

BIE5781

Aula 3 (cont.)

Distribuições de probabilidade contínuas: Uniforme, normal, lognormal

CONCEITOS

- Distribuição uniforme
- Distribuição normal
- Teorema central do limite
- Distribuição log-normal

Distribuição Uniforme

$$f(x) = \frac{1}{\max - \min}$$

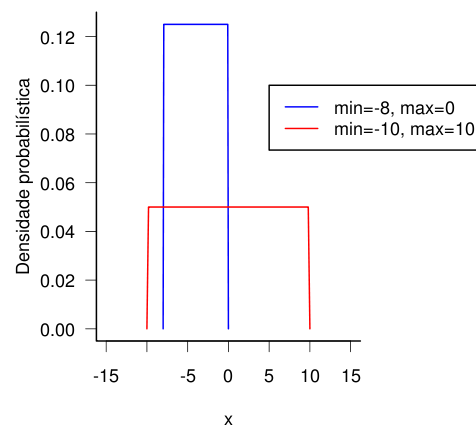
para $\min \leq x \leq \max$

min = valor mínimo (real)

max = valor máximo (real)

$$E[X] = \frac{\min + \max}{2}$$

$$\text{VAR}[X] = \frac{(\max - \min)^2}{12}$$



Distribuição Normal ou Gaussiana

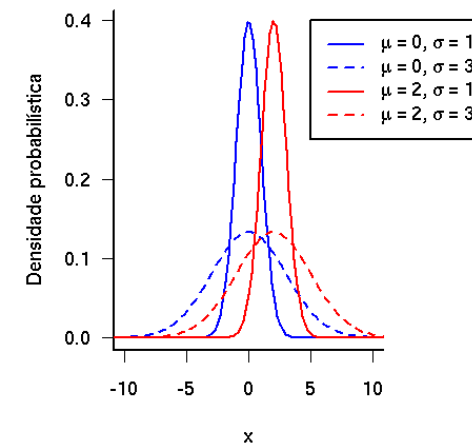
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

μ = média (real)

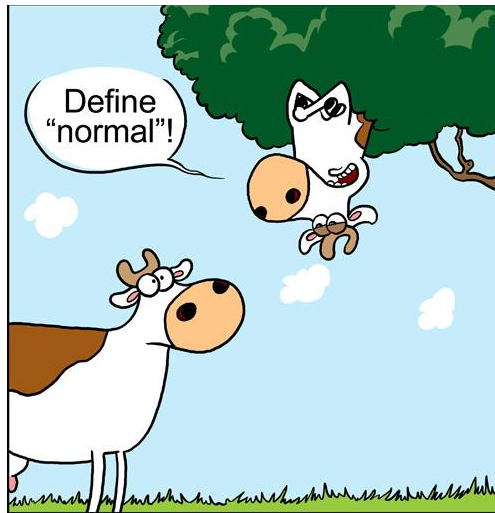
σ = desvio-padrão (real positivo)

$$E[X] = \mu$$

$$\text{VAR}[X] = \sigma^2$$



Teorema Central do Limite (TCL): De Perto Todo Mundo é Normal



Teorema Central do Limite

- Seja uma variável aleatória X com média $E[X]$ e variância $VAR[X]$
- Tome uma amostra de X de tamanho N
- Seja Y a média desta amostra.

Com N suficientemente grande ($N \rightarrow \infty$) :

- Y terá distribuição normal
- com média $E[Y] = E[X]$
- e variância $VAR[Y] = VAR[X]/N$

TCL: exemplo



DESCICLOPÉDIA
A enciclopédia livre de conteúdo

Instituto de Biociências da Universidade de São Paulo

“Esse departamento é conhecido por avaliar os graduandos com o famoso "método da escada". Os professores amassam todas as provas e fazem pequenas bolas com elas, arremessando-as escada acima. Dependendo do degrau que a bolinha de papel parar, eles atribuem uma nota.”

http://desciclopedia.org/wiki/Instituto_de_Biociências_da_Universidade_de_São_Paulo

TCL: exemplo



DESCICLOPÉDIA
A enciclopédia livre de conteúdo

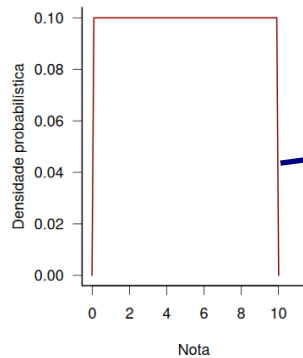
Instituto de Biociências da Universidade de São Paulo

A média das notas de provas de uma turma de 50 estudantes foi de 4.25.

