Probabilidade e variáveis aleatórias (revisão)

CE219 - Controle Estatístico de Qualidade

Prof. Cesar Taconeli
taconeli@ufpr.br

Prof. Walmes Zeviani
walmes@ufpr.br

Laboratório de Estatística e Geoinformação
Departamento de Estatística
Universidade Federal do Paraná

Métodos estatísticos para a análise da qualidade

- O objetivo aqui é fazer uma (breve) revisão de **modelos probabilísticos** e **métodos estatísticos** com aplicação na descrição, modelagem e produção de inferências para processos.
- Serão abordados:
  1. Métodos de análise descritiva.
  2. **Probabilidade e principais modelos probabilísticos**.
  3. Inferência estatística aplicada à qualidade do processo.
- As próximas aulas serão intercaladas com ilustrações no R e os scripts disponibilizados na página da disciplina.

Probabilidade e principais modelos probabilísticos

Para que servem?

- Um objetivo frequente na análise de dados é a inferência estatística, que consiste em produzir conclusões (estimar parâmetros, testar hipóteses, tomar decisões) para **populações** com base em **amostras**.
- No contexto de controle estatístico de processos, o usual é assumir como população a distribuição dos valores de todos os itens produzidos por um processo para certa característica da qualidade (variável aleatória).
- A forma mais usual de proceder é assumir um **modelo probabilístico** adequado à distribuição do processo.
- Um modelo probabilístico é um modelo matemático que associa probabilidades aos valores de uma variável aleatória (ou de múltiplas variáveis).

Quanto o tipo de valor da variável aleatória

- A literatura dispõe de muitas (muitas mesmo!) alternativas de modelos probabilísticos para variáveis aleatórias discretas, contínuas ou até mesmo mistas (parte contínua, parte discreta).
- Uma variável aleatória é definida **discreta** caso tenha probabilidades não nulas associadas a um conjunto enumerável de valores. São exemplos o número de itens defeituosos num lote de 100 peças, o número de clientes atendidos por hora em um banco, o número de itens produzidos até a verificação da primeira unidade defeituosa.
- Uma variável aleatória é definida **contínua** se pode gerar resultados em algum conjunto contínuo de valores. São exemplos o tempo de vida de equipamentos, a resistência de moldes de aço, as dimensões de placas de alumínio...

Média e variância para VA discreta

- Seja X uma variável aleatória discreta, cuja distribuição é definida pela função de probabulidades:

$$\Pr(X = x_i) = p_i, \quad i = 1, 2, 3, \ldots$$

- A média (esperança) e a variância de X ficam definidas, respectivamente, por:
  - $\mu = \mathrm{E}(X) = \sum_i x_i p_i$.
  - $\sigma^2 = \mathrm{Var}(X) = \sum_i (x_i - \mu)^2 p_i$.

Exercício

EXERCÍCIO: Variáveis aleatórias discretas

A variável aleatória X assume os valores 1, 2 e 3 com probabilidades $(1 + 3k)/3$, $(1 + 2k)/3$ e $(0.5 + 5k)/3$, respectivamente.

1. Ache o valor apropriado de k.
2. Ache a média e a variância de x.
3. Apresente a função de distribuição acumulada.

Média e variância para VA contínua

- Seja X uma variável aleatória contínua, cuja distribuição é definida pela **função densidade de probabilidades** $f(x)$, tal que:

$$\Pr(a < X < b) = \int_a^b f(x)\mathrm{d}x$$

- A média (esperança) e a variância de X são dadas, respectivamente, por:
  - $\mu = E(X) = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x$.
  - $\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\mathrm{d}x$.

Exercício

EXERCÍCIO: Variáveis aleatórias contínuas

A distribuição de probabilidade de X é
$f(x) = k \exp\{-x\}, \quad 0 < x < \infty$. Qual o valor apropriado de k? Qual a média e a variância de X?

Modelos probabilísticos para variáveis aleatórias discretas

Distribuição hipergeométrica

- A principal aplicação da distribuição hipergeométrica é em situações envolvendo amostragem aleatória simples sem reposição.
- Na área de controle de qualidade, aplica-se em problemas de amostragem de aceitação.
- A título de ilustração, poderíamos ter X como o número de ítens defeituosos em uma amostra aleatória de n ítens selecionada, sem reposição, de um lote de tamanho N.
- Neste caso, D e $N - D$ seriam os números de itens defeituosos e bons no lote, respectivamente.

Distribuição hipergeométrica

Uma variável aleatória Y tem distribuição hipergeométrica de parâmetros D, N e n se sua função de probabilidades é dada por:

$$P\left(X=x\right)=\frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}, \quad x=0,1,2,\ldots,\min\left(n,D\right).$$

com média e variância:

- $\mu=\frac{nD}{N}$.
- $\sigma^2=\frac{nD}{N}\binom{N-D}{N}\binom{N-n}{N-1}$.

Distribuição hipergeométrica

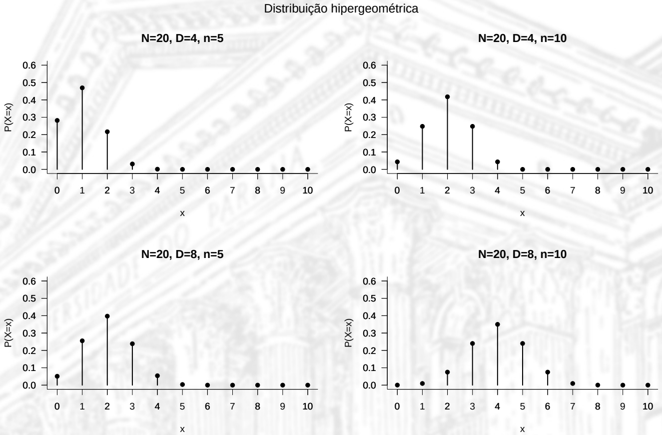


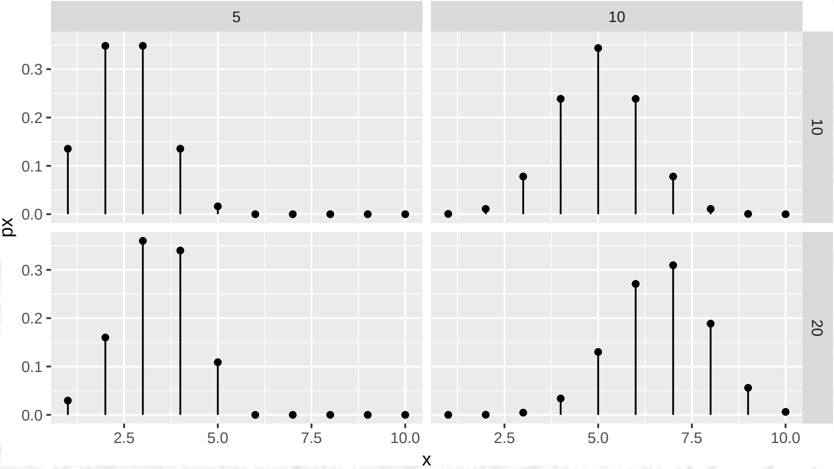
Figura 1. Distribuição hipergeométrica.

