Apostila Inferência Estatística

Yasmin Ferreira Cavaliere
Luis Guillermo Coca Velarde

Sumário

Capítulo 1

Distribuições Amostrais

A inferência estatística está interessada em tomar decisões sobre uma população, baseando-se apenas na informação contida em uma amostra aleatória (a.a.) da população de interesse. Para isto, usar uma a.a. de tamanho n implicará trabalhar num espaço n-dimensional o que aumenta o trabalho matemático a realizar. Porém, uma forma de diminuir esta dimensão consiste em usar estatísticas que resumem a informação da amostra. A seguir são apresentadas algumas definições e resultados úteis para implementar os mecanismos de inferência.

Definição 1.1 *Seja X uma v.a. com função de densidade ou função de probabilidade $f_X(x)$. Uma amostra aleatória (a.a.) de tamanho n de X é um conjunto de n variáveis aleatórias $X_1, \ldots, X_n$ com as seguintes condições:*

- *Cada uma das v.a's X_i, $i = 1, 2, \ldots, n$ tem a mesma distribuição de X_i*

$$f_{X_i}(x_i) = f_X(x)$$

- *As v.a's são independentes, logo*

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \ldots f_{X_n}(x_n)$$

A partir da Definição **??** pode-se concluir que dada uma amostra aleatória $\underset{\sim}{X} = (X_1, \ldots, X_n)$ de uma variável X com média μ e variância σ^2 então $E(X_i) = \mu$ e $Var(X_i) = \sigma^2$ para todo $i = 1, 2, \ldots, n$.

Definição 1.2 *Seja $X_1, \ldots, X_n$ uma a.a. de tamanho n. A distribuição da amostra $X_1, \ldots, X_n$ é a distribuição conjunta de $X_1, \ldots, X_n$.*

Exemplo 1.3 *Seja X_1, X_2 com função densidade de probabilidade (f.d.p.) $f_X(x)$. Então $f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$.*

Definição 1.4 *A distribuição de probabilidade de um estimador é chamada de distribuição amostral.*

Exemplo 1.5 *A distribuição de probabilidade de $\bar{X}$ é chamada de distribuição amostral da média. Portanto, dado que $\bar{X}$ tem uma distribuição de probabilidade pode-se calcular $P(10 < \bar{X} < 30)$ bastando para isso conhecer a forma da distribuição de $\bar{X}$.*

1.1 Estatísticas e momentos amostrais

Usualmente, é impraticável observar toda uma população, seja pelo custo alto seja por dificuldades operacionais. Examina-se então uma amostra, de preferência bastante representativa, para que os resultados obtidos possam ser generalizados para toda a população. Um experimento pode ter por finalidade a determinação da estimativa de uma função do parâmetro. Toda conclusão tirada por amostragem, quando generalizada para a população, apresentará um grau de incerteza. Inferência Estatística se refere ao conjunto de técnicas e procedimentos que permitem dar ao pesquisador um grau de confiabilidade nas afirmações que faz para a população, baseadas nos resultados das amostras.

Definição 1.6 *Estatística é uma função de v.a's observáveis que não contém parâmetro desconhecido.*

Exemplo 1.7 *Seja uma a.a.* $X_1, \ldots, X_n$. *A média amostral* $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ *é uma estatística.*

Definição 1.8 *Seja* $X_1, \ldots, X_n$ *uma a.a. de* $f_X(.)$. *O r-ésimo momento amostral em relação a zero, é definido por:*

$$M'_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r \tag{1.1}$$

Logo, o r-ésimo momento amostral em relação à media é dado por

$$M_r = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X}_n)^r \tag{1.2}$$

Observe que momentos amostrais são exemplos de estatísticas.

Definição 1.9 *O r-ésimo momento populacional de* X *é definido por:*

$$\mu'_r = E(X^r) \tag{1.3}$$

Com base nas definições **??** e **??** pode-se formular o seguinte Teorema.

Teorema 1.10 *Seja* $X_1, \ldots, X_n$ *uma a.a. de* $f_X(x)$, *então:*

- $E(M'_r) = \mu' r$, *se* μ'_r *existir.*
- $Var(M'_r) = \frac{\mu'_{2r} - (\mu'_r)^2}{n}$

Demonstração 1.11 *Primeiramente veja a demonstração do primeiro item:*

- $E[M'_r] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^r\right] = \frac{1}{n}E[X_1^r + \ldots + X_n^r] = \frac{1}{n}[E(X_1^r) + \ldots + E(X_n)^r] = \frac{1}{n}(nE(X^r)) = E(X^r) = \mu'_r$

  Seguindo, vamos demonstrar o segundo item.

- $E[(M'_r)^2] = E\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i^r\right)^2\right] = \frac{1}{n^2}E\left[\sum_{i=1}^{n} X_i^{2r} + \sum_{i\neq j}\sum X_i^r X_j^r\right] =$
$\frac{1}{n^2}\left[\sum_{i=1}^{n} E(X_i^{2r}) + \sum_{i\neq j}\sum E(X_i^r)E(X_j^r)\right] = \frac{1}{n^2}n\mu'_{2r} + \frac{1}{n^2}n(n-1)(\mu'_r)^2 = \frac{\mu'_{2r}}{n} + \frac{n-1}{n}(\mu'_r)^2$
Por propriedade da variância, sabemos que:

$$Var(\mu'_r) = E((\mu'_r)^2) - E^2(\mu'_r)$$

Logo, $Var(\mu'_r) = \frac{\mu'_{2r}}{n} + \frac{n-1}{n}(\mu'_r)^2 - (\mu'_r)^2 = \frac{\mu'_{2r} - (\mu'_r)^2}{n}$

1.2 Distribuição Amostral da média

A média populacional μ representa a média de todos os indivíduos ou objetos que estão sendo estudados. Mas usualmente, nem todos os indivíduos podem ser medidos. Em geral, somente uma amostra de todos os indivíduos está disponível e a média baseada nesta amostra, $\overline{X}$, é usada para estimar a média populacional μ. Um problema de fundamental importância é saber se a média amostral $\overline{X}$ é um bom estimador da média populacional μ.

Teorema 1.12 *Seja $X_1, \ldots, X_n$ a.a. de $f_X(.)$ com média μ e variância σ^2. Além disso, defina $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Então $E(\bar{X}) = \mu$ e $Var(\bar{X}) = \frac{\sigma^2}{n}$.*

Observação: Quanto maior o tamanho da amostra, menor a variabilidade de $\bar{X}$.

Demonstração 1.13

- $E(\bar{X}) = E\left[\frac{\sum_{i=1}^n X_i}{n}\right] = \frac{1}{n}E[\sum_{i=1}^n X_i] = \frac{1}{n}n\mu = \mu$
- $Var(\bar{X}) = Var\left[\frac{\sum_{i=1}^n X_i}{n}\right] = \frac{1}{n^2}Var\left[\sum_{i=1}^n X_i\right] = \frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$

Seja $f(\cdot, \theta)$ função densidade de probabilidade de uma variável aleatória X cujo valor esperado é denotado por μ. Suponha que queremos estimar μ. Porém em qualquer problema real, podemos somente observar um número finito de valores da variável aleatória X. Será então que, usando somente um número finito de valores de X (uma amostra aleatória de tamanho n), podemos realizar inferências confiáveis sobre μ? A resposta é sim e isto é o que veremos como aplicação da lei fraca dos grandes números.

A lei fraca dos grandes números estabelece que, para quaisquer dois números suficientemente pequenos ϵ e δ, com $\epsilon > 0$ e $0 < \delta < 1$, existe um número inteiro n tal que, se uma amostra aleatória de tamanho n ou maior que n é obtida de f, a média amostral está "próxima"de μ com probabilidade maior ou igual que $1 - \delta$.

Teorema 1.14 ***Lei Fraca dos Grandes Números***

Seja X uma v.a. com função de densidade $f(\cdot)$, média μ e variância σ^2. Seja $\bar{X}$ a média de uma a.a. de tamanho n de X. Sejam ϵ e δ quaisquer dois números especificados que satisfazem $\epsilon > 0$ e $0 < \delta < 1$. Se n é qualquer inteiro maior que $\frac{\sigma^2}{\epsilon^2\delta}$, então

$$\mathbb{P}[-\epsilon < \bar{X} - \mu < \epsilon] \geq 1 - \delta.$$

Demonstração 1.15 *Sabemos que*

$$\mathbb{P}(g(X) \geq k) \leq \frac{\mathbb{E}(g(x))}{k}$$

para todo $k > 0$*, para toda variável aleatória* X *e toda função não negativa* g*. De maneira equivalente, temos que*

$$\mathbb{P}(g(x) < k) \geq 1 - \frac{\mathbb{E}(g(X))}{k}.$$

Vamos tomar $g(X) = (\bar{X} - \mu)^2$ *e* $k = \epsilon^2$*. Então*

$$\mathbb{P}[|\bar{X} - \mu|^2 < \epsilon^2] = \mathbb{P}[|\bar{X} - \mu| < \epsilon] = \mathbb{P}[-\epsilon < \bar{X} - \mu < \epsilon].$$

Mas

$$\mathbb{P}[|\bar{X} - \mu|^2 < \epsilon^2] \geq 1 - \frac{E(\bar{X} - \mu)^2}{\epsilon^2} = 1 - \frac{(1/n)\sigma^2}{\epsilon^2} \geq 1 - \delta.$$

O que é verdade se $\frac{\frac{1}{n}\sigma^2}{\epsilon^2\delta} \leq \delta$ *ou* $n \geq \frac{\sigma^2}{\epsilon^2\delta}$.

Exemplo 1.16 *Qual é o tamanho de amostra necessário para ter* 99% *de certeza de que* $\bar{X}$ *está* $0,5\sigma$ *ao redor da média* μ*?*

Sabemos que, $\delta = 1 - 0,99 = 0,01$*,* $\epsilon = 0,5\sigma$*. Pelo Teorema* **??***, temos que:*

$$\mathbb{P}[-0,5\sigma < \bar{X} - \mu < 0,5\sigma] \geq 0,01$$

Então, n *é qualquer número inteiro que satisfaz:*

$$n > \frac{\sigma^2}{(0,5\sigma)^2 0,01} = \frac{1}{(0,5)^2 0,01} = 400$$

Assumir que os dados tem uma distribuição normal é altamente conveniente tanto do ponto de vista teórico como do ponto de vista computacional. Mas isto deixa um problema de fundamental importância: sob quais circunstâncias é razoável assumir que a distribuição normal pode ser usada? Gauss trabalhou neste problema por muito tempo, mas é um resultado de Laplace que é utilizado hoje. Anunciado em 1810, Laplace chamou este resultado de Teorema do Limite Central, que diz que sob a hipótese de amostragem aleatória, quando o tamanho da amostra aumenta, a distribuição de probabilidade da média amostral se aproxima de uma distribuição normal com média μ e variância σ^2/n, ou seja, se o tamanho amostral é suficientemente grande, podemos assumir que a média amostral tem uma distribuição normal.

Teorema 1.17 ***Teorema do Limite Central***
Seja X *uma v.a. com densidade* f *com média* μ *e variância* σ^2*. Seja* $\overline{X}$ *a média de uma amostra aleatória de tamanho* n *de* f*. Então, a distribuição da variável aleatória* Z_n *definida por*

$$Z_n = \frac{\overline{X} - E(\overline{X})}{\sqrt{Var(\overline{X})}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}.$$

se aproxima da distribuição normal padrão quando n *tende ao infinito. Em outras palavras,* Z_n *tem distribuição aproximadamente normal com média* 0 *e variância* 1 *quando* n *tende ao infinito.*

Observação: A qualidade da aproximação normal para a distribuição amostral da média dependerá do tamanho da amostra e da distribuição da população de onde foi retirada a amostra. Em muitos casos de interesse prático, se $n > 30$ a aproximação normal será satisfatória, independente da distribuição da população.

Se X não tiver distribuição normal então pelo Teorema Central do Limite segue que a distribuição da média amostral será aproximadamente normal com média μ e variância $\frac{\sigma^2}{n}$ quando $n \to \infty$. Como é especificado no seguinte corolário.

Corolário 1.18 *Seja $\bar{X}_n$ a média de uma a.a. de uma $N(\mu, \sigma^2)$. Então $\bar{X}_n \sim N(\mu, \sigma^2/n)$ quando $n \to \infty$.*

1.3 Distribuição Qui-Quadrado

A distribuição qui-quadrado possui numerosas aplicações importantes em inferência estatística. Devido a sua importância a distribuição qui-quadrado está tabulada para diferentes valores do parâmetro n.