Qual a probabilidade disso acontecer se as notas são sorteadas de uma uniforme entre 0 e 10?

TCL: exemplo

50 realizações de uma distribuição uniforme com entre 0 e 10



$\{9.33, 9.82, 9.76, 5.39, 8.22, 9.30, 9.50, 0.73, \dots\}$

Média da amostra = 5.04

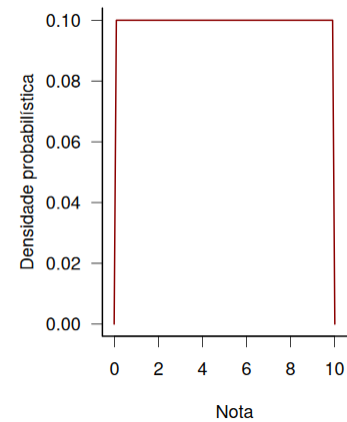
Se repetimos o procedimento muitas vezes:

Médias: $\{4.937344, 5.431391, 4.841943, 5.361601, 4.619908, \dots\}$

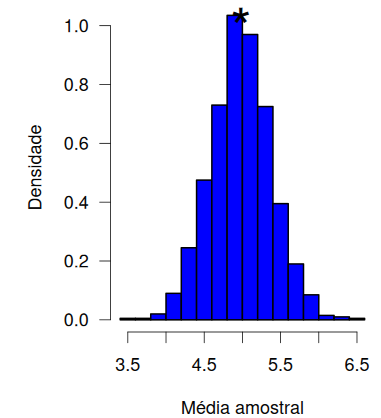
Qual a distribuição destas médias amostrais?

TCL: exemplo

Se repetimos 1000 vezes ...



1000 médias



* = média = 4.983

TCL: exemplo

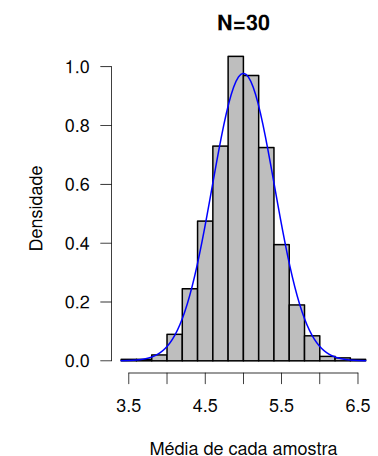
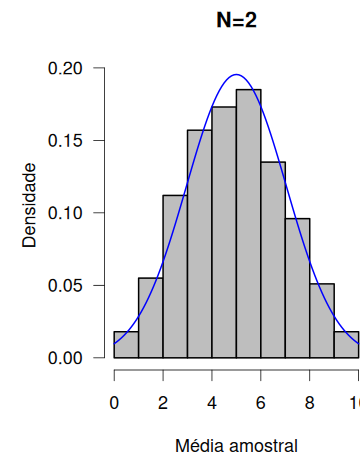
Distribuição normal das médias amostrais em nosso exemplo:

- Média da distribuição amostrada: $E[X] = 5$
- Variância da distribuição amostrada: $VAR[X] = 8.33$

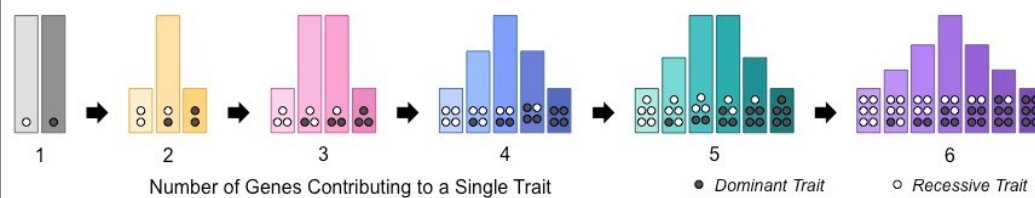
LOGO:

- As médias amostrais Y tendem a uma normal com:
- $E[Y] = E[X] = 5$
- $VAR[Y] = VAR[X] / N = 0.167$
- Sob esta distribuição $P(Y < 4.25) = 0.033$

Note bem: do Limite

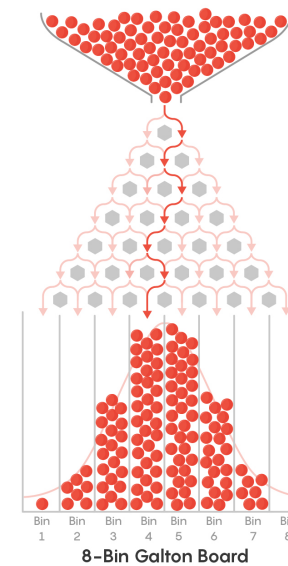


Gaussianas por todos os lados



Caracteres poligênicos

Gaussianas por todos os lados



Tabuleiro de Galton (Quincux)

<https://www.youtube.com/watch?v=YINTTVjBrY4>

Distribuição Log-normal

Medidas cujos logaritmos seguem distribuição Gaussiana.

