Redução de dados

Prof. Caio Azevedo

Introdução

- De posse de uma amostra aleatória $\boldsymbol{X} = (X_1, .., X_n)'$ de $X \sim F_x(, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta \subseteq \mathcal{R}^k$, desejamos construir funções de $\boldsymbol{X}$ que preservem (ou contenham) o máximo de informação possível (se possível tudo) sobre $\boldsymbol{\theta}$.
- Deseja-se que tais funções conduzam a conclusões confiáveis (acuradas) sobre o parâmetro de interesse $\boldsymbol{\theta}$.

Introdução

- Def: Uma função $T = t(\boldsymbol{X}) : \mathcal{X} \rightarrow \mathcal{R}$ é dita ser uma estatística. Ela não pode depender do parâmetro $\boldsymbol{\theta}$.
- A informação que a estatística T contem sobre θ é induzida, essencialmente, pela verossimilhança (que depende do(s) procedimento(s) amostral(is)/experimental(is) utilizado(s)) associada à $\boldsymbol{X}$ (aa) e a eventuais suposições adicionais.

Introdução

- Def: A verossimilhança associada à uma aa $\boldsymbol{X} = (X_1, ..., X_n)'$ de $X \sim F_X(.; \boldsymbol{\theta})$ (com respectiva fdp $f_X(.; \boldsymbol{\theta})$) é dada por

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

- Exemplo: Seja $X_i \overset{iid}{\sim}$ Bernoulli(θ), $\theta \in (0, 1)$ e defina $T = \sum_{i=1}^{n} X_i$. Considere n = 3, assim

$$\mathcal{X} = \{(0,0,0), (1,0,0), (0,1,0), (0,0,1), (1,1,0), (1,0,1), (0,1,1), (1,1,1)\}$$

Cont. do Exemplo

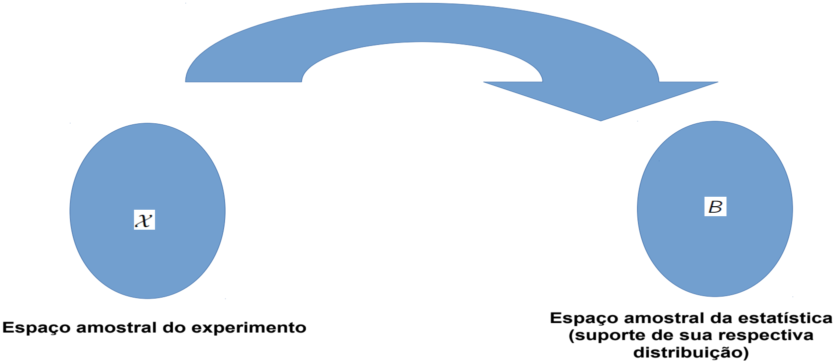
- Defina o seguinte conjunto $A_t = \{\boldsymbol{x} \in \mathcal{X} : t(\boldsymbol{x}) = t\}$.
- Observe que os conjuntos $A_t's$ $(A_t \subset \mathcal{X})$ formam uma partição de $\mathcal{X}$ pois

$$A_t \cap A_{t'} = \emptyset, \forall t \neq t'; \cup_{t \in B} A_t = \mathcal{X}$$

em que $B = \{t \in \mathcal{R}, t(\boldsymbol{x}) = t\}, \forall \boldsymbol{x} \in \mathcal{X}$.

- Note que: $A_0 = \{(0,0,0)\}$, $A_1 = \{(1,0,0),(0,1,0),(0,0,1)\}$, $A_2 = \{(1,1,0),(1,0,1),(0,1,1)\}$ e $A_3 = \{(1,1,1)\}$.

Representação gráfica (Estatística)



Outros exemplos

- μ (parâmetro de localização):
  - $T_1 = t_1(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$.
  - $T_2 = t_2(\boldsymbol{X}) = \text{med}(\boldsymbol{X})$, med: mediana.
  - $T_3 = Y_1 + \frac{Y_n}{2}$, em que Y_i é a i-ésima estatística de ordem da amostra $\boldsymbol{X}$, ou seja $Y_1 \leq Y_2 \leq ... \leq Y_n \equiv$ aos valores ordenados de forma crescente, de $\boldsymbol{X}$.

Outros exemplos

- σ^2 (parâmetro de escala).
  - $T_1 = t_1(\boldsymbol{X}) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$.
  - $T_2 = t_2(\boldsymbol{X}) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.
  - $T_3 = \frac{1}{n}\sum_{i=1}^{n}|X_i - \overline{X}|$.
- Obs: Se $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ são tais que $t(\boldsymbol{x}) = t(\boldsymbol{y})$, então as inferências a respeito de θ, considerando $\boldsymbol{x}$ ou $\boldsymbol{y}$, serão equivalentes entre si.

Suficiência

- Em geral, ou conceitos vistos nesta parte, aplicam-se, com as devidas adaptações, ao caso em que T é um vetor, ou seja $\boldsymbol{T} = (T_1, ..., T_r)'$, $T_j = t_j(\boldsymbol{X}), j = 1, 2, ..., r$, bem como quando θ for um vetor ($\boldsymbol{\theta}$).
- Def: Uma estatística T é dita ser suficiente para $\theta \in \Theta \subseteq \mathcal{R}$ se a distribuição conjunta da amostra condicionada em T, não depende de θ, ou seja se

$$f_{\boldsymbol{X}}(\boldsymbol{x}|t;\theta) = f_{\boldsymbol{X}}(\boldsymbol{x}|t), \forall \boldsymbol{x} \in \mathcal{X}.$$

Suficiência

- Em outras palavras, dado o conhecimento de T, a amostra não contem nenhuma informação relevante (adicional) a respeito de θ (embora ela ainda possa conter informação relevante, e geralmente contem, a respeito do modelo estatístico $F_{\boldsymbol{X}}(.,\theta)$).
- No caso discreto, a definição anterior traduz-se em

$$P_{\theta}(\boldsymbol{X} = \boldsymbol{x} | T = t) = P(\boldsymbol{X} = \boldsymbol{x} | T = t), \forall \boldsymbol{x} \in \mathcal{X}.$$

Suficiência

- Exemplo: Seja $X_i \overset{iid}{\sim}$ Bernoulli$(\theta), \theta \in (0,1)$. Defina $T = \sum_{i=1}^{n} X_i$. Nesse caso, (notando que $T \sim$ binomial(n,θ)), se $P(T(\boldsymbol{X}) = t(\boldsymbol{x})) \neq 0$, temos que:

$$
\begin{aligned}
P_\theta(\boldsymbol{X} = \boldsymbol{x} | T = t) &= \frac{P_\theta(\boldsymbol{X} = \boldsymbol{x}, T = t)}{P_\theta(T = t)} \\
&= \frac{\theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \mathbb{1}_{\{0,1,2,\ldots,n\}}(x_i)}{\binom{n}{t}\theta^t(1-\theta)^{n-t}\mathbb{1}_{\{0,1,2,\ldots,n\}}(t)} \\
&= \frac{1}{\binom{n}{t}} \prod_{i=1}^{n} \mathbb{1}_{\{0,1,2,\ldots,n\}}(x_i).
\end{aligned}
$$

Suficiência

- Por outro lado, se $P(T(\boldsymbol{X}) = t(\boldsymbol{x})) \equiv 0$, teremos

$$P_\theta(\boldsymbol{X} = \boldsymbol{x} | T = t) = \frac{P_\theta(\boldsymbol{X} = \boldsymbol{x}, T = t)}{P_\theta(T = t)} = 0.$$

- Logo, como a distribuição de $\boldsymbol{X}|T = t$ não depende de θ, T é uma estatística suficiente para esse parâmetro.

Suficiência

- Exemplo: Seja $X_i \overset{iid}{\sim} \text{Poisson}(\theta), \theta \in (0, \infty)$. Defina $T = \sum_{i=1}^{n} X_i$. Nesse caso, (notando que $T \sim \text{Poisson}(n\theta)$), se $P(T(\boldsymbol{X}) = t(\boldsymbol{x})) \neq 0$, temos que:

$$
\begin{aligned}
P_\theta(\boldsymbol{X} = \boldsymbol{x} | T = t) &= \frac{P_\theta(\boldsymbol{X} = \boldsymbol{x}, T = t)}{P_\theta(T = t)} \\
&= \frac{\frac{e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \prod_{i=1}^{n} \mathbb{1}_{\{0,1,2,\ldots\}}(x_i)}{\frac{e^{-n\theta}(n\theta)^t}{t!} \mathbb{1}_{\{0,1,2,\ldots,\}}(t)} \\
&= \frac{t!}{n^t \prod_{i=1} x_i!} \prod_{i=1}^{n} \mathbb{1}_{\{0,1,2,\ldots\}}(x_i)
\end{aligned}
$$

Suficiência

- Por outro lado, se $P(T(\boldsymbol{X}) = t(\boldsymbol{x})) \equiv 0$, teremos

$$P_\theta(\boldsymbol{X} = \boldsymbol{x} | T = t) = \frac{P_\theta(\boldsymbol{X} = \boldsymbol{x}, T = t)}{P_\theta(T = t)} = 0.$$

- Logo, como a distribuição de $\boldsymbol{X}|T = t$ não depende de θ, T é uma estatística suficiente para esse parâmetro.

Suficiência

- O procedimento anterior é válido quando T é uma vad. Se T for uma vac, temos que

$$P_\theta(T = t) = 0, \forall t \in \mathcal{T}(\text{espaço associado a distribuição T}), \forall \theta \in \Theta.$$

- Teorema: Seja $q_T(t;\theta)$ a fdp de $T = t(\boldsymbol{X})$. Então T é suficiente para θ, se se somente se $\frac{f_{\boldsymbol{X}}(\boldsymbol{x};\theta)}{q_T(t;\theta)}$ não depender de θ, sendo T discreta ou contínua.
- Prova para o caso discreto (para o caso contínuo, veja Testing Statistical Hypothesis (2005)).

