Inferência Estatística I

Estimação Pontual

Prof. Paulo Cerqueira Jr
Faculdade de Estatística - FAEST
Instituto de Ciências Exatas e Naturais - ICEN

https://github.com/paulocerqueirajr

Modelos Estatísticos

Modelos Estatísticos

- Nosso ponto de partida será um estudo empírico (pode ser experimental ou observacional) que irá fornecer certo conjunto de dados (amostra) que denotamos por $\mathbf{x}$. Nos casos mais simples, $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$.
- **Suposição fundamental:** considere $\mathbf{x}$ como um valor obtido de uma vetor aleatório X.
- Nosso objetivo é usar $\mathbf{x}$ para tirar conclusões sobre a distribuição desconhecida $F(\cdot)$ de X.
- Nossas conclusões sobre $F(\cdot)$ estão sujeitas à incerteza dado a aleatoriedade governando X (que irá produzir $\mathbf{x}$). Devemos certificar que:
  - O nível de incerteza é o menor possível, considerando a aleatoriedade de X.
  - Somos capazes de avaliar o nível de incerteza em nossas conclusões.

Modelos Estatísticos

- A natureza física do fenômeno que gera $\mathbf{x}$, o esquema de amostragem, e outras informações, irão colocar limites no conjunto de possíveis escolhas para $F(\cdot)$. Este conjunto (denotado por $\mathcal{F}$) é chamado de modelo estatístico.
- É intuitivo pensar que nossa inferênias serão mais precisas de formos capazes de selecionar o conjunto $\mathcal{F}$ menor possível, sob o requerimento de que $F \in \mathcal{F}$.
- Em alguns casos, podemos assumir que X é uma a. a. com componentes independentes e identicamente distribuídos. Neste caso, dizemos que $\mathbf{x}$ é uma a.a. simples de X.

Modelos paramétricos

- A princípio, $\mathcal{F}$ pode ser qualquer conjunto de funções de distribuições, mas existe uma categoria de tais conjuntos que possui importante papel, tanto do ponto de vista teórico quando aplicado.
- Este caso ocorre todos os elementos de $\mathcal{F}$ são funções com a mesma formulação matemática, identificadas apenas pelas diferentes especificações de θ, que varia em $\Theta \in \mathbb{R}^k$,

$$\mathcal{F} = \left\{ F(\cdot \mid \theta) : \theta \in \Theta \subseteq \mathbb{R}^k \right\}$$

Modelos paramétricos

- Na grande maioria dos casos (em todos os casos que iremos considerar neste curso), toda a função de distribuição menbro de $\mathcal{F}$ refere-se a v.a. discretas ou contínuas.
- Então $\mathcal{F}$ pode ser definida usandoas f.p. ou f.d. correspondentes.
- Podemos definir um modelo estatístico $\mathcal{F}$ (caso contínuo) como um conjunto de f.d's

$$\mathcal{F} = \left\{ f(\cdot \mid \theta) : \theta \in \boldsymbol{\Theta} \subseteq \mathbb{R}^k \right\}$$

θ : parâmetro.

$\boldsymbol{\Theta}$: espaço paramétrico.

- $\mathcal{F}$, indicado acima, é chamado de classe paramétrica ou modelo paramétrico.

Modelos paramétricos

- Portanto, os elementos de $\mathcal{F}$ estão associados aos elementos de Θ.
- Em particular, existe um valor $\theta_* \in \Theta$, associado a $F(\cdot)$, chamado de **valor real** do parâmetro, e nossas inferências serão sobre θ_*

Modelos paramétricos

Espaço amostral: é o conjunto $\mathcal{X}$ de todos os possíveis resultados x compatíveis com o modelo paramétrico dado.

- Formalmente denotado por $\mathcal{X}_\theta$ o suporte (domínio) da densidade $f(\cdot;\theta)$, o espaço amostral é dado por $\mathcal{X} = \bigcup_{\theta \in \Theta} \mathcal{X}_\theta$.
- Frequentemente, entretanto, $\mathcal{X}_\theta$ é o mesmo que as possíveis escolhas de θ, e este conjunto coincidirá com $\mathcal{X}$.

Modelos paramétricos

`Exemplo:` Se dois valores são amostrados independentemente da $N(\theta, 1)$, então $\mathbf{y} = (y_1, y_2)^\top$ onde $y_i \in \mathbb{R}$ $(i = 1, 2)$,

$$\mathcal{Y} = \mathbb{R} \times \mathbb{R}, \quad Y \sim N\left[\begin{matrix}\theta \\ \theta\end{matrix}, I_2\right],$$

em que,

$$f(y; \theta) = \phi(y_1 - \theta)\phi(y_2 - \theta)$$

- Se não houver qualquer restrição para θ, temos $\Theta = \mathbb{R}$.
- Se existir restrição (ex. sabemos que $\theta > 0$)

Famílias de locação e escala

Famílias de locação e escala

- Aqui discutiremos três técnicas para construir famílias de distribuições.
- As famílias resultantes possuem interpretações físicas diretas que as tornam úteis para modelagem, além de apresentarem propriedades matemáticas convenientes. Considere apenas o caso contínuo.
- Os 3 tipos de famílias são: (i) locação, (ii) escala e (iii) locação e escala.
- Cada família é construída pela especificação de uma f.d $f(x)$ chamada de densidade padrão da família.
- Todas as outras densidades da família podem ser geradas pela transformação da densidade padrão.

Famílias de locação e escala

`Teorema 11:` Seja $f(x)$ qualquer f.d. e considere μ e $\sigma > 0$ como constantes conhecidas. Então, a função $g(x|\mu,\sigma) = \dfrac{1}{\sigma} f\left(\dfrac{x-\mu}{\sigma}\right)$ é uma f.d.

Prova: Para verificar que a transformação produziu um f.d. legítima, precisamos verificar que $\dfrac{1}{\sigma} f\left(\dfrac{x-\mu}{\sigma}\right)$ é: i) não negativa e ii) integra 1.

i. $f(x)$ é uma f.d.$\Rightarrow$ $f(x) > 0, \forall x$, então $\dfrac{1}{\sigma} f\left(\dfrac{x-\mu}{\sigma}\right) > 0$, para todos os valores de x, μ e σ.

Famílias de Locação e Escala.

ii.

$$\int_{-\infty}^{\infty} \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right) dx, \qquad y = \frac{x-\mu}{\sigma} \text{ e } dy = \frac{1}{\sigma}.$$

Logo,

$$\int_{-\infty}^{\infty} \frac{1}{\sigma} f(y)\, \sigma dy = \int_{-\infty}^{\infty} f(y)\, dy = 1.$$

Famílias de Locação e Escala.

Definição 8: Seja $f(x)$ qualquer f.d., então a família de densidades $f(x-\mu)$ indexada pelo parâmetro real μ é chamada de **família de locação (localização)** com densidade padrão $f(x)$. O μ é conhecido como **parâmetro de localização da família**.

- O parâmetro μ simplesmente desloca a densidade $f(x)$ de maneira que o formato do gráfico não é alterado, mas o ponto do gráfico de $f(x)$ que estava acima de $x=0$, estará agora acima de $x=\mu$ para $f(x-\mu)$.

Exemplo: Se $\sigma > 0$ é especificado e definimos

$$f(x) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} I_{(-\infty,\infty)}(x),$$

então a família de localização com densidade padrão $f(x)$ é o conjunto de distribuições Normais com média μ desconhecida e variância σ^2 conhecida.

