Capítulo 1

Estimação Pontual

1.1 Introdução

Inferência Estatística, ou ”Aprendizado”como é chamado em Ciência da Computação, é o processo de buscar dados para inferir a distribuição que gera esses dados. Uma típica questão de Inferência Estatística é a seguinte:

Dado uma amostra $X_l, ..., X_n \sim F$, como inferimos F?

Em alguns casos, podemos querer inferir apenas algumas características de F, como sua média por exemplo.

Um **modelo estatístico** $\mathcal{F}$ é um conjunto de distribuições (ou densidades ou funções de regressão).

Um **modelo paramétrico** é um conjunto $\mathcal{F}$ que pode ser parametrizado por um número finito de parâmetros.
Por exemplo, se nós assumimos que os dados provém de uma distribução Normal, então o modelo é:

$$\mathcal{F} = \left\{ f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}, \mu \in \mathbb{R}, \sigma > 0 \right\}$$

Este é um modelo de dois parâmetros. Escrevemos a densidade como $f(x; \mu, \sigma)$ para mostrar que x é um valor da variável aleatória enquanto μ, σ são parâmetros.

Em geral, um modelo paramétrico assume a forma:

$$\mathcal{F} = \{f(x;\theta), \theta \in \Theta\}$$

onde θ é um parâmetro desconhecido (ou vetor de parâmetros) que pode levar valores no **espaço de parâmetros** Θ. Se θ é um vetor, mas estamos apenas interessados em um componente de θ, chamamos os parâmetros restantes de **parâmetros incômodos**.

Um **modelo não paramétrico** é um conjunto $\mathcal{F}$ que não pode ser parametrizado por um número finito de parâmetros.

Estimadores e Estimativas Dado um parâmetro $\theta \in \Theta$*umestimador Muitosproblemasinferenciaisp*
Estimação, IntervalosdeconfiançaaeTestedehipóteses.Vamostratardetodosesseproblemasemdetalh

Estimação Pontual Paramétrica*ou***Estimação Pontual***éumatécnicaquetentafornecerumaúnica*"m

Nestecapítulo, vamosabordarateoriasobreesteassunto.Pararesumir, usaremosotêrmoestimativaer

1.2 A Ideia básica da Estimação Pontual

O problema aqui, resumidamente declarado, é o seguinte. Seja X um v.a. com um f.d.p. f que, no entanto, envolve um parâmetro. Este é o caso, por exemplo, no Distribuição binomial $B(1, p)$, a distribuição de Poisson $P(\lambda)$, a Exponencial negativa $f(x) = \lambda e^{-\lambda x}, x > 0$ a distribuição Uniforme $U(0, \alpha)$, e a distribuição normal $N(\mu, \sigma^2)$ com uma das quantidades μ e σ^2 conhecido. O parâmetro é geralmente denotado por θ, e o conjunto de seus valores possíveis, denotado por Θ é chamado de espaço de parâmetro. A fim de enfatizar o fato de que a f.d.p. depende de θ, escrevemos $f(;\theta)$. Assim, nas distribuições mencionadas acima, temos para as respectivos f.d.p.'s e os espaços de parâmetros:

$f(x;\theta) = \theta^x(1-\theta)^{1/x}, x = 0, 1, \theta \in \Theta = (0, 1)$.
$f(x;\theta) = e^{-\theta}\frac{\theta^x}{x!}, x = 0, 1, ..., \theta \in \Theta = (0, \infty)$.
$f(x;\theta) = \theta e^{-\theta x}, x > 0, \theta \in \Theta = (0, \infty)$.

$$f(x;\theta) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta \\ & \theta \in \Theta = (0, \infty) \\ 0 & \text{Caso contrário} \end{cases}$$

Distribuições normais são adequadas para modelar as situações descritas em Exemplos 16 e 17 do Capítulo 1.

Nosso objetivo é desenhar uma amostra aleatória de tamanho n, $X_1, \ldots, X_n$, do distribuição subjacente e, com base nela, construir uma estimativa pontual (ou estimador) para θ, ou seja, uma estatística $\widehat{\theta} = \widehat{\theta}, X_1, \ldots, X_n$, que é usada para estimar θ, onde uma estatística é uma função conhecida da amostra aleatória $X_1, \ldots, X_n$. Se $x_1, \ldots, x_n$ são os valores realmente observados de v.a.'s $X_1, \ldots, X_n$, respectivamente, então o valor observado de nossa estimativa tem o valor numérico $\widehat{\theta}(x_1, \ldots, x_n)$. Os valores observados $x_1, \ldots, x_n$ também são chamados de dados. Então, com base nos dados disponíveis, é declarado que o valor de θ é $\widehat{\theta}(x_1, \ldots, x_n)$ dentre todos os pontos possíveis em . Uma estimativa pontual é muitas vezes referida apenas como uma estimativa, e a notação $\widehat{\theta}$ é usada indiscriminadamente, tanto para a estimativa $\widehat{\theta}$ (X1,..., Xn) (que é um r.v.) e para seu valor observado $\widehat{\theta}$ (x1,..., Xn) (que é apenas um número).

A única restrição óbvia em θ é $\widehat{\theta}(x_1, \ldots, x_n)$ é que se encontra em para todos valores possíveis de $X_1, \ldots, X_n$. Além disso, há inúmeras estimativas pode-se construir portanto, a necessidade de assumir certos princípios e/ou inventar métodos para construir $\widehat{\theta}$. Talvez, o princípio mais amplamente aceito seja o chamado princípio da Máxima Verossimilhança (MV). Este princípio dita que formamos a junta p.d.f. dos xi's, para os valores observados dos $X_i's$, olhe para esta junta f.d.p. como uma função de θ (e chame-a de função de verossimilhança), e maximizar a função de verossimilhança em relação a θ. O ponto de maximização (assumindo que existe e é único) é uma função de $(x_1, \ldots, x_n)$, e é o que nós chamamos de estimativa de máxima verossimilhança (MLE) de θ. A notação usada para o a função de verossimilhança é $L(\theta|x_1, ..., x_n)$. Então, temos isso:

$L(\theta|x_1, ..., x_n) = f(x_1;\theta) \ldots f(x_n;\theta), \theta \in \Theta$.

O MLE será estudado extensivamente no Capítulo 9.

Outro princípio frequentemente usado na construção de uma estimativa para θ é o prin- princípio de imparcialidade. Neste contexto, uma estimativa geralmente é denotada por $U = U(X_1, \ldots, X_n)$. Então, o princípio da imparcialidade dita que U deve ser construído de modo a ser imparcial; ou seja, sua expectativa (valor médio) deve ser sempre θ, não importa qual seja o valor de θ . Mais formalmente, $E_\theta U = \theta$ para todos $\theta \in \Theta$. (No sinal de expectativa E, o parâmetro θ foi inserido para indicar que esta expectativa depende de θ, uma vez que é calculada usando o f.d.p. $f(\cdot;\theta)$.) Agora, é intuitivamente claro que, ao comparar duas estimativas imparciais companheiros, um escolheria o outro com a pequena servidão, uma vez que seria mais fortemente concentrado em torno de sua média θ. Visualize o caso que, dentro da classe de todas as estimativas imparciais, existe uma que tem a menor variância (e isso é verdade para todos os $\theta \in Theta$). Essa estimativa é chamada de mínimo uniforme Estimativa de variância imparcial (UMVU) e é, claramente, uma estimativa desejável. Dentro No próximo capítulo, veremos como fazemos para construir essas estimativas.

O princípio (ou melhor, o método) baseado em momentos amostrais é outra forma de construir estimativas. O método dos momentos, no caso mais simples, dita para formar a média da amostra $\overline{X}$ e igualá-la com a média (teórica) $E_\theta X$. Em seguida, resolva para θ (assumindo que pode ser feito, e, de fato, exclusivamente) na ordem para chegar a uma estimativa de momento de θ. Um método muito mais sofisticado de construção de estimativas de θ é o o chamado método teórico da decisão. Este método exige a introdução de uma série de conceitos, terminologia e notação, e será retomada na Próximo Capítulo.

Finalmente, outro método relativamente popular (em particular, no contexto de certos modelos) é o método dos mínimos quadrados (LS). O método de ligações LS para a construção de uma estimativa para θ, a Estimativa de Mínimos Quadrados (LSE) de θ, por meio de uma minimização (em relação a θ) da soma de certos quadrados. Esta soma dos quadrados representa os desvios quadrados entre o que realmente observar depois que a experimentação for concluída e o que esperaríamos ter com base em um modelo presumido. Mais uma vez, os detalhes serão apresentados mais tarde em, mais especificamente, no Capítulo 13.

Em toda a discussão anterior, foi assumido que o p.d.f. subjacente. dependia de um único parâmetro, denotado por θ. Pode muito bem ser caso haja dois ou mais parâmetros envolvidos. Isso pode acontecer, por exemplo, na distribuição uniforme $U(\alpha,\beta), -\infty < \alpha < \beta < \infty$, onde ambos α e β são desconhecidos; a distribuição normal, $N(\mu, \sigma^2)$, onde ambos μ e σ^2 são desconhecidos; e isso acontece na distribuição Multinomial, onde o o número de parâmetros é $k, p_1, \ldots, p_k$(ou mais precisamente, k - 1, uma vez que o k-ésimo parâmetro, por exemplo, $p_k = 1 - p_1 - \cdots - p_{k-1}$). Por exemplo, Exemplos 20 e 21 do Capítulo 1 referem-se a situações em que uma distribuição Multinomial é apropriado. Em tais casos de multiparâmetros, simplesmente se aplica a cada parâmetro separadamente o que foi dito acima para um único parâmetro. A opção alternativa de usar a notação vetorial para os parâmetros envolvidos simplifica as coisas de certa forma, mas também introduz algumas complicações de outras maneiras.

Seja X uma v.a. com f.d.p. $f(;\theta)$, onde θ é um parâmetro que está em um espaço de parâmetro Θ. Supõe-se que a forma funcional da f.d.p. é completamente conhecido. Portanto, se θ fosse conhecido, a f.d.p. seria conhecido, e consequentemente, poderíamos calcular, em princípio, todas as probabilidades relacionadas a X, a esperança de X, sua variância, etc. O problema, no entanto, é que na maioria das vezes na prática (e no presente contexto) θ não é conhecido. Então o objetivo é estimar θ com base em uma amostra aleatória de tamanho n de $f(;\theta), X_1, \ldots, X_n$. Então, substituindo θ em $f(;\theta)$ por uma estimativa "boa"dele, seria de esperar ser capaz de usar a f.d.p. resultante para os fins descritos acima em um grau satisfatório.

1.3 Estimação de Máxima Verossimilhança: Motivação e Exemplos

O seguinte exemplo simples destina-se a esclarecer o que é intuitivo, mas bastante lógico, o princípio da Estimativa de Máxima Verossimilhança.