Gráfico da distribuição hipergeométrica com o R

```r
library(tidyverse)

tb <- crossing(x = 1:10,
               brancas = c(10, 20),
               pretas = 10,
               retiradas = c(5, 10))
tb <- tb %>%
    mutate(px = dhyper(x = x,
                       m = brancas,
                       n = pretas,
                       k = retiradas))

ggplot(tb, aes(x, px)) +
    facet_grid(facets = brancas ~ retiradas) +
    geom_point() +
    geom_segment(aes(y = 0, xend = x, yend = px))
```


5
10
10
20
px
0.3
0.2
0.1
0.0
0.3
0.2
0.1
0.0
2.5
5.0
7.5
10.0
2.5
5.0
7.5
10.0
x

Exercício

Um lote de tamanho $N = 30$ contém 10 unidades não conformes.

- Qual é a probabilidade de que uma amostra de cinco unidades selecionadas aleatoriamente e sem reposição do lote conter duas unidades não conformes?
- Nas mesmas condições do item anterior, qual a probabilidade do lote conter no máximo duas unidades não conformes?

Distribuição binomial

- Suponha n eventos independentes, cada um com dois resultados possíveis (*sucesso* ou *fracasso*).
- Vamos admitir igual probabilidade de sucesso (p) a cada um dos n eventos.
- A variável aleatória que conta o número de sucessos produzidos nos n eventos tem distribuição binomial, sendo n e p os seus parâmetros.
- A distribuição binomial serve como aproximação para a distribuição hipergeométrica para grandes populações ($N >>> n$).

Distribuição binomial

A variável aleatória X com distribuição binomial de parâmetros n e p tem sua função de probabilidades dada por:

$$\Pr(X = x) = \binom{n}{x} p^x (1-p)^{(n-x)}, \quad x = 0, 1, 2, \ldots, n.$$

com média e variância:

- $\mu = np$.
- $\sigma^2 = np(1-p)$.

Distribuição binomial

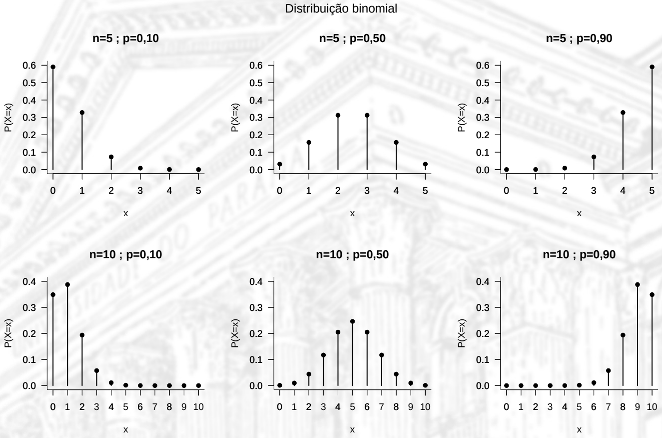
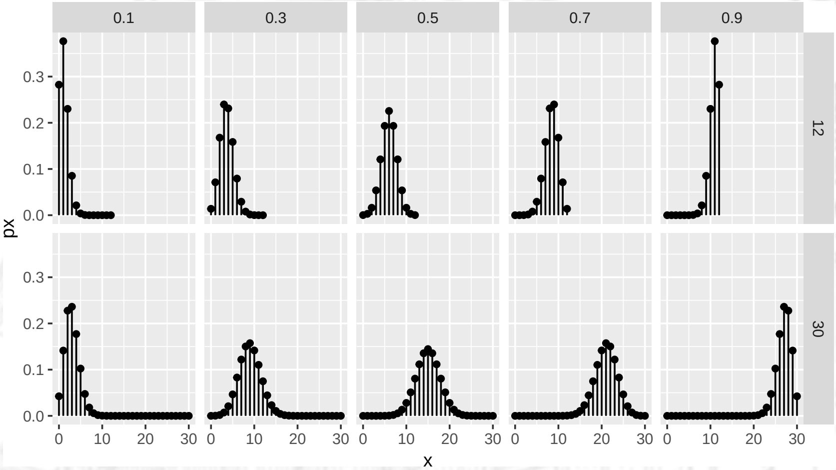


Figura 2. Distribuição binomial.

Gráfico da distribuição binomial com o R

```r
tb <- crossing(prob = seq(0.1, 0.9, by = 0.2),
               size = c(12, 30)) %>%
    mutate(x = map(size, seq, from = 0, by = 1)) %>%
    unnest(x) %>%
    mutate(px = dbinom(x, size = size, prob = prob))

ggplot(tb, aes(x, px)) +
    facet_grid(facets = size ~ prob) +
    geom_point() +
    geom_segment(aes(y = 0, xend = x, yend = px))
```


0.1
0.3
0.5
0.7
0.9
12
30
px
x
0.0
0.1
0.2
0.3
0
10
20
30

Distribuição binomial

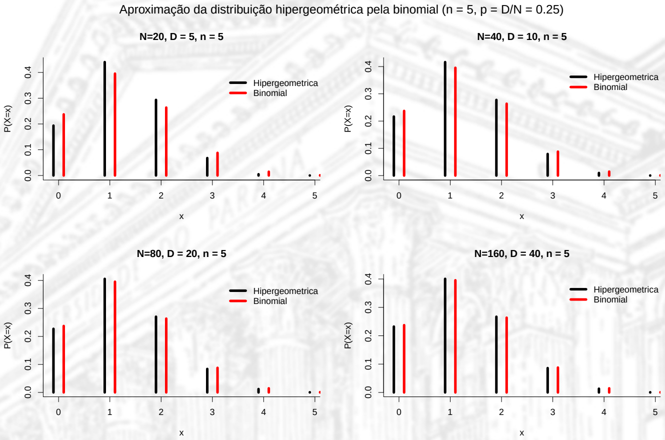


Figura 3. Aproximação da hipergeométrica pela binomial.

Exercício

Um processo de produção opera a uma taxa de 5% de peças produzidas não conformes. A cada hora uma amostra de 25 unidades do produto é retirada, e o número de peças não conformes (defeituosas) registrado.

- Qual a probabilidade de duas peças defeituosas serem verificadas numa realização da inspeção?
- A empresa será multada se três ou mais itens defeituosos forem verificados na próxima inspeção. Qual a probabilidade de multa?