Famílias de Locação e Escala.

- A família Cauchy com σ (conhecido) e μ (desconhecido) é outro exemplo de família de locação.
- O ponto principal da `Definição 8` é que podemos iniciar com qualquer densidade $f(x)$ e gerar uma família de densidades com a introdução do parâmetro de localização.
- Se X é uma variável aleatória com densidade $f(x-\mu)$, então X pode ser representada como $X = Z + \mu$, onde Z é variável aleatória com densidade $f(z)$.

`Exemplo:` Família de locação exponencial

Seja $f(x) = e^{-x}$ para todo $x \geq 0$ e $f(x) = 0$ para $x < 0$.

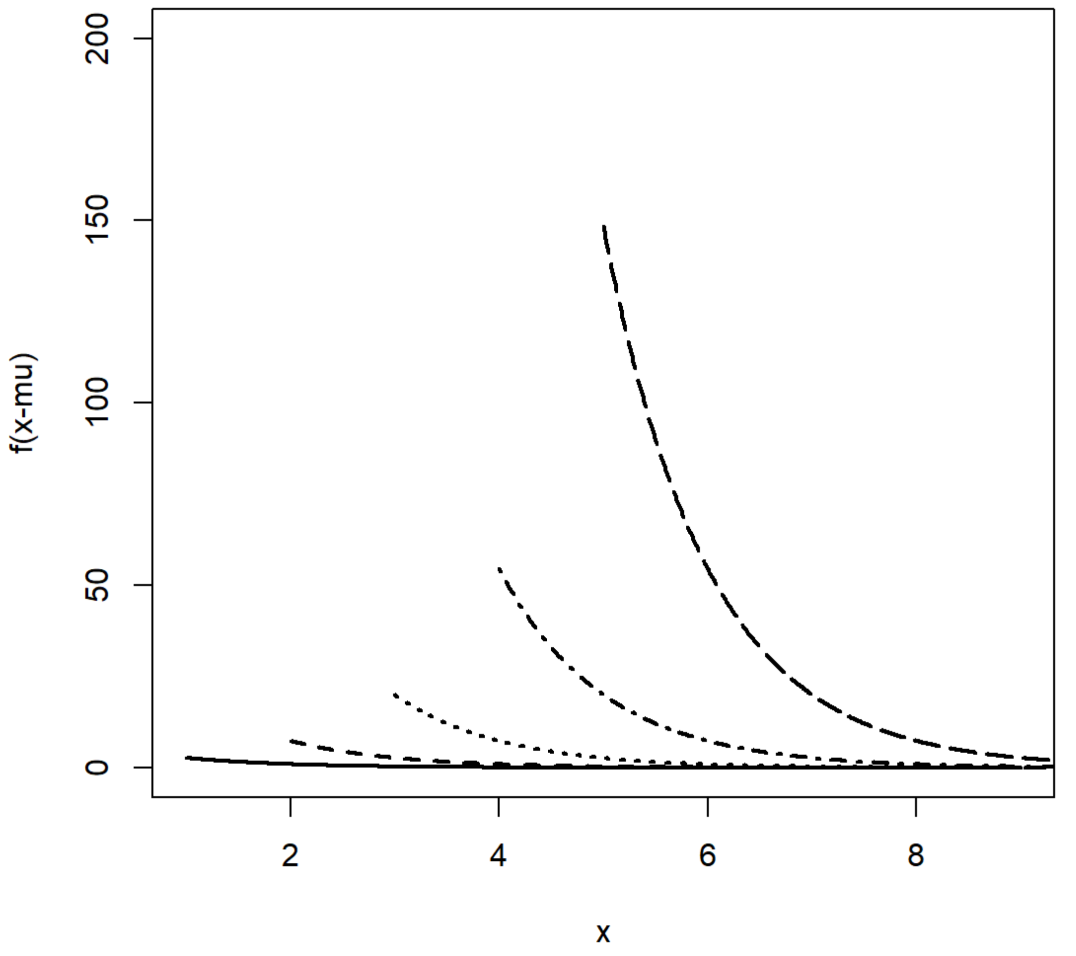
Para formar uma família de locação devemos substituir x po $x - \mu$

$$f(x) = \begin{cases} e^{-(x-\mu)} & x-\mu \geq 0 \\ 0 & x-\mu < 0 \end{cases} = \begin{cases} e^{-(x-\mu)} & x \geq \mu \\ 0 & x < \mu \end{cases}$$

Famílias de Locação e Escala.

```r
f <- function(x, mu){
  return(exp(-(x-mu)))}
mu.val <- c(1,2,3,4,5)
x.val <- seq(0,8, length.out=1000)

plot(x.val+mu.val[1], f(x=x.val,mu=1), type="l", ylim=c(0,200), ylab="f(x-mu)",
     xlab="x", lty=1, lwd=2)
lines(x.val+mu.val[2], f(x=x.val,mu=2), type="l", lty=2, lwd=2)
lines(x.val+mu.val[3], f(x=x.val,mu=3), type="l", lty=3, lwd=2)
lines(x.val+mu.val[4], f(x=x.val,mu=4), type="l", lty=4, lwd=2)
lines(x.val+mu.val[5], f(x=x.val,mu=5), type="l", lty=5, lwd=2)
```



