

Estatística Matemática

Verossimilhança



Prof. Paulo Cerqueira Jr - cerqueirajr@ufpa.br

Faculdade de Estatística - FAEST

Programa de Pós-Graduação em Matemática e Estatística - PPGME

Instituto de Ciências Exatas e Naturais - ICEN

<https://github.com/paulocerqueirajr> 

Modelos estadísticos

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{y} . Nos casos mais simples, $\mathbf{y} = (y_1, y_2, \dots, y_n)'$.

Suposição fundamental: considere \mathbf{y} como um valor obtido de uma vetor aleatório Y .

- Nosso objetivo é usar \mathbf{y} para tirar conclusões sobre a distribuição desconhecida $F(\cdot)$ de Y .
- Nossas conclusões sobre $F(\cdot)$ estão sujeitas à incerteza dado a aleatoriedade governando Y (que irá produzir \mathbf{y}). Devemos certificar que:
 - O nível de incerteza é o menor possível, considerando a aleatoriedade de Y .
 - Somos capazes de avaliar o nível de incerteza em nossas conclusões.

Modelos Estatísticos

- A natureza física do fenômeno que gera \mathbf{y} , 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 nossas inferências serão mais precisas de formas capazes de selecionar o conjunto \mathcal{F} menor possível, sob o requerimento de que $F \in \mathcal{F}$.
- Em alguns casos, podemos assumir que Y é uma a. a. com componentes independentes e identicamente distribuídos. Neste caso, dizemos que \mathbf{y} é uma a.a. simples de Y .

Modelos paramétricos

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 $\Theta \in \mathbb{R}^k$,

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

Modelos paramétricos

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

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

θ : parâmetro.

Θ : espaço paramétrico.

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

Definição 1 (Espaço amostral) É o conjunto \mathcal{Y} de todos os possíveis resultados y compatíveis com o modelo paramétrico dado.

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

Exemplo 1 Assuma que $Y \sim \text{Binomial}(n, \theta)$. Se $\theta \in [0, 1]$, \mathcal{Y}_θ será o mesmo para todo θ . \mathcal{Y}_θ coincidirá com o espaço amostral $\mathcal{Y} = \{0, 1, 2, \dots, n\}$.

- Se $\theta \in [0, 1]$, então $\mathcal{Y}_{\theta=0} = \{0\}$, $\mathcal{Y}_{\theta=1} = \{n\}$ e $\mathcal{Y}_{\theta \in (0,1)} = \{0, 1, \dots, n\}$.
- Neste caso, $\mathcal{Y} = \bigcup_{\theta \in [0,1]} \mathcal{Y}_\theta = \{0, 1, \dots, n\}$.

Modelos paramétricos

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

$$\mathcal{Y} = \mathbb{R} \times \mathbb{R}, \quad Y \sim N \begin{bmatrix} \theta \\ \theta \end{bmatrix}, I_2,$$

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. Iscala 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 1 Seja $f(x)$ qualquer f.d. e considere μ e $\sigma > 0$ como constantes conhecidas.

Então, a função $g(x|\mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right)$ é uma f.d.

Prova: Para verificar que a transformação produziu um f.d. legítima, precisamos verificar que

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

se é:

1. não negativa;

2. integra 1.

- Logo, em relação a 1., tem-se

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

Famílias de Locação e Escala.

- Referente a 2.

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

Logo,

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

Famílias de Locação e Escala.

Definição 2 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 3 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 Localização e Escala.

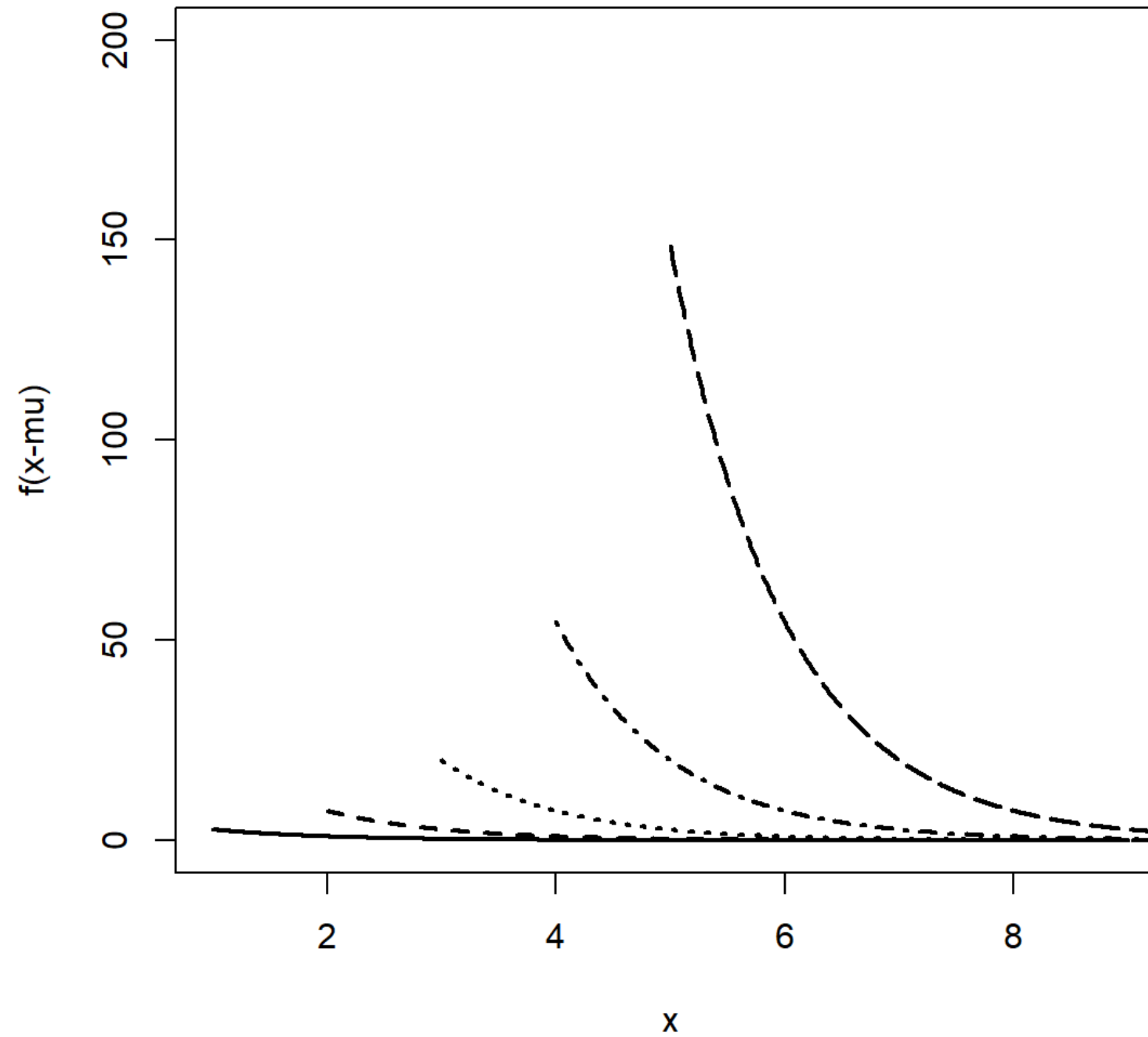
- A família Cauchy com σ (conhecido) e μ (desconhecido) é outro exemplo de família de localização.
- O ponto principal da [Definição 2](#) é que podemos iniciar com qualquer densidade $f(x)$ e gerar uma família de densidades com a introdução do parâmetro de localização.
- Se X é uma variável aleatória com densidade $f(x - \mu)$, então X pode ser representada como $X = Z + \mu$, onde Z é variável aleatória com densidade $f(z)$.

