Universidade Federal Fluminense

Instituto de Matemática e Estatística

Departamento de Estatística uff

Apostila de Inferência Estatística

Material elaborado para a disciplina
GET00135 - Inferência

Jessica Quintanilha Kubrusly

Março / 2022

Prefácio

Devido a pandemia de COVID-19 as universidades funcionaram por dois no formato remoto, com aulas à distância. Nesse novo cenário as notas de aula, que antes eram escritas no quadro e copiadas nos cadernos, precisaram ganhar um formato mais dinâmico, que pudessem chegar facilmente a todos os alunos. Daí surgiu a necessidade de organizar este texto didático.

Entre março de 2020 e março de 2022 foi o período em que este material foi elaborado. O texto foi organizado e escrito de forma direcionada para os alunos da disciplina GET00135 - Inferência. Junto com esse texto também foram gravadas videoaulas para que o aluno pudesse realizar seus estudos de forma assíncrona, alternando, da forma que preferissem, entre a leitura de texto e aula a ser assistida.

Parte dos exercícios propostos foram elaborados pela autora e outra parte retiradas de livros. Aqueles retirados de livros estão referenciados e as referências completas apresentadas ao final deste texto. Aos leitores, espera-se que tenham feito um curso de Probabilidade para que estejam aptos para o conteúdo apresentado.

Jessica Kubrusly

Sumário

Capítulo 1

Uma Breve Revisão: Distribuições Normal e Gama

Antes de apresentar os conceitos iniciais de um curso de inferência, será feita uma revisão de alguns itens de probabilidades muito usados ao longo de todo esse livro. A revisão concentra-se basicamente nas distribuições Normal e Gama.

1.1 A Distribuição Normal Padrão

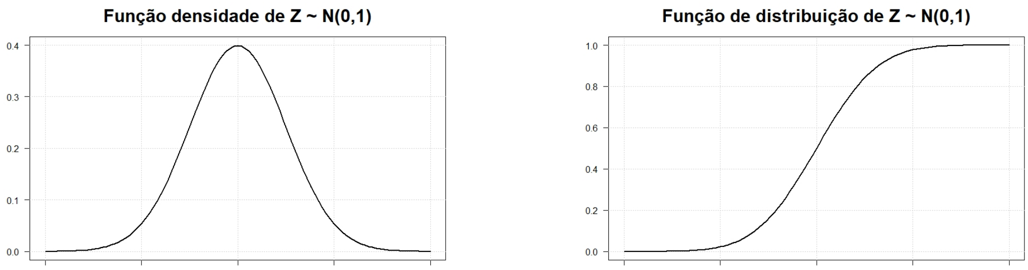
Todos os resultados para a distribuição normal são mais fáceis de serem deduzidos a partir dos resultados da distribuição normal padrão. Por isso, primeiro vamos rever a distribuição normal padrão para depois chegarmos à distribuição normal de parâmetros μ e σ^2.

Definição 1.1. *Uma variável aleatória contínua Z tem distribuição Normal Padrão se a sua função densidade for definida por:*

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}.$$

Nesse caso dizemos que $Z \sim N(0,1)$.

A partir do uso de coordenadas polares é possível mostrar que $\int_{-\infty}^{\infty} \varphi(z)\,\mathrm{d}z = 1$, e com isso mostramos que de fato $\varphi(z)$ definida acima é uma função densidade. Entretanto, a função $\varphi(z)$ não possuí antiderivada, o que dificulta os cálculos de probabilidades com a variável aleatória $Z \sim N(0,1)$.



(a) Função densidade

(b) Função distribuição

Figura 1.1: Função densidade e função de distribuição da Normal Padrão.

Vamos definir $\Phi(z)$ como a função de distribuição acumulada de Z, isto é,

$$\Phi(z) = \mathrm{P}(Z < t) = \int_{-\infty}^{z} \varphi(t)\,\mathrm{d}t.$$

Veja que não conhecemos a expressão de $\Phi(z)$, uma vez que $\varphi(z)$ não possuí antiderivada. Mas podemos, a partir de cálculos numéricos, encontrar aproximações para $\Phi(z)$ qualquer que seja $z \in \mathbb{R}$. É dessa forma que calculamos probabilidades envolvendo uma variável aleatória $Z \sim N(0,1)$. Os valores estimados para $\Phi(z)$ podem ser encontrados em tabelas estatísticas, como a Tabela A.1 do Apêndice A, ou podem ser calculados em *softwares* estatísticos como o Programa R. No Programa R a função `pnorm(z)` retorna o valor estimado para $\Phi(z)$.

Proposição 1.1. *Seja* $Z \sim N(0,1)$*. Então,* $\mathrm{E}(Z) = 0$ *e* $\mathrm{Var}(Z) = 1$.

Demonstração. Podemos perceber que a função φ é uma função par, ou seja, simétrica em torno de zero. Isso nos leva a concluir, de imediato, que $\mathrm{E}(Z) = 0$.

Dessa forma, $\mathrm{Var}(Z) = \mathrm{E}(Z^2)$ e

$$\mathrm{Var}(Z) = \mathrm{E}(Z^2) = \int_{-\infty}^{+\infty} z^2 \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \mathrm{d}z = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} z^2 e^{-z^2/2} \mathrm{d}z,$$

uma vez que a função a ser integrada é par. Esta integral é calculada usando o método de integração por partes. Faça $u = z$ e $\mathrm{d}v = ze^{-z^2/2}\mathrm{d}z$, então $\mathrm{d}u = \mathrm{d}z$ e $v = -e^{-z^2/2}$. Seguindo com as contas,

$$\begin{aligned}
\mathrm{Var}(Z) &= \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} z^2 e^{-z^2/2} \mathrm{d}z = \frac{2}{\sqrt{2\pi}} \left(\underbrace{-ze^{-z^2/2}\Big|_0^{+\infty}}_{0} + \int_0^{+\infty} e^{-z^2/2} \mathrm{d}z \right) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-z^2/2} \mathrm{d}z = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \mathrm{d}z = 1.
\end{aligned}$$

□

Por fim vamos encontrar a função geradora de momentos de uma variável aleatória normal padrão, representada por M_Z. Vale lembrar que, assim como a função densidade e a função de distribuição, a função geradora de momentos também é uma função que identifica a distribuição de uma variável aleatória. Isso significa que se duas variáveis aleatórias têm funções geradoras de momentos iguais, então elas seguem a mesma distribuição.

Proposição 1.2. *Seja* $Z \sim N(0,1)$*. Então,* $M_Z(t) = e^{t^2/2}$.

Demonstração.

$$\begin{aligned}
M_Z(t) = \mathrm{E}\left(e^{tZ}\right) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\left(\frac{x^2 - 2tx}{2} \right) \right] \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\left(\frac{x^2 - 2tx + t^2 - t^2}{2} \right) \right] \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\left(\frac{x^2 - 2tx + t^2}{2} \right) + \left(\frac{t^2}{2} \right) \right] \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\left(\frac{x^2 - 2tx + t^2}{2} \right) \right] \exp\left(\frac{t^2}{2} \right) \mathrm{d}x \\
&= \exp\left(\frac{t^2}{2} \right) \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x-t)^2}{2} \right] \mathrm{d}x}_{1} = e^{t^2/2}.
\end{aligned}$$

Para ver que a última integral resulta em 1 basta fazer a troca de variável $y = x - t$ e perceber que se chega na densidade de uma normal padrão. □

Veja que poderíamos usar a função geradora de momentos para encontrar a esperança e variância. Para isso é necessário calcular as derivadas primeira e segunda:

$$M'_Z(z) = tM_Z(t) \quad \text{e} \quad M''_Z(z) = M_Z(t) + tM'_Z(z) = (1+t^2)M_Z(t).$$

A partir das derivadas obtemos os resultados já vistos:

$$\mathrm{E}(Z) = M'_Z(0) = 0, \quad \mathrm{E}(Z^2) = M''_Z(0) = 1 \quad \Rightarrow \quad \mathrm{Var}(Z) = 1.$$

Exercícios da Seção 1.1

1.1.1 Seja $Z \sim N(0,1)$. Mostre que se $X = -Z$ então $X \sim N(0,1)$.

1.1.2 Seja $Z \sim N(0,1)$. A partir da Tabela A.1 encontre os valores numéricos das probabilidades pedidas a seguir. Em seguida repita o exercício usando a função `pnorm` do programa R em vez da tabela.

(a) $\mathrm{P}(Z > 2,43)$	(b) $\mathrm{P}(Z > -1,81)$	(c) $\mathrm{P}(Z < 2,76)$
(d) $\mathrm{P}(Z < -0,68)$	(e) $\mathrm{P}(0 < Z < 1,31)$	(f) $\mathrm{P}(-0,45 < Z < 0)$
(g) $\mathrm{P}(-1,4 < Z < 2,3)$	(h) $\mathrm{P}(1,21 < Z < 1,96)$	(i) $\mathrm{P}(-2,02 < Z < -0,06)$

1.1.3 Seja $Z \sim N(0,1)$. Encontre o valor de c tal que:
(a) $\mathrm{P}(Z < c) = 0,99$ (b) $\mathrm{P}(Z > c) = 0,15$ (c) $\mathrm{P}(Z > c) = 0,85$.

1.2 A Distribuição Normal

Defina agora uma nova variável aleatória X como função linear de $Z \sim N(0,1)$:

$$X = \mu + \sigma Z,$$

onde $\mu \in \mathbb{R}$ e $\sigma > 0$. Vamos encontrar a função densidade de X, para isso usaremos o método da função de distribuição.

$$F_X(x) = \mathrm{P}(X \leq x) = \mathrm{P}(\mu + \sigma Z \leq x) = \mathrm{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

Logo,

$$f_X(x) = \frac{\mathrm{d}F_X}{\mathrm{d}x}(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Uma variável aleatória com essa função densidade é denominada Normal, como definido abaixo.

Definição 1.2. *Diz-se que, uma variável aleatória contínua X, definida para todos os valores em $\mathbb{R}$, tem* ***distribuição normal*** *com parâmetros μ e σ^2, onde $-\infty < \mu < \infty$ e $0 < \sigma^2 < \infty$, se sua função densidade de probabilidade é dada por*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad , \qquad -\infty < x < \infty.$$

Usaremos a notação $X \sim N(\mu, \sigma^2)$ para indicar que X tem distribuição normal com parâmetros μ e σ^2. Se $\mu = 0$ e $\sigma^2 = 1$ chegamos na distribuição normal padrão, logo, esta é um caso particular da distribuição normal.

É importante destacar que se $X \sim N(\mu, \sigma^2)$ e $Z \sim N(0,1)$ então

$$\mu + \sigma Z \stackrel{\mathrm{d}}{=} X \qquad \text{e} \qquad \frac{X-\mu}{\sigma} \stackrel{\mathrm{d}}{=} Z, \tag{1.1}$$

onde $\stackrel{d}{=}$ é o símbolo de "identicamente distribuído". Esse é um resultado imediato a criação de X como função linear de Z, feita no início desta seção. E essa relação facilitará muito os resultados com a distribuição normal, a começar pelo cálculo de probabilidades envolvendo uma variável aleatória normal qualquer.

Assim como acontece com a distribuição Normal Padrão, a função densidade de uma variável aleatória normal não possui anti-deriva. Dessa forma o cálculo de probabilidades envolvendo uma variável aleatória normal não pode ser feito pela integração da sua densidade, e nem pela sua função de distribuição acumulada, uma vez que esta não exite. Para realizar esses cálculos vamos transformar a variável aleatória normal em uma normal padrão e consultar a tabela estatística, como feito na Seção 1.1. Veja o plo 1.1 a seguir.

Exemplo 1.1. *Seja* $X \sim N(1,4)$ *e* $Y \sim N(-3,2)$. *Calcule* $\mathrm{P}(X < 0)$ *e* $\mathrm{P}(Y < -1)$.

Solução:

Primeiro vamos encontrar $\mathrm{P}(X < 0)$. Como $X \stackrel{d}{=} \mu + \sigma Z = 1 + 2Z$ podemos afirmar que

$$\mathrm{P}(X < 0) = \mathrm{P}\left(1 + 2Z < 0\right) = \mathrm{P}\left(Z < -\frac{1}{2}\right), \quad \text{onde } Z \sim N(0,1).$$

Veja que o problema se transformou em um problema de encontrar probabilidades envolvendo uma variável aleatória normal padrão. Aplicando a propriedade de simetria da densidade normal padrão, a de complementariedade e usando a tabela estatística, ou um *software* como o R, chegamos em

$$\mathrm{P}(X < 0) = 1 - \Phi\left(\frac{1}{2}\right) = 1 - 0,6915 = 0,3085.$$

O cálculo de $\mathrm{P}(Y < -1)$ será feito de forma semelhante. Como $Y \stackrel{d}{=} \mu + \sigma Z = -3 + \sqrt{2}Z$ podemos afirmar que

$$\mathrm{P}(Y < -1) = \mathrm{P}\left(-3 + \sqrt{2}Z < -1\right) = \mathrm{P}\left(Z < \frac{2}{\sqrt{2}}\right) = \mathrm{P}\left(Z < \sqrt{2}\right), \quad \text{onde } Z \sim N(0,1).$$

Novamente o problema virou em um problema de encontrar probabilidades envolvendo uma variável aleatória normal padrão. Usando a tabela estatística, ou um *software* como o R, chegamos em

$$\mathrm{P}(Y < -1) = \Phi(\sqrt{2}) = 0,9213.$$

$\|$

Proposição 1.3. *Seja* $X \sim N(\mu, \sigma^2)$. *Então,* $\mathrm{E}(X) = \mu$ *e* $\mathrm{Var}(X) = \sigma^2$.

Demonstração. A maneira mais fácil de demonstrar esses resultados é usando novamente a relação entre X e Z, onde $X \sim N(\mu, \sigma^2)$ e $Z \sim N(0,1)$. Veja que como $X \stackrel{d}{=} \mu + \sigma Z$ temos $\mathrm{E}(X) = \mathrm{E}(\mu + \sigma Z)$ e $\mathrm{Var}(X) = \mathrm{Var}(\mu + \sigma Z)$.

Primeiro vamos encontrar $\mathrm{E}(X)$. Usando a afirmação acima, as propriedades de linearidade da esperança e o resultado $\mathrm{E}(Z) = 0$, chegamos em:

$$\mathrm{E}(X) = \mathrm{E}(\mu + \sigma Z) = \mu + \sigma\,\mathrm{E}(Z) = \mu.$$

Agora vamos encontrar $\mathrm{Var}(X)$. Usando a afirmação acima, as propriedades da variância e o resultado $\mathrm{Var}(Z) = 1$, chegamos em:

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

$\square$

Proposição 1.4. *Seja $X \sim N(\mu, \sigma^2)$. Então, $M_X(t) = e^{t\mu}e^{t^2\sigma^2/2}$.*

Demonstração. Seja $Z \sim N(0,1)$.

$$\begin{aligned} M_X(t) &= \mathrm{E}\left(e^{tX}\right) = \mathrm{E}\left(e^{t(\sigma Z+\mu)}\right) && \text{, uma vez que } X \stackrel{d}{=} \sigma Z + \mu \text{ (Equação 1.1),} \\ &= \mathrm{E}\left(e^{t\sigma Z}e^{t\mu}\right) = e^{t\mu}\,\mathrm{E}\left(e^{t\sigma Z}\right) \\ &= e^{t\mu}M_Z(t\sigma) = e^{t\mu}e^{\frac{t^2\sigma^2}{2}} && \text{, uma vez que } M_Z(t) = e^{t^2/2} \text{ (Proposição 1.2).} \end{aligned}$$

□

Proposição 1.5. *Seja $X \sim N(\mu, \sigma^2)$ e sejam $a, b \in \mathbb{R}$ com $a \neq 0$. Defina $Y = aX + b$. Então, $Y \sim N(a\mu + b, a^2\sigma^2)$.*

Demonstração. A demonstração será feita a partir da função geradora de momentos de Y. Vamos então encontrá-la.

$$M_Y(t) = \mathrm{E}\left(e^{tY}\right) = \mathrm{E}\left(e^{t(aX+b)}\right) = \mathrm{E}\left(e^{atX+bt}\right) = \mathrm{E}\left(e^{atX}e^{bt}\right) = e^{bt}\,\mathrm{E}\left(e^{atX}\right) = e^{bt}M_X(at).$$

Sabemos que a $M_X(at) = e^{t\mu}e^{t^2\sigma^2/2}$, logo

$$M_Y(t) = e^{bt}M_X(at) = e^{bt}e^{at\mu}e^{a^2t^2\sigma^2/2} = e^{t(a\mu+b)}e^{t^2(a^2\sigma^2)/2}.$$

Veja que essa é a função geradora de momentos de uma variável aleatória com distribuição $N(a\mu + b, a^2\sigma^2)$. Logo, pela unicidade da função geradora de momentos podemos concluir que $Y \sim N(a\mu + b, a^2\sigma^2)$. □

O mais importante da Proposição 1.5 é a conclusão de que qualquer função linear de uma variável aleatória normal é uma variável aleatória também com distribuição normal. Ou seja, se multiplicarmos uma normal por uma constante e/ou somarmos a outra constante, continuamos na distribuição normal.

O próximo resultado nos mostra que a distribuição normal também se mantém quando somamos duas variáveis aleatórias normais independente. A independência entre elas é uma hipótese importante, o resultado não vale para variáveis normais dependentes.

Proposição 1.6. *Sejam $X \sim N(\mu_X, \sigma_X^2)$ e $Y \sim N(\mu_Y, \sigma_Y^2)$ variáveis aleatórias independentes. Defina $W = X + Y$. Então, $W \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.*

Demonstração. Novamente usaremos a função geradora de momentos para a demonstração. Lembre-se que a função geradora de momentos da soma de duas variáveis aleatórias independentes é o produto entre as funções geradoras de momentos de cada uma delas.

$$M_W(t) = M_{X+Y}(t) = M_X(t)M_Y(t) = e^{t\mu_X}e^{t^2\sigma_X^2/2}e^{t\mu_Y}e^{t^2\sigma_Y^2/2} = e^{t(\mu_X+\mu_Y)}e^{t^2(\sigma_X^2+\sigma_Y^2)/2}.$$

Veja que o resultado é justamente a função geradora de momentos de uma $N(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2)$, logo esta é a distribuição de W. □

O último resultado desse seção é uma generalização do anterior. Veremos que qualquer combinação linear de variáveis aleatórias normais independentes resulta numa variável aleatória também com distribuição normal.

Proposição 1.7. *Sejam $X_i \sim N(\mu_i, \sigma_i^2)$ variáveis aleatórias independentes e $a_i \in \mathbb{R}$, com pelo menos um $a_i \neq 0$, $i = 1, \ldots, n$. Defina $W_n = \sum_{i=1}^n a_iX_i$. Então, $W_n \sim N\left(\sum_{i=1}^n a_i\mu_i, \sum_{i=1}^n a_i^2\sigma_i^2\right)$.*

Demonstração. A demonstração será feita por indução.

Primeiro vamos mostrar que o resultado é verdadeiro para $n = 1$, ou seja, que $W_1 = a_1 X_1 \sim N\left(a_1\mu_1, a_1^2\sigma_1^2\right)$ se $a_1 \neq 0$. Essa é uma aplicação imediata da Proposição 1.5, que nos mostra que qualquer função linear de uma variável aleatória normal também tem distribuição normal.

Em seguida, por hipótese de indução temos que $W_{n-1} \sim N\left(\sum_{i=1}^{n-1} a_i\mu_i, \sum_{i=1}^{n-1} a_i^2\sigma_i^2\right)$. Precisamos então mostrar que $W_n \sim N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$.

Veja que $W_n = W_{n-1} + a_n X_n$, sendo $W_{n-1} \sim N\left(\sum_{i=1}^{n-1} a_i\mu_i, \sum_{i=1}^{n-1} a_i^2\sigma_i^2\right)$ e $a_n X_n \sim N(a_n\mu_n, a_n^2\sigma_n^2)$. Veja também que W_{n-1} e $a_n X_n$ são variáveis aleatórias independentes, uma vez que W_{n-1} é função de variáveis todas independentes de X_n. Então, pela Proposição 1.6, que nos diz que a soma de variáveis aleatórias normais independentes também é uma variável aleatória normal, podemos concluir que

$$W_n \sim N\left(\sum_{i=1}^{n-1} a_i\mu_i + a_n\mu_n, \sum_{i=1}^{n-1} a_i^2\sigma_i^2 + a_n^2\sigma_n^2\right) = N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right).$$

□

Exercícios da Seção 1.2

1.2.1 Seja $X \sim N(8, 16)$. Calcule:
(a) $\mathrm{P}(X < 14)$ (b) $\mathrm{P}(X > 17)$ (c) $\mathrm{P}(X > 5)$ (d) $\mathrm{P}(X < 3)$ (e) $\mathrm{P}(4 < X < 16)$.

1.2.2 Seja $X \sim N(-1, 4)$. Encontre o valor de c tal que:
(a) $\mathrm{P}(X < c) = 0,85$ (b) $\mathrm{P}(X > c) = 0,10$ (c) $\mathrm{P}(X > c) = 0,95$.

1.2.3 Sejam $X_1 \sim N(1, 1)$, $X_2 \sim N(3, 4)$ e $X_3 \sim N(-1, 1)$ variáveis aleatórias independentes. Defina $Y = 2X_1 - X_2 + 3X_3$. Usando o resultado da Proposição 1.7, defina a distribuição de Y.

1.2.4 Seja $X \sim N(\mu, \sigma^2)$. Qual a probabilidade de X se distanciar da sua média mais de σ unidades?

1.3 A Distribuição Normal Multivariada

Definição 1.3. *Seja $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$ um vetor aleatório. Então a sua esperança é o vetor de esperanças $\mathrm{E}(\mathbf{X})$ e a sua variância é a matriz real simétrica $\mathrm{Var}(\mathbf{X})$ de dimensão $n \times n$ definidos por:*

$$\mathrm{E}(\mathbf{X}) = \begin{pmatrix} \mathrm{E}(X_1) \\ \mathrm{E}(X_2) \\ \vdots \\ \mathrm{E}(X_n) \end{pmatrix} \quad e \quad \mathrm{Var}(\mathbf{X}) = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \mathrm{Cov}(X_1, X_3) & \ldots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \mathrm{Cov}(X_2, X_3) & \ldots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \mathrm{Cov}(X_n, X_3) & \ldots & \mathrm{Var}(X_n) \end{pmatrix}.$$

Proposição 1.8. *Seja $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$ um vetor aleatório, um vetor $\mathbf{v} \in \mathbb{R}^m$ e uma matriz real A de dimensão $m \times n$. Então, $\mathbf{Y} = A\mathbf{X} + \mathbf{v}$ é um vetor aleatório com média e variância dadas por:*

(i) $\mathrm{E}(\mathbf{Y}) = A\,\mathrm{E}(\mathbf{X}) + \mathbf{v}$;

(ii) $\mathrm{Var}(\mathbf{Y}) = A\,\mathrm{Var}(\mathbf{X})A^T$.

Demonstração. Seja $\mathbf{Y} = A\mathbf{X} + \mathbf{v} = (Y_1, Y_2, \ldots, Y_n)^T$, $a_i = (A_{i,1}, A_{i,2}, \ldots, A_{i,n})$ a i-ésima linha da matriz A e v_i o i-ésimo elemento do vetor v. Então, $Y_i = a_i\mathbf{X} + v_i = \sum_{j=1}^{n} A_{i,j}X_j + v_i$. Essa informação será usada na demonstração dos dois itens.

(i) Como $Y_i = \sum_{j=1}^n A_{i,j}X_j + v_i$, usando as propriedade da esperança concluímos que $\mathrm{E}(Y_i) = \sum_{j=1}^n A_{i,j}\,\mathrm{E}(X_j) + v_i$. Assim podemos escrever o vetor esperança como:

$$\mathrm{E}(\mathbf{Y}) = \begin{pmatrix} \mathrm{E}(Y_1) \\ \mathrm{E}(Y_2) \\ \vdots \\ \mathrm{E}(Y_m) \end{pmatrix} = \begin{pmatrix} A_{1,1}\,\mathrm{E}(X_1) + A_{1,2}\,\mathrm{E}(X_2) + \ldots + A_{1,n}\,\mathrm{E}(X_n) & + & v_1 \\ A_{2,1}\,\mathrm{E}(X_1) + A_{2,2}\,\mathrm{E}(X_2) + \ldots + A_{2,n}\,\mathrm{E}(X_n) & + & v_2 \\ \vdots & & \\ A_{m,1}\,\mathrm{E}(X_1) + A_{m,2}\,\mathrm{E}(X_2) + \ldots + A_{m,n}\,\mathrm{E}(X_n) & + & v_m \end{pmatrix}$$

$$= \begin{pmatrix} A_{1,1} & A_{1,2} & \ldots & A_{1,n} \\ A_{2,1} & A_{2,2} & \ldots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1,1} & A_{1,2} & \ldots & A_{1,n} \end{pmatrix} \begin{pmatrix} \mathrm{E}(X_1) \\ \mathrm{E}(X_2) \\ \vdots \\ \mathrm{E}(X_n) \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = A\,\mathrm{E}(\mathbf{X}) + \mathbf{v}.$$

(ii) Seja $\mathrm{Var}(\mathbf{Y})_{i,l}$ a posição (i,l) da matriz de covariância do vetor aleatório $\mathbf{Y}$. Veja agora que

$$\mathrm{Var}(\mathbf{Y})_{i,l} = \mathrm{Cov}(Y_i, Y_l) = \mathrm{Cov}\left(\sum_{j=1}^n A_{i,j}X_j + v_i, \sum_{k=1}^n A_{l,k}X_k + v_l\right) = \sum_{j=1}^n \sum_{k=1}^n A_{l,k}A_{i,j}\,\mathrm{Cov}(X_j, X_k).$$

Vamos encontrar também a posição (i,l) da matriz $A\,\mathrm{Var}(\mathbf{X})A^T$, denotada por $A\,\mathrm{Var}(\mathbf{X})A^T_{i,l}$.

$$A\,\mathrm{Var}(\mathbf{X})A^T =$$

$$= \begin{pmatrix} \longleftarrow a_1 \longrightarrow \\ \longleftarrow a_2 \longrightarrow \\ \vdots \\ \longleftarrow a_i \longrightarrow \\ \vdots \\ \longleftarrow a_n \longrightarrow \end{pmatrix} \begin{pmatrix} \mathrm{Cov}(X_1,X_1) & \mathrm{Cov}(X_1,X_2) & \ldots & \mathrm{Cov}(X_1,X_n) \\ \mathrm{Cov}(X_2,X_1) & \mathrm{Cov}(X_2,X_2) & \ldots & \mathrm{Cov}(X_2,X_n) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{Cov}(X_n,X_1) & \mathrm{Cov}(X_n,X_2) & \ldots & \mathrm{Cov}(X_n,X_n) \end{pmatrix} \begin{pmatrix} \uparrow & \uparrow & & \uparrow & & \uparrow \\ a_1 & a_2 & \ldots & a_l & \ldots & a_n \\ \downarrow & \downarrow & & \downarrow & & \downarrow \end{pmatrix}$$

$$= \begin{pmatrix} \longleftarrow a_1 \longrightarrow \\ \longleftarrow a_2 \longrightarrow \\ \vdots \\ \longleftarrow a_i \longrightarrow \\ \vdots \\ \longleftarrow a_n \longrightarrow \end{pmatrix} \begin{pmatrix} \ldots & \sum_{j=1}^n A_{l,j}\,\mathrm{Cov}(X_1, X_j) & \ldots \\ \ldots & \sum_{j=1}^n A_{l,j}\,\mathrm{Cov}(X_2, X_j) & \ldots \\ \vdots & \vdots & \vdots \\ \ldots & \underbrace{\sum_{j=1}^n A_{l,j}\,\mathrm{Cov}(X_n, X_j)}_{\text{coluna } l} & \ldots \end{pmatrix}$$

Assim, multiplicando a linha i da primeira matriz pela coluna l da segunda encontramos $A\,\mathrm{Var}(\mathbf{X})A^T_{i,l}$:

$$A\,\mathrm{Var}(\mathbf{X})A^T_{i,l} = \sum_{j=1}^n \sum_{k=1}^n A_{i,k}A_{l,j}\,\mathrm{Cov}(X_k, X_j).$$

Como $\mathrm{Var}(\mathbf{Y})_{i,l} = A\,\mathrm{Var}(\mathbf{X})A^T_{i,l}$ para todo $1 \le i \le m$ e $1 \le l \le m$, podemos concluir que

$$\mathrm{Var}(\mathbf{Y}) = A\,\mathrm{Var}(\mathbf{X})A^T.$$

□

Proposição 1.9. *Seja* $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$ *um vetor aleatório com* $\mathrm{E}(\mathbf{X}) = \boldsymbol{\mu} \in \mathbb{R}^n$ *e* $\mathrm{Var}(\mathbf{X}) = \Sigma$. *As seguintes afirmações são equivalentes:*

(i) *Qualquer combinação linear das variáveis* $X_1, X_2, \ldots, X_n$ *é uma variável aleatória normal. Isto é, se* $\alpha_i \in \mathbb{R}$, $i = 1, 2, \ldots, n$, $\boldsymbol{\alpha} \neq \mathbf{0}$ *e* $Y = \sum_{i=1}^{n} \alpha_i X_i$, *então* Y *é variável aleatória normal.*

(ii) *A função geradora de momentos conjunta de* $\mathbf{X}$ *é:*

$$M_{\mathbf{X}}(t_1, t_2, \ldots, t_n) = e^{\mathbf{t}^T \boldsymbol{\mu}} e^{\frac{1}{2}\mathbf{t}^T \Sigma \mathbf{t}} \ , \qquad \forall\ \mathbf{t} \in \mathbb{R}^n$$

sendo Σ *uma matriz real simétrica positiva definida.*

(iii) *Existe uma variável aleatória* $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_m)^T$ *tal que* $Z_i \sim N(0,1)$, Z_i *é independente de* Z_j *para todo* $i \neq j$, *e existe uma matriz real* A, $n \times m$, *tal que*

$$\mathbf{X} = A\mathbf{Z} + \boldsymbol{\mu}.$$

Isto é, $\mathbf{X}$ *é uma função linear de vetores aleatórios independentes com distribuição normal padrão.*

A demonstração da Proposição 1.9 será omitida nesse trabalho. Esse resultado pode ser encontrado, por exemplo, em [Bickel e Doksum, 2001].

Definição 1.4. *Seja* $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$ *um vetor aleatório com* $\mathrm{E}(\mathbf{X}) = \boldsymbol{\mu} \in \mathbb{R}^n$ *e* $\mathrm{Var}(\mathbf{X}) = \Sigma$. *Se uma das afirmações descritas na Proposição 1.9 for satisfeita, dizemos que o vetor aleatório* $\mathbf{X}$ *tem distribuição Normal Multivariada com parâmetros* $\boldsymbol{\mu}$ *e* Σ. *Nesse caso, todas as afirmações descritas na Proposição 1.9 serão satisfeitas. Notação:* $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$.

O importante de ser entendido é a junção da Proposição 1.9 com a Definição 1.4. Para qualquer vetor aleatório com distribuição normal multivariada, isto é $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$, podemos assumir que são verdades as três afirmações apresentadas na Proposição 1.9. Com isso podemos demonstrar mais alguns resultados, como os apresentados a seguir.

Proposição 1.10. *Seja* $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T \sim N_n(\boldsymbol{\mu}, \Sigma)$. *Então,*

(i) *Cada* $X_i \sim N(\mu_i, \Sigma_{i,i})$.

(ii) *Seja* A *uma matriz real* $m \times n$ *e* $\mathbf{v} \in \mathbb{R}^m$. *Então,* $\mathbf{Y} = A\mathbf{X} + \mathbf{v} \sim N_m(A\boldsymbol{\mu} + \mathbf{v}, A\Sigma A^T)$.

(iii) X_i *e* X_j *são independentes se e somente se* $\Sigma_{i,j} = 0$, *isto é,* $\mathrm{Cov}(X_i, X_j) = 0$.

Demonstração. (i) Seja $A = (0, \ldots, 0, 1, 0, \ldots, 0)$ uma matriz $1 \times n$ com $A_{1,i} = 1$ e todas as demais posições são nulas . Veja que

$$A\mathbf{X} = \begin{pmatrix} 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_{i-1} \\ X_i \\ X_{i+1} \\ \vdots \\ X_n \end{pmatrix} = X_i.$$

Assim podemos ver que X_i é uma combinação linear de $\mathbf{X}$ e, pelo item i da Proposição 1.9, X_i tem distribuição normal. Vamos encontrar a esperança e variância de X_i. Para isso usaremos as propriedades de esperança e variância de vetores aleatórios:

$$\mathrm{E}(A\mathbf{X} + v) = A\,\mathrm{E}(\mathbf{X}) + \mathbf{v} \quad \text{ e } \quad \mathrm{Var}(A\mathbf{X} + v) = A\,\mathrm{Var}(\mathbf{X})A^T.$$

Primeiro fazendo as contas para a esperança.

$$\mathrm{E}(X_i) = \mathrm{E}\,(A\mathbf{X}) = A\,\mathrm{E}(\mathbf{X}) = \begin{pmatrix}0 & \dots & 0 & 1 & 0 & \dots & 0\end{pmatrix}\begin{pmatrix}\mu_1\\ \vdots \\ \mu_{i-1}\\ \mu_i \\ \mu_{i+1} \\ \vdots \\ \mu_n\end{pmatrix} = \mu_i$$

Agora encontrando a variância.

$$\begin{aligned}
\mathrm{Var}(X_i) &= \mathrm{Var}(A\mathbf{X}) = A\,\mathrm{Var}(\mathbf{X})A^T \\
&= A\begin{pmatrix}\mathrm{Var}(X_1) & \dots & \mathrm{Cov}(X_1,X_i) & \dots & \mathrm{Cov}(X_1,X_n)\\ \mathrm{Cov}(X_2,X_1) & \dots & \mathrm{Cov}(X_2,X_i) & \dots & \mathrm{Cov}(X_2,X_n)\\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathrm{Cov}(X_n,X_1) & \dots & \mathrm{Cov}(X_n,X_i) & \dots & \mathrm{Var}(X_n)\end{pmatrix}\begin{pmatrix}0\\ \vdots \\ 1 \\ \vdots \\ 0\end{pmatrix} \\
&= \begin{pmatrix}0 & \dots & 0 & 1 & 0 & \dots & 0\end{pmatrix}\begin{pmatrix}\mathrm{Cov}(X_i,X_1)\\ \vdots \\ \mathrm{Cov}(X_i,X_{i-1}) \\ \mathrm{Var}(X_i) \\ \mathrm{Cov}(X_i,X_{i+1}) \\ \dots \\ \mathrm{Cov}(X_i,X_n)\end{pmatrix} = \mathrm{Var}(X_i).
\end{aligned}$$

Assim concluímos que

$$X_i \sim N(\mu_i, \Sigma_{i,i})$$

(ii) Primeiro vamos ver que $\mathrm{E}(\mathbf{Y}) = A\boldsymbol{\mu} + \mathbf{v}$ e que $\mathrm{Var}(\mathbf{Y}) = A\Sigma A^T$. Em seguida veremos que $\mathbf{Y}$ tem distribuição conjunta normal multivariada.

Calculando a esperança de $\mathbf{Y}$ a partir das propriedades de linearidade do operador E.

$$\begin{aligned}
\mathrm{E}\,(\mathbf{Y}) &= \mathrm{E}\,(A\mathbf{X} + \mathbf{v}) \\
&= A\,\mathrm{E}\,(\mathbf{X}) + \mathbf{v} \\
&= A\boldsymbol{\mu} + \mathbf{v}.
\end{aligned}$$

Agora o cálculo da $\mathrm{Var}(\mathbf{Y})$.

$$\begin{aligned}
\mathrm{Var}\,(\mathbf{Y}) &= \mathrm{Var}\,(A\mathbf{X} + \mathbf{v}) \\
&= A\,\mathrm{Var}\,(\mathbf{X})\,A^T \\
&= A\Sigma A^T.
\end{aligned}$$

Seja $\mathbf{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$. Pelo item (iii) da Proposição 1.9 sabemos que existe $\mathbf{Z} = (Z_1, \dots, Z_m)^T$ tal que $Z_i \sim N(0,1)$, Z_i é independente de Z_j para todo $i \neq j$, e existe uma matriz real M, $n \times m$, tal que

$$\mathbf{X} = M\mathbf{Z} + \boldsymbol{\mu}.$$

Como $\mathbf{Y} = A\mathbf{X} + \mathbf{v}$ podemos concluir que

$$\mathbf{Y} = A\,(M\mathbf{Z} + \boldsymbol{\mu}) + \mathbf{v} \Rightarrow \mathbf{Y} = AM\mathbf{Z} + A\boldsymbol{\mu} + \mathbf{v} \Rightarrow \mathbf{Y} = AM\mathbf{Z} + \mathrm{E}\,(\mathbf{Y})\,.$$

Ou seja, existe $\mathbf{Z}$ tal que $Z_i \sim N(0,1)$, Z_i é independente de Z_j para todo $i \neq j$, e existe uma matriz real $B = AM$, $n \times n$, tal que

$$\mathbf{Y} = B\mathbf{Z} + \mathrm{E}(\mathbf{Y}).$$

E, novamente usando o item (iii) da Proposição 1.9, concluímos que $\mathbf{Y}$ é um vetor aleatório com distribuição normal multivariada.

(iii) Já sabemos que se X_i e X_j são independentes, então $\mathrm{Cov}(X_i, X_j) = 0$. Esse resultado é verdade para qualquer par de variáveis aleatórias. Veremos agora que para o caso de duas variáveis aleatórias com distribuição conjunta normal multivariada a recíproca também é verdadeira. Então, por hipótese, temos $\mathrm{Cov}(X_i, X_j) = 0$.

Primeiro vamos encontrar a função geradora de momentos conjunta de X_i e X_j. Já sabemos que X_i e X_j têm distribuição conjunta normal bivariada (item anterior). Então, pelo item (ii) da Proposição 1.9,

$$M_{X_i,X_j}(t,s) = e^{(t,s)\boldsymbol{\mu}} e^{((t,s)\Sigma(t,s)^T)/2}\ , \text{ onde } \boldsymbol{\mu} = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix} \text{ e } \Sigma = \begin{pmatrix}\mathrm{Var}(X_i) & 0\\ 0 & \mathrm{Var}(X_j)\end{pmatrix}.$$

Assim chegamos em

$$M_{X_i,X_j}(t,s) = e^{t\mu_i + s\mu_j} e^{(t^2\sigma_i^2 + s^2\sigma_j^2)/2} = e^{t\mu_i} e^{s\mu_j} e^{t^2\sigma_i^2/2} e^{s^2\sigma_j^2/2}.$$

Agora considere $Y \sim N(\mu_j, \sigma_j^2)$ independente de X_i. Vamos encontrar a função geradora de momentos conjunta de Y e X_i. Lembre-se que se as variáveis são independentes, a função geradora de momentos conjunta é o produto das funções geradoras de momento marginais.

$$M_{X_i,Y}(t,s) = M_{X_i}(t) M_Y(s) = e^{t\mu_i} e^{t^2\sigma_i^2/2} e^{s\mu_j} e^{s^2\sigma_j^2/2}.$$

Veja que $M_{X_i,Y} = M_{X_i,X_j}$. Logo, pela unicidade da função geradora de momentos, $Y \stackrel{\mathrm{d}}{=} X_j$, o que nos faz concluir que X_j é independente de X_i.

□

Exercícios da Seção 1.3

1.3.1 Seja (X, Y) vetor aleatório com distribuição normal bivariada de parâmetros $\mu_X = 3$, $\mu_Y = -2$, $\sigma_X^2 = 4$, $\sigma_Y^2 = 9$ e $\rho = -0.25$. Calcule $P(X < Y)$.
Dica: primeiro encontre a distribuição de $X - Y$ e depois calcule $P(X - Y < 0)$.

1.3.2 Sejam X e W variáveis aleatórias de média nula e distribuição conjunta normal bivariada tais que $\sigma_X^2 = 4$, $\sigma_W^2 = 17/9$ e $\mathrm{E}(XW) = 2$. Defina $Y = 2X - 3W$ e encontre:

(a) a distribuição de Y;

(b) a distribuição conjunta de (X, Y);

(c) a distribuição de $X|Y = y$.

Dica: $\mathrm{Cov}(X, W) = \mathrm{E}(XW) - \mathrm{E}(X)\,\mathrm{E}(W)$ e $\mathrm{Var}(Y) = \mathrm{Cov}(Y, Y)$.

1.3.3 Seja $X = (X_1, X_2, X_3)$ vetor aleatório com distribuição normal multivariada de parâmetros $\mu = (0, 1, -1)$ e $\Sigma = \begin{pmatrix}3 & 0 & 1\\ 0 & 1 & 0\\ 1 & 0 & 2\end{pmatrix}$.

(a) Defina as distribuições marginais de X_1, X_2 e X_3.

(b) Defina a distribuição conjunta de (X_1, X_2).

(c) Defina a distribuição de $X_1|X_3 = x_3$.

(d) As variáveis aleatórias X_1 e X_2 são independentes? Justifique sua resposta.

(e) As variáveis aleatórias X_1 e X_3 são independentes? Justifique sua resposta.

(f) As variáveis aleatórias $X_1 + X_2$ e $X_1 - X_2$ são independentes? Justifique sua resposta.

(g) Seja $Y = (Y_1, Y_2)$ vetor aleatório tal que $Y = AX + a$ com $A = \begin{pmatrix} 1 & 1 & 0 \\ 2 & -1 & 1 \end{pmatrix}$ e $a = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Encontre a distribuição conjunta de (Y_1, Y_2).

1.3.4 Seja X_1, X_2, ..., X_n variáveis aleatórias independentes tais que cada $X_i \sim N(\mu_i, \sigma_i^2)$. Mostre que o vetor aleatório $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ tem distribuição normal multivariada. Quem são os vetores média e a matriz de covariância de $\mathbf{X}$

1.4 A Função Gama

Para começar a revisão da distribuição vamos lembrar da função Gama e suas propriedades.

Definição 1.5. *Chama-se* ***Função Gama*** *a função de variável real estritamente positiva,* $\Gamma : \mathbb{R}^+ \to \mathbb{R}$*, definida por:*

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \mathrm{d}x, \qquad \alpha > 0.$$

Veja que $\Gamma(\alpha) \geq 0$ qualquer que seja $\alpha > 0$, uma vez que, para qualquer $\alpha > 0$ e $x > 0$, temos $x^{\alpha-1}e^x > 0$. Vamos mostrar que $\Gamma(\alpha)$ converge, isto é, é um número real qualquer que seja $\alpha > 0$.

Proposição 1.11. *Γ é uma função real, isto é, $\Gamma(\alpha) < \infty$, $\forall\ \alpha > 0$.*

Demonstração. Dado $\alpha > 0$ existe $a > 0$ tal que, $\forall\ x > a$, $\frac{x^{\alpha+1}}{e^x} < 1$, uma vez que $\frac{x^{\alpha+1}}{e^x} \to 0$ qualquer que seja $\alpha > 0$. Como

$$\frac{x^{\alpha+1}}{e^x} < 1 \Rightarrow \frac{x^{\alpha-1}x^2}{e^x} < 1 \Rightarrow \frac{x^{\alpha-1}}{e^x} < \frac{1}{x^2} \Rightarrow x^{\alpha-1}e^{-x} < \frac{1}{x^2}\ ,$$

podemos afirmar que, dado $\alpha > 0$ existe $a > 0$ tal que $\forall\ x > a$ é verdade que $x^{\alpha-1}e^{-x} < \frac{1}{x^2}$.

Então, dado $\alpha > 0$ seja $a > 0$ o número real com tal característica. Reescreva $\Gamma(\alpha)$ da seguinte maneira,

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\mathrm{d}x = \underbrace{\int_0^a x^{\alpha-1}e^{-x}\mathrm{d}x}_{I_1} + \underbrace{\int_a^\infty x^{\alpha-1}e^{-x}\mathrm{d}x}_{I_2} = I_1 + I_2.$$

Vamos mostrar que $I_1 < \infty$ e $I_2 < \infty$, consequentemente $\Gamma(\alpha) < \infty$.

$$I_1 = \int_0^a x^{\alpha-1}e^{-x}\mathrm{d}x < \int_0^a x^{\alpha-1}\mathrm{d}x = \left.\frac{x^\alpha}{\alpha}\right|_0^a = \frac{a^\alpha}{\alpha} < \infty.$$

$$I_2 = \int_a^\infty x^{\alpha-1}e^{-x}\mathrm{d}x < \int_a^\infty \frac{1}{x^2}\mathrm{d}x = -\left.\frac{1}{x}\right|_a^\infty = \frac{1}{a} < \infty.$$

□

Vejamos agora algumas propriedades da função Gama, apresentadas nas proposições a seguir.

Proposição 1.12. *(Função Gama Modificada) Dados $\alpha > 0$ e $\lambda > 0$, então*

$$\int_0^\infty x^{\alpha-1}e^{-\lambda x}\mathrm{d}x = \frac{\Gamma(\alpha)}{\lambda^\alpha}.$$

Demonstração. Resolva a integral pela seguinte troca de variável: $y = \lambda x \Rightarrow x = y/\lambda \Rightarrow \mathrm{d}x = \mathrm{d}y/\lambda$.

$$\int_0^\infty x^{\alpha-1}e^{-\lambda x}\mathrm{d}x = \int_0^\infty \left(\frac{y}{\lambda}\right)^{\alpha-1} e^{-y}\frac{1}{\lambda}\mathrm{d}y = \frac{1}{\lambda^\alpha}\int_0^\infty y^{\alpha-1}e^{-y}\mathrm{d}y = \frac{\Gamma(\alpha)}{\lambda^\alpha}.$$

□

Proposição 1.13. *Se $\alpha > 1$, então $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$.*

Demonstração. Resolvendo a integral por partes, escolha $u = x^{\alpha-1}$ e $dv = e^{-x}\mathrm{d}x$.

$$\begin{aligned}\Gamma(\alpha) &= \int_0^\infty x^{\alpha-1}e^{-x}\mathrm{d}x = \left(x^{\alpha-1}(-1)e^{-x}\right)\Big|_0^\infty + \int_0^\infty (\alpha-1)x^{\alpha-2}e^{-x}\mathrm{d}x \\ &= 0 + (\alpha-1)\int_0^\infty x^{\alpha-2}e^{-x}\mathrm{d}x = (\alpha-1)\Gamma(\alpha-1).\end{aligned}$$

□

Proposição 1.14. *Dado $n \in \mathbb{N}^+ = \{1, 2, \ldots\}$, $\Gamma(n) = (n-1)!$.*

Demonstração. Faremos pro indução. Primeiro veja que a afirmação é verdadeira para $n = 1$.

$$\Gamma(1) = \int_0^\infty e^{-x}\mathrm{d}x = 1 = 0!.$$

Agora, supondo que a afirmação é verdade para n, vamos mostrar que ela também é verdade para $n+1$. Ou seja, queremos mostrar que se $\Gamma(n) = (n-1)!$ então $\Gamma(n+1) = n!$.

Pela Proposição 1.13 podemos escrever

$$\Gamma(n+1) = n\Gamma(n).$$

Usando agora a hipótese de indução, $\Gamma(n) = (n-1)!$, podemos concluir que

$$\Gamma(n+1) = n\Gamma(n) = n(n-1)! = n!.$$

Logo a afirmação é verdadeira para qualquer $n \in \mathbb{N}^+$. □

Proposição 1.15. $\Gamma(1/2) = \sqrt{\pi}$.

Demonstração.

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty x^{\frac{1}{2}-1}e^{-x}\mathrm{d}x = \int_0^\infty x^{-\frac{1}{2}}e^{-x}\mathrm{d}x = \int_0^\infty \frac{1}{\sqrt{x}}e^{-x}\mathrm{d}x$$

Vamos resolver a partir da seguinte troca de variável: $u = \sqrt{2x} \Rightarrow x = u^2/2 \Rightarrow \mathrm{d}x = u\mathrm{d}u$.

$$\begin{aligned}\Gamma\left(\frac{1}{2}\right) &= \int_0^\infty \frac{\sqrt{2}}{u}e^{-\frac{u^2}{2}}u\mathrm{d}u = \sqrt{2}\int_0^\infty e^{-\frac{u^2}{2}}\mathrm{d}u = \frac{\sqrt{2}}{2}\int_{-\infty}^\infty e^{-\frac{u^2}{2}}\mathrm{d}u \\ &= \frac{\sqrt{2}}{2}\sqrt{2\pi}\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\mathrm{d}u = \sqrt{\pi}.\end{aligned}$$

□

Exercícios da Seção 1.4

1.4.1 Encontre os valores de $\Gamma(5)$ e $\Gamma(3,5)$. Dica: reveja as Proposições 1.13, 1.14 e 1.15.

1.4.2 Calcule as integrais abaixo usando o resultado da Proposição 1.12. Quando possível encontre a resposta numérica. Quando isso não for possível, reduza ao máximo o argumento da função Gama.

$$\text{(a)} \int_0^\infty x^3 e^{-2x} dx \qquad \text{(b)} \int_0^\infty x^{-\frac{1}{3}} e^{-\frac{x}{4}} dx \qquad \text{(c)} \int_0^\infty x^{\frac{9}{2}} e^{-\frac{1}{2}x} dx$$

1.4.3 Seja $Z \sim N(0;1)$ e $X \sim Gama\left(\frac{1}{2}, 3\right)$. Apresente uma transformação de Z que seja identicamente distribuída a X.

1.4.4 Mostre que para qualquer $\alpha \in \mathbb{N}^\star$ tem-se $\lim\limits_{n\to\infty} \dfrac{n^\alpha \Gamma(n-\alpha)}{\Gamma(n)} = 1$.

1.5 A Distribuição Gama

Agora que já revisamos a função gama estamos prontos para a definição da distribuição gama.

Definição 1.6. *Uma variável aleatória contínua X tem distribuição Gama, de parâmetros $\alpha > 0$ e $\lambda > 0$, se a sua função densidade for definida por:*

$$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \qquad x > 0.$$

Nesse caso dizemos que $X \sim Gama(\alpha, \lambda)$.

Veja que de fato f_X definida acima é função densidade, uma vez que $f_X(x) \geq 0 \ \forall\ x \in \mathbb{R}$ e $\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1$. Para concluir esse último resultado basta usar a função Gama Modificada, apresentada na Proposição 1.12, para resolver a integral. Veja como.

$$\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = \int_0^\infty \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}\mathrm{d}x = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-\lambda x}\mathrm{d}x = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha)}{\lambda^\alpha} = 1.$$

Proposição 1.16. *Seja $X \sim Gama(\alpha, \lambda)$, então* $\mathrm{E}(X) = \frac{\alpha}{\lambda}$ *e* $\mathrm{Var}(X) = \frac{\alpha}{\lambda^2}$.

Demonstração. Vamos primeiro encontrar a expressão para o k-ésimo momento da distribuição Gama. Novamente faremos o uso da função Gama Modificada.

$$\mathrm{E}\left(X^k\right) = \int_0^\infty x^k \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}\mathrm{d}x = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha+k-1} e^{-\lambda x}\mathrm{d}x = \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+k)}{\lambda^{\alpha+k}}.$$

Dessa forma concluímos que:

$$\begin{aligned}
\mathrm{E}(X) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} = \frac{1}{\Gamma(\alpha)} \frac{\alpha\Gamma(\alpha)}{\lambda} = \frac{\alpha}{\lambda}, \\
\mathrm{E}(X^2) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+2)}{\lambda^{\alpha+2}} = \frac{1}{\Gamma(\alpha)} \frac{(\alpha+1)\alpha\Gamma(\alpha)}{\lambda^2} = \frac{(\alpha+1)\alpha}{\lambda^2}, \\
\mathrm{Var}(X) &= \mathrm{E}(X^2) - \mathrm{E}(X)^2 = \frac{(\alpha+1)\alpha}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}.
\end{aligned}$$

□

Podemos também encontrar a função geradora de momentos de uma variável aleatória com distribuição Gama.

Proposição 1.17. *Seja $X \sim Gama(\alpha, \lambda)$, então* $M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha \;,\quad t < \lambda.$

Demonstração.

$$M_X(t) = \mathrm{E}\left(e^{tX}\right) = \int_0^\infty e^{tx}\frac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}\mathrm{d}x = \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{\alpha-1}e^{-(\lambda-t)x}\mathrm{d}x = \frac{\lambda^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha)}{(\lambda-t)^\alpha},$$

sendo que a última igualdade só é verdadeira se $\lambda - t > 0$, ou seja, se $t < \lambda$. Com isso chegamos ao resultado que queremos mostrar,

$$M_X(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha)}{(\lambda-t)^\alpha} = \left(\frac{\lambda}{\lambda - t}\right)^\alpha \;, \qquad t < \lambda.$$

□

Vejamos agora alguns resultados importantes sobre a distribuição gama.

Proposição 1.18. *Seja $X \sim Gama(\alpha, \lambda)$, $c \in \mathbb{R}^+$ e $Y = cX$. Então $Y \sim Gama(\alpha, \lambda/c)$.*

Demonstração. Faremos a demonstração pela função geradora de momentos, mas poderia ser pelo método da função de distribuição também.

$$\begin{aligned} M_Y(t) = \mathrm{E}\left(e^{tY}\right) = \mathrm{E}\left(e^{tcX}\right) = M_X(tc) &= \left(\frac{\lambda}{\lambda - tc}\right)^\alpha \;,\quad tc < \lambda \\ &= \left(\frac{\lambda/c}{\lambda/c - t}\right)^\alpha \;,\quad t < \lambda/c. \end{aligned}$$

Chegamos na função geradora de momentos de uma variável aleatória $Gama(\alpha, \lambda/c)$. Pela unicidade da função geradora de momentos, concluímos que esta é a distribuição de Y. □

Proposição 1.19. *Sejam $X \sim Gama(\alpha_X, \lambda)$ e $Y \sim Gama(\alpha_Y, \lambda)$ variáveis aleatórias independentes. Defina $W = X + Y$. Então $W \sim Gama(\alpha_X + \alpha_Y, \lambda)$.*

Demonstração. Novamente faremos uso da função geradora de momentos para a demonstração. Lembre-se que a função geradora de momentos da soma de variáveis aleatórias independentes é o produto das funções geradoras de momentos de cada uma das variáveis aleatórias.

$$\begin{aligned} M_W(t) = M_X(t)M_Y(t) &= \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_X}\left(\frac{\lambda}{\lambda - t}\right)^{\alpha_Y} \;,\quad t < \lambda \\ &= \left(\frac{\lambda}{\lambda - t}\right)^{\alpha_X + \alpha_Y} \;,\quad t < \lambda. \end{aligned}$$

Veja que esta é a função geradora de momentos de uma variável aleatória com distribuição $Gama(\alpha_X + \alpha_Y, \lambda)$. Logo, pela unicidade da função geradora de momentos, esta é a distribuição de W, como queríamos mostrar. □

Esse resultado pode ser generalizado para o caso de n variáveis aleatórias com distribuição gama, todas independentes e com mesmo parâmetro λ. Veja a Proposição 1.20 a seguir.

Proposição 1.20. *Sejam $X_i \sim Gama(\alpha_i, \lambda)$ variáveis aleatórias independentes. Defina $W_n = \sum_{i=1}^n X_i$. Então $W_n \sim Gama(\sum_{i=1}^n \alpha_i, \lambda)$.*

Demonstração. A prova será por indução.

Primeiro veja que o resultado é verdadeiro para $n = 2$, esse é justamento uma aplicação do resultado da Proposição 1.19. Com isso mostramos o passo básico.

Para demostrar o passo indutivo vamos supor que $W_{n-1} = \sum_{i=1}^{n-1} X_i \sim Gama(\sum_{i=1}^{n-1} \alpha_i, \lambda)$ e concluir que $W_n = \sum_{i=1}^{n} X_i \sim Gama(\sum_{i=1}^{n} \alpha_i, \lambda)$.

$$W_n = \sum_{i=1}^{n} X_i = \sum_{i=1}^{n-1} X_i + X_n = W_{n-1} + X_n$$

e pela hipótese de indução sabemos que $W_{n-1} \sim Gama(\sum_{i=1}^{n-1} \alpha_i, \lambda)$. Além disso, $X_n \sim Gama(\alpha_n, \lambda)$. Veja também que W_{n-1} e X_n são variáveis aleatórias independentes, uma vez que W_{n-1} é função somente de variáveis aleatórias independentes à X_n. Então, usando o resultado da Proposição 1.19, concluímos que

$$W_n = W_{n-1} + X_n \sim Gama\left(\sum_{i=1}^{n-1} \alpha_i + \alpha_n, \lambda\right) = Gama\left(\sum_{i=1}^{n} \alpha_i, \lambda\right).$$

□

O último resultado dessa seção mostra a relação entre as distribuições gama e normal, definindo assim uma nova distribuição: a distribuição qui-quadrado.

Proposição 1.21. *Sejam Z_1, Z_2, ..., Z_n variáveis aleatórias independentes tais que cada $Z_i \sim N(0, 1)$. Defina $X = \sum_{i=1}^{n} Z_i^2$. Então $X \sim Gama(n/2, 1/2)$.*

Demonstração. Essa demonstração será dividida em duas partes. Primeiro vamos mostrar que $Z^2 \sim Gama\,(1/2, 1/2)$, sempre que $Z \sim N(0, 1)$. Para isso usaremos o método da função de distribuição acumulada.

$$\begin{aligned}
F_{Z^2}(x) = \mathrm{P}(Z^2 \leq x) &= \begin{cases} 0 & , \ x < 0 \\ \mathrm{P}(-\sqrt{x} \leq Z \leq \sqrt{x}) & , \ x \geq 0 \end{cases} \\
&= \begin{cases} 0 & , \ x < 0 \\ \mathrm{P}(Z \leq \sqrt{x}) - \mathrm{P}(Z \leq -\sqrt{x}) & , \ x \geq 0 \end{cases} \\
&= \begin{cases} 0 & , \ x < 0 \\ F_Z(\sqrt{x}) - F_Z(-\sqrt{x}) & , \ x \geq 0 \end{cases}
\end{aligned}$$

Para encontrar a densidade de Z^2 basta derivar da sua função de distribuição.

$$\begin{aligned}
f_{Z^2}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{Z^2}(x) &= \begin{cases} 0 & , \ x < 0 \\ \dfrac{\mathrm{d}}{\mathrm{d}x} F_Z(\sqrt{x}) - \dfrac{\mathrm{d}}{\mathrm{d}x} F_Z(-\sqrt{x}) & , \ x \geq 0 \end{cases} \\
&= \begin{cases} 0 & , \ x < 0 \\ f_Z(\sqrt{x}) \dfrac{1}{2\sqrt{x}} + f_Z(-\sqrt{x}) \dfrac{1}{2\sqrt{x}} & , \ x \geq 0 \end{cases} \\
&= \begin{cases} 0 & , \ x < 0 \\ f_Z(\sqrt{x}) \dfrac{1}{\sqrt{x}} & , \ x \geq 0 \end{cases} \\
&= \begin{cases} 0 & , \ x < 0 \\ \dfrac{1}{\sqrt{2\pi}\sqrt{x}} e^{-\sqrt{x}^2/2} & , \ x \geq 0 \end{cases}
\end{aligned}$$

Assim chegamos em,

$$f_{Z^2}(x) = \frac{(1/2)^{1/2}}{\sqrt{\pi}} x^{-1/2} e^{-x/2} \ , \ x \geq 0$$

que é justamente a função densidade de uma variável aleatória com distribuição $Gama(1/2, 1/2)$ (lembre-se que $\Gamma(1/2) = \sqrt{\pi}$, Proposição 1.15). Com isso concluímos que $Z^2 \sim Gama(1/2, 1/2)$.

A segunda parte da demonstração é responsável por encontrar a distribuição de $\sum_{i=1}^{n} Z_i^2$. Para isso veja que cada $Z_i^2 \sim Gama(1/2, 1/2)$. Além disso, as variáveis $Z_i^2, \ldots, Z_n^2$ são independentes. Usando o resultado da Proposição 1.20, concluímos que

$$X = \sum_{i=1}^{n} Z_i^2 \sim Gama\left(\sum_{i=1}^{n} 1/2, 1/2\right) = Gama(n/2, 1/2).$$

□

Definição 1.7. *Seja* $X \sim Gama(n/2, 1/2)$*. Então dizemos que* X *tem distribuição qui-quadrado com* n *graus de liberdade e usamos a notação* $X \sim \chi_n^2$*.*

Assim como acontece com a distribuição normal padrão, os valores da distribuição qui-quadrado foram tabelados. Mas diferente da normal padrão, a tabela da qui-quadrada apresenta os quantis para algumas probabilidades pré-estabelecidas. A Tabela A.2 do Apêndice A apresenta tais valores para n variando de 1 até 30.

Exemplo 1.2. *Encontre o valor de* x *tal que* $\mathrm{P}(X \leq x) = 0,25$*, onde* $X \sim \chi_8^2$*. Use o Programa R ou a tabela da distribuição qui-quadrado (Tabela A.2 do Apêndice A).*

Solução:

A solução será feita pela tabela. Procure na tabela a coluna referente a probabilidade $0,25$ e a linha referente ao número de graus de liberdade, no caso do exemplo, $n = 8$. O valor nesta linha e coluna é o que queremos encontrar. No caso deste exemplo encontramos $x = 5,0706$, ou seja,

$$\mathrm{P}(X \leq 5,0706) = 0,25.$$

||

Exercícios da Seção 1.5

1.5.1 Seja $W \sim Gama(\alpha, \lambda)$. Encontre $\mathrm{E}\left(\frac{1}{W}\right)$ e $\mathrm{Var}\left(\frac{1}{W}\right)$ em função dos parâmetros α e λ.

1.5.2 Seja $W \sim Gama(3, 2)$. Calcule $\mathrm{P}(1 \leq W \leq 2)$ a partir da tabela da distribuição qui-quadrado.

1.5.3 Suponha que o tempo (em horas) até a ocorrência de um problema na n-ésima bomba de combustível de um certo tipo de aeronave seja uma variável aleatória contínua com distribuição Gama($\alpha = n, \lambda = 1/100$). Se ocorre problema em uma bomba, esta é desligada e outra bomba é automaticamente acionada.

Em uma dessas aeronaves foram instaladas duas bombas de combustível. Quando ocorre um problema na primeira bomba, a segunda bomba é automaticamente acionada. Se ocorrer um problema na segunda bomba durante um voo não há mais bombas para serem acionadas e por isso é necessário realizar um pouso de emergência.

Considerando que essa aeronave irá realizar um voo com duração de 50 horas, responda:

(a) Qual a probabilidade do voo terminar sem que a segunda bomba tenha sido acionada?

(b) Qual a probabilidade de ser necessário realizar um pouso de emergência devido a problemas com a bomba de combustível?

(c) Qual o tempo médio de voo dessa aeronave (com duas bombas) desde a sua decolagem até a realização de um pouso de emergência?

(d) Quantas bombas deveriam ter a aeronave para que a probabilidade de realizar um pouso de emergência devido a problemas com a bomba de combustível seja menor que 1%?

1.6 Mais Alguns Exercícios do Capítulo 1

1.1. Vimos que se $X \sim Gama\left(\frac{1}{2}, \frac{1}{2}\right)$ então $X \stackrel{d}{=} Z^2$. Usando essa relação e as tabelas da distribuição normal padrão, calcule $\mathrm{P}(X < 1)$ a partir de cada uma das duas tabelas.

1.2. Seja $X \sim Gama\left(\frac{1}{2}, 3\right)$. A partir da tabela normal, encontre $\mathrm{P}\left(\frac{1}{2} < X < 1\right)$.

1.3. Seja $Z \sim N(0;1)$ e $X \sim Gama\left(\frac{1}{2}, 3\right)$. Apresente uma transformação de Z que seja identicamente distribuída a X.

1.4. Cada item a seguir apresenta uma função de densidade. Identifique quais são funções de densidade de uma variável aleatória com distribuição gama ou normal, e no caso de ser uma dessas duas, indique os parâmetros da distribuição. Se quiser relembrar ainda mais os conceitos de probabilidade, identifique todas as distribuições.

(a) $f_X(x) = 1/2, \ -1 < x < 1$

(b) $f_X(x) = \dfrac{1}{3\sqrt{2\pi}} e^{-(x-1)^2/18}, x \in \mathbb{R}$

(c) $f_X(x) = \dfrac{e^{-x/3}}{3}, \ x > 0$

(d) $f_X(x) = 4x^2 e^{-2x}, \ x > 0$

(e) $f_X(x) = \dfrac{4(1-x)}{9\sqrt[3]{x^2}}, \ 0 < x < 1$

(f) $f_X(x) = \dfrac{2e^{2x^2}}{\sqrt{2\pi}}, x \in \mathbb{R}$

(g) $f_X(x) = \dfrac{32}{x^3}, \ x \geq 4$

(h) $f_X(x) = \dfrac{e^{-x}}{\sqrt{\pi x}}, \ x > 0$

1.5. Cada item a seguir apresenta uma função geradora de momentos. Identifique quais são funções geradoras de momentos de uma variável aleatória com distribuição gama ou normal, e no caso de ser uma dessas duas, indique os parâmetros da distribuição. Se quiser relembrar ainda mais os conceitos de probabilidade, identifique todas as distribuições.

(a) $M_X(t) = \sqrt{\dfrac{3}{1-3}}, \ t < 3$

(b) $M_X(t) = \left(\dfrac{1+e^t}{2}\right)^4$

(c) $M_X(t) = e^{2t} e^{4t^2}$

(d) $M_X(t) = e^{e^t - 1}$

(e) $M_X(t) = \dfrac{1}{1-t}, \ t < 1$

(f) $M_X(t) = \dfrac{1}{3 - 2e^t}$

(g) $M_X(t) = e^{t^2}$

(h) $M_X(t) = \left(\dfrac{1}{1-2t}\right)^5, \ t < 1/2$

Capítulo 2

Distribuições Amostrais

Nesse capítulo serão definidos novos termos usados no curso de Inferência, como por exemplo: amostra aleatória, estatística e distribuição amostral. Também serão apresentadas algumas distribuições novas. Para terminar, veremos resultados importantes para o caso da população normal.

2.1 Conceitos Gerais

Definição 2.1. *Dizemos que $X_1, X_2, \ldots, X_n$ é uma amostra aleatória de tamanho n da população X se*

(i) $X_1, X_2, \ldots, X_n$ são variáveis aleatórias independentes;

(ii) $X_i \stackrel{d}{=} X$ para todo $i = 1, 2, \ldots, n$.

É importante destacar que esse conceito é diferente do conceito de amostra aleatório apresentado em um curso de amostragem, por exemplo. Em Inferência Estatística toda vez que um vetor aleatório ou um conjunto de variáveis aleatórias formam uma amostra aleatória entende-se que essas variáveis aleatórias são independentes e são identicamente distribuídas.

Também é importante destacar que em Inferência Estatística o termo população é usado para a distribuição de uma amostra aleatória, o que pode confundir novamente com o mesmo termo usando em amostragem. Mesmas palavras com diferentes significados para diferentes contextos.

Definição 2.2. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória da população X, cuja distribuição depende dos parâmetros desconhecidos $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)^T$. Seja $T(X_1, X_2, \ldots, X_n)$, ou simplesmente $T(\mathbf{X})$, ou até mesmo T, qualquer transformação da amostra aleatória que não depende explicitamente dos parâmetros $\boldsymbol{\theta}$. Então a variável aleatória T é chamada de estatística. A distribuição de probabilidade de uma estatística T é chamada de distribuição amostral.*

Exemplo 2.1. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória da população X. Apresente transformações da amostra que são exemplos de estatísticas e outras que não são.*

Solução:

São exemplos de estatísticas as seguintes transformações da amostra aleatória:

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n} \ , \ S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \ , \ \hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \ ,$$
$$X_{(1)} = \min_{i=1}^n X_i \ , \ X_{(n)} = \max_{i=1}^n X_i \ .$$

Não são exemplos de estatísticas as seguintes transformações da amostra aleatória:

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \quad , \quad \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}.$$

$\|$

Resumindo, em Inferência uma estatística é qualquer variável aleatória definida como uma transformação de uma amostra aleatória que não depende de parâmetros desconhecidos da população.

Veremos agora alguns resultados de estatísticas e distribuições amostrais para as populações normal e gama.

Proposição 2.1. *Seja $X_1, X_2, \ldots, X_n$ amostra aleatória da população $X \sim N(\mu, \sigma^2)$ e $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$. Então, $\bar{X}$ é estatística e $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.*

Demonstração. Como $\bar{X}$ não depende nem de μ e nem de σ^2 fica claro que $\bar{X}$ é uma estatística.

Vamos agora encontrar a distribuição amostral de $\bar{X}$. Pela Proposição 1.6 sabemos que soma de normais independentes tem distribuição normal, aplicando essa proposição chegamos em

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2).$$

A Proposição 1.5 mostra que uma função linear de uma variável aleatória normal também é normal. Veja que $\bar{X}$ é uma função linear de $\sum_{i=1}^{n} X_i$. Então,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\frac{1}{n}n\mu, \frac{1}{n^2}n\sigma^2\right) \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

$\square$

Proposição 2.2. *Seja $X_1, X_2, \ldots, X_n$ amostra aleatória da população $X \sim Gama(\alpha, \lambda)$ e $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$. Então, $T(\mathbf{X})$ é estatística e a sua distribuição amostral é $T(\mathbf{X}) \sim Gama(n\alpha, \lambda)$.*

Demonstração. Podemos concluir que $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ é estatística uma vez que é uma transformação da amostra que não depende dos parâmetros desconhecidos α e λ.

Vamos agora encontrar a distribuição amostral de $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$. Essa demonstração será simplesmente uma aplicação da Proposição 1.20, a qual nos diz que a soma de gamas independentes e de mesmo parâmetros λ, isto é, $Gama(\alpha_i, \lambda)$, é uma variável aleatória $Gama(\sum_{i=1}^{n} \alpha_i, \lambda)$.

Veja que $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ é soma de variáveis aleatórias com distribuição gama, de mesmo parâmetro λ e independentes. Veja que como as variáveis formam uma amostra aleatória elas tem inclusive o mesmo parâmetro α. Então, aplicando a proposição citada, concluímos que

$$T(\mathbf{X}) = \sum_{i=1}^{n} X_i \sim Gama\left(\sum_{i=1}^{n} \alpha_i, \lambda\right) \sim Gama\left(\sum_{i=1}^{n} \alpha, \lambda\right) \sim Gama\left(n\alpha, \lambda\right).$$

$\square$

Corolário 2.1. *Seja $X_1, X_2, \ldots, X_n$ amostra aleatória da população $X \sim Exp(\lambda)$ e $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$. Então, $T(\mathbf{X})$ é uma estatística e $T(\mathbf{X}) \sim Gama(n, \lambda)$.*

Demonstração. Podemos concluir que $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ é estatística uma vez que é uma transformação da amostra que não depende do parâmetro desconhecido λ.

Vamos agora encontrar a distribuição amostral de $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$. Veja que esse resultado é consequência imediata da Proposição 2.2, uma vez que a distribuição exponencial é um caso particular da distribuição gama: se $X \sim Exp(\lambda) \Rightarrow X \sim Gama(1, \lambda)$. Aplicando a Proposição 2.2 para uma amostra de $X \sim Gama(1, \lambda)$ concluímos que $T(\mathbf{X}) \sim Gama(n, \lambda)$. $\square$

Corolário 2.2. *Seja $X_1, X_2, \ldots, X_n$ amostra aleatória da população $X \sim \chi^2_m$ e $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$. Então, $T(\mathbf{X})$ é estatística e $T(\mathbf{X}) \sim \chi^2_{nm}$.*

Demonstração. Podemos concluir que $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ é estatística uma vez que é uma transformação da amostra que não depende do parâmetro desconhecido m.

Vamos agora encontrar a distribuição amostral de $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$. Veja que se $X \sim \chi^2_m \Rightarrow X \sim Gama\,(m/2, 1/2)$. Aplicando novamente a Proposição 2.2 para uma amostra de $X \sim Gama(m/2, 1/2)$ concluímos que

$$T(\mathbf{X}) \sim Gama\left(\sum_{i=1}^{n} \frac{m}{2}, \frac{1}{2}\right) \Rightarrow T(\mathbf{X}) \sim Gama\left(\frac{nm}{2}, \frac{1}{2}\right) \Rightarrow T(\mathbf{X}) \sim \chi^2_{nm}.$$

$\square$

Proposição 2.3. *Seja $X_1, X_2, \ldots, X_n$ amostra aleatória da população $X \sim Exp(\lambda)$ e $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$. Então, $\bar{X}$ é estatística e $\bar{X} \sim Gama(n, n\lambda)$.*

Demonstração. Podemos concluir que $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ é estatística uma vez que é uma transformação da amostra que não depende do parâmetro desconhecido λ.

Vamos agora encontrar a distribuição amostral de $\bar{X}$. Pelo Corolário 2.1 já sabemos que $T(\mathbf{X}) = \sum_{i=1}^{n} X_i \sim Gama(n, \lambda)$. Veja que $\bar{X} = \frac{1}{n} T(\mathbf{X})$. Aplicando então a Proposição 1.18 podemos concluir que $\bar{X} \sim Gama(n, n\lambda)$, como queríamos demostrar. $\square$

Como veremos ao longo do curso, é bem comum precisarmos da distribuição amostral das estatísticas de ordem, principalmente $X_{(1)}$ ou $X_{(n)}$. Isso acontece quando $Im(X)$ depende do parâmetro da distribuição, como nos exemplos que seguem. Mas antes dos exemplos, vamos relembrar as distribuições das estatísticas de ordem $X_{(1)}$ ou $X_{(n)}$.

Proposição 2.4. *Seja X_1, X_2, $\ldots$, X_n amostra aleatória de uma população X com função de distribuição F_X e função densidade f_X. Defina*

$$X_{(n)} = \max\{X_i\}_{i=1}^{n}.$$

Então a variável aleatória $X_{(n)}$ tem função de distribuição e função densidade, em termos de F_X e f_X, dadas por:

$$F_{X_{(n)}}(x) = (F_X(x))^n \qquad e \qquad f_{X_{(n)}}(x) = n(F_X(x))^{n-1} f_X(x).$$

Demonstração. Faremos a demonstração pelo método da função de distribuição.

$$F_{X_{(n)}}(x) = \mathrm{P}(X_{(n)} \le x) = \mathrm{P}(X_1 \le x \cap X_2 \le x \cap \ldots \cap X_n \le x).$$

Como as variáveis aleatórias X_i, $i = 1, \ldots, n$, são independentes os eventos $\{X_i \le x\}$, $i = 1, \ldots, n$, também são independentes. Assim a probabilidade da interseção desses eventos pode ser escrita como o produto das probabilidade.

$$F_{X_{(n)}}(x) = \prod_{i=1}^{n} \mathrm{P}(X_i \le x) = \prod_{i=1}^{n} F_{X_i}(x) = \prod_{i=1}^{n} F_X(x) = (F_X(x))^n.$$

Chegamos então na expressão,

$$f_{X_{(n)}}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{X_{(n)}}(x) = n(F_X(x))^{n-1} f_X(x).$$

□

Proposição 2.5. *Seja X_1, X_2, …, X_n amostra aleatória de uma população X com função de distribuição F_X e função densidade f_X. Defina*

$$X_{(1)} = \min \{X_i\}_{i=1}^n .$$

Então a variável aleatória $X_{(1)}$ tem função de distribuição e função densidade, em termos de F_X e f_X, dadas por:

$$F_{X_{(1)}}(x) = 1 - (1 - F_X(x))^n \qquad e \qquad f_{X_{(1)}}(x) = n\,(1 - F_X(x))^{n-1} f_X(x).$$

Demonstração. Faremos a demonstração pelo método da função de distribuição.

$$F_{X_{(1)}}(x) = \mathrm{P}(X_{(1)} \le x) = 1 - \mathrm{P}(X_{(1)} > x) = 1 - \mathrm{P}(X_1 > x \cap X_2 > x \cap \ldots \cap X_n > x).$$

Como as variáveis aleatórias X_i, $i = 1, \ldots, n$, são independentes os eventos $\{X_i \le x\}$, $i = 1, \ldots, n$, também são independentes. Assim a probabilidade da interseção desses eventos pode ser escrita como o produto das probabilidade.

$$\begin{aligned} F_{X_{(1)}}(x) &= 1 - \prod_{i=1}^n \mathrm{P}(X_i > x) = 1 - \prod_{i=1}^n (1 - \mathrm{P}(X_i \le x)) = 1 - \prod_{i=1}^n (1 - F_{X_i}(x)) \\ &= 1 - \prod_{i=1}^n (1 - F_X(x)) = 1 - (1 - F_X(x))^n . \end{aligned}$$

Chegamos então na expressão,

$$f_{X_{(1)}}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{X_{(1)}}(x) = n\,(1 - F_X(x))^{n-1} f_X(x).$$

□

Veja que as estatísticas de ordem são transformações da amostra que não dependem dos parâmetros da população. Dessa forma, as estatísticas de ordem são sempre estatísticas. Seguimos com dois exemplos.

Exemplo 2.2. *Seja $X_1, X_2, \ldots, X_n$ amostra aleatória da população $X \sim U[0, \theta]$. Encontre a distribuição amostral de $X_{(n)}$.*

Solução:

Vamos aplicar o resultado da Proposição 2.4 para o caso da distribuição uniforme. Se $X \sim U[0, \theta]$ temos

$$f_X(x) = \frac{1}{\theta},\ 0 \le x \le \theta \qquad \text{e} \qquad F_X(x) = \begin{cases} 0 & , \ x < 0 \\ \dfrac{x}{\theta} & , \ 0 \le x \le \theta \\ 1 & , \ x > 1 \end{cases}$$

Então,

$$\begin{aligned} f_{X_{(n)}}(x) = n(F_X(x))^{n-1} f_X(x) &= n \left(\frac{x}{\theta}\right)^{n-1} \frac{1}{\theta},\ 0 \le x \le \theta \\ &= \frac{n}{\theta^n} x^{n-1},\ 0 \le x \le \theta. \end{aligned}$$

‖

Exemplo 2.3. *Seja* $X_1, X_2, \dots, X_n$ *amostra aleatória da população* X *cuja função densidade é definida por*

$$f(x) = \lambda e^{-\lambda(x-\theta)}, \qquad x \geq \theta, \ \lambda > 0 \ e \ \theta \in \mathbb{R}.$$

Encontre a distribuição amostral de $X_{(1)}$.

Solução:

Para aplicar o resultado da Proposição 2.5 precisamos primeiro conhecer F_X. Se $x < \theta$, $F_X(x) = 0$. Se $x \geq \theta$,

$$F_X(x) = \int_\theta^x \lambda e^{-\lambda(t-\theta)}\mathrm{d}t = \int_0^{x-\theta} \lambda e^{-\lambda u}\mathrm{d}u = -e^{-\lambda u}\Big|_0^{x-\theta} = \left(1 - e^{-\lambda(x-\theta)}\right).$$

Agora podemos aplicar o resultado da Proposição 2.5 e concluir que

$$\begin{aligned} f_{X_{(1)}}(x) &= n\left(1 - \left(1 - e^{-\lambda(x-\theta)}\right)\right)^{n-1} \lambda e^{-\lambda(x-\theta)}, x \geq \theta \\ &= n\left(e^{-\lambda(x-\theta)}\right)^{n-1} \lambda e^{-\lambda(x-\theta)}, x \geq \theta \\ &= n\lambda e^{-n\lambda(x-\theta)}, x \geq \theta \end{aligned}$$

||

Exercícios da Seção 2.1

2.1.1 Seja $X_1, X_2, \dots, X_n$ amostra aleatória da população $X \sim Gama(3, \lambda)$. Diga se $T(\mathbf{X}) = \bar{X}$ é estatística e, independente da resposta, encontre a distribuição de T.

2.1.2 Seja $X_1, X_2, \dots, X_n$ amostra aleatória da população $X \sim Exp(\lambda)$. Diga se $T(\mathbf{X}) = 2n\lambda\bar{X}$ é estatística e, independente da resposta, encontre a distribuição de T.

2.1.3 ([Larson, 1982] - Capítulo 6) Seja (X_1, X_2) uma amostra aleatória de tamanho 2 de uma população normal de média 0 e variância σ^2. Calcule:

(a) $\mathrm{P}(X_1^2 + X_2^2 \leq \sigma^2)$

(b) $\mathrm{P}(X_1^2 + X_2^2 \leq 2\sigma^2)$

Dica: Qual a distribuição de $(X_1^2 + X_2^2)/\sigma^2$?

2.1.4 Considere uma amostra aleatória (X_1, X_2, X_3, X_4) de uma população $X \sim N(0, \sigma^2)$. Calcule $\mathrm{P}(\sum_{j=1}^4 X_j^2 \geq 1,92\sigma^2)$. Use uma tabela ou o computador para achar a resposta numérica.

2.1.5 Seja uma amostra aleatória $X_1, X_2, \dots, X_n$ de uma população $X \sim N(\mu_0, \sigma^2)$, onde μ_0 é uma constante conhecida. Defina $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu_0)^2$.

(a) $\hat{\sigma}^2$ é uma estatística? Justifique.

(b) Mostre que $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_n$.

(c) Encontre a distribuição amostral de $\hat{\sigma}^2$.

2.1.6 Sejam duas amostras aleatórias independentes $(X_1, \dots, X_n)$ e $(Y_1, \dots, Y_m)$ de populações $X \sim Poisson(\lambda_x)$ e $Y \sim Poisson(\lambda_y)$, respectivamente. Mostre que

$$n\bar{X} + m\bar{Y} \sim Poisson(n\lambda_x + m\lambda_y).$$

Dica: Primeiro mostre que $n\bar{X}$ e $m\bar{Y}$ são *Poisson*. Quais os parâmetros de cada uma dessas v.a.? Depois mostre que soma de v.a. independentes com distribuição de Poisson também é uma Poisson. Aplique esse último resultado para terminar a demonstração.

2.1.7 ([Larson, 1982] - Capítulo 6) Considere uma amostra aleatória $X_1, \ldots, X_n$ de uma população $X \sim Exp(\lambda)$. Encontre a distribuição de:

(a) $\sum_{i=1}^{n} X_i$

(b) $2\lambda n\bar{X}$

2.1.8 Seja X_1, X_2, ..., X_n uma amostra aleatória de $X \sim Exp(\lambda)$.

(a) $X_{(1)}$ é uma estatística?

(b) Encontre a distribuição de $X_{(1)}$.

2.1.9 ([Larson, 1982] - Capítulo 5) Seja $U_1, U_2, \ldots, U_n$ uma amostra aleatória de $U \sim U(0, 1)$.

(a) Encontre a densidade de $U_{(1)}$.

(b) Encontre a densidade de $U_{(n)}$.

(c) Calcule $\mathrm{P}(U_{(1)} \leq \frac{1}{2})$ e $\mathrm{P}(U_{(n)} \leq \frac{1}{2})$.

2.1.10 Seja $X_1, X_2, \ldots, X_n$ amostra aleatória da população X cuja função densidade está definida abaixo. Diga se $T(\mathbf{X}) = X_{(n)}$ é estatística e, independente da resposta, encontre a distribuição de T.

$$f(x) = \frac{2x}{\theta}\ , \qquad 0 \leq x \leq \theta.$$

2.1.11 Seja X_1, ..., X_n amostra aleatória de X com função densidade $f_X(x) = \theta x^{\theta-1}$, $0 < x < 1$ e $\theta > 0$.

(a) Mostre que se $Y = -\ln(X)$ então $Y \sim Exp(\theta)$.

(b) Encontre a distribuição amostral de $-\sum_{i=1}^{n} \ln(X_i)$.

2.2 A Distribuição t-Student

Nessa seção veremos uma nova distribuição de variável aleatória, a distribuição t-Student. Começamos com a definição da sua função densidade e seguimos para alguns resultados que relacionam essa nova distribuição de probabilidade com outros modelos probabilísticos já conhecidos.

Definição 2.3. *Uma variável aleatória contínua X tem distribuição t-Student (ou simplesmente t) com n graus de liberdade, $n \in \mathbb{N}$, se a sua função densidade for definida por*

$$f_X(x) = \frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)} \frac{1}{\sqrt{n\pi}} \left(\frac{x^2}{n} + 1\right)^{-\dfrac{n+1}{2}}.$$

Nesse caso usamos a notação $X \sim t_n$.

A Figura 2.1 apresenta em preto o gráfico da densidade de uma variável aleatória com distribuição Normal Padrão e nas outras cores o gráfico da densidade de variáveis aleatória t-Student para diferentes valores para n: $n = 1$ para a curva em vermelho, $n = 2$ para a curva em laranja, $n = 3$ para a curva em azul e $n = 4$ para a curva em verde. Veremos na Proposição 2.7 que quando $n \to \infty$ a densidade de uma t-Student com n graus de liberdade se aproxima da densidade de uma normal padrão.

Outra característica da curva da função densidade de uma variável aleatória t-Student é a sua simetria em relação ao eixo y. Podemos analisar a expressão da densidade apresentada

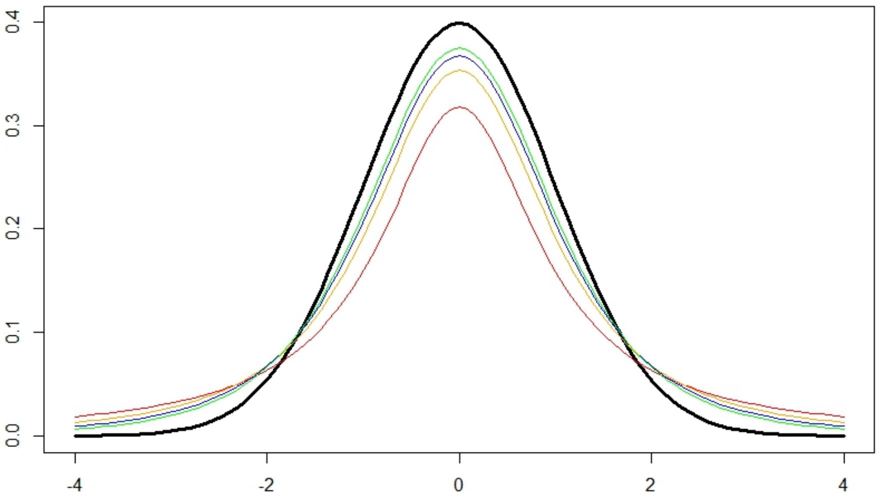


Figura 2.1: Comparação entre a densidade de uma $N(0,1)$ e a densidade de uma t-Student

na Definição 2.3 e perceber que f_X é sempre uma função par, isto é, $f_X(x) = f_X(-x)$, o que comprova tal simetria.

O resultado apresentado no Teorema 2.1 é provavelmente mais importante que a própria definição de uma t-Student pela sua função densidade. Na grande maioria das vezes esse será o resultado usado para identificar uma variável aleatória como seguindo a distribuição t.

Teorema 2.1. *Sejam $Z \sim N(0,1)$ e $W \sim \chi^2_n$ variáveis aleatórias independentes. Então a variável aleatória*

$$T = \frac{Z}{\sqrt{\frac{W}{n}}} \sim t_n,$$

isto é, tem distribuição t com n graus de liberdade.

Demonstração. A demonstração desse teorema se resume a um exercício de Método Jacobiano, isto é, definida uma variável aleatória $T = \frac{Z}{\sqrt{\frac{W}{n}}}$, com função de $Z \sim N(0,1)$ e $W \sim \chi^2_n \sim$ *Gama* $(n/2, 1/2)$ vamos usar o Método Jacobiano para encontrar a densidade marginal de T e comparar com a função densidade apresentada na Definição 2.3.

Como Z e W são independentes, por hipótese, temos que a função densidade conjunta delas é dada pelo produto das marginais. Assim,

$$f_{Z,W}(z,w) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \left(\frac{1}{2}\right)^{\frac{n}{2}} \frac{1}{\Gamma(n/2)} w^{n/2-1} e^{-w/2}\ , \quad w > 0.$$

Considere as seguintes transformações de Z e W:

$$T = \frac{Z}{\sqrt{\frac{W}{n}}} = g_1(Z,W) \qquad \text{e} \qquad V = \sqrt{\frac{W}{n}} = g_2(Z,W).$$

Veja que $Z = TV = h_1(T,V)$ e $W = nV^2 = h_2(T,V)$. Então, segundo o Método Jacobiano, a densidade conjunta entre T e V é dada por

$$f_{T,V}(t,v) = f_{Z,W}(h_1(t,v),h_2(t,v))|J(t,v)|, \quad \text{onde} \quad J(t,v) = \det\begin{pmatrix} \frac{\mathrm{d}}{\mathrm{d}t}h_1(t,v) & \frac{\mathrm{d}}{\mathrm{d}v}h_1(t,v) \\ \frac{\mathrm{d}}{\mathrm{d}t}h_2(t,v) & \frac{\mathrm{d}}{\mathrm{d}v}h_2(t,v) \end{pmatrix}.$$

Fazendo as devidas substituições chegamos em $J(t,v) = \det\begin{pmatrix} v & t \\ 0 & 2nv \end{pmatrix} = 2nv^2$ e

$$\begin{aligned} f_{T,V}(t,v) &= \frac{1}{\sqrt{2\pi}} e^{-(tv)^2/2} \left(\frac{1}{2}\right)^{\frac{n}{2}} \frac{1}{\Gamma(n/2)} (nv^2)^{n/2-1} e^{-(nv^2)/2} 2nv^2 \ , \quad nv^2 > 0 \\ &= \frac{2}{\sqrt{2\pi}} \left(\frac{n}{2}\right)^{\frac{n}{2}} \frac{1}{\Gamma(n/2)} v^n e^{-v^2(t^2+n)/2} \ , \quad v > 0 \end{aligned}$$

Agora vamos encontrar a marginal da variável aleatória T integrando a conjunta $f_{T,V}$.

$$f_T(t) = \int_0^\infty f_{T,V}(t,v)\mathrm{d}v = \frac{2}{\sqrt{2\pi}} \left(\frac{n}{2}\right)^{\frac{n}{2}} \frac{1}{\Gamma(n/2)} \int_0^\infty v^n e^{-v^2(t^2+n)/2}\mathrm{d}v$$

Faremos a substituição $u = v^2$, assim temos $v = \sqrt{u}$ e $\mathrm{d}u = 2v\mathrm{d}v$.

$$\begin{aligned} f_T(t) &= \frac{2}{\sqrt{2\pi}} \left(\frac{n}{2}\right)^{\frac{n}{2}} \frac{1}{\Gamma(n/2)} \int_0^\infty u^{n/2} e^{-u(t^2+n)/2} \frac{1}{2\sqrt{u}}\mathrm{d}u \\ &= \frac{1}{2}\frac{2}{\sqrt{2\pi}} \left(\frac{n}{2}\right)^{\frac{n}{2}} \frac{1}{\Gamma(n/2)} \int_0^\infty u^{(n-1)/2} e^{-u(t^2+n)/2}\mathrm{d}u \\ &= \frac{1}{\sqrt{2\pi}} \left(\frac{n}{2}\right)^{\frac{n}{2}} \frac{1}{\Gamma(n/2)} \Gamma\left(\frac{n+1}{2}\right) \left(\frac{2}{t^2+n}\right)^{\frac{n+1}{2}} \\ &= \frac{1}{\sqrt{2\pi}} \left(\frac{n}{2}\right)^{\frac{n}{2}} \frac{1}{\Gamma(n/2)} \Gamma\left(\frac{n+1}{2}\right) \left(\frac{2}{t^2+n}\right)^{\frac{n+1}{2}} \\ &= \frac{1}{\sqrt{\pi}} n^{\frac{n}{2}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{1}{t^2+n}\right)^{\frac{n+1}{2}} \\ &= \frac{1}{\sqrt{\pi}} n^{\frac{n+1}{2}} n^{-\frac{1}{2}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{1}{t^2+n}\right)^{\frac{n+1}{2}} \\ &= \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{n}{t^2+n}\right)^{\frac{n+1}{2}} \\ &= \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{t^2+n}{n}\right)^{-\frac{n+1}{2}} \\ &= \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{t^2}{n}+1\right)^{-\frac{n+1}{2}} . \end{aligned}$$

Comparando a equação final com a densidade apresentada na Definição 2.3, podemos concluir que $T \sim t_n$. □

Alguns livros definem a distribuição t-Student como aquela que resulta da transformação definida pela razão de uma normal padrão com a raiz de uma qui-quadrado dividida pelos seus graus de liberdades, independentes, como apresenta o Teorema 2.1. Em seguida esses livros deduzem a função densidade a partir do método jacobiano. Um caminho um pouco diferente do feito neste texto, mas absolutamente equivalente.

Todos os resultados deduzidos da distribuição t-Student serão feitos a partir do Teorema 2.1. Vamos começar pelos seus momentos.

Proposição 2.6. *Seja* $T \sim t_n$. $\mathrm{E}(T^k) < \infty$ *se e somente se* $k < n$.

Demonstração. Como $T \sim t_n$ sabemos que $T \overset{d}{=} \dfrac{Z}{\sqrt{W/n}}$ sendo $Z \sim N(0,1)$, $W \sim \chi^2_n$, Z e W independentes. Então,

$$\mathrm{E}(T^k) = \mathrm{E}\left(\left(\frac{Z}{\sqrt{W/n}}\right)^k\right) = \mathrm{E}\left(Z^k \frac{n^{k/2}}{W^{k/2}}\right) = n^{k/2}\,\mathrm{E}\left(Z^k W^{-k/2}\right) = n^{k/2}\,\mathrm{E}\left(Z^k\right)\mathrm{E}\left(W^{-k/2}\right).$$

Veja que $\mathrm{E}\left(Z^k\right) < \infty$ para todo $k \in \mathbb{N}$, uma vez que $Z \sim N(0,1)$ possui todos os momentos, veja pela sua função geradora de momentos.

Vamos então calcular $\mathrm{E}\left(W^{-k/2}\right)$ e definir para que valores de k temos $\mathrm{E}\left(W^{-k/2}\right) < \infty$. Lembre-se que W é uma variável aleatória com distribuição gama. Então a esperança dessa função de W pode ser encontrada a partir da expressão abaixo.

$$\begin{aligned}
\mathrm{E}\left(W^{-k/2}\right) &= \int_0^\infty w^{-k/2}\frac{(1/2)^{n/2}}{\Gamma(n/2)} w^{n/2-1} e^{-w/2}\mathrm{d}w \\
&= \frac{(1/2)^{n/2}}{\Gamma(n/2)}\int_0^\infty w^{(n-k)/2-1}e^{-w/2}\mathrm{d}w \\
&= \frac{(1/2)^{n/2}}{\Gamma(n/2)}\frac{\Gamma((n-k)/2)}{(1/2)^{(n-k)/2}}, \quad \text{se} \quad (n-k)/2 > 0.
\end{aligned}$$

Veja que a condição para a última integral convergir é $(n-k)/2 > 0$, ou seja, $n > k$. Nessa condição podemos dizer que $\mathrm{E}\left(W^{-k/2}\right) < \infty$, caso contrário essa esperança não converge.

