de equação volumétrica de um arvoredo de *Eucalyptus grandis* na Estação Experimental de Ciências Florestais de Itatinga (ESALQ-USP). A *validação cruzada uma-a-uma* foi realizada árvore-a-árvore, enquanto que a *validação cruzada k-a-k* foi implementada parcela-a-parcela.

CLASSE DE DAP	NÚMERO DE ÁRVORES	MODELOS											
		HUSCH		SPURR		SCHUMACHER-HALL		BADEN-WÜRT					
		SEE (m^3)	PB ($10^{-3}m^3$)										
Resíduos do Ajuste													
(10,12]	15	0,005885	-1,249	0,005345	1,899	0,005497	1,939	0,004713	1,055	0,004713	1,055		
(12,14]	27	0,014045	-3,199	0,009484	-2,434	0,009537	-2,510	0,009295	-1,831	0,009295	-1,831		
(14,16]	56	0,021748	5,173	0,014517	1,527	0,014354	1,280	0,014255	1,841	0,014255	1,841		
(16,18]	56	0,022097	0,610	0,019085	-0,823	0,019120	-0,849	0,019495	-1,392	0,019495	-1,392		
(18,22]	20	0,030167	-13,488	0,019942	0,339	0,019750	1,286	0,018873	0,290	0,018873	0,290		
Validação Cruzada Árvore-a-árvore													
(10,12]	15	0,006165	-1,259	0,005528	1,971	0,005713	2,008	0,005138	1,055	0,005138	1,055		
(12,14]	27	0,014218	-3,250	0,009607	-2,473	0,009762	-2,566	0,009605	-1,901	0,009605	-1,901		
(14,16]	56	0,021881	5,205	0,014607	1,536	0,014518	1,302	0,014402	1,851	0,014402	1,851		
(16,18]	56	0,022293	0,614	0,019261	-0,830	0,019361	-0,865	0,019934	-1,440	0,019934	-1,440		
(18,22]	20	0,030795	-13,785	0,020269	0,354	0,020221	1,267	0,019699	0,221	0,019699	0,221		
Validação Cruzada Parcela-a-parcela													
(10,12]	15	0,007458	-2,402	0,005920	1,097	0,006141	1,153	0,005605	0,334	0,005605	0,334		
(12,14]	27	0,014954	-3,536	0,010189	-2,665	0,010157	-2,525	0,010248	-2,023	0,010248	-2,023		
(14,16]	56	0,021696	4,933	0,014459	1,341	0,014530	1,189	0,014455	1,464	0,014455	1,464		
(16,18]	56	0,022235	0,383	0,019219	-0,978	0,019355	-1,258	0,020293	-2,159	0,020293	-2,159		
(18,22]	20	0,033930	-14,126	0,021733	-0,064	0,021842	0,923	0,020888	-1,221	0,020888	-1,221		

SEE — Standard Error of Estimate: erro padrão da estimativa

PB — Prediction Bias: vies de predição

Tabela 9.5: Mudança nos valores do AIC entre a análise do resíduo e a validação cruzada para modelos de equação volumétrica de um arvoredo de *Eucalyptus grandis* na Estação Experimental de Ciências Florestais de Itatinga (ESALQ-USP).

MODELO	K	ANÁLISE		VALIDAÇÃO CRUZADA			
		DE RESÍDUO	ÁRVORE			PARCELA	
			AIC _R	AIC _A	Δ_{A-R}	AIC _P	Δ_{P-R}
Spurr	3	-986,65	-980,84	5,81	-972,55	14,10	8,29
Schumacher-Hall	4	-985,15	-977,36	7,79	-969,23	15,92	8,13
Baden-Würt	6	-985,84	-974,43	11,41	-963,71	22,13	10,72
Husch	3	-874,11	-868,82	5,29	-859,34	14,77	9,48

K - número de parâmetros no modelo; $\Delta_{A-R} = AIC_A - AIC_R$; $\Delta_{P-R} = AIC_P - AIC_R$.

Análise Gráfica e Validação Cruzada nos Modelos de Predição

Gráfico de Dispersão do Resíduo e do Erro de Predição

Tradicionalmente, em modelos de predição de regressão linear e não-linear, o julgamento da adequação de um modelo inclui algumas avaliações gráficas do comportamento dos resíduos do modelo ajustado. Essas geralmente são (i) o gráfico de dispersão do resíduo contra os valores ajustados, (ii) o gráfico da raiz quadrada do valor absoluto do resíduo padronizado contra os valores ajustados e (iii) o gráfico quantil-quantil dos quantis empíricos dos resíduos contra os quantis da distribuição Gaussiana padronizada. O primeiro gráfico permite verificar tendências em relação a superestimação ou subestimação em função do valor ajustado. O segundo gráfico é utilizado para verificar se a variância do resíduo é constante e independente do valor ajustado, enquanto que o terceiro gráfico verifica a suposição de “normalidade” dos resíduos.

Em princípio, estes mesmos três gráficos podem ser utilizados para avaliar os modelos de predição através da validação cruzada. Neste caso, os resíduos são substituídos pelos erros de predição, enquanto que os valores preditos tomam o lugar dos valores ajustados. Embora o gráfico da homogeneidade de variância seja de pouca importância no caso de predições, o gráfico de dispersão dos erros de predição é importante para detectar tendências de viés que variam com a magnitude do valor predito. Também a suposição de “normalidade” não é muito relevante no caso de predições, embora tenha uma forte influência sobre os valores do AIC. Mas o gráfico quantil-quantil permite observar assimetrias sutis na distribuição dos erros de predição, que é um elemento importante a ser avaliado num modelo de predição.

A comparação do comportamento de um dado modelo de predição nos gráficos

de resíduos em relação aos mesmos gráficos com os erros de predição também é muito informativa. Bons modelos candidatos não devem ter o seu comportamento muito alterado de uma situação para outra e uma alteração marcante pode ser suficiente para descartar um modelo candidato.



Exemplo 9-9: Equação de Volume para um Povoamento II

Em sequência ao exemplo 10-8, os modelos de equação volumétrica para de um povoamento de *Eucalyptus grandis* da Estação Experimental de Ciências Florestais de Itatinga (ESALQ-USP) são comparados graficamente.

Gráficos de Dispersão. Em todos os modelos candidatos, os gráficos de dispersão dos erros de predição na validação cruzada parcela-a-parcela revela o mesmo padrão que pode ser observado nos gráficos de dispersão dos resíduos envolvendo todas as observações. A figura 9.3 ilustra este resultado apenas para os modelos de Husch e Spurr.

Como observados nos critérios de seleção de modelos, a maior distinção ocorre entre a equação local (Husch) e as equações de dupla entrada. As equações de dupla entrada se mostram muito semelhantes entre si, com ligeira vantagem para o modelo de Spurr. Este mesmo padrão se repete nos gráficos de quantil-quantil para avaliação da pressuposição de “normalidade” e nos gráficos do resíduo padronizado e erros de predição padronizados para verificação da pressuposição da homoscedasticidade. Os erros de predição tendem a ter uma amplitude de variação maior do que os resíduos, mas o padrão observado nos gráficos é muito semelhante, de forma que a validação cruzada árvore-a-árvore ou parcela-a-parcela não revela nenhuma informação nova.



Gráfico da Log-Verossimilhança dos Erros de Predição

Já o critério de informação de Akaike (AIC) não é geralmente associado a nenhum tipo de apresentação gráfica. Contudo, as predições de um modelo de predição ajustado por máxima verossimilhança podem ser consideradas funções das estimativas de máxima verossimilhança dos parâmetros do modelo, sendo elas mesmas “*predições de máxima verossimilhança*”. Essa condição pode ser esten-

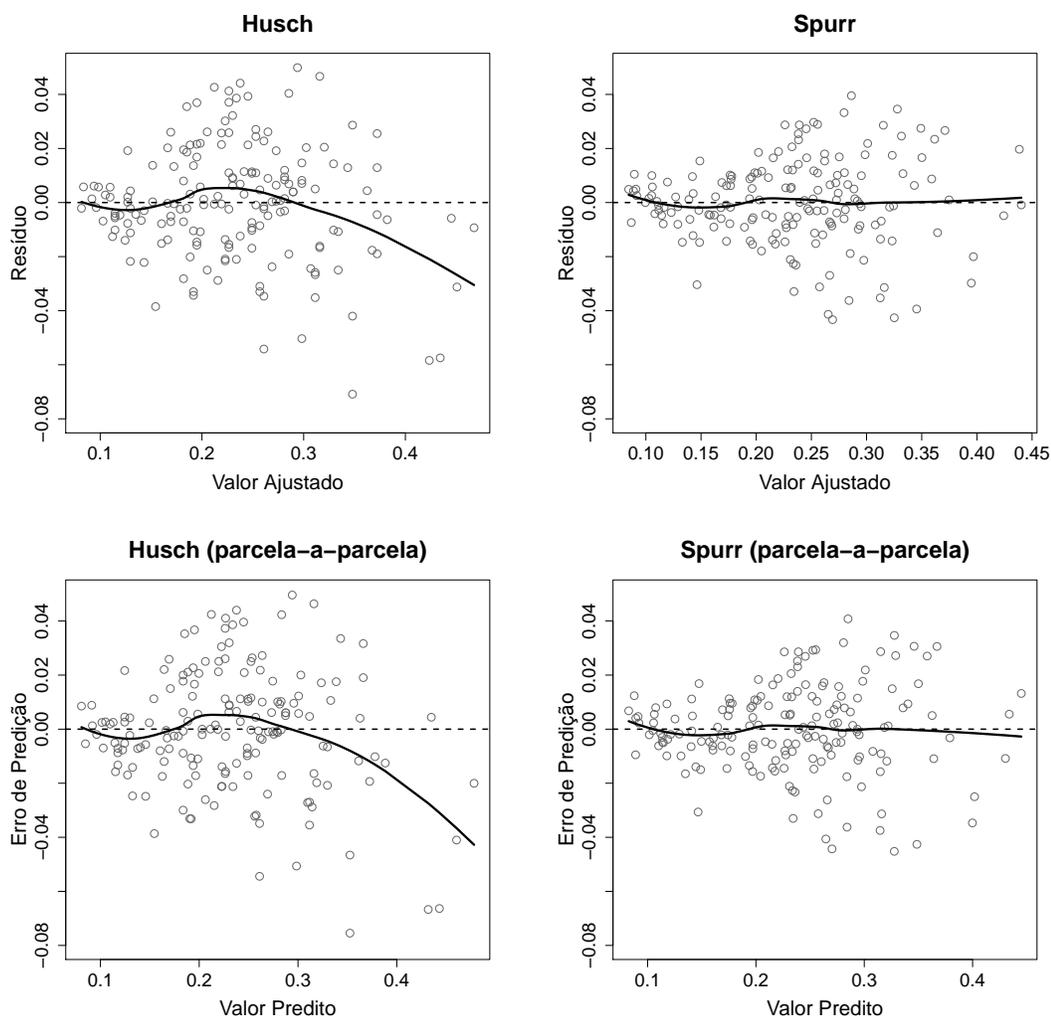


Figura 9.3: Gráficos de dispersão do resíduo contra o valor ajustado (alto) e do erro de predição contra o valor predito na validação cruzada parcela-a-parcela (baixo) para dois modelos (Husch e Spurr) de equações volumétrica das árvores num arvoredo de *Eucalyptus grandis* da Estação Experimental de Ciências Florestais de Itatinga (ESALQ-USP).

dida aos erros de predição, uma vez que estes são uma função linear das predições. Portanto, a cada erro de predição pode ser associada uma log-verossimilhança que pode ser calculada com base no modelo candidato. Como a log-verossimilhança não interpretável diretamente, a *log-verossimilhança relativa* de cada erro de predição pode ser obtida subtraindo-se o seu valor de log-verossimilhança do valor

máximo encontrado nos dados.

Como os modelos de predição são baseados na distribuição Gaussiana, espera-se que os erros próximos a zero tenham maior log-verossimilhança relativa que os grandes erros negativos e positivos. Assim, espera-se que, num gráfico de dispersão dos erros de predição (abcissas) pela log-verossimilhança relativa (ordenadas), as observações se distribuam como uma *nuvem de ponto* na forma de um curva parabólica côncava com ponto de máximo no valor zero. Obviamente, cada modelo candidato terá a sua própria nuvem de pontos, cuja dispersão e curvatura ilustram graficamente o desempenho preditivo do modelo e, conseqüentemente, as diferenças entre as nuvem de pontos ilustram graficamente as diferenças de AIC encontradas entre os modelos candidatos.

A comparação gráfica das nuvens de pontos de vários modelos candidatos seria, no entanto, um procedimento muito ineficiente, pois consistiriam na comparação de padrões em vários gráficos diferentes. Para que a comparação se torne eficiente, as nuvens de pontos de cada modelo candidato podem ser substituídas pela curva de suavização (*smoothing*) da log-verossimilhança em função do erro de predição. Assim, as curvas de suavização dos vários modelos candidatos podem ser colocadas num mesmo gráfico em termos de sua dispersão e curvatura, tornando possível a comparação gráfica das discrepâncias ente os modelos candidatos.



Exemplo 9-10: Equação de Volume para um Povoamento III

Em seqüência ao exemplo 10-9, os modelos de equação volumétrica para um povoamento de *Eucalyptus grandis* da Estação Experimental de Ciências Florestais de Itatinga (ESALQ-USP) são comparados graficamente.

Gráficos da Log-Verossimilhança. O gráfico da log-verossimilhança (figura 9.4) evidencia as mesmas pequenas diferenças numéricas da log-verossimilhança e do AIC entre os modelos. As curvas de log-verossimilhança dos modelos de dupla entrada são, em termos práticos, indistinguíveis entre si, enquanto que a curva da equação local (Husch) possui uma curvatura claramente menor que as curvas dos modelos de dupla entrada.

O gráfico da log-verossimilhança dos erros de predição permite observar que as curvas são aproximadamente simétricas, mas a equação local (Husch) é um pouco menos simétrica que os modelo de dupla entrada. Considerando o valor canônico para diferenças relevantes de log-verossimilhança ($\ln(8) = 2,079$), o gráfico mostra que o intervalo de log-verossimilhança para os erros de predição nos modelos

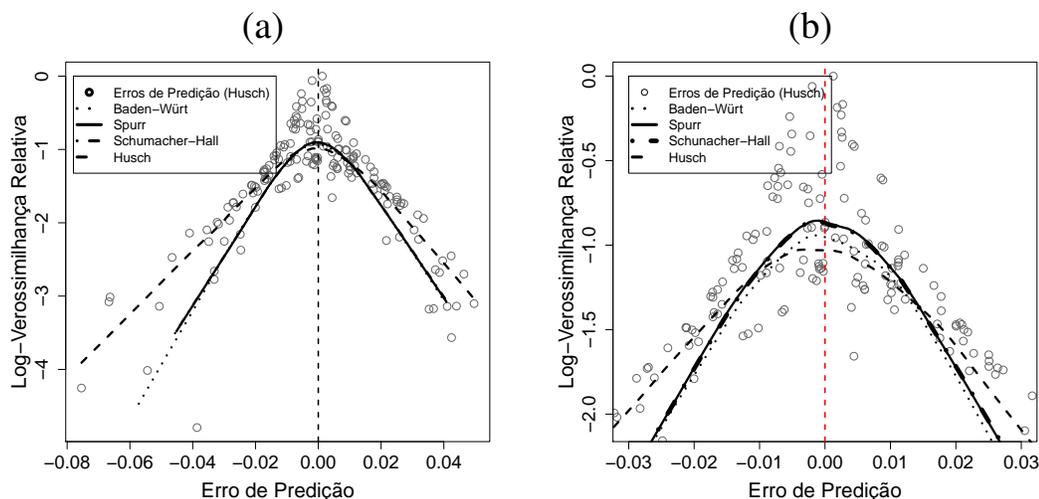


Figura 9.4: Log-verossimilhança da predição de cada observação na validação cruzada parcela-a-parcela grafada contra os erros de predição (m^3) para o volume de madeira de árvores num arvoredo de *Eucalyptus grandis* da Estação Experimental de Ciências Florestais de Itatinga (ESALQ-USP). Os pontos representam os erros de predição segundo o modelo de equação volumétrica de Husch (equação local), enquanto que as linhas de cada cor representam as curvas de log-verossimilhança dos modelos de equação volumétrica segundo a legenda no gráfico. As curvas foram geradas pela suavização da relação do log-verossimilhança e erro de predição das observações individuais. Em (a) o gráfico apresenta todos as observações, em (b) o gráfico ressalta uma região mais próxima do valor zero para enfatizar as diferenças sutis das curvas dos modelo de equação volumétrica de dupla entrada.

de dupla entrada seria aproximadamente $\pm 0,04 m^3$, enquanto que no modelo de Husch seria aproximadamente $[-0,06; 0,05] m^3$. Tais valores são compatíveis com os valores esperados para o intervalo de log-verossimilhança que seria duas vezes o erro padrão da estimativa ($2 SEE$).



9.6 Avaliação de Modelos Florestais

Vanclay e Skovsgaard (1997) discutiram a questão da avaliação de modelos florestais e propuseram um conjunto de procedimentos para sua realização. Este trabalho se tornou referência para vários outros tanto aqueles voltados para construção de modelos florestais (Soares e Tome, 2001; Soares e Tome, 2002; Namaalwa, Eid e Sankhayan, 2005; Barrio-Anta, Castedo-Dorado, Dieguez-Aranda, Alvarez-Gonzalez, Parresol e Rodriguez Soalleiro, 2006; Bi, Long, Turner, Lei, Snowdon, Li, Harper, Zerihun e Ximenes, 2010; Vanclay, 2010; Rathbun, LeMay e Smith, 2011; Lekwadi, Nemesova, Lynch, Phillips, Hunter e Mac Siurtain, 2012) quanto para aqueles voltados à avaliação de modelos (Soares, Tomé, Skovsgaard e Vanclay, 1995; Vospernik, Monserud e Sterba, 2010). Mas na sua discussão, Vanclay e Skovsgaard procuraram englobar os vários tipos de modelos florestais, tanto os modelos empíricos tradicionais quanto os modelos de processo e os simuladores de vegetação. Face a grande amplitude dos modelos discutidos, os procedimentos indicados por estes autores servem como linhas gerais de orientação, embora, com frequência, enumerem numa única lista processos e procedimentos de natureza muito distinta. Contudo, estes autores deram grande destaque à análise das propriedades estatísticas e à avaliação empírica dos modelos e discutiram em detalhe a questão da *partição dos dados* para avaliação dos modelos. Um ponto importante do trabalho de Vanclay e Skovsgaard (1997) é a clareza com que demonstram que a forma de partição dos dados depende da natureza e tipo de modelo sendo avaliado.

A questão da abordagem estatística apropriada para avaliar empiricamente os modelos florestais também foi objeto de estudo em vários trabalhos. A abordagem inicial foi a da inferência clássica segundo o teste de hipótese, tendo sido proposta por Freese (1960). Num primeiro momento, a ideia foi reformulada e adaptada para incluir testes paramétricos ou não-paramétricos (Reynolds Jr., 1984; Gregoire e Reynolds, 1988; Robinson e Froese, 2004), mas ela se mostrou empiricamente ineficiente (Yang, Monserud e Huang, 2004).

Assim, os critérios de validação dominantes na literatura florestal tem sido o erro padrão da estimativa (*SEE – standard error of estimate*), o viés (*bias*), o coeficiente de determinação (R^2) e suas variantes, e a raiz do quadrado médio do resíduo (*RMSE – root mean square error*). Tais critérios são geralmente utilizados sem aplicação de qualquer teste estatístico, utilizando-se um julgamento florestal para avaliar a sua magnitude e comparar os modelos.

Também a relevância da validação cruzada foi seriamente questionada. Kozak e Kozak (2003) num estudo empírico utilizando equação de volume e de relação hipsométrica compararam a validação cruzada contra a avaliação baseada apenas nos resíduos obtidos no procedimento de ajuste dos modelos. Foram testados vários tamanhos de amostras de validação com base em 1000 subamostras de va-

lidação selecionadas aleatoriamente e investigados tanto critérios globais quanto critérios condicionados por classes de DAP. Eles concluíram que a validação cruzada não acrescenta informação adicional além daquelas obtidas no procedimento de ajuste dos modelos ao conjunto de dados completos. Os exemplos apresentados acima relativos às equações de volume para um povoamento de *Eucalyptus grandis* indicam resultados semelhantes aos encontrados por Kozak e Kozak (2003).

Um aspecto importante da seleção dos modelos é o nível de complexidade ideal para se solucionar um dado problema prático. O nível de complexidade envolve tanto a complexidade da forma funcional dos modelos (Astrup, Coates e Hall, 2008) quanto o grau de detalhamento da estrutura hierárquica das observações e dos dados que é incorporado à estrutura dos modelos. Nos modelos biométricos florestais utilizados na Mensuração Florestal, este aspecto é particularmente importante, pois em geral o nível hierárquico de ajuste e predição dos modelos não é o nível hierárquico de aplicação final das predições dos modelos. Por exemplo, equações de volume são ajustadas e fazem predições no nível das árvores individuais, mas a aplicação final das predições é a produção de volume de madeira da floresta por unidade de área, seja como medida no nível da parcela, seja como estimativa no nível do povoamento ou da floresta como um todo.

A estrutura hierárquica das observações, contudo, varia com os casos particulares, pois cada caso tem a sua estrutura específica que demanda uma abordagem própria. Embora, dados de uma mesma situação geral, como por exemplo florestas plantadas, tendam a ter a mesma estrutura em linhas gerais, a heterogeneidade global e em cada nível dessa estrutura varia de caso para caso. Por isto, é muito difícil estabelecer uma abordagem ou procedimento padrão para a construção e seleção de modelos biométricos voltados à Mensuração Florestal nas situações gerais.

9.7 Conclusão

Neste capítulo, tratou-se da questão de especificação e seleção de modelos biométricos, sendo apresentados modelos úteis à Mensuração Florestal. Este tratamento completa o desenvolvimento de uma abordagem de inferência baseada na verossimilhança voltada às necessidades da Biometria Florestal, que é o tema da segunda parte deste trabalho. Resta agora, apresentar como essa abordagem pode ser utilizada na prática da Biometria Florestal voltada aos problemas da Mensuração Florestal.

Parte III

Aplicações da Inferência Segundo o Axioma da Verossimilhança na Biometria Florestal

Capítulo 10

Modelagem da Distribuição de Diâmetros Representando a Estrutura da Floresta

10.1 Introdução

A modelagem da distribuição do diâmetro das árvores de florestas e povoamentos florestais é um problema clássico da Mensuração Florestal, sendo, neste capítulo, tratado na perspectiva da Inferência baseada no Axioma da Verossimilhança. Inicialmente, realiza-se uma revisão concisa sobre o tema, seguindo-se de uma discussão sobre os problemas a ele relacionados. O capítulo é concluído com um exemplo de aplicação que trata da modelagem da distribuição de diâmetro das árvores de uma floresta nativa tropical.

10.2 Modelagem da Distribuição de DAP

A Importância da Distribuição de DAP

Na Mensuração e Manejo Florestal, reconhecem-se três razões para a representação da estrutura da floresta por meio da distribuição do DAP das árvores: (1) a influência da estrutura da floresta nas prescrições de manejo florestal, (2) a necessidade de uma técnica quantitativa para comparar e discriminar povoamentos florestais e (3) a representação da estrutura da floresta como um meio de se obter a predição da produção ou do crescimento da floresta.

Em 1898, François de Liocourt reconheceu a importância da frequência de árvores por classes de tamanho para o manejo de florestas dissetâneas (de Liocourt,

2001). Ele examinou a distribuição de DAP de sete povoamentos de abeto (*Abies alba* Mill.) e concluiu que a distribuição da frequência nas diferentes classes de DAP era semelhante nos povoamentos e podia ser aproximada por uma série geométrica de números. A partir deste resultado, de Liocourt propôs um método prático para o manejo de povoamentos dissetâneos baseado na colheita de árvores em todas as classes de diâmetro, de modo a manter a mesma distribuição de frequência à medida que as árvores crescem.

De Liocourt fincou um marco no manejo florestal e seu trabalho foi a base para o desenvolvimento de toda uma linha de pesquisa, tanto no manejo de florestas dissetâneas, quanto na predição de produção e crescimento de povoamentos florestais (Kerr, 2014). Além disto, este trabalho foi a origem de uma série de trabalhos voltados à modelagem da distribuição de DAP em povoamentos dissetâneos, destacando-se, na América do Norte, o trabalho de Meyer e Stevenson (1943) que já reconheciam a série geométrica de de Liocourt como a distribuição exponencial, dando um tratamento matemático mais apropriado à sua aplicação. Destacam-se também os trabalhos iniciais de aplicação de Meyer (1952) e Schmelz e Lindsey (1965), a apresentação do método de de Liocourt como método padrão de manejo de florestas dissetâneas por Meyer *et al.* (1961), no livro clássico “*Forest Management*”, o tratamento matemático detalhado apresentado por Leak (1965) e a ótima interpretação da distribuição fornecida por Zeide (1984).