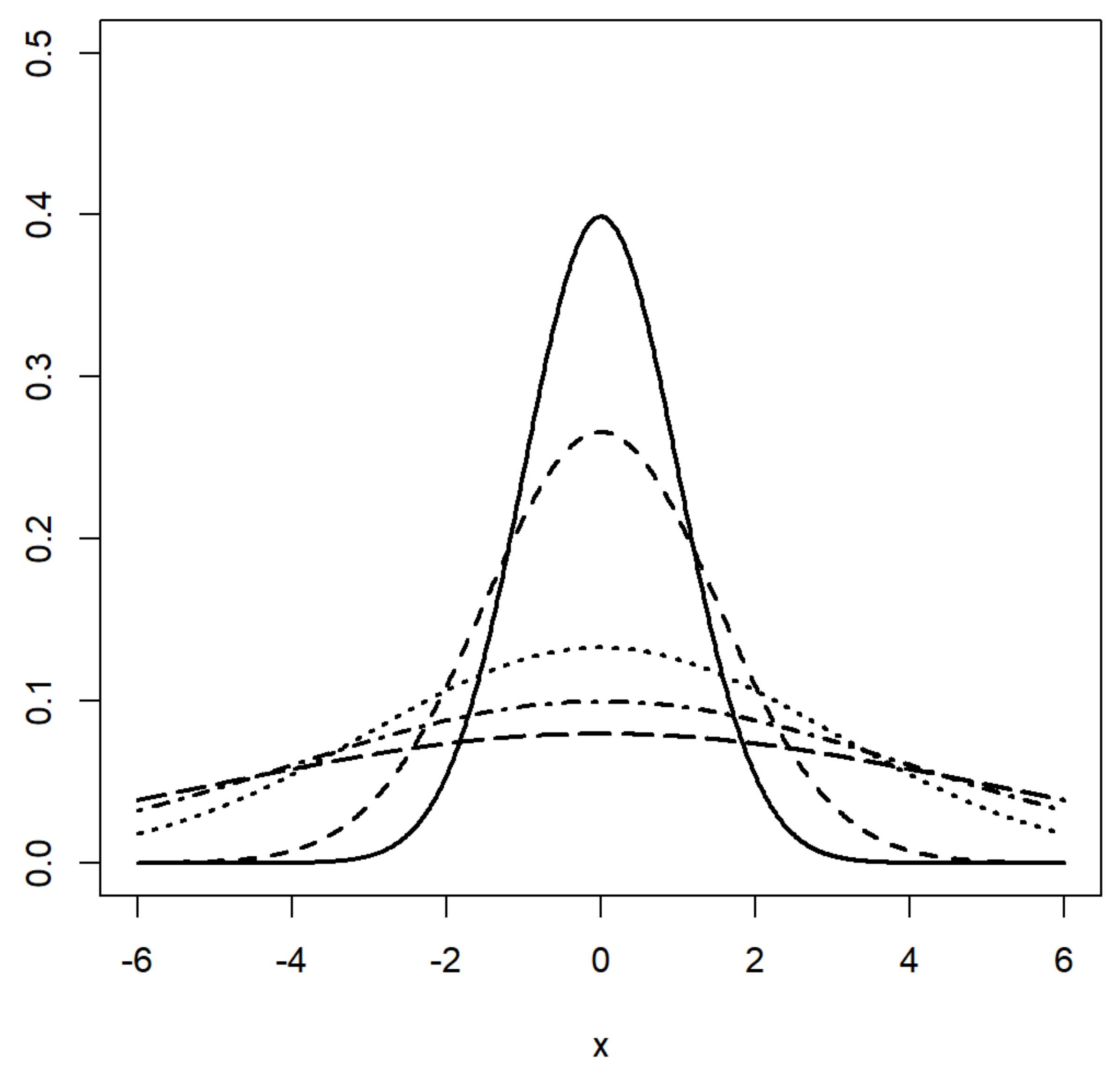
Famílias de Locação e Escala

`Definição 9:` Seja $f(x)$ qualquer f.d., então para qualquer $\sigma > 0$, a família de densidades $1/\sigma f[x/\sigma]$ indexada pelo parâmetro σ é chamada de **família de escala** com densidade padrão $f(x)$ e parâmetro de escala σ.

- O efeito de introduzir σ é tanto esticar ($\sigma > 1$) quanto contrair ($\sigma < 1$) o gráfico $f(x)$ a forma básica é mantida.

Famílias de Locação e Escala

```
sigma.val <- c(1,1.5,3,4,5)
x.val <- seq(-6,6, length.out=1000)
plot(x.val, dnorm(x=x.val,sd=sigma.val[1]), type="l", ylim=c(0,.5), ylab="", xlab='
     lty=1, lwd=2)
lines(x.val, dnorm(x=x.val,sd=sigma.val[2]), lty=2, lwd=2)
lines(x.val, dnorm(x=x.val,sd=sigma.val[3]), lty=3, lwd=2)
lines(x.val, dnorm(x=x.val,sd=sigma.val[4]), lty=4, lwd=2)
lines(x.val, dnorm(x=x.val,sd=sigma.val[5]), lty=5, lwd=2)
```



Famílias de Locação e Escala

`Exemplo:` $Ga\left(\alpha, \beta = \frac{1}{\sigma}\right)$ com α conhecido e $\beta = \frac{1}{\sigma}$ onde σ é desconhecido.

A densidade padrão $Ga(\alpha, \beta = 1)$

$$f(x|\alpha, \beta = 1) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-x\} I_{(-\infty,\infty)}(x).$$

Logo,

$$f(x/\sigma) = \frac{1}{\Gamma(\alpha)} \frac{x^{\alpha-1}}{\sigma^{\alpha-1}} \exp\left\{-\frac{x}{\sigma}\right\} I_{(-\infty,\infty)}(x)$$

$$1/\sigma f(x/\sigma) = \frac{1}{\sigma^{\alpha}} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} \exp\left\{-\frac{x}{\sigma}\right\} I_{(-\infty,\infty)}(x)$$

Famílias de Locação e Escala

`Exemplo:` Família Normal com $\mu = 0$ e σ^2 desconhecido.

A densidade padrão $N(0,1)$

$$f(x) = 1 * (2\pi)^{-1/2} \exp\left\{-\frac{1}{2(1)}x^2\right\} I_{(-\infty,\infty)}(x),$$

Logo,

$$\begin{aligned} f(x/\sigma) = &\quad (2\pi)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} I_{(-\infty,\infty)}(x) \\ 1/\sigma f(x/\sigma) = &\quad (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} I_{(-\infty,\infty)}(x) \end{aligned}$$

Famílias de Locação e Escala

Definição 10: Seja $f(x)$ qualquer f.d., então para qualquer μ real e qualquer $\sigma > 0$ a família de densidades

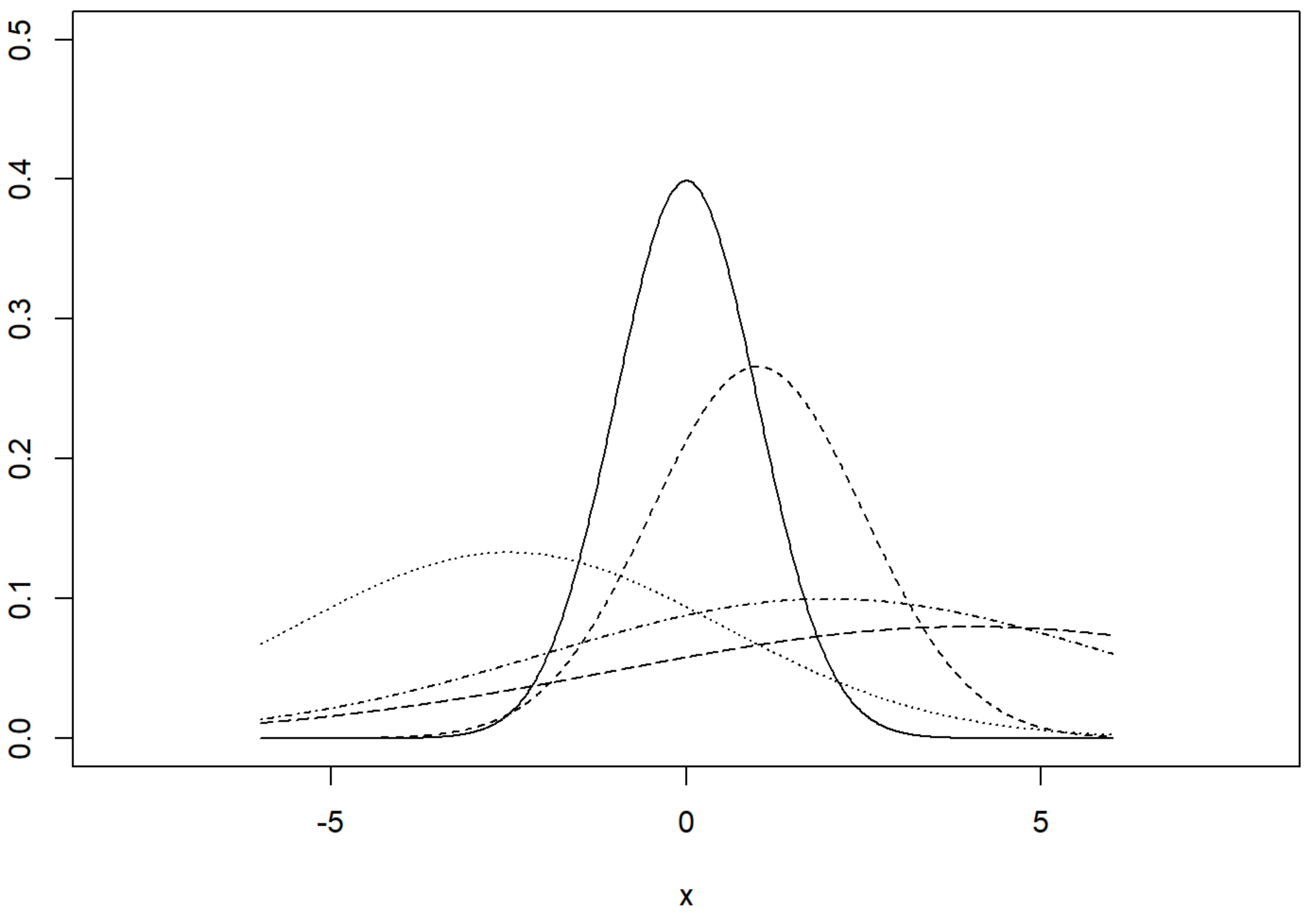
$$\frac{1}{\sigma} f\left[\frac{(x-\mu)}{\sigma}\right],$$