Então, $\mathrm{E}(T^k) < \infty$ se e somente $k < n$, como queríamos demonstrar. □

A Proposição 2.6 nos mostra que não existe todos os momentos de uma variável aleatória $T \sim t_n$. Existem apenas os momentos k tais que $k < n$, ou seja, se $T \sim t_2$ existem o primeiro momento mas não existe o segundo. Além disso, como não existem todos os momentos, a distribuição t-Student é um exemplo de distribuição que não possui função geradora de momentos.

Corolário 2.3. *Se* $T \sim t_n$ *e* $n \geq 3$, *então* $\mathrm{E}(T) = 0$ *e* $\mathrm{Var}(T) = \dfrac{n}{n-2}$.

Demonstração. Da demonstração da Proposição 2.6 tiramos que:

$$\mathrm{E}(T^k) = n^{k/2}\,\mathrm{E}\left(Z^k\right)\frac{(1/2)^{n/2}}{\Gamma(n/2)}\frac{\Gamma((n-k)/2)}{(1/2)^{(n-k)/2}}, \quad \text{se} \quad k < n.$$

Então, supondo $n \geq 3$, fazendo $k = 1$ concluímos que $\mathrm{E}(T) = 0$, uma vez que $\mathrm{E}(Z) = 0$.

Para encontrar a variância de T, que coincide com o segundo momento, já que $\mathrm{E}(T) = 0$, vamos substituir na expressão acima k por 2.

$$\mathrm{Var}(T) = n\,\mathrm{E}\left(Z^2\right)\frac{(1/2)^{n/2}}{\Gamma(n/2)}\frac{\Gamma((n-2)/2)}{(1/2)^{(n-2)/2}} = \frac{n}{2}\frac{\Gamma(n/2-1)}{\Gamma(n/2)} = \frac{n}{2}\frac{\Gamma(n/2)}{(n/2-1)\Gamma(n/2)} = \frac{n}{n-2},$$

como queríamos demostrar. □

Proposição 2.7. *Seja* $\{T_n\}_{n=1}^{\infty}$ *uma sequência de variáveis aleatórias independentes tais que* $T_n \sim t_n$. *Então* $T_n \xrightarrow[n\to\infty]{d} Z$ *onde* $Z \sim N(0,1)$.

Demonstração. A demonstração dessa proposição utiliza resultados de convergência de sequências de variáveis aleatórias.

Como cada $T_n \sim t_n$ sabemos que $T_n \stackrel{d}{=} \dfrac{Z_n}{\sqrt{W_n/n}}$ sendo $Z_n \sim N(0,1)$, $W_n \sim \chi^2_n$, Z_n e W_n independentes.

Como $\{Z_n\}_{n=1}^{\infty}$ é uma sequência constante de variáveis aleatórias, podemos afirmar que $Z_n \xrightarrow[n\to\infty]{d} Z \sim N(0,1)$.

Usando a desigualdade de Chebyshev podemos concluir que $\dfrac{W_n}{n} \xrightarrow[n\to\infty]{P} 1$. Lembre-se que $W_n \sim \chi^2_n$, $\mathrm{E}\,(W_n) = n$ e $\mathrm{Var}\,(W_n) = 2n$. Veja como:

$$\mathrm{P}\left(\left|\frac{W_n}{n} - 1\right| \geq \varepsilon\right) = \mathrm{P}\left(|W_n - n| \geq n\varepsilon\right) \underbrace{\leq}_{\text{Chebyshev}} \frac{\mathrm{Var}(W_n)}{n^2\varepsilon^2} = \frac{2n}{n^2\varepsilon^2} \to 0\ ,\ \forall\, \varepsilon > 0.$$

Usando o resultado sobre preservação de convergência por funções contínuas, podemos afirmar que

$$\frac{W_n}{n} \xrightarrow[n\to\infty]{P} 1 \quad \Rightarrow \quad \sqrt{\frac{W_n}{n}} \xrightarrow[n\to\infty]{P} \sqrt{1} = 1.$$

Usando o Teorema de Slutsky podemos afirmar que:

$$\frac{Z_n}{\sqrt{W_n/n}} \xrightarrow[n\to\infty]{d} \frac{Z}{1} = Z \sim N(0,1),$$

ou seja, $T_n \xrightarrow[n\to\infty]{d} Z \sim N(0,1)$. □

Para facilitar as contas de probabilidades envolvendo uma variável aleatória t_n, os principais quantis da distribuição t-Student estão apresentados na Tabela A.4 do Apêndice A. Por exemplo, considerando $T \sim t_8$, se queremos calcular $\mathrm{P}(T \leq 1,4)$ pela tabela podemos verificar que para $n = 8$ temos $\mathrm{P}(T \leq 1,396815) = 0,9$, logo $\mathrm{P}(T \leq 1,4) \approx 0,9$.

Como a densidade da distribuição t-Student é uma função par, só estão apresentados os quantis maiores que 0,5. A partir deles conseguimos determinar alguns quantis menores que 0,5. Por exemplo, supondo novamente $n = 8$, diretamente pela tabela não podemos encontrar t tal que $\mathrm{P}(T \leq t) = 0,25$. Mas sabemos, pelo fato da densidade ser uma função par, que $\mathrm{P}(T \leq t) = \mathrm{P}(T \geq -t) = 1 - \mathrm{P}(T \leq -t)$. Então,

$$\mathrm{P}(T \leq t) = 0,25 \quad \Rightarrow \quad 1 - \mathrm{P}(T \leq -t) = 0,25 \quad \Rightarrow \quad \mathrm{P}(T \leq -t) = 0,75.$$

Aí, em uma consulta diretas da tabela, descobrimos que para $n = 8$ temos $-t = 0,706387 \Rightarrow t = -0,706387$.

Também é possível usar os softwares estatísticos para esse cálculo. No R, por exemplo, o comando `pt(0.7,df=8)` fornece o valor $\mathrm{P}(T \leq 0,7)$, onde $T \sim t_8$ (`df` significa *degrees of freedom*, graus de liberdade em inglês). Para mais detalhes consulte `help(pt)` no R.

Exercícios da Seção 2.2

2.2.1 Seja $X \sim t_{12}$. Usando preferencialmente o Programa R, ou a partir da Tabela A.4, encontre:

(a) $\mathrm{P}(X \leq 1)$.
(b) $\mathrm{P}(X \geq 2,5)$.
(c) $\mathrm{P}(X \leq -0,7)$.
(d) $\mathrm{P}(-1 \leq X \leq 2)$.

Obs: Também é possível usar a Tabela A.4, mas em alguns casos será preciso fazer aproximações. Por isso prefiro que seja usado o Programa R.

2.2.2 Seja $X \sim t_5$. Usando preferencialmente o Programa R, ou a partir da Tabela A.4, encontre:

(a) o valor de x tal que $\mathrm{P}(X \leq x) = 0,95$.
(b) o valor de x tal que $\mathrm{P}(X \leq x) = 0,10$.
(c) o valor de x tal que $\mathrm{P}(-x \leq X \leq x) = 0,80$.
(d) o valor de x tal que $\mathrm{P}(X > x) = 0,88$.

Obs: Também é possível usar a Tabela A.4, mas em alguns casos será preciso fazer aproximações. Por isso prefiro que seja usado o Programa R.

2.2.3 Seja $T \sim t_9$. Calcule $\mathrm{P}(T^2 < 2,50)$. Use preferencialmente o Programa R, ou a Tabela A.4.

2.2.4 ([Larson, 1982] - Capítulo 5) Seja X uma variável aleatória normal padrão e Y uma qui-quadrado com $\nu = 4$ graus de liberdade, sendo X e Y independentes. Qual o valor de a tal que $\mathrm{P}(|X| \leq a\sqrt{Y}) = 0,9$. Use preferencialmente o Programa R, ou a Tabela A.4.

2.2.5 Suponha que $X_1, \ldots, X_{10}$ formem uma amostra aleatória de $X \sim N(0, \sigma^2)$. Determine os valores de c e k tais que

$$\frac{c\,(X_1 + X_2)}{\left(\sum_{i=3}^{10} X_i^2\right)^{\frac{1}{2}}} \sim t_k.$$

2.3 A Distribuição F-Snedecor

Essa é mais uma seção com a apresentação de uma nova distribuição, a distribuição F-Snedecor. Assim como foi feito na seção anterior, começaremos com a definição da sua função densidade e seguimos para alguns resultados que relacionam essa nova distribuição de probabilidade com outros modelos probabilísticos já conhecidos.

Definição 2.4. *Uma variável aleatória contínua X tem distribuição F (ou F-Snedecor) com n e m graus de liberdade, $n, m \in \mathbb{N}$, se a sua função densidade de probabilidade for*

$$f_X(x) = \frac{\Gamma\left(\dfrac{n+m}{2}\right)\left(\frac{n}{m}\right)^{n/2}}{\Gamma\left(\dfrac{n}{2}\right)\Gamma\left(\dfrac{m}{2}\right)} x^{n/2-1}\left(\frac{n}{m}x + 1\right)^{-(n+m)/2}, \qquad x > 0.$$

Nesse caso usamos a notação $X \sim F_{n,m}$.

A Figura 2.2 apresenta algumas curvas de funções densidades da distribuição F para diferentes valores de n e m. Podemos ver que o formato dessa curva varia bastante. Por exemplo,

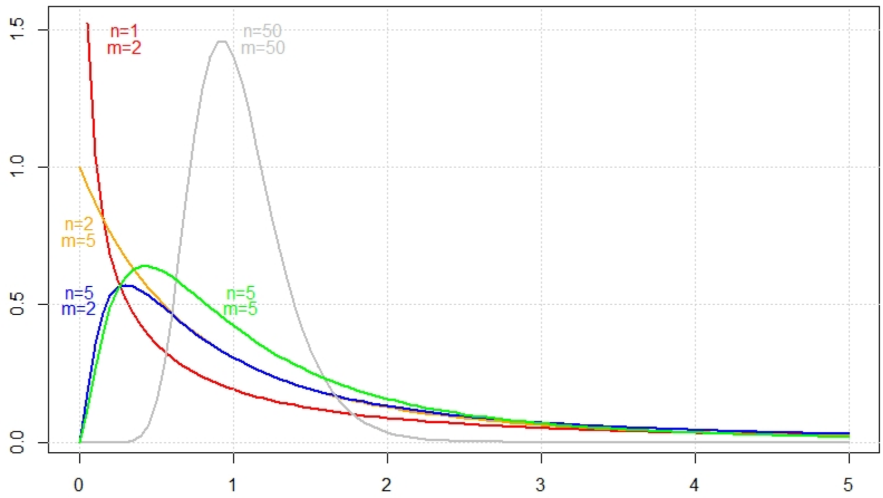


Figura 2.2: Curvas de densidade da distribuição F

analisando a curva em vermelho, com $n = 1$ e $m = 2$, percebemos que $f_X(x) \to \infty$ quando $x \to 0$, o que não ocorre com as demais curvas. Além disso, algumas curvas possuem um ponto de máximo, curvas em azul, verde e cinza, e as outras não.

O Teorema 2.2 a seguir apresenta uma definição alternativa para a distribuição F, como uma transformação de variáveis aleatórias independentes com distribuição qui-quadrado.

Teorema 2.2. *Sejam $W \sim \chi^2_n$ e $V \sim \chi^2_m$ variáveis aleatórias independentes. Então,*

$$F = \frac{W/n}{V/m} \sim F_{n,m},$$

isto é, tem distribuição F com n e m graus de liberdade.

Demonstração. Essa é mais uma demonstração que se resume a um exercícios do Método Jacobiano.

Por hipótese temos $W \sim \chi^2_n$ e $V \sim \chi^2_m$ variáveis aleatórias independentes. Então, a densidade conjunta dessas variáveis é definida pelo produto das densidades marginais.

$$\begin{aligned} f_{W,V}(w,v) &= \frac{(1/2)^{n/2}}{\Gamma(n/2)} w^{(n/2)-1} e^{-w/2} \frac{(1/2)^{m/2}}{\Gamma(m/2)} v^{(m/2)-1} e^{-v/2}, \qquad w > 0,\ v > 0. \\ &= \frac{(1/2)^{(n+m)/2}}{\Gamma(n/2)\Gamma(m/2)} w^{(n/2)-1} v^{(m/2)-1} e^{-(w+v)/2}, \qquad w > 0,\ v > 0. \end{aligned}$$

Considere as seguintes transformações de W e V:

$$U = \frac{W}{V} = g_1(W,V) \qquad \text{e} \qquad Y = V = g_2(W,V).$$

Veja que $W = UY = h_1(U,Y)$ e $V = Y = h_2(U,Y)$. Então, segundo o Método Jacobiano, a densidade conjunta entre U e Y é dada por

$$f_{U,Y}(u,y) = f_{W,V}(h_1(u,y), h_2(u,y))|J(u,y)|, \quad \text{onde} \quad J(u,y) = \det \begin{pmatrix} \frac{\mathrm{d}}{\mathrm{d}u} h_1(u,y) & \frac{\mathrm{d}}{\mathrm{d}y} h_1(u,y) \\ \frac{\mathrm{d}}{\mathrm{d}u} h_2(u,y) & \frac{\mathrm{d}}{\mathrm{d}y} h_2(u,y) \end{pmatrix}.$$

Fazendo as devidas substituições chegamos em $J(u,y) = \det \begin{pmatrix} y & u \\ 0 & 1 \end{pmatrix} = y$ e

$$\begin{aligned} f_{U,Y}(u,y) = & \frac{(1/2)^{(n+m)/2}}{\Gamma(n/2)\Gamma(m/2)}(uy)^{(n/2)-1}y^{(m/2)-1}e^{-(uy+y)/2}y\ , \qquad uy>0,\ y>0 \\ = & \frac{(1/2)^{(n+m)/2}}{\Gamma(n/2)\Gamma(m/2)}u^{(n/2)-1}y^{((m+n)/2)-1}e^{-y(u+1)/2}\ , \qquad u>0,\ y>0 \end{aligned}$$

Agora vamos encontrar a marginal da variável aleatória U integrando a conjunta $f_{U,Y}$.

$$f_U(t) = \int_0^\infty f_{U,Y}(u,y)\mathrm{d}y = \frac{(1/2)^{(n+m)/2}}{\Gamma(n/2)\Gamma(m/2)}u^{(n/2)-1}\underbrace{\int_0^\infty y^{((m+n)/2)-1}e^{-y(u+1)/2}\mathrm{d}y}_{\frac{\Gamma((m+n)/2)}{((u+1)/2)^{(m+n)/2)}}}$$

Chegamos então à expressão

$$f_U(t) = \left(\frac{1}{2}\right)^{(n+m)/2}\frac{\Gamma((m+n)/2)}{\Gamma(n/2)\Gamma(m/2)}u^{(n/2)-1}\left(\frac{2}{u+1}\right)^{(m+n)/2}, \qquad u>0$$

Veja que $F = \dfrac{m}{n}U$. Vamos usar o método da função de distribuição para encontrar a densidade de F.

$$F_F(f) = \mathrm{P}(F \leq f) = \mathrm{P}\left(\frac{m}{n}U \leq f\right) = \mathrm{P}\left(U \leq \frac{n}{m}f\right) = F_U\left(\frac{n}{m}f\right)$$

$$\begin{aligned} f_f(f) &= \frac{\mathrm{d}}{\mathrm{d}f}F_F(f) = \frac{n}{m}f_U\left(\frac{n}{m}f\right) \\ &= \frac{n}{m}\left(\frac{1}{2}\right)^{(n+m)/2}\frac{\Gamma((m+n)/2)}{\Gamma(n/2)\Gamma(m/2)}\left(\frac{n}{m}f\right)^{(n/2)-1}\left(\frac{2}{\frac{n}{m}f+1}\right)^{(m+n)/2}, \qquad \frac{n}{m}f>0 \\ &= \frac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{n}{m}\right)^{(n/2)}}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)}f^{(n/2)-1}\left(\frac{n}{m}f+1\right)^{-(m+n)/2}, \qquad f>0 \end{aligned}$$

Comparando a equação final com a densidade apresentada na Definição 2.4, podemos concluir que $F \sim F_{n,m}$. □

A partir de agora as demonstrações que envolvem uma variável aleatória com distribuição F serão, na grande maioria das vezes, feitas a partir do resultado do último Teorema 2.2.

Corolário 2.4. *São verdadeiras as seguintes relações.*

(i) Se $F \sim F_{n,m} \;\Rightarrow\; 1/F \sim F_{m,n}$.

(ii) Se $T \sim t_n \;\Rightarrow\; T^2 \sim F_{1,n}$.

Demonstração. A demonstração é simples, trata-se de uma aplicação direta dos resultados dos Teoremas 2.1 e 2.2. Inclusive aparecem como execícios da lista de exercícios desse capítulo.

(i) Se $F \sim F_{n,m}$ então $F \overset{d}{=} \dfrac{W/n}{V/m}$ onde $W \sim \chi^2_n$, $V \sim \chi^2_m$, sendo W e V independentes. Então, $\dfrac{1}{F} \overset{d}{=} \dfrac{V/m}{W/n}$, ou seja, $\dfrac{1}{F} \sim F_{m,n}$.

(ii) Se $T \sim t_n$ então $T \stackrel{d}{=} \dfrac{Z}{\sqrt{W/n}}$ onde $Z \sim N(0,1)$, $W \sim \chi^2_n$, sendo Z e W independentes.

Então, $T^2 \stackrel{d}{=} \dfrac{Z^2}{W/n}$. Mas veja que $Z^2 \sim \chi^2_1$, logo, $T^2 \sim F_{1,n}$.

□

Proposição 2.8. *Seja* $F \sim F_{n,m}$. $\mathrm{E}(F^k) < \infty$ *se e somente se* $k < \dfrac{m}{2}$.

Demonstração. Se $F \sim F_{n,m}$ então $F \stackrel{d}{=} \dfrac{W/n}{V/m} = \dfrac{mW}{nV}$ onde $W \sim \chi^2_n$, $V \sim \chi^2_m$, sendo W e V independentes. Então,

$$\mathrm{E}(F^k) = \mathrm{E}\left(\left(\frac{mW}{nV}\right)^k\right) = \left(\frac{m}{n}\right)^k \mathrm{E}\left(W^k\right)\mathrm{E}\left(\frac{1}{V^k}\right).$$

Veja que $\mathrm{E}\left(W^k\right)$ existe para todo k, uma vez que uma variável aleatória com distribuição qui-quadrado possui todos os momentos. Precisamos então entender sob quais condições $\mathrm{E}\left(\dfrac{1}{V^k}\right) < \infty$. Pela Demonstração do Teorema 2.6 tiramos que

$$\begin{aligned}
\mathrm{E}\left(V^{-k}\right) = \mathrm{E}\left(V^{-2k/2}\right) &= \frac{(1/2)^{m/2}}{\Gamma(m/2)}\frac{\Gamma((m-2k)/2)}{(1/2)^{(m-2k)/2}}, \quad \text{se} \quad (m-2k)/2 > 0.\\
&= \frac{1}{2^k}\frac{\Gamma(m/2-k)}{\Gamma(m/2)}, \quad \text{se} \quad k < m/2.
\end{aligned}$$

Então, se $k < \dfrac{m}{2}$ garantimos que $\mathrm{E}(F^k) < \infty$.

□

Corolário 2.5. *Se* $F \sim F_{n,m}$ *e* $m \geq 3$, *então* $\mathrm{E}(F) = \dfrac{m}{m-2}$.

Demonstração. Primeiro veja que se $F \sim F_{n,m}$ e $m \geq 3$, segundo a Proposição 2.8, existem todos os momentos k tais que $k < \dfrac{3}{2} = 1,5$. Ou seja, $\mathrm{E}(F)$ existe. Nesse caso, para calcular $\mathrm{E}(F)$ vamos usar o resultado extraído da demonstração da Proposição 2.8, onde $W \sim \chi^2_n$ e $V \sim \chi^2_m$,

$$\mathrm{E}(F) = \frac{m}{n}\mathrm{E}\left(W\right)\frac{1}{2}\frac{\Gamma(m/2-1)}{\Gamma(m/2)} = \frac{m}{n}n\frac{1}{2}\frac{\Gamma(m/2)}{(m/2-1)\Gamma(m/2)} = \frac{m}{2}\frac{2}{m-2} = \frac{m}{m-2}.$$

□

Para facilitar as contas de probabilidades envolvendo uma variável aleatória $F_{n,m}$, os principais quantis da distribuição F estão apresentados na Tabela A.5 do Apêndice A. Por exemplo, considerando $F \sim F_{2,5}$, se queremos calcular $\mathrm{P}(F \leq 5)$ pela tabela podemos verificar que para $n = 2$ e $m = 5$ temos $\mathrm{P}(F \leq 3,7797) = 0,9$ e $\mathrm{P}(F \leq 5,7861) = 0,95$, logo $0,90 \leq \mathrm{P}(F \leq 5) \leq 0,95$.

Já que $F \sim F_{n,m} \Rightarrow 1/F \sim F_{m,n}$, resultado apresentado no Corolário 2.4, a tabela da distribuição F só apresentada os quantis maiores que 0,5, mas pode ser usada para outros quantis também. Por exemplo, supondo novamente $n = 2$ e $m = 5$, diretamente pela tabela não podemos encontrar x tal que $\mathrm{P}(F \leq x) = 0,1$. Mas sabemos que $\mathrm{P}(F \leq x) = \mathrm{P}\left(\frac{1}{F} \geq \frac{1}{x}\right) = 1 - \mathrm{P}\left(\frac{1}{F} \leq \frac{1}{x}\right)$. Então,

$$\mathrm{P}(F \leq x) = 0,1 \quad \Rightarrow \quad 1 - \mathrm{P}\left(\frac{1}{F} \leq \frac{1}{x}\right) = 0,1 \quad \Rightarrow \quad \mathrm{P}\left(\frac{1}{F} \leq \frac{1}{x}\right) = 0,9.$$

Logo, pelo Corolário 2.4, $\frac{1}{F} \sim F_{5,2}$ e consultando a tabela temos $\frac{1}{x} = 9,2926$. Então, $x = 0,107612$.

Exercícios da Seção 2.3

2.3.1 Seja $X \sim F_{7,10}$. Fazendo o uso preferencialmente do Programa R, ou das Tabelas A.5-A.7, encontre o que se pede.

(a) $\mathrm{P}(X < 4)$

(b) $\mathrm{P}(X > 1,5)$

(c) $\mathrm{P}(X < 0,6)$

(d) $\mathrm{P}(0,37 < X < 0,97)$

(e) Qual o valor de x tal que $\mathrm{P}(X > x) = 0,95$?

2.3.2 Seja $T \sim t_9$. Calcule $\mathrm{P}(T^2 < 3,36)$ a partir da Tabela A.5, que apresenta os quantis da distribuição F.

2.3.3 ([Casella e Berger, 2002] - Capítulo 5) Demonstre as seguintes afirmações:

(a) Se $T \sim t_n \;\Rightarrow\; T^2 \sim F_{1,n}$

(b) Se $X \sim F_{n,m} \;\Rightarrow\; \frac{1}{X} \sim F_{m,n}$

(c) Se $X \sim F_{n,m} \;\Rightarrow\; \frac{\frac{n}{m}X}{1+\frac{n}{m}X} \sim Beta\left(\frac{n}{2}, \frac{m}{2}\right)$

2.3.4 Seja $X \sim F_{n,m}$ sendo $m > 4$. Encontre $\mathrm{Var}(X)$. Dica: reveja as demonstrações da Proposição 2.8 e do Corolário 2.5.

2.4 Alguns Resultados Importantes para a População Normal

Nessa seção veremos alguns resultados importantes para as estatísticas média e a variância amostral de uma população normal. A seção será composta por um único teorema e mais um corolário. O foco principal é na interpretação e demonstração desse teorema.

Teorema 2.3. *Seja $X_1, X_2, \ldots X_n$ amostra aleatória de $X \sim N(\mu, \sigma^2)$. Considere as estatísticas*

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \qquad e \qquad S^2 = \frac{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}{n-1}.$$

Então,

(i) $\bar{X}$ e S^2 são variáveis aleatórias independentes.

(ii) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.

(iii) $\dfrac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim t_{n-1}$.

Demonstração. Demonstraremos um item por vez, na ordem que eles foram apresentados.

(i) Para a demonstração desse item vamos usar alguns resultados de normal multivariada apresentados na Seção 1.3. De acordo com a Definição 1.4, podemos concluir que

$$\boldsymbol{X} = (X_1, X_2, \dots X_n)^T \sim N_n(\boldsymbol{\mu}, \Sigma),$$

uma vez que $\boldsymbol{X}$ é identicamente distribuído a uma função linear do vetor aleatório $Z = (Z_1, Z_2, \dots, Z_n)^T$, onde cada $Z_i \sim N(0,1)$ e as variáveis $Z_1, Z_2, \dots, Z_n$ são independentes:

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \stackrel{\text{d}}{=} \begin{pmatrix} \sigma & 0 & \dots & 0 \\ 0 & \sigma & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix} + \underbrace{\begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix}}_{\boldsymbol{\mu}}.$$

Veja agora que o vetor aleatório $\boldsymbol{Y} = (\bar{X}, X_1 - \bar{X}, X_2 - \bar{X}, \dots, X_n - \bar{X})$ tem dimensão $n+1$ e pode ser escrito como função linear de $\mathbf{X}$:

$$\boldsymbol{Y} = \begin{pmatrix} \bar{X} \\ X_1 - \bar{X} \\ X_2 - \bar{X} \\ \vdots \\ X_n - \bar{X} \end{pmatrix} = \underbrace{\begin{pmatrix} 1/n & 1/n & \dots & 1/n \\ 1 - 1/n & -1/n & \dots & -1/n \\ \vdots & \vdots & \ddots & \vdots \\ -1/n & -1/n & \dots & 1 - 1/n \end{pmatrix}}_{A} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}.$$

Ou seja, $\mathbf{Y} = A\mathbf{X}$ e a matriz A de dimensão $(n+1) \times n$ está apresentada acima. Assim, segundo a Proposição 1.10, $\boldsymbol{Y}$ também tem distribuição normal multivariada. Estamos interessados na matriz de covariância de $\boldsymbol{Y}$.

$$\begin{aligned} \operatorname{Var}(\mathbf{Y}) &= \operatorname{Var}(A\boldsymbol{X}) = A \operatorname{Var}(\mathbf{X})\, A^T = A\sigma^2 I A^T = \sigma^2 A A^T \\ &= \sigma^2 \begin{pmatrix} 1/n & 0 & \dots & 0 \\ 0 & 1 - 1/n & \dots & -1/n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & -1/n & \dots & 1 - 1/n \end{pmatrix} = \begin{pmatrix} \sigma^2/n & 0 & \dots & 0 \\ 0 & \sigma^2 - \sigma^2/n & \dots & -\sigma^2/n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & -\sigma^2/n & \dots & \sigma^2 - \sigma^2/n \end{pmatrix} \end{aligned}$$

A partir da matriz $\operatorname{Var}(\mathbf{Y})$ podemos concluir que, para todo i, $\operatorname{Cov}(\bar{X}, X_i - \bar{X}) = 0$, uma vez que a primeira linha (e primeira coluna) da matriz $\operatorname{Var}(\mathbf{Y})$ apresenta somente valores zeros, a menos da posição $(1,1)$, onde temos $\operatorname{Var}(\bar{X}) = \sigma^2/n$. Assim, novamente pela Proposição 1.10, concluímos que $\bar{X}$ e $X_i - \bar{X}$ são variáveis aleatórias independentes, qualquer que seja $i = 1, 2, \dots, n$, uma vez que para vetores aleatórios com distribuição normal multivariada variáveis aleatórias descorrelatadas são independentes.

Veja que $S^2 = \frac{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}{n-1}$ é função somente das variáveis aleatórias $X_i - \bar{X}$, $i = 1, 2, \dots, n$. Então S^2 também é variável aleatória independente de $\bar{X}$. E assim demonstramos o primeiro item.

(ii) Agora vamos mostrar que $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$. Para isso vamos escrever

$$\frac{(n-1)S^2}{\sigma^2} = \frac{n-1}{\sigma^2} \frac{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}{n-1} = \frac{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2.$$

Agora vamos desenvolver outra expressão, $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$.

$$\begin{aligned}
\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i-\bar{X}+\bar{X}-\mu\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left((X_i-\bar{X})+(\bar{X}-\mu)\right)^2 \\
&= \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 + 2\sum_{i=1}^{n}(X_i-\bar{X})(\bar{X}-\mu) + \sum_{i=1}^{n}\left(\bar{X}-\mu\right)^2\right) \\
&= \underbrace{\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2}_{\frac{(n-1)S^2}{\sigma^2}} + \frac{2}{\sigma^2}(\bar{X}-\mu)\underbrace{\sum_{i=1}^{n}(X_i-\bar{X})}_{0} + \underbrace{\sum_{i=1}^{n}\left(\frac{\bar{X}-\mu}{\sigma}\right)^2}_{n\left(\frac{\bar{X}-\mu}{\sigma}\right)^2}
\end{aligned}$$

Assim chegamos na expressão:

$$\underbrace{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2}_{W} = \underbrace{\frac{(n-1)S^2}{\sigma^2}}_{W_1} + \underbrace{\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2}_{W_2}.$$

Observe que:

- $W=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2_n$, uma vez que $\frac{X_i-\mu}{\sigma}\sim N(0,1)$ e são independentes;
- $W_2=\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2\sim\chi^2_1$, pois $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ tem distribuição normal padrão;
- $W_1=\frac{(n-1)S^2}{\sigma^2}$ e $W_2=\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$ são variáveis aleatórias independentes, uma vez que W_1 é função somente da variável aleatória S^2, W_2 é função somente da variável aleatória $\bar{X}$ e já vimos que S^2 e $\bar{X}$ são independentes.

Uma vez que escrevemos W como soma de variáveis aleatórias independentes, W_1 e W_2, podemos dizer que a função geradora de momentos de W é dada pelo produto das funções geradoras de momentos de W_1 e W_2, isto é,

$$M_W(t)=M_{W_1}(t)M_{W_2}(t) \qquad \Rightarrow \qquad M_{W_1}(t)=\frac{M_W(t)}{M_{W_2}(t)}.$$

Como conhecemos

$$M_W(t)=\left(\frac{1/2}{1/2-t}\right)^{n/2}=\left(\frac{1}{1-2t}\right)^{n/2} \text{ e } M_{W_2}(t)=\left(\frac{1/2}{1/2-t}\right)^{1/2}=\left(\frac{1}{1-2t}\right)^{1/2},$$

podemos obter

$$M_{W_1}(t)=\frac{\left(\frac{1}{1-2t}\right)^{n/2}}{\left(\frac{1}{1-2t}\right)^{1/2}}=\left(\frac{1}{1-2t}\right)^{(n-1)/2}.$$

Veja que $W_1=\frac{(n-1)S^2}{\sigma^2}$ tem função geradora de momentos de uma variável aleatória $Gama((n-1)/2,1/2)$. Então, pela unicidade da função geradora de momentos, concluímos que ela tem essa distribuição, ou seja, $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2_{n-1}$, como queríamos demonstrar.

(iii) O último item a ser demonstrado é o mais simples dos 3. Precisamos apenas escrever $\dfrac{\bar{X}-\mu}{\sqrt{S^2/n}}$ de acordo com a expressão apresentada no Teorema 2.1.

$$\frac{\bar{X}-\mu}{\sqrt{S^2/n}} = \sqrt{n}\frac{\bar{X}-\mu}{\sqrt{S^2}} = \sqrt{n}\frac{\dfrac{\bar{X}-\mu}{\sigma}}{\sqrt{\dfrac{S^2}{\sigma^2}}} = \frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{S^2}{\sigma^2}}} = \frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2}{(n-1)\sigma^2}}} = \frac{Z}{\sqrt{\dfrac{W_1}{(n-1)}}}$$

onde $Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ e $W_1 = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$, sendo Z e W_1 variáveis aleatórias independentes , uma vez que W_1 é função somente da variável aleatória S^2, W_2 é função somente da variável aleatória $\bar{X}$ e já vimos que S^2 e $\bar{X}$ são independentes. Então, de acordo com o Teorema 2.1,

$$\frac{\bar{X}-\mu}{\sqrt{S^2/n}} = \frac{Z}{\sqrt{\dfrac{W_1}{(n-1)}}} \sim t_{n-1}.$$

□

Corolário 2.6. *Seja $X_1, X_2, \ldots X_n$ amostra aleatória de $X \sim N(\mu, \sigma^2)$. Considere as estatísticas*

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \qquad e \qquad S^2 = \frac{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}{n-1}.$$

Então,

(i) $S^2 \sim Gama\left(\dfrac{n-1}{2}, \dfrac{n-1}{2\sigma^2}\right)$.

(ii) $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2}{n} \sim Gama\left(\dfrac{n-1}{2}, \dfrac{n}{2\sigma^2}\right)$.

(iii) $\bar{X}$ *e* $\hat{\sigma}^2$ *são variáveis aleatórias independentes.*

Demonstração. Essa demonstração é uma aplicação direta do Teorema 2.3

(i) Vimos na demonstração do Teorema 2.3 que $W_1 = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$, ou seja, $W_1 \sim Gama\left(\dfrac{n-1}{2}, \dfrac{1}{2}\right)$. Veja que $S^2 = \dfrac{\sigma^2}{n-1}W_1$. Logo, de acordo com o resultado da Proposição 1.18,

$$S^2 \sim Gama\left(\frac{n-1}{2}, \frac{n-1}{2\sigma^2}\right).$$

(ii) Veja que $\hat{\sigma}^2 = \dfrac{n-1}{n}S^2$. Como $S^2 \sim Gama\left(\dfrac{n-1}{2}, \dfrac{n-1}{2\sigma^2}\right)$, usando novamente a Proposição 1.18,

$$\hat{\sigma}^2 \sim Gama\left(\frac{n-1}{2}, \frac{n}{2\sigma^2}\right)$$

(iii) Para terminar, como podemos escrever $\hat{\sigma}^2$ como função somente da variável aleatória S^2, $\hat{\sigma}^2 = \dfrac{n-1}{n}S^2$, e S^2 é independente de $\bar{X}$, podemos concluir que $\hat{\sigma}^2$ também é independente de $\bar{X}$.

□

Exemplo 2.4. *Seja X_1, X_2, X_3, X_4, X_5 uma amostra aleatória de tamanho 5 de uma população $X \sim N(\mu, 1)$. Encontre as distribuições de W, Y, V, U e T definidas abaixo e diga quais dessas variáveis aleatórias são estatística e quais não são.*

$$W = \sum_{i=3}^{5}(X_i - \mu)^2 \ , \quad Y = X_1 - X_2 \ , \quad V = \frac{2}{3}\frac{W}{Y^2} \ , \quad U = \sum_{i=3}^{5}(X_i - \bar{X})^2 \ , \quad T = \frac{Y}{\sqrt{U}}$$

Solução:

Primeiro vejamos a variável $W = \sum_{i=3}^{5}(X_i - \mu)^2$. Veja que cada $(X_i - \mu) \sim N(0, 1)$ e são independentes sempre que $i \neq j$. Logo, $(X_i - \mu)^2 \sim \chi^2_1$ e são variáveis independentes quando $i \neq j$. Com isso, $W = \sum_{i=3}^{5}(X_i - \mu)^2 = (X_3 - \mu)^2 + (X_4 - \mu)^2 + (X_5 - \mu)^2 \sim \chi^2_3$, uma vez que é soma de 3 variáveis aleatórias independentes com distribuição dada pelo quadro de normais padrão. Além disso, W não é estatística pois depende do parâmetro desconhecido μ.

Agora vamos analisar $Y = X_1 - X_2$. Como já foi visto no capítulo de revisão, combinação linear de normais independentes também é uma normal e a esperança e a variância do resultado pode ser tirada pelas propriedades do operador esperança e variância (Proposição 1.7) . Por isso, $Y = X_1 - X_2 \sim N(0, 2)$. Como Y não depende de parâmetros desconhecidos, Y é estatística.

Seja $V = \frac{2}{3}\frac{W}{Y^2}$. Já vimos que $W \sim \chi^2_3$ e $Y \sim N(0, 2)$. Veja que W e Y são variáveis independentes, uma vez que $W = f(X_3, X_4, X_5)$ e $Y = g(X_1, X_2)$, sendo o vetor aleatório (X_1, X_2) independente de (X_3, X_4, X_5). Podemos então afirmar que $\frac{Y}{\sqrt{2}} \sim N(0, 1)$ e $\left(\frac{Y}{\sqrt{2}}\right)^2 = \frac{Y^2}{2} \sim \chi^2_1$. Veja que podemos escrever $V = \frac{W/3}{Y^2/2} \sim F_{3,1}$. V não é estatística pois depende do parâmetro desconhecido μ, já que W depende de μ.

Agora vejamos a variável $U = \sum_{i=3}^{5}(X_i - \bar{X})^2$. Considerando a amostra aleatória $\{X_3, X_4, X_5\}$, $S^2 = \frac{\sum_{i=3}^{5}(X_i - \bar{X})^2}{2}$. Uma aplicação direta do Teorema 2.3, podemos concluir que $2S^2 \sim \chi^2_2$. Veja que $U = \sum_{i=3}^{5}(X_i - \bar{X})^2 = 2S^2 \sim \chi^2_2$. A variável U é estatística uma vez que não depende de parâmetros desconhecidos.

Por fim vamos encontrar a distribuição de $T = \frac{Y}{\sqrt{U}}$, sendo que já sabemos que $Y \sim N(0, 2)$ e $U \sim \chi^2_2$. Podemos reescrever T da seguinte forma: $T = \frac{Y/\sqrt{2}}{\sqrt{U}/\sqrt{2}} = \frac{Y/\sqrt{2}}{\sqrt{U/2}} = \frac{Z}{\sqrt{U/2}}$, onde $Z = Y/\sqrt{2} \sim N(0, 1)$ e $U \sim chi^2_2$. Além disso, Z e U são independentes. Dito isso, aplicando o resultado do Teorema 2.1, podemos concluir que $T \sim t_2$. Como Y e U são estatísticas, T também será. ‖

Exercícios da Seção 2.4

2.4.1 ([Larson, 1982] - Capítulo 6) Seja $X_1, X_2, \ldots, X_{10}$ uma amostra aleatória da população $X \sim N(\mu, \sigma^2)$. Com o auxílio preferencialmente do computador, ou das tabelas, calcule:

(a) $\mathrm{P}(|\bar{X} - \mu| \geq \sigma)$ (b) $\mathrm{P}(|\bar{X} - \mu| \leq S)$ (c) $\mathrm{P}\left(\frac{2}{3}S^2 > \sigma^2\right)$

onde $\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i$ e $S^2 = \frac{1}{9}\sum_{i=1}^{10}(X_i - \bar{X})^2$.

2.4.2 Suponha que $X_1, \ldots, X_{16}$ sejam uma amostra aleatória de $N(\mu, \sigma^2)$. Com auxílio preferencialmente do computador, ou das tabelas, calcule as seguintes probabilidades:

(a) $\mathrm{P}\left(\frac{1}{2}\sigma^2 \leq \frac{1}{16}\sum_{i=1}^{16}(X_i - \mu)^2 < 2\sigma^2\right)$ (b) $\mathrm{P}\left(\frac{1}{2}\sigma^2 \leq \frac{1}{16}\sum_{i=1}^{16}(X_i - \bar{X})^2 < 2\sigma^2\right)$.

2.5 Família Exponencial

Muitas das distribuições de probabilidade conhecidas se enquadram em um padrão geral de distribuições, definido pela família exponencial. Para o caso das distribuições que dependem de um único parâmetro temos a definição da família exponencial unidimensional.

Definição 2.5. *Seja X uma variável aleatória cuja distribuição depende de um único parâmetro θ e seja f_X a sua função de densidade ou função de probabilidade. Dizemos que a distribuição de X pertence à família exponencial unidimensional se*

$$f_X(x) = e^{c(\theta)T(x)}e^{d(\theta)}e^{S(x)}, \quad x \in A,$$

onde c e d são funções reais de θ (não podem depender de x), T e S são funções reais de x (não podem depender de θ) e $A = Im(X)$ não depende de θ.

Exemplo 2.5. *Mostre que a distribuição de $X \sim Exp(\lambda)$ pertence à família exponencial.*
Solução: Se $X \sim Exp(\lambda)$, então $f_X(x) = \lambda e^{-\lambda x}$, $x > 0$. Para mostrar que X pertence à família exponencial unidimensional basta escrever f_X no padrão apresentado na Definição 2.5.

$$\begin{aligned} f_X(x) &= \lambda e^{-\lambda x},\ x > 0 \\ &= e^{\ln(\lambda)}e^{-\lambda x},\ x > 0 \\ &= e^{-\lambda x}e^{\ln(\lambda)}e^{0},\ x \in (0, \infty). \end{aligned}$$

Assim temos $c(\lambda) = -\lambda$, $T(x) = x$, $d(\lambda) = \ln(\lambda)$, $S(x) = 0$ e $A = (0, \infty)$. E com isso mostramos que X pertence à família exponencial unidimensional. ||

Exemplo 2.6. *Mostre que a distribuição de $X \sim Poisson(\lambda)$ pertence à família exponencial.*
Solução: Se $X \sim Poisson(\lambda)$, então $p_X(x) = e^{-\lambda}\dfrac{\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$. Para mostrar que X pertence à família exponencial unidimensional basta escrever f_X no padrão apresentado na Definição 2.5.

$$\begin{aligned} p_X(x) &= e^{-\lambda}\frac{\lambda^x}{x!},\ x = 0, 1, 2, \ldots \\ &= e^{-\lambda}e^{\ln\left(\frac{\lambda^x}{x!}\right)},\ x = 0, 1, 2, \ldots \\ &= e^{-\lambda}e^{\ln(\lambda^x) - \ln(x!)},\ x = 0, 1, 2, \ldots \\ &= e^{-\lambda}e^{x\ln(\lambda)}e^{-\ln(x!)},\ x = 0, 1, 2, \ldots \\ &= e^{x\ln(\lambda)}e^{-\lambda}e^{-\ln(x!)},\ x \in \mathbb{N} \end{aligned}$$

Assim temos $c(\lambda) = \ln(\lambda)$, $T(x) = x$, $d(\lambda) = -\lambda$, $S(x) = -\ln(x!)$ e $A = \mathbb{N}$. E com isso mostramos que X pertence à família exponencial unidimensional. ||

Para o caso das distribuições que dependem de mais de um parâmetro, temos a definição da família exponencial multidimensional.

Definição 2.6. *Seja X uma variável aleatória cuja distribuição depende de um vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ e seja f_X a sua função densidade, caso X seja contínua, ou a sua função de probabilidade, caso X seja discreta. Dizemos que a distribuição de probabilidade de X pertence à família exponencial multidimensional se*

$$f_X(x) = e^{\sum_{j=1}^{k} c_j(\boldsymbol{\theta})T_j(x)}e^{d(\boldsymbol{\theta})}e^{S(x)}, \quad x \in A,$$

onde c e d são funções reais de $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ (não podem depender de x), T e S são funções reais de x (não podem depender de $\boldsymbol{\theta}$) e $A = Im(X)$ não depende de $\boldsymbol{\theta}$.

Exemplo 2.7. *Mostre que a distribuição de $X \sim N(\mu, \sigma^2)$ pertence à família exponencial bi-dimensional.*

Solução: Se $X \sim N(\mu, \sigma^2)$ então $f_X(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$. Para mostrar que X pertence à família exponencial bi-dimensional basta escrever f_X no padrão apresentado na Definição 2.6.

$$
\begin{aligned}
f_X(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \\
&= e^{\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)} e^{-\frac{1}{2\sigma^2}(x^2-2x\mu+\mu^2)} \\
&= e^{-\ln\left(\sqrt{2\pi\sigma^2}\right)} e^{-\frac{x^2}{2\sigma^2}+\frac{2x\mu}{2\sigma^2}} e^{-\frac{\mu^2}{2\sigma^2}} \\
&= e^{-\frac{1}{2\sigma^2}x^2+\frac{2\mu}{2\sigma^2}x} e^{-\ln\left(\sqrt{2\pi\sigma^2}\right)-\frac{\mu^2}{2\sigma^2}}, \ x \in \mathbb{R}
\end{aligned}
$$

Assim temos $c_1(\mu,\sigma^2) = -\frac{1}{2\sigma^2}$, $T_1(x) = x^2$, $c_2(\mu,\sigma^2) = \frac{2\mu}{2\sigma^2}$, $T_2(x) = x$, $d(\mu,\sigma^2) = -\ln\left(\sqrt{2\pi\sigma^2}\right) - \frac{\mu^2}{2\sigma^2}$, $S(x) = 0$ e $A = \mathbb{R}$. E com isso mostramos que X pertence à família exponencial bi-dimensional. $\|$

O que vamos observar ao longo dos capítulos é que, supondo $\{X_1, X_2, \ldots, X_n\}$ uma amostra aleatória da população X, se X pertence à família exponencial unidimensional será de grande interesse conhecer a distribuição amostral da estatística $\sum_{i=1}^n T(X_i)$. Se X pertence à família exponencial multidimensional de dimensão k, o nosso interesse será no vetor k-dimensional de estatísticas definido por $(\sum_{i=1}^n T_1(X_i), \sum_{i=1}^n T_2(X_i), \ldots, \sum_{i=1}^n T_k(X_i))$.

Exercícios da Seção 2.5

2.5.1 Seja X uma variável aleatória contínua cuja função densidade depende do parâmetro $\theta > 0$ e é definida por

$$f_X(x) = \frac{\theta 4^\theta}{x^{\theta+1}}\ , \quad x > 4.$$

Veja que $X \sim Pareto(\theta, 4)$.

(a) Verifique se a distribuição de X pertence ou não à família exponencial unidimensional.

(b) Encontre a distribuição de $Y = \ln\left(\dfrac{X}{4}\right)$ e identifique o seu modelo probabilístico. Faças as contas para encontrar a função densidade de Y.

(c) Seja $X_1, \ldots, X_n$ uma amostra aleatória da população X. Encontre a distribuição amostral da estatística $T(\mathbf{X}) = \sum_{i=1}^n \ln\left(\dfrac{X}{4}\right)$ e identifique o seu modelo probabilístico.

2.5.2 Seja X uma variável aleatória contínua cuja função densidade depende do parâmetro $\theta > 0$ e é definida por

$$f_X(x) = \frac{3\theta^3}{x^4}\ , \quad x > \theta.$$

Veja que $X \sim Pareto(3, \theta)$.

(a) Verifique se a distribuição de X pertence ou não à família exponencial unidimensional.

(b) Suponha uma amostra aleatória de X de tamanho n. Encontre a distribuição amostral da estatística $T(\mathbf{X}) = X_{(1)}$ e identifique o seu modelo probabilístico.

2.5.3 Seja X uma variável aleatória contínua cuja função densidade depende do parâmetro $\theta > 0$ e é definida por

$$f_X(x) = 3\theta^3 x^2 e^{-(\theta x)^3} \ , \quad x > 0.$$

Veja que $X \sim Weib(3, 1/\theta)$, de acordo com a parametrização apresentada no Apêndice B.2

(a) Verifique se a distribuição de X pertence ou não à família exponencial unidimensional.

(b) Encontre a distribuição de $Y = X^3$ e identifique o seu modelo probabilístico. Faças as contas para encontrar a função densidade de Y.

(c) Seja $X_1, \ldots, X_n$ uma amostra aleatória da população X. Encontre a distribuição amostral da estatística $T(\mathbf{X}) = \sum_{i=1}^n X_i^3$ e identifique o seu modelo probabilístico.

2.5.4 Seja X uma variável aleatória contínua cuja função densidade depende do parâmetro $\theta > 0$ e é definida por

$$f_X(x) = 2^\theta \theta x^{\theta-1}, \quad 0 < x < \frac{1}{2}.$$

Veja que $2X \sim Beta(\theta, 1)$.

(a) Verifique se a distribuição de X pertence ou não à família exponencial unidimensional.

(b) Encontre a distribuição de $Y = -\ln(2X)$ e identifique o seu modelo probabilístico. Faças as contas para encontrar a função densidade de Y.

(c) Suponha uma amostra aleatória de X de tamanho n. Encontre a distribuição amostral da estatística $T(\mathbf{X}) = \sum_{i=1}^n -\ln(2X_i)$ e identifique o seu modelo probabilístico.

2.5.5 Seja X uma variável aleatória discreta cuja função de probabilidade depende do parâmetro $0 < p < 1$ e é definida por

$$p_X(x) = \binom{15}{x} p^x (1-p)^{15-x}, \quad x = 0, 1, \ldots, 15$$

Veja que $X \sim Binom(15, p)$.

(a) Verifique se a distribuição de X pertence ou não à família exponencial unidimensional.

(b) Suponha uma amostra aleatória de X de tamanho n. Encontre a distribuição amostral da estatística $T(\mathbf{X}) = \sum_{i=1}^n X_i$ e identifique o seu modelo probabilístico.

2.5.6 Mostre que os seguintes modelos probabilísticos admitem distribuições pertencentes à família exponencial bidimensional:
(a) Gama(α, λ) (b) Beta(α, β).

2.6 Mais Alguns Exercícios para o Capítulo 2

2.1. Suponha que $X_1, \ldots, X_n$ sejam i.i.d $Exp(\lambda_x)$, que $Y_1, \ldots, Y_m$ sejam i.i.d $Exp(\lambda_y)$ e que $(X_1, \ldots, X_n)$ e $(Y_1, \ldots, Y_m)$ sejam vetores aleatórios independentes.

(a) Encontre as distribuições amostrais das estatísticas $\frac{\lambda_x \bar{X}}{\lambda_y \bar{Y}}$ e $\frac{\lambda_y \bar{Y}}{\lambda_x \bar{X}}$.

(b) Encontre o erro na frase anterior.

2.2. Seja X uma variável aleatória contínua cuja função densidade depende do parâmetro $\theta < 5$ e é definida por

$$f_X(x) = \frac{1}{5-\theta} \ , \quad \theta < x < 5.$$

Veja que $X \sim U(\theta, 5)$.

(a) Verifique se a distribuição de X pertence ou não à família exponencial unidimensional.

(b) Seja $X_1, \ldots, X_n$ uma amostra aleatória da população X. Encontre a distribuição amostral da estatística $T(\mathbf{X}) = X_{(1)}$.

2.3. Para cada afirmação abaixo indique se ela é verdadeira ou falsa. Para a falsa, apresente um contra-exemplo, isto é, um exemplo que satisfaz a hipótese mas não satisfaz a conclusão. Dica: ao buscar o contra-exemplo comece a pensar nos exemplos mais simples.

(a) Toda estatística é uma transformação de uma amostra aleatória.

(b) Toda transformação de uma amostra aleatória é uma estatística.

(c) Se $X \sim Gama(\alpha, \lambda)$ e f é a sua função densidade, então $f(x) = f(-x)\ \forall\ x \in \mathbb{R}$.

(d) Se $X \sim Gama(\alpha, \lambda)$ então $\bar{X} \sim Gama(n\alpha, n\lambda)$

(e) Se $X \sim t_n$ e f é a sua função densidade, então $f(x) = f(-x)$.

(f) Se $X \sim Gama(\alpha, \lambda)$ então $\mathrm{E}\left(\frac{1}{X}\right) = \frac{\lambda}{\alpha}$.

(g) Se $X \sim F_{n,2}$ então não existe $\mathrm{Var}(X)$.

(h) Se $X \sim F_{n,m}$ e f é a sua função densidade, então $f(x) = f(-x)\ \forall\ x \in \mathbb{R}$.

(i) Se $X \sim t_2$ então não existe $\mathrm{Var}(X)$.

(j) Se $X \sim F_{n,5}$ então existe $\mathrm{Var}(X)$.

Capítulo 3

Estimação Pontual

Nesse capítulo veremos o conceito de estimadores. Como encontrar estimadores para parâmetros desconhecidos e quais as propriedades esperadas em bons estimadores.

3.1 Conceitos Gerais

Neste texto, em nos demais cursos de Inferência Estatística, o objetivo é estudar como fazer inferências sobre os parâmetros desconhecidos da distribuição da população. Vamos assumir conhecida a distribuição da população, isto é, conhecemos o seu modelo probabilístico, e assumir também ser possível de se observar uma amostra aleatória dessa população.

Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de tamanho n da população X. Lembre-se, cada X_i é uma variável aleatória. Considere $\{x_1, x_2, \ldots, x_n\}$ uma observação dessa amostra aleatória, então cada x_i é uma realização da variável aleatória X_i. Repare que a notação usual é letras maiúsculas para representar a amostra aleatória (variáveis aleatórias) e letras minúsculas para representar a observação da amostra aleatória (constantes).

Veremos agora o conceito de estimador, que é a primeira forma de fazer inferência sobre um parâmetro desconhecido. A fim de tornar o texto menos repetitivo, para todas as definições, proposições, teoremas e exemplos daqui por diante, considere $X_1, X_2, \ldots, X_n$ uma amostra aleatória da população X, cuja distribuição é conhecida e depende do parâmetro desconhecido θ, ou, para o caso mais geral, depende do vetor de parâmetros desconhecidos $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)^T$.

Definição 3.1. *Seja θ um parâmetro desconhecido da distribuição de uma população X e Ψ seu espaço paramétrico. Um estimador para θ é qualquer estatística T cuja imagem está contida em Ψ. Em geral usamos $\hat{\theta}$ para representar o estimador de θ.*

De acordo com a Definição 3.1, um estimador é uma estatística, ou seja, ele é uma variável aleatória. Para cada observação da amostra o estimador assume um valor diferente, e o valor que ele assume é o que chamamos de estimativa do parâmetro. Muitas vezes usamos a mesma notação $\hat{\theta}$ para representar tanto o estimador quanto a estimativa, o que as vezes pode confundir os conceitos. Mas são conceitos distintos: o estimador é uma variável aleatória e a estimativa um número real, que é o valor assumido pelo estimador para uma certa observação da amostra.

Se você achou a Definição 3.1 vaga saiba que em muitos livros, como por exemplo [Larson, 1982] e [Casella e Berger, 2002], a definição de estimador é ainda mais vaga. Nesses livros um estimador para um parâmetro é definido simplesmente como qualquer estatística da amostra. Ou seja, nem é exigido que a imagem da estatística esteja contida no espaço paramérico do parâmetro a ser estimado.

Na prática encontramos estimadores, considerados razoáveis, como veremos em alguns exemplos, cuja imagem realmente não está contida no espaço paramétrico de θ. Isso significa que em alguns casos, raros, é verdade, dependendo da observação da amostra, a estimativa para o parâmetro pode ser um valor impossível para o parâmetro. Mas como a probabilidade disso acontecer é bem pequena, podemos aceitar a definição mais relaxada e entender que qualquer estatística pode ser considerada um estimador para qualquer parâmetro. O que vai nos fazer escolher essa ou aquela estatística para ser o estimador serão algumas medidas de qualidade, apresentadas na Seção 3.4.

Exemplo 3.1. *Seja X o tempo de vida de um certo tipo de dispositivo eletrônico e assuma conhecida a sua distribuição: $X \sim Exp(\lambda)$. Porém λ é um parâmetro desconhecido. Como podemos estimar o tempo de vida médio desses dispositivos a partir da observação de uma amostra da população X? Considere observados os seguintes tempos de vida, em dias, de 15 dispositivos eletrônicos:*

$$
\begin{array}{ccccccc}
12,38 & 11,03 & 51,84 & 50,25 & 98,38 & 433,65 & 20,13
\end{array}
$$
$$
\begin{array}{cccccccc}
36,19 & 3,63 & 155,78 & 80,31 & 28,51 & 169,01 & 271,21 & 41,58
\end{array}
$$

Solução:

Queremos encontrar uma estimativa para o parâmetro $\mu = \mathrm{E}(X) = \dfrac{1}{\lambda}$. Como λ é desconhecido, μ também é um parâmetro desconhecido.

Intuitivamente podemos usar a média amostral para estimar μ,

$$\hat{\mu} = \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}.$$

Nesse caso, $\hat{\mu} = \bar{X}$ é um estimador para μ. Veja que $\bar{X}$ é simplesmente uma estatística usada para estimar o valor desconhecido do parâmetro, e isso faz dele um estimador.

A partir da observação da amostra apresentada podemos encontrar uma estimativa para μ:

$$\hat{\mu} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{12,38 + 11,03 + 51,84 + \ldots + 271,21 = 41,58}{15} = 97,592,$$

o que significa que o tempo de vida médio para dispositivos eletrônicos desse tipo é estimado estar entre 97 e 98 dias.

Repare que $\hat{\mu} = \bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$ é o estimador para o parâmetro μ, uma variável aleatória. Já $\hat{\mu} = \bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n} = 97,592$ é a estimativa para o parâmetro μ, um número real. ||

3.2 Método dos Momentos

Veremos nessa seção o primeiro método para encontrar estimadores: o Método dos Momentos. Este método consiste em escolher como estimador para o parâmetro θ, ou para o vetor de parâmetros $\boldsymbol{\theta}$, a estatística que faz com que momentos populacionais sejam iguais aos respectivos momentos amostrais.Antes de ver como isso acontece, vamos formalizar o conceito de momentos populacionais e amostrais.

Definição 3.2. *Seja X uma população. Chamamos de j-ésimo momento populacional (ou momento teórico) de X a esperança da j-ésima potência de X, isto é,*

$$\mathrm{E}(X^j).$$

Veja que o momento populacional é uma constante, que pode ser conhecida ou não. Muitas vezes é uma função dos parâmetros da distribuição de X.

Definição 3.3. *Seja X_1, X_2, ..., X_n uma amostra aleatória da população X. Chamamos de j-ésimo momento amostral a estatística definida pela média amostral da j-ésima potência de X, isto é,*

$$M_j(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^{n} X_i^j.$$

Já o momento amostral é uma estatística, ou seja, uma variável aleatória. Para cada amostra observada da população X, o momento amostral assume um valor diferente.

O estimador pelo Método dos Momentos para um parâmetro desconhecido θ é encontrado depois que igualamos momentos populacionais aos momentos amostrais.

Antes de apresentarmos um algoritmo para o método vejamos alguns exemplos.

Exemplo 3.2. *Assim como no Exemplo 3.1, seja X o tempo de vida de um certo tipo de dispositivo eletrônico, $X \sim Exp(\lambda)$, sendo λ é um parâmetro desconhecido. Como podemos estimar λ a partir do Método dos Momentos?*

Veja que o 1º momento populacional de X é definido por: $\mathrm{E}(X) = \dfrac{1}{\lambda}$, que é função do parâmetro desconhecido para o qual queremos o estimador. Já o 1º momento amostral é a média amostral: $M_1(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$. O estimador para λ pelo método dos momentos será o resultado obtido quando encontramos o valor de λ que esses dois momentos iguais:

$$\frac{1}{\hat{\lambda}} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \Rightarrow \quad \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} X_i}.$$

Considerando novamente os tempos de vida, em dias, de 15 desses dispositivos eletrônicos, como no Exemplo 3.1.

$$\begin{gathered} 12,38 \quad 11,03 \quad 51,84 \quad 50,25 \quad 98,38 \quad 433,65 \quad 20,13 \\ 36,19 \quad 3,63 \quad 155,78 \quad 80,31 \quad 28,51 \quad 169,01 \quad 271,21 \quad 41,58 \end{gathered}$$

A estimativa para λ obtida através do método dos momentos é:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{15}{12,38 + 11,03 + 51,84 + \ldots + 271,21 + 41,58} = 0,0102.$$

Exemplo 3.3. *Uma máquina produz peças com 1cm de espessura nominal. Na prática exitem variações na espessura das peças produzidas, que podem ser maior ou menor que 1cm. Considere que a diferença entre a espessura real e a nominal seja uma variável aleatória X tal que $X \sim U[-\theta, \theta]$, com θ desconhecido.*

A fim de estimar o valor θ foram produzidas n dessas peças e, para cada uma delas, foram medidas as espessuras reais com um equipamento de alta precisão. Dessa forma foi possível observar uma amostra de X, a diferença entre a espessura real e a nominal. Considerando que a diferença entre as medidas em diferentes peças sejam variáveis aleatórias independentes. Como podemos estimar θ a partir do Método dos Momentos?

Para encontrar o estimador de θ pelo método dos momentos é preciso encontrar um momento populacional que dependa de θ e então igualar ele ao momento amostral equivalente. Veja que o 1º momento amostral de X é dado por $\mathrm{E}(X) = 0$, e não é função de θ. Como o 1º momento amostral não depende de θ precisamos procurar outro momento, esse não poderá ser usado. O 2º momento de X é dado por $\mathrm{E}(X^2) = \dfrac{4\theta^2}{12}$, que depende de θ. Vamos igualá-lo ao 2º momento amostral: $M_2(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i^2$.

Assim o estimador para λ pelo Método dos Momentos é dado por:

$$\frac{4\hat{\theta}^2}{12} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \quad \Rightarrow \quad \hat{\theta}^2 = \frac{12}{4n}\sum_{i=1}^{n} X_i^2 \quad \Rightarrow \quad \hat{\theta} = \pm\sqrt{\frac{12}{4n}\sum_{i=1}^{n} X_i^2}$$

Veja que o espaço paramétrico para θ é $\Psi = \mathbb{R}^+$, então vamos escolher o estimador

$$\hat{\theta} = \sqrt{\frac{12}{4n}\sum_{i=1}^{n} X_i^2}.$$

Se forem observados as seguintes diferenças entre as espessuras reais e nominais, em mm, para 18 peças produzidas,

$$\begin{array}{ccccccccc} 0,84 & -1,30 & -1,50 & 1,20 & -1,40 & 1,10 & -1,40 & 2,00 & -0,63 \\ -0,18 & -1,50 & 0,96 & 0,63 & -0,09 & -0,29 & -0,40 & 1,90 & -1,60. \end{array}$$

é possível encontrar uma estimativa para o parâmetro θ a partir do Método dos Momentos. A estimativa será

$$\hat{\theta} = \sqrt{\frac{12}{4n}\sum_{i=1}^{n} x_i^2} = \sqrt{\frac{12}{4\times 18}\left((0,84)^2 + (-1,30)^2 + \ldots + (-1,60)^2\right)} = 4,2726.$$

Veja que no Exemplo 3.2 acima logo o primeiro momento pode ser escrito como função do parâmetro desconhecido. Já no Exemplo 3.3 isso não aconteceu, e então foi preciso procurar outro momento populacional que dependa do parâmetro para o qual queremos encontrar o estimador.

Uma outra questão que pode aparecer é quando existe mais de um parâmetro desconhecido. Nesse caso, para encontrar o estimador pelo Método dos Momentos, é preciso escrever mais de um momento populacional como função dos parâmetros desconhecidos. No final é montado um sistema onde as incógnitas são os parâmetros desconhecidos e, para que esse sistema tenha solução, é preciso ter tantas equações quantos incógnitas. Veja o Exemplo 3.4 a seguir.

Exemplo 3.4. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória da população $X \sim Bin(N,p)$, sendo N e p parâmetros desconhecidos. Encontre o estimador para o vetor de parâmetros (N,p) pelo Método dos Momentos.*

Como são dois parâmetros desconhecidos é preciso encontrar dois momentos populacionais que dependam de N e p. Vejamos os dois primeiros momentos populacionais:

$$\mathrm{E}(X) = Np \quad e \quad \mathrm{E}(X^2) = N^2p^2 + Np(1-p).$$

Como encontramos dois momentos populacionais que dependem dos parâmetros desconhecidos vamos igualá-los aos respectivos momentos amostrais e resolver o sistema. A solução será estimador para o vetor de parâmetros pelo método dos momentos.

$$\begin{cases} \hat{N}\hat{p} & = \bar{X} \\ \hat{N}^2\hat{p}^2 + \hat{N}\hat{p}(1-\hat{p}) & = \dfrac{\sum_{i=1}^n X_i^2}{n} \end{cases}$$

$$\hat{N}\hat{p} = \bar{X} \quad \Rightarrow \quad \bar{X}^2 + \bar{X}(1-\hat{p}) = \frac{\sum_{i=1}^n X_i^2}{n} \quad \Rightarrow \quad \hat{p} = 1 - \frac{\sum_{i=1}^n X_i^2}{n\bar{X}} + \bar{X}$$

$$\hat{N} = \frac{\bar{X}}{\hat{p}} \quad \Rightarrow \quad \hat{N} = \frac{\bar{X}}{1 - \frac{\sum_{i=1}^n X_i^2}{n\bar{X}} + \bar{X}} \quad \Rightarrow \quad \hat{N} = \frac{n\bar{X}^2}{n\bar{X} - \sum_{i=1}^n X_i^2 + n\bar{X}^2}$$

Se os valores abaixo correspondem a uma amostra de tamanho 16 de X podemos encontrar as estimativas para N e p.

6	5	5	3	5	6	6	6	4	7	4	6	4	4	4	1

Veja que $\sum_{i=1}^{16} x_i = 76$, $\bar{x} = \frac{76}{16} = 4,75$ e $\sum_{i=1}^{16} x_i^2 = 394$. Então, as estimativas para p e N são:

$$\hat{p} = 1 - \frac{\sum_{i=1}^{16} x_i^2}{16\bar{x}} + \bar{x} = 0,566 \quad e \quad \hat{N} = \frac{16\bar{x}^2}{16\bar{x} - \sum_{i=1}^{16} x_i^2 + 16\bar{x}^2} = 8,39$$

Com isso estimamos que p seja aproximadamente $0,566$ e que N seja aproximadamente 8, já que N é um parâmetro que assume apenas valores inteiros. Veja que esse é um exemplo onde a imagem do estimador $\hat{N}$ não está contida no espaço paramétrico de N, uma vez que o estimador pode assumir valores não naturais.

Vejamos agora um algoritmos geral, onde a partir dele é possível encontrar o estimador pelo Método dos Momentos.

Algoritmo para o Método dos Momentos

Seja X_1, X_2, ..., X_n uma amostra aleatória de uma população X cuja distribuição depende do vetor de parâmetros desconhecido $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)^T$.

1. Defina k momentos populacionais que sejam funções de $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)^T$. Isto é, vamos definir k equações do tipo
$$\mathrm{E}(X^j) = g_j(\theta_1, \theta_2, \ldots, \theta_k).$$
2. Defina agora os respectivos momentos amostrais. Ou seja, para os mesmos valores de j do passo anterior definimos outras k equações do tipo
$$M_j(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i^j.$$
3. Monte um sistema $k \times k$ formado pelas equações
$$g_j(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k) = \frac{1}{n}\sum_{i=1}^{n} X_i^j.$$
4. Resolva o sistema e a solução será o estimador pelo Método dos Momentos para $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_{\boldsymbol{MM}}$.

Exemplo 3.5. *Uma turma tem 30 alunos e cada aluno lança uma moeda, não necessariamente justa, N vezes. Todos lançam o mesmo número de vezes a moeda. Ao final do experimento cada aluno passa para o professor o número de caras observadas.*

Considere que os seguintes valores forma informados ao professor por cada um dos 30 alunos:

2	5	4	5	3	6	6	3	4	3	3	3	6	1	5
4	3	5	1	4	5	5	6	2	3	3	1	2	3	4

$\left(\sum_{i=1}^{30} x_i = 110 \text{ e } \sum_{i=1}^{30} x_i^2 = 470\right)$.

(a) Sabendo que cada aluno jogou a moeda 10 *vezes, encontre uma estimativa para a probabilidade de sair cara ao lançar essa moeda?*

(b) Sem saber o número de vezes que a moeda foi lançada por cada aluno, encontre uma estimativa para a probabilidade de sair cara ao lançar essa moeda?

Veja que a amostra que temos é de uma variável aleatória $X \sim Bin(N, p)$, onde p é a probabilidade de sair cara com a moeda lançada e N o número de vezes que a moeda foi lançada por cada aluno.

No item (a) queremos encontrar uma estimativa para p supondo $N = 10$. Então precisamos de um momento populacional que dependa de p. Veja que $\mathrm{E}(X) = Np = 10p$, logo,

$$10\hat{p} = \bar{X} \quad \Rightarrow \quad \hat{p} = \frac{\bar{X}}{10}.$$

Então, para o item (a) chegamos na seguinte estimativa para p: $\hat{p} = \dfrac{\bar{x}}{10} = \dfrac{110}{30 \times 10} = 0,367.$

Para o item (b) não conhecemos N, então queremos encontrar uma estimativa para p supondo N desconhecido. O que vamos fazer é encontrar o estimador pelo método dos momentos para o vetor de parâmetros (N, p) e para isso precisamos de dois momentos populacionais que dependam desses parâmetros. Isso é exatamente o que já fizemos no Exercício 3.4, onde encontramos

$$\hat{p} = 1 - \frac{\sum_{i=1}^{n} X_i^2}{n\bar{X}} + \bar{X}.$$

A partir desse estimador chegamos na seguinte estimativa para p: $\hat{p} = 1 - \dfrac{470}{110} + \dfrac{110}{30} = 0,394.$

Exemplo 3.6. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória da população $X \sim N(\mu, \sigma^2)$. Encontre o estimador para σ^2 a partir do Método dos Momentos, supondo μ um parâmetro desconhecido.*

Vamos encontrar o estimador pelo método dos momentos para o parâmetro σ^2 de uma população $X \sim N(\mu, \sigma^2)$. Como μ é um parâmetro desconhecido, na prática encontraremos o estimador pelo métodos dos momentos para o vetor de parâmetros (μ, σ^2). Vamos então procurar dois momentos populacionais que dependem desses parâmetros e em seguida igualá-los aos respectivos momentos amostrais: $\mathrm{E}(X) = \mu$ *e* $\mathrm{E}(X^2) = \mu^2 + \sigma^2$. *Assim chega-se ao seguinte sistema,*

$$\begin{cases} \hat{\mu} & = \bar{X} \\ \hat{\mu}^2 + \hat{\sigma^2} & = \dfrac{\sum_{i=1}^n X_i^2}{n} \end{cases} \Rightarrow \quad \hat{\sigma^2} = \frac{\sum_{i=1}^n X_i^2}{n} - \left(\bar{X}\right)^2$$

Veremos que

$$\hat{\sigma^2} = \frac{\sum_{i=1}^n X_i^2}{n} - \left(\bar{X}\right)^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n},$$

mas essa igualdade é mais fácil de mostrar pelo caminho contrário:

$$\begin{aligned} \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} &= \frac{\sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2)}{n} = \frac{\sum_{i=1}^n X_i^2 - 2\bar{X}\sum_{i=1}^n X_i + \sum_{i=1}^n \bar{X}^2}{n} \\ &= \frac{\sum_{i=1}^n X_i^2 - 2n\bar{X}\bar{X} + n\bar{X}^2}{n} = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n} = \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2 = \hat{\sigma^2} \end{aligned}$$

Exemplo 3.7. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória da população $X \sim Pois(\lambda)$.*

(a) encontre o estimador para λ pelo Método dos Momentos;

(b) encontre o estimador para $\eta = \lambda^2 + \lambda$ pelo Método dos Momentos.

Para resolver o item (a) precisamos encontrar um momento populacional que dependa de λ, e temos direto $\mathrm{E}(X) = \lambda$. *Então,*

$$\hat{\lambda} = \bar{X}.$$

Para encontrar o estimador para η precisamos encontrar um momento populacional que seja escrito como função de η. Veja que η é exatamente o segundo momento, ou seja, $\mathrm{E}(X^2) = \lambda^2 + \lambda = \eta$, *assim chegamos em*

$$\hat{\eta} = \frac{\sum_{i=1}^n X_i^2}{n}.$$

O objetivo principal desse exercícios é destacar que, pelo Método dos Momentos, não necessariamente temos $\hat{\eta} \neq \hat{\lambda}^2 + \hat{\lambda}$, mesmo que $\eta = \lambda^2 + \lambda$. Quando procuramos o estimado pelo método dos momentos para η não podemos usar o estimados pelo método dos momentos para o parâmetro da distribuição, no caso, λ. Precisamos encontrar um momento populacional que seja função do parâmetro η em questão.

Exercícios da Seção 3.2

3.2.1 Para cada item abaixo, considere $X_1, \ldots, X_n$ amostra aleatória de X definida no item. Encontre o estimador de cada parâmetro distribucional pelo método dos momentos. Verifique em cada item se a imagem do estimador coincide com o espaço paramétrico. (a) $X \sim Geo(p)$ (b) $X \sim Geo^*(p)$ (b) $X \sim Gama(\alpha, \lambda)$ (c) $X \sim Lognormal(\mu, \sigma^2)$

Dica: Sejam $X \sim Lognormal(\mu, \sigma^2)$ e $W \sim N(\mu, \sigma^2)$. Então $X \stackrel{d}{=} e^W$ ou, o que é equivalente, $ln(X) \stackrel{d}{=} W$. Use essa informação para encontrar os momentos da Lognormal.

3.2.2 Seja X uma variável aleatória de Bernoulli com parâmetro p. Dada uma amostra aleatória com n observações de X, encontre o estimador para p pelo método dos momentos.

3.2.3 Em um município com 10.000 habitantes foi retirada uma amostra, com reposição, de 100 habitantes, com o objetivo de estimar o número de pessoas que votariam a favor de uma discutível obra proposta pelo prefeito. Sessenta e sete pessoas responderam "sim". A partir do método dos momentos, forneça uma estimativa do total de pessoas do município que votariam a favor da obra.

3.2.4 ([Larson, 1982] - Capítulo 7) João realizou o seguinte experimento: a partir da linha de lance livre lançou uma bola de basquete até fazer uma cesta. Esse experimento foi realizado 5 vezes e João precisou dos seguintes números de arremessos em cada um das vezes: 5, 1, 7, 4 e 9. Encontre a estimativa pelo método dos momentos para a probabilidade de João acertar a cesta em um arremesso a partir da linha de lance livre.

3.2.5 Considere que em uma cidade existam N táxis circulando sendo que, registrados e com a documentação em dia, são $N_0 = 4.532$ táxis. Com o intuito de checar se existem muitos táxis não regularizados, o órgão de fiscalização precisa estimar o valor de N. Par isso foi feita uma blitz e checadas documentações de 100 táxis, selecionados ao acaso e com reposição. Entre eles, 62 estavam devidamente registrados enquanto os outros 38 não. Encontre uma estimativa pelo método dos momentos para o número total de táxis nessa cidade, isto é, para N.

3.3 Método da Máxima Verossimilhança

A ideia por trás Método da Máxima Verossimilhança é escolher como estimativa para o parâmetro desconhecido o valor que torna a amostra observada a mais verossímil possível. Isto é, o valor estimado para o parâmetro desconhecido será aquele que atribui maior probabilidade de ocorrer uma amostra como a observada. Vejamos um exemplo ilustrativo da intuição desse método.

Exemplo 3.8. *Suponha que uma urna tenha algumas bolas brancas e outras pretas. Não sabemos quantas bolas de cada cor existem na urna, mas queremos estimar a probabilidade de sair uma bola branca, chamada de p (parâmetro desconhecido). Para isso foram recolhidas, com reposição, 5 bolas da urna. A amostra encontrada foi (P, P, B, P, B), onde B indica a ocorrência de uma bola branca e P de uma bola preta.*

Você acha razoável, depois de observar essa amostra, imaginar $p = 0,9$? Veja que se $p = 0,9$ a probabilidade de ocorrer uma amostra como a observada é bem baixa: $0,1 \times 0,1 \times 0,9 \times 0,1 \times 0,9 = 0,00081$, intuitivamente não vamos escolher esse valor para p. E imaginar $p = 0,5$, você acha razoável? Vamos fazer a mesma conta, a probabilidade de ocorrer uma amostra como a observada supondo $p = 0,5$ é: $0,5^5 = 0,03125$, que já é bem maior do que $0,00009$. Podemos repetir a conta para $p = 0,4$, a probabilidade de ocorrer uma amostra como a observada supondo $p = 0,4$ é $0,6 \times 0,6 \times 0,4 \times 0,6 \times 0,4 = 0,03456$.

É possível replicar essa conta para qualquer valor de $p \in [0,1]$. O Método da Máxima Verossimilhança busca o valor de p que atribui maior probabilidade para ocorrer uma amostra como a observada.

Definição 3.4. *Seja f_X a função densidade ou função de probabilidade da população X, dependendo se X for contínua ou discreta, respectivamente. Suponha que f_X dependa dos parâmetros desconhecidos $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)^T$ e que Ψ seja o espaço paramétrico de $\boldsymbol{\theta}$. A função de verossimilhança $L : \Psi \to \mathbb{R}$ é a função nas variáveis $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)^T$ definida por*

$$L(\boldsymbol{\theta}|\mathbf{x}) = \prod_{i=1}^{n} f_X(x_i|\boldsymbol{\theta})$$

É importante destacar que L não é a função de densidade conjunta da amostra, definida por $f_{\mathbf{X}} = \prod_{i=1}^{n} f_X(x_i|\boldsymbol{\theta})$. Veja que L é função de k variáveis enquanto a função $f_{\mathbf{X}}$ é função de n variáveis. Além disso, para a função L as variáveis são os parâmetros desconhecidos θ_j, $j = 1, \ldots, k$, e os valores de x_i, $i = 1, \ldots, n$, são considerados constantes. Já para a função $f_{\mathbf{X}}$ as variáveis são x_i, $i = 1, \ldots, n$, e os valores de θ_j, $j = 1, \ldots, k$, são considerados constantes.

Definição 3.5. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória da população X cuja distribuição depende do vetor de parâmetros desconhecidos $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)^T$. O estimador para $\boldsymbol{\theta}$ pelo Método da Máxima Verossimilhança é definido por*

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \Psi} L(\boldsymbol{\theta}|\mathbf{x}).$$

Veja que o estimador é o argumento que maximiza L, isto é, o ponto de máximo da função L, restrito ao espaço paramétrico de θ. Então, a estimativa por máxima verossimilhança será sempre um valor dentro do espaço paramétrico do parâmetro que queremos estimar. Esse estimador será encontrado a partir de técnicas de otimização aprendidas em cursos de cálculo.

Muitas vezes, pela própria estrutura da função L, é mais fácil encontrar o ponto de máximo da função $\ln(L(.))$ do que da própria L. Como as funções L e $\ln(L(.))$ têm os mesmos pontos de máximo, como mostra a Proposição 3.1 a seguir, torna-se possível trabalhar com essa função alternativa.

Proposição 3.1. *Seja $l(\boldsymbol{\theta}) = \log(L(\boldsymbol{\theta}))$, denominada função de Log-Verossimilhança.*

$$\boldsymbol{\theta}^\star = \arg\max_{\boldsymbol{\theta} \in \Psi} l(\boldsymbol{\theta}|\mathbf{x}) \quad \textit{se e somente se} \quad \boldsymbol{\theta}^\star = \arg\max_{\boldsymbol{\theta} \in \Psi} L(\boldsymbol{\theta}|\mathbf{x}).$$

Demonstração. Primeiro veja que $L(\boldsymbol{\theta}|\mathbf{x}) \geq 0 \ \forall \ \boldsymbol{\theta} \in \ \boldsymbol{\Psi}$, uma vez que $\prod_{i=1}^{n} f_X(x_i|\boldsymbol{\theta}) \geq \mathbf{0} \ \forall \ \boldsymbol{x} \in \mathbb{R}^n$. Então faz sentido falar em $\log(L(\boldsymbol{\theta}))$.

Veja também que $\log(.)$ é função estritamente crescente, o que nos permite afirmar que

$$\boldsymbol{x < y \Leftrightarrow \log(x) < \log(y)}.$$

Seja $\boldsymbol{\theta}^\star$ o ponto de máximo da função $\boldsymbol{l}$:

$$\boldsymbol{\theta}^\star = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Psi}} \boldsymbol{l}(\boldsymbol{\theta}|\mathrm{x}).$$

Então $\boldsymbol{l(\theta^\star) \geq l(\theta)} \ \forall \ \boldsymbol{\theta} \in \ \boldsymbol{\Psi}$, ou de forma equivalente, $\log(\boldsymbol{L}(\boldsymbol{\theta}^\star)) \geq \log(\boldsymbol{L}(\boldsymbol{\theta})) \ \forall \ \boldsymbol{\theta} \in \ \boldsymbol{\Psi}$. Como $\log$ é uma função monótona crescente,

$$\log(\boldsymbol{L}(\boldsymbol{\theta}^\star)) \geq \log(\boldsymbol{L}(\boldsymbol{\theta})) \Rightarrow \text{Ł}(\boldsymbol{\theta}^\star) \geq \boldsymbol{L}(\boldsymbol{\theta}) \quad , \quad \forall \quad \boldsymbol{\theta} \in \boldsymbol{\Psi},$$

e assim $\boldsymbol{\theta}^\star$ também é o ponto de máximo da função $\boldsymbol{L}$.

De forma equivalente vamos mostrar que se $\boldsymbol{\theta}^\star$ for ponto de máximo da função $\boldsymbol{L}$ também será da função $\boldsymbol{l}$ e assim terminar a demonstração.

Seja $\boldsymbol{\theta}^\star$ o ponto de máximo da função $\boldsymbol{L}$:

$$\boldsymbol{\theta}^\star = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Psi}} \boldsymbol{L}(\boldsymbol{\theta}|\mathrm{x}).$$

Então $\boldsymbol{L(\theta^\star) \geq L(\theta)} \ \forall \ \boldsymbol{\theta} \in \ \boldsymbol{\Psi}$. Como $\log$ é uma função monótona crescente,

$$\log(\boldsymbol{L}(\boldsymbol{\theta}^\star)) \geq \log(\boldsymbol{L}(\boldsymbol{\theta})) \Rightarrow \boldsymbol{l}(\boldsymbol{\theta}^\star) \geq \boldsymbol{l}(\boldsymbol{\theta}) \quad , \quad \forall \quad \boldsymbol{\theta} \in \boldsymbol{\Psi},$$

e assim $\boldsymbol{\theta}^\star$ também é o ponto de máximo da função $\boldsymbol{l}$. □

O logaritmo que define a função l na Proposição 3.1 pode ser em qualquer base. Pela presença constante de exponenciais nas funções densidade usuais será conveniente trabalhar com ln, o logaritmo neperiano.

Repare que para encontrar o estimador de máxima verossimilhança pode-se optar por encontrar o ponto de máximo da função L ou l, segundo a Proposição 3.1. Mas na maioria das vezes essa tarefa será mais fácil de fazer com a função l.

Para determinar os pontos de máximo de uma função um dos caminhos é buscar os pontos críticos dessa função. Em seguida é realizado o teste da segunda derivada, que nos indica se o pronto crítico encontrado representa um ponto de máximo ou não. Se o domínio da função não for um conjunto aberto, o ponto de máximo também pode estar no extremos do domínio.

Sobre como encontrar os pontos críticos e o teste da segunda derivada. Para o caso de uma única variável θ, os pontos críticos θ^* são aqueles que anulam a primeira derivada de l (ou L) em relação à θ e o teste da segunda derivada consiste em verificar se $\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} l(\theta^*) < 0$ (ou $\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} L(\theta^*) < 0$). Se sim, o ponto crítico θ^* é ponto de máximo das funções l e L.

No caso de l (ou L) ser função de duas variáveis, o ponto crítico $\boldsymbol{\theta^*}= (\theta_1^*, \theta_2^*)$ será aquele tal que $\boldsymbol{\nabla} l(\theta_1^*, \theta_2^*) = (0,0)$, isto é, será o ponto que satisfaz

$$\frac{\partial}{\partial \theta_1} l(\theta_1^*, \theta_2^*) = 0 \quad \text{e} \quad \frac{\partial}{\partial \theta_2} l(\theta_1^*, \theta_2^*) = 0.$$

O teste da segunda derivada será feito com a matriz de segundas derivadas definida por

$$\mathrm{D}^2\, l(\boldsymbol{\theta^*}) = \mathrm{D}^2\, l(\theta_1^*, \theta_2^*) = \begin{pmatrix} \frac{\partial^2}{\partial \theta_1^2} l(\theta_1^*, \theta_2^*) & \frac{\partial^2}{\partial \theta_1 \partial \theta_2} l(\theta_1^*, \theta_2^*) \\ \frac{\partial^2}{\partial \theta_1 \partial \theta_2} l(\theta_1^*, \theta_2^*) & \frac{\partial^2}{\partial \theta_2^2} l(\theta_1^*, \theta_2^*) \end{pmatrix}.$$

Se $\det\left(\mathrm{D}^2\, l(\boldsymbol{\theta^*})\right) > 0$ e $\dfrac{\partial^2}{\partial \theta_1^2} l(\theta_1^*, \theta_2^*) < 0$ a matriz $\mathrm{D}^2\, l(\boldsymbol{\theta^*})$ é negativa definida e o ponto crítico $\boldsymbol{\theta^*}= (\theta_1^*, \theta_2^*)$ é ponto de máximo (idem para o caso da função L).

Para o caso geral os pontos críticos $\boldsymbol{\theta^*}$ são aqueles que anulam o gradiente de l (ou L), definido pelo vetor de derivadas parciais de primeira ordem em relação a cada variável. O teste da segunda derivada, conhecido como condição de segunda ordem, consiste em verificar se $\mathrm{D}^2\, l(\boldsymbol{\theta^*})$ é uma matriz negativa definida. Se sim, o ponto crítico $\boldsymbol{\theta^*}$ é ponto de máximo. Para saber mais sobre condições de segunda ordem e matriz negativa definida recomendo [Bortolossi, 2002].

Exemplo 3.9. *Seja $X \sim Exp(\lambda)$. Encontre o estimador de máxima verossimilhança para o parâmetro λ.*

Para encontrar o estimador de máxima verossimilhança precisa-se primeiro encontrar a função de verossimilhança, ou a função de log-verossimilhança. No caso da população $X \sim Exp(\lambda)$ temos:

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i}, \;\; \lambda > 0 \quad e \quad l(\lambda) = \ln\left(\prod_{i=1}^{n} \lambda e^{-\lambda x_i}\right), \;\; \lambda > 0$$

As funções L e l estão definidas para valores de λ dentro do seu espaço paramétrico. Por isso restringimos essa função para $\lambda > 0$. Veja que

$$l(\lambda) = \ln\left(\prod_{i=1}^{n} \lambda e^{-\lambda x_i}\right) = \sum_{i=1}^{n} \ln\left(\lambda e^{-\lambda x_i}\right) = \sum_{i=1}^{n} \left(\ln(\lambda) - \lambda x_i\right) = n\ln(\lambda) - \lambda \sum_{i=1}^{n} x_i,$$

e dessa forma fica mais fácil a busca pelos pontos críticos de l. Por isso vamos seguir com a função l e não com a função L.

Para encontrar o(s) ponto(s) crítico(s) primeiro precisa-se encontrar a derivada de l:

$$l(\lambda) = n\ln(\lambda) - \lambda \sum_{i=1}^{n} x_i \quad \Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}\lambda} l(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i.$$

O ponto crítico será a solução da equação $\frac{\mathrm{d}}{\mathrm{d}\lambda}l(\lambda^\star) = 0$*:*

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}l(\lambda^\star) = 0 \quad \Rightarrow \quad \frac{n}{\lambda^\star} - \sum_{i=1}^{n} x_i = 0 \quad \Rightarrow \quad \lambda^\star = \frac{n}{\sum_{i=1}^{n} x_i}.$$

Agora que já se conhece o ponto crítico da função de log-verossimilhança precisa-se verificar se trata-se de um ponto de máximo a partir da segunda derivada de l*.*

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}l(\lambda) = -\frac{n}{\lambda^2} \quad \Rightarrow \quad \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}l(\lambda^\star) < 0 \quad \Rightarrow \quad \lambda^\star \text{ é ponto de máximo.}$$

Assim chegamos ao estimador de máxima verissimilhança para o parâmetro λ*:*

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\bar{X}}.$$

Repare que na hora de apresentar o estimador usamos X_i *e não* x_i*, pois um estimador tem que ser uma estatística. Repare também que esse é o mesmo estimador para* λ *já encontrado pelo Método dos Momentos (Exemplo 3.2).*

Exemplo 3.10. *Seja* $X \sim Ber(p)$*. Encontre o estimador de máxima verossimilhança para o parâmetro* p*.*

Para encontrar o estimador de máxima verossimilhança precisa-se primeiro encontrar a função de verossimilhança, ou a função de log-verossimilhança. No caso da população $X \sim Ber(p)$ *temos:*

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}, \ 0 \le p \le 1 \qquad e \qquad l(p) = \ln\left(\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}\right), \ 0 \le p \le 1$$

Novamente seguiremos com a função de log-verossimilhança pela maior facilidade em derivar a soma do que derivar o produto. Para encontrar o(s) ponto(s) crítico(s) primeiro precisa-se encontrar a derivada de l*:*

$$\begin{aligned}
l(p) &= \sum_{i=1}^{n} \ln\left(p^{x_i}(1-p)^{1-x_i}\right) = \sum_{i=1}^{n} x_i \ln(p) + \sum_{i=1}^{n}(1-x_i)\ln(1-p) \\
&= \ln(p)\sum_{i=1}^{n} x_i + \ln(1-p)\sum_{i=1}^{n}(1-x_i) = \ln(p)\sum_{i=1}^{n} x_i + \ln(1-p)\left(n - \sum_{i=1}^{n} x_i\right) \\
&= \ln(p)\sum_{i=1}^{n} x_i + n\ln(1-p) - \ln(1-p)\sum_{i=1}^{n} x_i \\
\frac{\mathrm{d}}{\mathrm{d}p}l(p) &= \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{n}{1-p} + \frac{1}{1-p}\sum_{i=1}^{n} x_i
\end{aligned}$$

O ponto crítico será a solução da equação $\frac{\mathrm{d}}{\mathrm{d}p}l(p^\star) = 0$*:*

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}p}l(p^\star) = 0 \quad &\Rightarrow \quad \frac{1}{p^\star}\sum_{i=1}^{n} x_i - \frac{n}{1-p^\star} + \frac{1}{1-p^\star}\sum_{i=1}^{n} x_i = 0 \\
&\Rightarrow \quad \frac{(1-p^\star)\sum_{i=1}^{n} x_i - np^\star + p^\star\sum_{i=1}^{n} x_i}{p^\star(1-p^\star)} = 0 \\
&\Rightarrow \quad \frac{\sum_{i=1}^{n} x_i - np^\star}{p^\star(1-p^\star)} = 0
\end{aligned}$$

Atenção: $p^\star \neq 0$ *e* $p^\star \neq 1$*, condição para que o denominador da equação acima não seja nulo. Seguindo sob essa condição, a solução da equação acima será encontrar o valor de* $p\star$ *que anula o numerador:*

$$\sum_{i=1}^{n} x_i - np^\star = 0 \quad \Rightarrow \quad p^\star = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

O teste da segunda derivada irá indicar se o ponto crítico encontrado é um ponto de máximo.

$$\frac{\mathrm{d}^2}{\mathrm{d}p^2} l(p) = -\frac{1}{p^2}\sum_{i=1}^{n} x_i - \frac{n}{(1-p)^2} + \frac{1}{(1-p)^2}\sum_{i=1}^{n} x_i = -\frac{1}{p^2}\sum_{i=1}^{n} x_i - \frac{n - \sum_{i=1}^{n} x_i}{(1-p)^2}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}p^2} l(p^\star) = -\frac{1}{\bar{x}^2}\sum_{i=1}^{n} x_i - \frac{n - \sum_{i=1}^{n} x_i}{(1-\bar{x})^2} < 0$$

uma vez que a para uma amostra de $X \sim Ber(p)$ *temos* $0 < \bar{x} < 1$ *e* $\sum_{i=1}^{n} x_i < n$*, não considerando as amostras tais que* $\bar{x} = 0$ *ou* $\bar{x} = 1$*, uma vez que já foi indicado que* $p^\star \neq 0$ *e* $p^\star \neq 1$.

$$\frac{\mathrm{d}^2}{\mathrm{d}p^2} l(p^\star) < 0 \quad \Rightarrow \quad p^\star \text{ é ponto de máximo.}$$

Assim chegamos ao estimador de máxima verissimilhança para o parâmetro p*:*

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}.$$

Mas e se ocorrer uma amostra tal que $\bar{x} = 0$ *ou* $\bar{x} = 1$*? Repare que isso só acontece se* $x_i = 0$ *ou* $x_i = 1$ *para todo* $i = 1, 2, \ldots, n$*. Vamos procurar o ponto de máximo de* L *para essas duas amostras específicas. Nesse caso vai ser mais fácil trabalhar com a função* L*.*

Se $x_i = 0 \ \forall \ i$*,* $L(p) = \prod_{i=1}^{n} p^0(1-p)^{1-0} = (1-p)^n$*. Se procurarmos os pontos críticos de* L *vamos encontrar um único ponto crítico que é ponto de mínimo, e não de máximo, veja Figura 3.1(a). Como o espaço paramétrico* $\Psi = [0, 1]$ *é um conjunto fechado, nesse caso, o máximo local estará em um dos extremos do intervalo* $[0, 1]$ *. Novamente observando a Figura 3.1(a), pode-se concluir que o máximo de* L *ocorre em* $p = 0$ *quando* $x_i = 0 \ \forall \ i$*, então essa é a estimativa de máxima verossimilhança para* p *para essa observação de amostra. Veja que esse resultado é bastante intuitivo.*

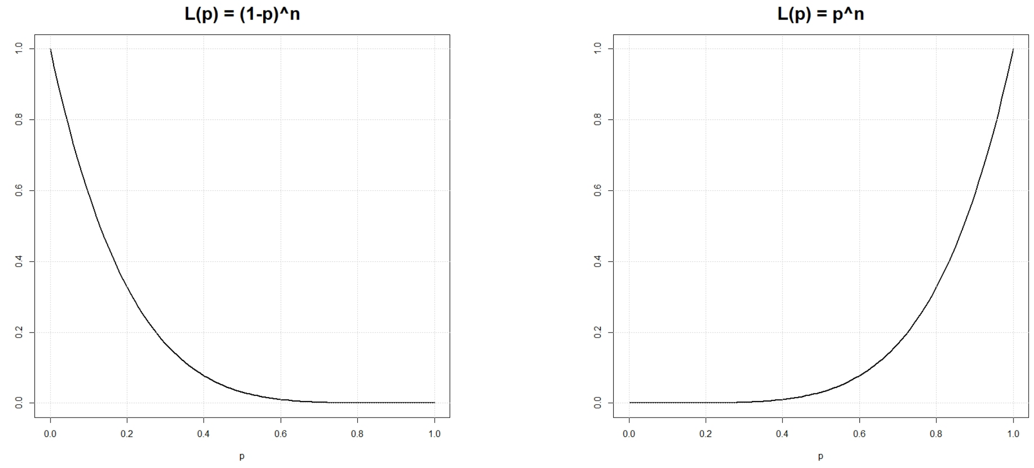
De forma análoga, Se $x_i = 1 \ \forall \ i$*,* $L(p) = \prod_{i=1}^{n} p^1(1-p)^{1-1} = p^n$*. Se procurarmos os pontos críticos de* L *vamos encontrar um único ponto crítico que é ponto de mínimo, e não de máximo, veja Figura 3.1(b). Como o espaço paramétrico* $\Psi = [0, 1]$ *é um conjunto fechado, nesse caso, o máximo local estará em um dos extremos do intervalo* $[0, 1]$ *. Novamente observando a Figura 3.1(b), pode-se concluir que o máximo de* L *ocorre em* $p = 1$ *quando* $x_i = 1 \ \forall \ i$*, então essa é a estimativa de máxima verossimilhança para* p *para essa observação de amostra. Mais uma vez, chegamos em um resultado intuitivo.*

O que é interessante é que se a amostra for uma dessas duas discutidas por último, as estimativas acabam coincidindo com a média amostral. Dessa forma podemos então concluir, de forma geral, para qualquer amostra observada, que o estimador de máxima verossimilhança para o parâmetro p *de* $X \sim Ber(p)$ *é*

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}.$$

Exemplo 3.11. *Seja* $X \sim N(\mu, \sigma^2)$*. Encontre o estimador de máxima verossimilhança para o vetor de parâmetros* (μ, σ^2)*.*

Nesse exemplo a distribuição da população depende de dois parâmetros desconhecidos, então o estimador do vetor de parâmetros (μ, σ^2) *será dado solução de um problema de maximização de uma função de duas variáveis.*



(a) $L(p) = (1-p)^n$ (b) $L(p) = p^n$

Figura 3.1: Gráficos da função L para o caso de $\bar{x} = 0$ ou $\bar{x} = 1$.