Definição 1.19 *Uma variável aleatória contínua X tem distribuição qui-quadrado com k graus de liberdade se sua função densidade for dada por:*

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} \exp\left(-\frac{x}{2}\right);\ k > 0,\ x > 0$$

sendo $\Gamma(\omega) = \displaystyle\int_0^\infty x^{\omega-1} e^{-x} dx,\ \omega > 0$. *Denotamos* $X \sim \chi^2_k$.

Observação: Se $X \sim \chi^2_k$, então temos que $X \sim Gamma(\frac{k}{2}, \frac{1}{2})$, ou seja a distribuição qui-quadrado é um caso particular da distribuição Gama.

Propriedades 1.20 *Se $X \sim \chi^2_k$, então:*

- $E(X) = k$
- $Var(X) = 2k$
- $M_X(t) = \left(\frac{1}{1-2t}\right)^{\frac{k}{2}}$

Teorema 1.21 *Se $X_i \sim N(\mu_i, \sigma_i^2)$ para $i = 1, \ldots, n$ independentes então $U = \sum_{i=1}^n \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \sim \chi^2_n$.*

Demonstração 1.22 *Defina $Z_i = \frac{X_i - \mu_i}{\sigma_i} \sim N(0,1)$. Como as variáveis aleatórias $Z_1, \cdots, Z_n$ são independentes, temos que*

$$M_U(t) = E[e^{tU}] = E\left(e^{t\left(\sum_{i=1}^n (Z_i^2)\right)}\right) = E\left(\prod_{i=1}^n e^{t(Z_i)^2}\right) = \prod_{i=1}^n E\left(e^{t(Z_i)^2}\right). \tag{1.4}$$

Mas

$$E\left(e^{t(Z_i)^2}\right) = \int e^{tz_i^2}\left(\frac{1}{\sqrt{2\pi}}\right) e^{-(1/2)z_i^2} dz_i = \int \frac{1}{\sqrt{2\pi}} e^{-(1/2)(1-2t)z_i^2} dz_i$$

e, portanto,

$$E\left(e^{t(Z_i)^2}\right) = \frac{1}{\sqrt{1-2t}}\int \frac{\sqrt{1-2t}}{\sqrt{2\pi}}e^{-(1/2)(1-2t)z_i^2}dz_i = \left(\frac{1}{1-2t}\right)^{\frac{1}{2}}$$

em que a última integral é igual a 1, pois se trata justamente de uma normal com média zero e variância $1/(1-2t)$*. Portanto, substituindo em* **??***,*

$$E\left(e^{tZ}\right) = \prod_{i=1}^{n} E\left(e^{tZ_i}\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{1-2t}} = \left(\frac{1}{1-2t}\right)^{n/2}.$$

que corresponde a função geradora de momentos da distribuição qui-quadrado com n *graus de liberdade. E o resultado segue.*

Teorema 1.23 *Se* $Z_1, \ldots, Z_n$ *é uma a.a. da* $N(0,1)$*, então:*

- $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i \sim N(0, 1/n)$.
- $\bar{Z}$ *e* $\sum_{i=1}^{n}(Z_i - \bar{Z})^2$ *são independentes.*
- $\sum_{i=1}^{n}(Z_i - \bar{Z})^2 \sim \chi^2_{(n-1)}$

Demonstração 1.24 *Façamos as contas...*

- $$M_{\bar{Z}}(t) = E[e^{t\bar{Z}}] = E\left(e^{\frac{t}{n}(\sum_{i=1}^{n} Z_i)}\right) = E\left(\prod_{i=1}^{n} e^{\frac{t}{n}Z_i}\right) = \prod_{i=1}^{n} E\left(e^{\frac{t}{n}Z_i}\right) = \prod_{i=1}^{n} M_Z\left(\frac{t}{n}\right) = e^{\frac{t^2}{2n^2}}.$$

  Logo, $\bar{Z} \sim N(0, 1/n)$.

- *Para* $n = 2$, $\bar{Z} = \frac{Z_1+Z_2}{2}$ *e,*

  $$(Z_i - \bar{Z}^2) = \left[Z_1 - \frac{Z_1+Z_2}{2}\right]^2 + \left[Z_2 - \frac{Z_1+Z_2}{2}\right]^2 = \left[\frac{2Z_1 - Z_1 - Z_2}{2}\right]^2 + \left[\frac{2Z_2 - Z_1 - Z_2}{2}\right]^2 =$$

  $$\left[\frac{Z_1 - Z_2}{2}\right]^2 + \left[\frac{Z_2 - Z_1}{2}\right]^2 = \left[\frac{Z_2 - Z_1}{2}\right]^2 + \left[\frac{Z_2 - Z_1}{2}\right]^2 = \frac{[Z_2 - Z_1]^2}{2}$$

  Seguindo,

  $$M_{Z_1+Z_2}(t_1) = E[e^{t(Z_1+Z_2)}] = E[e^{tZ_1}e^{tZ_2}] = E[e^{t_1 Z_1}]E[e^{t_1 Z_2}] = M_Z(t_1)M_Z(t_1) = e^{\frac{1}{2}t_1^2}e^{\frac{1}{2}t_1^2} = e^{t_1^2}$$

  De forma análoga,

  $$M_{Z_1-Z_2}(t_2) = e^{t_2^2}$$

  Portanto,

  $$M_{Z_1+Z_2, Z_1-Z_2}(t_1, t_2) = \mathbb{E}[e^{t_1(Z_1+Z_2)+t_2(Z_1-Z_2)}] = \mathbb{E}[e^{Z_1(t_1-t_2)}e^{Z_2(t_1+t_2)}] =$$

  $$M_{Z_1}(t_1 - t_2)M_{Z_2}(t_1 + t_2) = e^{\frac{(t_1-t_2)^2}{2}}e^{\frac{(t_1+t_2)^2}{2}}e^{\frac{(t_1-t_2)^2}{2}+\frac{(t_1+t_2)^2}{2}} =$$

  $$e^{t_1^2+t_2^2} = e^{t_1^2}e^{t_2^2} = M_{Z_1+Z_2}(t_1)M_{Z_1-Z_2}(t_2)$$

- *Temos que:*
$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n}(Z_i - \bar{Z} + \bar{Z})^2 = \sum_{i=1}^{n}(Z_i - \bar{Z})^2 + n\bar{Z}^2.$$

  Logo,
$$M_{\sum_{i=1}^n Z_i^2}(t) = M_{\sum_{i=1}^n (Z_i-\bar{Z})^2}(t) M_{n\bar{Z}^2}(t).$$

  Sabendo que $\sum_{i=1}^n Z_i^2 \sim \chi_n^2$ *e* $n\bar{Z}^2 \sim \chi_1^2$, *então:*
$$\left(\frac{1}{1-2t}\right)^{\frac{n}{2}} = M_{\sum_{i=1}^n (Z_i-\bar{Z})^2}(t) \left(\frac{1}{1-2t}\right)^{\frac{1}{2}}.$$

  Portanto,
$$M_{\sum_{i=1}^n (Z_i-\bar{Z})^2}(t) = \left(\frac{1}{1-2t}\right)^{\frac{n-1}{2}}$$

  ou seja, $\sum_{i=1}^n (Z_i - \bar{Z})^2 \sim \chi_{n-1}^2$

Teorema 1.25 *Se* $X_1, \ldots, X_n$ *é uma a.a. de* $N(\mu, \sigma^2)$ *e* $Z_i = \frac{X_i-\mu}{\sigma} \sim N(0,1)$, *então:*

- $\frac{1}{n}\sum_{i=1}^n \left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{\bar{X}-\mu}{\sigma} \sim N(0,1).$
- $\bar{X}$ *e* $\sum_{i=1}^n (X_i - \bar{X})^2$ *são independentes.*
- $\sum_{i=1}^n \left(\frac{X_i-\bar{X}}{\sigma}\right)^2 \sim \chi_{n-1}^2.$

Corolário 1.26 *Se* $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ *é a variância amostral de uma a.a. da* $N(\mu, \sigma^2)$, *então*
$$U = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Exercício 1.27 *Mostre que se* $X_1, \ldots, X_n$ *é uma a.a. de uma distribuição normal padronizada, então:*

1. $X^2 \sim \chi_n^2$.
2. $Z = \sum_{i=1}^n X_i^2$ *segue uma distribuição qui-quadrado com n graus de liberdade.*
3. *Sejam* $U_1, U_2, \ldots, U_k$ *variáveis aleatórias independentes com distribuição qui-quadrado com* $n_1, n_2, \ldots, n_k$ *graus de liberdade respectivamente. Então a soma* $W = U_1 + U_2 + \ldots + U_k$ *tem distribuição qui-quadrado com* $n_1 + n_2 + \ldots + n_k$ *graus de liberdade.*

1.4 Distribuição t-Student

A distribuição t de Student é uma das distribuições mais utilizadas na estatística, com aplicações que vão desde a modelagem estatística até testes de hipóteses.

Definição 1.28 *Dizemos que X tem distribuição t de Student com k graus de liberdade se sua função densidade de probabilidade é dada por*

$$f(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)}\left(1+\frac{x^2}{k}\right)^{-\left(\frac{k+1}{2}\right)} \qquad x \in (-\infty, \infty).$$

Utilizamos a notação $X \sim t_k$.

Propriedades 1.29 *Se $X \sim T_k$, então:*

- *$E(X) = 0$ para $k > 1$.*
- *$Var(X) = \frac{k}{k-2}$ se $k > 2$.*
- *A função geradora de momentos da t de Student não está definida para todos os graus de liberdade, entretanto podemos encontrar os momentos da função t de Student para alguns graus de liberdade.*

  Desta forma seja $X \sim t_\nu$, então

$$\mathbb{E}\left(X^k\right) = \begin{cases} 0, \ \textit{se } 0 \leq c < k \textit{ e } c \textit{ é ímpar} \\ \frac{1}{\sqrt{\pi}\Gamma\left(\frac{k}{2}\right)}\left[\Gamma\left(\frac{c+1}{2}\right)\Gamma\left(\frac{k-c}{2}\right)k^{c/2}\right], \ \textit{se } 0 \leq c \leq k \textit{ e } c \textit{ é par;} \\ \textit{indefinido}, \ \textit{se } c \geq k \textit{ e } c \textit{ é ímpar} \\ \infty, \ \textit{se } c \geq k \textit{ e } c \textit{ é par.} \end{cases}$$

Principais Características:

1. Cada número de graus de liberdade dá origem a uma distribuição t diferente.
2. A função densidade tem a mesma forma em sino da distribuição Normal, mas reflete uma maior variabilidade (com curvas mais alargadas) que é de se esperar em amostras pequenas.
3. A distribuição t-Student se aproxima da normal quando aumenta o número de graus de liberdade.
4. A curva é simétrica em torno do zero.

Teorema 1.30 *Considere $Z \sim N(0,1)$ e $U \sim \chi^2_k$ duas variáveis aleatórias independentes. Defina X como sendo uma variável aleatória de tal forma que*

$$X = \frac{Z}{\sqrt{U/k}}.$$

Temos que a variável aleatória X tem distribuição t de Student com k graus de liberdade.

Demonstração 1.31 *A função densidade de probabilidade conjunta de Z e U é dada por*

$$f_{Z,U}(z,u) = \frac{1}{\sqrt{2\pi}}\frac{1}{\Gamma(k/2)}\left(\frac{1}{2}\right)^{k/2} u^{(k/2)-1}e^{-u/2}e^{-z^2/2} \qquad u \in (0, \infty).$$

Considerando a transformação

$$X = \frac{Z}{\sqrt{U/k}} \quad e \quad Y = U$$

o jacobiano é $\sqrt{y/k}$ *e então*

$$f_{X,Y}(x,y) = \sqrt{\frac{y}{k}} \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(k/2)} \left(\frac{1}{2}\right)^{k/2} y^{(k/2)-1} e^{-y/2} e^{-x^2 y/2k} \qquad y \in (0,\infty)$$

e

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \frac{1}{\sqrt{2k\pi}} \frac{1}{\Gamma(k/2)} \left(\frac{1}{2}\right)^{k/2} \int_0^{\infty} y^{k/2-1+1/2} e^{-1/2(1+x^2/k)y} dy.$$

Na sequência, ao fazermos a mudança de variável

$$w = y\frac{1}{2}\left(1 + \frac{x^2}{k}\right),$$

obtemos que

$$f_X(x) = \frac{\Gamma[(k+1)/2]}{\Gamma(k/2)} \frac{1}{\sqrt{k\pi}} \frac{1}{(1+x^2/k)^{(k+1)/2}}$$

que é a função densidade de probabilidade de uma distribuição t *com* k *graus de liberdade.*

Corolário 1.32 *Seja* $X_1, \ldots, X_n$ *uma a.a.de uma* $N(\mu, \sigma^2)$*, então:*

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

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

Exercício 1.33 *Demonstre o Corolário* **??**. *(Dica: Este fato é decorrente do Teorema* **??** *acima).*

1.5 Distribuição F

A distribuição F de Snedecor também conhecida como distribuição de Fisher é frequentemente utilizada na inferência estatística para análise da variância.