indexadas pelos parâmetros (μ, σ), é chamada de família de locação e escala com densidade padrão $f(x)$. Neste caso, μ é o parâmetro de localização e σ é o parâmetro de escala.

- Efeito da inclusão dos parâmetros:
  - μ irá deslocar o gráfico de maneira que o ponto que estava acima de 0, agora fica acima de μ.
  - σ irá esticar ($\sigma > 1$) ou contrair ($\sigma < 1$) o gráfico de $f(x)$.

Famílias de Locação e Escala

```r
sigma.val <- c(1,1.5,3,4,5)
x.val <- seq(-6,6, length.out=1000)
plot(x.val, dnorm(x=x.val, mean = 0,sd=sigma.val[1]), type="l", xlim=c(-8,8),
     ylim=c(0,.5), ylab="", xlab="x", lty=1, lwd=1)
lines(x.val, dnorm(x=x.val, mean = 1,sd=sigma.val[2]), lty=2, lwd=1)
lines(x.val, dnorm(x=x.val, mean = -2.5,sd=sigma.val[3]), lty=3, lwd=1)
lines(x.val, dnorm(x=x.val, mean = 2,sd=sigma.val[4]), lty=4, lwd=1)
lines(x.val, dnorm(x=x.val, mean = 4,sd=sigma.val[5]), lty=5, lwd=1)
```



Famílias de Locação e Escala

- O seguinte teorema relaciona a transformação da f.d. $f(x)$, que define uma família de locação e escala, com a transformação da variável aleatória Z com densidade $f(z)$.

`Teorema 12:` Seja $f(\cdot)$ qualquer f.d. e considere $\mu \in \mathbb{R}$ e $\sigma \in \mathbb{R}^+$. Então X é uma v.a. com densidade $1/\sigma f[(x-\mu)/\sigma]$, se e somente se, existe uma v.a. Z com densidade $f(z)$ e $X = \sigma Z + \mu$.

- No `Teorema 12`:
  - Se $\sigma = 1$: família de locação (apenas).
  - Se $\mu = 0$: família de escala (apenas).

Famílias de Locação e Escala

- Fato importante a ser extraído do `Teorema 12` é que $Z = \dfrac{X - \mu}{\sigma}$, tem f.d.

$$f_Z(z) = \frac{1}{1} f\left(\frac{z - 0}{1}\right) = f(z),$$

isto é, a distribuição de Z é membro da família de locação escala com $\mu = 0$ e $\sigma = 1$.

- Frequentemente, cálculos são desenvolvidos para a v.a. padrão Z com f.d $f(z)$ e então o resultado correspondente para a v.a. X com f.d. $1/\sigma f[(x - \mu)/\sigma]$ pode ser facilmente derivado.

Famílias de Locação e Escala

Teorema 13: Seja Z uma v.a. com f.d. $f(z)$. Suponha que $E(Z)$ e $Var(Z)$ existem. Se X é uma v.a. com densidade $1/\sigma f(x/\sigma)$, então,

$$E(X) = \sigma E(Z) + \mu \text{ e } Var(X) = \sigma^2 Var(Z)$$

.

- Em particular, se $E(Z) = 0$ e $Var(Z) = 1$, então, $E(X) = \mu$ e $Var(X) = \sigma^2$.
- Probabilidades para qualquer membro da família de locação escala pode ser calculada em termos da variável padrão Z.

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

Estimação Pontual

Introdução

- Na população temos uma característica que denotamos por X;
- E dentro da inferência paramétrica associamos uma $f(\cdot \mid \theta)$ (função densidade ou probabilidade);
- θ: é uma quantidade fixa e desconhecida;

> O nosso maior interesse consiste em estimar o valor de θ.

Definições

Definição 1 O conjunto Θ em que θ toma valores é denominado Espaço paramétrico.

Exemplo 1 Seja $(X_1, X_2, \ldots, X_n)$ uma a.a. de $X \sim N(\mu, \sigma^2)$.

1. se $\sigma^2 = 1$, então $\theta = \mu$ é o parâmetro desconhecido.

$$\Theta = \{\mu : -\infty < \mu < \infty\} = \mathbb{R}.$$

2. se $\mu = 0$, então $\theta = \sigma^2$ é o parâmetro desconhecido.

$$\Theta = \left\{\sigma^2 : \sigma^2 > 0\right\} = \mathbb{R}^+.$$

3. se σ^2 e μ são desconhecidos, então $\theta = (\mu, \sigma^2)$ é o vetor de parâmetros desconhecido.

$$\Theta = \left\{(\mu, \sigma^2) : -\infty < \mu < \infty \quad \text{e} \quad \sigma^2 > 0\right\}.$$

Definições importantes

Definição 2 Estimador para θ: $\hat{\theta}$.

- Qualquer estatística qual assuma valores em Θ é um estimador para θ.

Definição 3 Estimativas para θ: Usando $\hat{\theta}$.

As estimativas são dos valores obtidos pelos estimadores $\hat{\theta}$

Definição 4 Estimador para $g(\theta)$.

- Qualquer estatística que assuma valores apenas no conjunto dos possíveis valores de $g(\Theta)$ é um estimador para $g(\theta)$.

Exemplos de estimadores:

Estimador para a média amostral:

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

Estimador para a variância:

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

Estimador para a uma função $g(\theta)$:

$$g(\hat{\theta}) = \frac{\hat{\theta}}{1 - \hat{\theta}}.$$

Estimadores

- Como saber se nosso estimador é um `bom estimador`?
- Precisamos de ferramentas/propriedades matemáticas para avaliar a qualidade do mesmo.
- Algumas são:
  - Viés do estimador;
  - Variância do estimador;
  - Erro quadrático médio;
  - Consistência;
  - Eficiência.

