Distribuições de Probabilidade Discretas

Paulo Inácio K.L. Prado e João L.F. Batista

BIE-5781 Modelagem Estatística em Ecologia e Recursos Naturais

Pós-Graduação em Ecologia – USP, Novembro de 2018

Conceitos

- Distribuição Bernoulli
- Distribuição Binomial
- Distribuição Poisson
- Distribuição Geométrica
- Distribuição Binomial Negativa

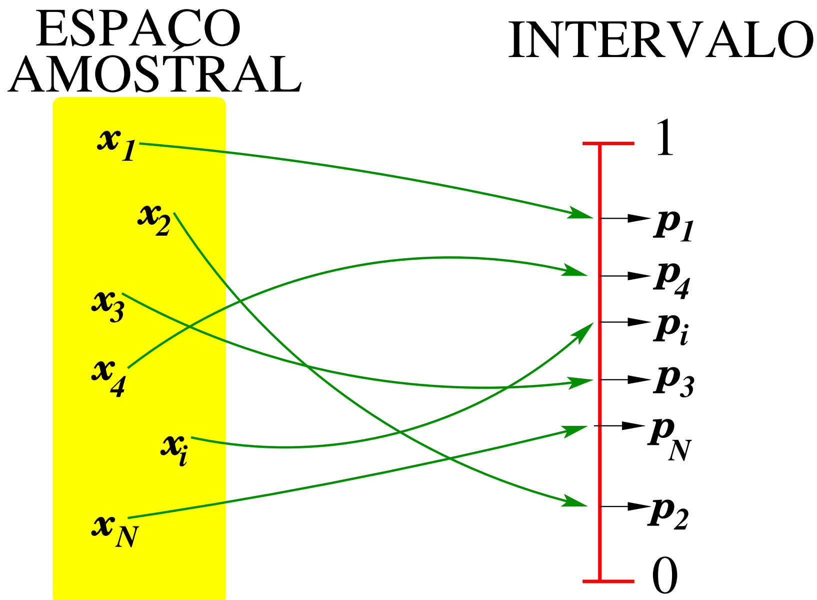
Distribuição de Probabilidades

Definição

uma função matemática

atribui probabilidades a valores do espaço amostral

Esquema



Classes de Variáveis Quantitativas

Variáveis Discretas

Os valores *possíveis* estão no conjunto dos **NÚMEROS NATURAIS (INTEIROS)**.

Variáveis obtidas por:

- enumeração: *e.g.* número de fêmeas em uma ninhada de 3 filhotes
- contagem: *e.g.* número de árvores em uma parcela amostral

Distribuição Bernoulli

Ensaio

- *Bernoulli* é a distribuição de probabilidades mais simples dentre as variáveis discretas.
- Mais que uma distribuição, é interessante vê-la como um ensaio com resultado estocástico.
- A partir desse ensaio aleatório pode se conceituar diversas distribuições de probabilidades de variáveis discretas.

Variável Binária

- Dois resultados possíveis:
  - *fracasso* $(X = 0)$: o evento não ocorre.
  - *sucesso* $(X = 1)$: o evento ocorre.
- Parâmetro: a probabilidade de sucesso (p).

Distribuição Bernoulli: Apresentação Formal

Descrição

Função de Densidade:

$$f(x) = p^x (1-p)^{1-x}, \qquad x = 0, 1.$$

Resultados

- Fracasso: $f(0) = P(X = 0) = (1 - p)$
- Sucesso: $f(1) = P(X = 1) = p$

Função de Distribuição

$$F(x) = (1-p)^{1-x}, \qquad x = 0, 1$$

Momentos

- Esperança: $\mathbf{E}[X] = p$
- Variância: $\mathbf{V}ar[X] = p\,(1-p)$

Distribuição Geométrica

Situação

Sequência de ensaios Bernoulli independentes.

- O único parâmetro é a probabilidade de sucesso por ensaio (p, constante).

Variável

- $X =$*número de fracassos* até o primeiro sucesso
- espaço amostral: $x = 0, 1, 2, \dots, n$

Exemplos:

- Tempo de vida em unidades discretas
- N de inspeções até encontrar algo
- N de carnavais até o divórcio