Definição 1.34 *Uma variável aleatória contínua* X *tem distribuição F de Snedecor com* m *graus de liberdade no numerador e* n *graus de liberdade no denominador se sua função densidade de probabilidade é definida por*

$$f(x) = \frac{\Gamma\left[\frac{m+n}{2}\right] \left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1}}{\Gamma\left[\frac{m}{2}\right] \Gamma\left[\frac{n}{2}\right] \left[\left(\frac{m}{n}\right)x + 1\right]^{\frac{m+n}{2}}} \qquad x \in [0,\infty)$$

Neste caso, utilizamos a notação $X \sim F(m,n)$.

Propriedades 1.35 *Se* $X \sim F(m,n)$*, então:*

- $\mathbb{E}(X) = \frac{n}{n-2}$ *para* $n > 2$.

- $Var(X) = \frac{2n^2(m+n-2)}{m(n-4)(n-2)^2}$ *para* $n > 4$
- *Não existe função geradora de momentos para a distribuição F de Snedecor.*

Principais Características:

1. Cada par de graus de liberdade dá origem a uma distribuição F diferente.
2. A distribuição F depende de dois parâmetros. O primeiro (m) é o grau de liberdade do numerador e o segundo (n) do denominador.
3. A variável aleatória F é não-negativa, e a distribuição é assimétrica à direita.
4. A distribuição F se parece com a distribuição qui-quadrado, no entanto, os parâmetros m e n fornecem flexibilidade extra em relação à forma.

Teorema 1.36 *Seja* $U \sim \chi^2_m$ *e* $V \sim \chi^2_n$*, ambas independentes, então:*

$$F = \frac{U/m}{V/n} \sim F(m,n).$$

Demonstração 1.37 *Seja X uma variável aleatória positiva com função densidade de probabilidade f_X e Y uma variável aleatória com função densidade f_Y. Suponha que as variáveis aleatórias X e Y sejam independentes. Neste caso, a função densidade de probabilidade conjunta é dada por $f_{XY} = f_X f_Y$. Considere a fração $Z = \frac{Y}{X}$. Neste caso, a função densidade conjunta do quociente Z é dada por*

$$\mathbb{P}\left(Z \leq z\right) = \mathbb{P}\left(\frac{Y}{X} \leq z\right) = \mathbb{P}\left((X,Y) \in B_z\right),$$

em que $y \leq zx$. Assim temos que

$$F_Z(z) = \mathbb{P}\left((X,Y) \in B_z\right) = \int_0^\infty \int_{-\infty}^{zx} f_{XY}(x,y)dydx$$

Considerando a mudança de variável $y = xw$; temos que:

$$F_Z(z) = \int_0^\infty \int_{-\infty}^{z} x f_{XY}(x,xw)dwdx = \int_{-\infty}^{z} \int_0^\infty x f_{XY}(x,xw)dxdw.$$

Assim, a função densidade de probabilidade de Z é dada por

$$f_Z(z) = \int_0^\infty x f_{XY}(x,xz)dx.$$

Como X e Y são independentes, a distribuição conjunta do quociente é dada por

$$f_Z(z) = \int_0^\infty x f_X(x) f_Y(xz)$$

Portanto a distribuição do quociente $Z = \frac{Y}{X}$, com $Y \sim \chi^2_m$ e $X \sim \chi^2_n$ é dada por:

$$f_Z(z) = \int_0^\infty x \frac{x^{(n/2)-1}e^{-x/2}}{\Gamma\left(\frac{n}{2}\right)2^{n/2}} \frac{(xz)^{(m/2)-1}e^{-xz/2}}{\Gamma\left(\frac{m}{2}\right)2^{m/2}} dx$$

de onde concluímos que

$$f_Z(z) = \frac{z^{(m/2)-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)2^{(m+n)/2}} \int_0^\infty x^{(m+n)/2-1} e^{-x(z+1)/2} dx$$

lembrando que $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$*. Fazendo a substituição* $t = x\left(\frac{z+1}{2}\right)$ *e reorganizando a integral acima temos que:*

$$f_Z(z) = \frac{z^{(m/2)-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)2^{(m+n)/2}}\frac{1}{\left(\frac{z+1}{2}^{(m+n)/2}\right)}\int_0^\infty t^{(m+n)/2-1}e^{-t}dt = \frac{\Gamma\left(\frac{m+n}{2}\right)z^{(m/2)-1}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)(z+1)^{(m+n)/2}}.$$

Para finalizar, tomamos $W = \frac{Y/m}{X/n}$ *e, neste caso, temos que*

$$\mathbb{P}\left(W \leq w\right) = \mathbb{P}\left(\frac{Y/m}{X/m} \leq w\right) = \mathbb{P}\left(\frac{Y}{X} \leq \frac{m}{n}w\right) = \int_0^{\frac{m}{n}w} f_Z(z)dz.$$

Ao realizarmos a transformação de variáveis $t = (n/m)z$*, concluímos que*

$$\mathbb{P}\left(W \leq w\right) = \int_0^w f_Z(tm/n)(m/n)dt$$

Ao substituirmos, concluímos que W *segue uma distribuição* F *com* m *graus de liberdade no numerador e* n *graus de liberdade no denominador.*

Corolário 1.38 *Seja* $X_1, \ldots, X_{m+1}$ *a.a. de uma* $N(\mu_X, \sigma^2)$ *e* $Y_1, \ldots, Y_{n+1}$ *a.a. de uma* $N(\mu_Y, \sigma^2)$*. Se as duas amostras são independentes, então:*

- $\frac{1}{\sigma^2}\sum_{i=1}^{m+1}(X_i - \bar{X})^2 \sim \chi^2_m$
- $\frac{1}{\sigma^2}\sum_{i=1}^{n+1}(Y_i - \bar{Y})^2 \sim \chi^2_n$
- $\frac{\frac{\frac{1}{\sigma^2}\sum_{i=1}^{m+1}(X_i-\bar{X})^2}{m}}{\frac{\frac{1}{\sigma^2}\sum_{i=1}^{n+1}(Y_i-\bar{Y})^2}{n}} \sim F(m,n)$

Exercício 1.39 *Demonstre os três itens do Corolário* **??**.

Relações Importantes:

- Se $X \sim F_{1-\alpha,1,n} = t^2_{1-\alpha/2,n}$, onde $1-\alpha$ representa o quantil da distribuição F, e $m = 1$ e n, os graus de liberdade.
- $F_{\alpha,m,\infty} = \frac{\chi^2_{\alpha,m}}{m}$, onde α representa o quantil da distribuição F, e m e "$n = \infty$"($n \to \infty$), os graus de liberdade.
- Se $X \sim F(m,n) \Rightarrow \frac{1}{X} \sim F(n,m)$.
- Se $X \sim F(m,n) \Rightarrow \frac{\frac{m}{n}X}{1+\frac{m}{n}X} \sim Beta(\frac{m}{2}, \frac{n}{2})$.

Capítulo 2

Elementos de inferência

Muitas das distribuições estudadas até agora são membros da chamada classe ou família exponencial. Assim, por exemplo, o são as distribuições normal, binomial, binomial negativa, gama, Poisson e normal inversa. Esta classe de famílias de distribuições foi descoberta independentemente por Koopman, Pitman e Darmois através do estudo de propriedades de suficiência estatística. Posteriormente, muitos outros aspectos dessas famílias foram descobertos e tornaram-se importantes na teoria moderna de estatística.

2.1 Caso uniparamétrico

Definição 2.1 *A família uniparamétrica de densidade $f(x|\theta)$ que pode ser expressa como:*

$$f(x|\theta) = a(\theta)b(x)exp\{c(\theta)d(x)\}I_{(-\infty,\infty)}(x)$$

$\forall\theta \in \Theta$ e para uma certa escolha de funções $a(.)$, $b(.)$, $c(.)$ e $d(.)$ é definida como pertencente à família ou classe exponencial.

Exemplo 2.2 *Seja $X \sim Exp(\lambda)$ tal que $f(x|\lambda) = \lambda e^{-\lambda x} I_{(0,\infty)}(x)$. Logo, temos que:*

$$f(x|\lambda) = I_{(0,\infty)}(x)\lambda exp\{-\lambda x\},$$

onde $a(\lambda) = \lambda$, $b(x) = I_{(0,\infty)}(x)$, $c(\lambda) = -\lambda$ e $d(x) = x$. Portanto, X pertence a família exponencial.

Exemplo 2.3 *Seja X uma v.a. com distribuição $f(x|\theta) = \frac{\theta^x log(\theta)}{\theta-1}$, para $0 < x < 1$. Veja que:*

$$f(x|\theta) = \frac{exp(log\theta^x)log\theta}{\theta - 1} I_{(0,1)}(x)$$

$$f(x|\theta) = exp(xlog\theta)\frac{log\theta}{\theta - 1} I_{(0,1)}(x)$$

.

Ao definir $a(\theta) = \frac{log\theta}{\theta-1}$, $b(x) = I_{(0,1)}(x)$, $c(\theta) = log\theta$ e $d(x) = x$, podemos verificar que X pertence a família exponencial.

Exercício 2.4 *Seja $X \sim N(\mu, \sigma^2)$ com σ conhecido. Verificar se X pertence a família exponencial uniparamétrica.*

2.2 Caso k-paramétrico

Definição 2.5 *Uma família de densidades $f(x|\theta_1,\ldots,\theta_k)$ é da família exponencial de k parâmetros se ela pode ser escrita como:*

$$f(x|\theta_1,\ldots,\theta_n)=a(\theta_1,\ldots,\theta_n)b(x)exp\{\sum_{j=1}^{k}c_j(\theta_1,\ldots,\theta_n)d_j(x)\}$$

para uma certa escolha de funções $a(\ldots)$, $b(.)$, $c_j(\ldots)$ e $d_j(.)$, $j=1,\ldots,k$. Vale ressaltar que existem k funções c_j e d_j.

Exemplo 2.6 *Seja $X\sim N(\mu,\sigma^2)$. Verificar se X pertence a família exponencial de dois parâmetros.*

Solução

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

.

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

Defina $a(\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}exp\left\{-\frac{\mu^2}{2\sigma^2}\right\}$, $b(x)=I_{(-\infty,\infty)}(x)$, $c_1(\mu,\sigma^2)=-\frac{1}{2\sigma^2}$, $c_2(\mu,\sigma^2)=\frac{\mu}{\sigma^2}$, $d_1(x)=x^2$ e $d_2(x)=x$ e veja que X pertence a família exponencial.

Exercício 2.7 *Seja $X\sim Gama(\alpha,\beta)$. Verificar que X pertence a família exponencial K-paramétrica. Qual o valor de K?*

Teorema 2.8 *Seja X uma v.a. cuja distribuição pertence a família exponencial, então:*

- $\mathbb{E}\left[\sum_{j=1}^{k}\frac{\partial}{\partial\theta_i}c_j(\theta_1,\ldots,\theta_n)d_j(x)\right]=-\frac{\partial}{\partial\theta_i}ln(a(\theta_1,\ldots,\theta_k))$;
- $Var\left[\sum_{j=1}^{k}\frac{\partial}{\partial\theta_i}c_j(\theta_1,\ldots,\theta_n)d_j(X)\right]=-\frac{\partial^2}{\partial\theta_i^2}ln(a(\theta_1,\ldots,\theta_k))-\mathbb{E}\left[\sum_{j=1}^{k}\frac{\partial^2}{\partial\theta_i^2}c_j(\theta_1,\ldots,\theta_n)d_j(X)\right]$

Embora essas equações possam parecer complicadas, quando aplicadas a casos específicos elas podem funcionar muito bem. Sua vantagem é que podemos substituir a integração ou soma pela diferenciação, que geralmente é mais direta. O exemplo a seguir ilustra isto.

Exemplo 2.9 *Seja $X\sim Bin(n,p)$, n conhecido. Determine $\mathbb{E}(X)$ e $Var(X)$ usando o Teorema* **??**.