A distribuição de DAP também foi utilizada para comparação e discriminação da estrutura de povoamentos florestais visando uma compreensão mais profunda a dinâmica de regeneração das florestais. Hough (1932) discute quais informações sobre a floresta que a distribuição fornece, no caso de povoamentos equiâneos de *Pinus strobus* (white pine) na Pensilvânia. Ele conclui que a distribuição de DAP revela pouco sobre a idade de povoamentos mistos e que a distribuição do diâmetro de tocos de árvores em florestas virgens sugere que as árvores tinham aproximadamente a mesma idade e, conseqüentemente, aponta para possibilidade de que a origem da floresta foi um evento catastrófico.

Concomitante a este desenvolvimento, ocorreu o estabelecimento do conceito de “tabela de povoamento” (*stand table*), em que o número de árvores por unidade de área e outros atributos do povoamento são apresentados por classes de DAP. No início do século XX, Graves (1906) já apresentava o método da tabela de povoamento (“*arbitrary group method*”) como a técnica adequada para a determinação do volume de madeira em povoamentos em que o diâmetro das árvores “varia consideravelmente”. Chapman (1921, 1924) dá seqüência à abordagem de Graves, acrescentando a utilização das tabelas de povoamento para a projeção da produção futura dos povoamentos nas duas edições do seu livro sobre mensuração florestal. As tabelas de povoamentos estão presentes também nos livros textos de Belyea (1931) e de Bruce e Schumacher (1935), como técnica para mensuração e manejo de povoamentos.

Mas, nessa época, as distribuições de diâmetro já estavam associadas às tabelas de povoamento, pois Schnur (1934) inicia o seu trabalho sobre a distribuição de diâmetros em povoamentos de *Pinus taeda* em Maryland, afirmando que “*embora os florestais europeu têm feito uso das distribuições de frequência de diâmetro, ou tabelas de povoamento, pelos últimos 50 anos, foi somente na última década que os mensuracionistas americanos focaram sua atenção nelas*”. Embora a apresentação do tema por Schnur (1934) seja predominante gráfica, ele já faz uso de várias famílias de distribuições estocásticas diferentes, comparando-as na busca da distribuição mais apropriada.

Famílias de Distribuições Estocásticas

As abordagens iniciais para representação matemática da distribuição de DAP foram semelhantes à abordagem usada por de Liocourt e comumente utilizadas pelos matemáticos no século XIX, que representavam uma dada família estocástica por uma série de somas lineares consistindo de uma função geradora e suas sucessivas derivadas (Samuelson, 1943). Schnur (1934) utilizou as *curvas* de Pearson tipo I, que é uma generalização da família beta, e tipo III, que é a família de Quiquadrado, a qual por sua vez é um caso particular da família gama. Mas, Schnur também utiliza as *curvas* de Gram-Charlier tipo A, que é uma série numérica cuja função geradora é a função de densidade da família Gaussiana (normal), e curva tipo B, que é uma série cuja função geradora é a densidade da família Poisson (Samuelson, 1943). Ainda em 1952, Meyer, ao representar a estrutura de povoamentos dissetâneos “balanceados” mistos da Pensilvânia, explicou a família exponencial como uma série geométrica.

Uma outra abordagem era utilizar funções com contra-domínio no intervalo $[0, 1]$ para representar a função de distribuição de uma família estocástica. Para construir tabelas de povoamentos para florestas equiâneas, Osborne e Schumacher (1935) utilizaram a *curva de crescimento populacional Pearl-Reed*, que é o nome que se dava no início de século XX à curva logística (Lloyd, 1967). Estes métodos matematicamente mais complexos provavelmente se faziam necessários numa época em que não havia equipamentos com grande capacidade de computação e todos os cálculos eram necessariamente realizados por operação humana.

Desde do início do estudo da estrutura de povoamentos florestais através de modelos matemáticos, várias famílias estocásticas foram aplicadas em muitas situações particulares. Partindo do trabalho de de Liocourt, a família exponencial foi amplamente discutida e aplicada em vários povoamentos e florestas dissetâneos (Meyer e Stevenson, 1943; Meyer, 1952; Meyer et al., 1961; Leak, 1965; Schmelz e Lindsey, 1965; Moser, 1972; Moser, 1976; Chapman e Weatherhead, 1984; Zeide, 1984; Chevrou, 1990; Muller-Landau et al., 2006; Rubin et al., 2006; Taubert et al., 2013; Kerr, 2014).

Muitas outras famílias estocásticas foram estudadas e aplicadas às mais variadas situações particulares de florestas e povoamentos florestais. Dentre elas, estão: a família Guassiana (Hough, 1932; Schnur, 1934), a família gama (Schnur, 1934; Nelson, 1964), a família log-normal (Bliss e Reinker, 1964; Nanang, 1998), a família beta (Schnur, 1934), a família Weibull (Bailey e Dell, 1973; Dell, 1973; Schreuder e Swank, 1974; Ek et al., 1975; Maltamo et al., 2000; Nord-Larsen e Cao, 2006; Westphal et al., 2006; Nord-Larsen e Cao, 2006; Cao, 2004; Zhang et al., 2001; Zhang et al., 2003; Muller-Landau et al., 2006; Palahí et al., 2007; Taubert et al., 2013), a família beta especial de Johnson (*Johnson's S_b distribution*) (Hafley e Schreuder, 1977; Schreuder e Hafley, 1977; Hafley e Buford, 1985; Knoebel e Burkhart, 1991; Zhang et al., 2003; Rennolls e Wang, 2005), a família de Burr (Lindsay et al., 1996) e a família Birnbaum-Saunders (Podlaski, 2008).

Particularmente importante para o desenvolvimento da modelagem da distribuição de DAP e sua incorporação nos modelos de crescimento e produção foi a família Weibull. No início da década de 1970, Bailey e Dell (Bailey e Dell, 1973; Dell, 1973) propuseram a Weibull como um modelo vantajoso para a modelagem da estrutura das florestas, argumentando de modo convincente e mostrando que essa família além de ser flexível para representar povoamentos com estruturas bastante variadas, pois comportava distribuições com diferentes formas, também possuía vantagens muito importantes na prática, como a possibilidade de ser ajustada por diferentes métodos (método da máxima verossimilhança, método dos momentos, método dos percentis, método de regressão linear) e a simplicidade na sua manipulação matemática para aplicação às circunstâncias práticas, como, por exemplo, no cálculo da frequência esperada por classes de DAP.

Devido às suas vantagens, a família Weibull se tornou praticamente uma *família estocástica de referência* na modelagem da distribuição de DAP das florestas, seja para sua representação quantitativa, para investigação dos fatores ecológicos e ambientais que influenciam a estrutura das florestas, ou para o desenvolvimento de modelos de crescimento e produção. Ela foi aplicada a uma ampla variedade de situações e tipos de floresta na América do Norte (Bailey e Dell, 1973; Schreuder e Swank, 1974; Ek et al., 1975; Little, 1983; Burk e Burkhart, 1984; Knoebel e Burkhart, 1991; Zhang et al., 2001; Zhang et al., 2003; Rennolls e Wang, 2005), na Europa (Maltamo et al., 2000; Nanos e Montero, 2002; Nord-Larsen e Cao, 2006; Westphal et al., 2006; Palahí et al., 2007; Podlaski, 2008), na África (Nanang, 1998; Maltamo et al., 2000; Bonou et al., 2009; Kakaï et al., 2011), na Ásia (Wang et al., 2009; Lai et al., 2013), na Oceania (Coomes e Allen, 2007) e na América do Sul (Couto, 1980; Campos e Turnbull, 1981; Silva, 1986; Batista, 1989; Neeff et al., 2003; Schneider et al., 2008; Machado et al., 2009; Leite et al., 2010).

Ajuste e Aplicação das Famílias Estocásticas

Numa primeira etapa do estudo da estrutura das florestas, por meio do ajuste de famílias estocásticas à distribuição do DAP das árvores, grande esforço foi dispendido, discutindo-se a melhor técnica de ajuste, isto é, de estimação dos seus parâmetros. A família Weibull foi alvo de vários estudos neste sentido. A literatura técnica florestal se baseou em grande medida na literatura de engenharia de materiais que já vinha sendo desenvolvida nas décadas de 1960 e 1970. A partir do trabalho de divulgação da família pelo próprio Waloddi Weibull (Weibull, 1951), vários trabalhos focaram nas técnicas e métodos de ajuste (Menon, 1963; Cohen, 1965; Harter e Moore, 1965; Bain e Antle, 1967; Dubey, 1967; Thoman et al., 1969; Johnson e Kotz, 1970; Wingo, 1972; Wingo, 1973; Berger e Lawrence, 1974; Gross, 1977).

Estes trabalhos foram, numa boa medida, repetidos e adaptados às condições particulares do ajuste das distribuições de diâmetro dos povoamentos florestais (Bailey e Dell, 1973; Schreuder e Swank, 1974; Ek et al., 1975; Garcia, 1981; Zutter et al., 1982; Burk e Newberry, 1984; Krug et al., 1984; Shiver, 1985; Zarnoch e Dell, 1985; van Deusen, 1986; Batista, 1989). Mas, de certo modo, a questão do ajuste da família Weibull ainda é uma questão corrente, havendo trabalhos mais recentes voltados a este problema (Zhang et al., 2003; Cao, 2004; Podlaski, 2008; Schneider et al., 2008; Leite et al., 2010; Taubert et al., 2013).

Concomitantemente à aplicação das várias família estocástica, vários aspectos práticos da utilização da distribuição do DAP foram sendo desenvolvidos. Os principais desenvolvimentos se relacionam à associação da distribuição do DAP com: (1) as predições da produção florestal (Osborne e Schumacher, 1935; Campos e Turnbull, 1981; Burk e Burkhart, 1984; Hafley e Buford, 1985; van Deusen, 1986), (2) a modelagem do crescimento das árvores e da floresta (Meyer e Stevenson, 1943; Nelson, 1964; Bailey, 1980; Burkhart et al., 1981; Clutter et al., 1983; Qin et al., 2007; Wang et al., 2009; Nascimento et al., 2012), (3) a modelagem da mortalidade e sobrevivência (Cao, 1997; Tang et al., 1997; Muller-Landau et al., 2006; Qin et al., 2007; Coomes e Allen, 2007; Wang et al., 2009; Leiva et al., 2012; Lai et al., 2013) e (4) a distribuição da altura das árvores para compor um modelo bivariado da estrutura da floresta (Hafley e Schreuder, 1977; Schreuder e Hafley, 1977; Hafley e Buford, 1985; Knoebel e Burkhart, 1991).

Com o objetivo de obter não só a predição da produção total da floresta, mas também a produção por classe de DAP para poder ser realizar o sortimento da produção para diferentes usos, uma série de técnicas foi desenvolvida para se incorporar a distribuição de DAP aos modelos de crescimento e produção. Já em 1983, Clutter e colaboradores (1983), no seu livro clássico *Timber Management*, descrevem como procedimento comum os métodos de “predição implícita da produção” (*implicit prediction of yield*) baseados na predição dos parâmetros de uma

família estocástica que permite recompor a distribuição de DAP e, a partir dela, a produção madeireira para diferentes usos. Dois métodos básicos são descritos. O método da *predição dos parâmetros* associa as estimativas dos parâmetros em cada povoamento aos atributos agregados destes povoamentos, a idade inclusive, geralmente num modelo linear. Utilizando o modelo linear, os parâmetros do povoamento podem ser preditos a partir de atributos com densidade de estade, área basal, DAP médio, etc. O segundo método consiste em associar alguns percentis da distribuição de DAP aos atributos agregados dos povoamentos, novamente nos modelos lineares. Dessa forma, os percentis são *preditos* e, por meio do método de estimação por percentis, os parâmetros são obtidos, por isto, este método ficou conhecido como método da *recuperação dos parâmetros*.

Knoebel e Burkhart (1991) compararam estes dois métodos com o método que utiliza como modelo uma a família beta especial bivariada (*Johnson's bivariate S_b distributions*) com dois pontos no tempo, de modo que a distribuição de DAP no futuro é determinada a partir de uma distribuição inicial. Mas, várias outras formas de obtenção da distribuição do DAP foram desenvolvidas. Cao (2004) comparou seis métodos diferentes de determinação dos parâmetros da Weibull a partir de atributos dos povoamentos: (1) o método da predição dos parâmetros; (2) o método da estimação dos momentos da distribuição; (3) o método da recuperação dos parâmetros pelos percentis estimados; (4) método híbrido, que utiliza o método dos momentos para o parâmetro de locação e o de percentis para os parâmetros de escala e forma; (5) o método do estimador de máxima verossimilhança, em que as funções lineares são incluídas na função de log-verossimilhança e seus parâmetros estimados minimizando essa função; (6) o método da “regressão” da função de distribuição, em que os parâmetros das funções lineares são estimados minimizando o quadrado da diferença entre as funções de distribuição observada e esperada. Ele concluiu que os métodos (5) e (6) se mostraram superiores aos demais.

Outros Métodos e Aplicações da Distribuição de DAP

A aplicação da distribuição de DAP na Mensuração e no Manejo Florestal possui um grande número de possibilidades e muitos métodos novos de aplicação estão sempre em desenvolvimento, tornando imensa a literatura técnica dedicada a este tema. A título de ilustração, seguem-se alguns exemplos mais recentes.

A explicitação da família estocástica para projeção da distribuição de DAP ao longo do tempo não é uma exigência incontornável. Foram sugeridos alguns métodos não paramétricos para projeção da distribuição de DAP que não requerem a definição explícita de uma família estocástica. Exemplos são o método da predição dos percentis no tempo, seguida da recomposição direta da distribuição do DAP (Maltamo et al., 2000) e o método da integral de Lebesgue-Stieltjes

que recompõe a distribuição de DAP a partir das funções de sobrevivência e de crescimento das árvores individuais (Tang et al., 1997). Por outro lado, Lane e colaboradores (2010) estudaram um procedimento de partição recursiva designado por *functional regression tree method* (FRT) que permite a representação da distribuição de DAP sem a que uma família estocástica seja especificada.

Um requisito importante para utilização realista da distribuição de DAP na medição ou predição da produção florestal é que a distribuição de DAP gerada pelo método seja compatível com a densidade de estande (árvores por hectare) e área basal conhecidas. Kangas e Maltamo (2000) utilizaram o método amostral da “estimação calibrada” (*calibration estimation*) para obter essa compatibilidade na escala de povoamentos particulares.

Se uma única distribuição não resulta na qualidade desejada para a resrepresentação da distribuição de DAP, pode-se utilizar uma mistura de um número finito de distribuições (*finite mixture of distributions*), não necessariamente da mesma família estocástica, para a modelagem de distribuições de diâmetro muito complexas, como aquelas de povoamentos dissetâneos (Zhang et al., 2001).

A tabela de povoamento é um instrumento de representação de povoamentos em uso na Mensuração e Manejo Florestal há muito tempo. Por exemplo, Graves (1906) já descrevia a sua elaboração e o seu uso na quantificação da produção de um povoamento, enquanto que Chapman (1921, 1924) apresenta o seu uso para determinação do crescimento das árvores. O método da “projeção da tabela de povoamento” já é um método clássico para determinação do crescimento de povoamentos. Mais recentemente, a projeção da tabela de povoamento foi recuperada como método de projeção da distribuição de diâmetros ao longo do tempo, por meio de algoritmos mais sofisticados de projeção do crescimento (Nepal e Somers, 1992; Cao e Baldwin, 1999; Trincado V. et al., 2003).

A escala de observação da distribuição dos diâmetros das árvores é a escala das parcelas ou unidades amostrais, mas a escala de manejo da florestal é normalmente o povoamento florestal. Considerando tal discrepância de escala entre os modelos e o manejo, Borders e colaboradores (2008) discutiram a questão do escalonamento das distribuições da escala de parcela para a escala de povoamento, considerando a variabilidade interna natural dos povoamentos.

Uma outra abordagem é a utilização dos modelos de efeitos mistos, que combina efeitos fixos e efeitos aleatórios, para modelagem da distribuição de diâmetros segundo as diferentes escalas em que a estrutura da floresta pode ser considerada. Os modelos de efeitos mistos são capazes de representar a estrutura hierárquica dos dados, que usualmente são compostos pela medida do DAP das árvores agregadas em parcelas (ou outras formas de unidade amostral), as quais foram locadas em povoamentos florestais, que, por sua vez, compõem plantações ou florestas, as quais podem ser consideradas como componentes de toda uma região florestal. Essas diferentes escalas podem ser consideradas como efeitos aleatórios

no modelo de distribuição de diâmetros, fazendo o valor dos parâmetros variar aleatoriamente. As aplicações de tal abordagem se iniciaram nos anos 2000, como por exemplo, os trabalhos de Robinson (2004), de Fortin *et al.* (2006), de Mehtälö *et al.* (2008), de Siipilehto (2009), de Stankova e Diéguez-Aranda (2010) e de Juntunen (2010).

Apesar da grande quantidade de trabalhos discutindo o desempenho de diferentes famílias estocástica e comparando os vários métodos de ajuste dessas famílias, aparentemente a palavra final ainda não foi dada neste tópico e, talvez, nunca seja dada. Taubert e colaboradores (2013) consideram que o ajuste de famílias estocástica para representar a estrutura de tamanho das árvores nas floresta nativas ainda é um desafio na Ecologia. Eles estudaram o ajuste das famílias da série de potência, exponencial e Weibull, comparando o método de ajuste pela máxima verossimilhança (MLE) com dois métodos modificados de MLE com correções para amplitude de classe de diâmetro e para erros de medição. Poudel e Cao (2013) discutiram diferentes métodos de ajuste da Weibull, considerando tanto o método da predição dos parâmetros como o método da recuperação dos parâmetros. Eles compararam o método que estima os parâmetros pelo método SUR (*seemingly unrelated regression*) com o método CDFR (*cumulative distribution function regression* — regressão da função de distribuição acumulada). O método CDFR se mostrou superior ao SUR segundo três dos quatro critérios de comparação utilizados.

10.3 Problemas na Modelagem de Distribuições

Pelo que foi apresentado sobre a modelagem da distribuição de diâmetros de florestas e povoamentos florestais, pode-se concluir que uma grande quantidade de esforço intelectual teórico e prático foi dispendido no estudo deste tema. No entanto, longe de se ter chegado a uma conclusão ou fechamento do tema, inúmeras possibilidades permanecem abertas. Ao invés de se chegar a procedimentos estáveis e estabelecidos, a grande quantidade de pesquisa publicada parece indicar exatamente o contrário, isto é, a ausência de métodos e procedimentos padrão.

A principal razão para inexistência de métodos e procedimentos padrão e para a existência de uma profusão de métodos alternativos possíveis é a ausência da aceitação de critérios tanto de ordem teórica quanto de ordem prática que sejam tomados como referência para discriminar, comparar e rejeitar ou aceitar os métodos propostos. A ausência de critérios de aceitação geral, por sua vez, parece resultar de uma filosofia em que qualquer forma de inferência é válida, desde que gere o resultado desejável. Infelizmente, nunca fica nem mesmo razoavelmente claro o que é um resultado desejável, pois isto parece variar com as circunstâncias

particulares de cada problema.

É importante, portanto, discutir-se alguns dos principais problemas da modelagem da estrutura das florestas que decorrem da carência de critérios teóricos e práticos de ampla aceitação entre os pesquisadores e profissionais florestais. O primeiro aspecto é a própria questão dos critérios de comparação de modelos e métodos de ajuste.

Cr terios para Comparação de Modelos e Métodos de Ajuste

No capítulo 9 (seção 9.3), foi apresentado um quadro conceitual para a seleção de modelos. Inicialmente, definiu-se o modelo operante, que é o modelo subjacente ao processo estocástico sendo modelado, seguindo-se a definição das famílias candidatas, que são modelos propostos para modelar o processo estocástico em questão. No caso da estrutura de florestas, o processo estocástico a ser modelado é a distribuição de diâmetros das árvores numa dada floresta ou povoamento florestal. Obviamente, o modelo operante é, e sempre será, desconhecido, enquanto os modelos candidatas são as diversas famílias estocásticas consideradas.

Como foi visto, o ideal seria poder discriminar as famílias candidatas em termos da sua discrepância de aproximação. Na prática, contudo, não é possível comparar as famílias, mas somente as distribuições ajustadas aos dados disponíveis, de modo que a discrepância de aproximação (das famílias candidatas) se confunde com a discrepância de estimação (das distribuições candidatas ajustadas). Qualquer critério de comparação e seleção de modelos é, portanto, uma discrepância global, que combina discrepância de aproximação e discrepância de estimação.

Por outro lado, os critérios de seleção de modelos são geralmente coincidentes com os critérios de ajuste de modelos, isto é, os critérios para estimação dos parâmetros dos modelos. Por exemplo, a log-verossimilhança e o critério de informação de Akaike (AIC) se fundamentam na ideia de que a log-verossimilhança é uma indicação adequada do contraste entre modelo e os dados observados, de forma que quanto maior a log-verossimilhança que um dado modelo candidato apresenta para uma amostra, melhor o seu ajuste aos dados dessa amostra. Mas, o método de estimação dos parâmetros por máxima verossimilhança se fundamenta nessa mesma ideia, pois ele gera as estimativas que maximizam a função de log-verossimilhança. A consequência é que as estimativas de máxima verossimilhança (MLE) sempre resultarão no melhor ajuste quando o critério de comparação dos métodos de ajuste for a log-verossimilhança ou o AIC.

Outro exemplo é apresentado por Cao (2004) no trabalho em que comparou seis métodos de ajuste da distribuição Weibull a dados da distribuição de diâmetros e sua relação com os atributos dos povoamentos. Um dos métodos estimava os parâmetros de relações lineares entre os parâmetros da Weibull e atributos dos

povoamentos minimizando o quadrado da diferença entre o valor da função de distribuição observada e esperada:

$$\sum_{i=1}^n \frac{\sum_{j=1}^{m_i} [F(x_{ij}) - \widehat{F}(x_{ij})]^2}{m_i},$$

onde i é o índice que indica as parcelas ($i = 1, 2, \dots, n$), j é o índice que indica as árvores dentro da parcela i ($j = 1, 2, \dots, m_i$), x_{ij} é o DAP da árvores j da parcela i , $F(\cdot)$ é o valor da função de distribuição observada e $\widehat{F}(\cdot)$ é o valor da função de distribuição esperada.

Este método de ajuste se mostrou como o melhor método quando o critério de comparação foi o teste de Kolmogorov-Smirnov. Este resultado é natural, uma vez que o teste de Kolmogorov-Smirnov usa como medida de discrepância a maior diferença absoluta entre os valores observados e valores esperados da função de distribuição. Dessa forma, a discrepância de Kolmogorov-Smirnov em cada parcela i é dada por:

$$D_{KS,i} = \max_j |F(x_{ij}) - \widehat{F}(x_{ij})|.$$

Da mesma forma, neste trabalho de Cao (2004) o critério de informação de Akaike (AIC) apontou o método da máxima verossimilhança como o melhor método de ajuste. Desta forma, quanto um dado critério de seleção de modelos está associado a um método de estimação dos parâmetros, o critério sempre indicará o método a ele associado como o “melhor” método de ajuste.

Uma outra possibilidade, é incorporar ao critério de seleção aspectos de ordem prática, de modo que a relevância do critério não seja meramente estatística. Reynolds e colaboradores (1988) propuseram considerar a utilização que será dada à distribuição de diâmetros após o seu ajuste como componente de um *índice de erro de predição* para comparação dos modelos. Estes autores fizeram a seguinte consideração: se a distribuição de diâmetros for utilizada na predição da produção de madeira de um povoamento, a relevância do erro de predição da frequência de árvores numa dada classe de diâmetro dependerá da classe em questão. As consequência do erro de predição será menor numa classe de DAP de 5 a 7 cm, que numa classe de 20 a 25 cm, pois, nessa última o volume de madeira de cada árvore individualmente é maior. Neste caso, o erro de predição da frequência nas classes de DAP deve ser ponderado pelo volume das árvores individuais. Generalizado, o índice de erro proposto por estes autores deve ser calculado para as parcelas individualmente e tem a seguinte forma:

$$e = N \sum_{j=1}^k \left| \int_{I_j} w(x) d\widehat{F}(x) - \int_{I_j} w(x) dF^*(x) \right| \quad (10.1)$$

na qual N é o número de árvores na parcela (ha^{-1}), j é o índice das classes de DAP ($j = 1, 2, \dots, k$), I_j denota o intervalo da classe de DAP j , $w(x)$ é a função de ponderação, x é o DAP como variável contínua, $\hat{F}(x)$ é a função de distribuição esperada, ou seja, aquela que é dada pela distribuição ajustada, e $F^*(x)$ é a função de distribuição empírica fornecida pelos dados. A expressão da função de distribuição empírica é obtida por

$$\int_{I_j} w(x) dF^*(x) = \sum_{i \in I_j} \frac{w(x_i)}{n},$$

onde x_i ($i = 1, 2, \dots, n$) são os DAP das n árvores observadas na parcela.