Gráfico: Distribuição Geométrica

Função de densidade

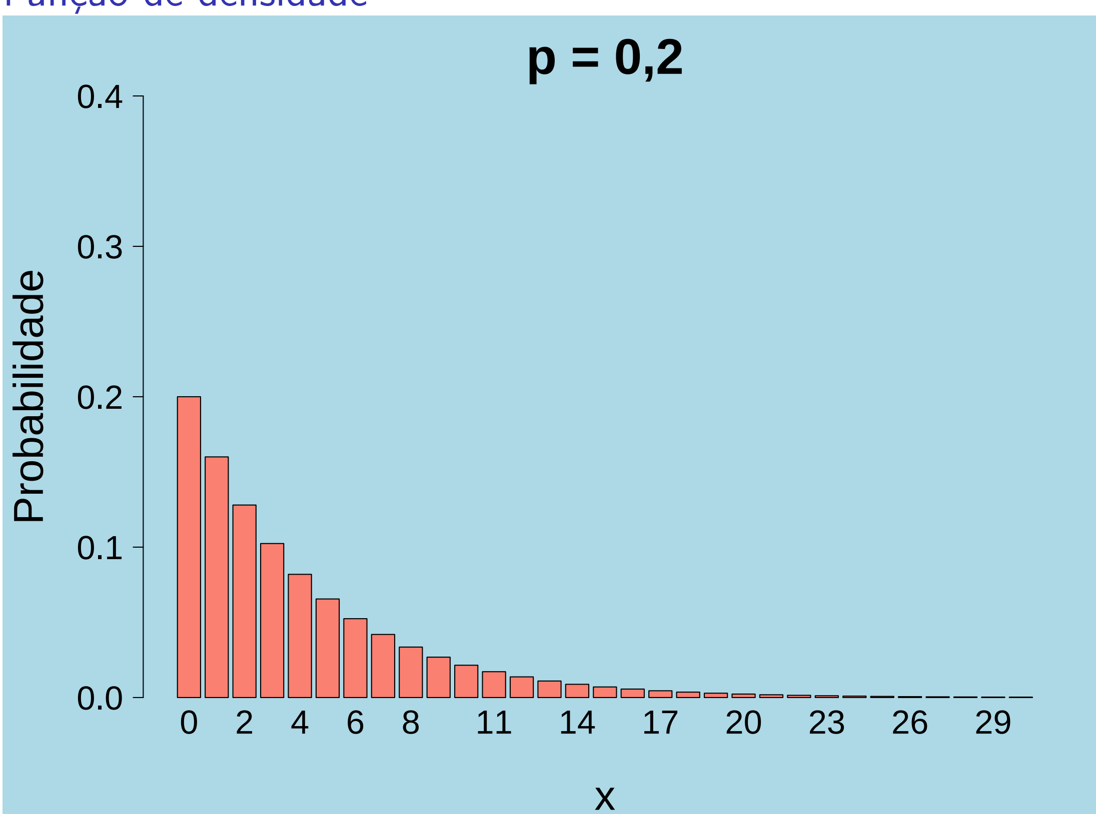


Gráfico: Distribuição Geométrica

Função de densidade

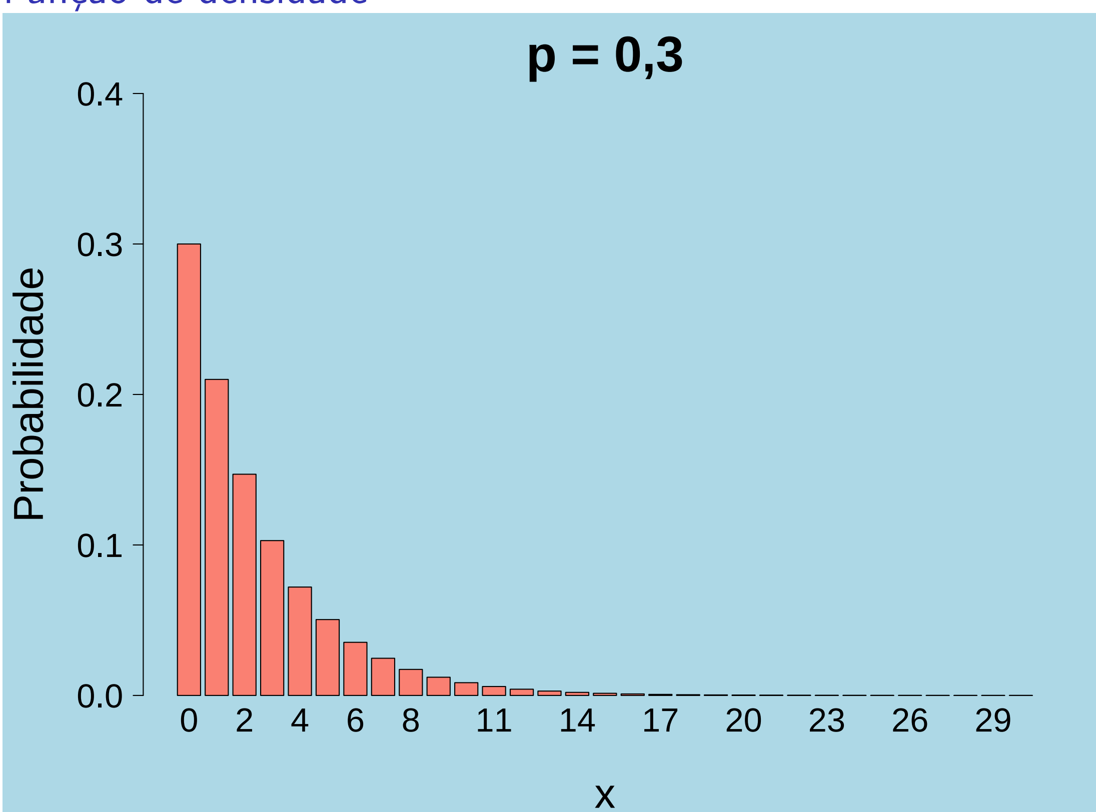
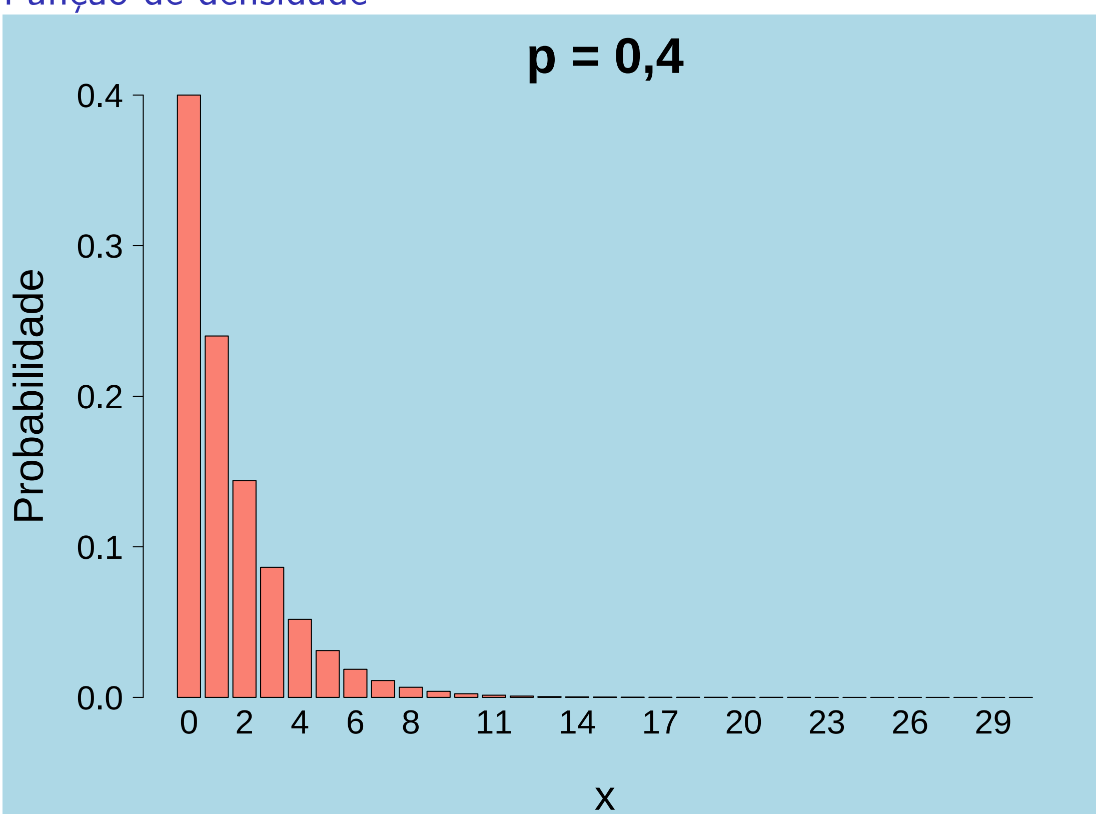