Exemplo 1 *Sejam* $X_1, \ldots, X_{10}$ *v.a's i.i.d. da distribuição* $B(1, \theta)$, $0 < \theta < 1$, *e se* $X_1, \ldots, X_{10}$ *são os respectivos valores observados. Por conveniência, defina* $t = x_1 + \ldots + x_{10}$. *Além disso, suponha que nas 10 tentativas, 6 resultaram em sucessos, de modo que t = 6. Então, a função de verossimilhança envolvida é:* $L(\theta|x) = \theta^6(1-\theta)^4, 0 < \theta < 1, \boldsymbol{x} = (x_1, ..., x_{10})$. *Assim,* $L(\theta|x)$ *é a probabilidade de observar exatamente 6 sucessos em 10 ensaios binomiais independentes, ocorrendo os sucessos nas tentativas para as quais* $x_i = 1, i = 1, \ldots, 10$; *esta probabilidade é uma função do parâmetro (desconhecido)* θ. *Vamos calcular os valores desta probabilidade para* θ *variando de 0,1 a 0,9. Observamos*

Valores de θ	**Valores de** $L(\theta\|x)$
0.1	0.0000006561
0.2	0.0000262144
0.3	0.0001750329
0.4	0.0005308416
0.5	0.0009765625
0.6	0.0011943936
0.7	0.0009529569
0.8	0.0004194304
0.9	0.0000531441

que os valores de $L(\theta|x)$, *são crescentes, atingem seu máximo valor em* $\theta = 0,6$, *e então os valores vão diminuindo. Portanto, se esses 9 valores fossem os únicos valores possíveis para* θ *(o que eles nãoo são!), alguém poderia raciocinar habilmente escolha o valor de 0,6 como o valor de* θ. *O valor* $\theta = 0,6$ *tem a distinção de maximizar (entre os 9 valores listados) a probabilidade de atingir os 6 sucessos já observados. Observamos que* $0,6 = \frac{6}{10} = \frac{t}{n}$ *onde n é o número de tentativas e t* *ó o número de sucessos. Veremos no Exemplo a seguir que o valor* $\frac{t}{n}$*, na verdade, maximiza a função de verossimilhança entre todos os valores de* θ *com* $0 < \theta < 1$*. Então* $\frac{t}{n}$ *será a estimativa de similaridade máxima de* θ *a ser denotada por* $\widehat{\theta}$*; ou seja,* $\widehat{\theta} = \frac{t}{n}$*.*

Definição 1 *Sejam* $X_1, \ldots, X_n$ *var. aleatórias independente e identicamente distribuídas(iid) com fdp* $f(\cdot;\theta)$ *com* $\theta \in \Theta$ *, e sejam* $x_1, \ldots, x_n$ *os respectivos valores observados e* $\boldsymbol{x} = (x_1, \ldots, x_n)$*. A* ***função de verossimilhança*** $L(\theta|\boldsymbol{x})$*, é definida por:*

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

o valor de θ *que maximiza* $L(\theta|\boldsymbol{x})$ *é chamado de* ***Estimativa de Máxima Verossimilhança*** *(****EMV****) de* θ*. Claramente, a* ***EMV*** *depende de* $\boldsymbol{x}$*, e geralmente escrevemos* $\widehat{\theta} = \widehat{\theta}(\boldsymbol{x})$*. Desse modo,*

$$L\left(\widehat{\theta}|\boldsymbol{x}\right) = \max\{L(\theta|\boldsymbol{x}), \theta \in \Theta\}.$$

Vamos enunciar agora um resultado muito importante na determinação das **EMV**'s

Proposição 1 *Dada* $f : \Omega \to \mathbb{R}$, $\Omega \subseteq \mathbb{R}^n$ *contínua. Se* $\log(f(x))$ *possui um máximo* x^* *em* Ω *então* x^* *também é máximo de* f*.*

Prova: Se x^* é um máximo de $\log(f(x))$ então $\log(f(x)) \leq \log(f(x^*))$ para todo $x \in \Omega$ como a função $\log(x)$ é estritamente crescente temos que $f(x) \leq f(x^*)$ para todo $x \in \Omega$ portanto x^* também é um máximo de f.

Exemplo 2 *Em termos de uma amostra aleatória de tamanho n,* $X_1, \ldots, X_n$ *de uma distribuição* $X \sim B(1, \theta)$ *com valores observados* $x_1, \ldots, x_n$*, determine a* ***EMV*** $\widehat{\theta} = \widehat{\theta}(\boldsymbol{x})$ *de* $\theta \in (0, 1)$, $\boldsymbol{x} = (x_1, ..., x_n)$*.*

Sendo $f(x_i, \theta) = \theta^{x_i}(1-\theta)^{1-x_i}$, $x_i = 0$ *ou* $1, i = 1, \ldots, n$*, a função de verossimilhança é dada por:*

$$L(\theta|\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^t(1-\theta)^{n-t}, \quad t = x_1 + \ldots + x_n$$

como $x_i = 0$ *ou* 1*, então* $t = 0, 1, \ldots, n$*. Portanto:*

$$log\,(L(\theta|\boldsymbol{x})) = tlog(\theta) + (n-t)log(1-\theta)$$

derivando com relação a θ *obtemos a equação que denominamos* ***equação de verossimilhança***

$$\frac{\partial log\,(L(\theta|\boldsymbol{x}))}{\partial\theta} = 0$$

obtendo para este caso:

$$\frac{\partial log\,(L(\theta|\boldsymbol{x}))}{\partial\theta} = \frac{t}{\theta} - \frac{n-t}{1-\theta} = 0$$

o que concluindo-se que $\theta = \frac{t}{n}$*. Fazendo-se agora o teste da derivada segunda temos:*

$$\frac{\partial^2 log\,(L(\theta|\boldsymbol{x}))}{\partial\theta^2} = -\frac{t}{\theta^2} - \frac{n-t}{(1-\theta)^2}$$

como $t \leq n$, $\dfrac{\partial^2 log\,(L(\theta|\boldsymbol{x}))}{\partial\theta^2} < 0$ *para todo* θ *em particular para* $\widehat{\theta} = \frac{t}{n}$.

Portanto o ***EMV*** *de* θ *é* $\widehat{\theta} = \frac{t}{n}$.

Exemplo 3 *Determine o* ***EMV*** $\widehat{\theta} = \widehat{\theta}(\boldsymbol{x})$ *de* $\theta \in (0, \infty)$ *na distribuição* $P(\theta)$ *(Poisson) em termos de uma amostra aleatória* $X_1, \ldots, X_n$ *com valores observados* $x_1, ..., x_n$*. Sendo* $f(x_i, \theta) = \dfrac{e^{-\theta}\theta^{x_i}}{x_i!}$, $x_i = 0, 1, \ldots, i = 1, \ldots, n$*, a função de verossimilhança é dada por:*

$$L(\theta|\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{x_i!} = e^{-n\theta} \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!},$$

$$log\,(L(\theta|\boldsymbol{x})) = -n\theta + nlog(\theta)\frac{\sum_{i=1}^{n} x_i}{n} - log\left(\prod_{i=1}^{n} x_i!\right)$$

derivando com relação a θ *obtemos a equação que denominamos* ***equação de verossimilhança***

$$\frac{\partial log\,(L(\theta|\boldsymbol{x}))}{\partial\theta} = 0$$

obtendo para este caso:

$$\frac{\partial log\,(L(\theta|\boldsymbol{x}))}{\partial\theta} = -n + n\frac{\overline{x}}{\theta} = 0$$

o que concluindo-se que $\theta = \overline{x}$. *Fazendo-se agora o teste da derivada segunda temos:*

$$\frac{\partial^2 log\,(L(\theta|\boldsymbol{x}))}{\partial\theta^2} = -n\frac{\overline{x}}{\theta^2}$$

como $x_i \geq 0, \overline{x} > 0$, *logo* $\dfrac{\partial^2 L(\theta|\boldsymbol{x})}{\partial\theta^2} < 0$ *para todo* θ *em particular para* $\widehat{\theta} = \overline{x}$. *Portanto o* ***EMV*** *de* θ *é* $\widehat{\theta} = \overline{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$.

Exemplo 4 *Determine o* ***EMV*** $\widehat{\theta} = \widehat{\theta}(\boldsymbol{x})$ *de* $\theta \in (0, \infty)$ *na distribuição Exponencial* $Exp(\theta)$ *em termos de uma amostra aleatória* $X_1, \ldots, X_n$ *com valores observados* $x_1, \ldots, x_n$.

Nesse caso $f(x_i, \theta) = \theta e^{-\theta x_i}, \;\; x_i > 0$.

$$L(\theta|\boldsymbol{x}) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-n\theta \frac{\sum i=1^n x_i}{n}} = \theta^n e^{-n\theta\overline{x}},$$

$$\log\,(L(\theta|\boldsymbol{x})) = n\log(\theta) - n\overline{x}\theta$$

$$\frac{\partial \log\,(L(\theta|\boldsymbol{x}))}{\partial\theta} = \frac{n}{\theta} - n\overline{x} = 0$$

conclui-se que $\theta = \dfrac{1}{\overline{x}}$

$$\frac{\partial^2 \log\,(L(\theta|\boldsymbol{x}))}{\partial\theta^2} = -\frac{n}{\theta^2} < 0 \forall\theta$$

em particular para $\widehat{\theta} = \dfrac{1}{\overline{x}}$. *Portanto o* ***EMV*** *de* θ *é* $\widehat{\theta} = \dfrac{1}{\overline{x}} = \dfrac{n}{\sum_{i=1}^{n} x_i}$.

Exemplo 5 *Seja $X_1, \dots, X_n$ uma amostra aleatória de uma população com distribuição $\mathcal{N}(\mu, \sigma^2)$, onde apenas um dos parâmetros é conhecido. Determine a* ***EMV*** *do outro (desconhecido) parâmetro.*

DISCUSSÃO *Com $x_1, \dots, x_n$ sendo valores observados de $X_1, \dots, X_n$, nós temos:*

$$L(\mu, \sigma; \boldsymbol{x}) = \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right] = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^2}}, \mu \in \mathbb{R}, \sigma > 0.$$

- *Se μ é desconhecido. Então:*

$$\log(L(\mu|\boldsymbol{x})) = -n \log\left(\sqrt{2\pi}\sigma\right) - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}.$$

Fazendo

$$\frac{\partial \log(L(\mu|\boldsymbol{x}))}{\partial \mu} = 0,$$

obtemos

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

e consequentemente

$$\mu = \overline{x}.$$

Como

$$\frac{\partial^2 \log(L(\mu|\boldsymbol{x}))}{\partial \mu^2} = -\frac{n}{\sigma^2} < 0 \text{ para todo } \mu,$$

concluímos que a ***EMV*** *para para μ sendo σ conhecido é $\widehat{\mu} = \overline{x}$.*

- *Se σ^2 é desconhecido. Então:*

$$\begin{aligned} \log\left(L(\sigma^2|\boldsymbol{x})\right) &= -n \log\left(\sqrt{2\pi\sigma^2}\right) - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2} \\ &= -\frac{n}{2} log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}. \end{aligned}$$

Fazendo

$$\frac{\partial \log\left(L(\sigma^2|\boldsymbol{x})\right)}{\partial \sigma^2} = 0,$$

obtemos

$$-\frac{n\sigma^2}{2\sigma^4} + \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^4} = 0.$$

e consequentemente

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n} = s^2.$$

Como

$$\begin{aligned}
\frac{\partial^2 \log\left(L(\sigma^2|\boldsymbol{x})\right)}{\partial (\sigma^2)^2} &= \frac{n}{2\sigma^4} - \frac{2\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^6} \\
&= \frac{n}{2(\sigma^2)^2} - \frac{2n}{2(\sigma^2)^3}\frac{\sum_{i=1}^n (x_i - \mu)^2}{n} \\
&= \frac{n}{2(\sigma^2)^2} - \frac{2n}{2(\sigma^2)^3}s^2 \\
\left.\frac{\partial^2 \log\left(L(\sigma^2|\boldsymbol{x})\right)}{\partial (\sigma^2)^2}\right|_{\sigma^2=s^2} &= \frac{n}{2(s^2)^2} - \frac{2n}{2(s^2)^3}s^2 \\
&= -\frac{n}{2(s^2)^2} < 0.
\end{aligned}$$

Portanto a ***EMV*** *para para*σ^2 *com conhecido é* $\widehat{\sigma^2} = \dfrac{\sum_{i=1}^n (x_i - \mu)^2}{n}$.

Exemplo 6 *Seja* $X_1, \ldots, X_n$ *uma amostra aleatória de uma população com distribuição* $\mathcal{U}(\alpha, \beta)$*, onde apenas um dos parâmetros é conhecido. Determine a* ***EMV*** *do outro (desconhecido) parâmetro.*

DISCUSSÃO *Com* $x_1, \ldots, x_n$ *sendo valores observados de* $X_1, \ldots, X_n$*, nós temos:*

$$L(\alpha, \beta|\boldsymbol{x}) = \frac{1}{(\beta-\alpha)^n}\prod_{i=1}^n I_{[\alpha,\beta]}(x_i) = \frac{1}{(\beta-\alpha)^n} I_{[\alpha,\beta]}(x_{min}) \times I_{[\alpha,\beta]}(x_{max})$$

onde

$$I_{[\alpha,\beta]}(x) = I(\alpha \leq x \leq \beta) = \begin{cases} 1, & se\ \alpha \leq x \leq \beta \\ \\ 0 & caso\ contrário \end{cases}$$

é a função indicadora do intervalo $[\alpha, \beta]$, $x_{min} = \min\{x_1, \ldots, n\}$ *e* $x_{max} = \max\{x_1, \ldots, x_n\}$.

- *Se* α *é desconhecido.*

$$\beta - \alpha \geq \beta - x_{min} \Rightarrow \frac{1}{(\beta-\alpha)^n} \leq \frac{1}{(\beta - x_{min})^n}$$

portanto $\alpha = x_{min}$ *e consequentemente* $\widehat{\alpha} = x_{min}$ *é a* ***EMV*** *de* α.