Exemplo 4 (Família de localização exponencial) Seja $f(x) = e^{-x}$ para todo $x \geq 0$ e $f(x) = 0$ para $x < 0$.

Para formar uma família de localização devemos substituir x por $x - \mu$

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

Famílias de Localização e Escala.

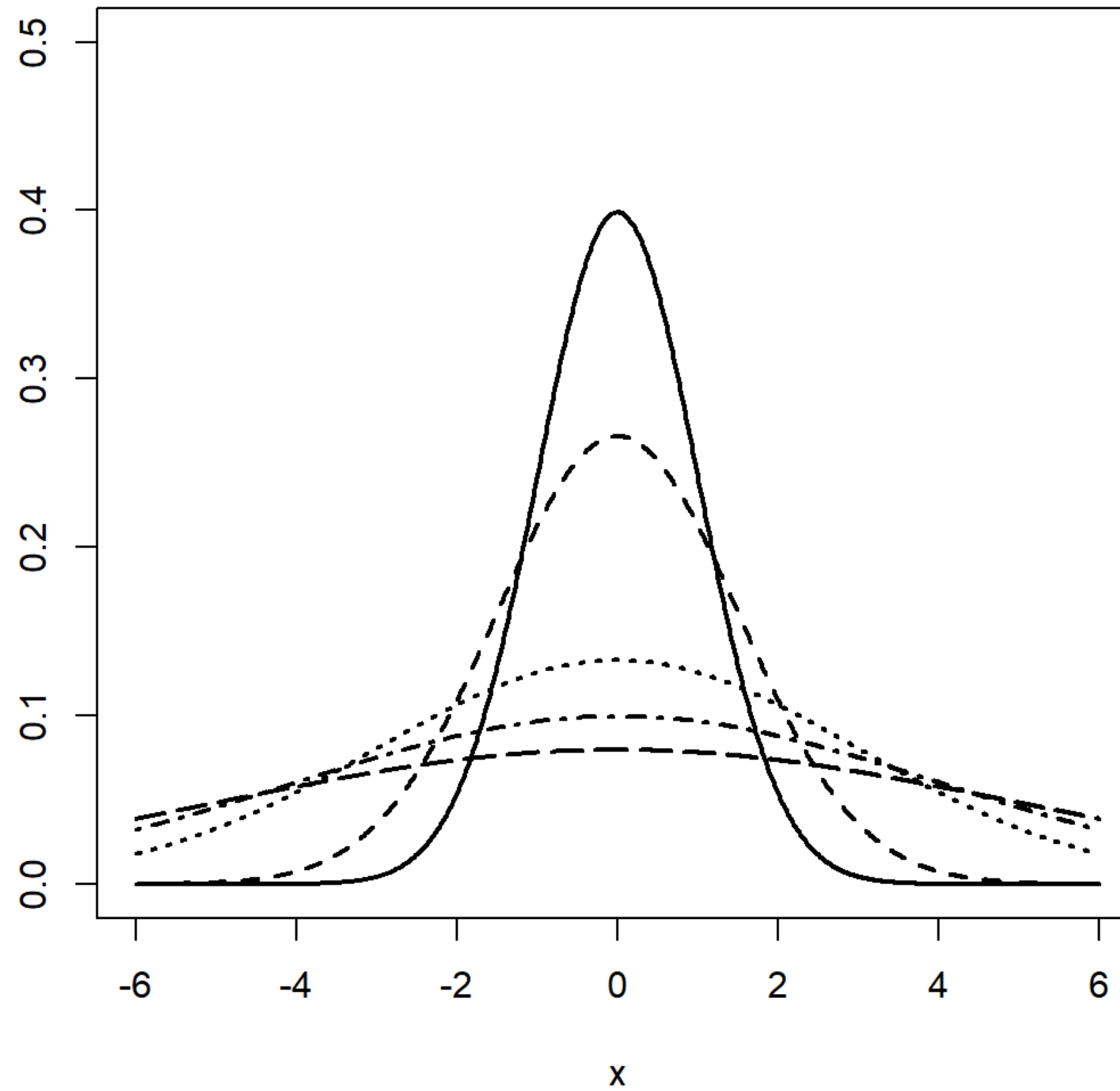


Famílias de Localização e Escala

Definição 3 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 Localização e Escala



Famílias de Locação e Escala

Exemplo 5 $Ga\left(\alpha, \beta = \frac{1}{\sigma}\right)$ 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 Localização e Escala