Suficiência (dem. do teorema)

- Temos que:

$$P_\theta(\boldsymbol{X} = \boldsymbol{x} | T = t) = \frac{P_\theta(\boldsymbol{X} = \boldsymbol{x}, T = t)}{P_\theta(T = t)} \stackrel{*}{=} \frac{P_\theta(\boldsymbol{X} = \boldsymbol{x})}{P_\theta(T = t)} = \frac{f_{\boldsymbol{X}}(\boldsymbol{x}; \theta)}{q_T(t; \theta)},$$

(*) pois $\{\omega \in \mathcal{X} : \boldsymbol{X}(\omega) = \boldsymbol{x}\} \subset \{\omega \in \mathcal{X} : T(\omega) = t\}$.

Suficiência

- Exemplo: Seja $X_i \overset{iid}{\sim} U[0,\theta], \theta \in \mathcal{R}^+$. Defina $Y_n = \max\{\boldsymbol{X}\}$.
- Primeiramente note que:

$$
\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x};\theta) &= \frac{1}{\theta^n}\prod_{i=1}^{n}\mathbb{1}_{[0,\theta]}(x_i) = \frac{1}{\theta^n}\mathbb{1}_{[0,\theta]}(y_n)\mathbb{1}_{[0,y_n]}(y_1) \\
&= \frac{1}{\theta^n}\mathbb{1}_{[y_1,\theta]}(y_n)\mathbb{1}_{[0,\theta]}(y_1)\,,
\end{aligned}
$$

em que $y_1 = \min\{\boldsymbol{x}\}$. Com efeito, $0 \leq x_i \leq \theta, \forall i \leftrightarrow 0 \leq y_1 < y_n \leq \theta$.

Suficiência: cont. exemplo

- Podemos ainda provar que:

$$q_{Y_n}(y_n;\theta) = \frac{n}{\theta^n} y_n^{n-1} \mathbb{1}_{(0,\theta)}(y_n) \text{ (exercício)}$$

- Portanto, temos que

$$\frac{f_{\boldsymbol{X}}(\boldsymbol{x};\theta)}{q_{Y_n}(y_n;\theta)} = \frac{\frac{1}{\theta^n}\mathbb{1}_{[0,\theta]}(y_n)\mathbb{1}_{[0,y_n]}(y_1)}{\frac{n}{\theta^n} y_n^{n-1}\mathbb{1}_{(0,\theta)}(y_n)} = \frac{1}{n y_n^{n-1}}\mathbb{1}_{[0,y_n]}(y_1)$$

Logo Y_n é uma estatística suficiente para θ.

Teorema da Fatoração (Neyman)

- Seja $f_{\boldsymbol{X}}(.,\theta)$ a fdp de $\boldsymbol{X} = (X_1, ...X_n)'$, $\theta \in \Theta \in \mathcal{R}$. A estatística $T(\equiv t(\boldsymbol{X}))$ é suficiente para θ se, se somente se, $\exists$ duas funções, digamos, g, h, $g : \mathcal{R}^n \to \mathcal{R}^+$, $h : \mathcal{R}^n \to \mathcal{R}^+$, tais que

$$f_{\boldsymbol{X}}(\boldsymbol{x};\theta) = g(t(\boldsymbol{x});\theta)h(\boldsymbol{x})$$

em que h não depende de θ e g depende de $\boldsymbol{x}$ apenas a através de t.

Teorema da Fatoração (Neyman)

- Prova (caso discreto): $(\rightarrow)$ Defina:

$$h(\boldsymbol{x}) = P(\boldsymbol{X} = \boldsymbol{x} | T = t); g(t;\theta) = P_\theta(T = t)$$

Por outro lado, temos que:

$$\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x};\theta) &= P_\theta(\boldsymbol{X} = \boldsymbol{x}) = P_\theta(\boldsymbol{X} = \boldsymbol{x}, T(\boldsymbol{X}) = t(\boldsymbol{x})) \\
&= P_\theta(\boldsymbol{X} = \boldsymbol{x} | T = t) P_\theta(T = t) \\
&\overset{(*)}{=} P(\boldsymbol{X} = \boldsymbol{x} | T = t) P_\theta(T = t) \\
&= h(\boldsymbol{x}) g(t;\theta)
\end{aligned}$$

Teorema da Fatoração (Neyman)

- Prova: $(\leftarrow)$. Temos que:

$$P_\theta(\boldsymbol{X} = \boldsymbol{x}) = g(t(\boldsymbol{x}); \theta) h(\boldsymbol{x}).$$

- Defina $q_T(t;\theta)$ a fdp de T e $A_t = \{\boldsymbol{x} \in \mathcal{X}; t(\boldsymbol{x}) = t\}$.
- Precisamos mostrar que: $\frac{f_{\boldsymbol{X}}(\boldsymbol{x};\theta)}{q_T(t;\theta)}$ não depende de θ, para que T seja suficiente para θ.

Teorema da Fatoração (Neyman)

- Por outro lado, temos que:

$$\begin{aligned} \frac{f_{\mathbf{X}}(\mathbf{x};\theta)}{q_T(t;\theta)} &= \frac{g(t(\mathbf{x});\theta)h(\mathbf{x})}{P_\theta(T=t)} = \frac{g(t(\mathbf{x});\theta)h(\mathbf{x})}{\sum_{\mathbf{y}\in A_t} P_\theta(\mathbf{X}=\mathbf{y})} \\ &= \frac{g(t(\mathbf{x});\theta)h(\mathbf{x})}{\sum_{\mathbf{y}\in A_t} g(t(\mathbf{y});\theta)h(\mathbf{y})} \stackrel{*}{=} \frac{g(t;\theta)}{g(t;\theta)} \frac{h(\mathbf{x})}{\sum_{\mathbf{y}\in A_t} h(\mathbf{y})} \\ &= \frac{h(\mathbf{x})}{\sum_{\mathbf{y}\in A_t} h(\mathbf{y})} \end{aligned}$$

que não depende de θ ((*) note que, neste caso, $t(\mathbf{y}) \equiv t$). Como o resultado vale $\forall A_t \subseteq \mathcal{X}$, então T é suficiente para θ.

Exemplo 1

- Seja $X_1, ..., X_n$ uma aa de $X \sim U(0,\theta), \theta > 0$. Temos que

$$
\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x};\theta) &= \frac{1}{\theta^n}\prod_{i=1}^{n} \mathbb{1}_{(0,\theta)}(x_i) = \frac{1}{\theta^n}\mathbb{1}_{(0,\theta)}(y_n)\mathbb{1}_{(0,y_n)}(y_1) \\
&= g(t(\boldsymbol{x});\theta)h(\boldsymbol{x})
\end{aligned}
$$

  em que $g(t(\boldsymbol{x});\theta) = \frac{1}{\theta^n}\mathbb{1}_{(0,\theta)}(y_n)$ e $h(\boldsymbol{x}) = \mathbb{1}_{(0,y_n)}(y_1)$ e Y_i é a i-ésima estatística de ordem.

- Logo pelo critério da fatoração, temos que Y_n é uma estatística suficiente para θ.

Exemplo 2

- Seja $X_1, ..., X_n$ uma aa de $X \sim N(\mu, \sigma^2), \mu \in \mathcal{R}, \sigma^2 \in \mathcal{R}^+$. Temos que

$$
\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta}) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}\prod_{i=1}^{n}\mathbb{1}_{\mathcal{R}}(x_i) \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}[(x_i-\overline{x})+(\overline{x}-\mu)]^2\right\}\mathbb{1}_{\mathcal{R}^n}(\boldsymbol{x}) \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\overline{x})^2+n(\overline{x}-\mu)^2\right]\right\}\mathbb{1}_{\mathcal{R}^n}(\boldsymbol{x}) \\
&= g(\boldsymbol{t};\boldsymbol{\theta})h(\boldsymbol{x})
\end{aligned}
$$

Exemplo 2 (cont.)

- Pois $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$, $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Além disso,

$$g(t(\boldsymbol{x});\theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i - \overline{x})^2 + n(\overline{x} - \mu)^2\right]\right\}$$

e

$$h(\boldsymbol{x}) = \mathbb{1}_{\mathcal{R}^n}(\boldsymbol{x}).$$

- Logo pelo critério da fatoração, temos que

$$\boldsymbol{T} = \left(\overline{X}, \sum_{i=1}^{n}(X_i - \overline{X})^2\right)'$$

é uma estatística suficiente para $\boldsymbol{\theta}$.

Suficiência

- Resultados
  a) Se $T_1 = g(T_2)$ e T_1 é uma estatística suficiente para θ, então T_2 também o é.
  b) Se a função $g(.)$, do item a) for 1 a 1, então T_1 é suficiente para θ, se e somente se T_2 o for.
  - Prova(s): exercício. Sugestão: critério da fatoração.
- Obs: Toda informação, a respeito de θ, contida em T_1, também está contida em T_2.
- Obs: As informações, a respeito de θ, contidas em T_1 e T_2, são equivalentes entre si.

Exemplo 3

- Seja $X_i \overset{iid}{\sim} U(\theta, \theta+1), \theta > 0$.
  - Nesse caso, pode-se provar que $(Y_1, Y_n)'$ é uma estatística suficiente para θ.
  - Seja $g : \mathcal{R}^2 \to \mathcal{R}$, $g(x,y) = y$, nesse caso $g(\boldsymbol{T}) = g(Y_1, Y_n) = Y_n$, não é uma estatística suficiente para θ (exercício).
  - OBS: Nem toda função de uma estatística suficiente é uma estatística suficiente.
  - OBS: Dada uma estatística suficiente, podemos encontrar várias outras estatísticas suficientes, como função desta.
  - OBS: Sempre existem estatísticas suficientes, p.e., $\boldsymbol{T}_1 = (X_1, ..., X_n)'$ e $\boldsymbol{T}_2 = (Y_1, .., Y_n)'$ (estatísticas de ordem). Prova: exercício. Estas são chamadas de estatísticas suficientes triviais.