- *Se* β *é desconhecido.*

$$\beta - \alpha \geq x_{max} - \alpha \Rightarrow \frac{1}{(\beta-\alpha)^n} \leq \frac{1}{(x_{max} - \alpha)^n}$$

portanto $\beta = x_{max}$ *e consequentemente* $\widehat{\beta} = x_{max}$ *é a* ***EMV*** *de* β.

Exemplo 7 *Seja* $X_1, \ldots, X_n$ *uma amostra aleatória de uma população com distribuição* $\mathcal{N}(\mu, \sigma^2)$*, onde os dois parâmetros são desconhecidos. Determine as* ***EMV's*** *desses parâmetros.*

$$\log\left(L(\mu, \sigma^2|\boldsymbol{x})\right) = -n\log\left(\sqrt{2\pi}\sigma\right) - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}.$$

$$\nabla \log\left(L(\mu, \sigma^2|\boldsymbol{x})\right) = \begin{pmatrix} \dfrac{\left(\sum_{i=1}^n x_i\right) - n\mu}{\sigma^2} \\ \\ -\dfrac{n\sigma^2}{2\sigma^4} + \dfrac{\sum_{i=1}^n (x_i-\mu)^2}{2\sigma^4} \end{pmatrix}$$

Fazendo

$$\nabla \log\left(L(\mu, \sigma^2|\boldsymbol{x})\right) = \mathbf{0}$$

Obtemos a única solução (μ, σ^2) *deste sistema de equações não lineares dada por:*

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \text{ e } \sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

A matriz Hessiana de $\log\left(L(\mu, \sigma^2|\boldsymbol{x})\right)$ *é dada por:*

$$\nabla^2 \log\left(L(\mu, \sigma^2|\boldsymbol{x})\right) = \begin{pmatrix} -\dfrac{n}{\sigma^2} & 2\dfrac{\left(\sum_{i=1}^n x_i - n\mu\right)}{2\sigma^4} \\ -\dfrac{\sum_{i=1}^n x_i - n\mu}{(\sigma^2)^2} & \dfrac{n}{2(\sigma^2)^2} - \dfrac{2n}{2(\sigma^2)^3}s^2 \end{pmatrix}$$

Substituindo nos valores obtidos de $\widehat{\mu}$ *e* $\widehat{\sigma^2}$ *obtemos:*

$$\nabla^2 \log\left(L(\widehat{\mu}, \widehat{\sigma^2}|\boldsymbol{x})\right) = \begin{pmatrix} -\dfrac{n}{\sigma^2} & 0 \\ 0 & -\dfrac{n}{2(s^2)^2} \end{pmatrix}$$

Como $\nabla^2 \log\left(L(\widehat{\mu}, \widehat{\sigma^2}|\boldsymbol{x})\right)$ *é é uma matriz definida negativa o par* $(\widehat{\mu}, \widehat{\sigma^2})$ *é uma* ***EMV*** *para* (μ, σ^2)*.*

Exemplo 8 *Um experimento multinomial é realizado independentemente n vezes, de modo que a função de verossimilhança é dada por:*

$$L(\theta_1, \ldots, \theta_r|\boldsymbol{x}) = \frac{n!}{\prod_{i=1}^n x_i!}\theta_1^{x_1}, \ldots, \theta_r^{x_r},$$

onde $x_i \geq 0$ *e inteiro ,* $\displaystyle\sum_{i=1}^r x_i = n,\ 0 < p_i < 1,\ \sum_{i=1}^r \theta_i = 1,\ \ i = 1, \ldots, r$*. Determine as* ***EMV's*** *para* $\boldsymbol{p}$*.*

DISCUSSÃO: *Nesse caso trata-se de um problema de otimização com uma restrição de igualdade:*

$$\begin{aligned} &\textit{Maximize} \quad \log(L(\theta_1, \ldots, \theta_r|\boldsymbol{x})) = \sum_{i=1}^r x_i \log(\theta_i) - n! \log\left(\prod_{i=1}^n x_i!\right) \\ &\textit{Sujeito à:} \quad \sum_{i=1}^r \theta_i = 1, \ \ \theta_i > 0 \end{aligned}$$

Cuja função Lagrangeana associada $\mathcal{L} : \mathbb{R}^r \times \mathbb{R} \to \mathbb{R}$ *é dada por:*

$$\mathcal{L}(\boldsymbol{\theta}, \lambda) = \sum_{i=1}^r x_i \log(\theta_i) - n! \log\left(\prod_{i=1}^n x_i!\right) - \lambda\left(\sum_{i=1}^r \theta_i - 1\right)$$

com as seguintes condições de otimalidade de $1^{a.}$ *ordem:*

$$\nabla\mathcal{L}(\boldsymbol{\theta},\lambda) = \mathbf{0} \quad \Leftrightarrow \tfrac{x_i}{\theta_i} - \lambda = 0,\ i = 1,\dots,r.$$
$$\sum_{i=1}^{r}\theta_i = 1.$$

$\frac{x_i}{\theta_i} - \lambda = 0,\ i = 1,\dots,r.$ *nos diz que* $\theta_i = \frac{x_i}{\lambda}$, *como* $\sum_{i=1}^{r}\theta_i = 1$. *substituindo os valores de* θ_i *nessa equação obtemos* $\lambda = \sum_{i=1}^{r} x_i = n$ *obtendo finalmente* $\theta_i = \frac{x_i}{n}$. *Portanto* $\widehat{\boldsymbol{\theta}} = \frac{\boldsymbol{x}}{n}$, *uma vez que a matriz Hessiana da função* $\nabla^2 L(\boldsymbol{\theta})$ *dada por:*

$$\nabla^2 L(\boldsymbol{\theta}) = \begin{pmatrix} -\frac{x_1}{\theta_1^2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & -\frac{x_r}{\theta_r^2} \end{pmatrix} \text{é Semidefinida Negativa isto é:}$$

$$\sum_{i=1}^{r}\sum_{j=1}^{r}\frac{\partial^2 L(\boldsymbol{\theta},\lambda)}{\partial\theta_i\partial\theta_j}v_i v_j = -\sum_{i=1}^{r}\frac{x_i}{\theta_i^2}v_i^2 < 0 \text{ para todo } \boldsymbol{v} \in \mathbb{R}^r.$$

1.3.1 A Família Exponencial e os Estimadores de Máxima Verossimilhança

Definição 2 *(**Família Exponencial**: O caso Uniparamétrico)*
Uma fdp $f(;\theta), \theta \in \Theta \subseteq \mathbb{R}$ *pertence a* Família Exponencial*, se* $f(x;\theta)$ *é da forma:*

$$f(x;\theta) = C(\theta)e^{Q(\theta)T(x)} \times h(x)$$

onde Q *é estritamente monótona e* h *não envolve* $\theta, C(\theta)$ *é simplesmente uma constante de normalização.*

Vejamos alguns exemplos de membros dessa família

Exemplo 9 *Se* $X \sim \mathcal{B}(n,\theta)$

$$f(x;\theta) = \frac{n!}{(n-x)!x!}\theta^k(1-\theta)^{n-x}I_A(x)$$

onde $A = \{1,2,\dots,n\}$. *Podemos escrever esta fdp da seguinte forma:*

$$
\begin{aligned}
f(x;\theta) &= (1-\theta)^n \frac{n!}{(n-x)!x!}\left(\frac{\theta}{1-\theta}\right)^x I_A(x) \\
&= (1-\theta)^n \frac{n!}{(n-x)!x!} e^{\ln\left[\left(\frac{\theta}{1-\theta}\right)^x\right]} I_A(x) \\
&= (1-\theta)^n e^{\ln\left(\frac{\theta}{1-\theta}\right)x} \frac{n!}{(n-x)!x!} I_A(x)
\end{aligned}
$$

Fazendo $C(\theta) = (1-\theta)^n$, $Q(\theta) = \ln\left(\frac{\theta}{1-\theta}\right)$, $T(x) = x$ *e* $h(x) = \frac{n!}{(n-x)!x!}$ *vemos que* $f(x;\theta)$ *é um membro da família exponencial.*

Exemplo 10 *Se* $X \sim \mathcal{N}(\mu, \sigma^2)$ *e* μ *é conhecido então* $\sigma^2 = \theta$*, logo*

$$
f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty
$$

Se σ^2 *é conhecido então* $\mu = \theta$*, logo*

$$
\begin{aligned}
f(x;\theta) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}}, -\infty < x < \infty \\
&= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2-2\theta x+\theta^2}{2\sigma^2}}, -\infty < x < \infty \\
&= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\theta^2}{2\sigma^2}} e^{\frac{\theta}{\sigma^2}x} e^{-\frac{x^2}{2\sigma^2}}, -\infty < x < \infty
\end{aligned}
$$

Se μ *é conhecido então* $\sigma^2 = \theta$*, logo*

$$
\begin{aligned}
f(x;\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty \\
f(x;\theta) &= \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x-\mu)^2}{2\theta}} \\
&= \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2\theta}(x-\mu)^2}
\end{aligned}
$$

Fazendo $C(\theta) = \dfrac{1}{\sqrt{2\pi\theta}}$, $Q(\theta) = e^{\frac{1}{2\theta}}$, $T(x) = e^{-(x-\mu)^2}$ *e* $h(x) = 1$, *vemos que* $f(x;\theta)$ *é um membro da* família exponencial.

Definição 3 *(**Família Exponencial**: O caso Multiparamétrico)*
Sejam $X_1, dots, X_r$ *var aleat. iid e* $\boldsymbol{X} = (X_1, \ldots, X_r)^t$ *, dizemos que sua fdp conjunta pertence a a* $r-$ paramétrica família exponencial *se ela puder ser escrita da seguinte forma:*

$$f(\boldsymbol{x}; \boldsymbol{\theta}) = C(\boldsymbol{\theta}) \exp\left[\sum_{j=1}^{r} Q_j(\boldsymbol{\theta}) T_j(\boldsymbol{x})\right] h(\boldsymbol{x})$$

onde $\boldsymbol{x} = (x_1, \ldots, x_n)^t$, $x_i \in \mathbb{R}$ $i = 1, \ldots, r$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^t \in \Theta \subset \mathbb{R}^r$, $C(\boldsymbol{\theta}) > 0$ *e* $h(\boldsymbol{x}) > 0$ *para* $\boldsymbol{x} \in S$ *sendo* S *o conjunto onde* f *é positiva.*

Exemplo 11 *A distribuição Multinomial*

$$f(\boldsymbol{x}) = \frac{n!}{\prod_{j=1}^{r} x_j!} \prod_{j=1}^{r} p_j^{x_j}, \quad \sum_{j=1}^{r} p_j = 1, \quad p_j > 0.$$

$$\begin{aligned}
f(\boldsymbol{x}; \boldsymbol{\theta}) &= \frac{n!}{\prod_{j=1}^{r} x_j!} \prod_{j=1}^{r} \theta_j{}^{x_j} \quad \sum_{j=1}^{r} \theta_j = 1, \quad \theta_j > 0. \\
&= \frac{n!}{\prod_{j=1}^{r} x_j!} \theta_r^r \prod_{j=1}^{r} \left(\frac{\theta_j}{\theta_r}\right)^{x_j} \\
&= \frac{n!}{\prod_{j=1}^{r} x_j!} (1 - \theta_1 - \theta_2 - \ldots \theta_{r-1})^r \prod_{j=1}^{r} \left(\frac{\theta_j}{1 - \theta_1 - \theta_2 - \ldots \theta_{r-1}}\right)^{x_j} \\
&= (1 - \theta_1 - \theta_2 - \ldots \theta_{r-1})^r \exp\left[\sum_{j=1}^{r} x_j \log\left(\frac{\theta_j}{1 - \theta_1 - \theta_2 - \ldots \theta_{r-1}}\right)\right] \frac{n!}{\prod_{j=1}^{r} x_j!}
\end{aligned}$$

$C(\boldsymbol{\theta}) = (1 - \theta_1 - \theta_2 - \ldots \theta_{r-1})^r$, $Q_j(\boldsymbol{\theta}) = \log\left(\frac{\theta_j}{1-\theta_1-\theta_2-\ldots\theta_{r-1}}\right)$, $T(\boldsymbol{x}) = \boldsymbol{x}$ *e*
$h(\boldsymbol{x}) = \frac{n!}{\prod_{j=1}^{r} x_j!}$

Exemplo 12 *A distribuição Normal. Se $X \sim \mathcal{N}(\theta_1, \theta_2)$ vamos mostrar que X também é um membro de uma 2 paramétrica família exponencial*

$$\begin{aligned} f(x;\theta_1,\theta_2) &= \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{(x-\theta_1)^2}{2\theta_2}}, -\infty < x < \infty \\ &= \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{x^2-2\theta_1 x+\theta_1^2}{2\theta_2}} = \frac{e^{-\frac{\theta_1^2}{2\theta_2}}}{\sqrt{2\pi\theta_2}} e^{-\frac{1}{2\theta_2}x^2+\frac{\theta_1^2}{2\theta_2}x} \end{aligned}$$

Fazendo $C(\boldsymbol{\theta}) = \frac{e^{-\frac{\theta_1^2}{2\theta_2}}}{\sqrt{2\pi\theta_2}}$, $Q(\boldsymbol{\theta}) = \left(\frac{1}{2\theta_2}, \frac{\theta_1^2}{2\theta_2}\right)^t$ *e* $T(x) = \left(-x^2, x\right)^t$ *podemos ver que trata-se de um membro da bi-paramétrica família exponencial.*

1.3.2 Exercícios

1.1 Se $X_1, \ldots, X_n$ são r.v. independentes distribuídos como $B(k,\theta), \theta \in \Theta = (0,1)$, com os respectivos valores observados $x_1 \ldots, x_n$, mostre que $\widehat{\theta} = \frac{\overline{x}}{k}$ é a **EMV** de θ, onde $\overline{x}$ é a média da amostra dos $x_i's$.