Exemplo 6 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,

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

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

Famílias de Localização e Escala

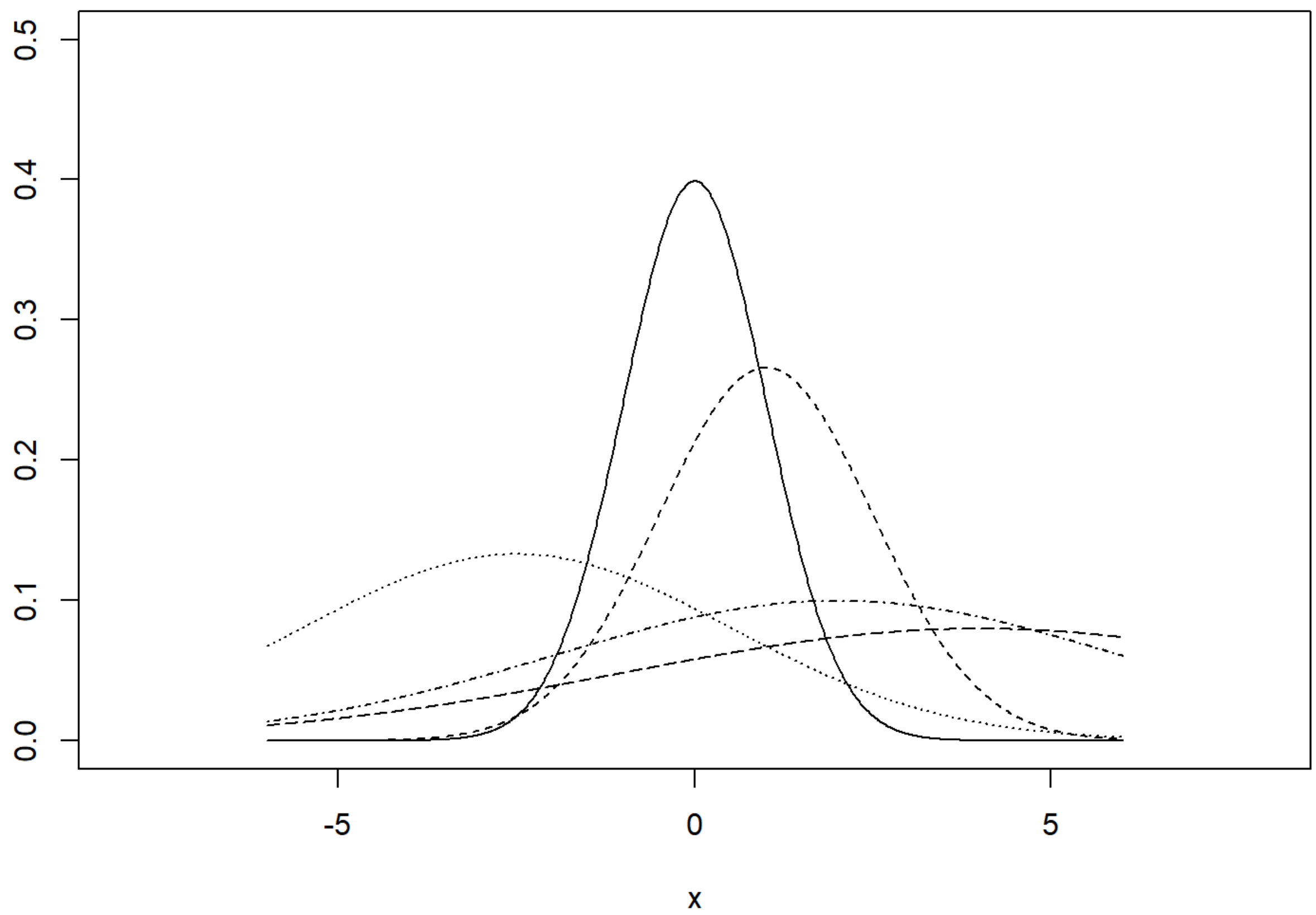
Definição 4 Seja $f(x)$ qualquer f.d., então para qualquer μ real e qualquer $\sigma > 0$ a família de densidades

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

indexadas pelos parâmetros (μ, σ) , é chamada de família de localização 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 Localização e Escala



Famílias de Localização e Escala

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

Teorema 2 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 2**:
 - Se $\sigma = 1$: família de localização (apenas).
 - Se $\mu = 0$: família de escala (apenas).

Famílias de Localização e Escala

- Fato importante a ser extraído do Teorema 12 é que $Z = \frac{X - \mu}{\sigma}$, 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 localização 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 Localização e Escala

Teorema 3 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 localização escala pode ser calculada em termos da variável padrão Z .

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

A função de verossimilhança

A função de verossimilhança

- Considere o dado modelo do tipo: $\mathcal{F} = \{f(\cdot \mid \theta) : \theta \in \Theta \subseteq \mathbb{R}^k\}$.
- Quando um valor amostral y é observado, o valor da f.d $f(y; \theta)$ **dependerá apenas** de θ .
- Esta função nos parece a f.d para observar aquilo que de fato observamos (y).
- Se precisamos estabelecer um *ranking* envolvendo dois elementos de θ (considere θ' e θ''), então uma quantidade relevante e útil para esta tarefa será a razão $f(y; \theta') / f(y; \theta'')$, desde que o denominador não seja zero.
- Como esta razão não muda caso ambos os termos sejam multiplicados por uma constante positiva C , independentemente de θ , então para comparar os elementos de Θ a quantidade relevante será proporcional a $f(y; \theta)$.

A função de verossimilhança

Definição 5 Para o modelo $\mathcal{F} = \{f(\cdot \mid \theta) : \theta \in \Theta \subseteq \mathbb{R}^k\}$ a partir do qual uma amostra $y \in \mathcal{Y}$ foi observada, usamos o termo **função de verossimilhança**, ou simplesmente **verossimilhança**, para a função de Θ para $\mathbb{R}^+ \cup \{0\}$ escrita como

$$L(\theta; y) = c(y)f(y; \theta),$$

em que $c(y)$: constante positiva independente de θ .

- A verossimilhança é uma função de θ .
- A notação $L(\theta; y)$ é usada para enfatizar que esta função depende de y , no sentido de que para uma amostra diferente y' obteremos uma verossimilhança diferente $L(\theta; y')$.

A função de verossimilhança

- Note que não importa se escrevemos c ou $c(y)$ na definição de verossimilhança, uma vez que a verossimilhança é uma função de θ .
- Mesmo que todo valor $L(\theta; y)$ seja determinado por distribuições de probabilidades, a função de verossimilhança **não é uma distribuição de probabilidade**.
- $L(\theta; y)$ é uma quantidade não negativa, e na maioria dos casos é positiva para todo θ . Sendo assim, definimos a função de log-verossimilhança como:

$$\ell(\theta; y) = \ln L(\theta; y) = c + \ln f(y; \theta),$$

com a convenção de que $\ell(\theta; y) = -\infty$ se $L(\theta; y) = 0$.

A função de verossimilhança

Exemplo 7 Considere uma a.a. $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ da v.a. $N(\mu, \sigma^2)$, onde $\theta = (\mu, \sigma^2)$ varia no espaço $\Theta = \mathbb{R} \times \mathbb{R}^+$.

- Devido à independência das componentes, temos,

$$\begin{aligned} L(\theta; \mathbf{y}) &= c \prod_{i=1}^n (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} (y_i - \mu)^2\right\} \\ &= c (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right\} \\ &= c (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i + n\mu^2\right]\right\} \end{aligned}$$

- Qualquer constante não dependente de θ , por exemplo, $c = 1$.
- Função de log-verossimilhança

$$\ell(\theta; \mathbf{y}) = c - \frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \left[\sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i + n\mu^2 \right]$$

A função de verossimilhança

Exemplo 8 Considere uma a.a. $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ da v.a. $U(0, \theta)$, onde $\theta > 0$. A f.d. associada a x_i é $f(x_i) = \frac{1}{\theta}$, se $x_i \in (0, \theta)$.