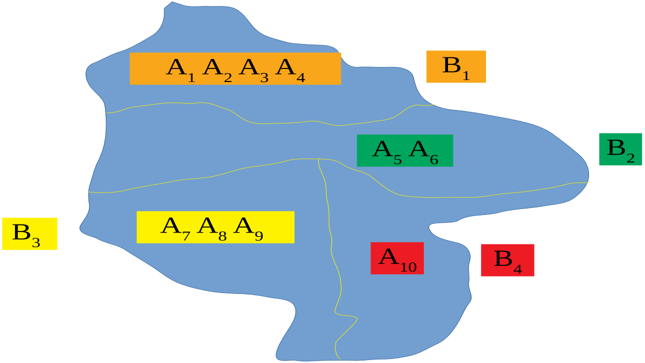
Minimalidade

- Pergunta: É possível encontrar uma estatística suficiente que consiga uma redução máxima dos dados, sem perda de informação sobre θ?

- Sejam P_1 e P_2 duas partições de $\mathcal{X}$, $P_1 = \{A_\alpha : \alpha \in I\}$ e $P_2 = \{B_\gamma : \gamma \in I'\}$, em que I e I' são índices convenientes.

- Assuma que $\forall B_\gamma \in P_2$, $\exists$ elementos de P_1, $B_\gamma = \dot\cup_{\alpha' \in J} A_{\alpha'}, J \subset I$. Ou seja (próxima figura):

Representação gráfica (partição)



Minimalidade

- Seja $T_2 = g(T_1)$, T_1 e T_2, estatísticas.
- Sejam $\boldsymbol{x}, \boldsymbol{y} \in A_1$. Então, se $T_1(\boldsymbol{x}) = T_1(\boldsymbol{y}) \rightarrow T_2(\boldsymbol{x}) = T_2(\boldsymbol{y})$.
- Sejam agora, $\boldsymbol{x} \in A_1$ e $\boldsymbol{y} \in A_4$, então:

$$T_2(\boldsymbol{x}) = T_2(\boldsymbol{y}) \nrightarrow T_1(\boldsymbol{x}) = T_1(\boldsymbol{y}).$$

- Def: Uma estatística, digamos T, é suficiente e minimal para θ, se T, além de ser suficiente, for função de toda outra estatística suficiente, digamos T'. Em outras palavras:
  - T é suficiente.
  - Se T' é suficiente $\exists$ g, $T = g(T')$.

Exemplo & Teorema

- Seja $X_i \overset{iid}{\sim} N(\theta, \sigma^2)$, σ^2 conhecido. Podemos provar que $T = \overline{X}$ é suficiente e minimal para θ.
- Seja, também, $T' = (\overline{X}, \sum_{i=1}^{n}(X_i - \overline{X})^2)$. Podemos provar que T' é suficiente para θ, contudo, $T = g(T')$.
- Teorema: Seja $f_{\boldsymbol{X}}(.;\theta)$ a fdp de $\boldsymbol{X} = (X_1, ..., X_n)'$ e $\exists$ **uma estatística T, a razão $\frac{f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})}{f_{\boldsymbol{X}}(\boldsymbol{y};\boldsymbol{\theta})}$ não depende de θ, se e somente se, $t(\boldsymbol{x}) = t(\boldsymbol{y})$, $\forall \theta \in \Theta$ e $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ (premissa)**. Então a estatística T é suficiente e minimal para θ.
- Antes de demonstrar o Teorema, discutiremos sua proposta através de um exemplo.

Exemplo

- Exemplo: Seja $X_1, .., X_n$ uma aa de $X \sim N(\mu, \sigma^2)$, temos que:

$$
\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta}) &= \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\underbrace{\sum_{i=1}^{n}(x_i - \overline{x})^2}_{s_x^2} + n(\overline{x} - \mu)^2\right]\right\} \mathbb{1}_{\mathcal{R}^n}(\boldsymbol{x}) \\
&= \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\left[s_x^2 + n\overline{x}^2 - 2n\overline{x}\mu + n\mu^2\right]\right\} \mathbb{1}_{\mathcal{R}^n}(\boldsymbol{x})
\end{aligned}
$$

- Então, $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$, temos que:

$$
\frac{f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})}{f_{\boldsymbol{X}}(\boldsymbol{y};\boldsymbol{\theta})} = \frac{\frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\left[s_x^2 + n\overline{x}^2 - 2n\overline{x}\mu + n\mu^2\right]\right\} \mathbb{1}_{\mathcal{R}^n}(\boldsymbol{x})}{\frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\left[s_y^2 + n\overline{y}^2 - 2n\overline{y}\mu + n\mu^2\right]\right\} \mathbb{1}_{\mathcal{R}^n}(\boldsymbol{y})}
$$

Exemplo

- Continuando

$$
\begin{aligned}
\frac{f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})}{f_{\boldsymbol{X}}(\boldsymbol{y};\boldsymbol{\theta})} &= \exp\left\{-\frac{1}{2\sigma^2}\left[s_x^2 - s_y^2 + n(\overline{x}^2 - \overline{y}^2) - 2\mu n(\overline{x} - \overline{y})\right]\right\} \\
&\times \mathbb{1}_{\mathcal{R}^n}(\boldsymbol{x})\mathbb{1}_{\mathcal{R}^n}(\boldsymbol{y})
\end{aligned}
$$

- Logo, $\frac{f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})}{f_{\boldsymbol{X}}(\boldsymbol{y};\boldsymbol{\theta})}$ não dependerá de $\boldsymbol{\theta}$, se e somente se, $s_x^2 = s_y^2$ e $\overline{x} = \overline{y}$.
- Portanto, $\boldsymbol{T} = \left(\overline{X}, \sum_{i=1}^{n}\left[X_i - \overline{X}\right]^2\right)'$ é uma estatística suficiente e minimal para $\boldsymbol{\theta}$.

Demonstração do Teorema (suficiência & minimalidade)

- 1) Suficiência: Sabemos que T é suficiente $\leftrightarrow$ $f_{\boldsymbol{X}}(\boldsymbol{x};\theta) = g(t;\theta)h(\boldsymbol{x})$. Além disso, a estatística T induz uma partição em $\mathcal{X}$.
- Com efeito, $P = \{A_t : A_t \subseteq \mathcal{X}, \forall t \in B\}$, $A_t = \{\boldsymbol{x} \in \mathcal{X} : t(\boldsymbol{x}) = t\}$ e $B = \{t \in \mathcal{R} : t(\boldsymbol{x}) = t$, para algum $\boldsymbol{x} \in \mathcal{X}\}$.
- Assim, para cada $A_t, \exists\, \boldsymbol{x}_t \in A_t$ e $\exists\, \boldsymbol{x} \in \mathcal{X}$, $t(\boldsymbol{x}) = t(\boldsymbol{x}_t)$.

Demonstração do Teorema (suficiência & minimalidade)

- Por outro lado, pelo teorema (suposição), temos que $\frac{f_{\boldsymbol{X}}(\boldsymbol{x};\theta)}{f_{\boldsymbol{X}}(\boldsymbol{x}_t;\theta)}$ não depende de θ.
- Defina, agora, $h(\boldsymbol{x}) = \frac{f_{\boldsymbol{X}}(\boldsymbol{x};\theta)}{f_{\boldsymbol{X}}(\boldsymbol{x}_t;\theta)}$ e $g(t;\theta) = f_{\boldsymbol{X}}(\boldsymbol{x}_t;\theta)$ (note que $g : B \rightarrow \mathcal{R}^+$, $t(\boldsymbol{x}_t) = t$, além do que se f não depender de $t(.)$, t não poderia ser uma estatística suficiente).
- Assim:

$$f_{\boldsymbol{X}}(\boldsymbol{x};\theta) = \frac{f_{\boldsymbol{X}}(\boldsymbol{x};\theta)}{f_{\boldsymbol{X}}(\boldsymbol{x}_t;\theta)} f_{\boldsymbol{X}}(\boldsymbol{x}_t;\theta) = h(\boldsymbol{x})g(t;\theta)$$

- Logo, pelo teorema da fatoração $T = t(\boldsymbol{X})$ é uma estatística suficiente para θ.
- Obs: note que, a rigor, $f_{\boldsymbol{X}}(\boldsymbol{x}_t;\theta) \equiv g^*(t;\theta)h^*(\boldsymbol{x}_t)$.

Demonstração do Teorema (sufiência & minimalidade)

- 2) Minimalidade: queremos provar que $\forall T'$ suficiente, $\exists g, T = g(T')$.
- Se T' é suficiente, $\exists\, g', h'$, funções tais que $f_{\boldsymbol{X}}(\boldsymbol{x};\theta) = g'(t';\theta)h'(\boldsymbol{x})$.
- Agora consideremos $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$, $t'(\boldsymbol{x}) = t'(\boldsymbol{y})$, então:

$$\frac{f_{\boldsymbol{X}}(\boldsymbol{x};\theta)}{f_{\boldsymbol{X}}(\boldsymbol{y};\theta)} = \frac{g'(t'(\boldsymbol{x});\theta)h'(\boldsymbol{x})}{g'(t'(\boldsymbol{y});\theta)h'(\boldsymbol{y})} = \frac{h'(\boldsymbol{x})}{h'(\boldsymbol{y})},$$

  que não depende de θ. Logo, pelo teorema (suposição), $t(\boldsymbol{x}) = t(\boldsymbol{y})$. Ou seja, se $t'(\boldsymbol{x}) = t'(\boldsymbol{y}) \rightarrow t(\boldsymbol{x}) = t(\boldsymbol{y})$. Portanto, $T = g(T')$.
- Portanto, por 1) e 2), temos que T é suficiente e minimal.