Como o índice de erro é baseado na diferença absoluta, uma sobrepredição numa dada classe de DAP não será corrigida por uma subpredição em outra classe. Bons índices de erro indicam uma predição adequada em todas as classes de DAP, dando maior relevância àquelas classes em que a função de ponderação ($w(x)$) é maior. Os autores também chamam a atenção que se a função de ponderação for o volume das árvores e se for utilizada uma única classe de DAP ($k = 1$), o índice mostrará o erro de predição do volume total da parcela.

Este índice de erro tem a vantagem de indicar os erros de predição da distribuição de DAP em termos de um atributo mensurável das parcelas, como o volume por unidade de área ($m^3 ha^{-1}$). Por outro lado, o fato de utilizar a diferença absoluta a interpretação se torna mais complexa. Por exemplo, se o erro de predição é expresso em termos relativos em relação ao volume da parcela, um erro de 10% não significa que o erro efetivo na predição do volume da parcela foi de 10%, exceto se somente uma classe de DAP for utilizada.

Para permitir uma comparação formal entre os modelos, Reynolds *et al.* (1988) propuseram o uso de um teste t para comparação dos índices de erro entre diferentes métodos de ajuste da distribuição de diâmetro, tomando o erro de predição em cada parcela como uma observação. O uso de um teste de comparação de médias, contudo, reduz a comparação entre os métodos de ajuste à comparação do comportamento médio dos métodos. Por outro lado, a aplicação do teste t requer a premissa de que os índices de erro de predição (e) têm distribuição Gaussiana.

Aparentemente, o índice de erro de predição teve uma boa aceitação na pesquisa quantitativa florestal. Uma busca de referência citada (*cited reference search*) na *Web of Science* em 23/08/2014, revelou que o trabalho de Reynolds *et al.* (1988) foi citado em 87 trabalhos de pesquisas em importantes periódicos científicos florestais, como: *Forest Science*, *Canadian Journal of Forest Research*, *Forest Ecology and Management*, *Forestry*, *Southern Journal of Applied Forestry*, *Annals of Forest Science*, *Silva Fennica*, *Scandinavian Journal of Forest Research*, e *European Journal of Forest Research*. A boa aceitação do índice de erro de predição não é, contudo, uma aceitação integral ou geral. Em

muitos trabalhos que o utilizam, ele é utilizado juntamente com outros critérios de comparação de modelos e métodos de ajuste.

A literatura científica mantém em aberto a questão que necessita ser respondida para a boa aplicação prática da Mensuração Florestal:

Qual o critério apropriado para comparar e selecionar as famílias candidatas a modelar uma dada distribuição de diâmetros observada e para comparar os diferentes métodos de ajuste e estimação dos parâmetros?

O capítulo 9 mostra que nossa resposta para essa questão é: a log-verossimilhança negativa (LVN) e, quando se deseja penalizar os modelos segundo o seu número de parâmetros, o critério de informação de Akaike (AIC). A justificativa para essa resposta é a abordagem de inferência estatística baseada no Axioma da Verossimilhança, conjugada com a solidez da fundamentação teórica tanto dessa inferência, quanto no método de estimação por máxima verossimilhança.

Estabelecido este critério, é óbvio que o método de referência para ajuste das famílias estocásticas é o método da máxima verossimilhança. Consequentemente, se todas as famílias candidatas a modelar uma dada distribuição de diâmetros são ajustadas por máxima verossimilhança, isto implica que o desempenho de todas as famílias candidatas foi maximizado perante os critérios de seleção. Logo, os critérios da LVN e do AIC constituem o melhor que se pode obter em termos de medida da discrepância de aproximação para comparação das famílias candidatas.

Os critérios LVN e AIC devem se constituir, portanto, nos critérios de referência no estudo das distribuições de DAP de florestas e povoamentos florestais, e o método da máxima verossimilhança como o método de ajuste ou estimação de referência. Outros critérios de seleção ou métodos de ajuste podem ser utilizados, desde que se forneçam justificativas de ordem teórica e prática para estes novos critérios e métodos em questões particulares. É importante ressaltar, contudo, que tais justificativas, tanto teóricas quanto práticas, não são de ordem estatística, nem muito menos de ordem especulativa. Elas devem ter origem nas áreas de aplicação, que no caso da modelagem da estrutura das florestas, são a Mensuração Florestal e o Manejo Florestal.

Problemas dos Testes de Aderência

Reynolds e colaboradores (1988) argumentaram que os “testes de aderência” (*goodness-of-fit tests*) são frequentemente inapropriados para a seleção de modelos para representar a distribuição de diâmetro de florestas e povoamentos florestais e propuseram o índice de erro de predição. A utilização dos testes de aderência, contudo, segue sendo talvez o critério mais utilizado nos trabalhos de

pesquisa sobre a modelagem da distribuição de diâmetros. Mas, os testes de aderência possuem problemas sérios tanto no que se refere à sua aplicação, quanto no que se refere à sua estrutura lógica.

Problemas na Aplicação dos Testes de Aderência

No que se refere a aplicação, um dos problemas é a definição de classes para agregação das observações individuais. Os testes de aderência são calculados, na maioria dos casos, com base na frequência observada em classes. No caso da distribuição de diâmetros, isto implica que as observações dos DAP das árvores individuais devem ser agregadas em classes de diâmetro para que o teste possa ser realizado. A definição das classes de diâmetro sempre tem um aspecto arbitrário, ainda que frequentemente haja um grande consenso entre os técnicos das regiões florestais particulares sobre as classes de diâmetro que devem ser utilizadas para se estudar a estrutura das florestas.

Outro aspecto problemático, é que alguns testes de aderência requerem que uma certa proporção das classes deve ter uma frequência mínima. Por exemplo, uma regra comumente utilizada para aplicação do teste de Qui-quadrado exige que no mínimo 80% das classes tenha frequência esperada igual ou superior a 5. É comum que as classes tenha que ser reorganizadas para atender essas exigências.

Talvez o aspecto mais inconveniente para aplicação dos testes de aderência é que eles não permitem a comparação direta entre dois modelos, uma vez que eles se baseiam no confronto entre a frequência observada e a frequência esperada segundo um único modelo. No caso de dois ou mais modelos se mostrarem *não significativos* num dado teste de aderência, nenhuma informação é fornecida sobre o desempenho comparativo sobre a representação da distribuição de diâmetros. Não há como se concluir qual dos modelos é o melhor.

Problemas Teóricos dos Testes de Aderência

Nessa situação, alguns pesquisadores tomam o valor numérico do valor-p (*p-value*) como medida do desempenho dos modelos, interpretando que o modelo com o maior valor-p seria aquele com “melhor aderência aos dados”. Este tipo de raciocínio, contudo, é totalmente equivocado. A única afirmação que a Teoria Estatística Clássica nos permite fazer sobre os valores-p de testes de hipóteses distintos é que, sendo a hipótese nula verdadeira em todos os testes, os valores-p observados representam números aleatórios que seguem a distribuição amostral da estatística utilizada no teste. É logicamente errado interpretar que os valores-p possam indicar que, num dado teste, a hipótese nula seja mais verdadeira do que em outro. O principal problema teórico dos testes de aderência, contudo, está na sua própria estrutura lógica.

O conceito de teste de aderência surgiu muito antes que o conceito de teste

estatístico de hipótese fosse estabelecido. Karl Pearson foi um dos primeiros a indicar a necessidade de se comparar as frequências observadas na distribuição de uma medida biológico ou demográfica com as frequências esperadas segundo uma distribuição teórica que se colocasse como candidata a representar o fenômeno estudado. Sua sugestão para uma “estatística de aderência” (*goodness-of-fit statistics*) foi a estatística Qui-quadrado:

$$\chi^2 = \sum_{i=1}^k \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i},$$

na qual o índice i indica as classes em que as frequências são computadas ($i = 1, 2, \dots, k$), f_i é a frequência observada na classe i e \hat{f}_i é a frequência esperada nessa mesma classe. O fato de que a estatística se baseia na soma do *quadrado* das diferenças entre as frequências observadas e esperadas evita que diferenças negativas e positivas se compensem, avaliando a “aderência” da distribuição teórica junto à distribuição observada.

Ao se definir uma estatística de discrepância entre a distribuição empírica (distribuição observada) e a distribuição teórica, seja ela a estatística Qui-quadrado ou qualquer outra, o problema é colocado da seguinte forma:

A distribuição teórica “adere” à distribuição empírica?

Logo, o teste de aderência realiza uma comparação absoluta entre a distribuição empírica e *uma* distribuição teórica. Ele não visa a comparação do desempenho de *dois ou mais* modelos teóricos.

Pode se argumentar que, sendo a estatística de aderência (Qui-quadrado ou outra) uma medida da *discrepância* entre observado e esperado, o modelo com *menor* valor da estatística indica uma *melhor aderência* aos valores observados. Este argumento é correto, mas também é problemático. A estatística de aderência é uma variável aleatória e, conseqüentemente, ela requer um quadro de referência estocástico para comparar os seus valores numéricos. Como determinar quando a diferença numérica entre duas estatísticas de aderência observadas para dois modelos candidatos é de fato relevante ou mero resultado aleatório?

Na Inferência Clássica, o quadro de referência estocástico seria composto não apenas da distribuição amostral da estatística de aderência (como a distribuição de Qui-quadrado), mas da distribuição amostral das *diferenças* entre as estatísticas de aderência (a distribuição das diferenças entre estatísticas Qui-quadrado). Este quadro de referência teria premissas muito mais exigentes e complexas que as premissas dos testes de aderência conhecidos. Por exemplo, uma dada diferença entre duas estatísticas de aderência pode se mostrar irrelevante em três situações completamente diferentes. Uma delas é o caso em que as hipóteses nulas de cada

estatística são verdadeiras, ou seja, ambos modelos não apropriados para representar a distribuição observada. Outra possibilidade é o caso em que ambas hipóteses nulas são falsas, isto é os dois modelos são inapropriados, mas estão próximos entre si. A terceira situação é aquela em que uma hipótese nula é verdadeira e a outra é falsa, mas os modelos estão próximos ao limite de rejeição das respectivas hipóteses nulas. Ou seja um modelo é apropriado mas o outro não. Mas, não é possível distinguir entre essas três situações na prática.

A estrutura teórica dos testes de aderências tradicionais, contudo, também tem problemas de inconsistência lógica. As *estatísticas de aderência* representam em si uma forma coerente de avaliação absoluta da adequação de uma distribuição teórica. Mas, os *teste de aderência* são problemáticos como quadros de referência estocásticos propostos pela Inferência Clássica para avaliar essas estatísticas.

A Teoria de Neyman-Pearson estabelece que, no teste estatístico de uma hipótese, pode-se incorrer em dois erros: o erro de se rejeitar uma hipótese nula verdadeira (erro tipo I) ou o erro de não se rejeitar uma hipótese nula falsa (erro tipo II). A construção do teste de hipóteses segundo uma regra de decisão tem como objetivo manter o erro tipo I dentro de uma margem aceitável, usualmente de 5% ($\alpha = 0.05$). A margem de erro tipo II permanece desconhecida e, pior ainda, ela varia em cada situação em que o teste é aplicado, pois depende do tamanho da amostra disponível para o teste. A interpretação dos testes de aderência, segundo a Teoria de Neyman-Pearson, deixa o pesquisador numa situação muito problemática, pois quando ele *não rejeita* a hipótese nula e toma um dado modelo teórico como apropriado ele não tem idéia da margem de erro associada a essa decisão. Assim a decisão que realmente importa no caso da modelagem da distribuição do diâmetro das árvores, que é a escolha de um modelo para representar a distribuição, é sempre “*tomada no escuro*”.

A interpretação do teste de aderência pela Teoria Fisheriana dos testes de significância também é problemática. No caso do teste de significância, a interpretação é que os valores-p pequenos são força de evidência contra a hipótese nula. Assim, nos testes de aderência, interpreta-se que quanto *maior* o valor-p, maior a força de evidência *em favor* da hipótese nula e, conseqüentemente, em favor do modelo teórico sendo testado. Mas, Royall (1997) demonstrou que essa interpretação é logicamente incoerente. Afirmar que o valor-p é força de evidência contra a hipótese nula é o mesmo que afirmar que a ocorrência de um evento (a estatística com valor-p pequeno), do qual se desconhece a frequência, mas do qual uma hipótese (a hipótese nula) diz que ele é raro, é evidência contra essa mesma hipótese que afirma a raridade do evento (a hipótese nula). Hacking (1965) e Royall (1997) argumentaram convincentemente que a evidência estatística é sempre uma evidência relativa, isto é, evidência em favor de uma hipótese contra (*vis-à-vis*) outra hipótese, e que a sua manifestação quantitativa mais coerente é a razão de verossimilhança.

Os problemas apresentados nos parecem suficientes para mostrar que o critério mais utilizado para seleção de modelos na modelagem da distribuição de diâmetro das árvores, os testes de aderência, está longe de ser uma abordagem teoricamente sólida e praticamente segura.

Não existe uma aceitação geral de critérios substitutos. A mensagem deste documento é que os critérios só terão validade universal, isto é, para qualquer situação de modelagem de distribuição de variáveis medidas nas Ciências Naturais, se forem baseados no Axioma da Verossimilhança. Consequentemente, os critérios candidatos a uso geral são a log-verossimilhança negativa e o critério de informação de Akaike.

Nas situações particulares de cada área de conhecimento ou de aplicação, podem existir critérios válidos e úteis, cuja a fundamentação não baseia necessariamente em uma “Teoria de Inferência”, mas nas particularidades da área. No caso da modelagem de distribuição de diâmetros de árvores em florestas e povoamentos florestais, o índice de erro de predição proposto por Reynolds e colaboradores (1988) nos parece ser um destes critérios. A sua validade e utilidade resulta da aplicação dos modelos de distribuição de diâmetro na modelagem do crescimento e produção da floresta. Neste contexto, a definição da função de ponderação (função $w(x)$ na expressão 10.1) surge naturalmente da aplicação da distribuição de diâmetros no processo de modelagem.

Contudo, a utilidade do índice de erro não está na comparação do desempenho relativo dos modelos candidatos, mas na validação absoluta de um dado modelo de distribuição, de modo a assegurar que seu desempenho tenha a qualidade mínima suficiente esperada na escala do modelo de crescimento e produção do qual faz parte. Ou seja, a utilidade do índice de erro de predição está na validação de um dado modelo no contexto do modelo maior (de crescimento e produção) no qual ele será utilizado. Infelizmente, ainda não se tem valores de referência que permitam que essa utilização seja realizada de modo coerente.

10.4 Aplicação:

Estrutura de Tamanho de Floresta Nativa

No exemplo de aplicação que se segue, a inferência com base no Axioma da Verossimilhança é utilizada na construção de um modelo de distribuição de diâmetros de uma dada floresta. Este exemplo está longe de ser completo ou exaustivo, mas ilustra como a Inferência baseada em Verossimilhança estabelece um caminho coerente para a Biometria Florestal na resolução de um tipo de problema que é clássico na Mensuração Florestal.

A Situação de Aplicação

No *Exemplo 9.7*, tratou-se da estrutura de tamanho de uma floresta nativa do município de Bom Jardim, MA, observada por meio de um levantamento florestal. Nessa aplicação, a análise da estrutura de tamanho dessa floresta será desenvolvida com maior detalhamento que o apresentado no exemplo. O objetivo dessa aplicação é encontrar a melhor forma de representar a estrutura de tamanho das árvores por meio da modelagem da distribuição de DAP. Este objetivo sempre possui dois aspectos interdependentes: (1) qual é o melhor modelo candidato e (2) qual é a escala ou nível de agregação dos dados mais apropriada à modelagem?

Dados do Levantamento Florestal

O levantamento florestal na floresta nativa de Bom Jardim foi constituído por 87 parcelas em faixa com 10 m de largura e de comprimento variável (150, 250 e 500 m), tendo sido medidas e identificadas 11972 árvores. O levantamento utilizou como método silvométrico a amostragem estratificada, sendo que os estratos foram tomados segundo três situações topográficas distintas da floresta: encosta, platô grande e platô pequeno. A figura 10.1 mostra que a grande maioria das parcelas foram parcelas de 5000 m² (500 m de comprimento).

Em cada parcelas, as árvores tiveram suas CAP e alturas comerciais medidas e foram identificadas segundo a espécie para maioria dos indivíduos, sendo a CAP mínima para medição de 45 cm. O DAP das árvores foi obtido como medida secundária pela razão da CAP e a constante π . A figura 10.2 mostra os atributos médios tanto da árvores individualmente quanto das parcelas. Nota-se que não há uma grande variação entre as situações topográficas.

Tabela 10.1: Número de parcelas em faixa no levantamento florestal de floresta nativa em função da situação topográfica e do comprimento da parcela.

SITUAÇÃO TOPOGRÁFICA	COMPRIMENTO (m)			TOTAL
	115	250	500	
ENCOSTA	0	4	18	22
PLATÔ GRANDE	0	4	30	34
PLATÔ PEQUENO	1	8	22	31
TOTAL	1	16	70	87
TOTAL EM ÁREA (ha)	0.115	4	35	39.115

Tabela 10.2: Alguns atributos médios das árvores individuais e das parcelas em cada situação topográfica.

ATRIBUTOS	SITUAÇÃO TOPOGRÁFICA		
	ENCOSTA	PLATÔ GRANDE	PLATÔ PEQUENO
<i>Atributos das Árvores</i>			
Número de Árvores	3360	4728	3884
DAP médio (<i>cm</i>)	26	26	25
Desvio padrão: DAP (<i>cm</i>)	13	13	13
Amplitude: DAP (<i>cm</i>)	110	105	187
Altura comercial média (<i>m</i>)	7.1	8.0	7.7
<i>Atributos das Parcelas</i>			
Número de Parcelas	22	34	31
Densidade de estande média (<i>ha</i> ⁻¹)	341	301	306
Área basal média (<i>m</i> ² <i>ha</i> ⁻¹)	22	19	20
DAP médio quadrático (<i>cm</i>)	22	22	22
Número médio de espécies	45	44	44

Modelos Candidatos

Os modelos candidatos para representação da distribuição de diâmetros das árvores (d) foram as famílias exponencial, Weibull e gama. Essas três famílias podem ser consideradas modelos apropriados para distribuição de diâmetros em florestas nativas. A família exponencial representa a distribuição definida por De Liocourt em 1898 (2001) para as florestas dissetâneas “balanceadas”. Já as famílias Weibull e gama possuem maior flexibilidade na sua forma, sendo, portanto, apropriadas para modelar distribuição de diâmetros “não balanceadas”.

Nas três famílias, assumiu-se que o *parâmetro de localização* é conhecido e igual ao DAP mínimo de medição:

$$\alpha = d_{\text{mínimo de medição}} = \frac{45}{\pi} = 14.32394 \text{ cm},$$

de modo que a variável efetivamente modelada foi:

$$x = d - \alpha.$$

Para tornar mais claros os métodos de ajuste utilizados, é melhor apresentar as

três famílias candidatas na forma das suas respectivas funções de log-verossimilhança:

$$\text{Exponencial: } \mathbf{L}\{\lambda\} = n \ln(\lambda) - \lambda \sum_{i=1}^n x_i \quad (10.2)$$

$$\begin{aligned} \text{Weibull: } \mathbf{L}\{\beta, \gamma\} = n \ln(\gamma) - n \gamma \ln(\beta) + \\ + (\gamma - 1) \sum_{i=1}^n \ln(x_i) - \frac{1}{\beta^\gamma} \sum_{i=1}^n x_i^\gamma \end{aligned} \quad (10.3)$$

$$\begin{aligned} \text{Gama: } \mathbf{L}\{\lambda, \gamma\} = n \ln(\lambda) - n \ln(\Gamma(\gamma)) + (\gamma - 1) \ln(\lambda) + \\ + (\gamma - 1) \sum_{i=1}^n \ln(x_i) - \lambda \sum_{i=1}^n x_i \end{aligned} \quad (10.4)$$

onde λ é o parâmetro de taxa, β é o parâmetro de escala e γ é o parâmetro de forma.

Estrutura dos Dados e Modelos Candidatos

Os dados do levantamento florestal têm uma estrutura hierárquica composta de três escalas: (1) a escala do levantamento, na qual todas as árvores são agregadas como se fossem um única floresta, ignorando-se variações internas na estrutura da floresta, (2) a escala da situação topográfica, na qual se ignora a variação da estrutura da floresta dentro de cada situação topográfica, e (3) a escala das parcelas, que é de fato a escala em que as árvores compõem os conglomerados de observações amostrados em campo.

Cada família estocástica candidata (exponencial, Weibull e gama) é um único modelo de aproximação, independentemente de quantas distribuições são ajustadas em função da escala dos dados em que o ajuste é realizado. No caso da escala de levantamento, cada modelo de aproximação (família estocástica) é representado por uma única distribuição ajustada às 11972 árvores do levantamento. Os modelos de aproximação neste caso, representados pelas suas respectivas funções de log-verossimilhança, são:

$$\text{Exponencial: } \mathbf{L}\{\lambda\}; \quad p = 1$$

$$\text{Weibull: } \mathbf{L}\{\beta, \gamma\}; \quad p = 2$$

$$\text{Gama: } \mathbf{L}\{\lambda, \gamma\}; \quad p = 2$$

de modo que o modelo exponencial possui um único parâmetro ($p = 1$), enquanto os modelos Weibull e gama possuem dois parâmetros ($p = 2$).

Na escala topográfica, cada modelo de aproximação é composto por três distribuições ajustadas a cada uma das situações topográficas. Como o levantamento

utilizou a amostragem estratificada, cada situação topográfica representa uma amostragem independente das demais situações e, portanto, o ajuste dos modelos também pode ser realizado de modo independente em cada uma delas. Assim, os modelos candidatos são:

$$\begin{array}{lll} \text{Exponencial:} & \mathbf{L}\{\lambda_k\}; & p = 3 \\ \text{Weibull:} & \mathbf{L}\{\beta_k, \gamma_k\}; & p = 6 \\ \text{Gama:} & \mathbf{L}\{\lambda_k, \gamma_k\}; & p = 6 \end{array}$$

sendo $k = 1, 2, 3$ o índice que representa as três situações topográficas: encosta, platô grande e platô pequeno, respectivamente.

Por fim, na escala das parcelas, o ajuste dos modelos candidatos é realizado parcela-a-parcela. Novamente, o fato da localização das parcelas em campo ser independente torna possível o ajuste independente dos modelos em cada parcela. Neste caso, os modelos candidatos se tornam:

$$\begin{array}{lll} \text{Exponencial:} & \mathbf{L}\{\lambda_j\}; & p = 87 \\ \text{Weibull:} & \mathbf{L}\{\beta_j, \gamma_j\}; & p = 174 \\ \text{Gama:} & \mathbf{L}\{\lambda_j, \gamma_j\}; & p = 174 \end{array}$$

sendo $j = 1, 2, \dots, 87$ o índice que representa cada uma das parcelas do levantamento florestal.

Para se estimar os parâmetros, utiliza-se o método da máxima verossimilhança nos três modelos candidatos e nas três escalas de ajuste.

Seleção dos Modelos

A seleção dos modelos implica não só a escolha de família de aproximação mais apropriada, mas também na escolha da escala mais apropriada para representar a estrutura da floresta. Dessa forma, o critério de seleção dos modelos deve possibilitar a comparação dos modelos nas diferentes escalas. Como as observações, no nível de cada parcela, são independentes, a *log-verossimilhança* de cada parcela pode ser utilizada tanto para somar os valores a partir das parcelas, para a escala topográfica e para a escala de levantamento, quanto para desdobrar os valores no sentido inverso da maior escala (levantamento) para o menor (parcela). Contudo, é mais conveniente que a comparação seja realizada com os valores da *log-verossimilhança negativa* (LVN), pois estes valores se tornam positivos e a sua interpretação é análoga à interpretação do critério de informação de Akaike (AIC), ou seja, quanto menor o valor, melhor a qualidade do ajuste do modelo.