- Quando multiplicamos tais f.d.'s para obter $L(\theta; x)$, não podemos simplesmente multiplicar o termo $1/\theta$ sem considerar a condição **se** $x_i \in (0, \theta)$.
- Portanto, escrevemos a densidade para uma única observação, como segue,

$$\frac{1}{\theta} I_{(0, \theta)}(x), \quad x \in \mathbb{R}.$$

A função de verossimilhança

Função de verossimilhança

$$\begin{aligned} L(\theta; \mathbf{x}) &= c \prod_{i=1}^n \frac{1}{\theta} I_{(0,\theta)}(x_i) \\ &= \frac{c}{\theta^n} \prod_{i=1}^n I_{(0,\theta)}(x_i) \end{aligned}$$

- Note que $\prod_{i=1}^n I_{(0,\theta)}(x_i) = 1$, se

$$\begin{array}{l} 0 < x_1 < \theta \\ 0 < x_2 < \theta \\ \vdots \\ 0 < x_n < \theta \end{array} \implies x_{(n)} < \theta$$

A função de verossimilhança

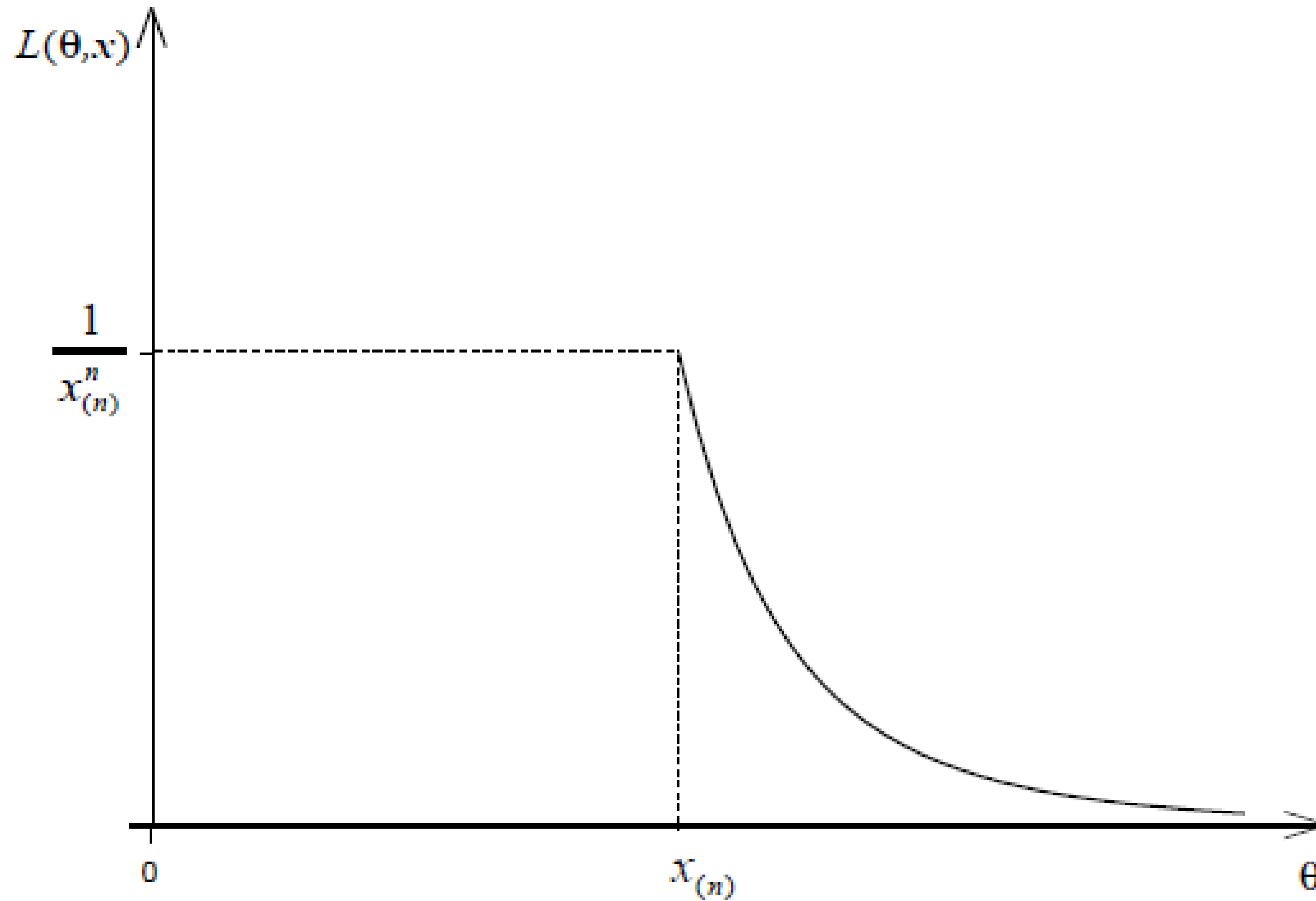
Assim,

$$\begin{aligned} L(\theta; \mathbf{y}) &= c \prod_{i=1}^n \frac{1}{\theta} I_{(0, \theta)}(x_i) \\ &= \frac{c}{\theta^n} \prod_{i=1}^n I_{(x_{(n)}, \infty)}(\theta) \end{aligned}$$

A log-verossimilhança:

$$\ell(\theta; \mathbf{y}) = \begin{cases} \infty, & \theta \leq x_{(n)} \\ c - n \ln \theta, & \theta > x_{(n)} \end{cases}$$

A função de verossimilhança



Princípio da verossimilhança.

- A função de verossimilhança conecta a informação pré-experimental (expressa pela escolha do modelo) com a informação experimental contida em y .
- Portanto, de certa forma, a verossimilhança contém tudo que sabemos sobre o problema de inferência em questão (sem levar em conta qualquer informação sobre θ que, por qualquer razão, não foi acomodada no modelo, tais como opiniões pessoais ou resultado de estudos relacionados).

Exemplo 9 O princípio da verossimilhança. Para um modelo estatístico $\{f(\cdot \mid \theta) : \theta \in \Theta\}$ dois pontos $y, z \in \mathcal{Y}$ tal que $L(\theta; y) \propto L(\theta; z)$ devem levar às mesmas conclusões inferenciais.

- Esta afirmação representa a versão mais fraca do princípio da verossimilhança.
- A seguir apresentamos uma versão mais forte que diz que as conclusões coincidem mesmo quando as duas observações se referem a modelos distintos e espaços amostrais distintos.

Princípio da verossimilhança.

Definição 6 Considere um experimento que consiste em lançar várias vezes uma moeda, de forma independente, e seja θ a probabilidade de ocorrer coroa \bar{C} e $1 - \theta$ a probabilidade de ocorrer cara C .

Suponha que o resultado do experimento foi:

$$\mathbf{x} = \{\bar{C}, \bar{C}, \bar{C}, \bar{C}, \bar{C}, C, \bar{C}, \bar{C}, \bar{C}, \bar{C}, C, C\}$$

Regras de parada:

1. Lançar a moeda 12 vezes (n^o de lançamentos fixado);
2. Lançar a moeda até aparecer 3 caras;
3. Lançar a moeda até aparecerem 2 caras consecutivas;
4. Lançar a moeda até o lançador ficar cansado.

Princípio da verossimilhança.

Em qualquer situação a verossimilhança é proporcional a

$$\theta^9 (1 - \theta)^3,$$

- Segundo o princípio da verossimilhança as inferências **devem ser a mesmas qualquer que tenha sido o processo experimental** (ou a regra de parada).