Erro quadrático médio

Erro quadrático médio

Definição 5 (Erro Quadrático Médio (EQM)) O erro quadrático médio (EQM) de um estimador $\hat{\theta}$ do parâmetro θ é dado por

$$EQM(\hat{\theta}) = E\left[\left(\hat{\theta} - \theta\right)^2\right].$$

- Podemos mostrar que,

$$EQM(\hat{\theta}) = Var(\hat{\theta}) + B(\hat{\theta})^2,$$

em que $B(\hat{\theta}) = E(\hat{\theta}) - \theta$.

Erro quadrático médio

- Algumas observações:

1. $B(\hat{\theta})$ é denominado de vício do estimador $\hat{\theta}$;
2. Dizemos que um estimador $\hat{\theta}$ é não viciado para θ se $E(\hat{\theta}) = \theta$, para todo $\theta \in \Theta$, ou seja, se $B(\hat{\theta}) = 0$, para todo $\theta \in \Theta$;
3. Se $\lim_{n\to\infty} B(\hat{\theta}) = 0$, para todo $\theta \in \Theta$, dizemos que $\hat{\theta}$ é assintoticamente não viciado para θ.
4. Se $\hat{\theta}$ é um estimador não viciado para θ, temos que $EQM(\hat{\theta}) = Var(\hat{\theta})$.

Erro Quadrático médio

Exemplo 2 Seja $(X_1, X_2, \ldots, X_n)$ uma a.a. de um população com variável X, com $E(X) = \mu$ e $Var(X) = \sigma^2$.

1. Tome $\hat{\theta} = \bar{X}$, mostre que é um estimador não viciado para μ.
2. Tome $\hat{\theta} = \hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$, mostre que é um estimador viciado σ^2.
3. Tome $\hat{\theta} = S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, mostre que é um estimador não viciado σ^2.

Comparação dos estimadores via EQM

1. $\hat{\theta}_1$ é melhor que $\hat{\theta}_2$ se $EQM(\hat{\theta}_1) \leq EQM(\hat{\theta}_2)$ para todo θ, com "$\leq$" substituído por "$<$" pelo menos para um valor de θ. Nesse caso, $\hat{\theta}_2$ é `dito ser inadmissível`.
2. Se existir um estimador $\hat{\theta}^*$ tal que para todo estimador $\hat{\theta}$ de θ, com $\hat{\theta} \neq \hat{\theta}^*$, o $EQM(\hat{\theta}^*) \leq EQM(\hat{\theta})$ para todo μ, com "$\leq$" substituído por "$<$" para pelo menos um valor de θ, então $\hat{\theta}^*$ é `dito ser ótimo` para θ.
3. Além disso, se em "2)" os estimadores são não viciados, então $\hat{\theta}^*$ é dito ser o `estimador não viciado de variância uniformemente mínima (ENVVUM)`, pois teremos

$$Var\left(\hat{\theta}^*\right) \leq Var\left(\hat{\theta}\right)$$

para todo θ, com "$\leq$" substituído por "$<$" para pelo menos um valor de θ.

Erro Quadrático Médio

Exemplo 3 Seja (X_1, X_2, X_3) uma a.a. de um população com variável X, com $E(X) = \theta$ e $Var(X) = 1$. Considere os estimadores:

$$\hat{\theta}_1 = \bar{X} = \frac{X_1 + X_2 + X_3}{n} \quad \text{e} \quad \hat{\theta}_2 = \frac{X_1}{2} + \frac{X_2}{4} + \frac{X_3}{4}.$$

Em termos do EQM qual o melhor estimador?

Estimador linear

Exemplo 4 Seja $(X_1, X_2, \ldots, X_n)$ uma a.a. de um população com variável X, com $E(X) = \theta$ e $Var(X) = \sigma^2$ (conhecido). Considere os estimadores lineares:

$$X_L = \sum_{i=1}^{n} l_i X_i,$$

com $l_i \geq 0, i = 1, \ldots, n$ constantes conhecidas. Que valores devem ter os l_i para que X_L seja um `estimador não viciado de variância mínima` para θ?

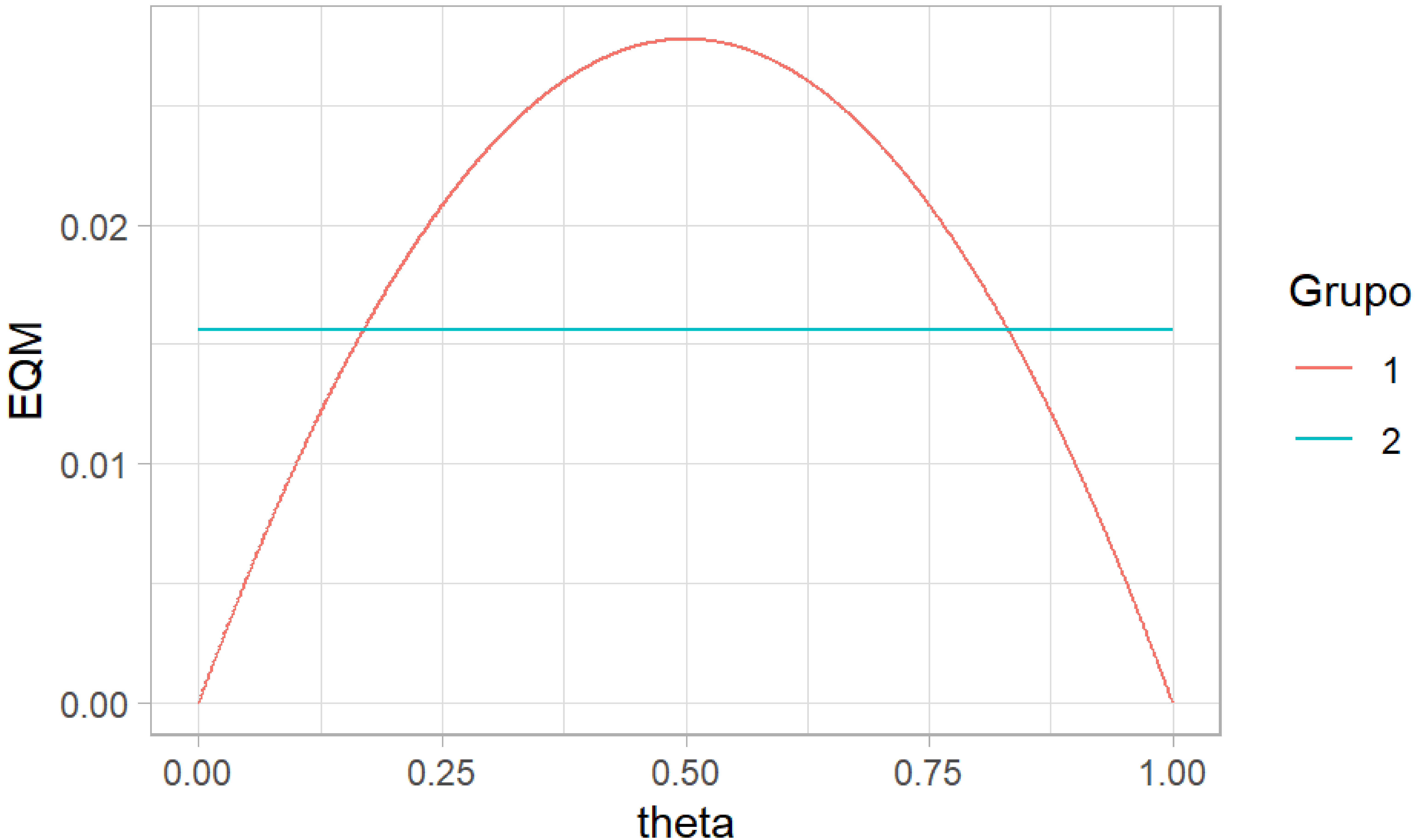
Exemplo

Exemplo 5 Seja $(X_1, X_2, \ldots, X_n)$ uma a.a. de um população com variável $X \sim Ber(\theta)$, com $E(X) = \theta$ e $Var(X) = \sigma^2$ (conhecido). Considere os estimadores:

$$\hat{\theta}_1 = \bar{X} = \frac{Y}{n} \quad \text{e} \quad \hat{\theta}_2 = \frac{Y + \frac{\sqrt{n}}{2}}{n + \sqrt{n}},$$

com $Y = \sum_{i=1}^{n} X_i$. Obtenha $EQM(\hat{\theta}_1)$ e o $EQM(\hat{\theta}_2)$.