O critério de informação de Akaike (AIC) também pode ser utilizado para comparação dos modelos, mas como ele penaliza o ajuste em função do número

de parâmetros do modelo, ele só pode ser calculado para a escala em que o modelo foi ajustado e para escala maiores, pela agregação dos valores. Por exemplo, os modelos ajustados na escala topográfica podem ter os valores de AIC calculados nessa escala e na escala de levantamento, por meio da soma dos AIC de cada situação topográfica. Mas, neste caso, não é possível obter os valores do AIC na escala de parcela.

Espera-se que à medida que a escala dos dados se torne menor, a qualidade do ajuste melhore, isto é, os valores da log-verossimilhança negativa (LVN) e do AIC se tornem menores. Isto deve acontecer porque ao se reduzir a escala de ajuste, por exemplo, da escala de levantamento para escala topográfica, a variação na estrutura da floresta que existe entre as situações topográficas é incorporada ao modelo candidato. O mesmo acontece quanto se passa da escala topográfica para a escala de parcela. Contudo o número de parâmetros nos modelos aumenta marcadamente e este aumento deve ser considerado.

Resultados dos Ajustes dos Modelos

A estimativas de máxima verossimilhança (MLE) encontradas no ajuste na escala de levantamento e na escala topográfica são apresentadas na tabela 10.3, com os seus respectivos erros padrão. Os pequenos valores dos erros padrão relativos às MLE indica que as estimativas foram obtidas com boa precisão. Observa-se pequena variação nas MLE dos parâmetros análogos nas diferentes distribuições e entre os mesmos parâmetros nas diferentes situações topográficas.

Critérios de Seleção

A tabela 10.4 apresenta a qualidade de ajuste das três famílias estocásticas em cada escala dos dados em que os modelos foram ajustados, discriminando os valores em cada situação topográfica. Observa-se que quando se passa da escala de levantamento para escala topográfica a melhora na qualidade do ajuste é irrelevante para a família exponencial, enquanto que, para as famílias Weibull e gama, a melhora foi relevante na situação de encosta e platô pequeno, mas não na situação de platô grande. A passagem do ajuste da escala topográfica para escala de parcela mostra melhoras na qualidade do ajuste bem mais marcantes que no caso anterior. Isto sugere que existe certa diferenciação na estrutura da floresta entre as situações topográficas, mas essa diferenciação é detectada apenas pelas famílias Weibull e gama. A diferenciação entre parcelas dentro das situações topográficas se mostra, contudo, de magnitude muito maior.

No que se refere às famílias estocástica, a família exponencial mostrou desempenho sempre inferior às famílias Weibull e gama, o que é esperado, uma vez que essas duas últimas contemplam um parâmetro de forma que lhes dá uma grande

flexibilidade. O ajuste dos modelos na escala de parcela mostra, contudo, que a situação não é tão simples assim. A figura 10.1 mostra as diferenças dos AIC

Tabela 10.3: Estimativas de máxima verossimilhança (MLE) para os parâmetros das distribuições Weibull, gama e exponencial ajustadas nas diferentes escalas dos dados. Valores entre parênteses são os erros padrões das estimativas.

DISTRIB.	ESTIMATIVAS DE MÁXIMA VEROS. (MLE)		
	Forma γ	Escala β	Taxa λ
<i>Escala de Levantamento</i>			
Weibull	0.89150 (0.006274)	10.5058 (0.113226)	
Gamma	0.83619 (0.0093626)	—	0.0752382 (0.0011291)
Exponential	—	—	0.0899794 (0.0008224)
<i>Escala Topográfica</i>			
Weibull	_____		
Encosta	0.931634 (0.0123118)	10.9927 (0.214094)	— —
Platô grande	0.889710 (0.0100236)	10.5886 (0.181916)	— —
Platô Pequeno	0.862166 (0.0106174)	9.99278 (0.195516)	— —
Gama	_____		
Encosta	0.899364 (0.0191354)	—	0.0791570 (0.00221561)
Platô Grande	0.830326 (0.0147842)	—	0.0740873 (0.00177131)
Platô Pequeno	0.795953 (0.0155752)	—	0.0738024 (0.00196126)
Exponencial	_____		
Encosta	—	—	0.0880155 (0.001518412)
Platô Grande	—	—	0.0892268 (0.00129765)
Platô Pequeno	—	—	0.0927214 (0.00148778)

Tabela 10.4: Desempenho dos modelos candidatos para representar a distribuição de DAP da floresta nativa de Bom Jardim, MA, nas três situações topográficas da floresta e nas três escalas hierárquicas dos dados. O desempenho é apresentado em termos da log-verossimilhança negativa (LVN) e do critério de informação de Akaike (AIC).

DISTRIB.	NÚMERO DE PARÂMETROS	SITUAÇÃO TOPOGRÁFICA						TOTAL	
		ENCOSTA		PLATÔ GRANDE		PLATÔ PEQUENO		LVN	AIC
		LVN	AIC	LVN	AIC	LVN	AIC		
<i>Escala de Levantamento</i>									
Exponencial	1	11526	—	16154	—	13122	—	40803	81607
Weibull	2	11517	—	16096	—	13047	—	40661	81325
Gama	2	11519	—	16096	—	13053	—	40669	81341
<i>Escala Topográfica</i>									
Exponencial	3	11526	23053	16154	32309	13121	26244	40800	81606
Weibull	6	11511	23026	16096	32197	13042	26088	40649	81310
Gama	6	11513	23030	16096	32196	13048	26101	40657	81327
<i>Escala de Parcela</i>									
Exponencial	87	11486	23015	16123	32313	13029	26101	40638	81430
Weibull	174	11459	23006	16049	32234	12951	25987	40459	81227
Gama	174	11461	23009	16045	32227	12951	25986	40457	81222

(delta AIC) das famílias estocásticas duas-a-duas para cada uma das parcelas do levantamento.

Nota-se que, apesar da Weibull e gama se mostrarem melhores que a exponencial, num grande número de parcelas não há diferença relevante em termos de AIC. Na prática, das 87 parcelas do levantamento, em 52 delas não há diferença marcante entre a família exponencial e Weibull, e dessas parcelas, em 40 delas o

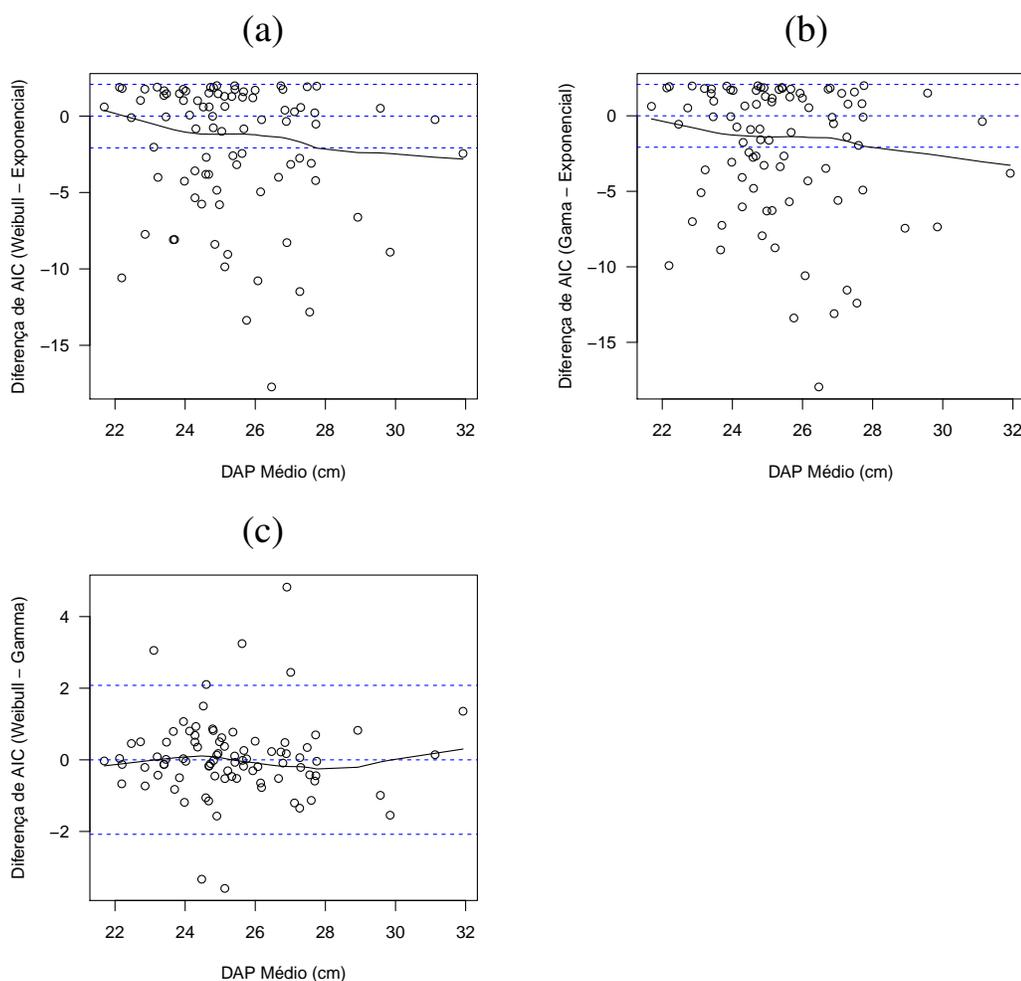


Figura 10.1: Diferenças do AIC grafadas contra o DAP médio das parcelas para as distribuições Weibull e exponencial (a), gama e exponencial (b) e Weibull e gama (c). As linhas horizontais dos gráficos indicam as posições de diferença de AIC iguais a $\ln(8)$, zero e $-\ln(8)$. As curvas em todos os gráficos são linhas de tendência de relação entre as variáveis encontradas por suavização.

AIC da exponencial chega a ser menor que o da Weibull, mas sem ultrapassar a diferença limite de $\log(8) = 2.0794$. Na comparação da exponencial com a gama estes números são: 53 parcelas sem diferenças relevantes e 37 em que exponencial tem AIC menor que a gama. Estes resultados indicam que “*lei de De Liocourt*” permanece como uma regra válida na maioria das parcelas do levantamento.

A comparação das famílias Weibull e gama (figura 10.1 c), ressalta a pouca diferença entre essas duas famílias estocásticas na modelagem da estrutura dessa floresta, com ligeira vantagem para família gama. O AIC indica superioridade da Weibull em apenas 2 parcelas e superioridade da gama de 4. Nas demais 81 parcelas a diferença de qualidade de ajuste destes modelos não se mostra relevante.

Perfis de Log-verossimilhança das MLE

Um aspecto importante para a avaliação dos modelos candidatos é a verificação se os parâmetros do modelo foram apropriadamente estimados pelo método da máxima verossimilhança. No caso das famílias Weibull e gama, em que não há solução analítica para as estimativas de máxima verossimilhança (MLE), essa verificação é particularmente importante.

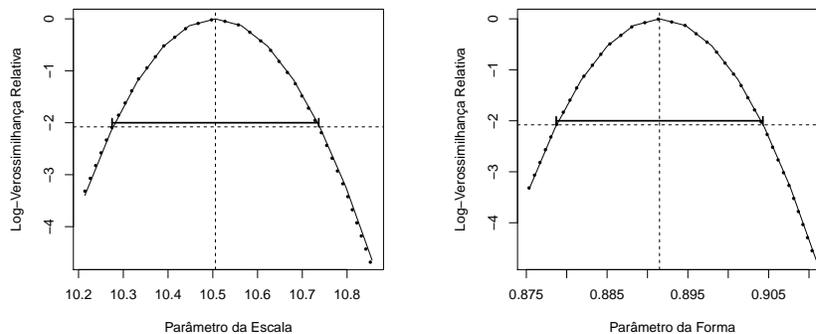
A figura 10.2 apresenta os perfis de log-verossimilhança para os parâmetros das três famílias estocásticas candidatas estimados na escala de levantamento. Como a base para estimação destes parâmetros são as 11972 árvores, os perfis atestam que o ajuste não só foi obtido com alta qualidade, mas que também a aproximação quadrática é perfeitamente válida.

Na escala topográfica, o número de árvores utilizados para ajuste das famílias estocástica também representa uma grande amostra, de modo que as estimativas de máxima verossimilhança obtidas (MLE) são de alta qualidade. A figura 10.3 apresenta, para fins de ilustração, os perfis de log-verossimilhança das MLE da Weibull em cada uma das situações topográficas. Nota-se que, nessa escala, a aproximação quadrática também se mostra adequada.

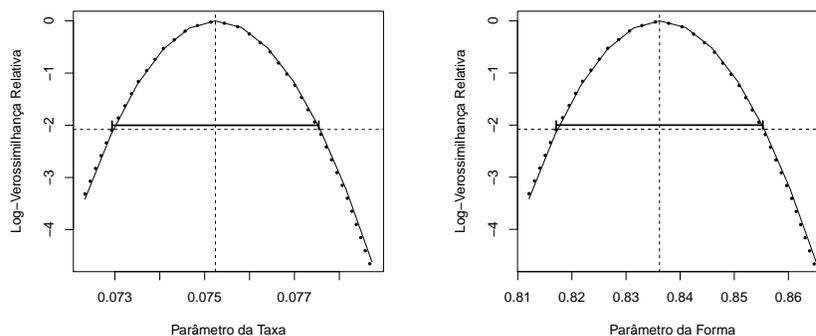
Na escala de parcela, embora possível, não é praticável se avaliar o perfil de log-verossimilhança das MLE para todas as 87 parcelas. Contudo, a qualidade das MLE é fortemente dependente do tamanho da amostra. Como o número de árvores nas parcelas varia de 62 (parcela 212, platô grande) a 169 (parcela 310, platô grande), a menor amostra para o ajuste das distribuições foi a parcela 212 com 62 árvores. Essa é uma indicação de que provavelmente o ajuste na escala de parcela resultou em boas MLE, pois 62 já constitui um tamanho de amostra bastante bom para o ajuste de distribuições estocásticas.

A título de ilustração, a figura 10.4 apresenta os perfis de log-verossimilhança para as MLE das distribuições ajustadas à parcela 212. Nota-se que neste caso, o perfil não é perfeitamente simétrico, de modo que a aproximação quadrática e o intervalo de verossimilhança aproximado estão ligeiramente deslocados para va-

Distribuição Weibull



Distribuição Gama



Distribuição Exponencial

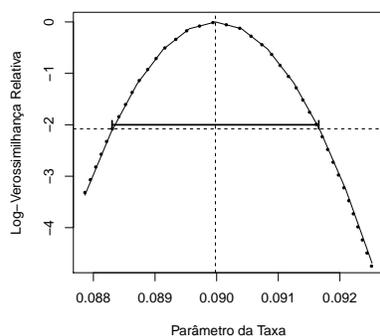


Figura 10.2: Perfis de log-verossimilhança relativa para as estimativas de máxima verossimilhança (MLE) dos parâmetros das distribuições Weibull, gama e exponencial ajustadas na escala de levantamento. A linha contínua é o perfil de log-verossimilhança relativa. A linha pontilhada é a aproximação quadrática para o perfil. O segmento de reta horizontal é o intervalo de log-verossimilhança segundo a aproximação quadrática. A linha tracejada vertical indica a posição da MLE. A linha tracejada horizontal indica o limite de $-\ln(8)$.

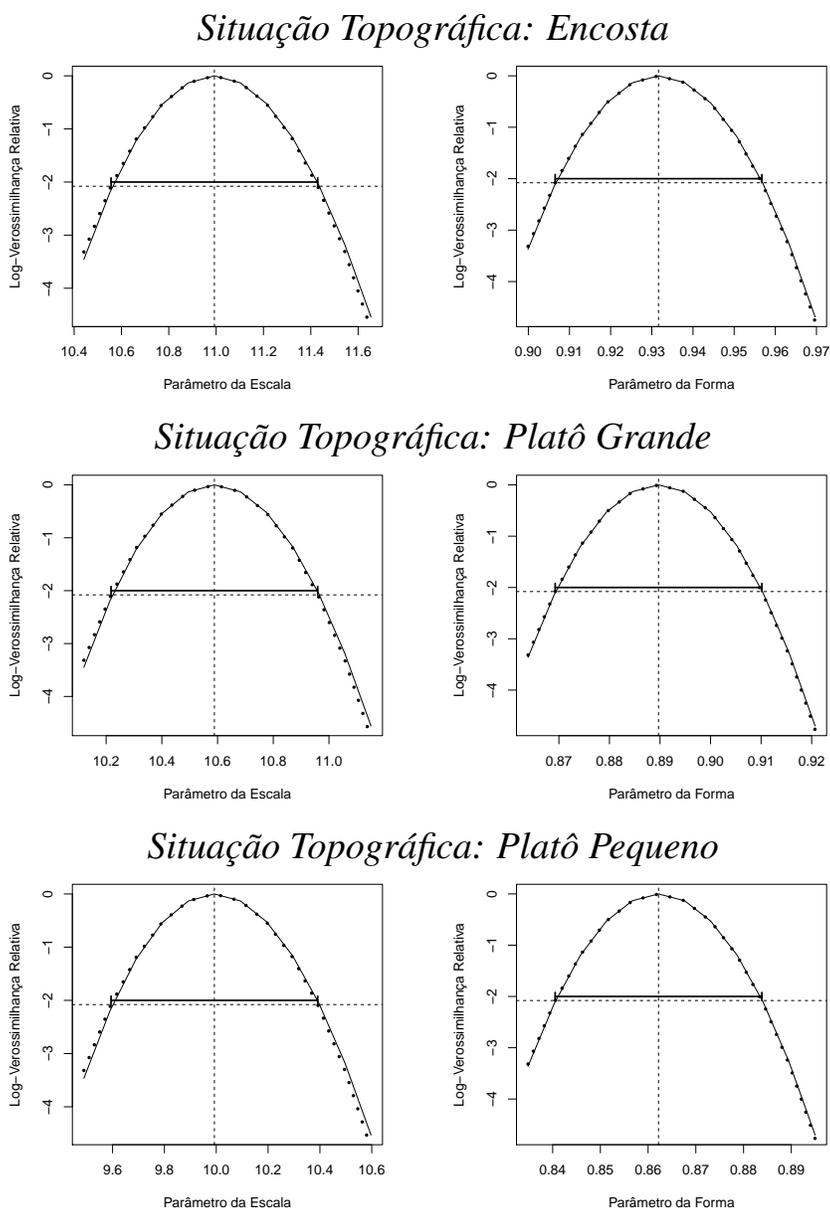


Figura 10.3: Perfis de log-verossimilhança relativa para as estimativas de máxima verossimilhança (MLE) dos parâmetros da família Weibull ajustada na escala topográfica. A linha contínua é o perfil de log-verossimilhança relativa. A linha pontilhada é a aproximação quadrática para o perfil. O segmento de reta horizontal é o intervalo de log-verossimilhança segundo a aproximação quadrática. A linha tracejada vertical indica a posição da MLE. A linha tracejada horizontal indica o limite de $-\ln(8)$.

lores menores da estimativa. Mas os perfis comprovam que as MLE obtidas são bastante estáveis, ainda que os seus intervalos de verossimilhança sejam relativamente grandes.

A parcela 212 é uma parcela em que não houve diferenças marcantes entre as três famílias candidatas e os perfis de log-verossimilhança ilustram este resultado. A figura 10.4 mostra que tanto na Weibull, quanto na gamma, o intervalo de verossimilhança do parâmetro da forma inclui o valor 1. Logo, a hipótese $\gamma = 1$ é uma hipótese plausível e, conseqüentemente, as distribuições Weibull e gama podem ser reduzidas à distribuição exponencial. Da mesma forma, os intervalos das MLE dos parâmetros de escala (Weibull) e taxa (gama e exponencial) se sobrepõem, mostrando que as estimativas para este parâmetro estão muito próximas.

Comparação Gráfica entre Distribuição Observada e Esperada

Os gráficos Quantil-Quantil possibilitam a comparação gráfica entre a distribuição observada e a distribuição esperada segundo cada modelo candidato. Ainda que não seja uma comparação formal, a comparação gráfica permite observar aspectos mais sutis da diferença entre distribuições que não podem ser observados nos critérios números de comparação (log-verossimilhança negativa e AIC). Contudo, nem sempre os gráficos Quantis-Quantis conseguem explicitar diferença marcantes indicadas pelos critérios numéricos.

A figura 10.5 apresenta os gráficos Quantis-Quantis das distribuições ajustadas na escala de levantamento. Estes gráficos explicitam a grande superioridade das distribuições Weibull e gama sobre a distribuição exponencial. Eles também permite verificar a superioridade da Weibull sobre a gama, que é uma diferença de ajuste mais sutil.

No caso do ajuste dos modelos na escala topográfica (figura 10.6), os gráficos Quantis-Quantis também exemplificam as diferenças de ajuste entre as distribuições. A inferioridade da exponencial em relação à Weibull e à gama também é ilustrada com clareza, mas as diferenças entre Weibull e gama nas três situações topográficas são de visualização mais difícil.

No caso do ajuste dos modelos na escala de parcela, a comparação gráfica se torna problemática devido ao grande número de parcelas. A figura 10.7 ilustra essa comparação para parcela 201 que é aquela com maior discrepância entre a exponencial e as distribuições Weibull e gama. A inferioridade da exponencial pode ser facilmente observada nos gráficos, mas eles também sugerem uma ligeira superioridade da Weibull sobre a gama, o que não é comprovado pelos critérios numéricos (LVN e AIC), pois a diferença entre essas distribuições nos critérios numéricos é desprezível (< 0.3).

Apesar de suas limitações, a análise gráfica por meio dos gráficos Quantil-Quantil é um instrumento importante para verificação do ajuste de distribuições

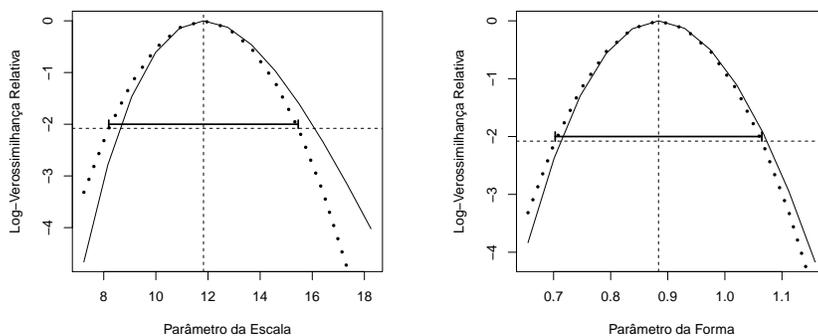
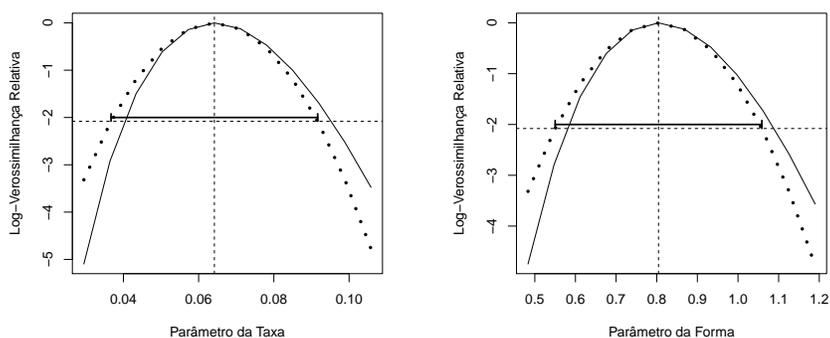
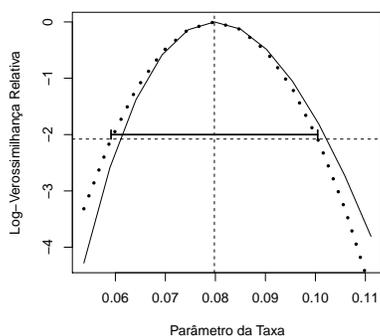
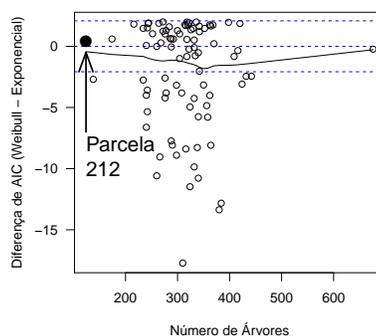
Distribuição Weibull*Distribuição Gama**Distribuição Exponencial**Parcela 212*

Figura 10.4: Perfis de log-verossimilhança relativa para as estimativas de máxima verossimilhança (MLE) dos parâmetros da família Weibull ajustada à parcela 212 (62 árvores). A linha contínua é o perfil de log-verossimilhança relativa. A linha pontilhada é a aproximação quadrática para o perfil. O segmento de reta horizontal é o intervalo de log-verossimilhança segundo a aproximação quadrática. A linha tracejada vertical indica a posição da MLE. A linha tracejada horizontal indica o limite de $-\ln(8)$.