Estatísticas Suficientes

Estatísticas Suficientes

- Em uma descrição mais simplificada da teoria estatística, alguns poderiam dizer que seu objetivo é selecionar as **operações** mais apropriadas para serem aplicadas aos dados.
- Visto que uma variedade destas operações serão consideradas, é necessário introduzir a seguinte definição:

Definição 7 Uma função $T(\cdot) : y \rightarrow \mathbb{R}^R$, para algum inteiro positivo r , tal que $T(y)$ não depende de θ , é chamada de **Estatística**, e o valor $t = T(y)$ correspondendo ao valor observado y é chamado de valor amostral.

- A condição de que $T(y)$ não dependa de θ é necessária para assegurar que a estatística seja calculável na presença dos dados.

Estatísticas Suficientes

Exemplo 10 Estatísticas para uma amostra (y_1, y_2, \dots, y_n) cujos elementos pertencem a \mathbb{R} :

$$T_1 = \sum_{i=1}^n y_i, \quad T_1(\cdot) : (y_1, y_2, \dots, y_n) \rightarrow \mathbb{R}$$

$$T_2 = \sum_{i=1}^n \exp\{y_i\}, \quad T_2(\cdot) : (y_1, y_2, \dots, y_n) \rightarrow \mathbb{R}^+$$

$$T_3 = \left(\sum_{i=1}^n y_i, \sum_{i=1}^n \exp\{y_i\} \right), \quad T_3(\cdot) : (y_1, y_2, \dots, y_n) \rightarrow \mathbb{R} \times \mathbb{R}^+$$

Obviamente, estes são apenas 3 exemplos entre inúmeros casos.

Estatísticas suficientes

- Algumas vezes devemos considerar a imagem inversa dos valores de t de uma estatística T , isto é, conjuntos do tipo:

$$A_t = \{y : y \in \mathcal{Y}; T(y) = t\},$$

formam uma partição do espaço amostral. Fazemos referência a esta partição induzida por $T(y)$.

- Por exemplo, se $T = \sum_{i=1}^n y_i$, os conjuntos $\{A_t\}$ serão hiperplanos paralelos uns aos outros.
- Um conjunto específico A_t é dado por todos os pontos $\mathbf{y} = (y_1, y_2, \dots, y_n)' \in \mathbb{R}$ satisfazendo a equação $y_1 + y_2 + \dots + y_n = t$.

Estatísticas suficientes

Exemplo 11 Entre os tipos de estatísticas que iremos considerar, alguns são usados com frequência e a eles são dados nomes específicos. Para a amostra $(y_1, y_2, \dots, y_n)'$ o r -ésimo momento amostral é a estatística de \mathcal{Y} para \mathbb{R} dada por

$$m_r = \frac{1}{n} \sum_{i=1}^n y_i^r.$$

- Em particular, $m_1 = \frac{1}{n} \sum_{i=1}^n y_i^1$ é a média amostral \bar{y} .
- Outro exemplo de estatística é o que definimos como variância amostral:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2,$$

que assume valores em $\mathbb{R}^+ \cup \{0\}$.

Estatísticas Suficientes

- No caso da distribuição $N(\mu, \sigma^2)$, temos a verossimilhança indicada abaixo para uma amostra aleatória simples $(y_1, y_2, \dots, y_n)' = \mathbf{y}$.
- Considere $\theta = (\mu, \sigma^2)$,

$$L(\theta; \mathbf{y}) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i + n\mu^2 \right] \right\}$$

Estatísticas Suficientes

- Veja que não é preciso conhecer todos os elementos individuais de (y_1, y_2, \dots, y_n) para escrever $L(\theta; \mathbf{y})$ do *slide* anterior, dado a quantidade $\left(\sum_{i=1}^n y_i, \sum_{i=1}^n y_i^2 \right)$.
- Essa função de verossimilhança é unicamente identificada, entre todas as possíveis funções de verossimilhança para o modelo estatístico escolhido, uma vez que dois valores $\left(\sum_{i=1}^n y_i, \sum_{i=1}^n y_i^2 \right)$ são dados.
- A questão agora é saber se tal situação favorável pode ser estendida em geral, ou pelo menos para algumas classes de modelos (neste caso, para quais classes?) A seguir, iremos examinar a natureza e propriedades daquelas estatísticas capazes de sumarizar **toda a informação** presente na função de verossimilhança.

Estatísticas Suficientes

Definição 8 Para o modelo $\mathcal{F} = \{f(\cdot \mid \theta) : \theta \in \Theta\}$, uma estatística $T(y)$ é dita **suficiente** para θ se assume valores em dois pontos do espaço amostral somente se estes pontos possuem verossimilhanças equivalentes. Isto é:

$$\forall y, z \in \mathcal{Y}, \quad T(y) = T(z) \Rightarrow L(\theta; y) \propto L(\theta; z), \forall \theta \in \Theta.$$

- Devemos ter em mente que a propriedade de suficiência está diretamente relacionada à escolha do modelo.
- Se o modelo é **alterado**, as estatísticas em questão podem não ser mais suficientes.
- Para qualquer modelo, sempre existirá uma estatística suficiente que será a própria amostra $\mathbf{y} = (y_1, \dots, y_n)'$, entretanto, esta estatística suficiente é uma escolha muito trivial e na prática é desconsiderada.

Exemplo 12 Suponha que θ pode assumir dois valores (0 e 1). As duas funções de verossimilhança correspondentes são fornecidas a seguir:

θ	$P(Y = 0)$	$P(Y = 1)$	$P(Y = 2)$
0	8/12	1/12	3/12
1	4/12	2/12	6/12

Note que:

$$\underbrace{L(\theta = 0; y = 2)}_{3/12} = 3 \underbrace{L(\theta = 0; y = 1)}_{1/12}$$

$$\underbrace{L(\theta = 1; y = 2)}_{6/12} = 3 \underbrace{L(\theta = 1; y = 1)}_{2/12}$$

Estatística:

$$T(y) = I_{\{0\}}(y) = \begin{cases} 1, & \text{se } y = 0 \\ 0, & \text{se } y \neq 0 \end{cases}$$

$$T(y = 1) = 0 = T(y = 2) \Rightarrow L(\theta; y = 1) \propto L(\theta; y = 2).$$

Estatísticas Suficientes

Exemplo 13 Considere $g(\cdot; \theta)$ a f.d. associada a uma a.a. simples $\mathbf{y} = (y_1, y_2, \dots, y_n)$.

- A função de verossimilhança pode ser escrita como

$$L(\theta; \mathbf{y}) = \prod_{i=1}^n g(y_i; \theta) = \prod_{i=1}^n g(y_{(i)}; \theta),$$

onde na última igualdade, os termos foram multiplicados após terem sido organizados de acordo com as estatísticas de ordem.

- Portanto, duas amostras com as mesmas estatísticas de ordem possuem as mesmas funções verossimilhança.
- Segue então que, para quaisquer f.d.'s $g(\cdot; \theta)$ as estatísticas de ordem são suficientes.

Fatoração de Neyman

Fatoração de Neyman

- Se $T(\cdot)$ é uma estatística suficiente, então $L(\theta; \mathbf{y})$ depende apenas através de $T(\mathbf{y})$. Isto significa que existe uma função g tal que $L(\theta; \mathbf{y}) \propto g(T(\mathbf{y})|\theta)$.
- Note que $L(\theta; \mathbf{y}) \propto f(\mathbf{y}|\theta)$, logo

$$\frac{f(\mathbf{y}|\theta)}{g(T(\mathbf{y})|\theta)},$$

não dependerá de θ .