Primeiramente, precisamos verificar que X pertence a família exponencial. Basta verificar que

$$f(x|n,p)=\binom{n}{x}p^x(1-p)^{n-x}I_{(0,1,\ldots,n)}(x)=\binom{n}{x}I_{(0,1,\ldots,n)}(x)(1-p)^n exp\left\{xln\left(\frac{p}{1-p}\right)\right\}.$$

Portanto, ao tomar $a(p)=(1-p)^n$, $b(x)=\binom{n}{x}I_{(0,1,\ldots,n)}(x)$, $c(p)=ln\left(\frac{p}{1-p}\right)$ e $d(x)=x$, verifica-se que X pertence a família exponencial. Com o intuito de utilizar o Teorema **??**, *vamos calcular as seguintes derivadas:*

- $\frac{\partial}{\partial p} lna(p) = \frac{\partial}{\partial p} ln[(1-p)^n] = n\frac{\partial}{\partial p} ln(1-p) = -\frac{n}{1-p}$;

- $\frac{\partial^2}{\partial p^2} lna(p) = \frac{\partial}{\partial p}\left(-\frac{n}{1-p}\right) = -n(1-p)^{-2}$;

- $\frac{\partial}{\partial p} c(p) = \frac{\partial}{\partial p} ln\left(\frac{p}{1-p}\right) = \frac{1}{p(1-p)}$;

- $\frac{\partial^2}{\partial p^2} c(p) = \frac{\partial}{\partial p}\left(\frac{1}{p(1-p)}\right) = \frac{2p-1}{p^2(1-p)^2}$.

Dessa forma, podemos calcular a esperança e a variância de X como segue-se:

(a)

$$\mathbb{E}\left[\frac{\partial}{\partial p} c(p)d(X)\right] = -\frac{\partial}{\partial p} lna(p)$$
$$\mathbb{E}\left[\frac{1}{p(1-p)}X\right] = -\left(-\frac{n}{1-p}\right)$$
$$\frac{1}{p(1-p)}\mathbb{E}\left[X\right] = \frac{n}{1-p}$$

$$\boxed{\mathbb{E}[X] = np}$$

(b)

$$Var\left[\frac{\partial}{\partial p} c(p)d(X)\right] = -\frac{\partial^2}{\partial p^2} lna(p) - \mathbb{E}\left[\frac{\partial^2}{\partial p^2} c(p)d(X)\right]$$
$$Var\left[\frac{1}{p(1-p)}X\right] = -\left(-n(1-p)^{-2}\right) - \mathbb{E}\left[\frac{2p-1}{p^2(1-p)^2}X\right]$$
$$\frac{1}{p^2(1-p)^2}Var[X] = -\left(-n(1-p)^{-2}\right) - \frac{2p-1}{p^2(1-p)^2}(np)$$

$$\boxed{Var[X] = np(1-p)}$$

Exercício 2.10 *Para cada uma das seguintes distribuições, verifique as equações para $\mathbb{E}[X]$ e $Var[X]$ dadas pelo Teorema* **??**.

1. *Verifique $Var[X]$ se X tem distribuição de Poisson(λ). (Dica: calcule $E[X(X-1)] = E[X^2] - E[X]$).*

2. *Verifique $Var[X]$ se X tem uma distribuição binomial negativa(r, p).*

3. *Verifique $E[X]$ e $Var[X]$ se X tem uma distribuição beta(α, β).*

Para obter mais que uma introdução às famílias exponenciais, veja Lehmann (1986, Seção 2.7).

2.3 Modelos de Locação e Escala

Nesta Seção abordaremos três técnicas para a construção de famílias de distribuições. As famílias resultantes têm prontas interpretações físicas que as tornam úteis para a modelagem, assim como têm propriedades matemáticas convenientes. Os três tipos de famílias são chamadas de locação, de escala e de locação-escala. Cada uma das famílias é construída pela especificação de uma única f.d.p.

Definição 2.11 *(**Modelo de locação**) Seja $f(x)$ qualquer fdp. Então, a família de fdps $f(x-\mu)$, indexada pelo parâmetro μ ($-\infty < \mu < \infty$), é chamada de família de locação com fdp padrão $f(x)$ e μ é chamado de parâmetro de locação do modelo.*

Observação: O efeito de mudar o valor de um parâmetro de locação μ é deslocar horizontalmente o gráfico de $f(x)$, mas sem alterar o formato do gráfico.

Exemplo 2.12 *Seja X uma v.a. com distribuição $N(\mu, \sigma^2)$ com σ^2 conhecida. Então,*

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$$

.

Tomando $u = x - \mu$, temos que $f(u) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\{-\frac{1}{2\sigma^2}(u)^2\}$ e dessa forma, X tem distribuição que é um modelo de locação com parâmetro de locação μ.

Definição 2.13 *(**Modelo de escala**) Seja $f(x)$ qualquer fdp. Então, para qualquer $\sigma > 0$, a família de fdps $(1/\sigma)f(x/\sigma)$, indexada pelo parâmetro σ, é chamada de família de escala com fdp padrão $f(x)$ e σ é chamado de parâmetro de escala do modelo.*

Observação: O efeito de mudar o valor de um parâmetro de escala σ é o de alongar ($\sigma > 1$) ou contrair ($\sigma < 1$) o gráfico de $f(x)$ enquanto ainda mantém o mesmo formato básico do gráfico.

Exemplo 2.14 *Diversas são as famílias que são de escala ou as têm como subfamílias. Como, por exemplo, as famílias gama se α for um valor fixo e β, o parâmetro de escala; normal, se μ conhecida e σ parâmetro de escala e a família exponencial.*

Definição 2.15 *(**Modelo de locação-escala**) Seja $f(x)$ qualquer fdp. Então, para qualquer μ ($-\infty < mu < \infty$) e qualquer $\sigma > 0$, a família de fdps $(1/\sigma)f((x-\mu)/\sigma)$, indexada pelos parâmetros (μ, σ), é chamada de família de locação-escala com fdp padrão $f(x)$; μ é chamado de parâmetro de locação e σ de parâmetro de escala.*

Observação: O efeito de introduzir os parâmetros de locação e de escala é o de alongar ($\sigma > 1$) ou contrair ($\sigma < 1$) o gráfico com o parâmetro de escala, e então, modificá-lo, de modo que o ponto que estava acima de 0 agora fique acima de μ.

Exemplo 2.16 *Seja $X \sim T_v(\mu, \sigma^2)$, v conhecido. Portanto,*

$$f(x|\mu, \sigma^2) = \frac{\Gamma(\frac{v+1}{2})v^{\frac{v}{2}}}{\Gamma(\frac{v}{2})\sqrt{\pi}} \frac{1}{\sigma} \left[v + \left(\frac{x-\mu}{\sigma}\right)^2 \right]^{-\frac{v+1}{2}}$$

.

Tomando $f(u) = \frac{\Gamma(\frac{v+1}{2})v^{\frac{v}{2}}}{\Gamma(\frac{v}{2})\sqrt{\pi}} \left[v + u^2\right]^{-\frac{v+1}{2}}$, temos que $f(x|\mu, \sigma^2) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, e portanto, este é um modelo de locação-escala com μ parâmetro de locação e σ parâmetro de escala.

Exercício 2.17 *Para cada um dos casos, verifique se o modelo é de locação, escala ou locação-escala.*

1. *Pareto(b,a), b conhecido.*

2. *$U[\theta, \theta + 1]$.*

2.4 O princípio da suficiência

Seja θ um parâmetro de interesse. O princípio da suficiência estabelece que uma estatística $T(\underline{X})$ é dita suficiente para o parâmetro θ se ela captura toda a informação sobre θ contida na amostra, ou seja, se $X_1, \ldots, X_n$ é uma amostra retirada da população, $T(\underline{X})$ é uma estatística suficiente para θ se qualquer inferência sobre θ depende da amostra $\underline{X}$ somente do valor $T(\underline{X})$, ou seja, se $\underline{X}$ e $\underline{Y}$ são duas amostras tais que $T(\underline{x}) = T(\underline{y})$, então a inferência sobre θ é a mesma independente da observação $\underline{X}$ ou $\underline{Y}$.

Nesta seção, investigaremos alguns aspectos de estatísticas suficientes e o princípio da suficiência.

Definição 2.18 ***(Estatística suficiente)*** *Seja $X_1, \ldots, X_n$ uma a.a. de $f(.|\theta)$, onde θ pode ser um vetor. Uma estatística $T = T(X_1, \ldots, X_n)$ é definida como estatística suficiente se e somente se a distribuição condicional de $X_1, \ldots, X_n$ dado $T = t$ não depende de θ para qualquer valor t de T.*

Exemplo 2.19 *Sejam X_1, X_2, X_3 uma amostra de tamanho 3 de uma distribuição de Bernoulli de parâmetro p. Considere as duas estatísticas $S = X_1 + X_2 + X_3$ e $T = X_1.X_2 + X_3$. Verifique se S e T são estatísticas suficientes.*

Solução:

	Valores de S	*Valores de T*	$f(x_1, x_2, x_3\|s)$	$f(x_1, x_2, x_3\|t)$
(0,0,0)	*0*	*0*	*1*	$\frac{1-p}{1+p}$
(0,0,1)	*1*	*1*	*1/3*	$\frac{1-p}{1+2p}$
(0,1,0)	*1*	*0*	*1/3*	$\frac{p}{1+p}$
(1,0,0)	*1*	*0*	*1/3*	$\frac{p}{1+p}$
(0,1,1)	*2*	*1*	*1/3*	$\frac{p}{1+2p}$
(1,0,1)	*2*	*1*	*1/3*	$\frac{p}{1+2p}$
(1,1,0)	*2*	*1*	*1/3*	$\frac{p}{1+2p}$
(1,1,1)	*3*	*2*	*1*	*1*

Para encontrarmos os valores associados às $f(x_1, x_2, x_3|s)$ e $f(x_1, x_2, x_3|t)$ basta fazermos as contas análogas ao caso abaixo:

- $$f_{(X_1,X_2,X_3|S=1)}(0,1,0|1) = P(X_1 = 0, X_2 = 1, X_3 = 0|S = 1)$$
$$= \frac{P(X_1 = 0; X_2 = 1; X_3 = 0; S = 1)}{P(S = 1)} = \frac{P(X_1 = 0)P(X_2 = 1)P(X_3 = 0)}{P(0,0,1) + P(0,1,0) + P(1,0,0)} = \frac{(1-p)p(1-p)}{3p(1-p)^2} = \frac{1}{3}$$

- $$f_{(X_1,X_2,X_3|T=0)}(0,1,0|0) = \frac{P(X_1 = 0, X_2 = 1, X_3 = 0|T = 0)}{P(T = 0)}$$
$$= \frac{P(X_1 = 0)P(X_2 = 1)P(X_3 = 0)}{P(0,0,0) + P(0,1,0) + P(1,0,0)} = \frac{(1-p)^2 p}{(1-p)^3 + 2(1-p)^2 p} = \frac{p}{1-p}$$

Uma vez que a distribuição condicional de $X_1, \ldots, X_n$ dado $S = s$ não depende de p, então S é uma estatística suficiente para p. Por outro lado, T não é estatística suficiente para p pois a distribuição condicional da amostra dado $T = t$ depende deste parâmetro.

Teorema 2.20 *Se $f(x_1, \ldots, x_n|\theta) = f(\underline{x}|\theta)$ é a densidaed conjunta de uma amostra aleatória $\underline{X} = (X_1, \ldots, X_n)$ e $g(t|\theta)$ é densidadde de $T = T(X_1, \ldots, X_n)$. T é uma estatística suficiente para θ se, para cada $\underline{x}$ no espaço amostral, a razão:*

$$\frac{f(\underline{x}|\theta)}{g(s|\theta)}$$

é constante como função de θ.

Exemplo 2.21 *Sejam $X_1, \ldots, X_n$ com distribuição $N(\mu, \sigma^2)$ i.i.d, com σ^2 conhecida. Queremos demonstrar que a média amostral, $T(\underline{X}) = \bar{X} = (X_1 + \ldots + X_n)/n$, é uma estatística suficiente para μ. A fdp conjunta da amostra X é:*

$$\begin{aligned}
f(x|\mu) &= \prod (2\pi\sigma^2)^{-1/2} exp\left\{-(x_i - \mu)^2/(2\sigma^2)\right\} \\
&= (2\pi\sigma^2)^{-n/2} exp\left\{-\sum (x_i - \mu)^2/(2\sigma^2)\right\} \\
&= (2\pi\sigma^2)^{-n/2} exp\left\{-\sum (x_i - \bar{x} + \bar{x} - \mu)^2/(2\sigma^2)\right\} \\
&= (2\pi\sigma^2)^{-n/2} exp\left\{-\left(\sum (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2\right)/(2\sigma^2)\right\}
\end{aligned}$$

Deste modo, a razão de fdps é

$$\begin{aligned}
\frac{f(\underline{x}|\theta)}{g(T(\underline{x})|\theta)} &= \frac{(2\pi\sigma^2)^{-n/2} exp\left\{-\left(\sum (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2\right)/(2\sigma^2)\right\}}{(2\pi\sigma^2/n)^{-1/2} exp\left\{-n(\bar{x} - \mu)^2/(2\sigma^2)\right\}} \\
&= n^{-1/2}(2\pi\sigma^2)^{-(n-1)/2} exp\left\{-\sum (x_i - \bar{x})^2/(2\sigma^2)\right\},
\end{aligned}$$

*que não depende de μ. Portanto, pelo Teorema **??**, a média amostral é uma estatística suficiente para μ.*

Exercício 2.22 *Sejam $X_1, \ldots, X_n$ variáveis aleatórias de Bernoulli i.i.d, com parâmetro θ ($0 < \theta < 1$). Mostre que $T(\underline{X} = X_1, \ldots, X_n) = \sum X_i$ é uma estatística suficiente para θ.*

2.5 Critério de fatoração

Teorema 2.23 ***(Fatoração de Neyman)*** *Seja $f(\underline{x}|\theta)$ a função densidade de probabilidade ou função de probabilidade conjunta de uma amostra $\underline{X}$. Uma estatística $T(\underline{X})$ é suficiente para θ se, e somente se, existem funções $g(t|\theta)$ e $h(\underline{x})$ tais que, para qualquer ponto amostral $\underline{x}$ e θ no espaço paramétrico, vale a igualdade*

$$f(\underline{x}|\theta) = g(T(\underline{x})|\theta)h(\underline{x}).$$

Para utilizar o Teorema da Fatoração a fim de encontrar uma estatística suficiente, fatoramos a fdp conjunta da amostra em duas partes, sendo que uma parte não depende de θ. Esta parte constitui a função $h(\underline{x})$. A outra, dependente de θ, e da amostra $\underline{x}$ somente por meio de alguma função $T(\underline{x})$. Esta função é uma estatística suficiente para θ. Isto é ilustrado como no exemplo a seguir.