1.2 Se as v.a's independentes. $X_1, \ldots, X_n$ têm distribuiçao geométrica com fdp:

$$f(x;\theta) = \theta(1-\theta)^{x-1}, \ x = 1, 2, \ldots, \ \theta \in \Theta = (0,1),$$

e respectivos valores observados $x_1, \ldots, x_n$, mostre que $\theta = \frac{1}{\overline{x}}$ é a **EMV** de θ.

1.3 Com base em uma amostra aleatória de tamanho n com fdp: $f(x;\theta) = (\theta+1)x^\theta, 0 < x < 1, \theta \in \Theta = (-1,\infty)$, encontre a **EMV** de θ.

1.4 Com base em uma amostra aleatória de tamanho n de de uma v.a com fdp. $f(x;\theta) = \theta x^{\theta-1}, \theta < x < 1, \theta \in \Theta = (0,\infty)$, encontre a **EMV** de θ.

1.5 **i-** Mostre que a função $f(x;\theta) = \frac{1}{2\theta}e^{-|x|/\theta}, x \in \mathbb{R}, \theta \in \Theta = (0,\infty)$ é uma fdp, e desenhe seu gráfico. (f é a chamada Exponencial Dupla),

ii- Com base em uma amostra aleatória desta fdp., encontre a **EMV** de θ.

1.6 **i -** Verifique que a função $f(x;\theta) = \theta^2 x e^{-\theta x}$, $x > 0$ $\theta \in \Theta = (0,\infty)$ é uma fdp, observando que a mesma é uma distribuição Gamma e exibindo seus parâmetros.

ii - Com base em uma amostra aleatória desta fdp., encontre a **EMV** de θ.

1.7 i - Mostre que a função $f(x;\alpha,\beta) = \dfrac{1}{\beta}e^{-\frac{(x-\alpha)}{\beta}},\ x \geq \alpha,\ \alpha \in \mathbb{R}\, \beta > 0$ é uma fdp. Com base em uma amostra aleatória desta fdp., encontre a **EMV** de:

ii - α quando β é conhecido.

iii- β quando α é conhecido.

iv - β e α quando ambos são desconhecidos.

1.8 Em cada um dos seguintes casos, mostre que a distribuição da v.a. X é um membro da *família exponencial* de um parâmetro e identificando as funções $C(\theta)$, $Q(\theta)$ e $h(x)$ que a caracterizam.

i $\mathcal{X}$ tem distribuição de Poisson;

ii $\mathcal{X}$ tem distribuição Binomial Negativa;

iii $\mathcal{X}$ tem distribuição Gamma com β conhecido;

iv $\mathcal{X}$ tem distribuição Gamma com α conhecido;

v $\mathcal{X}$ tem distribuição Beta com β conhecido;

vi $\mathcal{X}$ tem distribuição Beta com α conhecido;

1.9 Em cada um dos seguintes casos, mostre que a distribuição do v.a X e o vetor aleatório $\boldsymbol{X}$ são da *família exponencial* multiparamétrica e identificando as várias funções que os caracterizam.

i $\mathcal{X}$ tem distribuição Gamma;

ii $\mathcal{X}$ tem distribuição Beta;

iii $\boldsymbol{X} = (\mathcal{X}_1, \mathcal{X}_2)$ tem distribuição Normal Bivariada;

1.4 Propriedades das EMV's

Teorema 1 ***(Invariância)*** *Seja $\widehat{\theta} = \widehat{\theta}(\boldsymbol{x})$ a **EMV** de θ com base nos valores observados $x_1, \ldots, x_n$ da amostra aleatória $X_1, \ldots, X_n$ de uma fdp $f(\cdot;\theta), \theta \in \Theta \subseteq \mathbb{R}$. Também, seja $\theta^* = g(\theta)$ uma função injetiva definida de Θ para $\Theta^* \subseteq \mathbb{R}$. Então a **EMV** de $\theta^*, \widehat{\theta^*}$, é dada por $\widehat{\theta^*}(\boldsymbol{x}) = g\left[\widehat{\theta}(\boldsymbol{x})\right]$.*

Prova: Seja $\Theta^* = g(\Theta)$ e $L^* : \Theta^* \to \mathbb{R}$ definida por $L^*(\theta^*) = L(g^{-1}(\theta^*)|\boldsymbol{x})$, a injetividade de g garante a existência da $g^{-1} : \Theta^* \to \Theta$. Se $L(\theta|\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i;\theta)$ possui um máximo, então:

$$L(\widehat{\theta}|\boldsymbol{x}) = \max L(\theta|\boldsymbol{x}) = \max L^*(\theta^*|\boldsymbol{x}) = L^*(\widehat{\theta^*}|\boldsymbol{x})$$

Da construção de L^* temos: $L(\theta|\boldsymbol{x}) = L^*(g(\theta)|\boldsymbol{x})$ para todo $\theta \in \Theta$, como $\widehat{\theta^*} = \widehat{g(\theta)}$, segue que $\widehat{g(\theta)} = g(\widehat{\theta})$ provando o resultado.

Exemplo 13 *Se $\widehat{\theta}$ é a **EMV** de uma v.a. $X \sim \mathcal{P}(\theta)$ e $g(\theta) = exp(-\theta)$ como $\widehat{\theta} = \overline{x}$, $\widehat{g(\theta)} = exp(-\overline{x})$ para $h(\theta) = \frac{1}{\theta}$ $\widehat{h(\theta)} = \frac{1}{\overline{x}}$ pois ambas g e h são funções injetivas. da mesma forma quando temos $X \sim \mathcal{N}(\mu, \sigma^2), \widehat{\sigma^2} = \dfrac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$, fazendo tomando $h(\theta) = \sqrt{\theta}$ temos $\sigma = \widehat{\sqrt{\sigma^2}} = \sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}$.*

Teorema 2 *Seja $\widehat{\theta} = \widehat{\theta}(\boldsymbol{x})$ a **EMV** de θ com base nos valores observados $x_1, \ldots, x_n$ da amostra aleatória $X_1, \ldots, X_n$ de uma fdp $f(\cdot;\theta), \theta \in \Theta \subseteq \mathbb{R}$. Também, seja $\theta^* = g(\theta)$ definida de Θ para $\Theta^* \subseteq \mathbb{R}$ tal que a imagem de g seja Θ^*. Então a **EMV** de $\theta^*, \widehat{\theta^*}$, é dada por $\widehat{\theta^*}(\boldsymbol{x}) = g\left[\widehat{\theta}(\boldsymbol{x})\right]$.*

Prova: Diferente do teorema anterior aqui não temos a hipótese da injetividade, logo pode existir em Θ vários θ's tais que $g(\theta) = \theta^* \in \Theta^*$, nesse caso vamos definir para cada $\theta^* \in \Theta^*$ o conjunto de nível:

$$\Theta_{\theta^*} = \{\theta \in \Theta; g(\theta) = \theta^*\}$$

e em seguida definimos a função de verossimilhança induzida por g $L^* : \Theta_{\theta^*} \to \mathbb{R}$ dada por:

$$L^*(\theta^*) = \sup\{L(\theta) \;\; \theta \in \Theta_{\theta^*}\}$$

Se $\widehat{\theta}$ é a **EMV** de θ existe um único $\widehat{\theta^*} \in \Theta^*$ tal que $g(\widehat{\theta}) = \widehat{\theta^*}$ e

$$L^*(\widehat{\theta^*}) = \sup\left\{L(\theta); \;\; \theta \in \Theta_{\widehat{\theta^*}}\right\} = \sup\left\{L(\theta); \; \theta \in \Theta; \; g(\theta) = g(\widehat{\theta})\right\} = L(\widehat{\theta})$$

Agora, para todo $\theta^* \in \Theta^*$

$$L^*(\theta^*) = \sup\{L(\theta) \;\; \theta \in \Theta_{\theta^*}\} \leq \max\{L(\theta); \;\; \theta \in \Theta_{\theta}\} = L(\widehat{\theta}) = L^*(\widehat{\theta^*})$$

Esta última desigualdade mostra que $\widehat{\theta^*} = g(\widehat{\theta})$, provando assim o resultado.

Exemplo 14 *Em termos de uma amostra aleatória de tamanho n, $X_1, \ldots, X_n$ de uma distribuição $X \sim B(1, \theta)$ com valores observados $x_1, \ldots, x_n$, determine a **EMV** $\widehat{\theta} = \widehat{\theta}(\boldsymbol{x})$ de $\theta \in (0, 1), \boldsymbol{x} = (x_1, ..., x_n)$. Sendo $f(x_i, \theta) = \theta^{x_i}(1-\theta)^{1-x_i}, x_i =$ 0 ou $1, i = 1, \ldots, n$, a função de verossimilhança é dada por:*

$$L(\theta|\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^t(1-\theta)^{n-t}, \ \ t = x_1 + \ldots + x_n$$

,portanto o ***EMV*** *de θ é $\widehat{\theta} = \frac{t}{n}$. Se $g(\theta) = \theta(1-\theta)$ determine $\widehat{g(\theta)}$. Nesse caso a função $g : (0,1) \to (0, \frac{1}{4})$ não é injetiva mas pelo teorema acima:*

$$\widehat{g(\theta)} = g(\widehat{\theta}) = \widehat{\theta}(1-\widehat{\theta}) = \tfrac{t}{n}(1-\tfrac{t}{n})\,.$$

Definição 4 *(**Suficiência**)*
Sejam $X_1, \ldots, X_n$ é uma amostra aleatória com fdp $f(\cdot;\theta), \theta \in \Theta \subseteq \mathbb{R}$, e seja $T = T(X_1, \ldots, X_n)$ uma estatística. (uma função conhecida dos X_i's). Então, se as distribuições condicionais dos X_i's, dado $T = t$, não depende de θ dizemos que T é uma estatística suficiente para para θ.

Teorema 3 *(**Fatorização de Fisher-Neyman**) Sejam $X_1, \ldots, X_n$ é uma amostra aleatória com fdp $f(;\theta), \theta \in \Theta \subseteq \mathbb{R}$, e seja $T = T(X_1, \ldots, X_n)$ uma estatística. Então, T é uma estatística suficiente para para θ se e somente se a fdp conjunta dos X_i's pode ser escrita da seguinte forma:*

$$f_{X_1,\ldots,X_n}(x_1, \ldots .x_n; \theta) = g\left[T(X_1, \ldots, X_n; \theta)\right] \times h\left(x_1, \ldots, x_n\right)$$

Prova: Não será feita neste curso.

Exemplo 15 *Se $X_1, \ldots, X_n \sim \mathcal{B}(1,p)$ e são v.a. ind. com f.d.p's*

$$f(x;\theta) = \theta^{x_i}(1-\theta)^{1-x_i} I_{\{0,1\}}(x_i), \ i = 1, \ldots, n$$

a distribuição conjunta de f é

$$L(\boldsymbol{\theta}) = \theta^t(1-\theta)^{n-t} \times \textstyle\prod_{i=1}^{n} I_{\{0,1\}}(x_i), \ \textit{onde } t = \sum_{i=1}^{n} x_i.$$

Fazendo $g\left(T(\boldsymbol{x}), \boldsymbol{\theta}\right) = \theta^t(1-\theta)^{n-t}$, e $h(\boldsymbol{x}) = \prod_{i=1}^{n} I_{\{0,1\}}(x_i)$ vemos que $L(\boldsymbol{\theta}) = g(T(\boldsymbol{x}), \boldsymbol{\theta}) \times h(\boldsymbol{x})$ e h independe de $\boldsymbol{\theta}$ portanto $T = \sum_{i=1}^{n} X_i = n\overline{X}$é uma estatística suficiente para θ.

