Anotações de Inferência Frequentista

Aulas de Alexandre G. Patriota, digitado por Gustavo S. Garone

2025-06-25

Índice

Prefácio

Este é um livrete feito na plataforma Quarto para minhas anotações das disciplinas de inferência frequentista ou clássica MAE0225 e MAE0301, ambas ministradas pelo Professor Alexandre Galvão Patriota e de sua autoria.

Dúvidas, sugestões, críticas ou erros por favor contactar em gustavo.garone arroba usp.br ou abrir um Issue/PR no GitHub

Atribuição e Licença

Patriota, A.G., Garone, G.S. (2024, 2025). Notas do curso de inferência clássica (MAE0225, MAE0301) ministradas pelo Prof. A.G. Patriota digitadas por G.S. Garone.

Agradecimentos

Ao Andrey Sarmento por suas contribuições em revisar e melhorar o projeto.

Parte I

Introdução à Inferência Frequentista

1 Modelos Estatísticos na abordagem clássica

Em teoria de probabilidades, conhecemos a medida de probabilidade, logo, fazemos descrições probabilísticas.

$$(\Omega, \mathcal{A}, P) \xrightarrow{\text{X}} (\mathbb{R}, \mathcal{B}, P_X)$$

Na prática, contudo, não conhecemos a medida P.

Definimos então uma família de medidas de probabilidades que possivelmente descrevem o comportamento aleatório dos dados.

O *Modelo Estatístico* é definido pela trinca

$$(\Omega, \mathcal{A}, \mathcal{P})$$

em que Ω é o espaço amostral (evento certo), $\mathcal{A}$ é a Sigma álgebra, uma família de subconjuntos ou eventos em Ω, e $\mathcal{P}$ é uma família de medidas de probabilidade que possivelmente descrevem o comportamento aleatório dos dados ou eventos sob investigação.

1.1 Modelo Estatístico Paramétrico

Se $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, em que $\Theta \subseteq \mathbb{R}^p$ e $p \in \mathbb{N}$, então dizemos que $(\Omega, \mathcal{A}, \mathcal{P})$ é um modelo estatístico **Paramétrico**.

Caso não exista $\Theta \subseteq \mathbb{R}^p$ fixo, então dizemos que o modelo é não-paramétrico.

> **i** Observação
>
> Θ é o espaço paramétrico e θ é o vetor de parâmetros. θ **Não** é variável aleatória, apenas indexa as medidas de probabilidade.

1.1.1 Exemplos

1.1.1.1 Exemplo de Bernoulli

Considere um Ensaio de Bernoulli

$$\Omega = \{S, F\}, \mathcal{A} = 2^\Omega$$

Temos algum conhecimento prévio que sugere que as probabilidades de sucesso podem ser $0.1, 0.5, 0.9$

Nesse caso,

$$\mathcal{P} = \{P_1, P_2, P_3\}$$

em que

$$\begin{cases} P_1(\{S\}) = 0.1;\ P_1(\{F\}) = 0.9;\ P_1(\Omega) = 1 \\ P_2(\{S\}) = 0.5;\ P_2(\{F\}) = 0.5;\ P_2(\Omega) = 1 \\ P_3(\{S\}) = 0.9;\ P_3(\{F\}) = 0.1;\ P_3(\Omega) = 1 \end{cases}$$

Observe que

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$$

em que $\Theta = \{1, 2, 3\} \subseteq \mathbb{R}$ Portanto, $(\Omega, \mathcal{A}, \mathcal{P})$ é um modelo paramétrico.

Usando de variáveis aleatórias,

Seja $X : \Omega \to \mathbb{R}$ tal que

$$X(\omega) = \begin{cases} 1, \omega = \text{S} \\ 0, c.c. \end{cases}$$

$$E_\theta(X) = \sum_{x=0}^{1} x P_\theta(X = x)$$

$$\begin{cases} \theta = 1 \Rightarrow E_1(X) = 0.1 \\ \theta = 2 \Rightarrow E_2(X) = 0.5 \\ \theta = 3 \Rightarrow E_3(X) = 0.9 \end{cases}$$

1.1.1.2 Exemplo de Exponencial

Seja $\Omega = (0, \infty)$ e $\mathcal{A}$ uma sigma-álgebra de Ω (Sigma-Álgebra de Borel) Ω representa o tempo até a ocorrência de um evento (uma reclamação, por exemplo) Temos conhecimento prévio de que as funções densidade de probabilidade que possivelmente descrevem esse evento são:

$$f_1(\omega) = \begin{cases} \mathrm{e}^{-1\omega}, \omega > 0 \\ 0, c.c. \end{cases}$$

$$f_2(\omega) = \begin{cases} 2\mathrm{e}^{-2\omega}, \omega > 0 \\ 0, c.c. \end{cases}$$

$$f_3(\omega) = \begin{cases} \frac{1}{2}\mathrm{e}^{-\frac{1}{2}\omega}, \omega > 0 \\ 0, c.c. \end{cases}$$

$$f_4(\omega) = \begin{cases} \frac{1}{10}\mathrm{e}^{-\frac{1}{10}\omega}, \omega > 0 \\ 0, c.c. \end{cases}$$

e Ps,

$$P_1(A) = \int_A f_1(\omega)d\omega$$
$$P_2(A) = \int_A f_2(\omega)d\omega$$
$$P_3(A) = \int_A f_3(\omega)d\omega$$
$$P_4(A) = \int_A f_4(\omega)d\omega$$

Dessa forma,

$$P_\theta(A) = \int_A f_\theta(\omega)d\omega$$

e $\Theta = \{1, 2, 3, 4\}$ Seja $X : \Omega \to \mathbb{R}$ tal que

$$X(\omega) = \omega$$

Note que

$$E_\theta(X) = \int_{-\infty}^{\infty} x f_\theta(x)dx, \theta \in \{1, 2, 3, 4\}$$

$$E_\theta(X) = \begin{cases} 1, \ \theta = 1 \\ \frac{1}{2}, \ \theta = 2 \\ 2, \ \theta = 3 \\ 10, \ \theta = 4 \end{cases}$$

1.2 Principais Modelos Estatísticos

1.2.1 Modelo Estatístico de Bernoulli

Dizemos que X é uma variável aleatória com modelo de Bernoulli se, e somente se,

$$P_\theta(X = x) = \begin{cases} \theta^x \cdot (1-\theta)^{1-x}, x \in \{0, 1\} \\ 0, x \notin \{0, 1\} \end{cases}$$

O parâmetro é a probabilidade de sucesso

$$\begin{cases} E_\theta(X) = \theta \\ \text{Var}_\theta(X) = \theta(1-\theta) \\ P_\theta(X = 1) = \theta, P_\theta(X = 0) = 1 - \theta \end{cases}$$

Notação: $\text{Ber}(\theta)$ em que $\theta \in \Theta = (0, 1) \subseteq \mathbb{R}$

1.2.2 Modelo Estatístico Binomial

Dizemos que X é uma variável aleatória com modelo binomial se, e somente se,

$$P_\theta(X = x) = \begin{cases} \binom{n}{x} \cdot \theta^x \cdot (1-\theta)^{n-x}, x \in \{0, 1, \ldots, n\} \\ 0, x \notin \{0, 1, \ldots, n\} \end{cases}$$

$$\begin{cases} E_\theta(X) = n\theta \\ \mathrm{Var}_\theta(X) = n\theta(1-\theta) \\ P_\theta(X = 0) = (1-\theta)^n, \ldots, P_\theta(X = n) = \theta^n \end{cases}$$

Notação: $\mathrm{Bin}(n, \theta)$ em que n é conhecido e fixado previamente, θ é a probabilidade de sucesso (parâmetro do modelo) e $\Theta = (0, 1)$ é o espaço paramétrico

1.2.3 Modelo Estatístico Geométrico

Dizemos que X, representando o número de *fracassos* até o primeiro sucesso, é uma variável aleatória com modelo estatístico geométrico se, e somente se,

$$P_\theta(X = x) = \begin{cases} \theta(1-\theta)^{x-1}, x \in \{1, \ldots\} \\ 0, x \notin \{1, \ldots\} \end{cases}$$

$$\begin{cases} E_\theta(X) = \frac{1}{\theta} \\ \mathrm{Var}_\theta(X) = \frac{1-\theta}{\theta^2} \end{cases}$$

Notação: $\mathrm{Geom}(\theta)$ em que θ é o parâmetro do modelo (probabilidade de sucesso) e $\Theta = (0, 1)$ é o espaço paramétrico

1.2.4 Modelo de Poisson

Dizemos que X é uma variável aleatória com modelo estatístico Poisson, se, e somente se,

$$P_\theta(X = x) = \begin{cases} \mathrm{e}^{-\theta} \cdot \frac{\theta^x}{x!}, x \in \{0, 1, \ldots\} \\ 0, x \notin \{0, 1, \ldots\} \end{cases}$$

$$\begin{cases} E_\theta(X) = \theta \\ \mathrm{Var}_\theta(X) = \theta \end{cases}$$

Notação: $\mathrm{Poisson}(\theta)$ em que θ é a taxa média de ocorrência do evento (parâmetro do modelo) e $\Theta = (0, \infty)$, o espaço paramétrico.

1.2.5 Modelo Multinomial

Dizemos que $\boldsymbol{X} = (X_1, \dots, X_k)$ é um Vetor Aleatório com modelo estatístico Multinomial se, e somente se a função de probabilidade é

$$P_\theta(X_1 = x_1, \dots, X_k = x_k) = \begin{cases} \frac{n!}{x_1!x_2!\dots x_k!} \cdot \theta_1^{x_1} \cdot \dots \cdot \theta_k^{x_n} & x_1 + \dots + x_k = n \\ 0, c.c \end{cases}$$

$$\begin{cases} E_\theta(X_i) = n\theta_i \\ \mathrm{Var}_\theta(X_i) = n\theta_i(1-\theta_i) \\ \mathrm{Cov}(X_i X_j) = -n\theta_i\theta_j \end{cases}$$

Notação: Multinomial$(n, \theta_1, \theta_2, \dots, \theta_k)$ em que $\theta_1 + \dots + \theta_k = 1$ e $0 \leq \theta_i \leq 1$, $\forall i = 1, 2, \dots, k$, $\Theta = \{(\theta_1, , \dots, \theta_k) \in \mathbb{R}^k : 0 \leq \theta_i \leq 1, i = 1, \dots, k, \theta_1 + \dots + \theta_k = 1\}$

Esse modelo tem aplicação em modelos de linguagem como o ChatGPT. (k como tamanho do vocabulário, $n = 1$, $\theta_1 =$ probabilidade de escolher o primeiro elemento do vocabulário e assim por diante.)

1.2.6 Modelo Uniforme contínuo

Dizemos que X é uma variável aleatória contínua com modelo estatístico Uniforme em $(\theta_1, \theta_2), \theta_1 \leq x \leq \theta_2$, se, e somente se, a sua Função Densidade de Probabilidade é

$$f_\theta(x) = \begin{cases} \frac{1}{\theta_2 - \theta_1}, x \in (\theta_1, \theta_2) \\ 0, c.c. \end{cases}$$

$$\begin{cases} E_\theta(X) = \frac{\theta_2 + \theta_1}{2} \\ \mathrm{Var}_\theta(X) = \frac{(\theta_2 - \theta_1)^2}{12} \end{cases}$$

Notação: $X \sim U(\theta_1, \theta_2)$, em que $\theta = (\theta_1, \theta_2)$ é o vetor de parâmetros e $\Theta = \{\theta \in \mathbb{R}^2 : \theta_2 > \theta_1\}$ é o espaço paramétrico.

1.2.7 Modelo Exponencial

Dizemos que X é uma variável aleatória contínua com modelo estatístico Exponencial se, e somente se, a sua Função Densidade de Probabilidade é dada por

$$f_\theta(x) = \begin{cases} \theta \mathrm{e}^{-\theta x}, x > 0 \\ 0, c.c. \end{cases}$$

$$\begin{cases} E_\theta(X) = \frac{1}{\theta} \\ \mathrm{Var}_\theta(X) = \frac{1}{\theta^2} \end{cases}$$

Notação: $X \sim \mathrm{Exp}(\theta), \theta > 0$ em que $\theta > 0, \Theta = \{\theta \in \mathbb{R} : \theta > 0\}$

1.2.8 Modelo Normal

Dizemos que X é uma variável aleatória contínua com modelo estatístico Normal com média μ e variância σ^2 se, e somente se, a sua Função Densidade de Probabilidade é dada por

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu)^2}, x \in \mathbb{R}$$

em que $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}^+$

$$\begin{cases} E_\theta(X) = \mu \\ \mathrm{Var}_\theta(X) = \sigma^2 \end{cases}$$

Notação: $X \sim N(\mu, \sigma^2)$, em que $\theta = (\mu, \sigma^2)$ é o vetor de parâmetros.

2 População e Amostra

Veja: Modelo Estatístico para definições dos modelos estatísticos paramétricos.

2.1 Variável Populacional

Pela teoria estatística, população é o conjunto sob investigação de todos os potenciais elementos.

A *Variável Populacional* representa os valores numéricos de cada elemento da população:

$$X \sim f_{\theta,} \theta \in \Theta$$

em que f_θ é a Função Densidade de Probabilidade da Variável Aleatória populacional. θ é o vetor de parâmetros (desconhecido) e Θ é o espaço paramétrico

2.2 Amostra (Teórica)

Na estatística descritiva, a amostra é definida como um subconjunto da população.

Alguns livros utilizam o termo "amostra representativa" para representar a amostra confiável de outra amostra que carrega determinados viéses de seleção. O termo representativo é controverso na estatística teórica,

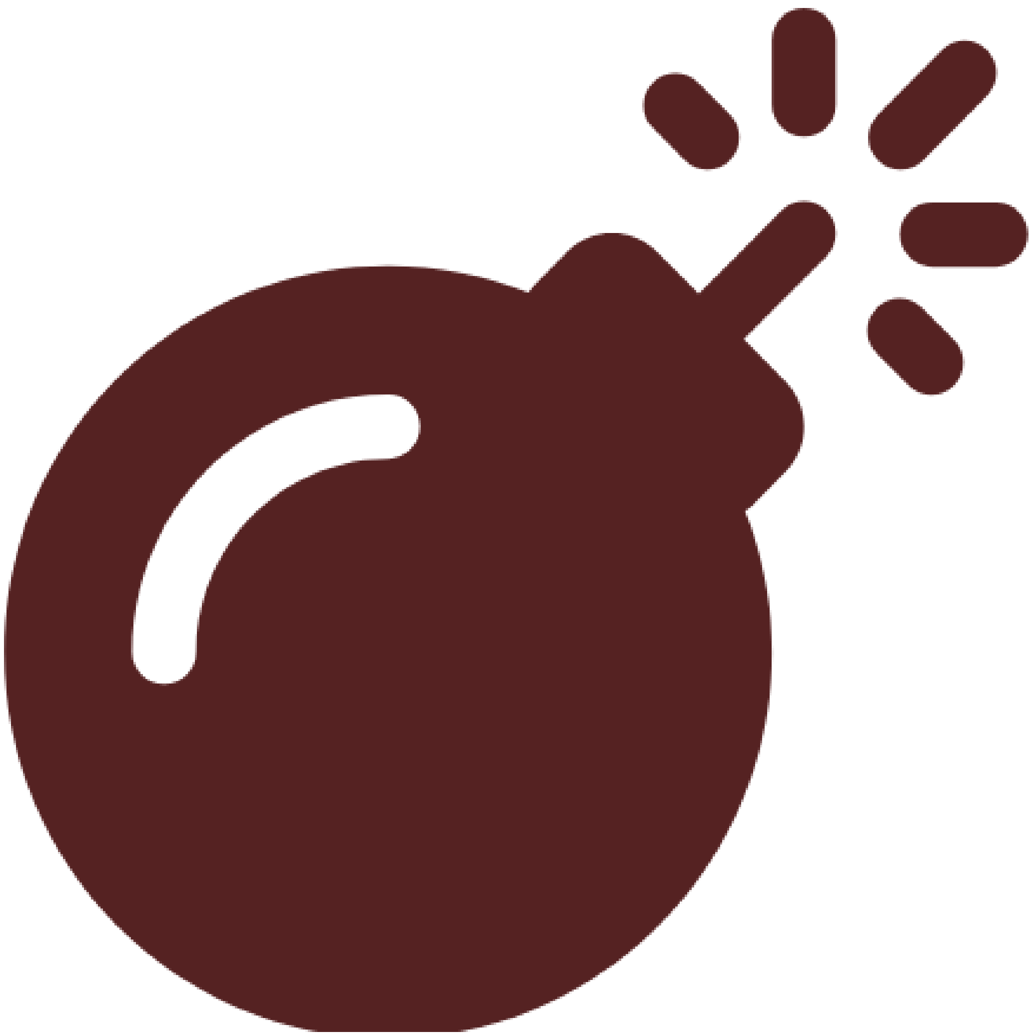
2.2.1 Amostra Aleatória

Na estatística teórica, dizemos que $(X_1, \dots, X_n)$ é uma amostra aleatória de X (v.a. populacional) se $X_1, \dots, X_n$ forem independentes e identicamente distribuídas de acordo com a distribuição da variável aleatória populacional X Ou seja,

$$\text{Independentes } \rightarrow \begin{cases} X_1 \sim f_{\theta,} \theta \in \Theta \\ . \\ . \\ . \\ X_n \sim f_\theta, \theta \in \Theta \end{cases}$$

Sendo assim, $(X_1, \dots, X_n)$ é amostra aleatória de $X \sim f_\theta, \theta in \Theta$ sempre que $(X_1, \dots, X_n)$ forem independentes para cada $P_\theta, \theta \in \Theta$ e $X_i \sim f_\theta, \theta \in \Theta, i = 1, 2, \dots, n$

2.2.1.1 Diagrama



2.3 Amostra (Observada)

É formada por valores numéricos após utilizar um procedimento de amostragem.

$$x_1, \dots, x_n$$

em que n é o tamanho amostral.

3 Quantidade de Interesse

É uma quantidade relacionada com a distribuição da variável aleatória populacional, ou seja, qualquer valor em função de θ.

$$g(\theta)$$

Como $g(\theta) = E_\theta(X), g(\theta) = \text{Var}_\theta(X), g(\theta) = P_\theta(X \geq 1), g(\theta) = \theta$ em que X é a variável aleatória populacional. Empregando outra variável populacional, como Y, conseguimos outras quantidades de interesse, como $g(\theta) = \text{Cov}_\theta(X, y)$ ou $P_\theta(X \in A|Y \in B)$

4 Distribuição Amostral

Seja $(X_1, \dots, X_n)$ amostra aleatória (a.a.) de $X \sim f_{\theta,}\theta \in \Theta$

a Função Densidade de Probabilidade conjunta de $(X_1, X_2, \dots, X_n)$ é $\forall \theta \in \Theta$, no caso discreto:

$$P_\theta(X_1 = k_1, X_2 = k_2, \dots, X_n = k_n) \stackrel{\text{ind}}{=} \prod_{i=1}^{n} P_\theta(X_i = k_i) \stackrel{\text{id}}{=} \prod_{i=1}^{n} P_\theta(X = k_i) \stackrel{\text{def}}{=} \prod_{i=1}^{n} f_\theta(k_i)$$

no caso contínuo:

$$f_\theta^{(X_1,\dots,X_n)}(k_1, k_2, \dots, k_n) \stackrel{\text{i.i.d}}{=} \prod_{i=1}^{n} f_\theta(k_i)$$

4.1 Exemplos

4.1.1 Exemplo um

Seja (X_1, X_2, X_3) uma a.a de $X \sim \text{Ber}(\theta), \theta \in (0,1)$ 1. Especifique o espaço paramétrico 2. Calcule a função de probabilidade da amostra 3. Encontre as seguintes quantidade de interesses em função de θ 1. $g(\theta) = E_\theta(X)$ 2. $g(\theta) = P_\theta(X = 0)$ 3. $g(\theta) = \text{CV}_\theta(X)$ Resolução 1. $\Theta = (0,1)$ 2. Duas resoluções possíveis 1. Dando valores à amostra

(X_1, X_2, X_3)	$P(X_1 = k_1, X_2 = k_2, X_3 = k_3) = \prod_{i=1}^{3} P_\theta(X = k_1)$
$(0,0,0)$	$(1-\theta)^3$
$(0,0,1)$	$(1-\theta)^2\theta$
$(0,1,0)$	$(1-\theta)^2\theta$
$(1,0,0)$	$(1-\theta)^2\theta$
$(0,1,1)$	$(1-\theta)\theta^2$
$(1,0,1)$	$(1-\theta)\theta^2$
$(1,1,0)$	$(1-\theta)\theta^2$
$(1,1,1)$	θ^3

2. Enunciando a função Observe que, se $k \in \{0,1\}$,

$$
\begin{aligned}
&P_\theta(X=k)=\theta^k(1-\theta)^{1-k}\cdot \mathbb{1}_{\{0,1\}}(k) \Rightarrow \\
&P_\theta(X_1=k,X_2=k_2,X_3=k_3) \overset{\text{i.i.d}}{=} \prod_{i=1}^{3}\{\theta^{k_i}(1-\theta)^{1-k_i}\mathbb{1}_{\{0,1\}}(k_i)\} = \\
&=\theta^{\sum\limits_{i=1}^{3} k_i}(1-\theta)^{3-\sum\limits_{i=1}^{3} k_i}\prod_{i=1}^{3}\mathbb{1}_{\{0,1\}}(k_i)
\end{aligned}
$$

3. Em função de θ: 1. $g(\theta)=E_\theta(X)=\theta$ 2. $g(\theta)=P_\theta(X=0)=1-\theta$ 3. $g(\theta)=\mathrm{CV}_\theta(X)=\frac{\sqrt{\theta(1-\theta)}}{\theta}$

4.1.2 Exemplo dois

Seja $(X_1,\dots,X_n)$ uma a.a de $X\sim \text{Pois}(\theta),\theta\in(0,\infty)$, encontre a f.p. conjunta da amostra. Como esse vetor é uma a.a. (ou seja, variáveis aleatórias independentes e identicamente distribuídas), temos que

$$P_\theta(X_1=k_1,\dots,X_n=k_n) \overset{\text{iid}}{=} \prod_{i=1}^{n} P_\theta(X=k_i)=\prod_{i=1}^{n}\{\mathrm{e}^{-\theta}\cdot\frac{\theta^{k_i}}{k_i!}\}$$

Sempre que $k_i\in\{0,1,\dots\},\forall i=1,\dots,n$

$$\Rightarrow P_\theta(X_1=k_1,\dots,X_n=k_n)=\mathrm{e}^{-n\theta}\cdot\frac{\theta^{\sum\limits_{i=1}^{n} k_i}}{\prod_{i=1}^{n}(k_i)!}$$

4.1.3 Exemplo contínuo um

Seja $(X_1,\dots,X_n)$ uma a.a de $X\sim \text{Exp}(\theta),\theta\in(0,\infty)$, encontre a função densidade de probabilidade (f.d.p.) conjunta da amostra.

$$
\begin{aligned}
f_\theta^{(X_1,\dots,X_n)}(k_1,\dots,k_n) &\overset{\text{iid}}{=} \prod_{i=1}^{n} f_\theta(k_i)=\prod_{i=1}^{n}\{\theta\mathrm{e}^{-\theta k_i}\cdot\mathbb{1}_{(0,\infty)}(k_i)\} \\
&\Rightarrow f_\theta^{(X_1,\dots,X_n)}(k_1,\dots,k_n)=\theta^n\cdot \mathrm{e}^{-\theta\sum\limits_{i=1}^{n} k_i}\cdot\prod_{i=1}^{n}\mathbb{1}_{(0,\infty)}(k_i)
\end{aligned}
$$

4.1.4 Exemplo contínuo dois

Seja $(X_1, \dots, X_n)$ uma a.a. (i.i.d) de $X \sim N(\mu, \sigma^2)$ em que $\theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+$. Considere $\underset{\sim}{x} = (x_1, \dots, x_n)$ a amostra observada.

$$
\begin{aligned}
L_{\underset{\sim}{X}}(\theta) &\overset{\text{iid}}{=} \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(x_i - \mu)^2\} \right\} \\
&= \frac{1}{(2\pi\sigma^2)^{\frac{x}{2}}} \cdot \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\}
\end{aligned}
$$

5 Função de Verossimilhança

Quando analisamos a distribuição conjunta da amostra em função de θ nos valores da amostra observada, temos a **Função de Verossimilhança**

$$\mathrm{L}_x(\theta) = P_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

em que $x = (x_1, x_2, \dots, x_n)$ é a amostra observada.

> **Observação**
>
> A função de verossimilhança, no caso discreto, é a probabilidade de observar a amostra observada.

5.1 Exemplo

Considere (X_1, X_2, X_3, X_4) a.a de $X \sim \mathrm{Ber}(\theta), \theta \in \{0.1, 0.5, 0.9\}$. Note que o espaço paramétrico é $\Theta = \{0.1, 0.5, 0.9\}$. Considere, ainda, que a amostra observada foi $(0, 1, 1, 1)$. Encontre a função de verossimilhança.

$$\begin{aligned} L_{x_n}(\theta) &= P_\theta(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4) \\ &= \theta^{\sum_{i=1}^{4} x_i} (1-\theta)^{4-\sum_{i=1}^{4} x_i} \cdot \cancelto{1}{\prod_{i=1}^{4} \mathbb{1}_{\{0,1\}}(x_i)} = \theta^3(1-\theta) \end{aligned}$$

6 Estatísticas

Funções da amostra que não dependem de $\theta \in \Theta$

6.1 Exemplo

Seja $(X_1, \dots, X_n)$ a.a. de $X \sim f_{\theta,}\theta \in \Theta$. São estatísticas: 1. $T_1(X_1, \dots, X_n) = X_1 + \dots + X_n$ 2. $T_2 = \bar{X} = \frac{X_1 + \dots + X_n}{n}$ 3. $T_3 = \max\{X_1, \dots, X_n\} = X_{(n)}$ 4. $T_4 = \min\{X_1, \dots, X_n\} = X_{(1)}$ 5. $T_5(X_1, \dots, X_n) = X_{(n)} - X_{(1)}$ 6. $T_6(X_1, \dots, X_n) = X_i, \text{para algum } i = 1, \dots, n$ 7. $T_7(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$ 8. $T_8(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ 9. $T_9(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} |X_i - \bar{X}|$ 10. $T_{10}(X_1, \dots, X_n) = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2}$ etc.

Observação

As estatísticas são variáveis aleatórias:

$$\boldsymbol{X}_n = (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$$
$$T(\boldsymbol{X}_n) = T \circ \boldsymbol{X}_n : \Omega \to \mathbb{R}$$

7 Estimadores

São estatísticas cujo objetivo é estimar uma quantidade de interesse. Portanto, estimadores são também variáveis aleatórias.

7.1 Estimativas

São os valores observados a partir da amostra observada dos estimadores. Portanto, *estimativas são valores numéricos*

Exemplos: 1. $\bar{X}$ é uma estatística, $\bar{X}$ é um estimador para $g(\theta) = E_\theta(X)$ 2. Observando $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ é uma estimativa Seja (X_1, X_2) a.a de $X \sim \text{Ber}(\theta), \theta \in (0,1)$. Considere as estatísticas e suas funções de probabilidade:

$$\begin{aligned}
&T_1(X_1, X_2) = X_1\\
&T_2(X_1, X_2) = X_2\\
&T_3(X_1, X_2) = X_1 + X_2\\
&T_4(X_1, X_2) = \max\{X_1, X_2\}\\
&T_5(X_1, X_2) = \min\{X_1, X_2\}\\
&T_6(X_1, X_2) = \frac{1}{2}[(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2] = S_n^2 = \frac{1}{2}S_{n-1}^2
\end{aligned}$$

(X_1, X_2)	$P_{\theta X_1, X_2}$	X_1	X_2	$X_1 + X_2$	T_4	T_5	$\bar{X}$	S_n^2	S_{n-1}^2
$(0,0)$	$(1-\theta)^2$	0	0	0	0	0	0	0	0
$(1,0)$	$\theta(1-\theta)$	1	0	1	1	0	0.5	0.25	0.5
$(0,1)$	$\theta(1-\theta)$	0	1	1	1	0	0.5	0.25	0.5
$(1,1)$	θ^2	1	1	2	1	1	1	0	0

Calcule para $T \in \{T_1, T_2, T_3, T_4, T_5, T_6\}$

a-) $E_\theta(T)$

$E_\theta(T_1(X_1, X_2)) = E_\theta(X_1) = (1-\theta)^2 + 1 \cdot \theta(1-\theta) + 0\theta(1-\theta) + 1\theta^2 = \theta$

O mesmo vale para T_2

$E_\theta(T_3(X_1,X_2)) = E_\theta(X_1+X_2) = 2\theta$ $E_\theta(T_4(X_1,X_2)) = E_\theta(\max\{X_1,X_2\}) = 0\cdot(1-\theta)^2+1\cdot[2\cdot\theta(1-\theta)+\theta^2] = 2\theta-\theta^2$ $E_\theta(T_5(X_1,X_2)) = E_\theta(\min\{X_1,X_2\}) = \theta^2$ $E_\theta(T_6(X_1,X_2)) = 2\theta(1-\theta)$

b-) $\text{Var}_\theta(T)$

Termine com os mesmos raciocínios

c-) $P_\theta(T=0)$

Termine com os mesmos raciocínios

Alguns resultados importantes:

$$\begin{aligned}
E_\theta(\bar{X}) &= E_\theta(\frac{X_1+X_2}{2}) = \theta, \theta \in (0,1)\\
E_\theta(\bar{X}^2) &= 0^2(1-\theta)^2 + \frac{2}{4}\theta(1-\theta) + 1^2\theta^2 = \frac{1}{2}\theta + \frac{1}{2}\theta^2, \theta \in (0,1)\\
\Rightarrow \text{Var}_\theta(\bar{X}) &= \frac{\theta+\theta^2}{2} - \theta^2 = \frac{\theta(1-\theta)}{2}, \theta \in (0,1)
\end{aligned}$$

Aplicando em nossa tabela (S^2_{n-1}):

$$\begin{aligned}
E_\theta([(X_1-\bar{X})^2 + (X_2-\bar{X})^2]) &= \theta(1-\theta)\\
E_\theta([(X_1-\bar{X})^2 + (X_2-\bar{X})^2]^2) &= \frac{\theta(1-\theta)}{2}\\
\Rightarrow \text{Var}_\theta([(X_1-\bar{X})^2 + (X_2-\bar{X})^2]) &= \frac{1}{2}\theta(1-\theta)[1-2\cdot\theta(1-\theta)]
\end{aligned}$$

7.2 Propriedades dos estimadores para quantidades de interesse

7.2.1 Estimados não-viciados ou não-enviesados

Seja $(X_1,\ldots,X_n)$ a.a. de $X \sim f_{\theta,}\theta \in \Theta$ e considere $T(X_1,\ldots,X_n) = \hat{\theta}_n$ um estimador para θ. Dizemos que $\hat{\theta}_n$ é *não-enviesado* para $\theta \Leftrightarrow$

$$E_\theta(\hat{\theta}_n) = \theta, \forall\theta \in \Theta$$

De forma geral, $T(X_1,\ldots,X_n)$ é um estimador não-viciado para $g(\theta) \Leftrightarrow$

$$E_\theta(T(X_1,\ldots,X_n)) = g(\theta), \forall\theta \in \Theta$$

Caso contrário, dizemos que $T(X_1,\ldots,X_n)$ é viciado ou enviesado para $g(\theta)$.

Dizemos que $\hat{\theta}_n$ é fracamente consistente para $\theta \Leftrightarrow$

$$\lim_{n\to\infty} P_\theta(|\hat{\theta}_n - \theta| > \epsilon) = 0, \forall\theta \in \Theta$$

e para cada $\epsilon > 0$ fixado.

7.2.1.1 Estimadores não viciados assintoticamente

Dizemos que $T(X_1, \dots, X_n)$ é um estimador assintoticamente não viciado para $g(\theta) \Leftrightarrow$

$$\lim_{n \to \infty} E_\theta(T(X_1, \dots, X_n)) = g(\theta), \forall \theta \in \Theta$$

8 Erro quadrático médio (EQM)

O erro quadrático médio (EQM) do estimador $T(X_1, \dots, X_n)$ com respeito a $g(\theta)$ é definido por

$$\text{EQM}(T, g(\theta)) = E_\theta((T(X_1, \dots, X_n) - g(\theta))^2)$$

> **ℹ Observação**
>
> Se $T(X_1, \dots, X_n)$ for não viciado para $g(\theta)$, então
>
> $$\text{EQM}(T, g(\theta)) = \text{Var}_\theta(T(X_1, \dots, X_n)) \forall \theta \in \Theta$$

8.1 Propriedades do EQM

Seja $T(X_1, \dots, X_n)$ um estimador para $g(\theta)$, seja $\mu_t = E_\theta(T(X_1, \dots, X_n))$

$$\begin{aligned}
&\text{EQM}(T, g(\theta)) \\
&= E_\theta[(T(X_1, \dots, X_n) - \mu_t + \mu_t - g(\theta))^2] \\
\\
&= E_\theta[((T(X_1, \dots, X_n) - \mu_t) + (\mu_t - g(\theta)))^2] \\
&= E_\theta[(T(X_1, \dots, X_n) - \mu_t)^2 + 2(T(X_1, \dots, X_n) - \mu_t)(\mu_t g(\theta)) + (\mu_t - g(\theta))^2] \\
&= \overbrace{E_\theta[(T(X_1, \dots, X_n) - \mu_t)^2]}^{\text{Var}_\theta(T(X_1, \dots, X_n))} + 2(\mu_t - g(\theta)) \cancelto{0}{E_\theta(T(X_1, \dots, X_n) - \mu_t)} + (\mu_t - g(\theta))^2 \\
&= \text{Var}_\theta(T(X_1, \dots, X_n)) + (\mu_t - g(\theta))^2
\end{aligned}$$

Portanto,

$$\text{EQM}(T, g(\theta)) = \text{Var}_\theta(T(X_1, \dots, X_n)) + (\mu_t - g(\theta))^2$$

8.2 Viés

Denotamos de viés de $T(X_1, \dots, X_n)$ com respeito a $g(\theta)$ por

$$\text{Viés}(T, g(\theta)) = E_\theta(T(X_1, \dots, X_n)) - g(\theta), \forall \theta \in \Theta$$

Dessa forma, temos que

$$\text{EQM}(T, g(\theta)) = \text{Var}_\theta(T(X_1, \dots, X_n)) + [\text{Viés}(T(X_1, \dots, X_n), g(\theta))]^2$$

8.3 Exemplo

Seja $(X_1, \dots, X_n)$ uma amostra aleatória, ou seja, independentes e identicamente distribuídas (i.i.d.), de $X \sim \text{Ber}(\theta)$ em que $\theta \in \Theta = (0,1)$. Calcule o viés e o EQM de $\bar{X}_n$ com respeito a $g(\theta) = P_\theta(X = 1)$

O estimador é, então, $T(X_1, \dots, X_n) = \bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ para $g(\theta) = P_\theta(X = 1) = \theta$ (pelo modelo de Bernoulli).

$$\begin{aligned}
E_\theta(\bar{X}_n) &= E_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E_\theta(X_i) \overset{id.dist.}{\Rightarrow} \\
E_\theta(\bar{X}_n) &= \frac{1}{n}\sum_{i=1}^{n} E_\theta(X) \\
&= \frac{n}{n}\theta = \theta, \forall \theta \in \Theta
\end{aligned}$$

Portanto, $\bar{X}_{\theta,n}$ é não enviesado para $g(\theta) = \theta$.

$$\Rightarrow \text{Viés}(\bar{X}_n, g(\theta)) = 0, \forall \theta \in \Theta$$

Para o EQM,

$$\begin{aligned}
\text{EQM}(\bar{X}_n, g(\theta)) = \text{Var}_\theta(\bar{X}_n) - 0^2 &= \text{Var}_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\text{Var}_\theta\left(\sum_{i=1}^{n} X_i\right) \\
&\overset{\text{ind}}{=} \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}_\theta(X_i) \overset{\text{ind. dist.}}{=} \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}_\theta(X), \\
&= \frac{n\theta(1-\theta)}{n^2} = \frac{\theta(1-\theta)}{n}, \forall \theta \in \Theta
\end{aligned}$$

9 Simulações de Monte Carlo

Tem como objetivo replicar artificialmente os dados de um modelo estatístico para estudar o comportamento de estatísticas e estimadores (ou qualquer procedimento estatístico)

9.1 Método

1. Defina o modelo estatístico: "Seja $(X_1, \dots, X_n)$" a.a. de $X \sim f_{\theta,}\theta \in \Theta$.
2. Escolha $\theta_0 \in \Theta$ e considere-o fixado daqui em diante.
3. Para n fixado, gere $(x_1, x_2, \dots, x_n)$ a amostra observada de $X \sim f_{\theta_0}$
4. Armazene a amostra observada
5. Repita 3. e 4. $M = 10000$ vezes (ou quantas vezes desejar, a depender do caso)

10 Estimação via máxima verossimilhança (EMV)

O valor numérico $\hat{\theta}_n$ que maximiza a função de verossimilhança, ou seja, $L_{\underset{\sim}{x}}(\hat{\theta}_n) \geq L_{\underset{\sim}{x}}(\theta) \forall \theta \in \Theta$ é dito ser uma estimativa de máxima verossimilhança (MV) para θ. Observe que $\hat{\theta}_n$ depende da amostra observada e portanto: $\hat{\theta}_n = \hat{\theta}(x_1, x_2, \dots, x_n)$.

O estimador de máxima verossimilhança é obtido substituindo $(x_1, \dots, x_n)$ por $(X_1, \dots, X_n)$, ou seja, $\hat{\theta}(X_1, \dots, X_n)$ é o Estimador de Máxima Verossimilhança (EMV)

10.1 Exemplo

Seja $(X_1, \dots, X_n)$ amostra aleatória (a.a.) de $X \sim f_\theta, \theta \in \{\frac{1}{3}, \frac{1}{2}\}$ em que f_θ é uma função de probabilidade que satisfaz:

$X = x:$	0	1	2
$f_\theta(x):$	θ	θ^2	$1-\theta-\theta^2$

Considere que a amostra observada é $\underset{\sim}{x} = (0, 0, 1)$.

a-) Encontre a estimativa da máxima verossimilhança

Sabemos que

$$f_\theta(x) = \theta^{\mathbb{1}_{\{0\}}(x)} \cdot (\theta^2)^{\mathbb{1}_{\{1\}}(x)} \cdot (1-\theta-\theta^2)^{\mathbb{1}_{\{2\}}(x)} \forall \theta \in \Theta$$

portanto,

$$L_{\underset{\sim}{x}}(\theta) \overset{\text{iid}}{=} \prod_{i=1}^{n} f_\theta(x_i) = \theta^{\sum\limits_{i=1}^{n} \mathbb{1}_{\{0\}}(x_i)} \cdot (\theta^2)^{\sum\limits_{i=1}^{n} \mathbb{1}_{\{1\}}(x_i)} \cdot (1-\theta-\theta^2)^{\sum\limits_{i=1}^{n} \mathbb{1}_{\{2\}}(x_i)} \forall \theta \in \Theta$$

Para $\underset{\sim}{x} = (0, 0, 1)$,

$$L_{\underset{\sim}{x}}(\theta) = \theta^2 \cdot \theta^2 \cdot (1-\theta-\theta^2)^0 = \theta^4 \forall \theta \in \Theta$$

Substituindo $\forall \theta \in \Theta$:

$$\theta = \frac{1}{2} \Rightarrow L_{\underset{\sim}{x}}\left(\frac{1}{2}\right) = \frac{1}{16} \qquad \theta = \frac{1}{3} \Rightarrow L_{\underset{\sim}{x}}\left(\frac{1}{3}\right) = \frac{1}{81}$$

Portanto, $\hat{\theta}_n = \frac{1}{2}$ é a estimativa de máxima verossimilhança.

10.1.1 Invariância dos EMVs

Teorema. Se $\hat{\theta}_{(X_1,\ldots,X_n)}$ for EMV para θ, então $g(\hat{\theta}_{(X_1,\ldots,X_n)})$ é o EMV para $g(\theta)$, ou seja, $g(\hat{\theta}_n)$ é a estimativa de máxima verossimilhança para $g(\theta)$

Mais um exemplo:

Seja $(X_1,\ldots,X_n)$ a.a. de $X \sim N(\mu,\sigma^2)$ em que $\theta = (\mu,\sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}^+$ Assuma que $\underset{\sim}{x} = (x_1,\ldots,x_n)$ é a amostra observada. Lembrando que estaremos chamando $\theta = (\mu,\sigma^2)$, mas estes são parâmetros genéricos. Poderíamos, por exemplo, chamá-los de $\theta = (\theta_1,\theta_2)$, o que pode facilitar a visualizar algumas derivadas.

a-) Encontre as estimativas de máxima verossimilhança para $\theta = (\mu,\sigma^2)$:

A Função de Verossimilhança é:

$$
\begin{aligned}
L_{\underset{\sim}{x}}(\theta) &\overset{\text{iid}}{=} \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{ \frac{-1}{2} \cdot \frac{(x_i-\mu)^2}{\sigma^2} \right\} \right\} \\
&= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \cdot \exp\left\{ \frac{-1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2 \right\}
\end{aligned}
$$

Podemos derivar para encontrar o máximo da FMV. Para isso, derivaremos e igualamos a zero primeiro em relação a μ, então a σ^2 (podemos aplicar o logaritmo para facilitar as operações.)

$$
\begin{aligned}
\frac{\partial \ln(L_{\underset{\sim}{x}})}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i-\mu) = 0 \Rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i \\
\frac{\partial \ln(L_{\underset{\sim}{x}})}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i-\mu)^2 = 0 \\
&\therefore
\end{aligned}
$$

$$
\begin{aligned}
\text{Estimativas MV} &= \begin{cases} \mu = \bar{x} \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum\limits_{i=1}^{n} (x_i-\mu)^2 = 0 \end{cases} \\
&\Leftrightarrow \begin{cases} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = \frac{1}{n} \sum\limits_{i=1}^{n} (x_i-\bar{x})^2 \end{cases}
\end{aligned}
$$

Estes são os pontos que maximizam a Função de Máxima Verossimilhança. (Provados em cálculo), ou seja, são as estimativas de máxima verossimilhança para μ, σ^2 respectivamente, e

$\hat{\mu}(X_1, \dots, X_n) = \bar{X}, \sigma^2(X_1, \dots, X_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ são os estimadores de máxima verossimilhança.

Pela propriedade de invariância podemos encontrar o EMV para $g(\theta) = \frac{\sqrt{\text{Var}_\theta(X)}}{E_{\theta(X)}}$:

$$\widehat{g(\theta)} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})}}{\bar{X}}$$

Observação

Seja $(X_1, \dots, X_n)$ a.a. de $X \sim N(\mu, \sigma^2)$. Então,

1. $\bar{X} \underset{\text{Exata!}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right) \forall \mu, \sigma \in \mathbb{R} : \sigma^2 > 0$ e $n \geq 1$

2. $\sum_{i=1}^{n} \frac{(x_1 - \bar{X})^2}{\sigma^2} \underset{\text{Exata!}}{\sim} \chi^2_{(n-1)}$

em que χ^2_k representa a Distribuição Qui-Quadrado com k grau de liberdade, cuja função densidade de probabilidade é:

$$f(x) = \frac{1}{\Gamma(\frac{k}{2})2^{\frac{k}{2}}} \cdot x^{\frac{k}{2}-1} \cdot \exp\left\{\frac{-x}{2}\right\} \cdot \mathbb{1}_{(0,\infty)}(x).$$

Para qualquer outra distribuição, existe um resultado aproximado pelo Teorema do Limite Central.

11 Distribuição Qui-Quadrado

Se $Z \sim N(0,1)$, então $Z^2 \sim \chi^2$, como provado por transformação de variáveis aleatórias

Se $W \sim \chi^2_k$, então sua função densidade de probabilidade é dada por

$$\begin{cases} \frac{1}{\Gamma(\frac{k}{2})\cdot 2^{\frac{k}{2}}} \cdot w^{\frac{k}{2}-1} \cdot e^{-\frac{1}{2}w}, & w > 0 \\ 0, cc \end{cases}$$

Sendo assim,

$$\begin{cases} E(W) = k \\ \text{Var}(W) = 2k \end{cases}$$

Se $Z_1, Z_{2,} \dots, Z_N \overset{\text{iid}}{\sim} N(0,1)$, então

$$Z_1^2 + \cdots + Z_n^2 \sim \chi^2_n$$

Prova por função característica

Se $X_1, \dots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, então

$$\frac{(X_1-\mu)^2 + \cdots + (X_n-\mu)^2}{\sigma^2} \sim \chi^2_n$$

Se $X_1, \dots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$, então

$$\frac{(X_1-\bar{X})^2 + \cdots + (X_n-\bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

Ademais, se $Y \sim \chi^2_\nu$, então

$$\frac{Y-\nu}{\sqrt{2\nu}} \overset{a}{\approx} N(0,1)$$

para $\nu > 30$

12 Teorema do Limite Central

Também conhecido como Teorema Central do Limite, é fundamental para a teoria da probabilidade e a estatística.

Seja $X_1, \dots, X_n$ a.a de $X \sim f_\theta, \theta \in \Theta : E_\theta(X^2) < \infty$, então:

$$\bar{X} \overset{\text{Aproximadamente}}{\sim} N\left(E_\theta(X), \frac{Var_\theta X}{n}\right)$$

Formalmente, temos o enunciado do *Teorema do Limite Central*:

$$\frac{\sqrt{n}(\bar{X} - E_\theta(x))}{\sqrt{\text{Var}_\theta(X)}} \underset{n\to\infty}{\overset{\text{Distribuição}}{\to}} N \sim (0,1) \forall \theta \in \Theta$$

Se $X \sim N(\mu, \sigma)$, então a distribuição é exata.

Ademais, seja g uma função contínua e diferenciável tal que $g'(\theta) \neq 0$. Então,

$$g(\bar{X}) \overset{\text{Aproximadamente}}{\sim} N\left(g(E_\theta(X)), \frac{g'(E_\theta(X))^2 \text{Var}_\theta(X)}{n}\right)$$

12.1 Exemplo

Seja $(X_1, \dots, X_n)$ a.a de $X \sim Ber(\theta), \theta \in (0,1)$. Já vimos que EMV p/θ é

$$\bar{\theta}(X_1, \dots, X_n) = \bar{X}$$

e o EMV p/ $g(\theta) = \text{Var}_{\theta(x)} = \theta(1-\theta)$ é:

$$\widehat{g(\theta)} = \bar{X}(1 - \bar{X}).$$

Agora,

$$\bar{X} \overset{\text{approx.}}{\sim} N\left(\theta, \frac{\theta(1-\theta)}{n}\right)$$

$$g(\bar{X}) = \bar{X}(1-\bar{X}) \overset{\text{approx}}{\sim} N\left(\theta(1-\theta), \frac{[g'(\theta)]^2 \theta(1-\theta)}{n}\right)$$

$$\Rightarrow N\left(\theta(1-\theta), \frac{(1-2\theta)^2 \theta(1-\theta)}{n}\right)$$

13 Intervalo de Confiança ou Estimador Intervalar

Seja $(X_1, \ldots, X_n)$ a.a. de $X \sim f_\theta, \theta \in \Theta \subseteq \mathbb{R}$.

Um intervalo de confiança (IC) com coeficiente de confiança γ é um intervalo aleatório que satisfaz:

$$P_\theta(I_1(\boldsymbol{X}_n) \leq \theta \leq I_2(\boldsymbol{X}_n)) = \gamma \forall \theta \in \Theta$$

em que $\boldsymbol{X}_n = (X_1 \ldots, X_n)$

> **Observação**
>
> Se o intervalo aleatório $[I_1(\boldsymbol{X}_n), I_2(\boldsymbol{X}_n)$ satisfaz
>
> $$P_\theta(I_1(\boldsymbol{X}_n) \leq \theta \leq I_2(\boldsymbol{X}_n)) \geq \gamma \forall \theta \in \Theta$$
>
> Então o intervalo aleatório é um intervalo de confiança com pelo menos coeficiente de confiança γ

> **Observação**
>
> $I_1(\boldsymbol{X}_n)$ e $I_2(\boldsymbol{X}_n)$ são estatísticas e são tais que
>
> $$I_1(\boldsymbol{X}_n) \leq I_2(\boldsymbol{X}_n)$$

> **Observação**
>
> Quando substituímos $(\boldsymbol{X}_n)$ pela amostra observada $(\boldsymbol{x}_n)$ temos que
>
> $$P_\theta(I_1(\boldsymbol{x}_n) \leq \theta \leq I_2(\boldsymbol{x}_n)) = \begin{cases} 1, & \text{se } \theta \in [I_1(\boldsymbol{x}_n), I_2(\boldsymbol{x}_n)] \\ 0, & \text{caso contrário} \end{cases}$$

13.1 IC sob normalidade para μ com σ^2 conhecido

Em uma distribuição normal (θ, σ^2), por exemplo, conseguimos de forma genérica para qualquer $\gamma \in [0,1]$ encontrar pela tabela um valor de c_γ que satisfaça $P(-c_\gamma \leq Z \leq c_\gamma)$. Assim,

$$-c_\gamma \leq Z \leq c_\gamma \Leftrightarrow -c_\gamma \leq \frac{\sqrt{n}(\bar{X}-\theta)}{\sqrt{\sigma^2}} \leq c_\gamma \Leftrightarrow \bar{X} - c_\gamma \frac{\sqrt{\sigma^2}}{\sqrt{n}} \leq \theta \leq \bar{X} + c_\gamma \frac{\sqrt{\sigma^2}}{\sqrt{n}}$$

Portanto,

$$P_\theta\left(\bar{X} - c_\gamma \frac{\sqrt{\sigma^2}}{\sqrt{n}} \leq \theta \leq \bar{X} + c_\gamma \frac{\sqrt{\sigma^2}}{\sqrt{n}}\right) = \gamma \forall \theta \in \Theta$$

Dessa forma, $\bar{X} - c_\gamma \frac{\sqrt{\sigma^2}}{\sqrt{n}} \leq \theta \leq \bar{X} + c_\gamma \frac{\sqrt{\sigma^2}}{\sqrt{n}}$ é um intervalo de confiança cujo coeficiente de confiança é γ.

13.1.1 Exemplo

Considere que a amostra observada foi $1, 2.2, 3, 3.5$ de uma distribuição $N(\theta, 3)$.

Encontre o IC observado com coeficiente de confiança de 99%.

Primeiro encontramos a média: $\bar{x} = 2.42$.

Dessa forma, $c_{99\%} = 2.58$ e nosso intervalo de confiança é

$$\left[2.42 - 2.58\frac{\sqrt{3}}{\sqrt{4}}, 2.42 + 2.58\frac{\sqrt{3}}{\sqrt{4}}\right] = [0.18, 4.65]$$

Dessa forma, se repetirmos o experimento N vezes, esperamos que $\gamma = 99\%$ dos ICs observados contenham a quantidade de interesse.

13.1.1.1 Notação

$$\text{IC}(\theta, \gamma) = [I_1(\boldsymbol{X}_n), I_2(\boldsymbol{X}_n)]$$

Denotará o intervalo de confiança teórico

$$\text{IC}_{\text{Obs}}(\theta, \gamma) = [I_1(\boldsymbol{x}_n), I_2(\boldsymbol{x}_n)]$$

13.2 IC sob normalidade para μ quando σ^2 é desconhecido

Seja uma a.a de $X \sim N(\mu, \sigma^2)$ em que μ, σ^2 são desconhecidos. Então, o IC para μ com coeficiente de confiança γ é dado por

$$\text{IC}(\mu, \gamma) = \left[\bar{X} - t_{\gamma, n-1} \sqrt{\frac{s^2(\boldsymbol{X}_n)}{n}}, \bar{X} + t_{\gamma, n-1} \sqrt{\frac{s^2(\boldsymbol{X}_n)}{n}} \right]$$

Em que $s^2(\boldsymbol{x}_n) = \frac{1}{n-1} \sum_{i=1}^{k} (X_i - \bar{x})^2$ e $t_{y,(n-1)}$ deve ser calculado da tabela $P(-t_{\gamma,n-1} \leq T_{n-1} \leq t_{\gamma,n-1}) = \gamma$, $T_{n-1} \sim \text{t} - \text{Student}(n-1)$

Se a variância for desconhecida, substituímos σ^2 pelo estimador

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

e o valor de c_γ obtido de uma t-Student com $n-1$ graus de liberdade.

Justificativa:

Se uma a.a. de $X \sim N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$.

1. $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$
2. $\sqrt{n} \frac{\bar{X} - \mu}{\sqrt{\sigma^2}} \sim N(0, 1)$
3. $\sum\limits_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sqrt{\sigma^2}} \sim \chi^2_{n-1}$
4. $\sqrt{n} \frac{\bar{X} - \mu}{\sqrt{s^2}} = \sqrt{n} \frac{\bar{X} - \mu}{\sqrt{s^2}} \cdot \frac{\sqrt{\sigma^2}}{\sqrt{\sigma^2}} = \frac{\sqrt{n} \frac{\bar{X} - \mu}{\sqrt{\sigma^2}}}{\sqrt{\frac{\sum (X_i - \bar{X})^2}{\sigma^2} \cdot \frac{1}{n-1}}}$
5. Se $Z \sim N(0, 1)$ e $W \sim \chi^2_k$, então $t = \frac{Z}{\sqrt{\frac{W}{k}}} \sim t_k$
6. $\sqrt{n} \frac{(\bar{X} - \mu)}{\sqrt{s^2}} \sim t_{(n-1)}$

13.2.1 Exercício

Considere uma a.a de $X \sim N(\mu, \sigma^2)$, ambos parâmetros desconhecidos, em que X é o retorno de um ativo específico. Considere que a seguinte amostra foi observada:

$$(-1.3\%, 0.4\%, -1.7\%, 3.2\%, 0.7\%, -1.6\%, 1.0\%, 1.5\%, 1.2\%, -0.6\%)$$

Encontre o IC para μ, com coeficiente de confiança $\gamma = 99\%$

Temos que $s^2(\boldsymbol{x}_n) = 2.48$ e $\bar{x} = 0.28$. Da tabela, $t_{99\%,9} = 3.25$

Sendo assim,

$$\text{IC}_{\text{Obs}}(\mu, 99\%) = \left[0.28 - 3.25\sqrt{\frac{2.48}{10}}, 0.28 + 3.25\sqrt{\frac{2.48}{10}}\right] = [-1.34, 1.9]$$

13.3 IC sob normalidade para σ^2 com μ desconhecido

Seja uma a.a. de $X \sim N(\mu, \sigma^2)$ em que μ, σ^2 são desconhecidos.

O intervalo de confiança para σ^2 com coeficiente de confiança γ é dado por:

$$\text{IC}(\sigma^2, \gamma) = \left[\frac{(n-1)s^2(\boldsymbol{X}_n)}{q^{(2)}_{\gamma,n-1}}, \frac{(n-1)s^2(\boldsymbol{X}_n)}{q^{(1)}_{\gamma,n-1}}\right]$$

$$\text{IC}_{\text{Obs}}(\sigma^2, \gamma) = \left[\frac{(n-1)s^2(\boldsymbol{x}_n)}{q^{(2)}_{\gamma,n-1}}, \frac{(n-1)s^2(\boldsymbol{x}_n)}{q^{(1)}_{\gamma,n-1}}\right]$$

em que $s^2(\boldsymbol{X}_n) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ e $s^2(\boldsymbol{x}_n) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ E $q^{(1)}_{\gamma,n-1}, q^{(2)}_{\gamma,n-1}$ são obtidos calculando $P(q^{(1)}_{\gamma,n-1} \leq W \leq q^{(2)}_{\gamma,n-1}) = \gamma$ no qual $W \sim \chi^2_{n-1}$ e $P(\chi^2_{n-1} \leq q^{(1)}_{\gamma,n-1}) = \frac{1-\gamma}{2} = P(\chi^2_{n-1} \geq q^{(2)}_{\gamma,n-1})$

Demonstração:

Vamos construir um IC para a variância.

Note que

$$\sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sqrt{\sigma^2}} \sim \chi^2_{n-1}$$

Note ainda que $W \sim \chi^2_{n-1}$, ou seja, $P(q^{(1)}_{\gamma,n-1} \leq W \leq q^{(2)}_{\gamma,n-1}) = \gamma$. Assim, $W = \frac{(n-1)s^2}{\sigma^2}$.

Dessa forma,

$$\frac{1}{q^{(2)}_{\gamma,n-1}} \leq \frac{\sigma^2}{(n-1)s^2} \leq \frac{1}{q^{(1)}_{\gamma,n-1}} \Leftrightarrow \frac{(n-1)s^2}{q^{(2)}_{\gamma,n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{q^{(1)}_{\gamma,n-1}}$$

em que $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Portanto,

$$P_\theta\left(\frac{(n-1)s^2}{q^{(2)}_{\gamma,n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{q^{(1)}_{\gamma,n-1}}\right) = \gamma \forall \theta \in \Theta$$

Exercício:

Considere uma a.a de $X \sim N(0, \theta), \theta > 0$ em que X é o retorno de um ativo específico. Considere que a seguinte amostra foi observada:

$$(-1.3\%, 0.4\%, -1.7\%, 3.2\%, 0.7\%, -1.6\%, 1.0\%, 1.5\%, 1.2\%, -0.6\%)$$

Construa um intervalo de confiança para a variância (populacional) θ com coeficiente de confiança $\gamma = 95\%$.

O IC para a variância é

$$\text{IC}(\theta, \gamma) = \left[\frac{(n-1)s^2(\boldsymbol{x}_n)}{q^2_{\gamma,n-1}}, \frac{(n-1)s^2(\boldsymbol{x}_n)}{q^1_{\gamma,n-1}} \right]$$

em que $s^2(\boldsymbol{x}_n) = \frac{1}{n-1}\sum_{i=1}^k (X_i - \bar{x})^2$ e $q^1_{\gamma,n-1}, q^2_{\gamma,n-1}$ satisfazem as fronteiras que delimitam uma área de γ em torno da média. Nesse caso, como $n = 10$, $q^1_9 = 2.7, q^2_9 = 19.023$. Dessa forma, $\bar{x} = 0.28, \sum_i x_i^2 = 2.308 \Rightarrow s^2(\boldsymbol{x}_n) = \frac{10}{9}(2.308 - 0.28^2) = 2.48$

Sendo assim

$$\text{IC}_{\text{Obs}}(\theta, 95\%) = \left[\frac{9 \cdot 2.48}{19.023}, \frac{9 \cdot 2.48}{2.7} \right] = [1.17, 8.27]$$

Podemos concluir que, com uma confiança de 95%, a variância está nesse intervalo.

13.4 Intervalo de Confiança para a proporção

Seja $(\boldsymbol{x}_n)$uma amostra aleatória de $X \sim Ber(\theta)$ em que $\theta \in [0, 1]$. Sabemos que, pelo Teorema do Limite Central

$$\bar{X} \overset{a}{\approx} N\left(\theta, \frac{\theta(1-\theta)}{n}\right)$$

Dessa forma,

$$\frac{\sqrt{n}(\bar{X} - \theta)}{\sqrt{\theta(1-\theta)}} \overset{a}{\approx} N(0, 1)$$

Além disso, pelo Teorema de Slutsky

$$\frac{\sqrt{n}(\bar{X} - \theta)}{\sqrt{\bar{X}(1-\bar{X})}} \overset{a}{\approx} N(0, 1)$$

Observe que, se $Z \sim N(0,1)$ então $-c_\gamma \leq Z \leq c_\gamma \Leftrightarrow -c_\gamma \leq \frac{\sqrt{n}(\bar{X}-\theta)}{\sqrt{\bar{X}(1-\bar{X})}} \leq c_\gamma$

$$\Leftrightarrow \bar{X} - c_\gamma \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq \theta \leq \bar{X} + c_\gamma \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$$

Então, seja c_γ tal que

$$\begin{aligned} &P(-c_\gamma \leq N(0,1) \leq c_\gamma) = \gamma \\ \Rightarrow &P_\theta \left(\bar{X} - c_\gamma \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \leq \theta \leq \bar{X} + c_\gamma \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \right) \approx \gamma \ \forall \theta \in \Theta = [0,1] \end{aligned}$$

que melhora conforme $n \to \infty$

Logo, um IC aproximado com coeficiente de confiança γ é dado por

$$IC(\theta, \gamma) = \left[\bar{X} - c_\gamma \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}, \bar{X} + c_\gamma \sqrt{\frac{\bar{X}(1-\bar{X})}{n}} \right] \cap \Theta$$

13.4.1 Exemplo

Seja $(\boldsymbol{x}_n)$ uma amostra aleatória de $X \sim \text{Ber}(\theta)$ em que $\theta \in [0,1]$.

$$X(w) = \begin{cases} 1, \text{ se w disser que vota no candidato} \\ \\ 0, c.c \end{cases}$$

A amostra observada foi $(0,0,0,1,0,0,0,1)$. Encontre o IC para a proporção de intenção de votos no candidato considerando $\gamma = 99\%$

$\bar{x} = \frac{1}{4}, \bar{x}(1-\bar{x}) = \frac{3}{16}, n = 8, c_\gamma = 2.58$

$$\begin{aligned} \text{IC}_{\text{Obs}}(\theta, 99\%) &= \left[0.25 - 2.58 \frac{\sqrt{3}}{4\sqrt{8}}, 0.25 - 2.58 \frac{\sqrt{3}}{4\sqrt{8}} \right] \cap [0,1] \\ &= [0.25 - 0.39, 0.25 + 0.39] \cap [0,1] \\ &= [0, 0.64] \end{aligned}$$

13.4.2 Intervalos conservadores e otimistas

Um intervalo de confiança de proporção é dito ser conservador quando o calculamos tomando θ ***Na variância*** $(\theta(1-\theta))$ como o valor mais alto possível. No exemplo Bernoulli, o valor máximo para θ (derivando $\theta(1-\theta)$ e igualando a 0, $1-2\theta=0$) é $\theta=0.5$. Dessa forma, o IC conservador é calculado usando $\theta=0.5$.

Por sua vez, um IC otimista é calculado usando o valor de θ obtido através do EMV para θ, no caso Bernoulli, usaríamos essa variância como $\bar{X}(1-\bar{X})$

13.5 Como interpretar intervalos de confiança

Importante:

Na estatística clássica (frequentista), devemos interpretar um intervalo de confiança $[a,b]$ com $\gamma=0.95$ da seguinte forma:

"Com 95% de confiança, o intervalo $[a,b]$ conterá o valor da quantidade de interesse".

Isso é importante para diferenciar a interpretação frequentista (Theta do espaço paramétrico) da Bayesiana (Theta como variável aleatória). Dessa forma, estaria ***incorreto*** na estatística clássica dizer que

"O intervalo $[a,b]$ conterá a quantidade de interesse com probabilidade 95%" ou "O intervalo $[a,b]$ conterá a quantidade de interesse 95% das vezes"

14 Teste de Hipótese simples

Um dos principais objetivos da estática é testar hipóteses. Veja algumas dessas hipóteses potenciais:

1. A moeda é honesta?
2. O medicamento proposto é melhor que o vendido no mercado?
3. O número médio de acidentes aumentou em relação ao ano passado?
4. A altura interfere na performance num determinado esporte?
5. Um suspeito é culpado?
6. O mercado financeiro está em equilíbrio?
7. Dona Maria terá dinheiro para comprar o pão do próximo mês?

14.1 Etapas de um teste de hipótese

1. Formular as hipóteses de interesse;
    1. Na estatística clássica, pela abordagem Fisheriana (uma hipótese) ou Neyman-Pearson (mais de uma hipótese).
2. Observar dados experimentais do estudo relacionado ao problema;
3. Elaborar uma conclusão utilizando um procedimento estatístico.

14.2 Exemplo da hipótese 5.

Considere uma pessoa que está sendo acusada de ter cometido um crime.

As duas hipóteses envolvidas aqui são (abordagem de Neyman-Pearson):

1. "O suspeito não é culpado" -> h_0 Hipótese Nula ou de não-efeito;
2. "O suspeito é culpado" -> h_1 Hipótese alternativa ou hipótese que contém o efeito.

- Após coletar as evidências, dizemos que, se houver evidências de que o suspeito cometeu o crime, a pessoa é culpada.
- Se não, concluímos que não é culpado.

Contudo, devemos nos atentar aos erros de decisão

14.2.1 Erros de decisão

	H_0	H_1
Decisão	Não cometeu o crime	Cometeu o crime
Inocente	Acerto	Erro Tipo II
Culpado	Erro Tipo I	Acerto

- Erro Tipo I: Decidir que o acusado é culpado quando na verdade é inocente (Rejeitar H_0). - Erro Tipo II: Decidir que o acusado é inocente quando na verdade é culpado (Rejeitar H_1).

14.3 Exemplo da hipótese 1.

Estamos interessados em verificar se uma moeda é honesta. Executaremos n experimentos de Bernoulli e verificaremos se a face voltada para cima após o lançamento é cara. Dessa forma, sendo X o resultado dum lançamento, teremos a a.a ($\boldsymbol{X}_n$) de $X \sim \text{Ber}(\theta), \theta \in \Theta = [0, 1]$. Suspeitamos que a moeda é honesta ou que $\theta = 0.9$.

Nossas hipóteses são

1. H_0 -> $\theta = 0.5$
2. H_1 -> $\theta = 0.9$

14.3.1 Erros Tipo I e II

Note que $\bar{X}$ é um estimador para θ e $\bar{x}$ é uma estimativa. - Se $\bar{x} > 0.7$, rejeitaremos a hipótese nula h_0, a moeda não seria honesta e haveria um viés agindo sobre seus lançamentos. - Se $\bar{x} \leq 0.7$, concluiremos que a moeda é honesta.

	H_0	H_1
Decisão	Honesta	Viesada
$\bar{x} < 0.7$	Acerto	Erro Tipo II
$\bar{x} \geq 0.7$	Erro Tipo I	Acerto

- Erro Tipo I: Rejeitar que a moeda é honesta (rejeitar h_0) quando na verdade é.
- Erro Tipo II: Rejeitar que a moeda é enviesada (rejeitar h_1) quando na verdade é.

Calculando a probabilidade dos erros

$$
\begin{aligned}
P(\text{Erro Tipo I}) &= P(\text{Probabilidade de rejeitar } h_0 | h_0 \text{ é verdadeiro}) \\
\text{Incorreto na Estatística Clássica} &= P(\bar{X} > 0.7 | \theta = 0.5) \\
\text{Correto na Estatística Clássica} &= P_{0.5}(\bar{X} > 0.7)
\end{aligned}
$$

Na segunda notação, correta na estatística frequentista, P está sob a hipótese nula h_0 verdadeira. Dessa forma

$$P(\text{Erro Tipo II}) = P_{0.9}(\bar{X} \leq 0.7)$$

Calcule as probabilidades dos erros considerando $n = 10$ e a aproximação pela distribuição normal.

14.3.1.1 Calculando Exato e pela aproximação do Teorema do Limite Central

Sabemos que $\sum_{i=1}^{n} X_i \sim \text{Bin}(n, \theta)$, logo

$$
\begin{aligned}
\alpha = P(\text{Erro Tipo I}) &\overset{\text{Sob } h_0}{=} P_{0.5}\left(\frac{1}{n}\sum_{i=1}^{n} X_i > 0.7\right) = P_{0.5}\left(\sum_{i=1}^{10} X_i > 7\right) \\
&= P_{0.5}\left(\sum_{i=1}^{10} X_i \geq 8\right) \\
&= \binom{10}{8} 0.5^8 \cdot 0.5^2 + \binom{10}{9} 0.5^9 \cdot 0.5^1 + \binom{10}{10} 0.5^{10} \approx 0.05469 \\
\alpha &= P_{0.5}\left(\sqrt{\frac{n}{0.25}}(\bar{X} - 0.5) > \sqrt{\frac{n}{0.25}}(0.7 - 0.5)\right) \\
&\approx P(N(0,1) > 1.26) \approx 0.103
\end{aligned}
$$

$$
\begin{aligned}
\beta = P(\text{Erro Tipo II}) &\overset{\text{Sob } h_1}{=} P_{0.9}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \leq 0.7\right) = P_{0.9}\left(\sum_{i=1}^{10} X_i \leq 7\right) \\
&= 1 - P_{0.9}\left(\sum_{i=1}^{10} X_i \geq 8\right) \\
&= 1 - \left(\binom{10}{8} 0.9^8 \cdot 0.9^2 + \binom{10}{9} 0.9^9 \cdot 0.9^1 + \binom{10}{10} 0.9^{10}\right) \approx 0.0702 \\
\alpha &= P_{0.9}\left(\sqrt{\frac{n}{0.09}}(\bar{X} - 0.9) \leq \sqrt{\frac{n}{0.09}}(0.7 - 0.9)\right) \\
&\approx P(N(0,1) \leq -2.1) \approx 0.018
\end{aligned}
$$

14.4 Poder do teste

Chamamos de poder do teste a probabilidade de rejeitar h_0 quando este é falso.

No exemplo anterior,

$$\pi = P_{0.9}(\bar{X} > 0.7) = 1 - P_{0.9}(\bar{X} \leq 0.7) = 1 - \beta = 92.92\%$$

Considere nesse exemplo uma a.a do lançamento de quatro moedas: $(\boldsymbol{x}_{10}) = (1, 0, 1, 0, 0, 1, 1, 0, 0, 0)$. Como $\bar{x} = 0.4 \leq 0.7$, não rejeitamos a hipótese nula h_0.

Em uma outra amostra, $(\boldsymbol{x}_{10}) = (0, 0, 1, 1, 1, 1, 1, 1, 1, 1)$. Como $\bar{x} = 0.8 > 0.7$, rejeitamos a hipótese nula h_0.

14.5 Um exemplo mais amplo (Diferença)

Seja (X_n) uma a.a. de $X \sim \text{Ber}(\theta)$, em que $\theta \in (0, 1)$. Considere as hipóteses:

$$\begin{cases} H_0 : \theta = 0.5 \\ H_1 : \theta \neq 0.5 \end{cases}$$

Decisões elaboradas:

1. Se $\bar{x} < 0.3$ ou $\bar{x} > 0.7$, rejeitamos H_0
2. Caso contrário, não rejeitamos H_0

Relembrando: $\alpha =$ Probabilidade do Erro Tipo I (Rejeitar um H_0 verdadeiro). β Probabilidade do Erro Tipo II (Rejeitar um $H1$ verdadeiro). $\pi =$ Poder do Teste. Lembre-se que $\sum_{i=1}^{n} X_i \sim \text{Bin}(n, \theta)$. Tome $n = 10$.

$$\begin{aligned} \alpha &= P_{\theta=0.5}(\bar{X} < 0.3 \text{ ou } \bar{X} > 0.7) = P_{0.5}(\bar{X} < 0.3) + P_{0.5}(\bar{X} > 0.7) \\ &= P(\text{Bin}(n, 0.5) < 3) + P(\text{Bin}(10, 0.5) > 7) \\ &= P(\text{Bin}(n, 0.5) \leq 2) + P(\text{Bin}(10, 0.5) \geq 8) \\ &= 0.055 + 0.055 = 0.11 \end{aligned}$$

Note que para o Erro Tipo II, não existe uma única probabilidade para o erro sob H_1. Optaremos por tentar calcular seu máximo. $\beta_{\max}$

$$\begin{aligned} \beta &= P_\theta(0.3 \leq \bar{X} \leq 0.7)\theta \in \Theta \setminus \{0.5\} \\ \beta_{\max} &= \sup_{\theta \in \Theta \setminus \{0.5\}} \beta(\theta) \end{aligned}$$

Para $n = 10$,

$$\beta(\theta) = P_\theta\left(3 \leq \sum_{i=1}^{n=10} X_i \leq 7\right) = P\left(3 \leq \text{Bin}(10, \theta) \leq 7\right), \theta \in \Theta \setminus \{0.5\}$$

Podemos encontrar o valor que maximiza $\beta(\theta)$, $\theta = 0.5$ derivando.