Gráfico: Distribuição Geométrica

Função de densidade



Distribuição Geométrica: Apresentação Formal

Função de densidade

$$f(x) = p(1-p)^x, \qquad x = 0, 1, 2, \ldots$$

Função de distribuição

$$F(x) = 1 - (1-p)^{x+1}, \qquad x = 0, 1, 2, \ldots$$

Momentos

- Esperança:

$$\mathbf{E}[X] = \frac{1}{p}$$

- Variância:

$$\mathbf{V}ar[X] = \frac{1-p}{p^2}$$

Exemplo de Distribuição Geométrica

Situação

Tabela de vida de uma coorte de *Vanellus vanellus* (Haldane 1953)

- 593 aves anilhadas, recontadas a cada ano
- Tempo médio de vida: 2,77 anos
- Probabilidade de morte: $2,77^{-1} \;\approx\; 0,361$ ano^{-1}
- Probabilidade de sobrevivência: $1 - 0,361 \;=\; 0,639$

Distribuição Binomial

Situação

- Uma série de n *de ensaios Bernoulli independentes*
- Parâmetro tamanho da série: n
- Parâmetro probabilidade de sucesso *constante*: p

Variável

- $X =$ o número de sucessos nos n ensaios
- espaço amostral: $x = 0, 1, 2, \ldots, n$

Exemplos:

- N de caras em 10 lançamentos de moeda
- N de mulheres em famílias de 5 filhos
- N de cobaias mortos em bioensaio (dose-resposta)