Exemplo 16 *Se $X \sim \mathcal{P}(\theta)$ com $f(x,\theta) = e^{-\theta}\frac{\theta^x}{x!}, \ x = 0, 1 \ldots, \ \theta > 0$.*

$$L(\boldsymbol{\theta}) = e^{-n\theta} \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} = e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i} \frac{1}{\prod_{i=1}^{n} x_i!}$$

Concluímos que $f(\boldsymbol{x}, \theta) = g\left(T(\boldsymbol{x}), \theta\right) \times h(\boldsymbol{x})$ onde $g\left(T(\boldsymbol{x}, \theta)\right) = e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i}$ e $h(\boldsymbol{x}) = \frac{1}{\prod_{i=1}^{n} x_i!}$ portanto $T = \sum_{i=1}^{n} X_i = n\overline{X}$ é uma estatística Suficiente para θ.

Exemplo 17 *Se* $X \sim \mathcal{E}xp(\theta)$ *com* $f(x_i, \theta) = \theta e^{-\theta x_i}, \quad x_i > 0.$

$$f(\boldsymbol{x}, \theta) = \prod_{i=1}^{n} f(x_i, \theta) = L(\boldsymbol{\theta}) = \theta^n e^{-\left(\theta \sum_{i=1}^{n} x_i\right)}$$

Concluímos que $f(\boldsymbol{x}, \theta) = g(T(\boldsymbol{x}), \theta) \times h(\boldsymbol{x})$ *onde* $g(T(\boldsymbol{x}, \theta)) = \theta^n e^{-\theta\left(\sum_{i=1}^{n} x_i\right)}$ *e* $h(\boldsymbol{x}) = 1$ *portanto* $T = \sum_{i=1}^{n} X_i = n\overline{X}$ *é uma estatística Suficiente para* θ.

Exemplo 18 *Seja* $X_1, \ldots, X_n$ *uma amostra aleatória de uma população com distribuição* $\mathcal{N}(\mu, \sigma^2)$:

$$L(\mu, \sigma; \boldsymbol{x}) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}}, \mu \in \mathbb{R}, \sigma > 0.$$

Se μ *é desconhecido*

$$L(\mu; \boldsymbol{x}) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{\sum_{i=1}^{n}\left(x_i^2 - 2x_i\mu + \mu^2\right)}{2\sigma^2}} = e^{-\frac{-2\mu\left(\sum_{i=1}^{n} x_i + \mu\right)}{2\sigma^2}} \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{\sum_{i=1}^{n} x_i^2}{2\sigma^2}}, \mu \in \mathbb{R}.$$

Observamos que $T(\boldsymbol{x}) = \sum_{i=1}^{n} x_i$ *e também* $\overline{X}$ *são estatísticas suficientes para* μ.

Se σ^2 *é desconhecido*

$$L(\sigma^2; \boldsymbol{x}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}} \times 1.$$

$T(\boldsymbol{x}) = \sum_{i=1}^{n}(x_i - \mu)^2$ *e também* $T(\boldsymbol{x}) = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$ *são estatísticas suficientes para* σ^2.

Finalmente, se μ *e* σ^2 *é desconhecidos*

$$\begin{aligned} L(\mu, \sigma^2; \boldsymbol{x}) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{\sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\mu)^2}{2\sigma^2}} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu)(\sum_{i=1}^{n} x_i - n\bar{x}) - n(\bar{x}-\mu)^2}{2\sigma^2}} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{\left(-\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{2\sigma^2} - n\frac{(\bar{x}-\mu)^2}{2\sigma^2}\right)} \end{aligned}$$

As estatísticas $\left(\overline{X}, \sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right)$ *formam um par de estatísticas suficientes para o par* (μ, σ^2).

Exemplo 19 *Com base em uma amostra aleatória de tamanho* $n, X_1, \ldots, X_n$, *de cada uma das f.d.p's dadas abaixo com valores observados* $x_1, \ldots, x_n$, *determine uma estatística suficiente estatística para* θ

i- $f(x;\theta) = \frac{\theta}{x^{\theta+1}},\ x \geq 1,\ \theta \in \Theta = (0,\infty)\,.$

ii- $f(x;\theta) = \frac{x}{\theta}e^{-\frac{x^2}{2\theta}},\ x > 0,\ \theta \in \Theta = (0,\infty)\,.$

iii- $f(x;\theta) = (1+\theta)\,x^{\theta},\ 0 < x < 1,\ \theta \in \Theta = (-1,\infty)\,.$

iv- $f(x;\theta) = \frac{\theta}{x^2},\ x \geq 0,\ \theta \in \Theta = (0,\infty)\,.$

.

Solução *Se* $x_{min} = \min\{x_1,\ldots,x_n\}$ *e* $x_{max} = \max\{x_1,\ldots,x_n\}$

i- $f(x;\theta) = \frac{\theta}{x^{\theta+1}}$ *logo*

$$\begin{aligned} f(\boldsymbol{x},\theta) &= \prod_{i=1}^{n} f(x_i,\theta) = \prod_{i=1}^{n}\frac{\theta}{x_i^{\theta+1}}\prod_{i=1}^{n} I_{(1,\infty)}(x_i) \\ &= \frac{\theta^n}{\prod_{i=1}^{n} x_i^{\theta}}\frac{1}{\prod_{i=1}^{n} x_i} I_{(1,\infty)}(x_{min}). \end{aligned}.$$

Portanto $T(\boldsymbol{X}) = \prod_{i=1}^{n} X_i$ *é uma estatística suficiente para* θ*.*

ii- *Se* $f(x;\theta) = \frac{x_i}{\theta}e^{-\frac{x_i^2}{2\theta}} I_{(1,\infty)}(x)$ *logo*

$$\begin{aligned} f(\boldsymbol{x};\theta) &= \prod_{i=1}^{n} f(x_i,\theta) \\ &= \prod_{i=1}^{n}\frac{x_i}{\theta}e^{-\frac{x_i^2}{2\theta}}\prod_{i=1}^{n} I_{(1,\infty)}(x_i) \\ &= \frac{e^{-\frac{\sum_{i=1}^{n} x_i^2}{2\theta}}}{\theta^n}\times\prod_{i=1}^{n} x_i I_{(1,\infty)}(x_{min}). \end{aligned}.$$

Portanto $T(\boldsymbol{X}) = \sum_{i=1}^{n} X_i^2$ *é uma estatística suficiente para* θ*.*

iii- *Se* $f(x;\theta) = (1+\theta)\,x^{\theta} I_{(0,1)}$ *logo*

$$\begin{aligned} f(\boldsymbol{x};\theta) &= \prod_{i=1}^{n} f(x_i,\theta) \\ &= \prod_{i=1}^{n}(1+\theta)^n\prod_{i=1}^{n} x_i^{\theta}\prod_{i=1}^{n} I_{(0,1)}(x_i) \\ &= \prod_{i=1}^{n}(1+\theta)^n\left(\prod_{i=1}^{n} x_i\right)^{\theta} I_{(0,1)}(x_{min})\times I_{(0,1)}(x_{max}). \end{aligned}.$$

Portanto $T(\boldsymbol{X}) = \prod_{i=1}^{n} X_i$ *é uma estatística suficiente para* θ*.*

iv- *Se* $f(x;\theta) = \frac{\theta}{x_i^2} I_{(\theta,\infty)}(x)$ *logo*

$$f(\boldsymbol{x};\theta) = \prod_{i=1}^{n} f(x_i,\theta) = \prod_{i=1}^{n} \frac{\theta}{x_i^2} \prod_{i=1}^{n} I_{(\theta,\infty)}(x_i) = \theta^n I_{(\theta,\infty)}(x_{min}) \frac{1}{\prod_{i=1}^{n} x_i^2}.$$

Portanto $T(\boldsymbol{X}) = X_{min} = \min\{X_1,\ldots,X_n\}$ *é uma estatística suficiente para* θ*.*

.

Teorema 4 *Se* $\widehat{\theta_n} = \widehat{\theta_n}(X_1,\ldots,X_n)$ *é a* ***EMV*** *de* $\theta \in \Theta \subseteq \mathbb{R}$ *baseada numa amostra aleatóra* $(X_1,\ldots,X_n)$ *com fdp* $f(\cdot;\theta)$*. Então, sob certas condições de regularidade* $\left\{\widehat{\theta_n}\right\}$ *é consistente no sentido de probablidade, isto é:* $\widehat{\theta_n} \to \theta$ *em* $P_\theta-$ *Probabilidade quando* $n \to \infty$

Prova: Não será feita neste curso.

Teorema 5 *Se* $\widehat{\theta_n} = \widehat{\theta_n}(X_1,\ldots,X_n)$ *é a* ***EMV*** *de* $\theta \in \Theta \subseteq \mathbb{R}$ *baseada numa amostara aleatóra* $X_1,\ldots,X_n$ *com fdp* $f(\cdot;\theta)$*. Então, sob certas condições de regularidade* $\left\{\widehat{\theta_n}\right\}$ *é assintoticamente normal, isto é:*

$$\sqrt{n}\left(\widehat{\theta_n} - \theta\right) \xrightarrow{d} \mathcal{N}(0,\sigma_\theta^2)$$

onde

$$\sigma_\theta^2 = 1/I(\theta) e I(\theta) = E_\theta\left[\frac{\partial log(f(X;\theta))}{\partial\theta}\right]^2, \quad X \sim f(;\theta).$$

Prova: Não será feita neste curso.

1.4.1 Exercícios

1- Se $X_1,\ldots,X_n$ são v.a's i.i.d tendo fdp com distribuição exponencial $f(x,\theta) = \theta e^{-\theta x},\ x > 0, \theta \in \Theta = (0,\infty)$. Então:

a Mostre que $1/\overline{X}$ é uma **EMV** de θ;

b Use Teorema 1 com o objetivo de provar que **EMV** de $\theta*$ na forma parametrizada $f(x;\theta*) = \frac{1}{\theta*} e^{-x/\theta*} x > 0$, é $\overline{X}$.

2- Seja X uma v.a que denota a vida útil de um equipamento. Então a confiabilidade do equipamento no tempo $x, R(x)$, é definido como a probabilidade de que $X > x$, isto é $R(x) = P(X > x)$. Suponha agora que X tem fdp com distribuição Exponencial $f(x, \theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}},\ x > 0, \theta \in \Theta = (0, \infty)$. Então:

a Calcule a confiabilidade de $R(x, \theta)$ com base nesta v.a X ;

b Use Teorema 1 com o objetivo de encontrar que **EMV** de $R(x, \theta)$ com base de uma amostra aleatória $X_1, \ldots, X_n$ da fdp. dada.

3- Seja X uma v.a. descrevendo a vida útil de um determinado equipamento, e suponha que a f.d.p. de X é $f(x, \theta) = \theta e^{-\theta x},\ x > 0, \theta \in \Theta = (0, \infty)$. Então:

a Mostre que a probabilidade de que X seja maior ou igual ao tempo t unidades é $g(\theta) = e^{-t\theta}$;

b Sabemos (ver Exercício 1) que a **EMV** de θ, com base em uma amostra de tamanho n da f.d.p. acima, é $\widehat{\theta} = 1/\overline{X}$. Então determine a **EMV** de $g(\theta)$.

4- Considere n variaveis aleatórias independentes e identicamentes distribuídas com fdp com distribuição Weibull $f(x, \theta) = \frac{\gamma}{\theta}x^{\gamma-1}e^{-\frac{x^\gamma}{\theta}},\ x > 0, \gamma > 0$ conhecido , $\theta \in \Theta = (0, \infty)$

a Mostre que a $\widehat{\theta \frac{\sum_{i=1}^n X_i^\gamma}{n}}$ é a **EMV** de θ ;

b Tome $\gamma = 1 = 1$ e relacione o resultado na parte (a) ao resultado do Exercício 1 item (b).