14.6 Hipóteses como subconjuntos do espaço paramétrico

Seja (X_n) uma a.a. de $X \sim \text{Ber}(\theta)$, em que $\theta \in (0, 1)$. Considere as hipóteses:

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1 \end{cases}$$

em que $\Theta_0 \cup \Theta_1 = \Theta, \Theta_0, \Theta_1 \neq \emptyset, \Theta_0 \cap \Theta_1 = \emptyset$.

Exemplos de decisões elaboráveis:

$$\begin{cases} \Theta_0 = \{0.5\} \\ \Theta_1 = \left(0, \frac{1}{2}\right) \cup \left(\frac{1}{2}, 1\right) \end{cases} \Rightarrow \begin{cases} H_0 : \theta = 0.5 \\ H_1 : \theta \neq 0.5 \end{cases} \quad \text{Hipótese alternativa bilateral}$$

$$\begin{cases} \Theta_0 = \left(0, \frac{1}{2}\right] \\ \Theta_1 = \left(\frac{1}{2}, 1\right) \end{cases} \Rightarrow \begin{cases} H_0 : \theta \leq 0.5 \\ H_1 : \theta > 0.5 \end{cases} \quad \text{Hipótese alternativa unilateral}$$

$$\begin{cases} \Theta_0 = \left[\frac{1}{2}, 1\right) \\ \Theta_1 = \left(0, \frac{1}{2}\right) \end{cases} \Rightarrow \begin{cases} H_0 : \theta \geq 0.5 \\ H_1 : \theta < 0.5 \end{cases} \quad \text{Hipótese alternativa uniteral}$$

14.7 Função poder

No caso geral, calculamos a função poder definida por

$$\pi(\theta) = P_\theta(\{\text{Rejeitar } H_0\}), \theta \in \Theta$$

em que “Rejeitar H_0” é o procedimento de decisão para rejeitar H_0.

A partir da função poder conseguimos calcular as probabilidades máximas de cometer os erros tipo I e II.

Probabilidade Máxima do Erro Tipo I:

$$\alpha_{\max} = \sup_{\theta \in \Theta_0} (\pi(\theta))$$

Probabilidade Máxima do Erro Tipo II:

$$\beta_{\max} = \sup_{\theta \in \Theta_1} [1 - \pi(\theta)]$$

14.8 Um exemplo do cálculo de erros com hipótese unilateral

Seja (X_n) uma a.a. de $X \sim \text{Ber}(\theta)$, em que $\theta \in (0,1) = \Theta$. Considere as hipóteses:

$$\begin{cases} H_0 : \theta \geq 0.6 \\ H_1 : \theta < 0.6 \end{cases}$$

Precisamos de decisões que fazem sentido. Uma delas seria

1. Se $\bar{x} < 0.4$, rejeitamos H_0
2. Se $\bar{x} \geq 0.4$, não rejeitamos H_0

Vamos calcular as probabilidades máximas dos erros I e II.

Primeiro, encontramos a função poder

$$\pi(\theta) = P_\theta(\bar{X} < 0.4)$$

Como $\sum_{i=1}^n X_i \sim \text{Bin}(n, \theta)$, temos que

$$\pi(\theta) = P_\theta\left(\sum_{i=1}^n X_i < 0.4 \cdot n\right) = P(\text{Bin}(n,\theta) < 0.4 \cdot n)$$

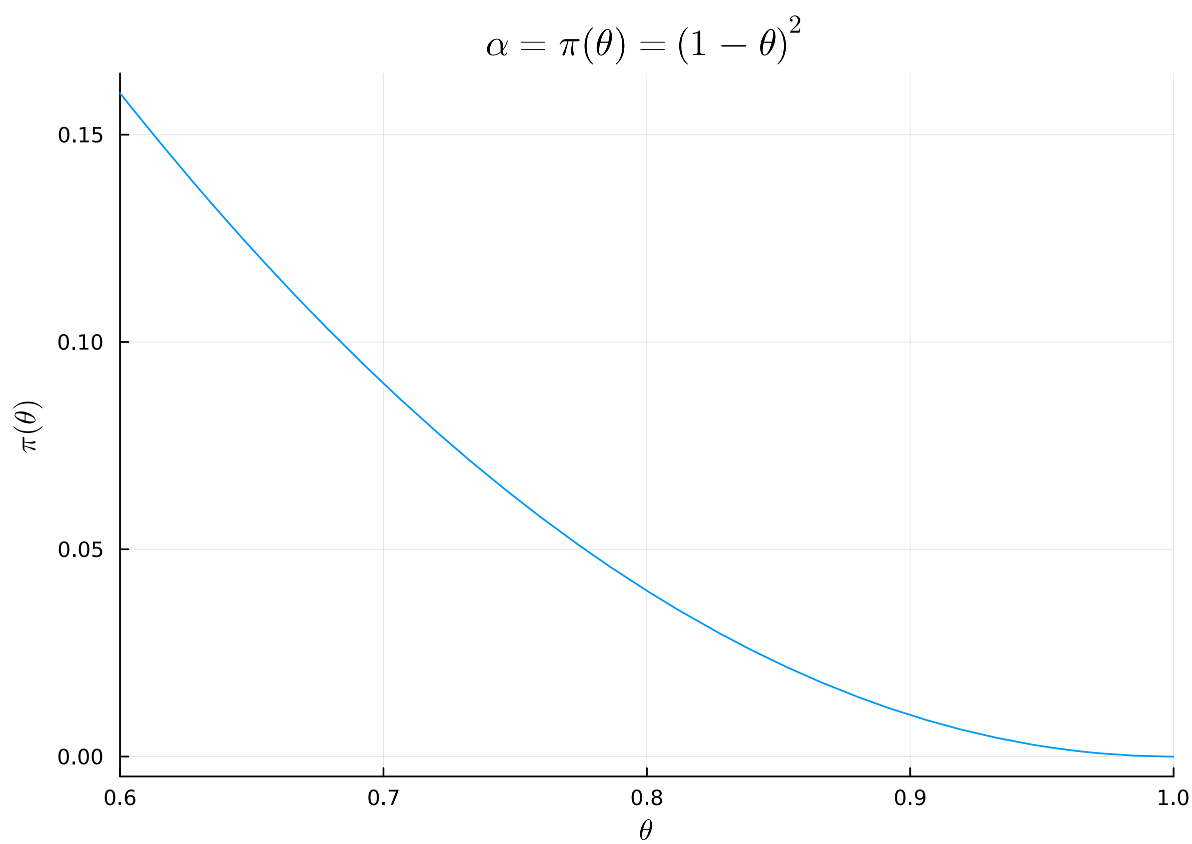
Relembrando:

$$\begin{aligned} \alpha_{\max} &= \sup_{\theta \in [0.6,1)} \pi(\theta) \\ &= \sup_{\theta \in [0.6,1)} P(\text{Bin}(n,\theta) < 0.4 \cdot n) \\ \beta_{\max} &= \sup_{\theta \in (0,0.6)} (1 - \pi(\theta)) \\ &= \sup_{\theta \in (0,0.6)} P(\text{Bin}(n,\theta) \geq 0.4 \cdot n) \end{aligned}$$

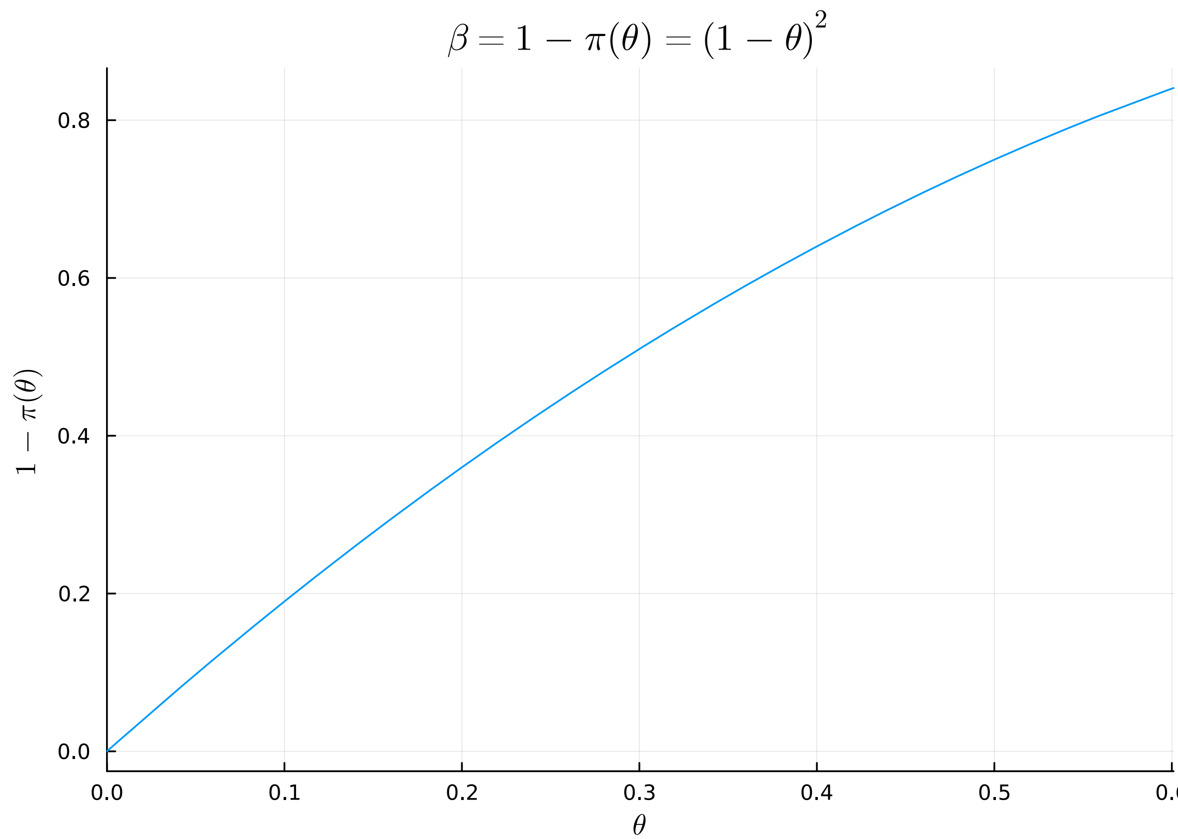
Para $n = 2$,

$$\begin{aligned}
\alpha_{\max} &= \sup_{\theta \in [0.6,1)} \pi(\theta) \\
&= \sup_{\theta \in [0.6,1)} P(\text{Bin}(2,\theta) < 0.4 \cdot 2) \\
&= \sup_{\theta \in [0.6,1)} P(\text{Bin}(2, \;) = 0) \\
\beta_{\max} &= \sup_{\theta \in (0,0.6)} (1 - \pi(\theta)) \\
&= \sup_{\theta \in (0,0.6)} P(\text{Bin}(2,\theta) \geq 0.8 \cdot n) \\
&= \sup_{\theta \in (0,0.6)} P(\text{Bin}(2,\theta) \geq 1) \\
&= \sup_{\theta \in (0,0.6)} [1 - P(\text{Bin}(2,\theta) = 0)]
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
\alpha_{\max} &= \sup_{\theta \in [0.6,1]} \left[\binom{2}{0} \theta^0 (1-\theta)^2 \right] \\
&= \sup_{\theta \in [0.6,1]} (1-\theta)^2 \\
\beta_{\max} &= \sup_{\theta \in (0,0.6)} (1 - (1-\theta)^2)
\end{aligned}$$

Podemos analisar os gráficos:



Como é uma função decrescente, seu supremo está no ponto 0.6



Como é uma função crescente, seu supremo está também no 0.6 Portanto,

$$
\begin{aligned}
\alpha_{\max} &= (1-0.6)^2 = 0.16 \\
\beta_{\max} &= (1-(1-0.6)^2) = 0.84
\end{aligned}
$$

14.9 Teste sob Normalidade

Seja (x_n) amostra aleatória de $X \sim N(\mu, \sigma^2)$ em que σ^2 é conhecido. Considere as hipóteses

$$
\begin{cases}
H_0 : \mu = \mu_0 \\
H_1 : \mu \neq \mu_0
\end{cases}
$$

com $\mu_0 \in \mathbb{R}$ e fixado.

Calcule as probabilidades (máximas) dos erros tipo I e II, para as seguintes decisões

1. Se $\bar{x} < \mu_0 - 1.96\sqrt{\frac{\sigma^2}{n}}$ ou $\bar{x} > \mu + 1.96\sqrt{\frac{\sigma^2}{n}}$, então *rejeitamos* H_0
2. Caso contrário, não rejeitamos H_0

Temos a função poder

$$\pi(\theta) = P_\theta(\text{Rejeitar} H_0) = P_\theta\left(\bar{X} < \mu_0 - 1.96\sqrt{\frac{\sigma^2}{n}}\right) + P_\theta\left(\bar{X} > \mu_0 + 1.96\sqrt{\frac{\sigma^2}{n}}\right)$$

em que $theta = \mu \in \mathbb{R}$. Portanto,

$$\alpha_{\max} = \sup_{\theta \in \Theta_0} \pi(\theta)$$

Como $H_0 = \mu = \mu_0 \Leftrightarrow H_0 : \theta \in \Theta$, em que $\Theta_0 = \{\mu_0\}$, logo, $\sup_{\theta \in \Theta_0} = \mu_0$ Portanto, temos que

$$\alpha_{\max} = \pi(\mu_0) = P_{\mu_0}\left(\bar{X} < \mu_0 - 1.96\sqrt{\frac{\sigma^2}{n}}\right) + P_{\mu_0}\left(\bar{X} > \mu_0 + 1.96\sqrt{\frac{\sigma^2}{n}}\right)$$

Sabemos que, pelo enunciado $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \forall \mu \in \mathbb{R}$ sob H_0, ou seja, quando $\mu = \mu_0$ temos que $\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$. Note que $P_{\mu_0}\left(\bar{X} < \mu_0 - 1.96\sqrt{\frac{\sigma^2}{n}}\right) = P_{\mu_0}\left(\frac{\bar{X}-\mu_0}{\sqrt{\frac{\sigma^2}{n}}}\right) = 2.5\%$ Pela simetria da distribuição normal, $P_{\mu_0}\left(\bar{X} > \mu_0 + 1.96\sqrt{\frac{\sigma^2}{n}}\right) = 2.5\%$ Portanto a probabilidade máxima do erro tipo 1 é

$$\alpha_{\max} = 2.5\% + 2.5\% = 5.0\%$$

Como $H_1 : \mu \neq \mu_0 \Leftrightarrow H_1 : \theta \in \Theta_1$, em que $\Theta_1 = \mathbb{R} \setminus \{\mu_0\}$, temos que

$$\beta_{\max} = \sup_{\theta \in \Theta_1} [1 - \pi(\theta)]$$
$$\pi(\theta) = P_\theta\left(\bar{X} < \mu_0 - 1.96\sqrt{\frac{\sigma^2}{n}}\right) + P_\theta\left(\bar{X} > \mu_0 + 1.96\sqrt{\frac{\sigma^2}{n}}\right)$$

Sabemos que $\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$ para todo $\mu \in \mathbb{R}$. Assim,

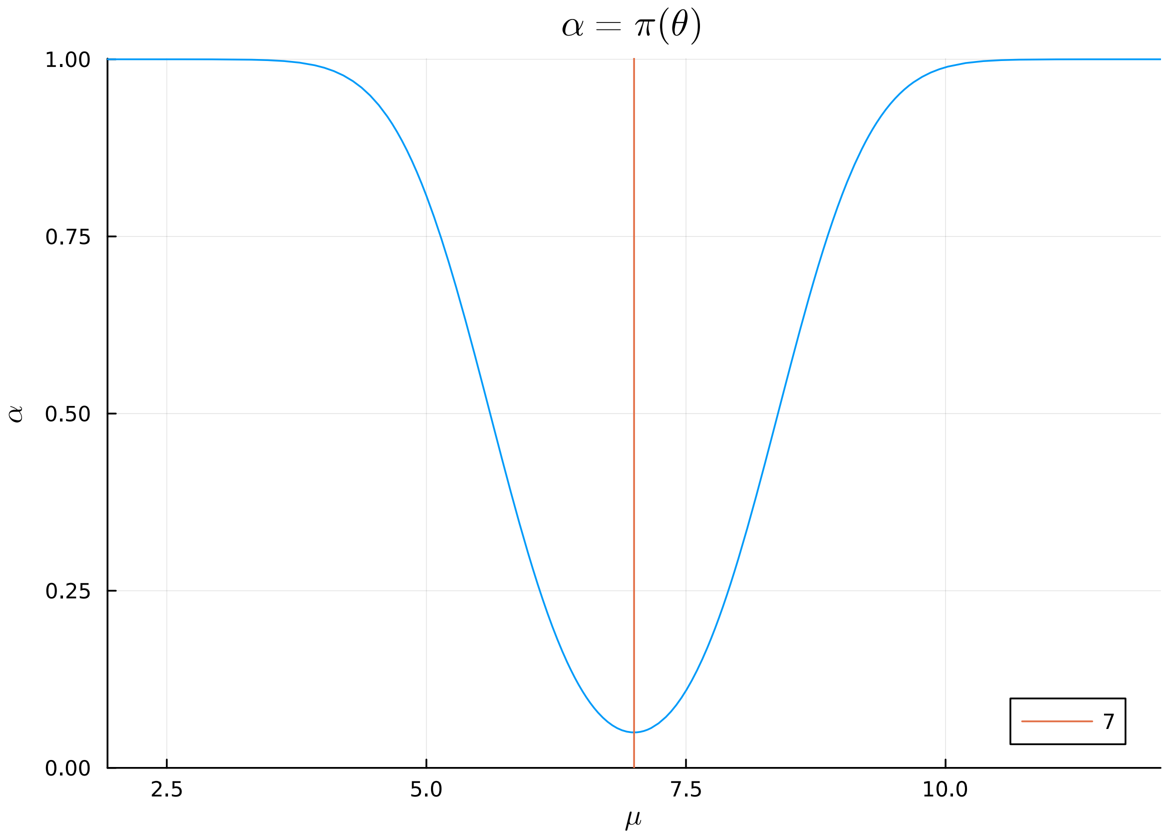
$$\begin{aligned}
\pi(\theta) &= P_\theta\left(\bar{X} < \mu_0 - 1.96\sqrt{\frac{\sigma^2}{n}}\right) + P_\theta\left(\bar{X} > \mu_0 + 1.96\sqrt{\frac{\sigma^2}{n}}\right) \\
&= P_\theta\left(\frac{\bar{X} - \theta}{\sqrt{\frac{\sigma^2}{n}}} < \frac{\mu_0 - \theta - 1.96\sqrt{\frac{\sigma^2}{n}}}{\sqrt{\frac{\sigma^2}{n}}}\right) + P_\theta\left(\frac{\bar{X} - \theta}{\sqrt{\frac{\sigma^2}{n}}} > \frac{\mu_0 - \theta + 1.96\sqrt{\frac{\sigma^2}{n}}}{\sqrt{\frac{\sigma^2}{n}}}\right)
\end{aligned}$$

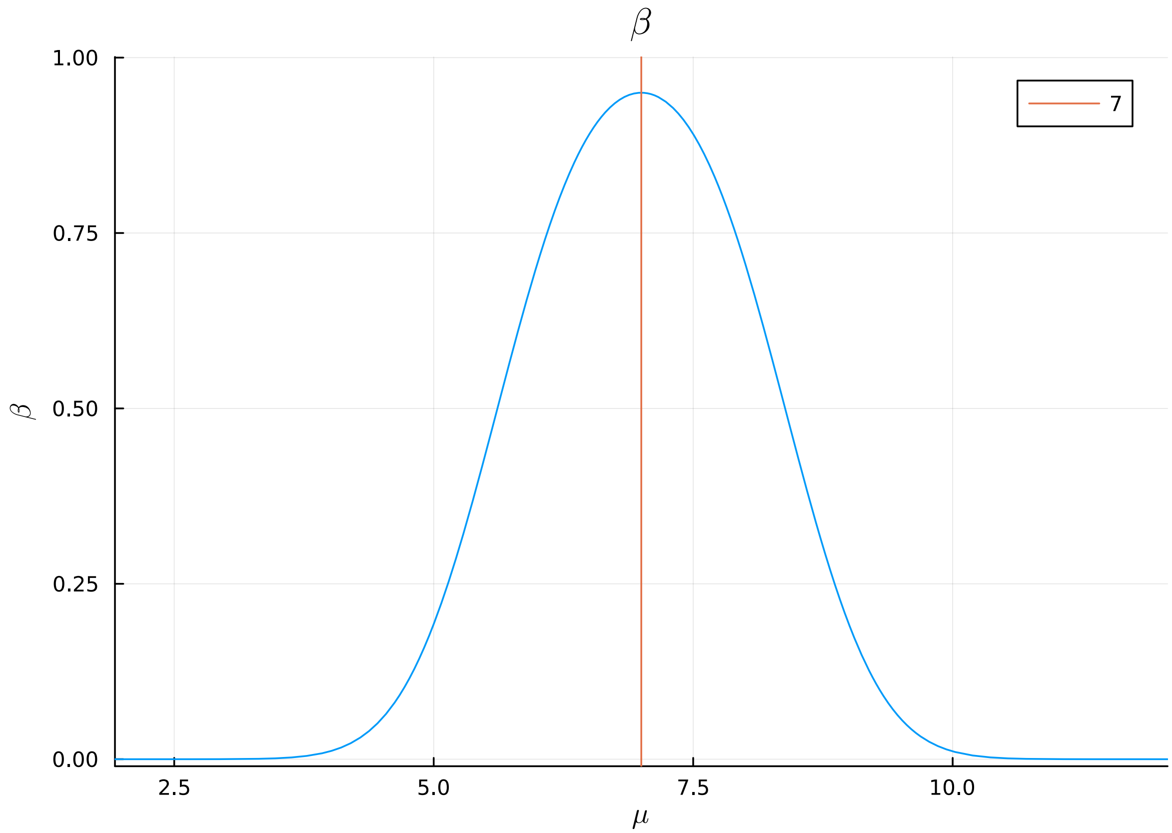
Dessa forma,

$$\beta_{max} = \sup_{\theta \in \Theta_1} [1 - \pi(\theta)] = 1 - \inf_{\theta \in \Theta_1} \pi(\theta)$$

Ou seja, o supremo dessa expressão é dado por 1 - o ínfimo da função poder, o que significa que queremos encontrar o valor de θ para o qual $P_\theta\left(\frac{\bar{X}-\theta}{\sqrt{\frac{\sigma^2}{n}}} < \frac{\mu_0-\theta-1.96\sqrt{\frac{\sigma^2}{n}}}{\sqrt{\frac{\sigma^2}{n}}}\right) + P_\theta\left(\frac{\bar{X}-\theta}{\sqrt{\frac{\sigma^2}{n}}} > \frac{\mu_0-\theta+1.96\sqrt{\frac{\sigma^2}{n}}}{\sqrt{\frac{\sigma^2}{n}}}\right)$ é o menor possível.

Vamos usar $\mu_0 = 7$ e $\sigma^2 = 5$ para visualizarmos o comportamento de α e β





14.9.1 Um outro exemplo

Seja (X_n) amostra aleatória de $X \sim \mathrm{N}(\mu, \sigma^2)$ com $\sigma^2 = 5, n = 10$ Com as hipóteses

$$\begin{cases} H_0 : \mu = 10 \\ H_1 : \mu \neq 10 \end{cases}$$

Com as decisões 1. Rejeitamos H_0 se $\bar{x} > 10 + 1.96\sqrt{\frac{5}{10}}$ ou $\bar{x} < 10 - 1.96\sqrt{\frac{5}{10}}$ Foram observados os seguintes valores

7.1	8.9	12	13	11.7
6.1	2.5	3.1	5.2	7

Temos então que $\bar{x} = 7.66$ que, como é abaixo de 8.6, rejeitamos a hipótese nula de que $\mu = 10$

15 Procedimento Geral para Testar Hipóteses (Método de Neyman-Pearson)

1. Definimos o Modelo Estatístico:
    1. "Seja (X_n) amostra aleatória de $x \sim f_\theta, \theta \in \Theta$"
2. Definir as hipóteses de interesse:
    1. $H_0 : \theta \in \Theta_0 \times H_1 : \theta \in \Theta_1$ em que $\Theta_0 \cap \Theta_1 = \emptyset, \Theta_0 \neq \emptyset, \Theta_1 \neq \emptyset, \Theta_0 \cup \Theta_1 = \Theta$
3. A partir da amostra observada, criamos uma regra de decisão para verificar a plausibilidade de H_0.
4. Definimos os pontos de corte da regra de decisão de forma que a probabilidade máxima do Erro Tipo I não ultrapasse um limite prefixado $\alpha \in [0, 1]$ (normalmente 5% ou 1%). Qualquer valor $\geq \alpha$ é dito ser um *nível de significância*.
5. Concluímos o Teste de Hipótese.
    1. Se H_0 for rejeitado, dizemos que "Há evidências para rejeitar H_0 a $\alpha \cdot$ [Valor]% de significância estatística".
    2. Se H_0 não for rejeitado, dizemos que "Não há evidências para rejeitar H_0 a $\alpha \cdot$ [Valor]% de significância estatística"

> **i Observação**
>
> Não rejeitar H_0 *não* indica evidência a favor de H_0, isto é, não sugere que H_0 seja verdadeiro, apenas que aquela amostra não apresentou evidências contrárias.

> **i Observação**
>
> Quanto menor o valor de α, mais forte será a significância estatística.

15.1 Exemplos

15.1.1 Eexemplo um

Seja (X_n) a.a de $X \sim \mathrm{N}(\mu, \sigma^2)$ em que σ^2 é conhecido. Considere (μ_0 fixado)

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

1. Construa uma decisão para rejeitar H_0 que produza no máximo $\alpha = 5\%$ (que tenha nível de significância de 5%)

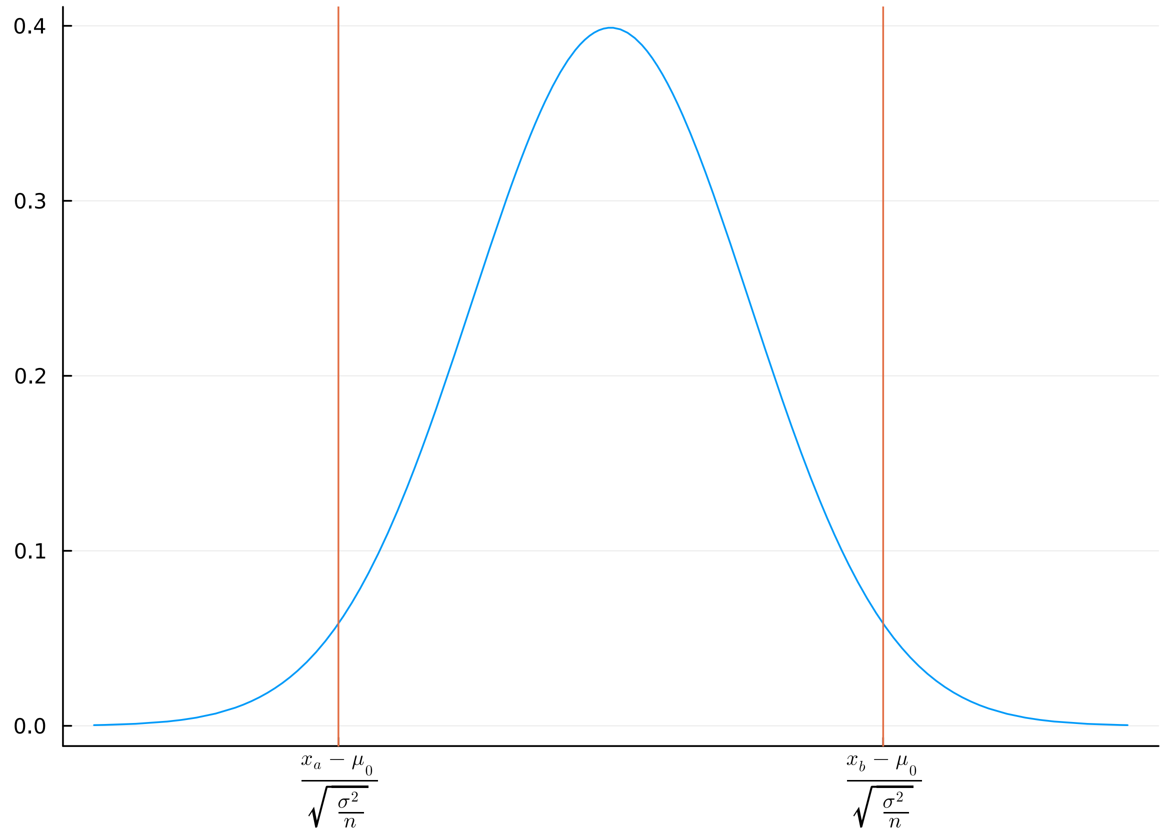
   Como a hipótese alternativa é bilateral, $H_1 : \mu \neq \mu_0$ e $\bar{x}$ é a EMV para o parâmetro μ - a esperança da distribuição Normal - definimos a regra: Se $\bar{x} < x_a$ ou $\bar{x} > x_b$, rejeitamos H_0. Caso contrário, não rejeitamos.

   $$\begin{aligned} \alpha_{\max} &= \sup_{\theta \in \Theta_0} P_\theta(\text{Rejeitar } H_0), \Theta_0 = \{\mu_0\} \\ &= P_{\mu_0}(\bar{X} < x_a) + P_{\mu_0}(\bar{X} > x_b) \leq 5\% \end{aligned}$$

   Note que $\bar{X} \sim \mathrm{N}\left(\mu_0, \frac{\sigma^2}{n}\right)$, sob H_0

   Logo,

   $$\alpha_{\max} = P\left(\mathrm{N}(0,1) < \frac{x_a - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}\right) + P\left(\mathrm{N}(0,1) > \frac{x_a - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}\right)$$



Tomando $\frac{x_a-\mu_0}{\sqrt{\frac{\sigma^2}{n}}} = -1.96$ e $\frac{x_b-\mu_0}{\sqrt{\frac{\sigma^2}{n}}} = 1.96$ (tabela normal padrão simétrica), temos que $\alpha_{\max} = 5\%$. Assim, resolvendo as equações,

$$\begin{cases} x_a = \mu - 1.96\sqrt{\frac{\sigma^2}{n}} \\ x_b = \mu + 1.96\sqrt{\frac{\sigma^2}{n}} \end{cases}$$

2. Considere $n = 100$, $\mu_0 = 1$, $\sigma^2 = 0.1$ e $\bar{x} = 0.99$. Conclua o teste considerando o mesmo nível de significância $\alpha = 5\%$.

   O pontos pontos de corte são $x_a = 0.93$ e $x_b = 1.069$. Como $0.93 \leq 0.99 \leq 1.069$, concluímos que não há evidências para rejeitarmos H_0 a 5% de significância.

3. Refaça considerando 15% de significância estatística.

   Usando os mesmos argumentos do item 1, podemos encontrar novos valores para x_a, x_b através da tabela da Normal-Padrão:

$$\begin{cases} x_a = \mu - 1.44\sqrt{\frac{\sigma^2}{n}} \\ x_b = \mu + 1.44\sqrt{\frac{\sigma^2}{n}} \end{cases}$$

Substituindo esses valores para os fornecidos em 2, temos que $0.95 \leq 0.99 \leq 1.045$. Portanto, continuaríamos a dizer que não há evidências para rejeitarmos H_0 a 15% de significância.

15.1.2 2

Seja (X_n) a.a de $X \sim \mathrm{N}(\mu, \sigma^2)$ em que σ^2 é conhecido.

Considere (μ_0 fixado)

$$\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

1. Construa uma decisão para rejeitar H_0 que produza no máximo $\alpha = 5\%$ (que tenha nível de significância de 5%)

   Pelos parâmetros e hipóteses envolvidos, $(\mu, \text{unilateral})$, podemos considerar a seguinte decisão:

   Se $\bar{x} < x_c$, rejeitamos H_0. Caso contrário, não rejeitamos.

$$\begin{aligned} \alpha_{\max} &= \sup_{\mu \geq \mu_0} P_\theta(\text{Rejeitar } H_0) \\ &= \sup_{\mu \geq \mu_0} P_\mu(\bar{X} < x_c) \\ \Rightarrow \alpha_{\max} &= \sup_{\mu \geq \mu_0} P_\mu\left(\mathrm{N}(0,1) < \frac{x_c - \mu}{\sqrt{\frac{\sigma^2}{n}}}\right) \end{aligned}$$

   Como essa função (acumulada) é decrescente em μ, temos que

$$\begin{aligned} \alpha_{\max} &= P_{\mu_0}(\text{Rejeitar } H_0) \\ &= P_{\mu_0}(\bar{X} < x_c) \\ \Rightarrow \alpha_{\max} &= P_{\mu_0}\left(\mathrm{N}(0,1) < \frac{x_c - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}\right) \leq 5\% \end{aligned}$$

   Logo, para encontrarmos x_c tal que $\frac{x_c - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} = -1.64$ (da tabela da normal padrão) $\Rightarrow$ $x_c = \mu_0 - 1.64\sqrt{\frac{\sigma^2}{n}}$

2. Considere $n = 100, \mu_0 = 1, \sigma^2 = 0.1, \bar{x} = 0.99$. Conclua o teste anterior a $\alpha = 5\%$ de significância.

   O ponto de corte é $x_c = 0.9836$. Como $0.99 \geq 0.9836$, concluímos que não há evidências para rejeitar a hipótese nula a 5% de significância.

15.1.3 3

Seja (X_n) amostra aleatória de $X \sim \mathrm{N}(\mu, \sigma^2)$ em que $\theta = (\mu, \sigma^2) = \mathrm{R} \times \mathrm{R}^+$, ou seja, ambos parâmetros são desconhecidos.

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases} \Rightarrow \begin{array}{l} \text{Decisão com significância } \alpha \\ \text{Rejeita } H_0 \text{ se} \\ \begin{cases} s^2 < c_{1c} \\ s^2 > c_{2c} \end{cases} \\ \text{Em que } c_{1c}, c_{2c} \text{ são tais que} \\ \sup\limits_{\theta \in \Theta_0} P_\theta(\text{Erro Tipo I}) = \alpha_{\max} = \alpha \\ \text{E } s^2 = \dfrac{1}{n-1} \displaystyle\sum_{i=1}^{n} (x_i - \bar{x})^2 \end{array}$$

Sabemos que $\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}, \forall \mu \in \mathbb{R}, \sigma^2 > 0$. Em particular, sob H_0

$$\begin{aligned} &\frac{(n-1)s^2(\underset{\sim}{X_n})}{\sigma_0^2} \sim \chi^2_{n-1} \\ \Rightarrow &\alpha_{\max} = \sup_{\theta \in \Theta_0} \left\{ P_\theta(s^2(\underset{\sim}{X_n}) < c_{1c}) + P_\theta(s^2(\underset{\sim}{X_n}) > c_{2c}) \right\} \end{aligned}$$

Além disso, note que $\Theta_0 = \{(\mu, \sigma^2) \in \Theta : \sigma^2 = \sigma_0^2\}$. Portanto, temos que

$$\begin{aligned} \alpha_{\max} &= \sup_{\theta \in \Theta_0} \underbrace{\left\{ P\left(\chi^2_{n-1} < \frac{c_{1c}(n-1)}{\sigma_0^2}\right) + P\left(\chi^2_{n-1} > \frac{c_{2c}(n-1)}{\sigma_0^2}\right) \right\}}_{\text{Não depende de } \theta} \\ \Rightarrow \alpha_{\max} &= P\left(\chi^2_{n-1} < \frac{c_{1c}(n-1)}{\sigma_0^2}\right) + P\left(\chi^2_{n-1} > \frac{c_{2c}(n-1)}{\sigma_0^2}\right) \end{aligned}$$

Fixando $\alpha_{\max} = \alpha$ (significância), encontramos pela tabela os valores de $q^{(1)}_{\frac{\alpha}{2}, n-1}, q^{(2)}_{\frac{\alpha}{2}, n-1}$ tais que dividam a distribuição χ^2_{n-1} criando duas seções de $\frac{\alpha}{2}$ de área. Portanto, Decisão com significância α

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases} \Rightarrow \begin{array}{l} \text{Rejeita } H_0 \text{ se} \\ \begin{cases} s^2 < q^{(1)}_{\frac{\alpha}{2},n-1} \cdot \frac{\sigma_0^2}{(n-1)} \\ s^2 > q^{(2)}_{\frac{\alpha}{2},n-1} \cdot \frac{\sigma_0^2}{(n-1)} \end{cases} \end{array}$$

15.1.4 Exemplo

Seja (X_n) amostra aleatória de $X \sim N(\mu, \sigma^2)$ em que X é o peso do pacote de café. Colheu-se uma amostra de $n = 16$ pacotes e observou-se uma variância de $s^2 = 169g^2$. O processo de fabricação diz que a média dos pacotes é $500g$ e desvio-padrão 10 gramas ($\sigma_0^2 = 100g^2$).

Queremos verificar se há alguma evidência de que o processo não esteja sendo cumprido com $\alpha = 5\%$ de significância

Decisão com significância 5%

$$\begin{cases} H_0 : \sigma^2 = 100 \\ H_1 : \sigma^2 \neq 100 \end{cases} \Rightarrow \begin{array}{l} \text{Rejeita } H_0 \text{ se} \\ \begin{cases} s^2 < q^{(1)}_{2.5\%,15} \cdot \frac{100}{15} \\ s^2 > q^{(2)}_{2.5\%,15} \cdot \frac{100}{15} \end{cases} \end{array}$$

Da tabela Qui-quadrado, temos $q^{(1)}_{2.5\%,15} = 6.26$ e $q^{(2)}_{2.5\%,15} = 27.49$.

Como $41.73 < 100 < 183.26$, concluímos que não há evidências para rejeitar a hipótese nula a 5% de significância

16 Fórmulas

16.1 Sob normalidade, variância conhecida

Seja (X_n) amostra aleatória de $X \sim \mathrm{N}(\mu, \sigma^2)$ em que σ^2 é conhecido.

Decisão com significância α

1. $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \Rightarrow$ Rejeita H_0 se $\begin{cases} \bar{x} < \mu_0 - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}} \\ \bar{x} > \mu_0 + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}} \end{cases}$

   Em que $z_{\frac{\alpha}{2}}$ é tal que $P\left(\mathrm{N}(0,1) < z_{\frac{\alpha}{2}}\right) = \dfrac{\alpha}{2}$

2. $\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \Rightarrow$ Rejeita H_0 se $\left\{ \bar{x} < \mu_0 - z_{\alpha} \sqrt{\frac{\sigma^2}{n}} \right.$

   Em que z_{α} é tal que $P\left(\mathrm{N}(0,1) \leq z_{\alpha}\right) = \alpha$

3. $\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \Rightarrow$ Rejeita H_0 se $\left\{ \bar{x} > \mu_0 + z_{\alpha} \sqrt{\frac{\sigma^2}{n}} \right.$

   Em que z_{α} é tal que $P\left(\mathrm{N}(0,1) \geq z_{\alpha}\right) = \alpha$

16.2 Sob normalidade, variância desconhecida

Seja (X_n) amostra aleatória de $X \sim \mathrm{N}(\mu, \sigma^2)$ em que $\theta = (\mu, \sigma^2) = \mathrm{R} \times \mathrm{R}^+$, ou seja, ambos parâmetros são desconhecidos.

Decisão com significância α

1. $$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \\ \Rightarrow \Theta_0 = \{(\mu, \sigma^2) \in \Theta : \mu = \mu_0\} \\ \Rightarrow \Theta_1 = \{(\mu, \sigma^2) \in \Theta : \mu \neq \mu_0\} \end{cases} \Rightarrow$$

Rejeita H_0 se

$$\begin{cases} \frac{\bar{x}-\mu_0}{\sqrt{\frac{s^2}{n}}} < -t_{\frac{\alpha}{2},n-1} \\ \frac{\bar{x}-\mu_0}{\sqrt{\frac{s^2}{n}}} > t_{\frac{\alpha}{2},n-1} \end{cases}$$

Em que $t_{\frac{\alpha}{2},n-1}$ é tal que

$$P\left(t_{n-1} < -t_{\frac{\alpha}{2},n-1}\right) = \frac{\alpha}{2}$$

2. $$\begin{cases} H_0 : \mu \geq \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \Rightarrow$$

Rejeita H_0 se

$$\begin{cases} \frac{\bar{x}-\mu_0}{\sqrt{\frac{s^2}{n}}} < -t_{\alpha,n-1} \end{cases}$$

Em que $t_{\alpha,n-1}$ é tal que

$$P\left(t_{n-1} \leq -t_{\alpha,n-1}\right) = \alpha$$

3. $$\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \Rightarrow$$

Rejeita H_0 se

$$\begin{cases} \frac{\bar{x}-\mu_0}{\sqrt{\frac{s^2}{n}}} > -t_{\alpha,n-1} \end{cases}$$

Em que $t_{\alpha,n-1}$ é tal que

$$P\left(t_{n-1} > t_{\alpha,n-1}\right) = \alpha$$

Em que $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ é a variância amostral (não enviesada)

16.3 Sob normalidade, para a variância.

Seja (X_n) amostra aleatória de $X \sim \mathrm{N}(\mu, \sigma^2)$ em que $\theta = (\mu, \sigma^2) = \mathrm{R} \times \mathrm{R}^+$, ou seja, ambos parâmetros são desconhecidos.

Decisão com significância α

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases} \Rightarrow$$

Rejeita H_0 se

$$\begin{cases} s^2 < q^{(1)}_{\frac{\alpha}{2},n-1} \cdot \frac{\sigma_0^2}{(n-1)} \\ s^2 > q^{(2)}_{\frac{\alpha}{2},n-1} \cdot \frac{\sigma_0^2}{(n-1)} \end{cases}$$

Em que $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ é a variância amostral (não enviesada) e $q^{(1)}_{\frac{\alpha}{2},n-1}, q^{(2)}_{\frac{\alpha}{2},n-1}$ tais que dividam a distribuição χ^2_{n-1} criando duas seções de $\frac{\alpha}{2}$ de área.

17 Teste de Hipótese para duas populações

Sejam X, Y duas variáveis de interesse representando duas sub-populações. Estamos interessados em verificar se a média populacional de X é menor, maior ou igual à de Y. Sendo assim, precisamos considerar os casos em que X é independente de Y e o caso em que não são independentes (pareados).

17.1 Dados independentes e dependentes

Um pesquisador propôs um novo método de investimento para aumentar o rendimento mensal. Selecionou 20 investidores aleatoriamente de um universo de investidores cadastrados. Em um primeiro momento, o pesquisador deixou os investidores investirem do jeito que sabem e ao final verificou a renda obtida. X é o rendimento dos investidores sem ter o conhecimento do método.

Então, o pesquisador ensinou seu método aos investidores, onde Y passou a ser o rendimento dos investidores após a aplicação do método ensinado.

Claramente, X, Y são dependentes.

O mesmo pesquisador testará o mesmo método de forma diferente. Para testar o seu método, o pesquisador selecionou 20 indivíduos com características similares do universo de investidores, dos quais

1. 10 foram designados aleatoriamente a não receber o método X : rendimento de um indivíduo que não recebeu o método
2. 10 foram designados aleatoriamente a receberem o método Y : rendimento de um indivíduo que recebeu o método

Nesse caso, as variáveis X, Y são independentes.

Em ambas abordagens, temos as mesmas hipóteses de interesse:

$$1. \begin{cases} H_0 : \mu_X \geq \mu_y \\ H_1 : \mu_X < \mu_Y \end{cases} \qquad 2. \begin{cases} H_0 : \mu_X \leq \mu_Y \\ H_1 : \mu_X > \mu_Y \end{cases} \qquad 3. \begin{cases} H_0 : \mu_X = \mu_Y \\ H_1 : \mu_X \neq \mu_Y \end{cases}$$

Podemos definir $\mu_D = \mu_X - \mu_Y$ e reescrever as hipóteses

$$1. \begin{cases} H_0 : \mu_D \geq 0 \\ H_1 : \mu_D < 0 \end{cases} \qquad 2. \begin{cases} H_0 : \mu_D \leq 0 \\ H_1 : \mu_D > 0 \end{cases} \qquad 3. \begin{cases} H_0 : \mu_D = 0 \\ H_1 : \mu_D \neq 0 \end{cases}$$

Para os próximos exemplos, assumiremos normalidade para X, Y

17.1.1 Caso pareado (variáveis dependentes)

No caso em que X, Y são dependentes, as amostras são (X_n) amostras aleatórias de X e (Y_n) a.a de Y tais que X_i, Y_i são dependentes.

Para este caso, fazemos $D_i = X_i - Y_i, i = 1, 2, \dots, n$. Temos que (D_n) é uma amostra aleatória de $D = X - Y \sim N(\mu_D, \sigma_D^2) = N(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y)$

Observe ainda que

$$\bar{D}_{\text{Par}} = \sum_{i=1}^{n} \frac{D_i}{n} \sim N\left(\mu_D, \frac{\sigma_D^2}{n}\right)$$

Note ainda que as variância e covariância de X, Y estão embutidas em σ_D^2. Podemos usar

$$s_D^2(\underset{\sim}{D}) = \frac{1}{n-1} \sum_{i=1}^{n} (D_i - \bar{D}_{\text{Par}})^2$$

Para estimar σ_D^2 Podemos construir as decisões como já vimos anteriormente em testes sob normalidade com variância desconhecida.

17.1.2 Exemplo

Foram coletados os rendimentos (em mil reais) antes e após a aplicação o método para 12 investidores. Queremos verificar se o método aumentou o rendimento médio. Chamaremos de X o rendimento anterior ao treinamento e Y o rendimento após. Isto é, queremos verificar se $\mu_D = \mu_X - \mu_Y \leq 0$. Portanto, nossa hipótese nula é de que o treinamento não tem efeito positivo no rendimento:

$$\begin{cases} H_0 : \mu_D \geq 0 \\ H_1 : \mu_D < 0 \end{cases}$$

Indíce	Antes(X)	Depois(Y)	Dif(D)	Dif^2(D^2)
1	2.4	4.3	-1.9	3.61
2	2.8	3.4	-0.6	0.36
3	4.6	3.2	1.4	1.96
4	3.1	3.3	-0.2	0.04
5	3.1	3.3	-0.2	$0,04$
6	4.7	5.8	-1.1	1.21
7	3.5	3.8	-0.3	0.09
8	1.7	3.5	-1.8	3.24
9	2.3	3.2	-0.9	0.81
10	2.6	3.9	-1.3	1.69
11	4.2	3.6	0.6	0.36
12	3.4	4.3	-0.9	0.81
Média	3.2	3.8	-0.6	
s^2	0.87	0.55	0.9	

Temos que $s^2(\underset{\sim}{D}) = 0.9$. (Podemos calcular diretamente ou usando a coluna D^2 e substituindo no somatório $s_D^2 = \left(\frac{\sum_{i=1}^n d_i^2}{n} - \bar{d}^2\right)\frac{n}{n-1}$) Rejeitamos H_0 se $\frac{\bar{d}_{\text{Par}}}{\sqrt{\frac{s_D^2}{n}}} < -t_{\alpha,n-1}$ onde $t_{\alpha,n-1}$ é tal que $P(t_{n-1} < -t_{\alpha,n-1}) = \alpha$, onde t_{n-1} é a distribuição T de Student com $n-1$ graus de liberdade.

Temos que $\frac{\bar{d}_{\text{Par}}}{\sqrt{\frac{s_D^2}{n}}} = -2.19$ e, a $\alpha = 0.05$, $-t_{0.05,11} = -1.796$. Como $-2.19 < -1.796$, podemos dizer que há evidências de que o método aumenta o rendimento médio dos investidores a 5% de significância. Por outro lado, com $\alpha = 0.01, -t_{0.01,11} = -2.718$ e, por $-2.19 \geq -2.718$, dizemos que não há evidências para rejeitar a hipótese de que o método não aumenta o rendimento (rejeitar a hipótese nula) a 1% de significância estatística.

17.2 Caso de independência

Sejam X, Y variáveis aleatórias *independentes* tais que

$$\begin{cases} X \sim N(\mu_X, \sigma_X^2) \\ Y \sim (\mu_Y, \sigma_Y^2) \end{cases}$$

Considere (X_n) amostra aleatória de X e (Y_m). Estamos interessados em testar as hipóteses:

$$1. \begin{cases} H_0 : \mu_X \geq \mu_Y \\ H_1 : \mu_X < \mu_Y \end{cases} \qquad 2. \begin{cases} H_0 : \mu_X \leq \mu_Y \\ H_1 : \mu_X > \mu_Y \end{cases} \qquad 3. \begin{cases} H_0 : \mu_X = \mu_Y \\ H_1 : \mu_X \neq \mu_Y \end{cases}$$

Podemos definir $\mu_D = \mu_X - \mu_Y$ e obter a equivalência dessas hipóteses

$$1. \begin{cases} H_0 : \mu_D \geq 0 \\ H_1 : \mu_D < 0 \end{cases} \qquad 2. \begin{cases} H_0 : \mu_D \leq 0 \\ H_1 : \mu_D > 0 \end{cases} \qquad 3. \begin{cases} H_0 : \mu_D = 0 \\ H_1 : \mu_D \neq 0 \end{cases}$$

Considere

$$\bar{D}_{\text{NPar}} = \bar{X} - \bar{Y}$$

Como $\bar{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right), \bar{Y} \sim N\left(\mu_Y, \frac{\sigma_Y^2}{m}\right)$, temos que $\bar{D}_{\text{NPar}} \sim N\left(\mu_D, \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)$

17.2.1 Ambas variâncias conhecidas

Podemos substituir o valor numérico das variâncias na distribuição de $\bar{D}_{\text{NPar}}$, obtendo uma distribuição normal com pontos de corte para as decisões:

1. Rejeita H_0 se $\bar{d}_{\text{NPar}} < -z_\alpha \underbrace{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}}_{\text{Var}(\bar{D}_{\text{NPar}})}, \sigma_2 = \sigma_X^2 = \sigma_Y^2$
2. Rejeita H_0 se $\bar{d}_{\text{NPar}} > z_\alpha \sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}$
3. Rejeita H_0 se $\bar{d}_{\text{NPar}} < -z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}$ ou $\bar{d}_{NPar} > z_\alpha \sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}$

17.2.2 Variâncias desconhecidas e iguais

Temos que $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ é conhecido.

17.2.2.1 Estimando via t-Student

Através da distribuição t-Student com $n+m-2$ graus de liberdade, podemos estimar os pontos de corte. Temos nossas decisões:

1. Rejeita H_0 se $\bar{d}_{\text{NPar}} < -t_{n+m-2,\alpha} \sqrt{\frac{s_p^2}{n} + \frac{s_p^2}{m}}$
2. Rejeita H_0 se $\bar{d}_{\text{NPar}} > t_{n+m-2,\alpha} \sqrt{\frac{s_p^2}{n} + \frac{s_p^2}{m}}$
3. Rejeita H_0 se $\bar{d}_{\text{NPar}} < -t_{n+m-2,\frac{\alpha}{2}} \sqrt{\frac{s_p^2}{n} + \frac{s_p^2}{m}}$ ou $\bar{d}_{NPar} > t_{n+m-2,\frac{\alpha}{2}} \sqrt{\frac{s_p^2}{n} + \frac{s_p^2}{m}}$

Onde $s_p^2 = \frac{(n-1)s_X^2+(m-1)s_Y^2}{n+m-2}$, com s_X^2, s_Y^2 sendo os estimadores não enviesados para as variâncias de X e Y, respectivamente. (Ponderamos os estimadores com base no tamanho de sua amostra, assim favorecendo os estimadores mais precisos)

17.2.3 Variâncias desconhecidas e diferentes (caso geral)

Temos que σ_X^2, σ_Y^2 são desconhecidos e diferentes.

17.2.3.1 Estimando via t-Student

usaremos a distribuição t-Student com n' graus de liberdade. Temos nossas decisões:

1. Rejeita H_0 se $\bar{d}_{\text{NPar}} < -t_{n',\alpha}\sqrt{\frac{s_p^2}{n} + \frac{s_p^2}{m}}$
2. Rejeita H_0 se $\bar{d}_{\text{NPar}} > t_{n',\alpha}\sqrt{\frac{s_p^2}{n} + \frac{s_p^2}{m}}$
3. Rejeita H_0 se $\bar{d}_{\text{NPar}} < -t_{n',\frac{\alpha}{2}}\sqrt{\frac{s_p^2}{n} + \frac{s_p^2}{m}}$ ou $\bar{d}_{NPar} > t_{n',\frac{\alpha}{2}}\sqrt{\frac{s_p^2}{n} + \frac{s_p^2}{m}}$

Onde $s_p^2 = \frac{(n-1)s_X^2+(m-1)s_Y^2}{n+m-2}$.

17.2.3.1.1 Encontrando n'

Na fórmula acima, temos os graus de liberdade da t-Student dado por

$$n' \approx \frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2}{\frac{\left(\frac{s_X^2}{n}\right)^2}{n-1} + \frac{\left(\frac{s_Y^2}{m}\right)^2}{m-1}}$$

Esse valor, caso não inteiro, deverá ser arredondado.

17.2.3.2 Exemplo (Importante)

Queremos testar a resistência de dois tipos de viga de aço, A e B. Tomando-se $n = 15$ vigas do tipo A e $m = 20$ vigas do tipo B. de um teste f, conseguimos com 10% de significância que as variâncias não são iguais. Obtemos os valores da tabela:

Tipo	Média	Variância(s^2)
A	70.5	81.6
B	84.3	210.8
$\bar{d}_{\text{NPar}}$	-13.8	$--$

Teste a hipótese

$$\begin{cases} H_0 : \mu_X = \mu_Y \\ H_1 : \mu_X \neq \mu_Y \end{cases}$$

Com significância $\alpha = 0.05$ para os casos

17.2.3.2.1 Caso 1. Variâncias conhecidas

Temos do produtor que $\sigma_X^2 = 81, \sigma_Y^2 = 209$ $\bar{d}_{\text{NPar}} = 70.5 - 84.3 = -13.8$. Logo, $z_{\frac{\alpha}{2}} \sqrt{\frac{81}{15} + \frac{209}{20}} = 7.8$. Como $-13.8 < -7.8$, concluímos que há evidências para rejeitarmos a hipótese nula de que as resistências médias das vigas A, B são iguais a $\alpha = 5\%$ de significância estatística

17.2.3.2.2 Caso 2. Variâncias desconhecidas e iguais.

Para $\alpha = 0.05$, $t_{33,0.025} = 2.03$. Encontrando $s_P^2 = 155.988$. Finalmente, $t_{33,0.025}\sqrt{\frac{s_P^2}{n} + \frac{s_P^2}{m}} = 8.65$. Como $-13.8 < -8.65$, concluímos que há evidências para rejeitarmos a hipótese nula a $\alpha = 5\%$ de significância estatística.

17.2.3.2.3 Caso 3. Variâncias desconhecidas e diferentes

Primeiro calculamos $n' = 32.08 \overset{\text{Arrendonda}}{=} 32$. Assim, $t_{32,0.025} = 2.037$. Portanto, $t_{32,0.025}\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}} = 8.14$. Como $-13.8 < -8.15$, concluímos que há evidências para rejeitarmos a hipótese nula a $\alpha = 5\%$ de significância estatística.

18 Testes não Paramétricos para a média populacional

Alguns Testes não-paramétricos para a média populacional podem ser estudados nas seções 13.3.2 e 13.4.2 do livro “Estatística Básica” (7ª edição), Morettin e Bussab.

Essa seção será completada em um futuro.

19 Tabela de frequências

Sejam X, Y variáveis aleatórias cujos valores observados são $B_1, B_2, \ldots, B_l$ e A_1, A_2, A_k, respectivamente. Observam-se os seguintes dados

ind.	X	Y
1	B_2	A_1
2	B_7	A_3
$\vdots$	$\vdots$	$\vdots$
n	B_1	A_5

Colocamos nossos dados numa tabela de frequências absolutas observadas

$X \setminus Y$	A_1	A_2	...	A_k	Total X
B_1	O_{11}	O_{12}	...	O_{1k}	$O_{1\cdot}$
B_2	O_{11}	O_{12}	...	O_{1k}	$O_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
B_l	O_{l1}	O_{l2}	...	O_{lk}	$O_{l\cdot}$
Total Y	$O_{\cdot_1}$	$O_{\cdot_2}$	...	$O_{\cdot k}$	n

Temos nossa tabela de frequências esperadas [[Teste de Hipótese|sob]] H_0 (Independência)

$X \setminus Y$	A_1	A_2	...	A_k	Total X
B_1	E_{11}	E_{12}	...	E_{1k}	$O_{1\cdot}$
B_2	E_{11}	E_{12}	...	E_{1k}	$O_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
B_l	E_{l1}	E_{l2}	...	E_{lk}	$O_{l\cdot}$
Total Y	$O_{\cdot_1}$	$O_{\cdot_2}$	...	$O_{\cdot k}$	n

Em que

$$E_{ij} = \frac{O_{i\cdot} \cdot O_{\cdot j}}{n}$$

Note que, sob independência

$$\begin{aligned} P(B_i \cap A_j) = P(B_i) \cdot P(A_j) \\ E_{ij} = n \cdot P(B_i \cap A_j) \overset{\text{ind.}}{=} nP(B_i) \cdot P_{A_j} \end{aligned}$$

Estimando $P(B_i), P(A_j)$ temos

$$\widehat{P(B_i)} = \frac{O_{i\cdot}}{n}, \widehat{P(A_j)} = \frac{O_{\cdot j}}{n}$$

Logo, o valor esperado estimado é

$$\widehat{E_{ij}} = n \cdot \widehat{P(B_i)} \cdot \widehat{P(A_j)} = \frac{O_{i\cdot} \cdot O_{\cdot j}}{n}$$

Em ambos testes, usaremos a seguinte estatística para testar suas hipóteses (independência e homogeneidade)

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Sob H_0, ou seja,

$$\chi^2_{obs} \sim \chi^2_{(k-1)(l-1)}$$

Dessa forma, rejeitamos a hipótese H_0 a α graus de liberdade se

$$\chi^2_{obs} > c_p$$

em que c_p satisfaz $P(\chi^2_{(k-1)(l-1)} > c_p) = \alpha$.

> **Observação**
>
> Essa aproximação com a χ^2 só funciona de modo razoável quando cada $E_{ij} > 5$

20 Teste Qui-Quadrado e análise de aderência

A análise de aderência testa a distribuição dos dados:

$$\begin{cases} H_0 : P = P_0 \\ H_1 : P \neq P_0 \end{cases}$$

Em que P_0 é a medida de probabilidade especificada que governaria (sob H_0) os eventos observados.

Neste teste co,paramos a frequência observada com a frequência esperada em k eventos disjuntos e distintos observáveis.

Eventos	1	2	...	k
P_0	P_{01}	P_{02}	...	P_{0k}
E_i	E_1	E_2	...	E_k
O_i	O_1	O_2	...	O_k

Em que observou-se uma amostra de tamanho n. Temos também que E_i é o valor esperado do número de eventos i sob H_0

$$\text{Freq.Esperada} = E_i = P_{0i} \cdot n$$

e Freq.Observada $= O_i$ é o numero real de eventos i observados na amostra. A estatística para testar H_0 é

$$\chi^2 = \sum_{i=1}^{k} \frac{(E_i - O_i)^2}{E_i}$$

que, sob H_0 - ou seja, sob a hipótese de que P_0 é de fato a medida de probabilidade que governa o comportamento probabilístico do evento - é aproximadamente

$$\underbrace{\chi^2 \sim \chi^2_{(k-1)}}_{\text{Sob } H_0}$$

*Esse procedimento é confiável sempre que $E_i > 5 \forall i \in \{1, \ldots, k\}$

20.1 Exemplo

Considere que queremos verificar se os números sorteados nos concursos da Mega Sena são de fato uniformemente distribuídos. Nesse caso, analisaremos 60 eventos, cuja probabilidade de cada um seria, caso uniformemente distribuídos, $\frac{1}{60}$.

$$\begin{cases} H_0 : P = P_0 \\ H_1 : P \neq P_0 \end{cases}$$

Em que $P_0(\{i\}) = \frac{1}{60} \forall i \in \{1, 2, ... 60\}$

Vamos criar a tabela para as frequências. Consideraremos *a primeira bola* de todos os 2800 sorteios da Mega.

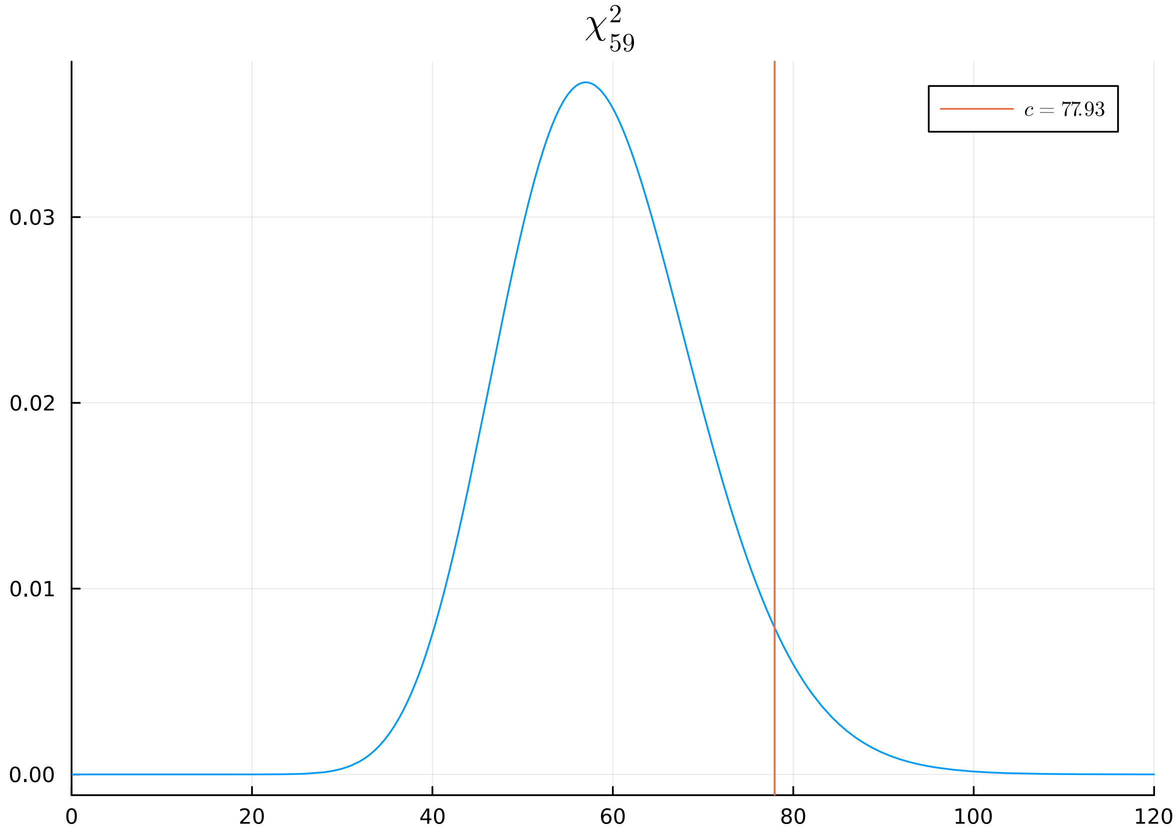
Eventos	1	2	...	60
P_0	$\frac{1}{60}$	$\frac{1}{60}$	...	$\frac{1}{60}$
E_i	$\frac{2800}{60}$	$\frac{2800}{60}$	...	$\frac{2800}{60}$
O_i	42	48	...	55

Portanto,

$$\chi^2 = \sum_{i}^{60} \frac{(46.7 - O_i)^2}{46.7} \overset{a}{\sim} \chi^2_{59}$$

Considerando um nível de significância de $\alpha = 5\%$, calculamos o ponto crítico c tal que

$$P(\chi^2_{59} > c) = 0.05$$



Pelo computador, encontramos $c = 77.93$ Logo, como $\chi^2 = 56.68 < 77.93$, concluímos que, sob H_0, não há evidências de que o modelo não seja equiprovável a 5% de significância de estatística.

20.2 K-Grupos

(Morettin, Pag.404 E.7) Considere os $n = 30$ dados abaixo que supostamente seguem uma distribuição normal $N(10, 25)$. (usando os dados do livro já em ordem)

$$\begin{matrix} 1.01 & 1.73 & 3.93 & 4.44 & 6.37 & 6.51 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 14.11 & 14.6 & 14.64 & 14.75 & 16.68 & 22.14 \end{matrix}$$

Queremos testar se os dados de fato se distribuem de acordo com $N(10, 25)$.

$$\begin{cases} H_0 : P = N(10, 25) \\ H_1 : P \neq N(10, 25) \end{cases}$$

Sob H_0, podemos dividir a distribuição normal em k blocos. Escolheremos $k = 4$ delimitado pelos *quartis* teóricos dessa distribuição normal. (Primeiro padronizamos, encontramos os valores pela tabela, então voltamos para nossa normal)

$$\begin{cases} q_1 = 6.63 \\ q_2 = 10 \\ q_3 = 13.3 \end{cases} \overset{\text{Intervalos}}{\Rightarrow} \begin{cases} 1.(-\infty, q_1) \\ 2.[q_1, q_2] \\ 3.(q_2, q_3] \\ 4.(q_3, \infty) \end{cases}$$

Podemos produzir uma tabela com as frequências por intervalo

Eventos	1.	2.	3.	4.
E_i	$0.25 \cdot 30 = 7.5$	7.5	7.5	7.5
O_i	6	9	9	6

$$\chi^2 = \sum_{i=1}^{4} \frac{(7.5 - O_i)^2}{7.5} = 1.2$$

Na χ^2_3 (número de nichos), com nível de significância $\alpha = 0.10$, $c = 6.25$. Como $\chi^2 = 1.2 < 6.25$, concluímos que não há evidências de que a distribuição dos dados difere de uma $N(10, 25)$ a $\alpha = 10\%$ de significância estatística

21 Testes de Independência e Homogeneidade

Com a ajuda da tabela de frequências, conseguimos testar independência entre eventos e homogeneidade em distribuição de eventos. Por mais que utilizem o mesmo mecanismo, os dois testes são interpretados de forma diferentes e, portanto, também apresentados individualmente nesta seção.

21.1 Teste de Homogeneidade

Usamos esse teste para verificar se as medidas de probabilidade de vários grupos diferentes são iguais (seguem uma mesma distribuição).

Os totais marginais para cada grupo devem ser fixados antes de executarmos o experimento.

$$\begin{cases} H_0 : \text{Os grupos são independentes} \\ H_1 : \text{Pelo menos um dos grupos não é indepndente} \end{cases}$$

21.2 Teste de independência

Usamos esse teste para verificar se os eventos são independentes.

Aqui, apenas o tamanho amostral (total dos totais) é fixado.

$$\begin{cases} H_0 : \text{Os grupos se distribuem de forma equivalente} \\ H_1 : \text{Pelo menos um dos grupos não se distribui de forma equivalente} \end{cases}$$

21.3 Exemplos

21.3.1 Primeiro exemplo (homogeniedade)

510 segurados foram amostrados, sendo 200 de São Paulo, 100 do Ceará e 210 de Pernambuco. O objetivo é verificar se o número de acidentes se distribui igualmente entre os estados.

Indivíduos	Estado	Sinistralidade
1	SP	1
⋮	⋮	⋮
200	SP	0
1	CE	1
⋮	⋮	⋮
100	CE	0
1	PE	1
⋮	⋮	⋮
210	PE	0

Tabela Observada

	Sinistralidade		
Estado	1	0	Total
SP	60	140	200
CE	10	90	100
PE	50	160	210
Total	120	390	510

Tabela esperada

	Sinistralidade		
Estado	1	0	Total
SP	47	153	200
CE	24	76	100
PE	49	161	210
Total	120	390	510

Temos nossa estatística qui-quadrado

$$\chi^2_{obs} = \frac{(60-47)^2}{47} + \frac{(140-153)^2}{153} + \frac{(10-24)^2}{24} + \frac{(90-76)^2}{76} + \frac{(50-49)^2}{49} + \frac{(160-161)^2}{161} = 15.47$$

Concluiremos o teste tomando $\alpha = 1\%$ de significância estatística Sabemos que, sob H_0,

$$\chi^2_{obs} \sim \chi^2_{(3-1)(2-1)}$$

Logo, devemos encontrar c_p tal que

$$P(\chi^2_2 > c_p) = 1\%$$

Pela tabela, $c_p = 9.21$ Como $15.47 > 9.21$, concluímos que, a sinistralidade não se distribui de forma homogênea entre os estados de SP, CE e PE a 1% de significância.

21.3.2 Outro Exemplo (independência)

Temos nossa tabela de valores observados:

Opinião	1ª Tent	2ª Tent	3ª Tent	Total
Excelente	62	36	12	110
Satisfatório	84	42	14	140
Insatisfatório	24	22	24	70
Total	170	100	50	320

Nossa tabela de valores esperados (arredondados):

Opinião	1ª Tent	2ª Tent	3ª Tent	Total
Excelente	58	34	17	110
Satisfatório	74	44	22	140
Insatisfatório	37	22	11	70
Total	170	100	50	320

$$\chi^2_{obs} = \frac{(62-58)^2}{58} + \frac{(36-34)^2}{34} + \frac{(12-17)^2}{17} + \frac{(84-74)^2}{74} + \frac{(42-44)^2}{44}$$
$$+ \frac{(14-22)^2}{22} + \frac{(24-37)^2}{37} + \frac{(22-22)^2}{22} + \frac{(24-11)^2}{11} = 26.14$$

Considerando $\alpha = 5\%$, precisamos encontrar c_p tal que

$$P(\chi^2_4 > c_p) = 5\% \Rightarrow c_p = 9.49$$

Como $26.14 > 9.49$, podemos concluir que, a 5% de significância estatística, existem evidências que o número da tentativa tem influência sobre a opinião do cliente.

22 Valor-p, p-Valor ou p-Value

Para todos os testes introduzidos até agora

Parte II

Inferência Frequentista

23 Família Exponencial (FE)

Podemos representar a função f_θ genericamente, incluindo funções de probabilidade e funções densidade de probabilidade, por meio da família exponencial.

23.1 Família Exponencial unidimensional

Dizemos que f_θ pertence à família exponencial unidimensional se, e somente se,

$$f_\theta(x) = \begin{cases} \mathrm{e}^{c(\theta)T(x)+d(\theta)+S(x)}, & x \in \mathfrak{X} \\ 0, & x \notin \mathfrak{X} \end{cases}$$

em que $c(\cdot)$ e $d(\cdot)$ são funções de "θ" cujas formas são conhecidas e $T(\cdot)$ e $S(\cdot)$ são funções de x com formas conhecidas, em que $\mathfrak{X} = \{x : f_\theta(x) > 0\}$ não depende de "θ". $\mathfrak{X}$ é dito ser o suporte de f_θ.