Cont.

- Obs: note que duas funções são iguais "quando têm o mesmo domínio, o mesmo contradomínio e a mesma regra de correspondência."
- No caso de razão de funções indicadoras, podemos utilizar a definição acima, por exemplo:

$$\frac{f_{\boldsymbol{X}}(\boldsymbol{x};\theta)}{f_{\boldsymbol{X}}(\boldsymbol{x}^*;\theta)} = \frac{I_{(g_1(\boldsymbol{x}),g_2(\boldsymbol{x}))}(\theta)}{I_{(g_1(\boldsymbol{x}^*),g_2(\boldsymbol{x}^*))}(\theta)}$$

não dependerá de θ se, e somente se

$$I_{(g_1(\boldsymbol{x}),g_2(\boldsymbol{x}))}(\theta) = I_{(g_1(\boldsymbol{x}^*),g_2(\boldsymbol{x}^*))}(\theta),$$

ou seja, se e somente se $g_i(\boldsymbol{x}) = g_i(\boldsymbol{x}^*), i = 1,2$ $(\forall\, \theta \in \Theta)$, em que $g_i(\boldsymbol{X})$ são estatísticas, $i - 1,2$.

Demonstração do Teorema (sufiência & minimalidade)

- OBS: Nem sempre existe uma estatística suficiente e minimal não trivial (a amostra completa, p.e.).
- Exemplo: Sejam X_1, X_2 uma *aa* de $X \sim \text{Cauchy}(\theta, 1)$.
- Se $\exists$ uma estatística, digamos T, suficiente e minimal então $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$, $t(\boldsymbol{x}) = t(\boldsymbol{y})$, teremos que:

$$\frac{f_{\boldsymbol{X}}(\boldsymbol{x};\theta)}{f_{\boldsymbol{X}}(\boldsymbol{y};\theta)} = Q(\boldsymbol{x},\boldsymbol{y}), \forall \theta \in \Theta.$$

Exemplo:

- Contudo, note que

$$\begin{aligned}\frac{f_{\boldsymbol{X}}(\boldsymbol{x};\theta)}{f_{\boldsymbol{X}}(\boldsymbol{y};\theta)} &= \frac{\frac{1}{1+(x_1-\theta)^2}\frac{1}{1+(x_2-\theta)^2}}{\frac{1}{1+(y_1-\theta)^2}\frac{1}{1+(y_2-\theta)^2}}Q(\boldsymbol{x},\boldsymbol{y}) \\ &\leftrightarrow (1+(y_1-\theta)^2)(1+(y_2-\theta)^2) \\ &- (1+(x_1-\theta)^2)(1+(x_2-\theta)^2)Q(\boldsymbol{x},\boldsymbol{y}) = 0.\end{aligned}$$

- Que tem a forma de um polinômio de 4^o grau em θ, ou seja, $a_0 + a_1\theta + a_2\theta^2 + a_3\theta^3 + a_4\theta^4 = 0$, em que $a_i = g(\boldsymbol{x},\boldsymbol{y}), i = 0, 1, 2, 3, 4$.
- Mas isso ocorre, no máximo, para 4 valores de θ, o que é uma contradição, pois deveria valer $\forall\theta \in \Theta$.
- OBS: Veremos uma forma conveniente de obter estatísticas suficientes e minimais, mais adiante.

Estatística ancilar

- Podemos também definir estatísticas com características opostas as estatísticas suficientes, ou seja, que não contenham informações (relevantes) sobre o parâmetro.
- Def: Dizemos que uma estatística $S = s(\boldsymbol{X}), s : \mathcal{R}^n \to \mathcal{R}$, é dita ancilar para θ, $\theta \in \Theta \subset \mathcal{R}$, se sua distribuição não depende de θ. Ou seja, se

$$S \sim F_\theta \equiv F.$$

- Exemplo: Seja $X_1, ..., X_n$ uma aa de $X \sim U(\theta, \theta + 1), \theta \in \mathcal{R}$ e defina $R = Y_n - Y_1$, Y_i é a i-ésima estatística de ordem.
- Vamos provar que R é uma estatística ancilar para θ.
- Defina $\boldsymbol{T} = \left(Y_n - Y_1, \frac{Y_1+Y_n}{2}\right)' = (R, M)'$.

Estatística ancilar

- Podemos mostrar (exerício) que se $\boldsymbol{Y} = (Y_1, Y_n)'$ então

$$f_{\boldsymbol{Y}}(\boldsymbol{y};\theta) = n(n-1)\left(y_n - y_1\right)^{n-2} \mathbb{1}_{(\theta,y_n)}(y_1)\mathbb{1}_{(\theta,\theta+1)}(y_n).$$

- Por outro lado, pelo método do Jacobiano, temos que

$$f_{(R,M)}(r,m;\theta) = n(n-1)r^{n-2}\mathbb{1}_{(0,1)}(r)\mathbb{1}_{\left(\theta+\frac{r}{2},\theta+1-\frac{r}{2}\right)}(m) \qquad (1)$$

- Além disso, temos (integrando (1) com relação à va M) que:

$$f_R(r) = n(n-1)r^{n-2}(1-r)\mathbb{1}(0,1)(r)$$

- Ou seja, $R \sim \text{beta}(n-1,2)$, que não depende de θ.

Estatística ancilar

- Outra forma de resolver: note que $X \sim U(\theta, \theta+1)$ pertence à família de localização. Assim, se $Z = X - \theta$, a distribuição de Z não depende de θ.
- Por outro lado, temos que:

$$
\begin{aligned}
R = Y_n - Y_1 &= \max(X_1, ..., X_n) - \min(X_1, ..., X_n) \\
&= \max(Z_1 + \theta, ..., Z_n + \theta) - \min(Z_1 + \theta, ..., Z_n + \theta) \\
&= \max(Z_1, ..., Z_n) + \theta - \min(Z_1, ..., Z_n) - \theta \\
&= Z_{(n)} - Z_{(1)}
\end{aligned}
$$

Estatística ancilar

- Exercício: Seja $X_1, ..., X_n$ uma aa de X, em que X pertence à família de escala. Mostre que:

$$S(\boldsymbol{X}) = f\left(\frac{X_1}{X_n}\right) + f\left(\frac{X_2}{X_n}\right) + ... + f\left(\frac{X_{n-1}}{X_n}\right)$$

é uma estatística ancilar para σ^2 (considere que f(.) é contínua).

- Em particular $S(\boldsymbol{X}) = \frac{X_1 + X_2 + ... + X_{n-1}}{X_n}$ é ancilar.
- Mostre também que $S = \frac{Y_n - Y_1}{Y_n + Y_1}$ é ancilar para σ^2.
- Exemplo: $X_1, ..., X_n$ é uma aa de X, $X \sim \exp(\theta)$, $f_X(x;\theta) = \frac{e^{-x/\theta}}{\theta} \mathbb{1}_{(0,\infty)}(x)$, Prove que $S = \frac{Y_n}{Y_1}$ é uma estatística ancilar para θ.
- Exercício: Seja $X_1, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, σ^2 conhecido. Encontre uma estatística ancilar para μ.

Estatística ancilar

- OBS: Informação (material) ancilar = informação irrelevante com respeito ao parâmetro.
- Vimos que é possível construir funções de estatísticas suficientes que são estatísticas ancilares.
- Isso quer dizer que, apesar de uma estatística suficiente conter toda informação relevante sobre o parâmetro, ela também (pode conter) contêm informação não relevante (em termos inferenciais) sobre o parâmetro.

Estatística ancilar

- É como se a informação relevante fosse “contaminada” pela informação ancilar.
- Isto, por sua vez, compromete propriedades ótimas que uma estatística suficiente (poderia apresentar) apresenta, no sentido de conduzir à obtenção de estimadores (pontuais e intervalares) e testes ótimos, como discutiremos mais adiante.

Estatística completa

- Def: Seja $f_T(.;\theta), \theta \in \Theta \subset \mathcal{R}$ a fdp de uma estatística T. A família de distribuições $f_T(.;\theta)$ é dita ser completa se:
    - $\mathcal{E}_\theta(g(T)) = 0, \forall \theta \in \Theta, \rightarrow P_\theta(g(T) = 0) = 1$, ou se,
    - $\mathcal{E}_\theta(g(T)) = 0, \forall \theta \in \Theta, \rightarrow g(t) = 0, \forall t \in B$ (suporte de T).
- Em particular, dizemos que T é uma estatística completa.
- Note que a definição implica na impossibilidade de se definir funções (estatísticas) ancilares a partir de T (funções de T).

Estatística completa

- Exemplo 1: Seja $X \sim N(0, \sigma^2)$, $\sigma^2 > 0$ e $T = X$.
    - Se $T = X \rightarrow \mathcal{E}(T) = \mathcal{E}(X) = 0$, mas $g(x) = x \neq 0, \forall x \neq 0$.
    - Nesse caso, X não é uma estatística completa.
- Exemplo 2: Sejam $X_1, X_2 \overset{iid}{\sim} N(\theta, 1)$, $\theta \in \mathcal{R}$. Defina $T_1 = X_1 + X_2$ e $T_2 = X_1 - X_2$.
- Neste caso, podemos provar que T_1 é uma estatística completa enquanto que T_2 não o é.