Como antes, primeiro precisa-se encontrar a função de verossimilhança, ou a função de log-verossimilhança. No caso da população $X \sim N(\mu, \sigma^2)$ temos:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/2\sigma^2}, \ -\infty < \mu < \infty \text{ e } \sigma^2 > 0$$
$$l(\mu, \sigma^2) = \ln\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/2\sigma^2}\right), \ -\infty < \mu < \infty \text{ e } \sigma^2 > 0$$

Como nos exemplos anteriores, será mais fácil buscar os pontos críticos pela função de log-verossimilhança.

$$l(\mu, \sigma^2) = \ln\left(\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/2\sigma^2}\right) = \sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/2\sigma^2}\right)$$
$$= -\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2} = -\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Para encontrar o(s) ponto(s) crítico(s) primeiro precisa-se encontrar as derivadas parciais de l:

$$l(\mu, \sigma^2) = -\frac{n}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \quad \Rightarrow \quad \frac{\partial}{\partial\mu} l(\mu, \sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} x_i - n\mu\right)$$
$$\Rightarrow \quad \frac{\partial}{\partial\sigma^2} l(\mu, \sigma^2) = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{2}{(2\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

O ponto crítico será a solução do sistema com equações $\frac{\partial}{\partial\mu} l(\mu^\star, \sigma^{2\star}) = 0$ e $\frac{\partial}{\partial\sigma^2} l(\mu^\star, \sigma^{2\star}) = 0$:

$$\begin{cases} \frac{1}{\sigma^{2\star}}\left(\sum_{i=1}^{n} x_i - n\mu^\star\right) & = \quad 0 \\ -\frac{n}{2}\frac{1}{\sigma^{2\star}} + \frac{2}{(2\sigma^{2\star})^2}\sum_{i=1}^{n}(x_i-\mu^\star)^2 & = \quad 0 \end{cases}$$

Da primeira equação do sistema tiramos:

$$\frac{1}{\sigma^{2}\star}\left(\sum_{i=1}^{n} x_i - n\mu^\star\right) = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} x_i - n\mu^\star = 0 \quad \Rightarrow \quad \mu^\star = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}.$$

Substituindo agora na segunda equação do sistema,

$$-\frac{n}{2}\frac{1}{\sigma^{2\star}}+\frac{2}{(2\sigma^{2\star})^2}\sum_{i=1}^{n}(x_i-\bar{x})^2=0 \;\Rightarrow\; \frac{2}{(2\sigma^{2\star})^2}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{n}{2}\frac{1}{\sigma^{2\star}} \;\Rightarrow\; \sigma^{2\star}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}.$$

Assim chega-se no ponto critico

$$(\mu^\star,\sigma^{2\star})=\left(\bar{x},\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}\right)$$

O próximo passo é verificar se o ponto crítico $(\mu^\star,\sigma^{2\star})$ encontrado é de fato um ponto de máximo. Para isso será feito o teste da segunda derivada para o problema de duas variáveis. As derivadas de segunda ordem são:

$$\frac{\partial^2}{\partial\mu^2}l(\mu,\sigma^2)=-\frac{n}{\sigma^2} \;\Rightarrow\; \frac{\partial^2}{\partial\mu^2}l(\mu^\star,\sigma^{2\star})=-\frac{n^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\frac{\partial^2}{\partial\mu\sigma^2}l(\mu,\sigma^2)=-\frac{1}{(\sigma^2)^2}\left(\sum_{i=1}^{n}x_i-n\mu\right) \;\Rightarrow\; \frac{\partial^2}{\partial\mu\sigma^2}l(\mu^\star,\sigma^{2\star})=-\frac{1}{(\sigma^{2\star})^2}\left(\sum_{i=1}^{n}x_i-n\bar{x}\right)=0$$

$$\frac{\partial^2}{\partial(\sigma^2)^2}l(\mu,\sigma^2)=\frac{n}{2}\frac{1}{(\sigma^2)^2}-\frac{1}{(\sigma^2)^3}\sum_{i=1}^{n}(x_i-\mu)^2 \;\Rightarrow\; \frac{\partial^2}{\partial(\sigma^2)^2}l(\mu^\star,\sigma^{2\star})=-\frac{n^3}{2(\sum_{i=1}^{n}(x_i-\bar{x})^2)^2}.$$

Agora já é possível montar a matriz de segunda derivada no ponto $(\mu^\star,\sigma^{2\star})$,

$$\mathrm{D}^2\,l(\mu^\star,\sigma^{2\star})=\begin{pmatrix}-\frac{n^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} & 0\\ 0 & -\frac{n^3}{2(\sum_{i=1}^{n}(x_i-\bar{x})^2)^2}\end{pmatrix}.$$

Como

$$\det\left(\mathrm{D}^2\,l(\mu^\star,\sigma^{2\star})\right)=\frac{n^5}{2\left(\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^3}>0 \quad e \quad -\frac{n^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}<0$$

para qualquer que seja a amostra observada, $\mathrm{D}^2\,l(\mu^\star,\sigma^{2\star})$ é uma matriz negativa definida o que garante que o ponto crítico encontrado é um ponto de máximo.

Sendo assim, o estimador para o vetor de parâmetros (μ,σ^2) pelo método da máxima verossimilhança é dado por:

$$(\hat{\mu},\hat{\sigma}^2)=\left(\bar{X},\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}\right).$$

Exemplo 3.12. *Seja $X\sim U[0,\theta]$. Encontre o estimador de máxima verossimilhança para o parâmetro θ.*

Esse é um exemplo interessante pois o método da máxima verossimilhança e o método dos momentos fornecem estimadores bem distintos, como veremos.

O primeiro passo para encontrar os estimadores de máxima verossimilhança, sempre, é buscar a função de verossimilhança L e a função de log-verossimilhança l. No caso da população $X\sim U[0,\theta]$ temos:

$$L(\lambda)=\prod_{i=1}^{n}\frac{1}{\theta},\;x_i\leq\theta \quad e \quad l(\lambda)=\ln\left(\prod_{i=1}^{n}\frac{1}{\theta}\right),\;x_i\leq\theta$$

Um detalhe desse exemplo que difere ele dos anteriores é a imagem de $X\sim U[0,\theta]$ depende de θ, uma vez que a função de densidade $f_X(x)=\dfrac{1}{\theta}$ somente para $x\leq\theta$. Por esse motivo *as funções L e l estão definidas para $x_i \leq \theta$, $\forall$ $i = 1, \ldots, n$. Veja que a condição $x_i \leq \theta$, $\forall$ $i = 1, \ldots, n$, é equivalente à $\max\{x_i\} \leq \theta$ $\forall$ $i = 1, \ldots, n$, ou ainda, $x_{(n)} \leq \theta$. Isso é verdade pois se θ é maior que todos os valores observados significa que θ é maior que o maior deles, e se θ é maior que o maior deles, θ será maior que todos eles.*

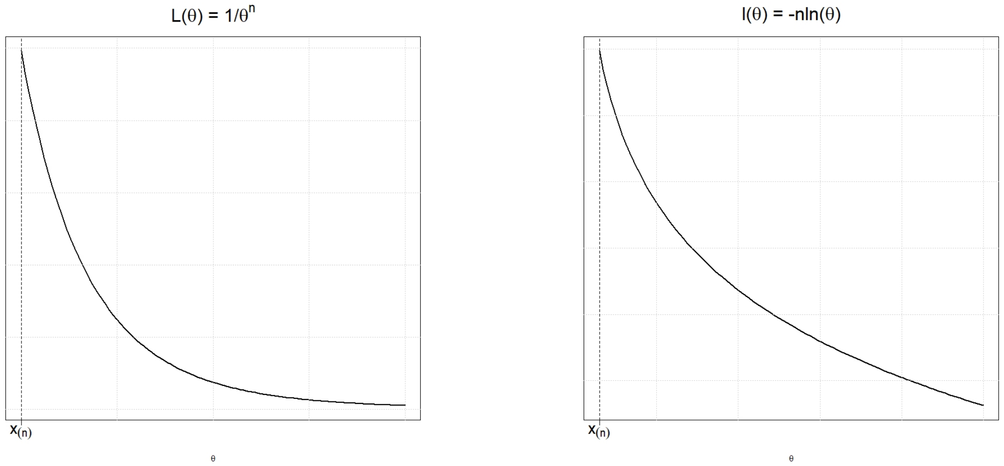
Resolvendo o somatório e produtório é possível chegar nas seguintes expressões para L e l

$$L(\lambda) = \prod_{i=1}^{n} \frac{1}{\theta} = \left(\frac{1}{\theta}\right)^{n}, \ x_{(n)} \leq \theta \quad \text{e} \quad l(\lambda) = -n\ln(\theta), \ x_{(n)} \leq \theta$$

Se tentarmos procurar os pontos críticos de L ou l veremos que não existem, uma vez que não existe $\theta^\star$ tal que $\frac{\mathrm{d}}{\mathrm{d}\theta}L(\theta^\star) = 0$ e nem $\frac{\mathrm{d}}{\mathrm{d}\theta}l(\theta^\star) = 0$. Mas podemos derivar e perceber que L e l são funções decrescentes, já que suas derivadas são negativas para qualquer valor de θ:

$$\frac{\mathrm{d}}{\mathrm{d}\theta}L(\theta) = -n\left(\frac{1}{\theta}\right)^{n+1}, \ x_i \leq \theta \quad \text{e} \quad \frac{\mathrm{d}}{\mathrm{d}\theta}l(\theta) = -n\frac{1}{\theta}, \ x_i \leq \theta.$$

Com isso podemos concluir que o máximo da função L (ou l) ocorre no menor valor possível para θ. No caso desse exemplo, o máximo de L e l ocorre em $x_{(n)}$. Isso também pode ser percebido pelo esboço do gráfico das funções L e l apresentados na Figura 3.2.



(a) $L(p) = 1/\theta^n$ (b) $l(p) = -n\ln(\theta)$

Figura 3.2: Gráficos das funções L e l para $X \sim Unif[0, \theta]$.

Sendo assim, o estimador pelo método da máxima verossimilhança para θ é a estatística de ordem n:

$$\hat{\theta} = X_{(n)}$$

Teorema 3.1. ***Propriedade de Invariância***

Seja $\hat{\theta}$ o estimador de máxima verossimilhança para θ e $\eta = \tau(\theta)$. Então $\hat{\eta} = \tau(\hat{\theta})$ é o estimador de máxima verossimilhança para η.

Demonstração. Vamos demonstrar apenas o caso particular em que τ é uma função inversível, isto é, quando existe τ^{-1} tal que $\theta = \tau^{-1}(\eta)$ e $\tau^{-1}(\tau(\theta)) = \theta$. Porém a propriedade de invariância vale mesmo quando τ é uma função não inversível. A demostração do caso geral pode ser encontrada em [Casella e Berger, 2002].

Seja L a função de máxima verossimilhança para o parâmetro θ e $\eta = \tau(\theta)$, sendo τ uma função inversível, então podemos escrever $\theta = \tau^{-1}(\eta)$. Veja que $L(\tau^{-1}(\cdot))$ é a função de máxima verossimilhança na variável η e, portanto, é a função de máxima verossimilhança para o parâmetro η. Queremos mostrar que $\hat{\eta} = \tau(\hat{\theta})$ é o estimador de máxima verossimilhança para η, isto é, queremos mostrar que $\hat{\eta} = \tau(\hat{\theta}) = \arg\max L(\tau^{-1}(\eta))$.

Suponha, por absurdo, que $\hat{\eta} = \tau(\hat{\theta})$ não seja o estimador de máxima verossimilhança para η. Então existe $\eta^\star \neq \tau(\hat{\theta})$ tal que

$$L(\tau^{-1}(\eta^\star)) > L(\tau^{-1}(\tau(\hat{\theta}))).$$

Como $L(\tau^{-1}(\tau(\hat{\theta}))) = L(\hat{\theta})$ concluímos que existe $\theta^\star = \tau^{-1}(\eta^\star) \neq \hat{\theta}$ tal que

$$L(\theta^\star) > L(\hat{\theta}),$$

o que é um absurdo, uma vez que, por hipótese, $\hat{\theta}$ é o estimador de máxima verossimilhança para θ. Então concluímos que $\tau(\hat{\theta})$ é o estimador de máxima verossimilhança para η.

□

O resultado do Teorema 3.1 só vem para facilitar a busca pelos estimadores de máxima verossimilhança. Na prática o que o teorema nos permite é encontrar o estimador de máxima verossimilhança para o parâmetro da distribuição e então usá-lo para encontrar o estimador de máxima verossimilhança para qualquer função desse parâmetro. A seguir alguns exemplos dessa aplicação.

Exemplo 3.13. *Seja $X \sim Ber(p)$. Encontre o estimador de máxima verossimilhança para* $\mathrm{Var}(X)$*.*

No Exercício 3.10 já encontramos o estimador de máxima verossimilhança para p,

$$\hat{p} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}.$$

Queremos agora o estimador de máxima verossimilhança para $\eta = \mathrm{Var}(X) = p(1-p) = \tau(p)$. Como η é função de p podemos aplicar a propriedade da invariância e concluir que o estimador de máxima verossimilhança para o parâmetro η é

$$\hat{\eta} = \tau(\hat{p}) = \bar{X}(1 - \bar{X}).$$

Exemplo 3.14. *Seja $X \sim N(\mu, \sigma^2)$. Encontre o estimador de máxima verossimilhança para σ.*

Já vimos no Exemplo 3.11 que o estimador de máxima verossimilhança para o parâmetro σ^2 é dado por

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}.$$

Queremos agora o estimador de máxima verossimilhança para o parâmetro $\sigma = \sqrt{\sigma^2} = \tau(\sigma^2)$. Como σ é função de σ^2 podemos aplicar a propriedade da invariância e concluir que o estimador de máxima verossimilhança para o parâmetro σ é

$$\hat{\sigma} = \tau(\hat{\sigma}^2) = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}.$$

Exemplo 3.15. *Suponha que o número de ocorrências de um certo evento em um determinado intervalo de tempo seja uma variável aleatória $X \sim Poisson(\lambda)$. Foram observados 50 intervalos de tempo distintos e constatou-se que em 20 deles o número de ocorrências foi nulo. Considere que os números de ocorrências do evento em intervalos distintos sejam variáveis aleatórias independentes. Como podemos estimar o número médio de ocorrências no determinado intervalo de tempo por máxima verossimilhança?*

Repare que queremos encontrar uma estimativa para o parâmetro $\mathrm{E}(X) = \lambda$ da população $X \sim Poisson(\lambda)$. Mas não temos uma amostra de X, apenas conhecemos o número de vezes *que não foi observado evento algum em 50 observações, ou seja, sabemos apenas quantas vezes observou-se* $X = 0$.

Veja que podemos construir a seguinte variável aleatória Y *como função de* X,

$$Y = \begin{cases} 1 & , X = 0 \\ 0 & , X \neq 0. \end{cases}$$

E essa é a população para a qual observou-se uma amostra aleatória, ou melhor, sabemos qual o valor de $\sum_{i=1}^{50} y_i = 20$ *para a amostra observada.*

Como $Y \sim Ber(p)$ *e o estimador por máxima verossimilhança para* p *é dado por*

$$\hat{p} = \bar{Y}$$

conseguimos encontrar uma estimativa para p *por máxima verossimilhança:*

$$\hat{p} = \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \quad \Rightarrow \quad \hat{p} = \frac{\sum_{i=1}^{50} y_i}{50} = \frac{20}{50} = 0,4.$$

Veja também que $p = \mathrm{P}(Y = 1) = \mathrm{P}(X = 0) = e^{-\lambda}$, *já que* $X \sim Poisson(\lambda)$. *Ou seja,* $p = e^{-\lambda}$ *e* $\lambda = -\ln(p)$. *Assim, aplicando a propriedade de invariância, podemos encontrar uma estimativa por máxima verossimilhança para o parâmetro* λ. *Uma vez que conseguimos escrever* λ *como função de* p *e temos uma estimativa por máxima verossimilhança para* p, *a estimativa por máxima verossimilhança para* λ *será:*

$$\hat{\lambda} = -\ln(\hat{p}) = -\ln(0,4) = 0,916.$$

Ou seja, estima-se que em média ocorrem $0,916$ *ocorrências por dia. Ou, em média, ocorrem um pouco menos de 1 ocorrência por dia.*

Em alguns casos existe a estimativa de máxima verossimilhança mas não é possível encontrar uma solução fechada para o estimador. Para esses casos podemos recorrer a métodos numéricos para encontrar a estimativa para uma dada amostra. Veja alguns exemplos.

Exemplo 3.16. *Suponha que a população* X *tenha densidade* f_X *definida abaixo, para* $-1 < \theta < 1$. *Encontre o estimador para* θ *pelo método da máxima verossimilhança.*

$$f_X(x) = \begin{cases} \frac{1}{2}(1 + \theta x) & , -1 \leq x \leq 1; \\ 0 & , \textit{caso contrário.} \end{cases}$$

Para encontrar o estimador de máxima verossimilhança para o parâmetro θ *precisa-se primeiro da função de verossimilhança ou função de log-verossimilhança.*

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{2}(1 + \theta x_i), \ -1 < \theta < 1 \qquad e \qquad l(\theta) = \ln\left(\prod_{i=1}^{n} \frac{1}{2}(1 + \theta x_i)\right), \ -1 < \theta < 1$$

Veja que

$$l(\theta) = \sum_{i=1}^{n} \ln\left(\frac{1}{2}(1 + \theta x_i)\right) = \sum_{i=1}^{n} \left(-\ln(2) + \ln(1 + \theta x_i)\right) = -n\ln(2) + \sum_{i=1}^{n} \ln(1 + \theta x_i).$$

Na busca pelos pontos críticos precisamos da derivada de l:

$$\frac{\mathrm{d}}{\mathrm{d}\theta} l(\theta) = \sum_{i=1}^{n} \frac{x_i}{1 + \theta x_i}.$$

Porém não é possível encontrar uma expressão para $\theta^\star$ *tal que* $\frac{\mathrm{d}}{\mathrm{d}\theta} l(\theta^\star) = 0$ *e por isso não encontramos os pontos críticos em função da amostra observada.*

Mas uma vez observada a amostra e conhecidos os valores de x_i *é possível aplicar métodos numéricos para encontrar as raízes da derivada. E caso essa derivada tenha raízes, isto é, caso exista* $\theta^\star$ *tal que* $\frac{\mathrm{d}}{\mathrm{d}\theta} l(\theta^\star) = 0$*, esse será um ponto de máximo uma vez que*

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} l(\theta) = -\sum_{i=1}^{n} \frac{x_i^2}{(1+\theta x_i)^2} < 0$$

qualquer que seja a amostra observada.

Exemplo 3.17. *Seja* $X \sim Gama(\alpha, \lambda)$*. Encontre o estimador de máxima verossimilhança para o vetor de parâmetros* (α, λ)*.*

Na busca pelo estimador de máxima verossimilhança para o vetor de parâmetros (α, λ) *precisa-se primeiro da função de verossimilhança ou função de log-verossimilhança, que para* $\alpha > 0$ *e* $\lambda > 0$ *são definida por*

$$L(\alpha, \lambda) = \prod_{i=1}^{n} \frac{\lambda^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\lambda x_i} \qquad e \qquad l(\theta) = \ln\left(\prod_{i=1}^{n} \frac{\lambda^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\lambda x_i}\right)$$

Vamos trabalhar com a função l *para buscar os pontos críticos.*

$$\begin{aligned} l(\alpha, \lambda) &= \ln\left(\prod_{i=1}^{n} \frac{\lambda^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\lambda x_i}\right) = \sum_{i=1}^{n} \ln\left(\frac{\lambda^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\lambda x_i}\right) \\ &= n\alpha \ln(\lambda) - n\ln(\Gamma(\alpha)) + (\alpha-1)\sum_{i=1}^{n} \ln(x_i) - \lambda \sum_{i=1}^{n} x_i \end{aligned}$$

O passo seguinte é encontrar as derivadas de primeira ordem em relação à cada uma das duas variáveis. Para isso, em algum momento, vai ser preciso encontrar a derivada da função Γ*, que vamos representar por* Γ'*, veja abaixo.*

$$\frac{\mathrm{d}}{\mathrm{d}\alpha} l(\alpha, \lambda) = n\ln(\lambda) - n\frac{1}{\Gamma(\alpha)}\Gamma'(\alpha) + \sum_{i=1}^{n} \ln(x_i) \qquad e \qquad \frac{\mathrm{d}}{\mathrm{d}\lambda} l(\alpha, \lambda) = n\alpha\frac{1}{\lambda} - \sum_{i=1}^{n} x_i$$

Igualando as duas derivadas a zero e buscando a solução do sistema encontramos

$$\lambda^\star = \frac{\alpha^\star}{\bar{x}} \qquad e \qquad n\ln(\lambda^\star) - n\frac{1}{\Gamma(\alpha^\star)}\Gamma'(\alpha^\star) + \sum_{i=1}^{n} \ln(x_i) = 0$$

sendo que essa última equação não tem solução, isto é, não conseguimos encontrar $\alpha^\star$ *escrito como função da amostra tal que* $(\alpha^\star, \lambda^\star)$ *seja solução do sistema. Mas novamente, uma vez observada a amostra, é possível encontrar a estimativa de máxima verossimilhança aplicando métodos numéricos para encontrar a solução do sistema para a amostra observada.*

Vale destacar que supondo $\alpha = \alpha_0$ *é um valor conhecido, existe o estimador de máxima verossimilhança para o parâmetro* λ *e ele será*

$$\hat{\lambda} = \frac{\alpha_0}{\bar{X}}.$$

Exemplo 3.18. *Suponha que a população* X *tenha densidade* f_X *definida abaixo. Encontre o estimador para* θ *pelo método da máxima verossimilhança.*

$$f_X(x) = \begin{cases} \dfrac{6x(\theta - x)}{\theta^3} & , 0 \leq x \leq \theta; \\ 0 & , \textit{caso contrário.} \end{cases}$$

Outra situação não comum, mas que pode acontecer com os estimadores de máxima verossimilhança, é a não unicidade dos estimadores. Ou seja, existem casos em que temos mais de um estimador de máxima verossimilhança para o mesmo parâmetro. Veja um exemplo.

Exemplo 3.19. *Seja* $X \sim U\left[\theta - \frac{1}{2}, \theta + \frac{1}{2}\right]$*. Encontre o estimador de máxima verossimilhança para* θ*.*

Veja que esse é mais um exemplo onde a imagem da variável aleatória depende do parâmetro. Nesse caso temos que ficar atentamos com a função de verossimilhança.

$$L(\theta) = \begin{cases} 1 & , \ \theta - \dfrac{1}{2} \leq x_i \leq \theta + \dfrac{1}{2} \\ 0 & , \ \textit{caso contrário.} \end{cases}$$

Podemos perceber que $x_i \leq \theta + \dfrac{1}{2} \ \ \forall\ i = 1, 2, \ldots, n \ \Rightarrow \ x_i - \dfrac{1}{2} \leq \theta \ \ \forall\ i = 1, 2, \ldots, n \ \Rightarrow \ \theta \geq x_{(n)} - \dfrac{1}{2}$*. De forma análoga,* $\theta - \dfrac{1}{2} \leq x_i \ \ \forall\ i = 1, 2, \ldots, n \ \Rightarrow \ \theta \leq x_i + \dfrac{1}{2} \leq \theta \ \ \forall\ i = 1, 2, \ldots, n \ \Rightarrow \theta \leq x_{(1)} + \dfrac{1}{2}$*. Assim fazendo a interseção dessas duas inequações é possível reescrever a função de verossimilhança como:*

$$L(\theta) = \begin{cases} 1 & , \ x_{(n)} - \dfrac{1}{2} \leq \theta \leq x_{(1)} + \dfrac{1}{2} \\ 0 & , \ \textit{caso contrário.} \end{cases}$$

Veja que L *é uma função constante na variável* θ *e por isso não tem uma máximo global. Por outro lado, todos os valores para* $\theta \in \left[x_{(n)} - \dfrac{1}{2}\ ,\ x_{(1)} + \dfrac{1}{2}\right]$ *são pontos de máximo local de* L *e podem ser um estimador de máxima verossimilhança. Por exemplo, o ponto médio desse intervalo* $\hat{\theta} = \dfrac{X_{(n)} + X_{(1)}}{2}$ *é o exemplo de estimador de máxima verossimilhança. Outro exemplo de estimador pode ser* $\hat{\theta} = \dfrac{1}{3}\left(X_{(n)} - \dfrac{1}{2}\right) + \dfrac{2}{3}\left(X_{(1)} + \dfrac{1}{2}\right)$*, que também pertence ao intervalo.*

Exercícios da Seção 3.3

3.3.1 ([Casella e Berger, 2002] - Capítulo 7)

Uma única observação é obtida de uma variável aleatória discreta X com função de probabilidade $p(.;\theta)$, com $\theta \in \psi = \{1, 2, 3\}$. O quadro ao lado apresenta as possíveis funções de probabilidades para cada $\theta \in \psi$. Encontre o estimador de máxima verossimilhança do parâmetro θ.

X	$p(x;1)$	$p(x;2)$	$p(x:3)$
0	1/3	1/4	0
1	1/3	1/4	0
2	0	1/4	1/4
3	1/6	1/4	1/2
4	1/6	0	1/4

3.3.2 Para os itens a seguir, considere $X_1, \ldots, X_n$ amostra aleatória de X. Encontre o estimador dos parâmetros distribucionais a partir do método da máxima verossimilhança. (a) $X \sim Geo^*(p)$ (b) $X \sim Bin(N_0, p)$, sendo N_0 conhecido (c) $X \sim U[-\theta, \theta]$ Atenção: Cuidado na resolução dos problemas de maximização. E também não esqueça de fazer o teste da 2ª derivada sempre que preciso.

3.3.3 ([Bolfarine e Sandoval, 2001] - Capítulo 3)
Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória da população $X \sim Exp(\lambda)$.

(a) Encontre o estimador para λ pelo método da máxima verossimilhança.

(b) Encontre o estimador para $\mu = \mathrm{E}(X)$ pelo método da máxima verossimilhança.

(c) Encontre o estimador para $\eta = \mathrm{P}(X > 1)$ pelo método da máxima verossimilhança.

(d) Os valores abaixo são a observação de uma amostra aleatória de $X \sim Exp(\lambda)$. Encontre as estimativas por máxima verossimilhança para os parâmetros λ, μ e η definidos acima.

$$0,081 \quad 0,637 \quad 0,183 \quad 0,743 \quad 0,207 \quad 0,318 \quad 0,096 \quad 0,389$$

3.3.4 ([Casella e Berger, 2002] - Capítulo 7 e [Bolfarine e Sandoval, 2001] - Capítulo 3) Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória da população X cuja função densidade é definida por:

$$f_X(x) = \theta x^{\theta-1}, \; 0 \leq x \leq 1 \;\text{ e } 0 \leq \theta \leq \infty.$$

(a) Encontre o estimador para θ pelo método dos momentos.

(b) Encontre o estimador para θ pelo método da máxima verossimilhança.

(c) Encontre o estimador para $\mu = \mathrm{E}(X)$ pelo método dos momentos.

(d) Encontre o estimador para $\mu = \mathrm{E}(X)$ pelo método da máxima verossimilhança.

3.3.5 ([Larson, 1982] - Capítulo 7) Assuma que o número de carros vendidos por dia por um negociante seja uma variável aleatória de Poisson com média μ. Encontre a estimativa por máxima verossimilhança para μ dado que:

(a) em 20 dias o negociante vendeu 30 carros;

(b) em 20 dos últimos 30 dias ele não fez venda alguma e nos outros 10 dias ele vendeu pelo menos um carro.

3.4 Propriedades dos Estimadores

Já vimos que qualquer estatística pode ser considerada um estimador para qualquer parâmetro. Mas como saber se um estimador é bom ou não? Será que os métodos estudados no Capítulo 3 fornece bons estimadores? Será que podemos criar outros melhores? O objetivo desse capítulo é apresentar as principais características desejadas de serem encontradas em um estimador, a fim de torná-lo uma boa opção para estimar um certo parâmetro.

Considere θ o parâmetro desconhecido da distribuição da população X. A fim de tornar os resultados apresentados mais gerais, vamos supor sempre que queremos estimar $\eta = \tau(\theta)$. Para o caso particular de querermos estimar o próprio θ basta considerar a função τ como a função identidade, isto é, $\tau(\theta) = \theta$.

3.4.1 Viés

A primeira característica esperada para um estimador é a ausência de viés, ou baixo viés. Antes de apresentar alguns exemplos vejamos a definição de viés para um estimador qualquer.

Definição 3.6. *Seja $\hat{\eta}$ um estimador para η. Dizemos que $\hat{\eta}$ é um estimador não tendencioso para η se $\mathrm{E}(\hat{\eta}) = \eta$. Caso contrário, $\hat{\eta}$ será um estimador tendencioso para η.*

A ideia é que um estimador não tendencioso é aquele que em média acerta o valor do parâmetro. Ou seja, se o estimador for usado muitas vezes com diferentes amostras, na situação assintótica de serem feitas infinitas estimativas, a média das estimativas fornecidas coincidiriam com o verdadeiro valor desconhecido do parâmetro.

Definição 3.7. *Seja $\hat{\eta}$ um estimador para η. O viés de $\hat{\eta}$, representado por $\mathrm{B}(\hat{\eta})$, é a diferença $\mathrm{B}(\hat{\eta}) = \mathrm{E}(\hat{\eta}) - \eta$.*

Veja que o viés de um estimador é um número real e que se $\hat{\eta}$ é não tendencioso temos $\mathrm{B}(\hat{\eta}) = 0$.

Exemplo 3.20. *Seja $X \sim Poisson(\lambda)$. Já vimos que tanto o Método dos Momentos quanto o Método da Máxima Verossimilhança resultam em*

$$\hat{\lambda} = \bar{X}.$$

Será que esse estimador é não tendencioso para o parâmetro λ?

Solução:

Para responder essa pergunta precisamos encontrar $\mathrm{E}(\hat{\lambda})$.

$$\mathrm{E}(\hat{\lambda}) = \mathrm{E}(\bar{X}) = \mathrm{E}\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{\sum_{i=1}^n \mathrm{E}(X_i)}{n} = \frac{\sum_{i=1}^n \mathrm{E}(X)}{n} = \frac{n\,\mathrm{E}(X)}{n} = \mathrm{E}(X) = \lambda.$$

Como $\mathrm{E}(\hat{\lambda}) = \lambda$ concluímos que o estimador é não tendencioso para o parâmetro λ. ||

Exemplo 3.21. *Seja X uma população qualquer. Verifique se $\bar{X}$, $\hat{\sigma}^2$ e S^2 são estimadores não tendenciosos para $\mathrm{E}(X)$, $\mathrm{Var}(X)$ e $\mathrm{Var}(X)$, respectivamente, independente da distribuição de X.*

Solução:

Do exercícios anteriores podemos tirar que $\mathrm{E}(\bar{X}) = \mathrm{E}(X)$ e portanto $\bar{X}$ sempre será não tendenciosos para $\mathrm{E}(X)$.

Vamos analisar agora $\mathrm{E}\left(\hat{\sigma}^2\right)$.

$$\begin{aligned}
\mathrm{E}\left(\hat{\sigma}^2\right) &= \mathrm{E}\left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}\right) = \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^n X_i^2 - \sum_{i=1}^n 2X_i\bar{X} + \sum_{i=1}^n \bar{X}^2\right) \\
&= \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^n X_i^2 - 2\bar{X}\sum_{i=1}^n X_i + n\bar{X}^2\right) = \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2\right) \\
&= \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right) = \frac{1}{n}\left(\mathrm{E}\left(\sum_{i=1}^n X_i^2\right) - \mathrm{E}\left(n\bar{X}^2\right)\right) = \frac{1}{n}\left(\sum_{i=1}^n \mathrm{E}\left(X_i^2\right) - n\,\mathrm{E}\left(\bar{X}^2\right)\right) \\
&= \frac{1}{n}\left(\sum_{i=1}^n \mathrm{E}\left(X^2\right) - n\,\mathrm{E}\left(\bar{X}^2\right)\right) = \frac{1}{n}\left(n\,\mathrm{E}\left(X^2\right) - n\,\mathrm{E}\left(\bar{X}^2\right)\right) = \mathrm{E}\left(X^2\right) - \mathrm{E}\left(\bar{X}^2\right) \\
&= \mathrm{Var}(X) + \mathrm{E}(X)^2 - \mathrm{Var}(\bar{X}) - \mathrm{E}(\bar{X})^2 = \mathrm{Var}(X) + \mathrm{E}(X)^2 - \frac{\mathrm{Var}(X)}{n} - \mathrm{E}(X)^2 \\
&= \mathrm{Var}(X) - \frac{\mathrm{Var}(X)}{n} = \frac{n-1}{n}\mathrm{Var}(X).
\end{aligned}$$

Como $\mathrm{E}(\hat{\sigma}^2) = \frac{n-1}{n}\mathrm{Var}(X)$, $\hat{\sigma}^2$ é um estimador tendencioso para $\mathrm{Var}(X)$.

Veja que $S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{n}{n-1}\hat{\sigma}^2$ então,

$$\mathrm{E}(S^2) = \frac{n}{n-1}\mathrm{E}(\hat{\sigma}^2) = \frac{n}{n-1}\frac{n-1}{n}\mathrm{Var}(X) = \mathrm{Var}(X).$$

Assim concluí-se que S^2 é um estimador não tendencioso para $\mathrm{Var}(X)$. ||

Mas será que $\hat{\sigma}^2$ é um estimador ruim para o parâmetro $\mathrm{Var}(X)$? Veremos na prática que exigir a não tendenciosidade pode ser de mais, e que muitos estimadores bons serão tendenciosos. Por isso é possível afrouxar um pouco essa exigência, o que motiva a Definição 3.8 a seguir.

Definição 3.8. *Seja $\hat{\eta}$ um estimador para η. Dizemos que $\hat{\eta}$ é um estimador assintoticamente não tendencioso para η se $\mathrm{E}(\hat{\eta}) \xrightarrow[n\to\infty]{} \eta$. Ou, o que é equivalente, se $\mathrm{B}(\hat{\eta}) \xrightarrow[n\to\infty]{} 0$.*

Exemplo 3.22. *Continuando o Exemplo 3.21 vamos mostrar que $\hat{\sigma}^2$ é um estimador assintoticamente não tendencioso para σ^2.*

Solução:

Já mostramos que $\mathrm{E}(\hat{\sigma}^2) = \dfrac{n-1}{n}\mathrm{Var}(X)$. Veja o que acontece quando $n \to \infty$:

$$\lim_{n\to\infty} \mathrm{E}(\hat{\sigma}^2) = \lim_{n\to\infty} \frac{n-1}{n}\mathrm{Var}(X) = \lim_{n\to\infty}\left(1 - \frac{1}{n}\right)\mathrm{Var}(X) = \lim_{n\to\infty}\mathrm{Var}(X) = \mathrm{Var}(X).$$

Ou seja, este é uma estimador assintoticamente não tendencioso para o parâmetro $\mathrm{Var}(X)$.

$\|$

Exemplo 3.23. *Seja $X \sim U[0,\theta]$. No Exemplo 3.12 foi encontrado o estimador pelo Método da Máxima Verossimilhança para o parâmetro θ: $\hat{\theta}_{MV} = X_{(n)}$. Podemos fazer as contas e encontrar o estimador pelo Método dos Momentos para θ: $\hat{\theta}_{MM} = 2\bar{X}$. Para cada um dos dois estimadores citados, verifique se ele é não tendencioso, assintoticamente não tendencioso ou tendencioso para θ.*

Solução:

Vamos começar pelo estimador pelo Método dos Momentos. Para fazer a verificação pedida precisa-se encontrar a esperança do estimador.

$$\mathrm{E}\left(\hat{\theta}_{MM}\right) = \mathrm{E}\left(2\bar{X}\right) = 2\,\mathrm{E}\left(\bar{X}\right) = 2\,\mathrm{E}\left(X\right) = 2\frac{\theta}{2} = \theta.$$

Logo, $\hat{\theta}_{MM}$ é um estimador não tendencioso para o parâmetro θ de $X \sim U[0,\theta]$.

Para encontrar $\mathrm{E}\left(\hat{\theta}_{MV}\right)$ precisamos conhecer a distribuição de $\hat{\theta}_{MV} = X_{(n)}$. Voltando ao Exemplo 2.2 podemos ver se $X \sim U[0,\theta]$ então a função densidade de $X_{(n)}$ é dada por

$$f_{X_{(n)}}(x) = \frac{n}{\theta^n}x^{n-1},\ 0 \leq x \leq \theta.$$

Agora já é possível encontrar a esperança de $\hat{\theta}_{MV}$.

$$\mathrm{E}\left(\hat{\theta}_{MV}\right) = \mathrm{E}\left(X_{(n)}\right) = \int_0^\theta x\frac{n}{\theta^n}x^{n-1}\,\mathrm{d}x = \frac{n}{\theta^n}\int_0^\theta x^n\,\mathrm{d}x = \frac{n}{\theta^n}\left.\frac{x^{n+1}}{n+1}\right|_0^\theta = \frac{n}{\theta^n}\frac{\theta^{n+1}}{n+1} = \frac{n}{n+1}\theta.$$

Como $\mathrm{E}\left(\hat{\theta}_{MV}\right) \neq \theta$ concluímos que o estimador não é não-tendencioso. Vejamos agora o limite quando $n \to \infty$.

$$\lim_{n\to\infty}\mathrm{E}\left(\hat{\theta}_{MV}\right) = \lim_{n\to\infty}\frac{n}{n+1}\theta = \lim_{n\to\infty}\frac{1}{1+1/n}\theta = \lim_{n\to\infty}\theta = \theta,$$

e assim verificamos que $\hat{\theta}_{MV}$ é um estimador assintoticamente não-tendencioso para o parâmetro θ de $X \sim U[0,\theta]$.

E será que esses resultados são suficientes para a gente concluir que $\hat{\theta}_{MM}$ é melhor que $\hat{\theta}_{MV}$?

$\|$

3.4.2 Erro Quadrático Médio

Outra característica esperada de um estimador é que ele tenha baixo valor para o Erro Quadrático Médio, definido a seguir.

Definição 3.9. *Seja $\hat{\eta}$ um estimador para η. O erro quadrático médio de $\hat{\eta}$, representado por* $\mathrm{EQM}(\hat{\eta})$*, é definido por:*

$$\mathrm{EQM}(\hat{\eta}) = \mathrm{E}\left((\hat{\eta} - \eta)^2\right).$$

Veja que o EQM é uma medida do quanto o estimador se afasta em média do parâmetro que ele pretende estimar. Essa medida vale para estimadores quaisquer, seja ele tendencioso ou não.

A Proposição 3.2 mostra que essa medida pode ser expressa pela soma da variância do estimador com o seu viés ao quadrado. Quanto menor a variância de um estimador, mais precisa será a sua estimativa, uma vez que ela se afasta pouco em relação ao valor médio estimado. Quando menor a variância de um estimador, menor será o seu EQM.

Já vimos que o viés de um estimador indica se ele retorna em média o valor real do parâmetro a ser estimado. Quanto menor o viés, mais perto do valor real do parâmetro são as estimativas, em média, melhor é a sua "mira". Esperamos encontrar estimadores com nenhum ou pequeno Viés. Quando menor é o viés de um estimador, menor será o seu EQM.

Dessa forma, informalmente, o EQM é uma medida que leva em consideração duas características dos estimadores: a precisão (dada pela variância) e a mira (dada pelo viés).

Proposição 3.2. $\mathrm{EQM}(\hat{\eta}) = \mathrm{Var}(\hat{\eta}) + \mathrm{B}^2(\hat{\eta})$.

Demonstração.

$$\begin{aligned}
\mathrm{EQM}(\hat{\eta}) &= \mathrm{E}\left((\hat{\eta} - \eta)^2\right) = \mathrm{E}\left(\hat{\eta}^2 - 2\hat{\eta}\eta + \eta^2\right) = \mathrm{E}\left(\hat{\eta}^2\right) - 2\eta\,\mathrm{E}\left(\hat{\eta}\right) + \eta^2 \\
&= \mathrm{E}\left(\hat{\eta}^2\right) - \mathrm{E}^2\left(\hat{\eta}\right) + \mathrm{E}^2\left(\hat{\eta}\right) - 2\eta\,\mathrm{E}\left(\hat{\eta}\right) + \eta^2 = \mathrm{Var}\left(\hat{\eta}\right) + \mathrm{E}^2\left(\hat{\eta}\right) - 2\eta\,\mathrm{E}\left(\hat{\eta}\right) + \eta^2 \\
&= \mathrm{Var}\left(\hat{\eta}\right) + \left(\mathrm{E}\left(\hat{\eta}\right) - \eta\right)^2 = \mathrm{Var}\left(\hat{\eta}\right) + \mathrm{B}^2(\hat{\eta}).
\end{aligned}$$

$\square$

Veja que se $\hat{\eta}$ é um estimador não-tendencioso, $\mathrm{EQM}(\hat{\eta}) = \mathrm{Var}(\hat{\eta})$. Além, disso o EQM acaba sendo uma medida de comparação entre estimadores.

Definição 3.10. *Sejam $\hat{\eta}_1$ e $\hat{\eta}_2$ dois estimadores para η. Dizemos que $\hat{\eta}_1$ é mais eficiente que $\hat{\eta}_2$ se* $\mathrm{EQM}(\hat{\eta}_1) \leq \mathrm{EQM}(\hat{\eta}_2)$.

Exemplo 3.24. *Seja $X \sim U[0, \theta]$. Sejam $\hat{\theta}_{MV} = X_{(n)}$ e $\hat{\theta}_{MM} = 2\bar{X}$ os estimadores para θ pelo Método da Máxima Verossimilhança e pelo Método dos Momentos, respectivamente. Encontre o Erro Quadrático Médio de cada um deles e compare os dois estimadores.*

Solução:

Vamos começar pelo $\hat{\theta}_{MM}$. Já vimos no Exemplo 3.23 que $\hat{\theta}_{MM}$ é um estimador não-tendencioso para o parâmetro θ, ou seja, $\mathrm{B}(\hat{\theta}_{MM}) = 0$. Então

$$\mathrm{EQM}(\hat{\theta}_{MM}) = \mathrm{Var}(\hat{\theta}_{MM}) = \mathrm{Var}(2\bar{X}) = 4\,\mathrm{Var}(\bar{X}) = 4\frac{\mathrm{Var}(X)}{n} = 4\frac{\theta^2}{12n} = \frac{\theta^2}{3n}.$$

Queremos encontrar agora $\mathrm{EQM}(\hat{\theta}_{MV}) = \mathrm{Var}(\hat{\theta}_{MV}) + \mathrm{B}^2(\hat{\theta}_{MV})$, sendo $\hat{\theta}_{MV} = X_{(n)}$. Já vimos que $\mathrm{E}\left(X_{(n)}\right) = \dfrac{n}{n+1}\theta$ (Exemplo 3.23), então

$$\mathrm{B}^2(\hat{\theta}_{MV}) = \left(\mathrm{E}(\hat{\theta}_{MV}) - \theta\right)^2 = \left(\frac{n}{n+1}\theta - \theta\right)^2 = \theta^2\left(\frac{n-n-1}{n+1}\right)^2 = \frac{\theta^2}{(n+1)^2}.$$

Precisamos agora encontrar $\mathrm{Var}(\hat{\theta}_{MV}) = \mathrm{Var}(X_{(n)})$, e para isso é necessário conhecer a distribuição amostral da estatística $X_{(n)}$. Voltando ao Exemplo 2.2 temos $f_{X_{(n)}}(x) = \dfrac{n}{\theta^n}x^{n-1}$, para $0 \leq x \leq \theta$. Como já se conhecer $\mathrm{E}(X_{(n)})$, vamos encontrar o segundo momento.

$$\mathrm{E}\left(X_{(n)}^2\right) = \int_0^\theta x^2\frac{n}{\theta^n}x^{n-1}\,\mathrm{d}x = \frac{n}{\theta^n}\int_0^\theta x^{n+1}\,\mathrm{d}x = \frac{n}{\theta^n}\left.\frac{x^{n+2}}{n+2}\right|_0^\theta = \frac{n}{\theta^n}\frac{\theta^{n+2}}{n+2} = \frac{n}{n+2}\theta^2.$$

Agora já é possível encontrar $\text{Var}(\hat{\theta}_{MV})$:

$$\text{Var}(\hat{\theta}_{MV}) = \text{E}\left(X_{(n)}^2\right) - \text{E}^2\left(X_{(n)}\right) = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 = n\theta^2\frac{(n+1)^2 - n(n+2)}{(n+2)(n+1)^2}$$
$$= n\theta^2\frac{n^2+2n+1-n^2-2n}{(n+2)(n+1)^2} = \frac{n\theta^2}{(n+2)(n+1)^2}.$$

Então.

$$\text{EQM}(\hat{\theta}_{MV}) = \text{Var}(\hat{\theta}_{MV}) + \text{B}^2(\hat{\theta}_{MV}) = \frac{n\theta^2}{(n+2)(n+1)^2} + \frac{\theta^2}{(n+1)^2} = \frac{n+(n+2)}{(n+2)(n+1)^2}\theta^2$$
$$= \frac{2n+2}{(n+2)(n+1)^2}\theta^2 = \frac{2(n+1)}{(n+2)(n+1)^2}\theta^2 = \frac{2}{(n+2)(n+1)}\theta^2.$$

Para comparar o EQM dois estimadores vamos comparar

$$\text{EQM}(\hat{\theta}_{MM}) = \frac{\theta^2}{3n} \quad \text{e} \quad \text{EQM}(\hat{\theta}_{MV}) = \frac{2}{(n+2)(n+1)}\theta^2$$

para diferentes valores de n. Veja que o denominador de $\text{EQM}(\hat{\theta}_{MV})$ é um polinômio de grau 2 enquanto o denominador de $\text{EQM}(\hat{\theta}_{MM})$ é um polinômio de grau 1, ambos com coeficientes positivos. Isso indica que a partir de um certo valor de n, isto é, a partir de um certo tamanho de amostra, $\hat{\theta}_{MV}$ é mais eficiente que $\hat{\theta}_{MM}$. De fato isso acontece para amostrar com $n \geq 3$.

$\|$

Seja $X \sim N(\mu, \sigma^2)$. Já vimos que

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

é o estimador para σ^2 tanto pelo Método dos Momentos (Exemplo 3.6) quanto pelo Método da Máxima Verossimilhança (Exemplo 3.11). É muito comum também usar o estimador

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

para esse parâmetro. Vimos também que $\hat{\sigma}^2$ é tendencioso para σ^2, embora seja assintoticamente não-tendencioso, e S^2 é não-tendencioso (Exemplo 3.21). Mas será que S^2 é tão melhor que $\hat{\sigma}^2$?

Exemplo 3.25. *Encontre o Erro Quadrático Médio de $\hat{\sigma}^2$ e S^2 e compare os resultados.*

Solução:

Para encontrar o *EQM* de cada estimador precisamos do Viés e da Variância de cada um deles.

$$B(\hat{\sigma}^2) = \text{E}(\hat{\sigma}^2) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = \sigma^2\left(\frac{n-1}{n} - 1\right) = -\frac{\sigma^2}{n}.$$

Para encontrar a variância vamos lembrar do resultado do Corolário 2.6 onde vimos que $\hat{\sigma}^2 \sim Gama\left(\frac{n-1}{2}, \frac{n}{2\sigma^2}\right)$. Assim,

$$\text{Var}(\hat{\sigma}^2) = \frac{(n-1)/2}{(n/2\sigma^2)^2} = \frac{n-1}{2}\frac{4(\sigma^2)^2}{n^2} = 2\frac{(n-1)}{n^2}(\sigma^2)^2.$$

$$\text{EQM}(\hat{\sigma}^2) = \text{Var}(\hat{\sigma}^2) + \text{B}^2(\hat{\sigma}^2) = 2\frac{(n-1)}{n^2}(\sigma^2)^2 + \left(-\frac{\sigma^2}{n}\right)^2$$
$$= 2\frac{(n-1)}{n^2}(\sigma^2)^2 + \frac{(\sigma^2)^2}{n^2} = \left(\frac{2n-2+1}{n^2}\right)(\sigma^2)^2 = \left(\frac{2n-1}{n^2}\right)(\sigma^2)^2$$

Vamos fazer as contas agora para o S^2. Como S^2 é não-tendencioso $\mathrm{B}(S^2) = 0$ e $\mathrm{EQM}(S^2) = \mathrm{Var}(S^2)$. A $\mathrm{Var}(S^2)$ é tirada também do Corolário 2.6 onde vimos que $S^2 \sim Gama\left(\frac{n-1}{2}, \frac{n-1}{2\sigma^2}\right)$. Assim,

$$\mathrm{EQM}(S^2) = \mathrm{Var}(S^2) = \frac{(n-1)/2}{((n-1)/(2\sigma^2))^2} = \frac{n-1}{2}\frac{(2\sigma^2)^2}{(n-1)^2} = \frac{2}{n-1}(\sigma^2)^2.$$

Fazendo as contas podemos ver que $\frac{2n-1}{n^2} < \frac{2}{n-1}$ para todo $n \geq 2$. Com isso concluímos que $\mathrm{EQM}(\hat{\sigma}^2) < \mathrm{EQM}(S^2)$ sempre que o tamanho da amostra n for tal que $n \geq 2$, ou seja, para $n \geq 2$ $\hat{\sigma}^2$ é mais eficiente que S^2. $\|$

Esse é um outro exemplo onde um estimador tendencioso ($\hat{\sigma}^2$) apresenta um erro quadrático médio menor que o de um estimador não-tendencioso (S^2). E agora, continuamos preferindo S^2 quando comparado com $\hat{\sigma}^2$? Na prática, para amostras não muito pequenas, esses dois estimadores fornecem estimativas muito parecidas e são dois bons estimadores para o parâmetro σ^2 de uma população $N(0,1)$.

3.4.3 Consistência

Já vimos que a ideia intuitiva de um estimador não-tendencioso é aquele que em média acerta o valor do parâmetro. Já a ideia intuitiva de um estimador com baixo valor de EQM é aquele que além de ter uma boa "mira", pois tem viés baixo, apresenta precisão na estimativa, uma vez que a variância também deve ser baixa.

Agora veremos outra característica que esperamos encontrar nos estimadores, a consistência. Como apresenta a Definição 3.11, a ideia intuitiva é que esperamos encontrar estimadores que fiquem mais precisos conforme o tamanho da amostra cresce.

Definição 3.11. *Seja $\hat{\eta}$ um estimador para η. Dizemos que $\hat{\eta}$ é um estimador consistente em erro quadrático médio se* $\mathrm{EQM}(\hat{\eta}) \xrightarrow[n\to\infty]{} 0$.

Jé foi visto que $\mathrm{EQM}(\hat{\eta}) = \mathrm{Var}(\hat{\eta}) + \mathrm{B}^2(\hat{\eta})$. Então se $\hat{\eta}$ é um eficiente em EQM,

$$\mathrm{EQM}(\hat{\eta}) \xrightarrow[n\to\infty]{} 0 \quad \Rightarrow \quad \mathrm{Var}(\hat{\eta}) + \mathrm{B}^2(\hat{\eta}) \xrightarrow[n\to\infty]{} 0.$$

Como $\mathrm{Var}(\hat{\eta}) > 0$ e $\mathrm{B}^2(\hat{\eta}) > 0$ podemos concluir que tanto $\mathrm{Var}(\hat{\eta}) \xrightarrow[n\to\infty]{} 0$ quanto $\mathrm{B}^2(\hat{\eta}) \xrightarrow[n\to\infty]{} 0$. Concluímos então que $\mathrm{B}(\hat{\eta}) \xrightarrow[n\to\infty]{} 0$, já que a função raiz quadrada é contínua. Logo $\mathrm{E}(\hat{\eta}) \xrightarrow[n\to\infty]{} \eta$.

Veja que vale o caminho contrário. Seja $\hat{\eta}$ um estimador tal que $\mathrm{E}(\hat{\eta}) \xrightarrow[n\to\infty]{} \eta$ e $\mathrm{Var}(\hat{\eta}) \xrightarrow[n\to\infty]{} 0$. Como $\mathrm{E}(\hat{\eta}) \xrightarrow[n\to\infty]{} \eta$, $\mathrm{B}(\hat{\eta}) \xrightarrow[n\to\infty]{} 0$. Então $\mathrm{Var}(\hat{\eta}) + \mathrm{B}^2(\hat{\eta}) \xrightarrow[n\to\infty]{} 0$ e, consequente, $\mathrm{EQM}(\hat{\eta}) \xrightarrow[n\to\infty]{} 0$. Logo $\hat{\eta}$ é um estimador consistente em EQM.

Com isso acabamos demonstrando a Proposição 3.3 a seguir.

Proposição 3.3. *Seja $\hat{\eta}$ um estimador para η. O estimador $\hat{\eta}$ é consistente em erro quadrático médio se, e somente se,* $\mathrm{E}(\hat{\eta}) \xrightarrow[n\to\infty]{} \eta$ *e* $\mathrm{Var}(\hat{\eta}) \xrightarrow[n\to\infty]{} 0$.

Perceba que a consistência é um conceito assintótico, que diz respeito a convergência do estimador. Como em probabilidade existem diferentes formas de convergência de variáveis aleatórias, que não serão abordadas nesse curso, existem também diferentes formas de consistência. Nesse texto veremos mais uma, a consistência fraca. Mais ainda existe, por exemplo, a consistência forte, que não trabalharemos aqui.

Definição 3.12. *Seja $\hat{\eta}$ um estimador para η. Dizemos que $\hat{\eta}$ é um estimador fracamente consistente se* $\mathrm{P}(|\hat{\eta} - \eta| \geq \varepsilon) \xrightarrow[n\to\infty]{} 0$.

Um estimador é fracamente consistente se ele converge em probabilidade para o parâmetro. Para algumas demonstrações de convergência fraca podem ser úteis a desigualdade de Markov, apresentada na Proposição 3.4, ou a versão fraca da Lei dos Grandes Números, apresentada no Teorema 3.2.

Proposição 3.4. ***Desigualdade de Markov***
Seja X uma variável aleatória qualquer. Então, para quaisquer t, $k > 0$, temos

$$\mathrm{P}(|X| \geq t) \leq \frac{\mathrm{E}(X^k)}{t^k}.$$

Para uma demonstração da Desigualdade de Markov, também conhecida como Desigualdade Básica de Chebyshev, veja o livro [Magalhães, 2011].

Teorema 3.2. ***Lei Fraca dos Grandes Números***
Seja $X_1, X_2, \ldots, X_n, \ldots$ uma sequência de variáveis aleatórias independentes tal que $\mathrm{E}(X_i) = \mu < \infty$ *e* $\mathrm{Var}(X_i) = \sigma^2 < \infty$. *Então, se $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$,*

$$\lim_{n\to\infty} \mathrm{P}\left(|\bar{X}_n - \mu| \geq \varepsilon\right) = 0,$$

isto é, $\bar{X}_n$ converge em probabilidade para μ quando $n \to \infty$.

Para uma demonstração da versão fraca da Lei dos Grandes Números veja o livro [Larson, 1982].

Antes de apresentarmos alguns exemplos, vamos estabelecer uma relação entre os dois tipos de consistência apresentados acima.

Proposição 3.5. *Seja $\hat{\eta}$ um estimador para η. Se $\hat{\eta}$ é consistente em erro quadrático médio então $\hat{\eta}$ também é fracamente consiste.*

Demonstração. Seja $\hat{\eta}$ um estimador consistente em erro quadrático médio. Então é verdade que $\mathrm{EQM}(\hat{\eta}) \xrightarrow[n\to\infty]{} 0$. Considere $X = |\hat{\eta} - \eta|$ e $k = 2$ na Desigualdade de Markov,

$$\mathrm{P}(|\hat{\eta} - \eta| \geq t) \leq \frac{\mathrm{E}(|\hat{\eta} - \eta|^2)}{t^2} = \frac{\mathrm{EQM}(\hat{\eta})}{t^2} \xrightarrow[n\to\infty]{} 0.$$

Logo, $\hat{\eta}$ um estimador fracamente consistente para o parâmetro η. □

Dessa forma podemos entender que a consistência em EQM é mais forte que a consistência fraca, por isso ela se chama "fraca". Se um estimador é consistente em EQM ele também será fracamente consistente. Mas a volta não é verdade. Existem estimadores fracamente consistentes que não são consistentes em EQM.

Exemplo 3.26. *Mostre que $\bar{X}$ é consistente em EQM para* $\mathrm{E}(X)$ *qualquer que seja X.*

Solução:
Primeiro veja que $\mathrm{E}(\bar{X}) = \mathrm{E}(X) \;\Rightarrow\; \mathrm{B}(\bar{X}) = 0$. Em seguida veja que $\mathrm{Var}(\bar{X}) \xrightarrow[n\to\infty]{} 0$:

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left(X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left(X\right)$$
$$= \frac{1}{n^2} n\,\mathrm{Var}\left(X\right) = \frac{1}{n}\mathrm{Var}\left(X\right) \xrightarrow[n\to\infty]{} 0.$$

Então, de acordo com a Proposição 3.3, $\bar{X}$ é consistente em EQM para o parâmetro $\mathrm{E}(X)$ qualquer que seja a população X. ‖

Exemplo 3.27. *Mostre que* $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$ *é fracamente consistente para* $\mathrm{Var}(X)$ *qualquer que seja* X.

Solução:
Primeiro veja que,

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n}\left(\sum_{i=1}^n X_i^2 - \sum_{i=1}^n 2X_i\bar{X} + \sum_{i=1}^n \bar{X}^2\right)\\
&= \frac{1}{n}\left(\sum_{i=1}^n X_i^2 - 2\bar{X}\sum_{i=1}^n X_i + n\bar{X}^2\right) = \frac{1}{n}\left(\sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2\right) = \frac{1}{n}\left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right)\\
&= \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2.
\end{aligned}$$

Aplicando a Lei Fraca dos Grandes Números (Teorema 3.2) na amostra de X podemos concluir que $\bar{X}$ converge fracamente para $\mathrm{E}(X)$. Como a convergência em probabilidade é preservada por funções contínuas, $-\bar{X}^2$ converge fracamente para $-\mathrm{E}^2(X)$.

Aplicando novamente a Lei Fraca dos Grandes Números, agora na amostra de $Y = X^2$, isto é, para $Y_i = X_i^2$ podemos concluir que $\bar{Y} = \dfrac{\sum_{i=1}^n X_i^2}{n}$ converge fracamente para $\mathrm{E}(Y) = \mathrm{E}(X^2)$.

Como a convergência em probabilidade é preservada pela soma, $\dfrac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2$ converge fracamente para $\mathrm{E}(X^2) - \mathrm{E}^2(X) = \mathrm{Var}(X)$. Logo, $\hat{\sigma}^2$ é fracamente consistente qualquer que seja a população X. $\|$

Exemplo 3.28. *Mostre que* $\hat{\sigma}^2$ *e* S^2 *são consistentes em EQM para o parâmetro* σ^2 *de* $X \sim N(\mu, \sigma^2)$.

Solução:
Quando $X \sim N(\mu, \sigma^2)$ conhecemos a distribuição amostral dos estimadores $\hat{\sigma}^2$ e S^2 e por isso é fácil encontrar a sua variância. Reveja o Exemplo 3.25 de onde tiramos os resultados a seguir.

$$\mathrm{EQM}(\hat{\sigma}^2) = \left(\frac{2n-1}{n^2}\right)(\sigma^2)^2\ , \quad \mathrm{EQM}(S^2) = \frac{2}{n-1}(\sigma^2)^2.$$

Veja que

$$\begin{aligned}
\mathrm{EQM}(\hat{\sigma}^2) &= \left(\frac{2n-1}{n^2}\right)(\sigma^2)^2 = \left(\frac{2}{n} - \frac{1}{n^2}\right)(\sigma^2)^2 \xrightarrow[n\to\infty]{} 0\\
\mathrm{EQM}(S^2) &= \frac{2}{n-1}(\sigma^2)^2 \xrightarrow[n\to\infty]{} 0
\end{aligned}$$

Assim, de acordo com a Definição 3.11, $\hat{\sigma}^2$ e S^2 são estimadores consistentes em EQM para o parâmetro σ^2 de $X \sim N(\mu, \sigma^2)$. $\|$

Exemplo 3.29. *Seja* $X \sim U[0, \theta]$. *Mostre que tanto* $\hat{\theta}_{MM} = 2\bar{X}$ *quanto* $\hat{\theta}_{MV} = X_{(n)}$ *são consistentes em EQM para o parâmetro* θ.

Solução:
Reveja o Exemplo 3.24 de onde tiramos os resultados a seguir.

$$\mathrm{EQM}(\hat{\theta}_{MM}) = \frac{\theta^2}{3n} \quad \text{e} \quad \mathrm{EQM}(\hat{\theta}_{MV}) = \frac{2}{(n+2)(n+1)}\theta^2.$$

Como $\mathrm{EQM}(\hat{\theta}_{MM}) \xrightarrow[n\to\infty]{} 0$ e $\mathrm{EQM}(\hat{\theta}_{MV}) \xrightarrow[n\to\infty]{} 0$, ambos os estimadores são consistentes em EQM para o parâmetro θ de $X \sim U[0, \theta]$. $\|$

Exercícios da Seção 3.4

3.4.1 Considere uma amostra aleatória $X_1, X_2, \ldots, X_n$ de $X \sim Ber(p)$. Mostre que o estimador de máxima verossimilhança para a variância de X é:

(a) tendencioso;

(b) assintoticamente não-tendencioso.

3.4.2 Considere uma amostra aleatória $X_1, X_2, \ldots, X_n$ de $X \sim N(\mu, \sigma^2)$ e o correspondente estimador de máxima verossimilhança para μ^2. Encontre o viés deste estimador e diga se ele é não-tendencioso ou assintoticamente não-tendencioso para μ^2?

3.4.3 Seja X_1, ..., X_n amostra aleatória de X com função densidade $f_X(x) = \theta x^{\theta-1}$, $0 < x < 1$ e $\theta > 0$. Seja $\hat{\theta}_{MV} = -n/\sum_{i=1}^n \ln(X_i)$ o estimador de máxima verossimilhança para θ, encontrado no Exercício 3.3.4(b).

(a) Encontre o viés de $\hat{\theta}_{MV}$ e diga se esse estimador é não-tendencioso, assintoticamente não-tendencioso ou tendencioso. (Dica: reveja o resultado do Exercício 2.1.11)

(b) Encontre a variância de $\hat{\theta}_{MV}$ e diga se esse estimador é consistente em EQM.

3.4.4 ([Larson, 1982] - Capítulo 7) Seja $\{X_1, X_2\}$ uma amostra aleatória de tamanho 2 de $X \sim Exp(\lambda)$. Sejam $T_1 = \sqrt{X_1 X_2}$ e $T_2 = \frac{X_1+X_2}{2}$ estimadores para a média de X. Mostre que T_1 é mais eficiente que T_2. Dica: Você vai precisar encontrar $\mathrm{E}(\sqrt{X})$.

3.4.5 ([Larson, 1982] - Capítulo 7 - modificado) Seja $X_1, \ldots, X_n$ uma amostra aleatória de $X \sim Ber(p)$. Seja $T = \frac{\sum_{i=1}^n X_i}{n+1}$ um estimador para p.

(a) Encontre o viés de T.

(b) T é um estimador não-tendencioso?

(c) T é um estimador assintoticamente não-tendencioso?

(d) Calcule o erro médio quadrático de T.

(e) T é consistente em EQM?

3.4.6 ([Larson, 1982] - Capítulo 7) Seja $X_1, \ldots, X_n$ uma amostra aleatória de $X \sim Exp(\lambda)$. Encontre o valor de a que minimiza $EQM(a\bar{X})$, considerando $a\bar{X}$ um estimador para $\mathrm{E}(X)$.

3.5 Mais Alguns Exercícios do Capítulo 3

3.1. ([Larson, 1982] - Capítulo 7) Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $X \sim N(\mu_0, \sigma^2)$, com μ_0 conhecido.

(a) Encontre o estimador para σ^2 pelo método dos momentos.

(b) Encontre o estimador para σ^2 pelo método da máxima verossimilhança.

(c) Os valores abaixo são a observação de uma amostra aleatória de $X \sim N(\mu_0 = 5, \sigma^2)$. Encontre as estimativas para σ^2 pelos métodos vistos nos itens (a) e (b).

$$6,22 \quad 6,62 \quad 2,57 \quad 3,85 \quad 3,06 \quad 3,96 \quad 6,95 \quad 4,42 \quad 6,25 \quad 7,27 \quad 3,23 \quad 4,87$$

(d) Repita o item 3.1c agora para a nova amostra de $X \sim N(\mu_0 = 5, \sigma^2)$, apresentada abaixo. O resultado é o que você esperava? Comente.

$$2,36 \quad 3,94 \quad 3,34 \quad 4,48 \quad 1,94 \quad 5,46 \quad 3,32 \quad 4,14 \quad 4,32 \quad 4,83 \quad 7,02 \quad 3,08$$

3.2. Seja $X_1, \ldots, X_n$ uma amostra aleatória da população $X \sim Poisson(\lambda)$. Seja $\eta = \mathrm{E}(X^2)$.

(a) Encontre o estimador para η pelo método dos momentos, vamos chamá-lo de $\hat{\eta}_{MM}$.

(b) Encontre o estimador para η pelo método da máxima verossimilhança, vamos chamá-lo de $\hat{\eta}_{MV}$.

Considere que os seguintes valores representam a observação de uma amostra de tamanho 6 de X: 1 3 3 1 3 0.

(c) Encontre as estimativas para $\hat{\eta}_{MM}$ e $\hat{\eta}_{MV}$.

3.3. Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória da população $X \sim Exp(\lambda)$.

(a) Encontre o estimador para λ pelo método dos momentos.

(b) Encontre o estimador para $\nu = \mathrm{Var}(X)$ pelo método dos momentos.

(c) Encontre o estimador para λ pelo método da máxima verossimilhança.

(d) Encontre o estimador para $\nu = \mathrm{Var}(X)$ pelo método da máxima verossimilhança.

3.4. ([Casella e Berger, 2002] - Capítulo 7 e [Bolfarine e Sandoval, 2001] - Capítulo 3)
Sejam $X_1, X_2, \ldots, X_n$ amostra aleatória de X com densidade descrita abaixo, sendo $\theta > 0$.

$$f_X(x) = \begin{cases} \frac{\theta}{x^2} & , \ x \geq \theta \\ 0 & , \ \text{caso contrário.} \end{cases}$$

(a) Mostre que não existe estimador para θ pelo método dos momentos.
Dica: mostre que não existe nenhum momento de X.

(b) Encontre o estimador para θ pelo método da máxima verossimilhança.

(c) Encontre o estimador para $\eta = \mathrm{E}(1/X)$ pelo método da máxima verossimilhança.

3.5. ([Larson, 1982] - Capítulo 7) Considere que o tempo de vida de um componente elétrico seja uma variável aleatória exponencial com o parâmetro λ. Dez destes componentes passaram pelo seguinte teste: os componentes foram ligados e após 100 horas foram observados quais permaneciam em funcionamento e quais falharam. Ou seja, o único registro de dados foi o número de componentes que falharam em menos de 100 horas versus o número que não falhou. Verificou-se que três falharam antes de 100 horas e os outros sete não. Considere que o teste garantiu independência entre os tempos de vidas de componentes diferentes. Encontre a estimativa por máxima verossimilhança para:

(a) O parâmetro λ. (b) O tempo médio de vida desse componente.

3.6. O número de acidentes por semana em trecho de um rodovia pode ser considerado uma variável aleatória com distribuição de Poisson. O Inspetor da Polícia Rodoviária responsável pela rodovia contratou você para um estudo. Ele gostaria de uma estimativa para a probabilidade de ocorrer nenhum acidente em uma semana. Você então observou os registros históricos do números de acidentes por semana das últimas 12 semanas, que apresentaram os seguintes valores:

1 3 3 1 5 5 1 4 2 3 3 1

(a) Baseados nas informações acima, apresente a amostra aleatória e a população (distribuição) que vai ser considerada nas suas análises. Apresente também o tamanho da amostra adotada e o(s) parâmetro(s) desconhecido(s). Por fim, apresente o parâmetro de interesse, aquele que é seu objetivo estimar.

(b) Considerando a amostra aleatória descrita no item anterior, apresente o estimador de máxima verossimilhança para o parâmetro de interesse.

(c) Para a amostra observada encontre a estimativa para o parâmetro de interesse.

(d) Formule uma frase resposta para ser entregue ao setor de controle de qualidade da fábrica.

3.7. O setor de controle de qualidade de uma fábrica contratou você para estimar a probabilidade de um lote com 12 produtos conter no máximo um produto defeituoso. Para fazer esse cálculo um funcionário foi diariamente para a linha de produção checar produtos de forma sequencial, até encontrar um com defeito. Essa atividade foi feita durante 25 dias, de forma independente.

(a) Baseados nas informações acima, apresente a amostra aleatória e a população (distribuição) que vai ser considerada nas suas análises. Apresente também o tamanho da amostra adotada e o(s) parâmetro(s) desconhecido(s). Por fim, apresente o parâmetro de interesse, aquele que é seu objetivo estimar.

(b) Considerando a amostra aleatória descrita no item anterior, apresente o estimador de máxima verossimilhança para o parâmetro de interesse.

(c) Os resultados encontrados pelo funcionário para o número de produtos inspecionados até que se encontre o primeiro com defeito foram:

14 6 8 53 17 27 25 29 16 8 14 3 5 8 2 48 32 15 54 12 24 6 22 6 30

Para essa amostra observada encontre a estimativa para o parâmetro de interesse.

(d) Formule uma frase resposta para ser entregue ao setor de controle de qualidade da fábrica.

3.8. Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $X \sim Poisson(\lambda)$ e $\eta = \tau(\lambda) = \lambda(\lambda + 1)$.

(a) Encontre os estimadores $\hat{\eta}_{MM}$ e $\hat{\eta}_{MV}$.

(b) Mostre que $\hat{\eta}_{MM}$ é não-tendencioso para η.

(c) Mostre que $\hat{\eta}_{MV}$ é assintoticamente não-tendencioso para η.

3.9. ([Larson, 1982] - Capítulo 7 - modificado) Seja X uma variável aleatória cuja densidade é definida por: $f(x) = e^{a-x},\ x \geq a$.

(a) Encontre o estimador de a pelo método dos momentos.

(b) Encontre o estimador de a pelo método da máxima verossimilhança.

(c) Determine se esses dois estimadores são ou não-tendenciosos.

(d) Qual deles é mais eficiente?

3.10. Seja X uma variável aleatória contínua cuja função densidade depende do parâmetro $\theta < 5$ e é definida por

$$f_X(x) = \frac{1}{5-\theta}\ , \quad \theta < x < 5.$$

Veja que $X \sim U(\theta, 5)$ e que esta distribuição é a mesma do Exercício 2.2.

(a) Encontre o estimador para θ pelo método dos momentos.

(b) Verifique se o estimador para θ pelo método dos momentos é consistente em EQM.

(c) Encontre o estimador para θ pelo método da máxima verissimilhança.

(d) Verifique se o estimador para θ pelo método da máxima verissimilhança é não-tendencioso, assintoticamente não-tendencioso ou tendencioso para θ.

(e) Considere os valores abaixo uma amostra da população X. Encontre as estimativas para o parâmetro θ tanto pelo método dos momentos quanto pelo método da máxima verossimilhança.

$$2,06 \quad 4,35 \quad 3,04 \quad 4,96 \quad 2,37 \quad 2,43 \quad 4,75 \quad 2,75 \quad 3,28 \quad 3,79 \quad 2,64 \quad 3,89$$

3.11. Seja X uma variável aleatória contínua cuja função densidade depende do parâmetro $\theta > 0$ e é definida por

$$f_X(x) = \frac{\theta 4^\theta}{x^{\theta+1}} \;, \quad x > 4.$$

Veja que $X \sim Pareto(\theta, 4)$ e que esta distribuição é a mesma do Exercício 2.5.1.

(a) Encontre o estimador para θ pelo método dos momentos.

(b) Encontre o estimador para θ pelo método da máxima verissimilhança.

(c) Verifique se o estimador para θ pelo método da máxima verissimilhança é não-tendencioso, assintoticamente não-tendencioso ou tendencioso para θ. Dica: no Exercício 2.1 foi encontrado a distribuição amostral de $T(\mathbf{X}) = \sum_{i=1}^n \ln(X_i/4)$.

(d) Considere os valores abaixo uma amostra da população X. Encontre as estimativas para o parâmetro θ tanto pelo método dos momentos quanto pelo método da máxima verossimilhança.

$$9,95 \quad 5,01 \quad 4,78 \quad 4,02 \quad 6,60 \quad 4,83 \quad 4,42 \quad 4,23 \quad 23,58 \quad 4,96 \quad 4,39 \quad 14,37 \quad 7,12 \quad 4,57$$

3.12. Seja X uma variável aleatória contínua cuja função densidade depende do parâmetro $\theta > 0$ e é definida por

$$f_X(x) = \frac{3\theta^3}{x^4} \;, \quad x > \theta.$$

Veja que $X \sim Pareto(3, \theta)$ e que esta distribuição é a mesma do Exercício 2.5.2.

(a) Encontre o estimador para θ pelo método dos momentos.

(b) Verifique se o estimador para θ pelo método dos momentos é consistente em EQM.

(c) Encontre o estimador para θ pelo método da máxima verissimilhança.

(d) Verifique se o estimador para θ pelo método da máxima verissimilhança é não-tendencioso, assintoticamente não-tendencioso ou tendencioso para θ. Dica: no Exercício 2.5.2 foi encontrado a distribuição amostral de $T(\mathbf{X}) = X_{(1)}$.

(e) Considere os valores abaixo uma amostra da população X. Encontre as estimativas para o parâmetro θ tanto pelo método dos momentos quanto pelo método da máxima verossimilhança.

$$1,11 \quad 1,17 \quad 1,33 \quad 2,22 \quad 1,08 \quad 2,14 \quad 2,62 \quad 1,43 \quad 1,39 \quad 1,02$$

3.13. Seja X uma variável aleatória contínua cuja função densidade depende do parâmetro $\theta > 0$ e é definida por

$$f_X(x) = 3\theta^3 x^2 e^{-(\theta x)^3} \;, \quad x > 0.$$

Veja que $X \sim Weib(3, 1/\theta)$ e que esta distribuição é a mesma do Exercício 2.5.3.

(a) Encontre o estimador para θ pelo método dos momentos.

(b) Encontre o estimador para θ pelo método da máxima verissimilhança.

(c) Considere os valores abaixo uma amostra da população X. Encontre as estimativas para o parâmetro θ tanto pelo método dos momentos quanto pelo método da máxima verossimilhança.

$$2,20 \quad 1,99 \quad 1,65 \quad 0,92 \quad 2,34 \quad 0,95 \quad 0,77 \quad 1,49 \quad 1,55$$

3.14. Seja X uma variável aleatória contínua cuja função densidade depende do parâmetro $\theta > 0$ e é definida por

$$f_X(x) = 2^\theta \theta x^{\theta-1}, \quad 0 < x < \frac{1}{2}.$$

Veja que $2X \sim Beta(\theta, 1)$ e que esta distribuição é a mesma do Exercício 2.5.4.

(a) Encontre o estimador para θ pelo método dos momentos.

(b) Encontre o estimador para θ pelo método da máxima verissimilhança.

(c) Verifique se o estimador para θ pelo método da máxima verissimilhança é não-tendencioso, assintoticamente não-tendencioso ou tendencioso para θ. Dica: no Exercício 2.5.4 foi encontrado a distribuição amostral de $T(\mathbf{X}) = \sum_{i=1}^{n} -\ln(2X_i)$.

(d) Considere os valores abaixo uma amostra da população X. Encontre as estimativas para o parâmetro θ tanto pelo método dos momentos quanto pelo método da máxima verossimilhança.

$$0,210 \quad 0,364 \quad 0,486 \quad 0,386 \quad 0,171 \quad 0,407 \quad 0,435 \quad 0,275$$

3.15. Seja X uma variável aleatória discreta cuja função de probabilidade depende do parâmetro $0 < p < 1$ e é definida por

$$p_X(x) = \binom{15}{x} p^x (1-p)^{15-x}, \quad x = 0, 1, \ldots, 15$$

Veja que $X \sim Binom(15, p)$ e que esta distribuição é a mesma do Exercício 2.5.5.

(a) Encontre o estimador para p pelo método dos momentos.

(b) Verifique se o estimador para θ pelo método dos momentos é consistente em EQM.

(c) Encontre o estimador para θ pelo método da máxima verissimilhança.

(d) Verifique se o estimador para θ pelo método da máxima verissimilhança é não-tendencioso, assintoticamente não-tendencioso ou tendencioso para θ.

(e) Considere os valores abaixo uma amostra da população X. Encontre as estimativas para o parâmetro θ tanto pelo método dos momentos quanto pelo método da máxima verossimilhança.

$$3 \quad 3 \quad 4 \quad 6 \quad 2 \quad 6 \quad 7 \quad 4 \quad 4 \quad 1 \quad 2 \quad 2 \quad 5 \quad 3 \quad 5 \quad 4 \quad 5 \quad 8 \quad 3 \quad 5 \quad 6 \quad 2 \quad 4 \quad 2 \quad 3$$

Capítulo 4

Estimadores Não-Viesados

Nesse capítulo vamos estudar com mais detalhes os estimadores não-tendencioso (não-viesados). A construção desse capítulo terá como objetivo responder a pergunta: Será que podemos afirmar que um certo estimador não-tendencioso é o mais eficiente entre todos os não-tendencioso? Ou seja, temos como afirmar que um estimador não-tendencioso é aquele com menor variância entre todos os não-tendencioso?

Veja que se estamos trabalhando somente com estimadores não-tendenciosos, comparar eficiência significa comparar as variâncias dos estimadores, uma vez que o viés deles é zero. Então, de fato, o que buscamos nesse capítulo é meios de encontrar um estimador não-tendencioso com a menor variância possível.

4.1 A Desigualdade de Cramér-Rao

O primeiro resultado que veremos para os estimadores não-tendenciosos é a Desigualdade de Cramér-Rao. Essa desigualdade nos fornece um limite inferior para a variância dos estimadores de um certo parâmetro. Ou seja, nenhum estimador não-tendencioso pode ter variância menor que o limite de Crámer-Rao, como mostra o Teorema 4.1 a seguir.

Teorema 4.1. ***Desigualdade de Cramér-Rao***

Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória da população X, cuja distribuição depende do parâmetro desconhecido θ. Suponha satisfeitas as condições de regularidade enunciadas a seguir. Então, a variância de qualquer estimador $\hat{\eta}$ não-tendencioso para o parâmetro $\eta = \tau(\theta)$ satisfaz a seguinte desigualdade,

$$\operatorname{Var}(\hat{\eta}) \geq \frac{\left(\dfrac{\mathrm{d}}{\mathrm{d}\theta}\tau(\theta)\right)^2}{nI_F(\theta)},$$

onde a quantidade $I_F(\theta)$ é denominada informação de Fisher de θ e é definida por:

$$I_F(\theta) = \mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\ln\left(f_X(X|\theta)\right)\right)^2\right],$$

sendo f_X a função densidade de X, para o caso de X contínua, ou a função de probabilidade de X, para o caso de X discreta. Vale observar que $\left(\dfrac{\partial}{\partial\theta}\ln\left(f_X(X|\theta)\right)\right)^2$ é função da variável aleatória X e por isso também é uma variável aleatória, então faz sentido calcular a sua esperança.

Condições de Regularidade:

(i) $Im(X)$ não depende de θ.

(ii) A função f_X é duas vezes diferenciável em relação à θ.

(iii) Vale a seguinte permutação entre derivada e integral,

$$\frac{\partial}{\partial\theta}\int\int\dots\int T(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int\int\dots\int\frac{\partial}{\partial\theta}T(\mathbf{x})f_{\mathbf{X}}(\mathbf{x})\,\mathrm{d}\mathbf{x},$$

para qualquer estatística $T(X_1,\dots,X_n)$.

Demonstração. Suponha X variável aleatória contínua que satisfaz as condições de regularidade. O caso discreto será análogo. Então f_X é a função densidade da população X e $f_{\mathbf{X}}$ é a função densidade conjunta de uma amostra de X e ambas dependem do parâmetro θ.

A demonstração deste teorema é uma aplicação direta da Desigualdade de Cauchy-Schwarz, a qual nos mostra que para quaisquer variáveis aleatórias W e Y é verdade que

$$\mathrm{Cov}^2(W,Y) \le \mathrm{Var}(W)\,\mathrm{Var}(Y).$$

Com isso podemos escrever que para quaisquer variáveis aleatórias W e Y

$$\mathrm{Var}(W) \ge \frac{\mathrm{Cov}^2(W,Y)}{\mathrm{Var}(Y)}. \tag{4.1}$$

Para essa demonstração vamos considerar $W = \hat{\eta}$ e $Y = \frac{\partial}{\partial\theta}\ln\left(f_{\mathbf{X}}(\mathbf{X},\theta)\right)$. Precisamos encontrar $\mathrm{Cov}(W,Y) = \mathrm{E}(WY) - \mathrm{E}(W)\,\mathrm{E}(Y)$ e $\mathrm{Var}(Y)$.

Veja que para qualquer estatística $T(\mathbf{X})$ e qualquer função densidade $f_{\mathbf{X}}$ que satisfaça as condições de regularidade, é verdade que

$$\begin{aligned}
\mathrm{E}\left(T(\mathbf{X})\left(\frac{\partial}{\partial\theta}\ln\left(f_{\mathbf{X}}(\mathbf{X},\theta)\right)\right)\right) &= \mathrm{E}\left(T(\mathbf{X})\frac{1}{f_{\mathbf{X}}(\mathbf{X},\theta)}\left(\frac{\partial}{\partial\theta}f_{\mathbf{X}}(\mathbf{X},\theta)\right)\right)\\
&= \int\cdots\int_{\mathbb{R}^n} T(\mathbf{x})\frac{1}{f_{\mathbf{X}}(\mathbf{x},\theta)}\left(\frac{\partial}{\partial\theta}f_{\mathbf{X}}(\mathbf{x},\theta)\right)f_{\mathbf{X}}(\mathbf{x},\theta)\mathrm{d}\mathbf{x}\\
&= \int\cdots\int_{\mathbb{R}^n} T(\mathbf{x})\left(\frac{\partial}{\partial\theta}f_{\mathbf{X}}(\mathbf{x},\theta)\right)\mathrm{d}\mathbf{x}\\
&= \int\cdots\int_{\mathbb{R}^n} \frac{\partial}{\partial\theta}T(\mathbf{x})f_{\mathbf{X}}(\mathbf{x},\theta)\mathrm{d}\mathbf{x}\\
\text{(condição de regularidade (iii))}\qquad &= \frac{\partial}{\partial\theta}\int\cdots\int_{\mathbb{R}^n} T(\mathbf{x})f_{\mathbf{X}}(\mathbf{x},\theta)\mathrm{d}\mathbf{x}\\
&= \frac{\partial}{\partial\theta}\mathrm{E}\left(T(\mathbf{X})\right).
\end{aligned}$$

Como essa igualdade vale para qualquer estatística $T(\mathbf{X})$ e qualquer densidade f que satisfaça as condições de regularidade. Em particular vale para $T(\mathbf{X}) = 1$, o que nos faz concluir que

$$\mathrm{E}\left(\left(\frac{\partial}{\partial\theta}\ln\left(f_{\mathbf{X}}(\mathbf{X},\theta)\right)\right)\right) = \frac{\partial}{\partial\theta}\mathrm{E}\left(1\right) = \frac{\partial}{\partial\theta}1 = 0. \tag{4.2}$$

Então $\mathrm{E}(Y) = 0$ e $\mathrm{Cov}(W,Y) = \mathrm{E}(WY)$. Considerando agora $T(\mathbf{X}) = \hat{\eta}$,

$$\mathrm{E}\left(\hat{\eta}\left(\frac{\partial}{\partial\theta}\ln\left(f_{\mathbf{X}}(\mathbf{x},\theta)\right)\right)\right) = \frac{\partial}{\partial\theta}\mathrm{E}\left(\hat{\eta}\right) = \frac{\partial}{\partial\theta}\eta = \frac{\partial}{\partial\theta}\tau(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}\tau(\theta),$$

já que $\hat{\eta}$ é um estimador não tendencioso para $\eta = \tau(\theta)$. Encontramos então

$$\text{Cov}(W,Y) = \text{E}(WY) - \text{E}(W)\,\text{E}(Y) = \text{E}(WY) = \text{E}\left(\hat{\eta}\left(\frac{\partial}{\partial\theta}\ln\left(f_{\mathbf{X}}(\mathbf{X},\theta)\right)\right)\right) = \frac{\mathrm{d}}{\mathrm{d}\theta}\tau(\theta).$$

Agora vamos encontrar $\text{Var}(Y)$. Como já vimos que $\text{E}(Y) = 0$, $\text{Var}(Y) = \text{E}(Y^2)$.

$$\begin{aligned}
\text{Var}(Y) = \text{E}(Y^2) &= \text{E}\left(\left(\frac{\partial}{\partial\theta}\ln\left(f_{\mathbf{X}}(\mathbf{X},\theta)\right)\right)^2\right) = \text{E}\left(\left(\frac{\partial}{\partial\theta}\ln\left(\prod_{i=1}^{n} f_X(X_i,\theta)\right)\right)^2\right) \\
&= \text{E}\left(\left(\frac{\partial}{\partial\theta}\sum_{i=1}^{n}\ln\left(f_X(X_i,\theta)\right)\right)^2\right) = \text{E}\left(\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln\left(f_X(X_i,\theta)\right)\right)^2\right) \\
&= \text{E}\left(\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta}\ln\left(f_X(X_i,\theta)\right)\right)^2 + \sum_{i=1}^{n}\sum_{j\neq i}\left(\frac{\partial}{\partial\theta}\ln\left(f_X(X_i,\theta)\right)\frac{\partial}{\partial\theta}\ln\left(f_X(X_j,\theta)\right)\right)\right) \\
&= \text{E}\left(\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta}\ln\left(f_X(X_i,\theta)\right)\right)^2\right) + \text{E}\left(\sum_{i=1}^{n}\sum_{j\neq i}\left(\frac{\partial}{\partial\theta}\ln\left(f_X(X_i,\theta)\right)\frac{\partial}{\partial\theta}\ln\left(f_X(X_j,\theta)\right)\right)\right) \\
&= \sum_{i=1}^{n}\text{E}\left(\left(\frac{\partial}{\partial\theta}\ln\left(f_X(X_i,\theta)\right)\right)^2\right) + \sum_{i=1}^{n}\sum_{j\neq i}\text{E}\left(\frac{\partial}{\partial\theta}\ln\left(f_X(X_i,\theta)\right)\frac{\partial}{\partial\theta}\ln\left(f_X(X_j,\theta)\right)\right) \\
&= \sum_{i=1}^{n}\text{E}\left(\left(\frac{\partial}{\partial\theta}\ln\left(f_X(X,\theta)\right)\right)^2\right) + \sum_{i=1}^{n}\sum_{j\neq i}\left[\underbrace{\text{E}\left(\frac{\partial}{\partial\theta}\ln\left(f_X(X_i,\theta)\right)\right)\text{E}\left(\frac{\partial}{\partial\theta}\ln\left(f_X(X_j,\theta)\right)\right)}_{0\qquad\text{(Equação 4.2)}}\right] \\
&= n\,\text{E}\left(\left(\frac{\partial}{\partial\theta}\ln\left(f_X(X,\theta)\right)\right)^2\right)
\end{aligned}$$

Juntando tudo temos

$$\text{Var}(\hat{\eta}) \geq \frac{\frac{\mathrm{d}}{\mathrm{d}\theta}\tau(\theta)}{n\,\text{E}\left(\left(\frac{\partial}{\partial\theta}\ln\left(f_X(X,\theta)\right)\right)^2\right)}$$

como queríamos demonstrar. □

Veja que a Desigualdade de Cramér-Rao fornece um limite inferior para a variância de estimadores não-tendenciosos para um parâmetro $\eta = \tau(\theta)$ qualquer. Veja que esse limite inferior é função de n e θ. Vamos chamar esse limite inferior V^{η}_{min}.

$$\text{V}^{\eta}_{min}(\theta) = \frac{\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\tau(\theta)\right)^2}{nI_F(\theta)}$$

Se encontrarmos um estimador não-tendencioso para η com variância igual a V^{η}_{min} estamos com o estimador não-tendencioso de menor variância entre todos os não-tendenciosos.

Veja que se o interesse for em um limite inferior para a variância dos estimadores não-tendenciosos para o próprio parâmetro da distribuição θ, basta considerar $\tau(\theta) = \theta$, ou seja, τ

serpa a função identidade. Nesse caso o numerador de V_{min} passa a ser 1 e a desigualdade de Cramér-Rao fica:

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{nI_F(\theta)} \tag{4.3}$$

A proposição a seguir apresenta um resultado bem útil.

Proposição 4.1. *Supondo satisfeitas as condições de regularidades apresentadas no Teorema 4.1, é verdade que*

$$I_F(\theta) = \mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\ln\left(f_X(X|\theta)\right)\right)^2\right] = -\,\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\ln\left(f_X(X|\theta)\right)\right].$$

Demonstração. Primeiro veja que:

$$\mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\ln\left(f_X(X|\theta)\right)\right)^2\right] = \mathrm{E}\left[\left(\frac{1}{f_X(X|\theta)}\left(\frac{\partial}{\partial\theta}f_X(X|\theta)\right)\right)^2\right] = \mathrm{E}\left[\frac{\left(\frac{\partial}{\partial\theta}f_X(X|\theta)\right)^2}{\left(f_X(X|\theta)\right)^2}\right].$$

Agora vamos desenvolver o outro lado da igualdade.

$$\begin{aligned}
-\,\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\ln\left(f_X(X|\theta)\right)\right] &= -\,\mathrm{E}\left[\frac{\partial}{\partial\theta}\left(\frac{1}{f_X(X|\theta)}\left(\frac{\partial}{\partial\theta}f_X(X|\theta)\right)\right)\right] \\
&= -\,\mathrm{E}\left[\frac{\partial}{\partial\theta}\left(\frac{\left(\frac{\partial}{\partial\theta}f_X(X|\theta)\right)}{f_X(X|\theta)}\right)\right] \\
&= -\,\mathrm{E}\left[\frac{\left(\frac{\partial^2}{\partial\theta^2}f_X(X|\theta)\right)f_X(X|\theta) - \frac{\partial}{\partial\theta}f_X(X|\theta)\frac{\partial}{\partial\theta}f_X(X|\theta)}{\left(f_X(X|\theta)\right)^2}\right] \\
&= -\,\mathrm{E}\left[\frac{\left(\frac{\partial^2}{\partial\theta^2}f_X(X|\theta)\right)f_X(X|\theta)}{\left(f_X(X|\theta)\right)^2} - \frac{\frac{\partial}{\partial\theta}f_X(X|\theta)\frac{\partial}{\partial\theta}f_X(X|\theta)}{\left(f_X(X|\theta)\right)^2}\right] \\
&= -\,\mathrm{E}\left[\frac{\left(\frac{\partial^2}{\partial\theta^2}f_X(X|\theta)\right)}{f_X(X|\theta)} - \frac{\left(\frac{\partial}{\partial\theta}f_X(X|\theta)\right)^2}{\left(f_X(X|\theta)\right)^2}\right] \\
&= \underbrace{-\,\mathrm{E}\left[\frac{\left(\frac{\partial^2}{\partial\theta^2}f_X(X|\theta)\right)}{f_X(X|\theta)}\right]}_{0} + \underbrace{\mathrm{E}\left[\frac{\left(\frac{\partial}{\partial\theta}f_X(X|\theta)\right)^2}{\left(f_X(X|\theta)\right)^2}\right]}_{\mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\ln(f_X(X|\theta))\right)^2\right]}
\end{aligned}$$

Agora veja porque $\mathrm{E}\left[\left(\frac{\partial^2}{\partial\theta^2}f_X(X|\theta)\right)/f_X(X|\theta)\right] = 0$, supondo X contínua. Se X for discreta o desenvolvimento é análogo.

$$\begin{aligned}
\mathrm{E}\left[\frac{\left(\frac{\partial^2}{\partial\theta^2}f_X(X|\theta)\right)}{f_X(X|\theta)}\right] &= \int\frac{\left(\frac{\partial^2}{\partial\theta^2}f_X(x|\theta)\right)}{f_X(x|\theta)}f_X(x|\theta)\,\mathrm{d}x \\
&= \int\frac{\partial^2}{\partial\theta^2}f_X(x|\theta)\,\mathrm{d}x = \frac{\partial^2}{\partial\theta^2}\underbrace{\int f_X(x|\theta)\,\mathrm{d}x}_{1} = \frac{\partial^2}{\partial\theta^2}1 = 0
\end{aligned}$$

□

Exemplo 4.1. *Seja $X \sim Poisson(\lambda)$. Encontre a variância mínima dada pelo limite de Cramér-Rao para os estimadores não-tendenciosos para o parâmetro λ.*

Solução:

Como no caso queremos a variância mínima para estimadores do parâmetro da distribuição (e não para um parâmetro definido como função dele), vamos usar a Equação 4.3:

$$V_{min}^{\lambda} = \frac{1}{nI_F(\theta)}$$

Vamos primeiro encontrar a Informação de Fisher.

$$\begin{aligned}
I_F(\theta) &= -\,\mathrm{E}\left(\frac{\partial^2}{\partial\lambda^2}\ln\left(f_X(X|\lambda)\right)\right) = -\,\mathrm{E}\left(\frac{\partial^2}{\partial\lambda^2}\left(\ln\left(\frac{\lambda^X}{X!}e^{-\lambda}\right)\right)\right) \\
&= -\,\mathrm{E}\left(\frac{\partial^2}{\partial\lambda^2}\left(X\ln(\lambda)-\lambda-\ln(X!)\right)\right) = -\,\mathrm{E}\left(-\frac{X}{\lambda^2}\right) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}
\end{aligned}$$

Então, qualquer que seja o estimador não-tendencioso $\hat{\lambda}$ para λ é verdade que

$$\mathrm{Var}(\hat{\lambda}) \geq \frac{1}{nI_F(\theta)} = \frac{1}{n(1/\lambda)} = \frac{\lambda}{n}.$$

‖

Quem é o estimador para λ tanto pelo método dos momentos quanto pelo método da máxima verossimilhança? $\hat{\lambda} = \bar{X}$. Ele é não-tendencioso? Sim. E qual é a variância desse estimador? $\mathrm{Var}(\hat{\lambda}) = \mathrm{Var}(\bar{X}) = \dfrac{\mathrm{Var}(X)}{n} = \dfrac{\lambda}{n}$. Então veja que este é um estimador não-tendencioso cuja variância coincide com o limite de Cramér-Rao. Isso significa que não existe outro estimador não-tendencioso com variância menor que a dele. Ele é o estimador mais eficiente entre todos os não tendenciosos.