23.1.1 Exemplos

23.1.1.1 Exemplo 1 - Exponencial

Considere $X \sim f_\theta, \theta \in (0, \infty)$ em que

$$f_\theta(x) = \begin{cases} \theta \mathrm{e}^{-\theta x}, & x > 0 \\ 0, & \text{c.c} \end{cases}$$

Ou seja, $X \sim \mathrm{Exp}(\theta), \theta > 0$. Mostre que f_θ pertence à família exponencial.

23.1.1.1.1 Resposta

Observe que $\mathfrak{X} = \{x : f_\theta(x) > 0\} = (0, \infty)$ não depende de "θ". Além disso, como $\theta > 0$, temos que $\theta = \mathrm{e}^{\ln(\theta)}$. Portanto, para $x > 0$, temos que

$$f_\theta(x) = \mathrm{e}^{-\theta x + \ln \theta}$$

Assim,

$$\begin{array}{rl} c(\theta) = -\theta, & d(\theta) = \ln \theta \\ T(x) = x, & S(x) = 0. \end{array}$$

Portanto, f_θ pertence à família exponencial.

23.1.1.2 Exemplo 2 - Bernoulli

Considere $X \sim \text{Ber}(\theta), \theta \in \Theta = (0, 1)$. Mostre que a sua função de probabilidade pertence à família exponencial.

23.1.1.2.1 Resposta

Observe que

$$f_\theta(x) = \begin{cases} \theta^x \cdot (1-\theta)^{1-x}, & x \in \{0,1\} \\ 0, & x \notin \{0,1\} \end{cases}$$

o suporte é $\mathfrak{X} = \{x : f_\theta(x) > 0\} = \{0, 1\}$ não depende de "θ".

Além disso, como $\theta^x(1-\theta)^{1-x} > 0$, temos que

$$\begin{aligned} \theta^x(1-\theta)^{1-x} &= \mathrm{e}^{\ln(\theta^x(1-\theta)^{1-x})} \\ &= \mathrm{e}^{x \ln \theta + (1-x)\ln(1-\theta)} \\ &= \mathrm{e}^{x \ln \theta + \ln(1-\theta) - x\ln(1-\theta)} \\ &= \mathrm{e}^{x(\ln \theta - \ln(1-\theta)) + \ln(1-\theta)} \\ &= \mathrm{e}^{\ln(\frac{\theta}{1-\theta})x + \ln(1-\theta)} \end{aligned}$$

logo, $f_\theta(x) = \mathrm{e}^{c(\theta)T(x) + d(\theta) + S(x)}, x \in \{0, 1\}$, em que

$$\begin{array}{rl} c(\theta) = \ln(\frac{\theta}{1-\theta}), & T(x) = x \\ d(\theta) = \ln(1-\theta), & S(x) = 0. \end{array}$$

Portanto, f_θ pertence à família exponencial.

23.1.1.3 Exemplo 3 - Normal (Média)

Considere que $X \sim N(\theta, 1), \theta \in \mathbb{R}$. Mostre que a sua função densidade de probabilidade pertence à FE unidimensional.

23.1.1.4 Resposta

Note que

$$f_\theta(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}(x-\theta)^2}, x \in \mathbb{R}$$

o suporte é $\mathfrak{X} = \{x : f_\theta(x) > 0\} = (-\infty, \infty)$ e não depende de "θ". Além disso, temos que

$$\begin{aligned} f_\theta(x) &= \mathrm{e}^{-\frac{1}{2}(x-\theta)^2 - \ln\sqrt{2\pi}} \\ &= \mathrm{e}^{-\frac{1}{2}(x^2-2x\theta+\theta^2) - \ln\sqrt{2\pi}} \\ &= \mathrm{e}^{\theta x - \frac{1}{2}\theta^2 - \frac{1}{2}x^2 - \ln\sqrt{2\pi}}. \end{aligned}$$

Portanto, f_θ pertence à FE:

$$\begin{aligned} c(\theta) &= \theta, & T(x) &= x \\ d(\theta) &= -\tfrac{1}{2}\theta^2, & S(x) &= -\tfrac{1}{2}x^2 - \ln\sqrt{2\pi}. \end{aligned}$$

23.1.1.5 Exempo 4 - Normal (Variância)

Considere que $X \sim N(0, \theta), \theta \in (0, \infty)$. Mostre que a sua função densidade de probabilidade pertence à FE unidimensional.

23.1.1.5.1 Resposta

Observe que

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\theta}} \mathrm{e}^{-\frac{1}{2\theta}x^2}, x \in \mathbb{R}$$

O suporte $\mathfrak{X} = \mathbb{R}$ não depende de "θ". Além disso,

$$\begin{aligned} f_\theta(x) &= \mathrm{e}^{-\frac{1}{2\theta}x^2 - \frac{1}{2}\ln(2\pi\theta)} \\ &= \mathrm{e}^{c(\theta)T(x) + d(\theta) + S(x)} \end{aligned}$$

em que

$$\begin{aligned} c(\theta) &= -\tfrac{1}{2\theta}, & T(x) &= x^2 \\ d(\theta) &= -\tfrac{1}{2}\ln(2\pi\theta), & S(x) &= 0. \end{aligned}$$

Portanto, f_θ pertence à FE.

23.1.1.6 Exemplo 5 - Normal (Média = Variância)

Seja $X \sim N(\theta, \theta)$, em que $\theta \in (0, \infty)$. Mostre que a sua função densidade de probabilidade pertence à FE de dimensão 1.

23.1.1.6.1 Resposta

A função densidade de probabilidade de X é

$$\begin{aligned} f_\theta(x) &= \frac{1}{\sqrt{2\pi\theta}} \mathrm{e}^{-\frac{1}{2\theta}(x-\theta)^2} \\ &= \mathrm{e}^{-\frac{1}{2\theta}(x^2-2\theta x+\theta^2)-\frac{1}{2}\ln(2\pi\theta)} \\ &= \mathrm{e}^{-\frac{x^2}{2\theta}+x-\frac{1}{2}\theta-\frac{1}{2}\ln(2\pi\theta)} \\ &= \mathrm{e}^{c(\theta)T(x)+d(\theta)+S(x)} \end{aligned}$$

em que

$$\begin{aligned} c(\theta) &= -\frac{1}{2\theta}, & T(x) &= x^2 \\ d(\theta) &= -\frac{1}{2}\theta - \frac{1}{2}\ln(2\pi\theta), & S(x) &= x. \end{aligned}$$

Portanto, f_θ pertence à FE.

23.1.1.7 Exemplo 6 - Exemplo negativo (Uniforme)

Se $X \sim U(0,\theta), \theta \in (0,\infty)$, então a sua função densidade de probabilidade não pertence à FE, pois o seu suporte depende de "θ".

$$f_\theta(x) = \begin{cases} \frac{1}{\theta}, & x \in (0,\theta) \\ 0, & x \notin (0,\theta) \end{cases}$$

Dessa forma, $\mathfrak{X} = \{x : f_\theta(x) > 0\} = (0,\theta)$.

23.1.1.8 Exemplo 7 - Exemplo negativo (Normal - Média e Variância)

Considere que $X \sim N(\mu,\sigma^2)$, em que $\theta = (\mu,\sigma^2) \in \mathbb{R}\times\mathbb{R}^+$. Então, pode-se mostrar que a sua função densidade de probabilidade não pertence à FE **unidimensional**.

Observe que

$$\begin{aligned} f_\theta(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{1}{2\sigma^2}(x-\mu)^2} \\ &= \mathrm{e}^{-\frac{1}{2\sigma^2}(x^2-2x\mu+\mu^2)-\frac{1}{2}\ln(2\pi\sigma^2)} \\ &= \mathrm{e}^{-\frac{1}{2\sigma^2}x^2+\frac{x\mu}{\sigma^2}-\frac{1}{2\sigma^2}\mu^2-\frac{1}{2}\ln(2\pi\sigma^2)} \end{aligned}$$

portanto, não é possível definir $c(\theta), T(x), d(\theta)$ e $S(x)$ tais que $c(\cdot), T(\cdot)$ representem $-\frac{1}{2\sigma^2}x^2 + \frac{x\mu}{\sigma^2}$.

23.1.2 Propriedades na integração

Como f_θ é uma função (densidade) de probabilidade, temos que

$$\int_{-\infty}^{\infty} f_\theta(x)dx = 1(\text{caso contínuo}), \forall\theta \in \Theta$$

Se f_θ pertence à FE unidimensional, então

$$\begin{aligned} &\int_{\mathfrak{X}} \mathrm{e}^{c(\theta)T(x)+d(\theta)+S(x)}dx = 1\forall\theta\in\Theta \\ \Rightarrow &\int_{\mathfrak{X}} \mathrm{e}^{c(\theta)T(x)+S(x)}dx = \mathrm{e}^{-d(\theta)}\forall\theta\in\Theta. \end{aligned}$$

23.2 Família Exponencial k-dimensional

Dizemos que a função (densidade) de probabilidade f_θ pertence à FE k-dimensional se, e somente se,

$$f_\theta(x) = \begin{cases} \mathrm{e}^{\sum_{j=1}^{k} c_j(\theta)T_j(x)+d(\theta)+S(x)}, & x \in \mathfrak{X} \\ 0, & x \notin \mathfrak{X} \end{cases}$$

em que

1. $\mathfrak{X} = \{x : f_\theta(x) > 0\}$ não depende de "θ";
2. As funções $c_1(\cdot), \dots, c_k(\cdot)$ e $d(\cdot)$ dependem apenas de "θ" (formas conhecidas) e
3. As funções $T_1(\cdot), \dots, T_l(\cdot)$ e $S(\cdot)$ dependem apenas de x (formas conhecidas).

Como f_θ é uma função (densidade) de probabilidade, temos que (para o caso contínuo)

$$\int_{-\infty}^{\infty} f_\theta(x)dx = 1\forall\theta\in\Theta.$$

Se f_θ pertence à FE k-dimensional, então

$$\begin{aligned} &\int_{\mathfrak{X}} \mathrm{e}^{\sum_{j=1}^{k} c_j(\theta)T_j(x)+d(\theta)+S(x)}dx = 1\forall\theta\in\Theta \\ \Rightarrow &\int_{\mathfrak{X}} \mathrm{e}^{\sum_{j=1}^{k} c_j(\theta)T_j(x)+S(x)}dx = \mathrm{e}^{-d(\theta)}\forall\theta\in\Theta. \end{aligned}$$

23.2.1 Exemplos

23.2.1.1 Exemplo 1 (Normal - Média e Variância)

Seja $X \sim N(\mu, \sigma^2)$, em que $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$. Mostre que a sua função densidade de probabilidade pertence à FE de dimensão 2.

23.2.1.1.1 Resposta

Note que

$$f_\theta(x) = \mathrm{e}^{-\frac{1}{2\sigma^2}x^2 + \frac{x\mu}{\sigma^2} - \frac{1}{2\sigma^2}\mu^2 - \frac{1}{2}\ln(2\pi\sigma^2)} \quad = \mathrm{e}^{c_1(\theta)T_1(x) + c_2(\theta)T_2(x) + d(\theta) + S(x)}$$

em que

$$\begin{array}{rl} c_1(\theta) = -\frac{1}{2\sigma^2}, & T_1(x) = x^2 \\ c_2(\theta) = \frac{\mu}{\sigma^2}, & T_2(x) = x \\ d(\theta) = -\frac{1}{2}\frac{\mu^2}{\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2), & S(x) = 0 \end{array}$$

23.2.2 Exercício

Refaça o exemplo 5 do caso univarido com $N(\theta, \theta^2)$ e mostre que pertence à FE de dimensão 2.

24 Distribuição Amostral - Aprofundamento

Seja $(Y_1, \dots, Y_n)$ uma amostra aleatória de $Y \sim f_\theta, \theta \in \Theta$. A função (densidade) de probabilidade da amostra aleatória é dada por

$$f_\theta^{(n)}(y_1, \dots, y_n) \overset{\text{iid}}{=} \prod_{i=1}^{n} f_\theta(y_i), \forall \theta \in \Theta$$

e $y_i \in \mathbb{R}, i = 1, \dots, n$.

24.1 Relação com a FE

24.1.1 Unidimensional

Se f_θ pertence à Família Exponencial Unidimensional, então

$$\begin{aligned} f_\theta^{(n)}(y_1, \dots, y_n) &= \begin{cases} \prod_{i=1}^{n} \mathrm{e}^{c(\theta)T(y_i)+d(\theta)+S(y_i)}, & \text{se } y_1, \dots, y_n \in \mathfrak{X} \\ 0, & \text{c.c.} \end{cases} \\ \Rightarrow f_\theta^{(n)}(\underbrace{y_1, \dots, y_n}_{\boldsymbol{y}_n}) &= \begin{cases} \mathrm{e}^{c(\theta)\sum_{i=1}^{n} T(y_i)+nd(\theta)+\sum_{i=1}^{n} S(y_i)}, & \text{se } \boldsymbol{y}_n \in \mathfrak{X}^n \\ 0, & \text{c.c.} \end{cases} \end{aligned}$$

Em que $\mathfrak{X}^n = \mathfrak{X} \times \cdots \times \mathfrak{X}$. Portanto, $f_\theta^{(n)}$ pertence à FE unidimensional.

24.1.2 k-dimensional

Se f_θ pertence à FE k-dimensional, então

$$\begin{aligned} f_\theta^{(n)}(\boldsymbol{y}_n) &= \begin{cases} \prod_{i=1}^{n} \mathrm{e}^{\sum_{j=1}^{k} c_j(\theta)T_j(y_i)+d(\theta)+S(y_i)}, & \text{se } \boldsymbol{y}_n \in \mathfrak{X}^n \\ 0, & \text{c.c.} \end{cases} \\ \Rightarrow f_\theta^{(n)}(\boldsymbol{y}_n) &= \begin{cases} \mathrm{e}^{\sum_{j=1}^{k} c_j(\theta)\sum_{i=1}^{n} T_j(y_i)+nd(\theta)+\sum_{i=1}^{n} S(y_i)}, & \text{se } \boldsymbol{y}_n \in \mathfrak{X}^n \\ 0, & \text{c.c.} \end{cases} \end{aligned}$$

Tome $T_j^*(y_n) = \sum_{i=1}^n T_j(y_i)$, $d^*(\theta) = nd(\theta)$, $S^*(y_n) = \sum_{i=1}^n S(y_i)$

$$\Rightarrow f_\theta^{(n)}(y_n) = \begin{cases} \mathrm{e}^{\sum_{j=1}^k c_j(\theta) T_j^*(y_n) + d^*(\theta) + S^*(y_n)}, & \text{se } y_n \in \mathfrak{X}^n \\ 0, & \text{c.c} \end{cases}$$

Portanto, $f_\theta^{(n)}$ pertence à FE k-dimensional.

24.1.3 Exemplos

24.1.3.1 Bernoulli

Se $X \sim \text{Ber}(\theta), \theta \in (0,1)$, então

$$f_\theta(y) = \begin{cases} \theta^y \cdot (1-\theta)^{1-y}, & y \in \{0,1\} \\ 0, & \text{c.c.} \end{cases}$$
$$\Rightarrow f_\theta(y) = \begin{cases} \mathrm{e}^{\ln\left(\frac{\theta}{1-\theta}\right) y + \ln(1-\theta)}, & y \in \{0,1\} \\ 0, & \text{c.c.} \end{cases}$$

A função probabilidade da amostra é

$$f_\theta(y_n) = \begin{cases} \mathrm{e}^{\ln\left(\frac{\theta}{(1-\theta)}\right) \sum_{i=1}^n y_i + n \ln(1-\theta)}, & y_n \in \{0,1\}^n \\ 0, & \text{c.c.} \end{cases}$$

24.1.3.2 Binomial

Se $X \sim \text{Bin}(m, \theta), \theta \in (0,1)$ e m fixado, então

$$f_\theta(y) = \begin{cases} \binom{m}{y} \theta^y \cdot (1-\theta)^{m-y}, & y \in \{0, 1, \dots, m\} \\ 0, & \text{c.c.} \end{cases}$$

Note que

$$\begin{aligned} \binom{m}{y} \theta^y (1-\theta)^{m-y} &= \mathrm{e}^{\ln \binom{m}{y} + y \ln \theta + (m-y) \ln(1-\theta)} \\ &= \mathrm{e}^{\ln \binom{m}{y} + y \ln(\theta) + m \ln(1-\theta) - y \ln(1-\theta)}. \end{aligned}$$

Portanto,

$$f_\theta(y) = \begin{cases} \mathrm{e}^{\ln\left(\frac{\theta}{1-\theta}\right) y + m \ln(1-\theta) + \ln \binom{m}{y}}, & y \in \{0, 1 \dots, m\} \\ 0, & \text{c.c.} \end{cases}$$

A função probabilidade da amostra é

$$f_\theta(y_n) = \begin{cases} \mathrm{e}^{\ln\left(\frac{\theta}{1-\theta}\right)\sum_{i=1}^n y_i + nm\ln(1-\theta) + \sum_{i=1}^n \ln\binom{m}{y_i}}, & y_n \in \mathfrak{X}^n \\ 0, & \text{c.c.} \end{cases}$$

em que $\mathfrak{X} = \{0, 1, \dots, m\}$.

24.1.3.3 Dirichlet

Se $X \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_k)$, então $X = (Y_1, \dots, Y_k)^T$ é um vetor (coluna) aleatório k-dimensional cuja função densidade de probabilidade é dada por

$$f_\theta(y) = \begin{cases} \frac{\Gamma(\sum_{j=1}^k \alpha_j)}{\prod_{j=1}^k \Gamma(\alpha_j)} \prod_{j=1}^k y_j^{\alpha_j - 1}, & y_j \in \mathfrak{X} \\ 0, & \text{c.c.} \end{cases}$$

em que $\mathfrak{X} = \{(y_1, \dots, y_k) \in (0,1)^k : \sum_{j=1}^k y_j = 1\}$ e o vetor de parâmetros é $\theta = (\alpha_1, \dots, \alpha_k) \in \mathbb{R}_+^k$.

A amostra aleatória é

$$X_1 = \begin{pmatrix} Y_{11} \\ \vdots \\ Y_{k1} \end{pmatrix}, X_2 = \begin{pmatrix} Y_{12} \\ \vdots \\ Y_{k2} \end{pmatrix}, \dots, X_n = \begin{pmatrix} Y_{1n} \\ \vdots \\ Y_{kn} \end{pmatrix}$$

e sua função densidade de probabilidade é dada por

$$f_\theta(y_n) \stackrel{\text{iid}}{=} \begin{cases} \left[\frac{\Gamma(\sum_{j=1}^k \alpha_j)}{\prod_{j=1}^k \Gamma(\alpha_j)}\right]^n \prod_{i=1}^n \left(\prod_{j=1}^k y_{ij}^{\alpha_j - 1}\right), & y_i \in \mathfrak{X}, \forall i = 1, \dots, n \\ 0, & \text{c.c.} \end{cases}$$

Tome

$$g(\theta) = \left[\frac{\Gamma(\sum_{j=1}^k \alpha_j)}{\prod_{j=1}^k \Gamma(\alpha_j}\right]^n$$

$$\Rightarrow f_\theta^{(n)}(y_n) = \begin{cases} g(\theta) \prod_{i=1}^n \prod_{j=1}^k y_{ij}^{\alpha_j - 1}, & y_1, \dots, y_n \in \mathfrak{X} \\ 0, & \text{c.c.} \end{cases}$$

Note que

$$\begin{aligned} g(\theta) \prod_{i=1}^n \prod_{j=1}^k y_{ij}^{\alpha_j - 1} &= \mathrm{e}^{\ln g(\theta) + \sum_{i=1}^n \sum_{j=1}^k (\alpha_j - 1) \ln y_{ij}} \\ &= \mathrm{e}^{\ln g(\theta) + \sum_{j=1}^k (\alpha_j - 1) \sum_{i=1}^n \ln y_{ij}} \end{aligned}$$

$$\Rightarrow f_\theta^{(n)}(y_n) = \begin{cases} e^{\sum_{j=1}^{k} c_j^*(\theta) T_j^*(y_n) + d^*(\theta) + S^*(y_n)}, & y_n \in \mathfrak{X}^n \\ 0, & \text{c.c.} \end{cases}$$

Portanto, pertence à FE k-dimensional.

Observação

A partir de $f_\theta^{(n)}$ conseguimos fazer inferência sobre a quantidade de interesse: podemos encontrar a distribuição de estatísticas e estimadores.

25 Estatísticas Suficientes

Seja $X_n(X_1, \dots, X_n)$ uma amostra aleatória de $X \sim f_\theta, \theta \in \Theta$. Dizemos que uma estatística $T(X_n)$ é suficiente para o modelo estatístico se, e somente se, a distribuição da amostra dado que $T(X_n) = t$ não depende de "θ". Ou seja,

$$P_\theta(X_1 \leq y_1, \dots, X_n \leq y_n | T(X_n) = t) \text{ Não depende de } \theta, \forall y_1, \dots, y_n \in \mathbb{R}$$
$$\text{E para todo valor de } t \text{ para o quais a distribuição de } T(X_n) \text{ exista}$$

Em outras palavras, a informação probabilística sobre "θ" da amostra aleatória está inteiramente contida no modelo induzido pela estatística.

No caso discreto, basta mostrar que $P_\theta(X_1 = y_1, \dots, X_n = y_n | T(X_n) = t)$ não depende de "θ" para todo $y_1, \dots, y_n \in \mathbb{R}$ e valores de t para os quais a distribuição de $T(X_n)$ exista.

No caso contínuo, basta mostrar que $f_\theta^{(n)}(y_1, \dots, y_n | t)$ não depende de θ para todo $y_1, \dots, y_n \in \mathbb{R}$ e valores de t para os quais a função densidade de probabilidade de $T(X_n)$ exista.

Como discutiremos adiante, podemos substituir a restrição "$\in \mathbb{R}$" pelo termo "quase certamente" (q.c.), isto é, para todos exceto um conjunto enumerável (de medida de probabilidade nula).

Também, $f_\theta^{(n)}$ será reescrita por $f_\theta^{X_n}$ para diferenciar a função (densidade) da amostra, da estatística e das condicionais.

Usaremos y no lugar de x para distinguir que y representa valores observados *em potencial*, e não realmente observados de uma amostra real colhida, como poderia estar subtendido com o uso de x.

25.1 Caso discreto

25.1.1 Exemplos

25.1.1.1 Exemplo 1

Seja $X_n = (X_1, \dots, X_n)$ uma a.a. de $X \sim \text{Ber}(\theta), \theta \in \Theta = (0, 1)$. Verifique se $T(X_n) = \sum_{i=1}^n X_i$ é suficiente para o modelo estatístico.

25.1.1.1.1 Resposta

Por definição,

$$P_\theta(X_1 = y_1, \dots, X_n = y_n | T(X_n) = t) = \frac{P_\theta(X_1 = y_1, \dots, X_n = y_n, T(X_n) = t)}{P_\theta(T(X_n) = t)}$$

Sabemos que $T(X_n) = \sum_{i=1}^n X_i \sim \text{Bin}(n, \theta)$ (por função geradora de momentos). Portanto,

$$P_\theta(T(X_n) = t) = \begin{cases} \binom{n}{t} \cdot \theta^t \cdot (1-\theta)^{n-t}, & \text{se } t \in \{0, 1, \dots, n\} \\ 0, & \text{c.c.} \end{cases}$$

Logo, $P_\theta(X_1 = y_1, \dots, X_n = y_n | T(X_n) = t)$ só está bem definida se $t \in \{0, 1, \dots, n\}$.

(Numerador)

$$\begin{aligned} & P_\theta(X_1 = y_1, \dots, X_n = y_n, T(X_n) = t) \\ \overset{\text{TPT}}{=} & \overbrace{P_\theta(X_1 = y_1, \dots, X_n = y_n)}^{A} \cdot \overbrace{P_\theta(T(X_n) = t | X_1 = y_1, \dots, X_n = y_n)}^{B} \end{aligned}$$

Observe que A é a função probabilidade da amostra e

$$B = \begin{cases} 1, & \text{se } t = \sum_{i=1}^n y_i \\ 0, & \text{c.c} \end{cases}$$

Para $t \in \{0, 1, \dots, n\}$

$$P_\theta(X_1 = y_1, \dots, X_n = y_n | T(X_n) = t) = \begin{cases} \frac{\prod_{i=1}^n \theta^{y_i} (1-\theta)^{1-y_i}}{\binom{n}{t} \cdot \theta^t \cdot (1-\theta)^{n-t}}, & \text{se } t = \sum_{i=1}^n y_i \\ 0, & \text{c.c} \end{cases} \quad \forall \theta \in \Theta$$

Portanto,

$$P_\theta(X_1 = y_1, \dots, X_n = y_n | T(X_n) = t) = \begin{cases} \frac{\theta^{\sum_{i=1}^n y_i} (1-\theta)^{\sum_{i=1}^n 1-y_i}}{\binom{n}{t} \cdot \theta^t \cdot (1-\theta)^{n-t}}, & \text{se } t = \sum_{i=1}^n y_i \\ 0, & \text{c.c} \end{cases} \quad \forall \theta \in \Theta$$

Concluímos que

$$\begin{aligned} & P_\theta(X_1 = y_1, \dots, X_n = y_n | T(X_n) = t) = \begin{cases} \frac{\prod_{i=1}^n \mathbb{1}_{\{0,1\}}(y_i)}{\binom{n}{t}}, & \text{se } t = \sum_{i=1}^n y_i \\ 0, & \text{c.c} \end{cases} \\ & \forall \theta \in \Theta, t \in \{0, 1, \dots, n\}, \forall y_1, \dots, y_n \in \mathbb{R} \end{aligned}$$

A estatística $T(X_n)$ é suficiente para o modelo estatístico Bernoulli.

25.1.1.2 Exemplo 2

Seja $X_n = (X_1, \dots, X_n)$ uma amostra aleatória de $X \sim \text{Pois}(\theta), \theta \in \Theta = (0, \infty)$. Verifique se $T(X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$ é uma estatística suficiente para o modelo estatístico.

25.1.1.2.1 Resposta

$$P_\theta(X_1 = y_1, \dots, X_n = y_n | T(X_n) = t) = \frac{P_\theta(X_1 = y_1, \dots, X_n = y_n, T(X_n) = t)}{P_\theta(T(X_n) = t)}$$

(Numerador)

$$\begin{aligned} &P_\theta(X_1 = y_1, \dots, X_n = y_n, T(X_n) = t) \\ &\overset{\text{TPT}}{=} \overbrace{P_\theta(X_1 = y_1, \dots, X_n = y_n)}^{A} \cdot \overbrace{P_\theta(T(X_n) = t | X_1 = y_1, \dots, X_n = y_n)}^{B} \end{aligned}$$

Observe que A é a função probabilidade da amostra e

$$B = \begin{cases} 1, & \text{Se } t = \frac{1}{n}\sum_{i=1}^{n} y_i \\ 0, & \text{c.c} \end{cases}$$

(Denominador)

Já sabemos que $\sum_{i=1}^{n} X_i \sim \text{Pois}(n\theta)$ (por função geradora de momentos)

$$P_\theta\left(\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{k}{n}\right) = \begin{cases} \mathrm{e}^{-n\theta} \cdot \frac{(n\theta)^k}{k!}, & k \in \{0, 1, 2, \dots\} \\ 0, & \text{c.c.} \end{cases}$$

Tome $t = \frac{k}{n}$, então $k = nt$

$$P_\theta\left(\sum_{i=1}^{n} X_i = k\right) = \begin{cases} \mathrm{e}^{-n\theta} \cdot \frac{(n\theta)^{nt}}{(nt)!}, & t \in \{0, \frac{1}{n}, \frac{2}{n}, \dots\} \\ 0, & \text{c.c.} \end{cases}$$

Portanto, para $t \in \{0, \frac{1}{n}, \frac{2}{n}, \dots\}$, temos que

$$P_\theta(X_1 = y_1, \dots, X_n = y_n | T(X_n) = t) =$$
$$= \begin{cases} \frac{\prod_{i=1}^n \left\{ \mathrm{e}^{-\theta} \cdot \frac{\theta^{y_i}}{y_i!} \cdot \mathbb{1}_{\{0,1,\dots\}}(y_i) \right\}}{\mathrm{e}^{-n\theta} \cdot \frac{(n\theta)^{nt}}{(nt)!}}, & t = \frac{1}{n}\sum_{i=1}^n y_i \\ 0, & \text{c.c.} \end{cases}$$
$$= \begin{cases} \frac{\mathrm{e}^{-n\theta} \cdot \theta^{\sum_{i=1}^n y_i} \cdot \prod_{i=1}^n \mathbb{1}_{\{0,1,\dots\}}(y_i)(nt!)}{\prod_{i=1}^n (y_i!) \cdot \mathrm{e}^{-n\theta} \cdot (n\theta)^{nt}}, & t = \frac{1}{n}\sum_{i=1}^n y_i \\ 0, & \text{c.c.} \end{cases}$$
$$= \begin{cases} \frac{\prod_{i=1}^n \mathbb{1}_{\{0,1,\dots\}}(y_i)(nt!)}{\prod_{i=1}^n (y_i!)(n)^{nt}}, & t = \frac{1}{n}\sum_{i=1}^n y_i \\ 0, & \text{c.c.} \end{cases}$$
$$\forall \theta \in \Theta$$

Não depende de "θ" para todo $y_1, \dots, y_n \in \mathbb{R}$ e $t \in \{0, \frac{1}{n}, \frac{2}{n}, \dots\}$.

25.2 Caso Contínuo

25.2.1 Exemplo (Normal)

Seja $X_n = (X_1, \dots, X_n)$ a.a. de $X \sim N(\theta, 1), \theta \in \mathbb{R}$. Verifique se $T(X_n) = \sum_{i=1}^n X_i$ é suficiente para o modelo estatístico.

25.2.1.1 Resposta

Por definição

$$f_\theta^{X_n|T(X_n)=t}(y_1, \dots, y_n) = \frac{f_\theta^{X_n, T(X_n)}(y_1, \dots, y_n, t)}{f_\theta^{T(X_n)}(t)}$$

(Denominador) Já sabemos que $T(X_n) = \sum_{i=1}^n X_i \sim N(n\theta, n)$

$$\Rightarrow f_\theta^{T(X_n)}(t) = \frac{1}{\sqrt{2\pi n}} \cdot \mathrm{e}^{-\frac{1}{2n} \cdot (t - n\theta)^2}, t \in \mathbb{R}$$

(Numerador)

$$f_\theta^{X_n, T(X_n)}(y_1, \dots, y_n, t) = f_\theta^{X_n}(y_1, \dots, y_n) \cdot f_\theta^{T(X_n)|X_n=(y_1,\dots,y_n)}(t)$$

Note que

$$f_\theta^{T(X_n)|X_n=(y_1,\dots,y_n} = \begin{cases} 1, & t=\sum_{i=1}^n y_1 \\ 0, & \text{c.c.} \end{cases}$$

$$\Rightarrow f_\theta^{X_n|T(X_n)=t}(y_1,\dots,y_n) = \begin{cases} \frac{f_\theta^{X_n}(y_1,\dots,y_n)}{f_\theta^{T(X_n)}(t)}, & t=\sum_{i=1}^n y_i \\ 0, & \text{c.c.} \end{cases}$$

Logo

$$f_\theta^{T(X_n)|X_n=(y_1,\dots,y_n)} = \begin{cases} \frac{\frac{1}{\sqrt{2\pi\cdot 1}^n}\cdot\exp\left\{-\frac{1}{2}\sum_i^n=1(y_i-\theta)^2\right\}}{\frac{1}{\sqrt{2\pi\cdot n}}\cdot\exp\left\{-\frac{1}{2n}(t-n\theta)^2\right\}}, & t=\sum y_i \\ 0, & \text{c.c.} \end{cases}$$

Note que

$$-\frac{1}{2}\sum(y_i-\theta)^2 = -\frac{1}{2}\left(\frac{t^2}{n}-2t\theta+n\theta^2\right)$$

logo $f_\theta^{X_n|T(X_n)=t}$ não depende de θ e $\sum X_i$ é suficiente para o modelo estatístico.

25.2.2 Problema das funções densidade de probabilidade

A função densidade de probabilidade não é única. Entretanto, é única para quase todo ponto (quase certamente).

Por exemplo, $X\sim\text{Exp}(\theta), \theta\in(0,\infty)$

$$f_\theta(x) = \begin{cases} \theta\mathrm{e}^{-\theta x}, & x\in(0,\infty) \\ 0, & x\notin(0,\infty) \end{cases} \quad \theta\in\Theta$$

$$P_\theta(X>2) = \int_2^\infty \theta\cdot\mathrm{e}^{-\theta x}dx$$

Se A é enumerável e defina

$$f_\theta^A(x) = \begin{cases} \theta\mathrm{e}^{-\theta x}, & x\in(0,\infty)\setminus A \\ 10, & x\in \mathrm{A} \\ 0, & x\notin(0,\infty) \end{cases} \quad \theta\in\Theta$$

Temos que

$$f_\theta(x) = f_\theta^A(x), \forall x\in\mathbb{R}\setminus A, \forall\theta\in\Theta$$

e

$$f_\theta(x) \neq f_\theta^A(x), \forall x\in A, \forall\theta\in\Theta$$

Note que f_θ e f_θ^A são diferentes, mas produzem as mesmas probabilidades. Dizemos portanto que f_θ e f_θ^A são iguais *quase certamente*, ou seja,

$$P_\theta\left(f_\theta(x)=f_\theta^A(x)\right) = 1, \forall\theta\in\Theta$$

Ou, de outra forma,

$$P_\theta\left(f_\theta(x) \neq f_\theta^A(x)\right) = 0, \forall\theta \in \Theta$$

Notação:

$$f_\theta(x) = f_\theta^A(x) \text{ q.c. } \forall\theta \in \Theta$$

No caso contínuo, portanto, a estatística $T(X_n)$ será suficiente mesmo se $f_\theta^{X_n|T(X_n)=t}(y_1, \ldots, y_n)$ depende de "θ" para $(y_1, \ldots, y_n) \in A$, **DESDE QUE** $P_\theta(X \in A) = 0, \forall\theta \in \Theta$. Ou seja, pode depender de "θ" em um conjunto com probabilidade zero.

25.3 Critério da Fatoração de Neyman-Fisher (Caso Simples)

Seja $X_n = (X_1, \ldots, X_n)$ uma amostra aleatória de $X \sim f_\theta, \theta \in \Theta \subseteq \mathbb{R}^p$ (sendo as $f_\theta, \theta \in \Theta$ do mesmo "tipo", formalmente, dominadas pela mesma medida), em que $p \in \{1, 2, \ldots\}$. Uma estatística $T(X_n)$ é suficiente para o modelo estatístico se, e somente se, existirem funções $h(\cdot) : \mathbb{R}^n \to \mathbb{R}, m(\cdot,\cdot) : \mathrm{Im}(T) \times \theta \to \mathbb{R}$ (mensuráveis) tais que:

$$f_\theta^{X_n}(y_1, \ldots, y_n) = h(y_1, \ldots, y_n) \cdot m\left(T(y_1, \ldots, y_n), \theta\right), \forall\theta \in \Theta \text{ q.c.}$$

Obs:

1. h não depende de "θ";
2. m depende de valores amostrais por meio da estatística $T(X_n)$;
3. $f_\theta^{X_n} = f_\theta^{(n)}$ é a função (densidade) de probabilidade da amostra aleatória.
4. Note que a função de verossimilhança é obtida calculando $f_\theta^{X_n}$ na amostra observada, ou seja,

$$L_{X_n}(\theta) = f_\theta^{X_n}\underbrace{(x_1, \ldots, x_n)}_{X_n}$$

5. Alguns livros usam a função de verossimilhança no critério da fatoração

$$L_{X_n}(\theta) = h(X_n) \cdot m(T(X_n), \theta) \text{ q.c.} \forall\theta \in \Theta$$

em que $X_n = (x_1, \ldots, x_n)$ da amostra observada

25.3.1 Prova (caso discreto)

25.3.1.1 $\Rightarrow$

Assuma que $T(X_n)$ seja suficiente. Por definição, $P_\theta(X_1 = y_1, \dots, X_n = y_n | T(X_n) = t)$ não depende de "θ". Logo, podemos escrever:

$$P_\theta(X_1 = y_1, \dots, X_n = y_n | T(X_n) = t) = h^*(y_1, \dots, y_n, t) \forall \theta \in \Theta \tag{25.1}$$

Note também que,

$$P_\theta(X_1 = y_1, \dots, X_n = y_n | T(X_n) = t) = \frac{P_\theta(X_1 = y_1, \dots, X_n = y_n, T(X_n) = t)}{P_\theta(T(X_n) = t)}$$

para valores de t em que a probabilidade condicional exista.

$$\begin{aligned} &P_\theta(X_1 = y_1, \dots, X_n = y_n, T(X_n) = t) \\ &= P_\theta(X_1 = y_1, \dots, X_n = y_n) \cdot P_\theta(T(X_n) = t | X_1 = y_1, \dots, X_n = y_n). \end{aligned}$$

Como, com $y_n = y_1, \dots, y_n$,

$$P_\theta(T(X_n) = t | X_1 = y_1, \dots, X_n = y_n) = \begin{cases} 1, & T(y_n) = t \\ 0, & \text{cc} \end{cases}$$

então

$$P_\theta(X_1 = y_1, \dots, X_n = y_n | T(X_n) = t) \quad = \frac{P_\theta(X_1 = y_1, \dots, X_n = y_n) \cdot \mathbb{1}(T(y_n) = t)}{P_\theta(T(X_n) = t)} \tag{25.2}$$

Por (**??**) e (**??**), temos que

$$h^*(y_1, \dots, y_n, t) \cdot P_\theta(T(X_n) = t) = P_\theta(X_1 = y_1, \dots, X_n = y_n | \mathbb{1}(T(X_n) = t)).$$

Para $T(X_n) = t$, temos que

$$P_\theta(X_1 = y_1, \dots, X_n = y_n) = h^*(y_1, \dots, y_n, T(y_n)) \cdot P_\theta(T(X_n) = T(y_n))$$

25.3.1.2 $\Leftarrow$

Assuma que existam h, m tais que

$$P_\theta(X_1 = y_1, \dots, X_n = y_n) = h(\boldsymbol{y}_n) m(T(\boldsymbol{y}_n, \theta))$$

Note que

$$P_\theta(X_1 = y_1, \dots, X_n = y_n | T(\boldsymbol{X}_n) = t) = \begin{cases} \frac{P_\theta(X_1 = y_1, \dots, X_n = y_n)}{P_\theta(T(\boldsymbol{X}_n) = t)}, & T(\boldsymbol{y}_n) = t \\ 0, & \text{c.c.} \end{cases}$$

Observe que

$$P_\theta(T(\boldsymbol{X}_n) = t) = \sum_{(y_1, \dots, y_n) : T(\boldsymbol{y})_n = t} P_\theta(X_1 = y_1, \dots, X_n = y_n)$$

Por suposição

$$\begin{aligned} P_\theta(T(\boldsymbol{X}_n) = t) &= \sum_{\boldsymbol{y}_n : T(\boldsymbol{y}_n) = t} h(\boldsymbol{y}_n) \cdot m(T(\boldsymbol{y}_n), \theta) \\ &= \sum_{\boldsymbol{y}_n : T(\boldsymbol{y}_n) = t} h(\boldsymbol{y}_n) \cdot m(t, \theta) \\ &= m(t, \theta) \cdot \sum_{\boldsymbol{y}_n : T(\boldsymbol{y}_n) = t} h(\boldsymbol{y}_n) \end{aligned}$$

Portanto

$$P_\theta(X_1 = y_1, \dots, X_n = y_n | T(\boldsymbol{X}_n) = t) = \begin{cases} \frac{h(\boldsymbol{y}_n) \cdot m(T(\boldsymbol{y}_n), \theta)}{m(t, \theta) \cdot \sum_{\boldsymbol{y}_n : T(\boldsymbol{y}_n) = t} h(\boldsymbol{y}_n)}, & T(\boldsymbol{y}_n) = t \\ 0, & \text{c.c.} \end{cases}$$

Logo,

$$P_\theta(X_1 = y_1, \dots, X_n = y_n | T(\boldsymbol{X}_n) = t) = \begin{cases} \frac{h(\boldsymbol{y}_n)}{\sum_{\boldsymbol{y}_n : T(\boldsymbol{y}_n) = t} h(\boldsymbol{y}_n)}, & T(\boldsymbol{y}_n) = t \\ 0, & \text{c.c.} \end{cases}$$

25.3.2 Exemplo (1 do caso discreto)

Seja $\boldsymbol{X}_n = (X_1, \dots, X_n)$ amostra aleatória de $X \sim \text{Ber}(\theta), \theta \in \Theta = (0, 1)$. Verifique se $T(\boldsymbol{X}_n) = \sum_{i=1}^n X_i$ é suficiente para o modelo estatístico.

25.3.2.1 Resposta

Observe que

$$
f_\theta^{X_n}(y_1, \dots, y_n) = \begin{cases} \prod_{i=1}^n \left\{\theta^{y_i} \cdot (1-\theta)^{1-y_i}\right\}, & y_i \in \{0,1\}, \forall i = 1, \dots, n \\ 0, & \text{c.c.} \end{cases}
$$

$$
\Rightarrow f_\theta^{X_n}(y_1, \dots, y_n) = \theta^{\sum y_i} \cdot (1-\theta)^{n-\sum y_i} \cdot \prod_{i=1}^n \mathbb{1}_{\{0,1\}}(y_1)
$$

Tome $h(y_1, \dots, y_n) = \prod_{i=1}^n \mathbb{1}_{\{0,1\}}(y_1)$ e $m(T(y_1, \dots, y_n), \theta) = \theta^{\sum y_i} \cdot (1-\theta)^{n-\sum y_i}$ em que $T(y_1, \dots, y_n) = \sum_{i=1}^n y_i$. Temos que

$$
f_\theta^{X_n}(y_1, \dots, y_n) = h(y_1, \dots, y_n) m(T(y_1, \dots, y_n), \theta), \forall \theta \in \Theta
$$

Pelo critério da fatoração, $T(X_n) = \sum_{i=1}^n X_i$ é suficiente para o modelo de Bernoulli.

25.3.3 Mais Exemplos

25.3.3.1 Exemplo a

Seja X_n amostra aleatória de $X \sim \text{Beta}(a,b), \theta = (a,b) \in \Theta = \mathbb{R}_+^2$. Encontre uma estatística suficiente para o modelo.

25.3.3.1.1 Resposta

A função densidade de probabilidade da amostra aleatória é

$$
f_\theta^{X_n}(y_1, \dots, y_n) \overset{\text{a.a.}}{=} \prod_{i=1}^n f_\theta(y_i) \overset{\text{Beta}}{=} \prod_{i=1}^n \left\{ \frac{1}{\beta(a,b)} y_i^{a-1} \cdot (1-y_i)^{b-1} \right\}
$$

Em que

$$
\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}
$$
$$
\Gamma(a) = \int_0^\infty x^{a-1} \mathrm{e}^{-x} dx
$$

$$
\Rightarrow f_\theta^{X_n}(y_n) = \begin{cases} \frac{1}{\beta(a,b)^n} \left(\prod_{i=1}^n y_1\right)^{a-1} \cdot \left(\prod_{i=1}^n (1-y_1)\right)^{b-1}, & y_n \in (0,1)^n \\ 0, & \text{c.c.} \end{cases}
$$

Tome $h(y_n) = \prod_{i=1}^n \mathbb{1}(y_i)_{(0,1)}$ e

$$m(t,\theta) = \frac{1}{\beta(a,b)^n} \cdot t_1^{a-1} \cdot t_2^{b-1}$$

em que $t = (t_1, t_2)$ e $t_1 = \prod_{i=1}^n y_i, t_2 = \prod_{i=1}^n (1-y_i)$. Ou seja,

$$T(X_n) = \left(\prod_{i=1}^n X_i, \prod_{i=1}^n (1-X_i) \right)$$

é uma estatística suficiente para o modelo pois

$$f_\theta^{X_n}(y_n) = m(T(y_n),\theta) \text{ q.c. } \forall \theta \in \Theta.$$

25.3.3.2 Exemplo b

Seja X_n amostra aleatória de $X \sim \mathrm{N}(\mu, \sigma^2), \theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+$. Encontre uma estatística suficiente para o modelo.

25.3.3.2.1 Resposta

A função densidade de probabilidade da amostra aleatória é

$$\begin{aligned}
f_\theta^{X_n}(y_1, \dots, y_n) &\overset{\text{a.a.}}{=} \prod_{i=1}^n f_\theta(y_i), \forall y_n \in \mathbb{R}^n \\
&= \prod_{i=1}^n \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{ -\frac{1}{2\sigma^2}(y_i - \mu)^2 \right\} \right\} \\
&= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \cdot \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_1 - \mu)^2 \right\}
\end{aligned}$$

Note que $(y_1 - \mu)^2 = y_i^2 - 2y_i\mu + \mu^2$, logo,

$$f_\theta^{X_n} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \cdot \exp\left\{ -\frac{1}{\sigma^2} \left(\sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i + n\mu^2 \right) \right\}$$

Tome $h(y_n) = \frac{1}{(2\pi\sigma^2)}$ e

$$m(t,\theta) = \frac{1}{(\sigma^2)^{\frac{n}{2}}} \cdot \exp\left\{ -\frac{1}{2\sigma^2} \left(t_2 - 2\mu t_1 - \mu^2 \right) \right\}.$$

Em que $t = (t_1, t_2)$ e $t_1 = \sum_{i=1}^n y_i, t_2 = \sum_{i=1}^n y_i^2$

Ou seja, $T(X_n) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X^2 \right)$ é uma estatística suficiente para o modelo pois

$$f_\theta^{X_n}(y_n) \cdot m(T(X_n),\theta) \text{ q.c. } \forall \theta \in \Theta.$$

25.3.3.3 Exemplo c

Seja X_n amostra aleatória de $X \sim \text{Unif}(0, \theta), \theta = (a, b) \in \Theta = \mathbb{R}_+$. Encontre uma estatística suficiente para o modelo.

25.3.3.3.1 Respostas

A função densidade de probabilidade da amostra aleatória

$$f_\theta^{X_n}(y_1, \dots, y_n) \overset{a.a.}{=} \prod_{i=1}^{n} f_\theta(y_1) \forall y_n \in \mathbb{R}^n$$

Note que

$$f_\theta(x) = \begin{cases} \frac{1}{\theta}, & x \in (0, \theta] \\ 0, & \text{c.c.} \end{cases} = \frac{1}{\theta} \mathbb{1}(x)_{(0,\theta]}$$

Logo,

$$f_\theta^{X_n}(y_n) \overset{\text{Unif}}{=} \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}_{(0,\theta]}(y_i) = \frac{1}{\theta^n} \cdot \prod_{i=1}^{n} \mathbb{1}_{(0,\theta]}(y_i)$$

Note que

$$\prod_{i=1}^{n} \mathbb{1}_{(0,\theta]}(y_i) = 1 \Leftrightarrow \begin{cases} 0 < y_1 \leq \theta \\ \vdots \\ 0 < y_n \leq \theta \end{cases} \Leftrightarrow \begin{cases} \min(y_n) > 0 \\ \max(y_n) \leq \theta \end{cases}$$

Logo

$$\prod_{i=1}^{n} \mathbb{1}_{(0,\theta]}(y_i) = 1 \Leftrightarrow \mathbb{1}_{(0,\infty)}(\min(y_n)) \cdot \mathbb{1}_{(0,\theta]}(\max(y_n)) = 1$$

e

$$f_\theta^{X_n}(y_n) = \frac{1}{\theta^n} \mathbb{1}_{(0,\infty)}(\min(y_n)) \cdot \mathbb{1}_{(0,\theta]}(\max(y_n))$$

Tome $h(y_n) = \mathbb{1}_{(0,\infty)}(\min(y_n))$ e $m(t, \theta) = \frac{1}{\theta^n} \mathbb{1}_{(0,\theta]}(t)$, em que $t = \max(y_n)$

Portanto, pelo critério da fatoração, $T(X_n) = \max(X_1, \dots, X_n)$ é uma estatística suficiente para o modelo em questão.

25.4 Teorema ("Invariância" da estatística suficiente)

Seja $T(X_n)$ uma estatística suficiente para o modelo estatístico. Então

$$G(X_n) = s(T(X_n))$$

é uma estatistica suficiente se $s(\cdot)$ for bijetora (só precisa ser injetora)

25.4.1 Prova

Como $T(X_n)$ é suficiente para o modelo temos, pelo critério da fatoração, que

$$f_\theta^{X_n}(y_n) = h(y_n) \cdot m(T(y_n), \theta) \text{ q.c. } \forall \theta \in \Theta$$

Como $s(\cdot)$ é bijetora, temos que sua inversa existe:

$$T(X_n) = s^{-1}(G(X_n)) \Rightarrow T(y_n) = s^{-1}(G(y_n))$$

Portanto,

$$f_\theta^{X_n}(y_n) = h(y_n) \cdot m(s^{-1}(G(X_n)), \theta) \text{ q.c. } \forall \theta \in \Theta$$

Tome $m^*(\cdot, \theta) = m(s^{-1}(\cdot), \theta)$. Substituindo, temos que

$$f_\theta^{X_n}(y_n) = h(y_n) \cdot m^*((G(X_n)), \theta) \text{ q.c. } \forall \theta \in \Theta$$

Logo, pelo critério da fatoração, $G(X_n)$ é suficiente para o modelo estatístico.

25.4.2 Exemplos

a) Se $T(X_n) = \sum_{i=1}^n X_i$ é suficiente para o modelo, então:

1. $G(X_n) = \frac{1}{n} \sum_{i=1}^n X_i$ é suficiente para o modelo;
2. $G(X_n) = \mathrm{e}^{\sum_{i=1}^n X_i}$ é suficiente para o modelo;
3. Para $\sum X_i \neq 0$ q.c., $G(X_n) = \frac{1}{\sum_{i=1}^n X_i}$ é suficiente para o modelo;
4. $G(X_n) = \left(\sum_{i=1}^n X_i\right)^2$ não é necessariamente suficiente para o modelo uma vez que $f(x) = x^2$ não é injetora.

b) Se $T(X_n)$ for suficiente para o modelo estatístico, então

1. $G(X_n) = (T(X_n), T(X_n))$ é suficiente para o modelo;
2. $G(X_n) = (T(X_n), X_1)$ é suficiente para o modelo;
3. $G(X_n) = (T(X_n), X_n)$ é suficiente para o modelo pois X_n é suficiente para o modelo;
4. $G(X_n) = (X_1 \dots, X_n)$ "quase" nunca será suficiente para o modelo;
5. $G(X_n) = (X_n, X_1)$ é suficiente para o modelo.
6. A amostra ordenada $X_{(1)} \leq \cdots \leq X_{(n)}$, denotada por $T^*(X_n) = (X_{(1)}, \dots, X_{(n)})$ é uma estatística suficiente para o modelo pelo critério da fatoração, no caso de variáveis aleatórias independentes e identicamente distribuídas. $T^*(X_n)$ é dita ser a *estatística de ordem*;

25.4.3 Corolário útil (Relação das suficientes com a FE)

Se f_θ pertencer à família exponencial k-dimensional, então $T(X_n) = \left(\sum_{i=1}^n T_1(X_i), \dots, \sum_{i=1}^n T_k(X_i)\right)$ é uma estatística suficiente para o modelo.

25.4.3.1 Prova:

$$
\begin{aligned}
f_\theta^{X_n}(y_n) &\overset{\text{iid}}{=} \prod_{i=1}^n f_\theta(y_i) = \prod_{i=1}^n \left\{ \exp\left\{ \sum_{j=1}^n c_j(\theta) T_j(y_i) + d(\theta) + S(y_i) \cdot \mathbb{1}_{\mathfrak{X}}(y_i) \right\} \right\} \\
&= \exp\left\{ \sum_{j=1}^n c_j(\theta) \cdot \sum_i T_j(y_i) + nd(\theta) + \sum_{i=1}^n S(y_i) \cdot \prod_{i=1}^n \mathbb{1}_{\mathfrak{X}}(y_i) \right\}
\end{aligned}
$$

Tome $h(y_n) = \prod \mathbb{1}_{\mathfrak{X}}(y_i) \cdot \mathrm{e}^{\sum S(y_i)}$, $m(t, \theta) = \exp\{c_1(\theta) t_1 + \cdots + c_k(\theta) t_k + nd(\theta)\}$, em que $t = (t_1, \dots, t_k)$ e

$$
\begin{aligned}
t_1 &= \sum T_1(y_i) \\
&\vdots \\
t_k &= \sum T_k(y_i)
\end{aligned}
$$

Pelo critério da fatoração,

$$
T(X_n) = \left(\sum_{i=1}^n T_1(X_i), \dots, \sum_{i=1}^n T_k(X_i) \right)
$$

Isso é útil para encontrar estatísticas completas e suficientes pelo Teorema das FEs.

25.5 Estatísticas Suficientes Minimais (SM)

Dizemos que $T(X_n)$ é uma estatística suficiente minimal para o modelo se, e somente se:

1. $T(X_n)$ é suficiente para o modelo
2. Para qualquer outra estatística suficiente $U(X_n)$, existe uma função H tal que

$$T(X_n) = H(U(X_n)), \text{ q.c.}$$

Obs: A σ-álgebra associada à estatística suficiente minimal é a menor σ-álgebra dentre aquelas associadas às estatísticas suficientes.

25.5.1 Teorema (1 das estatísticas SM)

Seja $X_n = (X_1, \dots, X_n)$ de $X \sim f_\theta, \theta \in \Theta_0 = \{\theta_0, \theta_1, \dots, \theta_p\}$ em que $\mathfrak{X} = \{x : f_\theta(x) > 0\}$ não depende de "θ". Então,

$$T(x) = \left(\frac{f_{\theta_1}^X(x)}{f_{\theta_0}^{X_n}(x)}, \dots, \frac{f_{\theta_p}^X(x)}{f_{\theta_0}^{X_n}(x)} \right)$$

em que $T : \mathfrak{X} \to \mathbb{R}^p$ é uma estatística suficiente minimal ($T(X_n)$) para o modelo estatístico.

Isso trata de *razões entre funções verossimilhança.*

25.5.1.1 Prova

Note que, $\forall y_n \in \mathfrak{X}$, temos que

$$f_{\theta_j}^{X_n}(y_n) = f_{\theta_0}^{X_n}(y_n) \cdot \frac{f_{\theta_j}^{X_n}(y_n)}{f_{\theta_0}^{X_n}(y_n)}$$

Tome $h(y_n) = f_{\theta_0}^{X_n}(y_n)$, não depende dos diferentes valores de "θ" e

$$m(T(X_n), \theta) = \begin{cases} T_1(y_n), & \theta = \theta_1 \\ T_2(y_n), & \theta = \theta_2 \\ \vdots & \\ T_p(y_n), & \theta = \theta_p \end{cases}$$

Em que $T(x) = (T_1(x), \dots, T_p(x)$ e $T_j = \frac{f_{\theta_j}^{X_n}(y_n)}{f_{\theta_0}^{X_n}(y_n)}$.

Logo, $T(X_n)$ é suficiente para o modelo pelo critério da fatoração.

Seja $U(X_n)$ uma estatística suficiente para o modelo. Então pelo critério da fatoração,

$$f_\theta^{X_n}(y_n) = h'(y_n) \cdot m'(U(y_n), \theta) \text{ q.c. } \forall \theta \in \Theta_0$$

Observe que $\forall y_n \in \mathfrak{X}$,

$$\frac{f_{\theta_j}^{X_n}(y_n)}{f_{\theta_0}^{X_n}(y_n)} = \frac{h'(y_n) \cdot m'(U(y_n), \theta_j)}{h'(y_n) \cdot m'(U(y_n), \theta_0)} \text{ q.c. } \forall \theta \in \Theta_0$$

Logo,

$$T_j(y_n) = \frac{f_{\theta_j}^{X_n}(y_n)}{f_{\theta_0}^{X_n}(y_n)} = \frac{m'(U(y_n), \theta_j)}{m'(U(y_n), \theta_0)},\ j = 1, \dots, p$$

Portanto, existe H tal que

$$T_j(X_n) = H(U(X_n))$$

basta tomar

$$H(u) = \left(\frac{m'(u, \theta_1)}{m'(u, \theta_0)}, \dots, \frac{m'(u\theta_p)}{m'(u, \theta_0)} \right)$$

25.5.1.2 Exemplo (Bernoulli)

Seja $X_n = (X_1, \dots, X_n)$ amostra aleatória de $X \sim \text{Ber}(\theta), \theta \in \Theta = \{0.1, 0, 5\}$. Encontre uma estatística suficiente minimal.

25.5.1.2.1 Resposta

Pelo teorema anterior,

$$T(y_n) = \frac{f_{\theta_1}^{X_n}(y_n)}{f_{\theta_0}^{X_n}(y_n)}$$

é suficiente minimal, em que $\theta_0 = 0.1, \theta_1 = 0.5$ e

$$\begin{gathered} f_{0.1}^{X_n}(y_n) = 0.1^{\sum y_i} \cdot 0.9^{n - \sum y_i} \\ f_{0.5}^{X_n}(y_n) = 0.5^{\sum y_i} \cdot 0.5^{n - \sum y_i} \\ \forall y_n \in \mathfrak{X} = \{0, 1\}^n \end{gathered}$$

Logo

$$\begin{aligned}
T(y_n) &= \frac{0.5^n}{0.1^{\sum y_i} \cdot 0.9^{n-\sum y_i}} \\
&= \frac{0.5^n}{\left(\frac{0.1}{0.9}\right)^{\sum y_i} \cdot 0.9^{\sum y_i}} \\
&= 9^{\sum y_i} \cdot \left(\frac{5}{9}\right)^n
\end{aligned}$$

Note que $T(y_n)$ é função 1:1 de $T'(y_n) = \sum_{i=1}^n y_i$. Logo,

$$T'(X_n) = \sum_{i=1}^n X_i$$

é também uma estatística suficiente minimal para o modelo.

25.5.1.3 Exemplo (Normal)

Seja $X_n = (X_1, \dots, X_n)$ amostra aleatória de $X \sim \mathrm{N}(\mu, \sigma^2), \theta = (\mu, \sigma^2) \in \Theta = \{(0,1), (1,1), (0,2)\}$. Encontre uma estatística suficiente minimal.

25.5.1.3.1 Resposta

Pelo teorema,

$$T(y_n) = \left(\frac{f_{\theta_1}^{X_n}(y_n)}{f_{\theta_0}^{X_n}(y_n)}, \frac{f_{\theta_2}^{X_n}(y_n)}{f_{\theta_0}^{X_n}(y_n)}\right)$$

Tomando $\theta_0 = (0,1), (1,1), (0,2)$, temos

$$\begin{aligned}
f_{\theta_0}^{X_n}(y_n) &= \frac{\exp\left\{-\frac{1}{2}\sum y_i^2\right\}}{(\sqrt{2\pi})^n} \\
f_{\theta_1}^{X_n}(y_n) &= \frac{\exp\left\{-\frac{1}{2}(\sum y_i^2 - 2\sum y_i + n)\right\}}{(\sqrt{2\pi})^n} \\
f_{\theta_2}^{X_n}(y_n) &= \frac{\exp\left\{-\frac{1}{4}\sum y_i^2\right\}}{(\sqrt{4\pi})^n}
\end{aligned}$$

Portanto,

$$\begin{aligned}
T(y_n) &= \left(\exp\left\{-\frac{1}{2}(n - 2\sum y_i)\right\}, \frac{\exp\left\{-\frac{1}{4}\sum y_i^2 + \frac{1}{2}\sum y_i^2)\right\}}{2^{\frac{n}{2}}},\right) \\
&= \left(\exp\left\{-\frac{1}{2}(n - 2\sum y_i)\right\}, \frac{\exp\left\{\frac{1}{4}\sum y_i^2)\right\}}{2^{\frac{n}{2}}},\right)
\end{aligned}$$

dessa forma, $T(X_n)$ é SM para o modelo.

Note que $T(y_n)$ é função 1:1 de $(\sum y_i, \sum y_i^2)$. Portanto,

$$T'(X_n) = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$$

é também SM para o modelo.

25.5.2 Teorema (2 das estatísticas SM)

Seja $X_n = (X_1, \dots, X_n)$ amostra aleatória de $X \sim f_\theta, \theta \in \Theta$. Considere $\Theta_0 \subseteq \Theta$ não vazio. Então, se $T(X_n)$ for SM para o modelo reduzido a Θ_0 e suficiente para Θ, então será suficiente minimal para Θ.

25.5.2.1 Prova

Como $T(X_n)$ é SM para o modelo restrito a $\Theta_0 \subseteq \Theta$, para qualquer estatística $U(X_n)$ suficiente para o modelo restrito a Θ_0, existe H tal que

$$T(X_n) = H(U(X_n)) \text{ q.c.}$$

Observe que todas as estatísticas suficientes *para o modelo completo* Θ são também suficientes *para os modelos restritos*. Como $T(X_n)$ também é, por hipótese, suficiente para o modelo completo, então é função de qualquer estatística suficiente minimal para o modelo completo.

25.5.2.2 Exemplo (Bernoulli)

Seja $X_n = (X_1, \dots, X_n)$ amostra aleatória de $X \sim \text{Ber}(\theta), \theta \in \Theta = (0, 1)$. Mostre que $T(X_n) = \sum X_i$ é uma estatística suficiente minimal para o modelo.

25.5.2.2.1 Resposta

Tome $\Theta_0 = \{0.1, 0.5\} \subseteq \Theta$. Já mostramos que $T'(X_n) = \sum_{i=1}^{n} X_i$ é suficiente minimal para o modelo reduzido a Θ_0. Além disso, $T'(X_n) = \sum X_i$ é suficiente para o modelo completo. Logo, pelo Teorema 2 para estatísticas suficientes minimais, concluímos que $T'(X_n) = \sum X_i$ é também SM para o modelo completo.

25.5.2.3 Exemplo (Normal)

Seja $X_n = (X_1, \dots, X_n)$ amostra aleatória de $X \sim \mathrm{N}(\mu, \sigma^2), \theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}^+$. Mostre que $T(X_n) = (\sum X_i, \sum X_i^2)$ é suficiente minimal para o modelo.

25.5.2.3.1 Resposta correta

Tome $\Theta_0 = \{(0,1),(1,1),(0,2)\}$. Já mostramos que essa estatística é suficiente minimal para o modelo reduzido a Θ_0. Além disso, já mostramos (lista) que é suficiente para o modelo completo. Logo, pelo teorema 2, é SM para o modelo completo.

25.5.2.3.2 Tentativa alternativa (frustrada)

Tome $\Theta_0 = \{(0,1),(1,1)\}$. Pelo teorema 1, temos que

$$T(X_n) = \frac{f_{\theta_1}^{X_n}(X_n)}{f_{\theta_0}^{X_n}(X_n)}$$

é SM para o modelo reduzido.

$$T(X_n) = \frac{\frac{1}{(2\pi)^{\frac{n}{2}}}\exp\left\{-\frac{1}{2}(\sum X_i^2 - 2\sum X_i + n)\right\}}{\frac{1}{(2\pi)^{\frac{n}{2}}}\exp\left\{-\frac{1}{2}\sum X_i^2\right\}} = \mathrm{e}^{\sum X_i - \frac{1}{2}n}$$

Como $\sum X_i$ é uma função 1:1 da estatística, temos que $\sum X_i$ é também SM para o modelo reduzido. Entretanto, $\sum X_i$ não é suficiente para o modelo completo.

25.5.2.3.3 Tentativa 2

Tome $\Theta_0 = \{(0,1),(1,1),(2,1)\}$. Então, pelo teorema 1, temos que

$$T(X_n) = \left(\frac{f_{\theta_1}^{X_n}(X_n)}{f_{\theta_0}^{X_n}(X_n)}, \frac{f_{\theta_2}^{X_n}(X_n)}{f_{\theta_0}^{X_n}(X_n)}\right) \overset{\text{Tent. Ant.}}{=} \left(\mathrm{e}^{\sum X_i - \frac{1}{2}n}, \frac{f_{\theta_2}^{X_n}(X_n)}{f_{\theta_0}^{X_n}(X_n)}\right)$$

Temos que

$$\frac{f_{\theta_2}^{X_n}(X_n)}{f_{\theta_0}^{X_n}(X_n)} = \frac{\frac{1}{(2\pi)^{\frac{n}{2}}}\exp\left\{-\frac{1}{2}(\sum X_i^2 - 4\sum X_i + 4n)\right\}}{\frac{1}{(2\pi)^{\frac{n}{2}}}\exp\left\{-\frac{1}{2}\sum X_i^2\right\}} = \mathrm{e}^{2\sum X_i - 2n}$$

Logo,

$$T(X_n) = \left(\mathrm{e}^{\sum X_i - \frac{1}{2}n}, \mathrm{e}^{2\sum X_i - 2n}\right)$$

Como $\sum X_i$ é função 1:1 de $T(X_n)$, temos que é SM para o modelo reduzido a Θ_0. Entretanto, não é suficiente para o modelo completo.

Observe que o espaço paramétrico do modelo completo é $\mathbb{R} \times \mathbb{R}^+$

25.5.2.4 Exemplo (Uniforme)

Seja $X_n = (X_1, \dots, X_n)$ amostra aleatória de $X \sim U(0, \theta), \theta \in \Theta = (0, \infty)$. Verifique se $X_{(n)} = \max\{X_1, \dots, X_n\}$ é SM para o modelo.

25.5.2.4.1 Resposta

Já sabemos que é uma estatística suficiente para o modelo. Com o suporte $\mathfrak{X}_\theta = (0, \theta]$ depende de "θ", não podemos usar o teorema anterior. Precisamos mostrar que, para qualquer estatística suficiente para o modelo $U(X_n)$ existe $H(\cdot)$ tal que

$$X_{(n)} = H(U(X_n)) \text{ q.c.}$$

Seja $f_\theta^{X_n}$ a função densidade probabilidade da amostra aleatória

$$\begin{aligned} f_\theta^{X_n}(y_n) &= \frac{1}{\theta^n} \prod \mathbb{1}_{(0,\theta]}(y_i) \\ &= \frac{1}{\theta^n} \mathbb{1}_{(0,\infty)}(\min\{y_1, \dots, y_n\}) \cdot \mathbb{1}_{(0,\theta]}(\max\{y_1, \dots, y_n\}) \end{aligned}$$

Note que podemos escrever

$$\max\{y_1, \dots, y_n\} = \inf\left\{\theta \in (0, \infty) : f_\theta^{X_n}(y_n) > 0\right\}$$

Logo, a estatística $X_{(n)}$ pode ser reescrita por

$$X_{(n)} = \inf\left\{\theta \in (0, \infty) : f_\theta^{X_n}(X_n) > 0\right\}$$

Seja $U(X_n)$ uma estatística suficiente qualquer. Então, pelo CF, temos que existem h' e m' tais que

$$f_\theta {y_n}^{X_n} = h'(y_n) \cdot m'(U(y_n), \theta), \text{ q.c. } \forall \theta > 0$$

Logo,

$$X_{(n)} = \inf\{\theta \in (0, \infty) : h'(X_n) \cdot m'(U(X_n), \theta) > 0\} \text{ q.c.}$$

Como $h'(y_n) > 0$ q.c. temos que

$$X_{(n)} = \inf\{\theta \in (0,\infty) : m'(U(X_n), \theta) > 0\} \text{ q.c.}$$

Portanto, $X_{(n)}$ é suficiente minimal para o modelo, pois existe $H(\cdot)$ tal que

$$X_{(n)} = H(U(X_n)) \text{ q.c.}$$

Basta tomar

$$H(u) = \inf\{\theta \in \Theta : m'(u, \theta) > 0\}$$

> **Observação**
>
> $$f_\theta(x) = \mathrm{e}^{-\theta} \cdot \mathbb{1}_{(0,\theta)}(x)$$
>
> Defina $g : \mathbb{R}^+ \to \mathbb{R}$
>
> $$\begin{aligned} g(u) &= \inf\{\theta \in (0,\infty) : f_\theta(u) > 0\} \\ &= \inf\left\{\theta \in (0,\infty) : \mathrm{e}^{-\theta} \cdot \mathbb{1}_{(0,\theta]}(u) > 0\right\} \\ \Rightarrow g(u) &= u \end{aligned}$$

25.5.2.5 Exemplo (Normal Curvada)

Seja $X_n = (X_1, \ldots, X_n)$ amostra aleatória de $X \sim \mathrm{N}(\theta, \theta^2), \theta \in \Theta = \mathbb{R}^+$. Encontre uma estatística SM para o modelo.