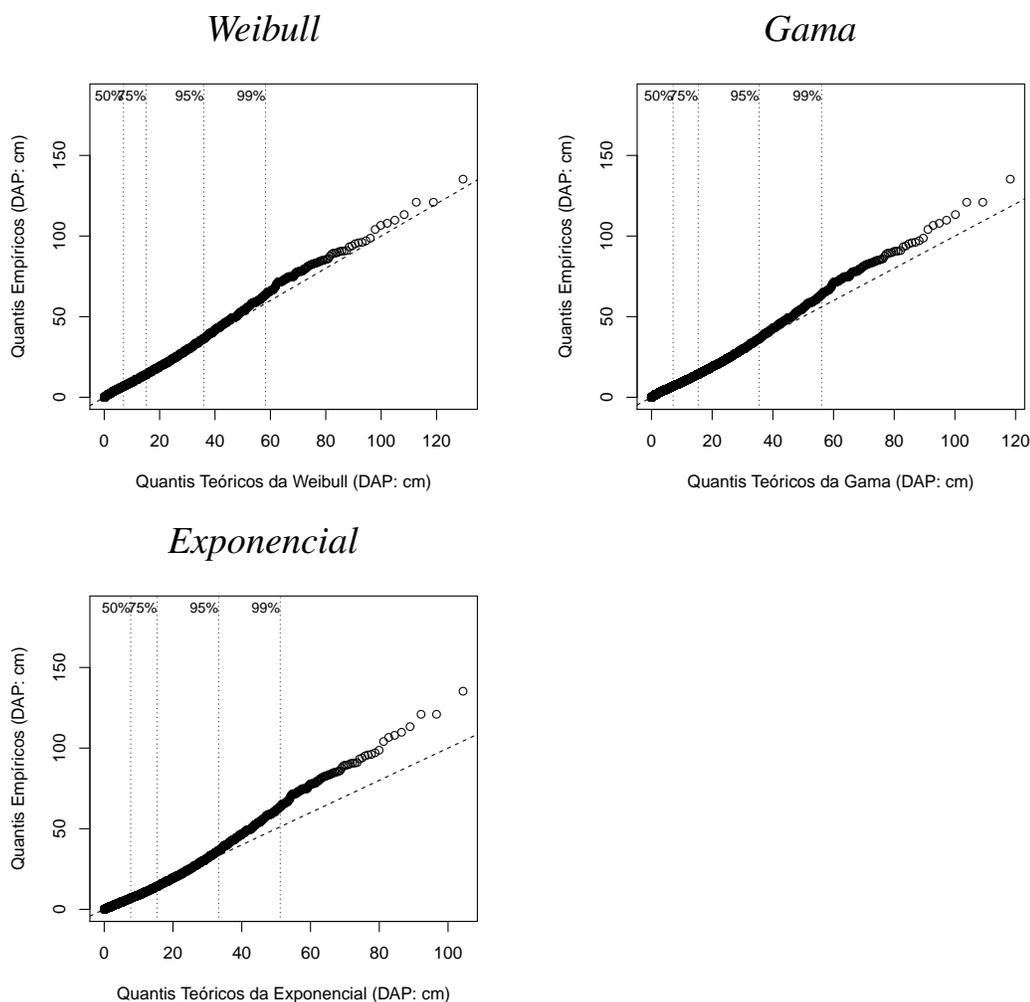


Figura 10.5: Gráficos Quantis-Quantis (gráfico QQ) comparando os quantis empíricos (ordenadas) com os quantis teóricos (abscissas) segundo as distribuições Weibull, gama e exponencial ajustadas na escala de levantamento.

estocásticas, particularmente distribuições de DAP. Enquanto os critérios numéricos são sempre critérios *relativos*, os gráficos Quantil-Quantil contrastam quantis empíricos com quantis teóricos e permitem avaliações absolutas. Por exemplo, na escala de levantamento (figura 10.5), a escolha da distribuição Weibull nos assegura coincidência de quase 99% entre os quantis empíricos e teóricos. Isto implica, que, caso o modelo Weibull seja adotado, discrepâncias marcantes entre observado e esperado ocorrerão apenas na região extrema da distribuição (1% da “calda” da distribuição). Já no caso do ajuste das distribuições para parcela

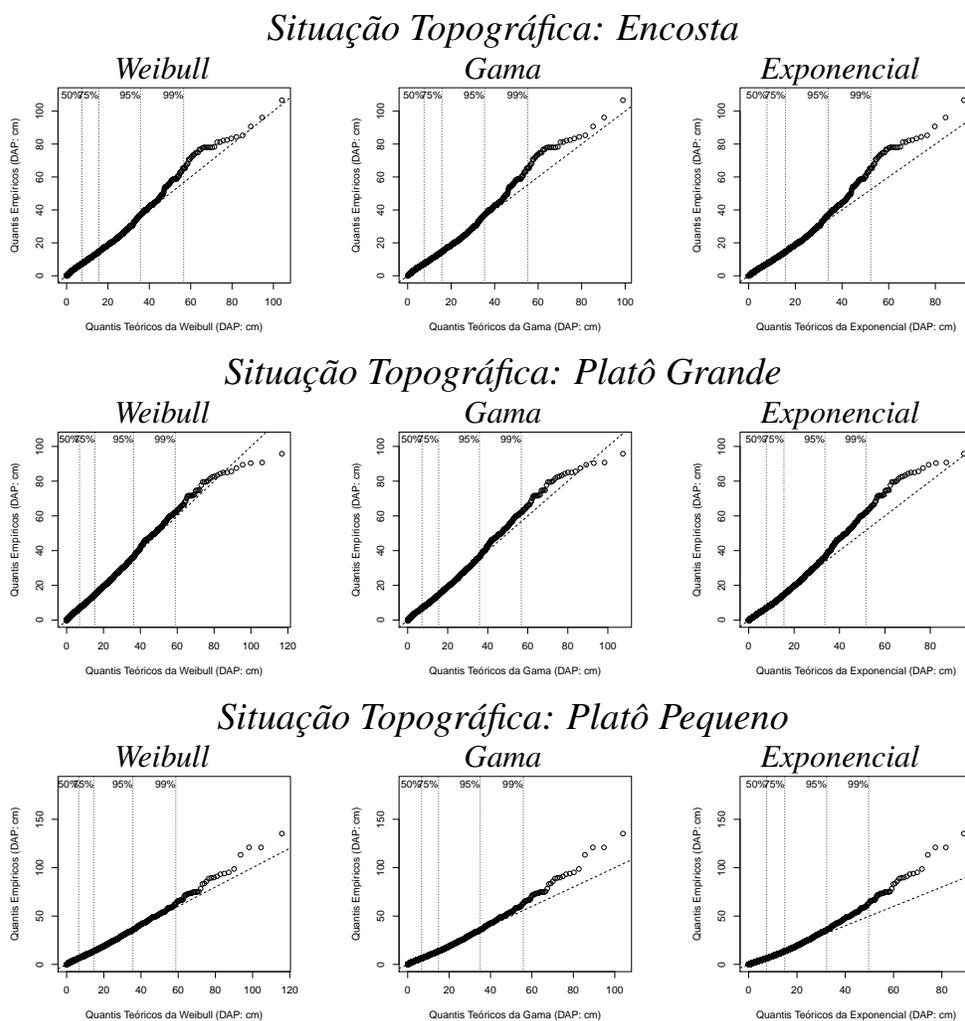


Figura 10.6: Gráficos Quantis-Quantis (gráfico QQ) comparando os quantis empíricos (ordenadas) com os quantis teóricos (abscissas) segundo as distribuições Weibull, gama e exponencial ajustadas na escala topográfica.

201 (figura 10.7) é possível verificar que a distribuição exponencial é um modelo absolutamente inapropriado, pois as discrepâncias entre os quantis empíricos e teóricos já se mostram abaixo da mediana (linha de 50%) e do terceiro quartil (linha de 75%), que é uma região com grande número de observações.

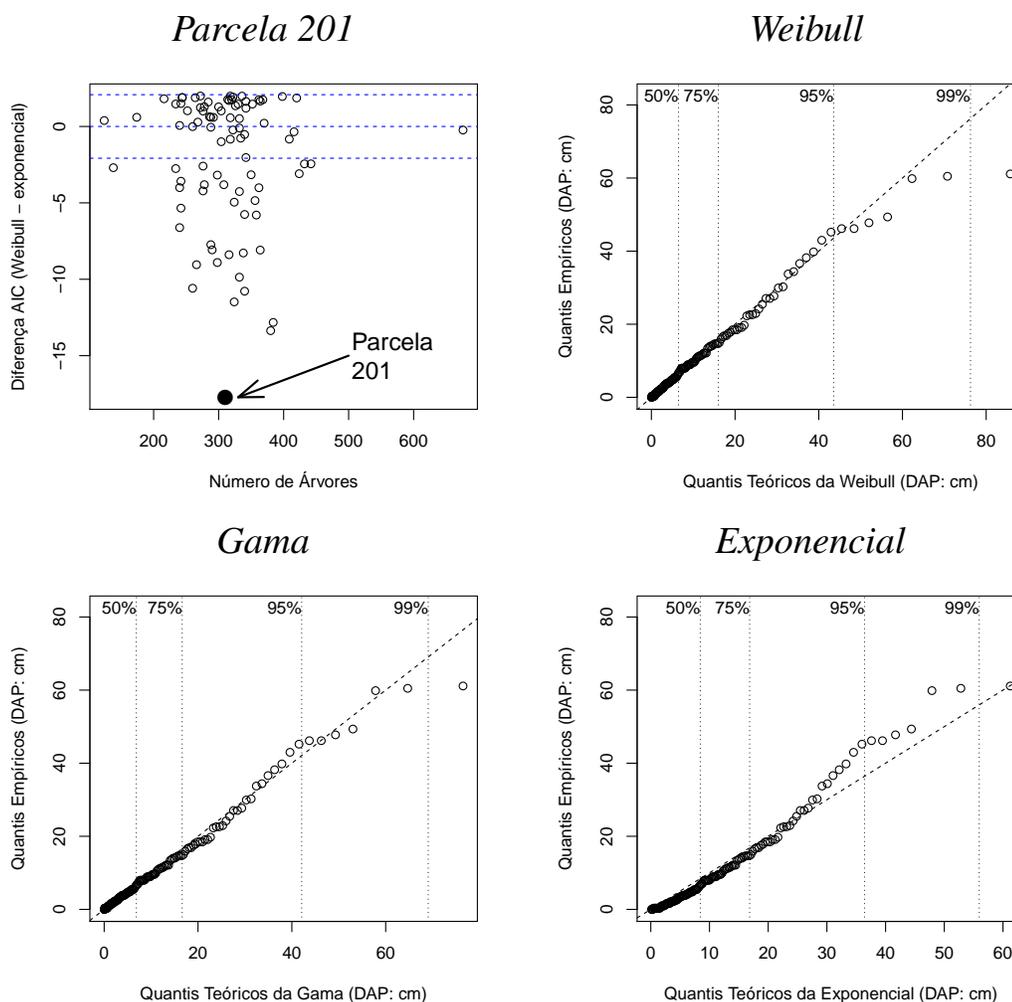


Figura 10.7: Gráficos Quantis-Quantis (gráfico QQ) comparando os quantis empíricos (ordenadas) com os quantis teóricos (abscissas) segundo as distribuições Weibull, gama e exponencial ajustadas para parcela 201.

Estrutura da Floresta como Função de Atributos Quantitativos

O ajuste das famílias estocásticas nas três escalas dos dados mostra que o modelo na escala mais detalhada (escala de parcela) tem o melhor desempenho (tabela 10.4). Em verdade, espera-se que este resultado seja sempre o resultado mais frequente, uma vez que a escala mais detalhada é aquela mais próxima do modo como observações foram realizadas no campo. O maior número de parâmetros a serem estimados nessa escala pode não penalizar suficientemente os modelos, de modo que o desempenho dos modelos na escala de maior detalhe seja superior aos

modelos ajustados em escalas maiores. É isto que se observa neste caso. Embora o número de parâmetros nos modelos na escala de parcela seja muito maior que nas outras escalas (escala de levantamento e topográfica), o melhor desempenho, em termos de log-verossimilhança negativa, mais que compensa o excesso de parâmetros, de modo que mesmo o critério de informação de Akaike, que penaliza o número de parâmetros, indica o ajuste na escala de parcela como sendo o melhor.

Distribuição de DAP e Atributos do Arvoredo

Contudo, a variação parcela-a-parcela do melhor modelo para representar a distribuição de DAP indica uma variação na estrutura da floresta que deve se mostrar em atributos agregados da parcela. Por exemplo, a distribuição exponencial tem um único parâmetro. Este parâmetro apresentado na forma de taxa (λ) é igual à recíproca da esperança matemática da distribuição. Caso o parâmetro seja apresentado como parâmetro de escala ($\beta = 1/\lambda$) ele é igual à da esperança matemática da distribuição (veja a tabela 7.1, no capítulo 7). Consequentemente, a MLE do parâmetro da distribuição exponencial (taxa ou escala) deve ter relação com o DAP médio da parcela.

Essa relação também deve estar presente nas distribuições Weibull e gama, com a complicação de que também é necessário estabelecer relações entre o parâmetro da forma e os atributos agregados das parcelas. Como a dispersão nessas distribuições é influenciada pelo parâmetro da forma (veja a tabela 7.1, no capítulo 7), espera-se que haja uma relação entre o parâmetro da forma e medidas amostrais da variabilidade do DAP nas parcelas. Na Teoria Estatística, a variabilidade de uma variável é tradicionalmente medida pela variância ou desvio padrão amostrais. Mas, na Mensuração Florestal é tradicional se representar a variabilidade dos diâmetros de uma floresta pela diferença entre o DAP médio quadrático (\bar{d}_q) e o DAP médio (\bar{d}). A diferença dos quadrados dessas medidas é a própria variância amostral dos DAP:

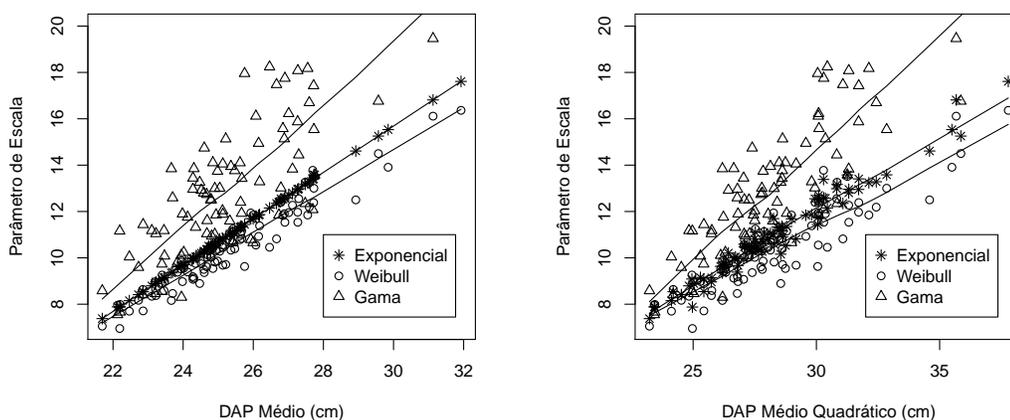
$$s_d^2 = \bar{d}_q^2 - \bar{d}^2 .$$

A relação empírica observada nos dados não segue necessariamente aquelas deduzidas teoricamente, por isto, a figura 10.8 mostra as relações observadas no levantamento da floresta de Bom Jardim entre as MLE dos parâmetros da escala e forma, com o DAP médio, DAP médio quadrático e a diferença entre eles. Essas foram as relações mais fortes observadas nos dados, quando se avaliou a relação das MLE dos parâmetros com vários atributos das parcelas: DAP médio, DAP médio quadrático, altura comercial média, densidade de estande (número de árvores por hectare), área basal e número de espécies na parcela.

Observou-se que a MLE do parâmetro de escala nas diversas parcelas tem estreita relação linear com o DAP médio da parcela. Ela também tem relação

linear com o DAP médio quadrático, embora essa relação seja menos estreita, principalmente no caso da distribuição exponencial. Já a MLE do parâmetro da forma apresentou a melhor relação linear com a recíproca da diferença entre o DAP médio quadrático e o DAP médio.

Parâmetro de Escala



Parâmetro de Forma

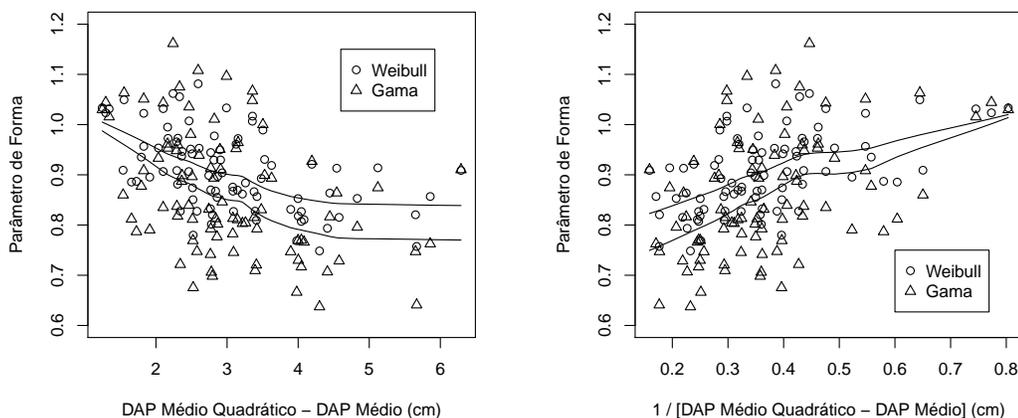


Figura 10.8: Relação entre as MLE dos parâmetros das distribuições (exponencial, Weibull e gama) ajustadas parcela-a-parcela com os atributos das parcelas: DAP médio e DAP médio quadrático. As linhas foram geradas por suavização, indicando a tendência da relação média.

Modelos Preditivos para Distribuição de DAP

Essas relações podem ser utilizadas para elaboração de modelos da distribuição de DAP parcela-a-parcela, mas sem a necessidade de utilizar um ou dois parâmetros para cada parcela. Os parâmetros da distribuição são substituídos por funções lineares do DAP médio e do DAP médio quadrático, gerando os seguintes modelos:

$$\begin{aligned} \text{Exponencial: } & \mathbf{L}\{\lambda(\bar{d}_j)\}; & \lambda(\bar{d}_j) &= \beta_{\lambda 0} + \beta_{\lambda 1} \bar{d}_j \\ \text{Weibull: } & \mathbf{L}\{\beta(\bar{d}_j); \gamma(\bar{d}_j, \bar{d}_{qj})\} & \beta(\bar{d}_j) &= \beta_{\beta 0} + \beta_{\beta 1} \bar{d}_j; \\ & & \gamma(\bar{d}_j, \bar{d}_{qj}) &= \beta_{\gamma 0} + \beta_{\gamma 1} 1/(\bar{d}_{qj} - \bar{d}_j) \\ \text{Gama: } & \mathbf{L}\{\lambda(\bar{d}_j); \gamma(\bar{d}_j, \bar{d}_{qj})\} & \lambda(\bar{d}_j) &= \beta_{\lambda 0} + \beta_{\lambda 1} \bar{d}_j \\ & & \gamma(\bar{d}_j, \bar{d}_{qj}) &= \beta_{\gamma 0} + \beta_{\gamma 1} 1/(\bar{d}_{qj} - \bar{d}_j) \end{aligned}$$

Os parâmetros dos modelos não são mais os parâmetros de taxa/escala e forma, mas os coeficientes das relações lineares ($\beta_{\lambda 0}$, $\beta_{\lambda 1}$, $\beta_{\gamma 0}$, $\beta_{\gamma 1}$). Estes modelos possibilitam a obtenção de uma distribuição particular para cada parcela, utilizando o DAP médio e o DAP médio quadrático como variáveis preditoras, mas com um número muito pequeno de parâmetros: 2 parâmetros para distribuição exponencial e 4 para as distribuições Weibull e gama.

Na prática, estes modelos são equivalentes à abordagem de *predição de parâmetros*, mas o ajuste não é realizado em duas etapas: (1) ajuste da distribuição parcela-a-parcela e (2) regressão linear das estimativas dos parâmetros da distribuição em função de variáveis preditoras. Nos modelos acima, a função de log-verossimilhança é expressa em termos das relações lineares com as variáveis preditoras, e a maximização da função de log-verossimilhança gera diretamente as MLE dos coeficientes das relações lineares. Este método, corresponde ao método (5) testado por Cao (2004).

Modelos Preditivos Ajustados e seu Desempenho

A tabela 10.5 apresenta as MLE e respectivos erros padrão obtidos no levantamento de Bom Jardim. Os gráficos dos perfis de log-verossimilhança dessas estimativas não são apresentados, mas eles revelaram que elas são estimativas estáveis e que a aproximação quadrática pode ser utilizada para gerar intervalos de log-verossimilhança coerentes. Os intervalos da tabela 10.5 foram calculados com base na equação 8.1 do capítulo 8.

É importante verificar se ao se utilizar estes modelos preditivos, em que os parâmetros da distribuição de DAP das parcelas são preditos como função linear dos atributos da floresta, obtem-se uma melhor representação da estrutura de DAP nas parcelas quando comparada com os modelos ajustados nas diferentes escalas

Tabela 10.5: Estimativas de máxima verossimilhança (MLE) para os modelos em que os parâmetros das distribuições estocásticas são os coeficientes de relação linear do DAP médio e do DAP médio quadrático. O intervalo de log-verossimilhança foi calculado para o limite canônico de $\ln(8)$.

DISTRIB.	PARÂMETRO	ESTIMATIVA (MLE)	ERRO PADRÃO	INTERVALO DE LOG-VEROSSIMILHANÇA
<i>Exponencial</i>	$\beta_{\lambda 0}$	-14.329999	1.4117754	± 2.87908
	$\beta_{\lambda 1}$	1.001057	0.0567037	± 0.11564
<i>Weibull</i>	$\beta_{\beta 0}$	-11.2348220	1.52195268	± 3.10377
	$\beta_{\beta 1}$	0.8578470	0.06127602	± 0.12396
	$\beta_{\gamma 0}$	0.7919923	0.01979824	± 0.04038
	$\beta_{\gamma 1}$	0.3133716	0.05297960	± 0.10804
<i>Gama</i>	$\beta_{\lambda 0}$	-22.9215663	2.26883495	± 4.62691
	$\beta_{\lambda 1}$	1.4179322	0.09204797	± 0.18772
	$\beta_{\gamma 0}$	0.7468314	0.02548235	± 0.05197
	$\beta_{\gamma 1}$	0.2901777	0.06698991	± 0.13661

dos dados. A tabela 10.6 apresenta os valores de log-verossimilhança negativa (LVN) e do critério de informação de Akaike (AIC) obtidos nos modelos por escala de dados e nos modelos preditivos.

Nota-se que, em termos da log-verossimilhança negativa, os modelos preditivos tem um desempenho muito melhor que os modelos ajustados na escala de levantamento e na escala topográfica, mas um desempenho inferior aos modelos ajustados na escala de parcela. O modelo exponencial é uma exceção neste caso, uma vez que a MLE do seu parâmetro de escala é o próprio DAP médio:

$$\hat{\beta} = \frac{1}{\hat{\lambda}} = \bar{x} = \bar{d} - d_{min}.$$

Por isto, no caso da distribuição exponencial o desempenho do modelo preditivo é igual ao modelo ajustado na escala de parcela.

Mas, há que considerar que o número de parâmetros no modelo ajustado na escala de parcela é muito grande: duas vezes o número de parcelas no caso das distribuições Weibull e gama. Se o número de parâmetros for penalizado no critério de ajuste, como acontece no AIC-total, os modelos preditivos se mostram superiores aos modelos ajustados na escala de parcela.

Tabela 10.6: Desempenho dos modelos candidatos para representar a distribuição de DAP da floresta nativa de Bom Jardim, MA, nas três situações topográficas da floresta, nas três escalas hierárquicas dos dados e na situação em que os parâmetros são tomados como funções lineares de variáveis predictoras. O desempenho é apresentado em termos da log-verossimilhança negativa (LVN) e do critério de informação de Akaike (AIC).