Exemplos:

- Tamanhos de populações de dinâmica estocástica
- Outras medidas que resultam de multiplicação de muitas variáveis aleatórias (e.g. fracionamento, cicatrização)
- Fenomenologicamente usada para medidas de valor positivo e com distribuição assimétrica à esquerda (e.g. diâmetros ou massas em populações naturais).

Distribuição Log-Normal

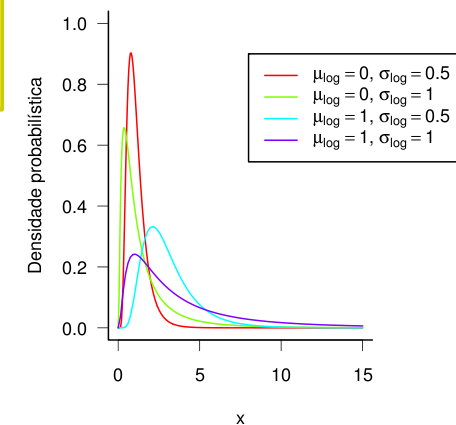
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

μ = média dos logaritmos (real)

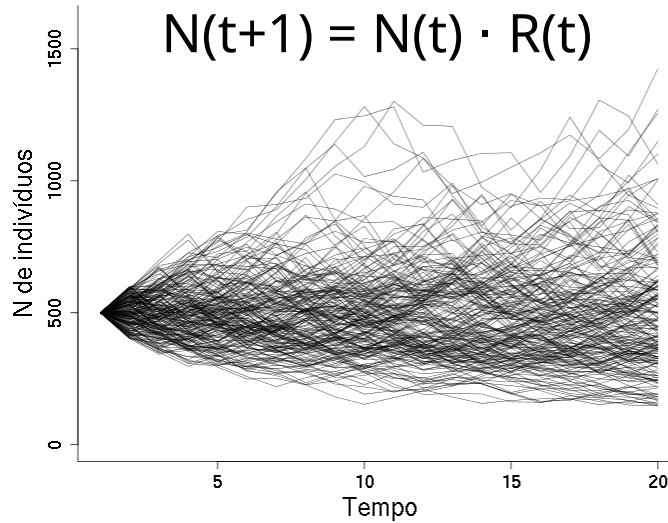
σ = desvio-padrão dos logaritmos (real positivo)

$$E[X] = e^{\mu + \sigma^2/2}$$

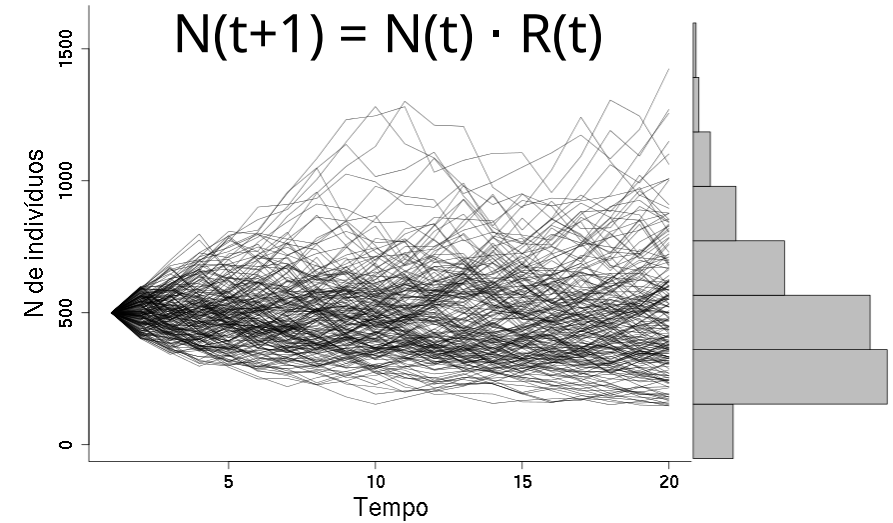
$$VAR[X] = (e^{\sigma^2} - 1) e^{2\mu + \sigma^2}$$