Exemplo 2.24 *Para o modelo normal com σ conhecido, descrito anteriormente, vimos que a fdp poderia ser fatorada como*

$$f(\underline{x}|\mu) = (2\pi\sigma^2)^{-n/2} exp\left\{-\sum (x_i - \bar{x}^2)\right\} exp\left\{-n\left(\bar{x} - \mu\right)^2)/(2\sigma^2)\right\}.$$

Podemos definir

$$h(\underline{x}) = (2\pi\sigma^2)^{-n/2} exp\left\{-\sum(x_i - \bar{x}^2)\right\},$$

que não depende do parâmetro desconhecido μ. O fator que contém μ depende da amostra $\underline{x}$ somente pela função $T(\underline{x}) = \bar{x}$, a média amostral. Assim, temos

$$g(t|\mu) = exp\left\{-n\left(t-\mu)^2\right)/(2\sigma^2)\right\}$$

e observamos que

$$f(\underline{x}|\mu) = h(\underline{x})g(T(\underline{x})|\mu).$$

Assim, de acordo com o Teorema da Fatoração de Neyman, $T(X) = \bar{X}$ é uma estatística suficiente para μ.

Exercício 2.25 *Seja uma a.a. $X_1, \ldots, X_n$ de uma população com função de densidade $f(x|\theta) = \theta x^{-2} I_{[\theta,\infty]}(x)$, onde $\theta > 0$. Testar se $Y_1 = min(X_1, \ldots, X_n)$ é estatística suficiente para θ.*

Teorema 2.26 *Sejam $X_1, \ldots, X_n$ observações i.i.d. a partir de uma fdp ou fp $f(.|\underline{\theta})$ que pertence a uma família exponencial dada por:*

$$f(.|\underline{\theta}) = a(\underline{\theta})b(\underline{x})exp\left\{\sum_{j=1}^{k} c_j(\underline{\theta})d_j(\underline{x})\right\},$$

onde $\underline{\theta} = (\theta_1, \ldots, \theta_k)$. Então,

$$T(\underline{X}) = \left(\sum_{i=1}^{n} d_1(X_i), \ldots, \sum_{i=1}^{n} d_k(X_i)\right)$$

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

Nesta seção, descobrimos uma estatística suficiente para cada modelo considerado. Na verdade, em qualquer problema existem muitas estatísticas suficientes. Por causa disso, podemos perguntar se uma estatística suficiente é, de algum modo, melhor que outra. Lembra-se de que o propósito de uma estatística suficiente é obter a redução de dados sem perder informações sobre o parâmetro θ; deste modo, uma estatística que consegue a maior redução de dados enquanto ainda mantém todas as informações sobre θ pode ser considerada preferível. A definição de tal estatística é formalizada a seguir.

Definição 2.27 ***(Estatística suficiente minimal)*** *Uma estatística suficiente $T(\underline{X})$ é chamada de estatística suficiente minimal se, para qualquer outra estatística suficiente $T'(\underline{X}), T(\underline{X})$ é uma função de $T'(\underline{X})$.*

Não é interessante utilizar a Definição **??** para encontrar uma estatística suficiente minimal, assim como não é prático utilizar a Definição **??** para encontrar uma estatística suficiente. O teorema abaixo nos fornece uma maneira prática para encontrar uma estatística suficiente minimal.

Teorema 2.28 *Seja $f(\underline{x}|\theta)$ a função de probabilidade ou função densidade de probabilidade de uma amostra $\underline{X}$. Suponha que exista uma função $T(\underline{x})$ tal que, para quaisquer $\underline{x}$ e $\underline{y}$, a razão $f(\underline{x}|\theta)/f(\underline{y}|\theta)$ é constante se, e somente se, $T(\underline{x}) = T(\underline{y})$. Então $T(\underline{X})$ é uma estatística suficiente minimal para θ.*

Exemplo 2.29 *Seja $X_1, \ldots, X_n$ uma amostra independente e igualmente distribuída de uma distribuição normal com média μ e variância σ^2 desconhecidas e sejam $\boldsymbol{x}$ e $\underline{y}$ dois pontos amostrais tais que $(\bar{x}, s^2_{\underline{x}})$ e $(\bar{y}, s^2_{\underline{y}})$ são as médias e variâncias correspondentes a $\underline{x}$ e $\underline{y}$ respectivamente. Neste caso, temos que*

$$\frac{f(\underline{x}|\mu,\sigma^2)}{f(\underline{y}|\mu,\sigma^2)} = \frac{(2\pi\sigma^2)^{-n/2}\exp\left(-[n(\bar{x}-\mu)^2+(n-1)s^2_{\boldsymbol{x}}]/(2\sigma^2)\right)}{(2\pi\sigma^2)^{-n/2}\exp\left(-[n(\bar{y}-\mu)^2+(n-1)s^2_{\boldsymbol{y}}]/(2\sigma^2)\right)}$$

de onde concluímos que

$$\frac{f(\underline{x}|\mu,\sigma^2)}{f(\boldsymbol{y}|\mu,\sigma^2)} = \exp\left([-n(\bar{x}^2-\bar{y}^2)+2n\mu(\bar{x}-\bar{y})-(n-1)(s^2_{\boldsymbol{x}}-s^2_{\boldsymbol{y}})]/(2\sigma^2)\right)$$

que é constante em relação aos parâmetros μ e σ^2 se, e somente se $\bar{x} = \bar{y}$ e $s^2_{\underline{x}} = s^2_{\underline{y}}$. Segue do Teorema de Lehmann-Scheffé que $(\bar{X}, S^2)$ é uma estatística suficiente minimal para (μ, σ^2).

2.6 O princípio da verossimilhança

Nesta seção, estudaremos uma estatística importante, específica, chamada de função de verossimilhança, que também pode ser utilizada para resumir dados. Veremos nesta seção que, se alguns princípios são aceitos, a função de verossimilhança deve ser utilizada como um dispositivo de redução dos dados.

Definição 2.30 ***(A função de verossimilhança)*** *Seja $f(\underline{x}|\theta)$ a função de probabilidade ou função densidade de probabilidade de uma amostra $\underline{X} = (X_1, \ldots, X_n)$. Então, dado que $\underline{X} = \underline{x}$ é observado, a função de θ definida por*

$$L(\theta|\underline{x}) = f(\underline{x}|\theta)$$

é chamada de função de verossimilhança.

Exemplo 2.31 *Considere 10 ensaios de Bernoulli independentes com parâmetro p, $(0 < p < 1)$ e a variável aleatória X definida como a soma dos valores observados nos ensaios. Neste caso, sabemos que X tem distribuição binomial com parâmetros 10 e p, isto é,*

$$X \sim \; \mathit{Binomial}(10, p)$$

Suponha que, a partir da amostra, temos que o valor observado seja $X = 3$. Então a função de verossimilhança é dada por

$$L(\theta|3) = \mathbb{P}(X=3) = \left(\begin{array}{c}10\\3\end{array}\right)p^3(1-p)^7.$$

Para o caso geral, temos que, se $X = x$ é observado, então a função de verossimilhança é

$$L(\theta|x) = \mathbb{P}(X=x) = \left(\begin{array}{c}10\\x\end{array}\right)p^x(1-p)^{10-x}.$$

Neste Caso foi observado apenas um valor de X, por tanto temos uma a.a. de tamanho 1.

O princípio da verossimilhança especifica como a função de verossimilhança deve ser usada para resumo dos dados.

<u>Observação:</u> Vale ressaltar que a diferença entre a função de probabilidade (ou função densidade de probabilidade) e a função de verossimilhança está justamente em qual é a variável da função. Quando consideramos a função de probabilidade ou função densidade de probabilidade, θ é fixo enquanto $\underline{x}$ é variável e quando consideramos a função de verossimilhança, a amostra observada $\underline{x}$ é fixa enquanto θ é variável em relação aos valores paramétricos possíveis.

Definição 2.32 ***(Princípio da verossimilhança)*** *Suponha que $\underline{x}$ e $\underline{y}$ sejam dois pontos amostrais tais que $L(\theta|\underline{x})$ é proporcional a $L(\theta|\underline{y})$, ou seja, existe uma constante $C(\underline{x},\underline{y})$ tal que*

$$L(\theta|\underline{x}) = C(\underline{x},\underline{y})L(\theta|\underline{y}), \ \textit{para todo } \theta$$

então as conclusões obtidas a partir de $\underline{x}$ e $\underline{y}$ devem ser idênticas.

Observação: No caso especial de $C(\underline{x},\underline{y}) = 1$, o Princípio da Verossimilhança define que se dois pontos amostrais resultam na mesma função de verossimilhança, então eles contém as mesmas informações sobre θ. Mas o Princípio da Verossimilhança vai além, pois estabelece que mesmo que dois pontos amostrais tenham somente verossimilhanças proporcionais, eles contém informações equivalentes sobre θ.

Capítulo 3

Estimação pontual

A lógica por trás da estimação pontual é bastante simples. Quando a amostragem é feita a partir de uma população descrita por uma fdp ou fp $f(x|\theta)$, o conhecimento de θ gera o da população inteira. Desse modo, é natural procurar um método para encontrar um bom estimador de θ, isto é, um bom estimador pontual. Também pode ser o caso de que alguma função de θ, digamos $\tau(\theta)$, seja de interesse. Os métodos descritos neste capítulo podem ser utilizados para se obter estimadores de $\tau(\theta)$.

Definição 3.1 *(**Estimador pontual**) Qualquer estatística cujos valores são usados para estimar $\tau(\theta)$, onde $\tau(.)$ é uma função do parâmetro θ, é definida como estimador pontual de $\tau(\theta)$.*

Existe uma distinção que deve ser esclarecida: a diferença entre uma estimativa e um estimador. Um **estimador** é uma função da amostra, ao passo que uma **estimativa** é o valor observado de um estimador (isto é, um número) que é obtido quando uma amostra é efetivamente obtida.

Em alguns casos existe um candidato natural ou obvio à estimador pontual de um parâmetro em particular porém, é útil ter algumas técnicas que possam fornecer alguns candidatos razoáveis a considerar. Estas técnicas não oferecem nenhuma garantia com relação aos estimadores por elas geradas, ainda é necessário avaliá-los antes de usá-los.

3.1 Métodos para encontrar estimadores

Em alguns casos, é uma tarefa fácil decidir como encontrar a estimativa de um parâmetro e, geralmente, a intuição, por si só, pode nos levar a estimadores muito bons. Em modelos mais complicados, aqueles que geralmente surgem na prática, precisamos de um meio mais metodológico para a estimação dos parâmetros. Nesta seção serão apresentados alguns métodos para encontrar estimadores, mais precisamente, três métodos.

3.1.1 Método dos momentos

Este método é baseado nos momentos teóricos e amostrais das variáveis aleatórias envolvidas. Ele tem a virtude de ser bastante simples em sua utilização e quase sempre gera algum tipo de estimativa que, mesmo precisando ser melhoradas em muitos casos, são um ponto de partida.