Exemplo 4.2. *Seja $X \sim U[0,\theta]$. Veja que a população X não satisfaz as condições de regularidade. Particularmente não ela não satisfaz a condição (i). Então a desigualdade de Cramér-Rao não se aplica para essa distribuição. Verifique que de fato isso é verdade: encontre um estimador não-tendencioso para θ cuja variância é menor que a variância mínima dada pelo limite de Cramér-Rao.*

Solução:

Vamos começar calculando a Informação de Fisher (mesmo que isso não faça sentido uma vez que ela não satisfaz as condições de regularidade):

$$\begin{aligned}
I_F(\theta) &= \mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\ln\left(f_X(X|\theta)\right)\right)^2\right] = \mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\left(\ln\left(\frac{1}{\theta}\right)\right)\right)^2\right] \\
&= \mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\left(-\ln(\theta)\right)\right)^2\right] = \mathrm{E}\left[\left(-\frac{1}{\theta}\right)^2\right] = \frac{1}{\theta^2}
\end{aligned}$$

Assim, se fosse aplicável a desigualdade de Cramér-Rao, seria verdade que todo estimador não-tendencioso para θ teria variância de no mínimo

$$V^{\theta}_{min} = \frac{1}{n(1/\theta^2)} = \frac{\theta^2}{n}.$$

Considere agora o estimador de máxima verossimilhança para θ, $\hat{\theta} = X_{(n)}$. Este é um estimador tendencioso tal que, segundo os Exemplos 3.23 e 3.24,

$$\mathrm{E}\left(X_{(n)}\right) = \frac{n}{n+1}\theta \qquad \text{e} \qquad \mathrm{Var}(X_{(n)}) = \frac{n\theta^2}{(n+2)(n+1)^2}.$$

Podemos definir um novo estimador, inspirando no de máxima verossimilhança, que seja não-tendencioso:

$$\hat{\theta}_{NOVO} = \frac{n+1}{n}X_{(n)}.$$

Esse novo estimador é tal que

$$\begin{aligned}
\mathrm{E}(\hat{\theta}_{NOVO}) &= \frac{n+1}{n}\,\mathrm{E}\left(X_{(n)}\right) = \frac{n+1}{n}\frac{n}{n+1}\theta = \theta \\
\mathrm{Var}(\hat{\theta}_{NOVO}) &= \left(\frac{n+1}{n}\right)^2 \mathrm{Var}(X_{(n)}) = \left(\frac{n+1}{n}\right)^2 \frac{n\theta^2}{(n+2)(n+1)^2} = \frac{\theta^2}{n(n+2)}.
\end{aligned}$$

Veja que $\hat{\theta}_{NOVO}$ é um estimador não-tendencioso tal que a sua variância é menor que a limite inferior dado por Cramér-Rao, o que comprova que a desigualdade não se aplica nesse caso. ‖

Exemplo 4.3. *Seja $X \sim Geo^*(p)$, definida pelo número de ensaios de Bernoulli de parâmetro p até a ocorrência do primeiro sucesso. Encontre a variância mínima dada pelo limite de Cramér-Rao para os estimadores não-tendenciosos para os parâmetros p e $\mu = \mathrm{E}(X) = \frac{1}{p}$.*

Solução:

Primeiro veja que X satisfaz as condições de regularidade, e por isso podemos aplicar a desigualdade de Cramér-Rao para buscar um limite inferior para a variância de estimadores de p ou qualquer outro parâmetro escrito como função de p.

Primeiro vamos encontrar a Informação de Fisher.

$$\begin{aligned}
I_F(p) &= -\,\mathrm{E}\left[\frac{\partial^2}{\partial p^2}\ln\left(f_X(X|p)\right)\right] = -\,\mathrm{E}\left[\frac{\partial^2}{\partial p^2}\ln\left(p(1-p)^{X-1}\right)\right] \\
&= -\,\mathrm{E}\left[\frac{\partial^2}{\partial p^2}\left(\ln(p) + (X-1)\ln(1-p)\right)\right] = -\,\mathrm{E}\left[-\frac{1}{p^2} - (X-1)\frac{1}{(1-p)^2}\right] \\
&= -\,\mathrm{E}\left[-\frac{1}{p^2} - \frac{X}{(1-p)^2} + \frac{1}{(1-p)^2}\right] = -\left(-\frac{1}{p^2} - \frac{1}{p(1-p)^2} + \frac{1}{(1-p)^2}\right) \\
&= -\left(\frac{-(1-p)^2 - p + p^2}{p^2(1-p)^2}\right) = \frac{1}{p^2(1-p)}
\end{aligned}$$

A variância mínima dada pelo limite Cramér-Rao para os estimadores não-tendenciosos para o parâmetro p é

$$V^{p}_{min} = \frac{1}{nI_F(p)} = \frac{1}{n(1/p^2(1-p)))} = \frac{p^2(1-p)}{n}.$$

Já para os estimadores não-tendenciosos para o parâmetro $\mu = \dfrac{1}{p} = \tau(p)$, a variância mínima dada pelo limite de Cramér-Rao é:

$$V_{min}^{\mu} = \frac{\left(\dfrac{\mathrm{d}}{\mathrm{d}p}\tau(p)\right)^2}{nI_F(p)} = \frac{\left(\dfrac{\mathrm{d}}{\mathrm{d}p}\dfrac{1}{p}\right)^2}{n(1/p^2(1-p))} = \frac{\left(-\dfrac{1}{p^2}\right)^2}{n(1/p^2(1-p))} = \frac{p^2(1-p)}{np^4} = \frac{(1-p)}{np^2}.$$

Conhecemos algum estimador não-tendencioso para μ? Sim, $\hat{\mu} = \bar{X}$. Veja que a sua variância coincide com a variância mínima dada pelo limite de Cramér-Rao. Ou seja, entre todos os estimadores não-tendencioso nenhum outro tem variância menor que a dele. $\|$

Exercícios da Seção 4.1

4.1.1 Considere uma amostra aleatória $X_1, \ldots, X_n$ de $X \sim Ber(p)$.

(a) Encontre a variância mínima dada pelo limite de Cramér-Ráo para um estimador não-tendencioso de p.

(b) Encontre a variância mínima dada pelo limite de Cramér-Ráo para um estimador não-tendencioso de $\eta = \mathrm{Var}(X)$.

4.1.2 ([Larson, 1982] - Capítulo 7) Seja $X_1, \ldots, X_n$ uma amostra aleatória de $X \sim N(\mu, \sigma^2)$. As condições de regularidade do limite de Cramér-Rao são satisfeitas para os dois parâmetros.

(a) Encontre o limite inferior para a variância de um estimador não-tendenciosos para μ a partir da desigualdade de Cramér-Rao.

(b) Encontre o limite inferior para a variância de um estimador não-tendenciosos para σ^2 a partir da desigualdade de Cramér-Rao.

4.2 Estimadores Eficientes

A variância mínima dada pelo limite de Cramér-Rao e a verificação de que alguns estimadores de fato possuem tal variância é motivação para a definição a seguir.

Definição 4.1. *Suponha satisfeitas as condições de desigualdades apresentadas no Teorema 4.1. Dizemos que o estimador $\hat{\eta}$ é eficiente para η se:*

(i) $\hat{\eta}$ é não-tendencioso para η, isto é, $\mathrm{E}(\hat{\eta}) = \eta$; e

(ii) $\mathrm{Var}(\hat{\eta})$ coincide com a variância mínima dada pela desigualdade de Cramér-Rao, isto é,

$$\mathrm{Var}(\hat{\eta}) = \frac{\left(\dfrac{\mathrm{d}}{\mathrm{d}\theta}\tau(\theta)\right)^2}{nI_F(\theta)}.$$

Vale destacar que nem sempre encontramos um estimador eficiente para um certo parâmetro. Além disso, podemos inclusive afirmar que alguns casos não existe esse estimador. Isso é uma aplicação do resultado a seguir.

Proposição 4.2. *Suponha satisfeitas as condições de regularidade apresentadas no Teorema 4.1. Suponha também que exista o estimador de máxima verossimilhança para θ, $\hat{\theta}_{MV}$, e que ele atende $\dfrac{d}{d\theta}L(\theta)\Big|_{\theta=\hat{\theta}_{MV}} = 0$, isto é, ele é ponto crítico da função de verossimilhança L. Se $\hat{\eta}$*

é um estimador eficiente para $\eta = \tau(\theta)$ então $\hat{\eta}$ é o estimador de máxima verossimilhança para η.

A demonstração dessa proposição será omitida, mas pode ser encontrada em [Mood, 1974] (Seção 5.1, página 320).

Talvez a afirmação contra-positiva da Proposição 4.2 seja mais usada do que o próprio resultado. A contra-positiva nos garante que, supondo satisfeitas as condições de regularidade, se $\hat{\eta}$ não for estimador de máxima verossimilhança então ele não será um estimador eficiente. Ou seja, os estimadores de máxima verossimilhança são os únicos candidatos a estimadores eficientes que temos.

Exemplo 4.4. *Seja $X \sim Exp(\lambda)$. Encontre a variância mínima dada pelo limite de Cramér-Rao para os estimadores não-tendenciosos para os parâmetros λ e $\mu = \mathrm{E}(X) = \dfrac{1}{\lambda}$. Verifique se existem estimadores eficientes para esses dois parâmetros.*

Solução:

Primeiro veja que X satisfaz as condições de regularidade. Vamos encontrar a Informação de Fisher.

$$\begin{aligned} I_F(\lambda) &= -\,\mathrm{E}\left[\frac{\partial^2}{\partial\lambda^2}\ln\left(f_X(X|\lambda)\right)\right] = -\,\mathrm{E}\left[\frac{\partial^2}{\partial\lambda^2}\ln\left(\lambda e^{-\lambda X}\right)\right] = -\,\mathrm{E}\left[\frac{\partial^2}{\partial\lambda^2}\left(\ln(\lambda) - \lambda X\right)\right] \\ &= -\,\mathrm{E}\left[-\frac{1}{\lambda^2}\right] = \frac{1}{\lambda^2} \end{aligned}$$

A variância mínima dada pela desigualdade de Cramér-Rao para os estimadores não-tendenciosos para λ é

$$V^{\lambda}_{min} = \frac{1}{nI_F(\lambda)} = \frac{1}{n1/\lambda^2} = \frac{\lambda^2}{n}.$$

Agora a variância mínima dada pela desigualdade de Cramér-Rao para os estimadores não-tendenciosos para $\mu = \dfrac{1}{\lambda} = \tau(\lambda)$ é

$$V^{\mu}_{min} = \frac{\left(\frac{\mathrm{d}}{\mathrm{d}\lambda}\tau(\lambda)\right)^2}{nI_F(\lambda)} = \frac{\left(\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{1}{\lambda}\right)^2}{n(1/\lambda^2)} = \frac{\left(-\frac{1}{\lambda^2}\right)^2}{n(1/\lambda^2)} = \frac{1}{n\lambda^2}.$$

Será que conseguimos encontrar estimadores eficientes para λ e μ, isto é, estimadores não-tendenciosos tais que a variância coincide com o limite de Cramér-Rao? Segundo a Proposição 4.2, se eles existem são os de máxima verossimilhança. Vamos então verificar se eles são eficientes.

O estimador de máxima verossimilhança para λ é $\hat{\lambda} = \dfrac{1}{\bar{X}}$ (Exercício 3.3). Podemos verificar que esse estimador nem é não-tendencioso. Para isso veja primeiro que $\sum_{i=1}^n X_i = Y \sim Gama(n, \lambda)$, então

$$\begin{aligned} \mathrm{E}(\hat{\lambda}) &= \mathrm{E}\left(\frac{1}{\bar{X}}\right) = \mathrm{E}\left(\frac{n}{\sum_{i=1}^n X_i}\right) = \mathrm{E}\left(\frac{n}{Y}\right) = \int_0^\infty \frac{n}{y}\frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y}\,\mathrm{d}y \\ &= n\frac{\lambda^n}{\Gamma(n)}\int_0^\infty y^{n-2}e^{-\lambda y}\,\mathrm{d}y = n\frac{\lambda^n}{\Gamma(n)}\frac{\Gamma(n-1)}{\lambda^{n-1}} = \frac{n}{(n-1)}\lambda \neq \lambda. \end{aligned}$$

Logo, o estimador de máxima verossimilhança $\hat{\lambda} = \dfrac{1}{\bar{X}}$ não é eficiente e por isso podemos afirmar que não existe estimador eficiente para λ.

Já o estimador por máxima verossimilhança para $\mu = \mathrm{E}(X)$ é $\hat{\mu} = \bar{X}$ (aplicando invariância). Veja que $\hat{\mu}$ é eficiente para μ:

$$\mathrm{E}(\hat{\mu}) = \mathrm{E}(\bar{X}) = \mathrm{E}(X) = \mu$$
$$\mathrm{Var}(\hat{\mu}) = \mathrm{Var}(\bar{X}) = \frac{\mathrm{Var}(X)}{n} = \frac{1}{n\lambda^2} = V^{\mu}_{min}.$$

$\|$

E se não existir estimador eficiente para um parâmetro, será que conseguimos encontrar o estimador não-tendencioso de menor variância entre todos os não-tendenciosos? Esse será o assunto da Seção 4.4, mas antes precisamos aprender um novo conceito, de Estatística Suficiente.

Exercícios da Seção 4.2

4.2.1 Continuação do Exercício 4.1.1. Considere uma amostra aleatória $X_1, \ldots, X_n$ de $X \sim Ber(p)$. Existe um estimador eficiente para p?

4.2.2 Continuação do Exercício 4.1.2. Seja $X_1, \ldots, X_n$ uma amostra aleatória de $X \sim N(\mu, \sigma^2)$. Você consegue encontrar um estimador não-tendencioso para σ^2 com a variância igual a variância mínima encontrada no Exercício 4.1.2?

4.2.3 Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $X \sim N(\mu_0, \sigma^2)$, com μ_0 conhecido.

(a) Mostre que $\hat{\sigma}^2_{MV}$ é eficiente para σ^2.

(b) Mostre que não existe estimador eficiente para $\sigma = \sqrt{\sigma^2}$.

4.2.4 ([Bolfarine e Sandoval, 2001] - Seção 2.6) Sejam $X_1, \ldots, X_n$ uma amostra aleatória da população X cuja função densidade de probabilidade é dada por

$$f(x|\theta) = \theta x^{\theta-1}, \quad 0 < x < 1,\ \theta > 0.$$

Verifique se existe estimador eficiente para θ. Dica: Reveja os Exercícios 2.1.11, 3.3.4 e 3.4.3.

4.2.5 Seja $X_1, \ldots, X_n$ amostra aleatória de $X \sim Gama(\alpha_0, \lambda)$, sendo α_0 conhecido.

(a) Encontre as variâncias mínimas dadas pelo limite de Cramér-Rao para estimadores não-tendenciosos de λ e $\theta = \frac{1}{\lambda}$.

(b) Mostre que não existe estimador eficiente para λ.

(c) Mostre que o estimador de máxima verissimilhança para $\theta = \frac{1}{\lambda}$ é eficiente.

4.3 Estatísticas Suficientes

Como veremos nessa seção, as estatísticas suficientes são aquelas que de alguma forma guardam toda a informação da amostra aleatória sobre o parâmetro de interesse. É como se pudéssemos trocar a observação dos n valores da amostra, $X_1, X_2, \ldots, X_n$, pela observação de um único valor, o valor da estatística T na amostra, $T(X_1, X_2, \ldots, X_n)$, sem perder a capacidade de realizar inferências sobre o parâmetro θ. Nesse caso, dizemos que a estatística T é suficiente para o parâmetro θ.

Em outra palavras, suponha $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ e $\mathbf{y} = \{y_1, y_2, \ldots, y_m\}$ duas observações diferentes de uma amostra aleatória da mesma população, cuja distribuição depende do parâmetro desconhecido θ. Se T é uma estatística suficiente para um parâmetro θ e $T(\mathbf{x}) = T(\mathbf{y})$, então a inferência sobre o parâmetro θ deve ser a mesma a partir de qualquer uma das duas observações, $\mathbf{x}$ ou $\mathbf{y}$. Vejamos esse princípio formalizado na definição a seguir.

Definição 4.2. *Dizemos que uma estatística* $T(\mathbf{X}) = T(X_1, X_2, \ldots, X_n)$ *é suficiente para o parâmetro* θ *se a distribuição condicional conjunta do vetor aleatório* $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$ *dado* $T(\mathbf{X}) = t$ *não depende de* θ*.*

A ideia dessa definição é a seguinte. Se a distribuição da população X depende do parâmetro desconhecido θ mas, uma vez conhecido o valor da estatística $T(\mathbf{X}) = t$, a distribuição conjunta da amostra não depende de θ, isso significa que toda a informação sobre o parâmetro θ está contida na estatística $T(\mathbf{X})$.

Verificar que uma estatística $T(\mathbf{X})$ é suficiente para o parâmero θ a partir da Definição 4.2 não é uma tarefa fácil. Ainda mais complicado é encontrar uma estatística suficiente para o parâmetro a partir dessa mesma definição. Para facilitar a busca por estatísticas suficientes temos o resultado do Teorema 4.2 a seguir.

Teorema 4.2. ***Critério da Fatoração de Neyman*** *Seja* $f_{\mathbf{X}}$ *a função densidade conjunta, se* X *for contínua, ou função de probabilidade conjunta, se* X *for discreta, da amostra* $\mathbf{X}$*. A estatística* $T(\mathbf{X})$ *é suficiente para* θ *se e somente se existem funções* h *e* g_θ *tais que*

$$f_{\mathbf{X}}(\mathbf{x}) = h(\mathbf{x}) g_\theta(T(\mathbf{x})),$$

sendo h *uma função que depende somente de* $x_1, x_2, \ldots, x_n$ *(*h *não depende de* θ*) e* g_θ *uma função que depende de* θ *e depende de* $x_1, x_2, \ldots, x_n$ *somente através de* $T(\mathbf{x})$*.*

A demonstração dessa proposição será omitida. A demonstração para o caso discreto pode ser encontrada em [Bolfarine e Sandoval, 2001] e para o caso geral, inclusive o contínuo, em [Casella e Berger, 2002]. Vejamos algumas aplicações diretas do Teorema 4.2.

Exemplo 4.5. *Seja* $X_1, X_2, \ldots, X_n$ *uma amostra aleatória de* $X \sim Bernoulli(p)$*. Encontre uma estatística suficiente para o parâmetro* p*.*

Solução: Vamos escrever a função de probabilidade conjunta da amostra de acordo com o padrão apresentado no Critério da Fatoração (Teorema 4.2).

$$\begin{aligned}
p_{\mathbf{X}}(\mathbf{x}) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}, x_i = 0,1 \\
&= p^{\sum_{i=1}^n x_i}(1-p)^{\sum_{i=1}^n (1-x_i)}, x_i = 0,1 \\
&= p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}, x_i = 0,1.
\end{aligned}$$

Seja $T(\mathbf{x}) = \sum_{i=1}^n x_i$, $g_p(T(\mathbf{x})) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$ e $h(\mathbf{x}) = \begin{cases} 1, & \text{se } x_i = 0,1 \\ 0, & \text{caso contrário} \end{cases}$. Veja que a função h não depende de p e que $p_{\mathbf{X}}(\mathbf{x}) = h(\mathbf{x}) g_p(T(\mathbf{x}))$. Logo, pelo Critério da Fatoração (Teorema 4.2), $T(\mathbf{X}) = \sum_{i=1}^n X_i$ é estatística suficiente para o parâmetro p. $\|$

Exemplo 4.6. *Seja* $X_1, X_2, \ldots, X_n$ *uma amostra aleatória de* $X \sim N(\mu, \sigma_0^2)$*, sendo* σ_0^2 *um parâmetro conhecido. Encontre uma estatística suficiente para o parâmetro* μ*.*

Solução: Vamos escrever a função de densidade conjunta da amostra de acordo com o padrão apresentado no Critério da Fatoração (Teorema 4.2).

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\dfrac{1}{2\sigma_0^2}(x_i-\mu)^2} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n e^{-\dfrac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i-\mu)^2} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n e^{-\dfrac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i^2-2x_i\mu+\mu^2)} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n e^{-\dfrac{1}{2\sigma_0^2}\sum_{i=1}^{n}x_i^2} e^{\dfrac{\mu}{2\sigma_0^2}\sum_{i=1}^{n}x_i} e^{-\dfrac{n\mu^2}{2\sigma_0^2}}
\end{aligned}
$$

Seja $T(\mathbf{x}) = \sum_{i=1}^{n} x_i$, $g_\mu(T(\mathbf{x})) = e^{\dfrac{\mu}{2\sigma_0^2}\sum_{i=1}^{n}x_i} e^{-\dfrac{n\mu^2}{2\sigma_0^2}}$ e $h(\mathbf{x}) = \left(\dfrac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n e^{-\dfrac{1}{2\sigma_0^2}\sum_{i=1}^{n}x_i^2}$.
Veja que a função h não depende de μ e que $f_{\mathbf{X}}(\mathbf{x}) = h(\mathbf{x})g_\mu(T(\mathbf{x}))$. Logo, pelo Critério da Fatoração (Teorema 4.2), $T(\mathbf{x}) = \sum_{i=1}^{n} x_i$ é estatística suficiente para o parâmetro μ.

$\|$

Exemplo 4.7. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $X \sim U[0,\theta]$. Encontre uma estatística suficiente para o parâmetro θ.*
Solução: Vamos novamente escrever a função de densidade conjunta da amostra de acordo com o padrão apresentado no Critério da Fatoração (Teorema 4.2).

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \begin{cases} \prod_{i=1}^{n} \dfrac{1}{\theta} & \text{, se } 0 \le x_i \le \theta \ \forall\ i = 1, \ldots, n \\ 0 & \text{, caso contrário.} \end{cases} \\
&= \begin{cases} \dfrac{1}{\theta^n} & \text{, se } 0 \le x_{(1)} \text{ e } x_{(n)} \le \theta \\ 0 & \text{, caso contrário.} \end{cases}
\end{aligned}
$$

Sejam $T(\mathbf{x}) = x_{(n)}$, $g_\theta(T(\mathbf{x})) = \begin{cases} \dfrac{1}{\theta^n} & \text{, se } x_{(n)} \le \theta \\ 0 & \text{, caso contrário.} \end{cases}$ e $h(\mathbf{x}) = \begin{cases} 1 & \text{, se } 0 \le x_{(1)} \\ 0 & \text{, caso contrário.} \end{cases}$.
Veja que a função h não depende de θ e que $f_{\mathbf{X}}(\mathbf{x}) = h(\mathbf{x})g_\theta(T(\mathbf{x}))$. Logo, pelo Critério da Fatoração (Teorema 4.2), $T(\mathbf{x}) = X_{(n)}$ é estatística suficiente para o parâmetro θ.

$\|$

Dizemos que duas estatísticas $T_1(\mathbf{X})$ e $T_2(\mathbf{X})$ são equivalentes se existe uma função inversível φ tal que $T_1(\mathbf{X}) = \varphi(T_2(\mathbf{X}))$. O Corolário 4.1 nos mostra que se $T(\mathbf{X})$ é uma estatística suficiente para um parâmetro θ então qualquer estatística equivalente a $T(\mathbf{X})$ também será suficiente para esse parâmetro.

Corolário 4.1. *(do Critério da Fatoração)*
Sejam $T_1(\mathbf{X})$ e $T_2(\mathbf{X})$ estatísticas equivalentes. Se $T_1(\mathbf{X})$ é suficiente para o parâmetro θ então $T_2(\mathbf{X})$ também será.

Demonstração. Seja $T_1(\mathbf{X})$ uma estatística suficiente para o parâmetro θ e $T_2(\mathbf{X})$ uma estatística equivalente a $T_1(\mathbf{X})$, isto é, existe uma função φ inversível tal que $T_1(\mathbf{X}) = \varphi(T_2(\mathbf{X}))$.

Como $T_1(\mathbf{X})$ é estatística suficiente para o parâmetro θ, pelo Critério da Fatoração, podemos escrever $f_{\mathbf{X}}(\mathbf{x}) = h(\mathbf{x})g_\theta(T_1(\mathbf{x}))$. Então,

$$f_{\mathbf{X}}(\mathbf{x}) = h(\mathbf{x})g_\theta(T_1(\mathbf{x})) = h(\mathbf{x})g_\theta(\varphi(T_2(\mathbf{X}))) = h(\mathbf{x})\tilde{g}_\theta(T_2(\mathbf{X})),$$

sendo $\tilde{g}_\theta$ a função definida pela combinação g com φ. Novamente aplicando o Critério da Fatoração, concluímos que $T_2(\mathbf{X})$ também é estatística suficiente para o parâmetro θ. □

Exemplo 4.8. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $X \sim Bernoulli(p)$. Já vimos no Exemplo 4.5 que $\sum_{i=1}^n X_i$ é estatística suficiente para o parâmetro p de X. Mostre que $\bar{X}$ também é suficiente para p.*

Solução: Vamos apenas mostrar que $\sum_{i=1}^n X_i$ e $\bar{X}$ são estatísticas equivalentes. Veja que $\bar{X} = \varphi\left(\sum_{i=1}^n X_i\right)$, sendo $\varphi(x) = \dfrac{x}{n}$. Veja que φ é função inversível: $\varphi^{-1}(x) = nx$. Assim, $\sum_{i=1}^n X_i$ e $\bar{X}$ são estatísticas equivalentes e portanto $\bar{X}$ também é estatística suficiente para p. ‖

Vejamos agora que as estatísticas suficientes para θ também são estatísticas suficientes para qualquer função $\tau(\theta)$.

Proposição 4.3. *Se $T(\mathbf{X})$ é uma estatística suficiente para o parâmetro θ então $T(\mathbf{X})$ também é suficiente para $\tau(\theta)$, qualquer que seja a função τ.*

Demonstração. Seja $T(\mathbf{X})$ é uma estatística suficiente para o parâmetro θ. Pela Definição 4.2, a distribuição condicional conjunta do vetor aleatório $\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$ dado $T(\mathbf{X}) = t$ não depende de θ. Veja que se essa distribuição não depende de θ, também não depende de $\tau(\theta)$. Logo essa estatística também é suficiente para o parâmetro $\tau(\theta)$. □

Exemplo 4.9. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $X \sim Bernoulli(p)$. Encontre uma estatística suficiente para $\eta = \mathrm{Var}(X)$.*

Solução: Queremos uma estatística suficiente para o parâmetro $\eta = \mathrm{Var}(X) = p(1-p) = \tau(p)$. Já vimos no Exemplo 4.5 que $\sum_{i=1}^n X_i$ é estatística suficiente para o parâmetro p. Pela Proposição 4.3, $\sum_{i=1}^n X_i$ também é suficiente para $\eta = \tau(p)$. ‖

A boa notícia é que para as distribuições pertencentes à família exponencial temos um resultado imediato, como mostra o Teorema 4.3 a seguir.

Teorema 4.3. *Seja X uma população cuja distribuição depende de um único parâmetro θ e seja f_X a sua função de densidade ou função de probabilidade. Se a distribuição de X pertence à família exponencial unidimensional, isto é, se*

$$f_X(x) = e^{c(\theta)T(x)}e^{d(\theta)}e^{S(x)}, \quad x \in A,$$

sendo A não dependente de θ (Definição 2.5), então a estatística $\sum_{i=1}^n T(X_i)$ é suficiente para o parâmetro θ.

Demonstração. Para essa demonstração vamos aplicar o Teorema 4.2. Como X pertence à família exponencial temos

$$f_X(x) = e^{c(\theta)T(x)}e^{d(\theta)}e^{S(x)}, \quad x \in A.$$

Com isso a função densidade conjunta da amostra $\mathbf{X}$, $f_{\mathbf{X}}$, ou função de probabilidade conjunta, pode ser escrita como

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \prod_{i=1}^n e^{c(\theta)T(x_i)}e^{d(\theta)}e^{S(x_i)}, && x_i \in A \\ &= e^{\sum_{i=1}^n c(\theta)T(x_i)}e^{\sum_{i=1}^n d(\theta)}e^{\sum_{i=1}^n S(x_i)}, && x_i \in A \\ &= e^{c(\theta)\sum_{i=1}^n T(x_i)}e^{nd(\theta)}e^{\sum_{i=1}^n S(x_i)}, && x_i \in A \end{aligned}$$

Considere agora

$$h(\mathbf{x}) = \begin{cases} e^{\sum_{i=1}^{n} S(x_i)}, & \text{se } x_i \in A \\ 0, & \text{caso contrário.} \end{cases} \qquad \text{e} \qquad g_\theta\left(\sum_{i=1}^{n} T(x_i)\right) = e^{c(\theta)\sum_{i=1}^{n} T(x_i)} e^{nd(\theta)}.$$

Veja que $f_{\mathbf{X}}(\mathbf{x}) = h(\mathbf{x}) g_\theta(\sum_{i=1}^{n} T(x_i))$. De acordo com o Teorema 4.2, $\sum_{i=1}^{n} T(X_i)$ é estatística suficiente para θ.

□

Os dois Exemplos a seguir podem ser resolvidos facilmente aplicando o Teorema 4.3.

Exemplo 4.10. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $X \sim Exp(\lambda)$. Encontre uma estatística suficiente para o parâmetro λ.*

Solução:

No Exemplo 2.5 já vimos que se $X \sim Exp(\lambda)$ a distribuição de X pertence à família exponencial e $f_X(x) = e^{-\lambda x} e^{\ln(\lambda)} e^0$, $x \in (0, \infty)$, ou seja, $c(\lambda) = -\lambda$, $T(x) = x$, $d(\lambda) = \ln(\lambda)$, $S(x) = 0$ e $A = (0, \infty)$. Então, pelo Teorema 4.3, $\sum_{i=1}^{n} T(X_i) = \sum_{i=1}^{n} X_i$ é estatística suficiente para o parâmetro λ. ‖

Exemplo 4.11. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $X \sim Poisson(\lambda)$. Encontre uma estatística suficiente para o parâmetro λ.*

Solução:

No Exemplo 2.6 já vimos que se $X \sim Poisson(\lambda)$ a distribuição de X pertence à família exponencial e $p_X(x) = e^{x\ln(\lambda)} e^{-\lambda} e^{-\ln(x!)}$, $x \in \mathbb{N}$, ou seja, $c(\lambda) = \ln(\lambda)$, $T(x) = x$, $d(\lambda) = -\lambda$, $S(x) = -\ln(x!)$ e $A = \mathbb{N}$. Então, pelo Teorema 4.3, $\sum_{i=1}^{n} T(X_i) = \sum_{i=1}^{n} X_i$ é estatística suficiente para o parâmetro λ. ‖

Estatísticas Conjuntamente Suficientes

Em alguns casos não conseguimos encontrar uma única estatística que seja suficiente para um parâmetro, ou para um vetor de parâmetros. Nesse caso vamos precisar de mais de uma estatística para que conjuntamente sejam suficientes.

Os resultados apresentados a seguir são praticamente uma generalização dos resultados anteriores, quando uma única estatística era suficiente para um único parâmetro.

Teorema 4.4. ***Critério da Fatoração de Neyman - versão Multiparamétrica***

Seja $f_{\mathbf{X}}$ a função densidade conjunta, se X for contínua, ou função de probabilidade conjunta, se X for discreta, da amostra $\mathbf{X}$. A estatística $\mathbf{T}(\mathbf{X}) = (T_1(\mathbf{X}), T_2(\mathbf{X}), \ldots, T_r(\mathbf{X}))$ é conjuntamente suficiente para $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ se e somente se existem funções g e h tais que

$$f_{\mathbf{X}}(\mathbf{x}|\boldsymbol{\theta}) = g(T_1(\mathbf{x}), T_2(\mathbf{x}), \ldots, T_r(\mathbf{x})|\boldsymbol{\theta}) h(\mathbf{x}).$$

Assim como no caso uniparamétrico, essa demonstração será omitida.

Exemplo 4.12. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $X \sim Normal(\mu, \sigma^2)$. Encontre estatísticas (conjuntamente) suficientes para o vetor de parâmetros (μ, σ^2).*

Solução: Vamos tentar escrever a função de densidade conjunta da amostra de acordo com o padrão apresentado no Teorema 4.4.

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\dfrac{1}{2\sigma^2}(x_i-\mu)^2} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\dfrac{1}{2\sigma^2}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\dfrac{1}{2\sigma^2}\displaystyle\sum_{i=1}^{n}(x_i^2-2x_i\mu+\mu^2)} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n e^{-\dfrac{1}{2\sigma^2}\displaystyle\sum_{i=1}^{n}x_i^2} e^{\dfrac{\mu}{2\sigma^2}\displaystyle\sum_{i=1}^{n}x_i} e^{-\dfrac{n\mu^2}{2\sigma^2}}
\end{aligned}
$$

Seja $T_1(\mathbf{x}) = \sum_{i=1}^n x_i$ e $T_2(\mathbf{x}) = \sum_{i=1}^n x_i^2$. Definimos

$$
g_{\mu,\sigma^2}(T_1(\mathbf{x}), T_2(\mathbf{x})) = \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n e^{-\dfrac{1}{2\sigma^2}\displaystyle\sum_{i=1}^{n}x_i^2} e^{\dfrac{\mu}{2\sigma^2}\displaystyle\sum_{i=1}^{n}x_i} e^{-\dfrac{n\mu^2}{2\sigma^2}}
$$

e $h(\mathbf{x}) = 1$. Veja que h não depende dos parâmetros (μ, σ^2) e $f_{\mathbf{X}}(\mathbf{x}) = h(\mathbf{x})g_\mu(T(\mathbf{x}))$. Pelo Teorema 4.4, $T_1(\mathbf{x}) = \sum_{i=1}^n x_i$ e $T_2(\mathbf{x}) = \sum_{i=1}^n x_i^2$ são estatísticas conjuntamente suficientes para o vetor de parâmetros (μ, σ^2). ‖

O mais comum é o número de estatísticas conjuntamente suficientes para um vetor de parâmetros desconhecidos coincidir com a dimensão desse vetor, ou seja, com o número de parâmetros desconhecidos. Mas isso não é regra. Se para o Exemplo 4.12 anterior fosse suposto μ conhecido e apenas o parâmetro σ^2 desconhecido, ainda assim seriam necessárias as duas estatísticas $T_1(\mathbf{x}) = \sum_{i=1}^n x_i$ e $T_2(\mathbf{x}) = \sum_{i=1}^n x_i^2$, que conjuntamente serão suficientes apenas para o parâmetro σ^2.

Exemplo 4.13. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $X \sim U[a, b]$. Encontre estatísticas (conjuntamente) suficientes para o vetor de parâmetros (a, b).*
Solução: Vamos novamente tentar escrever a função de densidade conjunta da amostra de acordo com o padrão apresentado no Teorema 4.4.

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \begin{cases} \prod_{i=1}^n \dfrac{1}{\theta} & \text{, se } a < x_i < b \ \forall\ i = 1, \ldots, n \\ 0 & \text{, caso contrário.} \end{cases} \\
&= \begin{cases} \dfrac{1}{\theta^n} & \text{, se } a < x_{(1)} \text{ e } x_{(n)} < b \\ 0 & \text{, caso contrário.} \end{cases}
\end{aligned}
$$

Sejam $T_1(\mathbf{x}) = x_{(n)}$ e $T_2(\mathbf{x}) = x_{(1)}$. Definimos

$$
g_{a,b}(T_1(\mathbf{x}), T_2(\mathbf{x})) = \begin{cases} \dfrac{1}{\theta^n} & \text{, se } a < x_{(1)} \text{ e } x_{(n)} < b \\ 0 & \text{, caso contrário} \end{cases}
$$

e $h(\mathbf{x}) = 1$. Veja que h não depende dos parâmetros (a, b) e $f_{\mathbf{X}}(\mathbf{x}) = h(\mathbf{x})g_\mu(T(\mathbf{x}))$. Pelo Teorema 4.4, $T_1(\mathbf{x}) = x_{(n)}$ e $T_2(\mathbf{x}) = x_{(1)}$ são estatísticas conjuntamente suficientes para o vetor de parâmetros (a, b). ‖

Sejam $\mathbf{T}(\mathbf{X}) = (T_1(\mathbf{X}), \ldots, T_r(\mathbf{X}))$ e $\mathbf{W}(\mathbf{X}) = (W_1(\mathbf{X}), \ldots, W_r(\mathbf{X}))$ dois vetores de estatísticas. Dizemos que $\mathbf{T}(\mathbf{X})$ e $\mathbf{W}(\mathbf{X})$ são equivalentes se existe uma função inversível $\varphi : \mathbb{R}^r \to \mathbb{R}^r$ tal que $\mathbf{T}(\mathbf{X}) = \varphi(\mathbf{W}(\mathbf{X}))$. Vejamos agora a generalização do resultado apresentado no Corolário 4.1, que nos diz que qualquer função inversível de estatísticas conjuntamente suficientes para um vetor de parâmetros também serão estatísticas conjuntamente suficientes para esse vetor.

Corolário 4.2. *(do Critério da Fatoração - versão Multiparamétrica)*
Sejam $\mathbf{T}(\mathbf{X}) = (T_1(\mathbf{X}), \ldots, T_r(\mathbf{X}))$ e $\mathbf{W}(\mathbf{X}) = (W_1(\mathbf{X}), \ldots, W_r(\mathbf{X}))$ dois vetores de estatísticas equivalentes. Se as estatísticas $(T_1(\mathbf{X}), \ldots, T_r(\mathbf{X}))$ são conjuntamente suficientes para o vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ então as estatísticas $(W_1(\mathbf{X}), \ldots, W_r(\mathbf{X}))$ também são.

Demonstração. Como as estatísticas $\mathbf{T}(\mathbf{X})$ são conjuntamente suficientes para $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$, pelo Critério da Fatoração Multiparamétrico,

$$f_{\mathbf{X}}(\mathbf{x}) = g(T_1(\mathbf{x}), T_2(\mathbf{x}), \ldots, T_r(\mathbf{x})|\boldsymbol{\theta})h(\mathbf{x}).$$

Como $\mathbf{T}(\mathbf{X}) = (T_1(\mathbf{X}), \ldots, T_r(\mathbf{X}))$ e $\mathbf{W}(\mathbf{X}) = (W_1(\mathbf{X}), \ldots, W_r(\mathbf{X}))$ são equivalentes, existe uma função inversível $\varphi : \mathbb{R}^r \to \mathbb{R}^r$ tal que

$$\mathbf{T}(\mathbf{X}) = \varphi(\mathbf{W}(\mathbf{X})) = \left(\varphi_1\left(\mathbf{W}(\mathbf{X})\right), \ldots, \varphi_r\left(\mathbf{W}(\mathbf{X})\right)\right)$$

Ou seja, exitem funções φ_i, $i = 1, \ldots, r$, tais que cada $T_i(\mathbf{X}) = \varphi_i(\mathbf{W}(\mathbf{X}))$.

Assim, podemos escrever,

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= g(\varphi_1(\mathbf{W}(\mathbf{X})), \varphi_2(\mathbf{W}(\mathbf{X})), \ldots, \varphi_r(\mathbf{W}(\mathbf{X}))))h(\mathbf{x}) \\ &= \tilde{g}(\mathbf{W}(\mathbf{X}))h(\mathbf{x}) \\ &= \tilde{g}(W_1(\mathbf{X}), W_2(\mathbf{X}), \ldots, W_r(\mathbf{X}))h(\mathbf{x}), \end{aligned}$$

sendo a função $\tilde{g} = g(\varphi(\mathbf{x}))$ definida como a função composta de $g : \mathbb{R}^r \to \mathbb{R}$ com $\varphi : \mathbb{R}^r \to \mathbb{R}^r$.

□

Exemplo 4.14. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $X \sim N(\mu, \sigma^2)$. Já vimos que $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ são conjuntamente suficientes para os parâmetros (μ, σ^2) (Exemplo 4.12). Mostre que $(\bar{X}, S^2)$ são também conjuntamente suficientes para (μ, σ^2).*

Solução:

Como já vimos que $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ são conjuntamente suficientes para os parâmetros (μ, σ^2) (Exemplo 4.12) precisamos apenas mostrar que existe uma função $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ inversível tal que $\varphi(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2) = (\bar{X}, S^2)$. Veja que

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{e} \quad \begin{aligned} \sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n (X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\ &= \sum_{i=1}^n X_i^2 - 2\bar{X}\sum_{i=1}^n X_i + n\bar{X}^2 \\ &= \sum_{i=1}^n X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 \\ &= \sum_{i=1}^n X_i^2 - n\bar{X}^2 \qquad \star \\ &= \sum_{i=1}^n X_i^2 - n\left(\frac{\sum_{i=1}^n X_i}{n}\right)^2 \\ &= \sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i\right)^2}{n} \end{aligned}$$

Então,

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad \text{e} \quad S^2 = \frac{\sum_{i=1}^{n} X_i^2}{n-1} - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)}.$$

Assim se definimos $\varphi(x,y) = \left(\frac{x}{n}, \frac{y}{n-1} - \frac{x^2}{n(n-1)}\right)$ é verdade que $\varphi(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2) = (\bar{X}, S^2)$. Basta verificar que φ é de fato uma função inversível. Para isso vamos encontrar a sua inversa, ou seja, encontrar uma função $\varphi^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$ tal que $\varphi(\varphi^{-1})(x,y) = (x,y)$ ou $\varphi^{-1}(\bar{X}, S^2) = (\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$.

Veja que

$$\sum_{i=1}^{n} X_i = n\bar{X} \quad \text{e} \quad \sum_{i=1}^{n} X_i^2 = (n-1)S^2 + n\bar{X}^2 \quad (\text{ veja linha } \star \text{ da equação acima}).$$

Assim podemos definir $\varphi^{-1}(x,y) = \left(nx, (n-1)y + nx^2\right)$. Verifique que $\varphi(\varphi^{-1})(x,y) = (x,y)$ e $\varphi^{-1}(\bar{X}, S^2) = (\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$. ‖

O último resultado a ser generalizado para o caso multi-paramétrico é o apresentado na Proposição 4.3, que nos mostra que uma estatística que é suficiente para um parâmetro também será suficiente para qualquer função desse parâmetro.

Proposição 4.4. *Sejam $\mathbf{T}(\mathbf{X}) = (T_1(\mathbf{X}), \ldots, T_r(\mathbf{X}))$ estatísticas conjuntamente suficientes para o vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$. Então as estatísticas $\mathbf{T}(\mathbf{X})$ também são conjuntamente suficientes para o vetor de parâmetros $\tau(\boldsymbol{\theta})$, qualquer que seja a função τ.*

Demonstração. A demonstração dessa proposição é uma generalização da demonstração da Proposição 4.3. Sejam $\mathbf{T}(\mathbf{X})$ estatísticas conjuntamente suficientes para o vetor de parâmetros $\boldsymbol{\theta}$. Então, então pela Definição 4.2, a distribuição condicional conjunta do vetor aleatório $\mathbf{X} = (X_1, \ldots, X_n)^T$ dado $\mathbf{T}(\mathbf{X}) = \mathbf{t}$ não depende de $\boldsymbol{\theta}$. Veja que se essa distribuição não depende de $\boldsymbol{\theta}$ ela também não depende de $\tau(\boldsymbol{\theta})$, logo essas estatísticas também são conjuntamente suficientes para o vetor de parâmetros $\tau(\boldsymbol{\theta})$. □

Exemplo 4.15. *Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de $X \sim N(\mu, \sigma^2)$. Encontre estatísticas suficientes para o segundo momento de X.*

Solução:

Queremos estatísticas suficientes para $\mathrm{E}(X^2) = \mu^2 + \sigma^2 = \tau(\mu, \sigma^2)$. Já vimos no Exemplo 4.14 que $(\bar{X}, S^2)$ são estatísticas conjuntamente suficientes para (μ, σ^2). Então $(\bar{X}, S^2)$ também são estatísticas conjuntamente suficientes para $\mathrm{E}(X^2)$. ‖

Assim como no caso uniparamétrico, para as distribuições pertencentes à família exponencial multidimensional são resolvidos de forma imediata.

Teorema 4.5. *Seja X uma variável aleatória cuja distribuição depende de um vetor de parâmetros $\boldsymbol{\theta}$ e seja f_X a sua função de densidade ou função de probabilidade. Se a distribuição de X pertence à família exponencial multidimensional, isto é, se*

$$f_X(x) = e^{\sum_{j=1}^{k} c_j(\boldsymbol{\theta}) T_j(x)} e^{d(\boldsymbol{\theta})} e^{S(x)}, \quad x \in A,$$

sendo A não dependente de θ (Definição 2.6), então as estatísticas $(\sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_k(X_i))$ são conjuntamente suficientes para o vetor de parâmetros $\boldsymbol{\theta}$.

Demonstração. Para essa demonstração vamos aplicar o Teorema 4.4. Como X pertence à família exponencial multidimensional temos

$$f_X(x) = e^{\sum_{j=1}^{k} c_j(\boldsymbol{\theta}) T_j(x)} e^{d(\boldsymbol{\theta})} e^{S(x)}, \quad x \in A,$$

Com isso a função densidade conjunta da amostra $\mathbf{X}$, $f_{\mathbf{X}}$, ou função de probabilidade conjunta, pode ser escrita como

$$\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= \prod_{i=1}^{n} e^{\sum_{j=1}^{k} c_j(\boldsymbol{\theta}) T_j(x)} e^{d(\boldsymbol{\theta})} e^{S(x)}, && x_i \in A \\
&= e^{\sum_{i=1}^{n}\sum_{j=1}^{k} c_j(\boldsymbol{\theta}) T_j(x_i)} e^{\sum_{i=1}^{n} d(\boldsymbol{\theta})} e^{\sum_{i=1}^{n} S(x_i)}, && x_i \in A \\
&= e^{\sum_{j=1}^{k}\sum_{i=1}^{n} c_j(\boldsymbol{\theta}) T_j(x_i)} e^{nd(\boldsymbol{\theta})} e^{\sum_{i=1}^{n} S(x_i)}, && x_i \in A \\
&= e^{\sum_{j=1}^{k} c_j(\boldsymbol{\theta}) \left(\sum_{i=1}^{n} T_j(x_i)\right)} e^{nd(\boldsymbol{\theta})} e^{\sum_{i=1}^{n} S(x_i)}, && x_i \in A \\
&= e^{c_1(\boldsymbol{\theta}) \left(\sum_{i=1}^{n} T_1(x_i)\right)} \dots e^{c_k(\boldsymbol{\theta}) \left(\sum_{i=1}^{n} T_k(x_i)\right)} e^{nd(\boldsymbol{\theta})} e^{\sum_{i=1}^{n} S(x_i)}, && x_i \in A
\end{aligned}$$

Considere agora $\mathbf{T}\left(\mathbf{x}\right) = \left(\sum_{i=1}^{n} T_1(x_i), \dots, \sum_{i=1}^{n} T_k(x_i)\right)$,

$$h(\mathbf{x}) = \begin{cases} e^{\sum_{i=1}^{n} S(x_i)}, & \text{se } x_i \in A \\ 0, & \text{caso contrário.} \end{cases}$$

$$g_{\boldsymbol{\theta}}(\mathbf{T}\left(\mathbf{x}\right)) = e^{c_1(\boldsymbol{\theta}) \left(\sum_{i=1}^{n} T_j(x_i)\right)} \ \dots \ e^{c_k(\boldsymbol{\theta}) \left(\sum_{i=1}^{n} T_j(x_i)\right)} e^{nd(\boldsymbol{\theta})}$$

Veja que $f_{\mathbf{X}}(\mathbf{x}) = h(\mathbf{x}) g_{\boldsymbol{\theta}}(\mathbf{T}(\mathbf{x}))$. De acordo com o Teorema 4.4, $\mathbf{T}\left(\mathbf{x}\right) = \left(\sum_{i=1}^{n} T_1(x_i), \dots, \sum_{i=1}^{n} T_k(x_i)\right)$ são estatísticas conjuntamente suficientes para o vetor de parâmetros $\boldsymbol{\theta}$.

$\square$

Exemplo 4.16. *Seja $X_1, X_2, \dots, X_n$ uma amostra aleatória de $X \sim Gama(\alpha, \lambda)$. Encontre uma estatística suficiente para o parâmetro vetor de parâmetros (α, λ).*
Solução:

Primeiro vamos ver que a distribuição de $X \sim Gama(\alpha, \lambda)$ pertence à família exponencial bidimensional.

$$\begin{aligned}
f_X(x) &= \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \ x > 0 \\
&= e^{\alpha \ln(\lambda) - \ln(\Gamma(\alpha))} \ e^{(\alpha-1)\ln(x)} \ e^{-\lambda x}, \ x > 0
\end{aligned}$$

Assim temos $c_1(\alpha, \lambda) = \alpha - 1$, $T_1(x) = \ln(x)$, $c_2(\alpha, \lambda) = -\lambda$, $T_2(x) = x$, $d(\alpha, \lambda) = \alpha \ln(\lambda) - \ln(\Gamma(\alpha))$, $S(x) = 0$ e $A = (0, \infty)$. E com isso mostramos que X pertence à família exponencial bi-dimensional.

Pelo Teorema 4.5, as estatísticas $\left(\sum_{i=1}^{n} T_1(X_i), \sum_{i=1}^{n} T_2(X_i)\right) = \left(\sum_{i=1}^{n} \ln(X_i), \sum_{i=1}^{n} X_i\right)$ são conjuntamente suficientes para o vetor de parâmetros (α, λ). $\|$

Exemplo 4.17. *Seja $X_1, X_2, \dots, X_n$ uma amostra aleatória de $X \sim Gama(\alpha, \lambda)$. Encontre uma estatística suficiente para $\mu = \mathrm{E}(X)$.*
Solução:

Queremos estatísticas suficientes para $\mu = \mathrm{E}(X) = \dfrac{\alpha}{\lambda} = \tau(\alpha, \lambda)$. Acabamos de ver no Exemplo 4.16 que $\left(\sum_{i=1}^{n} \ln(X_i), \sum_{i=1}^{n} X_i\right)$ são conjuntamente suficientes para o vetor de parâmetros (α, λ) de $X \sim Gama(\alpha, \lambda)$. Pela Proposição 4.3 podemos concluir que $\left(\sum_{i=1}^{n} \ln(X_i), \sum_{i=1}^{n} X_i\right)$ também são conjuntamente suficientes para o parâmetro μ. $\|$

Exercícios da Seção 4.3

4.3.1 Seja $X_1, \dots, X_n$ uma amostra aleatória da população X. Para cada item a seguir encontre uma estatística suficiente para o parâmetro da distribuição da população.

(a) $X \sim N(0, \sigma^2)$.

(b) $X \sim Binomial(5, \theta)$.

(c) X é tal que $f(x|\theta) = e^{-(x-\theta)}$, $x > \theta$.

4.4 O Estimador Não-Viesado de Variância Uniformemente Mínima (ENVVUM)

Podemos agora continuar a nossa busca. Já sabemos identificar quando um estimador é eficiente e quando não existe um estimador eficiente, isto é, quando todos os estimadores não-tendenciosos têm variância maior que a variância mínima dada pelo limite de Cramér-Rao. Nessa segunda situação, será que é possível encontrar o estimador não-tendencioso de menor variância entre todos os não-tendenciosos? Será essa a pergunta que vamos responder nessa seção.

Definição 4.3. *Seja $\hat{\eta}$ um estimador para $\eta = \tau(\theta)$. Dizemos que $\hat{\eta}$ é um estimador não viesado de variância uniformemente mínima (ENVVUM) para η se:*

(i) $\hat{\eta}$ é não-tendencioso, isto é, $\mathrm{E}(\hat{\eta}) = \eta$;

(ii) Dado qualquer outro estimador não-tendencioso $\tilde{\eta}$ para η temos $\mathrm{Var}(\hat{\eta}) \leq \mathrm{Var}(\tilde{\eta})$.

Veja que se $\hat{\eta}$ é um estimador eficiente para η então ele também é um ENVVUM. Mas a recíproca não é verdadeira. Existem ENVVUM que não são eficientes, isso acontece quando não existe um estimador eficiente para o parâmetro em questão.

Proposição 4.5. *Seja $\hat{\eta}$ um ENVVUM para $\eta = \tau(\theta)$, então $\hat{\eta}$ é único. Isto é, se existir outro ENVVUM para η, digamos $\tilde{\eta}$ temos $\tilde{\eta} = \hat{\eta}$.*

Demonstração. Vamos supor que existam dois ENVVUM para η, $\hat{\eta}$ e $\tilde{\eta}$, e vamos concluir que $\tilde{\eta} = \hat{\eta}$. Como $\hat{\eta}$ e $\tilde{\eta}$ são ambos ENVVUM, os dois são não tendenciosos e têm a menor variância entre todos os estimadores não tendenciosos:

$$\mathrm{E}(\hat{\eta}) = \eta,\ E(\tilde{\eta}) = \eta \text{ e } \mathrm{Var}(\hat{\eta}) = \mathrm{Var}(\tilde{\eta}) \leq \mathrm{Var}(\eta^\star)\ \forall\ \eta^\star \text{ tal que } \mathrm{E}(\eta^\star) = \eta.$$

Defina agora $W = \dfrac{\hat{\eta} + \tilde{\eta}}{2}$. Veja que W também será um ENVVUM para η.

$$\mathrm{E}(W) = E\left(\frac{\hat{\eta} + \tilde{\eta}}{2}\right) = \frac{E(\hat{\eta}) + E(\tilde{\eta})}{2} = \frac{\eta + \eta}{2} = \eta.$$

$$\begin{aligned}
\mathrm{Var}(W) = \mathrm{Var}\left(\frac{\hat{\eta} + \tilde{\eta}}{2}\right) &= \frac{\mathrm{Var}(\hat{\eta})}{4} + \frac{\mathrm{Var}(\tilde{\eta})}{4} + \frac{2\,\mathrm{Cov}(\hat{\eta}, \tilde{\eta})}{4} && \star \\
&\leq \frac{\mathrm{Var}(\hat{\eta})}{4} + \frac{\mathrm{Var}(\tilde{\eta})}{4} + \frac{\sqrt{\mathrm{Var}(\hat{\eta})\,\mathrm{Var}(\tilde{\eta})}}{2} && \star\star \\
&= \frac{\mathrm{Var}(\hat{\eta})}{4} + \frac{\mathrm{Var}(\hat{\eta})}{4} + \frac{\sqrt{\mathrm{Var}(\hat{\eta})\,\mathrm{Var}(\hat{\eta})}}{2} \\
&= \mathrm{Var}(\hat{\eta})
\end{aligned}$$

sendo a passagem para a linha $\star\star$ é justificada pelo fato de $\left|\dfrac{\mathrm{Cov}(\hat{\eta}, \tilde{\eta})}{\sqrt{\mathrm{Var}(\hat{\eta})\,\mathrm{Var}(\tilde{\eta})}}\right| \leq 1$. Veja que $\mathrm{Var}(W) \leq \mathrm{Var}(\hat{\eta}) \ \Rightarrow\ \mathrm{Var}(W) = \mathrm{Var}(\hat{\eta})$, uma vez que $\hat{\eta}$ é o ENVVUM.

Substituindo agora $\mathrm{Var}(W)$ e $\mathrm{Var}(\tilde{\eta})$ por $\mathrm{Var}\,(\hat{\eta})$ na linha $\star$ da equação acima temos:

$$\begin{aligned}
\mathrm{Var}\,(W) &= \frac{\mathrm{Var}\,(\hat{\eta})}{4} + \frac{\mathrm{Var}\,(\tilde{\eta})}{4} + \frac{2\,\mathrm{Cov}\,(\hat{\eta},\tilde{\eta})}{4} \\
\mathrm{Var}\,(\hat{\eta}) &= \frac{\mathrm{Var}\,(\hat{\eta})}{4} + \frac{\mathrm{Var}\,(\hat{\eta})}{4} + \frac{2\,\mathrm{Cov}\,(\hat{\eta},\tilde{\eta})}{4} \\
4\,\mathrm{Var}\,(\hat{\eta}) &= \mathrm{Var}\,(\hat{\eta}) + \mathrm{Var}\,(\hat{\eta}) + 2\,\mathrm{Cov}\,(\hat{\eta},\tilde{\eta}) \\
2\,\mathrm{Var}\,(\hat{\eta}) &= 2\,\mathrm{Cov}\,(\hat{\eta},\tilde{\eta}) \quad \Rightarrow \quad \mathrm{Cov}\,(\hat{\eta},\tilde{\eta}) = \mathrm{Var}\,(\hat{\eta}) \quad \Rightarrow \quad \rho_{\hat{\eta},\tilde{\eta}} = \frac{\mathrm{Cov}\,(\hat{\eta},\tilde{\eta})}{\sqrt{\mathrm{Var}\,(\hat{\eta})\,\mathrm{Var}\,(\tilde{\eta})}} = 1.
\end{aligned}$$

Dessa forma podemos garantir que existe uma relação linear entre $\hat{\eta}$ e $\tilde{\eta}$. Isto é, existem a e b tais que $\hat{\eta} = a\tilde{\eta} + b$.

Primeiro veja que $a = 1$:

$$\mathrm{Var}\,(\hat{\eta}) = \mathrm{Cov}\,(\hat{\eta},\tilde{\eta}) = \mathrm{Cov}\,(a\tilde{\eta}+b,\tilde{\eta}) = a\,\mathrm{Cov}\,(\tilde{\eta},\tilde{\eta}) = a\,\mathrm{Var}\,(\tilde{\eta}) = a\,\mathrm{Var}\,(\hat{\eta}) \;\Rightarrow\; a = 1.$$

Agora veja que $b = 0$:

$$\eta = \mathrm{E}\,(\hat{\eta}) = \mathrm{E}\,(a\tilde{\eta}+b) = \mathrm{E}\,(\tilde{\eta}+b) = \mathrm{E}\,(\tilde{\eta}) + b = \eta + b \;\Rightarrow\; b = 0.$$

E com isso concluímos que $\hat{\eta} = a\tilde{\eta} + b = \tilde{\eta}$. Ou seja, se existem dois ENVVUM, eles são iguais, trata-se do mesmo estimador. E por isso garantimos a unicidade dos ENVVUM. $\square$

Corolário 4.3. *Seja $\hat{\eta}$ um estimador eficiente para $\eta = \tau(\theta)$, então $\hat{\eta}$ é único.*

A demonstração do Corolário 4.3 fica como exercício. Ela é imediata, uma vez que todo estimador eficiente também é o ENVVUM.

Antes de apresentarmos o Teorema de Rao-Blackwell (Teorema 4.6) vamos fazer uma rápida revisão sobre alguns conceitos de probabilidade.

Sejam X e Y variáveis aleatórias. Durante o desenvolvimento vamos supor X e Y contínuas, mas os mesmos resultados se aplicam para variáveis discretas. Então,

$$\mathrm{E}(X|Y=y) = \int_{\mathbb{R}} x f_{X|Y}(x|y)\,\mathrm{d}x = g(y),$$

uma vez que a densidade $f_{X|Y}(x|y)$ muda quando muda o valor de y. Ou seja, $\mathrm{E}(X|Y=y)$ é função do valor y.

Por abuso de notação escrevemos $\mathrm{E}(X|Y)$ quando queremos escrever $g(Y)$, onde g é a função $g(y) = \mathrm{E}(X|Y=y)$, apresentada acima. Então podemos definir,

$$\mathrm{E}(X|Y) = g(Y) = \mathrm{E}(X|Y=y)\big|_{y=Y}.$$

Dessa forma, $\mathrm{E}(X|Y)$ é variável aleatória, pois é função de variável aleatória, no caso é função de Y. Então podemos calcular a sua esperança,

$$\begin{aligned}
\mathrm{E}\,(\mathrm{E}(X|Y)) &= \mathrm{E}\,(g(Y)) = \int_{\mathbb{R}} g(y) f_Y(y)\,\mathrm{d}y = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} x f_{X|Y}(x|y)\,\mathrm{d}x \right) f_Y(y)\,\mathrm{d}y \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X|Y}(x|y) f_Y(y)\,\mathrm{d}x\,\mathrm{d}y = \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x = \int_{\mathbb{R}} x \left(\int_{\mathbb{R}} f_{X,Y}(x,y)\,\mathrm{d}y \right) \mathrm{d}x \\
&= \int_{\mathbb{R}} x f_X(x)\,\mathrm{d}x = \mathrm{E}(X).
\end{aligned}$$

Ou seja, quaisquer que sejam as variáveis aleatórias X e Y,

$$\mathrm{E}\,(\mathrm{E}(X|Y)) = \mathrm{E}(X) \tag{4.4}$$

De forma análoga,

$$\begin{aligned}
\mathrm{Var}(X|Y=y) &= \mathrm{E}((X-\mathrm{E}(X|Y=y))^2|Y=y) \\
&= \mathrm{E}(X^2 - 2X\,\mathrm{E}(X|Y=y) + \mathrm{E}^2(X|Y=y)|Y=y) \\
&= \mathrm{E}(X^2) - 2\,\mathrm{E}(X|Y=y)\,\mathrm{E}(X|Y=y) + \mathrm{E}^2(X|Y=y) \\
&= \mathrm{E}(X^2|Y=y) - \mathrm{E}^2(X|Y=y) \\
&= h(y),
\end{aligned}$$

uma vez que $\mathrm{E}(X^2|Y=y)$ e $\mathrm{E}^2(X|Y=y)$ são funções de y. Ou seja, $\mathrm{Var}(X|Y=y)$ é função do valor y. Por abuso de notação também escrevemos $\mathrm{Var}(X|Y)$ quando queremos escrever $h(Y)$, onde h é a função $h(y) = \mathrm{Var}(X|Y=y)$, apresentada acima. Então podemos definir,

$$\mathrm{Var}(X|Y) = h(Y) = \left.\mathrm{Var}(X|Y=y)\right|_{y=Y}.$$

Veja que $\mathrm{Var}(X|Y)$ também é variável aleatória, já que é função de Y. É verdade inclusive que

$$\mathrm{Var}(X|Y) = \mathrm{E}(X^2|Y) - \mathrm{E}^2(X|Y).$$

Um outro resultado importante, que será usando na demonstração do Teorema de Rao-Blackwell vem do cálculo de $\mathrm{E}(\mathrm{Var}(X|Y))$ e $\mathrm{Var}(\mathrm{E}(X))$.

$$\begin{aligned}
\mathrm{E}\left(\mathrm{Var}(X|Y)\right) &= \mathrm{E}\left(\mathrm{E}(X^2|Y) - \mathrm{E}^2(X|Y)\right) \\
&= \mathrm{E}\left(\mathrm{E}(X^2|Y)\right) - \mathrm{E}\left(\mathrm{E}^2(X|Y)\right) \\
&= \mathrm{E}(X^2) - \mathrm{E}\left(\mathrm{E}^2(X|Y)\right). \qquad (4.5)
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}\left(\mathrm{E}(X|Y)\right) &= \mathrm{E}\left(\mathrm{E}^2(X|Y)\right) - \left(\mathrm{E}\left(\mathrm{E}(X|Y)\right)\right)^2 \\
&= \mathrm{E}\left(\mathrm{E}^2(X|Y)\right) - \left(\mathrm{E}(X)\right)^2 \\
&= \mathrm{E}\left(\mathrm{E}^2(X|Y)\right) - \mathrm{E}^2(X). \qquad (4.6)
\end{aligned}$$

Juntando as Equações 4.5 e 4.6 temos,

$$\begin{aligned}
\mathrm{E}\left(\mathrm{Var}(X|Y)\right) + \mathrm{Var}\left(\mathrm{E}(X|Y)\right) &= \mathrm{E}(X^2) - \mathrm{E}\left(\mathrm{E}^2(X|Y)\right) + \mathrm{E}\left(\mathrm{E}^2(X|Y)\right) - \mathrm{E}^2(X) \\
&= \mathrm{E}(X^2) - \mathrm{E}^2(X) \\
&= \mathrm{Var}(X).
\end{aligned}$$

Ou seja, quaisquer que sejam as variáveis aleatórias X e Y,

$$\mathrm{Var}(X) = \mathrm{E}\left(\mathrm{Var}(X|Y)\right) + \mathrm{Var}\left(\mathrm{E}(X|Y)\right). \qquad (4.7)$$

Estamos prontos agora para o Teorema de Rao-Blackwell, que nos mostra como partir de um estimador não tendencioso qualquer e de uma estatística suficiente e encontrar um outro estimador não tendencioso com variância menor ou igual a do estimador inicial.

Teorema 4.6. ***Rao-Blackwell***

Seja $\hat{\eta}$ um estimador não tendencioso para $\eta = \tau(\theta)$ e $T(\mathbf{X})$ uma estatística suficiente para θ. Defina

$$\hat{\eta}^\star = \mathrm{E}(\hat{\eta}|T(\mathbf{X})).$$

Então,

(i) $\hat{\eta}^\star$ é um estimador baseado na estatística $T(\mathbf{X})$;

(ii) $\hat{\eta}^\star$ é não tendencioso para $\eta = \tau(\theta)$;

(iii) $\mathrm{Var}(\hat{\eta}^\star) \leq \mathrm{Var}(\hat{\eta})$*, valendo a igualdade se e somente se* $\hat{\eta}^\star = \eta$*.*

Demonstração. Faremos um item por vez.

(i) Para demonstrar o primeiro item precisamos mostrar duas coisas: primeiro que $\hat{\eta}^\star$ é um estimador, isto é, ele é uma função da amostra que não depende do parâmetro desconhecido θ; e em seguida que $\hat{\eta}^\star$ é baseado na estatística $T(\mathbf{X})$, isto é, ele é função das variáveis X_1, X_2, ..., X_n via estatística $T(\mathbf{X})$.

Primeiro vejamos que de fato $\hat{\eta}^\star$ é uma função da amostra que não depende do parâmetro desconhecido θ. Como $\hat{\eta}$ é um estimador, vamos escrever $\hat{\eta} = W(X_1, X_2, \ldots, X_n)$ e deixar claro que $\hat{\eta}$ é uma transformação da amostra que não depende de θ. Então a transformação W não depende de θ. Assim podemos escrever

$$\hat{\eta}^\star = \mathrm{E}(\hat{\eta}|T(\mathbf{X})) = \mathrm{E}(W(X_1,\ldots,X_n)|T(\mathbf{X})) = \begin{cases} \sum \ldots \sum W(x_1,\ldots,x_n) p_{\mathbf{X}|T(\mathbf{X})}(\mathbf{x}) \\ \text{ou} \\ \int \ldots \int W(x_1,\ldots,x_n) f_{\mathbf{X}|T(\mathbf{X})}(\mathbf{x}) \mathrm{d}\mathbf{x} \end{cases}$$

E concluir que $\hat{\eta}^\star$ não depende de θ, uma vez que W não depende e também $p_{\mathbf{X}|T(\mathbf{X})}(\mathbf{x})$ ou $f_{\mathbf{X}|T(\mathbf{X})}(\mathbf{x})$ não dependem de θ, já que T é uma estatística suficiente para θ (Definição 4.2).

Pela revisão de probabilidade vimos que para quaisquer variáveis aleatórias X e Y temos $\mathrm{E}(X|Y) = g(Y)$. Então, numa aplicação direta desse resultado, $\hat{\eta}^\star = \mathrm{E}(\hat{\eta}|T(\mathbf{X})) = g(T(\mathbf{X}))$.

Assim concluímos que $\hat{\eta}^\star$ é um estimador baseado na estatística.

(ii) Nesse item precisamos mostrar que $\mathrm{E}(\hat{\eta}^\star) = \eta$. Veja que

$$\mathrm{E}(\hat{\eta}^\star) = \mathrm{E}(\mathrm{E}(\hat{\eta}|T(\mathbf{X}))) = \mathrm{E}(\hat{\eta}) = \eta,$$

já que $\hat{\eta}$ é um estimador não tendencioso para η.

(iii) Por fim, vamos mostrar que $\mathrm{Var}(\hat{\eta}^\star) \leq \mathrm{Var}(\hat{\eta})$, valendo a igualdade se e somente se $\hat{\eta}^\star = \eta$. Pela Equação 4.7,

$$\mathrm{Var}(X) = \mathrm{E}\left(\mathrm{Var}(X|Y)\right) + \mathrm{Var}\left(\mathrm{E}(X|Y)\right).$$

quaisquer que sejam X e Y. Fazendo $X = \hat{\eta}$ e $Y = T(\mathbf{X})$ temos,

$$\mathrm{Var}(\hat{\eta}) = \mathrm{E}\left(\mathrm{Var}(\hat{\eta}|T(\mathbf{X}))\right) + \mathrm{Var}\left(\mathrm{E}(\hat{\eta}|T(\mathbf{X}))\right) = \mathrm{E}\left(\mathrm{Var}(\hat{\eta}|T(\mathbf{X}))\right) + \mathrm{Var}\left(\hat{\eta}^\star\right).$$

Como $\mathrm{E}\left(\mathrm{Var}(\hat{\eta}|T(\mathbf{X}))\right) \geq 0$, já que $\mathrm{Var}(\hat{\eta}|T(\mathbf{X}))$ é uma variável aleatória não negativa, podemos concluir que

$$\mathrm{Var}(\hat{\eta}) \geq \mathrm{Var}\left(\hat{\eta}^\star\right),$$

valendo a igualdade somente se $\mathrm{E}\left(\mathrm{Var}(\hat{\eta}|T(\mathbf{X}))\right) = 0$. Veja que

$$\mathrm{Var}(\hat{\eta}|T(\mathbf{X})) = \mathrm{E}\left((\hat{\eta} - \mathrm{E}(\hat{\eta}|T(\mathbf{X})))^2 \,|T(\mathbf{X})\right) = \mathrm{E}\left((\hat{\eta} - \hat{\eta}^\star)^2 \,|T(\mathbf{X})\right).$$

Então,

$$\mathrm{E}\left(\mathrm{Var}(\hat{\eta}|T(\mathbf{X}))\right) = \mathrm{E}\left(\mathrm{E}\left((\hat{\eta} - \hat{\eta}^\star)^2 \,|T(\mathbf{X})\right)\right) = \mathrm{E}\left((\hat{\eta} - \hat{\eta}^\star)^2\right)$$

Assim, $\mathrm{Var}(\hat{\eta}) = \mathrm{Var}\left(\hat{\eta}^\star\right)$ se e somente se $\mathrm{E}\left(\mathrm{Var}(\hat{\eta}|T(\mathbf{X}))\right) = 0$, ou seja, se e somente se $\mathrm{E}\left((\hat{\eta} - \hat{\eta}^\star)^2\right) = 0$. Como $(\hat{\eta} - \hat{\eta}^\star)^2$ é uma variável aleatória não negativa,

$$\mathrm{E}\left((\hat{\eta} - \hat{\eta}^\star)^2\right) = 0 \iff \mathrm{P}((\hat{\eta} - \hat{\eta}^\star)^2 = 0) = 1 \iff \mathrm{P}(\hat{\eta} - \hat{\eta}^\star = 0) = 1 \iff \mathrm{P}(\hat{\eta} = \hat{\eta}^\star) = 1.$$

Isso significa que $\mathrm{Var}(\hat{\eta}) = \mathrm{Var}\left(\hat{\eta}^\star\right)$ se e somente se $\hat{\eta}$ e $\hat{\eta}^\star$ são iguais p.q.c.

□

Vejamos um exemplo da aplicação do Teorema de Rao-Blackwell.

Exemplo 4.18. *Seja* X_1, X_2, …, X_n *uma amostra aleatória de* $X \sim Bernoulli(p)$ *e* $\hat{p} = X_1$.

(a) Veja que $\hat{p}$ *é estimador não tendencioso para* p.

(b) Encontre uma estatística suficiente para p*, denominada* $T(\mathbf{X})$.

(c) Encontre $\hat{p}^\star = \mathrm{E}(\hat{p}|T(\mathbf{X}))$.

Solução:

(a) Veja que $\mathrm{E}(\hat{p}) = \mathrm{E}(X_1) = p$. Logo $\hat{p}$ é não tendencioso para p.

(b) Veja que a distribuição de X pertence à família exponencial.

$$\begin{aligned} p_X(x) &= p^x(1-p)^{1-x}, \ x = 0,1 \\ &= e^{x\ln(p)}e^{(1-x)\ln(1-p)}, \ x = 0,1 \\ &= e^{x\ln(p)}e^{ln(1-p)}e^{-x\ln(1-p)}, \ x = 0,1 \\ &= e^{x(\ln(p)-\ln(1-p))}e^{ln(1-p)}, \ x = 0,1 \end{aligned}$$

sendo, de acordo com a Definição 2.5, $c(p) = \ln(p) - \ln(1-p)$, $T(x) = x$, $d(p) = ln(1-p)$, $S(x) = 0$ e $A = \{0,1\}$. Então, $\sum_{i=1}^n X_i$ é uma estatística suficiente para o parâmetro p. Inclusive esse resultado já tinha sido visto no Exemplo 4.5.

(c) Seja $T(\mathbf{X}) = \sum_{i=1}^n X_i$.

$$\begin{aligned} \hat{p}^\star = \mathrm{E}(\hat{p}|T(\mathbf{X})) = \mathrm{E}\left(X_1 | \sum_{i=1}^n X_i\right) &= \mathrm{E}\left(X_1 | \sum_{i=1}^n X_i = t\right)\Bigg|_{t=\sum_{i=1}^n X_i} \\ &= \sum_{x_1=0}^{1} x_1 p_{X_1|\sum_{i=1}^n X_i}(x_1|t)\Bigg|_{t=\sum_{i=1}^n X_i} \\ &= \mathrm{P}\left(X_1 = 1 \mid \sum_{i=1}^n X_i = t\right)\Bigg|_{t=\sum_{i=1}^n X_i} \\ &= \frac{\mathrm{P}\left(X_1 = 1 \ \cap \ \sum_{i=1}^n X_i = t\right)}{\mathrm{P}\left(\sum_{i=1}^n X_i = t\right)}\Bigg|_{t=\sum_{i=1}^n X_i} \\ &= \frac{\mathrm{P}\left(\sum_{i=1}^n X_i = t \mid X_1 = 1\right)\mathrm{P}\left(X_1 = 1\right)}{\mathrm{P}\left(\sum_{i=1}^n X_i = t\right)}\Bigg|_{t=\sum_{i=1}^n X_i} \\ &= \frac{\mathrm{P}\left(\sum_{i=2}^n X_i = t-1\right)\mathrm{P}\left(X_1 = 1\right)}{\mathrm{P}\left(\sum_{i=1}^n X_i = t\right)}\Bigg|_{t=\sum_{i=1}^n X_i} \end{aligned}$$

Veja que $\sum_{i=1}^n X_i \sim Bin(n,p)$ e que $\sum_{i=2}^n X_i \sim Bin(n-1,p)$. Então, continuando,

$$\begin{aligned} \hat{p}^\star = & \frac{\binom{n-1}{t-1} p^{t-1}(1-p)^{n-1-t+1}p}{\binom{n}{t} p^t(1-p)^{n-t}}\Bigg|_{t=\sum_{i=1}^n X_i} & = \frac{\binom{n-1}{t-1}}{\binom{n}{t}}\Bigg|_{t=\sum_{i=1}^n X_i} \\ = & \frac{(n-1)!}{(t-1)!(n-1-t+1)!}\frac{t!(n-t)!}{n!}\Bigg|_{t=\sum_{i=1}^n X_i} & = \frac{t}{n}\Bigg|_{t=\sum_{i=1}^n X_i} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}. \end{aligned}$$

$\|$

No exemplo acima a partir da estatística suficiente $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ partimos de $\hat{p} = X_1$, um estimador não tendencioso bem ruim, e encontramos $\hat{p}^\star = \mathrm{E}(\hat{p}|T(\mathbf{X})) = \bar{X}$, outro estimador não tendencioso bem melhor que o inicial. E se quiséssemos realizar o procedimento novamente, partindo agora de $\hat{p}^\star$, será chegamos em um estimador ainda melhor? Ou seja, será que $\mathrm{E}(\hat{p}^\star|T(\mathbf{X}))$ tem menor variância que $\hat{p}^\star$? Na verdade $\mathrm{E}(\hat{p}^\star|T(\mathbf{X})) = \hat{p}^\star$, isso porque $\hat{p}^\star$ já é função de $T(\mathbf{X})$, então em nada melhoramos se realizamos o procedimento mais uma vez. Veja esse resultado de forma geral.

Seja $\hat{\theta}$ um estimador não tendencioso que é função de uma estatística suficiente $T(\mathbf{X})$, isto é, $\hat{\theta} = W(T(\mathbf{X}))$. Vamos encontrar $\hat{\theta}^\star = \mathrm{E}(\hat{\theta}|T(\mathbf{X}))$.

$$\begin{aligned}
\hat{\theta}^\star &= \mathrm{E}(\hat{\theta}|T(\mathbf{X})) = \mathrm{E}(W(T(\mathbf{X}))|T(\mathbf{X})) = \mathrm{E}(W(T(\mathbf{X}))|T(\mathbf{X}) = t)|_{t=T(\mathbf{X})} \\
&= \mathrm{E}(W(t)|T(\mathbf{X}) = t)|_{t=T(\mathbf{X})} = W(t)|_{t=T(\mathbf{X})} = W(T(\mathbf{X})) = \hat{\theta}.
\end{aligned}$$

O Teorema de Rao-Blackwell nos mostra que qualquer estimador não-tendenciosos que não é função de uma estatística suficiente pode ser melhorado a partir de uma estatística suficiente para o parâmetro que está sendo estimado. O Teorema de Lehmann-Scheffé, apresentado a seguir, completa o resultado ao mostrar que se a estatística suficiente $T(\mathbf{X})$ também for completa garantimos que o estimador $\hat{\theta}^\star = \mathrm{E}(\hat{\theta}|T(\mathbf{X}))$ é o ENVVUM, ou seja, chegamos na menor variância possível entre todos os estimadores não tendencioso.

Antes da apresentação desse último teorema precisamos definir mais um conceito, o de Estatística Completa.

Definição 4.4. *Seja X_1, X_2, ..., X_n uma amostra aleatória da população X, cuja distribuição depende do parâmetro θ. Seja $T(\mathbf{X})$ uma estatística e $f_T(t|\theta)$ a sua função de densidade ou função de probabilidade, que também depende do parâmetro θ. Dizemos que $T(\mathbf{X})$ é uma estatística completa se* $\mathrm{E}(g(T(\mathbf{X})) = 0 \ \forall\ \theta \in \Psi \ \Rightarrow\ \mathrm{P}\left(g(T(\mathbf{X})) = 0\right) = 1,\ \forall\ \theta \in \Psi$.

Segunda a definição, uma estatística $T(\mathbf{X})$ é completa se a única função g para a qual $\mathrm{E}(g(T(\mathbf{X})) = 0\ \forall\ \theta \in \Psi$ é a função nula. Repare que $\mathrm{E}(g(T(\mathbf{X}))$ depende de θ, uma vez que a distribuição de $T(\mathbf{X})$ depende de θ.

A boa notícia é que para a família exponencial unidimensional temos um resultado imediato: a mesma estatística $\sum_{i=1}^{n} T(X_i)$ que já vimos ser suficiente para o parâmetro da distribuição também será uma estatística completa. Veja Teorema 4.7.

Teorema 4.7. *Suponha que a distribuição da população X pertença à família exponencial unidimensional. Então, $\sum_{i=1}^{n} T(X_i)$ além de suficiente para o parâmetro θ também é estatística completa, desde que o conjunto $\{c(\theta)|\theta \in \Psi\}$ contenha um intervalo da reta.*

A demonstração do Teorema 4.7 será omitida, mas pode ser encontrada em [Lehmann e Romano, 2006]. A exigência de que o conjunto $\{c(\theta)|\theta \in \Psi\}$ contenha um intervalo da reta é necessária para evitar situações raras em que a estatística $\sum_{i=1}^{n} T(X_i)$ não é completa. Em todos os exemplos que veremos nesse texto o conjunto $\{c(\theta)|\theta \in \Psi\}$ contém um intervalo da reta.

Vejamos um exemplo fora da família exponencial.

Exemplo 4.19. *Seja X_1, X_2, ..., X_n uma amostra aleatória da população $X \sim U(0,\theta)$ e $T(\mathbf{X}) = X_{(n)}$. Mostre que T é uma estatística completa.*

Solução:

Para o exemplo $\Psi = \mathbb{R}^+$. Seja g uma função tal que $\mathrm{E}(g(X_{(n)})) = 0\ \forall\ \theta > 0$.

$$\mathrm{E}(g(X_{(n)})) = \int_0^\theta g(x) f_{X_{(n)}}(x)\mathrm{d}x = \int_0^\theta g(x)\frac{nx^{n-1}}{\theta^n}\mathrm{d}x = \frac{n}{\theta^n}\int_0^\theta g(x)x^{n-1}\mathrm{d}x = 0\ ,\ \forall\ \theta > 0$$

Então é verdade que $\int_0^\theta g(x)x^{n-1}\mathrm{d}x = 0 \ \forall\ \theta > 0$. Seja G a função tal que $G'(x) = g(x)x^{n-1}$. Logo,

$$\int_0^\theta g(x)x^{n-1}\mathrm{d}x = 0 \quad \Rightarrow \quad G(\theta) - G(0) = 0 \quad \Rightarrow \quad G(\theta) = G(0)\ ,\ \forall\ \theta > 0$$

Isto é, $G(\theta)$ é uma função constante e com isso $G'(\theta) = 0 \ \forall\ \theta > 0$. Ou seja, $g(x)x^{n-1} = 0 \ \forall\ x > 0$ $\Rightarrow g(x) = 0 \ \forall\ x > 0 \Rightarrow \mathrm{P}(g(T(\mathbf{X})) = 0) = 1$. Com isso mostramos que $X_{(n)}$ é uma estatística completa em relação ao parâmetro θ. ‖

Vejamos o último teorema desse capítulo, que nos mostra como identificar que chegamos ao ENVVUM.

Teorema 4.8. ***Lehmann-Scheffé***

Seja $\hat{\eta}$ um estimador não tendencioso para $\eta = \tau(\theta)$ e $T(\mathbf{X})$ uma estatística suficiente e completa para θ. Defina

$$\hat{\eta}^\star = \mathrm{E}(\hat{\eta}|T(\mathbf{X})).$$

Então, $\hat{\eta}^\star$ é o único estimador não tendencioso para $\eta = \tau(\theta)$ baseado na estatística $T(\mathbf{X})$. Além disso, ele é o estimador não viciado de variância uniformemente mínima (ENVVUM) para $\eta = \tau(\theta)$.

Demonstração. A demonstração do Teorema de Lehmann-Scheffé será dividida em duas partes: (1) mostrar que $\hat{\eta}^\star$ é o único estimador não tendencioso para $\eta = \tau(\theta)$ baseado na estatística $T(\mathbf{X})$; (2) mostrar que $\hat{\eta}^\star$ é o ENVVUM. Vamos começar com a parte 1.

Já sabemos pelo Teorema de Rao-Blackwell que $\hat{\eta}^\star = \mathrm{E}(\hat{\eta}|T(\mathbf{X}))$ é um estimador não tendencioso para η baseado na estatística $T(\mathbf{X})$. Isto é, $\mathrm{E}(\hat{\eta}^\star) = \eta$ e $\hat{\eta}^\star = g(T(\mathbf{X}))$. Vamos supor que exista outro estimador $\tilde{\eta}$ que seja não tendencioso e baseado na estatística suficiente e completa $T(\mathbf{X})$. Isto é, $\tilde{\eta} = h(T(\mathbf{X}))$ e $\mathrm{E}(\tilde{\eta}) = \eta$. Então,

$$0 = \mathrm{E}(\hat{\eta}^\star) - \mathrm{E}(\tilde{\eta}) = \mathrm{E}(\hat{\eta}^\star - \tilde{\eta}) = \mathrm{E}(g(T(\mathbf{X})) - h(T(\mathbf{X}))) = \mathrm{E}((g-h)(T(\mathbf{X})))$$

Como $T(\mathbf{X})$ é uma estatística completa, isso implica que a função $g - h$ é a função nula, ou seja, $g(T(\mathbf{X})) - h(T(\mathbf{X})) = 0 \Rightarrow g(T(\mathbf{X})) = h(T(\mathbf{X})) \Rightarrow \hat{\eta}^\star = \tilde{\eta}$. E assim mostramos que $\hat{\eta}^\star$ é o único estimador não tendencioso para $\eta = \tau(\theta)$ baseado na estatística $T(\mathbf{X})$.

Só falta mostrar que $\hat{\eta}^\star$ é o ENVVUM. Seja $\tilde{\eta}$ um estimador não tendencioso para η qualquer. Já sabemos pelo Teorema de Rao-Blackwell que $\mathrm{E}(\tilde{\eta}|T(\mathbf{X}))$ é um estimador não tendencioso baseado na estatística $T(\mathbf{X})$. Aplicando a parte 1 dessa demonstração, obrigatoriamente $\mathrm{E}(\tilde{\eta}|T(\mathbf{X})) = \hat{\eta}^\star$. Novamente pelo Teorema de Rao-Blackwell sabemos que $\mathrm{Var}(\hat{\eta}^\star) \leq \mathrm{Var}(\tilde{\eta})$. Dessa forma, $\hat{\eta}^\star$ é o ENVVUM. □

O Teorema de Lehmann-Scheffé nos mostra que se a estatística $T(\mathbf{X})$ for suficiente e completa, partindo de qualquer estimador não tendencioso para η, vamos chamá-lo de $\hat{\eta}$, $\hat{\eta}^\star = \mathrm{E}(\hat{\eta}|T(\mathbf{X}))$ é o ENVVUM. Mas ele nos diz outra coisa também. Esse teorema nos mostra que se encontramos algum estimador não tendencioso para η que seja baseado em uma estatística suficiente e completa para η já sabemos que ele é o ENVVUM. Isso devido ao fato do ENVVUM é o único estimador não tendencioso baseado em uma estatística suficiente e completa.

Exemplo 4.20. *Seja X_1, X_2, $\ldots$, X_n uma amostra aleatória de $X \sim U(0,\theta)$. Encontre o ENVVUM para θ.*

Solução:

Precisamos apenas encontrar um estimador não tendencioso baseado em uma estatística suficiente e completa para θ. Já vimos nos Exemplos 4.7 e 4.19 que $X_{(n)}$ é suficiente e completa para θ. Uma boa dica para buscar um estimador baseado em uma estatística suficiente e completa é pensar no estimador de máxima verossimilhança. Revisitando o Exemplo 3.23 temos o estimador de máxima verossimilhança para θ dado por

$$\hat{\theta}_{MV} = X_{(n)} \qquad \text{e} \qquad \mathrm{E}(\hat{\theta}_{MV}) = \mathrm{E}(X_{(n)}) = \frac{n}{n+1}\theta.$$

Com isso concluímos que o estimador de máxima verossimilhança é tendencioso, logo não pode ser o ENVVUM. Mas repare que podemos modificá-lo para troná-lo não tendencioso, basta multiplicarmos pela constante $\dfrac{n+1}{n}$:

$$\hat{\theta} = \frac{n+1}{n}X_{(n)} \quad \Rightarrow \quad \mathrm{E}(\hat{\theta}) = \mathrm{E}\left(\frac{n+1}{n}X_{(n)}\right) = \frac{n+1}{n}\,\mathrm{E}(X_{(n)}) = \frac{n+1}{n}\frac{n}{n+1}\theta = \theta.$$

Então, $\hat{\theta} = \dfrac{n+1}{n}X_{(n)}$ é um estimador não tendencioso baseado em uma estatística suficiente e completa para θ. Pelo Teorema de Lehmann-Scheffé, $\hat{\theta} = \dfrac{n+1}{n}X_{(n)}$ é o ENVVUM para θ. $\|$

Exemplo 4.21. *Seja X_1, X_2, ..., X_n uma amostra aleatória de $X \sim Poisson(\lambda)$. Encontre o ENVVUM para $\eta = \mathrm{P}(X = 0)$.*

Solução:

Primeiro veja que queremos o ENVVUM para o parâmetro $\eta = \mathrm{P}(X = 0) = e^{-\lambda}$. Então precisamos encontrar um estimador não tendencioso para η que seja baseado em uma estatística completa e suficiente para λ.

Como $X \sim Poisson(\lambda)$ pertence à família exponencial é fácil encontrar uma estatística suficiente e completa para λ. Já vimos que $\sum_{i=1}^n X_i$ é uma estatística suficiente e completa para λ, reveja os Exemplos 2.6, 4.3 e 4.7.

Nesse exemplo não vai ser útil pensar no estimador de máxima verossimilhança. Ele é tendencioso e não é fácil modificá-lo para que ele fique não tendencioso. Vamos então buscar um estimador não tendencioso qualquer e encontrar o ENVVUM pelo resultado de $\hat{\eta}^\star = \mathrm{E}(\hat{\eta}|T(\mathbf{X}))$, sendo $T(\mathbf{X}) = \sum_{i=1}^n X_i$ a estatística suficiente e completa já apresentada.

Veja que o estimador $\hat{\eta}$ apresentado a seguir é não tendencioso para $\eta = e^{-\lambda}$.

$$\hat{\eta} = \begin{cases} 1 & , X_1 = 0 \\ 0 & , \text{caso contrário.} \end{cases} \quad \Rightarrow \quad \mathrm{E}(\hat{\eta}) = \mathrm{P}(\hat{\eta} = 1) = \mathrm{P}(X_1 = 0) = e^{-\lambda}.$$

Vamos às contas para encontrar o ENVVUM para η:

$$\begin{aligned}
\hat{\eta}^\star &= \mathrm{E}(\hat{\eta}|T(\mathbf{X})) &&= \mathrm{E}\left(\hat{\eta}|\sum_{i=1}^n X_i = t\right)\Bigg|_{t=\sum_{i=1}^n X_i} \\
&= \mathrm{P}\left(\hat{\eta} = 1|\sum_{i=1}^n X_i = t\right)\Bigg|_{t=\sum_{i=1}^n X_i} &&= \mathrm{P}\left(X_1 = 0|\sum_{i=1}^n X_i = t\right)\Bigg|_{t=\sum_{i=1}^n X_i} \\
&= \frac{\mathrm{P}\left(X_1 = 0 \cap \sum_{i=1}^n X_i = t\right)}{\mathrm{P}\left(\sum_{i=1}^n X_i = t\right)}\Bigg|_{t=\sum_{i=1}^n X_i} &&= \frac{\mathrm{P}\left(\sum_{i=1}^n X_i = t|X_1 = 0\right)\mathrm{P}(X_1 = 0)}{\mathrm{P}\left(\sum_{i=1}^n X_i = t\right)}\Bigg|_{t=\sum_{i=1}^n X_i} \\
&= \frac{\mathrm{P}\left(\sum_{i=2}^n X_i = t\right)\mathrm{P}(X_1 = 0)}{\mathrm{P}\left(\sum_{i=1}^n X_i = t\right)}\Bigg|_{t=\sum_{i=1}^n X_i}.
\end{aligned}$$

Veja que $X_1 \sim Poissin(\lambda)$, $\sum_{i=2}^n X_i \sim Poisson((n-1)\lambda)$ e $\sum_{i=1}^n X_i \sim Poisson(n\lambda)$.

$$\begin{aligned}
\hat{\eta}^\star &= \left.\frac{\mathrm{P}\left(\sum_{i=2}^n X_i = t\right)\mathrm{P}\left(X_1 = 0\right)}{\mathrm{P}\left(\sum_{i=1}^n X_i = t\right)}\right|_{t=\sum_{i=1}^n X_i} &&= \left.\frac{(n-1)^t\lambda^t e^{-(n-1)\lambda}e^{-\lambda}}{t!}\frac{t!}{n^t\lambda^t e^{-n\lambda}}\right|_{t=\sum_{i=1}^n X_i} \\
&= \left.\frac{(n-1)^t e^{-n\lambda}e^{\lambda}e^{-\lambda}}{n^t e^{-n\lambda}}\right|_{t=\sum_{i=1}^n X_i} &&= \left.\frac{(n-1)^t}{n^t}\right|_{t=\sum_{i=1}^n X_i} = \left.\left(\frac{n-1}{n}\right)^t\right|_{t=\sum_{i=1}^n X_i} \\
&= \left(\frac{n-1}{n}\right)^{\sum_{i=1}^n X_i}.
\end{aligned}$$

Chegamos então à conclusão de que o ENVVUM para η é $\hat{\eta}^\star = \left(\frac{n-1}{n}\right)^{\sum_{i=1}^n X_i}$. Repara que verificar que se trata de um estimador não tendencioso não é fácil, mas o Teorema de Rao-Blackwell nos garante isso. E como ele é não tendencioso e baseado em uma estatística suficiente e completa para λ, ele é o ENVVUM. $\|$

Exemplo 4.22. *Seja X_1, X_2, ..., X_n uma amostra aleatória de $X \sim Exp(\lambda)$. Encontre os ENVVUM para λ e $\eta = P(X > 1)$.*

Solução:

Primeiro vamos encontrar o ENVVUM para λ. Como $X \sim Exp(\lambda)$ pertence à família exponencial, é fácil encontrar uma estatística suficiente e completa para λ. Podemos afirmar que $\sum_{i=1}^n X_i$ é uma estatística suficiente e completa para λ, reveja o Exemplo 2.5. Procuramos um estimador não tendencioso para λ baseado nessa estatística suficiente e completa. Revendo o Exemplo 3.9 percebemos que o estimador de máxima verossimilhança,

$$\hat{\lambda}_{MV} = \frac{1}{\bar{X}} = \frac{n}{\sum_{i=1}^n X_i},$$

é baseado na estatística suficiente e completa $\sum_{i=1}^n X_i$. Veja que conhecemos a distribuição amostral dessa estatística, $\sum_{i=1}^n X_i \sim Gama(n,\lambda)$. Vejamos agora se o estimador de máxima verossimilhança é ou não tendencioso.

$$\begin{aligned}
\mathrm{E}\left(\hat{\lambda}_{MV}\right) = \mathrm{E}\left(\frac{n}{\sum_{i=1}^n X_i}\right) &= \int_0^\infty \frac{n}{w}\frac{\lambda^n}{\Gamma(n)}w^{n-1}e^{-\lambda w}\mathrm{d}w \\
&= \frac{n\lambda^n}{\Gamma(n)}\int_0^\infty w^{n-2}e^{-\lambda w}\mathrm{d}w \\
&= \frac{n\lambda^n}{\Gamma(n)}\frac{\Gamma(n-1)}{\lambda^{n-1}} \\
&= \frac{n}{n-1}\lambda.
\end{aligned}$$

Então $\hat{\lambda}_{MV}$ é tendencioso e por isso ele não pode ser o ENVVUM. Mas é possível transformá-lo em um estimador não tendencioso.

$$\hat{\lambda}^\star = \frac{n-1}{n}\hat{\lambda}_{MV} = \frac{n-1}{n}\frac{n}{\sum_{i=1}^n X_i} = \frac{n-1}{\sum_{i=1}^n X_i} \Rightarrow \mathrm{E}(\hat{\lambda}^\star) = \frac{n-1}{n}\mathrm{E}\left(\hat{\lambda}_{MV}\right) = \frac{n-1}{n}\frac{n}{n-1}\lambda = \lambda.$$

Chegamos então à conclusão de que o ENVVUM para λ é $\hat{\lambda}^\star = \dfrac{n-1}{\sum_{i=1}^n X_i}$.