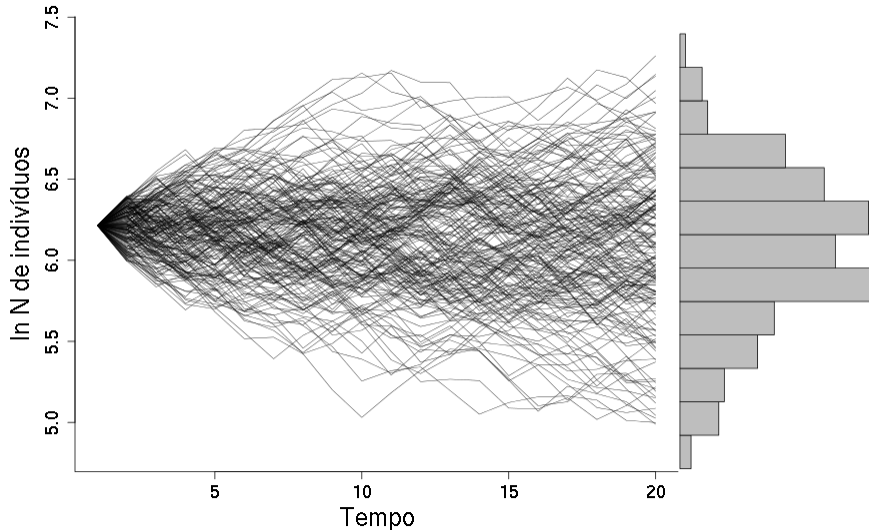
Exemplo de processo multiplicativo



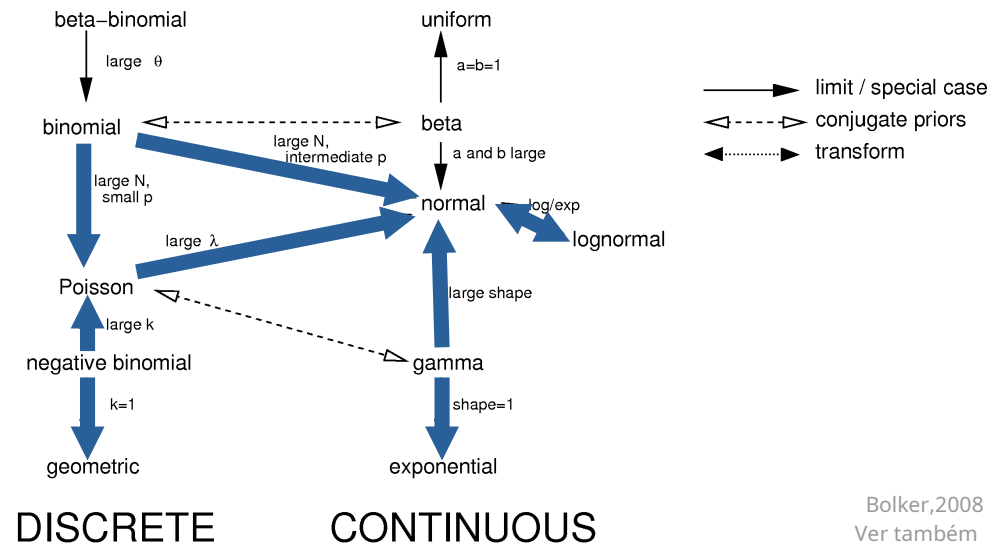
Distribuição de abundâncias



Distribuição de abundâncias



Relações entre as Distribuições



Bolker, 2008
 Ver também
http://www.johndcook.com/distribution_chart.html

RESUMINDO ...

- Algumas distribuições são casos especiais de outras.
- Algumas distribuições são casos-limite de outras.
- O caso-limite mais importante é o TCL
- A normal é um modelo probabilístico para a soma de muitas variáveis aleatórias (e.g., muitas fontes de erros aditivos, soma de realizações de variáveis i.i.d)
- A log-normal é um modelo probabilístico para a multiplicação de muitas variáveis aleatórias (e.g., muitas fontes de erros que se multiplicam, produto de realizações de variáveis i.i.d)

Leituras Recomendadas

- Bolker, B. (2008). Ecological Models and Data in R. Princeton, Princeton University Press. (Cap.4)
- Otto, S. P. & Day, T. (2007). A biologist's guide to mathematical modelling in ecology and evolution. Princeton, Princeton University Press. (Primer 3)
- Lindsey J. K. (2004) Introduction to applied statistics - a modelling approach. 2nd Ed, Oxford, Oxford University Press. (Cap.4)