Definição 3.2 *(Método dos momentos) Seja $X_1, \ldots, X_n$ uma amostra de uma população com fdp ou fp $f(x|\theta_1, \ldots, \theta_k)$. Estimadores pelo método de momentos são encontrados igualando-se os primeiros k momentos amostrais aos correspondentes momentos da população e resolvendo o sistema resultante de equações simultâneas. Definimos:*

- *Momento amostral:* $m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$.
- *Momento populacional:* $\mu'_k = \mathbb{E}[X^k]$.

Exemplo 3.3 *Seja $X_1, X_2, \ldots, X_m$ uma amostra aleatória de uma população com distribuição binomial de parâmetros n e p, ou seja,*

$$\mathbb{P}(X_i = x|n,p) = \left(\begin{array}{c} n \\ x \end{array}\right) p^x(1-p)^{n-x}, \quad x = 0, 1, \ldots, n.$$

Assumindo que os parâmetros n e p sejam desconhecidos, vamos encontrar estimadores para ambos os parâmetros a partir do método dos momentos. Usando o fato de que X tem média np e variância $np(1-p)$, temos que os dois primeiros momentos populacionais são dados, respectivamente, por

$$\mu_1 = \mathbb{E}(X) = np \quad e \quad \mu_2 = \mathbb{E}(X^2) = np(1-p) + n^2p^2$$

Igualando os dois primeiros momentos amostrais m_1 e m_2 aos dois primeiros momentos populacionais, temos o seguinte sistema de equações

$$\left\{ \begin{array}{l} \overline{X} = np \\ \frac{1}{m}\sum_{i=1}^{m} X_i^2 = np(1-p) + n^2p^2 \end{array} \right. .$$

Resolvendo em n e p, obtemos os seguintes estimadores pelo método dos momentos

$$\hat{n} = \frac{\overline{X}^2}{\overline{X} - (1/m)\sum_{i=1}^{m}(X_i - \overline{X})^2} \quad e \quad \hat{p} = \frac{\overline{X}}{\hat{n}}.$$

Este é um típico exemplo em que os estimadores não são os melhores para os parâmetros populacionais de interesse. Na verdade, utilizando estes estimadores, podemos ter estimativas negativas para n e p, o que não pode acontecer, já que estes devem ser números positivos.

Exercício 3.4 *Seja $X_1, \ldots, X_n$ i.i.d. com distribuição $N(\theta, \sigma^2)$. Achar o estimador de momentos para θ e σ^2*

Definição 3.5 *Dizemos que $\hat{\theta}_1, \ldots, \hat{\theta}_r$ são estimadores obtidos pelo método dos momentos se eles forem soluções das equações*

$$m_n = \mu_n, \quad n = 1, 2, \ldots, k.$$

Exemplo 3.6 *Seja X uma variável aleatória com média μ e variância σ^2. É conhecido que $\mu = \mathbb{E}(X)$ e $\sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$. Além disso, os dois primeiros momentos amostrais são dados, respectivamente, por*

$$m_1 = \frac{1}{k}\sum_{i=1}^{k} X_i = \overline{X}, \quad m_2 = \frac{1}{k}\sum_{i=1}^{k} X_i^2.$$

Neste caso, os estimadores para a média populacional μ e a variância populacional σ^2 obtidos pelo método dos momentos serão

$$\hat{\mu}_M = m_1 = \overline{X},$$

$$\hat{\sigma}^2_M = m_2 - m_1^2 = \frac{1}{k}\sum_{i=1}^{k} X_i^2 - \overline{X}^2$$

3.1.2 Método de máxima verossimilhança

O método de máxima verossimilhança é, de longe, a técnica mais popular para derivar estimadores. Consideremos uma população e uma variável aleatória X, relacionada a essa população, com função de probabilidade (se X é uma variável aleatória discreta) ou função densidade de probabilidade (se X é uma variável aleatória contínua) $f(x, \theta)$, sendo θ o parâmetro desconhecido. Retiremos uma amostra aleatória simples de X, de tamanho n, $X_1, \ldots, X_n$, e sejam $x_1, \ldots, x_n$ os valores efetivamente observados.

A função de verossimilhança L é definida por

$$L(\theta; x_1, \ldots, x_n) = f(x_1; \theta) \times \ldots \times f(x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta).$$

Definição 3.7 ***(Estimador de máxima verossimilhança)*** *Para cada ponto amostral $\underline{x}$, seja $\hat{\theta}(\underline{x})$ um valor de parâmetro no qual $L(\theta|\underline{x})$ atinge seu máximo como uma função de θ, com $\underline{x}$ mantido fixo. Um estimador de máxima verossimilhança (EMV) do parâmetro θ com base em uma amostra $\underline{X}$ é $\hat{\theta}(\underline{X})$.*

Observação: Em muitos casos, o estimador de máxima verossimilhança pode ser encontrado seguindo os passos abaixo:

1. encontrar a função de verossimilhança;
2. aplicar a função ln;
3. derivar em relação ao parâmetro $\underline{\theta}$;
4. igualar o resultado a zero, substituindo θ por $\hat{\theta}$;
5. colocar em evidência $\hat{\theta}$
6. verificar que este estimador é ponto de máximo global.

Observaçaão: Intuitivamente, o EMV é uma escolha razoável para um estimador. Ele é o valor do parâmetro para o qual a amostra observada é mais provável.

Exemplo 3.8 *Seja X uma variável aleatória com distribuição de Poisson e parâmetro λ. Tomemos uma amostra aleatória $X_1, \ldots, X_n$ de X. Qual é o estimador de máxima verossimilhança para λ?*

Solução: *Como $X \sim$ Poisson(λ), a função de probabilidade de X é*

$$f_\lambda(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad k \in \mathbb{N}.$$

Desta forma, a função de verossimilhança é dada por

$$L(\lambda; x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

Ou seja,

$$L(\lambda; x_1, \ldots, x_n) = \frac{1}{\prod_{i=1}^{n} x_i!} \lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}.$$

Para encontrar o estimador de máxima verossimilhança para λ, devemos encontrar o valor de λ para o qual a função de verossimilhança $L(\lambda; x_1, \ldots, x_n)$ é máxima.

Aplicamos a função logaritmo natural (ln) na função de verossimilhança $L(\lambda; x_1, \ldots, x_n)$. Desta forma, temos que

$$\ln L(\lambda; x_1, \ldots, x_n) = \ln\left(\frac{1}{\prod_{i=1}^{n} x_1!}\right) + \sum_{i=1}^{n} x_i \ln \lambda - n\lambda$$

e, derivando em relação a λ, segue que

$$\frac{d \ln L(\lambda; x_1, \ldots, x_n)}{d\lambda} = \frac{1}{\lambda} \sum_{i=1}^{n} x_i - n.$$

Igualando o resultado a zero, segue que

$$\frac{1}{\hat{\lambda}} \sum_{i=1}^{n} x_i - n = 0 \Leftrightarrow \hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}.$$

Neste caso, o possível estimador de máxima verossimilhança para o parâmetro λ é $\hat{\lambda} = \overline{X}$. Basta verificar se este ponto é realmente um ponto de máximo. Para isto, vamos calcular a segunda derivada de $\ln L(\lambda; x_1, \ldots, x_n)$.

$$\frac{d^2 \ln L(\lambda; x_1, \ldots, x_n)}{d\lambda^2} = -\frac{1}{\lambda^2} \sum_{i=1}^{n} x_i < 0.$$

Portanto, concluímos que $\hat{\lambda} = \overline{X}$ é um estimador de máxima verossimilhança para o parâmetro λ.

Exemplo 3.9 *Seja X uma variável aleatória com distribuição Normal com média μ e variância σ^2. Tomemos uma amostra aleatória $X_1, \ldots, X_n$ de X. Qual o estimador de máxima verossimilhança para $\theta = (\mu, \sigma^2)$?*

Solução: *Como $X \sim N(\mu, \sigma^2)$, a função densidade de X é*

$$f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad -\infty < x < \infty.$$

Assim, a função de verossimilhança é dada por

$$L(\mu, \sigma^2; x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right]$$

Ou seja,

$$L(\mu, \sigma^2; x_1, \ldots, x_n) = (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2\right].$$

Para encontrar o estimador de máxima verossimilhança para $\theta = (\mu, \sigma^2)$ devemos encontrar os valores de μ e σ^2 para os quais a função de verossimilhança, $L(\mu, \sigma^2; x_1, \ldots, x_n)$, é máxima.

Para isso primeiramente aplicaremos a função ln,

$$\ln L(\mu, \sigma^2; x_1, \ldots, x_n) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i - \mu)^2}{\sigma^2}.$$

Agora vamos derivar em relação a μ:

$$\frac{\partial L(\mu, \sigma^2; x_1, \ldots, x_n)}{\partial \mu} = -\frac{2}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)(-1) = \sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma^2}\right).$$

Igualando o resultado a zero obtemos:

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

E então, o possível estimador de máxima verossimilhança da média populacional μ é $\overline{X}$. Basta avaliar agora se realmente $\overline{x}$ é ponto de máximo. Para isto,

$$\frac{\partial^2 L(\mu, \sigma^2; x_1, \ldots, x_n)}{\partial \mu^2} = \frac{\partial^2}{\partial \mu^2}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu)\right] = -\frac{n}{\sigma^2} < 0.$$

Assim, concluimos que $\overline{x}$ é realmente um ponto de máximo e, portanto, o estimador de máxima verossimilhança para μ é $\hat{\mu} = \overline{X}$. Vamos agora encontrar o estimador de máxima verossimilhança para a variância σ^2. Para isso, derivamos a função em relação a σ^2:

$$\frac{\partial \ln L(\mu, \sigma^2; x_1, \ldots, x_n)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma^2}\right)^2.$$

Igualando a zero, temos que

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

substituindo μ pelo seu EMV, temos

$$\hat{\sigma}^2 = \frac{(n-1)}{n}s^2.$$

Como

$$\frac{\partial^2 L(\mu, \sigma^2; x_1, \ldots, x_n)}{\partial (\sigma^2)^2} = \frac{1}{(\sigma^2)^2}\left(\frac{n}{2} - \frac{(n-1)s^2}{\sigma^2}\right)$$

que, avaliado em $\hat{\sigma}^2 = \frac{(n-1)s^2}{n}$ é tal que

$$\frac{\partial^2 L(\mu, \hat{\sigma}^2; x_1, \ldots, x_n)}{\partial (\sigma^2)^2} = -\frac{n}{2}\frac{1}{\hat{\sigma}^2} < 0.$$

Portanto, o estimador de máxima verossimilhança para σ^2 é $\hat{\sigma}^2 = \dfrac{(n-1)}{n}s^2$, onde $s^2 = \dfrac{\sum_{i=1}^{n}(x_i - \mu)^2}{n-1}$.

Exercício 3.10 *Seja $X_1, \ldots, X_n$ i.i.d. com distribuição de Bernoulli(p). Provar que $\hat{p} = \bar{X}$ é EMV.*

Propriedades dos EMVs

1. **Invariância por transformações biunívocas**

   Se $\hat{\theta}$ é o EMV de θ, então, para qualquer função biunívoca $\tau(\theta)$ de θ, o EMV de $\tau(\theta)$ é $\tau(\hat{\theta})$.

   Exemplo 3.11 *Seja $X_1, \ldots, X_n$ uma a.a. com distribuição $N(\theta, \sigma^2)$. Sabemos que o EMV para θ é $\hat{\theta} = \bar{X}$. Dessa forma, se definirmos $\tau(\theta) = \theta + 3$, então $\tau(\hat{\theta}) = \bar{X} + 3$ é o EMV de $\tau(\theta)$.*

2. **O EMV não depende do plano amostral**

   Se experimentos diferentes fornecem verossimilhança proporcionais, os EMV serão idênticos.

3. **O EMV pode não existir**

   Exemplo 3.12 *Seja X uma variável aleatória uniformemente distribuída em $(0, \theta)$ e $X_1, \ldots, X_n$ uma amostra aleatória de X. Esta distribuição tem função densidade de probabilidade dada por*

   $$f(x|\theta) = \begin{cases} \frac{1}{\theta}, & se\ 0 < x < \theta, \\ 0 & caso\ contrário \end{cases}$$

   A função de verossimilhança é dada por

   $$L(\theta; X_1, \ldots, X_n) = \prod_{i=1}^{n} f(X_i|\theta) = \frac{1}{\theta^n} I_{(0,\theta)\}}(X_1, \ldots, X_n) = \frac{1}{\theta^n} I_{(\max\{X_1,\ldots,X_n\},\infty)}(\theta).$$

   e, como $\frac{1}{\theta^n}$ é uma função decrescente de θ, a estimativa será o menor valor possível de θ para o qual $\theta \geq X_i$ para $i = 1, \ldots, n$. Este valor é $\theta = \max\{X_1, \ldots, X_n\}$. Entretanto, como o intervalo é aberto, não existe máximo, logo não existe EMV.