Distribuição Poisson

- Aplicada na modelagem do número de eventos aleatórios verificados em unidades de tempo, área, volume, etc.
- Em CEQ, aplicada frequentemente na análise do número de defeitos (não conformidades) por unidade de produto.
- Aplicável caso os eventos ocorram aleatoriamente, de forma independente e sob taxa constante (processo de Poisson).
- Forma limite da distribuição binomial para p tendendo a 0 e n a infinito, mantendo constante $\lambda = np$.

Distribuição Poisson

A variável aleatória X com distribuição de Poisson de parâmetro (taxa) λ tem sua função de probabilidades dada por:

$$\Pr(X = x) = \frac{\exp -\lambda\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots, \quad \lambda > 0,$$

com média e variância dadas por:

- $\mathrm{E}(X) = \mu = \lambda$.
- $\mathrm{Var}(X) = \sigma^2 = \lambda$.

Distribuição Poisson

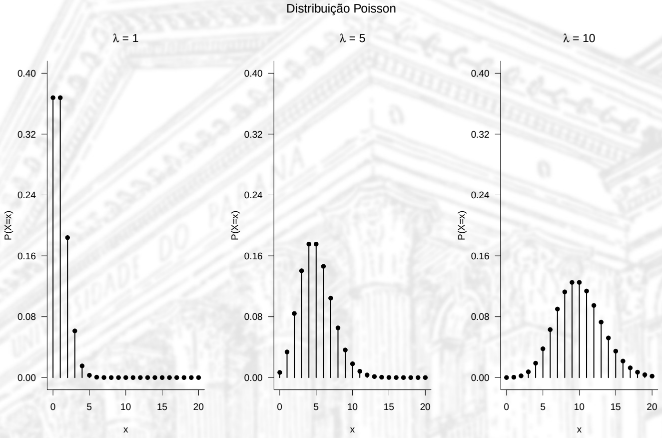


Figura 4. Distribuição Poisson.

Distribuição Poisson

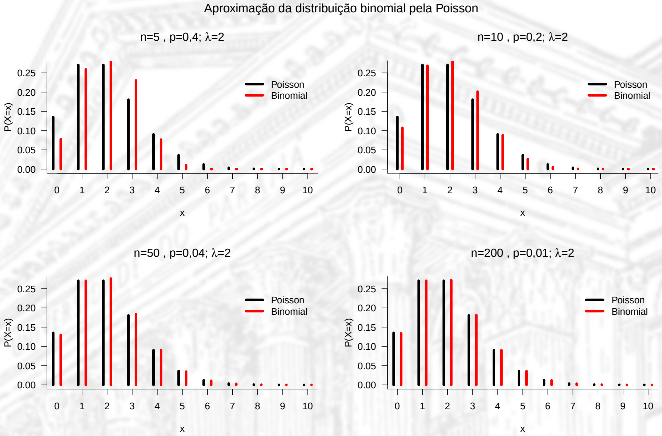
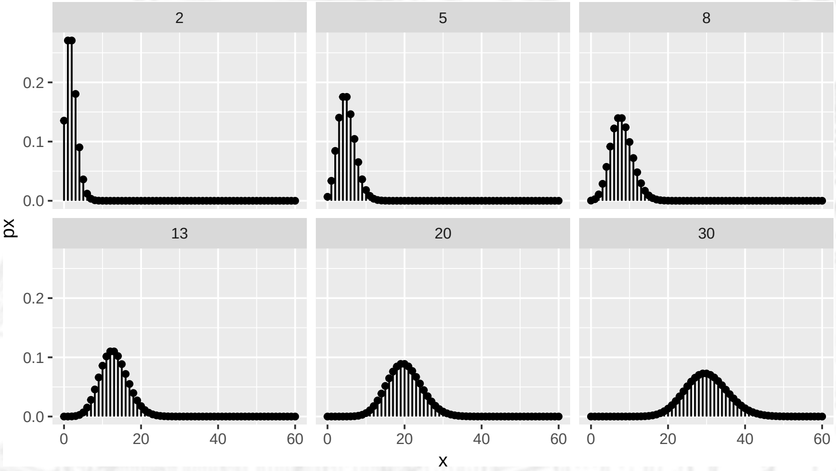


Figura 5. Aproximação da Poisson pela binomial.

Gráfico da distribuição Poisson com o R

```r
tb <- crossing(lambda = c(2, 5, 8, 13, 20, 30),
               x = 0:(2 * max(lambda))) %>%
    mutate(px = dpois(x, lambda = lambda))

ggplot(tb, aes(x, px)) +
    facet_wrap(facets = ~ lambda) +
    geom_point() +
    geom_segment(aes(y = 0, xend = x, yend = px))
```


2
5
8
13
20
30
0.0
0.1
0.2
px
0
20
40
60
x

Distribuição Poisson

- Como propriedade da distribuição Poisson, se a unidade de medida em que são contados os eventos for multiplicada por uma constante t, então a variável de contagem resultante terá distribuição Poisson com média (e variância) λt:

$$\Pr(X = x) = \frac{\exp -\lambda t (\lambda t)^x}{x!}, \quad x = 0, 1, 2, \ldots, \quad \lambda > 0,$$

- Padrões não aleatórios na ocorrência dos eventos (eventos agrupados, regularmente espaçados, etc) geram contagens que **não são adequadamente modeladas pela distribuição Poisson.**

Distribuição Poisson

Exercício

O departamento de cobrança de uma grande companhia de cartão de créditos tenta controlar os erros (administrativos, de digitação, etc) nas contas dos clientes. Suponha que tais erros ocorram aleatoriamente com uma taxa diária $\lambda = 0.8$.

1. Qual a probabilidade de ocorrerem dois erros num dia qualquer?
2. Qual a probabilidade de ocorrerem mais de dois erros num dia qualquer?
3. Qual a probabilidade de ocorrerem mais de dois erros nos próximos cinco dias?

Distribuições binomial negativa e geométrica

- Considere uma sequência de eventos independentes do tipo *sucesso ou fracasso*, cada um com probabilidade de sucesso p;
- Seja X a variável aleatória que conta o número de **eventos** ocorridos até o *r-ésimo* sucesso.
- Neste caso, X tem distribuição binomial negativa com parâmetros r e p.

Distribuições binomial negativa e geométrica

A variável aleatória X com distribuição binomial negativa de parâmetros r e p tem sua função de probabilidades dada por:

$$\Pr(X = x) = \binom{x-1}{r-1} p^r (1-p)^{(x-r)}, \quad x = r, r+1, r+2, \ldots$$

em que $r \geq 1$ inteiro, com média e variância dadas por:

- $\mu = \frac{r}{p}$.
- $\sigma^2 = \frac{r(1-p)}{p^2}$.