5- Se $X_1, \ldots, X_n$ é uma amostra aleatória de tamanho n de uma distribuição normal $\mathcal{N}\left(\mu, \sigma^2\right)$ onde μ, σ são desconhecidos. Se $\boldsymbol{\theta} = \left(\mu, \sigma^2\right)$ e $0 < p < 1$ é um número conhecido. Então:

a Mostre que o ponto c para o qual temos $P_\theta\left(\overline{X} \leq c\right) = p$ é dado por: $c = \mu + \frac{\sigma}{n}\phi^{-1}(p)$;

b Dado que as **EMV**'s de μ e σ^2 são $\widehat{\mu} = \overline{X}$ e $\widehat{\sigma^2} = S^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$, determine a **EMV** de c, chame ela de $\widehat{c}$.

c Expresse $\widehat{c}$ em termos de x_i se $n = 25$ e $p = .95$

6-

a Mostre que a $f(x; \theta) = \frac{\theta}{x^{\theta+1}},\ x \geq 1,\ \theta \in \Theta = (0, \infty)$. é uma fdp. ;

b Com base em uma amostra aleatória de tamanho n deste p.d.f., mostre que a estatística $\prod_{i=1}^n X_i$ é suficiente para θ, assim como a estatística $\sum_{i=1}^n \log(X_i)$

7- Se X é uma v.a. tendo fdp com dist. Geométrica $f(x;\theta) = \theta(1-\theta)^{x-1}, x = 1,2,\ldots,\theta \in \Theta = (0,1)$. Mostre que $\overline{X}$ é uma estatística suficiente para θ.

1.5 Estimativas Imparciais de Variancia Mínima Uniforme

Talvez o segundo método mais popular de estimar um parâmetro seja com base nos conceitos de imparcialidade e variância mínima. Este método será discutido aqui até certo ponto e também será ilustrado por exemplos.

Para iniciar, seja $X_1,\ldots,X_n$ seja uma amostra aleatória com fdp.$f(;\theta), \theta \in \Theta \subseteq \mathbb{R}$, e vamos introduzir a notação $U = U(X_1,\ldots,X_n)$ para uma estimativa de θ.

Definição 5
*Uma estimativa de U é dita **Imparcial** se $E_\theta U = \theta$ para todo $\theta \in \Theta$.*

A seguir apresesentamos alguns exemplos de estimativas imparciais.

Exemplo 20 *Sejam $X_1,\ldots,X_n$ uma amostra aleatória tendo uma das seguintes distribuições:*

(i) *$X_i \sim B(1,\theta), \theta \in (0,1)$, $\overline{X}$ é uma estimativa imparcial para θ.*

(ii) *$X_i \sim \mathcal{P}(\theta), \theta > 0$ $\overline{X}$ e $Var_\theta(X_1)$ são estimativas imparciais para θ.*

(iii) *Se $X_i \sim \mathcal{N}(\theta,\sigma^2)$, com σ conhecido , novamente $\overline{X}$ é uma estimativa imparcial para θ.*

(iv) *Se $X_i \sim \mathcal{N}(\mu,\theta)$, com $\sigma > 0$, μ conhecido, então a variância amostral $S^2 = \frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2$ é uma estimativa imparcial para θ.*

(v) *Se $X_i \sim \Gamma(\theta,\beta), \beta = 1$, $\overline{X}$ é uma estimativa imparcial para θ.*

(vi) *Se $X_i \sim \Gamma(\alpha,\theta), \alpha = 1$, $\overline{X}$ é uma estimativa imparcial para θ.*

Solução

(i) *Se $X_i \sim B(1,\theta), \theta \in (0,1)$, $E[X_i] = \theta$ logo*
$E_\theta\left[\overline{X}\right] = E_\theta\left[\frac{\sum_{i=1}^n X_i}{n}\right] = \frac{\sum_{i=1}^n E[X_i]}{n} = \frac{n\theta}{n} = \theta.$
Portanto $\overline{X}$ é uma estimativa imparcial para θ.

(ii) *Se $X_i \sim \mathcal{P}(\theta), \theta > 0$ $E[X_i] = \theta$ logo*
$E_\theta\left[\overline{X}\right] = E_\theta\left[\frac{\sum_{i=1}^n X_i}{n}\right] = \frac{n\theta}{n} = \theta$ *e $Var_\theta(X_1)$ são estimativas imparciais para θ.*

(iii) *Se* $X_i \sim \mathcal{N}(\theta, \sigma^2)$*, com* σ *conhecido, novamente* $E[X_i] = \theta$ *logo* $E_\theta\left[\overline{X}\right] = E_\theta\left[\frac{\sum_{i=1}^n X_i}{n}\right] = \frac{n\theta}{n} = \theta$*. Portanto* $\overline{X}$ *é uma estimativa imparcial para* θ*.*

(iv) *Se* $X_i \sim \mathcal{N}(\mu, \theta)$*, com* $\sigma > 0$*,* μ *conhecido, então* $\left(\frac{X_i-\mu}{\sqrt{\theta}}\right)^2 \sim \chi_1^2$ *consequentemente* $E_\theta\left[\left(\frac{X_i-\mu}{\sqrt{\theta}}\right)^2\right] = 1$ *logo*

$$\begin{aligned} E_\theta\left[S^2\right] &= E_\theta\left[\frac{\sum_{i=1}^n (X_i-\mu)^2}{n}\right] = \theta E_\theta\left[\frac{1}{n}\sum_{i=1}^n \left(\frac{X_i-\mu}{\sqrt{\theta}}\right)^2\right] \\ &= \frac{\theta}{n}\sum_{i=1}^n E_\theta\left[\left(\frac{X_i-\mu}{\sqrt{\theta}}\right)^2\right] = \frac{\theta n}{n} = \theta. \end{aligned}.$$

Portanto S^2 *é uma estimativa imparcial para* θ*.*

(v) *Se* $X_i \sim \Gamma(\theta, \beta), \beta = 1$*,* $\overline{X}$ *é uma estimativa imparcial para* θ*.*

(vi) *Se* $X_i \sim \Gamma(\alpha, \theta), \alpha = 1$*,* $\overline{X}$ *é uma estimativa imparcial para* θ*.*

Definição 6 *Uma estimativa* ***Imparcial*** $U = U(X_1, \ldots, X_n)$ *de* θ *é dita Imparcial de Variância Mínima Uniforme (EIVMU), se para qualquer outra estimativa* ***imparcial*** $V = V(X_1, \ldots, X_n)$ *a seguinte desigualdade é válida:*

$$Var_\theta(U) \leq Var_\theta(V) \text{ para todo } \theta \in \Theta.$$

Teorema 6 ***Desigualdade de Cramer-Rao*** *seja* $X_1, \ldots, X_n$ *seja uma amostra aleatória com fdp.* $f(;\theta), \theta \in \Theta \subseteq \mathbb{R}$ *e suponha que certas condições de regularidade são conhecidas. Então, para qualquer estimativa* ***imparcial*** $U = U(X_1, \ldots, X_n)$ *de* θ *a seguinte desigualdade é válida:*

$$Var_\theta(U) \geq \frac{1}{nI(\theta)} \text{ para todo } \theta \in \Theta. \tag{1.1}$$

onde

$$I(\theta) = E_\theta\left[\frac{\partial log(f(X;\theta))}{\partial\theta}\right]^2, \quad X \sim f(;\theta). \tag{1.2}$$

Observação 1 *As condições não especificadas mencionadas na formulação do teorema inclui a suposição de que o domínio de x na f.d.p.* $f(x;\theta)$ *não depende de* θ*; assim, a distribuição* $U(0,\theta)$*, por exemplo, é deixada de fora. Além disso, as condições incluem a validade de intercambiar as operações de diferenciação e integração em certas expressões.*

A quantidade $I(\theta)$ é denominada *Informação de Fisher* fornecida pela amostra aleatória $X_1, \ldots, X_n$ sobre o parâmetro θ. Uma justificativa para a "informação"decorre do fato de que $\sigma^2_{\widehat{\theta}} = 1/I(\theta)$, de modo que quanto maior $I(\theta)$ menor é a variância $\sigma^2_{\widehat{\theta}}$ e, portanto $\widehat{\theta}$ é mais concentrado em torno de θ. O oposto acontece para pequenos valores de $I(\theta)$.

Observação 2 *Pode-se mostrar que, em condições adequadas, a quantidade $I(\theta)$ em (1.2) também pode ser calculado da seguinte forma:*

$$I(\theta) = -E_\theta\left[\frac{\partial^2 log(f(X;\theta))}{\partial\theta^2}\right]. \tag{1.3}$$

Essa expressão geralmente é mais fácil de calcular. A desigualdade de Cramér-Rao é usada da seguinte maneira.

(i) *Calcule as informação de Fisher por meio de (1.2) ou por meio de (1.3);*

(ii) *Calcule o limite inferior de Cramér-Rao (C-R) apresentado em (8);*

(iii) *Tente identificar uma estimativa imparcial cuja variância é igual ao C-R limite inferior de C-R(para todos $\theta \in \Theta$). Se tal estimativa for encontrada;*

(iv) *Declare a estimativa descrita em (iii) como a estimativa UMVU de θ.*

O uso da desigualdade será ilustrado por dois exemplos; outros casos são deixados como exercícios.

Exemplo 21 *Sejam $X_1, \ldots, X_n, \sim B(1,\theta)$ $f(x_i,\theta) = \theta^{x_i}(1-\theta)^{1-x_i}, x_i = 0$ ou $1, i =$ $1, \ldots, n,$*
$(i) \log(f(X,\theta) = X\log(\theta) + (1-X)\log(1-\theta)$;

$$\frac{\partial log(f(X;\theta))}{\partial\theta} = \frac{X}{\theta} - \frac{1-X}{1-\theta} \; e \; \frac{\partial^2 log(f(X;\theta))}{\partial\theta^2} = -\frac{X}{\theta} - \frac{1-X}{(1-\theta)^2}$$

$$I(\theta) = -E_\theta\left[\frac{\partial^2 log(f(X;\theta))}{\partial\theta^2}\right] = \frac{X}{\theta^2} + \frac{1-X}{(1-\theta)^2} = \frac{1}{\theta(1-\theta)}$$

(ii) *O limite inferior de Cramér-Rao (C-R) é dado por $\frac{1}{nI(\theta)} = \frac{\theta(1-\theta)}{n}$;*

(iii) *Se $U = \overline{X}$ como $X \sim B(1,\theta)$ $E_\theta\left[\overline{X}\right] = \theta$, $\sigma^2\left[\overline{X}\right] = \frac{\sum_{i=1}^n \sigma^2[X_i]}{n^2} = \frac{n\theta(1-\theta)}{n^2} = \frac{1}{nI(\theta)}$*

(iv) *$U = \overline{X}$ é uma EIVMU de θ.*

Exemplo 22 *Sejam $X_1, \ldots, X_n$, v.a i.id $\sim P(\lambda), \lambda > 0$ fazendo $\lambda = \theta$ temos:*

$$f(x;\theta) = e^{-\theta}\frac{\theta^k}{k!}, k = 0, \ldots, \quad \log\left(f(x;\theta)\right) = -\theta + k\log(\theta) - \log(k!)$$

Então

$$\frac{\partial log(f(X;\theta))}{\partial\theta} = -1 + \frac{k}{\theta} \quad e$$

$$\left[\frac{\partial log(f(X;\theta))}{\partial\theta}\right]^2 = 1 + \frac{k^2}{\theta^2} - \frac{2k}{\theta}$$

Como $E_\theta\left[X\right] = \theta$ e $E_\theta\left[X^2\right] = \theta(1+\theta)$

$$I(\theta) = E_\theta\left[\frac{\partial log(f(X;\theta))}{\partial\theta}\right]^2 = E_\theta\left[1 + \frac{k^2}{\theta^2} - \frac{2k}{\theta}\right] = \frac{1}{\theta}$$

Como $\frac{1}{nI(\theta)} = \frac{\theta}{n}$ é o limite de Cramér-Rao, $\overline{X}$, é uma Estatística Suficiente para θ e $Var_\theta\overline{X} = \frac{\theta}{n}$, $\overline{X}$ é a EIVMU de θ.