Agora vamos encontrar o ENVVUM para o parâmetro $\eta = \mathrm{P}(X > 1) = 1 - F_X(1) = e^{-\lambda}$. Precisamos encontrar um estimador não tendencioso para η que seja baseado em uma estatística completa e suficiente para λ, que já vimos ser $\sum_{i=1}^n X_i$. Para esse caso, o estimador de máxima verossimilhança para η não será útil. Vamos então partir de um estimador não tendencioso qualquer, quanto mais simples mais fácil serão as contas. Considere

$$\hat{\eta} = \begin{cases} 1 & , X_1 > 1 \\ 0 & , \text{caso contrário.} \end{cases} \quad \Rightarrow \ \mathrm{E}(\hat{\eta}) = \mathrm{P}(\eta = 1) = \mathrm{P}(X_1 > 1) = \mathrm{P}(X > 1) = 1 - F_X(1) = e^{-\lambda}.$$

Vamos encontrar o ENVVUM para η pelo resultado de $\hat{\eta}^\star = \mathrm{E}(\hat{\eta}|T(\mathbf{X})) = \mathrm{E}(\hat{\eta}|T(\mathbf{X}) = t)|_{t=T(\mathbf{X})}$.

$$\begin{aligned}
\hat{\eta}^\star &= \mathrm{E}(\hat{\eta}|T(\mathbf{X}) = t)|_{t=T(\mathbf{X})} &&= \mathrm{P}(\hat{\eta} = 1|T(\mathbf{X}) = t)|_{t=T(\mathbf{X})} \\
&= \mathrm{P}\left(X_1 > 1 | \sum_{i=1}^n X_i = t\right)\Bigg|_{t=\sum_{i=1}^n X_i} &&= \int_1^t f_{X_1|\sum_{i=1}^n X_i}(x_1|t)\mathrm{d}x_1\Bigg|_{t=\sum_{i=1}^n X_i} \\
&= \int_1^t \frac{f_{X_1,\sum_{i=1}^n X_i}(x_1,t)}{f_{\sum_{i=1}^n X_i}(t)}\mathrm{d}x_1\Bigg|_{t=\sum_{i=1}^n X_i} &&= \int_1^t \frac{f_{\sum_{i=1}^n X_i|X_1}(t|x_1)f_{X_1}(x_1)}{f_{\sum_{i=1}^n X_i}(t)}\mathrm{d}x_1\Bigg|_{t=\sum_{i=1}^n X_i} \\
&= \int_1^t \frac{f_{\sum_{i=2}^n X_i}(t-x_1)f_{X_1}(x_1)}{f_{\sum_{i=1}^n X_i}(t)}\mathrm{d}x_1\Bigg|_{t=\sum_{i=1}^n X_i}
\end{aligned}$$

Veja que $X_1 \sim Exp(\lambda)$, $\sum_{i=2}^n X_i \sim Gama(n-1,\lambda)$ e $\sum_{i=1}^n X_i \sim Gama(n,\lambda)$. Podemos então continuar.

$$\begin{aligned}
\hat{\eta}^\star &= \int_1^t \frac{f_{\sum_{i=2}^n X_i}(t-x_1)f_{X_1}(x_1)}{f_{\sum_{i=1}^n X_i}(t)}\mathrm{d}x_1\Bigg|_{t=\sum_{i=1}^n X_i} \\
&= \int_1^t \lambda e^{-\lambda x_1}\frac{\lambda^{n-1}(t-x_1)^{n-2}e^{-\lambda(t-x_1)}}{\Gamma(n-1)}\frac{\Gamma(n)}{\lambda^n t^{n-1}e^{-\lambda t}}\mathrm{d}x_1\Bigg|_{t=\sum_{i=1}^n X_i} \\
&= \int_1^t \frac{\Gamma(n)}{\Gamma(n-1)}e^{-\lambda x_1}\frac{(t-x_1)^{n-2}e^{-\lambda t}e^{\lambda x_1}}{t^{n-1}e^{-\lambda t}}\mathrm{d}x_1\Bigg|_{t=\sum_{i=1}^n X_i} \\
&= \int_1^t (n-1)\frac{(t-x_1)^{n-2}}{t^{n-1}}\mathrm{d}x_1\Bigg|_{t=\sum_{i=1}^n X_i} \\
&= \frac{(n-1)}{t^{n-1}}\int_1^t (t-x_1)^{n-2}\mathrm{d}x_1\Bigg|_{t=\sum_{i=1}^n X_i} \\
&= \frac{(n-1)}{t^{n-1}}\int_{t-1}^0 y^{n-2}(-1)\mathrm{d}y\Bigg|_{t=\sum_{i=1}^n X_i} \\
&= \frac{(n-1)}{t^{n-1}}\left.\frac{-y^{n-1}}{(n-1)}\right|_{t-1}^0\Bigg|_{t=\sum_{i=1}^n X_i} \\
&= \frac{(n-1)}{t^{n-1}}\frac{(t-1)^{n-1}}{(n-1)}\Bigg|_{t=\sum_{i=1}^n X_i} \\
&= \left(\frac{t-1}{t}\right)^{n-1}\Bigg|_{t=\sum_{i=1}^n X_i} \\
&= \left(\frac{\sum_{i=1}^n X_i - 1}{\sum_{i=1}^n X_i}\right)^{n-1}
\end{aligned}$$

Chegamos então à conclusão de que o ENVVUM para $\eta = \mathrm{P}(X > 1)$ é $\left(\frac{\sum_{i=1}^n X_i - 1}{\sum_{i=1}^n X_i}\right)^{n-1}$.

$\|$

Exercícios da Seção 4.4

4.4.1 Esse exercício é continuação do Exercício 4.3.1. Seja $X_1, \ldots, X_n$ uma amostra aleatória da população X. Para cada item a seguir encontre o ENVVUM para o parâmetro da distribuição da população. Em seguida, indique se o ENVVUM é eficiente.

(a) $X \sim N(0, \sigma^2)$.

(b) $X \sim Binomial(5, \theta)$.

(c) X é tal que $f(x|\theta) = e^{-(x-\theta)}$, $x > \theta$. Obs: Assuma sabido que $X_{(1)}$ é uma estatística suficiente e completa para o parâmetro θ.

4.4.2 ([Bolfarine e Sandoval, 2001] - Seção 2.6) Sejam $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória $X \sim N(\mu, 1)$.

(a) Mostre que o estimador $\hat{\eta} = \bar{X}^2 - \frac{1}{n}$ é não viciado para $\eta = \mu^2$.

(b) Existe ENVVUM para μ^2?

(c) Encontre o limite inferior da variância dos estimadores não viciados de $\eta = \mu^2$ e verifique se $\hat{\eta}$ é eficiente.

4.4.3 ([Bolfarine e Sandoval, 2001] - Seção 2.6) Sejam $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória $X \sim Bernoulli(\theta)$. Obtenha o ENVVUM para $\eta = \theta(1-\theta) = \text{Var}(X)$. Sugestão: verifique se $S^2 = \dfrac{n}{n-1}\bar{X}(1-\bar{X})$ é não viciado para $\theta(1-\theta)$.

4.4.4 ([Bolfarine e Sandoval, 2001] - Seção 2.6) Sejam $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória X com distribuição geométrica alternativa com parâmetro θ, isto é, sua função de probabilidade é

$$p(x|\theta) = \theta(1-\theta)^x, \quad x = 0, 1, 2, \ldots \quad 0 < \theta < 1.$$

Encontre o ENVVUM para $1/\theta$.

4.4.5 Seja $X_1, \ldots, X_n$ uma amostra aleatória de $X \sim Exp(\lambda)$. Dica: Se preciso "Fabrique" um estimador não tendencioso a partir do estimador de máxima verossimilhança.

(a) Encontre o ENVVUM de λ. Esse estimador é eficiente?

(b) Encontre o ENVVUM de $\mu = \text{E}[X]$. Esse estimador é eficiente?

(c) Encontre o ENVVUM de $\theta = \text{Var}(X)$. Esse estimador é eficiente?

(d) Encontre o ENVVUM de $\eta = \text{P}(X > 2)$. Esse estimador é eficiente? (Dica: Reveja o Exemplo 4.22.)

4.4.6 ([Casella e Berger, 2002] - Capítulo 7) Considere uma amostra aleatória $X_1, \ldots, X_n$ de $X \sim Gama(\alpha_0, \lambda)$, com α_0 conhecido. Encontre o ENVVUM para $\eta = \frac{1}{\lambda}$.

4.4.7 Seja X_1, ..., X_n amostra aleatória de X com função densidade $f_X(x) = \theta x^{\theta-1}$, $0 < x < 1$ e $\theta > 0$. Encontre o ENVVUM para θ e diga se este estimador é eficiente.
Dica: Reveja os Exercícios 2.1.11, 3.3.4 e 3.4.3.

4.5 Mais Alguns Exercícios do Capítulo 4

4.1. Seja $X_1, \ldots, X_n$ uma amostra aleatória da população $X \sim Gama(\alpha_0, \lambda)$, sendo $\alpha_0 > 0$ um parâmetro conhecido. Defina $\mu = \text{E}(X)$ e $\eta = \text{Var}(X)$.

(a) Encontre a informação de Fisher para o parâmetro λ da distribuição de X.

(b) Encontre a variância mínima dada pelo limite de Cramér-Rao para: os estimadores não tendenciosos para λ; os estimadores não tendenciosos para μ; os estimadores não tendenciosos para η.

(c) Existe estimador eficiente para λ, para μ e para η?

(d) Encontre uma estatística suficiente e completa para λ.

(e) Encontre o ENVVUM para λ, o ENVVUM para μ e o ENVVUM para η.

(f) Considere $\alpha_0 = 3$ e a seguinte amostra da população X:

$$1,61 \qquad 5,04 \qquad 4,92 \qquad 3,20 \qquad 6,73 \qquad 3,80 \qquad 3,49 \qquad 2,04 \qquad 1,40$$

Encontre estimativas para os parâmetros λ, μ e η a partir dos estimadores encontrados nos itens acima.

4.2. Seja X uma variável aleatória discreta cuja função de probabilidade depende do parâmetro $0 < p < 1$ e é definida por

$$p_X(x) = \binom{15}{x} p^x (1-p)^{15-x}, \quad x = 0, 1, \ldots, 15$$

Veja que $X \sim Binom(15, p)$ e que esta distribuição é a mesma dos Exercícios 2.5.5 e 3.15. Defina $\mu = \mathrm{E}(X)$ e $\eta = \mathrm{Var}(X)$.

(a) Encontre a informação de Fisher para o parâmetro p da distribuição de X.

(b) Encontre a variância mínima dada pelo limite de Cramér-Rao para: os estimadores não tendenciosos para p; os estimadores não tendenciosos para μ; os estimadores não tendenciosos para η.

(c) Existe estimador eficiente para p, para μ e para η?

(d) Encontre uma estatística suficiente e completa para p.

(e) Encontre o ENVVUM para θ, o ENVVUM para μ e o ENVVUM para η.

(f) Considere a seguinte amostra da população X:

$$3 \quad 3 \quad 4 \quad 6 \quad 2 \quad 6 \quad 7 \quad 4 \quad 4 \quad 1 \quad 2 \quad 2 \quad 5 \quad 3 \quad 5 \quad 4 \quad 5 \quad 8 \quad 3 \quad 5 \quad 6 \quad 2 \quad 4 \quad 2 \quad 3$$

Encontre estimativas para os parâmetros p, μ e η a partir dos estimadores encontrados nos itens acima.

Capítulo 5

Estimação por Intervalo

Depois de já estudado estimação pontual, vamos aprender sobre a estimação por intervalos, ou intervalar. A ideia principal é fornecer uma estimativa em forma de intervalo, e não mais um número real. A estimativa por intervalo, apesar de perder a precisão, ela ganha confiança.

5.1 Conceitos Gerais

Ao estudar a estimação pontual, nos últimos dois capítulos, aprendemos como encontrar estimadores para parâmetros desconhecido e também como avaliar tais estimadores. Diante de um problema de estimação, podemos encontrar uma estimativa para o parâmetro desconhecido a partir de um estimador pontual, e essa estimativa será um número real. Mas qual a certeza que você tem sobre essa estimativa? Você acredita que o valor desconhecido do parâmetro coincide com o valor encontrado para a estimativa? Provavelmente não. Se recolheremos outra amostra e realizarmos as mesmas contas, a estimativa encontrada será outra.

A ideia de fornecer uma estimativa em forma de intervalo é justamente para associar a ela uma medida de confiança. Vamos trocar a estimativa pontual, que é um número real, sem confiança alguma por uma estimativa intervalar com alguma confiança associada.

Definição 5.1. *Duas estatísticas $L_1(\mathbf{X})$ e $L_2(\mathbf{X})$ definem um intervalo de confiança bilateral para o parâmetro θ com coeficiente de confiança $1-\alpha$, ou $100(1-\alpha)\%$, se*

$$\mathrm{P}(L_1(\mathbf{X}) \leq \theta \leq L_2(\mathbf{X})) \geq 1-\alpha.$$

Nesse caso escrevemos $IC_{\theta,100(1-\alpha)\%} = [L_1(\mathbf{X}), L_2(\mathbf{X})]$.

Se antes tínhamos um estimador para θ definido como uma estatística, $\hat{\theta} = T(\mathbf{X})$, agora temos um estimador definido como um intervalo de confiança para θ: $IC = [L_1(\mathbf{X}), L_2(\mathbf{X})]$, para o qual é possível indicar a probabilidade de $L_1(\mathbf{X}) \leq \theta$ e $L_2(\mathbf{X}) \geq \theta$. Atenção: as variáveis aleatórias em questão para as quais calcula-se as probabilidade são os limites dos intervalos, $L_1(\mathbf{X})$ e $L_2(\mathbf{X})$. Como o parâmetro θ não é variável aleatória é mais correto dizer "$\alpha - 1$ é a probabilidade do intervalo IC conter o real valor do parâmetro" do que dizer "$\alpha - 1$ é a probabilidade do parâmetro estar dentro do IC".

Para o caso bilateral, a amplitude do intervalo, definida por $L_2(\mathbf{X}) - L_1(\mathbf{X})$, pode ser entendida como uma medida de precisão: quanto menor a amplitude mais preciso é o intervalo. Já o valor de $1-\alpha$ é a confiança que temos no intervalo. Por exemplo, os estimadores pontuais podem ser entendidos como um intervalo de confiança de precisão máxima e confiança zero, e não são eles que estamos interessados nesse capítulo.

Se temos dois intervalos com mesma precisão, isto é, mesma amplitude, preferimos aquele com maior confiança. Se temos dois intervalos com mesma confiança, preferimos aqueles com maior precisão, isto é, menor amplitude.

A ideia de confiança no intervalo, que é a probabilidade do intervalo conter o valor do parâmetro, pode ser entendida da seguinte maneira. Para cada amostra recolhida é possível encontrar o valor observado de $L_1(\mathbf{x})$ e $L_2(\mathbf{x})$, e assim se obter uma estimativa para o intervalo de confiança: $[L_1(\mathbf{x}), L_2(\mathbf{x})]$, que é um intervalo numérico. Se for recolhida outra amostra, outra estimativa será encontrada. Não se conhece o valor real do parâmetro θ, mas se muitas amostras forem recolhidas e para cada uma delas for gerada uma estimativa para o intervalo de confiança, esperamos que $(1-\alpha)\%$ dos intervalos gerados contenha o valor θ.

Além do intervalo de confiança bilateral ainda é possível definir intervalos de confiança unilaterais, como apresentados nas Definições 5.2 e 5.3.

Definição 5.2. *A estatística $L_1(\mathbf{X})$ define um intervalo de confiança unilateral à direita para o parâmetro θ com coeficiente de confiança $1-\alpha$, ou $100(1-\alpha)\%$, se*

$$\mathrm{P}(L_1(\mathbf{X}) \leq \theta) \geq 1-\alpha.$$

Nesse caso escrevemos $IC_{\theta,100(1-\alpha)\%} = [L_1(\mathbf{X}), \infty)$.

Definição 5.3. *A estatística $L_2(\mathbf{X})$ define um intervalo de confiança unilateral à esquerda para o parâmetro θ com coeficiente de confiança $1-\alpha$, ou $100(1-\alpha)\%$, se*

$$\mathrm{P}(\theta \leq L_2(\mathbf{X})) \geq 1-\alpha.$$

Nesse caso escrevemos $IC_{\theta,100(1-\alpha)\%} = (-\infty, L_2(\mathbf{X})]$.

Para os intervalos unilaterais queremos apenas uma estatística, que será um dos limites do intervalo de confiança. Essa estatística é um limite inferior para θ, para o caso do intervalo unilateral à direita, ou um limite superior para θ, para o intervalo unilateral à esquerda. A ideia de amplitude (precisão) e de confiança se mentém também para os intervalos unilaterais.

Vamos para um primeiro exemplo, entes de formalizar um método para se obter intervalos de confiança.

Exemplo 5.1. *Seja $X \sim N(\mu, \sigma_0^2)$, com σ_0^2 conhecido. Deduza intervalos de confiança bilateral e unilaterais para o parâmetro μ. Em seguida, deduza estimativas para os intervalos encontrados considerando $\sigma_0^2 = 1$ e a seguinte amostra da população X:*

$$3,35 \quad 2,94 \quad 3,11 \quad 1,96 \quad 2,78 \quad 3,56 \quad 5,91 \quad 3,61 \quad 1,05 \quad 4,09.$$

Solução:

Primeiro veja que $Z = \dfrac{\bar{X}-\mu}{\sigma_0/\sqrt{n}} \sim N(0,1)$. Baseado nessa informação é possível afirmar que

$$\begin{aligned}
\mathrm{P}\left(-z_{1-\alpha/2} \leq Z \leq z_{1-\alpha/2}\right) &= 1-\alpha \quad \text{(veja Figura 5.1)}\\
\mathrm{P}\left(-z_{1-\alpha/2} \leq \frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}} \leq z_{1-\alpha/2}\right) &= 1-\alpha\\
\mathrm{P}\left(-\bar{X} - z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}} \leq -\mu \leq -\bar{X} + z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right) &= 1-\alpha\\
\mathrm{P}\left(\bar{X} - z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right) &= 1-\alpha,
\end{aligned}$$

onde $z_{1-\alpha/2} \in \mathbf{R}$ é tal que $\mathrm{P}(Z < z_{1-\alpha/2}) = 1-\alpha/2$, sendo $Z \sim N(0,1)$ (veja Figura 5.2).

De acordo com a Definição 5.1, chegamos em um intervalo de confiança bilateral para o parâmetro μ, definido por:

$$IC_{\mu,100(1-\alpha)\%} = \Bigg[\underbrace{\bar{X} - z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}}_{L_1(\mathbf{X})}\ ,\ \underbrace{\bar{X} + z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}}_{L_2(\mathbf{X})}\Bigg].$$

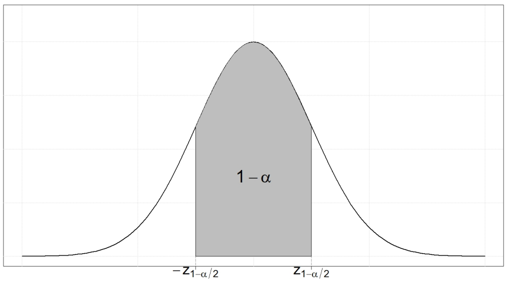


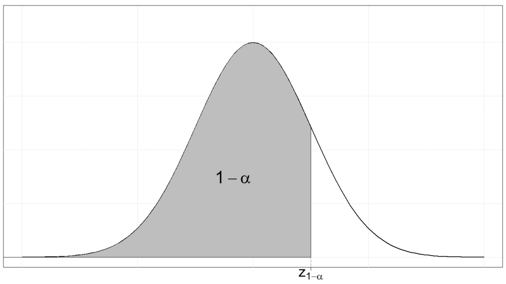
Figura 5.1: Quantis da Distribuição Normal - caso bilateral.

De forma análoga, novamente considerando $Z = \dfrac{\bar{X}-\mu}{\sigma_0/\sqrt{n}} \sim N(0,1)$, veja agora Figuras 5.2(a) e 5.2(b),

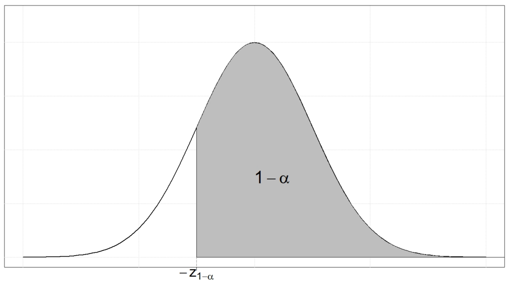
$$\begin{aligned}
\mathrm{P}\left(Z \leq z_{1-\alpha}\right) &= 1-\alpha \\
\mathrm{P}\left(\frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}} \leq z_{1-\alpha}\right) &= 1-\alpha \\
\mathrm{P}\left(-\mu \leq -\bar{X} + z_{1-\alpha}\frac{\sigma_0}{\sqrt{n}}\right) &= 1-\alpha \\
\mathrm{P}\left(\bar{X} - z_{1-\alpha}\frac{\sigma_0}{\sqrt{n}} \leq \mu\right) &= 1-\alpha \\
&\Downarrow \\
IC_{\mu,100(1-\alpha)\%} = [\underbrace{\bar{X} - z_{1-\alpha}\frac{\sigma_0}{\sqrt{n}}}_{L_1(\mathbf{X})}\ , \infty).
\end{aligned}$$

ou

$$\begin{aligned}
\mathrm{P}\left(-z_{1-\alpha} \leq Z\right) &= 1-\alpha \\
\mathrm{P}\left(-z_{1-\alpha} \leq \frac{\bar{X}-\mu}{\sigma_0/\sqrt{n}}\right) &= 1-\alpha \\
\mathrm{P}\left(-\bar{X} - z_{1-\alpha}\frac{\sigma_0}{\sqrt{n}} \leq -\mu\right) &= 1-\alpha \\
\mathrm{P}\left(\mu \leq \bar{X} + z_{1-\alpha}\frac{\sigma_0}{\sqrt{n}}\right) &= 1-\alpha \\
&\Downarrow \\
IC_{\mu,100(1-\alpha)\%} = (-\infty\ , \underbrace{\bar{X} + z_{1-\alpha}\frac{\sigma_0}{\sqrt{n}}}_{L_2(\mathbf{X})}].
\end{aligned}$$



(a)



(b)

Figura 5.2: Quantis da Distribuição Normal - casos unilaterais.

Pelas contas da esquerda chegamos à um intervalo de confiança unilateral à direita (Definição 5.2) e pelas contas da direita chegamos à um intervalo unilateral à esquerda (Definição 5.3).

Para encontrar as estimativas dos intervalos de confiança precisamos encontrar os valores das estatísticas L_1 e L_2 para a amostra observada. Como para a amostra observada temos $\bar{x} = 3,236$, considerando uma confiança de 95%, isto é, $\alpha = 0,05$, temos $z_{1-\alpha/2} = 1,96$ a estimativa para o intervalo bilateral será

$$ic_{\mu,95\%} = \left[3,236 - 1,96\frac{1}{\sqrt{10}}\ ,\ 3,236 + 1,96\frac{1}{\sqrt{10}}\right] = [2,616194\ ,\ 3,855806]$$

Já as estimativas para os intervalos unilaterais serão

$$ic_{\mu,95\%} = [3,236 - 1,64\frac{1}{\sqrt{10}} \ , \ \infty] = [2,717386 \ , \ \infty]$$
$$ic_{\mu,95\%} = [-\infty \ , \ 3,236 + 1,64\frac{1}{\sqrt{10}}] = [-\infty \ , \ 3,754614]$$

||

5.2 Quantidade Pivotal

Nessa seção veremos como encontrar intervalos de confiança para diferentes parâmetros de diferentes populações.

Definição 5.4. *Uma transformação $Q(\mathbf{X}, \theta)$ é uma quantidade pivotal para o parâmetro θ se:*

(i) A distribuição de Q não depende de θ e nem de qualquer outro parâmetro desconhecido.

(ii) Q é função não constante do parâmetro θ (isto é, depende de θ) e é função constante de qualquer outro parâmetro desconhecido da população (isto é, não depende de nenhum outro parâmetro desconhecido).

Veja que no Exemplo 5.1 temos uma quantidade pivotal para o parâmetro μ da população $X \sim N(\mu, \sigma_0^2)$, sendo σ_0^2 conhecido, $Z = \frac{\bar{X}-\mu}{\sqrt{\sigma_0^2/n}}$ (Verifique que Z é quantidade pivotal). A partir dessa quantidade pivotal construímos o intervalo de confiança encontrado no exemplo. Em geral esse será o procedimento para encontrar intervalos de confiança. Veja o passo-a-passo abaixo de como encontrar um $IC_{\theta,100(1-\alpha)\%}$ bilateral a partir de uma quantidade pivotal para θ.

Passo 1) Encontrar uma quantidade pivotal para θ. Vamos chamá-la de Q.

Passo 2) A partir da distribuição conhecida de Q, definir quantis q^1 e q^2 tais que

$$\mathrm{P}(q^1 \leq Q \leq q^2) = 1 - \alpha.$$

Passo 3) Realizar manipulações algébricas em $\mathrm{P}(q^1 \leq Q \leq q^2)$ até que ficar no padrão $\mathrm{P}(L_1(\mathbf{X}) \leq \theta \leq L_2(\mathbf{X}))$. Nesse caso temos,

$$\mathrm{P}(L_1(\mathbf{X}) \leq \theta \leq L_2(\mathbf{X})) = 1 - \alpha \ \Rightarrow \ IC_{\theta,100(1-\alpha)\%} = [L_1(\mathbf{X}), L_2(\mathbf{X})]$$

Se quisermos encontrar intervalos unilaterais podemos seguir o mesmo passo-a-passo, mas em vez de encontrar dois quantis, q^1 e q^2, vamos encontrar apenas um quantil, q, tal que

$$\mathrm{P}(q \leq Q) = 1 - \alpha \quad \text{ou} \quad \mathrm{P}(Q \leq q) = 1 - \alpha,$$

dependendo se queremos o intervalo unilateral à esquerda ou à direita.

Exemplo 5.2. *Considere a população $X \sim N(\mu, \sigma^2)$. Deduza um intervalo de confiança bilateral para μ a partir de uma quantidade pivotal. Em seguida, encontre a estimativa para o intervalo encontrado considerando a seguinte amostra da população X:*

$$3,35 \quad 2,94 \quad 3,11 \quad 1,96 \quad 2,78 \quad 3,56 \quad 5,91 \quad 3,61 \quad 1,05 \quad 4,09.$$

Solução:

Defina $Q = \dfrac{\bar{X} - \mu}{S/\sqrt{n}}$. A transformação Q depende do parâmetro desconhecido μ e de nenhum outro parâmetro desconhecido. Para que Q seja uma quantidade pivotal para μ basta a gente encontrar a sua distribuição e verificar que ela não depende de parâmetros desconhecidos. Pelo Teorema 2.3, temos que $Q = \dfrac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$, então de fato Q é uma quantidade pivotal para μ.

Seja $t_{p,n}$ o quantil da distribuição t tal que se $T \sim t_n$ então $\mathrm{P}(T < t_{p,n}) = p$. Seguindo essa notação e relembrando as propriedades da distribuição t, podemos escrever

$$\mathrm{P}\left(-t_{1-\alpha/2,n-1} \leq Q \leq t_{1-\alpha/2,n-1}\right) = 1-\alpha \ \Rightarrow \ \mathrm{P}\left(-t_{1-\alpha/2,n-1} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{1-\alpha/2,n-1}\right) = 1-\alpha.$$

Agora o objetivo é modificar a desigualdade de dentro da probabilidade até que o parâmetro μ fique isolado no meio da desigualdade.

$$\begin{aligned}
\mathrm{P}\left(-t_{1-\alpha/2,n-1} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{1-\alpha/2,n-1}\right) &= 1-\alpha \\
\mathrm{P}\left(-\frac{S}{\sqrt{n}}t_{1-\alpha/2,n-1} \leq \bar{X} - \mu \leq \frac{S}{\sqrt{n}}t_{1-\alpha/2,n-1}\right) &= 1-\alpha \\
\mathrm{P}\left(-\bar{X} - \frac{S}{\sqrt{n}}t_{1-\alpha/2,n-1} \leq -\mu \leq -\bar{X} + \frac{S}{\sqrt{n}}t_{1-\alpha/2,n-1}\right) &= 1-\alpha \\
\mathrm{P}\left(\bar{X} - \frac{S}{\sqrt{n}}t_{1-\alpha/2,n-1} \leq \mu \leq \bar{X} + \frac{S}{\sqrt{n}}t_{1-\alpha/2,n-1}\right) &= 1-\alpha
\end{aligned}$$

Logo,

$$IC_{\mu,100(1-\alpha)\%} = \left[\bar{X} - \frac{S}{\sqrt{n}}t_{1-\alpha/2,n-1}, \bar{X} + \frac{S}{\sqrt{n}}t_{1-\alpha/2,n-1}\right].$$

Para encontrar uma estimativa do intervalo de confiança encontrado precisamos dos valores de n, $\bar{x}$, s, α e $t_{1-\alpha/2,n-1}$.

Veja que $n = 10$, $\bar{x} = 3,236$ e $s = 1,284767$ (faça as contas, pode usar o programa R). Adotando o coeficiente de confiança de 95%, $\alpha = 0,05$ e $t_{1-\alpha/2,n-1} = t_{0.975,9} = 2,262157$ (consulta na Tabela A.4). Então,

$$\begin{aligned}
ic_{\mu,95\%} &= \left[\bar{x} - \frac{s}{\sqrt{10}}t_{0.975,9}\ ,\ \bar{x} + \frac{s}{\sqrt{10}}t_{0.975,9}\right] \\
&= \left[3,236 - \frac{1,284767}{\sqrt{10}}2,262157\ ,\ 3,236 + \frac{1,284767}{\sqrt{10}}2,262157\right] \\
&= [2,316933\ ,\ 4,155067].
\end{aligned}$$

||

Exemplo 5.3. *Considere a população $X \sim N(\mu, \sigma^2)$. Deduza um intervalo de confiança unilateral à direita para σ^2 a partir de uma quantidade pivotal. Em seguida, encontre a estimativa para o intervalo encontrado considerando a mesma amostra da população X do Exemplo 5.2:*

$$3,35 \quad 2,94 \quad 3,11 \quad 1,96 \quad 2,78 \quad 3,56 \quad 5,91 \quad 3,61 \quad 1,05 \quad 4,09.$$

Solução:

Novamente utilizando o resultado do Teorema 2.3, $Q = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ e então Q é uma quantidade pivotal para o parâmetro σ^2 (verifique!).

Seja $q_{p,n}$ o quantil da distribuição χ^2_n tal que se $W \sim \chi^2_n$ então $\mathrm{P}(W < q_{p,n}) = p$. Seguindo essa notação e lembrando que buscamos um intervalo unilateral, podemos escrever

$$\mathrm{P}\left(Q \leq q_{1-\alpha,n-1}\right) = 1-\alpha \;\Rightarrow\; \mathrm{P}\left(\frac{(n-1)S^2}{\sigma^2} \leq q_{1-\alpha,n-1}\right) = 1-\alpha.$$

Seguindo com as contas,

$$\begin{aligned}
\mathrm{P}\left(\frac{(n-1)S^2}{\sigma^2} \leq q_{1-\alpha,n-1}\right) &= 1-\alpha \\
\mathrm{P}\left(\frac{1}{\sigma^2} \leq \frac{q_{1-\alpha,n-1}}{(n-1)S^2}\right) &= 1-\alpha \\
\mathrm{P}\left(\frac{(n-1)S^2}{q_{1-\alpha,n-1}} \leq \sigma^2\right) &= 1-\alpha
\end{aligned}$$

Logo,

$$IC_{\sigma^2,100(1-\alpha)\%} = \left[\frac{(n-1)S^2}{q_{1-\alpha,n-1}}, \infty\right).$$

Observação: quando buscamos um intervalo unilateral podemos começar com $\mathrm{P}(q \leq Q) = 1-\alpha$ ou $\mathrm{P}(Q \leq q) = 1-\alpha$. Um deles resulta no intervalo unilateral à direita e o outro no à esquerda. A princípio não sabemos qual resulta em qual, depende das operações realizadas dentro da probabilidade. Então escolha um para começar e se você não chegar no intervalo unilateral desejado refaça as contas começando do outro jeito.

Falta apenas a estimativa intervalar. Veja que para a amostra apresentada $n = 10$. Adotando o coeficiente de confiança de 95%, temos $\alpha = 0,05$ e $q_{1-\alpha,n-1} = q_{0,95,9} = 16,9190$. Além disso, $\bar{x} = 3,236$ e $s^2 = 1,650627$. Assim, Logo,

$$ic_{\sigma^2,95\%} = \left[\frac{9 \times 1,650627}{16,9190}, \infty\right) = [0,878045\ ,\ \infty)$$

$\|$

Exemplo 5.4. *Considere a população $X \sim Exp(\lambda)$. Deduza um intervalo de confiança bilateral e unilateral à esquerda para λ a partir de uma quantidade pivotal. Em seguida, adotando uma confiança 99%, encontre as estimativas para os intervalos encontrados considerando a seguinte amostra da população X:*

$3,93 \quad 5,55 \quad 3,48 \quad 1,81 \quad 3,17 \quad 2,33 \quad 2,45 \quad 6,06 \quad 13,54 \quad 7,12 \quad 6,04 \quad 4,86 \quad 0,25 \quad 6,16 \quad 2,85.$

Solução:

O primeiro passo é encontrar uma quantidade pivotal para λ. Veja que $\sum_{i=1}^n X_i \sim Gama(n, \lambda)$, então $\lambda \sum_{i=1}^n X_i \sim Gama(n, 1)$ já é uma quantidade pivotal para λ. Mas vamos trabalhar com outra. A ideia é transformar a quantidade pivotal com distribuição gama em uma qui-quadrado para facilitar as contas, uma vez que a qui-quadrado tem valores tabelados. Então a quantidade pivotal escolhida para esse exercícios será

$$Q = 2\lambda \sum_{i=1}^n X_i \sim Gama\left(n, \frac{1}{2}\right) = \chi^2_{2n} \quad \text{ou} \quad Q = 2n\lambda\bar{X} \sim \chi^2_{2n}.$$

Para o intervalo bilateral serão escolhidos os quantis $q_{\frac{\alpha}{2},2n}$ e $q_{1-\frac{\alpha}{2},2n}$ da distribuição χ^2_{2n} tais que

$$\mathrm{P}(q_{\frac{\alpha}{2},2n} \leq Q \leq q_{1-\frac{\alpha}{2},2n}) = 1-\alpha.$$

Então,

$$\mathrm{P}\left(q_{\frac{\alpha}{2},2n} \leq 2n\lambda\bar{X} \leq q_{1-\frac{\alpha}{2},2n}\right) = 1-\alpha \quad \Rightarrow \quad \mathrm{P}\left(\frac{q_{\frac{\alpha}{2},2n}}{2n\bar{X}} \leq \lambda \leq \frac{q_{1-\frac{\alpha}{2},2n}}{2n\bar{X}}\right) = 1-\alpha.$$

Assim,

$$IC_{\lambda,100(1-\alpha)\%} = \left[\frac{q_{\frac{\alpha}{2},2n}}{2n\bar{X}}, \frac{q_{1-\frac{\alpha}{2},2n}}{2n\bar{X}}\right].$$

Para o intervalo unilateral à esquerda será usada a mesa quantidade pivotal, mas o quantil muda. Agora o quantil será $q_{1-\alpha,2n}$ da distribuição χ^2_{2n} tais que

$$\mathrm{P}\left(Q \leq q_{1-\alpha,2n}\right) = 1-\alpha \quad \Rightarrow \quad \mathrm{P}\left(2n\lambda\bar{X} \leq q_{1-\alpha,2n}\right) = 1-\alpha \quad \Rightarrow \quad \mathrm{P}\left(\lambda \leq \frac{q_{1-\alpha,2n}}{2n\bar{X}}\right) = 1-\alpha$$

e

$$IC_{\lambda,100(1-\alpha)\%} = \left(0, \frac{q_{1-\alpha,2n}}{2n\bar{X}}\right].$$

Um detalhe importante nesse exemplo é que como o espaço paramétrico para λ é $\mathbb{R}^+$, o limite inferior do intervalo unilateral à esquerda não será $-\infty$, mas sim 0.

Para terminar o exercício, as estimativas intervalares para a amostra observada. Para a amostra observada $n = 15$ e $\bar{x} = 4,64$. É pedido confiança de 99%, ou seja, $\alpha = 0,01$. Assim, $q_{\frac{\alpha}{2},2n} = q_{0,005,30} = 13,7867$, $q_{1-\frac{\alpha}{2},2n} = q_{0,995,30} = 53,6720$ e $q_{1-\alpha,2n} = q_{0,99,30} = 50,8922$. Então, os intervalos a seguir são estimativas intervalares para o parâmetro λ com 99% de confiança:

$$ic_{\lambda,99\%} = \left[\frac{13,7867}{30\times 4,64}\ ,\ \frac{53,6720}{30\times 4,64}\right] = [0,099\ ,\ 0,386]$$
$$ic_{\lambda,99\%} = \left(0\ ,\ \frac{50,8922}{30\times 4,64}\right] = (0\ ,\ 0,366]$$

$\|$

Exemplo 5.5. *Considere a população $X \sim U[0,\theta]$. Deduza um intervalo de confiança bilateral para θ a partir de uma quantidade pivotal. Em seguida, adotando uma confiança 90%, encontre as estimativas para os intervalos encontrados considerando a seguinte amostra da população X:*

$$1,149 \quad 0,083 \quad 0,085 \quad 1,160 \quad 1,294 \quad 0,874 \quad 1,297 \quad 1,016$$

Solução:

O primeiro passo sempre será buscar uma quantidade pivotal para o parâmetro da distribuição. Uma dica pode ser começar pelo estimador de máxima verossimilhança, ou alguma transformação dele que seja mais fácil de se obter a sua distribuição. A partir dele, ou de uma transformação sua, podemos realizar transformações para que essa estatística "vire" uma quantidade pivotal.

Já vimos que o estimador de máxima verossimilhança para θ é dado por $\hat{\theta} = X_{(n)}$ e que a sua distribuição amostral é (reveja Exemplos 3.12 e 2.2.)

$$f_{X_{(n)}}(x) = \frac{n}{\theta^n}x^{n-1},\ 0 \leq x \leq \theta.$$

Defina $Q = \dfrac{X_{(n)}}{\theta}$. Vejamos que Q é uma quantidade pivotal para o parâmetro θ. Primeiro, Q depende de θ. Então precisamos apenas encontrar a distribuição de Q e verificar que esta não depende de parâmetros desconhecidos. Para isso será usado o método da função de distribuição.

$$
\begin{aligned}
F_Q(x) &= \mathrm{P}(Q \leq x) = \mathrm{P}\left(\frac{X_{(n)}}{\theta} \leq x\right) = \mathrm{P}\left(X_{(n)} \leq x\theta\right) = F_{X_{(n)}}(x\theta) \\
f_Q(x) &= \frac{\mathrm{d}}{\mathrm{d}x} F_Q(x) = f_{X_{(n)}}(x\theta)\theta = \frac{n}{\theta^n}(\theta x)^{n-1}\theta,\ 0 \leq x\theta \leq \theta \\
&= \frac{n}{\theta^n} x^{n-1}\theta^n,\ 0 \leq x \leq 1 = nx^{n-1},\ 0 \leq x \leq 1
\end{aligned}
$$

Logo, a densidade de Q não depende de nenhuma parâmetro, $f_Q(x) = nx^{n-1},\ 0 \leq x \leq 1$ e por isso podemos considerar Q uma quantidade pivotal.

Nessa caso, por não ser uma distribuição tabelada, precisaremos encontrar os quantis manualmente. A Figura 5.3 um esboço do gráfico de f_Q. Vamos adotar a notação q_p para o quantil tal que $\mathrm{P}(Q < q_p) = p$. Repare que os quantis q^1 e q^2 não são únicos. Podemos escolher, por exemplo: $q^1 = q_{\frac{\alpha}{2}}$ e $q^2 = q_{1-\frac{\alpha}{2}}$; $q^1 = 0$ e $q^2 = q_{1-\alpha}$; ou $q^1 = q_\alpha$ e $q^2 = 1$. Para todos esses casos temos

$$
\mathrm{P}(q^1 \leq Q \leq q^2) = 1 - \alpha.
$$

E qual escolher? Uma opção é escolher a opção que resulta no intervalo de menor amplitude. E qual é essa opção, você sabe? Vamos usar $q^1 = q_\alpha$ e $q^2 = 1$. Mas qual o valor de q_α?

$$
\int_0^{q_\alpha} nx^{n-1}\,\mathrm{d}x = \alpha \quad \Rightarrow \quad x^n|_0^{q_\alpha} = \alpha \quad \Rightarrow \quad q_\alpha^n = \alpha \quad \Rightarrow \quad q_\alpha = \alpha^{1/n}.
$$

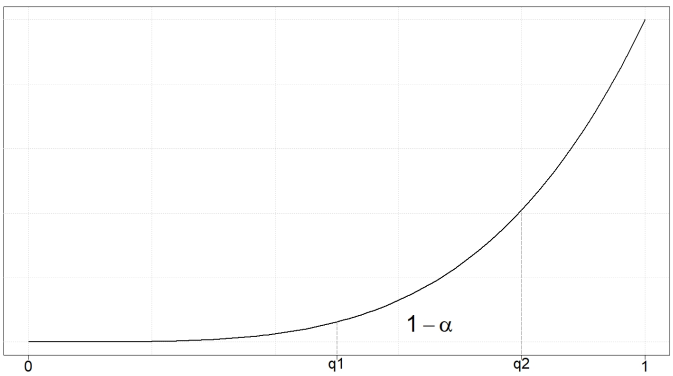


Figura 5.3: Esboço de $f_Q(x) = nx^{n-1},\ 0 \leq x \leq 1$.

Seguindo,

$$
\begin{aligned}
\mathrm{P}\left(\alpha^{1/n} \leq Q \leq 1\right) &= 1 - \alpha \\
\mathrm{P}\left(\alpha^{1/n} \leq \frac{X_{(n)}}{\theta} \leq 1\right) &= 1 - \alpha \\
\mathrm{P}\left(\frac{\alpha^{1/n}}{X_{(n)}} \leq \frac{1}{\theta} \leq \frac{1}{X_{(n)}}\right) &= 1 - \alpha \\
\mathrm{P}\left(X_{(n)} \leq \theta \leq \frac{X_{(n)}}{\alpha^{1/n}}\right) &= 1 - \alpha
\end{aligned}
$$

E então,

$$
IC_{\theta,100(1-\alpha)\%} = \left[X_{(n)}, \frac{X_{(n)}}{\alpha^{1/n}}\right]
$$

Para a amostra observada temos $n = 8$ e $x_{(8)} = 1,297$. Usando $\alpha = 0,1$ a estimativa intervalar para o parâmetro α com confiabilidade de 90% é:

$$ic_{\theta,90\%} = \left[1,297\ ,\ \frac{1,297}{(0,1)^{1/8}}\right] = [1,297\ ,\ 1,729577]$$

$\|$

Proposição 5.1. *Seja* $IC_{\theta,100(1-\alpha)\%} = [L_1(\mathbf{X}), L_2(\mathbf{X})]$ *e* $\eta = \tau(\theta)$.

(i) Se τ é uma função estritamente crescente, então $[\tau(L_1(\mathbf{X})), \tau(L_2(\mathbf{X}))]$ é um intervalo de confiança para η com coeficiente de confiança $1-\alpha$, isto é,

$$IC_{\eta,100(1-\alpha)\%} = [\tau(L_1(\mathbf{X})), \tau(L_2(\mathbf{X}))].$$

(ii) Se τ é uma função estritamente decrescente, então $[\tau(L_2(\mathbf{X})), \tau(L_1(\mathbf{X}))]$ é um intervalo de confiança para η com coeficiente de confiança $1-\alpha$, isto é,

$$IC_{\eta,100(1-\alpha)\%} = [\tau(L_2(\mathbf{X})), \tau(L_1(\mathbf{X}))].$$

Demonstração. Como $[L_1(\mathbf{X}), L_2(\mathbf{X})]$ é intervalo de confiança para θ é verdade que

$$\mathrm{P}\left(L_1(\mathbf{X}) \leq \theta \leq L_2(\mathbf{X})\right) = 1-\alpha.$$

Se τ é uma função estritamente crescente, então

$$\mathrm{P}\left(\tau(L_1(\mathbf{X})) \leq \tau(\theta) \leq \tau(L_2(\mathbf{X}))\right) = 1-\alpha,$$

logo $[\tau(L_1(\mathbf{X})), \tau(L_2(\mathbf{X}))]$ é intervalo de confiança para $\eta = \tau(\theta)$ com coeficiente de confiança $1-\alpha$.

De forma análoga, se τ é uma função estritamente decrescente, então

$$\mathrm{P}\left(\tau(L_1(\mathbf{X})) \geq \tau(\theta) \geq \tau(L_2(\mathbf{X}))\right) = 1-\alpha \quad \Rightarrow \quad \mathrm{P}\left(\tau(L_2(\mathbf{X})) \leq \tau(\theta) \leq \tau(L_1(\mathbf{X}))\right) = 1-\alpha$$

e $[\tau(L_2(\mathbf{X})), \tau(L_1(\mathbf{X}))]$ é intervalo de confiança para $\eta = \tau(\theta)$ com coeficiente de confiança $1-\alpha$. □

Exemplo 5.6. *Considere novamente a população $X \sim N(\mu, \sigma^2)$ e a amostra apresentada no Exemplo 5.2. Deduza um intervalo de confiança unilateral para σ a partir do intervalo já encontrado nesse exemplo. Em seguida, adotando a mesma confiança 95%, encontre a estimativa intervalar. Por fim, encontre a estimativa intervalar se for adotada uma confiança de 90%.*

Solução:

Veja que no Exemplo 5.3 foi encontrado o seguinte intervalo de confiança unilateral à direita para σ^2:

$$IC_{\sigma^2,100(1-\alpha)\%} = \left[\frac{(n-1)S^2}{q_{1-\alpha,n-1}}, \infty\right).$$

Queremos agora um intervalo de confiança unilateral para $\sigma = \sqrt{\sigma^2} = \tau(\sigma^2)$. Como τ é função estritamente crescente pode-se aplicar a Proposição 5.1, a qual garante que

$$IC_{\sigma,100(1-\alpha)\%} = \left[\sqrt{\frac{(n-1)S^2}{q_{1-\alpha,n-1}}}, \infty\right).$$

Novamente revisitando o Exemplo 5.3, para a amostra em questão, considerando uma confiança de 95%, foi encontrada a seguinte estimativa intervalar para σ^2:

$$ic_{\sigma^2,95\%} = [0,878045\ ,\ \infty).$$

Uma vez adotado o mesmo coeficiente de confiança do Exemplo 5.3, de 95%, de acordo com a Proposição 5.1,

$$ic_{\sigma,95\%} = \left[\sqrt{0,878045}\ ,\ \infty\right) = [0,9370406\ ,\ \infty)\,.$$

Se o coeficiente de confiança for outro é preciso encontrar os novos quantis para o novo valor de α. No caso do intervalo com 90% de confiança, $\alpha = 0,1$. Lembre-se que $s^2 = 1,650627$ e $n = 10$, então $q_{1-\alpha,n-1} = q_{0,9,9} = 14,6837$ e assim chega-se a seguinte estimativa:

$$ic_{\sigma,90\%} = \left[\sqrt{\frac{(n-1)s^2}{q_{0,9,9}}}, \infty\right) = \left[\sqrt{\frac{9 \times 1,650627}{14,6837}}, \infty\right) = [1,005838, \infty)\,.$$

Se forem comparados os dois intervalos unilaterais, 95% e 90% de confiança, é possível perceber que ao diminuir a confiança aumenta-se a precisão do intervalo, ou seja, diminui-se a sua amplitude. ||

Exemplo 5.7. *Considere novamente a população $X \sim Exp(\lambda)$ e a amostra apresentada no Exemplo 5.4. Deduza um intervalo de confiança bilateral e unilateral para $\mu = \mathrm{E}(X)$ a partir de dos intervalos já encontrados nesse exemplo. Em seguida, adotando a mesma confiança 99%, encontre as estimativas intervalares.*

Solução:

Veja primeiro que $\mu = \mathrm{E}(X) = \dfrac{1}{\lambda}$, então $\mu = \tau(\lambda)$ sendo τ uma função estritamente decrescente. Por esse motivo podemos aplicar o resultado da Proposição 5.1.

No Exemplo 5.4 foram encontrados os seguintes intervalos de confiança para λ

$$IC_{\lambda,100(1-\alpha)\%} = \left[\frac{q_{\frac{\alpha}{2},2n}}{2n\bar{X}}, \frac{q_{1-\frac{\alpha}{2},2n}}{2n\bar{X}}\right] \quad \text{e} \quad IC_{\lambda,100(1-\alpha)\%} = \left(0, \frac{q_{1-\alpha,2n}}{2n\bar{X}}\right]$$

Sendo $\tau(\lambda) = \frac{1}{\lambda}$, como resultado da Proposição 5.1 chegamos nos seguintes intervalos de confiança para μ

$$IC_{\mu,100(1-\alpha)\%} = \left[\frac{2n\bar{X}}{q_{1-\frac{\alpha}{2},2n}}, \frac{2n\bar{X}}{q_{\frac{\alpha}{2},2n}}\right] \quad \text{e} \quad IC_{\mu,100(1-\alpha)\%} = \left[\frac{2n\bar{X}}{q_{1-\alpha,2n}}, \infty\right).$$

Veja que o intervalo unilateral à esquerda para λ acabou se transformando em um intervalo unilateral à direita para μ, isso porque a função τ é decrescente.

Para encontrar as estimativas intervalares podemos usar os intervalos de confiança para μ recém deduzidos. Como queremos usar o mesmo coeficiente de confiança do Exemplo 5.4, podemos também simplesmente aplicar o resultado da Proposição 5.1 nas estimativas intervalares.

$$ic_{\lambda,99\%} = [0,099\ ,\ 0,386] \quad \Rightarrow \quad ic_{\mu,99\%} = \left[\frac{1}{0,386}\ ,\ \frac{1}{0,099}\right] = [2,59\ ,\ 10,1]$$

$$ic_{\lambda,99\%} = (0\ ,\ 0,366] \quad \Rightarrow \quad ic_{\mu,99\%} = \left[\frac{1}{0,366}\ ,\ \infty\right] = [2,73\ ,\ \infty]\,.$$

||

Exercícios da Seção 5.2

5.2.1 Seja $X_1, \ldots, X_n$ amostra aleatória de $X \sim Exp(\lambda)$.

(a) Encontre uma quantidade pivotal para λ e deduza um $IC_{\lambda,100(1-\alpha)\%}$ bilateral.

(b) Suponha que a amostra abaixo tenha sido retirada de uma população $X \sim Exp(\lambda)$. Considere o coeficiente de confiança de 95% e encontre as estimativas para os intervalos deduzidos nos itens acima.

$$0,07 \quad 3,11 \quad 1,61 \quad 0,57 \quad 3,38 \quad 5,42 \quad 0,83 \quad 0,02 \quad 1,76$$

5.2.2 No programa R podemos gerar amostras de variáveis aleatórias. Por exemplo, a função `rexp(n,rate)` gera uma amostra de tamanho `n` de uma variável aleatória exponencial com parâmetro λ igual a `rate`. Use o programa R para gerar 100 amostras distintas de tamanho $n = 9$ da população $X \sim Exp(\lambda = 1)$. Em seguida, para cada uma das 100 amostras geradas, encontre a estimativa do intervalo de confiança para λ, encontrado no Exercício 5.2.1, com confiabilidade de 95%. Contabilize quantos dos 100 intervalos contém o real valor de λ ($\lambda = 1$). Quantos intervalos você espera que contenha o valor 1?

5.2.3 ([Bolfarine e Sandoval, 2001] - Seção 5.6) Seja $X_1, \ldots, X_n$ uma amostra aleatória de $X \sim N(\theta, \theta)$. Sugira uma quantidade pivotal para e construa um intervalo de confiança para θ com coeficiente de confiança $1 - \alpha$.

5.2.4 ([Roussas, 2003] - Seção 10.1) Seja $X_1, \ldots, X_n$ uma amostra aleatória de X cuja função densidade é definida por:

$$f(x) = \frac{\lambda}{2} e^{-\lambda|x|} \ , \qquad x \in \mathbb{R} \ \text{ e } \ \lambda > 0.$$

(a) Mostre que $Y_i = |X_i| \sim Exp(\lambda)$.

(b) Mostre que $Q = 2\lambda n\bar{Y} = 2\lambda \sum_{i=1}^{n} Y_i \sim \chi^2_{2n}$.

(c) A partir de Q deduza um intervalo de confiança para λ.

(d) Suponha a amostra de X apresentada a seguir. Encontre estimativas para o intervalo de confiança deduzido acima para os diferentes níveis de confiança: 90%, 95% e 99%.

$$0,067 \quad 0,087 \quad -0,438 \quad -0,099 \quad -0,774 \quad -0,393 \quad -0,555 \quad 0,348 \quad -0,181 \quad 0,317$$

5.2.5 ([Roussas, 2003] - Seção 10.1) Seja $X_1, \ldots, X_n$ uma amostra aleatória de X cuja função densidade é definida por:

$$f_X(x) = \lambda e^{-\lambda(x-\theta)} \ , \qquad x > \theta \ , \ \lambda > 0 \ \text{ e } \ \theta \in \mathbb{R}.$$

Considerando λ um parâmetro conhecido, vamos buscar um intervalo de confiança para θ.

(a) Mostre que $F_X(x) = 1 - e^{-\lambda(x-\theta)}, \ x > \theta$.

(b) Veja que $\hat{\theta}_{MV} = X_{(1)}$. Mostre que $f_{X_{(1)}}(x) = n\lambda e^{-n\lambda(x-\theta)}, \ x > \theta$.

(c) Seja $Q = n\lambda(X_{(1)} - \theta)$. Mostre que $Q \sim Exp(1)$, logo Q é quantidade pivotal para θ.

(d) A partir de Q deduza um intervalo de confiança para θ com coeficiente de confiança de $1 - \alpha$. Busque o intervalo de menor amplitude.

(e) Considere os valores abaixo uma amostra de X para $\lambda = 1/2$. Encontre a estimativa para o intervalo de confiança deduzido acima com coeficiente de confiança de 95%.

$$7,79 \quad 5,94 \quad 7,07 \quad 5,49 \quad 10,13 \quad 8,12 \quad 5,19 \quad 5,62$$

5.2.6 ([Casella e Berger, 2002] - Capítulo 9) Seja $X_1, X_2, \ldots, X_n$ uma amostra aleatória de X cuja função de distribuição F está definida a seguir. Considere α_0 conhecido e $\beta > 0$.

$$F_X(x) = \begin{cases} 0 & , x < 0 \\ (x/\beta)^{\alpha_0} & , 0 \leq x \leq \beta \\ 1 & , x > \beta \end{cases}$$

(a) Encontre $\hat{\beta}_{MV}$ e, a partir dele, encontre uma quantidade pivotal para β.

(b) Construa um intervalo de confiança para β com coeficiente de confiança de 95%.

(c) O tamanho (em milimetros) de ovos de uma certa espécie de pássaros pode ser modelada por essa distribuição. Supondo a amostra abaixo e que $\alpha_0 = 12$, encontre a estimativa para o intervalo de confiança obtido no item acima.

$$22,0 \quad 23,9 \quad 20,9 \quad 23,8 \quad 25,0 \quad 24,0 \quad 21,7 \quad 23,8 \quad 22,8 \quad 23,1 \quad 23,1 \quad 23,5 \quad 23,0 \quad 23,0$$

5.2.7 Seja $X_1, \ldots, X_n$ amostra aleatória de X com função densidade $f_X(x) = \theta x^{\theta-1}$, $0 < x < 1$ e $\theta > 0$.

(a) Encontre uma quantidade pivotal para o parâmetro θ. Dica: reveja o resultado do Exercício 2.1.11.

(b) Encontre um intervalo de confiança bilateral para θ.

(c) Encontre um intervalo de confiança bilateral para $\mu = \mathrm{E}(X)$.

(d) Suponha que os valores abaixo formem uma amostra aleatória de X.

$$0,89 \quad 0,55 \quad 0,15 \quad 0,18 \quad 0,03 \quad 0,09 \quad 0,22 \quad 0,06 \quad 0,78 \quad 0,69 \quad 0,55 \quad 0,18$$

Encontre as estimativas para os intervalos de confiança para θ e $\mu = E(X)$ definidos nos itens acima. Considere o coeficiente de confiança no intervalo de 95%.

5.2.8 Seja X uma variável aleatória contínua cuja função densidade depende do parâmetro $\theta < 5$ e é definida por

$$f_X(x) = \frac{1}{5-\theta}\ , \quad \theta < x < 5.$$

Veja que $X \sim U(\theta, 5)$ e que esta distribuição é a mesma dos Exercícios 2.2 e 3.10.

(a) Mostre que $Q = \dfrac{5 - X_{(1)}}{5-\theta}$ é uma quantidade pivotal exata para θ.

(b) Encontre um intervalo de confiança exato para θ com confiabilidade de 95%.

(c) Considere os valores abaixo uma amostra da população X. Encontre a estimativa para o intervalo de confiança deduzido no item acima para o parâmetro θ.

$$2,06 \quad 4,35 \quad 3,04 \quad 4,96 \quad 2,37 \quad 2,43 \quad 4,75 \quad 2,75 \quad 3,28 \quad 3,79 \quad 2,64 \quad 3,89$$

5.3 Intervalos de Confiança Aproximados e Quantidades Pivotais Assintóticas

Algumas vezes não conseguimos uma quantidade pivotal para um certo parâmetro, e como consequência não encontramos um intervalo de confiança para ele. Mas no caso de grandes amostras podemos recorrer aos intervalos de confiança aproximados criados a partir de quantidades pivotais assintóticas.

Definição 5.5. *Duas estatísticas* $L_1(\mathbf{X})$ *e* $L_2(\mathbf{X})$ *definem um intervalo de confiança bilateral aproximado (ou assintótico) para o parâmetro* θ *com coeficiente de confiança* $1-\alpha$, *ou* $100(1-\alpha)\%$, *se*

$$\mathrm{P}(L_1(\mathbf{X}) \leq \theta \leq L_2(\mathbf{X})) \xrightarrow[n\to\infty]{} 1-\alpha.$$

Nesse caso escrevemos $IC^a_{\theta,100(1-\alpha)\%} = [L_1(\mathbf{X}), L_2(\mathbf{X})]$.

De forma análoga definimos os intervalos de confiança unilaterais aproximados.

Definição 5.6. *A estatística $L_1(\mathbf{X})$ define um intervalo de confiança unilateral à direita aproximado (ou assintótico) para o parâmetro θ com coeficiente de confiança $1-\alpha$, ou $100(1-\alpha)\%$, se*

$$\mathrm{P}(L_1(\mathbf{X}) \leq \theta) \xrightarrow[n\to\infty]{} 1-\alpha.$$

Nesse caso escrevemos $IC^a_{\theta,100(1-\alpha)\%} = [L_1(\mathbf{X}), \infty)$.

Definição 5.7. *A estatística $L_2(\mathbf{X})$ define um intervalo de confiança unilateral à esquerda aproximado (ou assintótico) para o parâmetro θ com coeficiente de confiança $1-\alpha$, ou $100(1-\alpha)\%$, se*

$$\mathrm{P}(\theta \leq L_2(\mathbf{X})) \xrightarrow[n\to\infty]{} 1-\alpha.$$

Nesse caso escrevemos $IC^a_{\theta,100(1-\alpha)\%} = (-\infty, L_2(\mathbf{X})]$.

Vejamos como encontrar esses intervalos de confiança aproximados. Para isso é preciso primeiro entender o conceito de quantidade pivotal assintótica.

Definição 5.8. *Uma transforma $Q(\mathbf{X},\theta)$ é uma quantidade pivotal assintótica para o parâmetro θ se:*

(i) A distribuição assintótica de Q não depende de θ e nem de qualquer outro parâmetro desconhecido.

(ii) Q é função não constante do parâmetro θ (isto é, depende de θ) e função constante de qualquer outro parâmetro desconhecido da população (isto é, não depende de nenhum outro parâmetro desconhecido).

O mesmo passo-a-passo apresentado na página 108 pode ser usado para encontrar intervalos de confiança aproximados, a única diferença é que nesse caso serão usadas quantidades pivotais assintóticas. Mas como encontrar quantidades pivotais assintóticas? Os dois teoremas a seguir ajudam nesse objetivo, eles apresentam uma alternativa a partir do estimador de máxima verossimilhança. Mas atenção, é preciso assumir satisfeitas as condições de regularidades apresentadas no Teorema 4.1.

Teorema 5.1. *Seja $\eta = \tau(\theta)$ e $\hat{\eta}$ o estimador de máxima verossimilhança para η. Se τ é uma função contínua e as condições de regularidades apresentadas no Teorema 4.1 são satisfeitas, o estimador de máxima verossimilhança é assintoticamente eficiente. Isto é,*

$$\sqrt{n}\,(\hat{\eta}-\eta) \xrightarrow{d} N\left(0, \frac{\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\tau(\theta)\right)^2}{I_F(\theta)}\right),$$

onde $I_F(\theta)$ é a informação de Fisher de θ, apresentada no Teorema 4.1, e θ o parâmetro da distribuição.

Esse resultado está apresentado em diversos livros de Inferência Estatística, como por exemplo [Bolfarine e Sandoval, 2001], [Roussas, 2003] ou [Casella e Berger, 2002]. Sua demonstração será omitida, mas pode ser encontrada em [Casella e Berger, 2002] (Teorema.10.1.12).

O Teorema 5.1 mostra que ou o estimador de máxima verossimilhança é eficiente ou ele está muito próximo de ser para grandes amostras. Isto é, se ele não for o estimador eficiente para o parâmetro η, para grandes amostras ele é aproximadamente não-tendencioso e sua variância é aproximadamente a variância mínima dada pelo limite de Cramér-Rao. Portanto, em grandes amostras, o estimador de máxima verossimilhança é aproximadamente eficiente.

O resultado do Teorema 5.1 junto com alguns resultados sobre convergência de variáveis aleatórias permite demonstrar a seguinte proposição.

Proposição 5.2. *Seja $\hat{\theta}$ o estimador de máxima verossimilhança para θ, $\eta = \tau(\theta)$ e $\hat{\eta}$ o estimador de máxima verossimilhança para η. Suponha satisfeitas as condições de regularidades apresentadas no Teorema 4.1 e que τ seja uma função contínua. Seja $\mathrm{V}^{\eta}_{min}(\theta)$ é o limite inferior para a variância de estimadores não tendenciosos para η fornecido pela Desigualdade de Cramér-Rao (Teorema 4.1). Então,*

$$Q = \frac{\hat{\eta} - \eta}{\sqrt{\mathrm{V}^{\eta}_{min}(\hat{\theta})}} \xrightarrow{d} N(0,1).$$

A demonstração dessa proposição será omitida e pode ser encontrada em [Casella e Berger, 2002] ou [Bolfarine e Sandoval, 2001], por exemplo.

Veja que Q definida na Proposição 5.2 acima é uma quantidade pivotal assintótica para η.

Veremos agora alguns exemplos de intervalos de confiança aproximados. Eles serão bem úteis para algumas distribuições, como alguns casos discretos. As vezes é difícil encontrar uma quantidade pivotal, ou mesmo que se encontre, fica difícil o cálculo dos quantis.

Exemplo 5.8. *Seja $X \sim Bernoulli(p)$. Deduza um intervalo de confiança aproximado bilateral para p. Em seguida encontre as estimativas intervalares aproximadas com 90% de confiança supondo uma amostra de tamanho 100 com onde ocorreram 24 sucessos e 76 fracassos.*

Solução: Esse é um exemplo em que é útil utilizar um intervalo de confiança aproximado. Procurar uma quantidade pivotal para o parâmetro p não é uma tarefa fácil, mas a aplicação da Proposição 5.2 nos permite encontrar uma quantidade pivotal assintótica para p.

Veja que $\hat{p} = \bar{X}$ é o estimador de máxima verossimilhança para p e que a distribuição de Bernoulli satisfaz as condições de regularidade. Vamos encontrar V^{p}_{min}.

$$\begin{aligned}
I_F(p) = -\,\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\ln\left(f_X(X|p)\right)\right] &= -\,\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\ln\left(p^X(1-p)^{(1-X)}\right)\right] \\
&= -\,\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\left(X\ln(p) + (1-X)\ln(1-p)\right)\right] \\
&= -\,\mathrm{E}\left[-\frac{X}{p^2} - \frac{1-X}{(1-p)^2}\right] \\
&= \frac{p}{p^2} + \frac{1-p}{(1-p)^2} \\
&= \frac{1}{p} + \frac{1}{1-p} \\
&= \frac{1}{p(1-p)}
\end{aligned}$$

Assim,

$$\mathrm{V}^{p}_{min}(p) = \frac{1}{nI_F(\theta)} = \frac{p(1-p)}{n}.$$

Podemos então aplicar o teorema e concluir que

$$Q = \frac{\bar{X} - p}{\sqrt{\mathrm{V}^{p}_{min}(\bar{X})}} = \frac{\bar{X} - p}{\sqrt{\bar{X}(1-\bar{X})/n}} \xrightarrow{d} N(0,1)$$

é uma quantidade pivotal assintótica. Usando os quantis da distribuição normal podemos escrever

$$
\begin{aligned}
\mathrm{P}\left(-z_{1-\frac{\alpha}{2}} \leq Q \leq z_{1-\frac{\alpha}{2}}\right) &\xrightarrow[n\to\infty]{} 1-\alpha \\
\mathrm{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X}-p}{\sqrt{\bar{X}(1-\bar{X})/n}} \leq z_{1-\frac{\alpha}{2}}\right) &\xrightarrow[n\to\infty]{} 1-\alpha \\
\mathrm{P}\left(-z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq \bar{X}-p \leq z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right) &\xrightarrow[n\to\infty]{} 1-\alpha \\
\mathrm{P}\left(\bar{X}-z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq p \leq \bar{X}+z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right) &\xrightarrow[n\to\infty]{} 1-\alpha
\end{aligned}
$$

$$
IC^a_{p,100(1-\alpha)\%} = \left[\bar{X}-z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\ ,\ \bar{X}+z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right]
$$

Para a amostra observada temos $n = 100$, $\bar{x} = \dfrac{24}{100} = 0,24$. Adotando um coeficiente de confiança de 90%, $\alpha = 0,1$ e $z_{1-\frac{\alpha}{2}} = 1,64$. Então a estimativa intervalar fica

$$
ic^a_{p,100(1-\alpha)\%} = \left[0,24 - 1,64\sqrt{\frac{0,24\times 0,76}{100}}\ ,\ 0,24 + 1,64\sqrt{\frac{0,24\times 0,76}{100}}\right] = [0,17\ ,\ 0,31]\,.
$$

$\|$

Exemplo 5.9. *Uma pesquisa de intensão de votos em um certo candidato será feita. Cada eleitor entrevistado vai responder se pretende ou não votar no candidato em questão. Quantos eleitores devem ser entrevistados para garantir uma confiança de 95% e uma margem de erro de no máximo 2%?*

Solução: Esse é um exemplo de aplicação direta do resultado do Exemplo 5.8 para o cálculo de tamanho de amostra. Veja que o problema vai levantar uma amostra de $X \sim Bernoulli(p)$, $X_i = 1$ se o eleitor i pretende votar no candidato e $X_i = 0$ se ele não pretende. Com essa amostra pretende-se construir um intervalo de confiança (aproximado) para p (proporção dos eleitores que pretendem votar no candidato) com confiabilidade de 95% e margem de erro, que é o raio do intervalo, de 2%, ou 0,02. A pergunta é: qual o tamanho da amostra que garante essa margem de erro com essa confiança?

No Exercício 5.8 encontramos o intervalo de confiança (aproximado) para o parâmetro p:

$$
IC^a_{p,100(1-\alpha)\%} = \left[\bar{X}-z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\ ,\ \bar{X}+z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right].
$$

Esse é o intervalo que o pesquisador vai construir depois da amostra recolhida, um intervalo centrado em $\bar{X}$ e com margem de erro (raio) igual a $z_{1-\frac{\alpha}{2}}\sqrt{\dfrac{\bar{X}(1-\bar{X})}{n}}$, veja Figura 5.4.

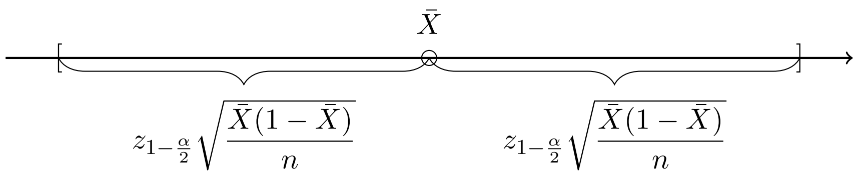


Figura 5.4: Intervalo de confiança aproximado bilateral para p de $X \sim Bernoulli(p)$.

Na prática queremos, para um dado valor de α, encontrar o tamanho da amostra n para o qual a margem de erro é aquela pré-estabelecida. Para isso precisaríamos saber o valor de $\bar{X}(1-\bar{X})$, que só será conhecido depois da amostra recolhida.

Mas perceba que $\bar{X}(1-\bar{X}) \leq \frac{1}{4}$, uma vez que o valor máximo da função $f(x) = x(1-x)$ é $\frac{1}{4}$, e esse valor ocorre quando $x = \frac{1}{2}$, veja Figura 5.5. Então $\bar{X} = \frac{1}{2}$ seria o valor que gera a maior margem de erro, qualquer outro valor observado para $\bar{X}$ resulta em uma margem de erro menor.

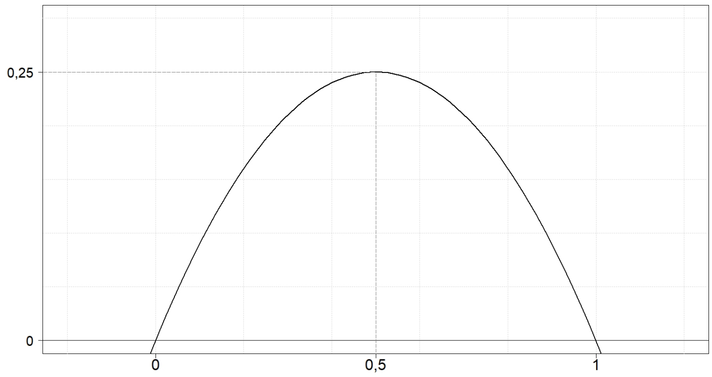


Figura 5.5: Gráfico da função $f(x) = x(1-x)$.

A ideia é encontrar o valor de n assumindo o pior caso possível, $\bar{x} = \frac{1}{2}$. Com isso será determinado um tamanho de amostra que garante a margem de erro quando $\bar{x} = \frac{1}{2}$, e consequentemente garante a margem de erro para qualquer outro valor de $\bar{x}$.

Vamos às contas. Considerando e a margem de erro e $1-\alpha$ a confiança do intervalo, queremos o valor de n tal que

$$z_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{4n}} = e \quad \Rightarrow \quad \frac{z_{1-\frac{\alpha}{2}}}{2 \times e} = \sqrt{n} \quad \Rightarrow \quad n = \left(\frac{z_{1-\frac{\alpha}{2}}}{2 \times e}\right)^2$$

Para o problema em questão: $e = 0,02$, $1-\alpha = 0,95$, $\alpha = 0,05$ e $z_{1-\frac{\alpha}{2}} = 1,96$.

$$\left(\frac{1,96}{2 \times 0,02}\right)^2 = 2.376,56 \quad \Rightarrow \quad n = 2.377.$$

Então, se 2.377 eleitores responderem a pesquisa garantimos que o intervalo de confiança gerado para p, com confiabilidade de 95%, não terá raio maior que $0,02$, ou seja, garantimos que a pesquisa retornará uma estimativa para p com confiança de 95% e margem de erro menor que 2%. $\|$

Exemplo 5.10. *Seja novamente $X \sim Bernoulli(p)$. Deduza agora um intervalo de confiança aproximado bilateral para $\eta = \text{Var}(X) = p(1-p)$. Encontre as estimativas intervalares aproximadas com 90% de confiança supondo uma amostra de X de tamanho 100 com onde ocorreram 24 sucessos e 76 fracassos.*

Solução: Ainda considerando a população $X \sim Bernoulli(p)$ vamos buscar agora um intervalo bilateral aproximado para $\eta = \text{Var}(X) = p(1-p)$. O que será feito é buscar uma quantidade pivotal assintótica para η a partir do resultado da Proposição 5.2. Para isso precisamos encontrar o estimador de máxima verossimilhança para η e o limite inferior para a variância de estimadores não tendenciosos para η dado pela desigualdade de Cramér-Rao.

Veja que como $\hat{p} = \bar{X}$ é o estimador de máxima verossimilhança para p, por invariância, podemos concluir que $\hat{\eta} = \bar{X}(1-\bar{X})$ é o estimador por máxima verossimilhança para η.

No Exemplo 5.8 já foi encontrada a informação de Fisher para o parâmetro p: $I_F(p) = \dfrac{1}{p(1-p)}$. Então a variância mínima dada pleo limite de Cramér-Rao é:

$$\mathrm{V}^{\eta}_{min}(p) = \frac{\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\left(p(1-p)\right)\right)^2}{nI_F(\theta)} = \frac{\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\left(p-p^2\right)\right)^2}{n(1/(p(1-p)))} = \frac{(1-2p)^2\,p(1-p)}{n}.$$

Então,

$$Q = \frac{\bar{X}(1-\bar{X})-\eta}{\sqrt{\dfrac{\left(1-2\bar{X}\right)^2\bar{X}(1-\bar{X})}{n}}} \xrightarrow{d} N(0,1)$$

é uma quantidade pivotal assintótica para η, o que significa que

$$\mathrm{P}\left(-z_{1-\frac{\alpha}{2}} \leq Q \leq z_{1-\frac{\alpha}{2}}\right) \xrightarrow[n\to\infty]{} 1-\alpha$$

$$\mathrm{P}\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X}(1-\bar{X})-\eta}{\sqrt{\frac{\left(1-2\bar{X}\right)^2\bar{X}(1-\bar{X})}{n}}} \leq z_{1-\frac{\alpha}{2}}\right) \xrightarrow[n\to\infty]{} 1-\alpha$$

E então chegamos ao seguinte intervalo de confiança aproximado para o parâmetro $\eta = \mathrm{Var}(X)$:

$$\left[\bar{X}(1-\bar{X}) - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\left(1-2\bar{X}\right)^2\bar{X}(1-\bar{X})}{n}}\ ,\ \bar{X}(1-\bar{X}) + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\left(1-2\bar{X}\right)^2\bar{X}(1-\bar{X})}{n}}\right].$$

Para a amostra observada temos $\bar{x} = 0,24$, $n = 100$, $z_{1-\frac{\alpha}{2}} = 1,64$. Então a estimativa para esse intervalo é:

$$0,24 \times 0,76 \pm 1,64\sqrt{\frac{(1-2\times 0,24)^2 \times 0,24 \times 0,76}{100}} = [0,146\ ,\ 0,219]$$

$\|$

Exemplo 5.11. *$X \sim Exp(\lambda)$. Deduza intervalos de confiança aproximados unilateral à esquerda para λ e unilateral à direita para $\mu = \mathrm{E}(X)$. Encontre as estimativas intervalares com 99% de confiança considerando a mesma amostra do Exercício 5.4. Compare com os resultados dos Exercícios 5.4 e 5.7.*

Solução: Já encontramos intervalos exatos equivalentes aos pedidos nesse exercício, reveja os Exemplos 5.4 e 5.7. Vamos agora encontrar intervalos aproximados. A ideia é comparar os resultados.

Para encontrar intervalos aproximados para λ precisamos de uma quantidade pivotal assintótica para esse parâmetro e para isso vamos usar o resultado apresentado na Proposição 5.2, já que as condições de regularidades são satisfeitas. Temos $\hat{\lambda} = \dfrac{1}{\bar{X}}$ o estimador de máxima verossimilhança. Precisamos apenas encontrar a variância mínima.

Aproveitando os resultados do Exemplo 4.4,

$$I_F(\lambda) = \frac{1}{\lambda^2} \quad \text{e} \quad V^{\lambda}_{min} = \frac{\lambda^2}{n}.$$

Então , segundo a Proposição 5.2,

$$Q = \frac{1/\bar{X} - \lambda}{\sqrt{\frac{1}{n\bar{X}^2}}} \xrightarrow{d} N(0,1)$$

é quantidade pivotal assintótica para λ e a partir dela podemos encontrar um intervalo de confiança aproximado.

$$\mathrm{P}\left(-z_{1-\alpha} \leq Q\right) \xrightarrow[n\to\infty]{} 1-\alpha$$

$$\mathrm{P}\left(-z_{1-\alpha} \leq \frac{1/\bar{X} - \lambda}{\sqrt{\frac{1}{n\bar{X}^2}}}\right) \xrightarrow[n\to\infty]{} 1-\alpha$$

$$\mathrm{P}\left(\lambda \leq \frac{1}{\bar{X}} + z_{1-\alpha}\sqrt{\frac{1}{n\bar{X}^2}}\right) \xrightarrow[n\to\infty]{} 1-\alpha$$

E então chegamos ao seguinte intervalo de confiança aproximado unilateral à direita para o parâmetro λ:

$$IC^a_{\lambda,100(1-\alpha)\%} = \left[0\ ,\ \frac{1}{\bar{X}} + z_{1-\alpha}\sqrt{\frac{1}{n\bar{X}^2}}\right).$$

Baseado na amostra do Exemplo 5.4, $n = 15$ e $\bar{x} = 4,64$. Usando confiança de 99%, $z_{1-\alpha} = 2,33$ e

$$ic^a_{\lambda,100(1-\alpha)\%} = \left[0\ ,\ \frac{1}{4,64} + 2,33\sqrt{\frac{1}{(15 \times 4,64^2)}}\right) = [0\ ,\ 0,345)\,.$$

No Exemplo 5.4 foi encontrada a estimativa para o intervalo exato unilateral à esquerda dado por $ic_{\lambda,99\%} = (0\ ,\ 0,366]$.

Para o parâmetro $\mu = \mathrm{E}(X)$, $\hat{\mu} = \bar{X}$ é o estimador de máxima verossimilhança e buscando novamente os resultados do Exemplo 4.4, já encontramos $V^{\mu}_{min} = \dfrac{1}{n\lambda^2}$. Então,

$$Q = \frac{\bar{X} - \mu}{\sqrt{\frac{\bar{X}^2}{n}}} \xrightarrow{d} N(0,1)$$

é uma quantidade pivotal assintótica para o parâmetro μ e a partir dela podemos construir um intervalo de confiança aproximado unilateral à direita:

$$\mathrm{P}\left(Q \leq z_{1-\alpha}\right) \xrightarrow[n\to\infty]{} 1-\alpha$$

$$\mathrm{P}\left(\frac{\bar{X} - \mu}{\sqrt{\frac{\bar{X}^2}{n}}} \leq z_{1-\alpha}\right) \xrightarrow[n\to\infty]{} 1-\alpha$$

$$\mathrm{P}\left(\bar{X} - z_{1-\alpha}\sqrt{\frac{\bar{X}^2}{n}} \leq \mu\right) \xrightarrow[n\to\infty]{} 1-\alpha$$

E então chegamos ao seguinte intervalo de confiança aproximado unilateral à direita para o parâmetro μ:

$$IC^a_{\mu,100(1-\alpha)\%} = \left[\bar{X} - z_{1-\alpha}\sqrt{\frac{\bar{X}^2}{n}}\ ,\ \infty\right).$$

Usando novamente $n = 15$, $\bar{x} = 4,64$, $\alpha = 0,01$ e $z_{1-\alpha} = 2,33$, a estimativa para esse intervalo é

$$ic^a_{\mu,99\%} = \left[4,64 - 2,33\sqrt{\frac{4,64^2}{15}}\ ,\ \infty\right) = [1,84856\ ,\ \infty)\,.$$

No Exemplo 5.7 foi encontrada a estimativa para o intervalo exato unilateral à esquerda dado por $ic_{\mu,99\%} = [2,73\ ,\ \infty)$.

$\|$

Exercícios da Seção 5.3

5.3.1 ([Larson, 1982] - Capítulo 7) O numero de casas vendidas por semana por uma firma de corretagem pode ser considerada uma variável aleatória de Poisson com parâmetro λ. Foram observados, em 15 semanas consecutivas, os totais de venda:

$$3 \quad 3 \quad 4 \quad 6 \quad 2 \quad 4 \quad 4 \quad 3 \quad 1 \quad 2 \quad 0 \quad 5 \quad 7 \quad 1 \quad 4.$$

Suponha que 15 seja "grande o suficiente" e que pode ser assumida independência entre as semanas. Considerando o coeficiente de confiança de 95%, encontre:

(a) um intervalo de confiança bilateral aproximado para o número médio de casas vendidas por semana;

(b) um limite superior para o número médio de casas vendidas por semana a partir de um intervalo de confiança unilateral aproximado.

5.3.2 O prefeito de uma cidade contratou um instituto de pesquisa para estudar sobre a aprovação de um certo projeto na população. Cada cidadão que participará da pesquisa vai responder se é ou não a favor do projeto do prefeito.

(a) Quantos cidadãos o instituto deve entrevistar para garantir uma confiança de 99% e uma margem de erro de no máximo 3%?

(b) Suponha que o instituto entrevistou exatamente o número de cidadãos encontrados no item (a) e que obteve 42% das respostas favoráveis ao projeto. Qual a conclusão da pesquisa? Verifique que a margem de erro foi de fato menor que 3%.

5.4 Intervalos de Confiança para 2 Amostras Independentes

Em algumas situações práticas definimos um parâmetro de interesse baseado em duas amostras distintas. Por exemplo, suponha que em uma fábrica duas máquinas realizam a mesma função de embalar um certo produto. Pode ser interessante estimar a diferença entre os pesos dos produtos embalados pela máquina 1 e pela máquina 2, a fim de comparar a produção das duas máquinas. Sendo assim, o parâmetro de interesse é $\theta = \mu_1 - \mu_2$, onde μ_1 e μ_2 são os pesos médios dos produtos embalados pelas máquinas 1 e 2, respectivamente.

Para esses casos também podemos encontrar intervalos de confiança para o parâmetro θ definido a partir de parâmetros de duas amostras. E o procedimento para a obtenção de tais intervalos será também a partir de quantidades pivotais.

Esse primeiro exemplo fornece um intervalo bilateral para a diferença entre as médias de duas populações normais de mesma variância. Mas poderia ser intervalos unilaterais que pouca coisa mudaria. O desafio da questão é encontrar uma quantidade pivotal para o parâmetro em questão.

Exemplo 5.12. *Sejam $X \sim N(\mu_x, \sigma^2)$ e $Y \sim N(\mu_y, \sigma^2)$ variáveis aleatórias independentes, com distribuição normal de mesma variância e médias distintas. Suponha uma amostra de tamanho n de X e outra de tamanho m de Y. Defina $\theta = \mu_x - \mu_y$. Encontre um intervalo de confiança bilateral para θ.*

Solução: Precisamos de uma quantidade pivotal para θ. Vamos mostrar que Q definida a seguir é uma quantidade pivotal para θ.

$$Q = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{S_p / \sqrt{\frac{nm}{n+m}}}$$

com $S_p^2 = \dfrac{1}{n+m-2}\left(\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2\right)$.

De imediato podemos perceber que Q é função de $\theta = \mu_x - \mu_y$ e não é mais função de qualquer outro parâmetro desconhecido. Então falta verificar que a sua distribuição não depende de parâmetros desconhecidos.

Para isso, primeiro veja que

$$\bar{X} \sim N\left(\mu_x, \frac{\sigma^2}{n}\right) \quad \text{e} \quad \bar{Y} \sim N\left(\mu_y, \frac{\sigma^2}{m}\right).$$

Além disso, $\bar{X}$ e $\bar{Y}$ são variáveis aleatórias independentes. Então, $\bar{X} - \bar{Y} \sim N\left(\mu_x - \mu_y, \frac{\sigma^2}{n} + \frac{\sigma^2}{m}\right)$. Por isso,

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} = \frac{(\bar{X} - \bar{Y}) - (\mu_x - \mu_y)}{\sigma\sqrt{\frac{m+n}{n \times m}}} \sim N(0,1).$$

Veja também que

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)S_x^2}{\sigma^2} \sim \chi^2_{n-1} \quad \text{e} \quad \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{\sigma^2} = \frac{(m-1)S_y^2}{\sigma^2} \sim \chi^2_{m-1}.$$

Como essas duas últimas variáveis aleatórias são independentes,

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^n (Y_i - \bar{Y})^2}{\sigma^2} \sim \chi^2_{n+m-2} \quad \Rightarrow \quad \frac{(n+m-2)S_p^2}{\sigma^2} \sim \chi^2_{n+m-2}.$$

Como

$$Q = \frac{Z}{\sqrt{S_p^2/\sigma^2}} \quad \Rightarrow \quad Q \sim t_{n+m-2}.$$

Assim chegamos a conclusão que Q é uma quantidade pivotal para $\theta = \mu_x - \mu_y$ e pode ser usada para encontrar intervalos de confiança para θ. Sendo q_α quantil da distribuição

$$P\left(t_{\frac{\alpha}{2},n+m-2} \leq \frac{(\bar{X}-\bar{Y})-(\mu_x-\mu_y)}{S_p/\sqrt{\frac{nm}{n+m}}} \leq t_{1-\frac{\alpha}{2},n+m-2}\right) = 1-\alpha$$

$$P\left(t_{\frac{\alpha}{2},n+m-2}\frac{S_p}{\sqrt{\frac{nm}{n+m}}} \leq (\bar{X}-\bar{Y})-(\mu_x-\mu_y) \leq t_{1-\frac{\alpha}{2},n+m-2}\frac{S_p}{\sqrt{\frac{nm}{n+m}}}\right) = 1-\alpha$$

$$P\left(-(\bar{X}-\bar{Y})+t_{\frac{\alpha}{2},n+m-2}\frac{S_p}{\sqrt{\frac{nm}{n+m}}} \leq -(\mu_x-\mu_y) \leq -(\bar{X}-\bar{Y})+t_{1-\frac{\alpha}{2},n+m-2}\frac{S_p}{\sqrt{\frac{nm}{n+m}}}\right) = 1-\alpha$$

$$P\left((\bar{X}-\bar{Y})-t_{\frac{\alpha}{2},n+m-2}\frac{S_p}{\sqrt{\frac{nm}{n+m}}} \geq \mu_x-\mu_y \geq (\bar{X}-\bar{Y})-t_{1-\frac{\alpha}{2},n+m-2}\frac{S_p}{\sqrt{\frac{nm}{n+m}}}\right) = 1-\alpha$$

$$P\left((\bar{X}-\bar{Y})-t_{1-\frac{\alpha}{2},n+m-2}\frac{S_p}{\sqrt{\frac{nm}{n+m}}} \leq \mu_x-\mu_y \leq (\bar{X}-\bar{Y})-t_{\frac{\alpha}{2},n+m-2}\frac{S_p}{\sqrt{\frac{nm}{n+m}}}\right) = 1-\alpha$$

E com isso chegamos ao seguinte intervalo de confiança bilateral (exato) para o parâmetro $\theta = \mu_x - \mu_y$:

$$IC_{\theta,100(1-\alpha)\%} = \left[(\bar{X}-\bar{Y})-t_{1-\frac{\alpha}{2},n+m-2}\frac{S_p}{\sqrt{\frac{nm}{n+m}}}\ ,\ (\bar{X}-\bar{Y})-t_{\frac{\alpha}{2},n+m-2}\frac{S_p}{\sqrt{\frac{nm}{n+m}}}\right].$$

$\|$

Ainda trabalhando com duas populações normais independentes, agora supondo variâncias não necessariamente iguais, podemos construir um intervalo de confiança para a razão entre as variâncias das duas populações.

Exemplo 5.13. *Sejam $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$variáveis aleatórias independentes. Suponha uma amostra de tamanho n de X e outra de tamanho m de Y. Defina $\theta = \sigma_x^2/\sigma_y^2$. Encontre um intervalo de confiança unilateral à direita para θ.*

Solução: Novamente o desafio será na construção de uma quantidade pivotal para θ. Depois é só seguir o passo-a-passo para construir o intervalo de confiança.

Veja que Q definida a seguir é quantidade pivotal para θ.

$$Q = \frac{S_y^2/\sigma_y^2}{S_x^2/\sigma_x^2} = \frac{S_y^2}{S_x^2}\theta.$$

De imediato podemos perceber que Q é função de $\theta = \sigma_x^2/\sigma_y^2$ e não é mais função de qualquer outro parâmetro desconhecido. Então falta verificar que a sua distribuição não depende de parâmetros desconhecidos.

Já sabemos que

$$\frac{(n-1)S_x^2}{\sigma_x^2} \sim \chi^2_{n-1} \quad \text{e} \quad \frac{(m-1)S_y^2}{\sigma_y^2} \sim \chi^2_{m-1},$$

além de serem variáveis aleatórias independentes. Então, de acordo com o Teorema 2.2,

$$Q = \frac{\frac{(n-1)S_x^2}{\sigma_x^2}/(n-1)}{\frac{(m-1)S_y^2}{\sigma_y^2}/(m-1)} = \frac{S_y^2/\sigma_y^2}{S_x^2/\sigma_x^2} \sim F_{m-1,n-1}.$$

Agora já vimos que Q é quantidade pivotal para $\theta = \sigma_x^2/\sigma_y^2$ e a partir dela vamos construir um intervalo de confiança unilateral à direita para θ.

$$\mathrm{P}\left(f_{\alpha,m-1,n-1} \leq \frac{S_y^2}{S_x^2}\theta\right) = 1-\alpha \quad \Rightarrow \quad \mathrm{P}\left(f_{\alpha,m-1,n-1}\frac{S_x^2}{S_y^2} \leq \theta\right) = 1-\alpha$$

E com isso o intervalo procurado é:

$$IC_{\theta,100(1-\alpha)\%} = \left[f_{\alpha,m-1,n-1}\frac{S_x^2}{S_y^2}\ ,\ \infty\right].$$

$\|$

O último exemplo dessa seção será de um intervalo para duas amostras independentes aproximado, obtido a partir de uma quantidade pivotal assintótica.

Exemplo 5.14. *Sejam $X \sim Bernoulli(p_x)$ e $Y \sim Bernoulli(p_y)$ variáveis aleatórias independentes. Suponha uma amostra de tamanho n de X e outra de tamanho m de Y. Defina $\theta = p_x - p_y$. Encontre um intervalo de confiança bilateral aproximado para θ.*
Solução: O enunciado do exemplo já indicou que este intervalo de confiança será aproximado e para isso precisamos de uma quantidade pivotal assintótica. Vamos construí-la seguindo a Proposição 5.2.

Como $\hat{p}_x = \bar{X}$ e $\hat{p}_y = \bar{Y}$ são os estimadores de máxima verossimilhança para os parâmetros p_x e p_y, respectivamente, aplicando invariância podemos chegar no estimador de máxima verossimilhança para θ: $\hat{\theta} = \bar{X} - \bar{Y}$.

De acordo com o Exemplo 5.8, a variância mínima para os estimadores não tendenciosos para p_x e p_y é

$$\mathrm{V}_{min}^{p_x}(p_x) = \frac{p_x(1-p_x)}{n} \quad \text{e} \quad \mathrm{V}_{min}^{p_y}(p_y) = \frac{p_y(1-p_y)}{m}.$$

Então, conhecendo o comportamento assintótico dos estimadores e máxima verossimilhança (Teorema 5.1), podemos afirmar que

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}(\bar{X}) + \mathrm{Var}(\bar{Y}) \to \frac{p_x(1-p_x)}{n} + \frac{p_y(1-p_y)}{m}.$$

E assim construímos uma quantidade pivotal assintótica ára θ,

$$Q = \frac{(\bar{X}-\bar{Y}) - (p_x - p_y)}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n} + \frac{\bar{Y}(1-\bar{Y})}{m}}} \xrightarrow{d} N(0,1)$$

A partir dela é possível construir o seguinte intervalo de confiança aproximado:

$$\mathrm{P}\left(-z_{1-\frac{1}{2}} \leq \frac{(\bar{X}-\bar{Y}) - (p_x - p_y)}{\sqrt{\frac{\bar{X}(1-\bar{X})}{n} + \frac{\bar{Y}(1-\bar{Y})}{m}}} \leq z_{1-\frac{1}{2}}\right) \xrightarrow[n\to\infty]{} 1-\alpha$$

Depois de algumas manipulações algébricas chegamos ao seguinte intervalo de confiança para $\theta = p_x - p_y$:

$$\left[(\bar{X}-\bar{Y}) - z_{1-\frac{1}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n} + \frac{\bar{Y}(1-\bar{Y})}{m}}\ ,\ (\bar{X}-\bar{Y}) + z_{1-\frac{1}{2}}\sqrt{\frac{\bar{X}(1-\bar{X})}{n} + \frac{\bar{Y}(1-\bar{Y})}{m}}\right].$$

$\|$

Exercícios da Seção 5.4

5.4.1 ([Mood, 1974] - Cap 8) Para testar duas novas linhas de híbridos em condições normais de cultivo, uma empresa de sementes selecionou oito fazendas ao acaso plantou em cada uma delas as duas linhas e híbridos. Os rendimentos (convertidos em alqueires por acre) para os oito locais foram:

Linha A:	86	87	56	93	84	93	75	79
Linha B:	80	79	58	91	77	82	74	66

Suponha os dois rendimentos conjuntamente normalmente distribuídos. Estime a diferença entre os rendimentos médios por um intervalo de confiança de 95%. Que conclusão pode ser tirada a partir do intervalo de confiança encontrado. Dica: Se X = rendimentos da linha A e Y = rendimentos da linha B, então $W = X - Y \sim N(\mu_x - \mu_y, \sigma^2)$. Nesse caso não vamos usar os intervalos para duas amostras independentes (até porque as amostras não são independentes), e sim construir um intervalo para a média de W.