25.5.2.5.1 Resposta

Sabemos que $(\sum X_i, \sum X_i^2)$ é suficiente para o modelo completo

Tome $\Theta_0 = \{1, 2, 3\}$. Pelo Teorema 1, temos que:

$$T'(X_n) = \left(\frac{f_{\theta_1}^{X_n}(X_n)}{f_{\theta_0}^{X_n}(X_n)}, \frac{f_{\theta_2}^{X_n}(X_n)}{f_{\theta_0}^{X_n}(X_n)} \right)$$

em que $\theta_0 = 1, \theta_1 = 2, \theta_2 = 3$. Dessa forma,

$$\begin{aligned} f_{\theta_0}^{X_n} &= \frac{\exp\left\{-\frac{1}{2}\sum(X_i - 1)^2\right\}}{(\sqrt{2\pi})^n} = \frac{\exp\left\{-\frac{1}{2}(\sum X_i^2 - 2\sum X_i + n)\right\}}{(\sqrt{2\pi})^n} \\ f_{\theta_1}^{X_n} &= \frac{\exp\left\{-\frac{1}{2\cdot 4}(\sum X_i^2 - 2 \cdot 2 \cdot \sum X_i + 4n)\right\}}{(\sqrt{2\pi \cdot 4})^n} \\ f_{\theta_2}^{X_n} &= \frac{\exp\left\{-\frac{1}{2\cdot 9}(\sum X_i^2 - 2 \cdot 3 \cdot \sum X_i + 9n)\right\}}{(\sqrt{2\pi \cdot 9})^n}. \end{aligned}$$

Portanto,

$$T'(X_n) = \left(\frac{\frac{\exp\{-\frac{1}{2}(\sum X_i^2 - 2\sum X_i + n)\}}{(\sqrt{2\pi})^n}}{\frac{\exp\{-\frac{1}{2\cdot 4}(\sum X_i^2 - 2\cdot 2\cdot \sum X_i + 4n)\}}{(\sqrt{2\pi\cdot 4})^n}}, \frac{\frac{\exp\{-\frac{1}{2}(\sum X_i^2 - 2\sum X_i + n)\}}{(\sqrt{2\pi})^n}}{\frac{\exp\{-\frac{1}{2\cdot 9}(\sum X_i^2 - 2\cdot 3\cdot \sum X_i + 9n)\}}{(\sqrt{2\pi\cdot 9})^n}} \right)$$

$$= \left(\frac{\exp\left\{-\frac{1}{2}\left(\frac{1}{4}\sum X_i^2 - \sum X_i^2 - 2(\frac{2}{4}\sum X_i - \sum X_i)\right)\right\}}{4^{\frac{n}{2}}}, \right.$$

$$\left. \frac{\exp\left\{-\frac{1}{2}\left(\frac{1}{9}\sum X_i^2 - \sum X_i^2 - 2(\frac{3}{9}\sum X_i - \sum X_i)\right)\right\}}{9^{\frac{n}{2}}} \right)$$

$$= \left(\frac{\exp\left\{-\frac{1}{2}\left(-\frac{3}{4}\sum X_i^2 - 2(-\frac{1}{2}\sum X_i)\right)\right\}}{4^{\frac{n}{2}}}, \right.$$

$$\left. \frac{\exp\left\{-\frac{1}{2}\left(-\frac{8}{9}\sum X_i^2 - 2(\frac{6}{9}\sum X_i)\right)\right\}}{9^{\frac{n}{2}}} \right)$$

$$= \left(\frac{\exp\left\{-\frac{3}{8}\sum X_i^2 - \frac{1}{2}\sum X_i)\right\}}{4^{\frac{n}{2}}}, \frac{\exp\left\{-\frac{4}{9}\sum X_i^2 - \frac{6}{9}\sum X_i\right\}}{9^{\frac{n}{2}}} \right).$$

Observe que $T'(X_n)$ é função 1:1 de $T(X_n)$. Pois,

$$\begin{cases} T(X_n) = (t_1, t_2) \\ \Rightarrow T'(X_n) = \left(\frac{\exp\{-\frac{3}{8}t_2 - \frac{1}{2}t_1)\}}{4^{\frac{n}{2}}}, \frac{\exp\{-\frac{4}{9}t_2 - \frac{6}{9}t_1\}}{9^{\frac{n}{2}}}, \right) \end{cases}$$

$$\begin{cases} T'(X_n) = (t_1', t_2') \\ t_1' = \frac{\exp\{-\frac{3}{8}t_2 - \frac{1}{2}t_1)\}}{4^{\frac{n}{2}}} \\ t_2' = \frac{\exp\{-\frac{4}{9}t_2 - \frac{6}{9}t_1\}}{9^{\frac{n}{2}}} \end{cases}$$

$$\Rightarrow \begin{cases} \ln(4^{\frac{n}{2}}t_1') = \frac{3}{8}t_2 - \frac{1}{2}t_1 \\ \ln(9^{\frac{n}{2}}t_2') = \frac{4}{9}t_2 - \frac{6}{9}t_1 \end{cases}$$

$$\Rightarrow \begin{cases} \ln(4^{\frac{n}{2}}t_1') = \frac{3}{8}t_2 - \frac{1}{2}t_1 \\ \ln(9^{\frac{n}{2}}t_2') = \frac{4}{9}t_2 - \frac{6}{9}t_1 \end{cases}$$

$$\Rightarrow \begin{cases} t_1 = -2\ln(4^{\frac{n}{2}}t_1') + \frac{3}{4}t_2 \\ \ln(9^{\frac{n}{2}}t_2') = \frac{4}{9}t_2 - \frac{6}{9}\left(\frac{3}{4}t_2 - 2\ln(4^{\frac{n}{2}}t_1')\right) \end{cases}$$

$$\Rightarrow \begin{cases} t_1 = \frac{3}{4}t_2 - 2\ln(4^{\frac{n}{2}}t_1') \\ \ln(9^{\frac{n}{2}}t_2') - \frac{12}{9}\ln(4^{\frac{n}{2}}t_1') = \left(\frac{4}{9} - \frac{18}{36}\right)t_2 \end{cases}$$

$$\Rightarrow t_2 = \frac{\ln(9^{\frac{n}{2}}t_2') - \frac{12}{9}(4^{\frac{n}{2}}t_1')}{\left(\frac{4}{9} - \frac{18}{36}\right)}$$

$$\Rightarrow \begin{cases} t_1 = \frac{\frac{3}{4}\left(\ln(9^{\frac{n}{2}}t_2') - \frac{12}{9}\ln(4^{\frac{n}{2}}t_1')\right)}{\left(\frac{4}{9} - \frac{18}{36}\right)} - 2\ln(4^{\frac{n}{2}}t_1') \\ t_2 = \frac{ln(9^{\frac{n}{2}}t_2') - \frac{12}{9}(4^{\frac{n}{2}}t_1')}{\left(\frac{4}{9} - \frac{18}{36}\right)} \end{cases}$$

Logo $T(X_n) = (\sum X_i, \sum X_i^2)$ é SM apara o modelo reduzido. Como é suficiente para o modelo completo, temos pelo Teorema 2 que também é SM para o modelo completo.

26 Estatísticas Ancilares

Dizemos que $U(X_n)$ é uma estatística ancilar ao modelo se, e somente se, sua distribuição não depende de "θ".

26.1 Exemplo 1 (Normal $\to \chi^2$)

Seja $X_n = (X_1, \dots, X_n)$ amostra aleatória de $X \sim \mathrm{N}(\theta, 2), \theta \in \Theta = \mathbb{R}$.

$$U(X_n) = \sum^n i = 1 \frac{(X_i - \bar{X})^2}{2} \sim \chi^2_{(n-1)}$$

é, portanto, ancilar ao modelo.

26.2 Exemplo 2 (Normal $\to$ t-student)

Seja $X_n = (X_1, \dots, X_n)$ amostra aleatória de $X \sim \mathrm{N}(\mu, \sigma^2), \theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}^+$.

Note que:

$$S(X_n) = \sum \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

não é uma estatística apesar de sua *distribuição* não depender de "θ", uma vez que a estatística depende de "θ".

Construiremos duas novas estatísticas

Sejam, com $k < n$,

$$S_1(X_n) = \sum_{i=1}^{k} \frac{(X_i - \bar{X}_1)^2}{k - 1}$$
$$S_2(X_n) = \sum_{i=k+1}^{n} \frac{(X_i - \bar{X}_2)^2}{n - k - 1}$$

em que

$$\begin{cases} \bar{X}_1 & = \frac{1}{k}\sum_{i=1}^{k} X_i \\ \bar{X}_2 & = \frac{1}{n-k}\sum_{i=k+1}^{n} X_i \end{cases}$$

Logo, $S_2(X_n)$ é independente de $S_2(X_n)$. Defina

$$S(X_n) = \frac{S_1(X_n)}{S_2(X_n)} = \frac{\sum_{i=1}^{k} \frac{(X_i-\bar{X}_1)^2}{k-1}}{\sum_{i=k+1}^{n} \frac{(X_i-\bar{X}_2)^2}{n-k-1}} \sim \mathrm{F}_{k-1,n-k-1}$$

Portanto, essa estatística $S(X_n)$ é ancilar ao modelo.

Observações: (serão transferidas para uma seção própria posteriormente)

Se $(X_1, \dots, X_n)$ é amostra aleatória de $X \sim \mathrm{N}(\mu, \sigma^2)$, então

1. $\bar{X} \sim \mathrm{N}(\mu, \frac{\sigma^2}{n})$ (por função geradora de momentos);
2. $\sum \frac{(X_i-\mu)^2}{\sigma^2} \sim \chi^2_n)$ (por função geradora de momentos e transformação de VAs);
3. $\sum \frac{(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2_{n-1})$ (por função geradora de momentos, álgebra linear e transformação de VAs);

Se $(X_1, \dots, X_n)$ é amostra aleatória de $X \sim \mathrm{N}(\mu, \sigma^2)$ e $(Y_1, \dots, Y_n)$ a.a. de $Y \sim \mathrm{N}(\mu, \sigma^2)$ são independentes, então

$$1. \begin{cases} \sum_{i=1}^{n_1} \frac{(X_i-\mu_X)^2}{\sigma_X^2} \sim \chi^2_{n_1}) \\ \sum_{i=1}^{n_2} \frac{(Y_i-\mu_X)^2}{\sigma_Y^2} \sim \chi^2_{n_2}) \end{cases}$$

$$2. \begin{cases} \sum_{i=1}^{n_1} \frac{(X_i-\bar{X})^2}{\sigma_X^2} \sim \chi^2_{n_1-1}) \\ \sum_{i=1}^{n_2} \frac{(Y_i-\bar{Y})^2}{\sigma_Y^2} \sim \chi^2_{n_2-1}) \end{cases}$$

Se $W \sim \chi^2_k$ e $M \sim \chi^2_m$ são independentes, então

$$\frac{\frac{W}{k}}{\frac{Q}{m}} \sim F_{(k,m)}$$

Se $Z \sim \mathrm{N}(0,1)$ e $W \sim \chi^2_k$, então

$$\frac{Z}{\sqrt{\frac{W}{k}}} \sim t_k$$

26.3 Ancilar de Primeira Ordem

Dizemos que $U(X_n)$ é uma estatística ancilar de primeira ordem se, e somente se, $E_\theta\left(U(X_n)\right)$ (o primeiro momento) não depende de "θ".

Toda estatística ancilar é também ancilar de primeira ordem, mas a recíproca não é verdadeira. Segue disso que uma estatística que não é ancilar de primeira ordem não é ancilar.

26.3.1 Exemplo 1 (Normal)

Seja X_n amostra aleatória de $X \sim \mathrm{N}(\mu, \sigma^2)$ e $U(X_n) = X_1 - \bar{X}$

$$E_\theta(U(X_n)) = E_\theta(X_1) - E_\theta(\bar{X}) = \mu - \mu = 0, \forall\theta \in \Theta$$

$$\begin{aligned}
\mathrm{Var}_\theta(U(X_n)) &= \mathrm{Var}_\theta(X_i - \bar{X}) \\
X_1 - \bar{X} &= X_1 - \frac{1}{n}X_1 - \cdots - \frac{1}{n}X_n \\
&= (1 - \frac{1}{n})X_1 - \frac{1}{n}X_2 - \cdots - \frac{1}{n}X_n \\
\Rightarrow \mathrm{Var}_\theta(U(X_n)) &= \left(1 - \frac{1}{n}\right)^2 \sigma^2 + \underbrace{\frac{1}{n^2}\sigma^2 + \cdots + \frac{1}{n^2}\sigma^2}_{n-1} \\
\Rightarrow \mathrm{Var}_\theta(U(X_n)) &= \left(1 - \frac{1}{n}\right)^2 \sigma^2 + \frac{n-1}{n^2}\sigma^2 \forall\theta \in \Theta
\end{aligned}$$

depende de "θ".

27 Estatísticas Completas

Dizemos que $T(X_n)$ é uma estatística completa se, e somente se, existe a função real $h(\cdot)$ tal que

$$E_\theta(h(T(X_n)) = 0, \forall \theta \in \Theta$$

é a função nula quase certamente, isto é,

$$P_\theta(h(T(X_n)) = 0) = 1, \forall \theta \in \Theta \text{ ou simplesmente } h(T(X_n)) = 0\text{q.c.}$$

27.1 Exemplo 1 (Bernoulli, Soma)

Seja $X_n = (X_1, \dots, X_n)$ amostra aleatória de $X \sim \mathrm{Ber}(\theta)$. Mostre que $T(X_n) = \sum X_i$ é uma estatistica completa para o modelo.

27.1.1 Resposta

Note que $T(X_n) \sim \mathrm{Binom}(n, \theta), \theta \in (0, 1)$

$$P_\theta(T(X_n) = k) \begin{cases} \binom{n}{k}\theta^k(1-\theta)^{n-k}, & k \in \{0, 1, \dots, n\} \\ 0, & \text{c.c.} \end{cases}$$

Seja $h(\cdot)$ uma função real (contradomínio em $\mathbb{R}$) tal que

$$
\begin{aligned}
& E_\theta(h(T(X_n))) = 0,\ \forall \theta \in \Theta = (0,1) \\
\overset{\text{a.a. } X\sim\text{Ber}}{\Longleftrightarrow} & \sum_{k=0}^{n} h(k) P_\theta(T(X_n) = k) = 0 \\
\Longleftrightarrow & \sum h(k) \binom{n}{k} \theta^k (1-\theta)^{n-k} = 0 \\
\Longleftrightarrow & \sum h(k) \binom{n}{k} \left(\frac{\theta}{(1-\theta)}\right)^k (1-\theta)^n = 0 \\
\Longleftrightarrow & \sum h(k) \binom{n}{k} \left(\frac{\theta}{(1-\theta)}\right)^k (1-\theta)^n = 0 \\
\Longleftrightarrow & \sum h(k) \binom{n}{k} \left(\frac{\theta}{(1-\theta)}\right)^k = 0 \\
\overset{\rho=\frac{\theta}{1-\theta}\in(0,\infty)}{\Longleftrightarrow} & \sum h(k) \binom{n}{k} \rho^k = 0, \forall \theta \in \Theta = (0,1), \rho \in (0,\infty).
\end{aligned}
$$

Observe que, tomando $a_k = h(k)\binom{n}{k}, k = 0, 1, \dots, n$, temos que

$$
a_0\rho^0 + a_1\rho^1 + \cdots + a_n\rho^n = 0,\ \forall \rho \in (0,\infty).
$$

Por igualdade de polinômios,

$$
\begin{cases} a_0 = 0 \\ \vdots \\ a_n = 0 \end{cases} \Rightarrow \begin{cases} h(0)\binom{n}{0} = 0 \\ \vdots \\ h(n)\binom{n}{n} = 0 \end{cases} \Rightarrow
$$

Como $\binom{n}{k} > 0\ \forall k \in \{0, 1, \dots, n\}$, temos que $h(k) = 0,\ \forall k \in \{0, 1, \dots, n\}$

$$
\Rightarrow P_\theta(h(T(X_n)) = 0) = 1,\ \forall \theta \in \Theta.
$$

Portanto, $T(X_n)$ é completa para o modelo.

27.2 Exemplo 2 (Uniforme)

Seja $X_n = (X_1, \dots, X_n)$ a.a. de $X \sim \text{Unif}(0,\theta), \theta \in \Theta = (0,\infty)$. Mostre que $T(X_n) = \max\{X_1, \dots, X_n\}$ é uma estatística completa para o modelo.

27.2.1 Resposta

Seja $h(\cdot)$ uma função tal que

$$E_\theta(h(X_n)) = 0,\ \forall\theta \in \Theta = (0, \infty)$$

Precisamos encontrar a função densidade de probabilidade de $X_(n)$ (máximo). Sabemos que a f.d.p é a derivada da função de distribuição acumulada. Isto é,

$$\begin{aligned}
P_\theta(X_{(n)} \le t) &= P_\theta(X_1 \le t, \dots, X_n \le t)\\
&\overset{\text{ind.}}{=} \prod_{i=1}^{n} P_\theta(X_i \le t) \overset{\text{id}}{=} \prod P_\theta(X \le t)\\
\Rightarrow P_\theta(X_{(n)} \le t) &\overset{\text{iid}}{=} [P_\theta(X \le t)]^n
\end{aligned}$$

em que

$$P_\theta(X \le t) = \begin{cases} 0, & t \le 0\\ \frac{t}{\theta}, & t \in (0, \theta]\\ 1, & t > \theta \end{cases}$$

Portanto,

$$\begin{aligned}
\frac{d}{dt}P_\theta(X_{(n)} \le t) &= n[P_\theta(X_{(n)} \le t)]^{n-1}\frac{d}{dt}P_\theta(X_{(n)} \le t)\\
\Rightarrow f_\theta^{X_{(n)}}(x) &= \left.\frac{dP_\theta(X_{(n)} \le t)}{dt}\right|_{t=x}\\
&= n[P_\theta(X_{(n)} \le x)]^{n-1} f_\theta(x)\\
&= \begin{cases} n\left(\frac{x}{\theta}\right)^{n-1}\frac{1}{\theta}, & x \in (0, \theta]\\ 0, & \text{c.c.} \end{cases}
\end{aligned}$$

Dessa forma,

$$\begin{aligned}
E_\theta(h(X_{(n)})) &= \int_{-\infty}^{\infty} h(x) f_\theta^{X_{(n)}}(x)dx,\\
&= \int_0^\theta h(x)\cdot n\left(\frac{x}{\theta}\right)^{n-1}\frac{1}{\theta}dx,\ \forall\theta \in \Theta.
\end{aligned}$$

Se

$$E_\theta(h(X_{(n)})) = 0,\ \forall\theta \in \Theta$$

então

$$
\begin{aligned}
&\int_0^\theta h(x)\cdot n\left(\frac{x}{\theta}\right)^{n-1}\frac{1}{\theta}dx = 0, \forall\theta > 0\\
\iff &\frac{n}{\theta^n}\int_0^\theta h(x)x^{n-1}dx = 0,\ \forall\theta > 0\\
\iff &\int_0^\theta h(x)x^{n-1}dx = 0,\ \forall\theta > 0.
\end{aligned}
$$

Pelo Teorema Fundamental do Cálculo: $\frac{d}{dt}\int_a^t f(x)dx = f(t)$, temos que

$$
\begin{aligned}
&\frac{d}{d\theta}\int_0^\theta h(x)x^{n-1}dx = 0\\
&\overset{\text{TFC}}{\Rightarrow} h(\theta)\theta^{n-1} = 0\\
&\iff h(\theta) = 0, \forall\theta > 0.
\end{aligned}
$$

Ou seja, $h(x) = 0, \forall x > 0$. Portanto, h é a função nula *quase certamente*. Logo, $X_{(n)}$ é completa para o modelo.

27.3 Exemplo negativo (Bernoulli, amostra)

Seja X_n amostra aleatória de $X \sim \text{Ber}(\theta), \theta \in \Theta = (0, 1)$. Mostre que $T(X_n) = X_n$ é suficiente porém não é completa para o modelo.

27.3.1 Resposta

Pelo critério da fatoração, já sabemos que X_n é suficiente para o modelo. Tome

$$E_\theta(h(T(X_n)) = E_\theta(X_1 - X_2) = \theta - \theta = 0, \forall\theta \in \Theta.$$

Portanto, $T(X_n)$ não é completa para o modelo.

27.4 Teorema 1 (da relação de ida com estatísticas suficientes minimais)

Toda estatística suficiente e completa é também suficiente minimal para o modelo estatístico.

27.5 Teorema 2 (da relação de volta com estatísticas suficientes minimais)

Se existir uma estatística suficiente e completa, então toda estatística suficiente minimal será também completa.

27.5.1 Exemplo negativo (Normal curvada)

Seja X_n a.a. de $X \sim \mathrm{N}(\theta, \theta^2), \theta \in \Theta = (0, \infty)$. Mostre que $T(X_n) = (\sum X_i, \sum X_i^2)$ não é completa para o modelo.

27.5.1.1 Resposta

Note que

$$\begin{aligned} E_\theta(\sum X_i) &= n\theta, \forall \theta \in \Theta \\ E_\theta(\sum X_i^2) &\overset{\text{iid}}{=} nE_\theta(X^2) = n(\theta^2 + \theta^2) = 2n\theta^2 \\ \Rightarrow E_\theta(T(X)n) = (n\theta, 2n\theta^2) \end{aligned}$$

Tentativas para encontrar $h(\cdot)$.

$$\begin{aligned} E_\theta\left(\left(\sum X_i\right)^2\right) &= E_\theta\left(\sum_{i=1}^n \sum_{j=1}^n X_i X_j\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n E_\theta(X_i \cdot X_j) \\ &= \sum_{i=1}^n E(X_i^2) + \sum\sum_{i \neq j} E_\theta(X_i X_j) \\ &\overset{\text{ind}}{=} 2n\theta^2 + \sum\sum_{i \neq j} E_\theta(X_i) E_\theta(X_j) \\ &\overset{\text{id}}{=} 2n\theta^2 + \sum\sum_{i \neq j} [E_\theta(X)]^2 \\ &= 2n\theta^2 + (n^2 - n)\theta^2 \\ &= (2n + n^2 - n)\theta^2 \\ &= (n^2 + n)\theta^2. \end{aligned}$$

Ou seja,

$$E_\theta\left(\frac{(\sum X_i)^2}{n^2+n}\right)=\theta^2, \forall\theta\in\Theta.$$

Note ainda que

$$E_\theta\left(\frac{\sum X_i^2}{2n}\right)=\theta^2, \forall\theta\in\Theta.$$

Logo, tome

$$h(T(X_n))=\frac{T_1(X_n)^2}{n^2+n}-\frac{T_2(X_n)}{2n}$$

em que

$$T_1(X_n)=\sum X_i$$
$$T_2(X_n)=\sum X_i^2.$$

Como $E_\theta(h(T(X_n))=0, \forall\theta>0$, concluímos que $T(X_n)$ não é conpleta.

> Nem toda SM é completa!
>
> Já demonstramos que $T(X_n)$ é SM. Portanto, nem toda estatística SM é também completa. Dessa forma, pelo Teorema 2, concluímos que não existe uma estatística completa para o modelo.

27.6 Teorema 3 (da relação com a família exponencial)

Seja X_n a.a. de $X\sim f_\theta, \theta\in\Theta$, em que f_θ pertence à família exponencial k-dimensional. Se "couber" um "retângulo" k-dimensional no conjunto

$$\mathcal{C}=\{(c_1(\theta),\dots,c_k(\theta):\theta\in\Theta\},$$

em que c_i representa a i-ésima componente c obtida do pertencimento à família exponencial. Então a estatística

$$T(X_n)=\left(\sum T_1(X_n),\dots,\sum T_k(X_n)\right),$$

em que T_i representa a i-ésima componente T obtida do pertencimento à FE, é suficiente *e completa* para o modelo.

Se $k=1$, então $\mathcal{C}$ deve conter um intervalo (aberto). Se $k=2$, $\mathcal{C}$ deve conter o interior de um retângulo. Se $k=3$, então deve conter o interior de um cubo, e assim por diante.

27.6.1 Exemplo (Normal livre)

Seja X_n a.a. de $X \sim \mathrm{N}(\mu, \sigma^2), \theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}^+$. Mostre que $T(X_n) = (\sum X_i, \sum X_i^2)$ é suficiente e completa para o modelo.

27.6.1.1 Resposta

Note que

$$f_\theta(x) = \mathrm{e}^{c_1(\theta)T_1(x)+c_2(\theta)T_2(x)+d(\theta)+S(x)}$$

em que

$$\begin{array}{rr} c_1(\theta) = -\frac{1}{2\sigma^2} & T_1(x) = X^2 \\ c_2(\theta) = \frac{\mu}{\sigma^2} & T_2(x) = x \\ d(\theta) = -\frac{\mu^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2) & S(x) = 0. \end{array}$$

Observe também que

$$\mathcal{C} = \{(c_1(\theta), c_2(\theta)) : \theta \in \Theta\} = \left\{\left(-\frac{1}{2\sigma^2}, \frac{\mu}{\sigma^2}\right)\right\}$$

Ou seja

$$\mathcal{C} = \mathbb{R}^- \times \mathbb{R}$$

Tome, por exemplo,

$$(-2, -1) \times (1, 2) \subseteq \mathcal{C}$$

Logo, concluímos que

$$T'(X_n) = (\sum X_i^2, \sum X_i)$$

é suficiente e completa para o modelo. Como $T(X_n) = (\sum X_i, \sum X_i^2)$ é função injetora (1:1) de $T'(X_n)$, concluímos que $T(X_n)$ também é suficiente e completa.

27.6.2 Exemplo (Exponencial)

Seja X_n a.a. de $X \sim \mathrm{Exp}(\theta), \theta \in \Theta = (0, \infty)$. Encontre uma estatística suficiente e completa.

27.6.2.1 Resposta

Note que

$$f_\theta(x) = \begin{array}{ll} \theta e^{-\theta x}, & x > 0 \\ 0, & \text{c.c.} \end{array}$$

Já mostramos que pertence à FE com $c_1(\theta) = -\theta$, e $T_1(x) = x$. Note que

$$\mathcal{C} = \{c_1(\theta) : \theta \in \Theta\} = (-\infty, 0)$$

Como $(-1, -0.5) \subseteq \mathcal{C}$, concluímos que $\sum T_1(X_i)$ é, pelo Teorema 3, suficiente e completa.

27.6.3 Exemplo negativo (Normal curvada)

Seja X_n a.a. de $X \sim \mathrm{N}(\theta, \theta^2), \theta \in \Theta = \mathbb{R}^+$. Verifique se o conjunto $\mathcal{C} = \{(c_1(\theta), c_2(\theta)) : \theta \in \mathbb{R}^+$.

27.6.3.1 Resposta

Note que

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\theta^2}} \exp\left\{-\frac{1}{2\theta^2}(x^2 - 2x\theta + \theta^2)\right\}$$

Tomando

$$\begin{array}{rl} c_1(\theta) = -\frac{1}{2\theta^2} & T_1(x) = x^2 \\ c_2(\theta) = \frac{1}{\theta} & T_2(x) = x \\ d(\theta) = -\ln(\sqrt{2\pi\theta^2}) & S(x) = -\frac{1}{2} \end{array}$$

Logo

$$\mathcal{C} = \{(c_1(\theta), c_2(\theta)) : \theta > 0\} = \left\{\left(-\frac{1}{2\theta^2}, \frac{1}{\theta}\right)\right\}$$

Como $c_1(\theta)$ e $c_2(\theta)$ estão relacionados funcionalmente (podemos escrever $c_1(\theta)$ como função de $c_2(\theta)$, temos uma curva ao invés de uma área e, portanto, não é possível encontrar um retângulo cujo interior esteja contido em $\mathcal{C}$.

Note que esse teorema não diz se não há uma estatística completa. Para isso, é necessário testar as estatísticas pela definição

28 Estimadores não-viciados com variância uniformemente mínima (ENVVUM)

Estimadores são estatísticas cujo objetivo é estimar uma quantidade de interesse $g(\theta)$.

Dizemos que $T(X_n)$ é um ENVVUM se, e somente se,

1. $T(X_n)$ é não-viciado para $g(\theta)$:

$$E_\theta(T(X_n)) = g(\theta), \forall \theta \in \Theta;$$

2. Para qualquer outro estimador não-viciado para $g(\theta)$, $U(X_n)$, a variância de $T(X_n)$ não é maior do que a variância de $U(X_n)$, ou seja,

$$\text{Var}_\theta(T(X_n)) \leq \text{Var}_\theta(U(X_n)), \forall \theta \in \Theta;$$

28.1 Teorema de Lehmamn-Scheffé

Seja $X_n = (X_1, \ldots, X_j)$ amostra aleatória de $X \sim f_\theta, \theta \in \Theta$, f_θ a distribuição amostral. Considere $T(X_n)$ uma estatística suficiente e completa para o modelo e $S(X_n)$ um estimador não-viciado para $g(\theta)$. Então,

$$\widetilde{T}(X_n) = E_\theta(S(X_n)|T(X_n))$$

é o ENVVUM para $g(\theta)$, quase certamente, baseado em $T(X_n)$.

Observação

1. $\widetilde{T}(X_n)$ não depende de "θ" pois $T(X_n)$ é suficiente para o modelo.
2. $E_\theta(\widetilde{T}(X_n)) = E_\theta(E_\theta(S(X_n)|T(X_n)) = E_\theta(S(X_n))$

28.1.1 Prova

Observe que

$$E_\theta(\widetilde{T}(X_n)) = E_\theta(E_\theta(S(X_n)|T(X_n)) = E_\theta(S(X_n)).$$

Como $S(X_n)$ é, por hipótese, não-viciado para $g(\theta)$, temos que

$$E_\theta(\widetilde{T}(X_n)) = g(\theta), \forall\theta \in \Theta.$$

Portanto, $\widetilde{T}(X_n)$ é não-viciado para $g(\theta)$.

Seja $U(X_n)$ um estimador não-viciado para $g(\theta)$ que dependa apenas de $T(X_n)$. Então, como $T(X_n)$ é, por hipótese, completa para o modelo, tome, como $\widetilde{T}(X_n)$ é uma função de $T(X_n)$,

$$\begin{aligned} h(T(X_n)) &= \widetilde{T}(X_n) - U(X_n) \\ \Rightarrow E_\theta(h(T(X_n))) &= 0, \forall\theta \in \Theta \\ \Rightarrow h(T(X_n)) &= 0, \text{q.c.} \end{aligned}$$

Portanto, $\widetilde{T}(X_n) = U(X_n), \text{q.c.}$

Seja $\widetilde{T}_1(X_n)$ um estimador não-viciado para $g(\theta)$. Então,

$$\begin{aligned} \text{Var}_\theta(\widetilde{T}_1(X_n)) &= E_\theta(\text{Var}_\theta(\widetilde{T}_1(X_n)|T(X_n)) + \text{Var}_\theta(\underbrace{E_\theta(\widetilde{T}_1(X_n)|T(X_n))}_{=\widetilde{T}(X_n),\text{q.c.}}) \\ &= \underbrace{E_\theta(\text{Var}_\theta(\widetilde{T}_1(X_n)|T(X_n))}_{\geq 0} + \text{Var}_\theta(\widetilde{T}(X_n)), \forall\theta \in \Theta \\ \Rightarrow \text{Var}_\theta(\widetilde{T}_1(X_n)) &\geq \text{Var}_\theta(\widetilde{T}(X_n)), \text{q.c. } \forall\theta \in \Theta \end{aligned}$$

Ou seja, $\widetilde{T}(X_n)$ é quase certamente *o* ENVVUM para $g(\theta)$ baseado em $T(X_n)$

28.2 Exemplo Bernoulli

Seja X_n amostra aleatória de $X \sim \text{Ber}(\theta), \theta \in \Theta = (0, 1)$.

a-) Encontre o ENVVUM para $g(\theta) = P_\theta(X = 1)$.

Já sabemos que $\sum X_i$ é suficiente e completa para o modelo de Bernoulli. Além disso,

$$E_\theta \left(\frac{1}{n} \sum X_i \right) = \theta, \forall \theta \in \Theta.$$

Ou seja, $T(X_n) = \frac{1}{n} \sum X_i$ é suficiente e completa, pois é função 1:1 de $\sum X_i$ e, além disso, é não viciada para $g(\theta) = P_\theta(X = 1) = \theta$. Portanto,

$$E_\theta(T(X_n)|T(X_n)) = T(X_n).$$

Logo, o ENVVUM é, pelo Teorema de Lehmamn-Scheffé,

$$\tilde{T}(X_n) = E_\theta(T(X_n)|T(X_n)) = T(X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Obs: Poderíamos iniciar com $S(X_n) = X_1$ pois $E_\theta(X_1) = \theta, \forall \theta \in \Theta$.

$$\tilde{T}(X_n) = E_\theta \left(S(X_n) | \sum X_i \right)$$

também é ENVVUM. Para demonstrar, precisamos calcular

$$E_\theta \left(X_1 | \sum X_i \right).$$

Uma estratégia é encontrar

$$E_\theta(S(X_n)| \sum X_i = t) = m(t), \forall \theta \in \Theta$$

e substituir t pela estatística suficiente e completa, neste caso $\sum X_i$. Note que, para $t = 0$

$$E_\theta(X_1 | \sum X_i = 0) = 0, \forall \theta \in \Theta.$$

Se $t \neq 0$, então

$$\begin{aligned}
\frac{t}{t}=1 &\iff E_\theta\left(\frac{t}{t}|\sum X_i=t\right)=1,\forall\theta\in\Theta\\
&\iff E_\theta(\frac{t}{t}|\sum X_i=t)=E_\theta\left(\frac{\sum X_i}{\sum X_i}|\sum X_i=t\right)=1,\forall\theta\in\Theta\\
&\iff \sum E_\theta\left(\frac{X_i}{\sum X_i}|\sum X_i=t\right)=1,\forall\theta\in\Theta\\
&\overset{\text{iid}}{\iff} nE_\theta\left(\frac{X_1}{\sum X_i}|\sum X_i=t\right)=1,\forall\theta\in\Theta\\
&\iff \frac{n}{t}E_\theta\left(X_1|\sum X_i=t\right)=1,\forall\theta\in\Theta\\
&\Rightarrow E_\theta\left(X_1|\sum X_i=t\right)=\frac{t}{n},\forall\theta\in\Theta.
\end{aligned}$$

Logo, substituindo t por $\sum X_i$, temos que

$$E_\theta\left(S(X_n)|\sum X_i\right)=\frac{1}{n}\sum X_i$$

b-) Encontre o ENVVUM para $g(\theta)=P_\theta(X=0)$.

Note que $g(\theta)=P_\theta(X=0)=1-\theta$. Como $\frac{1}{n}\sum X_i$ é suficiente e completa para o modelo e $1-\frac{1}{n}\sum X_i$ é não viciada para $g(\theta)$, temos que

$$\widetilde{T}(X_n)=E_\theta\left(1-\frac{1}{n}\sum X_i \mid \sum X_i\right)=1-\frac{1}{n}\sum X_i$$

Logo, pelo Teorema de Lehmamn-Scheffé, é o ENVVUM para $g(\theta)=1-\theta$.

28.3 Exemplo (Normal)

Seja X_n a.a. de $X\sim \mathrm{N}(\mu,\sigma^2),\theta=(\mu,\sigma^2)\in\Theta=\mathbb{R}\times\mathbb{R}_+$.

28.3.1 Item a (parâmetros)

Encontre o (quase certamente) ENVVUM para $g(\theta)=(\mu,\sigma^2)$.

Já sabemos que, pelo Teorema das FEs,

$$T(X_n)=\left(\sum X_i,\sum X_i^2\right)$$

é suficiente e completa para o modelo. Observe ainda que

$$T'(X_n) = \left(\frac{1}{n} \sum X_i, \frac{1}{n} \sum X_i^2 - \left(\frac{1}{n} \sum X_i \right)^2 \right)$$

é função 1:1 de $T(X_n)$. Portanto, $T'(X_n)$ é uma estatística suficiente e completa para o modelo.

Note que

$$E_\theta \left(\frac{1}{n} \sum X_i \right) = \mu, \forall \theta \in \Theta$$

e

$$E_\theta \left(\frac{1}{\sigma^2} \sum (X_i - \bar{X})^2 \right) \overset{\chi^2_{n-1}}{=} n-1$$
$$\Rightarrow E_\theta \left(\frac{1}{n} \sum (X_i - \bar{X})^2 \right) = \frac{n-1}{n} \sigma^2, \forall \theta \in \Theta.$$

Portanto,

$$T''(X_n) = \left(\bar{X}, S^2_{n-1} \right),$$

em que $\bar{X} = \frac{1}{n} \sum X_i$ e $S^2_{n-1}(X_n) = \frac{1}{n-1} \sum (X_i - \bar{X})^2$, é suficiente completa para o modelo não viciado para $g(\theta)$ pois

$$E_\theta(T''(X_n)) = (\mu, \sigma^2), \forall \theta \in \Theta.$$

Logo, $T''(X_n)$ é o ENVVUM quase certamente para $g(\theta) = \theta$.

28.3.2 Item b (DP)

Encontre o ENVVUM para $g(\theta) = \sigma$

Já temos, do item a, uma estatística suficiente completa para o modelo, como $T''(X_n)$.

Tentativa 1.

Intuitivamente, podemos esperar que $E_\theta(\sqrt{S^2_{n-1}(X_n)}) = c\sigma + b$. Se for, então bastaria tomar

$$S'(X_n) = \frac{\sqrt{S^2_{n-1}(X_n)} - b}{c}$$

para obtermos um estimador não viciado. Sob normalidade, $\bar{X}$ e $S^2_{n-1}(X_n)$ são independentes. Logo, $\bar{X}$ e S'_{n-1} também são independentes. Portanto,

$$E_\theta(S'(X_n)|T''(X_n)) \overset{\text{Ind.}}{=} E_\theta(S'(X_n)|S^2_{n-1}(X_n)) \overset{\text{Cond.}}{=} S'(X_n).$$

Verificando:

Observe que

$$W(\sigma^2) = \frac{(n-1)S^2_{n-1}(X_n)}{\sigma^2} \sim \chi^2_{n-1}.$$

Logo, para $n > 1$

$$\begin{aligned}
E_\theta(\sqrt{S^2_{n-1}(X_n)}) &= E_\theta\left(\frac{\sqrt{n-1}}{\sqrt{n-1}}\frac{\sqrt{\sigma^2}}{\sqrt{\sigma^2}}\sqrt{S^2_{n-1}(X_n)}\right)\\
&= E_\theta\left(\frac{\sqrt{\sigma^2}}{\sqrt{n-1}}\sqrt{\frac{(n-1)S^2_{n-1}(X_n)}{\sigma^2}}\right)\\
&= \frac{\sigma}{\sqrt{n-1}}E_\theta\left(\sqrt{W(\sigma^2)}\right).
\end{aligned}$$

Vamos calcular

$$E_\theta\left(\sqrt{W(\sigma^2)}\right) = \int_0^\infty \sqrt{x} f_\theta^{W(\sigma^2)}(x)dx$$

em que

$$f_\theta^{W(\sigma^2)}(x) = \frac{1}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})}x^{\frac{n-1}{2}-1}\mathrm{e}^{-\frac{1}{2}x}$$

Portanto,

$$\begin{aligned}
E_\theta\left(\sqrt{W(\sigma^2)}\right) &= \int_0^\infty \frac{x^{\frac{1}{2}}x^{\frac{n-1}{2}}\mathrm{e}^{-\frac{1}{2}x}}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}dx\\
&= \frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}\int_0^\infty x^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}dx\\
&\overset{x=2u}{=} \frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}\int_0^\infty (2u)^{\frac{n}{2}-1}\mathrm{e}^{-u}2du\\
&= \frac{2^{\frac{n}{2}}}{2^{\frac{n}{2}-\frac{1}{2}}\Gamma\left(\frac{n-1}{2}\right)}\int_0^\infty u^{\frac{n}{2}-1}\mathrm{e}^{-u}du\\
&= \frac{\sqrt{2}}{\Gamma\left(\frac{n-1}{2}\right)}
\end{aligned}$$

Dessa forma,

$$E_\theta\left(\sqrt{S^2_{n-1}(X_n)}\right) = \frac{\sigma}{\sqrt{n-1}}\frac{\sqrt{2}}{\Gamma\left(\frac{n-1}{2}\right)}\Gamma\left(\frac{n}{2}\right),$$

tomando

$$S'(X_n) = \frac{\sqrt{(n-1)S^2_{n-1}(X_n)}}{\sqrt{2}}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)},$$

é, como $S'(X_n) = c\sqrt{S^2_{n-1}(X_n)}$ é função 1:1 de $S^2_{n-1}(X_n)$, o ENVVUM para $g(\theta) = \sigma$.

28.4 Exemplo (Poisson)

Seja X_n a.a. de $X \sim \text{Poiss}(\theta), \theta \in (0, \infty)$.

Encontre o ENVVUM para $g(\theta) = P_\theta(X = 0)$.

Pelo Teorema da FE k-dimensional, temos que

$$T(X_n) = \sum X_i$$

é suficiente e completa para o modelo.

Note que

$$g(\theta) = P_\theta(X = 0) = \mathrm{e}^{-}$$

Um estimador não-viesado para essa quantidade de interesse é $\mathbb{1}_{\{0\}}(X_1)$, uma vez que

$$P_\theta(\mathbb{1}_{\{0\}}(X_1) = P_\theta(X_1 = 0) \overset{\text{a.a.}}{=} P_\theta(X = 0) = \mathrm{e}^{-\theta}$$

Portanto, pelo Teorema de Lehmamn-Scheffé,

$$\tilde{T} = E_\theta(S(X_n)|T(X_n))$$

é o ENVVUM para $g(\theta) = \mathrm{e}^{-\theta}$

uma estratégia é calcular $m(t) = E_\theta(S(X_n)|T(X_n) = t)$ e depois substituir t por $T(X_n)$ em $m(t)$. Note que, por definição,

$$\begin{aligned} E_\theta(S(X_n)|T(X_n) = t) &= \sum_{s \in \{0,1\}} s \cdot P_\theta(S(X_n) = s|T(X_n) = t) \\ &= P_\theta(S(X_n) = 1|T(X_n) = t) \\ &= \frac{P_\theta(S(X_n) = 1, T(X_n) = t)}{P_\theta(T(X_n) = t)} \end{aligned}$$

Note que, por FGM,

$$T(X_n) = \sum X_i \sim \text{Poiss}(n\theta), \theta \in (0, \infty)$$

e para $t \in \{0, 1, ...\}$

$$\text{"}S(X_n) = 1\text{"} \iff \text{"}X_1 = 0\text{"} \Rightarrow E_\theta(S(X_n)|T(X_n) = t) = \frac{P_\theta(X_1 = 0|\sum X_i = t)}{e^{-n\theta}\frac{(n\theta)^t}{t!}}$$

Note que

$$\begin{aligned} P_\theta(X_1 = 0|\sum X_i = t) &= P_\theta(\sum X_i = t|X_1 = 0) \cdot P_\theta(X_1 = 0) \\ &= P_\theta(X_1 + X_2 + \cdots + X_n = t|X_1 = 0) \cdot \mathrm{e}^{-\theta} \\ &= P_\theta(0 + X_2 + \cdots + X_n = t|X_1 = 0) \cdot \mathrm{e}^{-\theta} \\ &\overset{\text{iid}}{=} P_\theta\left(\sum_{i=2}^{n} X_i = t - 0\right) \cdot \mathrm{e}^{-\theta} \end{aligned}$$

Como, por FGM,

$$\sum_{i=2}^{n} X_i \sim \text{Poiss}((n-1)\theta), \theta \in (0, \infty)$$

temos que

$$P_\theta(X_1 = 0, \sum_{i=1}^{n} X_i = t) = \frac{\mathrm{e}^{-(n-1)\theta}((n-1)\theta)^t}{t!}\mathrm{e}^{-\theta}$$

Logo,

$$E_\theta(S(X_n)|T(X_n) = t) = \frac{(n-1)^t}{n^t} = \left(1 - \frac{1}{n}\right)^t$$

Portanto, substituindo t por $T(X_n)$, temos que

$$E_\theta(S(X_n)|T(X_n)) = \left(1 - \frac{1}{n}\right)^{T(X_n)}$$

28.5 Exemplo “patológico”

Sejam X_n a.a. de $X \sim \text{Pois}(\theta), \theta \in \Theta = \mathbb{R}_+$ e $g(\theta) = \mathrm{e}^{-2\theta}$.

Pode-se mostrar que

$$T(X_n) = \left(1 - \frac{2}{n}\right)^{\sum X_i}$$

é o ENVVUM para $g(\theta) = \mathrm{e}^{-2\theta}$. Para $n = 1$, temos que

$$T(X_n) = (-1)^{X_1}$$

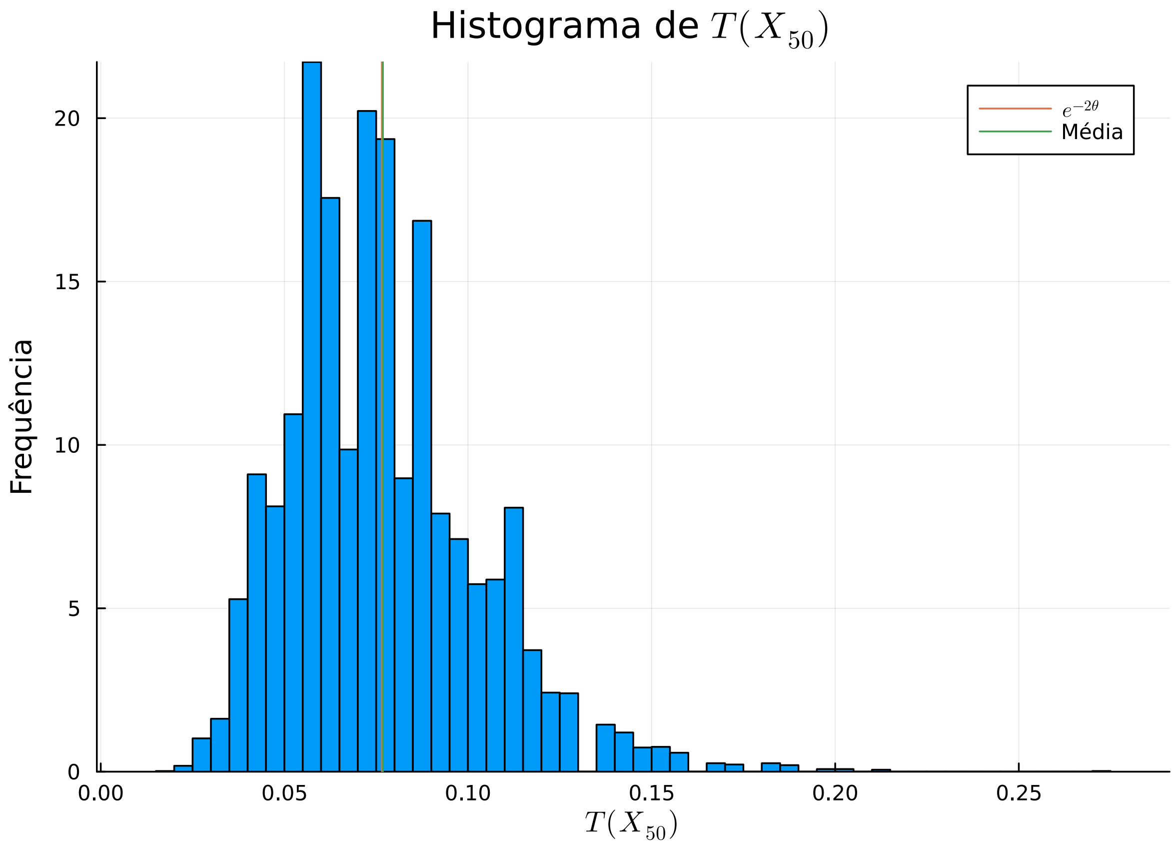
A quantidade de interesse está estritamente entre $(0, 1)$, mas o estimador pode apenas assumir valores -1 e 1.

Para $n = 2$, temos que

$$T(X_n) = 0$$

Ainda assim, em média, (como podemos verificar por simulação de Monte Carlo), ele estimará $g(\theta)$ sem viés com variância uniformemente mínima. Disso, temos que só deveríamos usar o ENVVUM para $n \geq 3$.

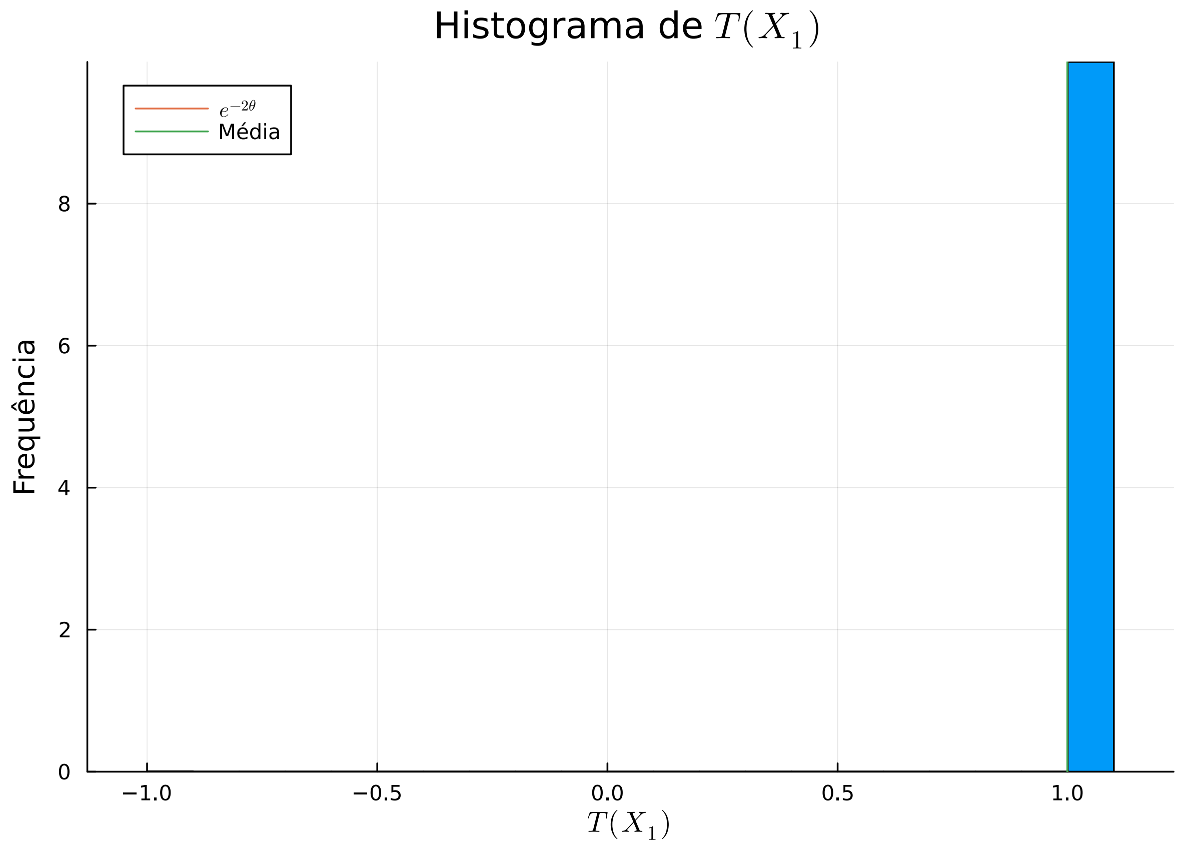
```julia
using Random, Distributions, Plots, LaTeXStrings
Random.seed!(15)
function simulacaoENVVUM(M=10000, n=50)
  # Seja X_n a.a. de X~Pois(theta), theta em (0, infinito). Vamos avaliar o
  # ENVVUM para g(theta) = e^{-2theta}, T(X_n) = (1-2/n)^{soma de X_i}
  theta = log(rand())^2 # fixo: número real (não uniformemente) aleatório
  d = Poisson(theta)
  estatisticas = []
  for i in 1:M
    amostra = rand(d, n)
    envvum = (1-2/n)^(sum(amostra))
    append!(estatisticas, envvum)
  end
  media = mean(estatisticas)
  e2theta = MathConstants.e^(-2*theta)
  h = histogram(estatisticas,
              title = LaTeXString("Histograma de \$T(X_{$n})\$"),
              normalize=:pdf,
              label="",
              xlabel=LaTeXString("\$T(X_{$n})\$"),
              ylabel="Frequência")
  vline!([e2theta],label=L"e^{-2\theta}")
  vline!([media],label="Média")
  display(h)
  display("Média: $media")
  display("e^{-2 * theta} = $e2theta")
end
simulacaoENVVUM()
```



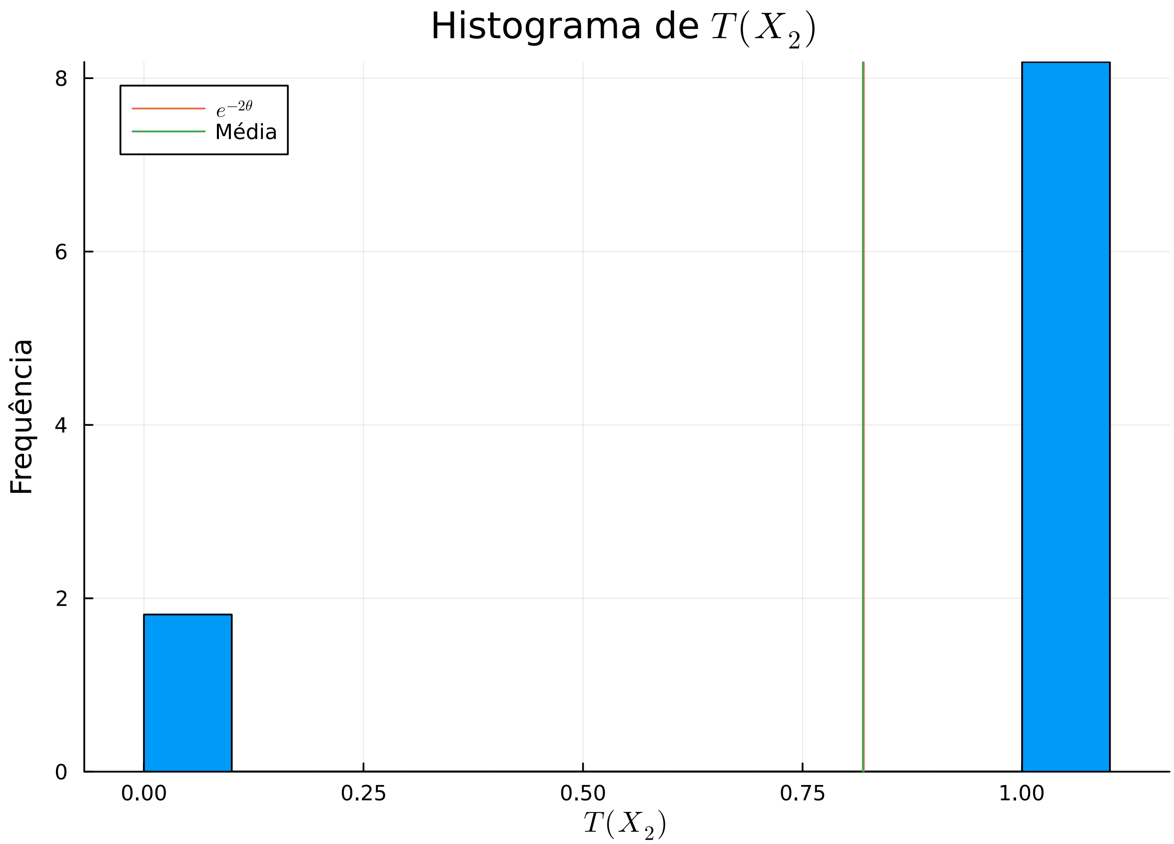

```
"Média: 0.07683263869408652"
```

```
"e^{-2 * theta} = 0.0765628110644349"
```

Podemos ainda avaliar os nossos casos patológicos.




```
"Média: 0.999"

"e^{-2 * theta} = 0.9993081423710376"
```




```
"Média: 0.8187"

"e^{-2 * theta} = 0.81973744188805"
```

29 Função Escore

Seja X_n amostra aleatória de $X \sim f_\theta, \theta \in \Theta$. Considere $x = (x_1, \dots, x_n)$ uma amostra observada. A função de verossimilhança $L_{x_n} : \Theta \to \mathbb{R}$ é definida por

$$L_{x_n}(\theta) = f_\theta^{X_n}(x_n)$$

em que $f_\theta^{X_n}$ é a função (densidade) de probabilidade da amostra. Assuma que f_θ é diferenciável com respeito a "θ", $\forall \theta \in \Theta$. A *função escore* é definida por

$$U : \mathfrak{X} \times \Theta \to \mathbb{R}^P, \Theta \subseteq \mathbb{R}^P$$

e $\mathfrak{X} = \{x : f_\theta(x) > 0\}$ é o suporte de X, tal que

$$U(x,\theta) = \frac{\partial \ln f_\theta(x)}{\partial \theta} = \begin{pmatrix} \frac{\partial \ln f_\theta(x)}{\partial \theta_1} \\ \vdots \\ \frac{\partial \ln f_\theta(x)}{\partial \theta_P} \end{pmatrix}$$

e a função escore da amostra é definida por

$$U_n(x_n,\theta) = \frac{\partial \ln L_{x_n}(\theta)}{\partial \theta} = \begin{pmatrix} \frac{\partial \ln L_{x_n}(\theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial \ln L_{x_n}(\theta)}{\partial \theta_P} \end{pmatrix}$$

Note que, para amostras aleatórias,

$$U_n(x_n,\theta) = \sum_{i=1}^{n} U(x_i,\theta)$$

pois, por definição,

$$\begin{aligned} U_n(x_n,\theta) = \frac{\partial \ln L_{x_n}(\theta)}{\partial \theta} &= \frac{\partial \ln}{\partial \theta} \prod f_\theta(x_i) \\ &= \frac{\partial}{\partial \theta} \sum \ln f_\theta(x_i) \\ &= \sum \frac{\partial \ln f_\theta(x_i)}{\partial \theta} \\ &= \sum U(x_i,\theta). \end{aligned}$$

29.1 Condições de regularidade (simples)

$C_1 : \mathfrak{X} = \{x : f_\theta(x) > 0\}$ não depende de "θ".

$C_2 : f_\theta$ é duas vezes diferenciável com respeito a "θ" e suas derivadas são contínuas.

C_3 : é possível trocar as derivadas pela integrais da seguinte forma:

$$\begin{cases} a)\ \dfrac{\partial}{\partial\theta}\displaystyle\int f_\theta(x)dx & = \displaystyle\int \dfrac{\partial}{\partial\theta} f_\theta(x)dx \\ b)\ \dfrac{\partial^2}{\partial\theta\partial\theta^T}\displaystyle\int f_\theta(x)dx & = \displaystyle\int \dfrac{\partial^2}{\partial\theta\partial\theta^T} f_\theta(x)dx \end{cases}$$

C_4 :

$$\begin{cases} \text{a) } E_\theta\left(\|\dfrac{\partial}{\partial\theta}\ln f_\theta(x)\|\right) < \infty,\ \forall\theta\in\Theta \\ \text{b) } E_\theta\left(\|\dfrac{\partial^2}{\partial\theta\partial\theta^T}\ln f_\theta(x)\|^2\right) < \infty,\ \forall\theta\in\Theta. \end{cases}$$

Para as duas últimas condições, substituímos no caso discreto as integrais por somatórios.

Nas provas e listas, se não afirmado o contrário (explicitamente ou com suporte dependendo de θ), os exemplos satisfazem as condições de regularidade.

29.1.1 Teorema (das condições do escore)

Seja X_n a.a. de $X \sim f_\theta, \theta \in \Theta \subseteq \mathbb{R}^P$. Se as condições $C_1 : C_4$ estiverem satisfeitas, então

$$E_\theta(U_n(X_n,\theta)) = 0,\ \forall\theta\in\Theta.$$

29.1.1.1 Prova

Note que $U_n(X_n,\theta) = \sum U(X_i,\theta)$. Por $C_4(a)$, podemos calcular a esperança:

$$\begin{aligned} E_\theta(U_n(X_n,\theta)) &= \sum E_\theta(U(X_i,\theta)) \\ &\overset{\text{id.}}{=} nE_\theta(U(X,\theta)) \\ \Rightarrow E_\theta(U(X,\theta)) &= \int_{\mathfrak{X}} U(x,\theta)f_\theta(x)dx \\ &= \int_{\mathfrak{X}} \frac{\partial}{\partial\theta}\ln f_\theta(x)\cdot f_\theta(x)dx \\ &= \int_{\mathfrak{X}} \frac{\partial}{\partial\theta} f_\theta(x)dx, \forall\theta\in\Theta \end{aligned}$$

Note que

$$\int_{\mathfrak{X}} f_\theta(x)dx = 1, \forall\theta \in \Theta$$

Logo, por C_3

$$\frac{\partial}{\partial\theta}\int_{\mathfrak{X}} f_\theta(x)dx = 0 = \int_{\mathfrak{X}} \frac{\partial}{\partial\theta} f_\theta(x)dx,\ \forall\theta \in \Theta$$

30 Informação de Fisher

Seja $U(X, \theta)$ função escore. A informação de Fisher é a variância da função escore:

$$I_1(\theta) = \mathrm{Var}_\theta(U(X, \theta))$$

A informação de Fisher **total** é a variância da função escore da amostra:

$$I_n(\theta) = \mathrm{Var}_\theta(U_n(X_n, \theta))$$

Note que, sob as condições $C_1 : C_4$, temos que

$$\begin{aligned}
&E_\theta(U(X, \theta)) = 0, \ \forall \theta \in \Theta \\
&E_\theta(U(X_n, \theta)) = 0, \ \forall \theta \in \Theta \\
\Rightarrow &\begin{cases} I_1(\theta) = E_\theta(U(X, \theta) \cdot U(X, \theta)^T) \\ I_n(\theta) = E_\theta(U_n(X_n, \theta) \cdot U_n(X_n, \theta)^T) \end{cases} \quad \forall \theta \in \Theta \\
&\text{Se } \Theta \in \mathbb{R}, \\
\Rightarrow &\begin{cases} I_1(\theta) = E_\theta(U(X, \theta)^2) \\ I_n(\theta) = E_\theta(U_n(X_n, \theta)^2) \end{cases} \quad \forall \theta \in \Theta
\end{aligned}$$

30.1 Teorema (da informação de Fisher sob as condições)

Sob as condições $C_1 : C_4$, temos que

$$\begin{cases} I_1(\theta) = -E_\theta\left(\frac{\partial}{\partial \theta^T} U(X, \theta)\right) \\ I_n(\theta) = -E_\theta\left(\frac{\partial}{\partial \theta^T} U_n(X_n, \theta)\right) \end{cases} \quad \forall \theta \in \Theta$$

30.1.1 Prova

Sem perda de generalidade, considere apenas o caso contínuo e $\Theta \subseteq \mathbb{R}$. Para $\Theta \subseteq \mathbb{R}^P$, basta utilizar da transposição de θ nas derivadas e $U(X, \theta)$.

Sabemos que $U(X,\theta) = \frac{\partial}{\partial\theta} f_\theta(x)$. Além disso,

$$\int_{\mathfrak{X}} f_\theta(x)dx = 1,\ \forall\theta \in \Theta$$

em que $\mathfrak{X} = \{x : f_\theta(x) > 0\}$ não depende de "θ".

Por $C_3(b)$,

$$\frac{\partial^2}{\partial\theta\partial\theta^T} \int_{\mathfrak{X}} f_\theta(x)dx = \int_{\mathfrak{X}} \frac{\partial^2}{\partial\theta\partial\theta^T} f_\theta(x)dx = 0, \forall\theta \in \Theta \tag{30.1}$$

Note que

$$\begin{aligned} I_1(\theta) &= E_\theta(U(X,\theta)^2) \\ &= \int_{\mathfrak{X}} U(x,\theta)^2 f_\theta(x)dx \\ &= \int_{\mathfrak{X}} \left[\frac{\partial \ln}{\partial\theta} f_\theta(x)\right]^2 f_\theta(x)dx,\ \forall\theta \in \Theta \end{aligned}$$

Observe que, por (**??**),

$$\begin{aligned} &\int_{\mathfrak{X}} \frac{\partial}{\partial\theta}\left[\frac{\partial f_\theta(x)}{\partial\theta}\frac{f_\theta(x)}{f_\theta(x)}\right]dx = 0 \\ \Rightarrow &\int_{\mathfrak{X}} \frac{\partial}{\partial\theta}\left[\frac{\partial \ln f_\theta(x)}{\partial\theta} f_\theta(x)\right]dx = 0 \\ \Rightarrow &\int_{\mathfrak{X}} \frac{\partial}{\partial\theta}\left[\frac{\partial^2 \ln f_\theta(x)}{\partial\theta^2} f_\theta(x) + \frac{\partial \ln f_\theta(x)}{\partial\theta}\frac{\partial f_\theta(x)}{\partial\theta}\right]dx = 0 \\ \Rightarrow &\int_{\mathfrak{X}} \frac{\partial^2 \ln f_\theta(x)}{\partial\theta^2} f_\theta(x) + \int_{\mathfrak{X}} \frac{\partial \ln f_\theta(x)}{\partial\theta}\frac{\partial f_\theta(x)}{\partial\theta}dx = 0 \\ \Rightarrow &E_\theta\left(\frac{\partial^2 \ln}{\partial\theta^2} f_\theta(X)\right) + \int_{\mathfrak{X}} \frac{\partial \ln f_\theta(x)}{\partial\theta}\frac{\partial \ln f_\theta(x)}{\partial\theta} f_\theta(x)dx = 0 \\ \Rightarrow &-E_\theta\left(\frac{\partial^2 \ln}{\partial\theta^2} f_\theta(X)\right) = E_\theta\left(\left[\frac{\partial \ln}{\partial\theta} f_\theta(X)\right]^2\right) \\ \Rightarrow &E_\theta(U(X,\theta)^2) = -E_\theta\left(\frac{\partial U(X,\theta)}{\partial\theta}\right) \\ \Rightarrow &I_1(\theta) = -E_\theta\left(\frac{\partial U(X,\theta)}{\partial\theta}\right),\ \forall\theta \in \Theta. \end{aligned}$$

Ademais, temos que

$$I_n(\theta) = \mathrm{Var}_\theta(U_n(X_n,\theta)) \overset{C_1:C_4}{=} E_\theta(U_n(X_n,\theta)^2)$$

Sabemos que $U_n(X_n, \theta) = \sum_{i=1}^{n} U(X_i, \theta)$. Portanto,

$$
\begin{aligned}
I_n(\theta) &= E_\theta \left[\left(\sum U(X_i, \theta) \right)^2 \right] \\
&= E_\theta \Big(\sum_i \sum_j U(X_i, \theta) U(X_j, \theta) \Big) \\
&= E_\theta \left(\sum_{i=1}^{n} U(X_i, \theta)^2 + \sum_{i \neq j} U(X_i, \theta) U(X_j, \theta) \right) \\
&= \sum E_\theta (U(X_i, \theta)^2) + \sum_{i \neq j} E_\theta (U(X_i, \theta) U(X_j, \theta)), \forall \theta \in \Theta.
\end{aligned}
$$

Como $X_1, \dots, X_n$ são i.i.d, temos que $U(X_1, \theta), \dots, U(X_n, \theta)$ também são i.i.d. Logo,

$$
I_n(\theta) = \sum E_\theta (U(X_i, \theta)^2) + \sum_{i \neq j} E_\theta (\cancelto{0}{U(X_i, \theta)}) E_\theta (\cancelto{0}{U(X_j, \theta)})
$$

Por ser i.i.d., temos portanto que

$$
I_n(\theta) = n I_1(\theta)
$$

Note também que

$$
\begin{aligned}
E_\theta \left(\frac{\partial}{\partial \theta} U(X_n, \theta) \right) &= E_\theta \left(\frac{\partial}{\partial \theta} \sum U(X_i, \theta) \right) \\
&= E_\theta \left(\sum \frac{\partial}{\partial \theta} U(X_i, \theta) \right) \\
&= n E_\theta \left(\frac{\partial}{\partial \theta} U(X_i, \theta) \right), \ \forall \theta \in \Theta.
\end{aligned}
$$

Como $I_1(\theta) = -E_\theta \left(\frac{\partial}{\partial} U(X, \theta) \right)$. Temos que

$$
\begin{aligned}
& - E_\theta \left(\frac{\partial}{\partial \theta} U(X_n, \theta) \right) = n I_1(\theta) \\
\Rightarrow & I_n(\theta) = n I_1(\theta) = -E_\theta \left(\frac{\partial}{\partial \theta} U(X_n, \theta) \right), \forall \theta \in \Theta. \blacksquare
\end{aligned}
$$

30.2 Exemplo

Seja X_n a.a de $X \sim f_\theta, \theta \in \mathbb{R}_+$, tal que

$$
f_\theta(x) = \begin{cases} \theta x^{\theta - 1}, & x \in (0, 1) \\ 0, & \text{c.c.} \end{cases}
$$

Encntre a função escore da amostra e a informação de Fisher total.

30.2.1 Resposta

Note que

$$L_{X_n}(\theta) = \prod^{n} \{\theta x_i^{\theta-1}\} = \theta^n (\prod x_i)^{\theta-1}, \ \forall\theta \in \Theta.$$

A função escore da amostra é

$$\begin{aligned} U_n(X_n, \theta) &= \frac{\partial}{\partial\theta} \ln L_{X_n}(\theta) \\ &= \frac{\partial}{\partial\theta} \left[n \ln\theta + (\theta - 1) \sum \ln x_i\right] \\ &= \frac{n}{\theta} + \sum \ln x_i, \forall\theta \in \Theta. \end{aligned}$$

A informação de Fisher total é, pelo teorema,

$$I_n(\theta) = -E_\theta \left(\frac{\partial}{\partial\theta} U_n(X_n, \theta)\right), \forall\theta \in \Theta.$$

Note que

$$\frac{\partial}{\partial\theta} U_n(X_n, \theta) = -\frac{n}{\theta^2}, \forall\theta \in \Theta.$$

Portanto,

$$I_n(\theta) = -\frac{n}{\theta^2}, \forall\theta \in \Theta.$$

30.3 Exemplo (Binomial)

Seja X_n a.a. de $X \sim \text{Bin}(3, \theta), \theta \in (0, 1)$. Encontre a função escore e a informação de Fisher total.

30.3.1 Resposta

Note que

$$\begin{aligned} L_{X_n}(\theta) &= \prod f_\theta(x) = \prod \left\{ \binom{3}{x_i} \theta^{x_i} (1-\theta)^{3-x} \right\} \\ \Rightarrow \ln L_{X_n}(\theta) &= \sum \ln \binom{3}{x_i} + \sum x_i \ln\theta + \sum (3 - x_i) \ln(1-\theta) \\ \Rightarrow U_n(x_n, \theta) &= \sum \frac{x_i}{\theta} + \sum \frac{(3 - x_i)}{1 - \theta} (-1) \\ &= \frac{x_i}{\theta} - \sum \frac{(3 - x_i)}{1 - \theta}, \ \forall\theta \in \Theta. \end{aligned}$$

A informação de Fisher total é, pelo teorema,

$$
\begin{aligned}
I_n(\theta) &= -E_\theta\left(\frac{\partial}{\partial\theta}U_n(X_n,\theta)\right)\\
&= -E_\theta\left(-\sum\frac{X_i}{\theta^2}+\sum\frac{(3-X)}{(1-\theta)^2}(-1)\right)\\
&= E_\theta\left(\sum\frac{X_i}{\theta^2}+\sum\frac{(3-X)}{(1-\theta)^2}\right)\\
&= \sum\frac{E_\theta(X_i)}{\theta^2}+\sum\frac{3-E_\theta(X)}{(1-\theta)^2}\\
&= \frac{n3\theta}{\theta^2}+\frac{n(3-3\theta)}{(1-\theta)^2}\\
&= \frac{3n}{\theta}+\frac{3n(1-\theta)}{(1-\theta)^2}\\
&= \frac{3n}{\theta}+\frac{3n}{(1-\theta)}\\
&= \frac{3n}{\theta(1-\theta)},\ \forall\theta\in\Theta.
\end{aligned}
$$

Podemos também fazer pela definição.

Observe que

$$
\begin{aligned}
E_\theta(U_n(X_n,\theta) &= E_\theta\left(\sum\frac{X_i}{\theta}-\sum\frac{3-X_i}{1-\theta}\right)\\
&= \sum\frac{E_\theta(X_i)}{\theta}-\frac{(3-E_\theta(X_i)}{1-\theta}\\
&= \frac{n3\theta}{\theta}-\frac{n(3-3\theta)}{1-\theta}\\
&= n3-n3\frac{1-\theta}{1-\theta}=0,\ \forall\theta\in\Theta.
\end{aligned}
$$

$$
\begin{aligned}
I_n(\theta) &= \mathrm{Var}_\theta(U_n(X_n,\theta)) \\
&= E_\theta(U_n(X,\theta)^2) \\
&= E_\theta\left(\left[\sum\left(\frac{X_i}{\theta}-\frac{(3-X_i)}{1-\theta}\right)\right]^2\right) \\
&= E_\theta\left(\sum_i\sum_j\left(\frac{X_i}{\theta}-\frac{(3-X_i)}{1-\theta}\right)\left(\frac{X_j}{\theta}-\frac{(3-X_j)}{1-\theta}\right)\right) \\
&= \sum_i E_\theta\left(\left(\frac{X_i}{\theta}-\frac{(3-X_i)}{1-\theta}\right)^2\right) \\
&\cdot \sum\sum_{i\neq j} E_\theta\left(\cancelto{0}{\frac{X_i}{\theta}-\frac{(3-X_i)}{1-\theta}}\right) E_\theta\left(\cancelto{0}{\frac{X_j}{\theta}-\frac{(3-X_j)}{1-\theta}}\right) \\
&= nE_\theta\left[\left(\frac{X}{\theta}-\frac{3-X}{1-\theta}\right)^2\right] \\
&= nE_\theta\left[\left(\frac{X(1-\theta)-(3-x)\theta}{\theta(1-\theta)}\right)^2\right] \\
&= \frac{nE_\theta\left[(X-\theta X-3\theta+\theta X)^2\right]}{\theta(1-\theta)} \\
&= \frac{nE_\theta\left[(X-3\theta)^2\right]}{\theta^2(1-\theta)^2} \\
&= \frac{n\mathrm{Var}_\theta}{\theta^2(1-\theta)^2} = \frac{n3\theta(1-\theta)}{\theta^2(1-\theta)^2} \\
&= \frac{3n}{\theta(1-\theta)},\ \forall\theta\in\Theta.
\end{aligned}
$$

30.4 Teorema (Desigualdade da informação, LICR)

Seja X_n a.a. de $X \sim f_\theta, \theta \in \Theta \subseteq \mathbb{R}$, em que f_θ satisfaz as condições $C_1 : C_4$. Considere $T(X_n)$ um estimador não-viciado para $g(\theta) = \theta$. Então,

$$\mathrm{Var}_\theta(T(X_n)) \geq \frac{1}{nI_1(\theta)} = \frac{1}{I_n(\theta)}, \forall\theta \in \Theta,$$

sempre que $E_\theta(T(X_n)^2) < \infty, \forall\theta \in \Theta$, em que $I_1(\theta)$ é a informação de Fisher e $I_n(\theta)$ é a informação de Fisher total.