DISTRIB.	NÚMERO DE PARÂMETROS	SITUAÇÃO TOPOGRÁFICA						TOTAL	
		ENCOSTA		PLATÔ GRANDE		PLATÔ PEQUENO		LVN	AIC
		LVN	AIC	LVN	AIC	LVN	AIC	LVN	AIC
<i>Escala de Levantamento</i>									
Exponencial	1	11526	—	16154	—	13122	—	40803	81607
Weibull	2	11517	—	16096	—	13047	—	40661	81325
Gama	2	11519	—	16096	—	13053	—	40669	81341
<i>Escala Topográfica</i>									
Exponencial	3	11526	23053	16154	32309	13121	26244	40800	81606
Weibull	6	11511	23026	16096	32197	13042	26088	40649	81310
Gama	6	11513	23030	16096	32196	13048	26101	40657	81327
<i>Escala de Parcela</i>									
Exponencial	87	11486	23015	16123	32313	13029	26101	40638	81430
Weibull	174	11459	23006	16049	32234	12951	25987	40459	81227
Gama	174	11461	23009	16045	32227	12951	25986	40457	81222
<i>Parâmetros como Funções Lineares</i>									
Exponencial	2	11486	—	16123	—	13029	—	40638	81280
Weibull	4	11478	—	16063	—	12970	—	40511	81031
Gama	4	11481	—	16066	—	12973	—	40520	81048

Conclusão

Este resultado mostra que a distribuição de DAP de arvoredos particulares, representados pelas parcelas do levantamento, pode ser adequadamente preditos em função de atributos agregados destes arvoredos, tomando os parâmetros das famílias estocásticas como funções lineares (ou não lineares) de atributos da estrutura do arvoredo. Estes modelos integram a abordagem tradicional da Biometria Florestal chamada de “*método da predição de parâmetros*”. A inferência segundo o Axioma da Verossimilhança, contudo, permite que a inferência seja realizada de forma integrada e fornece as ferramentas adequadas não só ajuste dos modelos, mas também para avaliação do desempenho e seleção dos modelos.

Apêndices

Apêndice A

Lista de Figuras

2.1	Distribuição probabilística da “estatística” t de Student com 5 graus de liberdade, mostrando a região de rejeição equivalente à probabilidade de 10%, a qual é definida pelo <i>valor crítico</i> de $t = 2,00 $	55
2.2	Distribuição probabilística da “estatística” t de Student com 5 graus de liberdade, mostrando a região correspondente ao valor-p para o valor observado de $t_{\text{obs}} = 2,75$ (valor-p = 4%).	57
6.1	Figuras ilustrativas dos resultados do cálculo de probabilidade utilizando diagramas de Venn.	118
7.1	Gráfico da função densidade da distribuição Bernoulli.	143
7.2	Gráfico da função densidade de uma distribuição uniforme discreta ($N = 12$).	144
7.3	Gráficos de duas distribuições da família binomial.	145
7.4	Gráficos da função densidade de duas distribuições Poisson.	147
7.5	Gráficos da função densidade de duas distribuições geométricas.	149
7.6	Gráficos da densidade de duas distribuições binomiais negativas.	150
7.7	Função de densidade ($f(x)$) e função de distribuição ($F(x)$) de uma distribuição retangular com parâmetros $a = 2$ e $b = 11$	153
7.8	Exemplo de padrão espacial completamente aleatório e histogramas das coordenadas cartesianas (x, y) da localização das árvores na parcela quadrada de 100 m de lado.	154
7.9	Gráficos da função densidade ($f(x)$) e da função de distribuição ($F(x)$) de três distribuições exponenciais.	155
7.10	Gráficos da função densidade (a-c) e da função de distribuição (d) de distribuições da família Weibull mostrando a influência dos parâmetros sobre o comportamento das distribuições.	158

- 7.11 (a) Variação dos parâmetros de escala e forma para as 60 espécies arbóreas indicando os grupos da sucessão ecológica. (b) Curvas da função de densidade para os valores médios de cada grupo. . . . 160
- 7.12 Gráficos da função densidade ($f(x)$) e da função de distribuição ($F(x)$) de algumas distribuições gama. 161
- 7.13 Distribuições gama como modelo das distribuições de diâmetros de árvores de uma mesma floresta plantada de *Eucalyptus grandis* em diferentes idades. 162
- 7.14 Exemplos de distribuições da família Gaussiana, mostrando-se os gráficos das funções de densidade (a,c) e das respectivas funções de distribuição (b,d). As influências da média (a,b) e do desvio padrão (c,d) se mostram totalmente independentes. 164
- 7.15 Distribuição Gaussiana Padronizada ilustrando a propriedade da família Gaussiana de probabilidades fixas para intervalos centrados na média com largura definida em proporções do desvio padrão. 166
- 7.16 Distribuição da área basal ($m^2 ha^{-1}$) de arvoredos de floresta tropical do Município de Bom Jardim, Maranhão, na escala original (a) e na forma de variável padronizada (b). A curva irregular (linha contínua) é a densidade estimada empiricamente por procedimento não paramétrico, enquanto a curva simétrica (linha tracejada) é o modelo Gaussiano. 167
- 7.17 Proporção de árvores dominadas em arvoredos de floresta plantada de eucalipto *urograndis*, na região central do estado de São Paulo. A curva irregular (linha sólida) é a densidade estimada empiricamente por procedimento não paramétrico, enquanto que a curva simétrica (linha tracejada) é o modelo Gaussiano. 168
- 7.18 Gráficos da função densidade (a,c) e a função de distribuição (b,d) de algumas distribuições log-normais. Os gráficos (a) e (b) mostram que o parâmetro μ influencia a dispersão dos valores, enquanto os gráficos (c) e (d) mostram a influência do parâmetro σ sobre a forma. 169
- 7.19 Densidade da abundância de espécies arbóreas numa floresta tropical na escala de número de árvores (a) e na escala de logaritmo do número de árvores (b). A curva irregular é a densidade estimada empiricamente por procedimento não paramétrico, enquanto que a curva regular é o modelo log-normal, o qual se mostra simétrico na escala logarítmica (gráfico (b)). 170

- 7.20 Densidade do volume de madeira de árvores individuais de *Eucalyptus saligna* de florestas plantadas na região central do estado de São Paulo. Em (a) o volume é apresentado na sua escala original (dm^3) e em (b) na escala logarítmica. A curva irregular é a densidade estimada empiricamente por procedimento não paramétrico, enquanto que a curva regular (simétrica no gráfico (b)) é o modelo log-normal. 171
- 7.21 Gráficos Quantil-Quantil (gráfico QQ) para verificar a distribuição de variáveis observadas. (a) Distribuição Exponencial: DAP (cm) de árvores de floresta tropical do município de Bom Jardim, Maranhão. (b) Distribuição Gaussiana: proporção (%) de árvores dominadas em povoaamentos de floresta plantada de eucalipto *urograndis*, na região central do estado de São Paulo. (c-d) Distribuição Log-normal: volume de madeira de árvores de floresta plantada de *Eucalyptus saligna* na região central do estado de São Paulo. Em (c) o volume foi transformado para escala logarítmica e comparado com os quantis teóricos da distribuição Gaussiana padronizada. Em (d) o volume é comparado com os quantis teóricos da distribuição log-normal. 173
- 7.22 Gráficos de simetria para medidas de árvores de florestas plantadas de *Eucalyptus saligna* da região central do estado de São Paulo: volume de madeira (a) na escala original e (b) na escala logarítmica; altura total (c) na escala original e (d) na escala logarítmica. 175
- 7.23 Padrões espaciais: (a) completamente aleatório, (b) agregado, (c) regular, e (d) agregado cujo índice de dispersão não consegue identificar. 181
- 7.24 Densidades empíricas para a média amostral de variáveis com diferentes distribuição estocásticas (X): (a) binomial ($n = 10, p = 0,15$), (b) Poisson ($\lambda = 50$), (c-f) Weibull com diferentes valores dos parâmetros. As curvas em cinza resultam de diferentes tamanhos de amostra, o menor e o maior tamanho de amostra são indicados no gráfico. As curvas com linhas de maior grossura indicam a distribuição estocástica dos dados. 186
- 7.25 Intervalo de 95% de probabilidade resultante da aproximação Gaussiana para (a) variável binomial e (b) média amostral resultante de uma variável com distribuição Weibull. A linha com maior irregularidade é a distribuição original, enquanto que alinha perfeitamente regular e simétrica é a aproximação Gaussiana. 188

- 8.1 Distribuição exponencial como modelo estocástico para distribuição de DAP de floresta tropical no município de Bom Jardim, Maranhão. Os gráficos mostram a curva de log-verossimilhança (linha tracejada) para a estimativa de máxima verossimilhança do parâmetro da distribuição exponencial ($\hat{\lambda}$) e sua aproximação quadrática (linha sólida). À medida que o gráfico retrata uma vizinhança mais próxima de $\hat{\lambda}$ (gráficos de (a) para (d)), a aproximação quadrática se mostra mais apropriada. 218
- 8.2 Exemplo ilustrativo do aumento da curvatura da curva de log-verossimilhança com o aumento do tamanho da amostra para distribuição binomial ($N = 50$ e $p = 0,20$). As curvas seguem, da menor curvatura para de maior curvatura, tamanhos de amostra de 20, 50, 100, 300 e 1000 observações independentes. São mostradas a curva de log-verossimilhança (linha sólida) e sua aproximação quadrática (linha tracejada). Os dois gráficos com vizinhanças de tamanho diferente mostram que à medida que o tamanho da amostra cresce a curvatura da curva de log-verossimilhança aumenta e a aproximação quadrática se torna mais apropriada. 221
- 8.3 Influência do tamanho da amostra sobre a qualidade do intervalo de log-verossimilhança gerado pela aproximação quadrática em função da curvatura da curva de log-verossimilhança. As curvas são referentes à distribuição binomial ($N = 30$ e $p = 0,02$), sendo que a de menor curvatura tem tamanho de amostra 20 e a de maior curvatura tamanho de amostra 100. São mostradas a curva de log-verossimilhança (linha tracejada) e sua aproximação quadrática (linha sólida). A linha horizontal representa a diferença de log-verossimilhança igual a $-\ln(8)$ 224
- 8.4 Superfície de log-verossimilhança da distribuição Gaussiana, para um tamanho de amostra $n = 100$. O gráfico (a) é um gráfico de contorno que apresenta as isolinhas de log-verossimilhança, a linha tracejada delimita a região de log-verossimilhança maior ou igual a $-\ln(8)$. O gráfico (b) é um gráfico de perspectiva que mostra uma representação tridimensional da superfície. 233
- 8.5 Gráficos de contorno da superfície de log-verossimilhança da distribuição Weibull para dois tamanhos de amostra. A linha tracejada indica a região de log-verossimilhança correspondente ao limite de $-\ln(8)$ 238

- 8.6 Gráficos da Weibull para dois valores diferentes do parâmetro da forma (γ) e mesmo valor do parâmetro da escala (β). Os gráficos no alto mostram a curva da função de densidade para as duas distribuições. Os gráficos de contorno respectivos mostram as isolinhas da superfície de log-verossimilhança (linhas contínuas), construída com uma amostra de 1000 observações ($n = 1000$), e as isolinhas da aproximação quadrática (linhas tracejadas). 240
- 8.7 Curva de log-verossimilhança para razão entre o número de mudas e de plântulas do palmitreiro (*Euterpe edulis* Martius) na região do Vale do Ribeira. Os gráficos apresentam a curva de log-verossimilhança (linha sólida), a aproximação quadrática (linha traço-ponto) e a linha que define a razão de verossimilhança de 8, em todo espaço paramétrico de θ (a) e na vizinhança da MLE (b). 245
- 8.8 Gráfico de contorno (a) mostrando a superfície de log-verossimilhança para uma distribuição Gaussiana com base numa amostra de 200 observações. A linha sólida corresponde à direção da MLE do desvio padrão ($\hat{\sigma}$). Gráfico da log-verossimilhança estimada da média (b) em que as linhas horizontais pontilhadas correspondentem às isolinhas da log-verossimilhança relativa do gráfico de contorno. . 247
- 8.9 Curvas de log-verossimilhança estimada (linha sólida) e de log-verossimilhança perfilhada (linha tracejada) para a média da distribuição Gaussiana em duas vizinhanças da estimativa de máxima verossimilhança (MLE) numa amostra de tamanho $n = 50$. A linha tracejada vertical indica a posição da MLE da média ($\hat{\mu}$), enquanto a linha tracejada horizontal indica a diferença de log-verossimilhança de $-\ln(8)$ 250
- 8.10 Gráfico de contorno da superfície de log-verossimilhança para a uma amostra de tamanho $n = 100$ da distribuição Gaussiana, mostrando as curvas da log-verossimilhança estimada (linha tracejada) e de log-verossimilhança perfilhada (linha pontilhada). A linha tracejada vertical indica a posição da MLE da média ($\hat{\mu}$), enquanto a região delimitada por linha sólida espessa corresponde à diferença de log-verossimilhança de $-\ln(8)$ 251
- 8.11 Curvas de log-verossimilhança estimada (linha sólida) e de log-verossimilhança perfilhada (linha tracejada) dos coeficientes de regressão (β_0 e β_1) do modelo linear simples clássico, com base em dados simulados com dois tamanhos de amostra: $n = 50$ e $n = 200$ observações. A linha pontilhada vertical indica a posição da MLE, enquanto que a linha pontilhada horizontal indica a diferença de log-verossimilhança de $-\ln(8)$ 254

- 8.12 Superfície de log-verossimilhança da distribuição Weibull para dados de DAP de árvores de floresta nativa do município de Bom Jardim, MA, apresentando duas vizinhanças das MLEs ((a) e (b)); as linhas tracejadas indicam a posição das MLEs. Curvas de log-verossimilhança estimada (linha tracejada) e perfilhada (linha sólida) para os parâmetros de escala (β) e forma (γ) nos mesmos dados ((c) e (d)). A linha pontilhada vertical indica a posição da MLE, enquanto que a linha pontilhada horizontal indica a diferença de log-verossimilhança de $-\ln(8)$ 257
- 9.1 Diferenças do AIC grafada contra o número de árvores por parcela (a, b, c) e o DAP médio (d, e, f), para as distribuições Weibull e exponencial (a, d), gama e exponencial (b, e) e Weibull e gama (c, f). As linhas horizontais dos gráficos indicam as posições de diferença de AIC iguais a $\ln(8)$, zero e $-\ln(8)$. As curvas em todos os gráficos são linhas de tendência de relação entre as variáveis encontradas por suavização. 289
- 9.2 Estimativas do parâmetro da forma (a, c) e do parâmetro de escala (b, d) em função do número de árvores na parcela (a, b) e do DAP médio da parcela (c, d). As curvas em todos os gráficos são linhas de tendência de relação entre as variáveis encontradas por suavização. 290
- 9.3 Gráficos de dispersão do resíduo contra o valor ajustado (alto) e do erro de predição contra o valor predito na validação cruzada parcela-a-parcela (baixo) para dois modelos (Husch e Spurr) de equações volumétrica das árvores num arvoredo de *Eucalyptus grandis* da Estação Experimental de Ciências Florestais de Itatinga (ESALQ-USP). 310

- 9.4 Log-verossimilhança da predição de cada observação na validação cruzada parcela-a-parcela grafada contra os erros de predição (m^3) para o volume de madeira de árvores num arvoredo de *Eucalyptus grandis* da Estação Experimental de Ciências Florestais de Itatinga (ESALQ-USP). Os pontos representam os erros de predição segundo o modelo de equação volumétrica de Husch (equação local), enquanto que as linhas de cada cor representam as curvas de log-verossimilhança dos modelos de equação volumétrica segundo a legenda no gráfico. As curvas foram geradas pela suavização da relação do log-verossimilhança e erro de predição das observações individuais. Em (a) o gráfico apresenta todas as observações, em (b) o gráfico ressalta uma região mais próxima do valor zero para enfatizar as diferenças sutis das curvas dos modelo de equação volumétrica de dupla entrada. 312
- 10.1 Diferenças do AIC grafadas contra o DAP médio das parcelas para as distribuições Weibull e exponencial (a), gama e exponencial (b) e Weibull e gama (c). As linhas horizontais dos gráficos indicam as posições de diferença de AIC iguais a $\ln(8)$, zero e $-\ln(8)$. As curvas em todos os gráficos são linhas de tendência de relação entre as variáveis encontradas por suavização. 340
- 10.2 Perfis de log-verossimilhança relativa para as estimativas de máxima verossimilhança (MLE) dos parâmetros das distribuições Weibull, gama e exponencial ajustadas na escala de levantamento. A linha contínua é o perfil de log-verossimilhança relativa. A linha pontilhada é a aproximação quadrática para o perfil. O segmento de reta horizontal é o intervalo de log-verossimilhança segundo a aproximação quadrática. A linha tracejada vertical indica a posição da MLE. A linha tracejada horizontal indica o limite de $-\ln(8)$ 342
- 10.3 Perfis de log-verossimilhança relativa para as estimativas de máxima verossimilhança (MLE) dos parâmetros da família Weibull ajustada na escala topográfica. A linha contínua é o perfil de log-verossimilhança relativa. A linha pontilhada é a aproximação quadrática para o perfil. O segmento de reta horizontal é o intervalo de log-verossimilhança segundo a aproximação quadrática. A linha tracejada vertical indica a posição da MLE. A linha tracejada horizontal indica o limite de $-\ln(8)$ 343

- 10.4 Perfis de log-verossimilhança relativa para as estimativas de máxima verossimilhança (MLE) dos parâmetros da família Weibull ajustada à parcela 212 (62 árvores). A linha contínua é o perfil de log-verossimilhança relativa. A linha pontilhada é a aproximação quadrática para o perfil. O segmento de reta horizontal é o intervalo de log-verossimilhança segundo a aproximação quadrática. A linha tracejada vertical indica a posição da MLE. A linha tracejada horizontal indica o limite de $-\ln(8)$ 345
- 10.5 Gráficos Quantis-Quantis (gráfico QQ) comparando os quantis empíricos (ordenadas) com os quantis teóricos (abscissas) segundo as distribuições Weibull, gama e exponencial ajustadas na escala de levantamento. 346
- 10.6 Gráficos Quantis-Quantis (gráfico QQ) comparando os quantis empíricos (ordenadas) com os quantis teóricos (abscissas) segundo as distribuições Weibull, gama e exponencial ajustadas na escala topográfica. 347
- 10.7 Gráficos Quantis-Quantis (gráfico QQ) comparando os quantis empíricos (ordenadas) com os quantis teóricos (abscissas) segundo as distribuições Weibull, gama e exponencial ajustadas para parcela 201. 348
- 10.8 Relação entre as MLE dos parâmetros das distribuições (exponencial, Weibull e gama) ajustadas parcela-a-parcela com os atributos das parcelas: DAP médio e DAP médio quadrático. As linhas foram geradas por suavização, indicando a tendência da relação média. 350

Apêndice B

Lista de Tabelas

7.1	Esperança matemática e variância das distribuições apresentadas.	180
8.1	Número de bolas brancas sucessivas observadas (n_b) no experimento canônico em função da razão de verossimilhança. Reproduzido de Royall (1997).	201
8.2	Estimadores de máxima verossimilhança (MLE) para algumas distribuições estocásticas apresentadas anteriormente, com base numa amostra de n observações independentes.	210
9.1	Critério de Informação de Akaike (AIC) para os modelos exponencial (E), Weibull (W) e gama (G), ajustados para o agregado de todas as árvores do levantamento (“Ajuste Geral”) e por parcela (“Ajuste por Parcela”). No ajuste geral, os AIC foram obtidos a partir da soma da log-verossimilhança por situação topográfica e no total, já os AIC do ajuste por parcela foram obtidos somando-se os AIC de cada parcela.	291
9.2	Critérios de <i>ajustados</i> modelos de equação de volume para <i>Eucalyptus grandis</i> na Estação Experimental de Ciências Florestais de Itatinga (ESALQ-USP). Estes critérios são baseados nos <i>resíduos</i> obtidos no ajuste dos modelos com todas as observações. O erro padrão da estimativa ($s_{\hat{y}}$) e o coeficiente de determinação ajustado (R_a^2) foram computados na escala de ajuste dos modelos ($\ln(m^3)$), enquanto que o critério de informação de Akaike (AIC) foi computado na escala de ajuste ($\ln(m^3)$) e de predição (m^3).	304

- 9.3 Critérios da validação cruzada dos modelos de equação volumétrica de um arvoredo de *Eucalyptus grandis* na Estação Experimental de Ciências Florestais de Itatinga (ESALQ-USP). A *validação cruzada uma-a-uma* foi realizada árvore-a-árvore, enquanto que a *validação cruzada k-a-k* foi implementada parcela-a-parcela. Todos os critérios foram calculados na escala de predição do volume das árvores, isto é, na escala de metros cúbicos. 305
- 9.4 Critérios da validação cruzada por classe de DAP aplicados tanto para os resíduos do ajuste quanto para os erros de predição dos modelos de equação volumétrica de um arvoredo de *Eucalyptus grandis* na Estação Experimental de Ciências Florestais de Itatinga (ESALQ-USP). A *validação cruzada uma-a-uma* foi realizada árvore-a-árvore, enquanto que a *validação cruzada k-a-k* foi implementada parcela-a-parcela. 307
- 9.5 Mudança nos valores do AIC entre a análise do resíduo e a validação cruzada para modelos de equação volumétrica de um arvoredo de *Eucalyptus grandis* na Estação Experimental de Ciências Florestais de Itatinga (ESALQ-USP). 308
- 10.1 Número de parcelas em faixa no levantamento florestal de floresta nativa em função da situação topográfica e do comprimento da parcela. 333
- 10.2 Alguns atributos médios das árvores individuais e das parcelas em cada situação topográfica. 334
- 10.3 Estimativas de máxima verossimilhança (MLE) para os parâmetros das distribuições Weibull, gama e exponencial ajustadas nas diferentes escalas dos dados. Valores entre parênteses são os erros padrões das estimativas. 338
- 10.4 Desempenho dos modelos candidatos para representar a distribuição de DAP da floresta nativa de Bom Jardim, MA, nas três situações topográficas da floresta e nas três escalas hierárquicas dos dados. O desempenho é apresentado em termos da log-verossimilhança negativa (LVN) e do critério de informação de Akaike (AIC). 339
- 10.5 Estimativas de máxima verossimilhança (MLE) para os modelos em que os parâmetros das distribuições estocásticas são os coeficientes de relação linear do DAP médio e do DAP médio quadrático. O intervalo de log-verossimilhança foi calculado para o limite canônico de $\ln(8)$ 352
- 10.6 Desempenho dos modelos candidatos para representar a distribuição de DAP da floresta nativa de Bom Jardim, MA, nas três situações topográficas da floresta, nas três escalas hierárquicas dos dados e na situação em que os parâmetros são tomados como funções lineares de variáveis preditoras. O desempenho é apresentado em termos da log-verossimilhança negativa (LVN) e do critério de informação de Akaike (AIC). 353

Apêndice C

Lista de Definições e Teoremas

5.1	Definição (Cenário Estocástico)	96
5.2	Definição (Processo Estocástico)	98
5.3	Definição (Modelagem Estocástica)	100
5.4	Definição (Modelo Estocástico)	101
5.5	Definição (Modelo Estocástico Empírico)	102
5.6	Definição (Predição)	103
5.7	Definição (Predição Interpolativa)	104
5.8	Definição (Predição Extrapolativa)	104
6.1	Definição (Espaço Amostral)	111
6.2	Definição (Evento)	113
6.3	Definição (Espaço de Eventos)	115
6.4	Definição (Sigma Algebra)	116
6.5	Definição (Probabilidade)	116
6.6	Definição (Espaço de Probabilidade)	116
6.7	Definição (Probabilidade Condicional)	120
6.8	Definição (Teroema de Bayes)	123
6.9	Definição (Eventos Independentes)	125
6.10	Definição (Multiplos Eventos Independentes)	126
7.1	Definição (Variável Aleatória)	132
7.2	Definição (Função de Distribuição)	134
7.3	Definição (Função de Distribuição II)	137
7.4	Definição (Função de Distribuição Discreta)	138
7.5	Definição (Função de Densidade Discreta)	138
7.6	Definição (Função de Densidade Discreta II)	139
7.7	Definição (Função de Distribuição Contínua)	140
7.8	Definição (Função de Densidade Contínua)	140
7.9	Definição (Função de Densidade Contínua II)	141