4. **O EMV pode não ser único**

   Exercício 3.13 *Para provar que existem casos em que o EMV não é único pense no caso em que $X \sim U[\theta, \theta + 1]$.*

3.1.3 Método de mínimos quadrados

O método de estimação por mínimos quadrados consiste em minimizar o quadrado das diferenças entre os valores observados de uma amostra e seus respectivos valores esperados. Este método é amplamente utilizado em situações em que são definidas associações entre variáveis, como é o caso dos modelos lineares.

Definição 3.14 ***(Estimadores de mínimos quadrados)*** *Seja $(x_1, Y_1), \ldots, (x_n, Y_n)$ uma a.a. de $Y = \alpha + \beta X + \epsilon$, onde α e β são constantes não conhecidas. As estimativas de mínimos quadrados dos parâmetros α e β são aqueles valore de α e β que minimizam*

$$\sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} [Y_i - (\alpha + \beta x_i)]^2,$$

portanto $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x}$ *e* $\hat{\beta} = \dfrac{\sum_{i=1}^{n}(Y_i - \bar{Y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$

Exercício 3.15 *Encontrar os estimadores de mínimos quadrados definidos acima.*

Exemplo 3.16 *Suponha que estamos interessados em estudar a resistência Y de uma cabo de aço em função de seu diâmetro X. A partir de uma amostra coletada, percebemos que as variáveis são, aproximadamente, proporcionais, isto é, $Y \approx \theta X$ em que θ é o coeficiente de proporcionalidade. O nosso objetivo é estimar o parâmetro θ, baseado nas medidas disponíveis em uma amostra de 10 unidades mostradas na tabela a seguir A partir dessas informações, po-*

x	0,50	0,60	0,75	0,80	0,90	1,05	1,20	1,30	1,50	1,65
y	2,07	2,24	3,28	3,35	3,81	4,14	4,64	5,13	6,05	6,57

demos concluir que, aparentemente, $\hat{\theta} = 4$ parece ser uma estimativa razoável para o parâmetros θ. Como podemos verificar a qualidade desta estimativa? Uma forma de fazer isso é verificar as diferenças entre os valores observados Y e os valores esperados utilizando a estimativa, ou seja, $4X$.

A ideia principal do método baseia-se em minimizar o erro quadrático total da amostra. Para a estimativa $\hat{\theta} = 4$, este erro é dado por 0,213, porém, pode ser que exista alguma outra estimativa com erro quadrático total menor do que 0,213. Desta forma, o objetivo é minimizar a função

$$S(\theta) = \sum_{i=1}^{10} (Y_i - \theta X_i)^2.$$

O mínimo da função é obtido derivando a função em relação a θ e igualando o resultado a zero, ou seja, encontrar $\hat{\theta}$ para o qual

$$\frac{dS(\theta)}{d\theta} = \sum_{i=1}^{10} (Y_i - \hat{\theta} X_i)(-2X_i) = 0.$$

E, resolvendo esta equação, obtemos o estimador

$$\hat{\theta}_{MQ} = \frac{\sum_{i=1}^{10} X_i Y_i}{\sum_{i=1}^{10} X_i^2}.$$

Utilizando os dados de X e Y, encontramos $\hat{\theta}_{MQ} = 4,011625$, ou seja, a estimativa que minimiza o erro quadrático total da amostra é dada por $\hat{\theta} = 4,011625$. De fato, utilizando este valor, temos que o erro quadrático total é 0,2114015.

Neste caso, estamos assumindo que, para um dado valor da variável X, os valores da variável Y seguem uma distribuição de probabilidade $f_Y(y)$ centrada em θX, o que é equivalente a dizer que, para cada X, o desvio $\epsilon = Y - \theta X$ segue uma distribuição centrada em zero e, desta forma, é comum escrever

$$Y = \theta x + \epsilon$$

com ϵ seguindo a distribuição $f_\epsilon(\cdot)$ com média zero. Desta forma, é razoável escolher θ que minimiza a soma dos quadrados dos erros

$$\sum_{i=1}^{10} \epsilon^2 = \sum_{i=1}^{10} (Y_i - \theta X_i)^2.$$

Observamos que o modelo pode ser generalizado. Isto é, podemos considerar funções mais gerais do parâmetros θ, ou seja,

$$Y = g(X, \theta) + \epsilon$$

e, da mesma forma do exposto acima, devemos encontrar o valor de θ que minimize a função

$$S(\theta) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (Y_i - g(X_i, \theta))^2,$$

para uma amostra $(X_1, Y_1), \ldots, (X_n, Y_n)$ das variáveis X e Y. A solução $\hat{\theta}_{MQ}$ é chamada de estimador de mínimos quadrados (EMQ) de θ.

3.2 Propriedades dos estimadores

Os método discutidos na seção anterior descrevem técnicas razoáveis para encontrar estimadores pontuais de parâmetros. Entretanto, uma vez que, geralmente, podemos aplicar mais do que um desses métodos em uma situação, surge a pergunta: dadas certas condições, qual o melhor estimador?

Nesta seção, apresentaremos alguns critérios básicos para a avaliação de estimadores.

3.2.1 Erro quadrático médio

Definição 3.17 ***(Erro quadrático médio)*** *Seja $T = T(X_1, \ldots, X_n)$ um estimador de $\tau(\theta)$. Define-se o erro quadrático médio de T, denotado por $EQM_T(\theta)$, como:*

$$EQM_T(\theta) = \mathbb{E}[(T - \tau(\theta))^2]$$
$$= \int \ldots \int (T(x_1, \ldots, x_n) - \tau(\theta))^2 f(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

Observe que o EQM mede a diferença quadrática média entre o estimador T e a função do parâmetro $\tau(\theta)$. Será o melhor estimador aquele que tiver o menor EQM.

Uma vantagem de utilizarmos o EQM é que podemos defini-lo como

$$EQM_T(\theta) = Var[T] + B(T)^2$$

onde definimos o viés de um estimador - $B(T)$ como segue.

Definição 3.18 ***(Viés de um estimador)*** *O viés de um estimador pontual T de uma função $\tau(\theta)$ do parâmetro θ é a diferença entre o valor esperado de T e $\tau(\theta)$, isto é, $B(T) = \mathbb{E}[T] - \tau(\theta)$.*

Exemplo 3.19 *Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição normal com média μ e variância σ^2 e considere o estimador de máxima verossimilhança para σ^2, ou seja, o estimador*

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

Temos que

$$\mathbb{E}(\hat{\sigma}^2) = \mathbb{E}\left(\frac{n-1}{n} s^2\right) = \frac{n-1}{n} \sigma^2,$$

de forma que $\hat{\sigma}^2$ é um estimador viciado de σ^2. Calculando a variância de $\hat{\sigma}^2$, temos que

$$Var(\hat{\sigma}^2) = Var\left(\frac{n-1}{n} s^2\right) = \left(\frac{n-1}{n}\right)^2 Var(s^2) = \frac{2(n-1)\sigma^4}{n^2},$$

de forma que o erro quadrático médio $EQM(\hat{\sigma}^2)$ é dado por

$$EQM(\hat{\sigma}^2) = \mathbb{E}(\hat{\sigma}^2 - \sigma^2)^2 = \frac{2(n-1)\sigma^4}{n^2} + \left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2 = \left(\frac{2n-1}{n^2}\right)\sigma^4.$$

Desta forma,

$$EQM(\hat{\sigma}^2) = \left(\frac{2n-1}{n^2}\right)\sigma^4 < \left(\frac{2}{n-1}\right)\sigma^4 = EQM(s^2),$$

e, portanto, $\hat{\sigma}^2$ possui um erro quadrático médio menor do que s^2.

3.3 Estimador não viciado

Definição 3.20 *Um estimador $T = T(X_1, X_2, \cdots, X_n)$ é dito não viciado (não viesado) para alguma função $\tau(\theta)$ do parâmetro populacional θ se*

$$\mathbb{E}(T) = \tau(\theta),$$

para todo θ. Se a igualdade acima ocorre, dizemos que $B(T) = 0$ e portanto, $EQM(T) = Var[T]$.

Observação Pode parecer intuitivo que, ao utilizar estimadores não-viciados, tenhamos um erro quadrático médio pequeno, porém nem sempre isso ocorre, ou seja, controlar o vício do estimador não garante um controle do erro quadrático médio. As vezes, um estimador com um pequeno aumento no vício pode gerar um grande decrescimento na variância e, consequentemente, um erro quadrático médio menor.

Exemplo 3.21 *Seja $X_1, X_2, \cdots, X_n$ uma amostra aleatória de uma população com distribuição normal $N(\mu, \sigma^2)$. O estimador $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$ é viesado. De fato, observe primeiramente que se $X_i \sim N(\mu, \sigma^2)$ então*

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

*Este fato decorre do Corolário **??**. Lembramos que a distribuição qui-quadrado é dada por*

$$f(x) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2} exp\left(-\frac{x}{2}\right), \quad x \in (0, \infty)$$

Com isso em mente vamos calcular o valor esperado de s para mostrarmos que $\mathbb{E}(s) \neq \sigma$. De fato,

$$\mathbb{E}(s) = \mathbb{E}\left(\sqrt{\frac{n-1}{n-1}\frac{\sigma^2}{\sigma^2}s^2}\right) = \sqrt{\frac{\sigma^2}{n-1}}\mathbb{E}\left(\sqrt{\frac{n-1}{\sigma}s^2}\right)$$

$$= \sqrt{\frac{\sigma^2}{n-1}}\int_{\infty}^{0}\sqrt{x}\frac{1}{2^{(n-1)/2}\Gamma(\frac{(n-1)}{2})}x^{((n-1)/2)-1}\exp(-x/2)dx$$

$$= \sqrt{\frac{\sigma^2}{n-1}}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\int_{0}^{\infty}\frac{1}{2^{(n-1)/2}\Gamma(n/2)}x^{(n/2)-1}\exp(-x/2)dx$$

$$= \sqrt{\frac{\sigma^2}{n-1}}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\frac{2^{n/2}}{2^{(n-1)/2}}\int_{0}^{\infty}\frac{1}{2^{n/2}\Gamma(n/2)}x^{(n/2)-1}\exp(-x/2)dx$$

$$= \sigma\sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)} \approx \sigma\left(1 - \frac{1}{4n} - \frac{7}{32n^2} - \frac{19}{128n^3} + o(n^{-4})\right) \approx \sigma\left(1 - \frac{1}{4n}\right)$$

Logo o viés é dado por

$$\sigma - \mathbb{E}(s) = \sigma\left(1 - \sqrt{\frac{2}{n-1}}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\right) \approx \frac{\sigma}{4n}$$

Portanto, de fato temos que o estimador s é viesado embora s^2 não seja, pois como sabemos $\mathbb{E}(s^2) = \sigma^2$.

3.4 Consistência e melhores estimadores

Definição 3.22 ***(Consistência em EQM)*** *Seja T_n uma sequência de estimadores de $\tau(\theta)$. Dizemos que esta sequência é consistente com relação ao EQM se e somente se:*

$$\lim_{n\to\infty}[\mathbb{E}(T_n - \tau(\theta))^2] = 0 \;\; \forall\theta$$

Teorema 3.23 *Se para um estimador $\hat{\theta} = T$ de um parâmetro populacional θ verifica-se as seguintes condições:*

1. $\lim_{n\to\infty} \mathbb{E}[\hat{\theta}] = \theta$
2. $\lim_{n\to\infty} Var[\hat{\theta}] = 0$

então $\hat{\theta}$ é um estimador consistente de θ.

Exemplo 3.24 *Seja $X_1, \ldots, X_n$ uma a.a. de uma distribuição com média μ e variância σ^2. Seja $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ uma sequência de estimadores de μ e $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ uma sequência de estimadores de σ^2. Verifique se as sequências de estimadores são consistentes com relação aos EQMs.*

Solução: *Como $\bar{X}_n$ é estimador não viciado de μ e $\lim_{n\to\infty}[\mathbb{E}(\bar{X}_n - \mu)^2] = \lim_{n\to\infty} Var[\bar{X}_n] = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$, então, a sequência $\bar{X}_n$ é consistente com relação ao EQM para μ.*

A segunda parte fica como ***exercício****.*

Definição 3.25 ***(Consistência fraca)*** *Seja T_n uma sequência de estimadores de $\tau(\theta)$. Dizemos que esta sequência é fracamente consistente se para cada $\epsilon > 0$, temos que:*

$$\lim_{n\to\infty}[\mathbb{P}|T_n - \tau(\theta)| > \epsilon] = 0 \;\; \forall\theta$$

Observação: Se um estimador é consistente em EQM, então é fracamente consistente.