> **ℹ Observação**
>
> A quantidade $\frac{1}{I_n(\theta)}$ é chamada de *Limite Inferior de Crammer Rao (LICR)*

30.4.1 Prova

Como $T(X_n)$ é não-viciado para $g(\theta) = \theta$, temos que

$$E_\theta(T(X_n)) = \theta, \forall \theta \in \Theta.$$

Por definição de esperança (para o caso contínuo),

$$E_\theta(T(X_n)) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} T(x_1, \dots, x_n) f_\theta^{X_n}(x_1, \dots, x_n) dx_1, \dots, dx_n \tag{30.2}$$

> **ℹ Observação**
>
> $$E((X_1 + X_2)^2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 + x_2)^2 f_\theta^{X_1, X_2}(x_1, x_2) dx_1 dx_2$$

Diferenciando (**??**) com relação a “θ”

$$\begin{aligned}
&\frac{\partial}{\partial \theta} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} T(x_n) f_\theta^{X_n} dx_n = 1 \\
&\overset{C_3}{\Rightarrow} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} T(x_n) \frac{\partial}{\partial \theta} f_\theta^{X_n}(x_n) = 1 \\
&\Rightarrow \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} T(x_n) \frac{\partial \ln}{\partial \theta} f_\theta^{X_n}(x_n) f_\theta^{X_n}(x_n) dx_n = 1, \forall \theta \in \Theta. \\
&\Rightarrow E_\theta(T(X_n) \cdot U_n(X_n, \theta)) = 1, \forall \theta \in \Theta.
\end{aligned}$$

Note que

$$\begin{aligned}
&\mathrm{Cor}_\theta(T(X_n), U_n(X_n, \theta))^2 \leq 1, \forall \theta \in \Theta \\
\iff &\frac{\mathrm{Cov}_\theta(T(X_n), U_n(X_n, \theta))^2}{\mathrm{Var}_\theta(T(X_n)) \mathrm{Var}_\theta(U_n(X_n, \theta))} \leq 1, \forall \theta \in \Theta.
\end{aligned} \tag{30.3}$$

Ademais,

$$\mathrm{Cov}_\theta(T(X_n), U_n(X_n, \theta)) = E_\theta(T(X_n \cdot U_n(X_n, \theta)) - \cancelto{\theta}{E_\theta(T(X_n)} \cdot \cancelto{0}{E_\theta(U_n(X_n, \theta))} = 1, \forall \theta \in \Theta$$

Substituindo em (**??**), temos que

$$\frac{1^2}{\text{Var}_\theta(T(X_n))\text{Var}_\theta(U_n(X_n,\theta))} \leq 1, \forall\theta \in \Theta.$$

Por definição de informação de Fisher total,

$$\frac{1}{I_n(\theta)} \leq \text{Var}_\theta(T(X_n)), \forall\theta \in \Theta.$$

30.5 Teorema (generalização)

Seja X_n a.a. de $X \sim f_\theta, \theta \in \Theta$, em que f_θ satisfaz as condições C_1, C_4. Considere $T(X_n)$ um estimador não-viciado para $g(\theta) \in \mathbb{R}$ cuja derivada com respeito a "θ" existe $\forall\theta \in \Theta$. Então,

$$\text{Var}_\theta(T(X_n)) \geq \frac{g'(\theta)^2}{I_n(\theta)}, \forall\theta \in \Theta.$$

sempre que $E_\theta(T(X_n)^2) < \infty, \forall\theta \in \Theta$.

30.5.1 Prova

Exercício!!!

31 Estimadores Eficientes

Dizemos que $T(X_n)$ é um estimador eificiente para $g(\theta)$ se, e somente se,

1. $T(X_n)$ é não-viesado para $g(\theta)$.
2. $\text{Var}_\theta(T(X_n)) = \frac{g'(\theta)^2}{I_n(\theta)}, \forall\theta \in \Theta$. Em que $I_n(\theta)$ é a informação de Fisher total.

31.1 Exemplo (Binomial)

Seja X_n amostra aleatória de $X \sim \text{Bin}(3,\theta), \theta \in \Theta = (0,1)$.

a-) Verifique se $\frac{\bar{X}}{3} = \frac{1}{n}\sum X_i \frac{1}{3}$ é suficiente para $g(\theta) = \theta$.

b-) Verifique se $\bar{X} = \frac{1}{n}\sum X_i$ é suficiente para $g(\theta) = 3\theta$.

31.1.1 Resposta a-)

Já sabemos que

$$E_\theta\left(\frac{\bar{X}}{3}\right) = \theta, \forall\theta \in \Theta.$$

$$\begin{aligned}
\text{Var}_\theta\left(\frac{\bar{X}}{3}\right) &= \frac{1}{9n^2}\text{Var}\left(\sum X_i\right) \\
&\stackrel{\text{iid}}{=} \frac{1}{9n^2} \cdot n\text{Var}_\theta(X) \\
&= \frac{1}{9n}3\theta(1-\theta),\ \forall\theta \in \Theta.
\end{aligned}$$

Já calculamos a informação de Fisher total.

$$I_n(\theta) = \frac{3n}{\theta(1-\theta)}$$

Como

$$\text{Var}_\theta\left(\frac{\bar{X}}{3}\right) = \frac{1}{I_n(\theta)}, \forall\theta \in \Theta,$$

$\frac{\bar{X}}{3}$ é um estimador eficiente para $g(\theta) = \theta$

31.1.2 Resposta b-)

Sabemos que

$$E_\theta(\bar{X}) = 3\theta, \ \forall \theta \in \Theta$$

$$\text{Var}_\theta(\bar{X}) = \frac{3}{n}\theta(1-\theta), \ \forall \theta \in \Theta.$$

Note que

$$\frac{g'(\theta)^2}{I_n(\theta)} = \frac{\theta(1-\theta)}{3n}[3]^2 = \frac{3\theta(1-\theta)}{n}$$

Logo, $\bar{X}$ é eficiente para $g(\theta) = 3\theta$.

31.2 Exemplo (Poisson)

Seja X_n a.a. de $X \sim \text{Poiss}(\theta), \theta \in \Theta = \mathbb{R}_+$

a-) Verifique se $\bar{X}$ é eficiente para $g(\theta) = \theta$.

b-) Verifique se $T(X_n) = \left(1 - \frac{1}{n}\right)^{\sum X_i}$ é eficiente para $g(\theta) = P_\theta(X = 0)$.

31.2.1 Resposta a-)

Já sabemos que

$$E_\theta(\bar{X}) = \theta, \ \forall \theta \in \Theta$$
$$\text{Var}_\theta(\bar{X}) = \frac{\theta}{n}, \ \forall \theta \in \Theta.$$

Como $T(X_n) = \bar{X}$ é não-viciado, precisamos apenas verificar se

$$\text{Var}_\theta(\bar{X}) = \frac{1}{I_n(\theta)}$$

Note que

$$
\begin{aligned}
L_{x_n}(\theta) = f_\theta^{X_n}(\theta) = \prod \left\{ e^{-\theta} \frac{\theta^{x_i}}{x_i!} \right\} \\
&\Rightarrow \ln L_{x_n}(\theta) = -n\theta + \left(\sum x_i\right) \ln\theta - \sum \ln(x_i) \\
&\Rightarrow \frac{\partial}{\partial\theta} \ln L_{x_n}(\theta) = -n + \frac{\sum x_i}{\theta} \\
&\Rightarrow \frac{\partial^2}{\partial\theta^2} \ln L_{x_n}(\theta) = -\frac{\sum x_i}{\theta^2}
\end{aligned}
$$

Pelo Teorema, como valem as condições $C_1 : C_4$,

$$
\begin{aligned}
I_n(\theta) &= -E_\theta \left(\frac{\partial^2 \ln L_X(\theta)}{\partial\theta^2} \right) \\
&= -E_\theta \left(-\sum \frac{X_i}{\theta^2} \right) \\
&\overset{\text{iid}}{=} \frac{1}{\theta^2} n E_\theta(X) = \frac{n}{\theta}, \forall\theta \in \Theta.
\end{aligned}
$$

Como $\mathrm{Var}_\theta(\bar{X}) = \frac{1}{I_n(\theta)}$, temos que $\bar{X}$ é eficiente para $g(\theta) = \theta$.

31.2.2 Resposta b-)

Note que $\sum X_i \sim \mathrm{Poiss}(n\theta), \forall\theta \in \Theta$ (por F.G.M.).

> **ℹ Nota**
>
> $$M_{\sum X_i}(t) = E_\theta \left(\mathrm{e}^{t \sum X_i} \right) = \mathrm{e}^{n\theta(e^t - 1)}$$

Observe que

$$
E_\theta \left(\left(1 - \frac{1}{n}\right)^{\sum X_i} \right) = E_\theta \left(\mathrm{e}^{(\sum X_i) \ln(1 - \frac{1}{n})} \right)
$$

pois $n > 1$.

Tome $t = \ln\left(1 - \frac{1}{n}\right)$.

$$
\begin{aligned}
E_\theta \left(\left(1 - \frac{1}{n}\right)^{\sum X_i} \right) = M_{\sum X_i}\left(\ln\left(1 - \frac{1}{n}\right) \right) &= \mathrm{e}^{n\theta\left(\mathrm{e}^{\ln\left(1 - \frac{1}{n}\right)} - 1\right)} \\
&= \mathrm{e}^{n\theta(1 - \frac{1}{n} - 1)} \\
&= \mathrm{e}^{-\theta}, \forall\theta \in \Theta.
\end{aligned}
$$

Portanto, $E_\theta(T(X_n)) = \mathrm{e}^{-\theta} = P_\theta(X = 0) = g(\theta)$, ou seja, é não-viciada.

Note que

$$\begin{aligned}\text{Var}_\theta(T(X_n)) &= E_\theta(T(X_n)^2) - E_\theta(T(X_n))^2\\ &= E_\theta(T(X_n)^2) - \mathrm{e}^{-2\theta}\end{aligned}$$

$$\begin{aligned}E_\theta(T(X_n)^2) &= E_\theta\left(\left(1-\frac{1}{n}\right)^{2\sum X_i}\right)\\ &= E_\theta\left(\mathrm{e}^{2(\sum X_i)\ln\left(1-\frac{1}{n}\right)}\right)\\ &= M_{\sum X_i}\left(2\ln\left(1-\frac{1}{n}\right)\right)\\ &= \mathrm{e}^{n\theta\left(\mathrm{e}^{2\ln\left(1-\frac{1}{n}\right)}-1\right)}\\ &= \mathrm{e}^{n\theta\left(\mathrm{e}^{\ln\left(1-\frac{1}{n}\right)^2}-1\right)}\\ &= \mathrm{e}^{n\theta\left(\left(1-\frac{1}{n}\right)^2-1\right)}\\ &= \mathrm{e}^{n\theta\left(1-\frac{2}{n}+\frac{1}{n^2}-1\right)}\\ &= \mathrm{e}^{\theta\left(\frac{1}{n}-2\right)} = \mathrm{e}^{-2\theta}\cdot\mathrm{e}^{\frac{\theta}{n}}.\end{aligned}$$

$$\begin{aligned}\Rightarrow \text{Var}_\theta(T(X_n)) &= \mathrm{e}^{-2\theta}\cdot\mathrm{e}^{\frac{\theta}{n}} - \mathrm{e}^{-2\theta}\\ &= \mathrm{e}^{-2\theta}\left(\mathrm{e}^{\frac{\theta}{n}}-1\right)\end{aligned}$$

Temos nosso LICR

$$\frac{g'(\theta)^2}{I_n(\theta)} = \frac{\theta\mathrm{e}^{-2\theta}}{n} = \text{LICR}$$

Como $\text{Var}_\theta \neq \frac{g'(\theta)}{I_n(\theta)}$ para algum $\theta \in \Theta$, como $\theta = 1$, temos que, apesar de ser ENVVUM, $T(X_n)$ não é eficiente para $g(\theta) = \mathrm{e}^{-\theta}$.

31.3 Eficiência assintótica para $g(\theta)$

Dizemos que $T(X_n)$ é um estimador assintoticamente eficiente se, e somente se,

1. $\lim\limits_{n\to\infty} E_\theta(T(X_n)) = g(\theta), \forall\theta\in\Theta$.
2. $\lim\limits_{n\to\infty} n\text{Var}_\theta(T(X_n)) = \frac{g'(\theta)^2}{I_1(\theta)}, \forall\theta\in\Theta$.

31.3.1 Exemplo (Poisson)

Seja X_n a.a. de $X \sim \text{Poiss}(\theta), \theta \in \mathbb{R}_+$.

a-) Verifique se $T(X_n) = \left(1 - \frac{1}{n}\right)^{\sum X_i}$

b-) Verifique se $T^*(X_n) = \frac{1}{n}\mathbb{1}(X_i = 0)$

c-) Verifique se $T_{MV}(X_n) = \mathrm{e}^{-\bar{X}}$ é eficiente para $g(\theta) = P_\theta(X = 0)$.

31.3.1.1 Resposta a-)

Já sabemos que

$$\text{Var}_\theta(T(X_n)) = \mathrm{e}^{-2\theta}(\mathrm{e}^{\frac{\theta}{n}} - 1), \ \forall \theta \in \Theta$$

$$E_\theta(T(X_n)) = \mathrm{e}^{-\theta} = g(\theta), \ \forall \theta \in \Theta.$$

Ou seja, é não-viciado, mas não eficiente, pois

$$\text{Var}_\theta(T(X_n)) \neq \frac{\theta \mathrm{e}^{-2\theta}}{n}$$

Precisamos calcular

$$\lim_{n\to\infty} n\text{Var}_\theta(T(X_n)) = \lim_{n\to\infty} n\mathrm{e}^{-2\theta}(\mathrm{e}^{\frac{\theta}{n}} - 1)$$

Note que

$$\mathrm{e}^x = \sum \frac{x^i}{i!} = 1 + x + \frac{x^2}{2!} + \cdots$$

Portanto,

$$\begin{aligned}
\mathrm{e}^{\frac{\theta}{n}} &= 1 + \frac{\theta}{n} + (\frac{\theta}{n})^2\frac{1}{x!} + ... \\
\iff \mathrm{e}^{\frac{\theta}{n}} - 1 &= \frac{\theta}{n} + \frac{\theta^2}{n^2 2!} + ... \\
\iff n(\mathrm{e}^{\frac{\theta}{n}} - 1) &= \theta + \frac{\theta^2}{n2!} + ... \\
\Rightarrow \lim_{n\to\infty} n(\mathrm{e}^{\frac{\theta}{n}} - 1) &= \theta, \ \forall \theta \in \Theta.
\end{aligned}$$

Dessa forma,

$$\lim_{n\to\infty} n\text{Var}_\theta(T(X_n)) = \mathrm{e}^{-2\theta}\theta = \frac{g'(\theta)^2}{I_1(\theta)}, \ \forall \theta \in \Theta.$$

Logo, $T(X_n)$ é assintoticamente eficiente para $g(\theta)$

31.3.1.2 Resposta b-)

Note que

$$
\begin{aligned}
E_\theta(T^*(X_n)) &= \frac{1}{n}\sum E_\theta(\mathbb{1}(X_i = 0)) \\
&\overset{\text{iid}}{=} \frac{1}{n}\cdot n \cdot P_\theta(X = 0) = \mathrm{e}^{-\theta},\ \forall \theta \in \Theta.
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}_\theta(T^*(X_n)) &\overset{\text{iid}}{=} \frac{1}{n^2}\cdot n \cdot \mathrm{Var}_\theta(\mathbb{1}(X = 0)), \\
&= \frac{1}{n} = \mathrm{e}^{-\theta}(1 - \mathrm{e}^{-\theta}),\ \forall \theta \in \Theta.
\end{aligned}
$$

Observe que o LICR é

$$
\frac{g'(\theta)^2}{nI_1(\theta)} = \frac{\theta \mathrm{e}^{-2\theta}}{n}.
$$

Como

$$
\mathrm{Var}_\theta(T^*(X_n)) \neq \frac{g'(\theta)^2}{nI_1(\theta)},
$$

concluímos que não é eficiente. Além disso,

$$
\lim_{n\to\infty} n\mathrm{Var}_\theta(T^*(X_n)) = \mathrm{e}^{-\theta}(1 - \mathrm{e}^{-\theta}) \neq \theta \mathrm{e}^{-2\theta}.
$$

Logo, não é assintoticamente eficiente.

31.3.1.3 Resposta c-)

Note que

$$
E_\theta(T_{MV}(X_n)) = E_\theta(\mathrm{e}^{-\bar{X}}) = E_\theta(\mathrm{e}^{-\frac{1}{n}\sum X_i})
$$

em que $\sum X_i \sim \text{Poiss}(n\theta), \theta \in \Theta$.

Observe que

$$
M_{\sum X_i}(t) = E_\theta\left(\mathrm{e}^{t\sum X_i}\right) = \mathrm{e}^{n\theta(\mathrm{e}^t - 1)}
$$

Tomando $t = -\frac{1}{n}$, temos que

$$
E_\theta(T_{MV}(X_n)) = M_{\sum X_i}(-\frac{1}{n}) = \mathrm{e}^{n\theta(\mathrm{e}^{-\frac{1}{n}} - 1)}
$$

Ademais,

$$
\begin{aligned}
\mathrm{e}^{-\frac{1}{n}} &= 1 - \frac{1}{n} + \left(-\frac{1}{n}\right)^2 \cdot \frac{1}{2} + ... \\
\iff \mathrm{e}^{-\frac{1}{n}} - 1 &= -\frac{1}{n} + \frac{1}{n^2 2!} + ... \\
\iff n\theta(\mathrm{e}^{-\frac{1}{n}} - 1) &= -\theta + \underbrace{\left(\frac{\theta}{2n} + ...\right)}_{h(\theta,n)} \\
\Rightarrow E_\theta(T_{MV}(X_n)) &= \mathrm{e}^{-\theta + h(\theta,n)}
\end{aligned}
$$

Como $h(\theta, n) \neq 0$,

$$E_\theta(T_{MV}(X_n)) \neq \mathrm{e}^{-\theta},$$

e

$$\lim_{n\to\infty} E_\theta(T_{MV}(X_n)) = \mathrm{e}^{-\theta},\ \forall \theta \in \Theta.$$

Apesar de $T_{MV}(X_n)$ ser viciado para $g(\theta) = \mathrm{e}^{\theta}$, ela é não-viciada assintoticamente.

Note ainda que

$$
\begin{aligned}
E_\theta(T_{MV}(X_n)^2) &= E_\theta\left(\mathrm{e}^{-\frac{2}{n}\sum X_i}\right) \\
&= M_{\sum X_i}(-\frac{2}{n}) \\
&= \mathrm{e}^{n\theta\left(\mathrm{e}^{-\frac{2}{n}} - 1\right)}
\end{aligned}
$$

Portanto,

$$\mathrm{Var}_\theta(T_{MV}(X_n)) = \mathrm{e}^{n\theta\left(\mathrm{e}^{-\frac{2}{n}} - 1\right)} - \mathrm{e}^{-2\theta + 2h(\theta,n)}$$

Seguindo,

$$
\begin{aligned}
\mathrm{e}^{-\frac{2}{n}} &= 1 + \left(-\frac{2}{n}\right) + \left(-\frac{2}{n}\right)^2 \cdot \frac{1}{2!} + ... \\
\iff n\theta(\mathrm{e}^{-\frac{2}{n}} - 1) &= -2\theta + \underbrace{n\theta\left(\frac{2}{n}\right)^2 + n\theta(...)}_{w(\theta,n)} \\
\Rightarrow \mathrm{Var}_\theta(T_{MV}(X_n)) &= \mathrm{e}^{-2\theta + w(\theta,n)} - \mathrm{e}^{-2\theta + 2h(\theta,n)} \\
&= \mathrm{e}^{-2\theta}\left(\mathrm{e}^{w(\theta,n)} - \mathrm{e}^{2h(\theta,n)}\right) \\
\Rightarrow \lim_{n\to\infty} n\mathrm{Var}_\theta(T_{MV}(X_n) &= \lim_{n\to\infty} n\mathrm{e}^{-2\theta}\left(\mathrm{e}^{w(\theta,n)} - \mathrm{e}^{2h(\theta,n)}\right) \\
&= \mathrm{e}^{-2\theta} \lim_{n\to\infty} n\left(\mathrm{e}^{w(\theta,n)} - \mathrm{e}^{2h(\theta,n)}\right)
\end{aligned}
$$

Note que

$$
\begin{aligned}
\mathrm{e}^{w(\theta,n)} &= 1 + w(\theta,n) + \frac{w(\theta,n)^2}{2!} + ... \\
\mathrm{e}^{2h(\theta,n)} &= 1 + 2h(\theta,n) + \frac{2h(\theta,n)^2}{2!} + ...
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow \mathrm{e}^{w(\theta,n)} - \mathrm{e}^{2h(\theta,n)} &= [w(\theta,n) - 2h(\theta,n)] + \left[\frac{w(\theta,n)^2}{2!} - \frac{2h(\theta,n)^2}{2!}\right] + ... \\
\Rightarrow n\left(\mathrm{e}^{w(\theta,n)} - \mathrm{e}^{2h(\theta,n)}\right) &= n[w(\theta,n) - 2h(\theta,n)] + \underbrace{n\left[\frac{w(\theta,n)^2}{2!} - \frac{2h(\theta,n)^2}{2!}\right] + ...}_{\overset{n\uparrow\infty}{\to 0}}
\end{aligned}
$$

Em que

$$
\begin{aligned}
2h(\theta,n) &= \frac{2\theta}{2n} - \frac{2\theta}{3!n^2} + \frac{2\theta}{4!n^3} + ... \\
w(\theta,n) &= \frac{4\theta}{2n} - \frac{8\theta}{3!n^2} + \frac{\theta}{4!n^3} + ...
\end{aligned}
$$

Agora,

$$
\begin{aligned}
& n\left[\left(\frac{4\theta}{2!n} - \frac{8\theta}{3!n^2} + ...\right) - \left(\frac{\theta}{n} - \frac{2\theta}{3!n^2} + ...\right)\right] \\
& = \left[\left(2\theta - \frac{8\theta}{3!n} + ...\right) - \left(\theta - \frac{2\theta}{3!n^2} + ...\right)\right] \\
\overset{n\uparrow\infty}{\to}\ & 2\theta - \theta = \theta
\end{aligned}
$$

Portanto,

$$
\lim_{n\to\infty} n\mathrm{Var}_\theta(T(X_n)) = \mathrm{e}^{-2\theta} \cdot \theta,\ \ \forall\theta \in \Theta.
$$

e concluímos que $T_{MV}(X_n) = \mathrm{e}^{-\bar{X}}$ é assintoticamente eficiente.

32 Consistência

Dizemos que $T(X_n)$ é um estimador (fracamente) consistente para $g(\theta)$ se, e somente se, para cada $\epsilon > 0$ fixo,

$$\lim_{n\to\infty} P_\theta\left(|T(X_n) - g(\theta)| > \epsilon\right) = 0, \forall\theta \in \Theta.$$

ou

$$T(X_n) \stackrel{P_\theta}{\to} g(\theta),\ \ \forall\theta \in \Theta.$$

> Cuidado
>
> Alguns livros escrevem
>
> $$T(X_n) \stackrel{p}{\to} g(\theta_0)$$
>
> em que θ_0 é o verdadeiro valor de θ se, e somente se, para cada $\epsilon > 0$ fixado:
>
> $$\lim_{n\to\infty} P\left(|T(X_n) - g(\theta_0)| > \epsilon\right) = 0.$$
>
> Apesar de ser uma forma comum de exposição para pessoas não treinadas em estatística, essa notação não está bem formalizada na estrutura do modelo estatístico clássico.

Dizemos $T(X_n)$ é *fortemente* consistente se, e somente se,

$$P_\theta\left(\lim_{n\to\infty} T(X_n) = g(\theta)\right) = 1,\ \ \forall\theta \in \Theta.$$

ou

$$T(X_n) \stackrel{\text{q.c.}[P_\theta]}{\to} g(\theta),\ \ \forall\theta \in \Theta.$$

> Cuidado
>
> Alguns livros escrevem
>
> $$T(X_n) \stackrel{\text{q.c.}}{\to} g(\theta_0)$$
>
> em que θ_0 é o verdadeiro valor de θ se, e somente se,
>
> $$P\left(\lim_{n\to\infty} T(X_n) = g(\theta_0)\right) = 1.$$

Apesar de ser uma forma comum de exposição para pessoas não treinadas em estatística, essa notação não está bem formalizada na estrutura do modelo estatístico clássico.

Nota

Quando se diz que $T(X_n)$ é consistente para $g(\theta)$, sem especificar a convergência, quer-se dizer que a convergência é fraca.

Nota

$$P_\theta(\underbrace{\lim_{n\to\infty} T(X_n) = g(\theta)}_{\{\omega\in\Omega:\ \lim_{n\to\infty} T(X_n(\omega))=g(\theta)\}})$$

32.1 Teorema (da relação com o EQM)

Seja $T(X_n)$ um estimador para $g(\theta) \in \mathbb{R}$ tal que o Erro Quadrático Médio

$$\text{EQM}_\theta(T(X_n), g(\theta)) = E_\theta[(T(X_n) - g(\theta))^2]$$

converge para 0 para cada $\theta \in \Theta$. Então, $T(X_n)$ é fracamente consistente para $g(\theta)$

32.1.1 Prova

Seja $\epsilon > 0$ fixado. Então,

$$\begin{aligned} P_\theta(|T(X_n) - g(\theta)| > \epsilon) &= P_\theta((T(X_n) - g(\theta))^2 > \epsilon^2) \\ &\overset{\text{Chebyshev}}{\leq} \frac{E_\theta((T(X_n - g(\theta))^2)}{\epsilon^2} = \frac{\text{EQM}_\theta(T(X_n), g(\theta))}{\epsilon^2} \\ \Rightarrow \lim_{n\to\infty} P_\theta(|T(X_n) - g(\theta)| > \epsilon) \leq 0,\ \forall\theta \in \Theta. \end{aligned}$$

Logo, $T(X_n) \overset{P_\theta}{\to} g(\theta),\ \forall\theta \in \Theta.$

Nota

Se $\text{EQM}_\theta(T(X_n), g(\theta)) \overset{n\uparrow\infty}{\to} 0, \forall\theta \in \Theta$, então

$$T(X_n) \overset{P_\theta}{\to} g(\theta),\ \forall\theta \in \Theta.$$

Usando desse recurso nos exercícios, basta mostrar que o EQM vai para zero quando n cresce.

Dica

Disso, temos que, se $T(\boldsymbol{X}_n)$ for assintóticamente eficiente, então será *fracamente* consistente.

32.2 Exemplos

32.2.1 Exemplo Bernoulli

Seja $\boldsymbol{X}_n$ amostra aleatória de $X \sim \text{Ber}(\theta), \theta \in \Theta = (0,1)$. Mostre que $T(\boldsymbol{X}_n)$ é consistente para $g(\theta) = \theta$.

32.2.1.1 Resposta

Como $E_\theta(\bar{X}) = \theta$ e $\text{Var}_\theta(\bar{X})\frac{\theta(1-\theta)}{n}, \forall\theta \in \Theta.$, temos que

$$\text{Viés}_\theta(\bar{X},\theta) = 0, \forall\theta \in \Theta,$$
$$\lim_{n\to\infty} \text{Var}_\theta(\bar{X}) = 0, \forall\theta \in \Theta,$$
$$\Rightarrow \text{EQM}_\theta(\bar{X},\theta) \to 0, \forall\theta \in \Theta \Rightarrow \bar{X} \xrightarrow{P_\theta} \theta, \forall\theta \in \Theta.$$

Dica

Poderíamos também, como $\boldsymbol{X}_n$ é uma amostra aleatória, utilizar a lei forte dos grandes números.

32.2.2 Exemplo Normal (média)

Seja $\boldsymbol{X}_n$ a.a. de $X \sim \text{N}(\mu,\sigma^2), \theta = (\mu,\sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+$. Mostre que $\bar{X}$ é consistente para $g(\theta) = \mu$.

32.2.2.1 Resposta

Como $E_\theta(\bar{X}) = \mu, \text{Var}_\theta(\bar{X}) = \frac{\sigma^2}{n}, \forall\theta \in \Theta$, temos que

$$
\begin{aligned}
\text{Viés}_\theta(\bar{X}, \mu) &= 0, \forall\theta \in \Theta, \\
\lim_{n\to\infty} \text{Var}_\theta(\bar{X}) &= 0, \forall\theta \in \Theta, \\
\Rightarrow \lim_{n\to\infty} \text{EQM}_\theta(\bar{X}, \theta) &= 0, \forall\theta \in \Theta \\
&\Rightarrow T(\boldsymbol{X}_n) = \bar{X} \text{ é fracamente consistente para } g(\theta) = \mu.
\end{aligned}
$$

32.2.3 Exemplo Normal (variância)

Seja $\boldsymbol{X}_n$ a.a. de $X \sim \text{N}(\mu, \sigma^2), \theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+$. Mostre que

$$T_k(\boldsymbol{X}_n) = \frac{1}{n-k} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

é consistente para $g(\theta) = \sigma^2$.

32.2.3.1 Resposta

Sabemos que

$$\sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{(n-1)}, \forall\theta \in \Theta.$$

Logo,

$$
\begin{aligned}
E_\theta(T_k(\boldsymbol{X}_n)) &= E_\theta\left(\frac{1}{n-k} \sum (X_i - \bar{X})^2\right) \\
&= \frac{\sigma^2}{n-k} E_\theta\left(\sum \frac{(X_i - \bar{X})^2}{\sigma^2}\right) \\
&= \frac{\sigma^2}{n-k}(n-1), \forall\theta \in \Theta.
\end{aligned}
$$

Ademais,

$$
\begin{aligned}
\text{Var}_\theta(T_k(\boldsymbol{X}_n)) &= \text{Var}_\theta\left(\frac{1}{n-k} \sum (X_i - \bar{X})^2\right) \\
&= \frac{\sigma^4}{(n-k)^2} \text{Var}_\theta\left(\sum \frac{(X_i - \bar{X})^2}{\sigma^2}\right) \\
&= \frac{\sigma^4}{(n-k)^2} 2(n-1), \forall\theta \in \Theta.
\end{aligned}
$$

Portanto,

$$\text{EQM}_\theta = \frac{2\sigma^2}{(n-k)^2}(n-1) + \left(\frac{\sigma^2}{n-k}(n-1) - \sigma^2\right)^2$$

Como

$$\lim_{n\to\infty} \frac{n-1}{n-k} = 1, \forall k \neq n, k \text{ fixado}$$

e

$$\lim_{n\to\infty} \frac{n-1}{(n-k)^2} = 0, \forall k \neq n, k \text{ fixado}$$

temos que

$$\lim_{n\to\infty} \text{EQM}_\theta(T_k(\boldsymbol{X}_n), \sigma^2) = 0, \forall \theta \in \Theta.$$

Logo, $T(\boldsymbol{X}_n)$ é consistente para $\sigma^2, \forall k \neq n,\ k$fixado

33 Escolha de Estimadores

O Erro Quadrático Médio (EQM) é o critério para escolher o estimador a ser utilizado.

Sejam $T_1(\boldsymbol{X}_n), T_2(\boldsymbol{X}_n)$ estimadores para a quantidade de interesse. Se $\text{EQM}_\theta(T_1(\boldsymbol{X}_n), g(\theta)) \leq \text{EQM}_\theta(T_2(\boldsymbol{X}_n), g(\theta))$ para cada $\theta \in \Theta$, dizemos que $T_1(\boldsymbol{X}_n)$ é melhor $T_2(\boldsymbol{X}_n)$.

Estimadores inadmissíveis

Se existir $\theta^* \in \Theta$ tal que

$$\text{EQM}_{\theta^*}(T_1(\boldsymbol{X}_n), g(\theta)) < \text{EQM}_{\theta^*}(T_2(\boldsymbol{X}_n), g(\theta))$$

dizemos que $T_2(\boldsymbol{X}_n)$ é inadmissível para $g(\theta)$.

Observação

Note que o ENVVUM é aquele que tem menor EQM dentre os estimadores não-viciados.

34 Estimador de Máxima Verossimilhança (EMV) - Aprofundamento

Seja $\mathcal{L}_{x_n}(\theta)$ a função de verossimilhança de um modelo estatístico. Dizemos que $\hat{\theta}_{\text{MV}}(X_n)$ é um estimador de máxima verossimilhança se, e somente se,

$$\mathcal{L}_{x_n}(\hat{\theta}_{\text{MV}}(x_n)) \geq \mathcal{L}_{x_n}(\theta), \text{ q.c. } \forall \theta \in \Theta.$$

em que x_n representa uma possível amostra observada.

> **Observação**
>
> $\hat{\theta}_{\text{MV}}(X_n)$ é um estimador (quase certamente, ele é único).
> $\hat{\theta}_{\text{MV}}(x_n)$ é uma estimativa.

34.1 Exemplos

34.1.1 Exemplo Uniforme

Seja X_n a.a. de $X \sim \text{U}(0, \theta), \theta > 0$. Encontre o EMV.

34.1.1.1 Resposta

A função de verossimilhança é dada por

$$\mathcal{L}_{x_n}(\theta) \overset{\text{iid}}{=} \prod_{i=1}^{n} f_\theta(x_i) = \prod \left\{ \frac{1}{\theta} \mathbb{1}_{(0,\theta]}(x_i) \right\} = \frac{1}{\theta^n} \prod \mathbb{1}_{(0,\theta]}(x_i)$$

Note que

$$\prod \mathbb{1}_{(0,\theta]}(x_i) = 1 \iff \mathbb{1}_{[0,\infty)}(\min x_n)\mathbb{1}_{(-\infty,\theta]}(\max x_n) = 1$$

Como $0 < \max x_n \leq \theta < \infty$,

$$\mathcal{L}_{x_n}(\theta) \frac{1}{\theta^n} \mathbb{1}_{[0,\infty)}(\min x_n) \mathbb{1}_{[\max x_n, \infty)}$$

Note que a função é *estritamente decrescente* para qualquer valor de $\theta \geq \max x_n$ e vale 0 para $\theta \leq \max x_n$, temos que $\theta = \max x_n$ é o ponto que maximiza a função de verossilhança, logo,

$$\hat{\theta}_{\text{MV}}(X_n) = \max X_n$$

é o estimador de máxima verossimilhança.

34.1.2 Exemplo Beta

Seja X_n a.a. de $X \sim f_\theta, \theta \in \Theta = (0, \infty)$ tal que

$$f_\theta(x) = \begin{cases} \theta x^{\theta-1}, x \in (0,1) \\ 0, \text{c.c} \end{cases}$$

Encontre o EMV para "θ".

34.1.2.1 Resposta

A função de verossimilhança é

$$\begin{aligned} \mathcal{L}_{x_n}(\theta) &\overset{\text{i.i.d.}}{=} \prod_{i=1}^{n} f_\theta(x_i) \\ &= \prod \{\theta x_i^{\theta-1}\} \\ &= \theta^n \left(\prod x_i\right)^{\theta-1} \end{aligned}$$

Como $\mathcal{L}_{x_n}$ é positiva $\forall \theta \in \Theta$, temos que,

$$\ln \mathcal{L}_{x_n}(\theta) = n \ln \theta + (\theta - 1) \sum \ln x_i$$

Ademais, $\ln(\cdot)$ é uma função estritamente crescente. Portanto, o valor que maximiza $\mathcal{L}_{x_n}(\cdot)$ é o mesmo que maximiza $\ln \mathcal{L}_{x_n}(\cdot)$.

Encontramos os pontos críticos de $\ln \mathcal{L}_{X_n}(\theta)$

$$\begin{aligned} &\frac{d \ln \mathcal{L}_{x_n}(\theta)}{d\theta} = 0 \\ &\Rightarrow \frac{n}{\theta} + \sum \ln x_i = 0 \\ &\iff \hat{\theta} = -\frac{n}{\sum \ln x_i} \text{ é um ponto crítico.} \end{aligned}$$

Para a segunda derivada,

$$\frac{d^2 \ln \mathcal{L}_{x_n}(\theta)}{d\theta^2} = -\frac{n}{\theta^2} < 0, \forall \theta \in \Theta.$$

Logo,

$$\hat{\theta}_{\mathrm{MV}}(X_n) = -\frac{n}{\sum \ln x_i}$$

é o EMV.

34.2 Caso Geral

Seja X_n a.a. de $X \sim f_\theta, \theta \in \Theta \subseteq \mathbb{R}^P$. Considere x_n uma possível amostra observada e considere as Condições $C_1 : C_4$ válidas. Então, o EMV pode ser obtido igualando a função escore a zero. Ou seja,

$$U_n(X_n, \hat{\theta}_{\mathrm{MV}}(X_n))$$

em que

$$U_n(x_n, \theta) = \frac{\partial \ln \mathcal{L}_{x_n}(\theta)}{\partial \theta}$$

desde que

$$\left.\frac{\partial U_n(x_n, \theta)}{\partial \theta^T}\right|_{\theta = \hat{\theta}_{\mathrm{MV}}(x_n)}$$

é negativa definida para quase todo x_n. Ou seja,

$$\left.\frac{\partial U_n(x_n, \theta)}{\partial \theta^T}\right|_{\theta = \hat{\theta}_{\mathrm{MV}}(x_n)} = \left.\frac{\partial^2 \ln \mathcal{L}_{x_n}}{\partial \theta \partial \theta^T}\right|_{\theta = \hat{\theta}_{\mathrm{MV}}(x_n)}$$

é negativa definida.

Matrizes positivas definidas

Dizemos que A é positiva definida se, e somente se,

$$x^T A x > 0, \forall x \in \mathbb{R}^P \setminus \{0\}$$

Teorema
A simétrica ($A = A^T$) é positiva definida se, e somente se, todos seus autovalores são positivos.

Matrizes negativas definidas

Dizemos que A é positiva definida se, e somente se,

$$x^T A x < 0, \forall x \in \mathbb{R}^P \setminus \{0\}$$

Teorema
A simétrica $(A = A^T)$ é negativa definida se, e somente se, todos seus autovalores são negativos.

Dica

Uma matriz A de dimensões 2×2 é dita positiva(negativa) definida se, e somente se, os elementos da diagonal são positivos(negativos) e o determinante de A é positivo.

34.2.1 Exemplos

34.2.1.1 Exemplo Normal

Seja X_n a.a. de $X \sim \mathrm{N}(\mu, \sigma^2), \theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}_+$

34.2.1.1.1 Resposta

A função de verossimilhança é

$$\begin{aligned} \mathcal{L}_{x_n}(\theta) &= \frac{1}{(\sqrt{2\pi\sigma^2})^n} \mathrm{e}^{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2} \\ \Rightarrow \ln \mathcal{L}_{x_n}(\theta) &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2. \end{aligned}$$

Logo,

$$U_n(x_n, \theta) = \frac{\partial \ln}{\partial \theta} \mathcal{L}_{x_n}(\theta) = \begin{pmatrix} \frac{\partial \ln}{\partial \mu} \mathcal{L}_{x_n}(\theta) \\ \frac{\partial \ln}{\partial \sigma^2} \mathcal{L}_{x_n}(\theta) \end{pmatrix}$$

em que

$$\frac{\partial \ln}{\partial \mu} \mathcal{L}_{x_n}(\theta) = -\frac{1}{2\sigma^2}\sum 2(x_i-\mu)(-1) = \frac{1}{\sigma^2}\sum(x_i-\mu)$$

e

$$\frac{\partial \ln}{\partial \sigma^2} \mathcal{L}_{x_n}(\theta) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum(x_i-\mu)^2$$

Logo,

$$
\begin{aligned}
U_n(x_n, \theta) = 0 \\
&\Longleftrightarrow \begin{cases} \frac{\partial \ln}{\partial \mu} \mathcal{L}_{x_n}(\theta) = \frac{1}{\sigma^2} \sum (x_i - \mu) & = 0 \\ \frac{\partial \ln}{\partial \sigma^2} \mathcal{L}_{x_n}(\theta) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (x_i - \mu)^2 & = 0 \end{cases} \\
&\Longleftrightarrow \begin{cases} \sum x_i - n\mu = 0 \\ -n + \frac{1}{\sigma^2} \sum (x_i - \mu)^2 = 0 \end{cases} \\
&\Longleftrightarrow \begin{cases} \mu = \sum \frac{x_i}{n} = 0 \\ \sigma^2 = \sum \frac{(x_i - \mu)^2}{n} \end{cases} \\
&\Rightarrow \begin{cases} \hat{\mu} = \bar{x} \\ \widehat{\sigma^2} = \frac{1}{n} \sum (x_i - \bar{x})^2 \end{cases}
\end{aligned}
$$

Ou seja, $\hat{\theta} = (\hat{\mu}, \widehat{\sigma^2})$ é o ponto crítico de $\ln \mathcal{L}_{x_n}(\theta)$.

Continuando,

$$\begin{aligned}
\frac{\partial^2 \ln}{\partial\theta\partial\theta^T}\mathcal{L}_{x_n}(\theta) &= \frac{\partial U_n(x_n,\theta)}{\partial\theta^T} \\
&= \frac{\partial}{\partial\theta^T}\left[\begin{array}{c} \frac{\partial \ln}{\partial\mu}\mathcal{L}_{x_n}(\theta) \\ \frac{\partial \ln}{\partial\sigma^2}\mathcal{L}_{x_n}(\theta) \end{array}\right] \\
&= \left[\begin{array}{cc} \frac{\partial^2 \ln}{\partial\mu^2}\mathcal{L}_{x_n}(\theta) & \frac{\partial^2 \ln}{\partial\mu\partial\sigma^2}\mathcal{L}_{x_n}(\theta) \\ \frac{\partial^2 \ln}{\partial\mu\partial\sigma^2}\mathcal{L}_{x_n}(\theta) & \frac{\partial^2 \ln}{\partial(\sigma^2)^2}\mathcal{L}_{x_n}(\theta) \end{array}\right] \\
\Rightarrow \frac{\partial^2 \ln}{\partial\mu^2}\mathcal{L}_{x_n}(\theta) &= -\frac{n}{\sigma^2} \\
\frac{\partial^2 \ln}{\partial\mu\partial\sigma^2}\mathcal{L}_{x_n}(\theta) &= -\frac{1}{\sigma^4}\sum(x_i-\mu) \\
\frac{\partial^2 \ln}{\partial(\sigma^2)^2}\mathcal{L}_{x_n}(\theta) &= \frac{n}{2\sigma^4} + \frac{1}{2}\sum(x_i-\mu)^2\cdot\frac{\partial(\sigma^2)^2}{\partial\sigma^2} \\
&= \frac{n}{2\sigma^4} - \sum\frac{(x_i-\mu)^2}{\sigma^6} \\
\Rightarrow \left.\frac{\partial^2 \ln}{\partial\theta\partial\theta^T}\mathcal{L}_{x_n}(\theta)\right|_{\theta=\hat{\theta}} &= \left[\begin{array}{cc} -\frac{n}{\hat{\sigma}^2} & -\frac{1}{(\hat{\sigma}^2)^2}\sum(x_i-\bar{x}) \\ -\frac{1}{(\hat{\sigma}^2)^2}\sum(x_i-\bar{x}) & \frac{n}{2(\hat{\sigma}^2)^2} - \sum\frac{(x_i-\bar{x})^2}{(\hat{\sigma}^2)^3} \end{array}\right]
\end{aligned}$$

Note que $\sum(x_i-\bar{x}) = \sum x_i - n\bar{x} = \sum x_i - \sum x_i = 0$

e que $\dfrac{n}{2(\hat{\sigma}^2)^2} - \dfrac{n\hat{\sigma}^2}{(\hat{\sigma}^2)^3} = -\dfrac{n}{2(\hat{\sigma}^2)^2}$

$$\Rightarrow \left.\frac{\partial^2 \ln}{\partial\theta\partial\theta^T}\mathcal{L}_{x_n}(\theta)\right|_{\theta=\hat{\theta}} = \left[\begin{array}{cc} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & \frac{-n}{(\hat{\sigma}^2)^2} \end{array}\right]$$

desde que $\sum(x_i-\bar{x})^2 > 0$.

Como $\left.\frac{\partial^2 \ln}{\partial\theta\partial\theta^T}\mathcal{L}_{x_n}(\theta)\right|_{\theta=\hat{\theta}}$ é negativa definida, concluímos que $\hat{\theta}$ é pelo menos máximo local. Dizemos que $\hat{\theta}_{\mathrm{MV}}(X_n) = (\bar{X}, S_n^2)$ é o EMV para $\theta = (\mu,\sigma^2)$, em que

$$S_n^2 = \frac{1}{n}\sum(X_i-\bar{X})^2$$

34.3 Relação com a Informação de Fisher

Note que a informação de Fisher total sob o modelo normal anterior é

$$
\begin{aligned}
I_n(\theta) &= -E_\theta\left(\frac{\partial^2 \ln}{\partial\theta\partial\theta^T}\mathcal{L}_{x_n}(\theta)\right)\\
&= -E_\theta\begin{pmatrix} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^4}\sum(X_i-\mu) \\ -\frac{1}{\sigma^4}\sum(X_i-\mu) & \frac{n}{2\sigma^4}-\sum\frac{(X_i-\mu)^2}{(\sigma^2)^3}\end{pmatrix}\\
&= E_\theta\begin{pmatrix} \frac{n}{\sigma^2} & \frac{1}{\sigma^4}\sum(X_i-\mu) \\ \frac{1}{\sigma^4}\sum(X_i-\mu) & -\frac{n}{2\sigma^4}+\sum\frac{(X_i-\mu)^2}{(\sigma^2)^3}\end{pmatrix}
\end{aligned}
$$

Note que $E_\theta[(X_i-\mu)^2] = \mathrm{Var}_\theta(X) = \sigma^2$

$$
\Rightarrow I_n(\theta) = E_\theta\begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4}\end{pmatrix}.
$$

34.4 Alguns exemplos

Seja $\boldsymbol{X}_n$ a.a. de $X \sim f_\theta$,

1. $X \sim \mathrm{Ber}(\theta), \theta \in (0,1)$, o EMV para "$\theta$" é $\widehat{\theta}_{\mathrm{MV}}(X_n) = \bar{X}$.
2. $X \sim \mathrm{Bin}(4,\theta), \theta \in (0,1)$, o EMV para "$\theta$" é $\widehat{\theta}_{\mathrm{MV}}(X_n) = \frac{\bar{X}}{4}$.
3. $X \sim \mathrm{Poiss}(\theta), \theta > 0$, o EMV para "$\theta$" é $\widehat{\theta}_{\mathrm{MV}}(X_n) = \bar{X}$.
4. $X \sim \mathrm{Exp}(\theta), \theta > 0$, o EMV para "$\theta$" é $\widehat{\theta}_{\mathrm{MV}}(X_n) = \frac{1}{\bar{X}}$.
5. $X \sim f_\theta, \theta > 0$, tal que

$$
f_\theta(x) = \begin{cases} \frac{1}{\Gamma(\theta)}x^{\theta-1}\mathrm{e}^{-x}, & x > 0 \\ 0, & \text{c.c.}\end{cases}
$$

com $\Gamma(\theta) = \int x^{\theta-1}\mathrm{e}^{-x}dx$, temos

$$\begin{aligned}
\mathcal{L}_{\boldsymbol{x}_n}(\theta) &\overset{\text{i.i.d.}}{=} \prod f_\theta(x_i) \\
&= \prod \left\{ \frac{1}{\Gamma(\theta)} x_i^{\theta-1}\mathrm{e}^{-x_i} \right\} \\
&= \frac{1}{\Gamma(\theta)^n} \left(\prod x_i\right)^{\theta-1} \mathrm{e}^{-\sum x_i} \\
\Rightarrow \ln \mathcal{L}_{\boldsymbol{x}_n}(\theta) &= -n \ln \Gamma(\theta) + (\theta - 1)\sum \ln x_i - \sum x_i
\end{aligned}$$

$$\Rightarrow \frac{d \ln}{d\theta}\mathcal{L}_{\boldsymbol{x}_n}(\theta) = 0 \Rightarrow -\frac{n}{\Gamma(\theta)}\Gamma'(\theta) + \sum \ln x_i = 0.$$

Como resolver:

$$\sum \ln x_i = n\frac{\Gamma'(\theta)}{\Gamma(\theta)}.$$

Posteriormente, encontraremos um método para encontrar essa solução.

6. $X \sim f_\theta, \theta \in \Theta$, em que f_θ pertence à Família Exponencial k dimensional, isto é,

$$f_\theta(x) = \begin{cases} \mathrm{e}^{\sum_{j=1}^{k} c_j(\theta)T_j(x) + d(\theta) + S(x)}, & x \in \mathfrak{X} \\ 0, & x \notin \mathfrak{X} \end{cases}$$

A função verossimilhança para os valores observados $x_i \in \mathfrak{X}$ é dada por

$$\begin{aligned}
\mathcal{L}_{\boldsymbol{x}_n}(\theta) &\overset{\text{i.i.d.}}{=} \prod \left\{ \mathrm{e}^{\sum_{j=1}^{k} c_j(\theta)T_j(x_i) + d(\theta) + S(x_i)} \right\} \\
&= \mathrm{Exp}\left\{ \sum_{j=1}^{k} c_j(\theta) \sum_{i=1}^{n} T_j(x_i) + nd(\theta) + \sum_{i=1}^{n} S(x_i) \right\} \\
\Rightarrow \ln \mathcal{L}_{\boldsymbol{x}_n}(\theta) &= \sum_{j=1}^{k} c_j(\theta) \sum_{i=1}^{n} T_j(x_i) + nd(\theta) + \sum_{i=1}^{n} S(x_i)
\end{aligned}$$

Se $c_1, \ldots, c_k$ forem diferenciáveis com respeito a "θ", temos que

$$\frac{\partial \ln}{\partial \theta}\mathcal{L}_{\boldsymbol{x}_n}(\theta) = \sum_{j=1}^{k} \frac{\partial c_j(\theta}{\partial \theta} \sum_{i=1}^{n} T_j(x_i) + n\frac{\partial d(\theta)}{\partial \theta}$$

igualando a zero, temos que

$$\sum_{j=1}^{k} \frac{\partial c_j(\theta)}{\partial \theta} \sum_{i=1}^{n} T(x_i) = -n\frac{\partial d(\theta)}{\partial \theta} \tag{34.1}$$

Em alguns casos, (**??**) não tem solução fechada e utilizaremos um método numério para encontrar as soluções.

35 Estimação pelo método de momentos (EMM)

Seja X_n uma a.a. de X tal que

$$E_\theta(|X|^k) < \infty, k \in \{1, \dots, p\}, \Theta \subseteq \mathbb{R}^P$$

O estimador obtido pelo método de momentos é aquele que satisfaz

$$E_\theta(X^k) = \frac{1}{n}\sum_{i=1}^{n} X_i^k, k = 1, \dots, p.$$

Note que a estimação pelo método de momentos *não* utiliza toda a informação contida na função de verossimilhança. Precisamos conhecer a forma dos primeiros p momentos.

35.1 Exemplos

35.1.1 Exemplo 1

Seja X_n a.a. de $X \sim U(0, \theta), \theta > 0$. Encontre o estimador pelo método de momentos.

Calcule o EQM do EMV e do estimador obtido pelo método de momentos.

35.1.1.1 Resposta

Note que $\Theta = (0, \infty) \subseteq \mathbb{R}$. Portanto, $p = 1$. Precisamos encontrar o valor de "θ" que satisfaz

$$E_\theta(X) = \frac{1}{n}\sum X_i$$

Sabemos que

$$E_\theta(X) = \int_0^\theta x\frac{1}{\theta}dx = \frac{\theta}{2} = \bar{X}$$

Logo, $\hat{\theta}_{\text{MM}}(X_n) = 2\bar{X}$ é o estimador obtido pelo método de momentos.

Sabemos que $\widehat{\theta}_{\text{MV}}(\boldsymbol{X}_n) = \max\{X_1, \dots, X_n\}$, logo,

$$\begin{aligned}\text{EQM}_\theta(\widehat{\theta}_{\text{MM}}(\boldsymbol{X}_n), \theta) &= E_\theta((\widehat{\theta}_{\text{MM}}(\boldsymbol{X}_n) - \theta)^2) \\ &= \text{Var}_\theta(\widehat{\theta}_{\text{MM}}(\boldsymbol{X}_n)) + \text{Viés}(\widehat{\theta}_{\text{MM}}(\boldsymbol{X}_n), \theta)^2\end{aligned}$$

$$\begin{aligned}E_\theta(\widehat{\theta}_{\text{MM}}(\boldsymbol{X}_n)) &= 2E_\theta(X) = 2\frac{\theta}{2} = \theta \\ \Rightarrow \text{Viés}_\theta(\widehat{\theta}_{\text{MV}}(\boldsymbol{X}_n), \theta) &= 0, \forall \theta \in \Theta.\end{aligned}$$

$$\begin{aligned}\text{Var}_\theta(\widehat{\theta}_{\text{MM}}(\boldsymbol{X}_n)) &= \text{Var}_\theta(2\bar{X}) \overset{\text{i.i.d.}}{=} 4\frac{\text{Var}_\theta(X)}{n} \overset{\text{Unif}}{=} \frac{\theta^2}{3n} \\ \Rightarrow \text{EQM}_\theta(\widehat{\theta}_{\text{MM}}(\boldsymbol{X}_n), \theta) &= \frac{\theta^2}{3n}\end{aligned}$$

Para o EMV,

$$\text{EQM}_\theta(\widehat{\theta}_{\text{MV}}(\boldsymbol{X}_n), \theta) = \text{Var}_\theta(\widehat{\theta}_{\text{MV}}(\boldsymbol{X}_n)) + \text{Viés}(\widehat{\theta}_{\text{MV}}(\boldsymbol{X}_n), \theta)^2$$

Precisaremos encontrar a distribuição do estimador,

$$P_\theta(\widehat{\theta}_{\text{MV}}(\boldsymbol{X}_n) \leq t) = P_\theta(\max \boldsymbol{X}_n \leq t) \overset{\text{iid}}{=} \prod P_\theta(X \leq t) = P_\theta(X \leq t)^n$$

note que

$$P_\theta(X \leq t) = \begin{cases} 0, & t \leq 0 \\ \frac{t}{\theta}, & 0 < t \leq \theta \\ 1, & t \geq \theta \end{cases} \Rightarrow P_\theta(\widehat{\theta}_{\text{MV}}(\boldsymbol{X}_n) \leq t)^n = \begin{cases} 0, & t \leq 0 \\ \left(\frac{t}{\theta}\right)^n, & 0 < t \leq \theta \\ 1, & t \geq \theta \end{cases}$$

Como $P_\theta(\widehat{\theta}_{\text{MV}}(\boldsymbol{X}_n))$ é absoulatemente contínua para todo $\theta > 0$, temos que $\widehat{\theta}_{\text{MV}}(\boldsymbol{X}_n)$ é uma variável aleatória contínua cuja f.d.p. é dada por

$$f_\theta^{\widehat{\theta}_{\text{MV}}(\boldsymbol{X}_n)}(x) = \frac{d}{dt} P_\theta(\widehat{\theta}_{\text{MV}}(\boldsymbol{X}_n) \leq t)\Big|_{t=x} = \begin{cases} \frac{nx^{n-1}}{\theta^n}, & 0 < x \leq \theta \\ 0, & \text{c.c.} \end{cases}.$$

Logo,

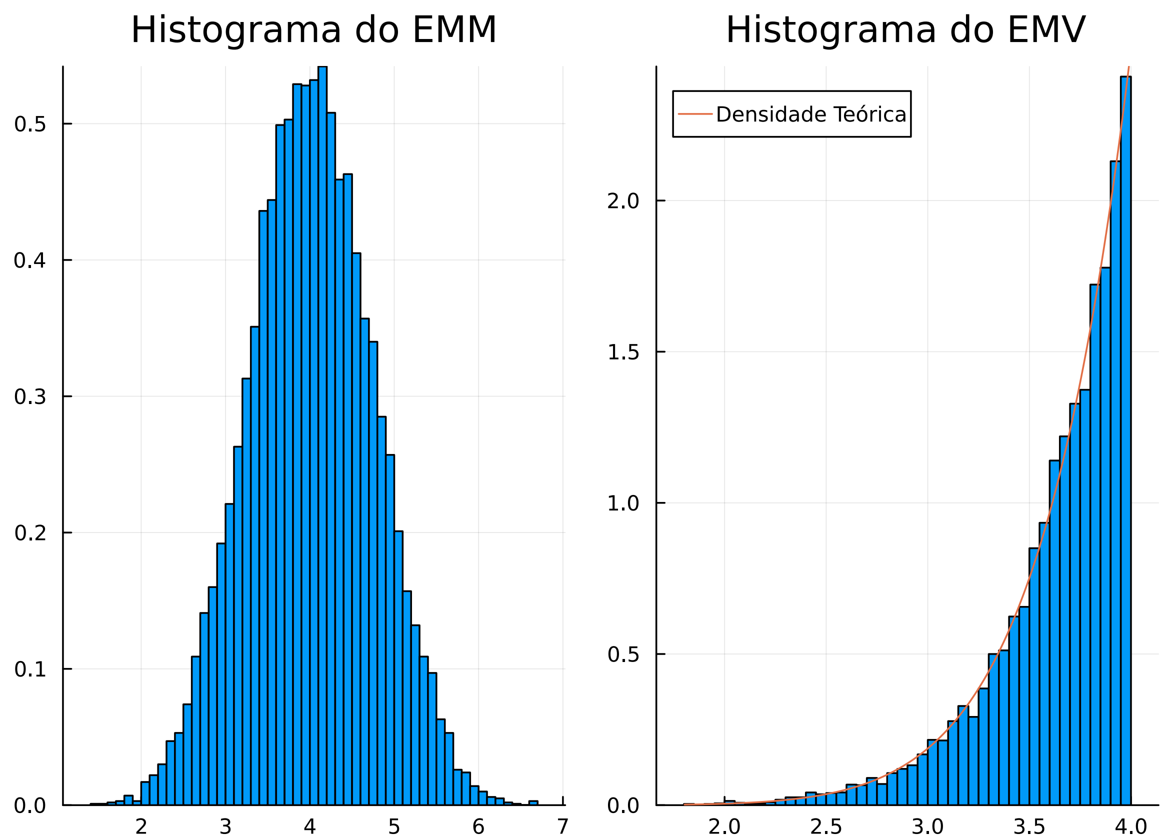
$$
\begin{aligned}
E_\theta(\hat{\theta}_{\text{MV}}(\boldsymbol{X}_n)) &= \int_{-\infty}^{\infty} w f_\theta^{\hat{\theta}_{\text{MV}}(\boldsymbol{X}_n)}(w)dw \\
&= \int_0^\theta w \frac{nw^{n-1}}{\theta^n} dw \\
&= \frac{n}{\theta^n}\int_0^\theta w^n dw = \frac{n}{\theta^n}\frac{w^{n+1}}{n+1}\bigg|_0^\theta = \frac{n}{n+1}\theta, \forall\theta\in\Theta, \\
\Rightarrow E_\theta(\hat{\theta}_{\text{MV}}(\boldsymbol{X}_n)^2) &= \frac{n\theta^2}{n+2}, \forall\theta\in\Theta, \\
\Rightarrow \text{Var}_\theta(\hat{\theta}_{\text{MV}}(\boldsymbol{X}_n)) &= \frac{n\theta^2}{n+2} - \left(\frac{n}{n+1}\right)^2\theta^2 \\
\Rightarrow \text{EQM}_\theta(\hat{\theta}_{\text{MV}}(\boldsymbol{X}_n),\theta) &= \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 + \left(\frac{n}{n+1}\theta - \theta\right)^2 \\
&= \theta^2\left(\frac{n}{n+2} - \frac{2}{n+1} + 1\right)
\end{aligned}
$$

35.1.1.2 Simulação

```julia
using Random, Distributions, StatsBase, Plots, LaTeXStrings

Random.seed!(76)

M = 10_000
n = 10
theta0 = rand(Poisson(4)) + 1

d = Uniform(0, theta0)

thetaMM = zeros(M)
thetaMV = zeros(M)
for i in 1:M
  x = rand(d, n)
  thetaMM[i] = 2*mean(x)
  thetaMV[i] = maximum(x)
end

# Comparar
println("Média simulada do EMM = $(mean(thetaMM)),
        calculada = $theta0")
```

```
println("Média simulada do EMV = $(mean(thetaMV)),
        calculada = $(10/11 * theta0)")
println("Variância simulada do EMM = $(var(thetaMM)),
        calculada = $(theta0^2 /(3*n))")
println("Variância simuladado EMV = $(var(thetaMV)),
        calculada = $(n * theta0^2 / (n+2) - (n/(n+1))^2 * theta0^2)")

densidade(x) = 0 < x < theta0 ? n * x^(n-1) / theta0^n : 0

pmm = histogram(thetaMM, label = "",
                title = "Histograma do EMM", normalize=:pdf)
pmv = histogram(thetaMV, label = "",
                title = "Histograma do EMV", normalize=:pdf)
plot!(minimum(thetaMV):0.001:maximum(thetaMV)-0.01, densidade,
      label="Densidade Teórica")
display(plot(pmm, pmv))
```

```
Média simulada do EMM = 4.012626175744703,
        calculada = 4
Média simulada do EMV = 3.6375235806062722,
        calculada = 3.6363636363636362
Variância simulada do EMM = 0.5384935090654293,
        calculada = 0.5333333333333333
Variância simuladado EMV = 0.11001021153100962,
        calculada = 0.11019283746556674
```



35.1.2 Exemplos adicionais

Seja X_n a.a. de $X \sim f_\theta, \theta \in \Theta$. Encontre o estimador pelo método de momentos para os casos abaixo:

1. $X \sim \text{Ber}(\theta), \theta \in (0,1)$
2. $X \sim \text{Bin}(m, \theta), \theta \in (0,1)$
3. $X \sim \text{Poiss}(\theta), \theta \in (0, \infty)$
4. $X \sim \text{N}(\mu, \sigma^2), \theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$
5. $X \sim \text{Exp}(\theta), \theta > 0$
6. $X \sim \text{Gamma}(\alpha, \beta), \theta = (\alpha, \beta) \in \mathbb{R}_+ \times \mathbb{R}_+$

35.1.2.1 Respostas

1. $E_\theta(X)=\theta$. Portanto, $\hat{\theta}_{\text{MM}}(X_n)=\bar{X}$.
2. $E_\theta(X)=m\theta$. Portanto, $\hat{\theta}_{\text{MM}}(X_n)=\frac{\bar{X}}{m}$.
3. $E_\theta(X)=\theta$. Portanto, $\hat{\theta}_{\text{MM}}(X_n)=\bar{X}$.
4. $E_\theta(X)=\theta, E_\theta(X^2)=\sigma^2+\mu^2$. Note que

$$
\begin{aligned}
&\begin{cases} \mu & =\bar{X} \\ \sigma^2+\mu^2 & =\frac{1}{n}\sum X_i^2 \end{cases} \\
\Rightarrow &\begin{cases} \hat{\mu}_{\text{MM}}(X_n) & =\bar{X} \\ \hat{\sigma^2}_{\text{MM}}(X_n)+\hat{\mu}_{\text{MM}}(X_n)^2 & =\frac{1}{n}\sum X_i^2 \end{cases} \\
\Rightarrow &\begin{cases} \hat{\mu}_{\text{MM}}(X_n) & =\bar{X} \\ \hat{\sigma^2}_{\text{MM}}(X_n)=\frac{1}{n}\sum X_i^2-\bar{X}^2 \end{cases} \\
\Rightarrow &\hat{\theta}_{\text{MM}}(X_n)=\left(\bar{X},\frac{1}{n}\sum X_i^2-\bar{X}^2\right)
\end{aligned}
$$

5. $E_\theta(X)=\frac{1}{\theta}$. Portanto, $\hat{\theta}_{\text{MM}}(X_n)=\frac{1}{\bar{X}}$
6.

> **Distribuição Gama**
>
> Se $X\sim\text{Gama}(\alpha,\beta)$, então
>
> $$\begin{cases} f_\theta^X(x)=\frac{\beta^\alpha x^{\alpha-1}\mathrm{e}^{-\beta x}}{\Gamma(\alpha)}, & x>0 \\ 0, & \text{c.c.} \end{cases}$$

Logo,

$$\begin{aligned} E_\theta(X^k) &= \int_0^\infty x^k \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\beta x} dx \\ &= \int_0^\infty \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha+k-1} \mathrm{e}^{-\beta x} dx \end{aligned}$$

$$y = \beta x, dy = \beta dx, x = \frac{y}{\beta}$$

$$\begin{aligned} \Rightarrow E_\theta(X^k) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{y}{\beta}\right)^{\alpha+k-1} \mathrm{e}^{-y} \frac{dy}{\beta} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{\beta^{\alpha+k}} \int_0^\infty y^{\alpha+k-1} \mathrm{e}^{-y} dy \\ &= \frac{1}{\Gamma(\alpha)\beta^k} \Gamma(\alpha+k) \\ \Rightarrow E_\theta(X) &= \frac{1}{\Gamma(\alpha)\beta} \Gamma(\alpha+1) = \frac{\alpha\Gamma(\alpha)}{\Gamma(\alpha)\beta} = \frac{\alpha}{\beta} \\ \Rightarrow E_\theta(X^2) &= \frac{1}{\Gamma(\alpha)\beta^2} \Gamma(\alpha+2) \\ &= \frac{(\alpha+1)\Gamma(\alpha+1)}{\Gamma(\alpha)\beta} = \frac{(\alpha+1)\alpha\Gamma(\alpha)}{\Gamma(\alpha)\beta^2} = \frac{\alpha(\alpha+1)}{\beta^2} \end{aligned}$$

Seguindo para as equações de momentos,

$$
\begin{aligned}
&\begin{cases} \frac{\alpha}{\beta} = \bar{X} \\ \frac{\alpha(\alpha+1)}{\beta^2} = \frac{1}{n}\sum X_i^2 \end{cases} \\
\Rightarrow &\begin{cases} \alpha = \bar{X}\beta \\ \frac{\bar{X}\beta(\bar{X}\beta+1)}{\beta^2} = \frac{1}{n}\sum X_i^2 \end{cases} \\
\Rightarrow &\begin{cases} \alpha = \bar{X}\beta \\ (\bar{X}\beta)^2 + \bar{X}\beta = \beta^2 \frac{1}{n}\sum X_i^2 \end{cases} \\
\Rightarrow &\begin{cases} \alpha = \bar{X}\beta \\ \beta^2\left(\frac{1}{n}\sum X_i^2\right) - (\beta\bar{X})^2 - \beta\bar{X} = 0 \end{cases} \\
\Rightarrow &\begin{cases} \alpha = \bar{X}\beta \\ \beta^2 \underbrace{\left(\frac{1}{n}\sum X_i^2 - \bar{X}^2\right)}_{S^2} - \beta\bar{X} = 0 \end{cases}
\end{aligned}
$$

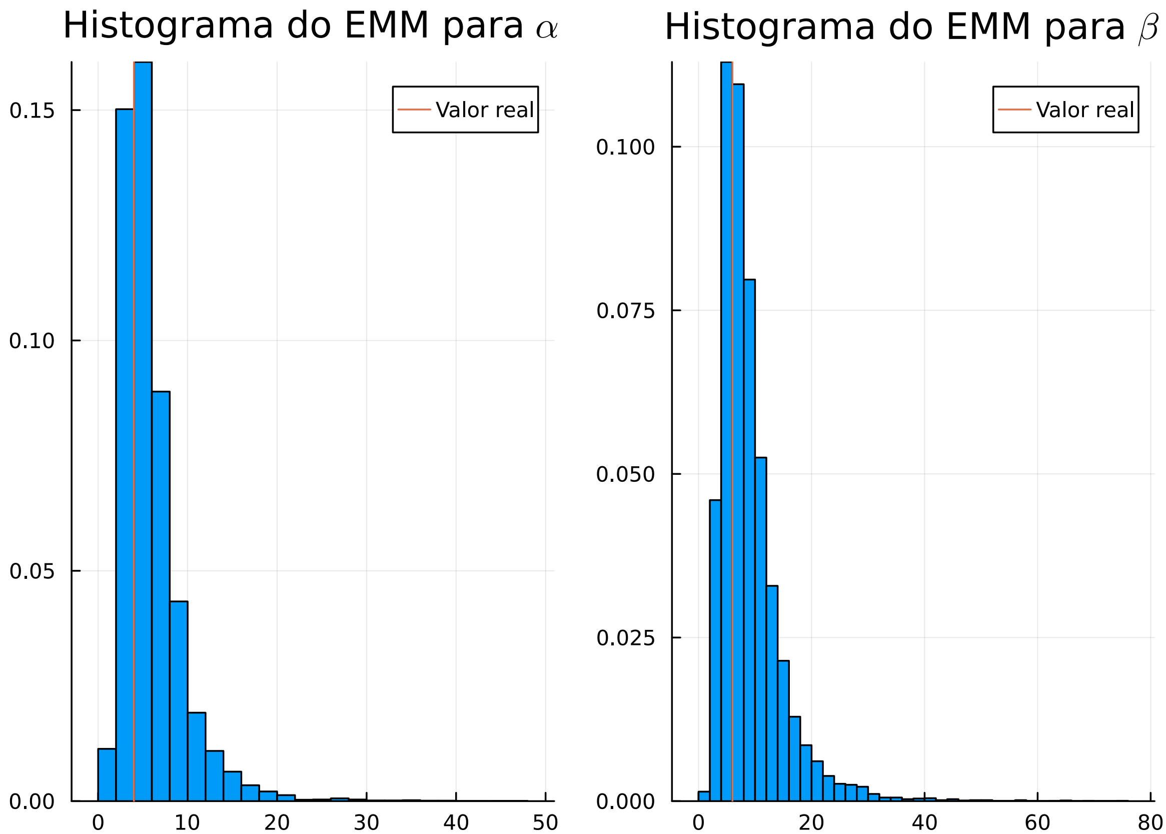
$$
\begin{aligned}
\frac{\bar{X} \pm \sqrt{\bar{X}^2 - 4 \cdot S^2 \cdot 0}}{2S^2} &= \frac{\bar{X} \pm \sqrt{\bar{X}^2}}{2S^2} \\
&= \begin{cases} 0 \\ \frac{\bar{X}}{S^2} \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow &\begin{cases} \alpha = \bar{X}\beta \\ \beta^2 S^2 = \beta\bar{X} \end{cases} \\
\Rightarrow &\begin{cases} \hat{\alpha}_{\text{MM}}(X_n) = \frac{\bar{X}^2}{S^2} \\ \hat{\beta}_{\text{MM}}(X_n) = \frac{\bar{X}}{S^2} \end{cases} \\
\Rightarrow &\hat{\theta}_{\text{MM}}(X_n) = \left(\frac{\bar{X}^2}{S^2}, \frac{\bar{X}}{S^2}\right).
\end{aligned}
$$

```julia
using Random, Distributions, StatsBase, Plots, LaTeXStrings

Random.seed!(31)

M = 10_000
n = 10 # Tamanho da amostra
alpha0 = 4
beta0 = 6

d = Gamma(alpha0, 1/beta0)

alphaMM = zeros(M)
betaMM = zeros(M)
```

```
for i in 1:M
  x = rand(d, n)
  s = var(x, corrected = false)
  alphaMM[i] = mean(x)^2/s
  betaMM[i] = mean(x)/s
end

# Comparar
println("Média simulada do EMM para alpha = $(mean(alphaMM)),
        real = $alpha0")
println("Média simulada do EMM para beta = $(mean(betaMM)),
        real = $beta0")

# Plotar
pmmalpha = histogram(alphaMM, label = "",
                title = "Histograma do EMM para " * L"\alpha", normalize=:pdf,
                     bins = 40)
vline!([alpha0], label="Valor real")
pmmbeta = histogram(betaMM, label = "",
                title = "Histograma do EMM para " * L"\beta", normalize=:pdf,
                    bins = 40)
vline!([beta0], label="Valor real")
display(plot(pmmalpha, pmmbeta))
```

```
Média simulada do EMM para alpha = 5.838811157394312,
        real = 4
Média simulada do EMM para beta = 8.994281492594121,
        real = 6
```



Histograma do EMM para α
Valor real
0.15
0.10
0.05
0.00
0
10
20
30
40
50
Histograma do EMM para β
Valor real
0.100
0.075
0.050
0.025
0.000
0
20
40
60
80

36 Método de Newton-Raphson

Nem sempre existem soluções fechadas para estimadores de máxima verossimilhança ou obtidos pelo método de momentos. Por exemplo, $U_n(X_n, \theta) = 0$ não tem solução fechada para o EMV e $E_\theta(X^k)\frac{1}{n}\sum X_i^k$ para o EMM.