Distribuição Binomial Negativa

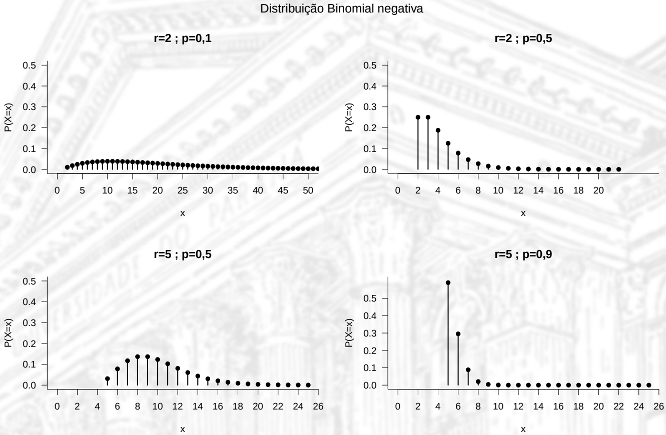
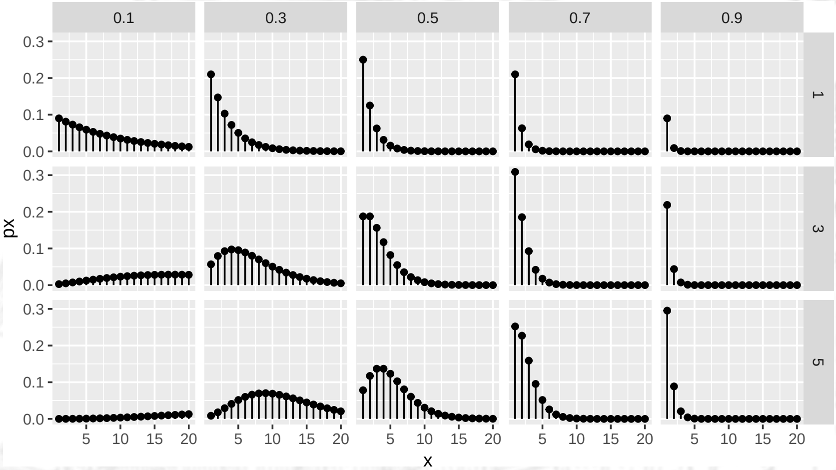


Figura 6. Distribuição ninomial negativa.

Gráfico da distribuição binomial negativa com o R

```r
tb <- crossing(x = 1:20,
               prob = seq(0.1, 0.9, by = 0.2),
               size = c(1, 3, 5)) %>%
    mutate(px = dnbinom(x, size = size, prob = prob))

ggplot(tb, aes(x, px)) +
    facet_grid(facets = size ~ prob) +
    geom_point() +
    geom_segment(aes(y = 0, xend = x, yend = px))
```


0.1
0.3
0.5
0.7
0.9
1
3
5
px
x
0.0
0.1
0.2
0.3
5
10
15
20

Distribuições binomial negativa e geométrica

- A distribuição binomial negativa tem importante aplicação como alternativa à distribuição de Poisson na modelagem de **contagens de eventos por unidade de tempo**, espaço, outra medida qualquer.
- Observe a **dualidade** entre as distribuições binomial e binomial negativa. Na primeira, o número de eventos é fixo e a variável é o número de sucessos. Na segunda, o número de sucessos é fixo, contando-se o número de eventos.
- Um importante **caso particular** da distribuição binomial negativa é a distribuição geométrica, para a qual $r = 1$ (X conta o número de eventos até se observar o primeiro sucesso).

Distribuições geométrica

A variável aleatória X tem distribuição geométrica de parâmetro p se sua função de probabilidades é dada por:

$$\Pr(X = x) = (1 - p)^{(x-1)} \quad 0 < p < 1, \quad x = 1, 2, \ldots$$

com média e variância dadas por:

- $\mu = \frac{1}{p}$.
- $\sigma^2 = \frac{1-p}{p^2}$.

Distribuição geométrica

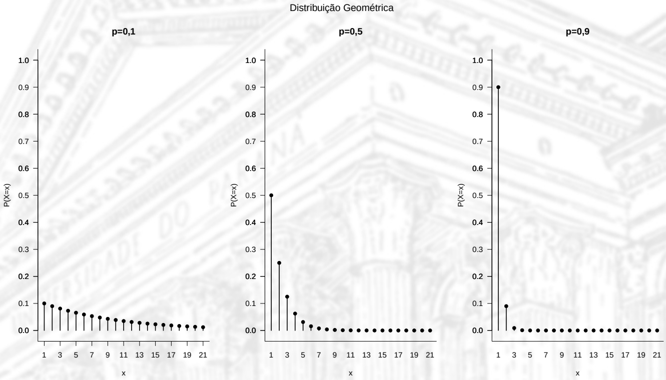
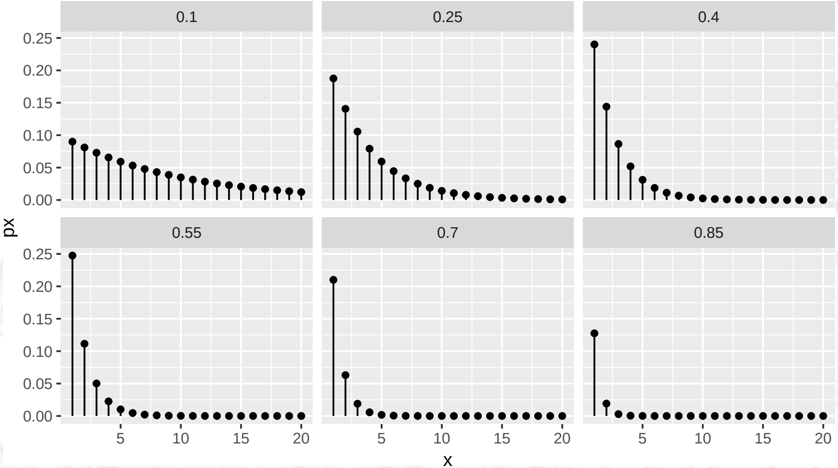


Figura 7. Distribuição geométrica.

Gráfico da distribuição geométrica com o R

```r
tb <- crossing(x = 1:20,
               prob = seq(0.1, 0.9, by = 0.15)) %>%
    mutate(px = dgeom(x, prob = prob))

ggplot(tb, aes(x, px)) +
    facet_wrap(facets = ~ prob) +
    geom_point() +
    geom_segment(aes(y = 0, xend = x, yend = px))
```


0.1
0.25
0.4
0.55
0.7
0.85
px
0.25
0.20
0.15
0.10
0.05
0.00
5
10
15
20
x

Distribuições binomial geométrica

- Uma das propriedades da distribuição geométrica é a **falta de memória**, segundo a qual:

$$\Pr(X > x + a | X > a) = \Pr(X > x), \text{para qualquer} \quad a > 0.$$

- Assim, sob distribuição geométrica, a probabilidade do próximo ítem defeituoso ocorrer daqui a cinco avaliações é a mesma se ainda não tivermos observado nenhum item, ou 10 ítens bons, ou 1000 ítens bons.
- A distribuição geométrica é o modelo adequado para um importante indicador de desempenho de gráficos de controle, o **comprimento médio de sequência**.