Gráfico: Distribuição Binomial

Função de densidade

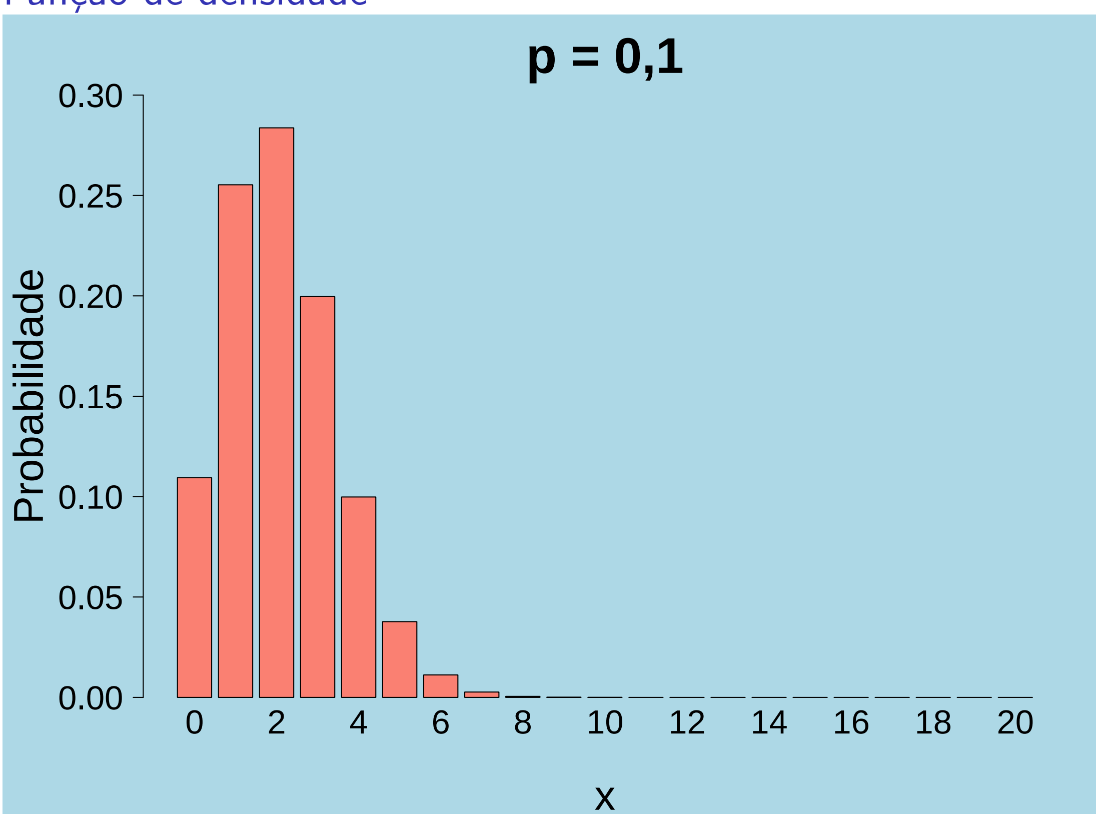


Gráfico: Distribuição Binomial

Função de densidade

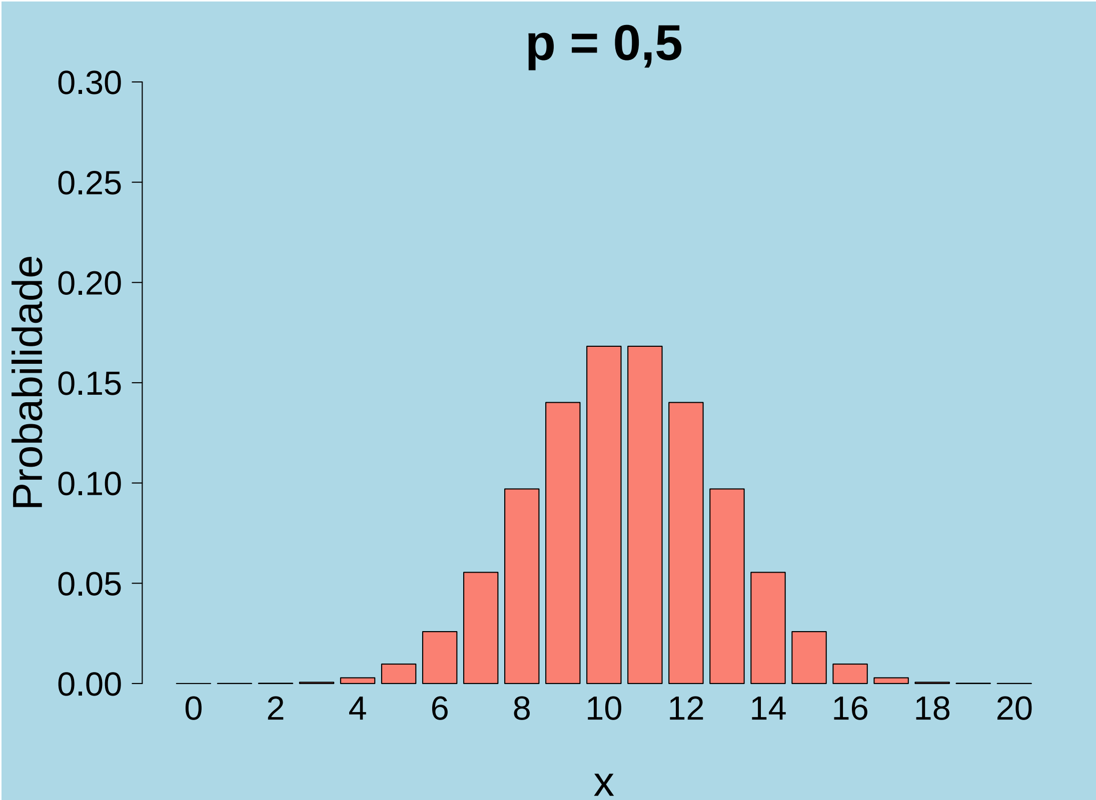
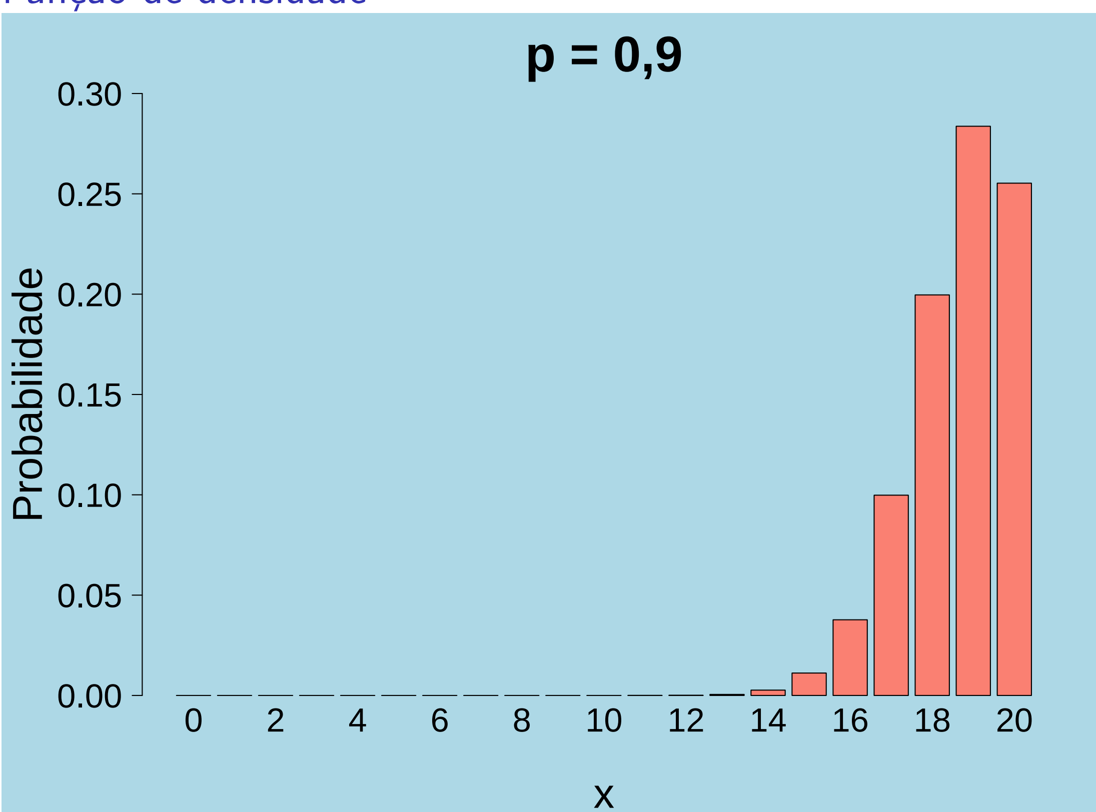


Gráfico: Distribuição Binomial

Função de densidade



Distribuição Binomial: Apresentação Formal

Função de densidade

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, \ldots, n$$

Parâmetros

- Tamanho da amostra: n (número de ensaios);
- Probabilidade de sucesso em cada ensaio: p.

Coeficiente binomial

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

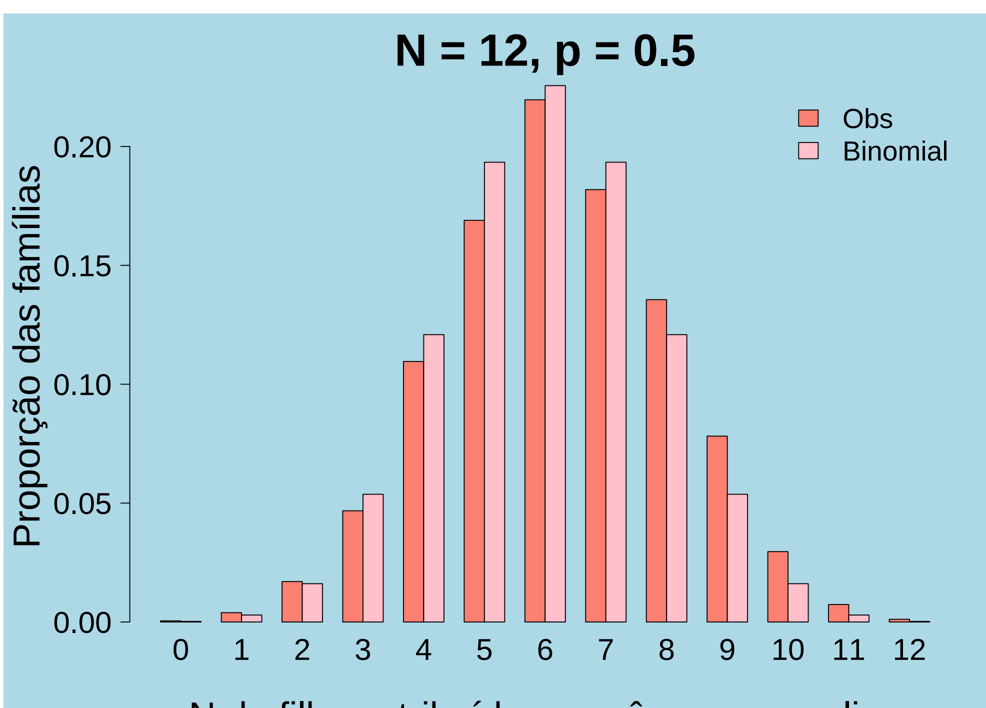
Momentos

- Esperança: $\mathbf{E}[X] = n\,p$
- Variância: $\mathbf{V}ar[X] = n\,p\,(1-p)$

Exemplo de Distribuição Binomial

Situação

6.115 famílias de 12 filhos na Saxônia: N de filhos atribuídos ao gênero masculino (Geisslller 1889):



Distribuição Poisson

Situação

- Contagem de eventos independentes.
- A contagem é *por unidade* de tempo e/ou espaço.
- Os eventos ocorrem a uma taxa constante por unidade.
- Esta taxa é o único parâmetro, λ.