Nos dois casos, estamos interessados em encontrar os valores de "θ" que "zeram" uma função $G(\theta)$,

1. $G(\theta) = U_n(X_n, \theta)$

2. $$G(\theta) = \begin{pmatrix} E_\theta(X) - \frac{1}{n}\sum X_i \\ \vdots \\ E_\theta(X^p) - \frac{1}{n}\sum X_i^p \end{pmatrix}$$

Consideraremos a seguir apenas o caso $p = 1$

Seja $G : \Theta \to \mathbb{R}$ uma função contínua e diferenciável, com derivadas contínuas, tal que $G'(\theta) \neq 0, \forall \theta \in \Theta$. Temos como objetivo encontrar $\hat{\theta} \in \Theta$ tal que $G(\hat{\theta}) = 0$.

Inicia-se o processo com um valor qualquer $\theta_0 \in \Theta$ e calculamos o valor seguinte por meio de:

$$\theta^{(j)} = \theta^{(j-1)} - \frac{G(\theta^{(j-1)})}{G'(\theta^{(j-1)})}, j = 1, 2, ...$$

Continua-se até que $|\theta^{(j)} - \theta^{(j-1)}| \leq \epsilon$ em que ϵ é um valor de erro pequeno fixado.

36.1 Exemplo

Quando tentamos encontrar o EMV para θ de $X \sim f_\theta, \theta \in \Theta$ com

$$f_\theta(x) = \begin{cases} \frac{1}{\Gamma(\theta)} x^{\theta-1} \mathrm{e}^{-x}, & x > 0 \\ 0, & \text{c.c.} \end{cases}$$

não conseguimos finalizar a expressão. Vamos relembrar o escore da amostra igualado a zero:

$$U_n(x_n, \theta) = \sum \ln x_i - \frac{n}{\Gamma(\theta)}\Gamma'(\theta) + \sum \ln x_i = 0$$

Considere os $n = 7$ valores observados, $x_7 = (3.1, 4.2, 5.7, 2.3, 7.7, 5.1, 3.5)$

$$\Rightarrow U_n(x_n, \theta) = \underbrace{\frac{7}{\Gamma(\theta)}\Gamma'(\theta)}_{\frac{\partial \ln}{\partial \theta}\Gamma(\theta)} + \sum \ln x_i = 0$$

Utilizando $\theta^{(0)} = 1$,

$$\begin{aligned}
\theta^{(j)} &= \theta^{(j-1)} - \frac{U_n(x_n, \theta^{(j-1)})}{U_n'(x_n, \theta^{(j-1)})} \\
&= \theta^{(j-1)} - \frac{-7\psi_1(\theta^{(j-1)}) + 7\sum \ln x_i}{-7\psi_2(\theta^{(j-1)})} \\
&= \theta^{(j-1)} - \frac{-\psi_1(\theta^{(j-1)}) + \sum \ln x_i}{-\psi_2(\theta^{(j-1)})}
\end{aligned}$$

em que $\psi_j(\theta) = \frac{\partial^j \ln \Gamma(\theta)}{\partial \theta^j}$

36.2 Algoritmos e outros métodos

Seja X_n a.a. de $X \sim \Gamma(\theta, 1)$ (exemplo anterior). Ainda considerando os mesmos valores amostrados, $x_7 = (3.1, 4.2, 5.7, 2.3, 7.7, 5.1, 3.5)$, utilize os métodos numéricos para encontrar a estimativa de máxima verossimilhança.

36.2.1 Algoritmo por Newton-Raphson

Considere um erro $\epsilon = 10^{-5}$ e $\hat{\theta}^{(0)} = 1$.

$$\theta^{(j+1)} = \theta^{(j)} - \frac{U_n(x_n, \theta^{(j)})}{U_n'(x_n, \theta^{(j)})}$$

em que $U_n(x_n, \theta) = -n\psi_1(\theta^{(j-1)}) + \sum \ln x_i$ e $U_n'(x_n, \theta) = -n\psi_2(\theta^{(j-1)})$ com $\psi_j(\theta) = \frac{\partial^j \ln \Gamma(\theta)}{\partial \theta^j}$ e $n = 7$.

Logo,

$$\theta^{(j+1)} = \theta^{(j)} - \frac{U_n(x_n, \theta^{(j)})}{U_n'(x_n, \theta^{(j)})}$$

36.2.1.1 Julia

```julia
using SpecialFunctions # Funções di e trigamma (derivadas do log da gama)


function main()

  x = [3.1, 4.2, 5.7, 2.3, 7.7, 5.1, 3.5]
  n = length(x)

  theta::Vector{Float64} = []
  theta0 = 1
  append!(theta, theta0)


  erromax = 10^(-5)
  erro = Inf
  i = 1
  # iteracoesMax = 6 # Podemos também definir apenas um erro máximo
  while erro > erromax # && i < iteracoesMax
    append!(theta, theta[i] - (sum(log.(x)) - n * digamma(theta[i]))/
                                        (-n*trigamma(theta[i])))
    erro = abs(theta[i+1] - theta[i])
    println("Erro na iteração $i: $erro")
    i += 1
  end
  println("Theta final: $(theta[length(theta)])")
  println("Total de iterações: $i")

end

main()
```

```
Erro na iteração 1: 1.2248511275743903
Erro na iteração 2: 1.5558203731733413
Erro na iteração 3: 0.8122385697386227
Erro na iteração 4: 0.10638335531193821
Erro na iteração 5: 0.0013809916201203976
Erro na iteração 6: 2.2502724394968254e-7
Theta final: 4.700674642445657
Total de iterações: 7
```

36.2.1.2 Python

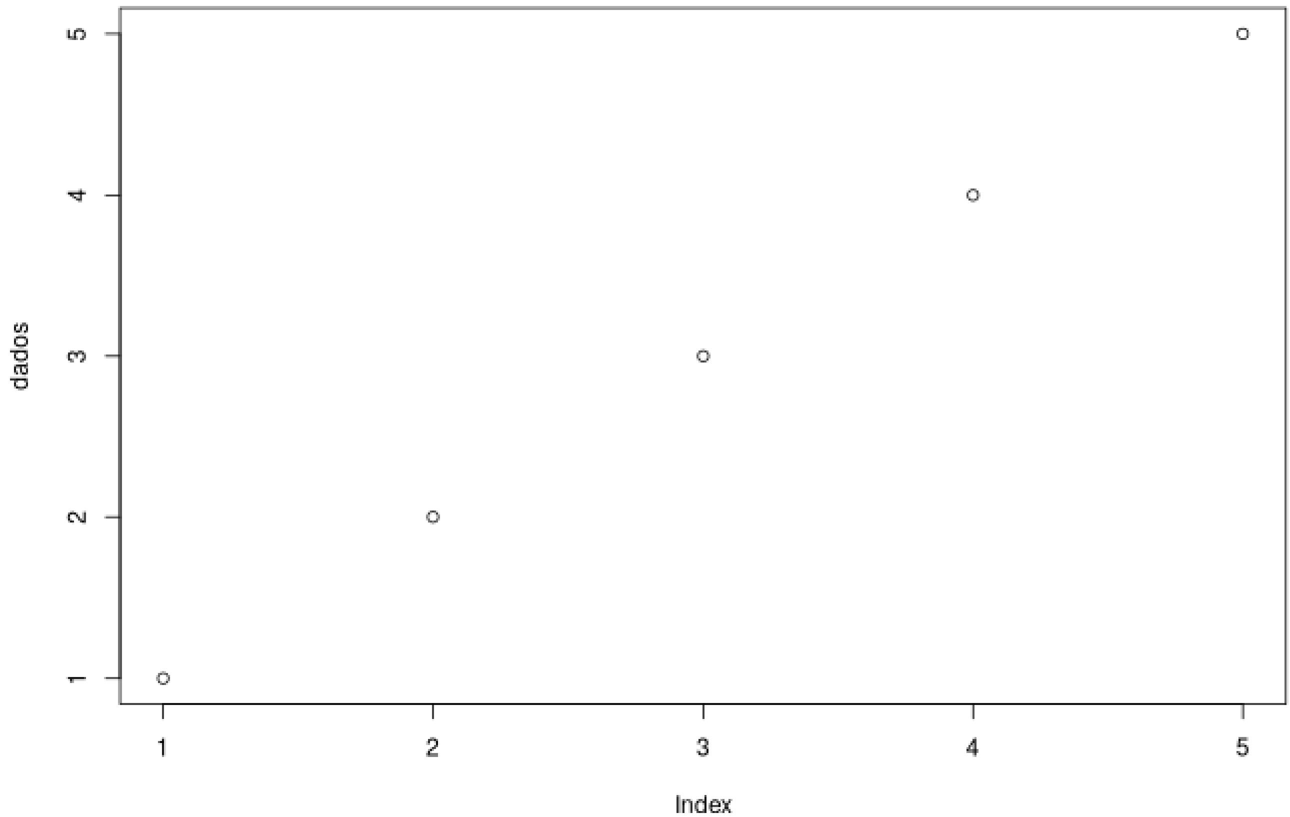
```python
print("Teste Python")
print(1+1)
```

```
Teste Python
2
```

36.2.1.3 R

```r
cat("Teste R")
dados<-c(1,2,3,4,5)
plot(dados)
```

```
Teste R
```



Note que, se $U_n'(\boldsymbol{x}_n, \widehat{\theta}^{(j+1)}) < 0$, a estimativa $\widehat{\theta}^{(j+1)}$ será a de MV.

Seja X_n a.a. de $X \sim \text{Exp}(\theta), \theta \in \Theta = (0, \infty)$. Encontre a estimativa de MV para θ pelo método de Newton-Raphson considerando $x_n = (2, 2.5, 3, 3.5, 2, 1.5)$ e $\widehat{\theta}^{(0)} = 0.5$ Com erro máximo $\epsilon = 10^{-5}$.

Note que

$$\begin{aligned}
f_\theta(x) &= \theta \mathrm{e}^{-\theta x} \\
\mathcal{L}_{x_n}(\theta) &= \theta^n \mathrm{e}^{-\theta \sum x_i} \\
\ln \mathcal{L}_{x_n}(\theta) &= n \ln \theta - \theta \sum x_i \\
U_n(x_n, \theta) &= \frac{n}{\theta} - \sum x_i \\
U_n'(x_n, \theta) &= -\frac{n}{\theta^2}
\end{aligned}$$

Logo

$$
\begin{aligned}
\widehat{\theta}^{(j+1)} &= \widehat{\theta}^{(j)} - \frac{\frac{n}{\widehat{\theta}^{(j)}} - \sum x_i}{-\frac{n}{(\widehat{\theta}^{(j)})^2}} \\
&= \widehat{\theta}^{(j)} - \frac{[n - \widehat{\theta}^{(j)} \sum x_i]\widehat{\theta}^{(j)}}{-n}
\end{aligned}
$$

```
function main()

  x = [2, 2.5, 3, 3.5, 2, 1.5]
  n = length(x)

  theta::Vector{Float64} = []
  theta0 = 0.5
  append!(theta, theta0)


  erromax = 10^(-5)
  erro = Inf
  i = 1
  while erro > erromax
    append!(theta, theta[i] - ((n - sum(x) *theta[i])*theta[i])/(-n))
    erro = abs(theta[i+1] - theta[i])
    println("Erro na iteração $i: $erro")
    i += 1
  end
  println("Theta final: $(theta[length(theta)])")
  println("Total de iterações: $i")

end

main()
```

```
Erro na iteração 1: 0.10416666666666669
Erro na iteração 2: 0.01718026620370372
Erro na iteração 3: 0.0007780354808986645
Erro na iteração 4: 1.468425129103057e-6
Theta final: 0.4137931034430648
Total de iterações: 5
```

Note que se $\Theta \subseteq \mathbb{R}^P$, o método é dado por

$$
\theta^{(j+1)} = \theta^{(j)} - \left[U_n'(x_n, \theta^{(j)})\right]^{-1} \cdot U_n(x_n, \widehat{\theta}^{(j)}), j = 0, 1, ...
$$

36.2.2 Método de Scoring de Fisher

Para $\Theta \subseteq \mathbb{R}$

$$\theta^{(j+1)} = \theta^{(j)} + \frac{U_n(x_n, \theta^{(j)})}{I_n(\theta^{(j)})}$$

Para $\Theta \subseteq \mathbb{R}^P$

$$\theta^{(j+1)} = \theta^{(j)} + \left[I_n(x_n, \theta^{(j)})\right]^{-1} \cdot U_n(x_n, \hat{\theta}^{(j)}), j = 0, 1, ...$$

36.2.3 Método de descida do gradiente

$$\theta^{(j+1)} = \theta^{(j)} + \delta \cdot U_n(x_n, \hat{\theta}^{(j)}), j = 0, 1, ...$$

em que $\delta \in (0, 1)$ é a *taxa de "aprendizado"*.

Para o exemplo da Gama,

```
using SpecialFunctions # Funções di e trigamma (derivadas do log da gama)


function main()

  x = [3.1, 4.2, 5.7, 2.3, 7.7, 5.1, 3.5]
  n = length(x)

  theta::Vector{Float64} = []
  theta0 = 1
  append!(theta, theta0)
  delta = 0.01


  erromax = 10^(-5)
  erro = Inf
  i = 1
  while erro > erromax
    append!(theta, theta[i] + delta *(sum(log.(x)) - n * digamma(theta[i])))
    erro = abs(theta[i+1] - theta[i])
    if i % 100 == 0
      println("Erro na iteração $i: $erro")
    end
    i += 1
  end
  println("Theta final: $(theta[end])")
```

```
  println("Total de iterações: $i")

end

main()
```

```
Erro na iteração 100: 0.007983394465004956
Erro na iteração 200: 0.0013675177845966502
Erro na iteração 300: 0.0002527165874308679
Erro na iteração 400: 4.731072948604975e-5
Theta final: 4.700082915774669
Total de iterações: 494
```

36.2.4 Descida estocástica do gradiente

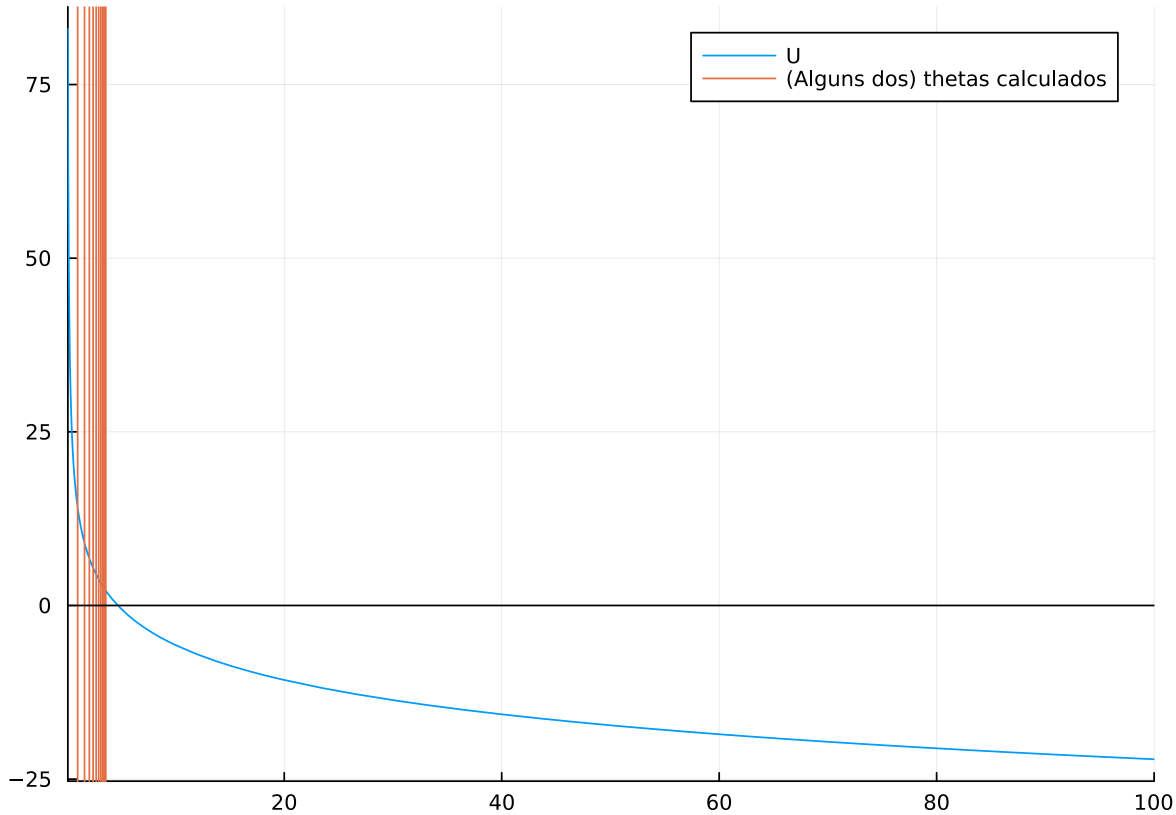
$$\theta^{(j+1,i)} = \theta^{(j,i)} + \delta \cdot U(x_i, \widehat{\theta}^{(j,i)}), j = 0, 1, \dots, i \in \{1, \dots, n\}$$

Em que $U(x, \widehat{\theta}^{(j)})$ é a função escore para uma observação. Usaremos o mesmo modelo gama.

```
using SpecialFunctions, Distributions, Plots

function descida_estocastica()

  x = [3.1, 4.2, 5.7, 2.3, 7.7, 5.1, 3.5]
  n = length(x)

  theta::Vector{Float64} = []
  append!(theta, 1)
  delta = 0.0001


  erro = Inf
  erromax = 10^(-5)
  j = 1
  epoca = 1
  while erro > erromax
    for i in 1:n
      append!(theta, theta[j] + delta * (log(x[i]) - digamma(theta[j])))
      erro = abs(theta[j+1] - theta[j])
      j += 1
    end
```

```
    epoca += 1
    if epoca % 500 == 0
      println("Erro na época $epoca: $erro")
    end
  end
  epoca -= 1
  println("Theta final: $(theta[end])")
  println("Total de epocas: $epoca \n \n")

  println("Verificando se é máximo local")
  println("-n * trigamma(theta calculado) = $(-n * trigamma(theta[end]))")

  U(y) = sum(log.(x)) - n*digamma(y)
  p = plot(U, xlims=(0.1, 100), label="U")
  vline!(theta[1:4000:end],label="(Alguns dos) thetas calculados")
  hline!([0],label="", color=:black)
  display(p)
end

descida_estocastica()
```

```
Erro na época 500: 0.00011562567419964864
Erro na época 1000: 8.455622767256088e-5
Erro na época 1500: 6.523785153733641e-5
Erro na época 2000: 5.164846886707153e-5
Erro na época 2500: 4.1416753598255696e-5
Erro na época 3000: 3.337361869748534e-5
Erro na época 3500: 2.685988163664277e-5
Erro na época 4000: 2.146877836084471e-5
Erro na época 4500: 1.6932425193516565e-5
Erro na época 5000: 1.306548867541224e-5
Theta final: 3.6540367477421754
Total de epocas: 5457


Verificando se é máximo local
-n * trigamma(theta calculado) = -2.201394773690926
```

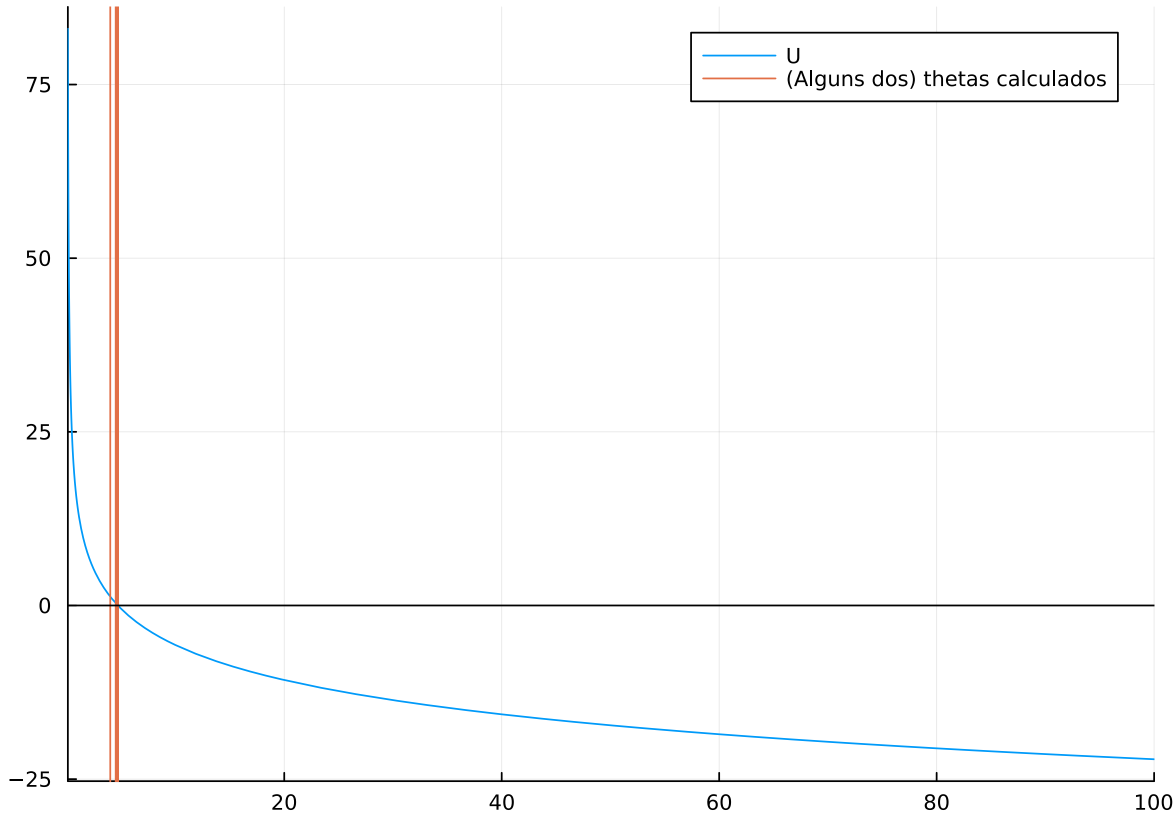


Note que esse método é falível. Caímos em um valor que não é máximo global e diferente dos obtidos pelos outros métodos. Aumentaremos o valor de δ e adicionaremos um máximo de 5000 épocas. Também escolheremos um valor mais alto para θ_0, como 4.

```
using SpecialFunctions, Distributions, Plots

function descida_estocastica_limitada()

  x = [3.1, 4.2, 5.7, 2.3, 7.7, 5.1, 3.5]
  n = length(x)

  theta::Vector{Float64} = []
  append!(theta, 4)
  delta = 0.01


  erro = Inf
  erromax = 10^(-5)
```

```
  j = 1
  epoca = 1
  while erro > erromax && epoca <= 5000
    for i in 1:n
      append!(theta, theta[j] + delta * (log(x[i]) - digamma(theta[j])))
      erro = abs(theta[j+1] - theta[j])
      j += 1
    end
    epoca += 1
    if epoca % 1000 == 0
      println("Erro na época $epoca: $erro")
    end
  end
  epoca -= 1
  println("Theta final: $(theta[end])")
  println("Total de epocas: $epoca \n \n")

  println("Verificando se é máximo local")
  println("-n * trigamma(theta calculado) = $(-n * trigamma(theta[end]))")

  U(y) = sum(log.(x)) - n*digamma(y)
  p = plot(U, xlims=(0.1, 100), label="U")
  vline!(theta[1:500:end],label="(Alguns dos) thetas calculados")
  hline!([0],label="", color=:black)
  display(p)
end

descida_estocastica_limitada()
```

```
Erro na época 1000: 0.0018561619919994499
Erro na época 2000: 0.001856162087572777
Erro na época 3000: 0.001856162087572777
Erro na época 4000: 0.001856162087572777
Erro na época 5000: 0.001856162087572777
Theta final: 4.70217698128991
Total de epocas: 5000


Verificando se é máximo local
-n * trigamma(theta calculado) = -1.6580912861780042
```



U
(Alguns dos) thetas calculados
75
50
25
0
−25
20
40
60
80
100

37 Condições de Regularidade

$C_1 : \mathfrak{X} = \{x : f_\theta(x) > 0\}$ não depende de "θ".

$C_2 : f_\theta$ é duas vezes diferenciável com respeito a "θ" e suas derivadas são contínuas.

C_3 : é possível trocar as derivadas pela integrais da seguinte forma:

$$\begin{cases} a) \ \dfrac{\partial}{\partial \theta} \displaystyle\int f_\theta(x)dx & = \displaystyle\int \dfrac{\partial}{\partial \theta} f_\theta(x)dx \\ b) \ \dfrac{\partial^2}{\partial \theta \partial \theta^T} \displaystyle\int f_\theta(x)dx & = \displaystyle\int \dfrac{\partial^2}{\partial \theta \partial \theta^T} f_\theta(x)dx \end{cases}$$

C_4 :

$$\begin{cases} \text{a) } E_\theta\left(\|\dfrac{\partial}{\partial \theta} \ln f_\theta(x)\|\right) < \infty, \ \forall \theta \in \Theta \\ \text{b) } E_\theta\left(\|\dfrac{\partial^2}{\partial \theta \partial \theta^T} \ln f_\theta(x)\|^2\right) < \infty, \ \forall \theta \in \Theta. \end{cases}$$

Para as duas últimas condições, substituímos no caso discreto as integrais por somatórios.

C_5: Θ é aberto e convexo, ou seja, se $\theta_1, \theta_2 \in \Theta$, então

$$\lambda \theta_1 + (1-\lambda)\theta_2 \in \Theta, \forall \lambda \in (0,1)$$

A condição (fraca) de identificabilidade C_6:

$$E_{\theta_1}(U_n(X_n, \theta_2)) = 0 \iff \theta_1 = \theta_2,$$

C_7: A informação de Fisher existe, é positiva (definida) e finita.

C_8:

$$\sup_{\theta_2 \in \Theta} \left| M(\theta, \theta_2) - \frac{1}{n} |U_n(X_n, \theta_2)| \right| \overset{P_\theta}{\to} 0$$

$$\forall \theta \in \Theta, M(\theta, \theta_2) = \frac{1}{n} \left| E_\theta(U_n(X_n, \theta_2)) \right|$$

Consistência e C_8

C_8 não é necessária se $\hat{\theta}_{\mathrm{MV}}(X_n)$ for consistente, ou seja,

$$\hat{\theta}_{\mathrm{MV}}(X_n) \overset{P_\theta}{\to} \theta, \forall \theta \in \Theta.$$

C_9:

$$\frac{1}{n}U_n'(X_n, \theta_n^*) - \frac{1}{n}U_n'(X_n, \theta) \overset{P_\theta}{\to} 0, \forall \theta \in \Theta.$$

em que θ_n^* é um estimador consistente qualquer para "θ".

Suficientes mas não necessárias

Nem toda condição de regularidade é necessária para provar algum teorema. Em muitos casos, por exemplo, podemos utilizar apenas as condições $C_1 : C_4$, enquanto resultados mais fortes podem precisar de mais condições.

38 Resultados utilizando o Teorema de Slutsky

Nota **??**: Método Delta

Seja $T(X_n)$ um estimador para "θ", $\theta \in \Theta \subseteq \mathbb{R}$, assintoticamente normal, ou seja,

$$\sqrt{n}(T(X_n) - \theta) \xrightarrow{\mathcal{D}} N(0, \mathrm{V}_\theta), \forall \theta \in \Theta.$$

Considere $g : \Theta \to \mathbb{R}$ uma função diferenciável com derivada contínua tal que $g'(\theta) \neq 0, \forall \theta \in \Theta$. Então,

$$\sqrt{n}(g(T(X_n)) - g(\theta)) \xrightarrow{\mathcal{D}} N(0, g'(\theta)^2 V_\theta), \forall \theta \in \Theta.$$

Seja

$$\frac{\sqrt{n}(T(X_n) - \theta)}{\sqrt{\mathrm{V}_\theta}} \xrightarrow{\mathcal{D}} N(0, 1)$$

Pelo teorema de Slutsky,

$$\frac{\sqrt{n}(T(X_n) - \theta)}{\sqrt{\mathrm{V}_{T(X_n)}}} \xrightarrow{\mathcal{D}} N(0, 1)$$

Ademais, seja

$$\frac{\sqrt{n}(g(T(X_n)) - g(\theta))}{\sqrt{g'(\theta)^2 \mathrm{V}_\theta}} \xrightarrow{\mathcal{D}} N(0, 1)$$

Pelo teorema de Slutsky,

$$\frac{\sqrt{n}(g(T(X_n)) - g(\theta))}{\sqrt{g'(T(X_n))^2 \mathrm{V}_{T(X_n)}}} \xrightarrow{\mathcal{D}} N(0, 1)$$

39 Propriedades dos estimadores de máxima verossimilhança e de momentos

39.1 Teorema (da invariância do EMV)

Seja X_n a.a. de $X \sim f_\theta, \theta \in \Theta$. Considere $g : \Theta \to \mathbb{R}^k$ uma função. Se existe o EMV para "θ", então $g(\hat{\theta}_{\mathrm{MV}}(X_n))$ é o EMV para $g(\theta)$.

39.1.1 Exemplo Bernoulli

$X \sim \mathrm{Ber}(\theta), \theta \in (0,1)$. $\hat{\theta}_{\mathrm{MV}}(X_n) = \bar{X}$.

Se $g(\theta) = \theta^2$, então $g(\bar{X}) = \bar{X}^2$ é o EMV para θ^2.

Se $g(\theta) = 1 - \theta$, então $g(\bar{X}) = 1 - \bar{X}$ é o EMV para $1 - \theta$.

39.1.2 Exemplo Poisson

$X \sim \mathrm{Pois}(\theta), \theta > 0$. $\hat{\theta}_{\mathrm{MV}}(X_n) = \bar{X}$.

Se $g(\theta) = P_\theta(X \geq 3)$, então

$$\begin{aligned} g(\hat{\theta}_{\mathrm{MV}}(X_n)) &= P_{\hat{\theta}_{\mathrm{MV}}(X_n)}(X \geq 3) \\ &= 1 - \left(+ \mathrm{e}^{-\hat{\theta}_{\mathrm{MV}}(X_n)} + \mathrm{e}^{-\hat{\theta}_{\mathrm{MV}}(X_n)} \hat{\theta}_{\mathrm{MV}}(X_n) + \mathrm{e}^{-\hat{\theta}_{\mathrm{MV}}(X_n)} \frac{\hat{\theta}_{\mathrm{MV}}(X_n)^2}{2!} \right) \end{aligned}$$

Se $g(\theta) = E_\theta(X^2)$, então $g(\hat{\theta}_{\mathrm{MV}}(X_n)) = E_{\hat{\theta}_{\mathrm{MV}}(X_n)}(X^2) = \hat{\theta}_{\mathrm{MV}}(X_n) + (\hat{\theta}_{\mathrm{MV}}(X_n))^2$

Se $g(\theta) = \frac{\sqrt{\mathrm{Var}_\theta(X)}}{E_\theta(X)} = \frac{\sqrt{\theta}}{\theta} = \frac{1}{\sqrt{\theta}}$, então $g(\hat{\theta}_{\mathrm{MV}}(X_n)) = \frac{1}{\sqrt{\hat{\theta}_{\mathrm{MV}}(X_n)}}$ é o EMV para $g(\theta)$.

39.2 Teorema (dos estimadores assintoticamente normais)

Dizemos que $T(X_n)$ é um estimador para "θ" assintoticamente normal se, e somente se, existir V_θ não negativo tal que

$$\sqrt{n}(T(X_n) - \theta) \xrightarrow{\mathcal{D}} N(0, \mathrm{V}_\theta), \forall \theta \in \Theta.$$

ou seja, converge em distribuição para uma distribuição Normal de média 0 e variância V_θ para todo θ no espaço paramétrico.

39.3 Teorema (do limite central para o EMV)

Seja X_n a.a. de $X \sim f_\theta$ tal que as condições de regularidade $C_1 : C_9$ estejam satisfeitas.

Portanto, com g diferenciável com derivada contínua tal que $g'(\theta) \neq 0$,

$$\sqrt{n}(g(\hat{\theta}_{\mathrm{MV}}(X_n)) - g(\theta)) \xrightarrow{\mathcal{D}} N(0, g'(\theta)^2 I_1(\theta)^{-1}), \forall \theta \in \Theta$$

Logo, pelo teorema de Slutsky

$$\sqrt{n} I_1(\hat{\theta}_{\mathrm{MV}}) \frac{(g(\hat{\theta}_{\mathrm{MV}}(X_n)) - g(\theta))}{\sqrt{g'(\hat{\theta}_{\mathrm{MV}})^2}} \xrightarrow{\mathcal{D}} N(0, 1), \forall \theta \in \Theta$$

EMVs são assintoticamente eficientes

Note que

$$E_\theta(\hat{\theta}_{\mathrm{MV}}(X_n)) \overset{n\uparrow\infty}{\rightarrow} \theta$$

$$n\mathrm{Var}_\theta(\hat{\theta}_{\mathrm{MV}}(X_n)) \overset{n\uparrow\infty}{\rightarrow} I_1(\theta)^{-1}$$

Ou seja, estimadores de máxima verossimilhança são assintoticamente eficientes sob as condições de regularidade.
A variância assintótica do estimador de máxima verossimilhança é denotada por

$$\mathrm{Var}_\theta^{(a)}(\hat{\theta}_{\mathrm{MV}}(X_n)) = \frac{I_1(\theta)^{-1}}{n}$$

Notação

$$\begin{cases} \hat{\theta}_{\text{MV}}(X_n) & \overset{a}{\approx} N(\theta, I_n(\theta)) \\ g(\hat{\theta}_{\text{MV}}(X_n)) & \overset{a}{\approx} N(g(\theta), g'(\theta)^2 I_n(\theta)^{-1}) \end{cases}$$

Podemos calcular probabilidades aproximadas mesmo sem saber a distribuição exata, para $\hat{\theta}_{\text{MV}}$

39.3.1 Exemplo

Se $\theta = 3$, então calcule:

$$\begin{aligned} P_\theta(\hat{\theta}_{\text{MV}}(X_n) \leq t) &= P_\theta(\hat{\theta}_{\text{MV}}(X_n) - \theta \leq t - \theta) \\ &= P_\theta\left(\frac{\hat{\theta}_{\text{MV}}(X_n) - \theta}{\sqrt{I_n(\theta)^{-1}}} \leq \frac{t-\theta}{\sqrt{I_n(\theta)^{-1}}}\right) \\ &\overset{n \text{ grande}}{\approx} P_\theta(N(0,1) \leq I_n(\theta)^{\frac{1}{2}}(t-\theta)) \\ &= P_\theta(N(0,1) \leq I_n(\theta)^{\frac{1}{2}}(t-3)) \end{aligned}$$

39.4 Teorema (do limite central para o estimador do método de momentos)

Seja X_n a.a. de X tal que $\Theta \subseteq \mathbb{R}$ e $E_\theta(|X|^k) < \infty$ para algum $k \in \mathbb{N} \setminus \{0\}$. Então, o estimador $\hat{\theta}_{\text{MM}}(X_n)$ tal que

$$E_{\hat{\theta}_{\text{MM}}(X_n)}(X^k) = \frac{1}{n} \sum X_i^k$$

é assintoticamente normal se as seguintes condições estiverem satisfeitas:

1. $E_\theta[(X^k - E_\theta(X^k))^2] < \infty$
2. $h(\theta) = \frac{\partial E_\theta(X^k)}{\partial \theta} \neq 0, \forall \theta \in \Theta$.

Com isso,

$$\sqrt{n}(\hat{\theta}_{\text{MM}}(X_n) - \theta) \xrightarrow{\mathcal{D}} N(0, V_\theta)$$

em que

$$V_\theta = \frac{E_\theta[(X^k - E_\theta(X^k))^2]}{\left(\frac{\partial E_\theta(X^k)}{\partial\theta}\right)^2} = \frac{\text{Var}_\theta(X^k)}{h(\theta)^2}$$

Além disso, se $g : \Theta \to \mathbb{R}$ for diferenciável com derivada contínua tal que $g'(\theta) \neq 0$, então

$$\sqrt{n}(g(\hat{\theta}_{\text{MM}}(X_n) - \theta) - g(\theta) \xrightarrow{\mathcal{D}} N\left(0, g'(\theta)^2 \frac{\text{Var}_\theta(X^k)}{h(\theta)^2}\right)$$

Notação

$$\begin{cases} \hat{\theta}_{\text{MM}}(X_n) & \overset{a}{\approx} N\left(\theta, \frac{\text{Var}_\theta(X^k)}{nh(\theta)^2}\right) \\ g(\hat{\theta}_{\text{MM}}(X_n)) & \overset{a}{\approx} N\left(g(\theta), g'(\theta)^2 \frac{\text{Var}_\theta(X^k)}{nh(\theta)^2}\right) \end{cases}$$

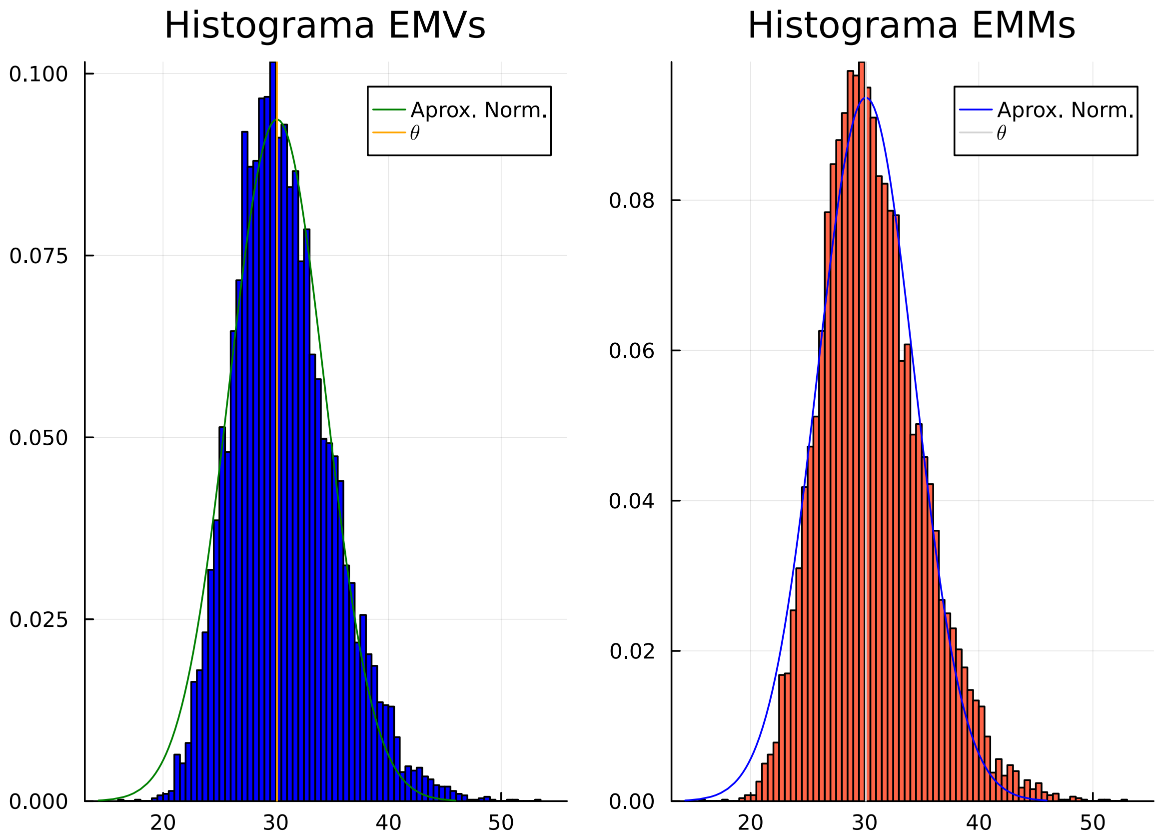
39.5 Algoritmos

Seja $X \sim \text{Beta}(\theta, 1)$.

```
using Distributions, Random, Plots, StatsPlots, StatsBase, LaTeXStrings



Random.seed!(10)
n = 50
M = 10_000

MVs = zeros(M)
MMs = zeros(M)
MV(obs) = -n/sum(log.(obs))
MM(obs) = mean(obs)/(1-mean(obs))
theta0 = rand(Uniform(0,100))
dsim = Beta(theta0, 1)
for i in 1:M
  xsim = rand(dsim, n)
  MVs[i] = MV(xsim)
  MMs[i] = MM(xsim)
end



VMV(theta) = theta^2/n
```

```
normMV = Normal(theta0, sqrt(VMV(theta0)))
VMM(theta) = theta*(theta+1)^2/((theta+2)*n)
normMM = Normal(theta0, sqrt(VMM(theta0)))

histmv = histogram(MVs, normalize=:pdf, label="", color=:blue,
                   title="Histograma EMVs")
plot!(normMV, label="Aprox. Norm.", color=:green)
vline!([theta0], label=L"\theta", color=:orange)
histmm = histogram(MMs, normalize=:pdf, label="", color=:tomato,
                   title="Histograma EMMs")
plot!(normMM, label="Aprox. Norm.", color=:blue)
vline!([theta0], label=L"\theta", color=:lightgrey)
p = plot(histmv, histmm)

display(p)
qMM = quantile(normMM, 0.25)
qMV = quantile(normMV, 0.25)
println("theta0: $theta0")
println("Viés simulado MM: $(mean(MMs)-theta0)")
println("EQM simulado MM: $(n *mean((MMs .- theta0).^2))")
println("Viés simulado MV: $(mean(MVs) - theta0)")
println("EQM simulado MV: $(n * mean((MVs .- theta0).^2))")
println("P(NormalMM > $(qMM)) = $(round(ccdf(normMM, qMM), digits=2))")
println("P(NormalMV > $(qMV)) = $(round(ccdf(normMM, qMV),digits=2))")
println("Real MM > 2.5 = $(mean(MMs .> qMM))")
println("Real MV > 2.5 = $(mean(MVs .> qMV))")
```

```
theta0: 30.104498710537307
Viés simulado MM: 0.5943591309955814
EQM simulado MM: 968.056873503618
Viés simulado MV: 0.612391481120099
EQM simulado MV: 968.182881983385
P(NormalMM > 27.2314280176146) = 0.75
P(NormalMV > 27.23291320813612) = 0.75
Real MM > 2.5 = 0.783
Real MV > 2.5 = 0.7843
```



Histograma EMVs
Aprox. Norm.
θ
0.100
0.075
0.050
0.025
0.000
20
30
40
50
Histograma EMMs
Aprox. Norm.
θ
0.08
0.06
0.04
0.02
0.00
20
30
40
50

40 Normal Multivariada

Dizemos que $X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ tem distribuição multivariada $N_d(\mu, \Sigma)$ se, e somente se, a sua função densidade de probabilidade é dada por:

$$f^X(x) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \text{Exp}\left\{ -\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right\}$$

em que $\mu = \begin{pmatrix} \mu_1, \vdots, \mu_n \end{pmatrix}$ e $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \cdots & \sigma_{dd-1} & \sigma_{dd} \end{bmatrix}$ é uma matriz simétrica positiva definida, ou seja, $y^T \Sigma y > 0, \forall y \in \mathbb{R}^d \setminus \{0\}$

Notação

$$X \sim N_d(\mu, \Sigma)$$

Obs 1: Caso particular: Normal univariada

Se $d = 1$, então $X = X_1$ e $X_1 \sim N_1(\mu_1, \sigma_{11})$

Obs 2: Esperança e (co)variância

Pode-se mostrar que:

$$E(X) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_d) \end{pmatrix} = \mu$$

e

$$\text{Var}(X) = E((X-\mu)(X-\mu)^T) = \Sigma$$

ou seja,

$$\text{Cov}(X_i, X_j) = \sigma_{i,j}, i, j = 1, \dots, d$$

Obs 3: Covariância implica independência para normais

Se

$$\Sigma = \text{diag}(\sigma_{11} \dots \sigma_{dd}) = \begin{bmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \sigma_{dd} \end{bmatrix}$$

então $X_1, \dots, X_d$ são independentes.
Ou seja, se $X_1, \dots, X_d$ forem não correlacionadas (covariância zero), então serão independentes. Em outras distribuições multivariadas, isso não é certo.

Obs 4: Caso de independência

Se $\Sigma = \text{diag}(\sigma_{11} \dots \sigma_{dd})$, então

$$|\Sigma| = \prod_{i=1}^{d} \sigma_{ii}, \Sigma^{-1} = \begin{bmatrix} \sigma_{11}^{-1} & 0 & \dots & 0 \\ 0 & \sigma_{22}^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \sigma_{dd}^{-1} \end{bmatrix}$$

e

$$\begin{aligned}
(x-\mu)^T \Sigma^{-1} (x-\mu) &= \begin{pmatrix} x_1 - \mu_1 \\ \vdots \\ x_d - \mu_d \end{pmatrix}^T \begin{bmatrix} \sigma_{11}^{-1} & 0 & \dots & 0 \\ 0 & \sigma_{22}^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \sigma_{dd}^{-1} \end{bmatrix} \begin{pmatrix} x_1 - \mu_1 \\ \vdots \\ x_d - \mu_d \end{pmatrix} \\
&= \left(\frac{(x_1 - \mu_1}{\sigma_{11}} \dots \frac{x_d - \mu_d}{\sigma_{dd}} \right) \begin{pmatrix} x_1 - \mu_1 \\ \vdots \\ x_d - \mu_d \end{pmatrix} \\
&= \frac{(x_1 - \mu_1)^2}{\sigma_{11}} + \dots + \frac{(x_d - \mu_d)^2}{\sigma_{dd}}
\end{aligned}$$

Portanto, a f.d.p de X quando $\Sigma = \text{diag}(\sigma_{11} \dots \sigma_{dd})$ é

$$\begin{aligned}
f^X(x) &= \frac{1}{(2\pi)^{\frac{d}{2}} \left(\prod \sigma_{ii}\right)^{\frac{1}{2}}} \text{Exp}\left\{ -\frac{1}{2} \sum^{d} \frac{(x_i - \mu_i)^2}{(\sigma_{ii})} \right\} \\
&= \prod \left\{ \frac{1}{(2\pi)^{\frac{d}{2}} \sigma_{ii}^{\frac{1}{2}}} \text{Exp}\left\{ -\frac{1}{2} \frac{(x_i - \mu_i)^2}{(\sigma_{ii})} \right\} \right\} \\
&= \prod_{i=1}^{d} f^{(i)}(x_i)
\end{aligned}$$

em que $f^{(i)}(x_i)$ é a f.d.p de $N(\mu_i, \sigma_{ii})$

Obs 5: Função geradora de momentos

A função geradora de momentos de $X \sim N_d(\mu\Sigma)$ é dada por

$$M_X(t) = \text{Exp}\left\{\mu^T t + \frac{1}{2}t^T\Sigma t\right\}, \forall t \in \mathbb{R}^d$$

Obs 6: Transformações para normais generalizadas

Se $Y = a + BX$, em que $a \in \mathbb{R}^m$ e B é uma matriz $m \times d$, em que $m \leq d$ cujas linhas são linearmente independentes, então,

$$Y \sim N_m(a + B\mu, B\Sigma B^T).$$

Prova por F.G.M:

$$\begin{aligned}
M_Y(t &= E(\mathrm{e}^{Y^T t})\\
&= E(\mathrm{e}^{a^T t + X^T B^T t})\\
&= \mathrm{e}^{a^T t}E(\mathrm{e}^{X^T B^T t})\\
&= \mathrm{e}^{a^T t}M_X(B^T t)\\
&= \mathrm{e}^{a^T t}\mathrm{e}^{\mu(B^T t) + \frac{1}{2}t^T B\Sigma B^T t}\\
&= \mathrm{e}^{(a+B\mu)^T t + \frac{1}{2}t^T B\Sigma B^T t}
\end{aligned}$$

Observações adicionais:
Se tomarmos $B = \Sigma^{-\frac{1}{2}}$, então $B \cdot B = \Sigma^{-1}$ e $B\Sigma B^T = I$. Ademais, tomando $a = -\Sigma^{-\frac{1}{2}}\mu$, temos que

$$Y = \Sigma^{-\frac{1}{2}}X - \Sigma^{-\frac{1}{2}}\mu = \Sigma^{-\frac{1}{2}}(X - \mu) \sim N_d(0, I)$$

Como Σ é uma matriz quadrada simétrica e positiva definida, todos seus autovalores são estritamente positivas. Logo,

$$\Sigma = \Gamma\Lambda\Gamma^T$$

em que Γ é a matriz diagonal de autovetores de Σ e Λ é a matriz diagonal de autovalores de Σ.
Pode-se mostrar que $\Sigma^{-1} = \Gamma\Lambda^{-1}\Gamma^T$ e $\Sigma^{-\frac{1}{2}} = \Gamma\Lambda^{-\frac{1}{2}}\Gamma^T$ em que

$$\Lambda^k = \begin{bmatrix} \lambda_1^k & 0 & \dots & 0 \\ 0 & \lambda_2^k & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_d^k \end{bmatrix}$$

Note ainda que

$$|\Sigma| = |\Gamma\Lambda\Gamma^T| = |\Lambda\Gamma^T\Gamma|$$

como $\Gamma^T\Gamma = I$,

$$|\Sigma| = |\Lambda| = \prod_{i=1}^{d} \lambda_i$$

Além disso,

$$\text{tr}\{\Sigma\} = \text{tr}\{\Gamma\Lambda\Gamma^T\} = \text{tr}\{\Lambda\Gamma^T\Gamma\} = \text{tr}\{\Lambda\} = \sum_{i=1}^{d} \lambda_i$$

i Obs 7: Normais marginais

Se $X \sim N_d(\mu, \Sigma)$, então $X_i \sim N(\mu_i, \sigma_{ii}), i = 1, \ldots d$
Demonstração:
Tomando $a = 0$ e $B_i = (0 \ldots 1 \ldots 0)$ (1 na i-ésima posição), temos que

$$a + B_i X = X_i \sim N_1(\mu_i, \sigma_{ii})$$

pelo resultado 6. Note que

$$B_i \Sigma B_i^T = \sigma_{ii}.$$

40.1 Normal bivariada $d = 2$

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_2(\mu, \Sigma)$$

em que $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ e $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$ é simétrica positiva definida.

$$|\Sigma| = \sigma_{11}\sigma_{22} - \sigma_{12}^2$$

$$\Sigma^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \begin{pmatrix} \sigma_{11} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{22} \end{pmatrix}$$

$$\Rightarrow f^X(x_1, x_2)$$

$$= \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22} - \sigma_{12}^2}} \text{Exp} \underbrace{\left\{ -\frac{1}{2} \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{pmatrix}^T \begin{pmatrix} \sigma_{11} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{22} \end{pmatrix} \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{pmatrix} \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} \right\}}_{*}$$

$$\begin{aligned} * &= [(x_1 - \mu_1)\sigma_{22} - (x_2 - \mu_2)\sigma_{12} - (x_1 - \mu_1)\sigma_{12} + (x_2 - \mu_2)\sigma_{11}] \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{pmatrix}^T \\ &= (x_1 - \mu_1)^2\sigma_{22} - 2(x_1 - \mu_1)(x_2 - \mu_2)\sigma_{12} + (x_2 - \mu_2)^2\sigma_{11} \end{aligned}$$

$$\Rightarrow \frac{*}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} = \frac{(x_1 - \mu_1)^2}{\frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{22}}} - \frac{2(x_1 - \mu_1)(x_2 - \mu_2)}{\frac{\sigma_{11}\sigma_{22} - \sigma 12^2}{\sigma_{22}}} + \frac{(x_2 - \mu_2)^2}{\frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{22}}}$$

Tome $\rho = \dfrac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$

$$\begin{aligned} A.\ \frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{22}} &= \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}} \\ &= \sigma_{11} - \sigma_{11}\frac{\sigma_{12}^2}{\sigma_{11}\sigma_{22}} \\ &= \sigma_{11} - \sigma_{11}\rho^2 \\ &= \sigma_{11}(1 - \rho^2) \end{aligned}$$

$$B.\ \frac{\sigma_{11}\sigma_{22} - \sigma_{12}^2}{\sigma_{11}} = \sigma_{22}(1 - \rho^2)$$

$$\begin{aligned} C.\ \frac{\sigma_{11}\sigma_{22}}{\sigma_{12}} - \sigma_{12} &= \sigma_{12}\frac{\sigma_{11}\sigma_{22}}{\sigma_{12}^2} - \sigma_{12} \\ &= \sigma_{12}(\rho^{-2} - 1) = \sigma_{12}\left(\frac{1 - \rho^2}{\rho^2}\right) \end{aligned}$$

$$\Rightarrow \frac{*}{\sigma_{11}\sigma_{22} - \sigma_{12}^2} = \frac{(x_1 - \mu_1)^2}{\sigma_{11}(1 - \rho^2)} - 2\rho^2\frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_{12}(1 - \rho^2)} + \frac{(x_2 - \mu_2)^2}{\sigma_{22}(1 - \rho^2)}$$

Além disso,

$$\sqrt{\sigma_{11}\sigma_{22} - \sigma_{12}^2} = \sqrt{\sigma_{11}\sigma_{22} - \rho^2\sigma_{11}\sigma_{22}} = \sqrt{\sigma_{11}\sigma_{22}}\sqrt{1 - \rho^2}$$

Assim temos que

$$f^X(x_1, x_2)$$
$$= \frac{1}{2\pi\sqrt{\sigma_{11}, \sigma_{22}}\sqrt{1-\rho^2}} \text{Exp}\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_{11}} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sqrt{\sigma_{11}\sigma_{22}}} + \frac{(x_2-\mu_2)^2}{\sigma_{22}}\right)\right\}$$

40.2 Modelo estatístico

Dizemos que $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$ é um vetor aleatório populacional normal multivariado se $X_n \sim N_d(\mu, \Sigma)$, em que o vetor de parâmetros é

$$\theta = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_d \\ \sigma_{11} \\ \vdots \\ \sigma_{d1} \\ \vdots \\ \sigma_{dd} \end{bmatrix} = \begin{bmatrix} \mu \\ \text{vech}\Sigma \end{bmatrix} = (\mu^T, (\text{vech}\Sigma)^T)^T$$

em que vechΣ é o vetor coluna com os elementos da triangular superior da matriz removidos:

$$\text{vech}\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_{11} \\ \sigma_{12} \\ \sigma_{22} \end{pmatrix}$$

40.3 Amostra aleatória de uma normal multivariada

Dizemos que $X_1, \ldots, X_n$ é uma amostra aleatória de $X \sim N_d(\mu, \Sigma)$ se, e somente se, $X_1, \ldots, X_n$ são independentes e $X_i \sim N_d(\mu, \Sigma), i = 1, \ldots, n$.

Notação

$$X_n^* = (X_1, \ldots, X_n) \text{ a.a. de } X \sim N_d(\mu, \Sigma)$$

40.4 Função de verossimilhança

Seja X^* a.a. de $X \sim N_d(\mu, \Sigma)$, em que

$$\begin{aligned}\theta &= (\mu^T, \text{vech}(\Sigma)^T)^T \\ &\in \Theta = \left\{(\mu^T, \text{vech}(\Sigma)^T) \in \mathbb{R}^d \times \mathbb{R}^{\frac{d(d+1)}{2}} : \Sigma \text{ é positiva definida (p.d.)}\right\} \subseteq \mathbb{R}^{d+\frac{d(d+1)}{2}}\end{aligned}$$

A função de verossimilhança é:

$$\begin{aligned}\mathcal{L}_{x^*}(\theta) &= \prod_{i=1}^{n} f_\theta^X(x_i) = \prod \left\{ \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \text{Exp}\left\{ -\frac{1}{2}(x_i - \mu)^T \Sigma^{-1} (x_i - \mu) \right\} \right\} \\ &= \frac{1}{(2\pi)^{\frac{nd}{2}} |\Sigma|^{\frac{1}{2}}} \text{Exp}\left\{ -\frac{1}{2} \underbrace{\sum^{n} (x_i - \mu)^T \Sigma^{-1} (x_i - \mu)}_{\dagger} \right\}\end{aligned}$$

Defina $\bar{x} = \frac{1}{n}\sum^n x_i$, então

$$\begin{aligned}\dagger &= \sum^{n} (x_i - \bar{x} + \bar{x} - \mu)^T \Sigma^{-1} (x_i - \bar{x} + \bar{x} - \mu) \\ &= \sum^{n} ((x_i - \bar{x}) - (\mu - \bar{x}))^T \Sigma^{-1} ((x_i - \bar{x}) - (\mu - \bar{x})) \\ &= \sum^{n} \left\{ (x - \bar{x})^T \Sigma^{-1} (x_i - \bar{x}) - \underbrace{(\mu - \bar{x})^T \Sigma^{-1} (x_i - \bar{x})}_{A} \right. \\ &\quad \left. - \underbrace{(x_i - \bar{x})^T \Sigma^{-1} (\mu - \bar{x})}_{B} + (\mu - \bar{x})^T \Sigma^{-1} (\mu - \bar{x}) \right\}\end{aligned}$$

Como $n\bar{x} = \sum^n x_i$, temos que

$$\begin{cases} A = \sum^n (\mu - \bar{x})^T \Sigma^{-1} (x_i - \bar{x}) = (\mu - \bar{x})^T \Sigma^{-1} (\sum^n x_i - \bar{x}) = 0 \\ B = \sum^n (x_i - \bar{x})^T \Sigma^{-1} (\mu - \bar{x}) = (\sum^n x_i - \bar{x})^T \Sigma^{-1} (\mu - \bar{x}) = 0 \end{cases}$$

Portanto,

$$\begin{aligned}\dagger &= \sum^{n} (x - \bar{x})^T \Sigma^{-1} (x_i - \bar{x}) + n(\mu - \bar{x})^T \Sigma^{-1} (\mu - \bar{x}) \\ &= \sum^{n} \text{tr}\{(x - \bar{x})^T \Sigma^{-1} (x_i - \bar{x})\} + n(\mu - \bar{x})^T \Sigma^{-1} (\mu - \bar{x}) \\ &\overset{\text{Prop. tr}}{=} \sum^{n} \text{tr}\{\Sigma^{-1} (x - \bar{x})(x_i - \bar{x})^T\} + n(\mu - \bar{x})^T \Sigma^{-1} (\mu - \bar{x})\end{aligned}$$

Note que $\sum^n \text{tr}\{Ay_i\} = \text{tr}\{A \sum y_i\}$, usando essa propriedade,

$$\dagger = n\text{tr}\{\Sigma^{-1}S^2(x_n^*)\} + n(\mu - \bar{x})^T\Sigma^{-1}(\mu - \bar{x})$$

em que $S^2(x_n^*) = \frac{1}{n}\sum^n (x_i - \bar{x})(x_i - \bar{x})^T$.

Dessa forma, podemos escrever a função de verossimilhança da seguinte forma:

$$\mathcal{L}_{x^*}(\theta) = \frac{1}{(2\pi)^{\frac{nd}{2}}|\Sigma|^{\frac{1}{2}}}\text{Exp}\left\{-\frac{n}{2}\left(\text{tr}\{\Sigma^{-1}S^2(x_n^*)\} + (\mu - \bar{x})^T\Sigma^{-1}(\mu - \bar{x})\right)\right\}$$

Pelo critério da fatoração de Neyman-Fisher, temos que $(\bar{X}, S^2(X_n^*))$ é uma estatística suficiente para o modelo normal multivariado. Como $S^2(X_n^*)$ é uma matriz simétrica, podemos considerar apenas $\text{vech}(S^2(X_n^*))$

Para maximizar a função de verossimilhança em relação a μ, note que basta minimizar

$$(\mu - \bar{x})^T\Sigma^{-1}(\mu - \bar{x})$$

note que, como Σ é positiva definida (p.d.), Σ^{-1} também é p.d. Portanto, pela definição, $y^T\Sigma^{-1}y > 0, \forall y \in \mathbb{R}^d \setminus \{0\}$ e $(\mu - \bar{x})^T\Sigma^{-1}(\mu - \bar{x})$ é mínimo quando $\mu = \bar{x}$.

Para encontrar Σ que maximiza a função de verossimilhança, substituímos μ por $\bar{x}$ e tentamos encontrar seu valor. Portanto, devemos maximizar

$$\frac{1}{(2\pi)^{\frac{nd}{2}}|\Sigma|^{\frac{1}{2}}}\text{Exp}\left\{-\frac{n}{2}\text{tr}\{\Sigma^{-1}S^2(x_n^*)\}\right\}$$

em relação a Σ. Aplicando ln, temos

$$-\frac{nd}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{n}{2}\text{tr}\{\Sigma^{-1}S^2(x_n^*)\} = c - \frac{n}{2}\left(\ln|\Sigma| + \text{tr}\{\Sigma^{-1}S^2(x_n^*)\}\right)$$

Note que, para qualquer $\lambda > 0$, temos

$$\lambda - \ln\lambda \geq 1.$$

Sejam $\lambda_1, \ldots, \lambda_d$ autovalores de uma matriz positiva definida M.

$$\begin{aligned}
&\lambda_i - \ln\lambda_i \geq 1 \forall i = 1, \ldots, d \\
\Rightarrow &\sum^d \lambda_i - \sum^d \ln_i \geq d \\
\iff &\sum^d \lambda_i - \ln(\prod^d \lambda_i) \geq d \\
\Rightarrow &\text{tr}\{M\} - \ln|M| \geq d
\end{aligned}$$

Tome $M = \Sigma^{-1}S^(x_n^*)$. Então, assumindo $S^2(x_n^*)$ positiva definida, M é positiva definida e

$$
\begin{aligned}
&\operatorname{tr}\{\Sigma^{-1}S^2(x_n^*)\} - \ln|\Sigma^{-1}S^2(x_n^*)| \geq d \\
\iff &\operatorname{tr}\{\Sigma^{-1}S^2(x_n^*)\} - \ln(|\Sigma^{-1}||S^2(x_n^*)|) \geq d \\
\iff &\operatorname{tr}\{\Sigma^{-1}S^2(x_n^*)\} + \ln|\Sigma| - \ln|S^2(x_n^*)| \geq d \\
\iff &\operatorname{tr}\{\Sigma^{-1}S^2(x_n^*)\} + \ln|\Sigma| \geq d + \ln|S^2(x_n^*)|
\end{aligned}
$$

Note que a igualdade (o menor valor possível) é atingida quando $\Sigma = S^2(x_n^*)$ pois

$$
\operatorname{tr}\{S^2(x_n^*)^{-1}S^2(x_n^*)\} = \operatorname{tr}\{I_d\} = d
$$

Logo,

$$
\begin{aligned}
\hat{\mu}_{\text{MV}}(X_n^*) &= \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \\
\hat{\Sigma}_{\text{MV}}(X_n^*) &= S^2(X_n^*) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T
\end{aligned}
$$

são os estimadores de máxima verossimilhança para μ e Σ, respectivamente. Dessa forma, podemos estimar usando a propriedade de invariância qualquer quantidade de interesse $g(\theta)$

40.4.1 Exemplo

Seja X_n^* a.a. de $X \sim N_2(\mu, \Sigma)$. Encontre o EMV para $g(\theta)$, em que

1. $g(\theta) = E_\theta(X)$
2. $g(\theta) = \operatorname{Var}_\theta(X)$
3. $g(\theta) = E_\theta(X)^T \operatorname{Var}_\theta(X)^{-1} E_\theta(X)$
4. $g(\theta) = P_\theta(X_1 \geq 4X_2), X = (X_1, X_2)^T$
5. Considerando os dados observados abaixo, apresente as estimativas para $g(\theta)$ dos itens anteriores:

40.4.1.1 Resposta

Já sabemos que os estimadores são dados por

$$
\begin{aligned}
\hat{\mu}_{\text{MV}}(X_n^*) &= \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \\
\hat{\Sigma}_{\text{MV}}(X_n^*) &= S^2(X_n^*) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^T \\
&= \frac{1}{n}\sum X_i X^T - \bar{X}\bar{X}^T
\end{aligned}
$$

e as estimativas por

$$\hat{\mu}_{\text{MV}}(x_n^*) = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\begin{pmatrix}\sum x_{1i}\\ \sum x_{2i}\end{pmatrix}$$

$$\begin{aligned}\hat{\Sigma}_{\text{MV}}(x_n^*) &= S^2(x_n^*) = \frac{1}{n}\sum x_i x^T - \bar{x}\bar{x}^T\\ &= \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix}(x_{1i}-\bar{x}_1)^2 & (x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)\\ (x_{2i}-\bar{x}_2)(x_{1i}-\bar{x}_1) & (x_{2i}-\bar{x}_2)^2\end{bmatrix}\end{aligned}$$

4. $P_\theta(X2 > 4X_1) = P_\theta(X_2 - 4X_1 > 0)$. Note que

$$X_2 - 4X_1 = BX$$

em que $B = (-4, 1)$. Então,

$$X_2 - 4X_1 \sim N(B\mu, B\Sigma B^T)$$

Desenvolvendo,

$$\begin{aligned}B\mu &= -4\mu_1 + \mu_2 = \mu_2 - 4\mu_1\\ B\Sigma B^T &= (-4,1)\begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{bmatrix}\begin{pmatrix}-4\\ 1\end{pmatrix}\\ &\overset{\sigma_{12}=\sigma_{21}}{=} 16\sigma_{11} + \sigma_{22} - 8\sigma_{12}\end{aligned}$$

Assim,

$$\begin{aligned}g(\theta) &= P_\theta(X_2 - 4X_1 > 0)\\ &= P_\theta\left(\frac{(X_2-4X_1)-(\mu_2-4\mu_1)}{\sqrt{16\sigma_{11}+\sigma_{22}-8\sigma_{12}}} > -\frac{\mu_2-4\mu_1}{\sqrt{16\sigma_{11}+\sigma_{22}-8\sigma_{12}}}\right)\\ &= P\left(Z > -\frac{\mu_2-4\mu_1}{\sqrt{16\sigma_{11}+\sigma_{22}-8\sigma_{12}}}\right)\\ &= 1-\Phi\left(-\frac{\mu_2-4\mu_1}{\sqrt{16\sigma_{11}+\sigma_{22}-8\sigma_{12}}}\right)\end{aligned}$$

40.5 Aplicações da normal multivariada

40.5.1 Teorema do limite central para o EMV com variáveis reais

Seja X_n a.a. de $X \sim f_\theta, \theta \in \Theta \subseteq \mathbb{R}^P, p \in \mathbb{N}$. Se as condições de regularidade estiverem satisfeitas, então,

$$\sqrt{n}(\hat{\theta}_{\text{MV}}(X_n) - \theta) \overset{\mathcal{D}}{\to} N_p(0, I_1(\theta)^{-1}),$$

ou, de outra forma,

$$I_n(\theta)^{\frac{1}{2}}(\hat{\theta}_{\text{MV}}(X_n) - \theta) \xrightarrow{\mathcal{D}} N_p(0, I).$$

Além disso, pelo Teorema de Slutsky,

$$I_n(\hat{\theta}_{\text{MV}}(X_n))^{\frac{1}{2}}(\hat{\theta}_{\text{MV}}(X_n) - \theta) \xrightarrow{\mathcal{D}} N_p(0, I)$$

em que I é a matriz identidade.

Notação

$$\hat{\theta}_{\text{MV}}(X_n) \overset{a}{\approx} N_p(\theta, I_n(\theta)^{-1})$$

Note que, se $\theta = (\theta_1, \theta_2)^T$, então

$$\hat{\theta}_{\text{MV}} = \begin{pmatrix} \hat{\theta}_{\text{MV}}^{(1)} \\ \hat{\theta}_{\text{MV}}^{(2)} \end{pmatrix} \overset{a}{\approx} N_2\left[\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}; I_n(\theta)^{-1}\right],$$

logo,

$$\hat{\theta}_{\text{MV}}(X_n)^{(i)} \overset{a}{\approx} N_p(\theta_i, V_i(\theta)), i = 1, 2$$

em que $V_i(\theta)$ é o elemento (i, i) da matriz $I_n(\theta)^{-1}, i = 1, 2$.