Exercício

Um operador de telemarketing trabalha com vendas. Historicamente, esse operador tem sucesso em 20% de suas tentativas.

1. Qual a probabilidade dele realizar sua primeira venda na terceira tentativa?
2. Qual a probabilidade dele realizar sua primeira venda até a terceira tentativa?
3. Sua meta é realizar três vendas. Qual a probabilidade dele atingir sua meta até a décima tentativa?
4. Qual o número esperado de tentativas até a primeira venda? E até atingir a meta?

Modelos probabilísticos para variáveis aleatórias contínuas

Distribuição Normal

A distribuição Normal é a mais importante distribuição contínua, dentre outros motivos por que:

- Modela adequadamente a distribuição de um grande número de variáveis aleatórias contínuas.
- Serve de aproximação para diversas outras distribuições contínuas e discretas.
- Tem papel importante na Teoria Estatística, fundamentando a obtenção de inferências em diferentes contextos.

Distribuição Normal

A variável aleatória X tem distribuição Normal com parâmetros μ e σ^2 se sua função densidade de probabilidade é dada por:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2\right\}, \quad x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma^2 > 0,$$

com média e variância dadas por:

- $\mathrm{E}(X) = \mu$
- $\mathrm{Var}(X) = \sigma^2$.

Distribuição Normal

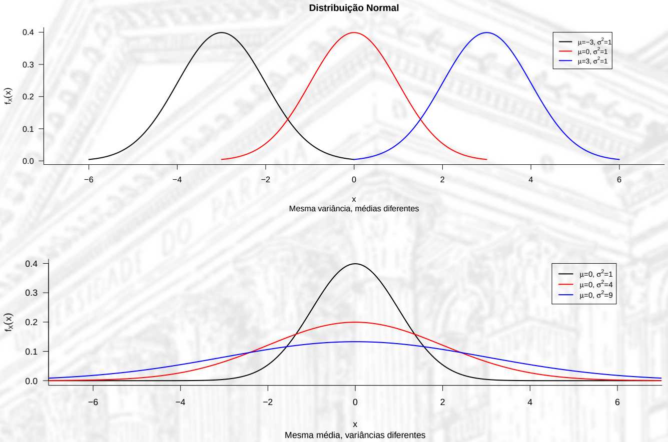
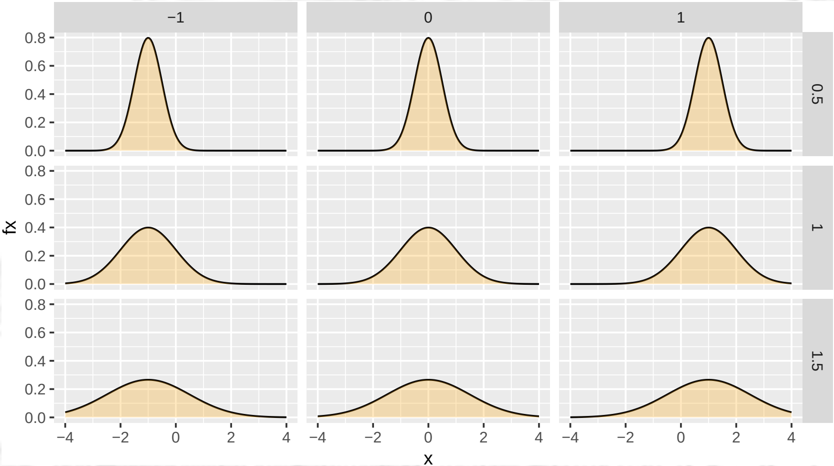


Figura 8. Distribuição Normal.

Gráfico da distribuição Normal com o R

```r
tb <- crossing(x = seq(-4, 4, length.out = 151),
               mean = -1:1,
               sd = c(0.5, 1, 1.5)) %>%
    mutate(fx = dnorm(x, mean = mean, sd = sd))

ggplot(tb, aes(x, fx)) +
    facet_grid(facets = sd ~ mean) +
    geom_line() +
    geom_area(alpha = 0.25, fill = "orange")
```


−1
0
1
0.5
1
1.5
fx
0.8
0.6
0.4
0.2
0.0
−4
−2
0
2
4
x

Distribuição Normal

- Probabilidades associadas à distribuição Normal não podem ser obtidas analiticamente (mas sim numericamente), pois a integral correspondente não tem forma fechada.
- Como recursos, temos os softwares estatísticos ou a consulta a tabelas da distribuição *Normal padrão* ($\mu = 0, \sigma^2 = 1$).
- A consulta a tabela da normal padrão se justifica pelo fato que se X tem distribuição Normal com média μ e variância σ^2, então:

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

Obs: Apenas lembrando que a notação $X \sim N(\mu, \sigma^2)$ lê-se: a variável aleatória X tem distribuição Normal de média μ e variância σ^2".

Exercício

A intensidade de luz resultante de uma lâmpada (X) tem distribuição Normal com média 5.000 *end foot candles* e desvio padrão 100 *end foot candles*.

1. Qual a probabilidade de $X > 4.800$?
2. Ache um limite de especificação tal que apenas $0,5\%$ das lâmpadas não excedam esse valor.
3. Qual deveria ser a média tal que $\Pr(X < 4.800) = 0.005$?

Distribuição Normal

Algumas propriedades da distribuição Normal:

- Se $X \sim N(\mu, \sigma^2)$, então:
  - $\Pr(\mu - \sigma < X < \mu + \sigma) \approx 0.6826$.
  - $\Pr(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.9546$.
  - $\Pr(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0.9973$.
- Por sua simetria em relação a μ:
  - $\Pr(X < \mu - \delta) = \Pr(X > \mu + \delta), \quad \text{para qualquer } \delta.$

Distribuição Normal

- Combinações lineares de variáveis aleatórias com distribuição Normal também têm distribuição Normal.
- Em particular, sejam

$$X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2), \cdots, X_n \sim N(\mu_n, \sigma_n^2)$$

variáveis aleatórias independentes, e $a_1, a_2, \cdots, a_n$ um conjunto de constantes. Então:

$$y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$$

tem distribuição Normal com média

$$\mu_y = a_1 \mu_1 + a_2 \mu_2 + \cdots + a_n \mu_n$$

e variância

$$\sigma_y^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \cdots + a_n^2 \sigma_n^2.$$

Exercício

Suponha que os ganhos mensais (em milhares de reais) de um investidor em três aplicações (1, 2 e 3) tenham distribuição $X_1 \sim N(10, 2)$, $X_2 \sim N(20, 5)$ e $X_3 \sim N(30, 9)$, independentes:

- Qual a probabilidade do rendimento acumulado em um mês qualquer ser superior a 65 mil reais?
- Qual a probabilidade do rendimento médio em um mês ser superior a 22 mil reais?
- Qual a distribuição de $2X_1 + 3X_2 - X_3$?