- Denote:

$$h(\mathbf{y}) = \frac{f(\mathbf{y}|\theta)}{g(T(\mathbf{y})|\theta)},$$

portanto, se $T(\cdot)$ é estatística suficiente, a seguinte relação será válida:
 $f(\mathbf{y}|\theta) = h(\mathbf{y})g(T(\mathbf{y})|\theta)$.

Fatoração de Neyman

Teorema 4 (Teorema da Fatoração de Neyman) Para o modelo

$\mathcal{F} = \{f(\cdot | \theta) : \theta \in \Theta \subseteq \mathbb{R}^k\}$ a estatística $T(\cdot)$ é suficiente para θ se e somente se $f(\mathbf{y}; \theta)$ pode ser escrita na forma $f(\mathbf{y}|\theta) = h(\mathbf{y})g(T(\mathbf{y})|\theta)$ para alguma função g e h .

i Uma forma alternativa de interpretar a definição:

Se conhecermos o valor amostral de $t = T(\mathbf{y})$ e escrevesseamos a verossimilhança $L_T(\theta; t)$ para o modelo estatístico associado a distribuição de T , então tal verossimilhança seria equivalente a $L(\theta; \mathbf{y})$.

Fatoração de Neyman

Exemplo 14 Considere $g(\cdot; \theta)$ a f.d. associada a uma a.a. simples $\mathbf{y} = (y_1, y_2, \dots, y_n)$ obtida de forma que $Y_i \sim \text{Bin}(1, \theta)$.

- Temos a verossimilhança:

$$\begin{aligned} L(\theta; \mathbf{y}) &= \prod_{i=1}^n \theta^{y_i} (1 - \theta)^{1-y_i} \\ &= \theta^{\sum_{i=1}^n y_i} (1 - \theta)^{n - \sum_{i=1}^n y_i} \\ &= \theta^{T(\mathbf{y})} (1 - \theta)^{n - T(\mathbf{y})} \\ &= \underbrace{1}_{h(\mathbf{y})} \underbrace{\theta^{T(\mathbf{y})} (1 - \theta)^{n - T(\mathbf{y})}}_{g(T(\mathbf{y})|\theta)} \end{aligned}$$

Fatoração de Neyman

- Temos então que $T(\mathbf{y}) = \sum_{i=1}^n y_i$ é uma estatística suficiente para θ e que $T(\mathbf{y}) \sim \text{Bin}(n, \theta)$, em que

$$\binom{n}{t} \theta^t (1 - \theta)^{n-t}.$$

- Uma outra forma de definir estatística suficiente pode ser expressa como segue:

Definição 9 Uma estatística $T(\cdot)$ é suficiente para θ se a distribuição condicional de Y dado o valor de $T(Y)$ não depende de θ . Em outras palavras:

$$P(Y = y | T(Y) = T(y), \theta) = P(Y = y | T(Y) = T(y)).$$

Fatoração de Neyman

Exemplo 15 Sejam (X_1, \dots, X_n) uma a.a. da v.a. $X \sim Ber(\theta)$. Verifique se $T = \sum_{i=1}^n X_i$ suficiente para θ .

Fatoração de Neyman

Exemplo 16 Considere $\mathbf{y} = (y_1, y_2, \dots, y_n)$ uma a.a. simples da $N(\mu, \sigma^2)$,

$$\begin{aligned} L(\theta; \mathbf{y}) &= \prod_{i=1}^n (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} (y_i - \mu)^2\right\} I_{(-\infty, \infty)}(y_i) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i + n\mu^2\right]\right\} \prod_{i=1}^n I_{(-\infty, \infty)}(y_i) \\ &= \underbrace{(2\pi)^{-n/2} \prod_{i=1}^n I_{(-\infty, \infty)}(y_i)}_{h(\mathbf{y})} \underbrace{(\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i + n\mu^2\right]\right\}}_{g(T(\mathbf{y})|\theta)} \end{aligned}$$

- Então $T(\mathbf{y}) = \left(\sum_{i=1}^n y_i, \sum_{i=1}^n y_i^2\right)$ é uma estatística suficiente para $\theta = (\mu, \sigma^2)$ de acordo com o método da fatoração de Neyman.

Fatoração de Neyman

Observação: Qualquer função 1 a 1 de uma estatística suficiente também é uma estatística suficiente.

- Suponha que $T(\mathbf{y})$ é estatística suficiente e defina $T^*(\mathbf{y}) = r(T(\mathbf{y}))$ para todo \mathbf{y}
- r é uma função 1 a 1 com inversa r^{-1} .
- Teorema da Fatoração, existe g e h tal que

$$f(\mathbf{y}|\theta) = h(\mathbf{y})g[r^{-1}(T^*(\mathbf{y}))|\theta]$$

- Defina $g^*(t|\theta) = g(r^{-1}(t)|\theta)$, então

$$f(\mathbf{y}|\theta) = h(\mathbf{y})g^*[T^*(\mathbf{y})|\theta]$$

e pelo Teorema da Fatoração $T^*(\mathbf{y})$ é uma estatística suficiente.

Fatoração de Neyman

- No caso anterior considere:

$$(t_1, t_2) = \left(\sum_{i=1}^n y_i, \sum_{i=1}^n y_i^2 \right) \text{ nossa estatística suficiente para } \theta = (\mu, \sigma^2).$$

- Veja que:

$$(\bar{y}, S^2) = \left(\frac{t_1}{n}, \frac{t_2 - (t_1^2/n)}{n-1} \right) = \left(\frac{\sum_{i=1}^n y_i}{n}, \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} \right)$$

que são função 1 a 1 de (t_1, t_2) com transformação inversa

$$t_1 = n\bar{y} \text{ e } t_2 = (n-1)S^2 + n\bar{y}^2.$$

- Logo, (\bar{y}, S^2) também é uma estatística suficiente para $\theta = (\mu, \sigma^2)$.

Fatoração de Neyman

Exemplo 17 Considere uma a.a. $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ da v.a. $U(\theta, 2\theta)$, onde $\theta > 0$. A f.d. associada a y_i é $f(y_i) = \frac{1}{\theta}$, se $y_i \in (\theta, 2\theta)$.