Observação

Observe que $V_i(\theta)$ pode depender do vetor "θ" inteiro. Veja o caso normal:

$$\bar{X} \sim N_1\left(\mu, \frac{\sigma^2}{n}\right)$$

40.5.2 Teorema do limite central para o EMM com variáveis reais

Seja X_n a.a. de $X \sim f_\theta, \theta \in \Theta \subseteq \mathbb{R}^P, p \in \mathbb{N}$ tal que

$$E_\theta(|X|^k) < \infty, k = 1, \dots, 2p.$$

Seja $h(\theta) = \begin{bmatrix} \frac{\partial E_\theta(X)}{\partial \theta^T} \\ \vdots \\ \frac{\partial E_\theta(X^p)}{\partial \theta^T} \end{bmatrix}$ e considere que $\det h(\theta) \neq 0$ e cada componente de $h(\theta)$ é uma função contínua. Então,

$$\sqrt{n}(\hat{\theta}_{\text{MM}}(X_n) - \theta) \xrightarrow{\mathcal{D}} N_p(0, V_\theta)$$

em que

$$V_\theta = h(\theta)^{-1}\text{Var}_\theta(Y)[h(\theta)^{-1}]^T$$

$$Y = \begin{bmatrix} X \\ X^2 \\ \vdots \\ X^p \end{bmatrix}$$

Notação

$$\hat{\theta}_{\text{MM}}(X_n) \overset{a}{\approx} N_p\left(\theta, \frac{h(\theta)^{-1}\text{Var}_\theta(Y)[h(\theta)^{-1}]^T}{n}\right)$$

Generalização do Método de Momentos

O método de momentos pode ser generalizado da seguinte forma:
Se $E_\theta(|X|^k) < \infty$ para $k \in \{k_1, k_2, \dots, k_p\} \subseteq N$ ou, se X for não negativa, $k \in \{k_1, \dots, k_p\} \subseteq \mathbb{Q}$. O EMM é obtido igualando

$$E_\theta(X^k) = \frac{1}{n}\sum X_i^k, k \in \{k_1, \dots, k_n\}$$

Seja

$$h(\theta) = \begin{bmatrix} \frac{\partial E_\theta(X^{k_1})}{\partial \theta^T} \\ \vdots \\ \frac{\partial E_\theta(X^{k_p})}{\partial \theta^T} \end{bmatrix}$$

Considere que

1. $E_\theta(|X|^{2k}) < \infty, \forall k \in \{k_1, \dots, k_n\}$
2. $\det h(\theta) \neq 0$
3. Cada componente de $h(\theta)$ é uma função contínua.

Então,

$$\sqrt{n}(\hat{\theta}_{\text{MM}}(X_n) - \theta) \overset{\mathcal{D}}{\to} N_p(0, V_\theta)$$

em que

$$V_\theta = h(\theta)^{-1}\text{Var}_\theta(Y)[h(\theta)^{-1}]^T$$

$$Y = \begin{bmatrix} X^{k_1} \\ X^{k_2} \\ \vdots \\ X^{k_p} \end{bmatrix}$$

40.5.3 Exemplos

40.5.3.1 Exemplo Beta

Seja X_n a.a. de $X \sim f_\theta, \theta = (\alpha, \beta) \in \Theta = \mathbb{R}^2_+$ tal que

$$f_\theta(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, x \in (0,1) \\ 0, \text{c.c.} \end{cases}$$

Encontre o EMV, EMM e suas distribuições assinstóticas para θ.

40.5.3.1.1 Resposta

Note que $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

$$\begin{aligned} \mathcal{L}_{x_n}(\theta) &= \frac{1}{B(\alpha,\beta)^n} (\prod x_i)^{\alpha-1} (\prod (1-x_i))^{\beta-1} \\ l_{x_n}(\theta) &= -n \ln \Gamma(\alpha) - n \ln \Gamma(\beta) + n \ln \Gamma(\alpha+\beta) \\ &\quad + (\alpha-1) \sum \ln x_i + (\beta-1) \sum \ln(1-x_i), \\ \Rightarrow \frac{\partial l_{x_n}}{\partial \alpha}(\theta) &= -n\psi_1(\alpha) + n\psi_1(\alpha+\beta) + \sum \ln x_i; \\ \Rightarrow \frac{\partial l_{x_n}}{\partial \beta}(\theta) &= -n\psi_1(\beta) + n\psi_1(\alpha+\beta) + \sum \ln(1-x_i) \end{aligned}$$

em que

$$\psi_k(a) = \frac{\partial^k \ln \Gamma(a)}{\partial a^k}$$

continuando,

$$\begin{aligned} \Rightarrow \frac{\partial^2 l_{x_n}}{\partial \alpha^2}(\theta) &= -n\psi_2(\alpha) + n\psi_2(\alpha+\beta); \\ \Rightarrow \frac{\partial^2 l_{x_n}}{\partial \beta^2}(\theta) &= -n\psi_2(\beta) + n\psi_2(\alpha+\beta); \\ \Rightarrow \frac{\partial^2 l_{x_n}}{\partial \beta \partial \alpha}(\theta) &= n\psi_2(\alpha+\beta). \end{aligned}$$

$$\begin{aligned} \Rightarrow \frac{\partial^2 l_{x_n}}{\partial \theta \partial \theta^T}(\theta) &= -n \begin{bmatrix} \psi_2(\alpha) - \psi_2(\alpha+\beta) & -\psi_2(\alpha+\beta) \\ -\psi_2(\alpha+\beta) & \psi_2(\beta) - \psi_2(\alpha+\beta) \end{bmatrix} \\ \Rightarrow I_n(\theta) &= n \begin{bmatrix} \psi_2(\alpha) - \psi_2(\alpha+\beta) & -\psi_2(\alpha+\beta) \\ -\psi_2(\alpha+\beta) & \psi_2(\beta) - \psi_2(\alpha+\beta) \end{bmatrix}. \end{aligned}$$

A estimativa de MV pode ser obtida numericamente via algoritmo de Newton-Raphson. Neste caso, o método é equivalente ao de escore de Fisher:

$$\hat{\theta}^{(j+1)} = \hat{\theta}^{(j)} + I_n(\hat{\theta}^{(j)})^{-1} U_n(x_n, \hat{\theta}^{(j)})$$

em que

$$U_n(x_n, \theta) = \begin{bmatrix} -n\psi_1(\alpha) + n\psi_1(\alpha+\beta) + \sum \ln x_i \\ -n\psi_1(\beta) + n\psi_1(\alpha+\beta) + \sum \ln(1 - x_i) \end{bmatrix}$$

Além disso,

$$\sqrt{n}(\hat{\theta}_{\mathrm{MV}} - \theta) \xrightarrow{\mathcal{D}} N_2(0, I_1(\theta)^{-1})$$

em que

$$\begin{aligned} I_1(\theta)^{-1} = & \frac{1}{(\psi_2(\alpha) - \psi_2(\alpha+\beta))(\psi_2(\beta) - \psi_2(\alpha+\beta)) - \psi_2(\alpha+\beta)^2} \\ & \cdot \begin{bmatrix} \psi_2(\alpha) - \psi_2(\alpha+\beta) & \psi_2(\alpha+\beta) \\ \psi_2(\alpha+\beta) & \psi_2(\beta) - \psi_2(\alpha+\beta) \end{bmatrix}. \end{aligned}$$

41 Intervalos de Confiança - Aprofundamento

Seja $X_n = (X_1, \dots, X_n)$ amostra aleatória e $X \sim f_\theta, \theta \in \Theta$.

Dizemos que $[I_1(X_n), I_2(X_n)]$ é um intervalo de confiança *exato* para $g(\theta)$ com coeficiente de confiança $\gamma \in (0,1)$ se, e somente se:

$$P_\theta\left(I_1(X_n) \leq g(\theta) \leq I_2(X_n)\right) = \gamma, \forall \theta \in \Theta.$$

em que $I_1(X_n), I_2(X_n)$ são estatísticas.

Observação

Se $P_\theta\left(I_1(X_n) \leq g(\theta) \leq I_2(X_n)\right) \geq \gamma, \forall \theta \in \Theta$, então $[I_1(X_n), I_2(X_n)]$ é um Intervalo de Confiança (IC) de *pelo menos* γ

Notação

$$\text{IC}(g(\theta), \gamma) = [I_1(X_n), I_2(X_n)]$$

Observe que $\text{IC}(g(\theta), \gamma)$ é um *intervalo aleatório* que não depende de "θ".

Na prática, observamos a amostra $x_n = (x_1, \dots, x_n)$ e calculamos o IC *observado*

Notação

$$\text{IC}_{\text{Obs}}(g(\theta), \gamma) = [I_1(x_n), I_2(x_n)]$$

Observe que este é um intervalo *numérico*

Portanto,

$$P_\theta\left(I_1(X_n) \leq g(\theta) \leq I_2(X_n)\right) = \begin{cases} 1, g(\theta) \in \text{IC}_{\text{Obs}}(g(\theta), \gamma) \\ 0, \text{c.c.} \end{cases}$$

41.1 Interpretação em termos de repetições:

Se repetirmos o experimento, mantendo as mesmas condições, então esperamos que em $\gamma \cdot 100\%$ dos experimentos os ICs contenham $g(\theta)$. Em outras palavras,

$$\# \frac{(g(\theta) \in \text{IC}_{\text{Obs}})}{N} \approx \gamma$$
$$\left[\frac{1}{N} \sum \mathbb{1}_{\{\text{IC}^{(i)}_{\text{Obs}}\}}(g(\theta)) \overset{N \uparrow \infty}{\to} \gamma \right]$$

Dizemos que

$$\frac{1}{N} \sum \mathbb{1}_{\{\text{IC}^{(i)}_{\text{Obs}}\}}(g(\theta))$$

é a cobertura de $\text{IC}(g(\theta), \gamma)$

O que dizer sobre IC_{Obs} em relação a $g(\theta)$?

Temos uma **confiança** de $\gamma \cdot 100\%$ de que $g(\theta) \in \text{IC}_{\text{Obs}}$.

41.2 Exemplos

41.2.1 Exemplo Normal

Seja X_n a.a. de $X \sim N(\mu, \sigma^2), \theta = (\mu, \sigma^2) \in \Theta = \mathbb{R} \times \mathbb{R}^+$. Encontre um IC com coeficiente de confiança $\gamma = 95\%$...

...para $g(\theta = \mu)$

Sabemos que

$$\bar{X} = \frac{1}{n} \sum X_i \sim N(\mu, \sigma^2/n).$$

Além disso,

$$\sum \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

Por definição de t-student, com as V.As das distribuições independentes,

$$\frac{N(0,1)}{\sqrt{\frac{\chi_k^2}{k}}} \sim t_k$$
$$\Rightarrow \frac{\sqrt{n}\frac{(\bar{X}-\mu)}{\sigma}}{\sqrt{\frac{\sum\frac{(X_i-\bar{X})^2}{\sigma^2}}{n-1}}} \sim t_{(n-1)}$$
$$\Rightarrow \frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{S_{n-1}^2(X_n)}} \sim t_{(n-1)}$$

logo, podemos sempre encontrar $c_{1,\gamma}, c_{2,\gamma}$ tais que

$$P_\theta\left(c_{2,\gamma} \leq \frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{S_{n-1}^2(X_n)}} \leq c_{1,\gamma}\right) = \gamma$$

Note que

$$c_{2,\gamma} \leq \frac{\sqrt{n}(\bar{X}-\mu)}{\sqrt{S_{n-1}^2(X_n)}} \leq c_{1,\gamma} \iff \bar{X} - c_{2,\gamma}\sqrt{\frac{S_{n-1}^2(X_n)}{n}} \leq \mu \leq \bar{X} - c_{1,\gamma}\sqrt{\frac{S_{n-1}^2(X_n)}{n}}$$

Portanto,

$$P_\theta\left(\bar{X} - c_{2,\gamma}\sqrt{\frac{S_{n-1}^2(X_n)}{n}} \leq g(\theta) \leq \bar{X} - c_{1,\gamma}\sqrt{\frac{S_{n-1}^2(X_n)}{n}}\right) = \gamma$$

Pela definição de Intervalo de Confiança,

$$\mathrm{IC}(\mu,\gamma) = \left[\bar{X} - c_{2,\gamma}\sqrt{\frac{S_{n-1}^2(X_n)}{n}}, \bar{X} - c_{1,\gamma}\sqrt{\frac{S_{n-1}^2(X_n)}{n}}\right]$$

em que

$$S_{n-1}^2(X_n) = \frac{1}{n-1}\sum(X_i - \bar{X})^2$$

e $c_{1,\gamma}, c_{2,\gamma}$ são os quantis obtidos da distribuição t-student com $n-1$ graus de liberdade que satisfaçam

$$P_\theta\left(\bar{X} - c_{2,\gamma}\sqrt{\frac{S_{n-1}^2(X_n)}{n}} \leq g(\theta) \leq \bar{X} - c_{1,\gamma}\sqrt{\frac{S_{n-1}^2(X_n)}{n}}\right) = \gamma$$

no caso simétrico (minimiza o IC para distribuições como a Normal), $c_{2,\gamma} = -c_{1,\gamma}$. Note que não é possível construir ICs simétricos dessa forma para distribuições estritamente positivas, como a qui-quadrado.

41.3 Quantidades Pivotais

Dizemos que $Q(g(\theta), X_n)$ é uma quantidade pivotal para $g(\theta)$ se, e somente se,

1. $Q(g(\theta), X_n)$ depende de $g(\theta)$
2. A distribuição de $Q(g(\theta), X_n)$ não depende de "θ"
3. Existem a_1, a_2, que não dependem de $g(\theta)$, tais que $c_1 \leq Q(g(\theta), X_n) \leq c_2 \iff a_1 \leq g(\theta) \leq a_2$

41.3.1 Exemplos

$X \sim N(\mu, \sigma^2)$

1. Se σ^2 é conhecido e $\theta = \mu.g(\theta) = \mu$, então uma quantidade pivotal é dada por

$$Q(\mu, X_n) = \sqrt{n}\frac{\bar{X} - \mu}{\sqrt{\sigma^2}} \sim N(0, 1)$$

$$\text{IC}(\mu, \gamma) = \bar{X} \mp c_\gamma \sqrt{\frac{\sigma^2}{n}}$$

2. Se σ^2 é desconhecido e $\theta = (\mu, \sigma^2), g(\theta) = \mu$, então uma quantidade pivotal é dada por

$$Q(\mu, X_n) = \sqrt{n}\frac{\bar{X} - \mu}{\sqrt{S^2_{n-1}(X_n)}} \sim t_{(n-1)}$$

$$\text{IC}(\mu, \gamma) = \bar{X} \mp c_\gamma \sqrt{\frac{S^2_{n-1}(X_n)}{n}}$$

essa é também uma quantidade pivotal para 1., mas o contrário não vale.

3. Se μ é conhecido e $\theta = \sigma^2, g(\theta) = \sigma^2$, então

$$Q(\sigma^2, X_n) = \sum \frac{(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_n$$

$$\text{IC}(\sigma^2, \gamma) = \left[\frac{\sum(X_i - \mu)^2}{c_{2,\gamma}}, \frac{\sum(X_i - \mu)^2}{c_{1,\gamma}}\right]$$

4. Se μ é desconhecido e $\theta = (\mu, \sigma^2), g(\theta) = \sigma^2$, então

$$Q(\sigma^2, X_n) = \sum \frac{(X_i - \bar{X})}{\sigma^2} \sim \chi^2_{(n-1)}$$

$$\text{IC}(\sigma^2, \gamma) = \left[\frac{\sum (X_i - \bar{X})^2}{c_{2,\gamma}}, \frac{\sum (X_i - \bar{X})^2}{c_{1,\gamma}} \right]$$

41.3.1.1 Exponencial

Seja $X \sim \text{Exp}(\theta), \theta > 0$. Encontre uma quantidade pivotal para θ.

Note que $\sum X_i \sim \text{Gama}(n, \theta)$ com F.G.M. dada por

$$M_{\sum X_i}(t) = \left(\frac{\theta}{\theta - t} \right)^n$$

note ainda que

$$\begin{aligned}
M_{\sum X_i}(t) &= E(\mathrm{e}^{t \sum X_i}) = \left(\frac{\theta}{\theta - t} \right)^n \\
\Rightarrow M_{\theta \sum X_i}(t) &= E(\mathrm{e}^{t\theta \sum X_i}) \\
&= M_{\sum X_i}(t\theta) = \left(\frac{\theta}{\theta - \theta t} \right)^n = \left(\frac{1}{1-t} \right)^n \\
\Rightarrow M_{2\theta \sum X_i}(t) &= \left(\frac{1}{1-2t} \right)^n \\
&\Rightarrow 2\theta \sum X_i \sim \chi^2_{(2n)}
\end{aligned}$$

Portanto, $Q(\theta, X_n) = 2\theta \sum X_i$ é uma quantidade de interesse para θ

$$\begin{aligned}
& c_{1,\gamma} \leq Q(\theta, X_n) \leq c_{2,\gamma} \\
\iff & c_{1,\gamma} \leq 2\theta \sum X_i \leq c_{2\gamma} \iff \frac{c_{1,\gamma}}{2 \sum X_i} \leq \theta \leq \frac{c_{2,\gamma}}{2 \sum X_i}
\end{aligned}$$

41.3.1.2 Uniforme

Seja $X \sim \text{Unif}(0, \theta), \theta > 0$. Encontre uma quantidade pivotal para θ.

Note que $X_{(n)} = \max X_n$ é uma estatística suficiente cuja f.d.p. é dada por

$$f_(\theta)^{X_{(n)}}(x) = \frac{nx^{n-1}}{\theta^n} 1_{(0,\theta]}(x)$$

Seja $Y = \frac{X_{(n)}}{\theta}$, então

$$f_\theta^Y(y) = f_\theta^{X_{(n)}}(y\theta) \cdot |J|$$

em que $J = \theta$ (determinante jacobiano)

$$f_\theta^Y(y) = \frac{n(y\theta)^{n-1}}{\theta^n}\theta \mathbb{1}_{(0,\theta]}(y\theta) = ny^{n-1}\mathbb{1}_{(0,1]}(y)$$

não depende de "θ"! Logo, $Q(\theta, X_n) = \frac{X_{(n)}}{\theta}$ é uma quantidade pivotal para θ.

$$\frac{X_{(n)}}{c_{2,\gamma}} \le \theta \le \frac{X_{(n)}}{c_{1,\gamma}} \Rightarrow \text{IC}(\theta,\gamma) = \left[\frac{X_{(n)}}{c_{2,\gamma}}, \frac{X_{(n)}}{c_{1,\gamma}}\right]$$

em que os quantis são obtidos da distribuição de Y, nesse caso, $Y \sim \text{Beta}(n,1)$:

$$\int_0^{c_{1,\gamma}} ny^{n-1}dy = \frac{1-\gamma}{2} \qquad \int_{c_{2,\gamma}}^1 ny^{n-1}dy = \frac{1-\gamma}{2}$$

$$\Rightarrow \begin{cases} y^n|_0^{c_{1,\gamma}} = \frac{1-\gamma}{2} \\ y^n|_{c_{2,\gamma}}^1 = \frac{1-\gamma}{2} \end{cases} \Rightarrow c_{1,\gamma} = \left(\frac{1-\gamma}{2}\right)^{1/2} \qquad c_{2,\gamma} = \left(\frac{1+\gamma}{2}\right)^{1/2}$$

Quantis que minimizam a amplitude do IC

Seja $Q(g(\theta), X_n)$ uma quantidade pivotal com função densidade de probabilidade f. $c_{1,\gamma}, c_{2,\gamma}$ devem ser obtidos

$$\int_{c_{1,\gamma}}^{c_{2,\gamma}} f(x)dx = \gamma \tag{41.1}$$

Note que em geral infinitas combinações desses quantis satisfazem (**??**). Podemos usar o par que satisfaz

$$\int_{-\infty}^{c_{1,\gamma}} f(y)dy = \frac{1-\gamma}{2} \quad \text{e} \quad \int_{c_{2,\gamma}}^{\infty} f(y)dy = \frac{1-\gamma}{2}$$

Esse método produz um intervalo de confiança simétrico, não necessariamente o de menor amplitude, mas é mais fácil de encontrar. O intervalo de confiança com menor amplitude com quantis que satisfaçam (**??**) é obtido minimizando $|c_{2,\gamma} - c_{1,\gamma}|$ sujeito a (**??**).
Se f for *unimodal* e *bicaudal*, pode-se demonstrar que $c_{1,\gamma}, c_2\gamma$ que produzem amplitude mínima e satisfazem (**??**) são tais que

$$f(c_{1,\gamma}) = f(c_{2,\gamma})$$

ou seja, tem mesma densidade.

```julia
using Distributions, Random, StatsBase, LaTeXStrings

Random.seed!(8)
```

```
theta0 = 4
n = 10
MC = 10000

I1 = []
I2 = []
for _ in 1:MC
  d = Exponential(1/theta0) # Parâmetro média => 1/theta parâmetro taxa
  x = rand(d,n)

  # Construindo um IC com 95% de confiança
  gamma = 0.95
  quiquadrado = Chisq(2*n)
  c1 = quantile(quiquadrado, (1-gamma)/2)
  c2 = quantile(quiquadrado, gamma + (1-gamma)/2)
  push!(I1, round(c1/(2*sum(x)), digits=4))
  push!(I2, round(c2/(2*sum(x)), digits=4))
end

acertos = [I1 .<= theta0 .<= I2]
display(L"\mathrm{IC}_{\mathrm{Obs (1)}}(\theta,0.95)=[%$(I1[1]), %$(I2[1])]")
display(L"\text{ICs que contém}\ \theta: %$((sum(acertos[1]))/MC * 100)\%")
```

$\mathrm{IC}_{\mathrm{Obs(1)}}(\theta, 0.95) = [2.7345, 9.7424]$

$\text{ICs que contém } \theta : 94.97\%$

41.3.2 Quantidades pivotais aproximadas ou assintóticas

Dizemos que $Q(g(\theta), X_n)$ é uma quantidade pivotal aproximada ou assintótica se, e somente se

1. $Q(g(\theta), X_n)$ depende de $g(\theta)$;
2. A distribuição assintótica de $Q(g(\theta), X_n)$ não depende de $g(\theta)$;
3. Existem a_1, a_2, que não dependem de $g(\theta)$, tais que $c_1 \leq Q(g(\theta), X_n) \leq c_2 \iff a_1 \leq g(\theta) \leq a_2$. Esses valores dependem apenas de X_n.

Podemos usar o TLC para estimadores:

Se $T(X_n)$ for um estimador para θ assintoticamente normal, então

$$\sqrt{n}(T(X_n) - \theta) \xrightarrow{\mathcal{D}} N_p(0, V_\theta), \forall \theta \in \Theta.$$

em que V_θ é uma matriz positiva definida. Para $p = 1$,

$$\sqrt{n}(T(X_n) - \theta) \xrightarrow{\mathcal{D}} N_1(0, V_\theta), \forall \theta \in \Theta.$$

em que $V_\theta > 0$.

Se $g : \Theta \to \mathbb{R}$ for uma função tal que $g'(\theta) \neq 0$ $g'(\theta)$ é contínua, então,

$$\sqrt{n}(g(T(X_n)) - g(\theta)) \xrightarrow{\mathcal{D}} N_1(0, g'(\theta)^2 V_\theta), \forall \theta \in \Theta.$$

Além disso,

1. $$\frac{\sqrt{n}(g(T(X_n)) - g(\theta))}{\sqrt{g'(\theta)^2 V_\theta}} \xrightarrow{\mathcal{D}} N_1(0, 1), \forall \theta \in \Theta.$$

2. Pelo teorema de Slutsky
$$\frac{\sqrt{n}(g(T(X_n)) - g(\theta))}{\sqrt{g'(T(X_n))^2 V_{T(X_n)}}} \xrightarrow{\mathcal{D}} N_1(0, 1), \forall \theta \in \Theta.$$

Ou seja,
$$Q(g(\theta), X_n) = \frac{\sqrt{n}(g(T(X_n)) - g(\theta))}{\sqrt{g'(T(X_n))^2 V_{T(X_n)}}}$$

é uma quantidade pivotal aproximada para $g(\theta)$. Note que, com $T(X_n) = \hat{(}\theta)$,

$$\begin{aligned}
&c_1 \leq Q(g(\theta), X_n) \leq c_2 \\
\iff &c_1 \leq \frac{\sqrt{n}(g(T(X_n)) - g(\theta))}{\sqrt{g'(T(X_n))^2 V_{T(X_n)}}} \leq c_2 \\
\iff &c_1 \sqrt{\frac{g'(\hat{\theta})^2}{n} V_{\hat{\theta}}} \leq g(\hat{\theta}) - g(\theta) \leq c_2 \sqrt{\frac{g'(\hat{\theta})^2}{n} V_{\hat{\theta}}} \\
\iff &g(\hat{\theta}) - c_2 \sqrt{\frac{g'(\hat{\theta})^2}{n} V_{\hat{\theta}}} \leq g(\theta) \leq g(\hat{\theta}) - c_1 \sqrt{\frac{g'(\hat{\theta})^2}{n} V_{\hat{\theta}}}
\end{aligned}$$

Tomando $c_1 = -c_2$, pois a distribuição assintótica é normal, temos

$$\mathrm{IC}^a(g(\theta), \gamma) = \left[g(\hat{\theta}) - c_2 \sqrt{\frac{g'(\hat{\theta})^2}{n} V_{\hat{\theta}}}, g(\hat{\theta}) + c_2 \sqrt{\frac{g'(\hat{\theta})^2}{n} V_{\hat{\theta}}} \right]$$

em que c_2 é obtido dos quantis da normal padrão tal que

$$P(-c_2 \leq N(0,1) \leq c_2) = \gamma$$

Erro padrão

A quantidade

$$\sqrt{\frac{g'(\hat{\theta})^2}{n} V_{\hat{\theta}}}$$

é o erro-padrão do estimador (seu desvio padrão).

Vamos conferir por simulações de monte carlo e o método de Newton-Raphson

a) Para θ
b) Para $g(\theta) = P(X < 950) \Rightarrow g'(\theta) = f_\theta(950)$
c) Para $g(\theta) = \ln f_\theta(x) g'(\theta) = -\psi_1(\theta) + \ln(\theta) - 1/\theta$
d) Para $g(\theta) = f_\theta(x)$

```
using SpecialFunctions, Distributions, Random, LaTeXStrings
# Modelo Gama( , 1)
# IC assintótico
# Calcule o IC
#a-) para
#b-) para g( ) = P(X < 950) => g'( ) = f_ (950)
#c-) para g_x( ) = ln f_ (x) g'( ) = -  _1( ) + ln( ) - 1/ . x =
#d-) para g_x( ) = f_\theta(x). x =


# Sabemos que  .MV ~a~ N( , I_n( )^-1)
# => IC^a( , ) =  .MV   c_2 * √(I_1( ))

Random.seed!(13)

function newton_raphson(x)

  theta::Vector{Float64} = []
  append!(theta, mean(x)) # Chute inicial = média
  erromax = 10^(-5)
  erro = Inf
  i = 1
  # iteracoesMax = 6 # Podemos também definir apenas um erro máximo
  while erro > erromax # && i < iteracoesMax
```

```
    append!(theta, theta[i] - (sum(log.(x)) - n * digamma(theta[i]))/
            (-n*trigamma(theta[i])))
    erro = abs(theta[i+1] - theta[i])
    # println("Erro na iteração $i: $erro")
    i += 1
  end
  # println("Theta final: $(theta[length(theta)])")
  # println("Total de iterações: $i")
  return theta[end]
end

# Resolução do item a
function monte_carlo_a(M,  , n, theta0)
  d = Gamma(theta0, 1)
  function mc()
    x = rand(d, n)
    EMV = newton_raphson(x)
    c2 = quantile(Normal(),   + (1- )/2)
    I1 = EMV - c2 * sqrt(1/(n*trigamma(EMV)))
    I2 = EMV + c2 * sqrt(1/(n*trigamma(EMV)))
    return I1, I2
  end

  inferiores = []; superiores = []
  for _ in 1:M
    intervalo = mc()
    push!(inferiores, intervalo[1])
    push!(superiores, intervalo[2])
  end

  acertos = inferiores .<= theta0 .<= superiores
  return mean(acertos)

end
```

Rodando o código para o item a, temos

Confiança para 10000 simulações com alvo $0.95, n = 100, \theta_0 = 100:$ 95.07%

Confiança para 10000 simulações com alvo $0.99, n = 100, \theta_0 = 100:$ 99.03%

Confiança para 10000 simulações com alvo $0.99, n = 5, \theta_0 = 75:$ 98.91%

Para o item b,

```
# Resolução do item b
function monte_carlo_b(M,  , n, theta0)
  d = Gamma(theta0, 1)
  g(a) = cdf(Gamma(a, 1), 950)
  g1(a) = pdf(Gamma(a,1), 950)
  function mc()
    x = rand(d, n)
    EMV = newton_raphson(x)
    c2 = quantile(Normal(),   + (1- )/2)
    I1 = g(EMV) - c2 * sqrt(g1(EMV)^2/(n*trigamma(EMV)))
    I2 = g(EMV) + c2 * sqrt(g1(EMV)^2/(n*trigamma(EMV)))
    return I1, I2
  end

  inferiores = []; superiores = []
  for _ in 1:M
    intervalo = mc()
    push!(inferiores, intervalo[1])
    push!(superiores, intervalo[2])
  end

  acertos = inferiores .<= g(theta0) .<= superiores
  return mean(acertos)

end
```

Rodando o código para o item b, temos

Confiança para 10000 simulações com alvo $0.95, n = 100, \theta_0 = 1000:$ 94.88%

Confiança para 10000 simulações com alvo $0.99, n = 100, \theta_0 = 951:$ 98.61%

Confiança para 10000 simulações com alvo $0.99, n = 5, \theta_0 = 951:$ 94.67%

Para o item c, precisaremos calcular as derivadas:

$$
\begin{aligned}
g(\theta) &= \ln f_\theta(\theta) = -\ln\Gamma(\theta) + (\theta-1)\ln(\theta) - \theta \\
\Rightarrow g'(\theta) &= -\psi_1(\theta) + \ln(\theta) + \frac{\theta-1}{\theta} - 1 \\
&= -\psi_1(\theta) + \ln(\theta) - \frac{-1}{\theta} = g1(\theta)
\end{aligned}
$$

```
# Resolução do item c
function monte_carlo_c(M,  , n, theta0)
  d = Gamma(theta0, 1)
  g(a) =  log(pdf(Gamma(a, 1), a))
  g1(a) = -digamma(a) + log(a) - 1/a
  function mc()
    x = rand(d, n)
    EMV = newton_raphson(x)
    c2 = quantile(Normal(),   + (1- )/2)
    I1 = g(EMV) - c2 * sqrt(g1(EMV)^2/(n*trigamma(EMV)))
    I2 = g(EMV) + c2 * sqrt(g1(EMV)^2/(n*trigamma(EMV)))
    return I1, I2
  end

  inferiores = []; superiores = []
  for _ in 1:M
    intervalo = mc()
    push!(inferiores, intervalo[1])
    push!(superiores, intervalo[2])
  end
  acertos = inferiores .<= g(theta0) .<= superiores
  return mean(acertos)
end
```

Rodando o código para o item c, temos

Confiança para 10000 simulações com alvo $0.95, n = 100, \theta_0 = 100:$ 94.77%

Confiança para 10000 simulações com alvo $0.99, n = 100, \theta_0 = 100:$ 99.06%

Confiança para 10000 simulações com alvo $0.99, n = 5, \theta_0 = 75:$ 98.97%

Para o item d, note que

$$g'(\theta) = \mathrm{e}^{\ln f_\theta(x)}(-\psi_1(\theta) + \ln(\theta) - 1/\theta)$$

```
# Resolução do item d
function monte_carlo_d(M,  , n, theta0)
  d = Gamma(theta0, 1)
  g(a) = exp(log(pdf(Gamma(a, 1), a))) # usado a expp(log), poderia usar direto
  g1(a) = exp(log(pdf(Gamma(a,1), a))) * (-digamma(a) + log(a) - 1/a)
  function mc()
    x = rand(d, n)
    EMV = newton_raphson(x)
```

```
    c2 = quantile(Normal(),   + (1- )/2)
    I1 = g(EMV) - c2 * sqrt(g1(EMV)^2/(n*trigamma(EMV)))
    I2 = g(EMV) + c2 * sqrt(g1(EMV)^2/(n*trigamma(EMV)))
    return I1, I2
  end

  inferiores = []; superiores = []
  for _ in 1:M
    intervalo = mc()
    push!(inferiores, intervalo[1])
    push!(superiores, intervalo[2])
  end
  acertos = inferiores .<= g(theta0) .<= superiores
  return mean(acertos)
end
```

Rodando o código para o item d, temos

Confiança para 10000 simulações com alvo 0.95, $n = 100, \theta_0 = 100:$ 95.28%

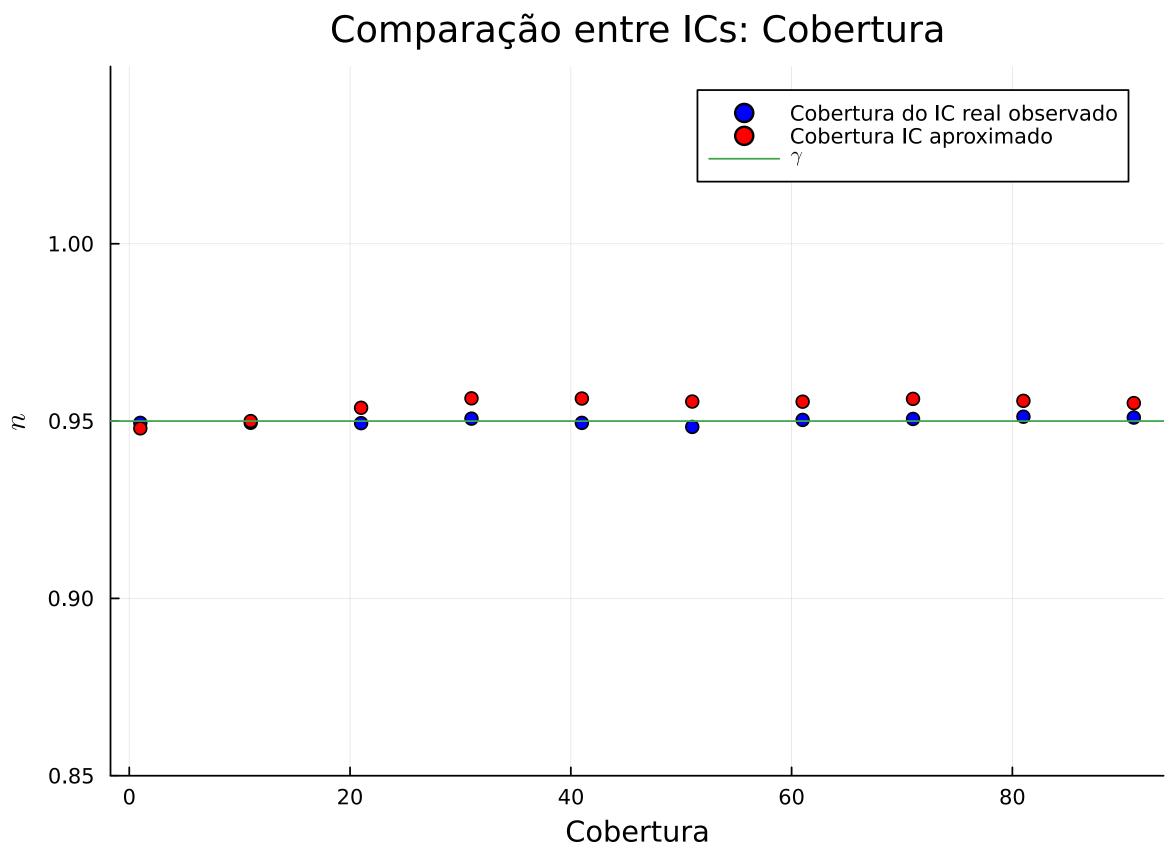
Confiança para 10000 simulações com alvo 0.99, $n = 100, \theta_0 = 100:$ 99.02%

Confiança para 10000 simulações com alvo 0.99, $n = 5, \theta_0 = 75:$ 98.83999999999999%

41.3.2.1 Comparação entre o exato e aproximado

```
using Plots, LaTeXStrings, Distributions, StatsBase, Random

function monte_carlo_realxaprox()
  theta0 = 10
  M = 100_000
  nn = 100

    = 0.95

  g(a) = a
  g1(a) = 1

  w = 0
  cober = zeros(nn)
  cobera = zeros(nn)
  for n in 1:nn
```

```
    w += 1
    I1 = zeros(M)
    Ia1 = zeros(M)
    I2 = zeros(M)
    Ia2 = zeros(M)

    for i in 1:M
      x = rand(Exponential(1/theta0), n)
      c1 = quantile(Chisq(2n), (1- )/2)
      c2 = quantile(Chisq(2n),   + (1- )/2)
      I1[i] = c1/(2*sum(x))
      I2[i] = c2/(2*sum(x))

      ca2 = quantile(Normal(),   + (1- )/2)
      EMV =  1/mean(x)
      Ia1[i] = g(EMV) - ca2 * sqrt(g1(EMV)^2 * EMV^2 /n)
      Ia2[i] = g(EMV) + ca2 * sqrt(g1(EMV)^2 * EMV^2 /n)
    end
    cober[w] = mean(I1 .<= theta0 .<= I2)
    cobera[w] = mean(Ia1 .<= g(theta0) .<= Ia2)
  end
  preal = scatter(1:10:100, cober, color="blue",
                  label="Cobertura do IC real observado",
                  title="Comparação entre ICs: Cobertura",
                  ylims=( -0.1,  +0.1),
                  xlabel="Cobertura",
                  ylabel=L"n")
  scatter!(1:10:100, cobera, color="red", label="Cobertura IC aproximado")
  hline!([ ], label=L"\gamma")
  return preal
end
```



41.3.2.2 Exemplo poisson

Note que0

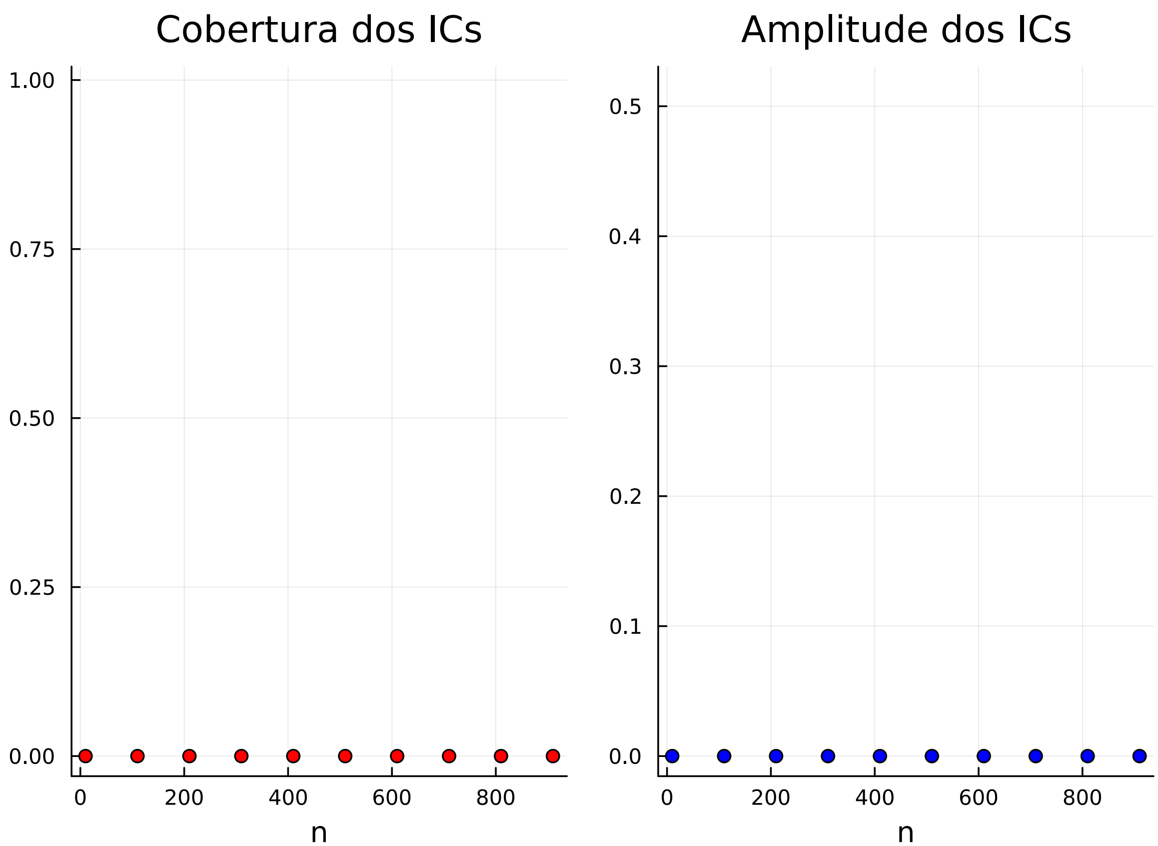
$$\hat{\theta}_{\mathrm{MV}}(X_n) = \bar{X}$$

$$\sqrt{n}\frac{\hat{\theta}_{\mathrm{MV}}(X_n) - \theta}{\sqrt{\theta}} \xrightarrow{\mathcal{D}} N(0,1)$$

$$\sqrt{n}\frac{\hat{\theta}_{\mathrm{MV}}(X_n) - \theta}{\sqrt{\hat{\theta}_{\mathrm{MV}}(X_n)}} \xrightarrow{\mathcal{D}} N(0,1)$$

```
function monte_carlo_poiss()
  theta0 = 10
  d = Poisson(theta0)
  g(a) = a
  g1(a) = 1
  nn = 1000
  amplitude = zeros(nn)
```

```
  cober = zeros(nn)
  M = 10_000
  for n in nn
    I1 = zeros(M)
    I2 = zeros(M)
    for i in 1:M
      x = rand(d, n)
      EMV = mean(x)
      c2 = quantile(Normal(),   + (1- )/2)
      I1[i] = EMV - c2 * sqrt(g1(EMV)^2 * EMV/n)
      I2[i] = EMV + c2 * sqrt(g1(EMV)^2 * EMV/n)
    end
    amplitude[n] = mean(I2 .- I1)
    cober[n] = mean(I1 .<= g(theta0) .<= I2)
  end
  amp = scatter(10:100:1000, amplitude, title="Amplitude dos ICs", label="",
              color=:blue, xlabel="n")
  cob = scatter(10:100:1000, cober, title="Cobertura dos ICs", label="",
              color=:red, xlabel="n")
  plt = plot(cob, amp)
  return plt
end
```



41.4 Regiões de confiança

Dizemos que $\text{RC}^{(a)}(g(\theta), \gamma)$ é uma região de confança para $g(\theta)$ com coeficiente de confiança γ se, e somente se,

$$P_\theta(\theta \in \text{RC}(g(\theta), \gamma)) = \gamma, \forall \theta \in \Theta.$$

Se $\lim_{n\to\infty} P_\theta(\theta \in \text{RC}(g(\theta), \gamma)) = \gamma, \forall \theta \in \Theta$, dizemos que $\text{RC}^{(a)}(g(\theta), \gamma)$ é uma região de confiança assintótica.

Observação

Se

$$P_\theta(\theta \in \text{RC}(g(\theta), \gamma)) \geq \gamma,$$

então dizemos que $\text{RC}(g(\theta), \gamma))$ tem confiança de pelo menos γ.

Observação

RC($g(\theta), \gamma$)) é uma região aleatória que depende apenas da amostra aleatória e do coeficiente γ

41.4.1 Construção da Região de Confiança

Seja $\hat{\theta}(X_n)$ um estimador para "θ" tal que

1. $$\sqrt{n}(\hat{\theta}(X_n) - \theta) \overset{\mathcal{D}}{\to} N_p(0, V_\theta), \forall \theta \in \Theta.$$

2. $$\sqrt{n} V_\theta^{-1/2}(\hat{\theta}(X_n) - \theta) \overset{\mathcal{D}}{\to} N_p(0, I), \forall \theta \in \Theta.$$

3. $$\sqrt{n} V_{\hat{\theta}(X_n)}^{-1/2}(\hat{\theta}(X_n) - \theta) \overset{\mathcal{D}}{\to} N_p(0, I), \forall \theta \in \Theta.$$

4. $$n(\hat{\theta}(X_n) - \theta)^T V_{\hat{\theta}(X_n)}^{-1}(\hat{\theta}(X_n) - \theta) \overset{\mathcal{D}}{\to} \chi_p^2, \forall \theta \in \Theta.$$

Podemos usar 4. para construir uma RC para "θ" assintótica:

$$\text{RC}^{(a)}(\theta, \gamma) = \left\{\theta \in \Theta : W_n(\theta) \leq q_\gamma\right\},$$

em que q_γ satisfaz

$$P(\chi_p^2 \leq q_\gamma) = \gamma.$$

Assim, por construição,

$$\lim_{n \to \infty} P_\theta(\theta \in \text{RC}^{(a)}(\theta, \gamma)) = \gamma, \forall \theta \in \Theta.$$

Para o caso $p = 1$, ou seja, $\Theta \in \mathbb{R}$, note que

$$n(\hat{\theta}(X_n) - \theta)^T V_{\hat{\theta}(X_n)}^{-1}(\hat{\theta}(X_n) - \theta) = \frac{n(\hat{\theta}(X_n) - \theta)^2}{V_{\hat{\theta}(X_n)}} = W_n(\theta).$$

Portanto,

$$
\begin{aligned}
&\mathrm{RC}(\theta,\gamma) = \left\{\theta\in\Theta : \frac{n(\hat\theta(X_n)-\theta)^2}{V_{\hat\theta(X_n)}} \leq q_\gamma\right\}\\
\iff &\mathrm{RC}(\theta,\gamma) = \left\{\theta\in\Theta : |\hat\theta(X_n)-\theta| \leq \sqrt{q_\gamma\cdot\frac{V_{\hat\theta(X_n)}}{n}}\right\}\\
\iff &\mathrm{RC}(\theta,\gamma) = \left\{\theta\in\Theta : -\sqrt{q_\gamma\cdot\frac{V_{\hat\theta(X_n)}}{n}} \leq \hat\theta(X_n)-\theta \leq \sqrt{q_\gamma\cdot\frac{V_{\hat\theta(X_n)}}{n}}\right\}\\
\iff &\mathrm{RC}(\theta,\gamma) = \left\{\theta\in\Theta : \hat\theta(X_n) - \sqrt{q_\gamma}\sqrt{\frac{V_{\hat\theta(X_n)}}{n}} \leq \theta \leq \hat\theta(X_n) + \sqrt{q_\gamma}\sqrt{\frac{V_{\hat\theta(X_n)}}{n}}\right\}
\end{aligned}
$$

em que q_γ é tal que $P(\chi^2_1 \leq q_\gamma) = \gamma$.

No caso em que Θ é uma semireta ou intervalo, a região de confiança coincide com o intervalo de confiança.

Observação

$$P(\chi^2_1 \leq q_\gamma) = P(-\sqrt{q_\gamma} \leq N(0,1) \leq \sqrt{q_\gamma}) = \gamma$$

Para o caso $p=2$, ou seja, $\Theta=\mathbb{R}^2$, a razão de confiança aproximada é

$$\mathrm{RC}(\theta,\gamma) = \left\{\theta\in\Theta : W_n(\theta) \leq q_\gamma\right\},$$

42 Teste de Hipótese - Aprofundamento

Seja $X_n = (X_1, \dots, X_n)$ amostra aleatória de $X \sim f_\theta, \theta \in \Theta$.

Note que, no caso discreto:

$$P_\theta(X \in A) = \sum_{x \in A} f_\theta(x), \forall \theta \in \Theta.$$

já no caso contínuo,

$$P_\theta(X \in A) = \int_A f_\theta(x), \forall \theta \in \Theta.$$

Temos uma família $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ de probabilidades que podem explicar o comportamento dos dados. Um dos objetivos do teste de hipótese é verificar se podemos reduzir $\mathcal{P}$ para uma família menor $\mathcal{P}_0 \subseteq \mathcal{P}$:

$$\mathcal{P}_0 = \{P_\theta : \theta \in \Theta_0\}, \Theta_0 \subseteq \Theta.$$

Definimos uma *hipótese estatística*

$$\mathcal{H}_0 : \theta \in \Theta_0 \iff \mathcal{H}_0 : P_\theta \in \mathcal{P}_0$$

$\mathcal{H}_0$ afirma que

> "A medida de probabilidade que explica os dados está em $\mathcal{P}_0$"

No caso em que $\Theta_0 = \{\theta_0\}$, então a hipótese $\mathcal{H}_0 : \theta = \theta_0 (\iff \mathcal{H}_0 : P_\theta = P_{\theta_0})$ afirma que P_{θ_0} explica o comportamento probabilístico dos dados observados.

$\mathcal{H}_0$ é chamada de *hipótese nula*.

Note que a negação de $\mathcal{H}_0$ é

> "Não é o caso que a família $\mathcal{P}_0$ contenha a medida que explica os dados".

Ou seja, a negação de $\mathcal{H}_0$ sugere que a medida que explica os dados pode, inclusive, não estar contida em $\mathcal{P}$.

Definimos também uma hipótese alternativa,

$$\mathcal{H}_1 : \theta \in (\Theta \setminus \Theta_0) \iff \mathcal{H}_1 : P_\theta \in (\mathcal{P} \setminus \mathcal{P}_0)$$

Observe que $\mathcal{H}_0$ e $\mathcal{H}_1$ não são exaustivas: na prática, ambas podem ser falsas.

Se o analista considerar que $\mathcal{P}$ contém a medida que explica os dados, então, condicional a essa crença, $\mathcal{H}_0$ e $\mathcal{H}_1$ são exaustivas.

42.1 Exemplo

Seja X_n a.a. de $X \sim N(\mu, \sigma^2), \theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$.

Considere as seguintes hipóteses

$$\begin{cases} \mathcal{H}_0 : \mu = 0 \\ \mathcal{H}_1 : \mu \neq 0. \end{cases}$$

Note que

$$\begin{cases} \mathcal{H}_0 : \theta \in \Theta_0 \\ \mathcal{H}_1 : \theta \in \Theta \setminus \Theta_0, \end{cases}$$

em que

$$\begin{aligned} \Theta_0 &= \{(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+ : \mu = 0\} \\ \Theta_1 = \Theta \setminus \Theta_0 &= \{(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+ : \mu \neq 0\}. \end{aligned}$$

Disso, temos que

$$\begin{aligned} \Theta &= \Theta_0 \cup \Theta_1 \\ \Theta_0 \cap \Theta_1 &= \emptyset. \end{aligned}$$

As hipóteses acima podem ser reescritas em termos das suas respectivas famílias de medidas de probabilidades da seguinte forma:

$$\begin{cases} \mathcal{H}_0 : P \in \mathcal{P}_0 \\ \mathcal{H}_1 : P \in \mathcal{P}_1 \end{cases}$$

em que $\mathcal{P}_0 = \{N(\mu, \sigma^2) : \ \mu = 0, \ \sigma^2 > 0\}$ e $\mathcal{P}_1 = (\mathcal{P} - \mathcal{P}_0) = \{N(\mu, \sigma^2) : \ \mu \neq 0, \ \sigma^2 > 0\}$

Observe que, se assumirmos que a distribuição normal explica os dados, então $\mathcal{H}_0$ e $\mathcal{H}_1$ são exaustivas. Caso contrário, existe uma terceira opção: $\neg(\mathcal{H}_0 \vee \mathcal{H}_1)$.

42.2 Hipóteses estatísticas

Toda hipótese estatística é uma interpretação de uma hipótese científica.

Hipótese científica	Hipótese estatística	Hipótese paramétrica
“A moeda é honesta”	$X \sim \text{Ber}(0.5)$	$\theta = 0.5$

As hipótese estatísticas são escritas em termos de medidas de probabilidade, mas podem também ser representadas em termos do espaço paramétrico:

$$\begin{cases} \mathcal{H}_0 : P_\theta \in \mathcal{P}_0 \\ \mathcal{H}_1 : P_\theta \in \mathcal{P}_1 = (\mathcal{P} \setminus \mathcal{P}_0), \end{cases}$$

em que $\mathcal{P}_0 = \{P_\theta : \theta \in \Theta_0\}$ e $\mathcal{P}_1 = \{P_\theta : \theta \in \Theta_1\}$ e $\Theta_1 = \Theta \setminus \Theta_0$.

Observe que $\mathcal{H}_0, \mathcal{H}_1$ estão sempre restritas a modelo estatístico. Entretanto, a negação de $\mathcal{H}_0$, $\neg\mathcal{H}_0$, não está restrita ao modelo adotado.

i Algumas considerações lógicas

1. Se o analista considerar que $\mathcal{H}_0$ ou $\mathcal{H}_1$ são verdadeiros, então ele está considerando que a medida que *explica* ou *gera* os dados está em $\mathcal{P}$. Está é a suposição de que o universo de possibilidades é fechado (suposição de Neyman-Pearson; Teoria da Decisão).
2. $\mathcal{H}_0$ e $\mathcal{H}_1$ não pode ser simultaneamente verdadeiras.
3. Se $\neg\mathcal{H}_0$ e $\neg\mathcal{H}_1$, então a medida que explica/gera os dados não está em $\mathcal{P}$.
4. $\neg\mathcal{H}_0 \nRightarrow \mathcal{H}_1$. Isto é, se provarmos que $\mathcal{H}_0$ é falsa, não necessariamente $\mathcal{H}_1$ é verdadeira.
5. $\mathcal{H}_1 \Rightarrow \neg\mathcal{H}_0$.

42.3 Tipos de hipótese

Sejam $\mathcal{H}_0, \mathcal{H}_1$ as hipóteses nula e alternativa, respectivamente, tais que

$$\begin{cases} \mathcal{H}_0 : \theta \in \Theta_0 \\ \mathcal{H}_1 : \theta \in \Theta_1 \end{cases}$$

em que $\Theta_0 \neq \emptyset, \Theta_0 \cup \Theta 1 = \Theta, \Theta_0 \cap \Theta_1 = \emptyset$. Dizemos que $\mathcal{H}_0$ é uma **hipótese simples** se $\#\Theta_0 = 1$. Caso contrário, dizemos que é uma **hipótese composta**.

42.3.1 Exemplos

1. Seja $X_n = (X_1, \ldots, X_n)$ a.a. de $X \sim \text{Ber}(\theta), \theta \in \{0.5, 0.9\}$. Considere as hipóteses

$$\begin{cases} \mathcal{H}_0 : \theta = 0.5 \\ \mathcal{H}_1 : \theta = 0.9 \end{cases} \quad \text{Ambas são simples!}$$

2. Seja $X_n = (X_1, \ldots, X_n)$ a.a. de $X \sim \text{Ber}(\theta), \theta \in (0, 1)$. Considere as hipóteses

$$\begin{cases} \mathcal{H}_0 : \theta = 0.5 & \text{Hipótese simples} \\ \mathcal{H}_1 : \theta \neq 0.5 & \text{Hipótese composta} \end{cases}$$

Note que

$$\begin{cases} \Theta_0 = \{0.5\} \\ \Theta_1 = (0, 0.5) \cup (0.5, 1) \end{cases}$$

3. Seja X_n a.a. de $X \sim N(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$

$$a) \begin{cases} \mathcal{H}_0 : \mu = 0.5 & \text{Hipótese composta!} \\ \mathcal{H}_1 : \mu \neq 0.5 & \text{Hipótese composta} \end{cases}$$

Note que como

$$\Big\{\Theta_0 = \{(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+ : \mu = 0\}\Theta_0 = \{(\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+ : \mu \neq 0\}$$

têm mais de um elemento, concluímos que ambas hipóteses são compostas.

$$b) \begin{cases} \mathcal{H}_0 : (\mu, \sigma^2) = (0, 1) & \text{Hipótese simples!} \\ \mathcal{H}_1 : (\mu, \sigma^2) \neq (0, 1) & \text{Hipótese composta} \end{cases}$$

42.3.2 Tipos de decisão sobre as hipóteses

1. Se o espaço de possibilidades for fechado, isto é, $\mathcal{H}_0$ ou $\mathcal{H}_1$ é verdadeira, então:

a) Aceitamos $\mathcal{H}_0$ (rejeitamos $\mathcal{H}_1$)
b) Aceitamos $\mathcal{H}_1$ (rejeitamos $\mathcal{H}_0$)

Observe que, neste caso, não há terceira opção (abordagem de Neyman-Pearson).

2. Se o espaço de possibilidades for aberto, isto é, temos a terceira opção de que o modelo está equivocado, então:

a) Não rejeitamos $\mathscr{H}_0$ (**Não** significa que aceitamos $\mathscr{H}_0$)
b) Rejeitamos $\mathscr{H}_0$ (**Não** significa aceitar $\mathscr{H}_1$, pela existência da terceira opção.)

Neste caso temos uma terceira opção (abordagem de Fisher).

Atualmente, dizemos apenas que há ou não há evidências para rejeitarmos $\mathscr{H}_0$. **Não** se diz aceitar "$\mathscr{H}_0$"

42.3.3 Tipos de Erro

Se o universo for aberto:

Erro tipo I: "Rejeitar $\mathscr{H}_0$ quando este é verdadeiro"

Erro tipo II: "Não rejeitar $\mathscr{H}_0$ quando este é falso"

$\mathscr{H}_0$	Não rejeitar	Rejeitar
Verdadeiro	Acerto	Erro tipo I
Falso	Erro tipo II	Acerto

Se o universo for fechado:

Erro tipo I: "Rejeitar $\mathscr{H}_0$ (ou aceitar $\mathscr{H}_1$) quando este é verdadeiro"

Erro tipo II: "Aceitar $\mathscr{H}_0$ (ou rejeitar $\mathscr{H}_1$ quando este é falso"

42.4 Teste de Hipótese de Neyman-Pearson

Considere o universo fechado.

42.4.1 Função teste

Definição. Seja $\delta : \mathfrak{X}^{(n)} \to \{0, 1\}$ uma função tal que

$$\delta(X_n) = \begin{cases} 0, & \text{Se não rejeitamos } \mathscr{H}_0 \\ 1, & \text{Se rejeitamos } \mathscr{H}_0 \end{cases}$$

em que $\mathfrak{X}^{(n)} = \{x \in \mathbb{R}^n : f_\theta^{X_n}(x) > 0\}$, $\mathscr{H}_0 : \theta \in \Theta_0$, $\mathscr{H}_1 : \theta \in \Theta_1$, $\Theta_0 \neq \emptyset, \Theta_1 \neq \emptyset, \Theta_0 \cap \Theta_1 = \emptyset, \Theta_0 \cup \Theta_1 = \Theta$.

Dizemos que $\delta(\cdot)$ é uma *função teste*.

Região de rejeição de $\mathcal{H}_0$:

$$S_\delta = \{x \in \mathfrak{X}^{(n)} : \delta(x) = 1\}$$

Região de não rejeição de $\mathcal{H}_0$:

$$S_\delta^c = \{x \in \mathfrak{X}^{(n)} : \delta(x) = 0\}$$

42.4.1.1 Exemplo

Seja X_n a.a. de $X \sim \text{Ber}(\theta), \theta \in \{0.1, 0.9\}$ e as hipóteses

$$\begin{cases} \mathcal{H}_0 : \theta = 0.1 \\ \mathcal{H}_1 : \theta = 0.9 \end{cases}$$

Defina $\delta_i : \{0,1\}^n \to \{0,1\}, i = 1,2$ tais que

$$\delta_1(x_n) = \begin{cases} 0, \bar{x} \leq 0.5 \\ 1, \bar{x} > 0.5 \end{cases}$$
$$\delta_2(x_n) = \begin{cases} 0, \bar{x} \leq 0.8 \\ 1, \bar{x} > 0.8 \end{cases}$$

42.4.2 Função Poder do teste δ

Definição. A função poder do teste δ é $\pi_\delta(\theta) = P_\theta(S_\delta)$

Observações

1. $$\pi_\delta(\theta), \forall \theta \in \Theta_0,$$
   são probabilidades de rejeitar $\mathcal{H}_0$ quando esta é verdadeira.

2. $$\pi_\delta(\theta), \forall \theta \in \Theta_1,$$
   são probabilidades de rejeitar $\mathcal{H}_0$ quando esta é falsa.

Definição. O *nível de significância* é qualquer valor $\alpha \in (0,1)$ tal que:

$$P_\theta(S_\delta) \leq \alpha, \forall \theta \in \Theta_0.$$

em termos de função poder,

$$\pi_\delta(\theta) \leq \alpha, \forall\theta \in \Theta_0$$

Definição. O tamanho do teste δ é

$$\tau_\delta = \sup_{\theta\in\Theta_0} P_\theta(S_\delta) \leq \alpha$$

em termos de função poder,

$$\tau_\delta = \sup_{\theta\in\Theta_0} \pi_\delta(\theta)$$

42.4.2.1 Probabilidade dos Erros I, II

> **Tipos de Erro**
>
> Relembrando os tipos de erro:
>
> $$\begin{cases} \text{Erro tipo I: Rejeitar } \mathcal{H}_0 \text{ quando esta é verdadeira} \\ \text{Erro tipo II: Não rejeitar } \mathcal{H}_0 \text{ quando esta é falsa} \end{cases}$$

$$\alpha_{\delta,\max} = \tau_\delta = \sup_{\theta\in\Theta_0} \pi_\delta(\theta)$$

é a probabilidade máxima de cometer o Erro tipo I e

$$\beta_{\delta,\max} = \sup_{\theta\in\Theta_1} (1 - \pi_\delta(\theta))$$

é a probabilidade máxima de cometer o Erro tipo II.

Quando $\mathcal{H}_0, \mathcal{H}_1$ são simples, temos

$$\begin{aligned} \alpha_{\delta,\max} &= \pi_\delta(\theta_0) = P_\theta(S_\delta) \\ \beta_{\delta,\max} &= 1 - \pi_\delta(\theta_1), \end{aligned}$$

com

$$\begin{aligned} \mathcal{H}_0 &: \theta = \theta_0 \\ \mathcal{H}_1 &: \theta \neq \theta_1. \end{aligned}$$

> **Notação (equivocada) em alguns materiais**
>
> Alguns livros, especialmente para iniciantes ou profissionais de outras áreas, podem es-

crever:

$$\begin{aligned}\alpha_{\delta,\max} &= P(\text{Erro tipo I})\\ &= P(\text{Rejeitar}|\mathcal{H}_0 \text{ é verdadeira})\\ &= P(\delta(X_n) = 1|\mathcal{H}_0 \text{ é verdadeira}),\end{aligned}$$

$$\begin{aligned}\beta_{\delta,\max} &= P(\text{Erro tipo II})\\ &= P(\text{Não rejeitar}|\mathcal{H}_0 \text{ é falsa})\\ &= P(\delta(X_n) = 0|\mathcal{H}_0 \text{ é falsa}).\end{aligned}$$

Note que, formalmente, isso **não** faz sentido na estatística clássica! O parâmetro em hipótese, θ, é desconhecido mas **não** aleatório.
Note que, na estatística clássica, as hipóteses são afirmações epistêmicas, e não experimentais (eventos do experimento). Portanto, não são elementos da σ-álgebra do modelo estatístico. Logo, a notação

$$P(S_\delta|\mathcal{H}_0 \text{ é verdadeira})$$

não está bem definida.

42.4.2.2 Poder do Teste

Se $\mathcal{H}_1$ for simples, isto é, $\mathcal{H}_1 : \theta = \theta_1$, então o poder do teste é definido

$$1 - \beta_{\delta,\max} = \pi_\delta(\theta_1)$$

⚠ Notação (equivocada) em alguns materiais

Alguns livros escrevem

$$\text{poder} = P(\text{Rejeitar } \mathcal{H}_0|\mathcal{H}_0 \text{ é falsa})$$

Isto está incorreto na interpretação clássica de estatística!

42.5 Lema de Neyman-Pearson

Estamos interessados em um teste δ^* que produza menor $\alpha_{\delta^*,\max}$ e maior poder possível. O Lema de Neyman-Pearson apresenta um teste que, dentre os testes de tamanho α, tem maior poder.

Seja X_n a.a. de $X \sim f_\theta, \theta \in \Theta = \{\theta_0, \theta_1\}$. Considere as hipóteses

$$\begin{cases}\mathcal{H}_0 : \theta = \theta_0\\ \mathcal{H}_1 : \theta = \theta_1\end{cases}$$

nula e alternativa. A função teste $\delta^*: \mathfrak{X}^{(n)} \to \{0,1\}$ tal que

$$\delta^*(x_n) = \begin{cases} 0, \text{ se } \mathcal{L}_{x_n}(\theta_1) < \eta \cdot \mathcal{L}_{x_n}(\theta_0) \\ 1, \text{ se } \mathcal{L}_{x_n}(\theta_1) \geq \eta \cdot \mathcal{L}_{x_n}(\theta_0), \end{cases}$$

em que η satisfaz $\underbrace{P_{\theta_0}(\delta^*(X_n) = 1)}_{\pi_{\delta^*}(\theta_0)} = \alpha$ para um α fixado apriori, é o teste mais poderoso dentre todo os testes de tamanho α.

42.5.1 Exemplo (Normal)

Seja X_n a.a. de $X \sim N(\mu, \sigma^2)$ em que $\theta = (\mu, \sigma^2) \in \{(0,1),(0,2)\}$. Considere

$$\begin{cases} \mathcal{H}_0 : \theta = (0,1) \\ \mathcal{H}_1 : \theta = (0,2) \end{cases}$$

as hipóteses nula e alternativa.

Encontre o teste mais poderoso de tamanho $\alpha = 5\%$.

42.5.1.1 Solução:

De acordo com o Lema de Neyman-Pearson, o teste mais poderoso de tamanho $\alpha = 5\%$ é

$$\delta^*(x_n) = \begin{cases} 0, \text{ se } \mathcal{L}_{x_n}((0,2)) < \eta \cdot \mathcal{L}_{x_n}((0,1)), \\ 1, \text{ se } \mathcal{L}_{x_n}((0,2)) \geq \eta \cdot \mathcal{L}_{x_n}((0,1)), \end{cases}$$

em que η satisfaz $\pi_{\delta^*}((0,1)) = 5\%$.

A função de verossimilhança é

$$\begin{aligned}
\mathcal{L}_{x_n}(\theta) &= \frac{1}{(\sqrt{2\pi\sigma^2})^n} \cdot \text{Exp}\left\{-\frac{1}{2}\sum \frac{(x_i-\mu)^2}{\sigma^2}\right\}, \theta \in \{(0,1),(0,2)\} \\
&= \frac{1}{(\sqrt{2\pi\sigma^2})^n} \cdot \text{Exp}\left\{-\frac{1}{2}\sum \frac{x_i^2}{\sigma^2}\right\}, \theta \in \{(0,1),(0,2)\} \\
\Rightarrow \mathcal{L}_{x_n}(\theta_0) &= \frac{1}{(\sqrt{2\pi})^n} \cdot \text{Exp}\left\{-\frac{1}{2}\sum x_i^2\right\} \\
\mathcal{L}_{x_n}(\theta_1) &= \frac{1}{(\sqrt{4\pi})^n} \cdot \text{Exp}\left\{-\frac{1}{2}\sum \frac{x_i^2}{2}\right\}
\end{aligned}$$

$$\begin{aligned}
&\Rightarrow \mathcal{L}_{x_n}((0,2)) \geq \eta \cdot \mathcal{L}_{x_n}((0,1)) \\
&\iff \frac{\mathcal{L}_{x_n}((0,2))}{\mathcal{L}_{x_n}((0,1))} \geq \eta \\
&\qquad \Rightarrow \frac{\frac{1}{(\sqrt{4\pi})^n} \cdot \text{Exp}\left\{-\frac{1}{2}\sum \frac{x_i^2}{2}\right\}}{\frac{1}{(\sqrt{2\pi})^n} \cdot \text{Exp}\left\{-\frac{1}{2}\sum x_i^2\right\}} \geq \eta \\
&\qquad \iff \frac{1}{2^{\frac{n}{2}}}\mathrm{e}^{\frac{1}{4}\sum x_i^2} \geq \eta \\
&\qquad \iff \mathrm{e}^{\frac{1}{4}\sum x_i^2} \geq 2^{n/2}\eta \\
&\qquad \iff \frac{1}{4}\sum x_i^2 \geq \ln(2^{n/2}\eta) \\
&\qquad \iff \sum x_i^2 \geq 4\ln(2^{n/2}\eta).
\end{aligned}$$

Logo, $\sum x_i^2 \geq 4\ln(2^{n/2}\eta)$ é equivalente a $\frac{\frac{1}{(\sqrt{4\pi})^n} \cdot \text{Exp}\left\{-\frac{1}{2}\sum \frac{x_i^2}{2}\right\}}{\frac{1}{(\sqrt{2\pi})^n} \cdot \text{Exp}\left\{-\frac{1}{2}\sum x_i^2\right\}}$.

Observe ainda que

$$\begin{aligned}
\pi_{\delta^*}((0,1)) &= P_{(0,1)}(\delta^*(X_n)=1) \\
&\overset{\text{Def.}}{=} P_{(0,1)}\left(\frac{\mathcal{L}_{x_n}((0,2))}{\mathcal{L}_{x_n}((0,1))} \geq \eta\right) \\
&\overset{\text{Desenv.}}{=} P_{(0,1)}\left(\sum X_i^2 \geq \eta^*\right),
\end{aligned}$$

em que $\eta^* = 4\ln(2^{n/2}\eta)$. Como, sob $\mathcal{H}_0, \sum X_i^2 \sim \chi_n^2$, temos que

$$\pi_{\delta^*}((0,1)) = P_{(0,1)}(\chi_n^2 \geq \eta^*) = 5\%.$$

Com $n=2$, temos que

$$\pi_{\delta^*}((0,1)) = P_{(0,1)}(\chi_n^2 \geq \eta^*) = 5\%. \Rightarrow \eta^* = 5.991.$$

Portanto, o teste mais poderoso é

$$\delta^*(x_n) = \begin{cases} 0, \text{ se } \sum x_i^2 < 5.991 \\ 1, \text{ se } \sum x_i^2 \geq 5.991. \end{cases}$$

42.5.2 Exemplo (Bernoulli)

Seja X_n a.a. de $X \sim \text{Ber}(\theta), \theta \in \{0.1, 0.9\}$. Considere

$$\begin{cases} \mathcal{H}_0 : \theta = 0.9 \\ \mathcal{H}_1 : \theta = 0.1 \end{cases}$$

as hipóteses nula e alternativa e $n = 10$. Encontre o Teste Mais Poderoso (TMP) de tamanho $\alpha = 10\%$.

42.5.2.1 Resposta

De acordo com o Lema de Neyman-Pearson, o teste mais poderoso de tamanho $\alpha = 5\%$ é

$$\delta^*(x_n) = \begin{cases} 0, \text{ se } \mathcal{L}_{x_n}(0.1) < \eta \cdot \mathcal{L}_{x_n}(0.9), \\ 1, \text{ se } \mathcal{L}_{x_n}(0.1) \geq \eta \cdot \mathcal{L}_{x_n}(0.9), \end{cases}$$

em que η satisfaz $\pi_{\delta^*}(0.9) = 10\%$.