5.4.2 Uma pesquisa foi realizada com o objetivo de verificar a eficácia de uma certa vacina no combate a uma certa doença. Dois grupos de pessoas, o primeiro com 195.365 pessoas e o segundo com 154.568 pessoas, participaram do estudo. A vacina foi aplicada nas pessoas do segundo grupo, enquanto que as pessoas do primeiro grupo receberam um placebo. Após a vacinação os casos da doença observados nos dois grupos foram 135 e 45, respectivamente. Estime a diferença de prevalência da doença entre os não vacinados e os vacinados, por um intervalo de confiança unilateral (aproximado) com coeficiente de confiança de 90%. Qual seria o intervalo mais adequado nesse caso, unilateral à direita ou à esquerda?

5.4.3 ([Bolfarine e Sandoval, 2001] - Seção 5.6 e [Roussas, 2003] - Seção 10.1)
Seja $X_1, \ldots, X_n$ uma amostra aleatória de $X \sim Exp(\theta_1)$ e $Y_1, \ldots, Y_m$ uma amostra aleatória de $Y \sim Exp(\theta_2)$. Assuma independência entre as duas amostras.

(a) Obtenha uma quantidade pivotal para construir um intervalo de confiança para θ_1/θ_2.

(b) Encontre um intervalo de confiança bilateral para θ_1/θ_2.

(c) Suponha as seguintes amostras de X e Y. Encontre a estimativa para o intervalo obtido no item (b). Obs: $\sum_{i=1}^{10} x_i = 6,646$ e $\sum_{i=1}^{15} y_i = 5,897$.

X:	0,199	0,257	0,639	0,405	0,412	0,480	0,826	2,039	0,763	0,626
Y:	0,119	0,337	0,167	0,750	0,592	0,019	0,666	0,015	0,101	0,248
	0,906	0,597	0,452	0,820	0,108					

5.5 Mais Alguns Exercícios do Capítulo 5

5.1. ([Larson, 1982] - Capítulo 7) A cada hora uma estação de rádio transmite um “beep”. A estação tem um quartz que é usado para engatilhar o instante do beep. A menos que o quartz seja de qualidade muito boa e mantido sob condições extremamente cuidadosas, este não é 100% acurado. Assumindo que a diferença entre a hora do “beep” e a hora exata é, em micro-segundos, uma v.a. $X \sim U[-\theta, \theta]$, e considerando os valores abaixo um amostra de X, construa um intervalo de confiança para θ. Dica: Se $X \sim U[-\theta, \theta] \Rightarrow Y = |X| \sim U[0, \theta]$.

$221 \quad 265 \quad -140 \quad 327 \quad -401 \quad 308 \quad -317 \quad 447 \quad -137 \quad -228 \quad -477 \quad 69 \quad 475 \quad 56 \quad -101.$

5.2. ([Larson, 1982] - Capítulo 7) O tempo até a falha de um equipamento eletrônico pode ser considerado uma variável aleatória exponencial com parâmetro λ (desconhecido), isto é,

$T \sim Exp(\lambda)$ onde T é o tempo até a ocorrência de uma falha. Foram testados 20 equipamentos e medidos os respectivos tempos (em horas) até a ocorrência de falha. Supondo independência entre tais observações, levantou-se a seguinte amostra após os testes:

377,6	590,8	72,9	69,9	218,0	1.447,5	614,8	269,8	478,3	73,5
695,4	381,0	618,8	2.212,0	527,3	517,6	938,0	327,4	168,5	294,2

Para as contas dos itens a seguir considere o coeficiente de confiança de 90% e, se preciso, $\sum_{i=1}^{20} t_i = 10.893,3$, onde t_i é o tempo até a falha do i-ésimo equipamento testado.

(a) Encontre a estimativa para o intervalo de confiança bilateral para o parâmetro $\mu = 1/\lambda$.

(b) Encontre a estimativa para o intervalo de confiança bilateral para o parâmetro $p = \mathrm{P}(T > 100)$, que é a confiabilidade do equipamento para um período de 100 horas.

5.3. Usando os mesmos dados do Exercício 5.2, defina limites inferiores para os parâmetros μ e p a partir de intervalos de confiança unilaterais para tais parâmetros. Use novamente coeficiente de confiança de 90%. Qual a interpretação para tais limites?

5.4. Repita o exercício 5.2 considerando agora intervalos assintóticos. Veja que para isso você precisa definir quantidades pivotais assintóticas para μ e p.

5.5. Seja X uma variável aleatória contínua cuja função densidade depende do parâmetro $\theta > 0$ e é definida por

$$f_X(x) = \frac{\theta 4^\theta}{x^{\theta+1}}\ , \quad x > 4.$$

Veja que $X \sim Pareto(\theta, 4)$ e que esta distribuição é a mesma dos Exercícios 2.5.1 e 3.11

(a) Encontre um intervalo de confiança exato unilateral à direita para θ com confiabilidade de 90% para uma amostra de tamanho 14.

(b) Encontre um intervalo de confiança exato unilateral à esquerda para $\eta = \mathrm{P}(X > 8)$ com confiabilidade de 90% para uma amostra de tamanho 14.

(c) Considere os valores abaixo uma amostra da população X. Encontre a estimativa para os intervalos de confiança deduzidos nos itens acima para os parâmetros θ e η.

$9,95 \quad 5,01 \quad 4,78 \quad 4,02 \quad 6,60 \quad 4,83 \quad 4,42 \quad 4,23 \quad 23,58 \quad 4,96 \quad 4,39 \quad 14,37 \quad 7,12 \quad 4,57$

5.6. Seja X uma variável aleatória contínua cuja função densidade depende do parâmetro $\theta > 0$ e é definida por

$$f_X(x) = 3\theta^3 x^2 e^{-(\theta x)^3}\ , \quad x > 0.$$

Veja que $X \sim Weib(3, 1/\theta)$ e que esta distribuição é a mesma dos Exercícios 2.5.3 e 3.13.

(a) Encontre um intervalo de confiança exato unilateral à esquerda para θ com confiabilidade de 95% para uma amostra de tamanho 9.

(b) Encontre um intervalo de confiança exato unilateral à direita para $\eta = \mathrm{P}(X > 1)$ com confiabilidade de 95% para uma amostra de tamanho 9.

(c) Considere os valores abaixo uma amostra da população X. Encontre a estimativa para os intervalos de confiança deduzidos nos itens acima para os parâmetros θ e η.

$2,20 \quad 1,99 \quad 1,65 \quad 0,92 \quad 2,34 \quad 0,95 \quad 0,77 \quad 1,49 \quad 1,55$

5.7. Seja X uma variável aleatória contínua cuja função densidade depende do parâmetro $\theta > 0$ e é definida por

$$f_X(x) = 2^\theta \theta x^{\theta-1}, \quad 0 < x < \frac{1}{2}.$$

Veja que $2X \sim Beta(\theta, 1)$ e que esta distribuição é a mesma dos Exercícios 2.5.4 e 3.14.

(a) Encontre um intervalo de confiança exato bilateral para θ com confiabilidade de 99% para uma amostra de tamanho 8.

(b) Encontre um intervalo de confiança exato bilateral para $\mu = \mathrm{E}(X)$ com confiabilidade de 99% para uma amostra de tamanho 8.

(c) Considere os valores abaixo uma amostra da população X. Encontre a estimativa para os intervalos de confiança deduzidos nos itens acima para os parâmetros θ e μ.

$$0,210 \quad 0,364 \quad 0,486 \quad 0,386 \quad 0,171 \quad 0,407 \quad 0,435 \quad 0,275$$

5.8. Seja X uma variável aleatória discreta cuja função de probabilidade depende do parâmetro $0 < p < 1$ e é definida por

$$p_X(x) = \binom{15}{x} p^x (1-p)^{15-x}, \quad x = 0, 1, \ldots, 15$$

Veja que $X \sim Binom(15, p)$ e que esta distribuição é a mesma dos Exercícios 2.5.5 e 3.15.

(a) Encontre um intervalo de confiança aproximado bilateral para p com confiabilidade de 85% para uma amostra de tamanho 25.

(b) Apresente uma quantidade pivotal assintótica para $\mu = \mathrm{E}(X)$.

(c) Encontre um intervalo de confiança aproximado bilateral para $\mu = \mathrm{E}(X)$ com confiabilidade de 85% para uma amostra de tamanho 25.

(d) Considere os valores abaixo uma amostra da população X. Encontre a estimativa para os intervalos de confiança deduzidos nos itens acima para os parâmetros p e μ.

3 3 4 6 2 6 7 4 4 1 2 2 5 3 5 4 5 8 3 5 6 2 4 2 3

5.9. O diretor de uma fábrica lhe deu a tarefa de estimar um intervalo de confiança para a probabilidade desta fábrica produzir um certo dispositivo com defeito. Para que você pudesse fazer os cálculos o diretor escalou um funcionário para ir diariamente a linha de produção selecionar de forma aleatória dispositivos até que fosse encontrado um com defeito. Esse funcionário registou a cada dia o número de dispositivos sem defeitos testados. A partir da amostra coletada faça o que se pede.

(a) Ao final de n dias de trabalho o funcionário entregou para você os registros obtidos. Baseados nesses dados, apresente a amostra aleatória, a população que você vai considerar nas suas análises e o parâmetro desconhecido e de interesse.

(b) Encontre uma quantidade pivotal assintótica para o parâmetro de interesse.

(c) Deduza um intervalo de confiança aproximado bilateral para o parâmetro em questão.

(d) Após 25 dias de trabalho o funcionário entregou os seguintes registros para o número de dispositivos sem defeitos testados por dia até que fosse encontrado o primeiro com defeito.

14 6 8 53 17 27 25 29 16 8 14 3 5 8 2 48 32 15 54 12 24 6 22 6 30

Apresente a estimativas do intervalo de confiança para a probabilidade de um dispositivo da fábrica ser produzido com defeito. Use a confiabilidade que achar adequada. Junto, formule uma frase resposta a ser entregue ao diretor da fábrica.

Capítulo 6

Teste de Hipótese

Como nos demais capítulos, considere uma população X com distribuição conhecida e de parâmetro(s) desconhecido(s). Nesse capítulo veremos como testar hipóteses levantadas sobre os parâmetros da distribuição da população a partir da observação de uma amostra de tamanho n da população X.

6.1 Conceitos Gerais

Para começar precisamos definir de forma clara o que é uma hipótese estatística.

Definição 6.1. *Uma hipótese estatística H, ou simplesmente hipótese H, é qualquer afirmação sobre a distribuição de probabilidade de uma população X. Se a hipótese H especifica completamente a distribuição de X, isto é, sendo H verdadeira a distribuição de X é conhecida, então dizemos que H é uma hipótese simples. Caso contrário, dizemos que H é uma hipótese composta.*

Exemplo 6.1. *Seja $X \sim N(\mu, 1)$. Considere as três hipóteses definidas a seguir.*

$$\text{(i) } H : \mu = 2 \qquad \text{(ii) } H : \mu \neq 2 \qquad \text{(iii) } H : \mu < 2.$$

Quais são simples e quais são compostas?

Solução:

A hipótese definida em (i) é uma hipótese simples, já as outras duas são compostas. ||

No contexto de Inferência Estatística clássica, as hipóteses formuladas sobre a distribuição da população X serão sempre referentes aos parâmetros desconhecidos dessa população. E tais hipóteses aparecem em dupla, ou seja, serão sempre apresentadas duas hipóteses: H_0 e H_1. Chamamos H_0 de hipótese nula e H_1 de hipótese alternativa. Caso H_0 seja falsa aceitamos como verdadeira a hipótese H_1.

Definição 6.2. *Um teste de hipótese (ou teste estatístico) é qualquer regra de decisão Γ para aceitar ou rejeitar a hipótese nula, H_0. Ou seja, Γ associa cada observação da amostra à uma das seguintes decisões: aceitar H_0 (e rejeitar H_1); ou rejeitar H_0 (e aceitar H_1).*

Exemplo 6.2. *Seja $X \sim N(\mu, 1)$. Considere as seguintes hipóteses:*

$$H_0 : \mu = 0 \qquad \textit{contra} \qquad H_1 : \mu \neq 0.$$

Apresente testes estatísticos para as hipóteses apresentadas.

Solução:

Γ_1: aceitar sempre H_0, independente da amostra observada.
Γ_2: aceitar H_0 se $-1 \leq \bar{x} \leq 1$. ||
Γ_3: aceitar H_0 se $-0,5 \leq \bar{x} \leq 0,5$.

6.1.1 Região Crítica e Região de Aceitação

Para conseguirmos comparar diferentes testes e escolher um adequado precisamos primeiro do conceito de região crítica e região de aceitação de um teste.

Defina S como o conjunto de todas as possíveis realizações de uma amostra $\mathbf{X}$ da população X.

$$S = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{X}(w) = \mathbf{x} \text{ para algum } w \in \Omega\}.$$

Qualquer teste estatístico Γ associa cada realização da amostra $\mathbf{x} \in S$ à decisão de aceitar ou rejeitar H_0, isto é,

$$\begin{array}{rccl} \Gamma: & S & \mapsto & \{\text{“aceita”}, \text{“rejeita”}\} \\ & \mathbf{x} & \to & \Gamma(\mathbf{x}). \end{array}$$

Então podemos dizer que qualquer teste Γ define uma partição de S formada pelos conjuntos A e R definidos por:

$$\begin{array}{rcl} A & = & \{\mathbf{x} \in S \mid \Gamma(\mathbf{x}) = \text{“aceita”}\} \\ R & = & \{\mathbf{x} \in S \mid \Gamma(\mathbf{x}) = \text{“rejeita”}\}. \end{array}$$

Veja que $A \cap R = \emptyset$ e que $A \cup R = S$, logo A e R de fato formam uma partição do conjunto S. O conjunto A é chamado de região de aceitação e representa todas as observações da amostra que levam à aceitação da hipótese nula. Já o conjunto R é chamado de região crítica, ou região de rejeição, e representa todas as observações da amostra que levam à rejeição da hipótese nula.

Exemplo 6.3. *Defina as regiões de aceitação e crítica para os três testes apresentados no Exemplo 6.2.*
Solução: Primeiro vamos encontrar as regiões para o teste Γ_1: aceitar sempre H_0, independente da amostra observada. Veja que qualquer que seja a observação da amostra o teste aceita H_0 sempre. Logo,

$$A_1 = S \qquad \text{e} \qquad R_1 = \emptyset.$$

Já para o teste Γ_2: aceitar H_0 se $-1 \leq \bar{x} \leq 1$,

$$A_2 = \{\mathbf{x} \in S \mid -1 \leq \bar{x} \leq 1\} \quad \text{e} \quad R_2 = \{\mathbf{x} \in S \mid \bar{x} < -1 \text{ ou } \bar{x} > 1\}.$$

De forma análoga, para o teste Γ_3: aceitar H_0 se $-0,5 \leq \bar{x} \leq 0,5$,

$$A_3 = \{\mathbf{x} \in S \mid -0,5 \leq \bar{x} \leq 0,5\} \quad \text{e} \quad R_3 = \{\mathbf{x} \in S \mid \bar{x} < -0,5 \text{ ou } \bar{x} > 0,5\}.$$

‖

6.1.2 Probabilidades de Erros Tipo I e II

Entre três testes apresentados no Exemplo 6.2, qual deles é mais interessante? O que seria um teste interessante? Veja que a definição de teste de hipótese é bem ampla, qualquer regra de decisão pode ser considerada um teste, mesmo aquelas mais estranhas como o teste Γ_1 do Exemplo 6.2. O que vai nos auxiliar na escolha de um teste adequado são algumas medidas de probabilidade de erros, apresentadas a seguir.

Qualquer que seja o teste criado temos sempre a chance de errar. Por exemplo, podemos aceitar H_0 quando na verdade H_0 é falsa, ou rejeitar H_0 quando esta for verdadeira. Queremos definir testes para os quais a probabilidade de cometer um erro não seja muito grande, ou pelo menos, seja conhecida e controlada. Vamos formalizar essa ideia.

Pensando em escolher testes com baixa probabilidade de errar, queremos um teste tal que a probabilidade de rejeitar H_0 seja baixa se a hipótese H_0 for verdadeira. Então vamos procurar testes com $\mathrm{P}(\mathbf{X} \in R \mid H_0 \text{ verdadeira})$ pequena.

Exemplo 6.4. *Calcule* $\mathrm{P}(\mathbf{X} \in R \mid H_0$ *verdadeira) para cada um dos três testes apresentados no Exemplo 6.2.*
Solução: Primeiro para o teste Γ_1.

$$\mathrm{P}(\mathbf{X} \in R_1 \mid H_0 \text{ verdadeira}) = \mathrm{P}(\mathbf{X} \in \emptyset \mid \mu = 0) = 0.$$

Vamos fazer as contas agora para o teste Γ_2.

$$\begin{aligned}
\mathrm{P}(\mathbf{X} \in R_2 \mid H_0 \text{ verdadeira}) &= 1 - \mathrm{P}(\mathbf{X} \in A_2 \mid H_0 \text{ verdadeira}) \\
&= 1 - \mathrm{P}(-1 \leq \bar{X} \leq 1 \mid \mu = 0) \\
&= 1 - \mathrm{P}\left(\frac{-1-0}{\sqrt{1/n}} \leq \frac{\bar{X}-0}{\sqrt{1/n}} \leq \frac{1-0}{\sqrt{1/n}} \;\middle|\; \mu = 0 \right) \\
&= 1 - \mathrm{P}\left(-\sqrt{n} \leq Z \leq \sqrt{n} \;\middle|\; Z \sim N(0,1)\right) \\
&= 1 - \left(2\Phi(\sqrt{n}) - 1\right) \\
&= 2 - 2\Phi(\sqrt{n}).
\end{aligned}$$

Para terminas, façamos as contas para Γ_3.

$$\begin{aligned}
\mathrm{P}(\mathbf{X} \in R_3 \mid H_0 \text{ verdadeira}) &= 1 - \mathrm{P}(\mathbf{X} \in A_3 \mid H_0 \text{ verdadeira}) \\
&= 1 - \mathrm{P}(-0,5 \leq \bar{X} \leq 0,5 \mid \mu = 0) \\
&= 1 - \mathrm{P}\left(\frac{-0,5-0}{\sqrt{1/n}} \leq \frac{\bar{X}-0}{\sqrt{1/n}} \leq \frac{0,5-0}{\sqrt{1/n}} \;\middle|\; \mu = 0 \right) \\
&= 1 - \mathrm{P}\left(-\frac{\sqrt{n}}{2} \leq Z \leq \frac{\sqrt{n}}{2} \;\middle|\; Z \sim N(0,1) \right) \\
&= 1 - \left(2\Phi\left(\frac{\sqrt{n}}{2}\right) - 1 \right) \\
&= 2 - 2\Phi\left(\frac{\sqrt{n}}{2}\right).
\end{aligned}$$

Se o critério de comparação entre os testes fosse apenas o valor de $\mathrm{P}(\mathbf{X} \in R \mid H_0$ verdadeira), o teste Γ_1 pareceria o mais interessante, já que é o com menor probabilidade, mas veremos em breve que ele não é tão interessante assim. Entre Γ_2 e Γ_3, o com menor valor de $\mathrm{P}(\mathbf{X} \in R \mid H_0$ verdadeira) é o Γ_2. Veja porque,

$$\mathrm{P}(\mathbf{X} \in R_2 \mid H_0 \text{ verdadeira}) = 2 - 2\Phi(\sqrt{n}) < 2 - 2\Phi\left(\frac{\sqrt{n}}{2}\right) = \mathrm{P}(\mathbf{X} \in R_3 \mid H_0 \text{ verdadeira}).$$

$\|$

Continuando na busca por calcular probabilidades de cometer erros, se for falsa H_0 queremos um teste tal que a probabilidade de aceitar H_0 seja baixa. Então vamos procurar testes com $\mathrm{P}(\mathbf{X} \in A \mid H_0$ falsa) pequena.

Exemplo 6.5. *Calcule* $\mathrm{P}(\mathbf{X} \in A \mid H_0$ *falsa) para cada um dos três testes apresentados no Exemplo 6.2.*
Solução: Primeiro para o teste Γ_1.

$$\mathrm{P}(\mathbf{X} \in A_1 \mid H_0 \text{ falsa}) = \mathrm{P}(\mathbf{X} \in S \mid \mu \neq 0) = 1.$$

O que nos mostra que esse teste não é interessante mesmo.

Não é possível encontrar o valor de $\mathrm{P}(\mathbf{X} \in A_2 \mid H_0 \text{ falsa})$ e nem de $\mathrm{P}(\mathbf{X} \in A_3 \mid H_0 \text{ falsa})$ pois H_0 falsa é o mesmo que H_1 verdadeira e H_1 é uma hipótese composta, por isso não conhecemos a distribuição de X e nem de $\bar{X}$ se H_0 for falsa. Mas podemos comparar esses valores mesmo sem conseguir calculá-los.

Veja que $A_3 \subset A_2$. Por isso, se $\{\mathbf{X} \in A_3\} \Rightarrow \{\mathbf{X} \in A_2\}$. Assim podemos estabelecer a seguinte relação entre esses dois eventos: $\{\mathbf{X} \in A_3\} \subset \{\mathbf{X} \in A_2\}$. Com isso temos

$$\mathrm{P}(\mathbf{X} \in A_2 \mid H_0 \text{ falsa}) > \mathrm{P}(\mathbf{X} \in A_3 \mid H_0 \text{ falsa}).$$

$\|$

A discussão em cima do Exemplo 6.2 já nos mostrou a ideia de possíveis erros cometidos por um teste. Vamos formalizar esse conceito.

Definição 6.3. *Seja Γ um teste estatístico para as hipóteses H_0 contra H_1.*

(i) Chama-se de erro tipo I o erro cometido quando Γ rejeita H_0, dado que H_0 é verdadeira.

(ii) Chama-se de erro tipo II o erro cometido quando Γ aceita H_0, dado que H_0 é falsa.

O que vamos usar para avaliar um teste é a probabilidade de cometer cada um desses dois erros. São elas,

$$\begin{aligned} \alpha &= \mathrm{P}(\text{erro tipo I}) = \mathrm{P}(\mathbf{X} \in R \mid H_0 \text{ verdadeira}). \\ \beta &= \mathrm{P}(\text{erro tipo II}) = \mathrm{P}(\mathbf{X} \in A \mid H_0 \text{ falsa}). \end{aligned}$$

Atenção, os erros tipo I e tipo II não são eventos complementares, ou seja, $\alpha + \beta$ não é necessariamente 1. O valor $1 - \beta$ é chamado de poder de um teste. O poder de um teste é a probabilidade de rejeitar H_0 quando esta é falsa,

$$\textit{poder} = \mathrm{P}(\mathbf{X} \in R \mid H_0 \text{ falsa}) = 1 - \beta.$$

Veremos com mais detalhes o conceito de poder de um teste Seção 6.3. Vamos continuar trabalhando com a população $X \sim N\mu, 1)$ e explorar mais um pouco os conceitos de erro tipo I e tipo II.

Exemplo 6.6. *Seja $X \sim N(\mu, 1)$. Considere as hipóteses*

$$H_0 : \mu = 2 \qquad \textit{contra} \qquad H_1 : \mu = 3$$

e o teste Γ definido pela região crítica

$$R = \{\mathbf{x} \in S \mid \bar{x} \geq c\}.$$

Encontre as probabilidades dos erros Tipo I e Tipo II em função das constante c e do tamanho da amostra n.

Solução: Vamos começar pela probabilidade do erro Tipo I, que pode ser calculada uma vez que H_0 é uma hipótese simples.

$$\alpha = \mathrm{P}(\text{erro Tipo I}) = \mathrm{P}(\mathbf{X} \in R \mid H_0 \text{ verdadeira}) = \mathrm{P}(\bar{X} \geq c \mid \mu = 2)$$

Veja a Figura 6.1, a área em cinza escuro representa o valor de α. Dado $\mu = 2$, $X \sim N(2, 1)$ e $\bar{X} \sim N\left(2, \frac{1}{n}\right)$. Então, seguindo,

$$\alpha = \mathrm{P}\left(\frac{\bar{X} - 2}{\sqrt{1/n}} \geq \frac{c - 2}{\sqrt{1/n}} \mid \mu = 2\right) = 1 - \mathrm{P}\left(\frac{\bar{X} - 2}{\sqrt{1/n}} \leq (c - 2)\sqrt{n} \mid \mu = 2\right) = 1 - \Phi\left((c - 2)\sqrt{n}\right).$$

Agora o cálculo da probabilidade do erro Tipo II, que pode ser calculado uma vez que H_1 é uma hipótese simples.

$$\beta = \mathrm{P}(\text{erro Tipo II}) = \mathrm{P}(\mathbf{X} \in A \mid H_0 \text{ falsa}) = \mathrm{P}(\mathbf{X} \in A \mid H_1 \text{ verdadeira}) = \mathrm{P}(\bar{X} < c \mid \mu = 3).$$

Veja a Figura 6.1, a área em cinza claro representa o valor de β. Dado $\mu = 3$, $X \sim N(3,1)$ e $\bar{X} \sim N\left(3, \frac{1}{n}\right)$. Então, seguindo,

$$\beta = \mathrm{P}\left(\frac{\bar{X}-3}{\sqrt{1/n}} < \frac{c-3}{\sqrt{1/n}} \mid \mu = 3\right) = \mathrm{P}\left(\frac{\bar{X}-3}{\sqrt{1/n}} \leq (c-3)\sqrt{n} \mid \mu = 3\right) = \Phi\left((c-3)\sqrt{n}\right).$$

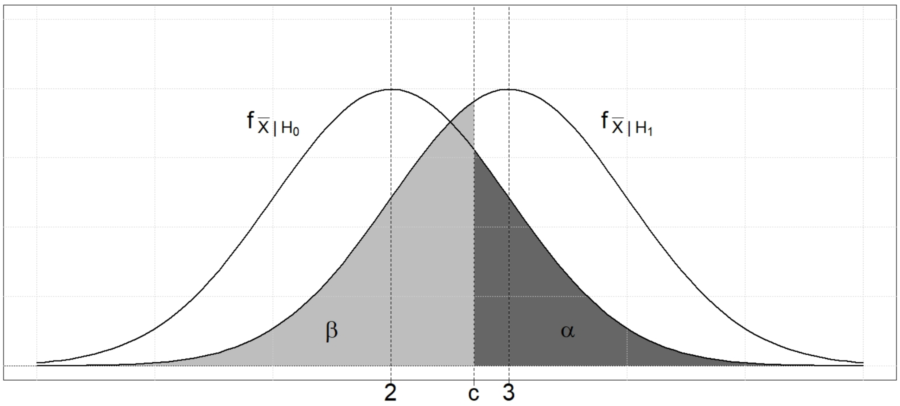


Figura 6.1: Erros Tipo I e II

Primeiro veja que uma escolha interessante para a constante c é $2 < c < 3$. Isso porque se $c < 2$ temos $\alpha > 0,5$. E se $c > 3$ temos $\beta > 0,5$. O que sem dúvida são valores grandes para as probabilidades dos erros. Para entender melhor, observe a Figura 6.1 e imagine como ficariam as áreas em cinza escuro (α) se $c < 2$ e como ficaria a área em cinza claro (β) se $c > 3$.

Assumindo então $2 < c < 3$, observe que conforme n cresce tanto α quanto β diminuem. Dessa forma a escolha do tamanho da amostra n é uma forma de controlar as probabilidades de erros do teste. ‖

Exemplo 6.7. *Continuação do exemplo com $X \sim N(\mu, 1)$, hipóteses*

$$H_0 : \mu = 2 \qquad \textit{contra} \qquad H_1 : \mu = 3$$

e teste Γ definido pela região crítica

$$R = \{\mathbf{x} \in S \mid \bar{x} \geq c\}.$$

Considerando uma amostra de tamanho $n = 9$, encontre o valor de c para o qual $\alpha = 0,05$. Nesse caso, qual será o valor de β? .

Solução: Já encontramos $\alpha = 1 - \Phi((c-2)\sqrt{n})$. Queremos o valor de c tal que $\alpha = 1 - \Phi((c-2)\sqrt{9}) = 0,05$, ou seja, queremos c tal que $\Phi(3(c-2)) = 0,95$. Após uma consulta na Tabela A.1 chegamos em

$$3(c-2) = 1,64 \qquad \Rightarrow \qquad c = \frac{1,64}{3} + 2 = 2,546667.$$

Nesse caso,

$$\beta = \Phi((2,546667 - 3)\sqrt{9}) = \Phi(-1,36) = 1 - \Phi(1,36) = 1 - 0,9131 = 0,0869.$$

Veja que fixado um $\alpha = 0,05$ para uma amostra de tamanho $n = 9$ resultou em uma probabilidade do erro tipo II de $0,0869$. Se quiséssemos uma probabilidade de erro tipo II também de no máximo 0,05, uma alternativa seria aumentar o tamanho da amostra e refazer as contas. Assim podemos encontrar um valor de n que nos garante tanto α quanto β menores que $0,05$, por exemplo. ‖

Exemplo 6.8. *Ainda uma continuação do exemplo com $X \sim N(\mu, 1)$, hipóteses*

$$H_0 : \mu = 2 \qquad \textit{contra} \qquad H_1 : \mu = 3$$

e teste Γ definido pela região crítica

$$R = \{\mathbf{x} \in S \mid \bar{x} \geq c\}.$$

Encontre um valor de c e um tamanho de amostra n para os quais podemos garantir que o teste tenha probabilidade dos erros Tipo I e II de no máximo $0,05$ cada.

Solução: Já vimos que com $n = 9$ não conseguimos o que se pede. Se refizermos as contas do Exemplo 6.7 para $n = 10$ veremos que ainda não resulta em $\beta \leq 0,05$, mas para $n = 11$ sim. Veja as contas.

Queremos o valor de c tal que $\alpha = 1 - \Phi((c-2)\sqrt{11}) = 0,05$, ou seja, queremos c tal que $\Phi((c-2)\sqrt{11}) = 0,95$. Após uma consulta na Tabela A.1 chegamos em

$$(c-2)\sqrt{11} = 1,64 \qquad \Rightarrow \qquad c = \frac{1,64}{\sqrt{11}} + 2 = 2,494479.$$

Nesse caso,

$$\beta = \Phi(\sqrt{11}(2,494479 - 3)) = \Phi(-1,676625) = 1 - \Phi(1.676625) = 1 - 0,9532 = 0,0468.$$

‖

Exercícios da Seção 6.1

6.1.1 ([Larson, 1982] - Capítulo 8) Seja $\{X_1, X_2, X_3, X_4\}$ amostra aleatória de $X \sim Ber(p)$. Suponha as hipóteses $H_0 : p = \frac{1}{4}$ contra $H_1 : p = \frac{3}{4}$. Considere o teste que rejeita H_0 somente se forem obtidos 4 sucessos na amostra. Calcule α e β para esse teste.

6.1.2 ([Larson, 1982] - Capítulo 8) Dado que $X \sim U[0, \theta]$, é de interesse testar $H_0 : \theta = 1$ contra $H_1 : \theta = 2$, a partir de uma amostra aleatória de tamanho $n = 2$ e com o teste Γ : “Rejeite-se H_0 se e só se $\bar{x} > 0.99$”. Calcule as probabilidades de erro tipo 1 e tipo 2 deste teste.

6.2 Teste para Hipóteses Nula e Alternativa Simples

Nessa seção veremos que para as hipóteses

$$H_0 : \theta = \theta_0 \quad \text{contra} \quad H_1 : \theta = \theta_1$$

é possível encontrar o teste com menor $\beta = \text{P(erro tipo II)}$ entre todos os testes com $\text{P(erro tipo I)} = \alpha$. Ou seja, fixado um valor de α encontramos o teste com maior poder (maior valor de $1-\beta$) entre todos os testes com $\text{P(erro tipo I)} = \alpha$. Por isso esse teste é chamado de teste mais poderoso para hipóteses nula e alternativa simples. Esse é o resultado do Lema de Neuman-Pearson, apresentado no Teorema 6.1 a seguir.

Teorema 6.1. ***Lema de Neuman-Pearson***

Considere as hipóteses $H_0 : \theta = \theta_0$ contra $H_1 : \theta = \theta_1$. Para essas hipóteses considere o teste Γ definido pela região crítica

$$R = \{\mathbf{x} \in S \mid L(\theta_1, \mathbf{x}) \geq KL(\theta_0, \mathbf{x})\},$$

onde $K > 0$ e L é a função de verossimilhança da amostra. Sejam $\alpha = \mathrm{P}(\mathbf{X} \in R \mid \theta = \theta_0)$ e $\beta = \mathrm{P}(\mathbf{X} \in A \mid \theta = \theta_1)$ as probabilidades dos erros do tipo I e II para o teste Γ, respectivamente.

Então, qualquer outro teste $\tilde{\Gamma}$ com probabilidade do erro tipo I dada por $\tilde{\alpha} = \alpha$ tem probabilidade do erro tipo II dada por $\tilde{\beta} \geq \beta$.

Demonstração. Seja $\tilde{\Gamma}$ um teste para as hipóteses H_0 e H_1 com Região Crítica $\tilde{R}$, Região de Aceitação $\tilde{A}$, probabilidade do erro tipo I igual a $\tilde{\alpha} = \alpha$ e probabilidade do erro tipo II igual a $\tilde{\beta}$. Vamos calcular $\tilde{\beta} - \beta$ e verificar que $\tilde{\beta} - \beta \geq 0$, logo, $\tilde{\beta} \geq \beta$.

$$\begin{aligned}
\tilde{\beta} - \beta &= \mathrm{P}(X \in \tilde{A} \mid \theta = \theta_1) - \mathrm{P}(X \in A \mid \theta = \theta_1) \\
&= (1 - \mathrm{P}(X \in \tilde{R} \mid \theta = \theta_1)) - (1 - \mathrm{P}(X \in R \mid \theta = \theta_1)) \\
&= \mathrm{P}(X \in R \mid \theta = \theta_1) - \mathrm{P}(X \in \tilde{R} \mid \theta = \theta_1) \\
&= \int_R L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} - \int_{\tilde{R}} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= \int_{R \cap S} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} - \int_{\tilde{R} \cap S} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= \int_{R \cap (\tilde{R} \cup \tilde{A})} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} - \int_{\tilde{R} \cap (R \cup A)} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= \int_{(R \cap \tilde{R}) \cup (R \cap \tilde{A})} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} - \int_{(\tilde{R} \cap R) \cup (\tilde{R} \cap A)} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= \int_{R \cap \tilde{R}} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} + \int_{R \cap \tilde{A}} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} - \int_{\tilde{R} \cap R} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} - \int_{\tilde{R} \cap A} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= \cancel{\int_{R \cap \tilde{R}} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x}} + \int_{R \cap \tilde{A}} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} - \cancel{\int_{\tilde{R} \cap R} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x}} - \int_{\tilde{R} \cap A} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= \int_{R \cap \tilde{A}} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} - \int_{\tilde{R} \cap A} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x}.
\end{aligned}$$

Sabemos que se $\mathbf{x} \in R$, então $L(\theta_1, \mathbf{x}) \geq KL(\theta_0, \mathbf{x})$. Logo, para $B \subset R$ é verdade que $\int_B L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} \geq \int_B KL(\theta_0, \mathbf{x})\,\mathrm{d}\mathbf{x}$. Em particular, $R \cap \tilde{A}$ é subconjunto de R e por isso podemos afirmar que $\int_{R \cap \tilde{A}} L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} \geq \int_{R \cap \tilde{A}} KL(\theta_0, \mathbf{x})\,\mathrm{d}\mathbf{x}$.

Analogamente, se $\mathbf{x} \in A$, então $L(\theta_1, \mathbf{x}) < KL(\theta_0, \mathbf{x})$. Logo, para $C \subset A$ é verdade que $\int_C L(\theta_1, \mathbf{x})\,\mathrm{d}\mathbf{x} \leq \int_C KL(\theta_0, \mathbf{x})\,\mathrm{d}\mathbf{x}$. Em particular, $\tilde{R} \cap A$ é subconjunto de A e então

$\int_{\tilde{R}\cap A} L(\theta_1, \mathbf{x})\, \mathrm{d}\mathbf{x} \leq \int_{\tilde{R}\cap A} KL(\theta_0, \mathbf{x})\, \mathrm{d}\mathbf{x}$, ou, $-\int_{\tilde{R}\cap A} L(\theta_1, \mathbf{x})\, \mathrm{d}\mathbf{x} \geq -\int_{\tilde{R}\cap A} KL(\theta_0, \mathbf{x})\, \mathrm{d}\mathbf{x}$. Assim,

$$
\begin{aligned}
\tilde{\beta} - \beta &= \int_{R\cap\tilde{A}} L(\theta_1, \mathbf{x})\, \mathrm{d}\mathbf{x} - \int_{\tilde{R}\cap A} L(\theta_1, \mathbf{x})\, \mathrm{d}\mathbf{x}. \\
&\geq \int_{R\cap\tilde{A}} KL(\theta_0, \mathbf{x})\, \mathrm{d}\mathbf{x} - \int_{\tilde{R}\cap A} KL(\theta_0, \mathbf{x})\, \mathrm{d}\mathbf{x} \\
&= K\left(\int_{R\cap\tilde{A}} L(\theta_0, \mathbf{x})\, \mathrm{d}\mathbf{x} - \int_{\tilde{R}\cap A} L(\theta_0, \mathbf{x})\, \mathrm{d}\mathbf{x}\right) \\
&= K\left(\int_{R\cap\tilde{A}} L(\theta_0, \mathbf{x})\, \mathrm{d}\mathbf{x} + \int_{R\cap\tilde{R}} L(\theta_0, \mathbf{x})\, \mathrm{d}\mathbf{x} - \int_{R\cap\tilde{R}} L(\theta_0, \mathbf{x})\, \mathrm{d}\mathbf{x} - \int_{\tilde{R}\cap A} L(\theta_0, \mathbf{x})\, \mathrm{d}\mathbf{x}\right) \\
&= K\left(\int_{(R\cap\tilde{A})\cup(R\cap\tilde{R})} L(\theta_0, \mathbf{x})\, \mathrm{d}\mathbf{x} - \int_{(\tilde{R}\cap R)\cup(\tilde{R}\cap A)} L(\theta_0, \mathbf{x})\, \mathrm{d}\mathbf{x}\right) \\
&= K\left(\int_{R\cap(\tilde{A}\cup\tilde{R})} L(\theta_0, \mathbf{x})\, \mathrm{d}\mathbf{x} - \int_{\tilde{R}\cap(R\cup A)} L(\theta_0, \mathbf{x})\, \mathrm{d}\mathbf{x}\right) \\
&= K\left(\int_{R} L(\theta_0, \mathbf{x})\, \mathrm{d}\mathbf{x} - \int_{\tilde{R}} L(\theta_0, \mathbf{x})\, \mathrm{d}\mathbf{x}\right) \\
&= K\left(\mathrm{P}(\mathbf{X} \in R \mid \theta = \theta_0) - \mathrm{P}(\mathbf{X} \in \tilde{R} \mid \theta = \theta_0)\right) = K\,(\alpha - \tilde{\alpha}) = 0
\end{aligned}
$$

Logo $\tilde{\beta} \geq \beta$ e o teste $\mathit{\Gamma}$ tem maior poder que o teste $\tilde{\mathit{\Gamma}}$, qualquer que seja o teste $\tilde{\mathit{\Gamma}}$ com probabilidade do erro tipo I igual a α. $\square$

Na prática a constante K é escolhida de forma a fazer com que o teste $\mathit{\Gamma}$ tenha a probabilidade de erro do tipo I igual ao valor escolhido para α. Para determinar K precisamos então conhecer α e n. O teste encontrado será aquele com menor probabilidade do erro tipo II, ou seja, maior poder, entre todos os testes com probabilidade do erro tipo I igual a α. Este é o teste mais poderoso para as hipóteses nula e alternativa simples.

Exemplo 6.9. *Seja $X \sim N(\mu, 1)$. Considere as hipóteses $H_0 : \mu = 2$ contra $H_1 : \mu = 3$.*

(a) *Encontre o teste mais poderoso para as hipóteses acima com $\alpha = \mathrm{P}(\textit{erro tipo I}) = 0,1$ supondo uma amostra de tamanho $n = 9$.*

(b) *Em seguida encontre o valor de $\beta = \mathrm{P}(\textit{erro tipo II})$ e o poder do teste.*

(c) *Supondo a amostra de X abaixo, qual decisão tomar em relação às hipóteses enunciadas?*

$$3,20 \quad 3,56 \quad 3,05 \quad 1,05 \quad 1,53 \quad 1,50 \quad 3,14 \quad 3,56 \quad 0,27$$

Solução:

(a) Segundo o Lema Neuman-Pearson (Teorema 6.1) o teste mais poderoso é aquele com região crítica dada por $R = \{\mathbf{x} \in S \mid L(\theta_1, \mathbf{x}) \geq KL(\theta_0, \mathbf{x})\}$. Precisamos primeiro da função de verossimilhança. Como $X \sim N(\mu, 1)$,

$$L(\mu, \mathbf{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i - \mu)^2} = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^2}$$

e dessa forma a região crítica é definida por:

$$
\begin{aligned}
R &= \{\mathbf{x} \in S \mid L(\mu_1, \mathbf{x}) \geq KL(\mu_0, \mathbf{x})\} \\
&= \left\{\mathbf{x} \in S \mid \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n (x_i-3)^2} \geq K \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n (x_i-2)^2}\right\} \\
&= \left\{\mathbf{x} \in S \mid e^{-\frac{1}{2}\sum_{i=1}^n (x_i-3)^2} \geq K e^{-\frac{1}{2}\sum_{i=1}^n (x_i-2)^2}\right\} \\
&= \left\{\mathbf{x} \in S \mid \frac{e^{-\frac{1}{2}\sum_{i=1}^n (x_i-3)^2}}{e^{-\frac{1}{2}\sum_{i=1}^n (x_i-2)^2}} \geq K\right\} \\
&= \left\{\mathbf{x} \in S \mid e^{-\frac{1}{2}\left(\sum_{i=1}^n (x_i-3)^2 - \sum_{i=1}^n (x_i-2)^2\right)} \geq K\right\} \\
&= \left\{\mathbf{x} \in S \mid -\frac{1}{2}\left(\sum_{i=1}^n (x_i-3)^2 - \sum_{i=1}^n (x_i-2)^2\right) \geq \ln(K)\right\} \\
&= \left\{\mathbf{x} \in S \mid \sum_{i=1}^n (x_i-3)^2 - \sum_{i=1}^n (x_i-2)^2 \leq -2\ln(K)\right\} \\
&= \left\{\mathbf{x} \in S \mid \sum_{i=1}^n (x_i^2 - 6x_i + 9) - \sum_{i=1}^n (x_i^2 - 4x_i + 4) \leq -2\ln(K)\right\} \\
&= \left\{\mathbf{x} \in S \mid -2\left(\sum_{i=1}^n x_i\right) + 5n \leq -2\ln(K)\right\} \\
&= \left\{\mathbf{x} \in S \mid -2\left(\sum_{i=1}^n x_i\right) \leq -2\ln(K) - 5n\right\} \\
&= \left\{\mathbf{x} \in S \mid \sum_{i=1}^n x_i \geq \ln(K) + \frac{5n}{2}\right\}.
\end{aligned}
$$

Na prática não importa muito o valor da constante K e sim a regra de decisão definida pela Região Crítica. Veja que podemos escrever

$$R = \left\{\mathbf{x} \in S \mid \sum_{i=1}^n x_i \geq c\right\} \quad \text{ou então} \quad R = \{\mathbf{x} \in S \mid \bar{x} \geq c\},$$

que é exatamente a região crítica apresentada no Exemplo 6.6. Vamos trabalhar com a região crítica escrita em termos da média amostral: $R = \{\mathbf{x} \in S \mid \bar{x} \geq c\}$.

Agora é preciso encontrar o valor de c considerando $\alpha = 0,1$ e $n = 9$, assim teremos a regra de decisão de forma clara. Lembre-se que $\alpha = \mathrm{P}(\mathbf{X} \in R \mid H_0 \text{ verdadeira})$. Queremos o valor de c tal que

$$\mathrm{P}(\bar{X} \geq c \mid \mu = 2) = 0,1 \qquad \Rightarrow \qquad \mathrm{P}(\bar{X} \leq c \mid \mu = 2) = 0,9.$$

Supondo $\mu = 2$ temos $\bar{X} \sim N(2, 1/9)$. Então, queremos o valor de c tal que

$$\mathrm{P}\left(\frac{\bar{X}-2}{1/3} \leq \frac{c-2}{1/3} \mid \mu = 2\right) = 0,9 \;\Rightarrow\; \Phi(3(c-2)) = 0,90 \;\Rightarrow\; 3(c-2) = 1,28,$$

ou seja, $c = \frac{1,28}{3} + 2 = 2,426667$. Logo o teste mais poderoso será aquele com região crítica definida por:

$$R = \{\mathbf{x} \in S \mid \bar{x} \geq 2,426667\}.$$

(b) Para esse teste temos

$$\begin{aligned}
\beta &= \mathrm{P}\left(\mathbf{X} \in A \mid H_1 \text{ verdadeira}\right) \\
&= \mathrm{P}\left(\bar{X} < 2,426667 \mid \mu = 3\right) \\
&= \mathrm{P}\left(\frac{\bar{X}-3}{1/3} < \frac{2,426667-3}{1/3} \mid \mu = 3\right) \\
&= \mathrm{P}\left(\frac{\bar{X}-3}{1/3} < -1,72 \mid \mu = 3\right) \\
&= 1 - \Phi\left(1,72\right) = 1 - 0,9573 = 0,0427.
\end{aligned}$$

E o poder do teste é $0,9573$.

(c) Para terminar o exemplo vejamos a decisão a ser tomada para a amostra observada. Veja que a amostra tem tamanho $n = 9$, então será usado exatamente o teste encontrado no item (a) definido pela região crítica

$$R = \left\{\mathbf{x} \in S \mid \bar{x} \geq 2,426667\right\}.$$

Veja que para a amostra em questão $\bar{x} = 2,317778$, então essa observação não pertence a região crítica e com isso a decisão é de não rejeitar a hipótese nula $H_0 : \mu = 2$.

$\|$

Vejamos dois outros exemplos com outras população.

Exemplo 6.10. *Seja $X \sim Exp(\lambda)$. Considere as hipóteses $H_0 : \lambda = 2$ contra $H_1 : \lambda = 4$.*

(a) Encontre o teste mais poderoso para as hipóteses acima com $\alpha = \mathrm{P}(\text{erro tipo I}) = 0,05$ supondo uma amostra de tamanho $n = 15$.

(b) Em seguida encontre o valor de $\beta = \mathrm{P}(\text{erro tipo II})$ e o poder do teste.

(c) Supondo a amostra de X abaixo, qual decisão tomar em relação às hipóteses enunciadas?

$$\begin{array}{cccccccc}
0,175 & 0,471 & 0,029 & 0,188 & 0,121 & 0,162 & 0,061 & 0,221 \\
0,082 & 0,719 & 1,151 & 0,215 & 0,332 & 0,516 & 0,175 &
\end{array}$$

Solução:

(a) Segundo o Lema Neuman-Pearson (Teorema 6.1) o teste mais poderoso é aquele com região crítica dada por $R = \{\mathbf{x} \in S \mid L(\theta_1, \mathbf{x}) \geq KL(\theta_0, \mathbf{x})\}$.

Precisamos primeiro da função de verossimilhança. Como $X \sim Exp(\lambda)$,

$$L(\lambda, \mathbf{x}) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

e dessa forma a região crítica é definida por:

$$\begin{aligned}
R &= \{\mathbf{x} \in S \mid L(\lambda_1, \mathbf{x}) \geq KL(\lambda_0, \mathbf{x})\} \\
&= \left\{\mathbf{x} \in S \mid 4^n e^{-4\sum_{i=1}^n x_i} \geq K 2^n e^{-2\sum_{i=1}^n x_i}\right\} \\
&= \left\{\mathbf{x} \in S \mid \frac{e^{-4\sum_{i=1}^n x_i}}{e^{-2\sum_{i=1}^n x_i}} \geq \frac{2^n}{4^n} K\right\} \\
&= \left\{\mathbf{x} \in S \mid -4\sum_{i=1}^n x_i + 2\sum_{i=1}^n x_i \geq \ln\left(2^{-n}K\right)\right\} \\
&= \left\{\mathbf{x} \in S \mid -2\sum_{i=1}^n x_i \geq \ln\left(2^{-n}K\right)\right\} \\
&= \left\{\mathbf{x} \in S \mid \sum_{i=1}^n x_i \leq -\frac{1}{2}\ln\left(2^{-n}K\right)\right\}.
\end{aligned}$$

Novamente o valor de K é o que menos importa. Queremos apenas uma expressão clara para a regra de decisão, que pode ser escrita como

$$R = \left\{\mathbf{x} \in S \mid \sum_{i=1}^n x_i \leq c\right\}$$

e a constante c depende de $\alpha = \mathrm{P}(\text{erro tipo I})$ e também de n. Vamos encontrar o valor de c para $n = 15$ e $\alpha = 0,05$, como sugere o enunciado. Queremos c tal que $\alpha = \mathrm{P}(\mathbf{X} \in R \mid H_0 \text{ verdadeira}) = 0,05$, ou seja, queremos c tal que

$$\mathrm{P}\left(\sum_{i=1}^{15} X_i \leq c \mid \lambda = 2\right) = 0,05.$$

Supondo $\lambda = 2$, a estatística $\sum_{i=1}^{15} X_i \sim Gama(15, 2)$. Assim já é possível encontrar o valor de c, mas vai ser mais fácil encontrar c se transformarmos essa estatística para uma qui-quadrado. Supondo $\lambda = 2$, $4\sum_{i=1}^{15} X_i \sim Gama(15, 1/2)$, logo, $4\sum_{i=1}^{15} X_i \sim \chi^2_{30}$. Assim, queremos o valor de c tal que

$$\mathrm{P}\left(\sum_{i=1}^{15} X_i \leq c \mid \lambda = 2\right) = 0,05 \quad \Rightarrow \quad \mathrm{P}\left(4\sum_{i=1}^{15} X_i \leq 4c \mid \lambda = 2\right) = 0,05$$

sendo $4\sum_{i=1}^{15} X_i \sim \chi^2_{30}$, consultando a tabela da qui-quadrado com 30 graus de liberdade chegamos em

$$4c = 18,4927 \qquad \Rightarrow \qquad c = 4,623175.$$

Assim, a região crítica do teste é:

$$R = \left\{\mathbf{x} \in S \mid \sum_{i=1}^n x_i \leq 4,623175\right\}.$$

(b) Para esse teste temos

$$\begin{aligned}
\beta &= \mathrm{P}(\mathbf{X} \in A \mid H_1 \text{ verdadeira}) \\
&= \mathrm{P}\left(\sum_{i=1}^{15} X_i > 4,623175 \mid \lambda = 4\right) \\
&= 1 - \mathrm{P}\left(\sum_{i=1}^{15} X_i \leq 4,623175 \mid \lambda = 4\right).
\end{aligned}$$

Supondo $\lambda = 4$ temos $\sum_{i=1}^{15} X_i \sim Gama(15,4)$ e $8\sum_{i=1}^{15} X_i \sim \chi^2_{30}$. Assim,

$$\begin{aligned}\beta &= 1 - \mathrm{P}\left(\sum_{i=1}^{15} X_i \leq 4,623175 \mid \lambda = 4\right)\\ &= 1 - \mathrm{P}\left(8\sum_{i=1}^{15} X_i \leq 8 \times 4,623175 \mid \lambda = 4\right)\\ &= 1 - \mathrm{P}\left(8\sum_{i=1}^{15} X_i \leq 36,9854 \mid \lambda = 4\right)\end{aligned}$$

sendo $8\sum_{i=1}^{15} X_i \sim \chi^2_{30}$. Para encontrar o valor de β foi usado o Programa R. O comando `1 - pchisq(36.9854,df=30)` retornou o valor de $\beta = 0,1775702$. Logo o poder do teste é $0,8224298$.

(c) Para terminar o exemplo vejamos a decisão a ser tomada para a amostra observada. Veja que a amostra tem tamanho $n = 15$, então será usado exatamente o teste encontrado no item (a) definido pela região crítica

$$R = \left\{\mathbf{x} \in S \mid \sum_{i=1}^{n} x_i \leq 4,623175\right\}.$$

Veja que para a amostra em questão $\sum_{i=1}^{15} x_i = 4,618$, então essa observação pertence à região crítica e com isso a decisão é de rejeitar a hipótese nula $H_0 : \lambda = 2$.

$\|$

Exercícios da Seção 6.2

6.2.1 Sabe-se que uma variável aleatória X tem distribuição Normal com média $\mu = 5$, mas variância σ^2 desconhecida.

(a) ([Larson, 1982] - Capítulo 8) Qual seria a região crítica do teste mais poderoso para as hipóteses $H_0 : \sigma^2 = 10$ contra $H_1 : \sigma^2 = 20$, a partir de uma amostra aleatória de X de tamanho n? Como ficaria a sua resposta se a hipótese alternativa fosse trocada para $H_1 : \sigma^2 = 5$?

(b) Apresente a região crítica dos dois testes encontrados no item (a) considerando $n = 15$ e $\alpha = 0,05$.

(c) Calcule o poder dos dois testes encontrados no item (b).

6.2.2 Assuma que temos uma amostra aleatória de tamanho n de uma v.a. contínua X e gostaríamos de testar a hipótese simples que a densidade de X é $f_X(x) = 2x$, $0 < x < 1$, contra a hipótese alternativa, também simples, de que a densidade de X é $f_X(x) = 1$, $0 < x < 1$.

(a) ([Larson, 1982] - Capítulo 8) Encontre o teste mais poderoso para tais hipóteses. Dica: Veja que sob H_0 temos $X \sim Beta(2,1)$ e sob H_1 temos $X \sim Beta(1,1)$.

(b) Apresente a região crítica considerando $\alpha = 0,05$ e $n = 10$. Dica: Se $X \sim Beta(\alpha,1)$ então $Y = -\ln(X) \sim Exp(\alpha)$.

(c) Qual decisão tomar se for recolhida a amostra de X a seguir?

$0,0472 \quad 0,821 \quad 0,241 \quad 0,543 \quad 0,00717 \quad 0,207 \quad 0,0301 \quad 0,191 \quad 0,319 \quad 0,779$

6.3 Função Poder e Nível de Significância

Já vimos que se H_0 for simples conseguimos calcular $\alpha = \mathrm{P}(\mathbf{X} \in R|H_0 \text{ verdadeira})$. Se H_1 for simples conseguimos calcular $\beta = \mathrm{P}(\mathbf{X} \in A|H_1 \text{ verdadeira})$. Porém, se H_0 for uma hipótese composta não teremos como calcular α e se H_1 for uma hipótese composta não teremos como calcular β. Essa seção vai tratar de conceitos mais gerais que as probabilidades dos erros, para que possamos avaliar testes com hipóteses compostas.

Definição 6.4. *A função poder de um teste Γ com região crítica R para as hipóteses H_0 contra H_1 é a função $\Pi : \Psi \mapsto [0,1]$ definida por:*

$$\Pi(\theta) = \mathrm{P}(\mathbf{X} \in R \mid \theta).$$

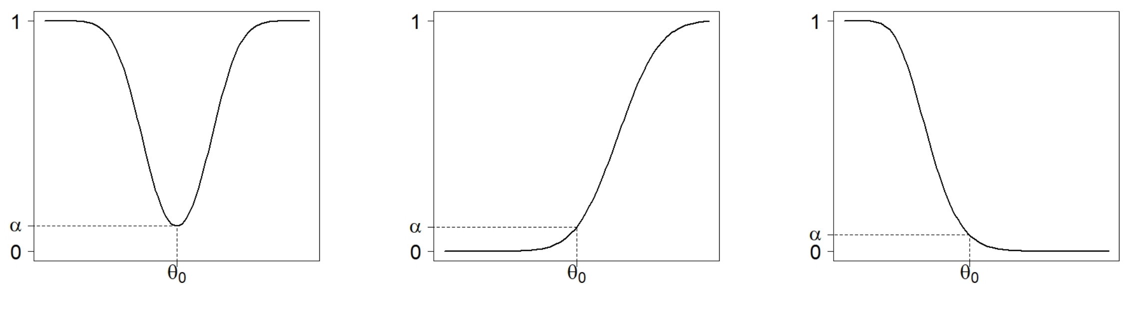
Ou seja, a função poder é definida pela probabilidade de rejeitar H_0 para cada possível valor de θ no espaço paramétrico.

Seja Ψ o espaço paramétrico do parâmetro a ser testado. Vamos chamar de Ψ_0 os valores de Ψ para os quais H_0 é verdadeira e de Ψ_1 os valores de Ψ para os quais H_1 é verdadeira. Esperamos que a função poder de um teste Γ, que retorna a probabilidade com que esse teste rejeita H_0 para cada valor de $\theta \in \Psi$, assuma valores "grandes" em Ψ_1 e "pequeno" em Ψ_0. Ou seja, seja grande a probabilidade de rejeitar H_0 caso $\theta \in \Psi_1$, isto é, caso H_1 seja verdadeira, e seja pequena a probabilidade de rejeitar H_0 caso $\theta \in \Psi_0$, isto é, caso H_0 seja verdadeira.

Por exemplo, supondo $\Psi = \mathbb{R}$,

Hipóteses	Ψ_0	Ψ_1
$H_0 : \theta = \theta_0 \times H_1 : \theta \neq \theta_0$	$\{\theta_0\}$	$(-\infty, \theta_0) \cup (\theta_0, \infty)$
$H_0 : \theta \leq \theta_0 \times H_1 : \theta > \theta_0$	$(-\infty, \theta_0]$	(θ_0, ∞)
$H_0 : \theta \geq \theta_0 \times H_1 : \theta < \theta_0$	$[\theta_0, \infty)$	$(-\infty, \theta_0)$

Se as hipóteses a serem testadas são do tipo $H_0 : \theta = \theta_0$ contra $H_1 : \theta \neq \theta_0$ seria interessante que a função poder fosse parecida com a função cujo gráfico está apresentado na Figura 6.2(a). Já se as hipóteses a forem do tipo $H_0 : \theta \leq \theta_0$ contra $H_1 : \theta > \theta_0$ seria interessante que a função poder fosse parecida com a função cujo gráfico está apresentado na Figura 6.2(b). Por fim, se as hipóteses a forem do tipo $H_0 : \theta \geq \theta_0$ contra $H_1 : \theta < \theta_0$ seria interessante que a função poder fosse parecida com a função cujo gráfico está apresentado na Figura 6.2(c).



(a) $H_0 : \theta = \theta_0 \times H_1 : \theta \neq \theta_0$ (b) $H_0 : \theta \leq \theta_0 \times H_1 : \theta > \theta_0$ (c) $H_0 : \theta \geq \theta_0 \times H_1 : \theta < \theta_0$

Figura 6.2: Gráficos de Algumas Funções Poder

Definição 6.5. *O nível de significância (ou tamanho) de um teste Γ para as hipóteses H_0 contra H_1 é definido por:*

$$\alpha = \sup_{\theta \in \Psi_0} \Pi(\theta) = \sup_{\theta \in \Psi_0} \mathrm{P}(\mathbf{X} \in R \mid \theta).$$

Interpretamos o nível de significância como a maior probabilidade de rejeitar H_0 sendo H_0 verdadeira. Veja que se H_0 for uma hipótese simples, $H_0 : \theta = \theta_0$, então $\Psi_0 = \{\theta_0\}$ e $\alpha = \Pi(\theta_0) = \mathrm{P}(\text{erro tipo 1})$.

Na Figura 6.2 também estão marcados os valores dos níveis de significância em cada um dos gráficos apresentados .

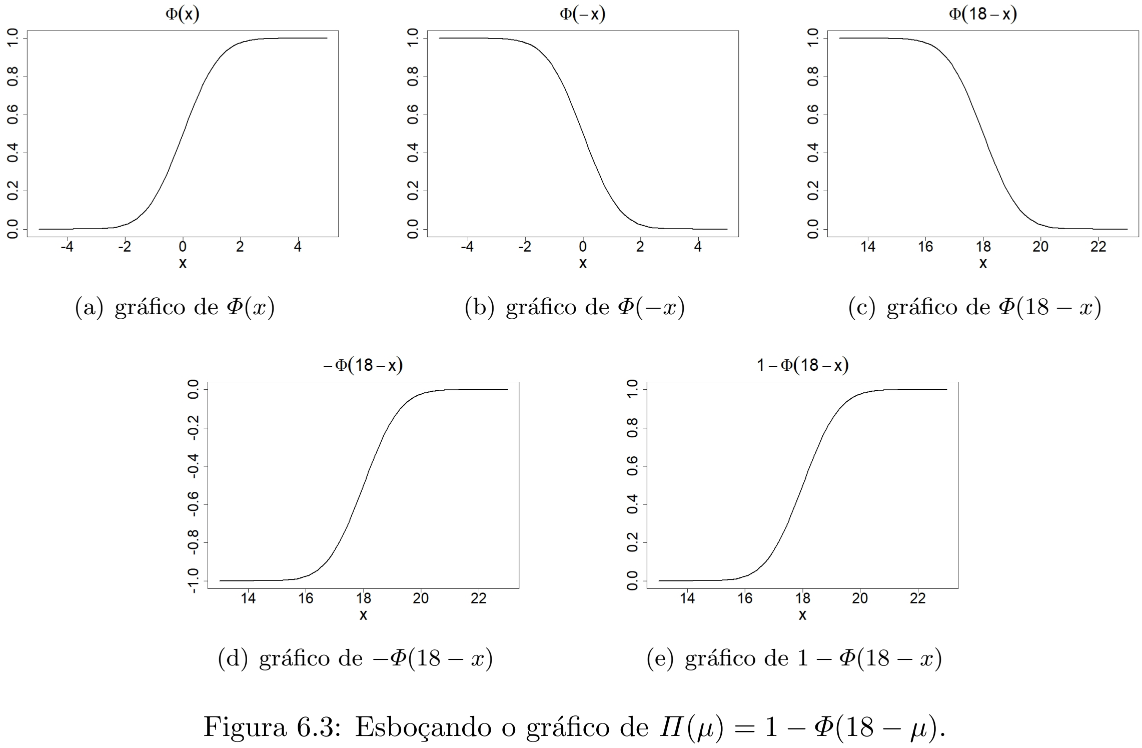
Exemplo 6.11. *Considere a população $X \sim N(\mu, 25)$ e as hipóteses $H_0 : \mu \leq 17$ contra $H_1 : \mu > 17$. Para uma amostra de tamanho $n = 25$ foi criado um teste estatístico definido pela região crítica: $R = \{\mathbf{x} \in S \mid \bar{x} \geq 18\}$. (a) Defina Ψ_0 e Ψ_1. (b) Encontre a função poder e faça um esboço do gráfico dela. (c) Calcule o nível de significância do teste.*

Solução:

(a) Para o parâmetro μ, $\Psi = \mathbb{R}$. Como $H_0 : \mu \leq 17$, $\Psi_0 = (-\infty, 17]$ e como $H_1 : \mu > 17$, $\Psi_1 = (17, \infty)$.

(b) $\Pi(\mu) = \mathrm{P}(\mathbf{X} \in R \mid \mu) = \mathrm{P}(\bar{X} \geq 18 \mid \mu) = 1 - \mathrm{P}(\bar{X} \leq 18 \mid \mu) = 1 - \mathrm{P}\left(\frac{\bar{X} - \mu}{1} \leq \frac{18 - \mu}{1} \mid \mu\right) = 1 - \Phi(18 - \mu)$.

Para fazer um esboço do gráfico de Π vamos partir do gráfico da função Φ que a gente conhece e usar propriedades de funções, como espelhamento e deslocamento no eixo.



(a) gráfico de $\Phi(x)$ (b) gráfico de $\Phi(-x)$ (c) gráfico de $\Phi(18 - x)$

(d) gráfico de $-\Phi(18 - x)$ (e) gráfico de $1 - \Phi(18 - x)$

Figura 6.3: Esboçando o gráfico de $\Pi(\mu) = 1 - \Phi(18 - \mu)$.

(c) O nível de significância é o maior valor da função poder para valores de Φ_0. Veja que $\Psi_0 = (-\infty, 17]$ e Π é uma função crescente (Figura 6.3(e)). Então o maior valor de $\Pi(\mu)$ para $\mu \in \Psi_0$ é: $\Pi(17) = 1 - \Phi(18 - 17) = 1 - \Phi(1) = 1 - 0,8413 = 0,1587$. Veja na Figura 6.4 a representação gráfica do nível de significância junto com a função poder.

$\|$

Exemplo 6.12. *Considere a população $X \sim Exp(\lambda)$ e as hipóteses $H_0 : \lambda \geq 2$ contra $H_1 : \lambda < 2$. Para uma amostra de tamanho $n = 20$ foi criado um teste estatístico definido pela região*

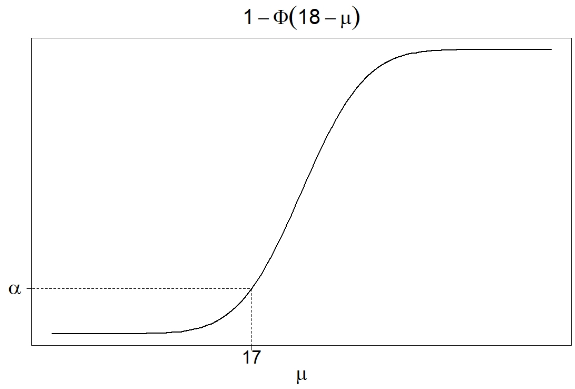


Figura 6.4: Função poder do Exemplo 6.11

crítica: $R = \{\mathbf{x} \in S \mid \bar{x} \geq 0,7\}$. *(a) Defina* Ψ_0 *e* Ψ_1. *(b) Encontre a função poder e faça um esboço do gráfico dela. (c) Calcule o nível de significância do teste.*

Solução:

(a) Para o parâmetro λ, $\Psi = (0, \infty)$. Como $H_0 : \lambda \geq 2$, $\Psi_0 = [2, \infty)$ e como $H_1 : \lambda < 2$, $\Psi_1 = (0, 2)$.

(b) $\Pi(\lambda) = \mathrm{P}(\mathbf{X} \in R \mid \lambda) = \mathrm{P}(\bar{X} \geq 0,7 \mid \lambda) = 1 - \mathrm{P}(\bar{X} \leq 0,7 \mid \lambda)$.

Veja que $\bar{X} \sim Gama(20, 20\lambda)$, então $40\lambda\bar{X} \sim \chi^2_{40}$. Continuando,
$\Pi(\lambda) = 1 - \mathrm{P}(40\lambda\bar{X} \leq 28\lambda \mid \lambda) = 1 - F_W(28\lambda)$, sendo F_W a função de distribuição acumulada de $W \sim \chi^2_{40}$.

Mesmo não sabendo exatamente como é a curva de F_W é possível fazer seu esboço, pois se trata de uma função de distribuição acumulada de uma variável aleatória contínua e por isso sabemos que é uma função crescente, $\lim_{x\to\infty} F_W(x) = 1$ e $\lim_{x\to-\infty} F_W(x) = 0$. Isso é suficiente para fazer um bom esboço do gráficos de Π.

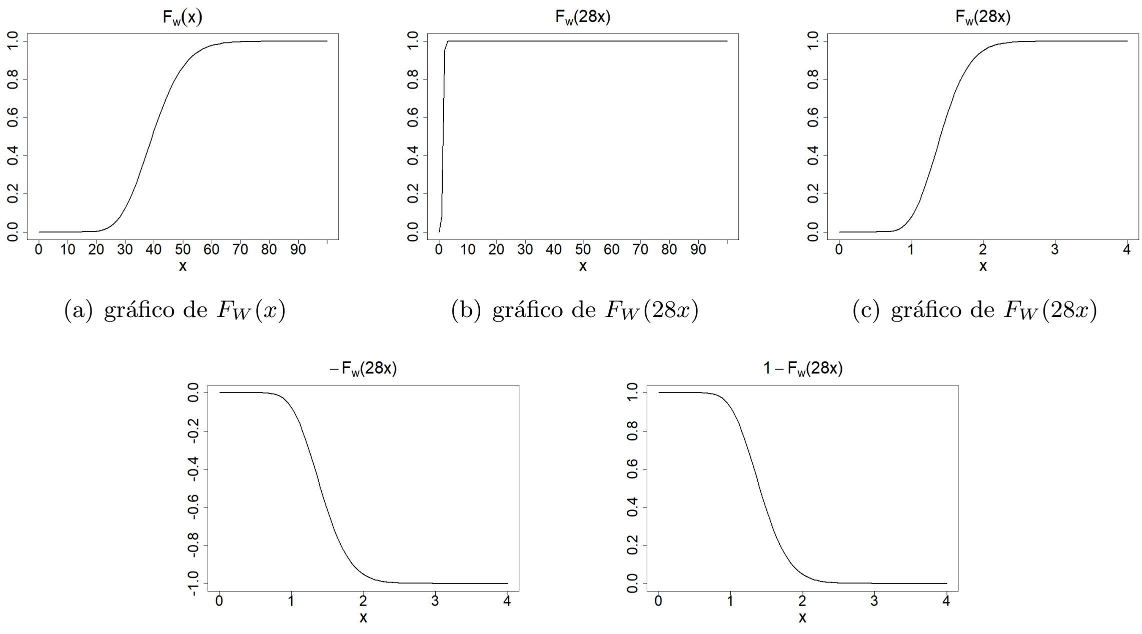
(c) O nível de significância é o maior valor da função poder para valores de Φ_0. Veja que $\Psi_0 = [2, \infty)$ e Π é uma função decrescente (Figura 6.5.6.5(e)). Então o maior valor de $\Pi(\lambda)$ para $\lambda \in \Psi_0$ é: $\Pi(2) = 1 - F_W(28 \times 2) = 1 - F_W(56) = 1 - 0,9521929 = 0,04780711$. Para encontrar o valor de $F_W(56)$ foi usado o comando `pchisq(56,df=40)` do Programa R. Veja na Figura 6.6 a representação gráfica do nível de significância junto com a função poder.

$\|$

Exemplo 6.13. *Considere a população* $X \sim Beta(\theta, 1)$ *e as hipóteses* $H_0 : \theta = 3$ *contra* $H_1 : \theta \neq 3$. *Para uma amostra de tamanho* $n = 35$ *foi criado um teste estatístico definido pela região crítica:* $R = \left\{ \mathbf{x} \in S \mid -\frac{\sum_{i=1}^{35} \ln(x_i)}{35} < 0,25 \text{ ou } -\frac{\sum_{i=1}^{35} \ln(x_i)}{35} > 0,44 \right\}$. *(a) Defina* Ψ_0 *e* Ψ_1. *(b) Encontre a função poder e use o Programa (R para fazer um esboço do gráfico dela. (c) Calcule o nível de significância do teste.*

Solução: Antes de começar veja que se $X \sim Beta(\theta, 1)$, $\theta > 0$ e a função densidade de X é dada por $f_X(x) = \theta x^{\theta-1}$, $0 < x < 1$. Já trabalhamos com essa distribuição em alguns exercícios dos capítulos anteriores.

(a) Para o parâmetro θ temos $\Psi = (0, \infty)$. Como $H_0 : \theta = 3$, $\Psi_0 = \{3\}$ e como $H_1 : \theta \neq 3$, $\Psi_1 = (0, 3) \cup (3, \infty)$.



(a) gráfico de $F_W(x)$ (b) gráfico de $F_W(28x)$ (c) gráfico de $F_W(28x)$

(d) gráfico de $-F_W(28x)$ (e) gráfico de $1 - F_W(28x)$

Figura 6.5: Esboçando o gráfico de $\mathit{\Pi}(\mu) = 1 - \Phi(18 - \mu)$.

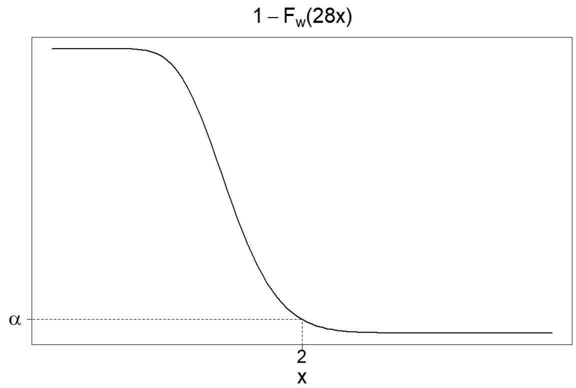


Figura 6.6: Função poder do Exemplo 6.12

(b) $$\mathit{\Pi}(\lambda) = \mathrm{P}(\mathbf{X} \in R \mid \theta) = \mathrm{P}\left(-\frac{\sum_{i=1}^{35} \ln(X_i)}{35} < 0,25 \text{ ou } -\frac{\sum_{i=1}^{35} \ln(X_i)}{35} > 0,44 \mid \theta\right).$$

Veja no Exercício 2.1.11 que $-\sum_{i=1}^{n} \ln(X_i) \sim Gama(n, \theta)$, então para calcular a probabilidade acima vamos trabalhar com essa estatística, já que conhecemos a sua distribuição. Ou ainda melhor, vamos modificá-la para uma qui-quadrado, ou seja, $-2\theta\sum_{i=1}^{35}\ln(X_i) \sim \chi^2_{70}$.

$$\begin{aligned}
\mathit{\Pi}(\theta) &= \mathrm{P}\left(-\frac{\sum_{i=1}^{35}\ln(X_i)}{35} < 0,25 \text{ ou } -\frac{\sum_{i=1}^{35}\ln(X_i)}{35} > 0,44 \mid \theta\right)\\
&= \mathrm{P}\left(-\sum_{i=1}^{35}\ln(X_i) < 8,75 \text{ ou } -\sum_{i=1}^{35}\ln(X_i) > 15,4 \mid \theta\right)\\
&= \mathrm{P}\left(-2\theta\sum_{i=1}^{35}\ln(X_i) < 17,5\theta \text{ ou } -2\theta\sum_{i=1}^{35}\ln(X_i) > 30,8\theta \mid \theta\right)\\
&= \mathrm{P}\left(-2\theta\sum_{i=1}^{35}\ln(X_i) < 17,5\theta \mid \theta\right) + \mathrm{P}\left(-2\theta\sum_{i=1}^{35}\ln(X_i) > 30,8\theta \mid \theta\right)\\
&= \mathrm{P}\left(-2\theta\sum_{i=1}^{35}\ln(X_i) < 17,5\theta \mid \theta\right) + 1 - \mathrm{P}\left(-2\theta\sum_{i=1}^{35}\ln(X_i) \leq 30,8\theta \mid \theta\right)
\end{aligned}$$

Logo, $\mathit{\Pi}(\theta) = F_W(17,5\theta) + 1 - F_W(30,8\theta)$, sendo $W = -2\theta\sum_{i=1}^{35}\ln(X_i) \sim \chi^2_{70}$.

Esboçar o gráfico da função $\mathit{\Pi}$ já é mais complicado. Mas usando o Programa R podemos fazer isso sem dificuldades. Rodando o código `curve(pchisq(17.5*x,df=70) + 1 - pchisq(30.8*x,df=70),lwd=2,xlim=c(0,6),cex.axis=2.2)` obtemos a Figura 6.7.

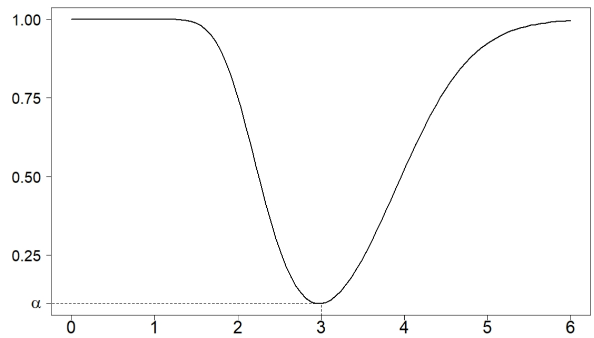


Figura 6.7: Função poder do Exemplo 6.13

(c) O nível de significância é o maior valor da função poder para valores de $\theta \in \mathit{\Phi}_0 = \{3\}$. Então, $\alpha = \mathit{\Pi}(3) = F_W(17,5{\times}3)+1-F_W(30,8{\times}3) = 0,0963917$, sendo $W \sim \chi^2_{70}$. Para encontrar o valor de $\mathit{\Pi}(3)$ foi usado o comando `pchisq(17.5*3,df=70) + 1 - pchisq(30.8*3,df=70)` do Programa R. Veja na Figura 6.7 a representação gráfica do nível de significância.

||

Exercícios da Seção 6.3

6.3.1 Seja $X \sim N(mu = 5, \sigma^2)$. Sejam as hipóteses $H_0 : \sigma^2 \leq 15$ contra $H_1 : \sigma^2 > 15$. Para uma amostra de tamanho $n = 38$, considere o teste definido pela região crítica

$$R = \left\{\mathbf{x} \mid \sum_{i=1}^{38}(x_i - 5)^2 \geq 800\right\}.$$

(a) Defina $\mathit{\Psi}_0$ e $\mathit{\Psi}_1$.

(b) Encontre a função poder e esboce seu gráfico no Programa R.

(c) Calcule o nível de significância do teste.

6.3.2 Seja $X \sim Gama(2, \lambda)$. Sejam as hipóteses $H_0 : \lambda = 1$ contra $H_1 : \lambda \neq 1$. Para uma amostra de tamanho $n = 15$, considere o teste definido pela região crítica

$$R = \{\mathbf{x} \mid \bar{x} < 1,3 \text{ ou } \bar{x} > 2,8\}.$$

(a) Defina Ψ_0 e Ψ_1.

(b) Encontre a função poder e esboce seu gráfico no Programa R.

(c) Calcule o nível de significância do teste.

6.4 O Teste Uniformemente Mais Poderoso

Os conceitos de função poder e nível de significância viabilizam a seguinte definição.

Definição 6.6. *Um teste Γ com função poder Π é o teste Uniformemente Mais Poderoso (UMP) com nível de significância α para as hipóteses $H_0 : \theta \in \Psi_0$ contra $H_1 : \theta \in \Psi_1$ se:*

(i) O teste Γ tem nível de significância α.

(ii) Dado qualquer outro teste $\tilde{\Gamma}$ com nível de significância $\tilde{\alpha} = \alpha$ e função poder $\tilde{\Pi}$ é verdade que $\Pi(\theta) \geq \tilde{\Pi}(\theta),\ \forall\ \theta \in \Psi_1$.

Corolário 6.1. *O teste apresentado no Lema de Neuman-Pearson (Teorema 6.1), definido pela região crítica $R = \{\mathbf{x} \in S \mid L(\theta_1, \mathbf{x}) \geq KL(\theta_0, \mathbf{x})\}$ é o teste UMP de tamanho $\alpha = \mathrm{P}(\mathbf{X} \in R \mid \theta = \theta_0)$ para as hipóteses $H_0 : \theta = \theta_0$ contra $H_1 : \theta = \theta_1$.*

A verdade é que nem sempre conseguimos encontrar o teste UMP e as vezes ele nem existe. Por exemplo, não existe teste UMP para hipóteses do tipo $H_0 : \theta = \theta_0$ conta $H_1 : \theta \neq \theta_0$, como mostra [Casella e Berger, 2002] em um exemplo. Mas em certos caso particulares podemos encontrá-los. Esses casos particulares são quando a distribuição da população pertence à família exponencial e as hipóteses forem do tipo: $H_0 : \theta \leq \theta_0$ conta $H_1 : \theta > \theta_0$; $H_0 : \theta \geq \theta_0$ conta $H_1 : \theta < \theta_0$; $H_0 : \theta = \theta_0$ conta $H_1 : \theta > \theta_0$; ou $H_0 : \theta = \theta_0$ conta $H_1 : \theta < \theta_0$. Sob essas condições o teste UMP existe e é definido pelos teoremas a seguir.

Teorema 6.2. *Suponha que a distribuição da população X depende do parâmetro desconhecido θ e pertence à família exponencial, isto é, $f_X(x) = e^{c(\theta)T(x)}e^{d(\theta)}e^{S(x)}, \quad x \in A$.*

(i) Se a função $c(\cdot)$ for estritamente crescente,

- *o teste com região crítica $R = \{\mathbf{x} \in S \mid \sum_{i=1}^{n} T(x_i) > K\}$ é o teste UMP com nível de significância $\alpha = \mathrm{P}(\mathbf{X} \in R \mid \theta = \theta_0)$ para as hipóteses $H_0 : \theta = \theta_0$ contra $H_1 : \theta > \theta_0$;*
- *o teste com região crítica $R = \{\mathbf{x} \in S \mid \sum_{i=1}^{n} T(x_i) < K\}$ é o teste UMP com nível de significância $\alpha = \mathrm{P}(\mathbf{X} \in R \mid \theta = \theta_0)$ para as hipóteses $H_0 : \theta = \theta_0$ contra $H_1 : \theta < \theta_0$.*

(ii) Se a função $c(\cdot)$ for estritamente decrescente,

- *o teste com região crítica $R = \{\mathbf{x} \in S \mid \sum_{i=1}^{n} T(x_i) < K\}$ é o teste UMP com nível de significância $\alpha = \mathrm{P}(\mathbf{X} \in R \mid \theta = \theta_0)$ para as hipóteses $H_0 : \theta = \theta_0$ contra $H_1 : \theta > \theta_0$;*
- *o teste com região crítica $R = \{\mathbf{x} \in S \mid \sum_{i=1}^{n} T(x_i) > K\}$ é o teste UMP com nível de significância $\alpha = \mathrm{P}(\mathbf{X} \in R \mid \theta = \theta_0)$ para as hipóteses $H_0 : \theta = \theta_0$ contra $H_1 : \theta < \theta_0$.*

A demonstração do Teorema 6.2 será omitida nesse texto, mas pode ser encontrada na Seção 8.3 do [Casella e Berger, 2002].

Além do resultado acima, também podemos aproveitar o Teorema 6.3 enunciado a seguir, retirado do [Bolfarine e Sandoval, 2001], para encontrar testes UMP para as hipóteses $H_0 : \theta = \theta_0$ contra $H_1 : \theta > \theta_0$ ou $H_0 : \theta = \theta_0$ contra $H_1 : \theta < \theta_0$.

Teorema 6.3. *Se a distribuição da população X depende do parâmetro desconhecido θ e pertence à família exponencial então:*

(i) O teste UMP de nível de significância α para $H_0 : \theta = \theta_0$ contra $H_1 : \theta > \theta_0$ também é o teste UMP de nível de significância α para $H_0 : \theta \leq \theta_0$ contra $H_1 : \theta > \theta_0$.

(ii) O teste UMP de nível de significância α para $H_0 : \theta = \theta_0$ contra $H_1 : \theta < \theta_0$ também é o teste UMP de nível de significância α para $H_0 : \theta \geq \theta_0$ contra $H_1 : \theta < \theta_0$.

Exemplo 6.14. *Seja $X \sim N(\mu, 1)$ e as hipóteses $H_0 : \mu \leq 0$ contra $H_1 : \mu > 0$.*

(a) Defina a região crítica do teste UMP de nível de significância $\alpha = 0,05$ para uma amostra de tamanho $n = 9$.

(b) Use o Programa R para desenhar o gráfico da função poder desse teste.

(c) Supondo os valores abaixo uma amostra de tamanho $n = 9$ da população X , qual decisão tomar em relação às hipóteses apresentadas?

$$0,886 \quad 0,930 \quad -0,758 \quad -1,177 \quad -0,970 \quad -0,043 \quad 0,585 \quad 0,376 \quad 0,413$$

Solução:

(a) Primeiro veja que a distribuição $N(\mu, 1)$ pertence à família exponencial. A $Im(X) = \mathbb{R}$ não depende de μ e a função densidade de X pode ser escrita por:

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\mu)^2} = e^{-\ln\left(\sqrt{2\pi}\right)} e^{-\frac{1}{2}(x^2-2\mu x+\mu^2)} = e^{-\ln\left(\sqrt{2\pi}\right)} e^{-\frac{x^2}{2}} e^{\mu x} e^{-\frac{\mu^2}{2}} \\ &= e^{-\frac{x^2}{2}-\ln\left(\sqrt{2\pi}\right)} e^{\mu x} e^{-\frac{\mu^2}{2}}. \end{aligned}$$

De acordo com a Definição 2.5, a distribuição pertence à família exponencial sendo $T(x) = x$ e $c(\mu) = \mu$. Então, segundo o Teorema 6.2, como c é uma função crescente, o teste UMP é aquele com região crítica definida por

$$R = \left\{ \mathbf{x} \in S \mid \sum_{i=1}^{9} T(x_i) > K \right\} = \left\{ \mathbf{x} \in S \mid \sum_{i=1}^{9} x_i > K \right\}$$

e a constante K depende de n e α. Vamos encontrar a região crítica para $\alpha = 0,05$ e $n = 9$. O nível de significância desse teste é $\alpha = \mathrm{P}\left(\sum_{i=1}^{9} X_i > K \mid \mu = 0\right)$ e a partir dessa expressão podemos encontrar K.

$$\begin{aligned} \alpha = 0,05 = \mathrm{P}\left(\sum_{i=1}^{9} X_i > K \mid \mu = 0\right) &= \mathrm{P}\left(\bar{X} > \frac{K}{9} \mid \mu = 0\right) = 1 - \mathrm{P}\left(\bar{X} \leq \frac{K}{9} \mid \mu = 0\right) \\ &= 1 - \mathrm{P}\left(\frac{\bar{X}}{1/3} \leq \frac{K}{9/3} \mid \mu = 0\right) = 1 - \Phi\left(\frac{K}{3}\right) \end{aligned}$$

Então,

$$1 - \Phi\left(\frac{K}{3}\right) = 0,05 \;\Rightarrow\; \Phi\left(\frac{K}{3}\right) = 0,95 \;\Rightarrow\; \frac{K}{3} = 1,644854 \;\Rightarrow\; K = 4,934561.$$

Com isso a região crítica do teste é definida por:

$$R = \left\{ \mathbf{x} \in S \mid \sum_{i=1}^{9} x_i > 4,934561 \right\}.$$

(b) Para fazer o gráfico da função poder precisa-se primeiro conhecer a sua expressão. A função poder desse teste é:

$$\begin{aligned} \Pi(\mu) &= \mathrm{P}\left(\sum_{i=1}^{9} X_i > 4,934561 \mid \mu\right) = \mathrm{P}\left(\bar{X} > \frac{4,934561}{9} \mid \mu\right) \\ &= \mathrm{P}\left(\frac{\bar{X}-\mu}{1/3} > 3\,(0,5482846 - \mu) \mid \mu\right) = 1 - \Phi\,(1,644854 - 3\mu)\,. \end{aligned}$$

e o seu gráfico pode ser obtido pelo Programa R pelo comando `curve(1 - pnorm(1.644854 - 3*x),xlim=c(-2,2))`, que produz o gráfico da Figura 6.8.

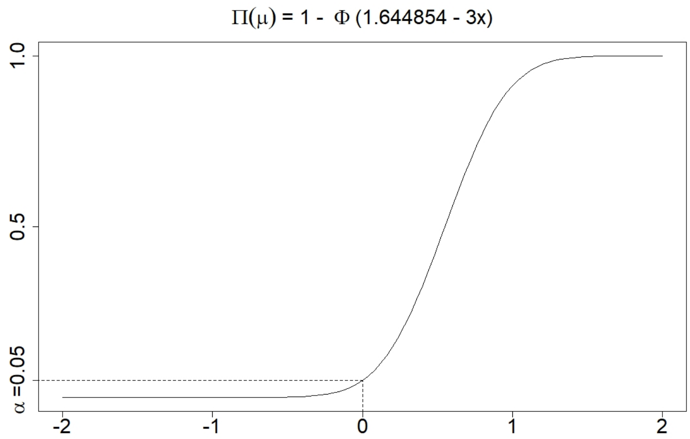


Figura 6.8: Função poder do teste do Exemplo 6.14

(c) Veja que para a amostra observada temos $\sum_{i=1}^{9} x_i = 0,242$. Logo, a amostra não pertence à região crítica e com isso não rejeitamos a hipótese $H_0 : \mu = 0$.

$\|$

Exemplo 6.15. *Seja $X \sim Ber(p)$ e as hipóteses $H_0 : p \geq 0,5$ contra $H_1 : p < 0,5$.*

(a) Defina a região crítica do teste UMP de tamanho $\alpha = 0,1$ para uma amostra com $n = 16$. Em seguida decida entre rejeitar ou não H_0 se forem observados 6 sucessos e 10 fracassos.

(b) Defina a região crítica do teste UMP de tamanho $\alpha = 0,1$ para uma amostra com $n = 160$. Em seguida decida entre rejeitar ou não H_0 se forem observados 60 sucessos e 100 fracassos.

Solução: Primeiro veja que a distribuição $Ber(p)$ pertence a família exponencial. A $Im(X) = \{0,1\}$ não depende de p e a função de probabilidade de X pode ser escrita como

$$\begin{aligned} f(x) &= p^x(1-p)^{1-x} = e^{\ln\left(p^x(1-p)^{1-x}\right)} = e^{x\ln(p)+(1-x)\ln(1-p)} = e^{x\ln(p)+\ln(1-p)-x\ln(1-p)} \\ &= e^{\ln(1-p)}e^{x(\ln(p)-\ln(1-p))} \end{aligned}$$

De acordo com a Definição 2.5, a distribuição pertence à família exponencial sendo $T(x) = x$ e $c(p) = \ln(p) - \ln(1-p)$. Veja que c é uma função crescente, uma vez que a sua derivada é $c'(p) = 1/p + 1/(1-p) > 0$ sempre que $0 < p < 1$.

(a) Segundo o Teorema 6.2, o teste UMP é aquele com região crítica definida por

$$R = \left\{ \mathbf{x} \in S \mid \sum_{i=1}^{16} x_i < K \right\}.$$

e a constante K depende de n e α. Vamos encontrar o K para $\alpha = 0,1$ e $n = 16$. O nível de significância desse teste é $\alpha = \mathrm{P}\left(\sum_{i=1}^{16} X_i < K \mid p = 0,5\right)$ e a partir dessa expressão podemos encontrar K. Veja que $W = \sum_{i=1}^{16} X_i \sim Bin(n = 16, p)$, então

$$\alpha = 0,1 = \mathrm{P}\left(\sum_{i=1}^{16} X_i < K \mid p = 0,5\right) = \mathrm{P}\left(W < K\right), \text{ sendo } W \sim Bin(n = 16, p = 0,5).$$

Veja que podemos nos restringir em encontrar K inteiro, porque se K não for inteiro, a conta acima resulta no mesmo valor quando avaliada no inteiro seguinte a K. O maior problema é que como a distribuição é discreta, nem sempre conseguimos um K que garante o nível significância exato, aí vamos ter que aproximar.

Para encontrar o K tal que $\mathrm{P}\left(W < K\right) \approx 0,1$ sendo $W \sim Bin(n = 16, p = 0,5)$ foi usado o Programa R e calculada a probabilidade acima para diferentes valores de K, começando com $K = 1$ e terminando com o K tal que $\mathrm{P}\left(W < K\right) = \mathrm{P}\left(W \leq K - 1\right) > 0,1$. Assim encontramos $K = 6$ e $\mathrm{P}\left(W < 6\right) = \mathrm{P}\left(W \leq 5\right) = 0,1050568$. Logo a região crítica será:

$$R = \left\{ \mathbf{x} \in S \mid \sum_{i=1}^{16} x_i < 6 \right\}.$$

Com a observação de 6 sucessos e 10 fracassos temos $\sum_{i=1}^{16} x_i = 6$ e por isso não rejeitamos $H_0 : p \geq 0,5$. Isso significa não rejeitar a hipótese de que a probabilidade de sucesso seja maior que a probabilidade de fracasso, mesmo observando mais fracassos na amostra.

Uma alternativa para encontrar o valor de K seria usar a aproximação de uma Binomial pela Normal. Segundo o Teorema Limite Central, se $X_i \sim Ber(p)$,

$$\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \xrightarrow[n \to \infty]{} N(0,1)$$

voltando para a equação que escreve α como função de K,

$$\begin{aligned} \alpha = 0,1 &= \mathrm{P}\left(\sum_{i=1}^{16} X_i < K \mid p = 0,5\right) \\ &= \mathrm{P}\left(\frac{\sum_{i=1}^{16} X_i - 16 \times 0,5}{\sqrt{16 \times 0,5 \times (1 - 0,5)}} < \frac{K - 16 \times 0,5}{\sqrt{16 \times 0,5 \times (1 - 0,5)}} \mid p = 0,5\right) \approx \Phi\left(\frac{K-8}{2}\right) \end{aligned}$$

consultando a Tabela A.1 ou o Programa R, encontramos $\dfrac{K-8}{2} = -1,28$ e com isso, $K = -1,28 \times 2 + 8 = 5,436897$. Veja que esse K resulta na mesma regra de decisão encontrada pela distribuição binomial.

(b) Segundo o Teorema 6.2, o teste UMP é aquele com região crítica definida por

$$R = \left\{ \mathbf{x} \in S \mid \sum_{i=1}^{160} x_i < K \right\}.$$

e a constante K depende de n e α. Vamos encontrar o K para $\alpha = 0,1$ e $n = 160$. O nível de significância desse teste é $\alpha = \mathrm{P}\left(\sum_{i=1}^{160} X_i < K \mid p = 0,5\right)$ e a partir dessa expressão podemos encontrar K. Veja que $W = \sum_{i=1}^{160} X_i \sim Bin(n = 160, p)$, então

$$\alpha = 0,1 = \mathrm{P}\left(\sum_{i=1}^{160} X_i < K \mid p = 0,5\right) = \mathrm{P}\left(W < K\right), \text{ sendo } W \sim Bin(n = 160, p = 0,5)$$

O mesmo processo do item anterior foi feito para a escola de K. Assim encontramos $K = 72$ e $\mathrm{P}\left(W < 72\right) = \mathrm{P}\left(W \leq 71\right) = 0,08938863$. Logo a região crítica do teste será

$$R = \left\{\mathbf{x} \in S \mid \sum_{i=1}^{160} x_i < 72\right\}.$$

Com a observação de 60 sucessos e 100 fracassos temos $\sum_{i=1}^{160} x_i = 60$ e por isso rejeitamos $H_0 : p \geq 0,5$.