Estatística completa

- Dois resultados que ajudam a provar que uma estatística é completa são:
  a) Seja $h(x;\theta)$ uma função diferenciável com relação à θ e $\frac{\partial h(x;\theta)}{\partial \theta}$ contínua em função de x e θ, então,

$$\begin{aligned}\frac{d}{d\theta}\int_{a(\theta)}^{b(\theta)} h(x;\theta)dx &= h(b(\theta);\theta)\frac{db(\theta)}{d\theta} \\ &- h(a(\theta),\theta)\frac{da(\theta)}{d\theta} + \int_{a(\theta)}^{b(\theta)} \frac{\partial}{\partial\theta}h(x;\theta)dx\end{aligned}$$

  Em particular $\frac{d}{d\theta}\int_0^{\theta} g(x)dx = g(\theta)$.

  b) Se $\sum_{k=1}^{\infty} c_k\theta^k = 0, \forall\theta \in \Theta \subseteq \mathcal{R}$, $\Theta \neq \emptyset$, então $c_k = 0$, $\forall k$.

Cont.

- Especificamente, para a verificação da completitude, as funções $g(.)$ devem ser tais que: $\mathcal{E}(g(T)) = \int_B g(t)f_T(t)dt = 0$, portanto $|g(t)f_T(t)| < c, \forall t \in B$.
- Para $B = (-\infty, b(\theta))$, temos que

$$\int_{-\infty}^{b(\theta)} g(t)f_T(t)dt \rightarrow \lim_{a \rightarrow -\infty} \int_{a}^{b(\theta)} g(t)f_T(t)dt$$

$$\rightarrow \quad \frac{d}{d\theta}\int_{-\infty}^{b(\theta)} g(t)f_T(t)dt = g(b(\theta))f_T(b(\theta))\frac{db(\theta)}{d\theta}$$

$$- \lim_{a \rightarrow -\infty} g(b(\theta))f_T(b(\theta))\frac{da}{d\theta} = g(b(\theta))f_T(b(\theta))\frac{db(\theta)}{d\theta}$$

- Analogamente, para $B = (a(\theta), \infty)$, temos que

$$\frac{d}{d\theta}\int_{a(\theta)}^{\infty} g(t)f_T(t)dt = -g(a(\theta))f_T(a(\theta))\frac{da(\theta)}{d\theta}$$

Estatística completa

- Exemplo 1: Seja $X_1, ..., X_n \overset{iid}{\sim}$ Bernoulli$(\theta), \theta \in (0,1)$ e defina $T = \sum_{i=1}^{n} X_i$. Mostre que T é uma estatística completa.
- Queremos provar que, se $\mathcal{E}_\theta(g(T)) = 0 \to g(T) = 0, \forall t \in B$.
- Sabemos que $T \sim$ binomial(n, θ). Então

$$
\begin{aligned}
\mathcal{E}_\theta(g(T)) &= 0 \to \sum_{t=0}^{n} g(t) \binom{n}{t} \theta^t (1-\theta)^{n-t} = 0 \\
&= \sum_{t=0}^{n} g(t) \binom{n}{t} \left(\frac{\theta}{1-\theta}\right)^t = 0 \\
&= \sum_{t=0}^{n} g(t) \binom{n}{t} \gamma^t = 0 \qquad (2)
\end{aligned}
$$

Estatística completa

- Pelo resultado b), se (2) vale, então $g(t)\binom{n}{t} = 0, \forall t \in B \leftrightarrow g(t) = 0, \forall t \in B$, pois $\binom{n}{t} > 0, \forall t \in B$.
- Logo T é uma estatística completa.

Estatística completa

- Exemplo 2: Seja $X_1, ..., X_n \overset{iid}{\sim} U(0, \theta), \theta \in \mathcal{R}^+$ e defina $T = \max(X_1, ..., X_n)$. Mostre que T é uma estatística completa.
- Sabemos que $h_T(t;\theta) = \frac{nt^{n-1}}{\theta^n} \mathbb{1}_{(0,\theta)}(t)$ (fdp). Assim (lembre que $(n, \theta) > 0$),

$$\mathcal{E}_\theta(g(T)) = 0 \rightarrow \int_0^\theta g(t) \frac{nt^{n-1}}{\theta^n} dt = 0 \rightarrow \int_0^\theta g(t) t^{n-1} dt = 0 \quad (3)$$

- Derivando (3) com relação θ, temos que

$$g(\theta)\theta^{n-1} = 0 \rightarrow g(\theta) = 0, \forall \theta \in \Theta \qquad (4)$$

- Note que $B \subseteq \Theta$. Assim, em particular, $g(t) = 0, \forall t \in B$.
- Portanto, T é uma estatística completa.

Teorema de Basu

- Seja T uma estatística suficiente e completa para θ. Então T é estatisticamente independente de qualquer estatística ancilar.
- Prova para o caso discreto: Queremos provar que $P_\theta(S = s, T = t) = P_\theta(S = s)P_\theta(T = t), \forall s, t, \theta$.
- Seja $S = s(\boldsymbol{X})$ uma estatística ancilar para θ. Logo $f_S(s; \theta) = P_\theta(S = s) \equiv P(S = s)$, não depende de θ.
- Defina, $\forall s \in \mathcal{S}$ (suporte de S), s fixado:

$$g(t) = P_\theta(S = s | T = t) - P_\theta(S = s)$$

que não depende de θ pois, $P_\theta(S = s | T = t) = \sum_{\boldsymbol{x} \in \mathcal{X} : s(\boldsymbol{x}) = s} P(\boldsymbol{X} = \boldsymbol{x} | T = t(\boldsymbol{x}))$, (esta última expressão não depende de θ, pois T é suficiente para θ).

Teorema de Basu

- Por outro lado ($\sum_t \equiv \sum_{t \in B} \equiv \sum_{\boldsymbol{x} \in \mathcal{X}: t(\boldsymbol{x})=t}$, B é o suporte de T)

$$
\begin{aligned}
\mathcal{E}_\theta(g(T)) &= \sum_t g(t) P_\theta(T=t) \\
&= \sum_t [P(S=s|T=t) - P(S=s)] P_\theta(T=t) \\
&= \sum_t P(S=s|T=t) P_\theta(T=t) - P(S=s) \underbrace{\sum_t P_\theta(T=t)}_{1} \\
&= \sum_t P_\theta(S=s, T=t) - P(S=s) \\
&= P(S=s) - P(S=s) = 0 \\
&\rightarrow \mathcal{E}_\theta(g(T)) = 0 \qquad (5)
\end{aligned}
$$

Teorema de Basu

- Como T é uma estatística suficiente e completa, (5) implica que :

$$
\begin{aligned}
g(T) &\equiv 0\,\forall t \in B, \text{ ou seja} \\
P(S = s|T = t) - P(S = s) &= 0 \\
\rightarrow P(S = s|T = t) &= P(S = s)\forall t \in B, \text{ e como s é arbitrário} \\
P(S = s|T = t) &= P(S = s), \forall t, s, \theta \\
&\quad \text{(ambas expressões independem de } \theta)
\end{aligned}
$$

- Logo, T e S são estatisticamente independentes.

Exemplos

- Exemplo 1: Seja $X_1, ..., X_n$ uma aa de $X \sim N(\mu, \sigma^2)$, σ^2 conhecido. Então $T = \overline{X}$ e $S = \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$ são independentes.
  - Nesse caso, $\overline{X}$ é suficiente e completa (temos que provar a completitude) para μ.
  - Por outro lado, $\frac{S}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \sim \chi^2_{(n-1)} \rightarrow S \sim \sigma^2 \chi^2_{(n-1)}$.
  - Assim, S é ancilar para μ. Portanto, pelo Teorema de Basu, $\overline{X}$ e S são estatisticamente independentes.
- Exemplo 2: Seja $X_1, ..., X_n$ uma aa de $X \sim U(0, \theta), \theta > 0$. Então Y_n (já provamos que é suficiente e completa) e $S = \frac{Y_n - Y_1}{Y_1 + Y_n}$ são estatisticamente independentes (exercício).

Voltando à Família Exponencial

- Teorema 1: Seja $X_1, ..., X_n$ uma aa de $X \sim FE_k(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subseteq \mathcal{R}^k$ (X pode ser um vetor aleatório). Então a estatística $\boldsymbol{T} = (T_1, ..., T_k)'$ é uma estatística sucifiente para $\boldsymbol{\theta}$ (este resultado também vale para a família exponencial curvada).

  Prova: Critério da Fatoração (exercício).

Voltando à Família Exponencial

- Teorema 2: Seja $X_1, ..., X_n$ uma aa de X cuja fdp pode ser escrita como $(FE_1(\theta), \theta \in \Theta \subseteq \mathcal{R})$:

$$
\begin{aligned}
f_X(x;\theta) &= h(x)\exp\{c(\theta)t(x) + d(\theta)\}\,\mathbb{1}_A(x) \\
&= h(x)\exp\{\eta t(x) + d_0(\eta)\}\,\mathbb{1}_A(x).
\end{aligned}
$$

Então se $\Gamma = \left\{\eta \in \mathcal{R} : \int_A h(x)\exp\{\eta t(x)\}\,dx < \infty\right\}$ contiver algum intervalo aberto em $\mathcal{R}$ então $T = t(\boldsymbol{X}) = \sum_{i=1}^{n} t(X_i)$ será completa e minimal.