Note que

$$\begin{aligned} \mathcal{L}_{x_n}(\theta) &= \theta^{\sum x_i}(1-\theta)^{n-\sum x_i} \\ \Rightarrow \frac{\mathcal{L}_{x_n}(0.1)}{\mathcal{L}_{x_n}(0.9)} &= \frac{0.1^{\sum x_i}(0.9)^{n-\sum x_i}}{0.9^{\sum x_i}(0.1)^{n-\sum x_i}} \\ &= \frac{0.1^{\sum x_i} \cdot 0.9^n \cdot 0.1^{\sum x_i}}{0.9^{\sum x_i} \cdot 0.9^{\sum x_i} \cdot 0.1^n} \\ &= \left(\frac{0.1 \cdot 0.1}{0.9 \cdot 0.9}\right)^{\sum x_i} \cdot \left(\frac{0.9}{0.1}\right)^n \geq \eta \\ \Leftrightarrow (\sum x_i) \ln\left(\frac{0.01}{0.81}\right) &+ n\ln(9) \geq \ln \eta \\ \Leftrightarrow (\sum x_i) \ln\left(\frac{1}{81}\right) &\geq \ln \eta - n\ln(9) \\ \Leftrightarrow (-\sum x_i) \ln(81) &\geq \ln \eta - n\ln(9) \\ \Leftrightarrow -\sum x_i &\geq \frac{\ln \eta - n\ln(9)}{\ln(81)} \\ \Leftrightarrow \sum x_i &\leq \frac{-\ln \eta + n\ln(9)}{\ln(81)} \end{aligned}$$

Logo, $\left(\frac{0.1\cdot 0.1}{0.9\cdot 0.9}\right)^{\sum x_i} \cdot \left(\frac{0.9}{0.1}\right)^n \geq \eta \iff \sum x_i \leq \frac{-\ln\eta + n\ln(9)}{\ln(81)}$.

$$\pi_{\delta^*}(0.1) = P_{0.1}(\delta^*(X_n) = 1) = P_{0.1}(\sum X_i \leq \eta^*) = \alpha = 10\%$$

Note que, sob $\mathcal{H}_0, \sum^1 0_{i=1} X_i \sim \text{Bin}(10, 0.9)$. Ademais,

$$P_{0.9}(\sum X \leq 7) = 0.0702, \quad P_{0.9}(\sum X \leq 8) = 0.2639.$$

Logo, não é possível encontrar um valor para η exato com esse tamanho do teste.

42.6 Testes uniformemente mais poderosos (TUMP)

Considere

$$\begin{cases} \mathcal{H}_0 : \theta \in \Theta_0 \\ \mathcal{H}_1 : \theta \in \Theta_1 \end{cases}$$

as hipóteses nula e alternativa em que $\Theta_0 \cup \Theta_1 = \Theta, \Theta_0 \cap \Theta_1 = \emptyset, \#(\Theta_0) \geq 1, \#(\Theta_1) \geq 1$. O teste de tamanho α $\delta^* : \mathfrak{X}^{(n)} \to \{0,1\}$ é uniformemente mais poderoso se, e somente se,

1. $\sup\limits_{\theta\in\Theta_0} \pi_{\delta^*}(\theta) = \alpha$;
2. Para qualquer outro teste δ tal que $\pi_\delta(\theta) \leq \alpha, \underbrace{\forall\theta \in \Theta_0}_{\text{Sob } \mathcal{H}_0}$, então

$$\pi_{\delta^*}(\theta) \geq \pi_\delta(\theta), \underbrace{\forall\theta \in \Theta_1}_{\text{Sob } \mathcal{H}_1}.$$

42.6.1 Teorema de Karlin-Rubin (I)

Uma generalização do Lema de Neyman-Pearson.

Seja X_n a.a. de $X \sim f_\theta, \theta \in \Theta \subseteq \mathbb{R}$. Considere

$$\begin{cases} \mathcal{H}_0 : \theta \leq \theta_0 \\ \mathcal{H}_1 : \theta > \theta_0 \end{cases}$$

em que θ_0 é um valor dado. Se a *razão de verossimilhanças*

$$\mathrm{R}_{\theta',\theta''}(T(x_n)) = \frac{\mathcal{L}_{x_n}(\theta')}{\mathcal{L}_{x_n}(\theta'')}$$

for *monótona não-decrescente* e $\theta' \geq \theta''$, em que $T(x_n)$ é uma estatística suficiente para o modelo, então o teste $\delta^* : \mathfrak{X}^{(n)} \to \{0,1\}$ tal que

$$\delta^*(x_n) = \begin{cases} 0, T(x_n) < c \\ 1, T(x_n) \geq c, \end{cases}$$

e c satisfaz

$$\pi_{\delta^*}(\theta_0) = P_{\theta_0}(\delta^*(x_n) = 1) = \alpha,$$

é o teste uniformemente mais poderoso de tamanho α.

42.6.1.1 Exemplo (Normal)

Seja X_n a.a. de $X \sim N(\theta, 2), \theta \in \Theta \subseteq \mathbb{R}$. Considere

$$\begin{cases} \mathcal{H}_0 : \theta \leq 1 \\ \mathcal{H}_1 : \theta > 1. \end{cases}$$

Encontre o TUMP de tamanho $\alpha = 10\%$.

42.6.1.1.1 Resposta

A função de verossimilhança é dada por

$$\begin{aligned} \mathcal{L}_{x_n}(\theta) &= \left(\frac{1}{\sqrt{4\pi}}\right)^n \operatorname{Exp}\left\{-\frac{1}{4}\sum (x_i - \theta)^2\right\} \\ &= \left(\frac{1}{\sqrt{4\pi}}\right)^n \operatorname{Exp}\left\{-\frac{1}{4}\sum x_i^2 - 2\theta \sum x_i + n\theta^2\right\}. \end{aligned}$$

Pelo Critério de Fatoração de Fisher, $T(X_n) = \sum x_i$ é uma estatística suficiente. Note que

$$\begin{aligned} \mathrm{R}_{\theta',\theta''}(T(X_n)) &= \frac{\left(\frac{1}{\sqrt{4\pi}}\right)^n \operatorname{Exp}\left\{-\frac{1}{4}\sum x_i^2 - 2\theta' \sum x_i + n\theta'^2\right\}}{\left(\frac{1}{\sqrt{4\pi}}\right)^n \operatorname{Exp}\left\{-\frac{1}{4}\sum x_i^2 - 2\theta'' \sum x_i + n\theta''^2\right\}} \\ &= \operatorname{Exp}\left\{-\frac{1}{4}(-\theta' T(x_n) + n\theta'^2 + 2\theta'' T(x_n) - n\theta''^2\right\} \\ &= \operatorname{Exp}\left\{\frac{1}{2}(\theta' - \theta'')T(x_n) - \frac{1}{4}n\theta''^2 + \frac{1}{4}\theta''^2\right\}. \end{aligned}$$

Como $\theta' \geq \theta''$, temos que $\theta' - \theta'' \geq 0$. Logo, $\mathrm{R}_{\theta',\theta''}(T(x_n))$ é uma função monótona *não-decrescente* em $T(x_n)$. Logo, pelo Teorema de Karlin-Rubin (I), temos que

$$\delta^*(x_n) = \begin{cases} 0, \sum x_i < c \\ 1, \sum x_i \geq c \end{cases}$$

em que c satisfaz

$$\pi_{\delta^*}(1) = P_1(\sum X_i \geq c) = \alpha.$$

Com $n = 5$, temos que, quando $\theta = 1$, $\sum X_i \sim N(1 \cdot 5, 5 \cdot 2)$. Assim,

$$\begin{aligned}
P_1(\sum X_i \geq c) = P_1\left(\frac{\sum x_i - 5}{\sqrt{10}} \geq \frac{c-5}{\sqrt{10}}\right) = 0.1 \\
\Rightarrow \frac{c-5}{\sqrt{10}} = 1.28 \\
\Rightarrow c = 1.28 \cdot \sqrt{10} + 5 = 9.064 \\
\Rightarrow \delta^*(x_n) = \begin{cases} 0, \sum x_i < 9.064 \\ 1, \sum x_i \geq 9.064 \end{cases}
\end{aligned}$$

Portanto, se $\sum x_i < 9.064$, dizemos que *não há evidencias* para rejeitarmos $\mathcal{H}_0$ a 10% de significância estatística. Se $\sum x_i \geq 9.064$, dizemos que *há evidencias* para rejeitarmos $\mathcal{H}_0$ a 10% de significância estatística.

42.6.1.2 Exemplo (Exponencial)

Seja X_n a.a. de $X \sim \text{Exp}(\theta), \theta \in \Theta \subseteq \mathbb{R}_+$. Considere

$$\begin{cases} \mathcal{H}_0 : \theta \leq 2 \\ \mathcal{H}_1 : \theta > 2. \end{cases}$$

Encontre o TUMP de tamanho $\alpha = 5\%$.

42.6.1.2.1 Resposta

A função de verossimilhança é dada por

$$\mathcal{L}_{x_n}(\theta) = \theta^n \cdot \mathrm{e}^{-\theta \sum x_i}.$$

Pelo Critério de Fatoração de Fisher, $T(X_n) = \sum X_i$ é uma estatística suficiente.

$$\begin{aligned}
\mathrm{R}_{\theta',\theta''}(T(X_n)) &= \frac{\theta'^n \cdot \mathrm{e}^{-\theta' \sum x_i}}{\theta''^n \cdot \mathrm{e}^{-\theta'' \sum x_i}} \\
&= \left(\frac{\theta'}{\theta''}\right)^n \mathrm{Exp}\left\{-\theta' T(x_n) + \theta'' T(x_n)\right\} \\
&= \left(\frac{\theta'}{\theta''}\right)^n \mathrm{Exp}\left\{(\theta'' - \theta') T(x_n)\right\}, \theta' \geq \theta''.
\end{aligned}$$

Como $\theta' \geq \theta''$, temos que $\mathrm{R}_{\theta',\theta''}(T(X_n))$ é decrescente em $T(x_n)$. Tome $T'(x_n) = -\sum x_i$, logo

$$\mathrm{R}_{\theta',\theta''}(T(X_n)) = \left(\frac{\theta'}{\theta''}\right)^n \mathrm{Exp}\left\{(\theta' - \theta'')T(x_n)\right\}, \theta' \geq \theta''.$$

é monótona não-decrescente em $T'(x_n)$. Portanto, pelo Teorema de Karlin-Rubin (I),

$$\delta^*(x_n) = \begin{cases} 0, \sum -x_i < c \\ 1, \sum -x_i \geq c \end{cases} \Rightarrow \delta^*(x_n) = \begin{cases} 0, \sum x_i > -c \\ 1, \sum x_i \leq -c \end{cases}$$

em que c satisfaz

$$\pi_{\delta^*}(1) = P_1\left(\sum X_i \leq -c\right) = \alpha.$$

Para $n = 10$, sob $\theta = 2, \sum X_i \sim \mathrm{Gama}(10, 2)$. Computacionalmente, o quantil 0.05 dessa distribuição é 2.7127. Logo, $-c = 2.7127 \Rightarrow c = -2.7127$. Dessa forma, se $\sum x_i > 2.7127$, dizemos que *não há evidencias* para rejeitarmos $\mathcal{H}_0$ a 5% de significância estatística. Se $\sum x_i \leq 2.7127$, dizemos que *há evidencias* para rejeitarmos $\mathcal{H}_0$ a 5% de significância estatística.

42.6.2 Teorema de Karlin-Rubin (II)

Seja X_n a.a. de $X \sim f_\theta, \theta \in \Theta \subseteq \mathbb{R}$. Considere

$$\begin{cases} \mathcal{H}_0 : \theta \geq \theta_0 \\ \mathcal{H}_1 : \theta < \theta_0 \end{cases}$$

em que θ_0 é um valor dado. Se a *razão de verossimilhanças*

$$\mathrm{R}_{\theta',\theta''}(T(x_n)) = \frac{\mathcal{L}_{x_n}(\theta')}{\mathcal{L}_{x_n}(\theta'')}$$

for *monótona não-crescente* e $\theta' \leq \theta''$, em que $T(x_n)$ é uma estatística suficiente para o modelo, então o teste $\delta^* : \mathfrak{X}^{(n)} \to \{0, 1\}$ tal que

$$\delta^*(x_n) = \begin{cases} 0, T(x_n) > c \\ 1, T(x_n) \leq c, \end{cases}$$

e c satisfaz

$$\pi_{\delta^*}(\theta_0) = P_{\theta_0}(\delta^*(x_n) = 1) = \alpha,$$

é o teste uniformemente mais poderoso de tamanho α.

42.6.2.1 Exemplo (Normal)

Seja X_n a.a. de $X \sim N(\theta, 2), \theta \in \Theta \subseteq \mathbb{R}$. Considere

$$\begin{cases} \mathcal{H}_0 : \theta \geq 1 \\ \mathcal{H}_1 : \theta < 1. \end{cases}$$

Encontre o TUMP de tamanho α.

42.6.2.1.1 Resposta

A função de verossimilhança é dada por

$$\begin{aligned} \mathcal{L}_{x_n}(\theta) &= \left(\frac{1}{\sqrt{4\pi}}\right)^n \text{Exp}\left\{-\frac{1}{4}\sum(x_i - \theta)^2\right\} \\ &= \left(\frac{1}{\sqrt{4\pi}}\right)^n \text{Exp}\left\{-\frac{1}{4}\sum x_i^2 - 2\theta\sum x_i + n\theta^2\right\}. \end{aligned}$$

Pelo Critério de Fatoração de Fisher, $T(X_n) = \sum x_i$ é uma estatística suficiente. Note que

$$\begin{aligned} \mathrm{R}_{\theta',\theta''}(T(X_n)) &= \frac{\left(\frac{1}{\sqrt{4\pi}}\right)^n \text{Exp}\left\{-\frac{1}{4}\sum x_i^2 - 2\theta'\sum x_i + n\theta'^2\right\}}{\left(\frac{1}{\sqrt{4\pi}}\right)^n \text{Exp}\left\{-\frac{1}{4}\sum x_i^2 - 2\theta''\sum x_i + n\theta''^2\right\}} \\ &= \text{Exp}\left\{-\frac{1}{4}(-\theta' T(x_n) + n\theta'^2 + 2\theta'' T(x_n) - n\theta''^2\right\} \\ &= \text{Exp}\left\{\frac{1}{2}(\theta' - \theta'')T(x_n) - \frac{1}{4}n'\theta^2 + \frac{1}{4}\theta''^2\right\}. \end{aligned}$$

Como $\theta' \leq \theta''$, temos que $\theta' - \theta'' \leq 0$. Logo, $\mathrm{R}_{\theta',\theta''}(T(x_n))$ é uma função monótona *não-crescente* em $T(x_n)$. Logo, pelo Teorema de Karlin-Rubin (II), temos que

$$\delta^*(x_n) = \begin{cases} 0, \sum x_i > c \\ 1, \sum x_i \leq c \end{cases}$$

em que c satisfaz

$$\pi_{\delta^*}(1) = P_1(\sum X_i \leq c) = \alpha.$$

42.6.3 Simulações

```
# Seja X a.a. de X ~ Exp()
# H_0    2; H_1   > 2
#  (x) = 1    x   2.71
using Distributions, Random, StatsBase
Random.seed!(24)

# Borda do Θ_0
 _0 = 2
# Valores do teste e amostra
n = 100
  = 0.05
MC = 10_000

function h0()
    # Sob H_0
     _00 = rand(Uniform(0,  _0))


    d = Exponential(1 /  _00)
    dsoma = Gamma(n, 1 /  _00)
    c = quantile(dsoma,  )

      = zeros(MC)
    for i in 1:MC
        x = rand(d, n)
         [i] = sum(x)   c
    end
    return println("Testes rejeitados sob H_0: $(mean( ) * 100)%")
end

function h1()
    # Sob H_0
     _11 = rand(Uniform( _0,  _0 + 10))


    d = Exponential(1 /  _11)
    dsoma = Gamma(n, 1 /  _11)
    c = quantile(dsoma,  )

      = zeros(MC)
    for i in 1:MC
        x = rand(d, n)
```

```
        [i] = sum(x)   c
    end
    return println("Testes rejeitados sob H_1 (Poder do teste): $(mean( ) * 100)")
end

h0()
h1()
```

```
Testes rejeitados sob H_0: 4.979999999999995%
Testes rejeitados sob H_1 (Poder do teste): 95.6
```

42.7 Testes de hipótese gerais

Considere

$$\begin{cases} \mathscr{H}_0 : \theta \in \Theta_0 \\ \mathscr{H}_1 : \theta \in \Theta_1 \end{cases}$$

as hipóteses nula e alternativa em que $\Theta_0 \cup \Theta_1 = \Theta, \Theta_0 \cap \Theta_1 = \emptyset, \#(\Theta_0) \geq 1, \#(\Theta_1) \geq 1$.

O teste da razão de verossimilhanças generalizada de tamanho α é $\delta : \mathfrak{X}^{(n)} \to \{0, 1\}$ tal que

$$\delta(X_n) = \begin{cases}0, \ x_n \notin A_c 1, \ x_n \in A_c\end{cases}$$

e $\sup\limits_{\theta \in \Theta_0} \pi_\delta(\theta) = \alpha$, em que

$$A_c = \left\{ x_n \in \mathfrak{X}^{(n)} : \frac{\sup\limits_{\theta \in \Theta_1} \mathscr{L}_{x_n}(\theta)}{\sup\limits_{\theta \in \Theta_0} \mathscr{L}_{x_n}(\theta)} \geq c \right\}$$

Constante c

A constante c é obtida resolvendo $\sup\limits_{\theta \in \Theta_0} \pi_\delta(\theta) = \alpha$

Teste de Neyman-Pearson

Se $\Theta_0 = \{\theta_0\}, \Theta_1 = \{\theta_1\}$, temos o teste mais poderoso de Neyman-Pearson.

TUMP de Karlin-Rubin I e II

Se $\Theta_0 = (-\infty\theta_0] \cap \Theta, \Theta_1 = (\theta_1, \infty) \cap \Theta$, temos o teste uniformemente mais poderoso de Karlin-Rubin (I). Por sua vez, se $\Theta_0 = [\theta_0, \infty) \cap \Theta, \Theta_1 = (-\infty\theta_1) \cap \Theta$, temos o teste uniformemente mais poderoso de Karlin-Rubin (II)

42.7.1 Estatística da razão de verossimilhanças generalizada

Se $\dim(\Theta) = \dim(\Theta_1) > \dim(\Theta_0)$ e $\mathcal{L}_{x_n}(\theta)$ for contínua para todo θ em Θ, então:

$$\sup_{\theta\in\Theta} \mathcal{L}_{x_n}(\theta) = \sup_{\theta\in\Theta_1} \mathcal{L}_{x_n}(\theta), \text{ q.c.}$$

Portanto,

$$\begin{aligned}\frac{\sup_{\theta\in\Theta_1} \mathcal{L}_{x_n}(\theta)}{\sup_{\theta\in\Theta_0} \mathcal{L}_{x_n}(\theta)} = \frac{\sup_{\theta\in\Theta} \mathcal{L}_{x_n}(\theta)}{\sup_{\theta\in\Theta} \mathcal{L}_{x_n}(\theta)} \cdot \frac{\sup_{\theta\in\Theta_1} \mathcal{L}_{x_n}(\theta)}{\sup_{\theta\in\Theta_0} \mathcal{L}_{x_n}(\theta)} = \frac{\sup_{\theta\in\Theta} \mathcal{L}_{x_n}(\theta)}{\sup_{\theta\in\Theta_0} \mathcal{L}_{x_n}(\theta)} &\geq c \\ \iff \frac{\sup_{\theta\in\Theta_0} \mathcal{L}_{x_n}(\theta)}{\sup_{\theta\in\Theta} \mathcal{L}_{x_n}(\theta)} &\leq \frac{1}{c} \\ \Rightarrow A_c = \left\{ \frac{\sup_{\theta\in\Theta_0} \mathcal{L}_{x_n}(\theta)}{\sup_{\theta\in\Theta} \mathcal{L}_{x_n}(\theta)} \leq \frac{1}{c} \right\}\end{aligned}$$

Dizemos que

$$\lambda(x_n; \Theta_0) = \frac{\sup_{\theta\in\Theta_0} \mathcal{L}_{x_n}(\theta)}{\sup_{\theta\in\Theta} \mathcal{L}_{x_n}(\theta)} \tag{42.1}$$

é a estatística da razão de verossimilhanças generalizada.

Denote por $\hat{\theta}_0$ o estimador para "θ" sob $\mathcal{H}_0$, ou seja,

$$\hat{\theta}_0 = \text{argmax}_{\theta\in\Theta_0} \mathcal{L}_{x_n}(\theta)$$

sempre que existir. Denote também

$$\hat{\theta}_{\text{MV}} = \text{argmax}_{\theta\in\Theta} \mathcal{L}_{x_n}(\theta)$$

o estimador de máxima verossimilhança não restrito a $\mathcal{H}_0$. A estatística da razão de verossimilhança generalizada pode ser reescrita:

$$\lambda(x_n;\Theta_0) = \frac{\mathcal{L}_{x_n}(\hat{\theta}_0)}{\mathcal{L}_{x_n}(\hat{\theta}_{\text{MV}})} \tag{42.2}$$

Note que a Equação **??** é bem definida sempre que $0 < \sup_{\theta\in\Theta} \mathcal{L}_{x_n}(\theta) < \infty$. Em alguns casos, a Equação **??** não pode ser resolvida por não existir um argumento que maximize a função de verossimilhança sob a hipótese nula.

Observe que $\mathcal{H}_0 : \theta = \theta_0$ contra $\mathcal{H}_1 : \theta = \theta_1$ e $\Theta \subseteq \mathbb{R}$,

$$\lambda(x_n;\Theta_0) = \frac{\mathcal{L}_{x_n}(\theta_0)}{\mathcal{L}_{x_n}(\hat{\theta}_{\text{MV}})}$$

Teorema. Sob as condições de regularidade e com $\dim(\Theta_0) < \dim(\Theta)$, temos que

$$-2\ln\lambda(X_n,\Theta_0) \xrightarrow{\mathcal{D}} \chi^2_s,$$

em que $s = \dim(\Theta) - \dim(\Theta_0)$. Supõe-se que o conjunto Θ_0 *não* contém singularidades.

O teste da Razão de Verossimilhança Generalizada (RVG) pode ser reescrito:

$$\begin{aligned}
\delta(x_n) &= \begin{cases} 0, & \frac{\mathcal{L}_{x_n}(\hat{\theta}_0)}{\mathcal{L}_{x_n}(\hat{\theta}_{\text{MV}})} > \frac{1}{c} \\ 1, & \frac{\mathcal{L}_{x_n}(\hat{\theta}_0)}{\mathcal{L}_{x_n}(\hat{\theta}_{\text{MV}})} \le \frac{1}{c} \end{cases} \\
\iff \delta(x_n) &= \begin{cases} 0, & \lambda(x_n,\Theta_0) > \frac{1}{c} \\ 1, & \lambda(x_n,\Theta_0) \le \frac{1}{c} \end{cases} \\
\iff \delta(x_n) &= \begin{cases} 0, & -2\ln\lambda(x_n,\Theta_0) < 2\ln c \\ 1, & -2\ln\lambda(x_n,\Theta_0) \ge 2\ln c \end{cases}
\end{aligned}$$

em que c deve satisfazer

$$\begin{aligned}
\sup_{\theta\in\Theta_0} \pi_\delta(\theta) &= \alpha \\
\iff \sup_{\theta\in\Theta_0} P_\theta(\delta(X_n) = 1) &= \alpha \\
\iff \sup_{\theta\in\Theta_0} P_\theta(-2\ln\lambda(X_n,\Theta_0) \ge 2\ln c) &= \alpha
\end{aligned} \tag{42.3}$$

1. Se a distribuição exata de $-2\ln\lambda(X_n,\Theta_0)$ for conhecida, entã basta encontrar $2\ln c$ que satisfaça a Equação **??**.

2. Caso contrário, utilizamos o teorema anterior:

   $$-2\ln\lambda(X_n,\Theta_0) \xrightarrow{\mathcal{D}} \chi^2_s,$$

   em que $s = \dim(\Theta) - \dim(\Theta_0)$

42.7.2 Exemplo (Normal)

Seja X_n a.a. de $X \sim N(\theta, 1), \theta \in \Theta \subseteq \mathbb{R}$, e considere

$$\begin{cases} \mathcal{H}_0 : \theta = 0 \\ \mathcal{H}_1 : \mu \neq 0, \end{cases}$$

as hipóteses nula e alternativa. Encontre o teste da razão de verossimilhanças generalizada de tamanho α.

42.7.2.1 Resposta

Já sabemos que $\hat{\theta}_{\text{MV}} = \bar{X}$ e

$$\begin{aligned} \mathcal{L}_{x_n}(\theta) &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \text{Exp}\left\{-\frac{1}{2}\sum (x_i - \theta)^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \text{Exp}\left\{-\frac{1}{2}\sum x_i^2 - 2\theta \sum x_i + n\theta^2\right\}. \end{aligned}$$

.

Assim,

$$\begin{aligned} \lambda(x_n; \Theta_0) &= \frac{\sup\limits_{\theta \in \Theta_0} \mathcal{L}_{x_n}(\theta)}{\sup\limits_{\theta \in \Theta} \mathcal{L}_{x_n}(\theta)} \\ &= \frac{\mathcal{L}_{x_n}(0)}{\mathcal{L}_{x_n}(\bar{x})} \\ &= \text{Exp}\left\{-\frac{1}{2}\sum x_i^2 + \frac{1}{2}\left(\sum x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right)\right\} = \mathrm{e}^{-\frac{1}{2}n\bar{x}^2} \end{aligned}$$

Portanto,

$$-2\ln\lambda(x_n, \Theta_0) = n\bar{x}^2.$$

Sob $\mathcal{H}_0$, $\theta = 0, \bar{X} \sim N(0, 1/n), \sqrt{n}\bar{X} \sim N(0,1) \Rightarrow n\bar{X}^2 \sim \chi_1^2$. Logo, o teste de RVG é

$$\delta(x_n) = \Big\{0, \quad n\bar{x}^2 < 2\ln c1, \quad n\bar{x}^2 \geq 2\ln c$$

em que c deve satisfazer

$$\sup_{\theta \in \Theta_0} P_\theta(n\bar{X}^2 \geq 2\ln c) = \alpha.$$

Note que $\forall \theta \in \Theta_0$, $n\bar{X}^2 \sim \chi_1^2$ (é ancilar ao modelo reduzido). Portanto,

$$\sup_{\theta \in \Theta_0} P_0(n\bar{X}^2 \geq c^*) = P(\chi_1^2 \geq c^*) = \alpha.$$

Tomando $c^* = 2\ln c$, podemos obter o valor para c^* de uma qui-quadradado com 1 grau de liberdade para qualquer valor de α. Por exemplo, se $\alpha = 10\%, c^* = 2.70$. Se $\alpha = 5\%, c^* = 3.84$, etc.

42.7.3 Exemplo (Poisson)

Seja X_n a.a. de $X \sim \text{Poiss}(\theta), \theta \in \Theta = (0, \infty)$. Considere

$$\begin{cases} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \mu \neq \theta_0, \end{cases}$$

as hipóteses nula e alternativa. Encontre o teste da razão de verossimilhanças generalizada de tamanho α. Use o resultado assintótico.

42.7.3.1 Resposta

Já sabemos que $\hat{\theta}_{\text{MV}} = \bar{X}$ e

$$\begin{aligned} \mathcal{L}_{x_n}(\theta) &= \mathrm{e}^{n\theta} \cdot \frac{\theta^{\sum x_i}}{\prod(x_i!)} \\ \Rightarrow \lambda(x_n, \Theta_0) &= \frac{\mathrm{e}^{n\theta_0} \cdot \frac{\theta_0^{\sum x_i}}{\prod(x_i!)}}{\mathrm{e}^{n\bar{x}} \cdot \frac{\bar{x}^{\sum x_i}}{\prod(x_i!)}} \\ &= \mathrm{e}^{n\theta_0 + n\bar{x}} \cdot \left(\frac{\theta_0}{\bar{x}}\right)^{\sum x_i} \end{aligned}$$

Sabemos que, sob $\mathcal{H}_0$, (i.e. $\forall \theta \in \Theta_0$), $-2 \ln \lambda(X_n, \Theta_0) \xrightarrow{\mathcal{D}} \chi^2_1$. Assim,

$$2n\theta_0 - 2n\bar{X} - 2n\bar{X} \ln\left(\frac{\theta_0}{\bar{X}}\right) = \underbrace{2n\theta_0 - 2n\bar{X} - 2n\bar{X} \ln \theta_0 + 2n\bar{X} \ln \bar{X}}_{T(X_n)} \xrightarrow{\mathcal{D}} \chi^2_1.$$

Logo, o teste de RVG é

$$\delta(x_n) = \begin{cases} 0, \quad T(x_n) < c^* 1, \quad T(x_n) \geq c^* \end{cases}$$

em que c deve satisfazer

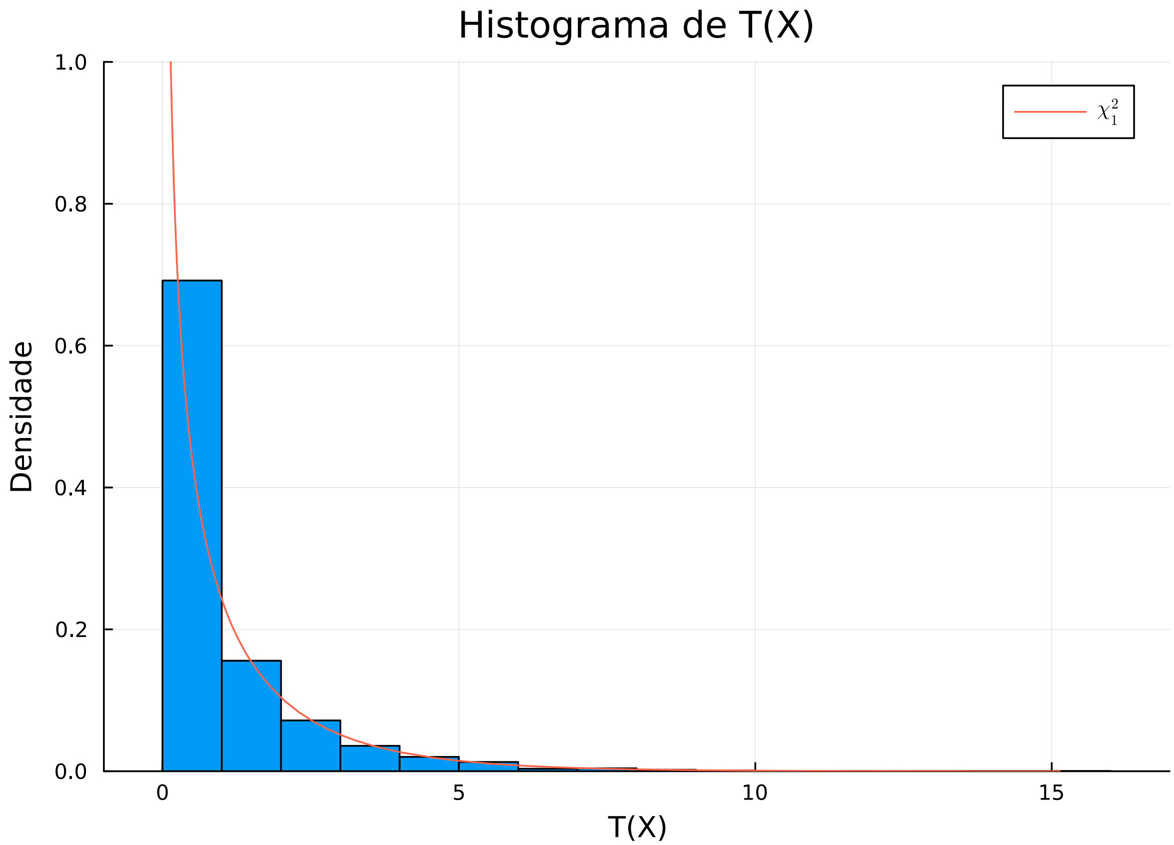
$$\sup_{\theta \in \Theta_0} P_\theta(n\bar{X}^2 \geq c^*) = P_{\theta_0}(T(X_n) \geq c^*) = \alpha.$$

Pela aproximação,

$$P_{\theta_0}(T(X_n) \geq c^*) \approx P(\chi^2_1 \geq c^*)\alpha.$$

```
# Seja Xn a.a. de X ~ Pois( )
# H0:   =  _0
# H1:     _0
using Random, Distributions, StatsBase, Plots, StatsPlots, LaTeXStrings
n = 100
 _0 = 3

# Gerando sob H_0
d = Poisson( _0)
MC = 10_000
TX = zeros(MC)
for i in 1:MC
    x = rand(d, n)
    TX[i] = 2 * n *  _0 - 2 * n * mean(x) - 2 * n * mean(x) * log( _0) + 2 * n * mean(x) * log
end
p = histogram(TX, normalize = true, title = "Histograma de T(X)", label = "", ylims = (0, 1)
plot!(Chisq(1), label = L"\chi^2_1", color = :tomato)

## Probabilidade do erro tipo I,   = 10%

quantil = quantile(Chisq(1), 0.9)
rejeita = TX .> quantil
println("Probilidade do Erro Tipo I com   = 10%: $(mean(rejeita) * 100)%")
display(p)
```

```
Probilidade do Erro Tipo I com   = 10%: 9.790000000000001%
```



42.7.4 Hipótese Linear Geral

Seja X_n a.a. de $X \sim f_\theta, \theta \in \Theta \subseteq \mathbb{R}^p$. Considere

$$\begin{cases} \mathcal{H}_0 : C \cdot \theta = d \\ \mathcal{H}_1 : C \cdot \theta \neq d \end{cases}$$

em que C é uma matriz $s \times p$ conhecida e d é um vetor de dimensão s conhecido. Podemos escrever as hipóteses em termos de Θ_0 e Θ_1:

$$\begin{cases} \mathcal{H}_0 : \theta \in \Theta_0 \\ \mathcal{H}_1 : \theta \in \Theta_1 \end{cases}$$

em que $\Theta_0 = \{\theta \in \Theta : C \cdot \theta = d\}$ e $\Theta_1 = \{\theta \in \Theta : C \cdot \theta \neq d\}$.

A hipótese $\mathcal{H}_0 : C \cdot \theta = d$ é conhecida como *hipótese linear geral*.

Note que, podemos testar as seguintes hipóteses como casos particulares: 1.

$$\begin{cases} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \theta \neq \theta_0 \end{cases} \iff C = I, d = \theta_0.$$

2.

$$\begin{cases} \mathcal{H}_0 : \theta_1 = \theta_2 \\ \mathcal{H}_1 : \theta_1 \neq \theta_2 \end{cases} \iff C = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \end{pmatrix}, d = 0, \quad \theta = (\theta_1, \theta_2 \dots, \theta_p).$$

3.

$$\begin{cases} \mathcal{H}_0 : \theta_1 = \theta_3 \text{ e } \theta_2 = \theta_4 \\ \mathcal{H}_1 : \text{Ao menos um diferente} \end{cases} \iff C = \begin{bmatrix} 1 & 0 & -1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & -1 & \cdots & 0 \end{bmatrix},$$

$$d = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

4.

$$\begin{cases} \mathcal{H}_0 : \sum_{i=1}^{p} c_i \theta_i = d \\ \mathcal{H}_1 : \sum_{i=1}^{p} c_i \theta_i \neq d \end{cases} \iff C = \begin{pmatrix} c_1 & c_2 & \dots & c_p \end{pmatrix}, d = d.$$

5.

$$\begin{cases} \mathcal{H}_0 : \sum_{i=1}^{p} c_i^{(1)} \theta_i = d^{(1)}, \dots, \sum_{i=1}^{p} c_i^{(p)} \theta_i = d^{(p)} \\ \mathcal{H}_1 : \text{Ao menos um diferente} \end{cases} \iff C = \begin{bmatrix} c_1^{(1)} & c_2^{(1)} & \dots & c_p^{(1)} \\ c_1^{(2)} & c_2^{(2)} & \dots & c_p^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ c_1^{(p)} & c_2^{(p)} & \dots & c_p^{(p)} \end{bmatrix},$$

$$d = \begin{bmatrix} d^{(1)} \\ d^{(2)} \\ \vdots \\ d^{(p)} \end{bmatrix}.$$

42.7.5 Estatística de Wald

Seja $\hat{\theta}$ um estimador para "θ" assintoticamente normal, ou seja,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\mathcal{D}} N_s(0, V_\theta)$$

Se C tiver posto linha completo, ou seja, todas suas linhas são linearmente independentes, então

$$\sqrt{n}(C\hat{\theta} - C\theta) \xrightarrow{\mathcal{D}} N_s(0, CV_\theta C)$$

$$\Rightarrow n(C\hat{\theta} - C\theta)^T [CV_{\hat{\theta}} C^T]^{-1} (C\hat{\theta} - C\theta) \xrightarrow{\mathcal{D}} \chi_s^2, \forall \theta \in \Theta.$$

Sob $\mathcal{H}_0, C\theta = d$. Disso, temos que

$$n(C\hat{\theta} - d)^T [CV_{\hat{\theta}} C^T]^{-1} (C\hat{\theta} - d) \xrightarrow{\mathcal{D}} \chi_s^2.$$

A estatística de Wald é definida por

$$W(X_n) = n(C\hat{\theta}(X_n) - d)^T [CV_{\hat{\theta}(X_n)} C^T]^{-1} (C\hat{\theta}(X_n) - d).$$

Sob $\mathcal{H}_0$, $W \xrightarrow{\mathcal{D}} \chi^2_s$, em que s é o número de linhas de C.

42.7.5.1 Exemplos

Considere as hipóteses com $\theta_{1,0}$ conhecido

$$\begin{aligned}
\Big\{\mathcal{H}_0 : \theta_1 = \theta_{1,0} \mathcal{H}_1 : \theta_1 \neq \theta_{1,0} \qquad & \Rightarrow C = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}, d = \theta_{1,0} \\
W &= n(\hat{\theta}_1 - \theta_{1,0})(V_{\hat{\theta}}^{(1,1)})^{-1}(\hat{\theta}_1 - \theta_{1,0}) \\
&= \frac{(\hat{\theta}_1 - \theta_{1,0})^2}{(V_{\hat{\theta}}^{(1,1)}/n)} \xrightarrow{\mathcal{D}} \chi^2_1, \text{ Sob } \mathcal{H}_0
\end{aligned}$$

Logo,

$$CV_{\hat{\theta}}C^T = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} \begin{bmatrix} V_{\hat{\theta}}^{(1,1)} & V_{\hat{\theta}}^{(1,2)} & \dots & V_{\hat{\theta}}^{(1,p)} \\ . & V_{\hat{\theta}}^{(2,2)} & \dots & V_{\hat{\theta}}^{(2,p)} \\ \vdots & \vdots & \ddots & . \\ . & . & . & V_{\hat{\theta}}^{(p,p)} \end{bmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = V_{\hat{\theta}}^{(1,1)}$$

Considere as hipóteses

$$\begin{aligned}
\Big\{\mathcal{H}_0 : \theta_1 = \theta_3 \mathcal{H}_1 : \theta_1 \neq \theta_3 \qquad & \Rightarrow C = \begin{pmatrix} 1 & 0 & -1 & 0 & \dots & 0 \end{pmatrix}, d = 0 \\
W &= n(\hat{\theta}_1 - \hat{\theta}_3)(CV_{\hat{\theta}}C^T)^{-1}(\hat{\theta}_1 - \hat{\theta}_3) \\
&= \frac{n(\hat{\theta}_1 - \hat{\theta}_3)^2}{CV_{\hat{\theta}}C^T}
\end{aligned}$$

Logo,

$$CV_{\hat{\theta}}C^T = \begin{pmatrix}1 & 0 & -1 & 0 & \dots & 0\end{pmatrix} \begin{bmatrix} V_{\hat{\theta}}^{(1,1)} & V_{\hat{\theta}}^{(1,2)} & \dots & V_{\hat{\theta}}^{(1,p)} \\ V_{\hat{\theta}}^{(2,1)} & V_{\hat{\theta}}^{(2,2)} & \dots & V_{\hat{\theta}}^{(2,p)} \\ \vdots & \vdots & \ddots & . \\ V_{\hat{\theta}}^{(p,1)} & V_{\hat{\theta}}^{(p,2)} & . & V_{\hat{\theta}}^{(p,p)} \end{bmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} V_{\hat{\theta}}^{(1,1)} - V_{\hat{\theta}}^{(3,1)} & V_{\hat{\theta}}^{(1,2)} - V_{\hat{\theta}}^{(3,2)} & \dots & V_{\hat{\theta}}^{(1,p)} - V_{\hat{\theta}}^{(3,p)} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$= V_{\hat{\theta}}^{(1,1)} - V_{\hat{\theta}}^{(3,1)} + V_{\hat{\theta}}^{(3,3)} - V_{\hat{\theta}}^{(1,3)}$$
$$= V_{\hat{\theta}}^{(1,1)} + V_{\hat{\theta}}^{(3,3)} - 2V_{\hat{\theta}}^{(1,3)}$$

O teste usando a estatística de Wald é definido por

$$\delta_W(x_n) = \begin{cases} 0, & W(x_n) < \eta \\ 1, & W(x_n) \geq \eta \end{cases}$$

em que η é obtido de

$$P(\chi_s^2 \geq \eta) = \alpha$$

para um teste de Wald de tamanho α.

Relação com as hipóteses

Sob $\mathcal{H}_0$, esperamos que W seja "pequeno", enquanto sob $\mathcal{H}_1$, esperamos que W seja "grande".

42.7.6 Estatística Escore (ou de Rao)

Considere as hipóteses

$$\begin{cases} \mathcal{H}_0 : \theta = \theta_0 \\ \mathcal{H}_1 : \theta \neq \theta_0 \end{cases}$$

em que θ_0 é conhecido e $\theta \in \theta \subseteq \mathbb{R}^p$. A estatística escore é definida por

$$R(X_n) = U_n(X_n, \theta_))^T I_n(\theta_0)^{-1} U_n(X_n, \theta_0)$$

em que U_n é o escore da amostra e I_n é a informação de fisher total.

Pode-se demonstrar que

$$R \xrightarrow{\mathcal{D}} \chi^2_p, \text{ sob } \mathcal{H}_0$$

O teste usando a estatística escore é definido por

$$\delta_R(x_n) = \begin{cases} 0, & R(x_n) < \eta \\ 1, & R(x_n) \geq \eta \end{cases}$$

em que η é obtido de

$$P(\chi^2_p \geq \eta) = \alpha$$

para um teste escore de tamanho α.

42.7.7 Comparando as estatísticas

```
# Razão de verossimilhanças
# Seja Xn a.a. de X ~ Pois( )
# H0:   =  _0
# H1:     _0
using Random, Distributions, StatsBase, Plots, StatsPlots, LaTeXStrings
Random.seed!(96)
n = 100
 = 0.1

MC = 10_000

# Razão de verossimilhanças
function razao(dist,  _0)
    TX = zeros(MC)
    for i in 1:MC
        x = rand(dist, n)
        TX[i] = 2 * n *  _0 - 2 * n * mean(x) - 2 * n * mean(x) * log( _0) +
            2 * n * mean(x) * log(mean(x))
    end
    return TX
end

# Estatística de Wald
function wald(dist,  _0)
    C = 1
    d =  _0
```

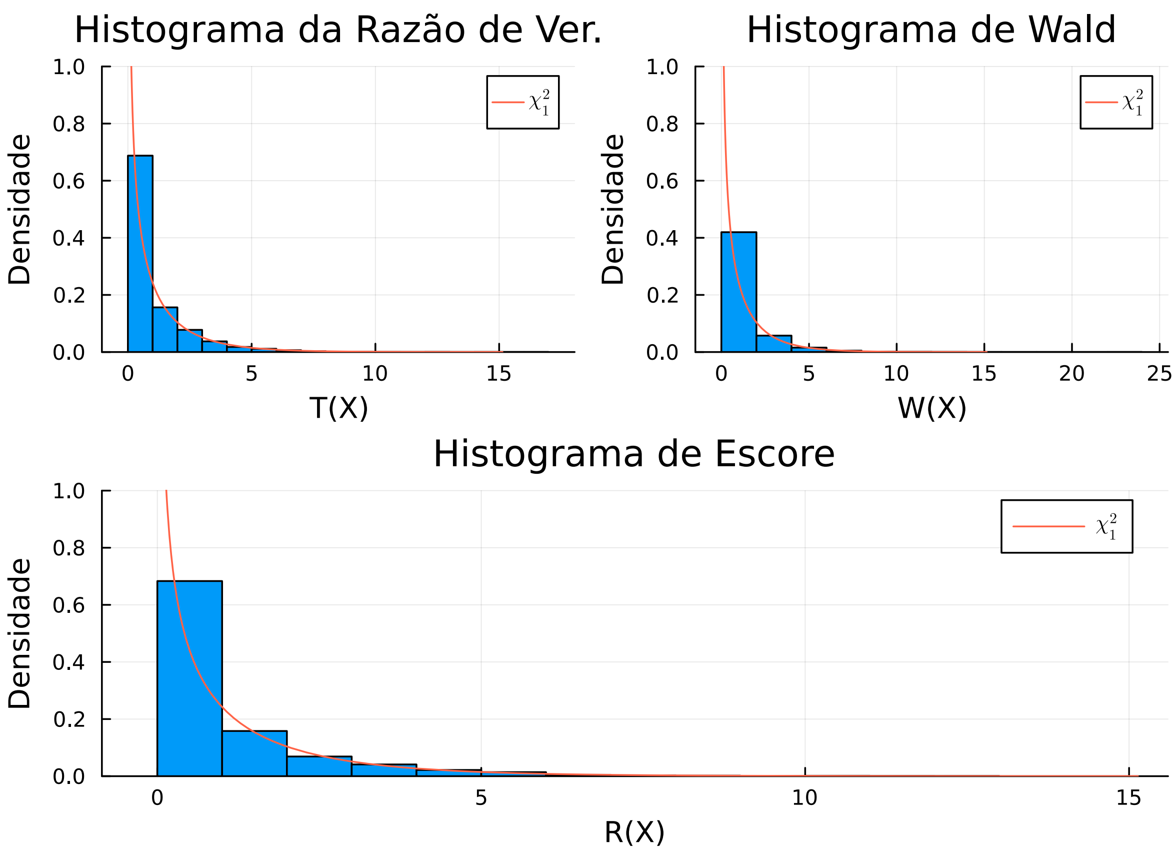
```
    W = zeros(MC)
    for i in 1:MC
        x = rand(dist, n)
        V = mean(x) # theta =emv> mean(x)
        W[i] = n * (C * mean(x) - d)^2 / V
    end
    return W
end

# Estatística de Escore
function escore(dist,  _0)
    I( ) = n /
    R = zeros(MC)
    for i in 1:MC
        x = rand(dist, n)
        U( ) = sum(x) /   - n
        R[i] = U( _0)^2 / I( _0)
    end
    return R
end
```

Geraremos sob $\mathscr{H}_0$:

```
# Gerando sob H_0
 _0 = 3
dist = Poisson( _0)
distassin = Chisq(1)
quantil = quantile(distassin, 1 -  )
## Probabilidade do erro tipo I,   = 10%, Razão

TX = razao(dist,  _0)
rejeita = TX .> quantil
println("Probilidade do Erro Tipo I com   = 10% (Razão de Ver.):
        $(mean(rejeita) * 100)%")


## Probabilidade do erro tipo I,   = 10%, Wald
W = wald(dist,  _0)
rejeita = W .> quantil
println("Probilidade do Erro Tipo I com   = 10% (Wald):
        $(mean(rejeita) * 100)%")
```

```
## Probabilidade do erro tipo I,   = 10%, Wald
R = escore(dist,  _0)
rejeita = R .> quantil
println("Probilidade do Erro Tipo I com   = 10% (Escore):
        $(mean(rejeita) * 100)%")

pv = histogram(
    TX, normalize = true, title = "Histograma da Razão de Ver.",
    label = "", ylims = (0, 1), bins = 20, xlabel = "T(X)",
    ylabel = "Densidade"
)
plot!(Chisq(1), label = L"\chi^2_1", color = :tomato)
pw = histogram(
    W, normalize = true, title = "Histograma de Wald", label = "",
    ylims = (0, 1), bins = 20, xlabel = "W(X)", ylabel = "Densidade"
)
plot!(Chisq(1), label = L"\chi^2_1", color = :tomato)
pr = histogram(
    R, normalize = true, title = "Histograma de Escore", label = "",
    ylims = (0, 1), bins = 20, xlabel = "R(X)", ylabel = "Densidade"
)
plot!(Chisq(1), label = L"\chi^2_1", color = :tomato)
l = @layout [pv pw; pr]
plot(pv, pw, pr, layout = l)
```

```
Probilidade do Erro Tipo I com   = 10% (Razão de Ver.):
        9.74%
Probilidade do Erro Tipo I com   = 10% (Wald):
        10.02%
Probilidade do Erro Tipo I com   = 10% (Escore):
        10.05%
```



Para comparar, podemos também gerar sob $\mathcal{H}_1$:

```
# Gerando sob H_1

rejeitaTX( ) = mean(razao(dist,  ) .> quantil)
rejeitaW( ) = mean(wald(dist,  ) .> quantil)
rejeitaR( ) = mean(escore(dist,  ) .> quantil)
Δ =  _0 * 0.3
# alance = length(θ_0-Δ:0.1:θ_0+Δ)
# poderTX = zeros(alcance)
# poderW = zeros(alcance)
# poderR = zeros(alcance)
# for i in θ_0-Δ:0.1:θ_0+Δ
#   poderTX[i] = rejeitaTX(i)
# end
pv = plot(
    rejeitaTX,
    xlims = ( _0 - Δ,  _0 + Δ),
```

```
    ylims = (0, 1),
    label = "",
    ylabel = "Poder Razão de ver.",
    xlabel = L" _1"
)
pw = plot(
    rejeitaW,
    xlims = ( _0 - Δ,  _0 + Δ),
    ylims = (0, 1),
    label = "",
    ylabel = "Poder Wald",
    xlabel = L" _1"
)
pr = plot(
    rejeitaR,
    xlims = ( _0 - Δ,  _0 + Δ),
    ylims = (0, 1),
    label = "",
    ylabel = "Poder Escore",
    xlabel = L" _1"
)
l = @layout [pv pw; pr]
plot(pv, pw, pr, layout = l)
```

42.7.8 Valor-p ou nível descritivo do teste

Seja X_n a.a. de $X \sim f_\theta, \theta \in \Theta \subseteq \mathbb{R}^p$. Considere

$$\mathcal{H}_0 : \theta \in \Theta_0$$

> **ℹ Suficiência da hipótese nula para o valor-p**
>
> A hipótese alternativa, $\mathcal{H}_1$, não é necessária para formular o valor-p.

Seja $T_{\mathcal{H}_0}(X_n)$ tal que, quanto maior for seu valor observado, maior é a discrepância entre $\mathcal{H}_0$ e os dados observados. Podemos usar a estatística da razão de verossimilhanças, Wald, ou de Escore. O valor-p é definido por

$$\text{valor}{-}p(x_n, \mathcal{H}_0) - \sup_{\theta \in \Theta_0} P_\theta(T_{\mathcal{H}_0}(X_n) \geq T_{\mathcal{H}_0}(x_n))$$

O valor-p é a probabilidade de observar uma estatística tão ou mais extrema que a observada sob o cenário mais favorável à $\mathcal{H}_0$.

Notação em alguns textos

Alguns livros, para fim didáticos, definem o valor-p como probabilidade condicional:

$$\text{valor}-p(x_n, \mathcal{H}_0) - \sup_{\theta \in \Theta_0} P_\theta(T_{\mathcal{H}_0}(X_n) \geq T_{\mathcal{H}_0}(x_n) | \mathcal{H}_0)$$

Isto, na estatística clássica, está equivocado uma vez que $\mathcal{H}_0$ é uma hipótese científica, não um evento definido na σ-álgebra, isto é, $P(A|B)$ só está bem definido se, e somente se, A e B estiverem na mesma σ-álgebra.

1. Observe que, quanto menor for o valor-p, há mais evidencias de que $\mathcal{H}_0$ é falsa.

Interpretação com níveis de significância

Caso valor$-p(x_n, \mathcal{H}_0) \leq \alpha$, então dizemos que há evidências para rejeitar $\mathcal{H}_0$ a $\alpha \cdot 100\%$ de significância estatística.

2. Se $T_{\mathcal{H}_0} \sim G$, sob $\mathcal{H}_0$, em que G não depende de "θ", então

$$\text{valor}-p(x_n, \mathcal{H}_0) = P(G \geq T_{\mathcal{H}_0}(x_n))$$

3. Se $T_{\mathcal{H}_0} \sim G$, sob $\mathcal{H}_0$, em que G não depende de "θ", então

$$\text{valor}-p(X_n, \mathcal{H}_0) \overset{\text{Sob } \mathcal{H}_0}{\sim} \text{Unif}(0, 1)$$

42.7.8.1 Valor-p assintótico

Seja $T_{\mathcal{H}_0}$ uma estatística tal que

$$T_{\mathcal{H}_0} \overset{\mathcal{D}}{\to} G, \quad \text{Sob } \mathcal{H}_0.$$

então, o valor-p assintótico é definido por

$$\text{valor}-p^{(a)}(x_n, \mathcal{H}_0) = P(G \geq T_{\mathcal{H}_0}(x_n)$$

Pode-se conduzir simulações de Monte Carlo para verificar se o tamanho amostral é grande o suficiente para fazer a aproximação assintótica:

1. Gerar os dados sob $\mathcal{H}_0$;

2. Calcular valor$-p^{(a)}$ e armazenar;
3. Repetir 1-2 M (tradicionalmente 10 mil) vezes;
4. Fazer um histograma do valor-p assintótico e comparar com $\text{Unif}(0,1)$.

42.7.8.2 Exemplo (Normal)

Seja X_n a.a. de $X \sim N(\theta, 1), \theta \in \Theta \subseteq \mathbb{R}$. Considere

$$\mathcal{H}_0 : \theta = 0.$$

Encontre um valor-p.

42.7.8.2.1 Resposta

Observe que $T_{H_0}(x_n) = n\bar{X}^2$ (essa também é a estatística de Wald) satisfaz o critério que, quanto maior o valor observado de T_{H_0}, mais evidencias contra $\mathcal{H}_0$. Além disso, sob $\mathcal{H}_0$,

$$T_{\mathcal{H}_0} \sim \chi_1^2 \Rightarrow \text{valor}-p(x_n, \mathcal{H}_0) = P(\chi_1^2 \geq n\bar{x}^2)$$

Se $n = 10$ e $\bar{x} = 1$, então

$$\text{valor}-p(x_n, \mathcal{H}_0) = P(\chi_1^2 \geq 10 \cdot 1^2) = 0.001$$

42.7.8.3 Exemplo (Poisson)

Seja X_n a.a. de $X \sim \text{Pois}(\theta), \theta \in \Theta = (0, \infty)$. Considere

$$\mathcal{H}_0 : \theta = 1.$$

Encontre o valor-p assintótico utilizando as três estatísticas (Razão de Verossimilhanças, Wald e Escore, $T'_{\mathcal{H}_0}(X_n), T''_{\mathcal{H}_0}(X_n), T'''_{\mathcal{H}_0}(X_n)$).

42.7.8.3.1 Resposta

Para a razão de verossimilhanças,

$$\begin{aligned} T'_{\mathcal{H}_0}(X_n) &= 2n \cdot 1 - 2n\bar{X} - 2n\bar{X}\ln\frac{1}{\bar{X}} \\ &= 2n - 2n\bar{X} + 2n\bar{X}\ln\frac{1}{\bar{X}} \xrightarrow{\mathcal{D}} \chi_1^2, \quad \theta = 1. \end{aligned}$$

Para a estatística de Wald, usaremos $\hat{\theta} = \bar{X}$ (que é o EMV, EMM e ENVVUM)

$$T''_{\mathcal{H}_0}(X_n) = \frac{n(\bar{X}-1)^2}{\bar{X}} \xrightarrow{\mathcal{D}} \chi_1^2, \quad \theta = 1.$$

pois $\bar{X} \overset{a}{\approx} N(\theta, \theta/n)$, sob $\mathcal{H}_0$, então, pelo Teorema de Slutsky,

$$\frac{\bar{X}-\theta}{\sqrt{\frac{\bar{X}}{n}}} \overset{a}{\approx} N(0,1) \Rightarrow n\left(\frac{\bar{X}-\theta}{\sqrt{\bar{X}}}\right)^2 \overset{a}{\approx} \chi_1^2, \forall\theta \in \Theta.$$

Logo, sob $\mathcal{H}_0 : \theta = 1$,

$$\frac{n(\bar{X}-1)^2}{\bar{X}} \approx \chi_1^2$$

Para a estatística de escore, temos que o escore é dado por

$$U_n(X_n, \theta) = -n + \frac{1}{\theta_0}\sum X_i$$

e a Informação de Fisher Total por

$$I_n(\theta_0) = -E_\theta\left(-\sum\frac{X_i}{\theta_0}\right) = \frac{n}{\theta_0},$$

portanto,

$$\begin{aligned} T'''_{\mathcal{H}_0}(X_n) &= \left(\frac{\sum X_i}{\theta_0} - n\right)^2 \cdot \frac{\theta_0}{n} \\ &= n\frac{(\bar{X}-\theta_0)^2}{\theta_0} = n(\bar{X}-1)^2. \end{aligned}$$

O valor-p assintótico é, portanto,

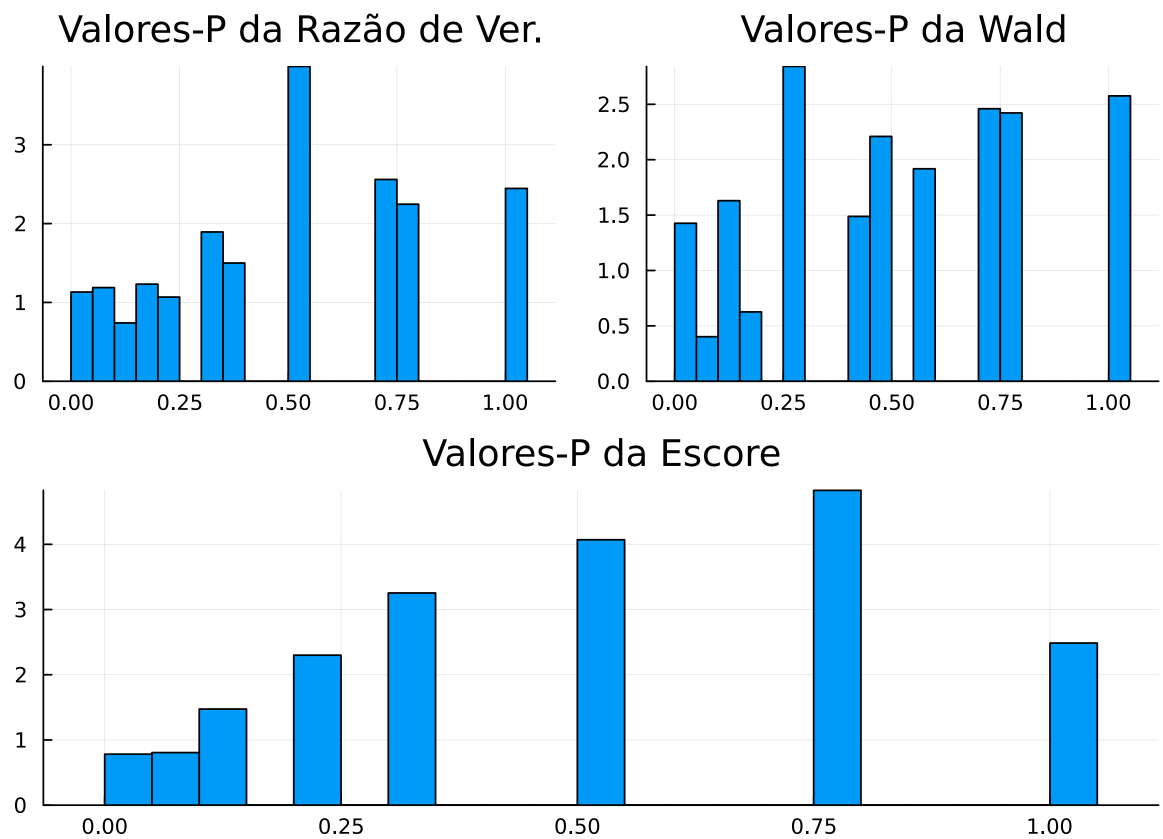
$$\begin{aligned} &\text{valor}-p_1(x_n, \mathcal{H}_0) = P_1(T'_{\mathcal{H}_0}(X_n) \geq T'_{H_0}(x_n)) \overset{a}{\approx} \text{val.}-p_1^{(a)}(x_n, \mathcal{H}_0) = P(\chi_1^2 \geq T'_{H_0}(x_n)) \\ &\text{valor}-p_2(x_n, \mathcal{H}_0) = P_1(T''_{\mathcal{H}_0}(X_n) \geq T''_{H_0}(x_n)) \overset{a}{\approx} \text{val.}-p_2^{(a)}(x_n, \mathcal{H}_0) = P(\chi_1^2 \geq T''_{H_0}(x_n)) \\ &\text{valor}-p_3(x_n, \mathcal{H}_0) = P_1(T'''_{\mathcal{H}_0}(X_n) \geq T'''_{H_0}(x_n)) \overset{a}{\approx} \text{val.}-p_3^{(a)}(x_n, \mathcal{H}_0) = P(\chi_1^2 \geq T'''_{H_0}(x_n)) \end{aligned}$$

Consideraremos $\bar{x} = 2$ e $n = 10$ para encontramos os valores numéricos dos valor-p:

$$
\begin{aligned}
& T'_{H_0}(x_n) = 2 \cdot 10 - 2 \cdot 10 \cdot 2 + 2 \cdot 10 \cdot 2 \ln 2 = 7.73 \\
\Rightarrow \text{valor}{-}p_1^{(a)}(x_n, \mathcal{H}_0) = 0.005, & \\
\\
& T''_{H_0}(x_n) = 10 \cdot \frac{(2-1)^2}{2} = 5 \\
\Rightarrow \text{valor}{-}p_2^{(a)}(x_n, \mathcal{H}_0) = 0.025, & \\
\\
& T'''_{H_0}(x_n) = 10 \cdot 1^2 = 10 \\
\Rightarrow \text{valor}{-}p_3^{(a)}(x_n, \mathcal{H}_0) = 0.002. &
\end{aligned}
$$

Note que, com $n = 10$, a aproximação pode não ser tão boa:

```julia
n = 10
p1 = @. ccdf(Chisq(1), razao(Poisson(1), 1))
p2 = @. ccdf(Chisq(1), wald(Poisson(1), 1))
p3 = @. ccdf(Chisq(1), escore(Poisson(1), 1))

hist1 = histogram(
    p1, title = "Valores-P da Razão de Ver.", label = "",
    normalize = true
)
hist2 = histogram(
    p2, title = "Valores-P da Wald", label = "",
    normalize = true
)
hist3 = histogram(
    p3, title = "Valores-P da Escore", label = "",
    normalize = true
)

plot(hist1, hist2, hist3, layout = l)
```



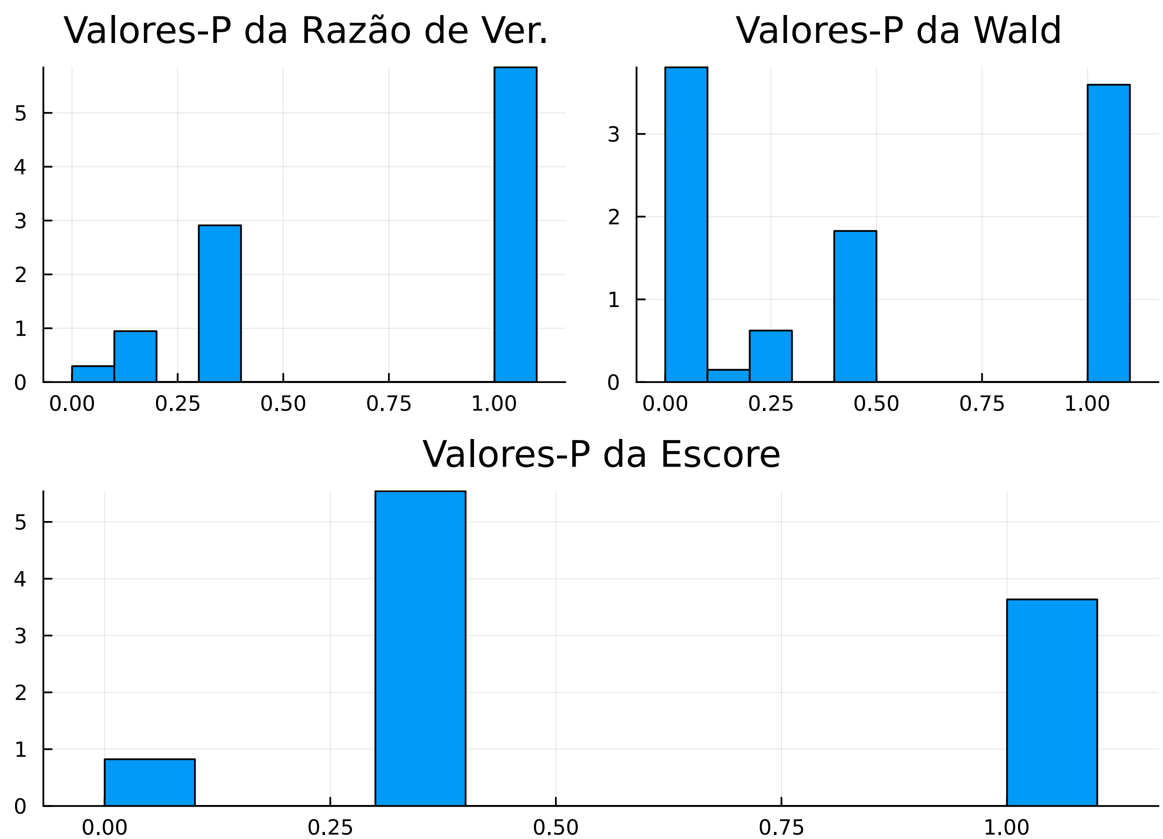
(Lembre-se que, assintoticamente, esperamos um histograma uniforme). Vamos medir:

$\hat{P}(\text{val.}-p_1 < 0.05) = 0.0566, \hat{P}(\text{val.}-p_2 < 0.05) = 0.0713, \hat{P}(\text{val.}-p_3 < 0.05) = 0.0391$

Na distribuição Poisson, espera-se que, com valores pequenos de θ, a aproximação piore pela assimetria. Vejamos com $\theta_0 = 0.01, n = 100$:

```
n = 100
p1 = @. ccdf(Chisq(1), razao(Poisson(0.01), 0.01))
p2 = @. ccdf(Chisq(1), wald(Poisson(0.01), 0.01))
p3 = @. ccdf(Chisq(1), escore(Poisson(0.01), 0.01))

hist1 = histogram(
    p1, title = "Valores-P da Razão de Ver.", label = "",
    normalize = true
)
hist2 = histogram(
    p2, title = "Valores-P da Wald", label = "",
    normalize = true
```

```
)
hist3 = histogram(
    p3, title = "Valores-P da Escore", label = "",
    normalize = true
)

plot(hist1, hist2, hist3, layout = l)
```



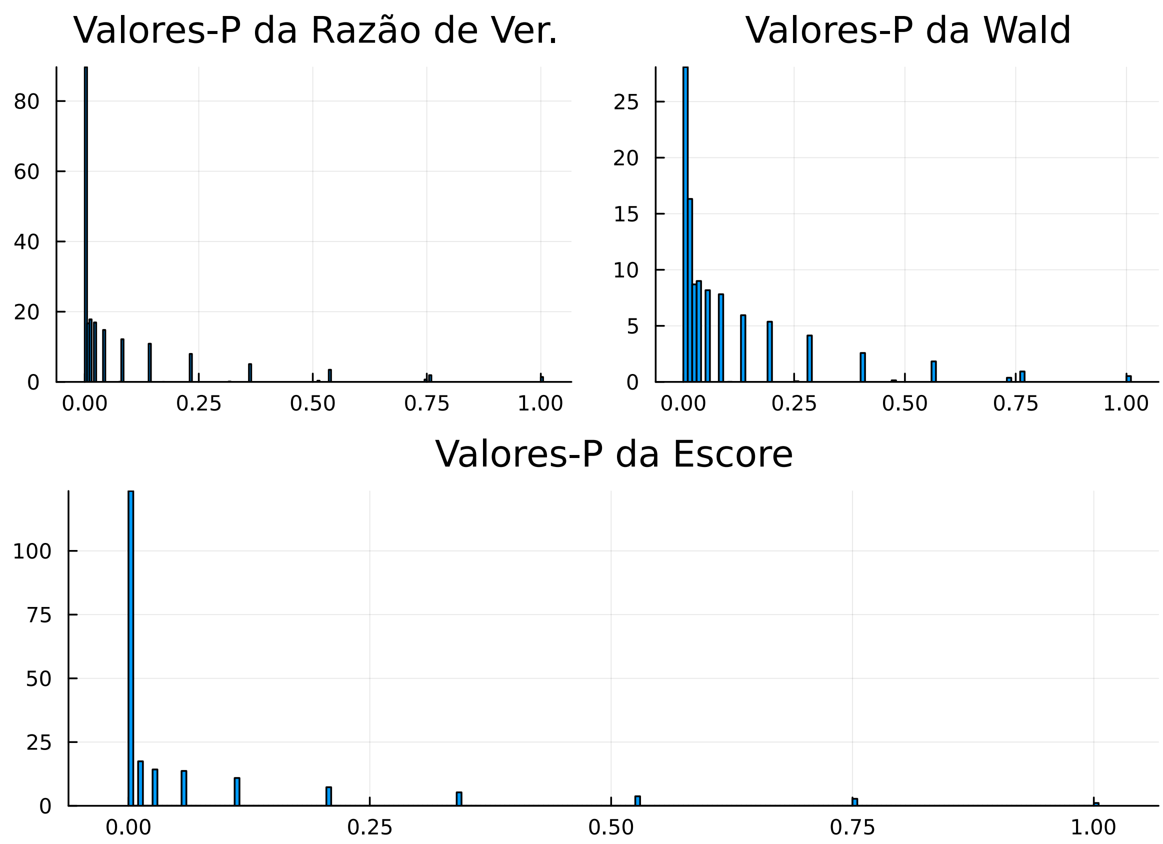
$\hat{P}(\text{val.}{-}p_1 < 0.05) = 0.019, \hat{P}(\text{val.}{-}p_2 < 0.05) = 0.3766, \hat{P}(\text{val.}{-}p_3 < 0.05) = 0.0823$

Mesmo com uma amostra maior, o efeito da assimetria fez com que a convergência assintótica fosse muito lenta.

Se gerarmos dados fora da hipótese nula, como $\theta_1 = 2$, será possível visualizar o erro no histograma:

```
n = 10
p1 = @. ccdf(Chisq(1), razao(Poisson(2), 1))
p2 = @. ccdf(Chisq(1), wald(Poisson(2), 1))
```

```
p3 = @. ccdf(Chisq(1), escore(Poisson(2), 1))

hist1 = histogram(
    p1, title = "Valores-P da Razão de Ver.", label = "",
    normalize = :pdf
)
hist2 = histogram(
    p2, title = "Valores-P da Wald", label = "",
    normalize = :pdf
)
hist3 = histogram(
    p3, title = "Valores-P da Escore", label = "",
    normalize = :pdf
)

plot(hist1, hist2, hist3, layout = l)
```



$\hat{P}(\text{val.}{-}p_1 < 0.05) = 0.7796, \hat{P}(\text{val.}{-}p_2 < 0.05) = 0.621, \hat{P}(\text{val.}{-}p_3 < 0.05) = 0.7761$