Exemplo 23 *Se $X \sim Exp(1/\theta)$ isto é $f(X,\theta) = \frac{1}{\theta}e^{-\frac{X}{\theta}}$, então $E_\theta\left[X\right] = \theta$ e $\sigma^2\left[X\right] = \theta^2$;*

$$\begin{aligned}
\textit{i}\ \log(f(X,\theta) &= -\log(\theta) - \tfrac{X}{\theta} \\
\frac{\partial log(f(X;\theta))}{\partial\theta} &= -\frac{1}{\theta} + \frac{X}{\theta^2}, \quad e \quad \frac{\partial^2\log(f(X;\theta))}{\partial\theta^2} = \frac{1}{\theta^2} - \frac{2X}{\theta^3} \\
\textit{ii}\ I(\theta) &= -E_\theta\left[\frac{\partial^2\log(f(X;\theta))}{\partial\theta^2}\right] \\
&= E_\theta\left[\frac{1}{\theta^2} - \frac{2X}{\theta^3}\right] = -\frac{1}{\theta^2} + \frac{2}{\theta^2} = \frac{1}{\theta^2}
\end{aligned}$$

iii *A cota inferior C-R é dada por $\frac{1}{nI(\theta)} = \frac{\theta^2}{n}$*

iv *Considere $U = \overline{X}$ $\sigma^2\left[U\right] = \frac{\theta^2}{n} = \frac{1}{nI(\theta)}$. Portanto $\overline{X}$ é uma EIVMU de θ.*

Existe uma maneira alternativa de procurar EIVMU, em particular, quando a abordagem por meio da desigualdade de Cramér-Rao deixa de produzir tal estimativa. Esta abordagem depende fortemente do conceito de suficiência já introduzido e também de um conceito técnico adicional, denominado completude.

Definição 7 *(*Completude*) Seja T uma var. aleatória com fdp $f_T(\cdot;\theta)$, $\theta \in \Theta \subseteq \mathbb{R}$, a família $\{f_T(\cdot;\theta), \theta \in \Theta\}$ ou a var. aleatória T é dita* **completa** *se para $h : \mathbb{R} \to \mathbb{R}$ tal que $E_\theta h(T) = 0$ para todo $\theta \in \Theta$ então $h(t) = 0$.*

Teorema 7 ***(Rao-Blackwell, Lehmann-Scheffé)*** *Se $X_1, \ldots, X_n$ é uma amostra aleatória com fdp $f(\cdot;\theta)$, $\theta \in \Theta \subseteq \mathbb{R}$ e $T = T(X_1, \ldots, X_n)$ é uma estatística suficiente para θ e completa. Se $U = U(X_1, \ldots, X_n)$ é uma estatística* imparcial *de θ defina a estatística*

$$\phi(T) = E_\theta (U|T)$$

Então

i *A var. aleat. $\phi(T)$ é função apenas da estatística suficiente T;*

ii *$\phi(T)$ é uma estatística* imparcial *de θ;*

iii $\sigma^2_\theta [\phi(T)] < \sigma^2_\theta [U]$, $\theta \in \Theta$ *provado que* $E_\theta U^2 < \infty$

iv *$\phi(T)$ é única EIVMUde θ.*

Os exemplos a seguir ilustram como aplicar o Teorema de **Rao-Blackwell, Lehmann-Scheffé** em casos concretos.

Exemplo 24 *Determine a EIVMUde θ com base na amostra aleatória $X_1, \ldots, X_n$ da distribuição $P(\theta)$.*

SOLUÇÃO: Já vimos nos exemplos 16 e 20 que se $X \sim P(\theta)$, $T = X_1 + \ldots + X_n$ e X_1 são estatistica suficiente e imparcial de θ respectivamente. Vamos construir agora a distribuição condicional de X_1 dado $T = t$.

$$
\begin{aligned}
P\left(X_1 = x \,\middle|\, T = t\right) &= P\left(X_1 = x \,\middle|\, X_2 + \ldots + X_n = t - x\right) \\
&= \frac{P\left(X_1 = x; X_2 + \ldots + X_n = t - x\right)}{P\left(X_1 + \ldots + X_n = t\right)} \\
&= \frac{P\left(X_1 = x; X_2 + \ldots + X_n = t - x\right)}{\sum_{x=0}^{t} P\left(X_1 = x\right) P\left(X_2 + \ldots + X_n = t\right)} \\
&= \frac{e^{-\theta}\theta^{x}}{x!} \times \frac{e^{-(n-1)\theta}\left((n-1)\theta\right)^{t-x}}{(t-x)!} \\
&\times \frac{t!}{t! \sum_{x=0}^{t} P\left(X_1 = x\right) P\left(X_2 + \ldots + X_n = t\right)} \\
&= \frac{e^{-\theta}\theta^{x}}{x!} \times \frac{e^{-(n-1)\theta}\left((n-1)\theta\right)^{t-x}}{(t-x)!} \\
&\times \frac{t!}{e^{-\theta}e^{-(n-1)\theta} \sum_{x=0}^{t} t! \frac{\theta^{x}}{x!} \frac{((n-1)\theta)^{t-x}}{(t-x)!}} \\
&= \frac{t!\theta^{x}\left((n-1)\theta\right)^{t-x}}{x!\,(t-x)!\,(n\theta)^{t}} = \frac{t!}{x!\,(t-x)!}\left(\frac{\theta}{n\theta}\right)^{x}\left(\frac{(n-1)\theta}{n\theta}\right)^{t-x} \\
&= \frac{t!}{x!\,(t-x)!}\left(\frac{1}{n}\right)^{x}\left(1 - \frac{1}{n}\right)^{t-x} \sim B\left(t, \frac{1}{n}\right)
\end{aligned}
$$

Logo $E_\theta\left(X_1 = x \,\middle|\, T = t\right) = \frac{t}{n}$ *então a* $\phi(T) = E_\theta\left(X_1 = x \,\middle|\, T = t\right) = \frac{t}{n} = \overline{X}$ *portanto* $\overline{X}$ *é a EIVMU de* θ*.*

Exemplo 25 *Seja* $X_1, \ldots, X_n$ *é uma amostra aleatória da distribuição* $N(\mu, \sigma^2)$*. Suponha que* σ^2 *é conhecido e façamos* $\mu = \theta$*, então*

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

e portanto

$$\log\left(f(x;\theta)\right) = \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{(x-\theta)^2}{2\sigma^2}.$$

Logo,

$$\frac{\partial \log(f(x;\theta))}{\partial\theta} = \frac{1}{\sigma}\frac{x-\theta}{\sigma},$$

$$\left[\frac{\partial \log(f(x;\theta))}{\partial\theta}\right]^2 = \frac{1}{\sigma^2}\left(\frac{x-\theta}{\sigma}\right)^2,$$

$$I(\theta) = E_\theta\left[\frac{\partial \log(f(X;\theta))}{\partial\theta}\right]^2 = \frac{1}{\sigma^2}E_\theta\left(\frac{X-\theta}{\sigma}\right)^2 = \frac{1}{\sigma^2}$$

Pois $\left(\frac{X-\theta}{\sigma}\right)^2 \sim \chi_1^2$ *logo* $E_\theta\left(\frac{X-\theta}{\sigma}\right)^2 = 1$*, consequentemente o limite inferior de Cramér-Rao é dado por* $\frac{\sigma^2}{n}$*. Calculando,* $Var\left(\overline{X}\right) = \frac{\sigma^2}{n}$*, como* $\overline{X}$ *é uma estatística imparcial para* θ *pelo Teorema Cramér-Rao* $\overline{X}$ *é uma EIVMU de* θ *quando* σ *é conhecido.*

Suponha que μ *é conhecido e façamos* $\sigma^2 = \theta$*, então*

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

e portanto

$$\log(f(x;\theta)) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\theta) - \frac{(x-\mu)^2}{2\theta}.$$

Logo,

$$\frac{\partial \log(f(x;\theta))}{\partial\theta} = -\frac{1}{2\theta} + \frac{(x-\mu)^2}{2\theta^2},$$

$$\begin{aligned}
\left[\frac{\partial \log(f(x;\theta))}{\partial\theta}\right]^2 &= \frac{1}{4\theta^2} - 2\frac{1}{2\theta}\frac{(x-\mu)^2}{2\theta^2} + \left(\frac{(x-\mu)^2}{2\theta^2}\right)^2 \\
&= \frac{1}{4\theta^2} - \frac{1}{2\theta^2}\left(\frac{x-\mu}{\sqrt{\theta}}\right)^2 + \frac{1}{4\theta^2}\left(\frac{x-\mu}{\sqrt{\theta}}\right)^4 \\
I(\theta) = E_\theta\left[\frac{\partial \log(f(X;\theta))}{\partial\theta}\right]^2 &= E_\theta\left[\frac{1}{4\theta^2} - \frac{1}{2\theta^2}\left(\frac{x-\mu}{\sqrt{\theta}}\right)^2 + \frac{1}{4\theta^2}\left(\frac{x-\mu}{\sqrt{\theta}}\right)^4\right] \\
&= \frac{1}{4\theta^2} - \frac{1}{2\theta^2}E_\theta\left(\frac{x-\mu}{\sqrt{\theta}}\right)^2 + \frac{1}{4\theta^2}E_\theta\left(\frac{x-\mu}{\sqrt{\theta}}\right)^4 \\
&= \frac{1}{4\theta^2} - \frac{1}{2\theta^2} + \frac{3}{4\theta^2} = \frac{1}{2\theta^2}
\end{aligned}$$

Pois $E_\theta\left(\frac{x-\mu}{\sqrt{\theta}}\right)^2 = 1$ *e* $E_\theta\left(\frac{x-\mu}{\sqrt{\theta}}\right)^4 = 3$ *devido ao fato que* $Z = \frac{x-\mu}{\sqrt{\theta}} \sim \mathcal{N}(0,1)$, $E\left[Z^{2n}\right] = \frac{(2n)!}{2^n(n!)}$. *Como* μ *é conhecido vamos considerar agora a estatística:*

$$W = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sqrt{\theta}}\right)^2, \quad W \sim \chi_n^2 \;\; E[W] = n, \;\; Var[W] = 2n$$

Se $U = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$, $U = \frac{\theta}{n}W$, $E[U] = \theta$ $Var[W] = \frac{\theta^2}{n^2}Var[W] = \frac{2\theta^2}{n}$. *Observe que* U *é imparcial e* $Var[W] = \frac{1}{nI(\theta)}$ *portanto* U *é uma EIVMU de* θ *quando* μ *é conhecido.*

Vejamos agora o caso onde μ *e* σ *são desconhecidos. Fazendo* $\mu = \theta_1$ *e* $\sigma^2 = \theta_2$. *Suponha que estejamos interessados ??em encontrar um estimador EIVMU de* θ_2. *Já vimos que* $(\overline{X}, S^2)$ *são estatística sificientes e imparaciais de* μ, σ *respectivamente. Calculando o limite de Cramér-Rao, considerando* θ_1 *constante temos:*

$$\frac{1}{nI(\theta)} = \frac{2\theta_2^2}{n}$$

Calculando

$$\begin{aligned} Var_{\theta_2}\left(S^2\right) &= Var_{\theta_2}\left(\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right) \\ &= \frac{1}{(n-1)^2}\sum_{i=1}^{n}Var\left(\left(X_i-\overline{X}\right)^2\right) \\ &= \frac{\theta_2}{(n-1)^2}Var\left(\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sqrt{\theta_2}}\right)^2\right) = \frac{2\theta_2^2(n-1)}{(n-1)^2} = \frac{2\theta_2^2}{(n-1)} > \frac{2\theta_2^2}{n} \end{aligned}$$

$$W = \textstyle\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sqrt{\theta_2}}\right)^2 \sim \chi_{n-1}^2 \text{ logo } Var(W) = 2(n-1)$$

Conclusão: S^2 *é uma EIVMU no entanto a desigualdade de Cramér Rao não é satisfeita como igualdade.*

1.6 Exercícios

Capítulo 2

Intervalos de Confiança e Regiões de Confiança

2.1 A Ideia Básica da Estimação Intervalar

Suponha que estejamos interessados em construir uma estimativa pontual da média μ em uma distribuição normal $N(\mu, \sigma^2)$ com variância conhecida; isso deve ser feito com base em uma amostra aleatória de tamanho $n, X_1, \ldots, X_n$, extraído da subjacente distribuição. Isso equivale a construir uma estatística adequada dos X_i, chamada $V = V(X_1, ..., X_n)$, que para os valores observados x_i de $X_i, i = 1, \ldots, n$ é uma entidade numérica e declara-a como o valor (desconhecido) de μ. Isso parece um tanto presunçoso, uma vez que, a partir do conjunto de valores possíveis para $\mu, -\infty < \mu < \infty$, apenas um é selecionado como seu valor. Pensando assim, pode ser mais razoável visar, em vez disso, um intervalo aleatório que conterá o valor (desconhecido) de μ com alta probabilidade (prescrita). Isso é exatamente o que um intervalo de confiança faz.