7.10	Definição (Família de Distribuições e Parâmetros)	142
7.11	Definição (Quantil)	171
7.12	Definição (Distribuição Simétrica)	174
7.1	Teorema (Transformação de Distribuição)	176
7.13	Definição (Esperança Matemática)	178
7.14	Definição (Esperança de Variável Transformada)	178
7.15	Definição (Variância de uma Distribuição)	179
7.2	Teorema (Lei Forte dos Grandes Números)	184
7.3	Teorema (Teorema Central do Limite)	185
8.1	Definição (Verossimilhança)	195
8.2	Definição (Verossimilhança II)	196
8.3	Definição (Razão de Verossimilhança)	196
8.1	Teorema (Combinação de Razões de Verossimilhança)	198
8.4	Definição (Função de Log-Verossimilhança)	203
8.5	Definição (Função de Log-Verossimilhança Relativa)	205
8.6	Definição (Lei da Verossimilhança)	207
8.7	Definição (Princípio da Verossimilhança)	207
8.8	Definição (Princípio da Verossimilhança II)	207
8.9	Definição (Estimativa de Máxima Verossimilhança (MLE))	208
8.10	Definição (Estimativa de Máxima Verossimilhança (MLE) II)	209
8.11	Definição (Curva de Log-verossimilhança)	214
8.12	Definição (Aproximação Quadrática da Curva de Log-Veros.)	215
8.13	Definição (Informação da Curva de Log-Verossimilhança)	219
8.14	Definição (Intervalo de Log-Verossimilhança)	223
8.15	Definição (Superfície de Log-verossimilhança)	228
8.16	Definição (Aproximação Quadrática da Superfície de Log-veros.)	228
8.17	Definição (Matrix de Informação Observada)	229
8.18	Definição (Parâmetros Ortogonais)	241
8.19	Definição (Verossimilhança Estimada)	245
8.20	Definição (Verossimilhança Perfilhada)	249
9.1	Definição (Modelo e Família Operante)	267
9.2	Definição (Modelos e Famílias de Aproximação)	268
9.3	Definição (Subfamílias de Aproximação)	269
9.4	Definição (Discrepâncias)	272
9.5	Definição (Discrepância Relativa de Aproximação)	277
9.6	Definição (Critério de Informação de Akaike)	280
9.7	Definição (Correção do AIC)	282
9.8	Definição (Validação Cruzada)	292
9.9	Definição (Validação Cruzada uma-a-uma)	293

9.10 Definição (Validação Cruzada k -a- k) 293

Referências Bibliográficas

- Akaike, H. Information Theory and an Extension of the Maximum Likelihood Principle. In Petrov, B. e Caski, F. (Eds.), *Proceedings of the Second International Symposium on Information Theory*, (pp. 267–281)., Budapest. Akademiai Kiado, 1973. Reimpressão: Akaike, H. 1992. Information Theory and an Extension of the Maximum Likelihood Principle. Pages 610-624 In: Kotz, S., and N.L. Johnson, editors. *Breakthroughs in Statistics Volume 1. Foundations and Basic Theory*. Springer Series in Statistics, Perspectives in Statistics. Springer-Verlag: New York.
- Akaike, H. Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, v.60, n.2, p.255-265, 1973.
- Akaike, H. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, v.AC-19, n.6, p.716-723, 1974.
- Akaike, H. This Week's Citation Classic. *CC/Social and Behavioral Sciences*, v.51, p.42, 1981.
- Akaike, H. Prediction and entropy. In A. Atkinson e S. Fienberg (Eds.), *A celebration of Statistics* chapter 1, (pp. 1–24). New York: Springer-Verlag, 1985.
- Akaike, H. Factor Analysis and AIC. *Psychometrika*, v.52, n.3, p.317-332, 1987.
- Aldrich, J. Fisher and Regression. *Statistical Science*, v.20, n.4, p.401-417, 2005.
- Allen, D. M. Mean Square Error of Prediction as a Criterion for Selecting Variables. *Technometrics*, v.13, n.3, p.469-475, 1971.
- Anderson, D., Burnham, K., e Thompson, W. Null hypothesis testing: problems, prevalence, and an alternative. *Journal of Wildlife Management*, v.64, n.4, p.921-923, 2000.
- Anderson, D. R. e Burnham, K. P. Avoiding Pitfalls When Using Information-Theoretic Methods. *The Journal of Wildlife Management*, v.66, n.3, p.912-918, 2002.

- Arlot, S. e Celisse, A. A survey of cross-validation procedures for model selection. *Statistics Surveys*, v.4, p.40-79, 2010.
- Arnold, S. F. *Mathematical Statistics*. Englewood Cliffs: Prentice-Hall, 1990. 636p.
- Astrup, R., Coates, K. D., e Hall, E. Finding the appropriate level of complexity for a simulation model: An example with a forest growth model. *Forest Ecology and Management*, v.256, p.1659-1665, 2008.
- Atkinson, C. e Mitchell, A. Rao's Distance Measure. *Sankhya: The Indian Journal of Statistics, Series A*, v.43, n.3, p.345-365, 1981.
- Avery, T. e Burkhart, H. *Forest measurements*. New York: McGraw-Hill, 1983.
- Bailey, R. Individual tree growth derived from diameter distribution models. *Forest Science*, v.26, n.4, p.626-632, 1980.
- Bailey, R. e Dell, T. Quantifying diameter distribution with the Weibull function. *Forest Science*, v.19, n.2, p.97-104, 1973.
- Bain, L. e Antle, C. Estimation of parameters in the Weibull distribution. *Technometrics*, v.9, n.4, p.621-627, 1967.
- Barrio-Anta, M., Castedo-Dorado, F., Dieguez-Aranda, U., Alvarez-Gonzalez, J. G., Parresol, B. R., e Rodriguez Soalleiro, R. Development of a basal area growth system for maritime pine in northwestern Spain using the generalized algebraic difference approach. *Canadian Journal of Forest Research*, v.36, n.6, p.1461-1474, 2006.
- Batista, J. A função Weibull como modelo para a distribuição de diâmetros de espécies arbóreas tropicais. Dissertação de Mestrado, Universidade de São Paulo (ESALQ), Piracicaba, SP, 1989. (Mestrado - dissertação - ESALQ - USP), 116p.
- Batista, J., do Couto, H. Z., e da Silva Filho, D. *Quantificação de Recursos Florestais: Árvores, Arvoredos e Florestas*. São Paulo: Oficina de Texto, 2014. NO PRELO.
- Batista, J., Vettorazzi, C., e Couto, H. Relatório da Fase III: Levantamento Regional do Estoque – Relatório Final. Projeto: Levantamento do Estoque de Palmiteiro (*Euterpe edulis*) na Região do Vale do Ribeira, Departamento de Ciências Florestais (ESALQ/USP) - Instituto de Pesquisas e Estudos Florestais (IPEF) - Fundação Florestal (Secretaria do Estado do Meio-Ambiente), Piracicaba, 2000. 40p.

- Belyea, H. C. *Forest Measurement*. New York/London: John Wiley & Sons - Chapman & Hall, 1931. 319p.
- Berger, J. O. e Wolpert, R. L. *The likelihood principle*, volume 6 of *Lecture Notes - Monograph Series*, Gupta, Shanti S. (ed.). Hayward: Institute of Mathematical Statistics, 1988.
- Berger, R. e Lawrence, K. Estimating Weibull parameters by linear and nonlinear regression. *Technometrics*, v.16, n.4, p.617-619, 1974.
- Bi, H., Long, Y., Turner, J., Lei, Y., Snowdon, P., Li, Y., Harper, R., Zerihun, A., e Ximenes, F. Additive prediction of aboveground biomass for *Pinus radiata* (D. Don) plantations. *FOREST ECOLOGY AND MANAGEMENT*, v.259, n.12, p.2301-2314, 2010.
- Birnbaum, A. On the foundation of statistical inference. *Journal of the American Statistical Association*, v.57, n.298, p.269-326, 1962.
- Birnbaum, A. More on Concepts of Statistical Evidence. *Journal of the American Statistical Association*, v.67, n.340, p.858- 861, 1972.
- Bitterlich, W. *The Relascope Idea: Relative Measurement in Forestry*. Slough, England: Commonwealth Agricultural Bureaux, 1984. 242p.
- Bjornstad, J. F. Predictive Likelihood: A Review. *Statistical Science*, v.5, n.2, p.242-254, 1990.
- Bliss, C. e Reinker, K. A lognormal approach to diameter distributions in even-aged stands. *Forest Science*, v.10, n.3, p.350-360, 1964.
- Bonou, W., R. Kakai, G., Assogbadjo, A., Fonton, H., e Sinsin, B. Characterisation of *Azelia africana* Sm. habitat in the Lama forest reserve of Benin. *Forest Ecology and Management*, v.258, n.7, p.1084-1092, 2009.
- Borders, B. E., Wang, M., e Zhao, D. Problems of Scaling Plantation Plot Diameter Distributions to Stand Level. *Forest Science*, v.54, n.3, p.349-355, 2008.
- Bozdogan, H. Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions. *Psychometrika*, v.52, n.3, p.345-370, 1987.
- Bozdogan, H. Akaike's Information Criterion and Recent Developments in Information Complexity. *Journal of Mathematical Psychology*, v.44, p.62-91, 2000.

- Bragg, C. e Henry, N. Modelling stand development for prediction and control in tropical forest management. In Shepherd, K. e Richter, H. (Eds.), *Managing the Tropical Forest*, (pp. 281–297)., Canberra. 1985.
- Bruce, D. e Schumacher, F. X. *Forest mensuration*. London: McGraw-Hill Book Company, 1935.
- Bruce, D. e Schumacher, F. X. *Forest mensuration*. New York; London: McGraw-Hill Book Company, 1942.
- Buckland, S. T., Burnham, K. P., e Augustin, N. H. Model Selection: An Integral Part of Inference. *Biometrics*, v.53, n.2, p.603-618, 1997.
- Burk, T. e Burkhart, H. Diameter distribution and yield of natural stands of loblolly pine. Relatório Técnico FWS-1-84, School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University, Blacksburg, 1984.
- Burk, T. e Newberry, J. A simple algorithm for moment-based recovery of Weibull distribution parameters. *Forest Science*, v.30, n.2, p.329-332, 1984.
- Burkhart, H. Data collection and modeling approaches for forest growth and yield prediction. In H. Chapell e D. M. (eds.) (Eds.), *Predicting forest growth and yield: current issues, future prospects* (pp. 3–16). College of Forest Resources, University of Washington, Institute of Forest Resources, Contribution Number 58, 1987. 95p.
- Burkhart, H., Cao, Q. V., e Ware, K. D. A comparison of growth and yield prediction models for loblolly pine. Relatório Técnico FWS-2-81, School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University, Blacksburg, 1981.
- Burnham, K. P. e Anderson, D. R. *Model selection and multimodel inference: a practical information-theoretic approach*. New York: Springer-Verlag, 2002.
- Burnham, K. P. e Anderson, D. R. Multimodel Inference Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, v.33, n.2, p.261-304, 2004.
- Burnham, P. e Anderson, D. Kullback-Leibler information as a basis for strong inference in ecological studies. *Wildlife Research*, v.28, p.111-119, 2001.
- Campos, J. e Turnbull, K. Um sistema para estimar a produção por classe de diâmetro e sua aplicação na interpretação de efeitos de desbaste. *Revista Árvore*, v.5, n.1, p.1-16, 1981.

- Cao, Q., Burkhart, H., e Max, T. Evaluation of two methods for cubic-volume prediction of Loblolly Pine to any merchantable limit. *Forest Science*, v.26, n.1, p.71-80, 1980.
- Cao, Q. V. A Method to Distribute Mortality in Diameter Distribution Models. *Forest Science*, v.43, n.3, p.435-442, 1997.
- Cao, Q. V. Predicting Parameters of a Weibull Function for Modeling Diameter Distribution. *Forest Science*, v.50, n.5, p.682-685, 2004.
- Cao, Q. V. e Baldwin, V. C. A New Algorithm for Stand Table Projection Models. *Forest Science*, v.45, n.4, p.506-511, 1999.
- Carmack, P. S., Spence, J. S., e Schucany, W. R. Generalised correlated cross-validation. *Journal of Nonparametric Statistics*, v.24, n.2, p.269-282, 2012.
- Chambers, J., Cleveland, W., Kleiner, B., e Tukey, P. *Graphical methods for data analysis*. Pacific Grove: Wadsworth & Brooks/Cole, 1983.
- Chapman, D. C. e Weatherhead, D. J. Some Mensurational Formulas Associated with Balanced Diameter Distributions. *Forest Science*, v.30, n.3, p.682-684, 1984.
- Chapman, H. H. *Forestry: an Elementary Treatise*. Chicago: American Lumberman, 1912.
- Chapman, H. H. *Forest valuation*. New York: John Wiley & Sons, 1915.
- Chapman, H. H. *Forest Mensuration* (2 ed.). New York / London: John Wiley & Sons / Chapman & Hall, 1921.
- Chapman, H. H. *Forest Mensuration*. New York: John Wiley & Sons, 1924.
- Chapman, H. H. e Demeritt, D. B. *Elements of forest mensuration*. Albany: J. B. Lyon Company, 1932.
- Chapman, H. H. e Meyer, W. H. *Forest mensuration*. New York: McGraw-Hill, 1949.
- Chevrou, R. de Lioucourt's law and the truncated law. *Canadian Journal of Forest Research*, v.20, p.1933-1946, 1990.
- Claeskens, G. e Hjort, N. L. *Model Selection and Model Averaging*. Cambridge Series In Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 2008. 312p.

- Clutter, J., Fortson, J., Pienaar, L., Brister, G., e Bailey, R. *Timber management: a quantitative approach*. New York: John Wiley & Sons, 1983.
- Cohen, A. Maximum likelihood estimation in the Weibull distribution based on complete and censored samples. *Technometrics*, n.7, p.570-601, 1965.
- Coomes, D. A. e Allen, R. B. Mortality and tree-size distributions in natural mixed-age forests. *Journal of Ecology*, v.95, n.1, p.27-40, 2007.
- Couto, H. *Distribuição de diamétrica em plantações de Pinus caribaea var. caribaea*. Tese de Doutorado, Universidade de São Paulo, Piracicaba, 1980. (Livre Docência - tese - ESALQ - USP).
- Cowles, M. e Davis, C. On the Origins of the .05 Level of Statistical Significance. *American Psychologist*, v.37, n.5, p.553-558, 1982.
- Cox, D. Biometrika: the first 100 years. *Biometrika*, v.88, n.1, p.3-11, 2001.
- Cramér, H. *Elementos da teoria de probabilidade e algumas de suas aplicações*. São Paulo: Editora Mestre Jou, 1955.
- Davies, S. L., Neath, A. A., e Cavanaugh, J. E. Cross validation model selection criteria for linear regression based on the Kullback-Leibler discrepancy. *Statistical Methodology*, v.2, p.249-266, 2005.
- de Liocourt, F. On the improvement of fir forests through selective management (De l'amenagement des sapinières). The School of Natural Resources, University of Missouri-Columbia, 2001. F. de Liocourt, De l'amenagement des sapinières, Bulletin trimestriel, Société forestière de Franche-Comté et Belfort, juillet 1898, pp. 396-409. English translation 2001 by Maria Nygren. Technical review by Pekka Nygren and David R. Larsen.
- de Vries, P. *Sampling Theory for Forest Inventory: A Teach-Yourself Course*. Berlin: Springer-Verlag, 1986. 399p.
- Dell, T. Why not the Weibull for diameter distribution modeling? In *Statistics in Forestry Research, Proceedings of the meeting of IUFRO Subject Group S6.02*, (pp. 49-50)., Vancouver. 1973.
- Dixon, R., Meldahl, R., Ruark, G., e Warren, W. *Process Modeling of Forest Growth Responses to Environmental Stress*. Portland: Timber Press, 1990. 441P.
- Draper, N. e Smith, H. *Applied Regression Analysis* (2 ed.). New York: John Wiley & Sons, 1981.

- Dubey, S. Some percentile estimators for Weibull parameters. *Technometrics*, v.9, n.1, p.119-129, 1967.
- Durant, W. *The Age of Voltaire. The Story of Civilization: Part IX*. New York: Simon and Schuster, 1965.
- Edwards, A. R.A. Fisher twice professor of Genetics: London and Cambridge or "a fairly well-known geneticist". *Biometrics*, v.40, p.897-904, 1990.
- Edwards, A. *Likelihood: Expanded Edition*. London/Baltimore: The John Hopkins University Press, 1992.
- Edwards, A. W. F. Fiducial Probability. *The Statistician*, v.25, n.1, p.15-35, 1976.
- Ek, A. R., Issos, J. N., e Bailey, R. L. Solving for Weibull Diameter Distribution Parameters to Obtain Specified Mean Diameters. *Forest Science*, v.21, n.3, p.290-292, 1975.
- Fernow, B. E. *A Brief History of Forestry in Europe, The United States, and Other Countries: a course lectures delivered before the Yale Forestry School*. New Haven: The Price, Lee & Adkins Co. Print, 1907.
- Fienberg, S. E. e Tanur, J. M. Reconsidering the Fundamental Contributions of Fisher and Neyman on Experimentation and Sampling. *International Statistical Review / Revue Internationale de Statistique*, v.64, n.3, p.237-253, 1996.
- Finger, C. *Fundamentos de Biometria Florestal*. Santa Maria: Universidade Federal de Santa Maria / CEPEF / FATEC, 1992. 269p.
- Fisher, R. Biometry. *Biometrics*, v.4, n.3, p.217-219, 1948.
- Fisher, R., Corbet, A., e Williams, C. The relation between number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, v.12, n.1, p.42-58, 1943.
- Fisher, R. A. On the Mathematical Foundations of Theoretical Statistics. *Philosophical Transactions of the Royal Society of London, Series A*, v.222, p.309-368, 1922.
- Fisher, R. A. *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd., 1925.
- Fisher, R. A. Theory of Statistical Estimation. *Proceedings of the Cambridge Philosophical Society*, v.22, p.700-725, 1925.

- Flewelling, J. e Pienaar, L. Multiplicative regression with lognormal errors. *Forest Science*, v.27, n.2, p.281-289, 1981.
- Forest History Society. Forestry Lectures of Carl Alwin Schenck, 2008. URL: http://www.foresthistory.org/ead/Schenck_Carl_Alwin_Forestry_Lectures.html, acesso em 14/04/2012.
- Forest History Society. Bernhard E. Fernow (1851-1923): 3rd Chief of the U.S. Division of Forestry (1886-1898), 2009. URL: <http://www.foresthistory.org/ASPNET/People/Fernow/Fernow.aspx>, acesso em 25/10/2012.
- Forest History Society Library and Archives. Guide to Forestry Lectures of Carl Alwin Schenck, 1904 - 1909, 2008. URL: http://www.foresthistory.org/research/Biltmore_Project/Schenck_Lectures.html, acesso em 25/10/2012.
- Forster, M. e Sober, E. How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions. *The British Journal for the Philosophy of Science*, v.45, n.1, p.1-35, 1994.
- Forster, M. R. Key Concepts in Model Selection: Performance and Generalizability. *Journal of Mathematical Psychology*, v.44, p.205-231, 2000.
- Fortin, M., Ung, C.-H., Archambault, L., e Bégin, J. Calibrating a Generalized Diameter Distribution Model with Mixed Effects. *Forest Science*, v.52, n.6, p.650-658, 2006.
- Freese, F. Testing Accuracy. *Forest Science*, v.6, n.2, p.139-145, 1960.
- Freese, F. *Elementary Forest Sampling*. USDA, Forest Service, Agricultural Handbook No. 232, 1962. 87p.
- Freese, F. *Elementary Statistical Methods for Foresters*. USDA, Forest Service, Agricultural Handbook No. 317, 1967. 91p.
- Furnival. An index for comparing equations used in constructing volume tables. *Forest Science*, v.7, n.4, p.337-341, 1961.
- Garcia, O. Simplified method-of-moments estimation for the Weibull distribution. *New Zealand Journal of Forestry Science*, v.11, n.3, p.304-6, 1981.
- Graves, H. S. *Forest Mensuration*. New York/London: John Wiley & Sons/Chapman & Hall, 1906.

- Greene, W. H. *Econometric Analysis*. Upper Saddle River: Prentice Hall, 2000. 1004p.
- Gregoire, T. G. e Reynolds, M. R. Accuracy Testing and Estimating Alternatives. *Forest Science*, v.34, n.2, p.302-320, 1988.
- Gross, A. Monte Carlo comparisons of parameter estimators of the 2-parameter Weibull distribution. *IEEE Transactions on Reliability*, v.R - 26, n.5, p.365-358, 1977.
- Hacking, I. *Logic of statistical inference*. Cambridge: Cambridge at the University Press, 1965.
- Hafley, W. e Schreuder, H. Statistical distributions for fitting diameter and height data in even-aged stands. *Canadian Journal of Forest Research*, v.7, n.3, p.481-487, 1977.
- Hafley, W. L. e Buford, M. A. A Bivariate Model for Growth and Yield Prediction. *Forest Science*, v.21, n.1, p.237-247, 1985.
- Hanisch, K. H. e Stoyan, D. Remarks on statistical inference and prediction for a hard-core clustering model. *Statistics*, v.14, p.559-567, 1983.
- Hann, D., Walters, D., e Scrivani, J. Incorporating crown ratio into prediction equations for Douglas-fir stem volume. *Canadian Journal of Forest Research*, v.17, p.17-22, 1987.
- Harter, H. e Moore, A. Maximum-likelihood estimation of the parameters of gamma and Weibull population from complete and from censored samples. *Technometrics*, v.7, n.4, p.639-643, 1965.
- Heilbron, J. Introductory Essay. In T. Frängsmyr, J. Heilbron, e R. E. Rider (Eds.), *The Quantifying Spirit in the 18th Century* (pp. 2-25). Berkeley: University of California Press, 1990.
- Hilborn, R. e Mangel, M. *The ecological detective: confronting models with data*. Number 28 in Monographs in Population Biology. Princeton: Princeton University Press, 1997.
- Hill, I. Statistical Society of London - Royal Statistical Society: The First 100 years: 1834-1934. *Journal of the Royal Statistical Society, Series A (General)*, v.147, n.2, The 150th Anniversary of the Royal Statistical Society (1984), p.130-139, 1984.

- Hobbs, N. e Hilborn, R. Alternatives to statistical hypothesis testing in Ecology: a guide to self teaching. *Ecological Applications*, v.16, n.1, p.5-19, 2006.
- Hocking, R. The Analysis and Selection of Variables in Linear Regression. *Biometrics*, v.32, n.1, p.1-49, 1976.
- Hoel, P. G., Port, S. C., e Stone, C. J. *Introduction do Probability Theory*. Atlanta: Houghton Mifflin, 1971. 258p.
- Hough, A. F. Some Diameter Distribution in Forest Stands of Nortwestern Pennsylvania. *Journal of Forestry*, v.30, n.8, p.933-943, 1932.
- Husch, B. *Forest mensuration and statistics*. New York: Ronald Press Co, 1963.
- Husch, B., Miller, C. I., e Beers, T. W. *Forest Mensuration*. New York: John Wiley & Sons, 1982.
- Israel, J. I. *Radical Enlightenment: Philosophy and the Making of Modernity 1650-1750*. Oxford: Oxford University Press, 2001.
- Johnson, J. B. e Omland, K. S. Model selection in ecology and evolution. *TRENDS in Ecology and Evolution*, v.19, n.2, p.101-108, 2004.
- Johnson, N. e Kotz, S. *Continous univariate distributions*, volume 2. Boston: Houghton Mifflin, 1970.
- Juntunen, M. Modelling Tree and Stand Characteristics of Lodgepole Pine (*Pinus contorta*) Plantations in Iceland. Dissertação de Mestrado, University of Eastern Finland, Faculty of Science and Forestry, Forest Planning and Economics, 2010.
- Kakaï, R. G., Akpona, T. J. D., Assogbadjo, A. E., Gaoué, O. G., Chakeredza, S., Gnanglè, P. C., Mensah, G. A., e Sinsin, B. Ecological adaptation of the shea butter tree (*Vitellaria paradoxa* C.F. Gaertn.) along climatic gradient in Bénin, West Africa. *African Journal of Ecology*, v.49, n.4, p.440-449, 2011.
- Kangas, A. e Maltamo, M. Calibrating Predicted Diameter Distribution with Additional Information. *Forest Science*, v.46, n.3, p.390-396, 2000.
- Kangas, A. e Maltamo, M. (Eds.). *Forest inventory : methodology and applications*. Dordrecht: Springer Verlag, 2006.
- Kempton, R. e Taylor, L. Log-Series and Log-Normal Parameters as Diversity Discriminants for the Lepidoptera. *Source: Journal of Animal Ecology*, v.43, n.2, p.381-399, 1974.