Exemplo



Consistência

Consistência

Definição 6 Uma sequência $\{\hat{\theta}_n, n = 1, 2 \ldots, \}$ de estimadores de um parâmetro θ é consistente se, para todos $\epsilon > 0$,

$$P\left(\left|\hat{\theta}_n - \theta\right| > \epsilon\right) \to 0, \quad \text{quando} \quad n \to \infty.$$

De forma equivalente, $\{\hat{\theta}_n, n = 1, 2 \ldots, \}$ é uma sequência consistente de estimadores de θ se

1. $\lim\limits_{n \to \infty} E(\hat{\theta}) = \theta$;
2. $\lim\limits_{n \to \infty} Var(\hat{\theta}) = 0$.

Consistência - Exemplos:

$\bar{X}_n$: estimador consistente da média populacional

1. $\lim_{n \to \infty} E(\bar{X}_n) = \lim_{n \to \infty} \mu = \mu$
2. $\lim_{n \to \infty} Var(\bar{X}_n) = \lim_{n \to \infty} \sigma^2/n = 0$

$\hat{p}_n$: estimador consistente da proporção populacional.

1. $\lim_{n \to \infty} E(\hat{p}_n) = \lim_{n \to \infty} \theta = \theta$
2. $\lim_{n \to \infty} Var(\hat{p}_n) = \lim_{n \to \infty} \theta(1 - \theta)/n = 0$

Consistencia

Exemplo

Exemplo 6 Deseja-se estimar a proporção de moradores de um determinado bairro favoráveis a um projeto municipal. Para isso, coleta-se uma amostra de n moradores e registra-se sua opinião. Seja S o no total de moradores favoráveis na amostra, e considere os dois estimadores:

$$\hat{p}_1 = \frac{S}{n} \quad \text{e} \quad \hat{p}_2 = \begin{cases} 1, & \text{se o prim. indivíduo da amostra é favorável} \\ 0, & \text{caso contrário} \end{cases}$$

Compare os dois estimadores. Por que $\hat{p}_2$ não é um bom estimador?

Exemplo

Exemplo 7 Foram sorteadas 15 famílias com filhos num certo bairro e observado o número de crianças de cada família, matriculadas na escola. Os dados foram: 1, 1, 2, 0, 2, 0, 2, 3, 4, 1, 1, 2, 0, 0 e 2. Obtenha as estimativas correspondentes aos seguintes estimadores do n^o médio de crianças na escola nesse bairro:

$$\hat{\theta}_1 = \frac{X_{(1)} + X_{(n)}}{2}, \quad \hat{\theta}_1 = \frac{X_1 + X_2}{2} \quad \hat{\theta}_3 = \bar{X}.$$

Obtenha as 3 estimativas e discuta qual delas é a melhor.

Estimador eficiente

Introdução

- Aprenderemos agora a noção de estimador eficiente.
- Tal estimador como sendo aquele que atinge o `limite inferior da variância dos estimadores não viciados`.
- Estimadores eficientes são obtidos apenas para distribuições que são membros de uma classe especial, que é a `família exponencial de distribuições`.

Estimadores eficientes

Definição 7 Chamamos de eficiência de um estimador $\hat{\theta}$, não viciado para o parâmetro θ, o quociente

$$e(\hat{\theta}) = \frac{LI(\theta)}{Var(\hat{\theta})},$$

onde $LI(\theta)$ é o limite inferior da variância dos estimadores não viciados de θ.

Estimadores eficientes

Notamos que:

1. $e(\hat{\theta}) = 1$, quando $LI(\theta) = Var(\hat{\theta})$ ou seja, quando a variância de $\hat{\theta}$ coincide com o limite inferior da variância dos estimadores não viciados de θ. Nesse caso, $\hat{\theta}$ é dito ser ;
2. Temos que,

$$LI(\theta) = \frac{1}{nE\left[\left(\frac{\partial \log f(X \mid \theta)}{\partial \theta}\right)^2\right]}$$

   quando certas condições de regularidade estão satisfeitas;
3. As condições de regularidade: que o suporte $A(x) = x,\, f(x|\theta) > 0$ seja independente de $\hat{\theta}$; possível a troca das ordens das operações de derivação e de integração sob a distribuição da variável aleatória X;
4. a não ser que mencionado o contrário, todo logaritmo utilizado no texto é calculado na base e.

Exemplo

Exemplo 8 Sejam $(X_1, \ldots, X_n)$ uma amostra aleatória da variável aleatória $X \sim N(\mu, \sigma^2)$, em que σ^2 é conhecido. Verifique se $\hat{\theta} = \bar{X}$ é um estimador eficiente para μ.

Estimadores eficientes

Definição 8 A quantidade

$$\frac{\partial \log f(X \mid \theta)}{\partial \theta},$$

é chamada de função escore.

- Como resultado temos,

$$E\left[\frac{\partial \log f(X \mid \theta)}{\partial \theta}\right] = 0.$$

Definição 9 A quantidade

$$I_F = E\left[\left(\frac{\partial \log f(X \mid \theta)}{\partial \theta}\right)^2\right],$$

é denominada de informação de Fisher de θ.

Estimadores eficientes

- Como resultado temos que,

$$I_F = Var\left[\frac{\partial \log f(X \mid \theta)}{\partial \theta}\right],$$

uma vez que para uma variável aleatória X qualquer com $E[X] = 0, Var[X] = E[X^2]$.

- Outro resultado importante:

$$E\left[\left(\frac{\partial \log f(X \mid \theta)}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \log f(X \mid \theta)}{\partial \theta^2}\right]$$

Estimadores eficientes

- Uma outra propriedade importante estabelece que para uma amostra aleatória, $X_1, \ldots, X_n$, da variável aleatória X com f.d.p (ou f.p.) $f(x \mid \theta)$ e informação de Fisher $I_F(\theta)$;
- A informação total de Fisher de θ correspondente á amostra observada é a soma da informação de Fisher das n observações da amostra, ou seja, sendo

$$E\left[\left(\frac{\partial \log L(\theta \mid \mathbf{X})}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \log L(\theta \mid \mathbf{X})}{\partial \theta^2}\right] = -E\left[\sum_{i=1}^{n} \frac{\partial^2 \log f(X_i \mid \theta)}{\partial \theta^2}\right] =$$

$$= \sum_{i=1}^{n} E\left[-\frac{\partial^2 \log f(X_i \mid \theta)}{\partial \theta^2}\right] = nI_F(\theta).$$

Estimadores eficientes

Teorema 1 (Desigualdade da Informação ou Desigualdade de Cramér-Rao) Quando as condições de regularidade estão satisfeitas, a variância de qualquer estimador não viciado $\hat{\theta}$ do parâmetro θ satisfaz

$$Var(\hat{\theta}) \geq \frac{1}{nI_F(\theta)}.$$

Exemplo

Exemplo 9 Sejam $(X_1, \ldots, X_n)$ uma a.a. da v.a. $X \sim Poisson(\theta)$. Obtenha um estimador eficiente para θ.