Teorema Central do Limite

O Teorema Central do Limite fundamenta o uso da distribuição Normal como aproximação em diversas situações.

Teorema Central do Limite (TCL)

Sejam $X_1, X_2, \cdots, X_n$ variáveis aleatórias independentes com média μ_i e variância σ_i^2, e $Y = \sum_{i=1}^{n} X_i$. Então, a distribuição de

$$\frac{Y - \sum_{i=1}^{n} \mu_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}$$

se aproxima da $N(0,1)$ quando $n \to \infty$, independente das distribuições individuais das variáveis.

Discutiremos adiante os tamanhos amostrais adequados para aplicação do TCL em problemas de Controle Estatístico de Qualidade.

Aproximação da distribuição binomial pela Normal

- Considere $Y \sim \text{binomial}(n, p)$. Podemos considerar $Y = X_1 + X_2 + \cdots + X_n$, em que cada X_i tem distribuição de Bernoulli com parâmetro p.
- Assim, o Teorema Central do Limite aplica-se à variável Y, garantindo que, para n suficientemente grande:

$$Y \sim \text{Normal}(\mu = np, \sigma^2 = np(1-p)),$$

ficando a aproximação melhor a medida que o n aumenta.

Aproximação da distribuição binomial pela Normal

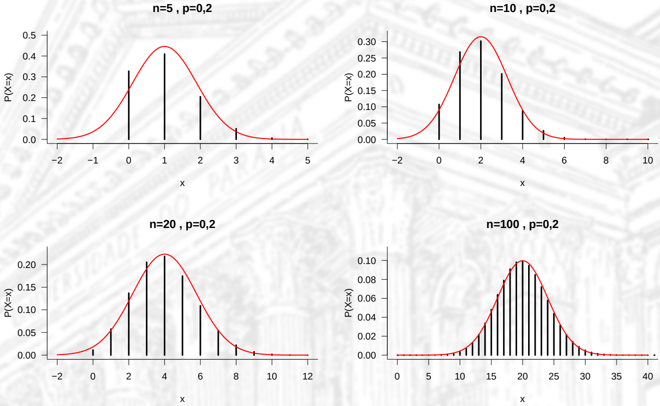


Figura 9. Aproximação da Normal pela binomial.

Distribuição logNormal

- Seja Y uma variável com distribuição Normal. Então, $X = \exp\{Y\}$ tem distribuição lognormal.
- Diferentemente da distribuição Normal, a distribuição lognormal tem seu suporte no conjunto dos reais positivos, assumindo formas assimétricas.
- A distribuição lognormal, assim como outras que estudaremos adiante, tem aplicações na modelagem de variáveis na área de confiabilidade, como tempo de vida e resistência de equipamentos, dentre outras.

Distribuição lognormal

Seja $Y \sim N(\theta, \omega^2)$. Então, $X = \exp\{Y\}$ tem distribuição lognormal com função densidade de probabilidade dada por:

$$f(x) = \frac{1}{x\omega\sqrt{2\pi}} \exp\left\{-\frac{(\ln(x) - \theta)^2}{2\omega^2}\right\}, \quad 0 < x < \infty,$$

com média e variância dadas por:

- $\mu = \exp\{\theta + \frac{\omega^2}{2}\}$.
- $\sigma^2 = \exp\{2\theta + \omega^2\}(\exp\{\omega^2\} - 1)$.

Distribuição log-normal

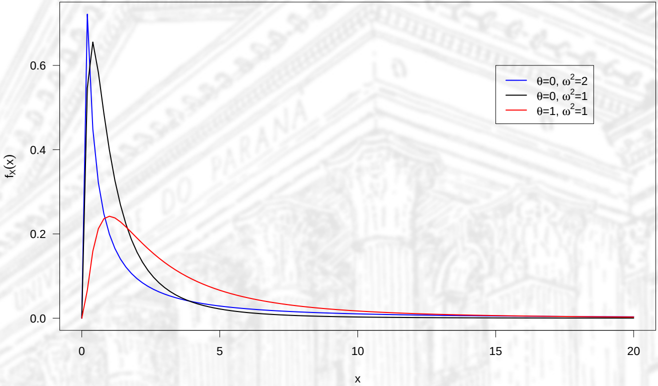
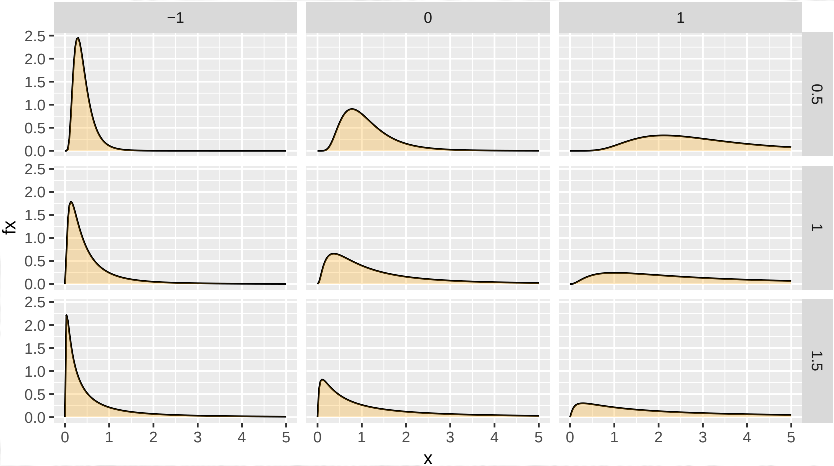


Figura 10. Distribuição lognormal.

Gráfico da distribuição lognormal com o R

```r
tb <- crossing(x = seq(0, 5, length.out = 151),
               meanlog = -1:1,
               sdlog = c(0.5, 1, 1.5)) %>%
    mutate(fx = dlnorm(x, meanlog = meanlog, sdlog = sdlog))

ggplot(tb, aes(x, fx)) +
    facet_grid(facets = sdlog ~ meanlog) +
    geom_line() +
    geom_area(alpha = 0.25, fill = "orange")
```



−1
0
1
0.5
1
1.5
fx
x
0.0
0.5
1.0
1.5
2.0
2.5
0
1
2
3
4
5

Distribuição exponencial

A variável aleatória X tem distribuição exponencial de parâmetro $\lambda > 0$ se sua função densidade de probabilidade é dada por:

$$f(x) = \lambda \exp\{-\lambda x\}, \quad x > 0.$$

A média e a variância de X são dadas, respectivamente, por:

- $\mu = \frac{1}{\lambda}$.
- $\sigma^2 = \frac{1}{\lambda^2}$.