Voltando à Família Exponencial

- Teorema 3: Seja $X_1, ..., X_n$ uma aa de X cuja fdp é da forma $(FE_K(\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta \subseteq \mathcal{R}^k)$:

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta}) &= h(\boldsymbol{x}) \exp\left\{\sum_{j=1}^{k} c_j(\boldsymbol{\theta}) t_j(\boldsymbol{x}) + d(\boldsymbol{\theta})\right\} \mathbb{1}_{A^*}(\boldsymbol{x}) \\ &= h(\boldsymbol{x}) \exp\left\{\sum_{j=1}^{k} \eta_j t_j(\boldsymbol{x}) + d_0(\boldsymbol{\eta})\right\} \mathbb{1}_{A^*}(\boldsymbol{x}). \end{aligned}$$

Então se $\Gamma = \left\{\eta \in \mathcal{R}^k : \int_{A^*} h(\boldsymbol{x}) \exp\left\{\eta t(\boldsymbol{x})\right\} d\boldsymbol{x} < \infty\right\}$ contiver algum "retângulo" aberto em $\mathcal{R}^k$, então $\boldsymbol{T} = (t_1(\boldsymbol{X}), ..., t_k(\boldsymbol{X}))' = \left(\sum_{i=1}^n t_1(X_1), ..., \sum_{i=1}^n t_k(X_k)\right)$ será completa e minimal, em que A^* é o suporte associado ao vea $\boldsymbol{X}$.

Prova(s): Lehmann & Romano (Testing Statistical Hypothesis).

Voltando à Família Exponencial

- Discussão sobre completitude: Seja $f_X(., \theta)$ na forma da família exponencial canônica ($\eta \in \mathcal{R}$).

- Seja T a estatística inerente à FE. Queremos provar que se (caso contínuo).

$$\mathcal{E}_\eta(g(T)) = \int_B g(t) \exp\left\{t\eta + d_0(\eta)\right\} h_0(t)dt = 0 \to g(t) = 0 \forall\, t \in B$$

- Seja $g(t) = g^+(t) - g^-(t), \forall t \in B$, em que $g^+(t) = g(t)\mathbb{1}_{\mathcal{R}^+}(g(t))$ e $g(t) = (-g(t))\mathbb{1}_{(\mathcal{R}^-)}(g(t))$. Assim

$$\mathcal{E}_\eta(g(T)) = 0 \to \int g^+(t) \exp\left\{t\eta\right\} h_0(t)dt = \int g^-(t) \exp\left\{t\eta\right\} h_0(t)dt$$

$\forall \eta \in N(\eta_0) = \{\eta \in \mathcal{R} : |\eta - \eta_0| < \epsilon\}$, para algum $\epsilon > 0$.

Voltando à Família Exponencial

- Em particular,

$$\rightarrow \int g^{+}(t) \exp\{t\eta\} h_0(t)dt = \int g^{-}(t) \exp\{t\eta\} h_0(t)dt = c \quad (6)$$

- Assim $\frac{1}{c}g^{+}(t) \exp\{t\eta\} h_0(t)$ e $\frac{1}{c}g^{-}(t) \exp\{t\eta\} h_0(t)$ são fdp's. Por (6) as fgm's dessa fdp's são iguais numa vizinhança de 0. Então, pelo teorema da unicidade, temos que

$$g^{+}(t) = g^{-}(t) \rightarrow g(t) = 0 \forall\, t \in B.$$

Voltando à Família Exponencial

- Com relação à minimalidade (de $\boldsymbol{T}$) devemos provar que $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ $\frac{f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta})}{f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta})} = Q(\boldsymbol{x}, \boldsymbol{y}) \leftrightarrow t(\boldsymbol{x}) = t(\boldsymbol{y})$.
- Com efeito, temos que:

$$
\begin{aligned}
Q(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\theta}) &= \exp\left\{\sum_{j=1}^{k} \eta_j t_j(\boldsymbol{x}) - \sum_{i=1}^{k} \eta_j t_j(\boldsymbol{y})\right\} \frac{h(\boldsymbol{x})}{h(\boldsymbol{y})} \\
&= \exp\left\{\sum_{j=1}^{k} \eta_j \left(t_j(\boldsymbol{x}) - t_j(\boldsymbol{y})\right)\right\} \frac{h(\boldsymbol{x})}{h(\boldsymbol{y})},
\end{aligned}
$$

o que prova a minimalidade de $\boldsymbol{T}$.

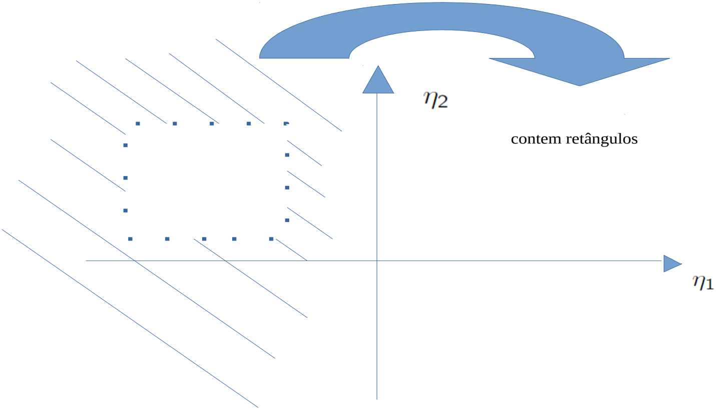
Voltando à Família Exponencial

- Exemplo 1: Seja $X_1, ..., X_n$ uma aa de X, $X \sim N(\mu, \sigma^2)$. Temos que:

$$
\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta}) &= \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^{n} x_i - n\frac{\mu}{2\sigma^2}\right\}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \mathbb{1}_{\mathcal{R}^n}(\boldsymbol{x}) \\
&= \exp\left\{\sum_{j=1}^{2}\eta_j t_j(\boldsymbol{x}) + d_0(\boldsymbol{\eta})\right\} h(\boldsymbol{x}),
\end{aligned}
$$

em que $\eta_1 = -\frac{1}{\sigma^2}$ e $\eta_2 = \frac{\mu}{\sigma^2}$. Graficamente, veja slide seguinte, o espaço paramétrico natural contem retângulos no $\mathcal{R}^2$. Assim, por esse fato e como $f_{\boldsymbol{X}}(.;\boldsymbol{\theta}) \in FE_2(\boldsymbol{\theta})$, temos que $\boldsymbol{T} = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)'$ é uma estatística suficiente, completa e minimal para $\boldsymbol{\theta} = (\mu, \sigma^2)'$.

Representação gráfica (espaço paramétrico)



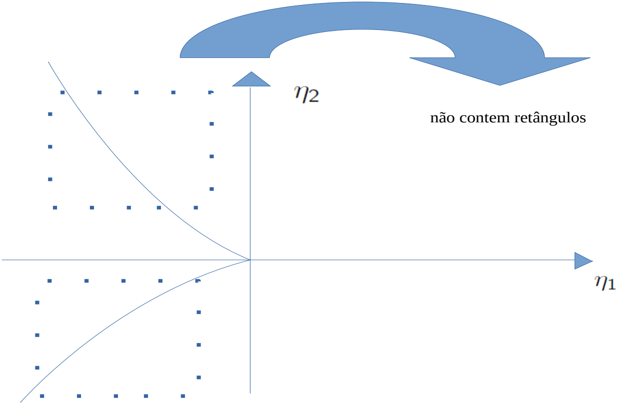
Voltando à Família Exponencial

- Exemplo 2: Seja $X_1, ..., X_n$ uma aa de X, $X \sim N(\theta, \theta^2)$. Temos que:

$$
\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x};\theta) &= \left(\frac{1}{\sqrt{2\pi}\theta}\right)^n \exp\left\{-\frac{1}{2\theta^2}\sum_{i=1}^n x_i^2 + \frac{1}{\theta}\sum_{i=1}^n x_i - n\right\} \mathbb{1}_{\mathcal{R}^n}(\boldsymbol{x}) \\
&= \exp\left\{\sum_{j=1}^2 \eta_j t_j(\boldsymbol{x}) + d_0(\boldsymbol{\eta})\right\} h(\boldsymbol{x}),
\end{aligned}
$$

em que $\eta_1 = -\frac{1}{\theta^2}$ e $\eta_2 = \frac{1}{\theta}$. Graficamente, veja slide seguinte, o espaço paramétrico natural não contem retângulos no $\mathcal{R}^2$. Assim, não podemos afirmar que a estatística $\boldsymbol{T} = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)'$ é uma completa e minimal (embora ela seja suficiente).

Representação gráfica (espaço paramétrico)



Mais exemplos

- Sejam $X_1, ..., X_n$ uma aa de X, em que

  a) $X \sim U(0, \theta)$.

  b) $X \sim U(-\theta, \theta)$.

  c) $X \sim U(\theta, \theta + 1)$.

- Note que nenhum dos três modelos pertencem à FE.

Mais exemplos

a) Relembrando:

$$f_{\boldsymbol{X}}(\boldsymbol{x};\theta) = \frac{1}{\theta^n}\prod_{i=1}^{n}\mathbb{1}_{(0,\theta)}(x_i)$$

Mas

$$\begin{aligned}\prod_{i=1}^{n}\mathbb{1}_{(0,\theta)}(x_i) = 1 \quad &\leftrightarrow \quad 0 < x_i < \theta, \forall i \leftrightarrow 0 < \min(\boldsymbol{x}) < \max(\boldsymbol{x}) < \theta \\ &\leftrightarrow \quad 0 < y_1 < y_n < \theta \leftrightarrow \mathbb{1}_{(0,\theta)}(y_1)\mathbb{1}_{(y_1,\theta)}(y_n) = 1 \\ &\leftrightarrow \quad \mathbb{1}_{(0,\theta)}(y_n)\mathbb{1}_{(0,y_n)}(y_1) = 1\end{aligned}$$

Mais exemplos

a) Já vimos que

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x};\theta) &= \frac{1}{\theta}\prod_{i=1}^{n}\mathbb{1}_{(-\theta,\theta)}(x_i) = \frac{1}{\theta^n}\mathbb{1}_{(0,\theta)}(y_n)\mathbb{1}_{(0,y_n)}(y_1) \\ &= \frac{1}{\theta^n}\mathbb{1}_{(0,\theta)}(y_1)\mathbb{1}_{(y_1,\theta)}(y_n) \end{aligned}$$

Também já vimos que Y_n é suficiente (pág. 18) e completa (pág. 50). Em relação à minimalidade, note que $\forall \mathbf{x}, \mathbf{x}^* \in \mathcal{X}$, temos que

$$\frac{f_{\mathbf{X}}(\mathbf{x};\theta)}{f_{\mathbf{X}}(\mathbf{x}^*;\theta)} = \frac{\frac{1}{\theta^n}\mathbb{1}_{(0,\theta)}(y_n)\mathbb{1}_{(0,y_n)}(y_1)}{\frac{1}{\theta^n}\mathbb{1}_{(0,\theta)}(y_n^*)\mathbb{1}_{(0,y_n^*)}(y_1^*)},$$

que não dependerá de $\theta \leftrightarrow y_n = y_n^*$, em que $y_1^* = \min(\mathbf{x}^*)$, $\mathbf{y}_n^* = \max(\mathbf{x}^*)$ e $\mathbf{x}^* = (x_1^*, ..., x_n^*)'$. Portanto, Y_n também é minimal.