Estatísticas suficientes

Estatísticas suficientes

`Estatísticas:` resumir a informação trazida pelos dados `sem perda de informação`.

Definição 10 (Estatísticas suficientes) Dizemos que a estatística $T = T(X_1, X_2, \ldots, X_n)$ é suficiente para θ quando a distribuição condicional de $X_1, X_2, \ldots, X_n$, dado T, for independente de θ.

> **Nota**
>
> Uma estatística é suficiente para θ se ela condensa toda a informação sobre θ contida na amostra.

Exemplo

Exemplo 10 Sejam $(X_1, \ldots, X_n)$ uma a.a. da v.a. $X \sim Ber(\theta)$. Verifique se $T = \sum_{i=1}^{n} X_i$ suficiente para θ.

Estatística suficientes

Um procedimento para a obtenção de estatísticas suficientes é o `critério da fatoração`.

Definição 11 (Critério da Fatoração de Neyman) Sejam $X_1, \ldots, X_n$ uma amostra aleatória da distribuição da variável aleatória X com função de densidade (ou de probabilidade) $f(x|\theta)$ e função de verossimilhança $L(\theta \mid \mathbf{x})$. Temos, então, que a estatística $T = T(X_1, \ldots, X_n)$ é suficiente para θ, se e somente se pudermos escrever

$$L(\theta \mid \mathbf{x}) = h(x_1, \ldots, x_n) g_\theta \left[T(x_1, \ldots, x_n)\right],$$

onde $h(x_1, \ldots, x_n)$ é uma função que depende de $x_1, \ldots, x_n$ e $g_\theta \left[T(x_1, \ldots, x_n)\right]$ uma função que depende de θ e $x_1, \ldots, x_n$ somente através de T.

Exemplo

Exemplo 11 Sejam $(X_1, \ldots, X_n)$ uma a.a. da v.a. $X \sim Pois(\theta)$. Determine a estatística suficiente para θ

Exemplo

Exemplo 12 Sejam $(X_1, \ldots, X_n)$ uma a.a. da v.a. $X \sim U(0, \theta)$. Determine a estatística suficiente para θ

Estatísticas conjuntamente suficientes e Família exponencial

Introdução

- Saímos de um cenário em que θ é uniparamétrico;
- Agora podemos ter situações em que $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ representando um vetor de parâmetros;
- Chamamos de caso `multiparamétrico`;
- Além disso podemos verificar se as distribuições de probabilidades pertencem a família de exponencial de distribuições;
- E as vantagens consistem em avaliar as propriedades já vistas.

Estatística conjuntamente suficientes

Teorema 2 (Critério da Fatoração para o Caso Multiparamétrico) Sejam $X_1, \ldots, X_n$ uma amostra aleatória da distribuição da variável aleatória X com função de verossimilhança $L(\theta \mid \mathbf{x})$. Temos, então, que a estatística $\mathbf{T} = (T_1, \ldots, T_r), T_i = T_i(X_1, \ldots, X_n)$ é suficiente para θ, se e somente se pudermos escrever

$$L(\theta \mid \mathbf{x}) = h(x_1, \ldots, x_n) g_\theta \left[T_1(\mathbf{x}), \ldots, T_r(\mathbf{x})\right],$$

onde $h(x_1, \ldots, x_n)$ é uma função que depende de $x_1, \ldots, x_n$ e $g_\theta \left[T_1(\mathbf{x}), \ldots, T_r(\mathbf{x}))\right]$ uma função que depende de θ e $x_1, \ldots, x_n$ somente através de $\mathbf{T}$.

Exemplo

Exemplo 13 Seja $X_1, \dots, X_n$ uma a.a. de $X \sim N(\mu, \sigma^2)$, onde μ e σ^2 são desconhecidos. Obtenha uma estatística conjuntamente suficiente para $\theta = (\mu, \sigma^2)$.

Estatísticas conjuntamente suficientes

Observação:

Duas estatísticas T_1 e T_2 são equivalentes se T_1 puder ser obtida a partir de T_2 e vice-versa. Nesse caso, se T_1 é suficiente para θ, então T_2 também é suficiente para θ.

No exemplo anterior, podemos verificar que $\mathbf{T}_1 = (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ e $\mathbf{T}_2 = (\bar{X}, S^2)$ são equivalentes.

Exemplo

Exemplo 14 Seja $X_1, \ldots, X_n$ uma a.a. de $X \sim Gama(\alpha, \gamma)$, onde α e γ são desconhecidos. Obtenha uma estatística conjuntamente suficiente para $\theta = (\alpha, \gamma)$.

Estatísticas suficientes mínimas

Definição 12 (Estatística Suficiente Mínima) Uma estatística conjuntamente suficiente é definida ser suficiente mínima se e somente se ela é uma função de todas as outras estatísticas suficientes.

> **Importante**
>
> A definição acima é de pouco uso prático para se obter estatísticas suficientes mínimas. Porém, se a densidade conjunta da amostra é corretamente fatorada, o critério da fatoração fornecerá estatísticas suficientes mínimas.

Família exponencial

Definição 13 Dizemos que a distribuição da v.a. X pertence à família exponencial unidimensional se pudermos escrever sua f.p. ou sua f.d.p. como

$$f(x \mid \theta) = \exp\{c(\theta)T(x) + d(\theta) + S(x)\}, \quad x \in A,$$

em que c e d são funções reais de θ, T e S são funções reais de x e A não depende de θ.

Exemplo 15 Verifique se $X \sim Ber(\theta)$, pertence à família exponencial.

Estatística conjuntamente suficientes

> Amostras aleatórias de famílias exponenciais são também membros da família exponencial

Teorema 3 Sejam $(X_1, \ldots, X_n)$ uma a.a. da v.a. X, que pertence à família exponencial unidimensional, ou seja,

$$f(x \mid \theta) = \exp\{c(\theta)T(x) + d(\theta) + S(x)\}, \quad x \in A.$$

Então, a função de verossimilhança é dada por,

$$L(\theta \mid \mathbf{x}) = f(x_1, \ldots, x_n \mid \theta) = \exp\left\{c^*(\theta)\sum_{i=1}^{n} T(x_i) + d^*(\theta) + S^*(\mathbf{x})\right\}$$

que também pertence à família exponencial com

$$T(\mathbf{x}) = \sum_{i=1}^{n} T(x_i), \;\; c^*(\theta) = c(\theta), \;\; d^*(\theta) = nd(\theta), \;\; S^*(\mathbf{x}) = \sum_{i=1}^{n} S(x_i)$$

Família exponencial e o critério da fatoração

Observe que pelo critério da fatoração de Neyman, temos,

$$L(\theta \mid \mathbf{x}) = h(x_1, \ldots, x_n) g_\theta \left[T_1(\mathbf{x}), \ldots, T_r(\mathbf{x})\right],$$

em que

$$h(x_1, \ldots, x_n) = \exp\{S^*(\mathbf{x})\} = \exp\left\{\sum_{i=1}^{n} S(x_i)\right\}$$

e

$$g_\theta \left[T_1(\mathbf{x}), \ldots, T_r(\mathbf{x})\right] = \exp\{c^*(\theta) T(\mathbf{x}) + d^*(\theta)\}$$

Portanto, pelo Critério da Fatoração, a estatística $T(\mathbf{x})$ é suficiente para θ.