Novamente as contas poderiam ser feitas pela aproximação da Binomial pela Normal.

$$\begin{aligned} 0,1 &= \mathrm{P}\left(\sum_{i=1}^{160} X_i < K \mid p = 0,5\right) \\ &= \mathrm{P}\left(\frac{\sum_{i=1}^{160} X_i - 160 \times 0,5}{\sqrt{160 \times 0,5 \times (1 - 0,5)}} < \frac{K - 160 \times 0,5}{\sqrt{160 \times 0,5 \times (1 - 0,5)}} \mid p = 0,5\right) \approx \Phi\left(\frac{K - 80}{\sqrt{40}}\right) \end{aligned}$$

Consultando a Tabela A.1, ou o Programa R, encontramos $\dfrac{K - 80}{\sqrt{40}} = -1,28$ e com isso, $K = -1,28 \times \sqrt{40} + 80 = 71,89476$. Veja que esse K resulta na mesma regra de decisão encontrada pela distribuição binomial.

‖

Exercícios da Seção 6.4

6.4.1 ([Bolfarine e Sandoval, 2001]-Capítulo 6) Seja $X_1, \ldots, X_n$ amostra aleatória de $X \sim N(0, \sigma^2)$. Considere as hipóteses:

$$H_0 : \sigma^2 \leq 9 \qquad \text{contra} \qquad H_1 : \sigma^2 > 9.$$

(a) Defina a região crítica do teste UMP para testar as hipóteses acima considerando $\alpha = 0,05$ e $n = 9$.

(b) Qual decisão tomar se for observada a seguinte amostra de X:
$-0,705 \quad 0,264 \quad -8,971 \quad -2,442 \quad -5,131 \quad 1,315 \quad 7,116 \quad -3,555 \quad 2,090.$

(c) Defina a função poder do teste encontrado no item (a) e faça seu gráfico. Use o Programa R para fazer o gráfico.

6.4.2 ([Bolfarine e Sandoval, 2001]-Capítulo 6) Seja $X_1, \ldots, X_n$ amostra aleatória de $X \sim Poisson(\lambda)$. Considere as hipóteses:

$$H_0 : \lambda = 1 \qquad \text{contra} \qquad H_1 : \lambda > 1.$$

(a) Defina a região crítica do teste UMP para testar as hipóteses acima considerando $\alpha = 0,05$ e $n = 25$. Dica: Use o TCL para aproximar a soma de variáveis aleatórias ou a média amostral por uma Normal.

(b) Qual decisão tomar se for observada a seguinte amostra de X:
0 2 0 1 2 1 1 2 2 1 1 0 2 0 1 1 0 3 2 2 1 2 1 2 0

(c) Defina a função poder do teste encontrado no item (a) e faça seu gráfico. Use a aproximação pela Normal para encontrar a função poder. Use o Programa R para fazer o gráfico.

6.5 Teste da Razão Verossimilhanças Generalizada

A seção anterior, apesar de apresentar o teste UMP, exige muitas restrições para a distribuição da população e para as hipóteses. Nessa seção veremos um teste mais geral, que pode ser utilizado em muitos caso. Ele é o Teste da Razão de Verossimilhanças Generalizada (TRVG).

Definição 6.7. *Considere as hipóteses*

$$H_0 : \theta \in \Psi_0 \qquad \textit{contra} \qquad H_1 : \theta \in \Psi_1,$$

sendo Ψ o espaço paramétrico para o parâmetro θ, $\Psi_0 \cup \Psi_1 = \Psi$, $\Psi_0 \cap \Psi_1 = \emptyset$, $\Psi_0 \neq \emptyset$ e $\Psi_1 \neq \emptyset$.

Chamamos de estatística da razão de verossimilhança generalizada a estatística

$$\lambda(\mathbf{X}) = \frac{\max_{\theta \in \Psi_0} L(\theta; \mathbf{x})}{\max_{\theta \in \Psi} L(\theta; \mathbf{x})}$$

e de Teste da Razão de Verossimilhanças Generalizada (TRVG) aquele cuja região crítica é definida por

$$R = \{\mathbf{x} \in S \mid \lambda(\mathbf{x}) \leq K\}.$$

Veja que $0 \leq \lambda(\mathbf{x}) \leq 1$. Se $\lambda(\mathbf{x}) \approx 0$, isto é, $\max_{\theta \in \Psi_0} L(\theta; \mathbf{x}) << \max_{\theta \in \Psi} L(\theta; \mathbf{x})$, acreditamos ser mais provável que a amostra recolhida tenha sido gerada por uma população de parâmetro $\theta \notin \Psi_0$. Nesse caso rejeitamos H_0. Por outro lado, se $\lambda(\mathbf{x}) \approx 1$, isto é, $\max_{\theta \in \Psi_0} L(\theta; \mathbf{x}) \approx \max_{\theta \in \Psi} L(\theta; \mathbf{x})$, acreditamos que a amostra gerada tenha alta probabilidade de ter sido gerada por uma população de parâmetro $\theta \in \Psi_0$, logo não rejeitamos H_0. O valor da constante K é determinado de forma a satisfazer o nível de significância escolhido para o teste dado um tamanho de amostra.

Esse teste é comum de ser usado quando as hipóteses são do tipo $H_0 : \theta = \theta_0$ conta $H_1 : \theta \neq \theta_0$. Veja que para esse caso a estatística $\lambda(\mathbf{x})$ pode ser reescrita por:

$$\lambda(\mathbf{x}) = \frac{\max_{\theta \in \Psi_0} L(\theta; \mathbf{x})}{\max_{\theta \in \Psi} L(\theta; \mathbf{x})} = \frac{\max_{\theta = \theta_0} L(\theta; \mathbf{x})}{\max_{\theta \in \Psi} L(\theta; \mathbf{x})} = \frac{L(\theta_0; \mathbf{x})}{L(\hat{\theta}_{MV}; \mathbf{x})},$$

o que torna o teste mais fácil de ser implementado.

Exemplo 6.16. *Seja $X \sim N(\mu, \sigma_0^2)$, com σ_0^2 conhecido. Considere as seguintes hipóteses: $H_0 : \mu = \mu_0$ conta $H_1 : \mu \neq \mu_0$. Mostrar que o TRVG de nível α é definido pela região crítica*

$$R = \left\{ \mathbf{x} \in S \mid \frac{|\bar{x} - \mu_0|}{\sqrt{\sigma_0^2/n}} \geq z_{1-\frac{\alpha}{2}} \right\}.$$

Solução: Para isso vamos primeiro encontrar a estatística $\lambda(\mathbf{x})$ e depois desenvolver a desigualdade $\lambda(\mathbf{x}) \leq K$ até no padrão da região crítica R definida acima.

Veja que como o teste é em relação ao parâmetro μ, $\Psi_0 = \{\mu_0\}$ e $\Psi = \mathbb{R}$. Então,

$$\lambda(\mathbf{x}) = \frac{\max_{\mu\in\Psi_0} L(\mu;\mathbf{x})}{\max_{\mu\in\Psi} L(\mu;\mathbf{x})} = \frac{\max_{\mu=\mu_0} L(\mu;\mathbf{x})}{\max_{\mu\in\mathbb{R}} L(\mu;\mathbf{x})} = \frac{L(\mu_0;\mathbf{x})}{L(\hat{\mu}_{MV};\mathbf{x})} = \frac{L(\mu_0;\mathbf{x})}{L(\bar{x};\mathbf{x})}$$

$$= \frac{\left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n e^{-\frac{\sum_{i=1}^n (x_i-\mu_0)^2}{2\sigma_0^2}}}{\left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n e^{-\frac{\sum_{i=1}^n (x_i-\bar{x})^2}{2\sigma_0^2}}} = e^{-\frac{1}{2\sigma_0^2}\left(\sum_{i=1}^n (x_i-\mu_0)^2 - \sum_{i=1}^n (x_i-\bar{x})^2\right)}$$

$$= e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n\left(\cancel{x_i^2}+\mu_0^2-2x_i\mu_0-\cancel{x_i^2}-\bar{x}^2+2x_i\bar{x}\right)} = e^{-\frac{1}{2\sigma_0^2}\left(n\mu_0^2-2\mu_0\sum_{i=1}^n x_i - n\bar{x}^2+2\bar{x}\sum_{i=1}^n x_i\right)}$$

$$= e^{-\frac{1}{2\sigma_0^2}\left(n\mu_0^2-2\mu_0 n\bar{x} - n\bar{x}^2+2\bar{x}n\bar{x}\right)} = e^{-\frac{n}{2\sigma_0^2}\left(\mu_0^2-2\mu_0\bar{x} + \bar{x}^2\right)} = e^{-\frac{n}{2\sigma_0^2}(\mu_0-\bar{x})^2}.$$

Então, $\lambda(\mathbf{x}) \leq K$ é equivalente a

$$\lambda(\mathbf{x}) \leq K \iff e^{-\frac{n}{2\sigma_0^2}(\mu_0-\bar{x})^2} \leq K \iff -\frac{n}{2\sigma_0^2}(\mu_0-\bar{x})^2 \leq \ln(K) \iff$$

$$\frac{(\mu_0-\bar{x})^2}{\sigma_0^2/n} \geq -2\ln(K) \iff \sqrt{\frac{(\mu_0-\bar{x})^2}{\sigma_0^2/n}} \geq \underbrace{\sqrt{-2\ln(K)}}_{K'} \iff \frac{|\mu_0-\bar{x}|}{\sqrt{\sigma_0^2/n}} \geq K'$$

Logo a região crítica do teste é definida por

$$R = \left\{\mathbf{x} \in S \mid \frac{|\mu_0-\bar{x}|}{\sqrt{\sigma_0^2/n}} \geq K'\right\}.$$

O valor de K' é encontrado de forma a garantir que o teste tenha o nível de significância desejado. Veja que $Z = \dfrac{\mu_0-\bar{x}}{\sqrt{\sigma_0^2/n}} \sim N(0,1)$, por isso

$$\begin{aligned}
\mathrm{P}\left(\frac{|\mu_0-\bar{x}|}{\sqrt{\sigma_0^2/n}} \geq K' \mid \mu=\mu_0\right) = \alpha &\Rightarrow \mathrm{P}\big(|Z| \geq K' \mid Z \sim N(0,1)\big) = \alpha \\
&\Rightarrow \mathrm{P}\big(Z \geq K' \mid Z \sim N(0,1)\big) = \frac{\alpha}{2} \\
&\Rightarrow \mathrm{P}\big(Z \leq K' \mid Z \sim N(0,1)\big) = 1-\frac{\alpha}{2} \\
&\Rightarrow K' = z_{1-\alpha/2}
\end{aligned}$$

como queríamos mostrar. ||

Exemplo 6.17. *(Teste t) Seja $X \sim N(\mu,\sigma^2)$. Considere as seguintes hipóteses: $H_0 : \mu = \mu_0$ conta $H_1 : \mu \neq \mu_0$. Mostre que o TRVG de nível α é definido pela região crítica*

$$R = \left\{\mathbf{x} \in S \mid \frac{|\bar{x}-\mu_0|}{\sqrt{s^2/n}} \geq t_{n-1,1-\frac{\alpha}{2}}\right\}.$$

Solução:

Para isso vamos primeiro encontrar a estatística $\lambda(\mathbf{x})$ e depois desenvolver a desigualdade $\lambda(\mathbf{x}) \leq K$ até no padrão da região crítica R definida acima.

Veja que novamente $\Psi_0 = \{\mu_0\}$ e $\Psi = \mathbb{R}$. A diferença é que agora σ^2 é desconhecido e então L também é função de σ^2. Precisamos então decidir o que fazer para que λ não dependa de σ^2 e seja de fato uma estatística.

Veja que no numerador da estatística $\lambda(\mathbf{X})$ temos

$$\max_{\mu\in\Psi_0} L(\mu;\sigma^2;\mathbf{x}) = \max_{\mu=\mu_0} L(\mu;\sigma^2;\mathbf{x}) = \max_{\mu=\mu_0} L(\mu_0;\sigma^2;\mathbf{x}) = L(\mu_0;\tilde{\sigma}^2;\mathbf{x})$$

onde $\tilde{\sigma}^2 = \dfrac{\sum_{i=1}^n (x_i-\mu_0)^2}{n}$ é o estimador de máxima verossimilhança para σ^2 supondo $\mu = \mu_0$.

Já o denominador da estatística $\lambda(\mathbf{X})$ fica

$$\max_{\mu\in\Psi} L(\mu;\sigma^2;\mathbf{x}) = \max_{\mu\in\mathbb{R}} L(\mu;\sigma^2;\mathbf{x}) = L(\bar{x};\hat{\sigma}^2;\mathbf{x})$$

onde $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^n (x_i-\bar{x})^2}{n}$, pois esse é o estimador para σ^2 supondo μ desconhecido. Assim,

$$\lambda(\mathbf{x}) = \frac{L(\mu_0;\tilde{\sigma}^2;\mathbf{x})}{L(\bar{x};\hat{\sigma}^2;\mathbf{x})} = \frac{\left(\dfrac{1}{\sqrt{\cancel{2\pi}\tilde{\sigma}^2}}\right)^n e^{-\frac{\sum_{i=1}^n (x_i-\mu_0)^2}{2\tilde{\sigma}^2}}}{\left(\dfrac{1}{\sqrt{\cancel{2\pi}\hat{\sigma}^2}}\right)^n e^{-\frac{\sum_{i=1}^n (x_i-\bar{x})^2}{2\hat{\sigma}^2}}}$$

$$= \left(\frac{\hat{\sigma}^2}{\tilde{\sigma}^2}\right)^{\frac{n}{2}} e^{-\frac{n\sum_{i=1}^n (x_i-\mu_0)^2}{2\sum_{i=1}^n (x_i-\mu_0)^2}+\frac{n\sum_{i=1}^n (x_i-\bar{x})^2}{2\sum_{i=1}^n (x_i-\bar{x})^2}} = \left(\frac{\hat{\sigma}^2}{\tilde{\sigma}^2}\right)^{\frac{n}{2}} = \left(\frac{\sum_{i=1}^n (x_i-\bar{x})^2}{\sum_{i=1}^n (x_i-\mu_0)^2}\right)^{\frac{n}{2}}$$

Então, $\lambda(\mathbf{x}) \le K$ é equivalente a:

$$\Leftrightarrow \quad \left(\frac{\sum_{i=1}^n (x_i-\bar{x})^2}{\sum_{i=1}^n (x_i-\mu_0)^2}\right)^{\frac{n}{2}} \le K \;\Leftrightarrow\; \frac{\sum_{i=1}^n (x_i-\bar{x})^2}{\sum_{i=1}^n (x_i-\mu_0)^2} \le K^{-\frac{n}{2}} \;\Leftrightarrow\; \frac{\sum_{i=1}^n (x_i-\mu_0)^2}{\sum_{i=1}^n (x_i-\bar{x})^2} \ge K^{\frac{n}{2}}$$

$$\Leftrightarrow \quad \frac{\sum_{i=1}^n (x_i^2+\mu_0^2-2x_i\mu_0)}{\sum_{i=1}^n (x_i-\bar{x})^2} \ge K^{\frac{n}{2}} \;\Leftrightarrow\; \frac{\sum_{i=1}^n (x_i^2+\mu_0^2-2x_i\mu_0+\bar{x}^2-\bar{x}^2+2x_i\bar{x}-2x_i\bar{x})}{\sum_{i=1}^n (x_i-\bar{x})^2} \ge K^{\frac{n}{2}}$$

$$\Leftrightarrow \quad \frac{\sum_{i=1}^n (x_i^2+\bar{x}^2-2x_i\bar{x})+\sum_{i=1}^n (\mu_0^2-2x_i\mu_0-\bar{x}^2+2x_i\bar{x})}{\sum_{i=1}^n (x_i-\bar{x})^2} \ge K^{\frac{n}{2}}$$

$$\Leftrightarrow \quad \frac{\sum_{i=1}^n (x_i-\bar{x})^2+n\mu_0^2-2\mu_0 n\bar{x}-n\bar{x}^2+2\bar{x}n\bar{x}}{\sum_{i=1}^n (x_i-\bar{x})^2} \ge K^{\frac{n}{2}} \;\Leftrightarrow\; 1+\frac{n(\bar{x}-\mu_0)^2}{\sum_{i=1}^n (x_i-\bar{x})^2} \ge K^{\frac{n}{2}}$$

$$\Leftrightarrow \quad \sqrt{\frac{(n-1)n(\bar{x}-\mu_0)^2}{\sum_{i=1}^n (x_i-\bar{x})^2}} \ge \underbrace{\sqrt{(n-1)(K^{\frac{n}{2}}-1)}}_{K'} \;\Leftrightarrow\; \frac{|\bar{x}-\mu_0|}{\sqrt{\sum_{i=1}^n (x_i-\bar{x})^2/(n-1)n}} \ge K'$$

$$\Leftrightarrow \quad \frac{|\bar{x}-\mu_0|}{\sqrt{s^2/n}} \ge K'$$

Logo a região crítica do teste é definida por

$$R = \left\{ \mathbf{x} \in S \mid \frac{|\bar{x}-\mu_0|}{\sqrt{s^2/n}} \ge K' \right\}.$$

O valor de K' é encontrado de forma a garantir que o teste tenha o nível de significância desejado. Ou seja, K' é tal que

$$\mathrm{P}\left(\frac{|\bar{x}-\mu_0|}{\sqrt{s^2/n}} \ge K' \mid \mu = \mu_0 \right) = \alpha.$$

Já vimos que se $\mu = \mu_0$ temos $T = \dfrac{\bar{x} - \mu_0}{\sqrt{s^2/n}} \sim t_{n-1}$, por isso

$$\begin{aligned}
\mathrm{P}\left(\frac{|\bar{x} - \mu_0|}{\sqrt{s^2/n}} \geq K' \mid \mu = \mu_0 \right) = \alpha &\Rightarrow \mathrm{P}\left(|T| \geq K' \mid T \sim t_{n-1})\right) = \alpha \\
&\Rightarrow \mathrm{P}\left(T \geq K' \mid T \sim t_{n-1})\right) = \frac{\alpha}{2} \\
&\Rightarrow \mathrm{P}\left(T \leq K' \mid T \sim t_{n-1})\right) = 1 - \frac{\alpha}{2} \\
&\Rightarrow K' = t_{n-1,1-\alpha/2}
\end{aligned}$$

como queríamos mostrar. ||

Exercícios da Seção 6.5

6.5.1 ([Roussas, 2003] - Seção 11.4 e [Bolfarine e Sandoval, 2001]-Capítulo 6) Seja $X_1, \ldots, X_n$ uma amostra aleatória de $X \sim N(\mu_0, \sigma^2)$, com μ_0 conhecido.

(a) Encontre o teste da razão de verossimilhança generalizado de nível de significância α para testar $H_0 : \sigma^2 = \sigma_0^2$ contra $H_1 : \sigma^2 \neq \sigma_0^2$. Dica: A função $g(y) = y^{n/2}e^{-y/2}$ é crescente para $y < n$ e decrescente para $y > n$, atingindo o máximo em $y = n$. Então $y^{n/2}e^{-y/2} < c$ se e somente se $y < c_1$ ou $y > c_2$.

(b) Reescreva a região crítica desse teste para $n = 40$ e $\alpha = 0.01$.

6.6 Teste Baseado em Intervalo de Confiança

Uma outra forma de criar testes para hipóteses bem gerais é a partir dos intervalos de confiança apresentados no Capítulo 5. Veremos como fazer isso nessa seção.

Suponha que queremos testar

$$H_0 : \theta \in \Psi_0 \quad \text{contra} \quad H_1 : \theta \in \Psi_1,$$

para isso encontra-se um intervalo de confiança para θ com confiabilidade $1-\alpha$, vamos chamá-lo de $IC_{\theta,100(1-\alpha),\%0}$ e a partir dele defini-se o teste Γ com região crítica

$$R = \{\mathbf{x} \in S \mid ic_{\theta,100(1-\alpha)\%} \cap \Psi_0 = \emptyset\}. \tag{6.1}$$

A ideia intuitiva por trás desse teste é acreditar que um intervalo de confiança para θ fornece valores próximos ao verdadeiro valor do parâmetro θ. Então, se para uma dada amostra a estimativa do intervalo de confiança não contém valor algum de Ψ_0, isto é, $ic_{\theta,100(1-\alpha)\%} \cap \Psi_0 = \emptyset$, entendemos que o real valor do parâmetro θ da distribuição da população que gerou a amostra tem pouca probabilidade de estar contido em Ψ_0 e por isso rejeitamos H_0. Os testes assim criados terão nível de significância α, onde $1 - \alpha$ é o coeficiente de confiabilidade do intervalo criado.

Uma questão importante é: qual o tipo de intervalo de confiança que devemos escolher? Bilateral, unilateral à direita ou à esquerda? A resposta dessa pergunta depende da hipótese alternativa H_1 e, consequentemente, do conjunto Ψ_1. Vamos escolher o tipo de intervalo que cria um teste cuja função poder é adequada para as hipóteses formuladas. Por isso vamos dividir essa seção por tipos de hipóteses e fazer as análises adequadas para cada um deles. Em todos os casos, sem perda de generalidade, considere $\Psi = \mathbb{R}$.

Testando as hipóteses $H_0 : \theta = \theta_0$ contra $H_1 : \theta \neq \theta_0$

Nesse caso temos

$$\Psi_0 = \{\theta_0\} \quad \text{e} \quad \Psi_1 = \mathbb{R} - \{\theta_0\}.$$

Queremos definir um teste com nível de significância α e com função poder parecida com a apresentada na Figura 6.2(a). Vamos verificar que isso acontece quando escolhemos um intervalo de confiança bilateral para θ com coeficiente de confiança $1-\alpha$, isto é,

$$IC_{\theta,100(1-\alpha)\%} = [L_1(\mathbf{X}), L_2(\mathbf{X})] \quad \text{e} \quad \mathrm{P}(L_1(\mathbf{X}) < \theta < L_2(\mathbf{X})) = 1-\alpha.$$

Veja que a região crítica do teste apresentada na Equação 6.1 pode ser reescrita como

$$R = \{\mathbf{x} \in S \mid \theta_0 < L_1(\mathbf{x}) \ \cup \ L_2(\mathbf{x}) < \theta_0 \ \}.$$

Essa equivalência é ilustrada pela Figura 6.9, que apresenta 10 intervalos de confiança bilaterais para θ. Sempre que há interseção de um intervalo com $\Psi_0 = \{\theta_0\}$ significa que a amostra que gerou o intervalo não está na região crítica e para essa amostra H_0 não é rejeitada. Caso contrário, se $\theta_0 < L_1(\mathbf{x})$ ou $L_2(\mathbf{x}) < \theta_0$, as outras amostras que geraram esses intervalos pertencem a região crítica e para elas H_0 é rejeitada.

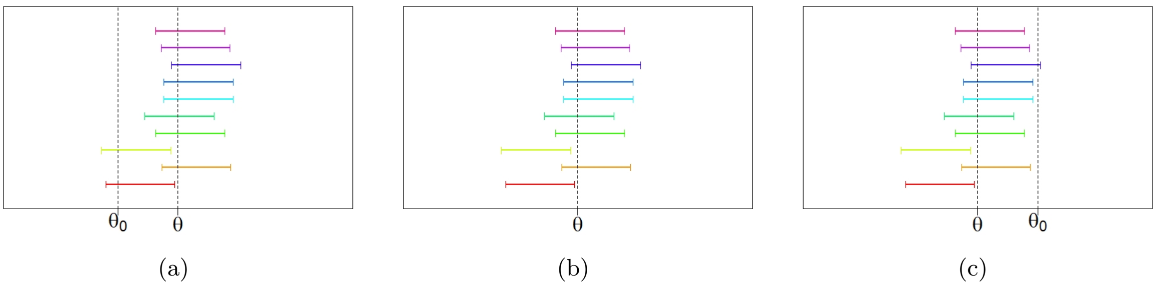


Figura 6.9: Intervalos de confiança bilateral para θ.

Vamos agora analisar o comportamento da função poder

$$\Pi(\theta) = \mathrm{P}(\mathbf{X} \in R \mid \theta) = \mathrm{P}(\theta_0 < L_1(\mathbf{X}) \ \cup \ L_2(\mathbf{X}) < \theta_0 \mid \theta).$$

Primeiro vejamos o valor de $\Pi(\theta_0)$, que no caso da hipótese nula ser simples também é o nível de significância do teste.

$$\begin{aligned}\Pi(\theta_0) &= \mathrm{P}(\theta_0 < L_1(\mathbf{X}) \ \cup \ L_2(\mathbf{X}) < \theta_0 \mid \theta = \theta_0)\\ &= 1 - \mathrm{P}(L_1(\mathbf{X}) < \theta_0 < L_2(\mathbf{X}) \mid \theta = \theta_0)\\ &= 1 - (1-\alpha) = \alpha.\end{aligned}$$

Agora veja que se $\theta_0 < \theta_1 < \theta_2$, $\Pi(\theta_1) \leq \Pi(\theta_2)$. Então Π é crescente para $\theta > \theta_0$.

$$\begin{aligned}\Pi(\theta_1) &= \mathrm{P}(\theta_0 < L_1(\mathbf{X}) \ \cup \ L_2(\mathbf{X}) < \theta_0 \mid \theta = \theta_1) = 1 - \mathrm{P}(L_1(\mathbf{X}) \leq \theta_0 \leq L_2(\mathbf{X}) \mid \theta = \theta_1)\\ &\leq 1 - \mathrm{P}(L_1(\mathbf{X}) \leq \theta_0 \leq L_2(\mathbf{X}) \mid \theta = \theta_2) = \mathrm{P}(\theta_0 < L_1(\mathbf{X}) \ \cup \ L_2(\mathbf{X}) < \theta_0 \mid \theta = \theta_2) = \Pi(\theta_2)\end{aligned}$$

E se $\theta_1 < \theta_2 < \theta_0$, $\Pi(\theta_1) \geq \Pi(\theta_2)$. Então Π é decrescente para $\theta < \theta_0$.

$$\begin{aligned}\Pi(\theta_1) &= \mathrm{P}(\theta_0 < L_1(\mathbf{X}) \ \cup \ L_2(\mathbf{X}) < \theta_0 \mid \theta = \theta_1) = 1 - \mathrm{P}(L_1(\mathbf{X}) \leq \theta_0 \leq L_2(\mathbf{X}) \mid \theta = \theta_1)\\ &\geq 1 - \mathrm{P}(L_1(\mathbf{X}) \leq \theta_0 \leq L_2(\mathbf{X}) \mid \theta = \theta_2) = \mathrm{P}(\theta_0 < L_1(\mathbf{X}) \ \cup \ L_2(\mathbf{X}) < \theta_0 \mid \theta = \theta_2) = \Pi(\theta_2)\end{aligned}$$

Além disso,

$$\lim_{\theta \to \infty} \Pi(\theta) = \lim_{\theta \to \infty} \mathrm{P}(\theta_0 < L_1(\mathbf{X}) \ \cup \ L_2(\mathbf{X}) < \theta_0 \mid \theta) = 1$$
$$\lim_{\theta \to -\infty} \Pi(\theta) = \lim_{\theta \to -\infty} \mathrm{P}(\theta_0 < L_1(\mathbf{X}) \ \cup \ L_2(\mathbf{X}) < \theta_0 \mid \theta) = 1$$

pois quando $\theta \to -\infty$ ou $\theta \to \infty$ aumenta a probabilidade do intervalo de confiança não conter o valor de θ_0 e com isso aumenta a probabilidade de rejeitar H_0. Assim a função poder do teste criado a partir do intervalo de confiança bilateral tem esboço apresentado na Figura 6.10, que é o padrão esperado para as hipóteses formuladas.

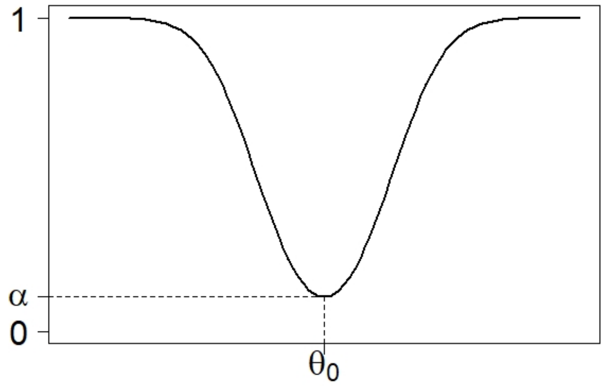


Figura 6.10: Função poder do teste de hipótese baseado em um intervalo de confiança bilateral.

Testando as hipóteses $H_0 : \theta \leq \theta_0$ (ou $\theta = \theta_0$) contra $H_1 : \theta > \theta_0$

Vamos assumir $H_0 : \theta \leq \theta_0$, o caso em que $H_0 : \theta = \theta_0$ é análogo. Nesse caso temos

$$\Psi_0 = (-\infty, \theta_0] \quad \text{e} \quad \Psi_1 = (\theta_0, \infty).$$

Queremos definir um teste com nível de significância α e com função poder parecida com a apresentada na Figura 6.2(b). Vamos verificar que isso acontece quando escolhemos o intervalo unilateral à direita para θ com coeficiente de confiança $1 - \alpha$, isto é,

$$IC_{\theta, 100(1-\alpha)\%} = [L_1(\mathbf{X}), \infty) \quad \text{e} \quad \mathrm{P}(L_1(\mathbf{X}) < \theta) = 1 - \alpha.$$

Veja que a região crítica do teste apresentada na Equação 6.1 pode ser reescrita como

$$R = \{\mathbf{x} \in S \mid \theta_0 < L_1(\mathbf{x})\}.$$

Essa equivalência é ilustrada pela Figura 6.11, que apresenta 10 intervalos de confiança unilaterais à direita para θ. Sempre que há interseção de um intervalo com $\Psi_0 = (-\infty, \theta_0]$, isto é, sempre que $L_1(\mathbf{x}) \leq \theta_0$, significa que a amostra que gerou o intervalo não está na região crítica e para essa amostra H_0 não é rejeitada. Já as outras amostras, que geraram intervalos com $L_1(\mathbf{x}) > \theta_0$ pertencem a região crítica e para elas H_0 é rejeitada.

Vamos analisar o comportamento da função poder

$$\Pi(\theta) = \mathrm{P}(\mathbf{X} \in R \mid \theta) = \mathrm{P}(\theta_0 < L_1(\mathbf{X}) \mid \theta).$$

Veja que se $\theta_1 < \theta_2$, $\Pi(\theta_1) \leq \Pi(\theta_2)$. Então Π é crescente em todo Ψ.

$$\Pi(\theta_1) = \mathrm{P}(\theta_0 < L_1(\mathbf{X}) \mid \theta = \theta_1) \leq \mathrm{P}(\theta_0 < L_1(\mathbf{X}) \mid \theta = \theta_2) = \Pi(\theta_2).$$

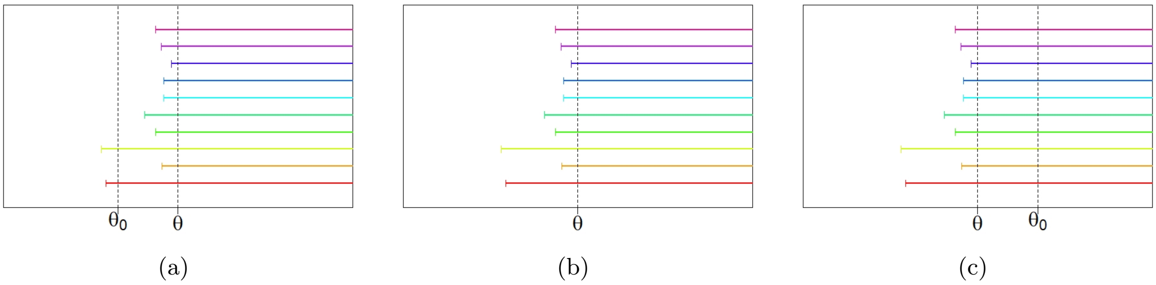


Figura 6.11: Intervalos de confiança unilateral à direita para θ.

E se $\theta = \theta_0$,

$$\Pi(\theta_0) = \mathrm{P}(\theta_0 < L_1(\mathbf{X}) \mid \theta = \theta_0) = 1 - \mathrm{P}(\theta_0 \geq L_1(\mathbf{X}) \mid \theta = \theta_0) = 1 - (1-\alpha) = \alpha.$$

Então o nível de significância do teste é α, pois a função poder é crescente e θ_0 é o maior valor de Ψ_0.

Além disso,

$$\lim_{\theta \to \infty} \Pi(\theta) = \lim_{\theta \to \infty} \mathrm{P}(\theta_0 < L_1(\mathbf{X}) \mid \theta) = 1 \quad \text{e} \quad \lim_{\theta \to -\infty} \Pi(\theta) = \lim_{\theta \to -\infty} \mathrm{P}(\theta_0 < L_1(\mathbf{X}) \mid \theta) = 0$$

pois quando $\theta \to \infty$, o limite inferior do intervalo de confiança fica cada vez maior e diminui a chance do intervalo conter θ_0 e com isso aumenta a chance de rejeitar H_0, isto é, $\Pi(\theta) \to 1$. Já se $\theta \to -\infty$ o limite inferior do intervalo de confiança fica cada vez menor e então aumenta a chance do intervalo conter θ_0 e com isso diminui a chance de rejeitar H_0, ou seja, $\Pi(\theta) \to 0$.

Podemos então concluir que a função poder do teste criado a partir do intervalo de confiança unilateral à direita tem esboço apresentado na Figura 6.12, que é o padrão esperado para as hipóteses formuladas.

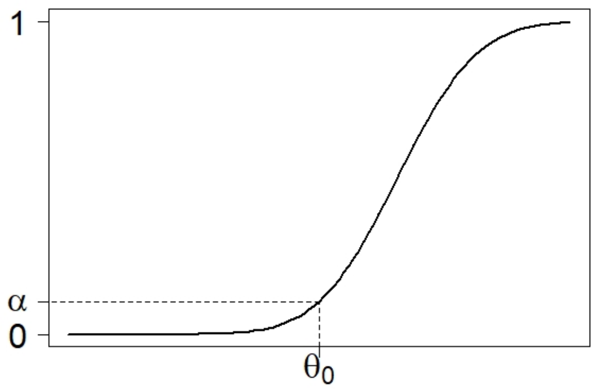


Figura 6.12: Função poder do teste de hipótese baseado em um intervalo de confiança unilateral à direita.

Testando as hipóteses $H_0 : \theta \geq \theta_0$ (ou $\theta = \theta_0$) contra $H_1 : \theta < \theta_0$

Vamos assumir $H_0 : \theta \geq \theta_0$, o caso em que $H_0 : \theta = \theta_0$ é análogo. Nesse caso temos

$$\Psi_0 = [\theta_0, \infty) \quad \text{e} \quad \Psi_1 = (-\infty, \theta_0).$$

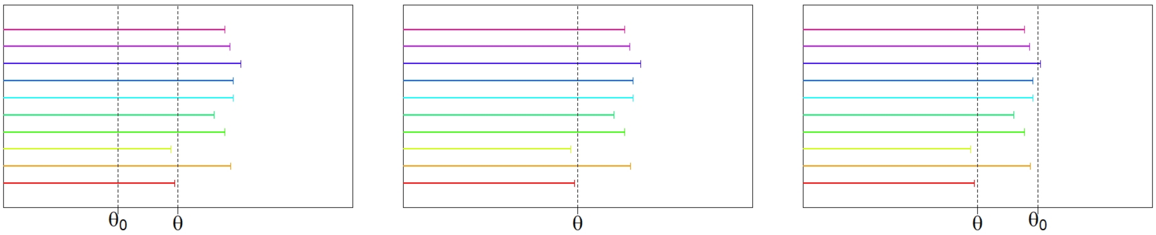
Queremos definir um teste com nível de significância α e com função poder parecida com a apresentada na Figura 6.2(c). Vamos verificar que isso acontece quando escolhemos o intervalo unilateral à esquerda para θ com coeficiente de confiança $1-\alpha$, isto é,

$$IC_{\theta,100(1-\alpha)\%} = (-\infty, L_2(\mathbf{X})] \quad \text{e} \quad \mathrm{P}(\theta < L_2(\mathbf{X})) = 1-\alpha.$$

Veja que a região crítica do teste apresentada na Equação 6.1 pode ser reescrita como

$$R = \{\mathbf{x} \in S \mid L_2(\mathbf{x}) < \theta_0\}.$$

Essa equivalência é ilustrada pela Figura 6.13, que apresenta 10 intervalos de confiança unilaterais à esquerda para θ. Sempre que há interseção de um intervalo com $\Psi_0 = [\theta_0, \infty)$, isto é, sempre que $\theta_0 \leq L_2(\mathbf{x})$ significa que a amostra que gerou o intervalo não está na região crítica e para essa amostra H_0 não é rejeitada. Já as outras amostras, que geraram intervalos com $L_2(\mathbf{x}) < \theta_0$ pertencem a região crítica e para elas H_0 é rejeitada.



(a) (b) (c)

Figura 6.13: Intervalos de confiança unilateral à direita para θ.

Vamos analisar o comportamento da função poder

$$\Pi(\theta) = \mathrm{P}(\mathbf{X} \in R \mid \theta) = \mathrm{P}(L_2(\mathbf{X}) < \theta_0 \mid \theta).$$

Veja que se $\theta_1 < \theta_2$, $\Pi(\theta_1) \geq \Pi(\theta_2)$. Então Π é decrescente em todo Ψ.

$$\Pi(\theta_1) = \mathrm{P}(L_2(\mathbf{X}) < \theta_0 \mid \theta = \theta_1) \geq \mathrm{P}(L_2(\mathbf{X}) < \theta_0 \mid \theta = \theta_2) = \Pi(\theta_2).$$

E se $\theta = \theta_0$,

$$\Pi(\theta_0) = \mathrm{P}(L_2(\mathbf{X}) < \theta_0 \mid \theta = \theta_0) = 1 - \mathrm{P}(L_2(\mathbf{X}) \geq \theta_0 \mid \theta = \theta_0) = 1-(1-\alpha) = \alpha.$$

Então o nível de significância do teste é α, pois a função poder é decrescente e θ_0 é o menor valor de Ψ_0.

Além disso,

$$\lim_{\theta\to\infty} \Pi(\theta) = \lim_{\theta\to\infty} \mathrm{P}(\theta_0 < L_1(\mathbf{X}) \mid \theta) = 0 \qquad \lim_{\theta\to-\infty} \Pi(\theta) = \lim_{\theta\to-\infty} \mathrm{P}(\theta_0 < L_1(\mathbf{X}) \mid \theta) = 1$$

pois quando $\theta \to \infty$, o limite superior do intervalo de confiança fica cada vez maior e aumenta a chance do intervalo conter θ_0. Assim diminui a chance de rejeitar H_0, isto é, $\Pi(\theta) \to 0$. Já se $\theta \to -\infty$ o limite superior do intervalo de confiança fica cada vez menor e então diminui a chance do intervalo conter θ_0. Assim aumenta a chance de rejeitar H_0, isto é, $\Pi(\theta) \to 1$.

Podemos então concluir que a função poder do teste criado a partir do intervalo de confiança unilateral à esquerda tem esboço apresentado na Figura 6.14, que é o padrão esperado para as hipóteses formuladas.

Resumindo, quaisquer que sejam as hipóteses, o teste baseado em um intervalo de confiança para θ é aquele definido pela região crítica

$$R = \{\mathbf{x} \in S \mid ic_{\theta,100(1-\alpha)\%} \cap \Psi_0 = \emptyset\}.\},$$

onde $IC_{\theta,100(1-\alpha)\%}$ é um intervalo de confiança para θ com coeficiente de confiança α. O tipo de intervalo escolhido depende da hipótese alternativa H_1, como mostra a Tabela 6.1.

Todos os testes assim criados tem nível de significância α. Além disso, a função poder tem comportamento adequado às hipóteses formuladas.

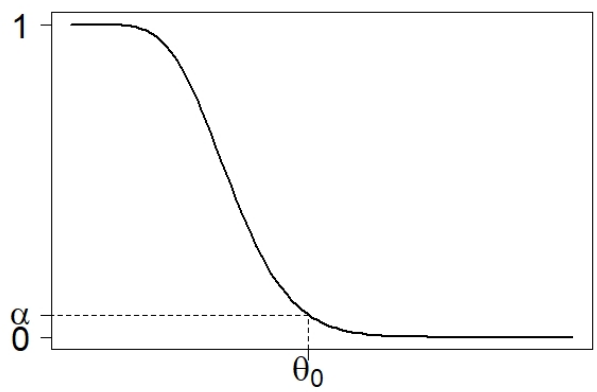


Figura 6.14: Função poder do teste de hipótese baseado em um intervalo de confiança unilateral à esquerda.

Tabela 6.1: Escolha do tipo de intervalo.

H_1 :	Ψ_1	$IC_{\theta,100(1-\alpha)\%}$	Tipo
$\theta \neq \theta_0$	$(-\infty, \theta_0) \cap (\theta_0, \infty)$	$[L_1(\mathbf{X}), L_2(\mathbf{X})]$	Bilateral
$\theta > \theta_0$	(θ_0, ∞)	$[L_1(\mathbf{X}), \infty)$	Unilateral à Direita
$\theta < \theta_0$	$(-\infty, \theta_0)$	$(-\infty, L_2(\mathbf{X})]$	Unilateral à Esquerda

Exemplo 6.18. *Considere a população $X \sim N(\mu, \sigma^2)$. Queremos testar*

$$H_0 : \sigma^2 \leq 2 \quad \textit{contra} \quad H_1 : \sigma^2 > 2.$$

(a) Defina a regra de decisão do teste, baseado em um intervalo de confiança para σ^2, para as hipóteses acima com nível de significância $\alpha = 0,05$ supondo uma amostra de tamanho $n = 10$.

(b) Use o Programa R para fazer o gráfico da função poder do teste criado no item anterior.

(c) Decida entre aceitar H_0 ou rejeitá-la para a amostra de X dada por:

$$0,066 \quad 1,638 \quad 0,184 \quad -2,494 \quad 0,926 \quad -0,159 \quad 2,171 \quad 1,176 \quad -1,738 \quad -1,786$$

Solução:

(a) Veja que nesse exemplo temos $\Psi_0 = (-\infty, 2]$ e $\Psi_1 = (, \infty)$. Então precisamos do intervalo de confiança unilateral à direita para criar a região crítica do teste. No Exemplo 5.3 esse intervalo foi criado:

$$IC_{\sigma^2,100(1-\alpha)\%} = \left[\frac{(n-1)S^2}{q_{1-\alpha,n-1}}, \infty\right).$$

Então a região crítica do nosso teste será:

$$R = \left\{\mathbf{x} \in S \mid \sigma_0^2 < \frac{(n-1)s^2}{q_{1-\alpha,n-1}}\right\}.$$

Como $\alpha = 0,05$ e $n = 10$, $q_{0,95,9} = 16,91898$. Além disso, $\sigma_0^2 = 2$ e $(n-1)s^2 = \sum_{i=1}^{10}(x_i - \bar{x})^2$. Assim a região crítica fica:

$$R = \left\{\mathbf{x} \in S \mid 2 < \frac{\sum_{i=1}^{10}(x_i - \bar{x})^2}{16,91898}\right\} = \left\{\mathbf{x} \in S \mid 33.83796 < \sum_{i=1}^{10}(x_i - \bar{x})^2\right\}$$

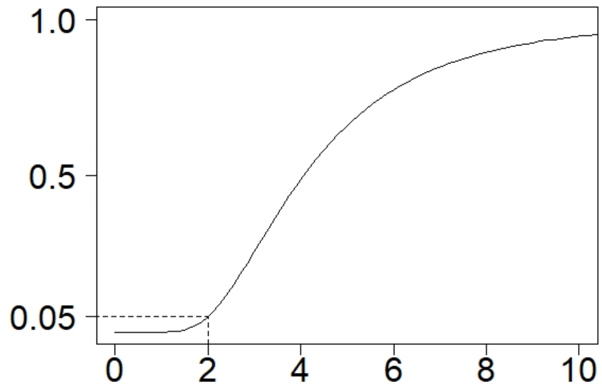


Figura 6.15: Função poder do teste do Exemplo 6.18.

(b) A função poder desse teste é

$$\Pi(\sigma^2) = \mathrm{P}\left(33.83796 < \sum_{i=1}^{10}(X_i - \bar{X})^2 \mid \sigma^2\right) = \mathrm{P}\left(\frac{33,83796}{\sigma^2} < \frac{\sum_{i=1}^{10}(X_i - \bar{X})^2}{\sigma^2} \mid \sigma^2\right)$$

e $W = \dfrac{\sum_{i=1}^{10}(X_i - \bar{X})^2}{\sigma^2} = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_9$. Logo $\Pi(\sigma^2) = 1 - \mathrm{P}\left(W \leq \dfrac{33.83796}{\sigma^2} \mid W \sim \chi^2_9\right)$. Usando o comando `curve(1 - pchisq(33.83796/x,df=9),xlim=c(0,10))` é possível gerar a curva da função poder como a da Figura 6.15.

(c) Para a amostra apresentada temos $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 22,13068$. Como esse valor é menor que $33,83796$, a amostra não está na região crítica e então H_0 não é rejeitada.

Outra maneira de chegar a essa conclusão seria criar o intervalo de confiança estimado a partir dessa amostra:

$$ic_{\sigma^2,100(1-\alpha)\%} = \left[\frac{(n-1)s^2}{q_{1-\alpha,n-1}}, \infty\right) = [1,308039\ ,\ \infty)\,.$$

Veja que o valor $\sigma_0^2 = 2$ pertence ao intervalo, então não rejeitamos H_0.

‖

Uma vantagem dos testes baseados em intervalos de confiança é que podemos também aproveitar os intervalos de confiança aproximados para encontrar testes aproximados e os intervalos para duas amostras independentes e criar testes para duas amostras independentes. Veja um exemplo.

Exemplo 6.19. *Sejam $X \sim N(\mu_x, \sigma^2)$ e $Y \sim N(\mu_y, \sigma^2)$ variáveis aleatórias independentes, com distribuição normal de mesma variância e médias distintas. Suponha uma amostra de tamanho n de X e outra de tamanho m de Y. Queremos testar*

$$H_0 : \mu_x = \mu_y \quad \text{contra} \quad H_1 : \mu_x \neq \mu_y.$$

(a) Defina um teste para as hipóteses acima a partir de um intervalo de confiança. Considere nível de significância de $\alpha = 0,1$ supondo uma amostras de tamanhos $n = 12$ e $m = 15$.

(b) Decida entre aceitar H_0 ou rejeitá-la para as amostras de X e Y apresentadas abaixo.

X: -0.486 1.311 1.062 0.802 -0.497 -0.824 -0.229 -0.794 -0.169 -2.674 -0.441 -1.417

Y: 1.471 0.976 1.671 1.663 1.183 0.329 -0.250 1.261 1.519 0.837 2.035 0.037 0.277 1.121 -0.655

Solução:

(a) Primeiro veja que testar as hipóteses acima é o mesmo que testar

$$H_0 : \mu_x - \mu_y = 0 \quad \text{contra} \quad H_1 : \mu_x - \mu_y \neq 0.$$

Então vamos definir $\theta = \mu_x - \mu_y$ e procurar um intervalo de confiança bilateral para θ. Veja que esse é o resultado do Exemplo 5.12, que tem como resultado o seguinte intervalo para θ com coeficiente de confiança $1 - \alpha$,

$$IC_{\theta,100(1-\alpha)\%} = \left[(\bar{X} - \bar{Y}) - t_{1-\frac{\alpha}{2},n+m-2} \frac{S_p}{\sqrt{\frac{nm}{n+m}}} \ , \ (\bar{X} - \bar{Y}) - t_{\frac{\alpha}{2},n+m-2} \frac{S_p}{\sqrt{\frac{nm}{n+m}}} \right].$$

com $S_p^2 = \dfrac{1}{n+m-2} \left(\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2 \right)$.

Como $\theta_0 = 0$, a região crítica do nosso teste será:

$$\begin{aligned}
R &= \left\{ \mathbf{x} \in S \mid 0 < (\bar{x} - \bar{y}) - t_{1-\frac{\alpha}{2},n+m-2} \frac{s_p}{\sqrt{\frac{nm}{n+m}}} \text{ ou } (\bar{x} - \bar{y}) - t_{\frac{\alpha}{2},n+m-2} \frac{s_p}{\sqrt{\frac{nm}{n+m}}} < 0 \right\} \\
&= \left\{ \mathbf{x} \in S \mid \ |(\bar{x} - \bar{y})| > t_{1-\frac{\alpha}{2},n+m-2} \frac{s_p}{\sqrt{\frac{nm}{n+m}}} \right\}, \ \text{pois } t_{1-\frac{\alpha}{2},n+m-2} = -t_{1-\frac{\alpha}{2},n+m-2} \\
&= \left\{ \mathbf{x} \in S \mid \ |(\bar{x} - \bar{y})| > t_{1-\frac{\alpha}{2},n+m-2} \frac{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2}}{\sqrt{\frac{nm(n+m-2)}{n+m}}} \right\} \\
&= \left\{ \mathbf{x} \in S \mid \ \left| \frac{\bar{x} - \bar{y}}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2}} \right| > \frac{t_{1-\frac{\alpha}{2},n+m-2}}{\sqrt{\frac{nm(n+m-2)}{n+m}}} \right\}
\end{aligned}$$

Supondo $\alpha = 0,1$, $n = 12$ e $m = 15$ temos $t_{0,95,25} = 1,708141$ e $\sqrt{\frac{nm(n+m-2)}{n+m}} = \sqrt{\frac{180*25}{27}} = 12,90994$. Então a região crítica fica:

$$R = \left\{ \mathbf{x} \in S \mid \ \left| \frac{\bar{x} - \bar{y}}{\sqrt{(\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2)}} \right| > 0,1323121 \right\}.$$

(b) Para a amostra observada temos

$$\left| \frac{\bar{x} - \bar{y}}{\sqrt{(\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^m (y_i - \bar{y})^2)}} \right| = 0,2702889$$

e por isso ela encontra-se na região crítica. Então rejeitamos H_0, ou seja, rejeitamos a hipótese das duas populações terem mesma média.

Poderíamos tomar a decisão baseados na estimativa para o intervalo supondo a amostra observada. O intervalo estimado fica assim:

$$ic_{\theta,100(1-\alpha)\%} = [-1,541771 \ , \ -0,9808958]$$

e como $\theta_0 = 0$ não está contido no intervalo estimados rejeitamos H_0.

$\|$

Exercícios da Seção 6.6

6.6.1 Seja $X_1, \ldots, X_n$ amostra aleatória de $X \sim Exp(\lambda)$ e média $\mu = 1/\lambda$. Para as hipóteses apresentadas nos itens (a), (b) e (c) a seguir defina a região crítica dos testes baseados em intervalos de confiança, considerando nível de significância $\alpha = 0.05$. Em seguida, tome a decisão entre rejeitar ou não H_0 caso a amostra recolhida para X seja:

$$0,983 \quad 10,273 \quad 4,336 \quad 2,064 \quad 7,183 \quad 3,274 \quad 8,733 \quad 6,376 \quad 0,380 \quad 1,250.$$

(a) $H_0 : \mu = 5 \times H_1 : \mu \neq 5$ (b) $H_0 : \mu = 5 \times H_1 : \mu < 5$ (c) $H_0 : \mu \leq 5 \times H_1 : \mu > 5$.

6.6.2 (Teste t) Seja $X \sim N(\mu, \sigma^2)$. Mostre que o teste baseado em um intervalo de confiança para as hipóteses

$$H_0 : \mu = \mu_0 \quad \text{contra} \quad H_1 : \mu \neq \mu_0.$$

resulta no Teste t, isto é, no teste de nível de confiança α cuja região crítica é definida por

$$R = \left\{ \mathbf{x} \in S \mid \frac{|\bar{x} - \mu_0|}{\sqrt{s^2/n}} \geq t_{n-1,1-\frac{\alpha}{2}} \right\}.$$

6.6.3 Seja $X \sim N(\mu_x, \sigma_x^2)$ e $Y \sim N(\mu_y, \sigma_y^2)$ variáveis aleatórias independentes. Suponha uma amostra de tamanho n de X e outra de tamanho m de Y. Considere as seguintes hipóteses

$$H_0 : \sigma_x^2 \geq \sigma_y^2 \quad \text{contra} \quad H_1 : \sigma_x^2 < \sigma_y^2.$$

(a) Encontre o teste para as hipóteses acima baseado em um intervalo de confiança.

(b) Suponha as amostras a seguir. Com nível de significância $\alpha = 0,1$, você acredita que a variância da população X é maior ou igual que a variância da população Y?

X:	1,326	-0,480	0,653	0,649	1,234	1,246	0,559	3,195		
Y:	-2,271	2,819	0,284	2,566	1,571	-0,393	-5,272	1,071	0,426	0,671

6.6.4 ([Bolfarine e Sandoval, 2001] - Capítulo 6) Sejam $X \sim N(\mu_X, 9)$ e $Y \sim N(\mu_Y, 25)$ variáveis aleatórias independentes. Suponha que queremos testar as seguintes hipóteses:

$$H_0 : \mu_X = \mu_Y \text{ contra } H_1 : \mu_X \neq \mu_Y$$

(a) Defina a regra de decisão para o teste de nível $\alpha = 0.05$ baseado em um intervalo de confiança.

(b) Suponha amostras de X e Y tais que: $n = 9$; $\sum_{i=1}^{9} x_i = 3,4$; $m = 16$ e $\sum_{i=1}^{16} y_i = 4,3$. Qual a sua conclusão sobre as hipóteses acima, ao nível de significância de 5%?

6.7 Valor-p

Quando realizamos um teste estatístico precisamos informar a decisão tomada de alguma maneira. O que vimos até agora é que escolhemos um nível de significância α e a partir dele junto com a observação de uma amostra observada, tomamos uma decisão entre rejeitar ou não H_0. Então uma maneira de reportar o resultado do teste é informar a decisão tomada junto com o nível de significância escolhido.

Repare que a escolha do nível de significância dá alguma informação sobre a credibilidade da nossa decisão. Por exemplo, se escolhemos um α muito pequeno e tomamos a decisão de rejeitar H_0, temos grande credibilidade nessa decisão, pois o α pequeno indica pequena probabilidade de rejeitar uma hipótese nula que é verdadeira. Por outro lado, se o valor de α for grande e a decisão é rejeitar H_0, temos pouca credibilidade nessa decisão. Pois o α grande indica grande probabilidade de rejeitar H_0 sendo ela verdadeira.

Vejamos um exemplo pra entender como o valor de α pode influencia na decisão tomada.

Suponha uma amostra aleatória de tamanho 9 de uma população $X \sim N(\mu, 1)$. Estamos interessados em testar

$$H_0 : \mu \leq 2 \quad \text{contra} \quad H_1 : \mu > 2.$$

Podemos verificar que o teste UMP para as hipóteses formuladas tem região crítica

$$R = \{\mathbf{x} \in S \mid \bar{x} > c\}$$

e a constante c é definida a partir do nível de significância α estabelecido para o teste.

$$\alpha = \mathrm{P}(\bar{X} > c \mid \mu = 2) = \mathrm{P}\left(\frac{\bar{X} - 2}{1/\sqrt{9}} > \frac{c-2}{1/\sqrt{9}} \mid \mu = 2\right) = \mathrm{P}\left(Z > 3(c-2)\right) = 1 - \Phi\left(3(c-2)\right).$$

$$3(c-2) = z_{1-\alpha} \quad \Rightarrow \quad c = \frac{z_{1-\alpha}}{3} + 2.$$

Assim, a região crítica do teste passa a ser:

$$R = \left\{\mathbf{x} \in S \mid \bar{x} > \frac{z_{1-\alpha}}{3} + 2\right\}.$$

Suponha agora a observação de uma amostra tal que $\bar{x} = 2,5$. Qual decisão tomar? Repare que essa decisão depende do nível de significância α escolhido. Vejamos algumas possibilidades.

$$\begin{array}{llll}
\text{Se } \alpha = 0,001 & \Rightarrow \quad \frac{z_{1-\alpha}}{3} + 2 = 3,030077 & \Rightarrow & \text{não rejeitamos } H_0. \\
\text{Se } \alpha = 0,01 & \Rightarrow \quad \frac{z_{1-\alpha}}{3} + 2 = 2,775449 & \Rightarrow & \text{não rejeitamos } H_0. \\
\text{Se } \alpha = 0,05 & \Rightarrow \quad \frac{z_{1-\alpha}}{3} + 2 = 2,548285 & \Rightarrow & \text{não rejeitamos } H_0. \\
\text{Se } \alpha = 0,08 & \Rightarrow \quad \frac{z_{1-\alpha}}{3} + 2 = 2,468357 & \Rightarrow & \text{rejeitamos } H_0. \\
\text{Se } \alpha = 0,10 & \Rightarrow \quad \frac{z_{1-\alpha}}{3} + 2 = 2,427184 & \Rightarrow & \text{rejeitamos } H_0. \\
\text{Se } \alpha = 0,15 & \Rightarrow \quad \frac{z_{1-\alpha}}{3} + 2 = 2,345478 & \Rightarrow & \text{rejeitamos } H_0.
\end{array}$$

Não é coincidência que para α muito pequeno não se rejeita H_0. Isso porque α pequeno indica uma probabilidade muito pequena de rejeitar H_0 sendo H_0 verdadeira. Então H_0 só é rejeitada quando há muita evidência para isso. Conforme o valor de α cresce, cresce a probabilidade de rejeitar H_0 sendo ela verdadeira e em algum momento, para a mesma amostra observada, mudamos de decisão: trocarmos de "não rejeitar H_0" para "rejeitar H_0".

E qual é o valor $\alpha^\star$ para o qual trocamos entre rejeitar e não rejeitar? Isto é, quem é o $\alpha^\star$ para o qual se escolheremos $\alpha < \alpha^\star$ tomamos a decisão de não rejeitar e se escolhermos $\alpha > \alpha^\star$ tomamos a decisão de rejeitar? Esse valor é o que chamamos de valor-p do teste. Vamos ver qual o valor-p para o exemplo em questão.

Temos que encontrar o α tal que

$$\frac{z_{1-\alpha}}{3} + 2 = 2,5 \;\Rightarrow\; z_{1-\alpha} = 3(2,5-2) = 1,5 \;\Rightarrow\; \mathrm{P}(Z < 1,5) = 1 - \alpha \;\Rightarrow$$
$$\alpha = 1 - \mathrm{P}(Z < 1,5) = 0,0668072$$

Então a amostra observada resulta em um o valor-p de $0,0668072$. E o que isso indica? Indica que se usarmos $\alpha < 0,0668072$ vamos decidir não rejeitar H_0 e se usarmos $\alpha > 0,0668072$ vamos decidir rejeitar H_0. Repare que a decisão não é muito certa, para o típico $\alpha = 0,05$ a decisão seria não rejeitar H_0, porém para $\alpha = 0,1$, que não é um valor grande de mais, a decisão já seria rejeitar H_0. Mas precisamos tomar uma decisão. Podemos escolher não rejeitar H_0 e informar o valor-p encontrado, assim indicamos essa falta de certeza na decisão tomada. Veja que nenhum valor de α foi escolhido.

Vamos formalizar esse conceito.

Definição 6.8. *Dado um teste estatístico e observada uma amostra da população, o valor-p é o maior valor de α para o qual decidimos não rejeitar H_0. Ele também pode ser definido como o menor valor de α para o qual decidimos rejeitar H_0.*

Se o valor-p de um teste for muito pequeno, temos fortes evidências para rejeitar H_0. Se o valor-p de um teste for grande, não temos evidências para rejeitar H_0. Na prática costuma-se usar o critério de rejeitar H_0 quando valor-p $< 0,05$.

Podemos entender também o valor-p como o nível de significância α que faz com que a amostra observada esteja no limite entre a região crítica e a região de aceitação.

Exemplo 6.20. *Considere Exemplo 6.18. Nesse exemplo foi criado um teste para as hipóteses*

$$H_0 : \sigma^2 \leq 2 \quad \textit{contra} \quad H_1 : \sigma^2 > 2.$$

com região crítica dada por

$$R = \left\{ \mathbf{x} \in S \mid \sigma_0^2 < \frac{(n-1)s^2}{q_{1-\alpha,n-1}} \right\}.$$

A amostra observada, de tamanho 10, gerou $(n-1)s^2 = \sum_{i=1}^{10}(x_i - \bar{x})^2 = 22,13068$. Qual o valor-p desse teste. Qual decisão tomar?

Solução: Para calcular o valor-p precisamos descobrir o valor de α que faz com que $\dfrac{(n-1)s^2}{q_{1-\alpha,n-1}} = \sigma_0^2$. Isto é, o α tal que

$$\frac{22,13068}{q_{1-\alpha,9}} = 2 \Rightarrow \frac{22,13068}{2} = q_{1-\alpha,9} \Rightarrow 1-\alpha = \mathrm{P}(W \leq 11,06534 \mid W \sim \chi^2_9)$$
$$\alpha = 1 - \mathrm{P}(W \leq 11,06534 \mid W \sim \chi^2_9)$$

A partir do Programa R podemos usar o comando `1 - pchisq(11.06534,df=9)` e chegar no resultado

$$\text{valor-p } = 0,2712568.$$

Veja que o valor-p é grande, bem maior que $0,05$. Ele indica que a gente só rejeitaria H_0 se usássemos níveis de significância maiores que $0,27$. Com isso a decisão é não rejeitar H_0 e essa decisão poderia ser reportada da seguinte maneira: "Não há evidências para rejeitar a hipótese H_0 uma vez que foi calculado um valor-p de $0,2712568$ para a amostra observada". Repare que a decisão foi tomada sem que tenha sido feita uma escolha de α. ‖

Exemplo 6.21. *Considere o Teste t apresentado no Exemplo6.17. Esse teste foi criado para uma população $X \sim N(\mu, \sigma^2)$ quando pretende-se testar $H_0 : \mu = \mu_0$ conta $H_1 : \mu \neq \mu_0$. A região crítica do teste é*

$$R = \left\{ \mathbf{x} \in S \mid \frac{|\bar{x} - \mu_0|}{\sqrt{s^2/n}} \geq t_{n-1,1-\frac{\alpha}{2}} \right\}.$$

Suponha $\mu_0 = 1$ e a amostra de X apresentada a seguir, tome uma decisão entre rejeitar ou não H_0 baseada no valor-p encontrado.

2,337	*2,544*	*2,201*	*2,294*	*0,921*	*2,695*	*0,760*
1,578	*0,875*	*3,154*	*1,595*	*3,344*	*0,830*	*2,342*

Solução: Para calcular o valor-p precisamos encontrar o α que faz com que

$$\frac{|\bar{x} - \mu_0|}{\sqrt{s^2/n}} = t_{n-1,1-\frac{\alpha}{2}}$$

Para a amostra apresentada podemos calcular $n = 14$, $\bar{x} = 1,962143$, $s^2 = 0,7640611$ e com isso podemos encontrar $\frac{|\bar{x}-\mu_0|}{\sqrt{s^2/n}} = 4,118505$. Então queremos o valor de α tal que

$$t_{13,1-\frac{\alpha}{2}} = 4,118505 \quad \Rightarrow \quad 1 - \frac{\alpha}{2} = \mathrm{P}\left(T \leq 4,118505 \mid T \sim t_{13}\right) \quad \Rightarrow$$
$$\alpha = 2\left(1 - \mathrm{P}\left(T \leq 4,118505 \mid T \sim t_{13}\right)\right).$$

A partir do Programa R podemos usar o comando `2*(1-pt(4.118505,df=13))` e encontrar

$$\text{valor-p } = 0,001210463.$$

Assim encontramos o valor-p bem pequeno, o que nos dá grande evidência em rejeitar H_0. Repare que H_0 seria rejeitada para qualquer escolha de $\alpha > 0,001210463$, ou seja, para todos os valores típicos de α tomaríamos a decisão de rejeitar H_0. Sendo assim o resultado do teste poderia ser reportado da seguinte maneira: "Decide-se rejeitar a hipótese de que a média da população X é igual a 1 após encontrar um valor-p de $0,001210463$ para a amostra observada". Repare que a decisão foi tomada sem que tenha sido feita uma escolha de α.

$\|$

Exercícios da Seção 6.7

6.7.1 Calcule o valor-p do teste realizado no Exercício 6.2.2.

6.7.2 Calcule o valor-p para o teste realizado no Exercício 6.4.1.

6.8 Mais Alguns Exercícios do Capítulo 6

6.1. Seja $X_1, \ldots, X_n$ amostra aleatória de X com função densidade $f_X(x) = \theta x^{\theta-1}$, $0 < x < 1$ e $\theta > 0$. Suponha que queremos testar as hipóteses

$$H_0 : \theta = 1 \qquad \text{contra} \qquad H_1 : \theta \neq 1.$$

(a) Apresente a região crítica do teste baseado em um intervalo de confiança para θ em função dos valores de α e n.

(b) Considere a amostra X apresentada a seguir. Baseado no valor-p do teste, qual decisão tomar em relação às hipóteses formuladas acima?

$$0,89 \quad 0,55 \quad 0,15 \quad 0,18 \quad 0,03 \quad 0,09 \quad 0,22 \quad 0,06 \quad 0,78 \quad 0,69 \quad 0,55 \quad 0,18$$

Dica: reveja o resultado dos Exercícios 2.1.11 e 5.2.7.

6.2. Seja X uma variável aleatória contínua cuja função densidade depende do parâmetro $\theta > 0$ e é definida por

$$f_X(x) = 3\theta^3 x^2 e^{-(\theta x)^3}\ , \quad x > 0.$$

Veja que $X \sim Weib(3, 1/\theta)$ e que esta distribuição é a mesma dos Exercícios 2.5.3, 3.13 e 5.6.

(a) Considere uma amostra de tamanho $n = 9$ e adote o nível de significância de $\alpha = 10\%$. Apresente a região crítica do teste Uniformemente Mais Poderoso para as hipóteses:

$$H_0 : \theta \leq 0,5 \qquad \times \qquad H_1 : \theta > 0,5$$

(b) Encontre a função poder do teste definido no item (a) e esboce seu gráfico. Use o R para fazer o gráfico. Aproveite para checar se de fato o teste tem o nível de significância adotado.

(c) Considere os valores abaixo uma amostra da população X. Baseado nesta amostra tome uma decisão entre rejeitar ou não H_0 para as hipóteses apresentadas.

$$2,20 \quad 1,99 \quad 1,65 \quad 0,92 \quad 2,34 \quad 0,95 \quad 0,77 \quad 1,49 \quad 1,55$$

6.3. Seja X uma variável aleatória contínua cuja função densidade depende do parâmetro $\theta > 0$ e é definida por

$$f_X(x) = \frac{\theta 4^\theta}{x^{\theta+1}}\ , \quad x > 4.$$

Veja que $X \sim Pareto(\theta, 4)$ e que esta distribuição é a mesma dos Exercícios 2.5.1, 3.11 e 5.5.

(a) Considere uma amostra de tamanho $n = 14$ e adote o nível de significância de $\alpha = 5\%$. Apresente a região crítica do teste baseado em um intervalo de confiança para as hipóteses:

$$H_0 : \theta = 1 \qquad \times \qquad H_1 : \theta \neq 1$$

(b) Considere os valores abaixo uma amostra da população X. Baseado nesta amostra tome uma decisão entre rejeitar ou não H_0 para as hipóteses apresentadas.

$$9,95 \quad 5,01 \quad 4,78 \quad 4,02 \quad 6,60 \quad 4,83 \quad 4,42 \quad 4,23 \quad 23,58 \quad 4,96 \quad 4,39 \quad 14,37 \quad 7,12 \quad 4,57$$

(c) Para a amostra apresentada no item anterior, encontre o valor-p do teste.

6.4. Seja X uma variável aleatória discreta cuja função de probabilidade depende do parâmetro $0 < p < 1$ e é definida por

$$p_X(x) = \binom{15}{x} p^x (1-p)^{15-x}, \quad x = 0, 1, \ldots, 15$$

Veja que $X \sim Binom(15, p)$ e que esta distribuição é a mesma dos Exercícios 2.5.5, 3.15 e 5.8. Para os itens a seguir considere uma amostra de tamanho $n = 25$ e as hipóteses

$$H_0 : p \geq 0,3 \qquad \times \qquad H_1 : p < 0,3$$

(a) A partir do teste Uniformemente Mais Poderoso e do Teorema Limite Central encontre uma região crítica com nível de significância aproximado de $\alpha = 5\%$ para as hipóteses enunciadas.

(b) A partir do teste baseado em um Intervalo de Confiança aproximado encontre uma região crítica com nível de significância aproximado de $\alpha = 5\%$ para as hipóteses enunciadas.

(c) Considere os valores abaixo uma amostra da população X. Baseado nesta amostra tome uma decisão entre rejeitar ou não H_0 para as hipóteses apresentadas.

$$3 \quad 3 \quad 4 \quad 6 \quad 2 \quad 6 \quad 7 \quad 4 \quad 4 \quad 1 \quad 2 \quad 2 \quad 5 \quad 3 \quad 5 \quad 4 \quad 5 \quad 8 \quad 3 \quad 5 \quad 6 \quad 2 \quad 4 \quad 2 \quad 3$$

Aviso importante para os próximos exercícios: Os exercícios a seguir são apresentados em forma de problema. Ao resolvê-los certifique-se de que está realizando os itens (a)-(e) descritos a seguir.

(a) Apresente a amostra aleatória e a população (distribuição) a ser considerada nas análises. Apresente também o tamanho da amostra adotada e o(s) parâmetro(s) desconhecido(s).

(b) Formule as hipóteses a serem testadas a fim de verificar a questão levantada no problema.

(c) Defina um teste para verificar as hipóteses formuladas com o nível de significância escolhido. Para isso apresente a região crítica do teste e/ou a regra de decisão.

(d) A partir da amostra observada tome a decisão entre rejeitar ou não H_0.

(e) Formule uma frase resposta de forma que traga interpretação para a sua decisão do intem anterior.

6.5. ([Larson, 1982] - Capítulo 8) Durante um período de 2 meses, 13 mulheres seguiram uma mesma dieta. O total de peso perdido por cada uma, em quilograma, foi

3,9 4,3 5,6 5,6 4,1 6,5 3,7 5,9 4,3 3,7 4,4 5,9 5,0.

Assumindo que essa é uma amostra aleatória de uma população normal, você aceitaria a hipótese de que as mulheres que seguem esta dieta durante 2 meses perdem em média pelo menos 4,5 quilos, com nível de significância $\alpha = 0,1$?

6.6. ([Casella e Berger, 2002] - Capítulo 8) Em 1.000 lançamentos de uma moeda saíram 560 caras e 440 coroas. É razoável assumir que essa moeda é justa?

6.7. Uma pesquisa foi realizada com o objetivo de verificar a eficácia de uma certa vacina no combate a uma certa doença. Dois grupos de pessoas, o primeiro com 195.365 pessoas e o segundo com 154.568 pessoas, participaram do estudo. A vacina foi aplicada nas pessoas do segundo grupo, enquanto que as pessoas do primeiro grupo receberam um placebo. Um período de tempo após a vacinação o número de casos da doença observados nos dois grupos foram: 135 e 45, respectivamente. Qual a conclusão sobre a eficácia dessa vacina ao nível de significância de 10%?

6.8. Uma seguradora oferece diversos tipos de seguros, entre eles o seguro para *notebooks* para empresas de médio porte. O diretor de riscos desta seguradora quer reavaliar o valor do prêmio deste seguro. O prêmio é o valor que a seguradora cobra do cliente e o sinistro o valor que a seguradora paga ao cliente em caso do seguro ser acionado.

O valor do prêmio deste seguro foi determinado no passado com base na informação de que ocorriam em média 42 sinistros por ano por empresa. O diretor afirma que hoje existem, em média, mais sinistros por ano por empresa do que no passado, o que justificaria o aumento no valor do prêmio. Para comprovar essa afirmação o diretor contratou você, que vai ter acesso aos dados da seguradora, para fazer as análises estatísticas necessárias.

Para realizar este trabalho você selecionou, de forma aleatória, 103 empresas de médio porte do porrifólio da seguradora e registrou o número de sinistros de cada uma delas ao longo do último ano. Nessa análise você contabilizou um total de 4.617 sinistros no último ano entre as 103 empresas observadas.

Considere que o número de sinistros relacionados a *notebooks* em cada empresa de médio porte ao longo do ano são variáveis aleatórias independentes e identicamente distribuídas e que seguem o modelo de Poisson. Qual a resposta que você dá ao diretor da seguradora ao nível de significância de 5%?

Apêndice A

Tabelas Estatísticas

Tabela A.1: Probabilidade Acumulada da Distribuição Normal Padrão: $\Phi(z) = P(Z \leq z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Tabela A.2: Quantis da Distribuição Qui-quadrado

	$P(X \leq x)$												
df	0.005	0.01	0.025	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.975	0.99	0.995
1	0.0000393	0.000157	0.000982	0.00393	0.0158	0.1015	0.4549	1.3233	2.7055	3.8415	5.0239	6.6349	7.8794
2	0.0100	0.0201	0.0506	0.1026	0.2107	0.5754	1.3863	2.7726	4.6052	5.9915	7.3778	9.2103	10.5966
3	0.0717	0.1148	0.2158	0.3518	0.5844	1.2125	2.3660	4.1083	6.2514	7.8147	9.3484	11.3449	12.8382
4	0.2070	0.2971	0.4844	0.7107	1.0636	1.9226	3.3567	5.3853	7.7794	9.4877	11.1433	13.2767	14.8603
5	0.4117	0.5543	0.8312	1.1455	1.6103	2.6746	4.3515	6.6257	9.2364	11.0705	12.8325	15.0863	16.7496
6	0.6757	0.8721	1.2373	1.6354	2.2041	3.4546	5.3481	7.8408	10.6446	12.5916	14.4494	16.8119	18.5476
7	0.9893	1.2390	1.6899	2.1673	2.8331	4.2549	6.3458	9.0371	12.0170	14.0671	16.0128	18.4753	20.2777
8	1.3444	1.6465	2.1797	2.7326	3.4895	5.0706	7.3441	10.2189	13.3616	15.5073	17.5345	20.0902	21.9550
9	1.7349	2.0879	2.7004	3.3251	4.1682	5.8988	8.3428	11.3888	14.6837	16.9190	19.0228	21.6660	23.5894
10	2.1559	2.5582	3.2470	3.9403	4.8652	6.7372	9.3418	12.5489	15.9872	18.3070	20.4832	23.2093	25.1882
11	2.6032	3.0535	3.8157	4.5748	5.5778	7.5841	10.3410	13.7007	17.2750	19.6751	21.9200	24.7250	26.7568
12	3.0738	3.5706	4.4038	5.2260	6.3038	8.4384	11.3403	14.8454	18.5493	21.0261	23.3367	26.2170	28.2995
13	3.5650	4.1069	5.0088	5.8919	7.0415	9.2991	12.3398	15.9839	19.8119	22.3620	24.7356	27.6882	29.8195
14	4.0747	4.6604	5.6287	6.5706	7.7895	10.1653	13.3393	17.1169	21.0641	23.6848	26.1189	29.1412	31.3193
15	4.6009	5.2293	6.2621	7.2609	8.5468	11.0365	14.3389	18.2451	22.3071	24.9958	27.4884	30.5779	32.8013
16	5.1422	5.8122	6.9077	7.9616	9.3122	11.9122	15.3385	19.3689	23.5418	26.2962	28.8454	31.9999	34.2672
17	5.6972	6.4078	7.5642	8.6718	10.0852	12.7919	16.3382	20.4887	24.7690	27.5871	30.1910	33.4087	35.7185
18	6.2648	7.0149	8.2307	9.3905	10.8649	13.6753	17.3379	21.6049	25.9894	28.8693	31.5264	34.8053	37.1565
19	6.8440	7.6327	8.9065	10.1170	11.6509	14.5620	18.3377	22.7178	27.2036	30.1435	32.8523	36.1909	38.5823
20	7.4338	8.2604	9.5908	10.8508	12.4426	15.4518	19.3374	23.8277	28.4120	31.4104	34.1696	37.5662	39.9968
21	8.0337	8.8972	10.2829	11.5913	13.2396	16.3444	20.3372	24.9348	29.6151	32.6706	35.4789	38.9322	41.4011
22	8.6427	9.5425	10.9823	12.3380	14.0415	17.2396	21.3370	26.0393	30.8133	33.9244	36.7807	40.2894	42.7957
23	9.2604	10.1957	11.6886	13.0905	14.8480	18.1373	22.3369	27.1413	32.0069	35.1725	38.0756	41.6384	44.1813
24	9.8862	10.8564	12.4012	13.8484	15.6587	19.0373	23.3367	28.2412	33.1962	36.4150	39.3641	42.9798	45.5585
25	10.5197	11.5240	13.1197	14.6114	16.4734	19.9393	24.3366	29.3389	34.3816	37.6525	40.6465	44.3141	46.9279
26	11.1602	12.1981	13.8439	15.3792	17.2919	20.8434	25.3365	30.4346	35.5632	38.8851	41.9232	45.6417	48.2899
27	11.8076	12.8785	14.5734	16.1514	18.1139	21.7494	26.3363	31.5284	36.7412	40.1133	43.1945	46.9629	49.6449
28	12.4613	13.5647	15.3079	16.9279	18.9392	22.6572	27.3362	32.6205	37.9159	41.3371	44.4608	48.2782	50.9934
29	13.1211	14.2565	16.0471	17.7084	19.7677	23.5666	28.3361	33.7109	39.0875	42.5570	45.7223	49.5879	52.3356
30	13.7867	14.9535	16.7908	18.4927	20.5992	24.4776	29.3360	34.7997	40.2560	43.7730	46.9792	50.8922	53.6720

Tabela A.3: Quantis da Distribuição Qui-quadrado (continuação)

	$P(X \le x)$												
df	0.005	0.01	0.025	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.975	0.99	0.995
31	14.4578	15.6555	17.5387	19.2806	21.4336	25.3901	30.3359	35.8871	41.4217	44.9853	48.2319	52.1914	55.0027
32	15.134	16.3622	18.2908	20.0719	22.2706	26.3041	31.3359	36.973	42.5847	46.1943	49.4804	53.4858	56.3281
33	15.8153	17.0735	19.0467	20.8665	23.1102	27.2194	32.3358	38.0575	43.7452	47.3999	50.7251	54.7755	57.6484
34	16.5013	17.7891	19.8063	21.6643	23.9523	28.1361	33.3357	39.1408	44.9032	48.6024	51.966	56.0609	58.9639
35	17.1918	18.5089	20.5694	22.465	24.7967	29.054	34.3356	40.2228	46.0588	49.8018	53.2033	57.3421	60.2748
36	17.8867	19.2327	21.3359	23.2686	25.6433	29.973	35.3356	41.3036	47.2122	50.9985	54.4373	58.6192	61.5812
37	18.5858	19.9602	22.1056	24.0749	26.4921	30.8933	36.3355	42.3833	48.3634	52.1923	55.668	59.8925	62.8833
38	19.2889	20.6914	22.8785	24.8839	27.343	31.8146	37.3355	43.4619	49.5126	53.3835	56.8955	61.1621	64.1814
39	19.9959	21.4262	23.6543	25.6954	28.1958	32.7369	38.3354	44.5395	50.6598	54.5722	58.1201	62.4281	65.4756
40	20.7065	22.1643	24.433	26.5093	29.0505	33.6603	39.3353	45.616	51.8051	55.7585	59.3417	63.6907	66.766
41	21.4208	22.9056	25.2145	27.3256	29.9071	34.5846	40.3353	46.6916	52.9485	56.9424	60.5606	64.9501	68.0527
42	22.1385	23.6501	25.9987	28.144	30.7654	35.5099	41.3352	47.7663	54.0902	58.124	61.7768	66.2062	69.336
43	22.8595	24.3976	26.7854	28.9647	31.6255	36.4361	42.3352	48.84	55.2302	59.3035	62.9904	67.4593	70.6159
44	23.5837	25.148	27.5746	29.7875	32.4871	37.3631	43.3352	49.9129	56.3685	60.4809	64.2015	68.7095	71.8926
45	24.311	25.9013	28.3662	30.6123	33.3504	38.291	44.3351	50.9849	57.5053	61.6562	65.4102	69.9568	73.1661
46	25.0413	26.6572	29.1601	31.439	34.2152	39.2197	45.3351	52.0562	58.6405	62.8296	66.6165	71.2014	74.4365
47	25.7746	27.4158	29.9562	32.2676	35.0814	40.1492	46.335	53.1267	59.7743	64.0011	67.8206	72.4433	75.7041
48	26.5106	28.177	30.7545	33.0981	35.9491	41.0794	47.335	54.1964	60.9066	65.1708	69.0226	73.6826	76.9688
49	27.2493	28.9406	31.5549	33.9303	36.8182	42.0104	48.335	55.2653	62.0375	66.3386	70.2224	74.9195	78.2307
50	27.9907	29.7067	32.3574	34.7643	37.6886	42.9421	49.3349	56.3336	63.1671	67.5048	71.4202	76.1539	79.49
51	28.7347	30.475	33.1618	35.5999	38.5604	43.8745	50.3349	57.4012	64.2954	68.6693	72.616	77.386	80.7467
52	29.4812	31.2457	33.9681	36.4371	39.4334	44.8075	51.3349	58.4681	65.4224	69.8322	73.8099	78.6158	82.0008
53	30.23	32.0185	34.7763	37.2759	40.3076	45.7412	52.3348	59.5344	66.5482	70.9935	75.0019	79.8433	83.2526
54	30.9813	32.7934	35.5863	38.1162	41.183	46.6755	53.3348	60.6	67.6728	72.1532	76.192	81.0688	84.5019
55	31.7348	33.5705	36.3981	38.958	42.0596	47.6105	54.3348	61.665	68.7962	73.3115	77.3805	82.2921	85.749
56	32.4905	34.3495	37.2116	39.8013	42.9373	48.546	55.3348	62.7294	69.9185	74.4683	78.5672	83.5134	86.9938
57	33.2484	35.1305	38.0267	40.6459	43.8161	49.4821	56.3347	63.7933	71.0397	75.6237	79.7522	84.7328	88.2364
58	34.0084	35.9135	38.8435	41.492	44.696	50.4188	57.3347	64.8565	72.1598	76.7778	80.9356	85.9502	89.4769
59	34.7704	36.6982	39.6619	42.3393	45.577	51.356	58.3347	65.9193	73.2789	77.9305	82.1174	87.1657	90.7153
60	35.5345	37.4849	40.4817	43.188	46.4589	52.2938	59.3347	66.9815	74.397	79.0819	83.2977	88.3794	91.9517

Tabela A.4: Quantis da Distribuição t-Student

	$P(T \leq t)$							
df	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995
1	1.000000	1.376382	1.962611	3.077684	6.313752	12.70621	31.82052	63.65674
2	0.816497	1.060660	1.386207	1.885618	2.919986	4.302653	6.964557	9.924843
3	0.764892	0.978472	1.249778	1.637744	2.353363	3.182446	4.540703	5.840909
4	0.740697	0.940965	1.189567	1.533206	2.131847	2.776445	3.746947	4.604095
5	0.726687	0.919544	1.155767	1.475884	2.015048	2.570582	3.364930	4.032143
6	0.717558	0.905703	1.134157	1.439756	1.943180	2.446912	3.142668	3.707428
7	0.711142	0.896030	1.119159	1.414924	1.894579	2.364624	2.997952	3.499483
8	0.706387	0.888890	1.108145	1.396815	1.859548	2.306004	2.896459	3.355387
9	0.702722	0.883404	1.099716	1.383029	1.833113	2.262157	2.821438	3.249836
10	0.699812	0.879058	1.093058	1.372184	1.812461	2.228139	2.763769	3.169273
11	0.697445	0.875530	1.087666	1.363430	1.795885	2.200985	2.718079	3.105807
12	0.695483	0.872609	1.083211	1.356217	1.782288	2.178813	2.680998	3.054540
13	0.693829	0.870152	1.079469	1.350171	1.770933	2.160369	2.650309	3.012276
14	0.692417	0.868055	1.076280	1.345030	1.761310	2.144787	2.624494	2.976843
15	0.691197	0.866245	1.073531	1.340606	1.753050	2.131450	2.602480	2.946713
16	0.690132	0.864667	1.071137	1.336757	1.745884	2.119905	2.583487	2.920782
17	0.689195	0.863279	1.069033	1.333379	1.739607	2.109816	2.566934	2.898231
18	0.688364	0.862049	1.067170	1.330391	1.734064	2.100922	2.552380	2.878440
19	0.687621	0.860951	1.065507	1.327728	1.729133	2.093024	2.539483	2.860935
20	0.686954	0.859964	1.064016	1.325341	1.724718	2.085963	2.527977	2.845340
21	0.686352	0.859074	1.062670	1.323188	1.720743	2.079614	2.517648	2.831360
22	0.685805	0.858266	1.061449	1.321237	1.717144	2.073873	2.508325	2.818756
23	0.685306	0.857530	1.060337	1.319460	1.713872	2.068658	2.499867	2.807336
24	0.684850	0.856855	1.059319	1.317836	1.710882	2.063899	2.492159	2.796940
25	0.684430	0.856236	1.058384	1.316345	1.708141	2.059539	2.485107	2.787436
26	0.684043	0.855665	1.057523	1.314972	1.705618	2.055529	2.478630	2.778715
27	0.683685	0.855137	1.056727	1.313703	1.703288	2.051831	2.472660	2.770683
28	0.683353	0.854647	1.055989	1.312527	1.701131	2.048407	2.467140	2.763262
29	0.683044	0.854192	1.055302	1.311434	1.699127	2.045230	2.462021	2.756386
30	0.682756	0.853767	1.054662	1.310415	1.697261	2.042272	2.457262	2.749996
31	0.682486	0.853370	1.054064	1.309464	1.695519	2.039513	2.452824	2.744042
32	0.682234	0.852998	1.053504	1.308573	1.693889	2.036933	2.448678	2.738481
33	0.681997	0.852649	1.052979	1.307737	1.692360	2.034515	2.444794	2.733277
34	0.681774	0.852321	1.052485	1.306952	1.690924	2.032245	2.441150	2.728394
35	0.681564	0.852012	1.052019	1.306212	1.689572	2.030108	2.437723	2.723806
36	0.681366	0.851720	1.051580	1.305514	1.688298	2.028094	2.434494	2.719485
37	0.681178	0.851444	1.051165	1.304854	1.687094	2.026192	2.431447	2.715409
38	0.681001	0.851183	1.050772	1.304230	1.685954	2.024394	2.428568	2.711558
39	0.680833	0.850935	1.050399	1.303639	1.684875	2.022691	2.425841	2.707913
40	0.680673	0.850700	1.050046	1.303077	1.683851	2.021075	2.423257	2.704459
45	0.679981	0.849682	1.048516	1.300649	1.679427	2.014103	2.412116	2.689585
50	0.679428	0.848869	1.047295	1.298714	1.675905	2.008559	2.403272	2.677793
55	0.678977	0.848205	1.046298	1.297134	1.673034	2.004045	2.396081	2.668216
60	0.678601	0.847653	1.045469	1.295821	1.670649	2.000298	2.390119	2.660283
65	0.678283	0.847186	1.044768	1.294712	1.668636	1.997138	2.385097	2.653604
70	0.678011	0.846786	1.044169	1.293763	1.666914	1.994437	2.380807	2.647905
80	0.677569	0.846137	1.043195	1.292224	1.664125	1.990063	2.373868	2.638691
90	0.677225	0.845633	1.042440	1.291029	1.661961	1.986675	2.368497	2.631565
100	0.676951	0.845230	1.041836	1.290075	1.660234	1.983972	2.364217	2.625891
∞	0.674490	0.841621	1.036433	1.281552	1.644854	1.959964	2.326348	2.575829

Tabela A.5: Quantis da Distribuição F-Snedecor

		n													
$\mathrm{P}(F_{n,m} < f)$	m	1	2	3	4	5	6	7	8	9	10	15	20	30	40
0.500	1	1.0000	1.5000	1.7092	1.8227	1.8937	1.9422	1.9774	2.0041	2.0250	2.0419	2.0931	2.1191	2.1452	2.1584
0.750		5.8284	7.5000	8.1999	8.5809	8.8198	8.9833	9.1021	9.1923	9.2631	9.3201	9.4934	9.5813	9.6698	9.7144
0.900		39.863	49.500	53.593	55.833	57.240	58.204	58.906	59.439	59.858	60.195	61.220	61.740	62.265	62.529
0.950		161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	245.95	248.01	250.10	251.14
0.975		647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28	968.63	984.87	993.10	1001.4	1005.6
0.990		4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1	6022.5	6055.8	6157.3	6208.7	6260.6	6286.8
0.500	2	0.6667	1.0000	1.1349	1.2071	1.2519	1.2824	1.3046	1.3213	1.3344	1.3450	1.3771	1.3933	1.4096	1.4178
0.750		2.5714	3.0000	3.1534	3.2321	3.2799	3.3121	3.3352	3.3526	3.3661	3.3770	3.4098	3.4263	3.4428	3.4511
0.900		8.5263	9.0000	9.1618	9.2434	9.2926	9.3255	9.3491	9.3668	9.3805	9.3916	9.4247	9.4413	9.4579	9.4662
0.950		18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385	19.396	19.429	19.446	19.462	19.471
0.975		38.506	39.000	39.165	39.248	39.298	39.331	39.355	39.373	39.387	39.398	39.431	39.448	39.465	39.473
0.990		98.503	99.000	99.166	99.249	99.299	99.333	99.356	99.374	99.388	99.399	99.433	99.449	99.466	99.474
0.500	3	0.5851	0.8811	1.0000	1.0632	1.1024	1.1289	1.1482	1.1627	1.1741	1.1833	1.2111	1.2252	1.2393	1.2465
0.750		2.0239	2.2798	2.3556	2.3901	2.4095	2.4218	2.4302	2.4364	2.4410	2.4447	2.4552	2.4602	2.4650	2.4674
0.900		5.5383	5.4624	5.3908	5.3426	5.3092	5.2847	5.2662	5.2517	5.2400	5.2304	5.2003	5.1845	5.1681	5.1597
0.950		10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7029	8.6602	8.6166	8.5944
0.975		17.443	16.044	15.439	15.101	14.885	14.735	14.624	14.540	14.473	14.419	14.253	14.167	14.081	14.037
0.990		34.116	30.817	29.457	28.710	28.237	27.911	27.672	27.489	27.345	27.229	26.872	26.690	26.505	26.411
0.500	4	0.5486	0.8284	0.9405	1.0000	1.0367	1.0617	1.0797	1.0933	1.1040	1.1126	1.1386	1.1517	1.1649	1.1716
0.750		1.8074	2.0000	2.0467	2.0642	2.0723	2.0766	2.0790	2.0805	2.0814	2.0820	2.0829	2.0828	2.0825	2.0821
0.900		4.5448	4.3246	4.1909	4.1072	4.0506	4.0097	3.9790	3.9549	3.9357	3.9199	3.8704	3.8443	3.8174	3.8036
0.950		7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.8578	5.8025	5.7459	5.7170
0.975		12.218	10.649	9.9792	9.6045	9.3645	9.1973	9.0741	8.9796	8.9047	8.8439	8.6565	8.5599	8.4613	8.4111
0.990		21.198	18.000	16.694	15.977	15.522	15.207	14.976	14.799	14.659	14.546	14.198	14.020	13.838	13.745
0.500	5	0.5281	0.7988	0.9071	0.9646	1.0000	1.0240	1.0414	1.0545	1.0648	1.0730	1.0980	1.1106	1.1234	1.1297
0.750		1.6925	1.8528	1.8843	1.8927	1.8947	1.8945	1.8935	1.8923	1.8911	1.8899	1.8851	1.8820	1.8784	1.8763
0.900		4.0604	3.7797	3.6195	3.5202	3.4530	3.4045	3.3679	3.3393	3.3163	3.2974	3.238	3.2067	3.1741	3.1573
0.950		6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6188	4.5581	4.4957	4.4638
0.975		10.007	8.4336	7.7636	7.3879	7.1464	6.9777	6.8531	6.7572	6.6811	6.6192	6.4277	6.3286	6.2269	6.1750
0.990		16.258	13.274	12.060	11.392	10.967	10.672	10.456	10.289	10.158	10.051	9.7222	9.5526	9.3793	9.2912

Tabela A.6: Quantis da Distribuição F-Snedecor (continuação)

		n													
$P(F_{n,m} < f)$	m	1	2	3	4	5	6	7	8	9	10	15	20	30	40
0.500	6	0.5149	0.7798	0.8858	0.9419	0.9765	1.0000	1.0169	1.0298	1.0398	1.0478	1.0722	1.0845	1.0969	1.1031
0.750		1.6214	1.7622	1.7844	1.7872	1.7852	1.7821	1.7789	1.776	1.7733	1.7708	1.7621	1.7569	1.7509	1.7477
0.900		3.7759	3.4633	3.2888	3.1808	3.1075	3.0546	3.0145	2.9830	2.9577	2.9369	2.8712	2.8363	2.8000	2.7812
0.950		5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	3.9381	3.8742	3.8082	3.7743
0.975		8.8131	7.2599	6.5988	6.2272	5.9876	5.8198	5.6955	5.5996	5.5234	5.4613	5.2687	5.1684	5.0652	5.0125
0.990		13.745	10.925	9.7795	9.1483	8.7459	8.4661	8.2600	8.1017	7.9761	7.8741	7.5590	7.3958	7.2285	7.1432
0.500	7	0.5057	0.7665	0.8709	0.9262	0.9603	0.9833	1.000	1.0126	1.0224	1.0304	1.0543	1.0664	1.0785	1.0846
0.750		1.5732	1.7010	1.7169	1.7157	1.7111	1.7059	1.7011	1.6969	1.6931	1.6898	1.6781	1.6712	1.6635	1.6593
0.900		3.5894	3.2574	3.0741	2.9605	2.8833	2.8274	2.7849	2.7516	2.7247	2.7025	2.6322	2.5947	2.5555	2.5351
0.950		5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.5107	3.4445	3.3758	3.3404
0.975		8.0727	6.5415	5.8898	5.5226	5.2852	5.1186	4.9949	4.8993	4.8232	4.7611	4.5678	4.4667	4.3624	4.3089
0.990		12.246	9.5466	8.4513	7.8466	7.4604	7.1914	6.9928	6.8400	6.7188	6.6201	6.3143	6.1554	5.9920	5.9084
0.500	8	0.4990	0.7568	0.8600	0.9146	0.9483	0.9711	0.9876	1.0000	1.0097	1.0175	1.0412	1.0531	1.0651	1.0711
0.750		1.5384	1.6569	1.6683	1.6642	1.6575	1.6508	1.6448	1.6396	1.635	1.6310	1.6170	1.6088	1.5996	1.5945
0.900		3.4579	3.1131	2.9238	2.8064	2.7264	2.6683	2.6241	2.5893	2.5612	2.5380	2.4642	2.4246	2.3830	2.3614
0.950		5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2184	3.1503	3.0794	3.0428
0.975		7.5709	6.0595	5.4160	5.0526	4.8173	4.6517	4.5286	4.4333	4.3572	4.2951	4.1012	3.9995	3.894	3.8398
0.990		11.259	8.6491	7.5910	7.0061	6.6318	6.3707	6.1776	6.0289	5.9106	5.8143	5.5151	5.3591	5.1981	5.1156
0.500	9	0.4938	0.7494	0.8517	0.9058	0.9391	0.9617	0.9780	0.9904	1.0000	1.0077	1.0311	1.0429	1.0548	1.0608
0.750		1.5121	1.6236	1.6315	1.6253	1.6170	1.6091	1.6022	1.5961	1.5909	1.5863	1.5705	1.5611	1.5506	1.5449
0.900		3.3603	3.0065	2.8129	2.6927	2.6106	2.5509	2.5053	2.4694	2.4403	2.4163	2.3396	2.2983	2.2547	2.2320
0.950		5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0061	2.9365	2.8637	2.8259
0.975		7.2093	5.7147	5.0781	4.7181	4.4844	4.3197	4.1970	4.1020	4.0260	3.9639	3.7694	3.6669	3.5604	3.5055
0.990		10.561	8.0215	6.9919	6.4221	6.0569	5.8018	5.6129	5.4671	5.3511	5.2565	4.9621	4.8080	4.6486	4.5666
0.500	10	0.4897	0.7435	0.8451	0.8988	0.9319	0.9544	0.9705	0.9828	0.9923	1.0000	1.0232	1.0349	1.0467	1.0526
0.750		1.4915	1.5975	1.6028	1.5949	1.5853	1.5765	1.5688	1.5621	1.5563	1.5513	1.5338	1.5235	1.5119	1.5056
0.900		3.2850	2.9245	2.7277	2.6053	2.5216	2.4606	2.4140	2.3772	2.3473	2.3226	2.2435	2.2007	2.1554	2.1317
0.950		4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.8450	2.7740	2.6996	2.6609
0.975		6.9367	5.4564	4.8256	4.4683	4.2361	4.0721	3.9498	3.8549	3.7790	3.7168	3.5217	3.4185	3.3110	3.2554
0.990		10.044	7.5594	6.5523	5.9943	5.6363	5.3858	5.2001	5.0567	4.9424	4.8491	4.5581	4.4054	4.2469	4.1653

Tabela A.7: Quantis da Distribuição F-Snedecor (continuação)

		n													
$P(F_{n,m} < f)$	m	1	2	3	4	5	6	7	8	9	10	15	20	30	40
0.500	15	0.4777	0.7262	0.8257	0.8783	0.9107	0.9327	0.9485	0.9605	0.9698	0.9773	1.0000	1.0114	1.0229	1.0287
0.750		1.4321	1.5227	1.5202	1.5071	1.4938	1.4820	1.4718	1.4631	1.4556	1.4491	1.4263	1.4127	1.3973	1.3888
0.900		3.0732	2.6952	2.4898	2.3614	2.2730	2.2081	2.1582	2.1185	2.0862	2.0593	1.9722	1.9243	1.8728	1.8454
0.950		4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4034	2.3275	2.2468	2.2043
0.975		6.1995	4.7650	4.1528	3.8043	3.5764	3.4147	3.2934	3.1987	3.1227	3.0602	2.8621	2.7559	2.6437	2.5850
0.990		8.6831	6.3589	5.4170	4.8932	4.5556	4.3183	4.1415	4.0045	3.8948	3.8049	3.5222	3.3719	3.2141	3.1319
0.500	20	0.4719	0.7177	0.8162	0.8683	0.9004	0.9221	0.9378	0.9496	0.9589	0.9663	0.9887	1.0000	1.0114	1.0171
0.750		1.4037	1.4870	1.4808	1.4652	1.4500	1.4366	1.4252	1.4153	1.4069	1.3995	1.3736	1.3580	1.3401	1.3301
0.900		2.9747	2.5893	2.3801	2.2489	2.1582	2.0913	2.0397	1.9985	1.9649	1.9367	1.8449	1.7938	1.7382	1.7083
0.950		4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.2033	2.1242	2.0391	1.9938
0.975		5.8715	4.4613	3.8587	3.5147	3.2891	3.1283	3.0074	2.9128	2.8365	2.7737	2.5731	2.4645	2.3486	2.2873
0.990		8.0960	5.8489	4.9382	4.4307	4.1027	3.8714	3.6987	3.5644	3.4567	3.3682	3.0880	2.9377	2.7785	2.6947
0.500	30	0.4662	0.7094	0.8069	0.8584	0.8902	0.9117	0.9272	0.9389	0.9480	0.9554	0.9776	0.9888	1.0000	1.0056
0.750		1.3761	1.4524	1.4426	1.4244	1.4073	1.3923	1.3795	1.3685	1.3590	1.3507	1.3213	1.3033	1.2823	1.2703
0.900		2.8807	2.4887	2.2761	2.1422	2.0492	1.9803	1.9269	1.8841	1.8490	1.8195	1.7223	1.6673	1.6065	1.5732
0.950		4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.0148	1.9317	1.8409	1.7918
0.975		5.5675	4.1821	3.5894	3.2499	3.0265	2.8667	2.7460	2.6513	2.5746	2.5112	2.3072	2.1952	2.0739	2.0089
0.990		7.5625	5.3903	4.5097	4.0179	3.6990	3.4735	3.3045	3.1726	3.0665	2.9791	2.7002	2.5487	2.3860	2.2992
0.500	40	0.4633	0.7053	0.8023	0.8536	0.8852	0.9065	0.9220	0.9336	0.9427	0.9500	0.9721	0.9832	0.9944	1.0000
0.750		1.3626	1.4355	1.4239	1.4045	1.3863	1.3706	1.3571	1.3455	1.3354	1.3266	1.2952	1.2758	1.2529	1.2397
0.900		2.8354	2.4404	2.2261	2.0909	1.9968	1.9269	1.8725	1.8289	1.7929	1.7627	1.6624	1.6052	1.5411	1.5056
0.950		4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772	1.9245	1.8389	1.7444	1.6928
0.975		5.4239	4.0510	3.4633	3.1261	2.9037	2.7444	2.6238	2.5289	2.4519	2.3882	2.1819	2.0677	1.9429	1.8752
0.990		7.3141	5.1785	4.3126	3.8283	3.5138	3.2910	3.1238	2.9930	2.8876	2.8005	2.5216	2.3689	2.2034	2.1142

Apêndice B

Algumas Famílias de Distribuições

B.1 Distribuições Discretas

Bernoulli

Notação:	$X \sim Ber(p)$ ou $X \sim Bernoulli(p)$.
Parâmetro(s):	$0 \leq p \leq 1$
Função de probabilidade:	$p_X(x) = p^x(1-p)^{1-x}, \quad x = 0$ ou $x = 1$.
Função de distribuição acumulada:	$F_X(x) = \begin{cases} 0 & \text{, se } x < 0 \\ 1-p & \text{, se } 0 \leq x < 1 \\ 1 & \text{, se } x \geq 1 \end{cases}$
Esperança e Variância:	$\mathrm{E}(X) = p$ e $\mathrm{Var}(X) = p(1-p)$.
Função Geradora de Momentos:	$M_X(x) = 1 - p + e^t p$.
Estimador de Máx. Verossimilhança:	$\hat{p} = \bar{X}$.
Informação de Fisher:	$I(p) = \dfrac{1}{p(1-p)}$.
Estatística Suficiente para p:	$\sum_{i=1}^n X_i \sim Binom(n, p)$
Observações:	• Se $X \sim Ber(p)$, $X \sim Binom(n = 1, p)$.
	• Se $X \sim Ber(p)$, $1 - X \sim Ber(1 - p)$.