- Quando multiplicamos tais f.d.'s para obter $L(\theta; \mathbf{y})$, não podemos simplesmente multiplicar o termo $1/\theta$ sem considerar a condição se $y_i \in (\theta, 2\theta)$.
- Portanto, escrevemos a densidade para uma única observação, como segue,

$$\frac{1}{\theta} I_{(\theta, 2\theta)}(y).$$

Fatoração de Neyman

Função de verossimilhança

$$\begin{aligned} L(\theta; \mathbf{y}) &= \prod_{i=1}^n \frac{1}{\theta} I_{(\theta, 2\theta)}(y_i) \\ &= \frac{1}{\theta^n} \prod_{i=1}^n I_{(\theta, 2\theta)}(y_i) \end{aligned}$$

- Note que $\prod_{i=1}^n I_{(\theta, 2\theta)}(y_i) = 1$, se

$$\begin{array}{l} \theta < y_1 < 2\theta \\ \theta < y_2 < 2\theta \\ \vdots \\ \theta < y_n < 2\theta \end{array} \implies \theta < y_{(1)} \text{ e } \frac{y_{(n)}}{2} < \theta \implies \left[\frac{y_{(n)}}{2} < \theta < y_{(1)} \right]$$

Fatoração de Neyman

Assim,

$$\begin{aligned} L(\theta; \mathbf{y}) &= \frac{1}{\theta^n} \prod_{i=1}^n I_{(\theta, 2\theta)}(y_i) \\ &= \frac{1}{\theta^n} \prod_{i=1}^n I_{(y_{(n)}/2, y_{(1)})}(\theta) \end{aligned}$$

- Pelo critério da Fatoração, o par $(y_{(1)}, y_{(n)})$ é uma estatística suficiente para θ .

Fatoração de Neyman

- Lembre que qualquer função 1 a 1 de uma estatística suficiente é também estatística suficiente.
- Desta forma, podemos definir inúmeras estatísticas suficientes para um dado problema.
- Poderíamos perguntar se uma estatística suficiente é melhor que as outras.
- O objetivo de uma estatística suficiente é atingir uma redução nos dados sem perder informação sobre o parâmetro θ .
- Iremos preferir a estatística que atinge a maior redução nos dados e mantenha toda informação sobre θ

Definição 10 Uma estatística suficiente $T(\mathbf{y})$ é chamada de **Estatística suficiente minimal** se, para qualquer outra estatística suficiente $T'(\mathbf{y})$, $T(\mathbf{y})$ é uma função de $T'(\mathbf{y})$. Dizer que $T(\mathbf{y})$ é função de $T'(\mathbf{y})$ significa que

$$T'(\mathbf{x}) = T'(\mathbf{y}) \Rightarrow T(\mathbf{x}) = T(\mathbf{y})$$

Fatoração de Neyman

Exemplo 18 Duas estatísticas suficientes (caso Normal).

Já vimos anteriormente que se $\mathbf{y} = (y_1, y_2, \dots, y_n)$ uma a.a. simples da $N(\mu, \sigma^2)$ com $\theta = (\mu, \sigma^2)$, temos $T'(\mathbf{y}) = (\bar{y}, S^2)$ como estatística suficiente para (μ, σ^2) .

Se σ^2 é conhecido, podemos usar a Fatoração de Neyman e obter

$$\begin{aligned} L(\theta; \mathbf{y}) &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i + n\mu^2 \right]\right\} \\ &= \underbrace{(2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n y_i\right\}}_{h(\mathbf{y})} \underbrace{\exp\left\{-\frac{n}{2} \left[\frac{\mu^2}{\sigma^2} - \frac{2}{\sigma^2} \bar{y} \right]\right\}}_{g(T(\mathbf{y})|\theta)} \end{aligned}$$

Portanto, $T(\mathbf{y}) = \bar{y}$ é estatística suficiente para μ .

Fatoração de Neyman

- Se σ^2 é conhecido, temos que $T'(\mathbf{y}) = (\bar{y}, S^2)$ e $T(\mathbf{y}) = \bar{y}$ são estatísticas suficientes para μ .
- Claramente, $T(\mathbf{y})$ atinge maior redução nos dados neste caso.
- Podemos escrever $T(\mathbf{y})$ como função de $T'(\mathbf{y})$ usando a seguinte igualdade: $r(a, b) = a$. Então, $T(\mathbf{y}) = \bar{y} = r(\bar{y}, S^2) = r(T'(\mathbf{y}))$.
- Como $T(\mathbf{y})$ e $T'(\mathbf{y})$ são ambas estatísticas suficientes, as duas possuem a mesmas informações sobre μ .
- Então a informação adicional S^2 não acrescenta nada ao nosso conhecimento de μ , dado que σ^2 é conhecido.
- Obviamente, se σ^2 é desconhecido, a estatística $T(\mathbf{y}) = \bar{y}$ deixa de ser suficiente e $T'(\mathbf{y}) = (\bar{y}, S^2)$ passa a conter mais informação sobre (μ, σ^2) do que $T(\mathbf{y})$.

Estatística Suficiente Minimal

Estatística Suficiente Minimal

- Usar a última definição para encontrar uma estatística suficiente minimal não é uma tarefa prática.
- Teríamos que adivinhar que $T(\mathbf{y})$ é uma estatística suficiente minimal e então verificar a condição dada na definição.
- Felizmente, o seguinte resultado de Lehmann e Scheffé (1950) fornece uma maneira mais fácil de encontrar uma estatística suficiente minimal.

Estatística Suficiente Minimal

Teorema 5 Seja $f(\mathbf{y})$ é uma f.d. associada com a amostra \mathbf{y} . Suponha que exista uma função $T(\mathbf{y})$ tal que para quaisquer dois pontos amostrais x e y , a razão $f(x|\theta)/f(y|\theta)$ é constante como função de θ se e somente se $T(x) = T(y)$. Então $T(\mathbf{y})$ é uma estatística suficiente minimal para θ .

Exemplo 19 Estatística suficiente minimal (caso Normal). Sejam (Y_1, Y_2, \dots, Y_n) uma a.a. simples da $N(\mu, \sigma^2)$ com $\theta = (\mu, \sigma^2)$ desconhecido.

Considere $\mathbf{x} = (x_1, x_2, \dots, x_n)$ e $\mathbf{y} = (y_1, y_2, \dots, y_n)$ são dois pontos amostrais.

(\bar{x}, S_x^2) é a média e a variância amostral de \mathbf{x}

(\bar{y}, S_y^2) é a média e a variância amostral de \mathbf{y}

Estatística Suficiente Minimal

- O Exemplo 19 solicita a seguinte razão:

$$\begin{aligned}
 L(\theta; \mathbf{y}) &= \frac{(2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2\right]\right\}}{(2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n y_i^2 - 2\mu \sum_{i=1}^n y_i + n\mu^2\right]\right\}} \\
 &= \frac{\exp\left\{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i + \bar{x} - \bar{x})^2 - 2n\mu\bar{x}\right]\right\} \exp\left\{-\frac{1}{2\sigma^2} n\mu^2\right\}}{\exp\left\{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (y_i + \bar{y} - \bar{y})^2 - 2n\mu\bar{y}\right]\right\} \exp\left\{-\frac{1}{2\sigma^2} n\mu^2\right\}} \\
 &= \frac{\exp\left\{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (x_i - \bar{x})^2 + n\bar{x}^2 - 2n\mu\bar{x}^2\right]\right\}}{\exp\left\{-\frac{1}{2\sigma^2} \left[\sum_{i=1}^n (y_i - \bar{y})^2 + n\bar{y}^2 - 2n\mu\bar{y}^2\right]\right\}}
 \end{aligned}$$

Estatística Suficiente Minimal

Assim,

$$\begin{aligned} L(\theta; \mathbf{y}) &= \exp \left\{ -\frac{1}{2\sigma^2} \left[(n-1)S_x^2 \bar{x}^2 (n-2n\mu) - (n-1)S_y^2 - \bar{y}^2 (n-2n\mu) \right] \right\} \\ &= \exp \left\{ -\frac{1}{2\sigma^2} \left[(n-1)(S_x^2 - S_y^2)(n-2n\mu)(\bar{x}^2 - \bar{y}^2) \right] \right\} \end{aligned}$$

- Este resultado será constante como função de μ e σ^2 se e somente se $\bar{x} = \bar{y}$ e $S_x^2 = S_y^2$.
Então pelo [Exemplo 19](#) (\bar{X}, S^2) é uma estatística suficiente minimal para (μ, σ^2) .