Distribuição exponencial

- A distribuição exponencial é mais uma empregada frequentemente na modelagem do tempo de vida de equipamentos, componentes, sistemas, etc.
- A distribuição exponencial está relacionada à distribuição de Poisson, de forma que se o número de eventos em um intervalo (de tempo, por exemplo) de tamanho t tem distribuição de Poisson(λt), então o tempo até o próximo evento tem distribuição exponencial (λ).
- A distribuição exponencial tem como característica a falta de memória, segundo a qual:

$$\Pr(X > t + \delta | X > t) = \Pr(X > \delta), \quad \text{para qualquer } \delta > 0.$$

- Assim, a distribuição exponencial somente é aplicável em problemas de confiabilidade se não houver efeito (desgaste) do tempo de operação.

Distribuição exponencial

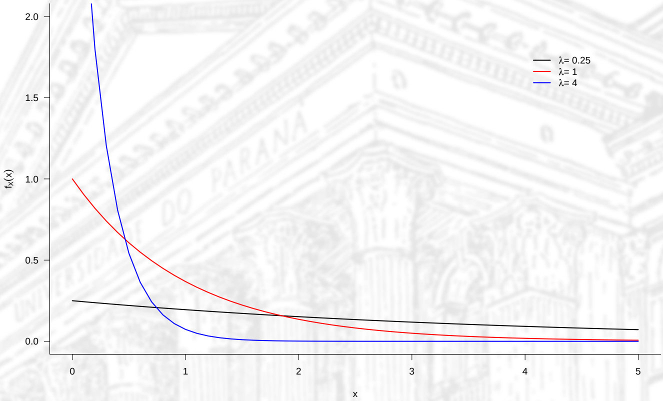
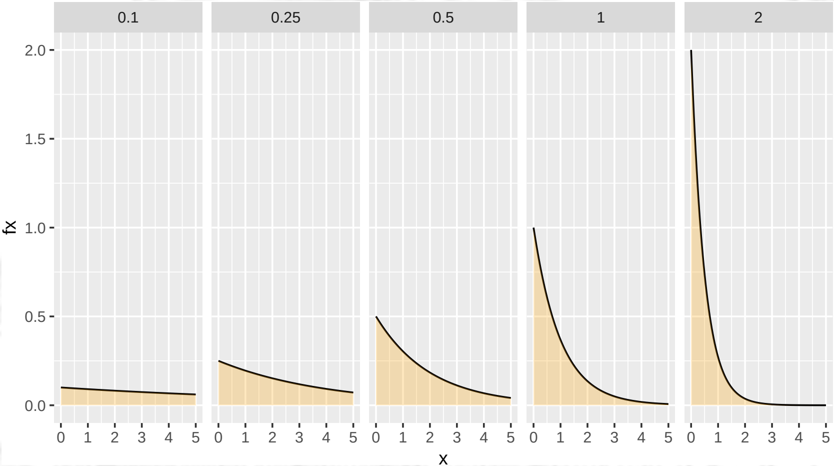


Figura 11. Distribuição exponencial.

Gráfico da distribuição exponencial com o R

```r
tb <- crossing(x = seq(0, 5, length.out = 151),
               rate = c(0.1, 0.25, 0.5, 1, 2)) %>%
    mutate(fx = dexp(x, rate = rate))

ggplot(tb, aes(x, fx)) +
    facet_grid(facets = ~ rate) +
    geom_line() +
    geom_area(alpha = 0.25, fill = "orange")
```


0.1
0.25
0.5
1
2
fx
2.0
1.5
1.0
0.5
0.0
0 1 2 3 4 5
x

Distribuição Gama

A variável aleatória X tem distribuição Gama de parâmetros $r > 0$ e $\lambda > 0$ se sua função densidade de probabilidade é dada por:

$$f(x) = \frac{\lambda}{\Gamma(r)}(\lambda x)^{(r-1)} \exp\{-\lambda x\}, \quad x > 0.$$

A média e a variância de X são dadas, respectivamente, por:

- $\mu = \frac{r}{\lambda}$.
- $\sigma^2 = \frac{r}{\lambda^2}$.

Distribuição Gama

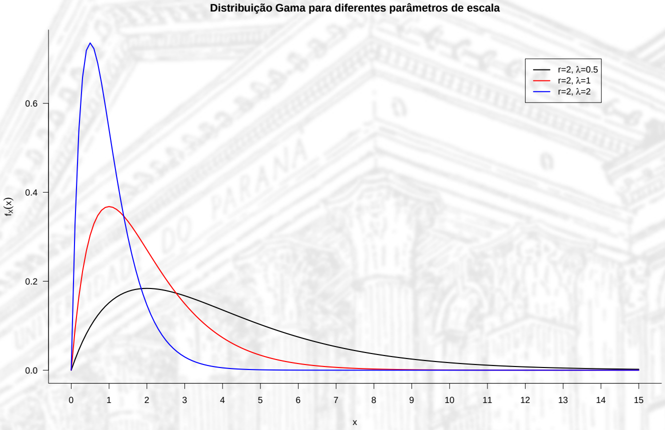


Figura 12. Distribuição gama I.

Distribuição Gama

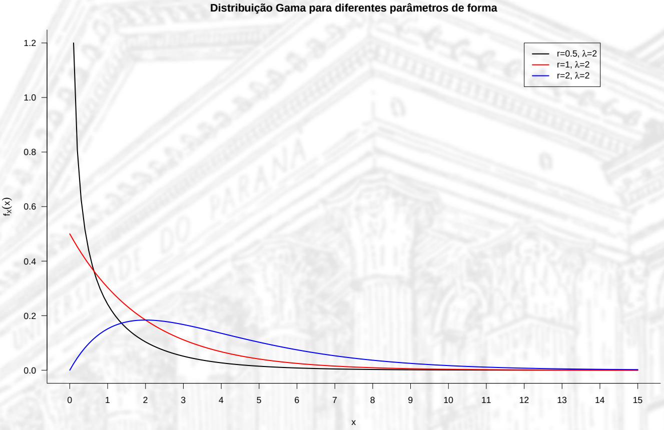
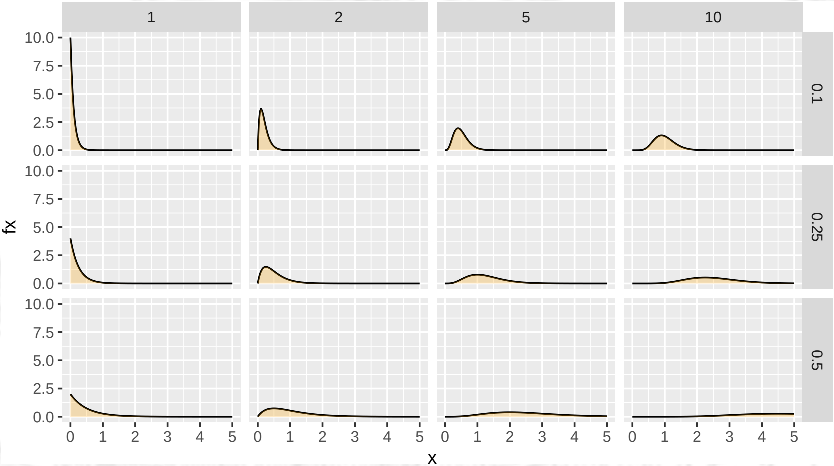


Figura 13. Distribuição gama II.