Variável

- $X =$ o número de eventos em cada unidade.
- espaço amostral: $x = 0, 1, 2, \ldots, n$

Exemplos:

- N de árvores por parcela
- N de capturas por unidade de tempo
- N bombas V1 por área em Londres [1]

[1] R.D. Clarke 1946, Journal of the Institute of Actuaries,72,481.

Gráfico: Distribuição Poisson

Função de densidade

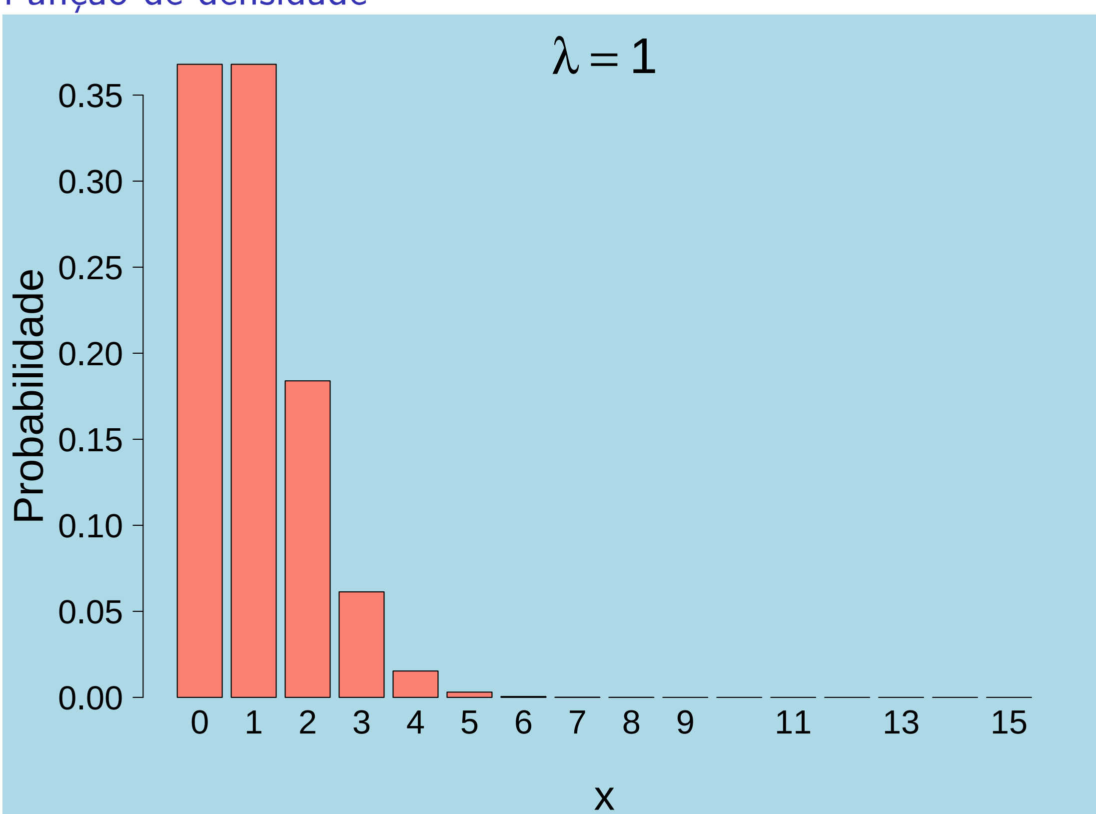


Gráfico: Distribuição Poisson

Função de densidade

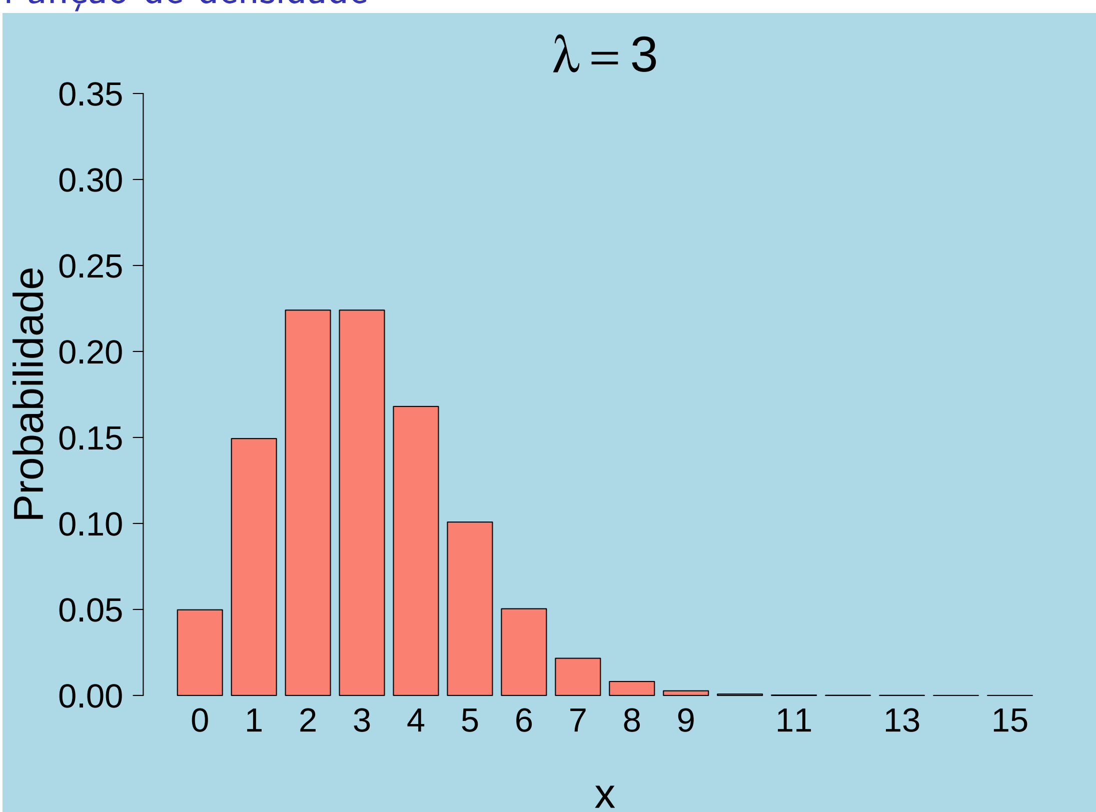
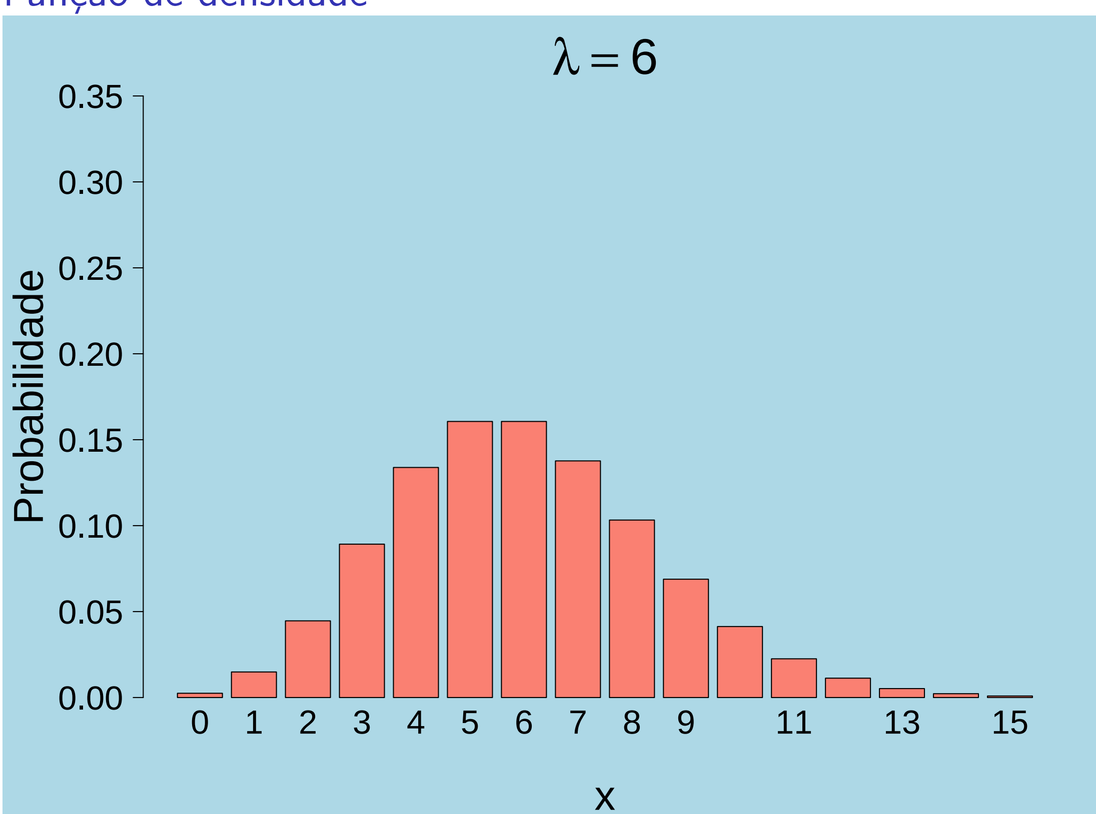


Gráfico: Distribuição Poisson

Função de densidade



Distribuição Poisson: Apresentação Formal

Função de densidade

$$f(x) = \frac{\lambda^x \, \mathrm{e}^{-\lambda}}{x!}, \qquad x = 0, 1, 2, \ldots$$

Parâmetro

λ: valor esperado da contagem