Para ser mais preciso, ao lançar o problema em um cenário geral,seja $X_1, \ldots, X_n$ uma amostra aleatória de uma f.d.p. $f(;\theta), \theta \in \Theta \subseteq \mathbb{R}$, $L = L(X_1, ..., X_n)$ e $U = U(X_1, ..., X_n)$ duas estatísticas de X_i, tais que $L < U$. Então o intervalo com limites L e U, $[L, U]$, é chamado de um intervalo aleatório. Seja α um pequeno número em (0, 1), como 0,005, 0,01, 0,05, e suponha que o intervalo aleatório $[L, U]$ contém θ com probabilidade igual para $1 - \alpha$ (como 0,995, 0,99, 0,95), independentemente do valor verdadeiro de θ. Em outras palavras, suponha que:

$$P\theta(L \leq \theta \leq U) = 1 - \alpha \text{ para todo } \theta \in \Theta\ . \ (1)$$

Se a relação (1) for mantida, então dizemos que o intervalo aleatório $[L, U]$, é um intervalo de confiança para θ com nível de confiança $1 - \alpha$.

A interpretação da significância de um intervalo de confiança é baseada na interpretação da frequência relativa do conceito de probabilidade, da seguinte forma: Suponha que n v.a's independentes sejam extraídos de uma f.d.p. $f(;\theta)$, e $x_1, ..., x_n$ sejam seus valores observados. Além disso, seja $[L_1, U_1]$ o intervalo resultante a partir dos valores observados de $L = L(X_1, \ldots, X_n)$ e $U = U(X_1, \ldots, X_n)$; isto é, $L_1 = L(x_1, \ldots, x_n)$ e $U_1 = U(x_1, \ldots, x_n)$. Prossiga para constuir de forma independente um segundo conjunto de r.v.a.'s como acima, e seja $[L_2, U_2]$ o intervalo resultante. Repita este processo independentemente um grande número de vezes, N, digamos, com o intervalo correspondente sendo $[L_N, U_N]$. Então, a interpretação de (1) é que, em média, cerca de $100(1-\alpha)\%$ dos N intervalos acima irão, na verdade, contém o valor verdadeiro de θ. Por exemplo, para $\alpha = 0,05$ e N = 1.000, a proporção de tais intervalos será de 95%; ou seja, seria de esperar 950 de dos 1.000 intervalos construídos como acima para conter o valor verdadeiro de θ. A evidência empírica mostra que tal expectativa é válida.

Também podemos definir um limite de confiança superior para θ, $U = U(X_1, ..., X_n)$, e um limite de confiança inferior para $\theta, L = L(X_1, \ldots, X_n)$, ambos com nível de confiança $1 - \alpha$, se, respectivamente, os intervalos $(-\infty, U]$ e $[L, \infty)$ são intervalos de confiança para θ com nível de confiança $1 - \alpha$. Quer dizer:

$$P_\theta(-\infty < \theta \leq U) = 1 - \alpha, \quad P_\theta(L \leq \theta < \infty) = 1 - \alpha \text{ para todo } \theta \in \Theta. \quad (2)$$

Existem algumas variações de (1) e (2). Por exemplo, quando a f.d.p. em questão é discreta, então igualdades em (1) e (2) raramente são obtidas para dados α e devem ser substituídos por desigualdades $\geq$. Além disso, exceto em casos especiais, igualdades em (1) e (2) são válidos apenas aproximadamente para grandes valores do tamanho da amostra n (mesmo nos casos em que os r.v. subjacentes são contínuos). Nesses casos, dizemos que os respectivos intervalos de confiança (limites de confiança) têm coeficiente de confiança aproximadamente $1 - \alpha$.

Finalmente, os parâmetros de interesse podem ser dois (ou mais) em vez de um, como presumimos até agora. Nesses casos, o conceito de intervalo de confiança é substituído por aquele de uma região de confiança (no parâmetros num espaço multidimensional). Este conceito será ilustrado por um exemplo ao estudarmos testes de hipótese, quando também expandiremos consideravelmente o que foi brevemente discutido aqui.

2.2 Intervalos de Confiança

Vamos formalizar na forma de uma definição alguns conceitos já introduzidos na seção anterior.

Definição 8 *Se $X_1, \dots, X_n$ é uma amostra aleatória com f.dp. $f(\cdot;\theta), \theta \in \Theta \subseteq \mathbb{R}$. Então:*

(i) *Um* ***intervalo aleatório*** *é um intervalo cujos limites são variaveis aleatórias.*

(ii) *Um* ***intervalo de confiança*** *para θ com* ***nível de confiança*** *$1-\alpha(0<\alpha<1, \alpha$ pequeno, é um intervalo aleatório cujos limites são estatisticas, digamos $L(X_1, \dots, X_n)$ e $U(X_1, \dots, X_n)$ tais que: $L(X_1, \dots, X_n) < U(X_1, \dots, X_n)$ e*

$$P_\theta\Big(L(X_1, \dots, X_n) \leq \theta \leq U(X_1, \dots, X_n)\Big) \geq 1-\alpha \text{ para todo } \theta \in \Theta$$

(iii) *A estatística $L(X_1, \dots, X_n)$ é dita* ***limite inferior de confiança*** *para θ com nível de confiança $1-\alpha$, se o intervalo $[L(X_1, \dots, X_n), \infty)$ é um intervalo de confiança para θ com nível de confiança $1-\alpha$. Da mesma forma $U(X_1, \dots, X_n)$ é dita* ***limite superior de confiança*** *para θ com nível de confiança $1-\alpha$, se o intervalo $(-\infty, U(X_1, \dots, X_n)]$ é um intervalo de confiança para θ com nível de confiança $1-\alpha$.*

Observação 3 *A significância de um intervalo de confiança provém da interpretação da frequência relativa da probabilidade. Assim, com base nos valores observados $x_1, \dots, x_n$ de $X_1, \dots, X_n$, constrói-se um intervalo com limites $L(X_1, \dots, X_n)$ e $U(X_1, \dots, X_n)$, e denota-se por $[L_1, U_1]$. Repete-se este experimento aleatório independentemente mais n vezes e da mesma forma, formando intervalo $[L_2, U_2]$. Repetindo-se este processo um grande número de vezes N independentemente cada vez, seja $[L_N, U_N]$ o intervalo correspondente. Então o fato de que $[L(X_1, \dots, X_n), U(X_1, \dots, X_n)]$ é um intervalo de confiança para θ com nível confiança $1-\alpha$ significa que aproximadamente $100(1-\alpha)\%$ dos N intervalos acima conterão θ, não importa qual seja seu valor.*

Observação 4 *Se $L(X_1, \dots, X_n)$ é um limite inferior e $U(X_1, \dots, X_n)$ é um limite superior de um intervalo de confiança para θ com nível de confiança $1-\frac{\alpha}{2}$ então $[L(X_1, \dots, X_n), U(X_1, \dots, X_n)]$ é um intervalo de confiança para θ com nível de confiança $1-\alpha$.*

Esta seção é concluída com a construção de intervalos de confiança em alguns exemplos concretos. Ao fazer isso, nos baseamos fortemente na teoria da distribuição e estimativas pontuais. Seria, talvez, útil delinear as etapas que geralmente segue na construção de um intervalo de confiança.

1. Encontre uma v.a. que contém o parâmetro θ, as v.a.s $U(X_1, \dots, X_n)$, de preferência na forma de uma estatística suficiente, e cuja distribuição é (exatamente ou pelo menos aproximadamente) conhecido.

2. Determine os pontos adequados $a < b$ de modo que a v.a. na etapa (a) encontre-se em $[a, b]$ com P_θ−probabilidade $\geq 1 - \alpha$.

3. Na expressão da etapa (b), reorganize os termos para chegar a um intervalo cujos limites são estatísticas e contendo θ.

4. O intervalo na etapa (c) é o intervalo de confiança necessário.

2.3 Alguns exemplos de Intervalos de Confiança

Vejamos agora alguns exemplos de Intervalo de confiança.

Exemplo 26 *Sejam $U(X_1, \ldots, X_n)$ var. aleat. iid com distribuição $\mathcal{N}(\mu, \sigma^2)$. Inicialmente considere σ conhecido nesse caso μ será o parâmetro e vamos considerar a seguinte estatística $T_n(\mu) = \frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}$, T_n depende das v.a's X_i através da estatística suficiente $\overline{X}$ e de μ, além disso $T_n \sim \mathcal{N}(0, 1)$ para todo $\mu \in \mathbb{R}$.*

A seguir vamos determinar $a < b$ tais que

$$P\left[a \leq \mathcal{N}(0,1) \leq b\right] = 1 - \alpha, \text{ (3) como } T_n \sim \mathcal{N}(0,1)$$

$$P\left[a \leq \frac{\sqrt{n}\left(\overline{X} - \mu\right)}{\sigma} \leq b\right] = 1 - \alpha$$

que é equivalente a

$$P\left[\overline{X} - b\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} - a\frac{\sigma}{\sqrt{n}}\right] = 1 - \alpha$$

Portanto

$$\left[\overline{X} - b\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} - a\frac{\sigma}{\sqrt{n}}\right] \quad (4)$$

É um intervalo de confiança para μ com nível de confiança $1-\alpha$, comprimento $(b - a)\frac{\sigma}{\sqrt{n}}$. Agora vamos determinar a e b de forma que entre todos os intervalos de confiança com nível de confiança $1 - \alpha$ que são da forma (4), o mais curto é aquele para o qual $b - a$ é o menor possível , onde a e b satisfazem (3). Isso acontece se $b = c(> 0)$ e $a = -c$, onde c é o quantil $\alpha/2$ superior da distribuição $\mathcal{N}(0, 1)$ que denotamos por $z_{\alpha/2}$. Portanto o menor intervalo de confiança para μ com nível de confiança $1 - \alpha$ (e que tem a forma (4)) é dado por

$$\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Em seguida, suponha que μ seja conhecido, de modo que σ^2 é o parâmetro, nesse caso considere a seguinte v.a:

$$\overline{T}_n\left(\sigma^2\right) = \tfrac{nS_n^2}{\sigma^2} \ \ onde\ S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$$

$$\overline{T}_n\left(\sigma^2\right) = \frac{n}{\sigma^2}\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_n^2$$

Agora vamos determinar dois números a e b tais que:

$$Prob\left(\left[a \leq \chi_n^2 \leq b\right]\right) = 1-\alpha$$

$$Prob\left(\left[a \leq \tfrac{nS_n^2}{\sigma^2} \leq b\right]\right) = 1-\alpha$$

$$Prob\left(\left[\tfrac{1}{b} \leq \tfrac{\sigma^2}{nS_n^2} \leq \tfrac{1}{a}\right]\right) = 1-\alpha$$

$$Prob\left(\left[\tfrac{nS_n^2}{b} \leq \sigma^2 \leq \tfrac{nS_n^2}{a}\right]\right) = 1-\alpha$$

O intervalo de confiança é dado por:

$$\left[\tfrac{nS_n^2}{b} \leq \sigma^2 \leq \tfrac{nS_n^2}{a}\right]$$

Agora utilizando a tabela da distribuição χ_n^2 encontramos $\chi^2_{n;1-\alpha/2}$ e $\chi^2_{n;\alpha/2}$ tais que $Prob(X \leq \chi^2_{n;1-\alpha/2}) = Prob(X \geq \chi^2_{n;\alpha/2}) = \alpha/2$ e temos o seguinte intervalo:

$$\left[\frac{nS_n^2}{\chi^2_{n;\alpha/2}} \leq \sigma^2 \leq \frac{nS_n^2}{\chi^2_{n;1-\alpha/2}}\right],$$

Se $n = 25$, $\sigma = 1$ e $1-\alpha = 0,95$ então $z_{\frac{\alpha}{2}} = 1,96$, $\chi^2_{n;1-\alpha/2} = 40,646$, $\chi^2_{n;\alpha/2} = 13,120$ nesse caso o Intervalo de confiança para a média é dado por:

$$IC_{\mu;1-\alpha} = \left[\overline{X} - 1,96\tfrac{1}{\sqrt{25}}, \overline{X} + 1,96\tfrac{1}{\sqrt{25}}\right]$$

O Intevalo de Confiança para σ^2 é dado por:

$$IC_{\sigma^2;1-\alpha} = \left[\tfrac{25S_n^2}{40{,}646} \leq \sigma^2 \leq \tfrac{25S_n^2}{13{,}120}\right],$$

2.4 Intervalos de Confiança na Presença de Parâmetros Incômodos