Gráfico da distribuição gama com o R

```r
tb <- crossing(x = seq(0, 5, length.out = 151),
               shape = c(1, 2, 5, 10),
               scale = c(0.1, 0.25, 0.5)) %>%
    mutate(fx = dgamma(x, shape = shape, scale = scale))

ggplot(tb, aes(x, fx)) +
    facet_grid(facets = scale ~ shape) +
    geom_line() +
    geom_area(alpha = 0.25, fill = "orange")
```


1
2
5
10
0.1
0.25
0.5
fx
x
10.0
7.5
5.0
2.5
0.0
0 1 2 3 4 5

Distribuição Gama

- A distribuição Gama tem como caso particular a distribuição exponencial (λ) ao fixarmos $r = 1$.
- Por apresentar um segundo parâmetro, a distribuição Gama proporciona funções com maior variedades de formas, permitindo modelar adequadamente um maior número de variáveis aleatórias.
- Se $X_1, X_2, \cdots, X_n$ são variáveis aleatórias independentes, com distribuição Gama de parâmetro λ, então:

$$Y = X_1 + X_2 + \cdots + X_n \sim \text{Gama}(n, \lambda).$$

Distribuição Weibull

A variável aleatória X tem distribuição Weibull de parâmetros $\theta > 0$ e $\beta > 0$ se sua função densidade de probabilidade é dada por:

$$f(x) = \frac{\beta}{\theta}\left(\frac{x}{\theta}\right)^{\beta-1} \exp\left\{-\left(\frac{x}{\theta}\right)^{\beta}\right\}, \quad x > 0.$$

A média e a variância de X são dadas, respectivamente, por:

- $\mu = \theta\Gamma\left(1 + \frac{1}{\beta}\right)$.
- $\sigma^2 = \theta^2\left\{\Gamma\left(1 + \frac{2}{\beta}\right) - \left\{\Gamma\left(1 + \frac{1}{\beta}\right)\right\}^2\right\}$.

Distribuição Weibull

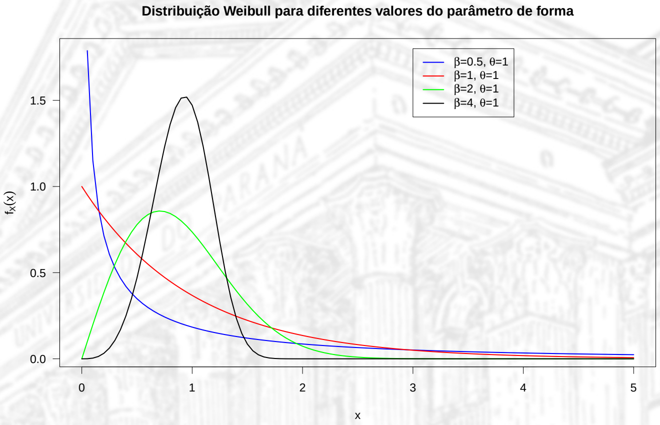


Figura 14. Distribuição Weibull I.

Distribuição Weibull

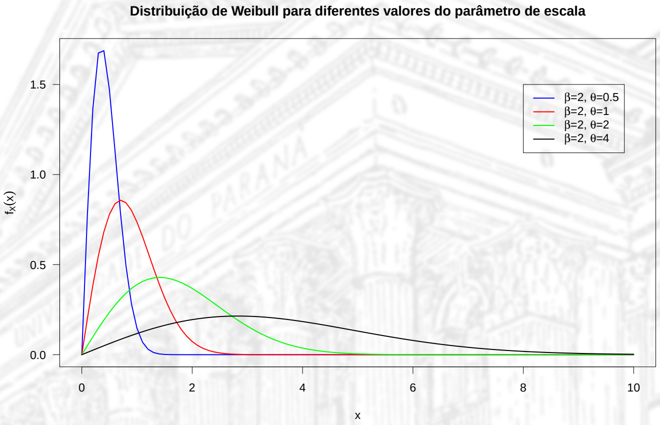
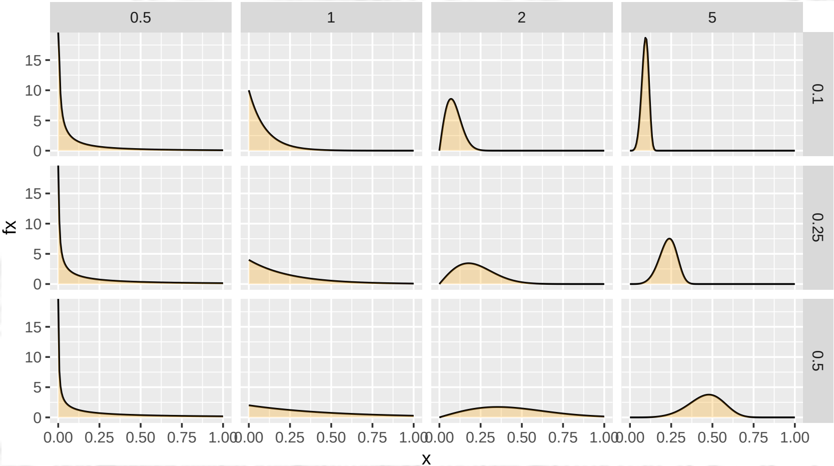


Figura 15. Distribuição Weibull II.

Gráfico da distribuição Weibull com o R

```r
tb <- crossing(x = seq(0, 1, length.out = 151),
               shape = c(0.5, 1, 2, 5),
               scale = c(0.1, 0.25, 0.5)) %>%
    mutate(fx = dweibull(x, shape = shape, scale = scale))

ggplot(tb, aes(x, fx)) +
    facet_grid(facets = scale ~ shape) +
    geom_line() +
    geom_area(alpha = 0.25, fill = "orange")
```


0.5
1
2
5
0.1
0.25
0.5
fx
x
15
10
5
0
0.00
0.25
0.50
0.75
1.00

Algumas distribuições de forma interativa

- Minha coleção: http://shiny.leg.ufpr.br/walmes/distribuicoes/.
- Coleção completa: https://statdist.ksmzn.com/.