Função de distribuição

$$F(x) = \sum_{k=0}^{x} \frac{\lambda^k \, \mathrm{e}^{-\lambda}}{k!}, \qquad x = 0, 1, 2, \ldots$$

Momentos

- Esperança: $\mathbf{E}[X] = \lambda$
- Variância: $\mathbf{V}ar[X] = \lambda$
- Parâmetro λ = Esperança = Variância

Exemplo de Distribuição Poisson

Situação

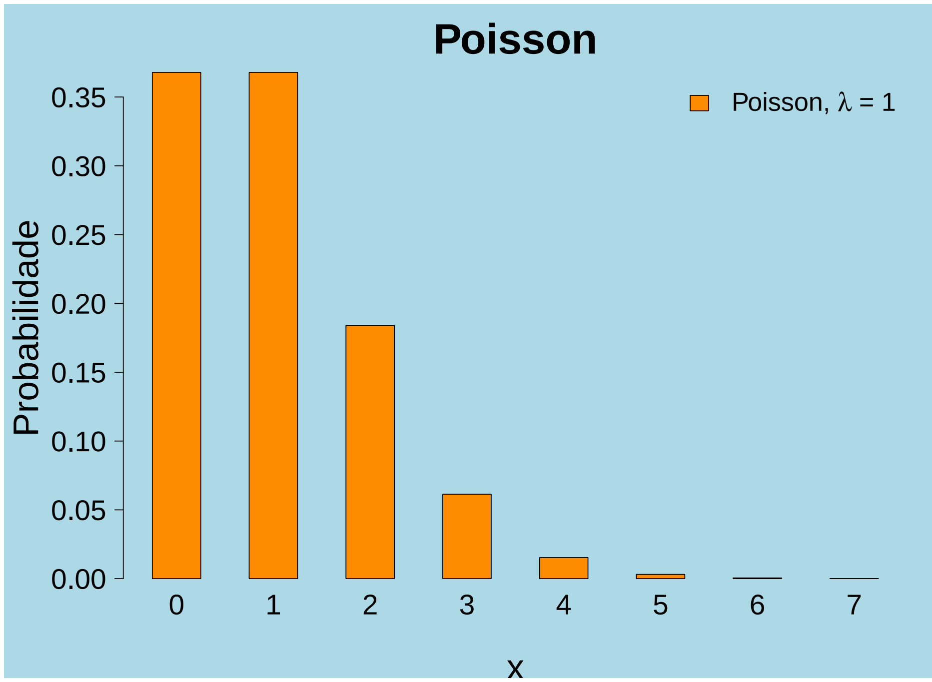
Número de bombas V1 que caíram no sul de Londres durante a II Guerra:

- $\lambda = 535 \text{ bombas}/576 \text{ quadrículas} \simeq 0,929/\text{quadr.}$

Caso Limite

A Poisson é um caso limite da distribuição binomial:

- quando o tamanho da amostra (n) tende a infinito ($n \to \infty$),
- a probabilidade de sucesso (p) tende a zero ($p \to 0$),
- mas o valor esperado permanece constante: $n\,p = \lambda$.