Exemplo

Exemplo 16 Seja $X_1, \ldots, X_n$ uma a.a. de $X \sim Ber(\theta)$. Obtenha uma estatística suficiente para θ.

Família exponencial

Definição 14 Dizemos que a distribuição da v.a. X pertence à família exponencial de dimensão k se pudermos escrever sua f.p. ou sua f.d.p. como

$$f(x \mid \theta) = \exp\left\{ \sum_{j=1}^{k} c_j(\theta) T_j(x) + d(\theta) + S(x) \right\}, \quad x \in A,$$

em que $c_j,\ T_j,\ d$ e S são funções reais de $\theta, j = 1, \ldots k$ e A não depende de θ.

Família exponencial

> Amostras de famílias exponenciais de dimensão k são também membros da família exponencial de dimensão k.

Para uma a.a. $X_1, \ldots, X_n$ de uma v.a. com f.d.p (ou f.p) na família exponencial multiparamétrica, temos que $\left(T_1^*(\mathbf{x}), T_2^*(\mathbf{x}), \ldots, T_k^*(\mathbf{x})\right)$ é conjuntamente suficiente para θ, com

$$T_j^*(\mathbf{x}) = \sum_{i=1}^{n} T_j^*(x_i)$$

Exemplo

Exemplo 17 Seja $X_1, \ldots, X_n$ uma a.a. de $X \sim N(\mu, \sigma^2)$. Verifique se X pertence à família exponencial bidimensional.

Resumo

População	Estatística Suficiente
$X \sim Ber(\theta)$	$T = \sum_{i=1}^{n} X_i$ é suficiente para θ
$X \sim U(0, \theta)$	$T = X_{(n)} = \max(X_1, \ldots, X_n)$ é suficiente para θ
$X \sim N(\mu, 1)$	$T = \sum_{i=1}^{n} X_i$ é suficiente para μ
$X \sim N(0, \sigma^2)$	$T = \sum_{i=1}^{n} X_i^2$ é suficiente para σ^2
$X \sim N(\mu, \sigma^2)$	$T = (\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2)$ é conjuntamente suficiente para $\theta = (\mu, \sigma^2)$
$X \sim Pois(\theta)$	$T = \sum_{i=1}^{n} X_i$ é suficiente para θ
$X \sim Exp(\theta)$	$T = \sum_{i=1}^{n} X_i$ é suficiente para θ
$X \sim Gama(\alpha, \gamma)$	$T = (\prod_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i)$ é conjuntamente suficiente para $\theta = (\alpha, \gamma)$
$X \sim Beta(a, b)$	$T = (\prod_{i=1}^{n} X_i, \prod_{i=1}^{n}(1 - X_i))$ é conjuntamente suficiente para $\theta = (a, b)$

Teorema de Rao-Blackwell

Teorema de Rao-Blackwell

Teorema 4 Sejam $X_1, \ldots, X_n$ uma amostra aleatória da distribuição da variável aleatória X com função de densidade $f(x \mid \theta)$. Temos, então, $T = T(X_1, \ldots, X_n)$ uma estatística suficiente para θ, e S um estimador não viciado para θ que não é função de T. Seja $\hat{\theta} = E(S \mid T)$, então:

1. $\hat{\theta}$ é uma estatística e é função da estatística suficiente T;
2. $E(\hat{\theta}) = \theta$;
3. $Var(\hat{\theta}) \leq Var(S)$, para todo θ, e $Var(\hat{\theta}) < Var(S)$ para pelo menos um valor de θ, a menos que $S = \hat{\theta}$, com probabilidade um.

`Interpretação:` Dado um estimador não-viciado, podemos obter outro estimador não-viciado que é função de uma estatística suficiente, e ele `terá variância menor`.

Teorema de Rao-Blackwell

Demonstração:

ii. $E(\hat{\theta}) = E(E(S \mid T)) = E(S) = \theta$ (Estimador não viciado!!)

iii. $Var(S) = E(Var(S \mid T)) + Var(E(S \mid T)) = \underbrace{E(Var(S \mid T))}_{\geq 0} + Var(\hat{\theta})$.

Portanto, $Var(\hat{\theta}) \leq Var(S)$.

Exemplo

Exemplo 18 Sejam $X_1, \ldots, X_n$ uma a.a. da v.a. $X \sim Ber(\theta)$. Vamos obter um estimador para θ que seja função de uma estatística suficiente.

Exemplo

Exemplo 19 Sejam $X_1, \ldots, X_n$ uma a.a. da v.a. $X \sim Pois(\theta)$. Vamos obter um estimador para $P(X = 0) = \exp\{-\theta\}$, que seja função de uma estatística suficiente.

Teorema de Rao-Blackwell

- Através do Teorema de Rao-Blackwell conseguimos melhorar um estimador não-viciado.
- Qual a relação entre $\hat{\theta} = E(S \mid T)$ e o `ENVVUM`?

Estatísticas completas

Definição 15 Uma estatística $T = T(X_1, \ldots, X_n)$ é dita ser completa em relalção à família $f(x \mid \theta), \theta \in \Theta$, se a única função real g, definida no domínio de T, tal que $E(g(T)) = 0$ para todo θ é a função nula, isto é, $g(T) = 0$ com probabilidade um.

- T é completa se, e somente se, $E(g(T)) = 0, \theta \in \Theta$, implicar que

$$P(g(T) = 0) = 1, \ \theta \in \Theta$$

.

Exemplo

Exemplo 20 Sejam $X_1, \dots, X_n$ uma a.a. da v.a. $X \sim Ber(\theta)$. Considere as estatísticas $T_1 = \sum_{i=1}^{n} X_i$ e $T_2 = X_1 - X_2$. Verifique se T_1 e T_2 são estatísticas completas.

Estatísticas completas pela família exponencial

Teorema 5 Suponha que X tenha distribuição na família exponencial $k-$dimensional. Então, a estatística

$$T(\mathbf{X}) = \left(\sum_{i=1}^{n} T_1(X_i), \ldots, \sum_{i=1}^{n} T_k(X_i) \right),$$

é suficiente para θ. $T(\mathbf{X})$ será também completa desde que o domínio de variação de $(c_1(\theta), \ldots, c_k(\theta))$ contenha um retângulo $k-$dimensional.

Lehmann-Scheffé

Teorema 6 Seja T uma estatística suficiente e completa, e seja S um estimador não-viciado para θ. Então, $\hat{\theta} = E(S \mid T)$ é o único estimador não-viciado para θ, baseado em T, e é o `Estimador não-viciado de variância uniformemente mínima (ENVVUM)` para θ.

Exemplo

Exemplo 21 Sejam $X_1, \ldots, X_n$ uma a.a. da v.a. $X \sim Pois(\theta)$. Obtenha um ENVVUM para θ.