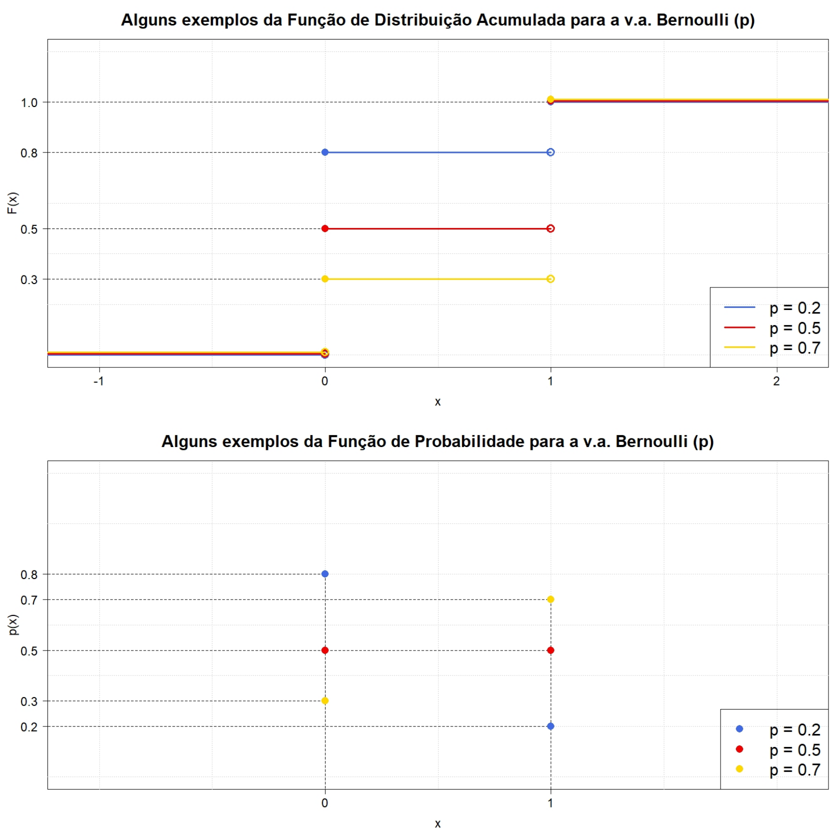


Figura B.1: Função de Distribuição Acumulada e de Probabilidade da Bernoulli

Binomial

Notação:	$X \sim Binom(N, p)$
Parâmetro(s):	$N \in \mathbb{N}^*$ e $0 \leq p \leq 1$.
Função de probabilidade:	$p_X(x) = \binom{N}{x} p^x (1-p)^{N-x}, \quad x = 0, 1, \ldots, N.$
Esperança e Variância:	$\mathrm{E}(X) = Np$ e $\mathrm{Var}(X) = Np(1-p)$.
Função Geradora de Momentos:	$M_X(x) = (pe^t + 1 - p)^N$.
Estimador de Máx. Verossimilhança:	$\hat{p} = \bar{X}/N$, supondo N conhecido.
Informação de Fisher:	$I(p) = \dfrac{N}{p(1-p)}$, supondo N conhecido.
Estatística Suficiente para p:	$\sum_{i=1}^n X_i \sim Binom(nN, p)$, supondo N conhecido.

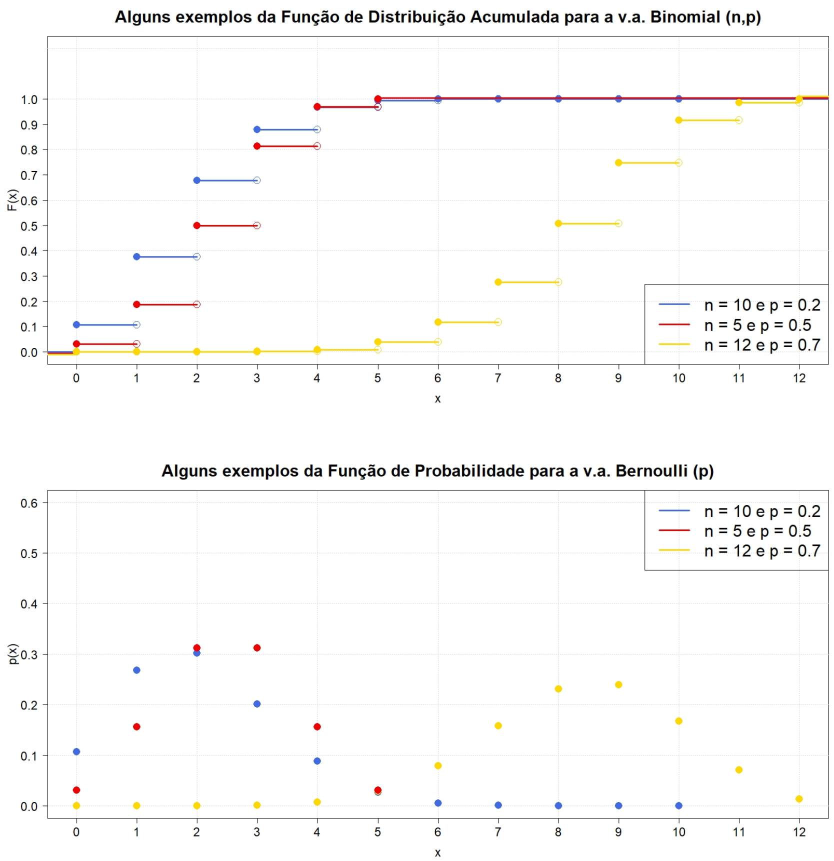
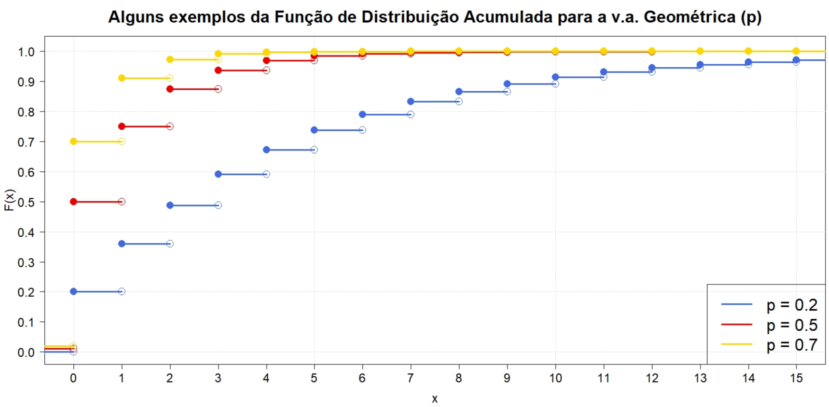


Figura B.2: Função de Distribuição Acumulada e de Probabilidade da Binomial

Geométrica

Notação:	$X \sim Geo(p)$
Parâmetro(s):	$0 \leq p \leq 1$
Função de probabilidade:	$p_X(x) = (1-p)^x p, \quad x = 0, 1, 2, \dots$
Função de distribuição acumulada:	$F(x) = \begin{cases} 0 & , \text{ se } x < 0 \\ 1-(1-p)^{\lfloor x \rfloor + 1} & , \text{ se } x \geq 0 \end{cases}$
Esperança e Variância:	$\mathrm{E}(X) = \dfrac{1-p}{p}$ e $\mathrm{Var}(X) = \dfrac{1-p}{p^2}$
Função Geradora de Momentos:	$M_X(x) = \dfrac{p}{1-e^t(1-p)}, \ t < -\ln(1-p)$
Estimador de Máx. Verossimilhança:	$\hat{p} = 1/(\bar{X}+1)$.
Informação de Fisher:	$I(p) = \dfrac{1}{p^2(1-p)}$.
Estatística Suficiente para p:	$\sum_{i=1}^{n} X_i \sim BinNeg(n,p)$.
Observações:	• X aqui definida indica o número de fracassos até o primeiro sucesso. • Se $X \sim Geo(p)$, $X+1 \sim Gem^*(p)$.



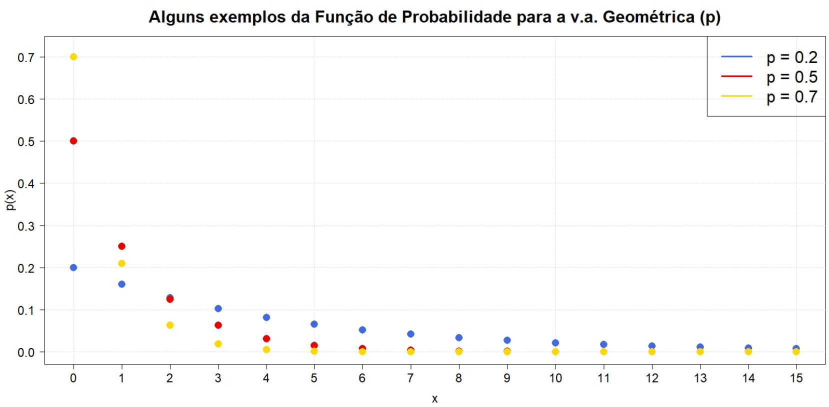
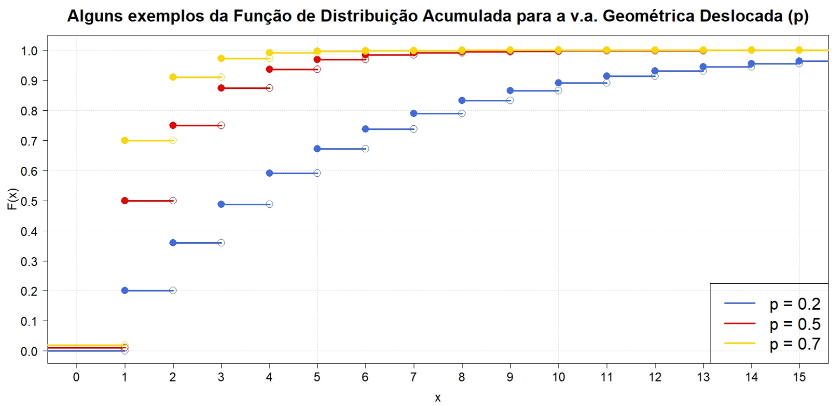


Figura B.3: Função de Distribuição Acumulada e de Probabilidade da Geométrica

Geométrica Deslocada

Notação:	$X \sim Geo^*(p)$
Parâmetro(s):	$0 \leq p \leq 1$
Função de probabilidade:	$p_X(x) = (1-p)^{x-1}p, \quad x = 1, 2, 3, \ldots$
Função de distribuição acumulada:	$F(x) = \begin{cases} 0 & , \text{ se } x < 1 \\ 1-(1-p)^{\lfloor x \rfloor} & , \text{ se } x \geq 1 \end{cases}$
Esperança e Variância:	$\mathrm{E}(X) = \dfrac{1}{p}$ e $\mathrm{Var}(X) = \dfrac{1-p}{p^2}$
Função Geradora de Momentos:	$M_X(x) = \dfrac{pe^t}{1-e^t(1-p)}, \ t < -\ln(1-p)$
Estimador de Máx. Verossimilhança:	$\hat{p} = 1/\bar{X}$
Informação de Fisher:	$I(p) = 1/\left(p^2(1-p)\right)$
Estatística Suficiente para p:	$\sum_{i=1}^n X_i \sim BinNeg^*(n,p)$
Observações:	• X aqui definida indica o número de tentativas até o primeiro sucesso. • Se $X \sim Geo^*(p)$, $X - 1 \sim Gem(p)$.



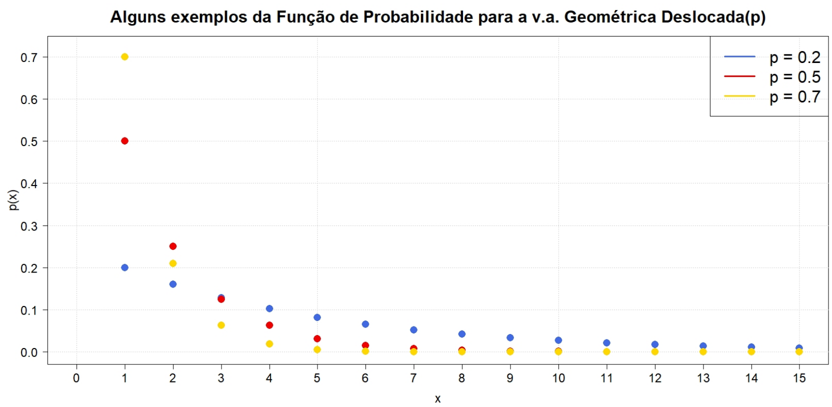


Figura B.4: Função de Distribuição Acumulada e de Probabilidade da Geométrica Deslocada

Binomial Negativa

Notação:	$X \sim BinNeg(r,p)$
Parâmetro(s):	$r \in \mathbb{N}^\star = \{1,2,3,\ldots\}$ e $0 \leq p \leq 1$
Função de probabilidade:	$p_X(x) = \binom{x+r-1}{x}(1-p)^x p^r, \quad x = 0,1,2,\ldots$
Esperança e Variância:	$\mathrm{E}(X) = \dfrac{r(1-p)}{p}$ e $\mathrm{Var}(X) = \dfrac{r(1-p)}{p^2}$
Função Geradora de Momentos:	$M_X(x) = \left(\dfrac{p}{1-e^t(1-p)}\right)^r, \; t < -\ln(1-p)$
Estimador de Máx. Verossimilhança:	$\hat{p} = r/(\bar{X}+r)$, supondo r conhecido.
Informação de Fisher:	$I(p) = \dfrac{r}{p^2(1-p)}$, supondo r conhecido..
Estatística Suficiente para p:	$\sum_{i=1}^n X_i \sim BinNeg(nr,p)$, supondo r conhecido.
Observações:	• X aqui definida indica o número de fracassos até o r-ésimo sucesso. • Se $X \sim BinNeg(r,p)(p)$, $X + r \sim BinNeg^*(r,p)$.

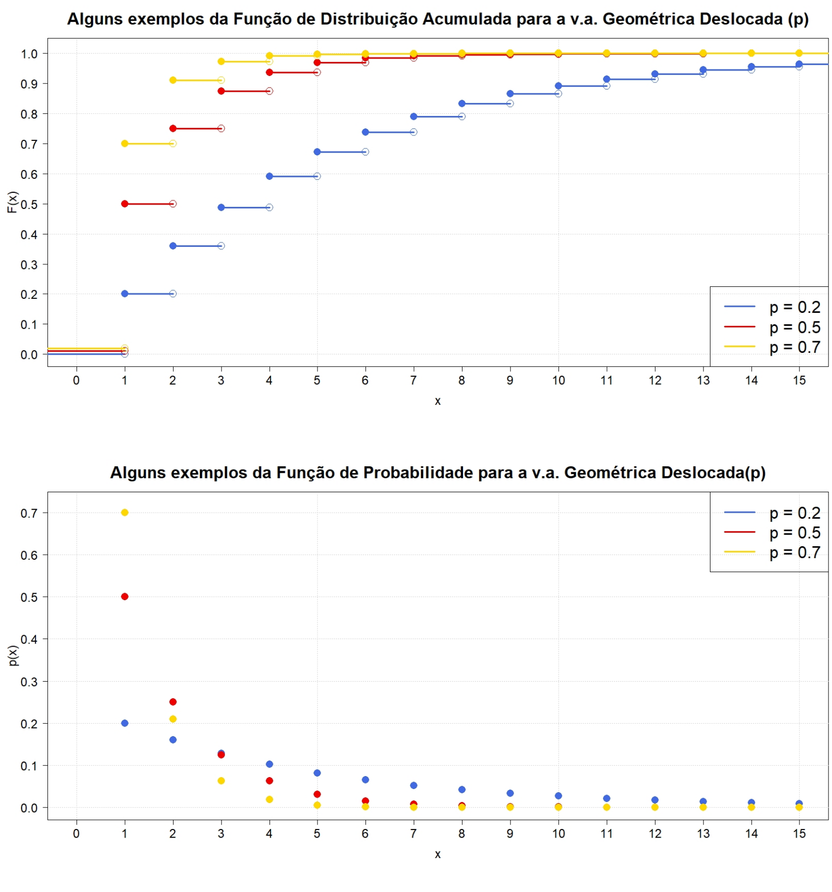
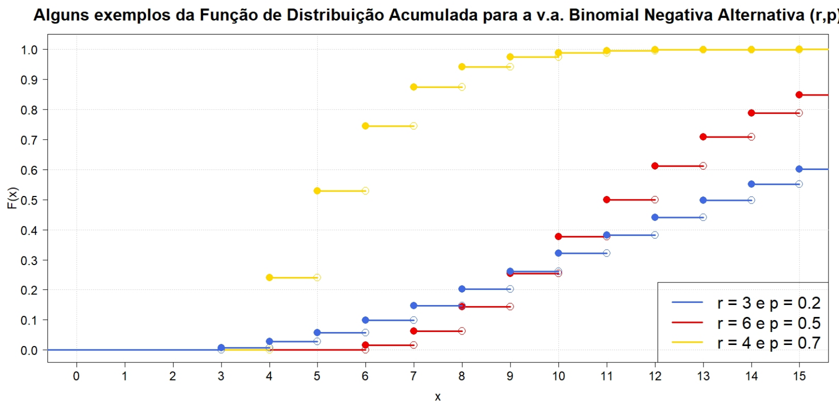


Figura B.5: Função de Distribuição Acumulada e de Probabilidade da Geométrica Deslocada

Binomial Negativa Alternativa (ou distribuição de Pascal)

Notação:	$X \sim BinNeg^*(r, p)$
Parâmetro(s):	$r \in \mathbb{N}^\star = \{1, 2, 3, \ldots\}$ e $0 \leq p \leq 1$
Função de probabilidade:	$p_X(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r, \quad x = r, r+1, r+2, \ldots$
Esperança e Variância:	$\mathrm{E}(X) = \dfrac{r}{p}$ e $\mathrm{Var}(X) = \dfrac{r(1-p)}{p^2}$
Função Geradora de Momentos:	$M_X(x) = \left(\dfrac{pe^t}{1-e^t(1-p)}\right)^r, \; t < -\ln(1-p)$
Estimador de Máx. Verossimilhança:	$\hat{p} = r/\bar{X}$, supondo r conhecido.
Informação de Fisher:	$I(p) = \dfrac{r}{p^2(1-p)}$, supondo r conhecido.
Estatística Suficiente para p:	$\sum_{i=1}^n X_i \sim BinNeg^*(nr, p)$, supondo r conhecido.
Observações:	• X aqui definida indica o número de tentativas até r-ésimo sucesso. • Se $X \sim BinNeg^*(r, p)$, $X - r \sim BinNeg(r, p)$.



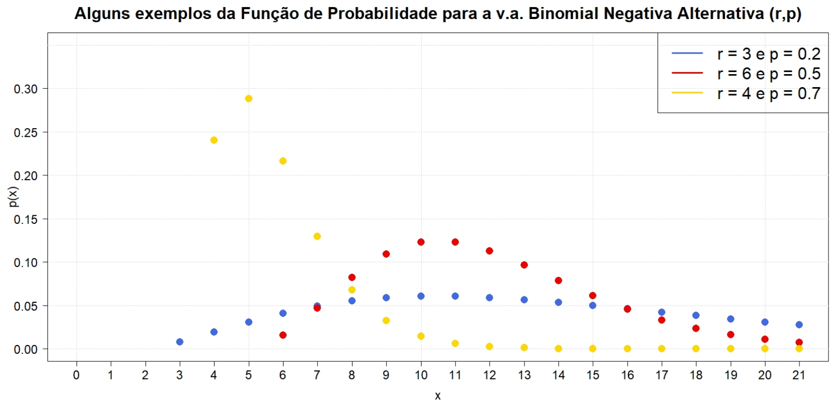
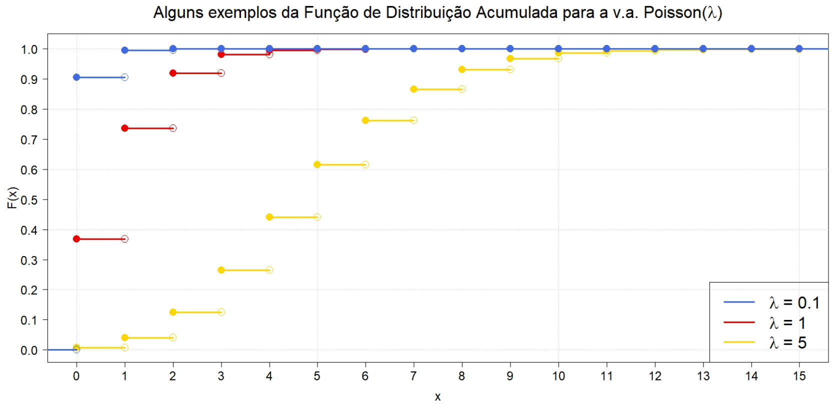


Figura B.6: Função de Distribuição Acumulada e de Probabilidade da Geométrica Deslocada

Poisson

Notação:	$X \sim Poi(\lambda)$
Parâmetro(s):	$\lambda > 0$
Função de probabilidade:	$p_X(x) = \dfrac{\lambda^x}{x!}e^{-\lambda}, \quad x = 0, 1, 2, \ldots$
Esperança e Variância:	$\mathrm{E}(X) = \lambda$ e $\mathrm{Var}(X) = \lambda$
Função Geradora de Momentos:	$M_X(x) = e^{\lambda(e^t - 1)}$.
Estimador de Máx. Verossimilhança:	$\hat{\lambda} = \bar{X}$.
Informação de Fisher:	$I(\lambda) = \dfrac{1}{\lambda}$.
Estatística Suficiente para λ:	$\sum_{i=1}^{n} X_i \sim Poi(n\lambda)$.
Observações:	• Se $X \sim Bin(N, p)$ com $N \to \infty$, $p \to 0$ e $np \to \lambda \in \mathbb{R}^*$, então $P(X = x) \approx \dfrac{\lambda^x}{x!}e^{-\lambda}$, sendo $\lambda = np$.



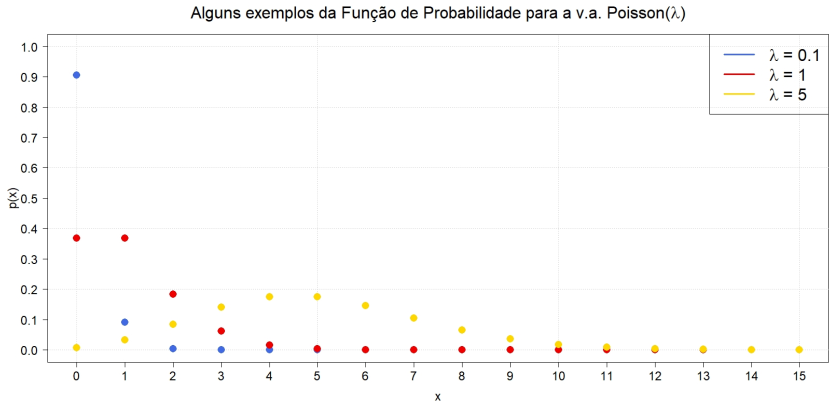


Figura B.7: Função de Distribuição Acumulada e de Probabilidade da Geométrica Deslocada

B.2 Distribuições Contínuas

Uniforme

Notação:	$X \sim U[a,b]$
Parâmetro(s):	$a, b \in \mathbb{R},\ a < b$
Função de densidade:	$f_X(x) = \dfrac{1}{b-a}, \quad a \leq x \leq b$
Função de distribuição acumulada:	$F_X(x) = \begin{cases} 0 & , \quad x < a \\ \frac{x}{b-a} & , \quad a \leq x < b \\ 1 & , \quad x \geq b. \end{cases}$
Esperança e Variância:	$\mathrm{E}(X) = \dfrac{b+a}{2}$ e $\mathrm{Var}(X) = \dfrac{(b-a)^2}{12}$.
Função Geradora de Momentos:	$M_X(x) = \begin{cases} \frac{e^{tb}-e^{ta}}{t(b-a)} & , \quad t \neq 0 \\ 1 & , \quad t = 0. \end{cases}$
Estimador de Máx. Verossimilhança:	$(\hat{a}, \hat{b}) = (X_{(1)}, X_{(n)})$
Estatística Suficiente para (a,b):	$\left(X_{(1)}, X_{(n)}\right)$
Observações:	• Se $X \sim U[a,b]$, $Y = \dfrac{X-a}{b-a} \sim U[0,1]$.
	• Se $X \sim U[0,1]$, $Y = \ln(X) \sim Exp(1)$.

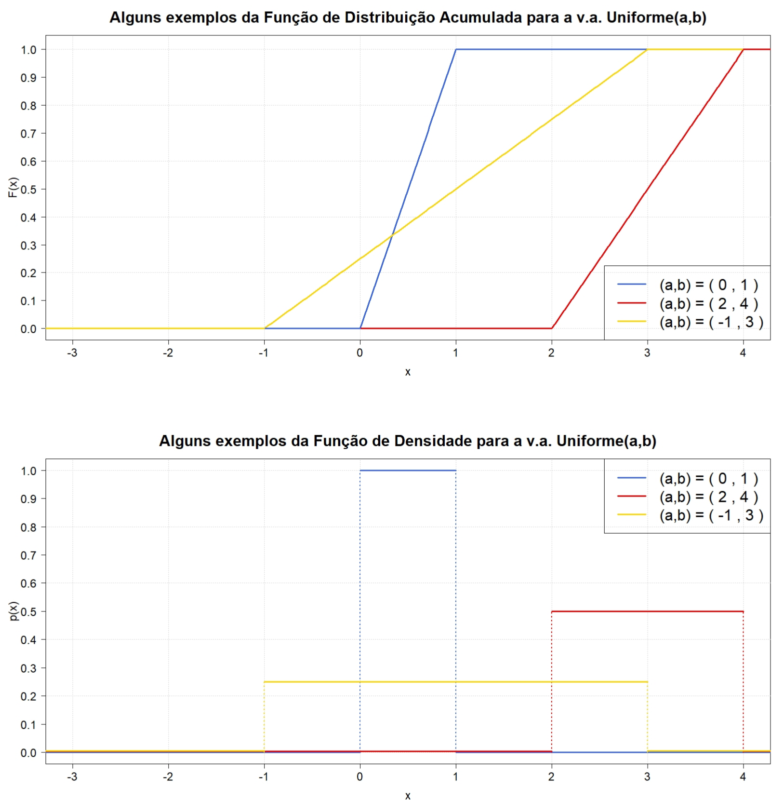
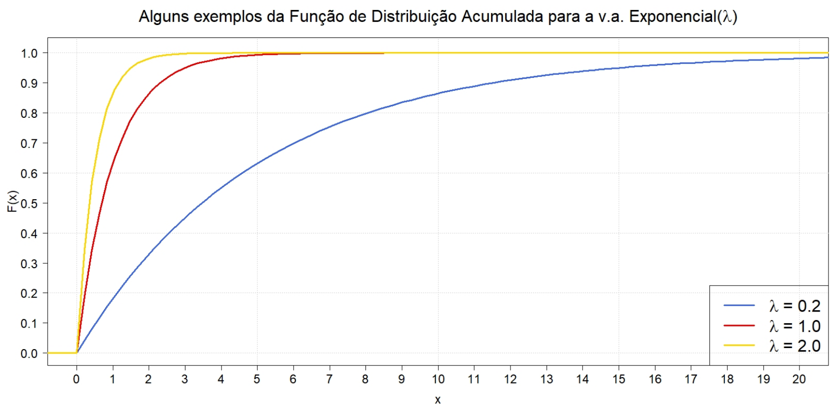


Figura B.8: Função de Distribuição Acumulada e de Probabilidade da Geométrica Deslocada

Exponencial

Notação:	$X \sim Exp(\lambda)$.
Parâmetro(s):	$\lambda > 0$.
Função de densidade:	$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0$.
Função de distribuição acumulada:	$F_X(x) = 1 - e^{-\lambda x}, \quad x \geq 0$.
Esperança e Variância:	$\mathrm{E}(X) = 1/\lambda$ e $\mathrm{Var}(X) = 1/\lambda^2$.
Função Geradora de Momentos:	$M_X(x) = \dfrac{\lambda}{\lambda - t}, \quad t < \lambda$.
Estimador de Máx. Verossimilhança:	$\hat{\lambda} = 1/\bar{X}$
Informação de Fisher:	$I(\lambda) = 1/\lambda^2$.
Estatística Suficiente para λ	$\sum_{i=1}^n X_i \sim Gama(n, \lambda)$
Observações:	• Se $X \sim Exp(\lambda)$, $Y = e^{-X/\lambda} \sim U[0,1]$. • Se $X \sim Exp(\lambda)$ e $a > 0$, $Y = X^{1/a} \sim Weibull(a, \lambda)$. • Satisfaz a propriedade de falta de memória: para $a < b$, $\mathrm{P}(X > b \mid X > a) = \mathrm{P}(X > b - a)$.



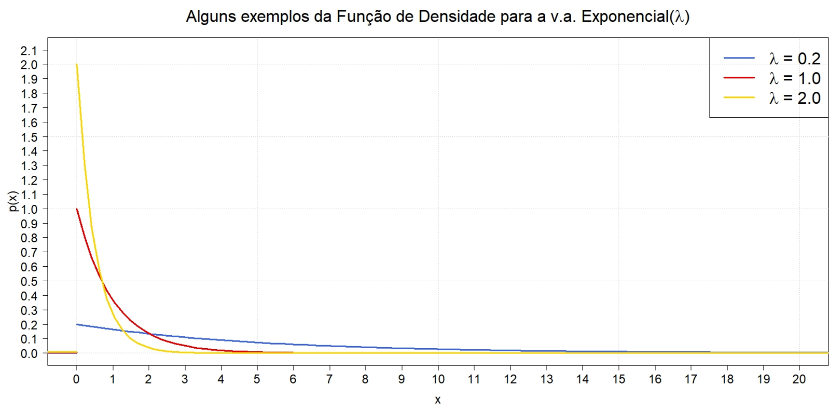
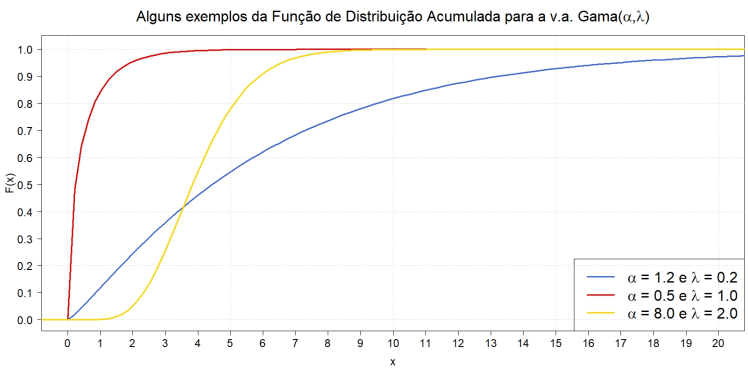


Figura B.9: Função de Distribuição Acumulada e de Probabilidade da Geométrica Deslocada

Gama

Notação:	$X \sim Gama(\alpha, \lambda)$.
Parâmetro(s):	$\alpha > 0$ e $\lambda > 0$.
Função de densidade:	$f_X(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x \geq 0.$
Esperança e Variância:	$\mathrm{E}(X) = \frac{\alpha}{\lambda}$ e $\mathrm{Var}(X) = \frac{\alpha}{\lambda^2}$.
Função Geradora de Momentos:	$M_X(x) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha, \quad t < \lambda.$
Estimador de Máx. Verossimilhança:	$\hat{\lambda} = \frac{\alpha}{\overline{X}}$, supondo α conhecido.
Informação de Fisher:	$I(\lambda) = \frac{\alpha}{\lambda^2}$, supondo α conhecido.
Estatística Suficiente para λ:	$\sum_{i=1}^n X_i \sim Gama(n\alpha, \lambda)$, supondo α conhecido.
Estatística Suficiente para α:	$\sum_{i=1}^n \ln(X_i)$, supondo λ conhecido.
Observações:	• Se $X \sim Gama(1, \lambda)$, $X = Exp(\lambda)$.
	• Se $X \sim Gama(n/2, 1/2)$, sendo $n \in \mathbb{N}^*$, $X = \chi^2_n$.
	• Se $X \sim Gama(\alpha, \lambda)$ e $c > 0$, $Y = cX \sim Gama(\alpha, \lambda/c)$.
	• Se $X_i \sim Gama(\alpha_i, \lambda)$, $\sum_{i=1}^n X_i \sim Gama(\sum_{i=1}^n \alpha_i, \lambda)$.



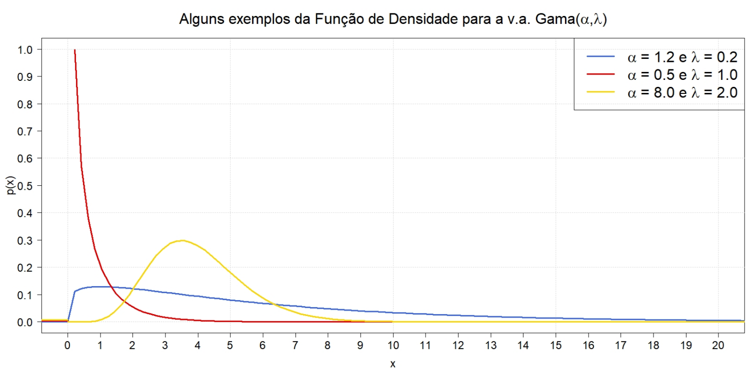
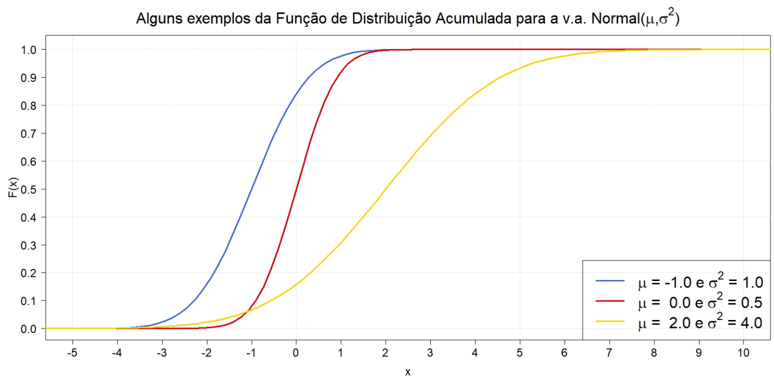


Figura B.10: Função de Distribuição Acumulada e de Probabilidade da Geométrica Deslocada

Normal

Notação:	$X \sim N(\mu, \sigma^2)$.
Parâmetro(s):	$\mu \in \mathbb{R}$ e $\sigma^2 > 0$.
Função de densidade:	$f_X(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$.
Esperança e Variância:	$\mathrm{E}(X) = \mu$ e $\mathrm{Var}(X) = \sigma^2$.
Função Geradora de Momentos:	$M_X(x) = e^{\mu t} e^{\sigma^2 t^2/2}$.
Estimador de Máx. Verossimilhança:	$(\hat{\mu}, \hat{\sigma}^2) = \left(\bar{X}, \dfrac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}\right)$..
Informação de Fisher:	$I(\mu) = \dfrac{1}{\sigma^2}$, supondo σ^2 conhecido. $I(\sigma^2) = \dfrac{1}{(\sigma^2)^2}$, supondo μ conhecido.
Estatística Suficiente para μ:	$\sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$, supondo σ^2 conhecido.
Estatística Suficiente para σ^2:	$(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$, supondo μ conhecido.
Observações:	• Se $X \sim N(0,1)$, $Y = \sigma X + \mu \sim N(\mu, \sigma^2)$. • Se $X \sim N(0,1)$, $Y = X^2 \sim Gama(1/2, 1/2) \sim \chi^2_1$.



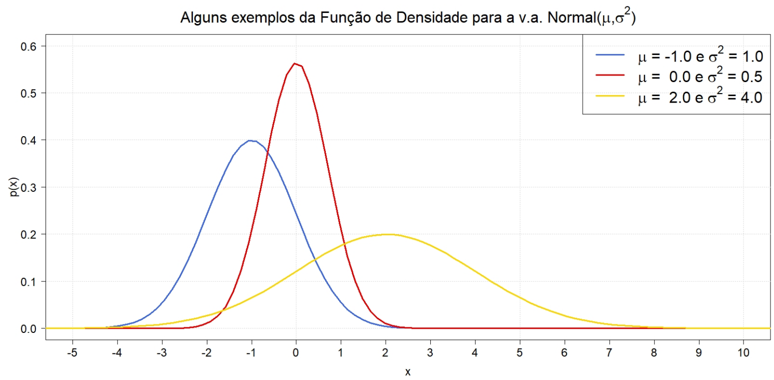
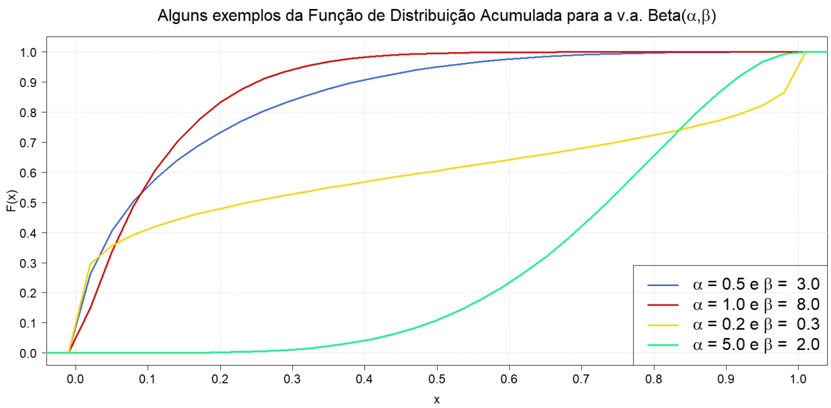


Figura B.11: Função de Distribuição Acumulada e de Probabilidade da Geométrica Deslocada

Beta

Notação:	$X \sim Beta(\alpha, \beta)$.
Parâmetro(s):	$\alpha > 0$ e $\beta > 0$.
Função de densidade:	$f_X(x) = \dfrac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathrm{B}(\alpha,\beta)}, 0 \leq x \leq 1.$
Esperança e Variância:	$\mathrm{E}(X) = \dfrac{\alpha}{\alpha+\beta}$ e $\mathrm{Var}(X) = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.
Estatística Suficiente para α:	$\sum_{i=1}^n \ln(X_i)$, supondo β conhecido.
Estatística Suficiente para β:	$\sum_{i=1}^n \ln(1-X_i)$, supondo α conhecido.
Observações:	• $\mathrm{B}(\alpha,\beta) = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.
	• $X \sim Beta(1,1) \sim U[0,1]$.
	• $X \sim Beta(\alpha,\beta)$, $Y = 1-X \sim Beta(\beta,\alpha)$.
	• $X \sim Beta(\alpha,1)$, $Y = -\ln(X) \sim Exp(\alpha)$.



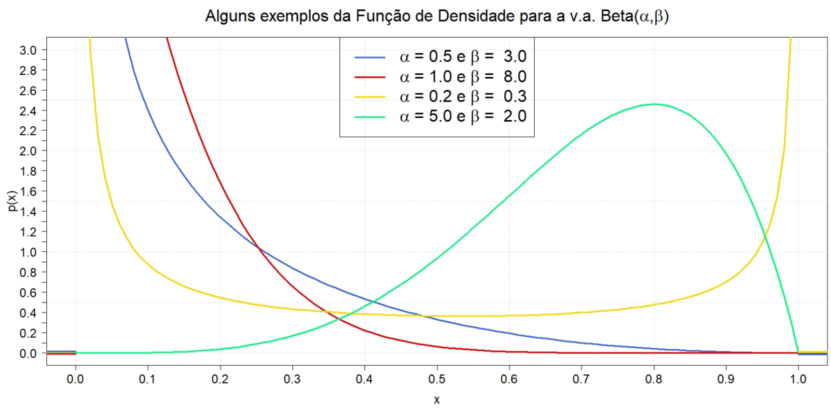
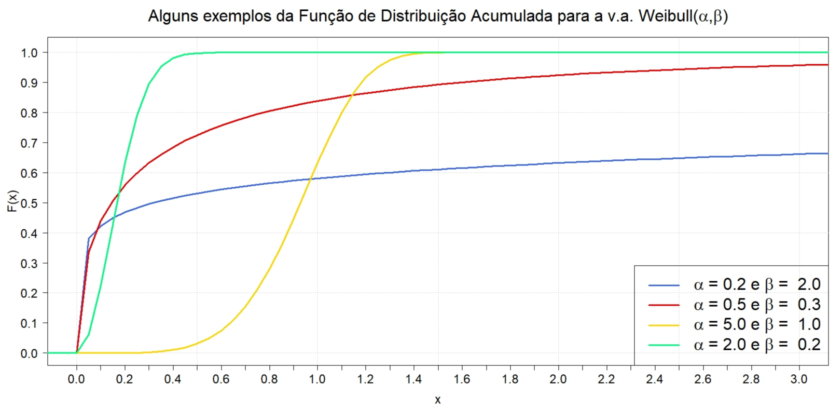


Figura B.12: Função de Distribuição Acumulada e de Probabilidade da Geométrica Deslocada

Weibull

Notação:	$X \sim Weib(\alpha, \beta)$.
Parâmetro(s):	$\alpha > 0$ e $\beta > 0$.
Função de densidade:	$f(x) = \dfrac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^\alpha}$, $x > 0$.
Função de distribuição acumulada:	$F_X(x) = \begin{cases} 0 & , \ x \leq 0 \\ 1 - e^{-(x/\beta)^\alpha} & , \ x > 0 \end{cases}$
Esperança e Variância:	$\mathrm{E}(X) = \beta \Gamma\left(\frac{\alpha+1}{\alpha}\right)$ e $\mathrm{Var}(X) = \beta^2 \left[\Gamma\left(\frac{\alpha+2}{\alpha}\right) - \left(\Gamma\left(\frac{\alpha+1}{\alpha}\right)\right)^2\right]$.
Estimador de Máx. Verossimilhança:	$\hat{\beta} = \left(\dfrac{\sum_{i=1}^n X_i^\alpha}{n}\right)^{1/\alpha}$, supondo α conhecido.
Informação de Fisher:	$I(\beta) = \dfrac{\alpha^2}{\beta^2}$, supondo α conhecido.
Estatística Suficiente para β:	$\sum_{i=1}^n X^\alpha \sim Gama(n, 1/\beta^\alpha)$, supondo α conhecido.
Observações:	• $X \sim Weib(\alpha, \beta)$, $Y = X^\alpha \sim Exp(1/\beta^\alpha)$.



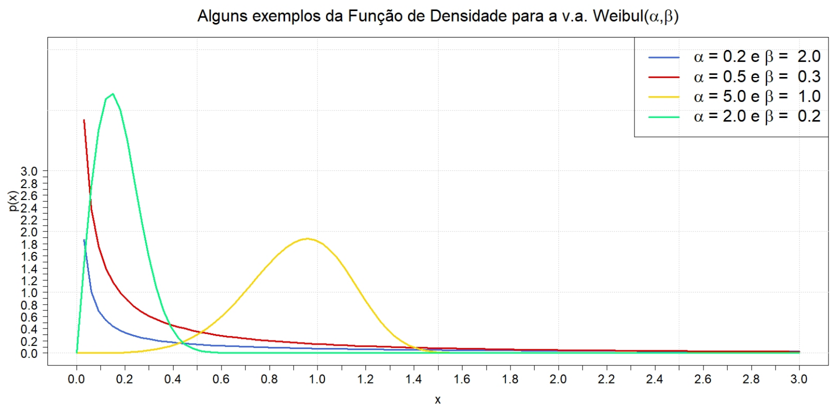
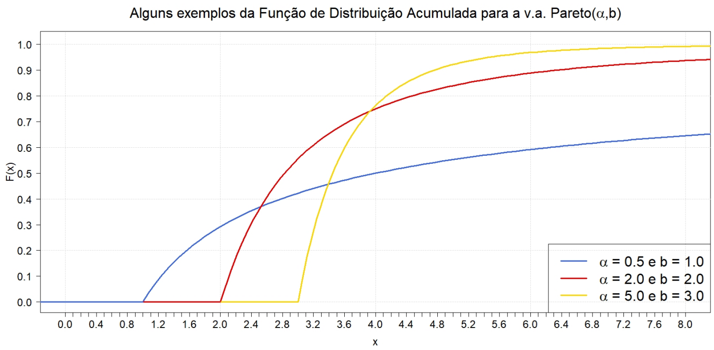


Figura B.13: Função de Distribuição Acumulada e de Probabilidade da Geométrica Deslocada

Pareto

Notação:	$X \sim Pareto(\alpha, b)$.
Parâmetro(s):	$\alpha > 0$ e $b > 0$.
Função de densidade:	$f(x) = \frac{\alpha}{b}\left(\frac{b}{x}\right)^{\alpha+1} \quad x \geq b$
Função de distribuição acumulada:	$F_X(x) = \begin{cases} 0 & \text{, se } x < b \\ 1 - \left(\frac{b}{x}\right)^{\alpha} & \text{, se } x \geq b \end{cases}$
Esperança e Variância:	$\mathrm{E}(X) = \frac{\alpha}{\alpha - 1}b$, se $\alpha > 1$ (caso contrário, não existe) $\mathrm{Var}(X) = \frac{\alpha b^2}{(\alpha-1)^2(\alpha-2)}$, se $\alpha > 2$ (caso contrário, não existe).
Estimador de Máx. Verossimilhança:	$\hat{\alpha} = \frac{n}{\sum_{i=1}^{n} \ln(X_i/b)}$, supondo b conhecido. $\hat{\beta} = X_{(1)}$, supondo α conhecido.
Informação de Fisher:	$I(\alpha) = \frac{1}{\alpha^2}$, supondo b conhecido.
Estatística Suficiente para α:	$\sum_{i=1}^{n} \ln(X_i/b) \sim Gama(n, \alpha)$, supondo b conhecido.
Observações:	• $X \sim Pareto(\alpha, b)$, $Y = \ln(X/b) \sim Exp(\alpha)$. • $X \sim Pareto(\alpha, b)$, $X_{(1)} \sim Pareto(n\alpha, b)$



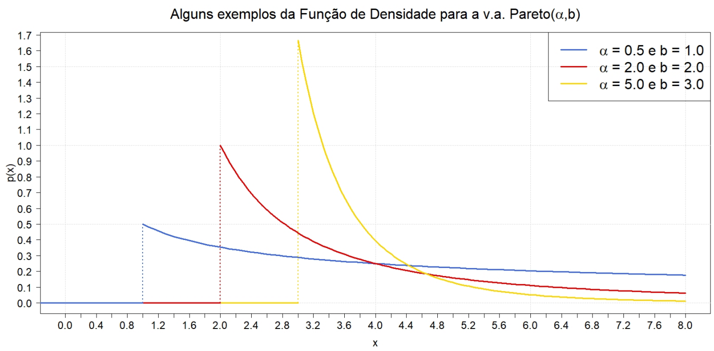


Figura B.14: Função de Distribuição Acumulada e de Probabilidade da Geométrica Deslocada

Apêndice C

Respostas dos Exercícios

Exercícios da Seção 1.1

1.1.2. (a) $\mathrm{P}(Z > 2,43) = 0,0075$, (b) $\mathrm{P}(Z > -1,81) = 0,9649$,
(c) $\mathrm{P}(Z < 2,76) = 0,9971$, (d) $\mathrm{P}(Z < -0,68) = 0,2483$, (e) $\mathrm{P}(0 < Z < 1,31) = 0,4049$,
(f) $\mathrm{P}(-0,45 < Z < 0) = 0,1736$, (g) $\mathrm{P}(-1,4 < Z < 2,3) = 0,9085$,
(h) $\mathrm{P}(1,21 < Z < 1,96) = 0,0881$, (i) $\mathrm{P}(-2,02 < Z < -0,06) = 0,4544$.

1.1.3. (a) (b) (c)

Exercícios da Seção 1.2

1.2.1. (a) 0,9332 (b) 0,0122 (c) 0,7734 (d) 0,1056 (e) 0,8185.

1.2.2. (a) 1,08 (b) 1,56 (c) -4,3.

1.2.3. $Y \sim N(-4, 17)$.

1.2.4. 0,3174.

Exercícios da Seção 1.3

1.3.1. 0,1056.

1.3.2. (a) $Y \sim N(0, 9)$ (b) (X, Y) tem distribuição Normal Bivariada com parâmetros $\mu_X = \mu_Y = 0$, $\sigma_X^2 = 4$, $\sigma_Y^2 = 9$ e $\rho = 1/3$ (c) $X|Y = y \sim N(2y/9, 32/9)$.

1.3.3. (a) (b) (c) (d) (e) (f) (g)

1.3.4. $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^T$ e Σ é a matriz diagonal com valores σ_1^2, σ_2^2, $\ldots$, σ_n^2 na diagonal principal

Exercícios da Seção 1.4

1.4.1. .

1.4.2. .

1.4.3. $Z^2/6 \sim Gama\left(\frac{1}{2}, 3\right)$.

Exercícios da Seção 1.5

1.5.1. .

1.5.2. .

1.5.3. (a) (b) (c) (d) .

Mais Alguns Exercícios do Capítulo 1

1.1. 0,6826.

1.2. 0,0694.

1.3. $Z^2/6 \sim Gama\left(\frac{1}{2}, 3\right)$.

1.4. (a) Não é gama nem normal, $X \sim U[-1, 1]$ (b) $X \sim N(1, 9)$ (c) $X \sim Gama(1, 1/3)$ (d) $X \sim Gama(3, 2)$ (e) Não é gama nem normal, $X \sim Beta(1/3, 2)$ (f) $X \sim N(0, 1/4)$ (g) Não é gama nem normal, $X \sim Pareto(2, 4)$ (h) $X \sim Gama(1/2, 1)$.

1.5. (a) $X \sim Gama(1/2, 3)$ (b) Não é gama nem normal, $X \sim Binomial(1/2, 4)$ (c) $X \sim N(2, 8)$ (d) Não é gama nem normal, $X \sim Poisson(1)$ (e) $X \sim Gama(1, 1)$ (f) Não é gama nem normal, $X \sim Geo(1/3)$ (g) $X \sim N(0, 4)$ (h) $X \sim Gama(5, 1/2)$.

Exercícios da Seção 2.1

2.1.1.

2.1.2.

2.1.3. (a) 0,3935 (b) 0,6321.

2.1.4. 0,75.

2.1.5. (a) Sim (c) $\hat{\sigma}^2 \sim Gama\big(n/2, n/(2\sigma^2)\big)$.

2.1.7. (a) $\sum_{i=1}^n X_i \sim Gama(n, \lambda)$ (b) $2\lambda n\bar{X} \sim \chi^2_{2n}$.

2.1.8. (a) Sim (b) $X_{(1)} \sim Exp(n\lambda)$.

2.1.9. (a) $f(u) = n(1-u)^{n-1}, 0 < u < 1$ (b) $f(u) = nu^{n-1}, 0 < u < 1$ (c) $\mathrm{P}(U_{(1)} \leq \frac{1}{2}) = 1 - 1/2^n$ e $\mathrm{P}(U_{(n)} \leq \frac{1}{2}) = 1/2^n$.

2.1.10.

2.1.11. (b) $T(\mathbf{X}) \sim Gama(n, \theta)$.

Exercícios da Seção 2.2

2.2.1.

2.2.2.

2.2.3.

2.2.4. $a = 1,066$.

2.2.5. $c = 2$ e $k = 8$.

Exercícios da Seção 2.3

2.3.1. (a) $\mathrm{P}(X < 4) \approx 0,975$ (b) $\mathrm{P}(X > 1,5) \approx 0,25$ (c) $\mathrm{P}(X < 0,6) \approx 0,25$
(d) $\mathrm{P}(0,37 < X < 0,97) \approx 0,4$ (e) $x = 0,2749$.

2.3.2.

2.3.3.

2.3.4. $\mathrm{Var}(X) = \dfrac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$.

Exercícios da Seção 2.4

2.4.1. (a) $0,001565402$ (b) $0,988492$ (c) $0,1412558$.

2.4.2 (a) 0,9389 (b) 0,9174.

Exercícios da Seção 2.5

2.5.1. (a) Pertence à família exponencial unidimensional (b) $Y = \ln\left(\dfrac{X}{4}\right) \sim Exp(\theta)$
(c) $T(\mathbf{X}) = \sum_{i=1}^{n} \ln\left(\dfrac{X}{4}\right) \sim Gama(n, \theta)$

2.5.2. (a) Não pertence à família exponencial unidimensional (b) $T(\mathbf{X}) = X_{(1)} \sim Pareto(3n, \theta)$

2.5.3. (a) Pertence à família exponencial unidimensional (b) $Y = X^3 \sim Exp(\theta^3)$
(c) $T(\mathbf{X}) = \sum_{i=1}^{n} X_i^3 \sim Gama(n, \theta^3)$

2.5.4. (a) Não pertence à família exponencial unidimensional (b) $Y = -\ln(2X) \sim Exp(\theta)$
(c) $T(\mathbf{X}) = \sum_{i=1}^{n} -\ln(2X_i) \sim Gama(n, \theta)$.

2.5.5. (a) Pertence à família exponencial unidimensional (b) $T(\mathbf{X}) = \sum_{i=1}^{n} X_i \sim Binom(15n, p)$.

Exercícios do Capítulo 2

2.1. (a) $\frac{\lambda_x \bar{X}}{\lambda_y \bar{Y}} \sim F_{2n,2m}$ e $\frac{\lambda_y \bar{Y}}{\lambda_x \bar{X}} \sim F_{2m,2n}$.

2.2. (a) Não pertence à família exponencial unidimensional
(b) $f_{X_{(1)}}(x) = \frac{n}{(5-\theta)^n}(5-x)^{n-1},\ \theta < x < 5$.

2.3. (a) V (b) F, contra exemplo: Seja $X \sim N(\mu, \sigma^2)$, $\sum_{i=1}^{n}(X_i - \mu)^2/n$ é uma transformação da amostra e não é uma estatística. (c) F, contra exemplo: Seja $X \sim Gama(1,1)$, $f(x) = e^{-x}, x > 0$. Para $x = 1$ temos $f(1) = 1/e$ e $f(-1) = e$. Logo, $f(1) \neq f(-1)$. (d) V

(e) V (f) F, contra exemplo: Seja $X \sim Gama(2,1)$, $E(1/X) = 1 \neq 1/2$ (faça as contas).

(g) V (h) F, contra exemplo (quase demonstração): Veja que $f(x) = 0$ para $x \leq 0$. Então, seja $X \sim F(1,1)$ e $x > 0$ qualquer, $f(x) > 0$ e $f(-x) = 0 \Rightarrow f(x) \neq f(-x)$. (i) V (j) V.

Exercícios da Seção 3.2

3.2.1. (a) $\hat{p} = 1/(\bar{X}+1)$ (b) $\hat{p} = 1/\bar{X}$ (c) $\hat{\alpha} = \bar{X}^2/(M_2 - \bar{X}^2)$ e $\hat{\lambda} = \bar{X}/(M_2 - \bar{X}^2)$ (d) $\hat{\mu} = 2\ln(\bar{X}) - \ln(M_2)/2$ e $\hat{\sigma}^2 = \ln(M_2) - 2\ln(\bar{X})$, onde $M_2 = \sum_{i=1}^n X_i^2/n$.

3.2.2. $\hat{p} = \bar{X}$.

3.2.3. 6.700 pessoas votariam a favor da obra.

3.2.4. $0,192$.

3.2.5. $\hat{N} = 7.310$ táxis.

Exercícios da Seção 3.3

3.3.1. Se $X_1 = 0$ ou $X_1 = 1$, $\hat{\theta} = 1$; se $X_1 = 2$, $\hat{\theta} = 2$ ou 3; se $X_1 = 3$ ou $X_1 = 4$, $\hat{\theta} = 3$.

3.3.2. (a) $\hat{p} = 1/\bar{X}$ (b) $\hat{p} = \bar{X}/N_0$ (c) $\hat{\theta} = \max\{-X_{(1)}, X_{(n)}\}$.

3.3.3. (a) $\hat{\lambda} = 1/\bar{X}$ (b) $\hat{\mu}) = \bar{X}$ (c) $\hat{\eta} = e^{-1/\bar{X}}$ (d) $\hat{\lambda} = 3,01$, $\hat{\mu} = 0,331$ e $\hat{\eta} = 0,0491$.

3.3.4. (a) $\hat{\theta}_{MM} = \bar{X}/(1-\bar{X})$ (b) $\hat{\theta}_{MV} = -n/(\sum_{i=1}^n \ln(X_i))$ (c) $\hat{\mu}_{MM} = \bar{X}$ (d) $\hat{\mu}_{MV} = n/\left((\sum_{i=1}^n \ln(X_i)) - n\right)$.

3.3.5. (a) 1,5 (b) 0,405.

Exercícios da Seção 3.4

3.4.2. $B(\hat{\theta}) = \sigma^2/n$ e $\hat{\theta}$ é assintoticamente não-tendencioso.

3.4.3. (a) $B(\hat{\theta}_{MV}) = \theta/(n-1)$ e $\hat{\theta}_{MV}$ é assintoticamente não-tendencioso. (b) $\mathrm{Var}(\hat{\theta}_{MV}) = (n^2/(n-1)^2(n-2))\theta^2$ e $\hat{\theta}_{MV}$ é consistente em EQM.

3.4.4. $EQM(T_1) = (4-\pi)/2\lambda^2$ e $EQM(T_2) = 1/2\lambda^2$, como $4 - \pi < 1$ $\Rightarrow EQM(T_1) < EQM(T_2)$.

3.4.5. (a) $B(T) = -p/(n+1)$ (b) não (c) sim (d) $EQM(T) = (np(1-p) + p^2)/(n+1)^2$ (e) sim.

3.4.6. $a = n/(n+1)$.

Mais Alguns Exercícios do Capítulo 3

3.1. (a) $\hat{\sigma}^2_{MM} = \sum_{i=1}^n X_i^2/n - \mu_0^2$ (b) $\hat{\sigma}^2_{MV} = \sum_{i=1}^n (X_i - \mu_0)^2/n$ (c) $\hat{\sigma}^2_{MM} = 1,907$ e $\hat{\sigma}^2_{MV} = 2,516$ (d) $\hat{\sigma}^2_{MM} = -7,099$ e $\hat{\sigma}^2_{MV} = 2,709$.

3.2. (a) $\hat{\eta}_{MM} = M_2 = \sum_{i=1}^n X_i^2/n$ (b) $\hat{\eta}_{MV} = \bar{X}^2 + \bar{X}$ (c) $\hat{\eta}_{MM} = 4,83$ e $\hat{\eta}_{MV} = 5,19$.

3.3. (a) $\hat{\lambda} = 1/\bar{X}$ (b) $\hat{\nu} = (\sum_{i=1}^n X_i^2)/2n$. (c) $\hat{\lambda} = 1/\bar{X}$ (d) $\hat{\nu} = \bar{X}^2$

3.4. (b) $\hat{\theta} = X_{(1)}$ (c) $\hat{\eta} = 1/2X_{(1)}$.

3.5. (a) $\hat{\lambda} = 0,00357$ (b) $\hat{\mu} = 280,4$ horas.

3.8. (a) $\hat{\eta}_{MM} = \sum_{i=1}^{n} X_i^2/n$ e $\hat{\eta}_{MV} = \bar{X}(\bar{X}+1)$.

3.9. (a) $\hat{a}_{MM} = \bar{X}-1$ (b) $\hat{a}_{MV} = X_{(1)}$ (c) $\hat{a}_{MM}$ é não-tendencioso e $\hat{a}_{MV}$ é assintoticamente não-tendencioso. (d) Para $n > 2$ $\hat{a}_{MV}$ é mais eficiente que $\hat{a}_{MM}$

3.10. (a) $\hat{\theta}_{MM} = 2\bar{X} - 5$ (b) $\hat{\theta}_{MM}$ é consistente em EQM (c) $\hat{\theta}_{MV} = X_{(1)}$ (d) $\hat{\theta}_{MV}$ é assintoticamente não-tendencioso (e) $\hat{\theta}_{MM} = 1,718$ e $\hat{\theta}_{MV} = 2,06$.

3.11. (a) $\hat{\theta}_{MM} = 4/(\bar{X} - 4)$ (b) $\hat{\theta}_{MV} = n/\sum_{i=1}^{n} \ln(X_i/4)$ (c) $\hat{\theta}_{MV}$ é assintoticamente não-tendencioso (d) $\hat{\theta}_{MM} = 1,196$ e $\hat{\theta}_{MV} = 1,443$.

3.12. (a) $\hat{\theta}_{MM} = 2\bar{X}/3$ (b) $\hat{\theta}_{MM}$ é consistente em EQM (só pq o parâmetro conhecido $\alpha = 3 > 2$) (c) $\hat{\theta}_{MV} = X_{(1)}$ (d) $\hat{\theta}_{MV}$ é assintoticamente não-tendencioso. (e) $\hat{\theta}_{MM} = 1,034$ e $\hat{\theta}_{MV} = 1,02$.

3.13. (a) $\hat{\theta}_{MM} = \Gamma(4/3)/\bar{X} = 0,8929/\bar{X}$ (b) $\hat{\theta}_{MV} = \left(n/\sum_{i=1}^{n} X_i^3\right)^{1/3}$ (c) $\hat{\theta}_{MM} = 0,579$ e $\hat{\theta}_{MV} = 0,585$.

3.14. (a) $\hat{\theta}_{MM} = 2\bar{X}/(1-2\bar{X})$ (b) $\hat{\theta}_{MV} = n/(-\sum_{i=1}^{n} \ln(2X_i)$ (c) $\hat{\theta}_{MV}$ é assintoticamente não-tendencioso. (d) $\hat{\theta}_{MM} = 2,16$ e $\hat{\theta}_{MV} = 2,29$.

3.15. (a) $\hat{p}_{MM} = \bar{X}/15$ (b) $\hat{p}_{MM}$ é consistente em EQM (c) $\hat{p}_{MV} = \bar{X}/15$ (d) $\hat{p}_{MV}$ é não-tendencioso. (e) $\hat{p}_{MM} = 0,264$ e $\hat{\theta}_{MV} = 0,264$.

Exercícios da Seção 4.1

4.1.1. (a) $p(1-p)/n$ (b) $(1-2p)^2 p(1-p)/n$.

4.1.2. (a) σ^2/n (b) $2\sigma^4/n$.

Exercícios da Seção 4.2

4.2.1. sim, $\bar{X}$.

4.2.2. não.

4.2.4. Não existe estimador eficiente para θ, pois o estimador de MV não é eficiente.

4.2.5. (a) $V_{min}^{\lambda} = \lambda^2/(n\alpha_0)$ $V_{min}^{\theta} = 1/(n\alpha_0\lambda^2)$.

Exercícios da Seção 4.3

4.3.1. (a) $\sum_{i=1}^{n} X_i^2$ (b) $\sum_{i=1}^{n} X_i$ (c) $X_{(1)}$

Exercícios da Seção 4.4

4.4.1. (a) $\hat{\sigma}^2 = \sum_{i=1}^{n} X_i^2/n$. Sim, o ENVVUM é eficiente. (b) $\hat{\theta} = \sum_{i=1}^{n} X_i/5n$. Sim, o ENVVUM é eficiente. (c) $\hat{\theta} = X_{(1)} - 1/n$. Não, o ENVVUM não é eficiente.

4.4.2. (b) Sim, $\hat{\eta}$ (b) $\hat{\eta}$ não é eficiente pois não é estimador de MV, $V_{min} = 4\mu^2/n$.

4.4.3. $n\bar{X}(1-\bar{X})/(n-1)$ é o ENVVUM para $\theta(1-\theta)$.

4.4.4. $1+\bar{X}$ é o ENVVUM para $1/\theta$.

4.4.5. (a) $\hat{\lambda}=(n-1)/\sum_{i=1}^{n}X_i$, $\hat{\lambda}$ não é eficiente (b) $\hat{\mu}=\bar{X}$, $\hat{\mu}$ é eficiente (c) $\hat{\theta}=n\bar{X}^2/(n+1)$, $\hat{\theta}$ não é eficiente. (d) $\hat{\eta}=\left(\dfrac{\sum_{i=1}^{n}X_i-2}{\sum_{i=1}^{n}X_i}\right)^{n-1}$, $\hat{\eta}$ não é eficiente.

4.4.6. $\hat{\eta}=\bar{X}/\alpha$.

4.4.7. $\hat{\theta}_{ENVVUM}=-(n-1)/\sum_{i=1}^{n}\ln(X_i)$ e o $\hat{\theta}_{ENVVUM}$ não é eficiente.

Mais Alguns Exercícios do Capítulo 4

4.1. (a) $I(\lambda)=\alpha_0/\lambda^2$ (b) $V^{\lambda}_{min}=\lambda^2/(n\alpha_0)$, $V^{\mu}_{min}=\alpha_0/(n\lambda^2)$ e $V^{\eta}_{min}=4\alpha_0/(n\lambda^4)$

(c) Não existe estimador eficiente para λ e nem para η e $\hat{\mu}=\bar{X}$ é eficiente para μ.

(d) $T(\mathbf{X})=\sum_{i=1}^{n}X_i$ (e) $\hat{\lambda}=(n\alpha_0-1)/\sum_{i=1}^{n}X_i$, $\hat{\mu}=\bar{X}$ e $\hat{\eta}=\left(\sum_{i=1}^{n}X_i\right)^2/(n(1+n\alpha_0))$

(f) $\hat{\lambda}=0,806$, $\hat{\mu}=3,581$ e $\hat{\eta}=4,122$.

4.2. (a) $I(p)=15/(p(1-p))$ (b) $V^{p}_{min}=p(1-p)/(15n)$, $V^{\mu}_{min}=15p(1-p)/n$ e $V^{\eta}_{min}=60p(1-p)/n$ (c) $\hat{p}=\bar{X}/15$ é eficiente para p, $\hat{\mu}=\bar{X}$ é eficiente para μ e não existe estimador eficiente para η (d) $T(\mathbf{X})=\sum_{i=1}^{n}X_i$ (e) $\hat{p}=\bar{X}/15$, $\hat{\mu}=\bar{X}$ e $\hat{\eta}=\bar{X}(1-\bar{X}/15)(15n/(15n-1))$ (f) $\hat{p}=0,264$, $\hat{\mu}=3,96$ e $\hat{\eta}=2,92$.

Exercícios da Seção 5.2

5.2.1. (a) $IC_{exa}=\left(q_1/2n\bar{X}\ ,\ q_2/2n\bar{X}\right)$, onde q_1 e q_2 são quantis da distribuição χ^2_{2n}

(b) $(0.2454\ ,\ 0.9399)$.

5.2.3. $IC_\theta=\bar{X}\pm t_{n-1,1-\alpha/2}\sqrt{S^2/n}$ ou $IC_\theta=\left((n-1)S^2/q_{1-\alpha/2,n-1}\ ,\ (n-1)S^2/q_{\alpha/2,n-1}\right)$.

5.2.4. (c) $IC=(q_1/2\sum_{i=1}^{n}|X_i|\ ,\ q_2/2\sum_{i=1}^{n}|X_i|)$, onde q_1 e q_2 são quantis da distribuição χ_{2n} (d) $IC_{\lambda,90\%}=(1.66\ ,\ 4.82)$, $IC_{\lambda,95\%}=(1.47\ ,\ 5.25)$ e $IC_{\lambda,99\%}=(1.14\ ,\ 6.14)$.

5.2.5. (d) $IC_\theta=\left(X_{(1)}+\ln(\alpha)/n\lambda\ ,\ X_{(1)}\right)$ (e) $IC_\theta=(4.4411\ ,\ 5.19)$.

5.2.6. (a) $\hat{\beta}_{MV}=X_{(n)}$ e $Q=X_{(n)}/\beta$ é quantidade pivotal para β.

(b) $IC_{\beta,95\%}=(X_{(n)},X_{(n)}/(0,05^{1/n\alpha_0}))$ (c) $IC_{\beta,95\%}=(25.00\ ,\ 25.45)$.

5.2.7. (a) $Q=-2\theta\sum_{i=1}^{n}\ln(X_i)\sim\chi^2_{2n}$

(b) $IC_\theta=[-q_{\alpha/2,2n}/2\sum_{i=1}^{n}\ln(X_i)\ ,\ -q_{1-\alpha/2,2n}/2\sum_{i=1}^{n}\ln(X_i)]$

(c) $IC_\mu=[q_{\alpha/2,2n}/(q_{\alpha/2,2n}-2\sum_{i=1}^{n}\ln(X_i))\ ,\ q_{1-\alpha/2,2n}/(q_{1-\alpha/2,2n}-2\sum_{i=1}^{n}\ln(X_i))]$

(d) $ic_{\theta,95\%}=[0,3543086\ ,\ 1,124656]$ e $ic_{\mu,95\%}=[0,2616159\ ,\ 0,5293356]$.

5.2.8. (a) $f_Q(x)=n(1-x)^{n-1}$, $0<x<1$ (b) $IC_\theta=\left[5-(5-X_{(1)})/q_1\ ,\ 5-(5-X_{(1)})/q_2\right]$, escolhendo $q_2=1$, $IC_\theta=\left[5-(5-X_{(1)})/(1-0,95)^{1/n}\ ,\ X_{(1)}\right]$ (c)$ic_\theta=[1.226\ ,\ 2.06]$.

Exercícios da Seção 5.3

5.3.1. (a) $IC_{\lambda,95\%} = [2,35\ ,\ 4,18] \Rightarrow$ com 95% de confiança são vendidas em média entre 2,35 e 4,18 casas por semana (b) $IC_{\lambda,95\%} = (-\infty\ ,\ 4,03] \Rightarrow$ com 95% de confiança são vendidas em média menos de 4 casas por semana.

5.3.2. (a) $n = 1.849$ (usando valores da tabela com 2 casas decimais de precisão)
(b) o intervalo de confiança para a proporção de cidadãos que apoiam o projeto é $[0,39\ ,\ 0,45]$ e com isso podemos afirmar com 99% de certeza que a maioria dos cidadãos não apoiam o projeto.

Exercícios da Seção 5.4

5.4.1. O intervalo de confiança unilateral à direita para $\mu_x - \mu_y$ é $ic_{\mu_x-\mu_y} = [2,320454\ ,\ \infty)$. Com 95% de certeza acreditamos que o rendimento da linha A é superior ao da linha B.

5.4.2. A partir do IC unilateral aproximado podemos encontrar um limite inferior para $p_1 - p_2$: $IC_{p_1-p_2} = [0.0003055526, \infty)$ e concluir, com 90% de confiança, que $p_1 - p_2 > 0.0003055526$, ou seja, a prevalência da doença é maior entre os não vacinados.

5.4.3. (a) $Q = \theta_1\bar{X}/\theta_2\bar{Y} \sim F_{2n,2m}$ (b) $IC = \left[f_1\bar{Y}/\bar{X}\ ,\ f_2\bar{Y}/\bar{X}\right]$, onde f_1 e f_2 são quantis da distribuição $F_{2n,2m}$. (c) $ic_{95\%} = [0,2518\ ,\ 1,2985]$.

Mais Alguns Exercícios do Capítulo 5

5.1. A resposta depende da escola dos quantis e do α. Veja algumas possibilidades:
$ic_{\theta,90\%} = [477\ ,\ 556,14]$, $ic_{\theta,90\%} = [478,63\ ,\ 582,44]$, $ic_{\theta,99\%} = [477,16\ ,\ 679,08]$.

5.2. (a) $ic_\mu = [390,73\ ,\ 821,85]$ (b) $ic_p = [0,77\ ,\ 0,89]$.

5.3. Com 90% de confiança podemos afirmar que o tempo médio até a ocorrência de uma falha é maior que 420,54 horas e que a probabilidade da falha ocorrer depois de 100 horas de funcionamento é maior que 78,8%.

5.4. $ic_{\mu,90\%} = [344,337\ ,\ 744,993]$ e $ic_{p,90\%} = [0,7760681\ ,\ 0,8884707]$.

5.5. (a) $Q = 2\theta\sum_{i=1}^n \ln\left(\frac{X_i}{4}\right) \sim \chi^2_{2n}$ e $IC_\theta = [q_{\alpha,2n}/(2\sum_{i=1}^n \ln(X_i/4), \infty]$
$= [18,93924/(2\sum_{i=1}^{14} \ln(X_i/4), \infty]$ (b) $IC_\eta = [0\ ,\ (1/2)^{(18,93924/\left(2\sum_{i=1}^{14}\ln(X_i/4)\right))}]$
(c) $ic_\theta = [1,518647\ ,\ \infty]$ e $ic_\eta = (0\ ,\ 0,3490131)$

5.6. (a) $Q = 2\beta^3\sum_{i=1}^n X_i^3$ e $IC_\theta = [0\ , q_{1-\alpha,2n}/(2\sum_{i=1}^n X_i^3)] =$
$[0\ , (28.8693/(2\sum_{i=1}^9 X_i^3))^{1/3}]$ (b) $IC_\eta = [e^{-(28.8693/(2\sum_{i=1}^9 X_i^3))}\ , 1\ ]$ (c) $ic_\theta = [0\ ,\ 0,6847514]$ e $ic_\eta = [0,5042156\ ,\ 1]$.

5.7. (a) $Q = -2\theta\sum_{i=1}^n \ln(2X_i) \sim \chi^2_{2n}$ e $IC_\theta = \left(\frac{q_{\alpha/2,2n}}{-2\sum_{i=1}^n \ln(2X_i)}\ ,\ \frac{q_{1\alpha/2,2n}}{-2\sum_{i=1}^n \ln(2X_i)}\right) =$
$\left(\frac{5,142205}{-2\sum_{i=1}^8 \ln(2X_i)}, \frac{34,26719}{-2\sum_{i=1}^8 \ln(2X_i)}\right)$ (b) $IC_\mu = \left(\frac{q_{\alpha/2,2n}}{2(q_{\alpha/2,2n} - 2\sum_{i=1}^n \ln(2X_i))}, \frac{q_{1-\alpha/2,2n}}{(2(q_{1-\alpha/2,2n} - 2\sum_{i=1}^n \ln(2X_i))}\right)$
$= \left(\frac{5,142205}{2(5,142205 - 2\sum_{i=1}^8 \ln(2X_i))}\ ,\ \frac{34,26719}{2(34,26719 - 2\sum_{i=1}^8 \ln(2X_i))}\right)$ (c) $ic_\theta = (0,7371358\ ,\ 4,912205)$ e $ic_\mu = (0,2121699\ ,\ 0,4154292)$.

5.8. (a) $Q^a = (\bar{X} - 15p)/\sqrt{\frac{\bar{X}}{n}\left(1 - \frac{\bar{X}}{15}\right)}$ e $IC_p = \frac{1}{15}\left(\bar{X} \pm z_{1-\alpha/2}\sqrt{\frac{\bar{X}}{n}\left(1 - \frac{\bar{X}}{15}\right)}\right) =$

$\frac{1}{15}\left(\bar{X} \pm 1,439531\sqrt{\frac{\bar{X}}{25}\left(1-\frac{\bar{X}}{15}\right)}\right)$ (b) $Q = (\bar{X}-\mu)/\sqrt{\frac{\bar{X}}{n}\left(1-\frac{\bar{X}}{15}\right)}$

(c) $IC_\mu = \bar{X} \pm z_{1-\alpha/2}\sqrt{\frac{\bar{X}}{n}\left(1-\frac{\bar{X}}{15}\right)} = \bar{X} \pm 1,439531\sqrt{\frac{\bar{X}}{25}\left(1-\frac{\bar{X}}{15}\right)}$

(d) $ic_p = [0,2312323\ ,\ 0,2967677]$ e $ic_\mu = [3,468484\ , 4,451516\ ]$.

5.9. (a) Seja $X_i =$ o número de dispositivos testados sem defeitos antes de encontrar um com defeito no dia i; $X_1,\dots,X_n$ é amostra aleatória de $X \sim geo(p)$ onde p é o parâmetro de interesse, desconhecido que representa a probabilidade de um dispositivo qualquer ser produzido com defeito nessa fábrica. (b) $Q^a = (\frac{1}{\bar{X}+1} - p)/(\sqrt{\frac{1}{n}\left(1-\frac{1}{\bar{X}+1}\right)^2\left(\frac{1}{\bar{X}+1}\right)})$

(c) $IC_p = \left(\frac{1}{\bar{X}+1}\right) \pm z_{1-\alpha/2}\left(\frac{1}{\bar{X}+1}\right)\sqrt{\frac{1}{n}\left(1-\frac{1}{\bar{X}+1}\right)}$ (d) $ic_{p,95\%} = [0,0303416\ ,\ 0,06789023]$. Estima-se que a probabilidade de um dispositivo da fábrica ser produzido com defeito esteja entre $3,03\%$ e $6,79\%$, com confiabilidade de 95%.

Exercícios da Seção 6.1

6.1.1. $\alpha = 0.0039$ e $\beta = 0.684$.

6.1.2. $\alpha = 0.0002$ e $\beta = 0.49$.

Exercícios da Seção 6.2

6.2.1. (a) $R_A = \{\mathbf{x}| \sum_{i=1}^n (x_i-5)^2 \geq c_A\}$, $R_B = \{\mathbf{x}| \sum_{i=1}^n (x_i-5)^2 \leq c_B\}$

(b) $R_A = \{\mathbf{x}| \sum_{i=1}^n (x_i-5)^2 \geq 249,96\}$, $R_B = \{\mathbf{x}| \sum_{i=1}^n (x_i-5)^2 \leq 72,61\}$

(c) $\beta_A = 0,3589898$ e $\beta_B = 0,4863637$ (use o R).

6.2.2. (a) $R = \{\sum_{i=1}^n \ln(x_i) \leq c\}$ (b) $R = \{\sum_{i=1}^n \ln(x_i) \leq -7,85\}$

(c) Como $\sum_{i=1}^{10} \ln(x_i) = -18.348 < -7,85$, rejeita H_0.

Exercícios da Seção 6.3

6.3.1. (a) $\Psi_0 = (0,15]$ e $\Psi_1 = (15,\infty)$ (b) $\Pi(\sigma^2) = 1 - F_W(800/\sigma^2)$, $W \sim \chi^2_{38}$ e para o esboço use o comando `curve(1-pchisq(800/x,df=38),xlim=c(0,20))`

(c) $\alpha = \Pi(15) = 0,05047629$.

6.3.2. (a) $\Psi_0 = \{1\}$ e $\Psi_1 = (0,1)\cup(1,\infty)$ (b) $\Pi(\lambda) = F_W(39\lambda) + 1 - F_W(84\lambda)$, $W \sim \chi^2_{60}$ e para o esboço use o comando

```
curve(pchisq(39*x,df=60) +1 -pchisq(84*x,df=60),xlim=c(0,2))
```

(c) $\alpha = \Pi(1) = 0,03836988$.

Exercícios da Seção 6.4

6.4.1. (a) $R = \{\sum_{i=1}^{9} x_i^2 > 152,271\}$ (b) Como $\sum_{i=1}^{9} x_i^2 = 182,7089$, rejeitamos H_0.

(c) $\Pi(\sigma^2) = 1 - \mathrm{P}\left(W \leq 152,271/\sigma^2\right)$ sendo $W \sim \chi_9^2$. Use esse comando para fazer gráfico da função poder: `curve(1 - pchisq(152.271/x,df=9),xlim=c(0,50))`.

6.4.2. (a) $R = \{\sum_{i=1}^{25} x_i > 33,2\}$ (b) Como $\sum_{i=1}^{25} x_i = 30$, não rejeitamos H_0.

(c) $\pi(\lambda) \approx 1 - \Phi\left(\left(\frac{33,2}{25} - \lambda\right)\frac{5}{\sqrt{\lambda}}\right)$ ou $\pi(\lambda) \approx 1 - \Phi\left(\frac{33,2-25\lambda}{25\sqrt{\lambda}}\right)$ e o comando `curve(1 - pnorm((33.2 - 25*x)/(5*sqrt(x))),xlim=c(0,5))` gera o gráfico aproximado.

Exercícios da Seção 6.5

6.5.1. (a) $R = \left\{\underline{x} | \sum_{i=1}^{n}\left(\frac{x_i-\mu_0}{\sigma_0}\right)^2 \leq \chi^2_{n,1-\alpha/2} \text{ ou } \sum_{i=1}^{n}\left(\frac{x_i-\mu_0}{\sigma_0}\right)^2 \geq \chi^2_{n,\alpha/2}\right\}$

(b) $R = \left\{\underline{x} | \sum_{i=1}^{40}\left(\frac{x_i-\mu_0}{\sigma_0}\right)^2 \leq 20,707 \text{ ou } \sum_{i=1}^{40}\left(\frac{x_i-\mu_0}{\sigma_0}\right)^2 \geq 66,766\right\}$.

Exercícios da Seção 6.6

6.6.1. (a) $RC = \{\mathbf{x} \in S | 5 < \frac{2n\bar{X}}{q_{1-\alpha/2,2n}}$ ou $5 > \frac{2n\bar{X}}{q_{\alpha/2,2n}}\}$, para a amostra apresentada não rejeitamos H_0. (b) (c)

6.6.3. (a) $R = \{\mathbf{x} \in S \mid f_{1-\alpha,m-1,n-1} < s_Y^2/s_X^2\}$ (b) Para o amostra temos: $f_{1-\alpha,m-1,n-1} = 2,724678$ e $s_Y^2/s_X^2 = 5,261332$. Por isso rejeitamos H_0, ou seja, rejeitamos a hipótese de que $\sigma_X^2 \geq \sigma_Y^2$.

6.6.4. (a) " Rejeita se $|\bar{x} - \bar{y}| > 1.96\sqrt{\frac{9}{n} + \frac{25}{m}}$ " (b) Não rejeitamos H_0, isto é, não rejeitamos a hipótese de que $\mu_X = \mu_Y$.

Exercícios da Seção 6.7

6.7.1. valor-p $= 5,04 \times 10^{-8}$.

6.7.2. valor-p $= 0,0161431$.

Mais Alguns Exercícios do Capítulo 6

6.1. (a) $R = \{-2\sum_{i=1}^{n} \ln(x_i) < q_{\alpha/2,2n}$ ou $-2\sum_{i=1}^{n} \ln(x_i) > q_{1-\alpha/2,2n}$

(b) valor-p $= 0,1367725$, logo não rejeita H_0. .

6.2. (a) $R = \{x \in S \mid \sum_{i=1}^{9} x_i^3 < 43,46\}$ (b) $\Pi(\theta) = F_W(86,92\theta^3)$, onde F_W é a função de distribuição acumulada e $X \sim \chi^2_{18}$. (c) $\sum_{i=1}^{9} x_i^3 = 44,95805 > 43,46$, logo não há evidência estatística para rejeitar H_0.

6.3 (a) $R = \{x \in S \mid \sum_{i=1}^{14} \ln\left(\frac{x_i}{4}\right) < 7,65$ ou $\sum_{i=1}^{14} \ln\left(\frac{x_i}{4}\right) > 22,23\}$ (b) Com $\sum_{i=1}^{14} \ln\left(\frac{x_i}{4}\right) = 6,235565 < 7,65$ rejeita-se H_0. (c) valor-p $= 0,0100662$. Como valor-p $<$ o nível de significância, confirmamos a decisão de rejeitar H_0.

6.4. (a) $R = \{x \in S \mid \sum_{i=1}^{25} X_i < 25 \times 15 \times 0,3 + z_{0,05}\sqrt{25 \times 15 \times 0,3 \times 0,7}\} =$

$\{x \in S \mid| \sum_{i=1}^{25} X_i < 97,90\}$. (b) $R = \{x \in S \mid 4,5 > \bar{X} - z_{0,05}\sqrt{\frac{\bar{X}}{25}\left(1 - \frac{\bar{X}}{15}\right)}\}$ (c) Pela regra de decisão do teste do item (a), $\sum_{i=1}^{25} X_i = 99 > 97,90$ logo não rejeitamos H_0. Pela regra de decisão do item (b) $\bar{X} = 3,96$ e $\bar{X} - z_{0,05}\sqrt{\frac{\bar{X}}{25}\left(1 - \frac{\bar{X}}{15}\right)} = 4,521622 > 4,5$, logo não rejeitamos H_0.

6.5. $R = \{\underline{x} | 4,5 > \bar{x} + \frac{s}{\sqrt{13}} 1,356217\} = \{\underline{x} | 4,5 > \bar{x} + s0,376\}$, não rejeita H_0.

6.6. Temos $X \sim Ber(p)$, $H_0 : p = 0.5$ contra $H_1 : p > 0.5$. Usando o teste baseado no IC assintótico unilateral com limite inferior para p encontramos $IC = (0.5399, \infty)$ e decidimos por rejeitar H_0 com α =5%.

6.7. Queremos testar $H_0 : p_1 = p_2$ contra $H_1 : p_1 > p_2$. Usando o teste baseado no IC assintótico unilateral com limite inferior para $p_1 - p_2$ encontramos $IC = (0.0003055526, \infty)$ e decidimos por rejeitar H_0 e concluir que a vacina é eficaz.

6.8. (a) É recolhida uma amostra aleatória $X_1, \ldots, X_n$ da população $X \sim Poisson(\lambda)$, sendo X_i = número de sinistros da empresa i no último ano, com $n = 103$. λ é o parâmetro desconhecido e representa o número médio de sinistros ocorridos em um ano. (b) $H_0 : \lambda \leq 42$ contra $H_1 : \lambda > 42$ (c) A partir do teste baseado no IC unilateral à direita aproximado para λ, a regra de decisão do teste será: " Rejeita H_0 se $42 < \bar{x} - z1 - \alpha\sqrt{\frac{\bar{x}}{n}}$". (d) Usando o nível de significância de 5%, calculamos $\bar{x} - z_{1-\alpha}\sqrt{\frac{\bar{x}}{n}} = 4.617 - 1,644854\sqrt{\frac{4.617}{103}}$ $= 43,04041 > 42$, logo a hipótese nula é rejeitada. (e) Baseado na amostra levantada, com significância de 5%, rejeitamos a hipótese de que o número de sinistros ocorridos por empresa por ano seja menor ou igual a 42. Com isso há evidência estatística para acreditar que o número de sinistros por ano por empresa seja maior que 42.

Referências Bibliográficas

[Bickel e Doksum, 2001] Bickel, P. J. e Doksum, K. A. (2001). *Mathematical Statistics Basic Ideas and Selected Topics*, volume I. Pearson Prentice Hall, second edição.

[Bolfarine e Sandoval, 2001] Bolfarine, H. e Sandoval, M. (2001). *Introdução à inferência estatística*. Coleção Matematica Aplicada. SBM.

[Bortolossi, 2002] Bortolossi, H. J. (2002). *Cálculo Diferencial à várias variáveis*. Edicoes Loyola.

[Casella e Berger, 2002] Casella, G. e Berger, R. L. (2002). *Statistical Inference*. Duxbury, 2ª edição.

[Larson, 1982] Larson, H. J. (1982). *Introduction to Probability Theory and Statistical Inference*. John Wiley & Sons, 3ª edição.

[Lehmann e Romano, 2006] Lehmann, E. L. e Romano, J. P. (2006). *Testing statistical hypotheses*. Springer Science & Business Media.

[Magalhães, 2011] Magalhães, M. N. (2011). *Probabilidade e Variáveis Aleatórias*. Edusp, 3ª edição.

[Mood, 1974] Mood, A. M. (1974). *Introduction to the theory of statistics*. McGraw-Hill, 3ª edição.

[Roussas, 2003] Roussas, G. (2003). *An Introduction to Probability and Statistical Inference*. Statistics/Mathematics. Academic Press.