Distribuição Binomial Negativa

Situação

Sequência de ensaios Bernoulli independentes (generaliza a geométrica).

Variável

$X =$ *número de fracassos* até o $n^{\text{ésimo}}$ sucesso.

Parâmetros

- n: número de sucessos
- p: probabilidade de sucesso (constante)

Exemplos:

- N de tentativas até conseguir duas caras
- N de tentativas até seu personagem morrer no *game*
- Em biologia mais usada para descrever agregações (em breve)

Gráfico: Distribuição Binomial Negativa

Função de densidade

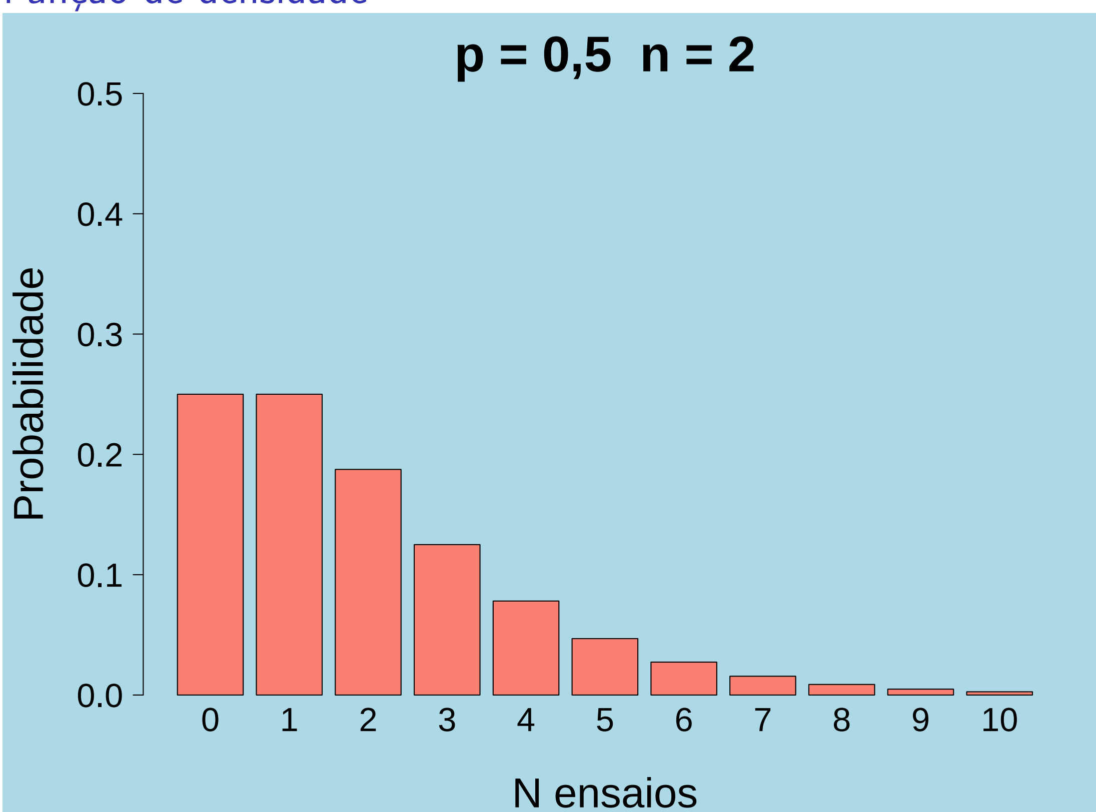


Gráfico: Distribuição Binomial Negativa

Função de densidade

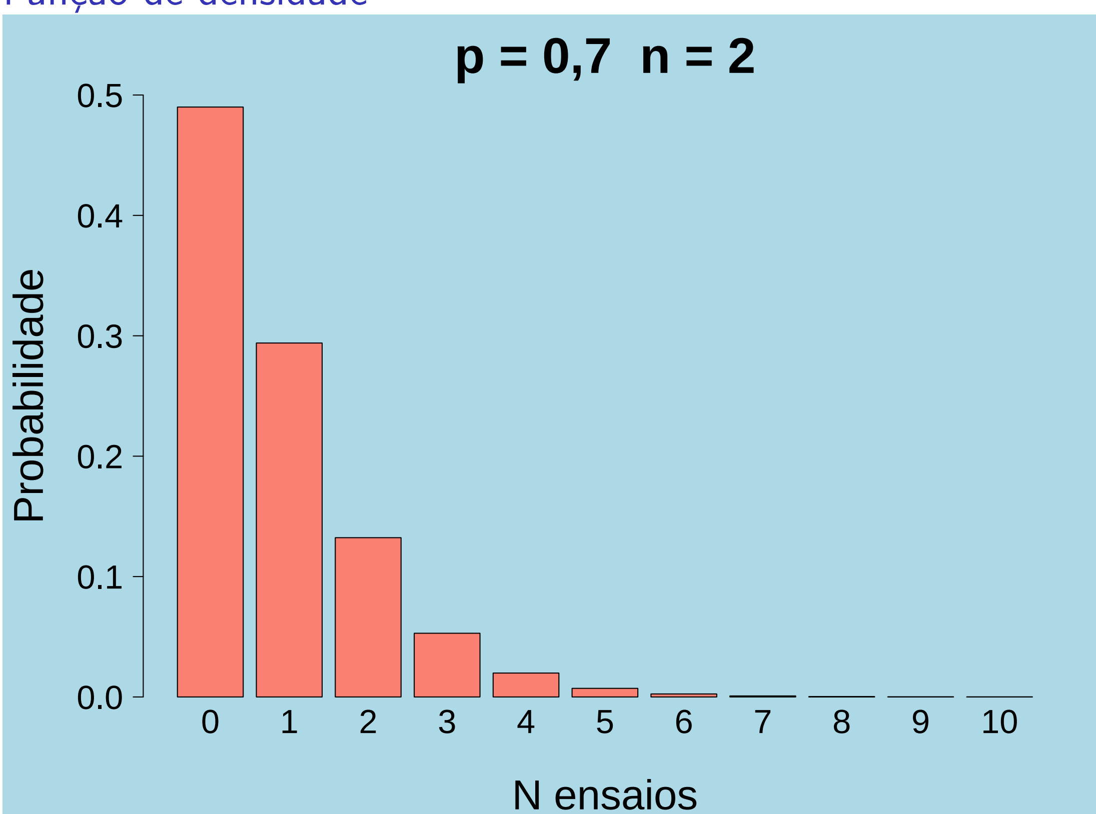
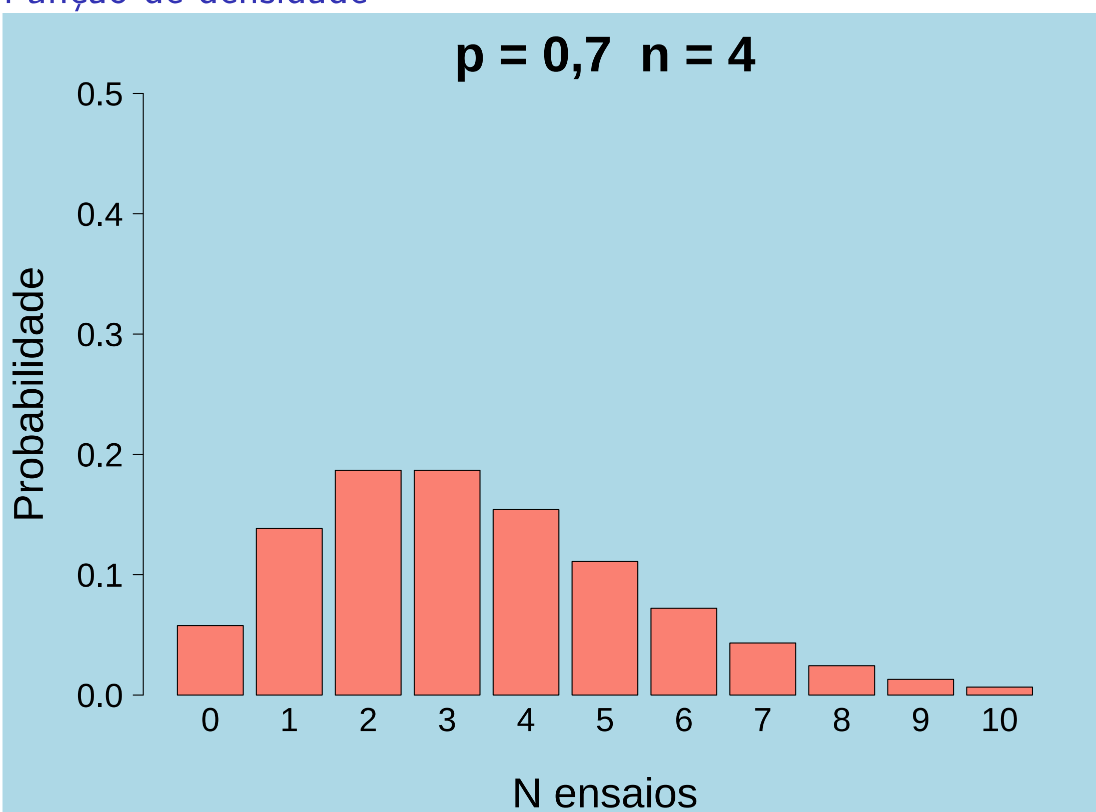


Gráfico: Distribuição Binomial Negativa

Função de densidade



Distribuição Binomial Negativa: Apresentação Formal

Função de densidade

$$f(x) = \binom{n+x-1}{x} p^n (1-p)^x, \qquad x = 0, 1, 2, \ldots$$

Função de distribuição

$$F(x) = \sum_{k=0}^{x} \binom{n+k-1}{k} p^n (1-p)^k, \qquad x = 0, 1, 2, \ldots$$

Distribuição Binomial Negativa: Apresentação Formal

Momentos

- Esperança:

$$\mathbf{E}[X] = \frac{n(1-p)}{p}$$

- Variância:

$$\mathbf{V}ar[X] = \frac{n(1-p)}{p^2}$$

Binomial Negativa $\times$ Geométrica

Caso Particular

A distribuição geométrica é um caso particular da distribuição binomial negativa.

Função de densidade quando $n = 1$:

$$f(x) = \binom{n+x-1}{x} p^n (1-p)^x \;=\; p(1-p)^x$$

Binomial Negativa: Parametrização Alternativa

Situação

Dados de contagem de eventos agregados

- Agregação: variância maior que a média
- Exemplos:
  - número de árvores por parcela
  - número de plântulas por parcela
  - capturas por armadilha

Parâmetros

- valor esperado (contagem média) $\mathbf{E}[X] = \mu$