- Kempton, R. A. A Generalized form of Fisher's Logarithmic Series. *Biometrika*, v.62, n.1, p.29-38, 1975.
- Kendall, M. G. Where shall the history of statistics begin! *Biometrika*, v.47, p.447-49, 1960.
- Kendall, M. Ronald Aymer Fisher. *Biometrika*, v.50, n.1/2, p.1-15, 1963.
- Kerr, G. The management of silver fir forests: de Liocourt (1898) revisited. *Forestry*, v.87, n.1, p.29-38, 2014.
- Knoebel, B. R. e Burkhart, H. E. A Bivariate Distribution Approach to Modeling Forest Diameter Distributions at Two Points in Time. *Biometrics*, v.47, p.243-251, 1991.
- Kozak, A. e Kozak, R. Does cross validation provide additional information in the evaluation of regression models? *Canadian Journal of Forest Research*, v.33, p.976-987, 2003.
- Krishnamoorthy, K. *Handbook of Statistical Distributions with Applications*. Boca Raton: Chapman & Hall, 2006.
- Krug, A., Nordheim, E., e Giese, R. Determining initial values for parameter of Weibull model: a case study. *Forest Science*, v.30, n.3, p.573-581, 1984.
- Kruskal, W. e Mosteller, F. Representative Sampling, IV: The History of the Concept in Statistics, 1895-1939. *International Statistical Review / Revue Internationale de Statistique*, v.48, n.2, p.169-195, 1980.
- Kuha, J. AIC and BIC : Comparisons of Assumptions and Performance. *Sociological Methods & Research*, v.33, n.2, p.188-229, 2004.
- Kullback, S. *Information Theory and Statistics*. New York/London: John Wiley & Sons/Chapman & Hall, 1959.
- Kullback, S. e Leibler, R. A. On Information and Sufficiency. *The Annals of Mathematical Statistics*, v.22, n.1, p.79-86, 1951.
- Lai, J., Coomes, D. A., Du, X., Hsieh, C.-f., Sun, I.-F., Chao, W.-C., Mi, X., Ren, H., Wang, X., Hao, Z., e Ma, K. A general combined model to describe tree-diameter distributions within subtropical and temperate forest communities. *Oikos*, v.122, n.11, p.1636-1642, 2013.
- Lane, S. E., Robinson, A. P., e Baker, T. G. The functional regression tree method for diameter distribution modelling. *Canadian Journal of Forest Research*, v.40, p.1870-1877, 2010.

- Lappi, J. Calibration of height and volume equations with random parameters. *Forest Science*, v.37, n.3, p.781-801, 1991.
- Lappi, J. e Bailey, R. A height prediction model with random stand and tree parameters: an alternative to traditional site index methods. *Forest Science*, v.34, n.4, p.907-927, 1988.
- Leak, W. B. The J-shaped Probability Distribution. *Forest Science*, v.11, n.4, p.405-409, 1965.
- Lehmann, E. The Fisher, Neyman-Pearson theories of testing hypotheses: one theory or two? *Journal of the American Statistical Association*, v.88, n.424, p.1242-1249, 1993.
- Lehmann, E. L. *Fisher, Neyman, and the Creation of Classical Statistics*. New York: Springer, 2011.
- Leite, H. G., Binoti, D. H. B., Guimarães, D. P., da Silva, M. L. M., e Garcia, S. L. R. Avaliação do ajuste das funções Weibull e hiperbólica a dados de povoamentos de eucalipto submetidos a desbaste. *Revista Árvore*, v.34, n.2, p.305-311, 2010.
- Leiva, V., Ponce, M. G., e Marchant, Bustos, C. O. Fatigue Statistical Distributions Useful for Modeling Diameter and Mortality of Trees: Distribuciones estadísticas de fatiga útiles para modelar diámetro y mortalidad de árboles. *Revista Colombiana de Estadística*, v.35, n.3, p.349-370, 2012.
- Lekwadi, S. O., Nemesova, A., Lynch, T., Phillips, H., Hunter, A., e Mac Siurtain, M. Site classification and growth models for Sitka spruce plantations in Ireland. *FOREST ECOLOGY AND MANAGEMENT*, v.283, p.56-65, 2012.
- Lele, S. R. *The Nature of Scientific Evidence: Statistical, Philosophical, and Empirical Considerations*, chapter Evidence Functions and the Optimality of the Law of Likelihood, (pp. 191–216). Chicago: The University of Chicago Press. 2004.
- Lindsay, S. R., Wood, G. R., e Woollons, R. C. Modelling the diameter distribution of forest stands using the Burr distribution. *Journal of Applied Statistics*, v.23, n.6, p.609-619, 1996.
- Linhart, H. e Zucchini, W. *Model Selection*. New York: John Wiley & Sons, 1986.
- Little, S. Weibull diameter distributions for mixed stands of Western conifers. *Canadian Journal of Forest Research*, v.13, n.1, p.85-88, 1983.

- Lloyd, P. J. American, German and British Antecedents to Pearl and Reed's Logistic Curve. *Population Studies*, v.21, n.2, p.99-108, 1967.
- Lowood, H. H. The Calculating Forester: Quantification, Cameral Science, and the Emergence of Scientific Forestry Management. In T. Frängsmyr, J. Heilbron, e R. E. Rider (Eds.), *The Quantifying Spirit in the 18th Century* (pp. 315–342). Berkeley: University of California Press, 1990.
- Machado, S. d. A., Augustynczyk, Alan Lessa Derci and Rodrigo Nascimento, G. M., Figura, M. A., da Silva, L. C. R., Miguel, E. P., e Téo, S. J. Distribuição diamétrica de *Araucaria angustifolia* (Bert.) O. Ktze. em um Fragmento de floresta ombrófila mista. *Scientia Agraria*, v.10, n.2, p.103-110, 2009.
- Magurran, A. *Ecological Diversity and its Measurement*. London: Croom Helm, 1988.
- Maltamo, M., Kangas, A., Uuttera, J., Torniainen, T., e Saramäki, J. Comparison of percentile based prediction methods and the Weibull distribution in describing the diameter distribution of heterogeneous Scots pine stands. *Forest Ecology and Management*, v.133, p.263-274, 2000.
- McMullen, L. Student as a Man. *Biometrika*, v.30, n.3/4, p.205-210, 1939.
- Mehtätalo, L., Gregoire, T. G., e Burkhart, H. E. Comparing strategies for modeling tree diameter percentiles from remeasured plots. *Environmetrics*, v.19, n.5, p.529-548, 2008.
- Menon, M. Estimation of the shape and scale parameters of the Weibull distribution. *Technometrics*, v.5, n.2, p.175-182, 1963.
- Meyer, H. A. Structure, Growth, and Drain in Balanced Uneven-Aged Forests. *Journal of Forest*, v.50, n.2, p.85-92, 1952.
- Meyer, H. A., Recknagel, A. B., Stevenson, D. D., e Bartoo, R. A. *Forest Management* (2 ed.). New York: Ron, 1961.
- Meyer, H. A. e Stevenson, D. D. The Structure and Growth of Virgin Beech-Birch-Maple-Hemlock Forests in Northern Pennsylvania. *Journal of agricultural research*, v.67, n.12, p.465-484, 1943.
- Mood, A., Graybill, F., e Boes, D. *Introduction to the theory of statistics*. Singapore: McGraw-Hill, 1974. 564p.
- Moser, J. Dynamics of an uneven-aged forest stand. *Forest Science*, v.18, n.3, p.184-191, 1972.

- Moser, J. Specification of density for the inverse J-shaped diameter distribution. *Forest Science*, v.22, n.2, p.177-180, 1976.
- Mosteller, F. e Tukey, J. W. *Data Analysis and Regression*. Addison-Wesley, 1977.
- Muller-Landau, H. C., Condit, R. S., Harms, K. E., Marks, C. O., Thomas, S. C., Bunyavejchewin, S., e outros. Comparing tropical forest tree size distributions with the predictions of metabolic ecology and equilibrium models. *Ecology Letters*, v.9, p.589-602, 2006.
- National Wildlife Federation. Conservation Hall of Fame: Gifford Pinchot, 2012. URL: <http://www.nwf.org/About/History-and-Heritage/Conservation-Hall-of-Fame/Pinchot.aspx>, acesso em 14/04/2012.
- Namaalwa, J., Eid, T., e Sankhayan, P. A multi-species density-dependent matrix growth model for the dry woodlands of Uganda. *FOREST ECOLOGY AND MANAGEMENT*, v.213, n.1-3, p.312-327, 2005.
- Nanang, D. M. Suitability of the Normal, Log-normal and Weibull distributions for fitting diameter distributions of neem plantations in Northern Ghana. *Forest Ecology and Management*, v.103, n.1, p.1-7, 1998.
- Nanos, N. e Montero, G. Spatial prediction of diameter distribution models. *Forest Ecology and Management*, v.161, n.1-3, p.147-158, 2002.
- Nascimento, R. G. M., Machado, S. d. A., Figueiredo Filho, A., e Higuchi, N. Modelo de projeção por classe diamétrica para florestas nativas: enfoque na função probabilística de Weibull. *Pesquisa Florestal Brasileira*, v.32, n.70, p.209-219, 2012.
- Neeff, T., Dutra, L. V., dos Santos, J. R., Freitas, C. d. C., e Araujo, L. S. Tropical forest stand table modelling from SAR data. *Forest Ecology and Management*, v.186, n.1-3, p.159-170, 2003.
- Nelson, T. Diameter distribution and growth of Loblolly Pine. *Forest Science*, v.10, n.1, p.105-114, 1964.
- Nepal, S. K. e Somers, G. L. A Generalized Approach to Stand Table Projection. *Forest Science*, v.38, n.1, p.120-133, 1992.
- Neter, J., Wasserman, W., e Kutner, M. *Applied Linear Statistical Models*. Homewood: Richard D. Irwin, 1990. 1181p.

- Neyman, J. On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society*, v.97, n.4, p.558-625, 1934.
- Neyman, J. e Pearson, E. On the problem of the most efficient test of statistical hypotheses. *Philosophical Transactions of the Royal Statistical Society A*, v.231, p.289–337, 1933.
- Nord-Larsen, T. e Cao, Q. V. A diameter distribution model for even-aged beech in Denmark. *Forest Ecology and Management*, v.213, p.218-225, 2006.
- Osborne, J. e Schumacher, F. X. The construction of normal-yield and stand table for even-aged stands. *Journal of Agricultural Research*, v.51, n.6, p.547-564, 1935.
- Palahí, M., Pukkala, T., Blasco, E., e Trasobares, A. Comparison of beta, Johnson's SB, Weibull and truncated Weibull functions for modeling the diameter distribution of forest stands in Catalonia (north-east of Spain). *European Journal of Forest Research*, v.167, n.4, p.563-571, 2007.
- Pearson, E. Student as a Statistician. *Biometrika*, v.30, n.3/4, p.210-250, 1939.
- Pearson, E. S. Studies in the History of Probability and Statistics XIV Some Incidents in the Early History of Biometry and Statistics. *Biometrika*, v.52, n.1/2, p.3-18, 1965.
- Pinchot, G. e Graves, H. S. *The White Pine: a study with tables of volume and yield*. The Century Co., 1896. 102p.
- Pinder, J., Wiener, J., e Smith, M. The Weibull distribution: a new method of summarizing survivorship data. *Ecology*, v.59, n.1, p.175-179, 1978.
- Podlaski, R. Characterization of diameter distribution data in near-natural forests using the Birnbaum-Saunders distribution. *Canadian Journal of Forest Research*, v.38, p.518-527, 2008.
- Poudel, K. P. e Cao, Q. V. Evaluation of Methods to Predict Weibull Parameters for Characterizing Diameter Distributions. *Forest Science*, v.59, n.2, p.243-252, 2013.
- Prodan, M. *Forest Biometrics*. Oxford; New York: Pergamon Press, 1968.
- Qin, J., Cao, Q. V., e Blouin, D. C. Projection of a diameter distribution through time. *Canadian Journal of Forest Research*, v.37, p.188-194, 2007.

- Rao, C. R. R.A. Fisher: the founder of Modern Statistics. *Statistical Science*, v.7, n.1, p.34-48, 1992.
- Rathbun, L. C., LeMay, V., e Smith, N. Diameter growth models for mixed-species stands of Coastal British Columbia including thinning and fertilization effects. *ECOLOGICAL MODELLING*, v.222, n.14, SI, p.2234-2248, 2011. Conference of the International-Society-for-Ecological-Modelling, Quebec City, CANADA, OCT 06-09, 2009.
- Recknagel, A. B. e Bentley, J. *Forest Management*. New York/London: John Wiley & Sons/Chapman & Hall, 1919.
- Rennolls, K. e Wang, M. A new parameterization of Johnson's SB distribution with application to fitting forest tree diameter data. *Canadian Journal of Forest Research*, v.35, n.3, p.575-579, 2005.
- Reynolds, M., Burk, T., e Huang, W. Goodness-of-fit tests and model selections procedures for diameter distribution models. *Forest Science*, v.34, n.2, p.373-399, 1988.
- Reynolds Jr., M. R. Estimating the Error in Model Predictions. *Forest Science*, v.30, n.2, p.454-469, 1984.
- Richards, S. A. Testing Ecological Theory using the Information-Theoretic Approach: Examples and Cautionary Results. *Ecology*, v.86, n.10, p.2805-2810, 2005.
- Robinson, A. Preserving correlation while modelling diameter distributions. *Canadian Journal of Forest Research*, v.21, p.221-232, 2004.
- Robinson, A. P. e Froese, R. E. Model validation using equivalence tests. *Ecological Modelling*, v.176, p.349-358, 2004.
- Ronan, C. A. *História Ilustrada da Ciência da Universidade de Cambridge: da Renascença à Revolução Científica*. História Ilustrada da Universidade de Cambridge, vol. III. São Paulo: Círculo do Livro, 1983.
- Ross, S. *Introduction to Probability Models* (4 ed.). Boston: Academic Press, 1989.
- Royall, R. *Statistical evidence: a likelihood paradigm*. Boca Raton: Chapman & Hall, 1997.

- Rubin, B. D., Manion, P. D., e Faber-Langendoen, D. Diameter distributions and structural sustainability in forests. *Forest Ecology and Management*, v.222, p.427-438, 2006.
- Sakamoto, Y., Ishiguro, M., e Kitagawa, G. *Akaike information criterion statistics*. Mathematics and Its Applications. Tokyo/Dordrecht: KTK Scientific Publishers/D. Reidel Publishing Company, 1986.
- Samuelson, P. A. Fitting General Gram-Charlier Series. *The Annals of Mathematical Statistics*, v.14, n.2, p.179-187, 1943.
- Schenck, C. *Forest Mensuration*. The University Press of Swannee Tennessee, 1905.
- Schmelz, D. V. e Lindsey, A. A. Size-class Structure of Old-growth Forests in Indiana. *Forest Science*, v.11, n.3, p.258-264, 1965.
- Schneider, P. R., Finger, C. A. G., Bernett, L. G., Schneider, P. S. P., e Fleig, F. D. Estimativa dos parâmetros da função de densidade probabilística de weibull por regressão aninhada em povoamento desbastado de Pinus taeda L. *Ciência Florestal*, v.18, n.3, 2008.
- Schnur, G. L. Diameter Distributions for Old-Field Loblolly Pine Stands in Maryland. *Journal of Agricultural Research*, p.731-743, 1934.
- Schreuder, H., Gregoire, T., e Wood, G. *Sampling methods for multiresource forest inventory*. New York: John Wiley & Sons, 1993.
- Schreuder, H. T. e Hafley, W. L. A Useful Bivariate Distribution for Describing Stand Structure of Tree Heights and Diameters. *Biometrics*, v.33, n.3, p.471-478, 1977.
- Schreuder, H. T. e Swank, W. Coniferous stands characterized with the Weibull distribution. *Canadian Journal of Forest Research*, v.4, n.4, p.518-523, 1974.
- Schumacher, F. X. e Chapman, R. *Sampling methods in forestry and range management*. Bulletin (Duke University. School of Forestry) 7. Durham: Duke University. School of Forestry, 1942. 213p.
- Schumacher, F. X. e Hall, F. d. Logarithmic expression of timber-tree volume. *Journal of Agricultural Research*, v.47, n.9, p.719-734, 1933.
- Schwarz, G. Estimating the Dimension of a Model. *The Annals of Statistics*, v.6, n.2, p.461-464, 1978.

- Seal, H. L. The Historical Development of The Gauss Linear Model. *Biometrika*, v.54, p.1-24, 1967.
- Shannon, C. E. A Mathematical Theory of Communication. *The Bell System Technical Journal*, v.22, p.379-423, 623-656, 1948.
- Shapiro, S. e Gross, A. *Statistical modeling techniques*. New York: Marcel Dekker, 1981.
- Shiver, B. Sample size for estimating Weibull function parameters for Slash Pine plantations diameter distributions. Relatório Técnico 1985 - 4, Plantation Mangement Research Cooperative, Technical Report, 1985. 21p.
- Shiver, B. e Borders, B. *Sampling techniques for forest resource inventory*. New York: John Wiley & Sons, 1996. 356p.
- Siipilehto, J. Modelling stand structure in young Scots pine dominated stands. *Forest Ecology and Management*, v.257, n.1, p.223-232, 2009.
- Silva, J. *Dynamics of stand structure in fertilized Slash Pine plantations*. Tese de Doutorado, University of Georgia, Athens, 1986. 139p.
- Soares, J., Tomé, M., Skovsgaard, J., e Vanclay, J. Evaluating a growth model for forest management using continuous forest inventory data. *Forest Ecology and Management*, v.71, p.251-265, 1995.
- Soares, P. e Tome, M. A tree crown ratio prediction equation for eucalypt plantations. *ANNALS OF FOREST SCIENCE*, v.58, n.2, p.193-202, 2001.
- Soares, P. e Tome, M. Height-diameter equation for first rotation eucalypt plantations in Portugal. *FOREST ECOLOGY AND MANAGEMENT*, v.166, n.1-3, p.99-109, 2002.
- Sober, E. Instrumentalism, Parsimony, and the Akaike Framework. *Philosophy of Science*, v.69, p.S112-S123, 2002.
- Sober, E. Likelihood, Model Selection, and the Duhem-Quine Problem. *The Journal of Philosophy*, v.C1, n.5, p.221-241, 2004.
- Spurr, S. H. *Forest Inventory*. New York: The Ronald Press, 1952. 476p.
- Stankova, T. e Diéguez-Aranda, U. Diameter Distribution model for Scots Pine Plantations in Bulgaria. *Forestry Ideas*, v.16, n.2, p.155-162, 2010.
- Stigler, S. The problematic unit of Biometrics. *Biometrics*, v.56, p.653-658, 2000.

- Stigler, S. M. *The History of Statistics: The Measurement of Uncertainty before 1900*. Cambridge: The Belknap Press of Harvard University Press, 1986.
- Stigler, S. M. The History of Statistics in 1933. *Statistical Science*, v.11, n.3, p.244-252, 1996.
- Stigler, S. M. The pedigree of International Biometric Society. *Biometrics*, v.63, p.317-321, 2007.
- Stone, M. Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society, Series B (Methodological)*, v.36, n.2, p.111-147, 1974.
- Stone, M. An Asymptotic Equivalence of Choice of Model by Cross-Validation and Akaike's Criterion Author(s): M. Stone. *Journal of the Royal Statistical Society, Series B (Methodological)*, v.39, n.1, p.44-47, 1977.
- Tang, S., Wang, Y., Zhang, L., e Meng, C.-H. A Distribution-Independent Approach to Predicting Stand Diameter Distribution. *Forest Science*, v.43, n.4, p.491-500, 1997.
- Taubert, F., Hartig, F., Dobner, H.-J., e Huth, A. On the Challenge of Fitting Tree Size Distributions in Ecology. *PLoS ONE*, v.8, n.2, p.e58036, 2013. 9p.
- Taylor, L. Aggregation, Variance and Mean. *Nature*, v.189, n.476, p.732-&, 1961.
- Taylor, L. A Power Law Transformation for Aggregated Populations. *Biometrics*, v.17, n.3, p.498-&, 1961.
- Taylor, L. Assessing and Interpreting the Spatial Distribution of insect Populations. *Annual Review of Entomology*, v.29, p.321-357, 1984.
- Taylor, L., Perry, J., Woiwod, I., e Taylor, R. Specificity of the Spatial Power-Law Exponent in Ecology and Agriculture. *Nature*, v.332, n.6166, p.721-722, 1988.
- Thoman, D., Bain, L., e Antle, C. Inferences on the parameters of the Weibull distribution. *Technometrics*, v.11, n.3, p.445-460, 1969.
- Trincado V., G., Quezada, R., e von Gadow, K. A comparison of two stand table projection methods for young *Eucalyptus nitens* (Maiden) plantations in Chile. *Forest Ecology and Management*, v.180, p.443-451, 2003.
- van Deusen, P. C. Fitting Assumed Distributions to Horizontal Point Sample Diameters. *Forest Science*, v.34, n.1, p.146-148, 1986.

- Vanclay, J. e Skovsgaard, J. Evaluating forest growth models. *Forest Ecology and Management*, v.98, p.1-12, 1997.
- Vanclay, J. K. Robust relationships for simple plantation growth models based on sparse data. *FOREST ECOLOGY AND MANAGEMENT*, v.259, n.5, p.1050-1054, 2010.
- Vospersnik, S., Monserud, R. A., e Sterba, H. Do individual-tree growth models correctly represent height:diameter ratios of Norway spruce and Scots pine? *FOREST ECOLOGY AND MANAGEMENT*, v.260, n.10, p.1735-1753, 2010.
- Wang, X., Hao, Z., Zhang, J., Lian, J., Li, B., Ye, J., e Yao, X. Tree size distributions in an old-growth temperate forest. *Oikos*, v.118, n.1, p.25-36, 2009.
- Ward, E. J. A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools. *Ecological Modelling*, v.211, p.1-10, 2008.
- Weibull, W. A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, v.18, n.3, p.293-297, 1951.
- Weisstein, E. W. Log-Series Distribution, sd. MathWorld—A Wolfram Web Resource. <http://mathworld.wolfram.com/Log-SeriesDistribution.html>. Acesso em 20/02/2013.
- Westphal, C., Tremer, N., von Oheimb, G., Hansen, J., von Gadow, K., e Härdtle, W. Is the reverse J-shaped diameter distribution universally applicable in European virgin beech forests? *Forest Ecology and Management*, v.223, n.1-3, p.75-83, 2006.
- Wingo, D. Maximum likelihood estimation of the parameters of the Weibull distribution by modified quasilinearization. *IEEE Trans. Rel. Vol.*, v.R-21, n.2, p.89-93, 1972.
- Wingo, D. Solution of the three-parameter Weibull equations by constrained modified quasilinearization (progressively censored samples). *IEEE Trans. Rel. Vol.*, v.R-22, n.2, p.96-102, 1973.
- Winkenwerder, H. e Clark, E. T. *Handbook of Field and Office Problems in Forest Mensuration*. New York: John Wiley & Sons/Chapman & Hall, 1922.
- Yanagihara, H. e Fujisawa, H. Iterative Bias Correction of the Cross-Validation Criterion. *Scandinavian Journal of Statistics*, v.39, p.116-130, 2012.

- Yang, Y., Monserud, R. A., e Huang, S. An evaluation of diagnostic tests and their roles in validating forest biometric models. *Canadian Journal of Forest Research*, v.34, p.619-629, 2004.
- Yule, G. U. *An Introduction to the Theory of Statistics*. London: C. Griffin, 1919.
- Zabell, S. L. R. A. Fisher and Fiducial Argument. *Statistical Science*, v.7, n.3, p.369-387, 1992.
- Zarnoch, S. e Dell, T. An evaluation of percentile and maximum likelihood estimators of Weibull parameters. *Forest Science*, v.31, n.1, p.260-268, 1985.
- Zeide, B. Exponential diameter distribution: interpretation of coefficients. *Forest Science*, v.30, n.4, p.907-912, 1984.
- Zhang, L., Gove, J. H., Liu, C., e Leak, W. B. A finite mixture of two Weibull distributions for modeling the diameter distributions of rotated-sigmoid, uneven-aged stands. *Canadian Journal of Forest Research*, v.31, n.9, p.1654-1659, 2001.
- Zhang, L., Packard, K. C., e Liu, C. A comparison of estimation methods for fitting Weibull and Johnson's SB distributions to mixed spruce fir stands in northeastern North America. *Canadian Journal of Forest Research*, v.33, n.7, p.1340-1347, 2003.
- Zucchini, W. An Introduction to Model Selection. *Journal of Mathematical Psychology*, v.44, p.41-61, 2000.
- Zutter, B., Oderwald, R., Farrar, R., e Murphy, P. Weibull - a program to estimate parameters of forms of Weibull distribution using complete, censored and truncated data. Relatório Técnico Publication FWS-3-82, School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University, Blacksburg, 1982.