Mais exemplos

b) Neste caso, temos uma estrutura semelhante ao caso anterior (item a)), com efeito:

$$
\begin{aligned}
f_{\boldsymbol{X}}(\boldsymbol{x};\theta) &= \frac{1}{(2\theta)^n}\prod_{i=1}^{n}\mathbb{1}_{(-\theta,\theta)}(x_i) = \frac{1}{(2\theta)^n}\prod_{i=1}^{n}\mathbb{1}_{(0,\theta)}(|x_i|) \\
&= \frac{1}{(2\theta)^n}\mathbb{1}_{(0,\theta)}(r_n)\mathbb{1}_{(0,r_n)}(r_1) = \frac{1}{(2\theta)^n}\mathbb{1}_{(0,\theta)}(r_1)\mathbb{1}_{(r_1,\theta)}(r_n) \\
&= h(\boldsymbol{x})g(r_n;\theta),
\end{aligned}
$$

em que $R_n = \max(|X_1|,...,|X_n|)$ e $R_1 = \min(|X_1|,...,|X_n|)$. Assim, pelo critério da fatoração, temos que R_n é uma estatística suficiente para θ.

Mais exemplos

b) Em relação à minimalidade, note que $\forall \boldsymbol{x}, \boldsymbol{x}^* \in \mathcal{X}$, temos que

$$\frac{f_{\boldsymbol{X}}(\boldsymbol{x};\theta)}{f_{\boldsymbol{X}}(\boldsymbol{x}^*;\theta)} = \frac{\frac{1}{(2\theta)^n}\mathbb{1}_{(0,\theta)}(r_n)\mathbb{1}_{(0,r_n)}(r_1)}{\frac{1}{(2\theta)^n}\mathbb{1}_{(0,\theta)}(r_n^*)\mathbb{1}_{(0,r_n^*)}(r_1^*)}$$

que não dependerá de $\theta \leftrightarrow r_n = r_n^*$, em que $r_1^* = \min(\boldsymbol{x}^*)$, $\boldsymbol{r}_n^* = \max(\boldsymbol{x}^*)$ e $\boldsymbol{x}^* = (x_1,^*, ..., x_n^*)'$. Portanto, R_n também é minimal.

Mais exemplos

b) Com relação à completitude, analogamente ao caso anterior (item a)), temos que:

- Sabemos que, se $R_n = T$, então $h_T(t;\theta) = \frac{nt^{n-1}}{\theta^n} \mathbb{1}_{(0,\theta)}(t)$ (fdp) (provar). Assim (lembre que $(n,\theta) > 0$),

$$\mathcal{E}_\theta(g(T)) = 0 \rightarrow \int_0^\theta g(t)\frac{nt^{n-1}}{\theta^n} dt = 0 \rightarrow \int_0^\theta g(t)t^{n-1} dt = 0 \quad (7)$$

- Derivando (7) com relação θ, temos que

$$g(\theta)\theta^{n-1} = 0 \rightarrow g(\theta) = 0, \forall \theta \in \Theta$$

- Note que $B \subseteq \Theta$. Assim, em particular, $g(t) = 0, \forall t \in B$.
- Portanto, R_n também é uma estatística completa.

Mais exemplos

c) Já vimos que

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x};\theta) &= \prod_{i=1}^{n} \mathbb{1}_{(\theta,\theta+1)}(x_i) = \mathbb{1}_{(\theta,\theta+1)}(y_1)\mathbb{1}_{(y_1,\theta+1)}(y_n) \\ &= \mathbb{1}_{(\theta,\theta+1)}(y_n)\mathbb{1}_{(\theta,y_n)}(y_1) = h(\boldsymbol{x})g(\boldsymbol{t};\theta). \end{aligned}$$

Logo, pelo critério da fatoração, $\boldsymbol{T} = (Y_1, Y_n)'$ é uma estatística suficiente. Com relação à minimalidade, $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$, note que:

$$\frac{\mathbb{1}_{(\theta,\theta+1)}(y_n)\mathbb{1}_{(\theta,y_n)}(y_1)}{\mathbb{1}_{(\theta,\theta+1)}(y_n^*)\mathbb{1}_{(\theta,y_n^*)}(y_1^*)},$$

que não depende de $\theta \leftrightarrow \boldsymbol{T} = \boldsymbol{T}^*$, em que $\boldsymbol{T}^* = (Y_1^*, Y_n^*)$, $y_1^* = \min(x_1^*, ..., x_n^*)$ e $y_n^* = \max(x_1, ..., x_n)$.

Mais exemplos

c) Com relação à completitude, já vimos que

$$R = Y_n - Y_1 \sim \text{beta}(n-1, 2).$$

Defina, agora, $R^* = R - \mathcal{E}(R) = R - \frac{n-1}{n+1}$. Portanto, $\mathcal{E}(R^*) = \mathcal{E}(R) - \mathcal{E}(R) = 0$ mas $R^* \neq 0$. Assim, $\boldsymbol{T}$ não é completa.

Invariância

- O conceito de invariância (na estatística) está relacionado com mudanças específicas na distribição de estatísticas sob certos tipo de transformações.
- Essencialmente, veremos os conceitos de invariância por localização e por escala.
- Sua utilidade está relacionada a demonstração de certos resultados importantes.
- Para sua distinção em relação ao conceito de **equivariância**, veja aqui.

Invariância por localização

- Dizemos que uma estatística, digamos $T = t(\boldsymbol{X})$, é invariante por localização se $T = t(X_1 + c, ..., X_n + c) = T + c$, $c \in \mathcal{R}$ (constante)
- Exemplos
  - $T_1 = \overline{X}$.
  - $T_2 = Y_1 = \min(\boldsymbol{X})$.
  - $T_3 = Y_n = \max(\boldsymbol{X})$.
  - $T_4 = \frac{Y_1 + Y_n}{2}$.
- Teorema: Seja $X_1, ..., X_n$ uma aa de X, com fdp $f_X(;\theta)$, $\theta \in \mathcal{R}$. Se θ for uma parâmetro de localização, então

$$T_P = t_P(\boldsymbol{X}) = \frac{\int_{-\infty}^{\infty} \theta f_{\boldsymbol{X}}(\boldsymbol{X};\theta)d\theta}{\int_{-\infty}^{\infty} f_{\boldsymbol{X}}(\boldsymbol{X};\theta)d\theta} = \frac{\int_{-\infty}^{\infty} \theta \prod_{i=1}^{n} f_X(X_i;\theta)d\theta}{\int_{-\infty}^{\infty} \prod_{i=1}^{n} f_X(X_i;\theta)d\theta},$$

é uma estatística invariante por localização.

Invariância por localização

- OBS:
  1. T_p é uma estatística (depende somente de $\boldsymbol{X}$).
  2. T_p é o estimador de Pitman (critério para obtenção de estimadores) para θ (θ parâmetro de localização) (pesquisar na literatura do curso).
  3. T_p é o estimador bayesiano esparença a posteriori (mais do que um critério para obtenção de estimadores) quando $f(\theta) \propto \mathbb{1}_{\mathcal{R}}(\theta)$ (veremos mais adiante).

Invariância por localização

- Demonstração: Temos (considere $-\gamma = c - \theta$) (lembre-se de que aqui página 4) que:

$$
\begin{aligned}
& t_p(x_1 + c, ..., x_n + c) \\
= \quad & \frac{\int_{-\infty}^{\infty} \theta \prod_{i=1}^{n} f_X(x_i + c; \theta) d\theta}{\int_{-\infty}^{\infty} \prod_{i=1}^{n} f_X(x_i + c; \theta) d\theta} = \frac{\int_{-\infty}^{\infty} \theta \prod_{i=1}^{n} g(x_i + c - \theta) d\theta}{\int_{-\infty}^{\infty} \prod_{i=1}^{n} g(x_i + c - \theta) d\theta} \\
= \quad & \frac{\int_{-\infty}^{\infty} (c + \gamma) \prod_{i=1}^{n} g(x_i - \gamma) d\gamma}{\int_{-\infty}^{\infty} \prod_{i=1}^{n} g(x_i - \gamma) d\gamma} = c + \frac{\int_{-\infty}^{\infty} \gamma \prod_{i=1}^{n} f_X(x_i; \gamma) d\gamma}{\int_{-\infty}^{\infty} \prod_{i=1}^{n} f_X(x_i; \gamma) d\gamma} \\
= \quad & c + t_p.
\end{aligned}
$$

Invariância por localização

- Exemplo Seja $X_1, .., X_n$ uma aa de $X \sim U(\theta - a, \theta + a)$, $\theta \in \mathcal{R}$, $a \in \mathcal{R}^+$ (conhecido). Temos que

$$
\begin{aligned}
T_p &= \frac{\int_{-\infty}^{\infty} \theta \frac{1}{(2a)^n} \prod_{i=1}^{n} \mathbb{1}_{(\theta-a,\theta+a)}(x_i) d\theta}{\int_{-\infty}^{\infty} \frac{1}{(2a)^n} \prod_{i=1}^{n} \mathbb{1}_{(\theta-a,\theta+a)}(x_i) d\theta} = \frac{\int_{-\infty}^{\infty} \theta \mathbb{1}_{(y_n-a,y_1+a)}(\theta) d\theta}{\int_{-\infty}^{\infty} \mathbb{1}_{(y_n-a,y_1+a)}(\theta) d\theta} \\
&= \frac{\frac{\theta^2}{2}\Big|_{y_n-a}^{y_1+a}}{\theta\Big|_{y_n-a}^{y_1+a}} = \frac{1}{2}\frac{(y_1+a)^2 - (y_n-a)^2}{y_1 - y_n + 2a} \\
&= \frac{1}{2}\frac{y_1^2 - y_n^2 + 2y_1 a + 2y_n a + a^2 - a^2}{y_1 - y_n + 2a} = \frac{1}{2}\frac{(y_1+y_n)(y_1 - y_n + 2a)}{y_1 - y_n + 2a} \\
&= \frac{y_1 + y_n}{2}.
\end{aligned}
$$

Temos que $T_p = \frac{Y_1 + Y_n}{2}$ é o estimador de Pitman para θ. Exercício: calcular $\mathcal{E}(T_p)$ e $\mathcal{V}(T_p)$.