- parâmetro de dispersão: k

Relações

$$n = k \; ; \; p = \frac{k}{k+\mu} \quad \Longleftrightarrow \quad k = n \; ; \; \mu = \frac{n(1-p)}{p}$$

Binomial Negativa: Parametrização Alternativa (cont.)

Função de densidade

$$f(x) = \frac{\Gamma(k+x)}{\Gamma(k)x!} \left(\frac{k}{k+\mu}\right)^k \left(\frac{\mu}{k+\mu}\right)^x, \; x = 0, 1, 2, \ldots$$

Momentos

Esperança: $\mathbf{E}[X] = \mu$

Variância:

$$\mathbf{V}ar[X] = \frac{n(1-p)}{p^2} = \mu + \frac{\mu^2}{k}$$

Exemplo de Distribuição Binomial Negativa

Situação

Número de adutos de Palmito-Jussara em parcelas em sítio de mata atlântica:

Poisson: $\lambda = \frac{4242 \text{ adultos}}{568 \text{ parcelas}} = 7,468/\text{parcela}.$

Binomial Negativa: $\mu = 7,468/\text{parcela}$, $k = 0.642$

Distribuições de probabilidade no R

Binomial

- `dbinom(x, size, prob, log = FALSE)`
- `pbinom(q, size, prob, lower.tail = TRUE, log.p = FALSE)`
- `qbinom(p, size, prob, lower.tail = TRUE, log.p = FALSE)`
- `rbinom(n, size, prob)`

Poisson

- `dpois(x, lambda, log = FALSE)`
- `ppois(q, lambda, lower.tail = TRUE, log.p = FALSE)`
- `qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)`
- `rpois(n, lambda)`

...

Resumo

- Distribuições de Probabilidade são *funções* que associam os valores de uma variável quantitativa com probabilidades.
- Os parâmetros controlam o comportamento da distribuição.
- Média e variância *não são* parâmetros de todas as distribuições mas podem ser expressas como funções desses.

Resumo (cont.)

- Algumas distribuições são casos especiais de outras.
- Algumas distribuições são casos limites de outras.
- Uma mesma distribuição pode ter aplicações muito diferentes daquela que a originou.