Estatística Suficiente Minimal

Exemplo 20 (Estatística suficiente minimal (caso Uniforme).) Sejam (X_1, X_2, \dots, X_n) uma a.a. simples da $Uniforme(\theta, \theta + 1)$ com $-\infty < \theta < \infty$.

Considere $\mathbf{x} = (x_1, x_2, \dots, x_n)$ e $\mathbf{y} = (y_1, y_2, \dots, y_n)$ são dois pontos amostrais.

- A densidade para uma única observação, como segue,

$$f(x \mid \theta) = I_{(\theta, \theta+1)}(x), \quad \theta \in \mathbb{R}.$$

Estatística Suficiente Minimal

Assim,

$$\begin{array}{l} \theta < x_1 < \theta + 1 \\ \theta < x_2 < \theta + 1 \\ \vdots \\ \theta < x_n < \theta + 1 \end{array} \implies \theta < x_{(1)} \text{ e } x_{(n)} - 1 < \theta \implies [x_{(n)} - 1 < \theta < x_{(1)}]$$

- Logo,

$$f(x \mid \theta) = I_{(x_{(n)}-1, x_{(1)})}(\theta), \quad \theta \in \mathbb{R}.$$

Estatística Suficiente Minimal

- Então,

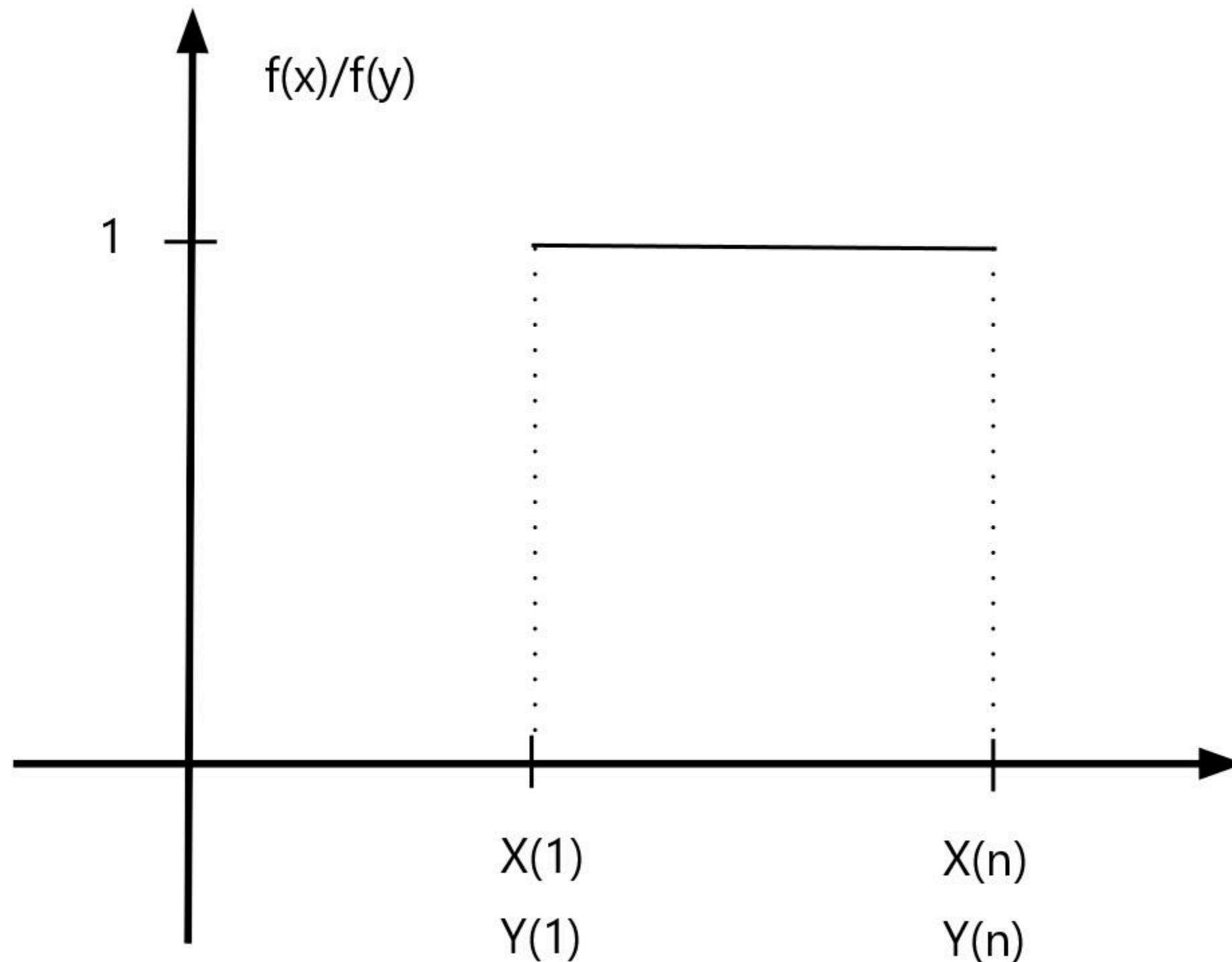
$$f(x \mid \theta) = I_{(x_{(n)}-1, x_{(1)})}(\theta) = \begin{cases} 1 & \text{se } x_{(n)} - 1 < \theta < x_{(1)} \\ 0 & \text{c.c.} \end{cases}$$

- De forma similar:

$$f(y \mid \theta) = I_{(y_{(n)}-1, y_{(1)})}(\theta) = \begin{cases} 1 & \text{se } y_{(n)} - 1 < \theta < y_{(1)} \\ 0 & \text{c.c.} \end{cases}$$

Estatística Suficiente Minimal

A razão $\frac{f(x | \theta)}{f(y | \theta)}$ será constante como função de θ se e só $x_{(1)} = y_{(1)}$ e $x_{(n)} = y_{(n)}$.



Conclusão: $T(X) = (x_{(1)}, x_{(n)})$ é uma estatística suficiente minimal para θ . Neste caso, note que a dimensão é diferente da dimensão do parâmetro.