Invariância por escala

- Def: Uma estatística $T = t(\boldsymbol{X})$ é invariante por escala, se $T^* = t(cX_1, .., cX_n) = ct(\boldsymbol{X})$, em que $c \in \mathcal{R}$ (constante).
- Exemplos:
  - $T_1 = \overline{X}$.
  - $T_2 = S_X = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$
  - $T_3 = Y_n - Y_1$.
  - $T_4 = \frac{S_X^2}{\overline{X}}$.
- Teorema: Seja $X_1, ..., X_n$ uma aa de X, com fdp $f_X(.;\theta)$, $\theta \in \mathcal{R}^+$, θ parâmetro de escala. Então

$$T_p = \frac{\int_0^{\infty} \frac{1}{\theta^2} f_{\boldsymbol{X}}(\boldsymbol{x};\theta)d\theta}{\int_0^{\infty} \frac{1}{\theta^3} f_{\boldsymbol{X}}(\boldsymbol{x};\theta)d\theta} = \frac{\int_0^{\infty} \frac{1}{\theta^2} \prod_{i=1}^{n} f_X(X_i;\theta)d\theta}{\int_0^{\infty} \frac{1}{\theta^3} \prod_{i=1}^{n} f_X(X_i;\theta)d\theta},$$

é uma estatística invariante por escala.

Invariância por escala

- Dem: Temos que:

$$t_p(\boldsymbol{x}) = \frac{\int_0^\infty \frac{1}{\theta^2}\prod_{i=1}^n f_X(x_i;\theta)d\theta}{\int_0^\infty \frac{1}{\theta^3}\prod_{i=1}^n f_X(x_i;\theta)d\theta}$$

- Assim, fazendo $(\gamma = \frac{\theta}{c} \to d\gamma = \frac{d\theta}{c})$, vem que:

$$\begin{aligned} t_p(cx_1,...,cx_n) &= \frac{\int_0^\infty \frac{1}{\theta^2}\prod_{i=1}^n f_X(cx_i;\theta)d\theta}{\int_0^\infty \frac{1}{\theta^3}\prod_{i=1}^n f_X(cx_i;\theta)d\theta} = \frac{\int_0^\infty \frac{1}{\theta^2}\prod_{i=1}^n \frac{1}{\theta} h\left(\frac{cx_i}{\theta}\right)d\theta}{\int_0^\infty \frac{1}{\theta^3}\prod_{i=1}^n \frac{1}{\theta} h\left(\frac{cx_i}{\theta}\right)d\theta} \\ &= c\frac{\int_0^\infty \frac{1}{\gamma^2}\prod_{i=1}^n \frac{1}{\gamma} h\left(\frac{x_i}{\gamma}\right)d\gamma}{\int_0^\infty \frac{1}{\gamma^3}\prod_{i=1}^n \frac{1}{\gamma} h\left(\frac{x_i}{\gamma}\right)d\gamma} = C\frac{\int_0^\infty \frac{1}{\theta^2}\prod_{i=1}^n f_X(x_i;\theta)d\theta}{\int_0^\infty \frac{1}{\theta^3}\prod_{i=1}^n f_X(x_i;\theta)d\theta} \\ &= ct_p(\boldsymbol{x}). \end{aligned}$$

Logo, $T_p = t_p(\boldsymbol{X})$ é uma estatística invariante por escala.

Invariância por escala

- Exemplo 2: Seja $X_1, ..., X_n$ uma aa de $X \sim U(0, \theta)$, $\theta \in \mathcal{R}^+$. Temos que:

$$
\begin{aligned}
T_p &= \frac{\int_0^\infty \frac{1}{\theta^2}\frac{1}{\theta^n}\mathbb{1}_{(y_n,\infty)}(\theta)\mathbb{1}_{(0,y_n)}(y_1)d\theta}{\int_0^\infty \frac{1}{\theta^3}\frac{1}{\theta^n}\mathbb{1}_{(y_n,\infty)}(\theta)\mathbb{1}_{(0,y_n)}(y_1)d\theta} = \frac{\int_{y_n}^\infty \frac{1}{\theta^{n+2}}d\theta}{\int_{y_n}^\infty \frac{1}{\theta^{n+3}}d\theta} = \frac{-\frac{\theta^{-n-1}}{n+1}\Big|_{y_n}^\infty}{-\frac{\theta^{-n-2}}{n+2}\Big|_{y_n}^\infty} \\
&= \frac{y_n^{-n-1}/(n+1)}{y_n^{-n-2}/(n+2)} = y_n\frac{n+2}{n+1}
\end{aligned}
$$

Assim, $Y_n\frac{n+2}{n+1}$ é uma estatística invariante por escala. Exercício: Calcular $\mathcal{E}(t_p(\boldsymbol{X}))$ e $\mathcal{V}(t_p(\boldsymbol{X}))$.

Obs: Se $\exists$ uma estatística suficiente, digamos $T = t(\boldsymbol{X})$, então $T_p = t_p(\boldsymbol{X})$ será função de T, em ambos os casos (localização e escala).

Prova: exercício.

Princípio da invariância

- Seja $\boldsymbol{X} = (X_1, ..., X_n)'$ uma aa de $X \sim f_X(., \theta)$, $\theta \in \Theta$ e seja $\boldsymbol{Y} = g(\boldsymbol{X})$ tem distribuição na mesma classe (por exemplo, ambas tem distribuição normal). Então dizemos que $\boldsymbol{X}$ é invariante pela transformação g, $g : \mathcal{X} \to \mathcal{X}$.
- Exemplo 1: Seja $X \sim N(\mu, \sigma^2)$, $Y = g(X) = aX + b$ $\sim N(a\mu + b, a^2\sigma^2)$. Então a família $X(f_X(.; \boldsymbol{\theta}) \sim N(\mu, \sigma^2))$ é invariante a transformação lineares.
- Exemplo 2: Seja $X \sim \text{binomial}(n, \theta)$ e $Y = n - X \sim \text{binomial}(n, 1 - \theta)$. Então a família $X(f_X(.; \theta) \sim \text{binomial}(n, \theta))$ é invariante à transformação $Y = n - X$.

Princípio da invariância

- Def: A classe G de transformações $g : \{g : \mathcal{X} \to \mathcal{X}\}$ é chamada de grupo de transformações de $\mathcal{X}$ se
  1. $\forall g \in G, \exists g' \in G, (g \circ g')(x) = g(g'(x)) = x$ inversa
  2. $\forall g, g' \in G, (g \circ g') \in G$ (fechada por composição).
  3. $\exists g_0 \in G, g_0(x) = x$ (identidade).
- Def: Seja $F = \{f_{\boldsymbol{X}}(;\theta), \theta \in \Theta\}$, o conjunto de distribuições de $\boldsymbol{X}$ e seja G um grupo de transformação sobre $\mathcal{X}$. Dizemos que F é invariante sob G, se $\forall \theta \in \Theta$ e $g \in G$, $Y = g(\boldsymbol{X}) \sim f_{\boldsymbol{Y}}(\boldsymbol{y}; \theta'), \theta' \in \Theta$, desde que $\boldsymbol{X} \sim f_{\boldsymbol{X}}(\boldsymbol{x}; \theta)$.

Princípio da invariância

- Exemplo: Seja $X_1, ..., X_n$ uma aa de $X, X \sim N(\mu, \sigma^2)$. Defina $G = \{g : \mathcal{X} \to \mathcal{X}\}$, $g(\boldsymbol{X}) = (aX_1 + b, ..., aX_n + b)$. Então G satisfaz as condições de grupo e $F = \{f_{\boldsymbol{X}}(;\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$, $f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta}) \sim N(\mu, \sigma^2)$ é invariante sob G.
- Note que:

$$f_{\boldsymbol{X}}(\boldsymbol{x};\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\} \mathbb{1}_{\mathcal{R}^n}(\boldsymbol{x})$$

Princípio da invariância

- É possível demostrar (pela fgm), que:

$$f_{\boldsymbol{Y}}(\boldsymbol{y};\boldsymbol{\theta}) = \frac{1}{(2\pi a^2\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2a^2\sigma^2}\sum_{i=1}^{n}(y_i - (a\mu + b))^2\right\} \mathbb{1}_{\mathcal{R}^n}(\boldsymbol{y})$$

- Exercício: Verificar que as condições 1), 2) e 3) são satisfeitas, neste caso.