Definição 3.26 ***(Melhores estimadores assintoticamente normais)*** *Uma sequência de estimadores T_n^* de $\tau(\theta)$ é definida como melhor estimador assintoticamente normal, se e somente se, satisfaz as seguintes condições :*

1. $\sqrt{n}(T_n^* - \tau(\theta)) \approx N(0, \sigma^{*2}(\theta))$;
2. *Para cada $\epsilon > 0$ e para todo θ,* $\lim_{n\to\infty}[\mathbb{P}|T_n - \tau(\theta)| > \epsilon] = 0$;
3. *Seja T_n outra sequência de estimadores fracamente consistentes para os quais $\sqrt{n}(T_n - \tau(\theta)) \approx N(0, \sigma^2(\theta))$, então $\sigma^{*2}(\theta) \le \sigma^2(\theta)$.*

3.5 Estimadores não viciados de mínima variância - ENVMV

O objetivo dessa seção é encontrar um "melhor" estimador não viciado, que definimos da seguinte maneira:

Definição 3.27 ***(Estimador não viciado de mínima variância)*** *Seja $X_1, \ldots, X_n$ uma a.a. de $f(x|\theta)$, o estimador T_n^* de $\tau(\theta)$ é definido como ENVMV de $\tau(\theta)$ se:*

1. $\mathbb{E}[T_n^*] = \tau(\theta)$;
2. $Var[T_n^*] \leq Var[T_n]$ *para qualquer T_n estimador não viciado de $\tau(\theta)$.*

3.5.1 Desigualdade de Cramér-Rao

Suponha que, para a estimação de uma função $\tau(\theta)$ do parâmetro θ de uma distribuição $f(x|\theta)$, podemos especificar um limite inferior, digamos $I(\theta)$, sobre a variância de qualquer estimador não viciado de $\tau(\theta)$. Se pudermos, então, encontrar um estimador não viesado T^* que satisfaz $Var[T^*] = I(\theta)$, encontramos um melhor estimador não viciado. Esta é uma abordagem adotada com o uso do limite inferior de Cramér-Rao.

Teorema 3.28 ***(Desigualdade de Cramér-Rao)*** *Seja $X_1, \ldots, X_n$ um vetor de v.a.'s com função conjunta $f(x_1, \ldots, x_n|\theta)$, um estimador não viciado de $\tau(\theta)$. Se as condições de regularidade (ver Anexo A) são satisfeitas, então temos que:*

$$Var[T(X_1, \ldots, X_n)] \geq \frac{\left[\frac{\partial}{\partial\theta}\tau(\theta)\right]^2}{\mathbb{E}\left[\left(\frac{\partial}{\partial\theta} ln f(X_1, \ldots, X_n|\theta)\right)^2\right]}$$

Observações:

- A desigualdade de Cramér-Rao também se aplica para o caso de variáveis aleatórias discretas. Neste caso, utilizamos a função de probabilidade ao invés da função densidade de probabilidade e, observamos que basta substituir a integral pelo somatório. Apesar de $f(x|\theta)$ não ser diferenciável em x, ela o é em θ.
- A quantidade $\mathbb{E}\left(\left(\frac{\partial}{\partial\theta}\log f(\underline{X}|\theta)\right)^2\right)$ é chamada de informação de Fisher. A informação de Fisher recebe este nome pois fornece um limitante para a variância do ENVMV de θ. Conforme a informação aumenta e temos mais informação sobre o parâmetro θ, temos um menor limitante para a variância do ENVMV.

Corolário 3.29 ***(Desigualdade Cramér-Rao, caso a.a.)*** *Se as suposições do Teorema* **??** *são satisfeitas e, além disso, se $X_1, \ldots, X_n$ é uma a.a. de uma população com fdp $f(x|\theta)$, então*

$$Var[T(X_1, \ldots, X_n)] \geq \frac{\left[\frac{\partial}{\partial\theta}\tau(\theta)\right]^2}{n\mathbb{E}\left[\left(\frac{\partial}{\partial\theta} ln f(X)\right)^2\right]}$$

Antes de ver alguns exemplos, apresentamos um resultado computacional que ajuda na aplicação do teorema.

Proposição 3.30 *Se $f(x, \theta)$ satisfaz*

$$\frac{d}{d\theta}\mathbb{E}\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right) = \int \frac{\partial}{\partial\theta}\left[\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right) f(x|\theta)\right] dx$$

então

$$\mathbb{E}\left(\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right) = -\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right).$$

Exemplo 3.31 *Seja $X_1,\ldots,X_n$ uma a.a. com distribuição de Poisson (λ), achar a cota inferior de Cramér-Rao.*
Solução: *Temos da Proposição **??** que*

$$\mathbb{E}\left(\left(\frac{\partial}{\partial\lambda}\log f(X|\lambda)\right)^2\right) = -\mathbb{E}\left(\frac{\partial^2}{\partial\lambda^2}\log f(X|\lambda)\right) = -\mathbb{E}\left(\frac{\partial^2}{\partial\lambda^2}\log\left(\frac{e^{-\lambda}\lambda^X}{X!}\right)\right)$$

de onde segue que

$$\mathbb{E}\left(\left(\frac{\partial}{\partial\lambda}\log f(X|\lambda)\right)^2\right) = -\mathbb{E}\left(-\frac{X}{\lambda^2}\right) = \frac{1}{\lambda}.$$

*Então, pela desigualdade de Cramer-Rao no caso a.a. (Corolário **??**), segue que, para qualquer estimador não-viciado $T(\boldsymbol{X})$ vale a desigualdade*

$$Var(T(\boldsymbol{X}) \geq \frac{\left(\frac{d}{d\lambda}\mathbb{E}\left(T(\boldsymbol{X})\right)\right)^2}{n\mathbb{E}\left(\left(\frac{\partial}{\partial\lambda}\log f(X|\lambda)\right)^2\right)} = \frac{\lambda}{n}.$$

Como $Var(\bar{X}) = \lambda/n$, segue que $\bar{X}$ é o ENVMV de λ, além de ser o estimador de máxima verossimilhança de λ.

Observação: É importante lembrar que as condições de regularidade são chaves para o teorema de Cramér-Rao e que elas se aplicam para a família exponencial, porém, de forma geral é necessário checá-las. Vejamos agora, um exemplo em que o Teorema de Crámer-Rao não se aplica.

Exemplo 3.32 *Seja $X_1,\ldots,X_n$ uma amostra aleatória de uma distribuição uniforme no intervalo $[0,\theta]$. Neste caso, temos que a função densidade de probabilidade $f(x|\theta)$ é dada por*

$$f(x|\theta) = \frac{1}{\theta}I_{[0,\theta]}(x).$$

Segue então que

$$\frac{\partial}{\partial\theta}\log(f(x|\theta)) = \frac{\partial}{\partial\theta}\left[-\log(\theta)\right] = -\frac{1}{\theta}$$

de forma que

$$\mathbb{E}\left(\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right) = \frac{1}{\theta^2}.$$

Desta forma, se T é um estimador não-viciado do parâmetro θ, segue da desigualdade de Cramer-Rao que

$$Var(T) \geq \frac{\theta^2}{n}.$$

*Vamos considerar o estimador de máxima verossimilhança de θ dado por $T = \max\{X_1,\ldots,X_n\}$ (verifique o Exemplo **??**). Inicialmente, observamos que T tem função densidade de probabilidade dada por*

$$f_T(t|\theta) = \frac{nt^{n-1}}{\theta^n} \text{ se } 0 \leq t \leq \theta,$$

de modo que

$$\mathbb{E}(T) = \int_0^\theta \frac{nt^n}{\theta^n} dt = \frac{n}{n+1}\theta.$$

Portanto, $T = \max\{X_1, \ldots, X_n\}$ *é um estimador viciado do parâmetro* θ*. Porém, o estimador* $T' = \frac{n+1}{n}T$ *é, obviamente, não-viciado. Além disso, temos que*

$$Var\left(\frac{n+1}{n}T\right) = \frac{1}{n(n+2)}\theta^2,$$

que é uniformemente menor do que θ^2/n*. Isto indica que a desigualdade de Cramér-Rao não se aplica a esta função densidade de probabilidade, isto acontece pois o parâmetro de interesse está inserido no intervalo de definição da densidade uniforme, indo de encontro às condições de regularidade.*

Corolário 3.33 *Se* $f(x|\theta)$ *satisfaz as condições de regularidade, e além disso, seja* $L(\theta|\underline{x}) = \prod_{i=1}^n f(x_i|\theta)$*. Se* $T(\underline{X})$ *é estimador não viciado de* $\tau(\theta)$*, então* $T(\underline{X})$ *atinge a cota inferior de Cramér-Rao, se e somente se,*

$$\frac{\partial}{\partial\theta} lnL(\theta|\underline{x}) = a(\theta)\left[T(\underline{X} - \tau(\theta))\right]$$

Propriedades resultantes da desigualdade de Cramér-Rao

Vamos considerar a matriz de informação de Fisher dada por $I(\theta) = \mathbb{E}\left(\left(\frac{\partial}{\partial\theta}\log f(\underline{X}|\theta)\right)^2\right)$.

1. Sob condições de regularidade satisfeitas, se $T(\underline{X})$ é um estimador não-viciado de $\tau(\theta)$ e se $\mathrm{Var}(T(\underline{X})) = \frac{[\frac{\partial}{\partial\theta}\tau(\theta)]^2}{I(\theta)}$, então $T(\underline{X})$ é um ENVMV de $\tau(\theta)$.

2. Sob condições de regularidade satisfeitas, se $T(\underline{X})$ é um estimador não-viciado de θ e se $\mathrm{Var}(T(\underline{X})) = \frac{1}{I(\theta)}$, então $T(\underline{X})$ é um ENVMV para θ. Esta é uma consequência direta do item anterior. Basta observar que, se $T(\underline{X})$ é um estimador não-viciado de θ, então, neste caso, $\tau(\theta) = \theta$, de forma que $\frac{\partial}{\partial\theta}\tau(\theta) = 1$ e assim, se $\mathrm{Var}(T(\underline{X})) = \frac{1}{I(\theta)}$, utilizando a Desigualdade de Cramer-Rao, temos que esta é a menor variância possível para um estimador não-viciado de θ.

3. Se $f(\underline{x}|\theta)$ pertence a família exponencial

$$f(\underline{x}|\theta) = a(\theta)b(\underline{x})exp\{c(\theta)d(\underline{x})\}I_{(-\infty,\infty)}(\underline{x})$$

   e se $c(\cdot)$ tem derivada contínua não nula sobre Θ, então

$$\mathrm{Var}\theta(T(\underline{X})) = \frac{[\frac{\partial}{\partial\theta}\tau(\theta)]^2}{I(\theta)}$$

   em que $\tau(\theta) = \mathbb{E}(T(\underline{X}))$ e $T(\underline{X})$ é ENVMV para $\tau(\theta)$.

4. A quantidade $\frac{[\frac{\partial}{\partial\theta}\tau(\theta)]^2}{I(\theta)}$ é chamada de limite inferior de Cramer-Rao para a variância do estimador não-viciado de $\tau(\theta)$ (apenas sob condições de regularidade satisfeitas).

3.5.2 Suficiência e não viés

Na seção anterior, o conceito de suficiência não foi utilizado em nossa busca por estimativas não viciadas. Veremos agora que a consideração de suficiência é, na verdade, uma importante ferramenta.

Teorema 3.34 *(**Rao-Blackwell**) Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma população com função de probabilidade ou função densidade de probabilidade $f(\underline{x}|\theta)$, $T = T(X_1, \ldots, X_n)$ uma estatística suficiente para θ e seja $S = S(X_1, \ldots, X_n)$ um estimador não viciado qualquer de $\tau(\theta)$. Tomando $\tau(T) = T^*(\underline{X}) = \mathbb{E}(S(\underline{X})|T(\underline{X}))$, temos que $T^*(\underline{X})$ é independente de θ, pois $T(\underline{X})$ é suficiente para θ. Além disso, $\mathbb{E}(T^*(\underline{X})) = \tau(\theta)$ e $Var(T^*(\underline{X})) \leq Var(S(\underline{X}))$ para todo θ. Isto é, $T^*(\underline{X})$ é um estimador não-viciado uniformemente melhor de $\tau(\theta)$.*

Demonstração 3.35 *De fato, temos que*

$$\mathbb{E}(T^*(\underline{X})) = \mathbb{E}\left(\mathbb{E}(S(\underline{X})|T(\underline{X}))\right) = \mathbb{E}(S(\underline{X})) = \tau(\theta).$$

e, portanto, $T^*(\underline{X})$ é um estimador não viciado para $\tau(\theta)$. Além disso,

$$Var(S(\underline{X})) = Var\left[\mathbb{E}(S(\underline{X})|T(\underline{X}))\right] + \mathbb{E}\left[var(S(\underline{X})|T(\underline{X}))\right]$$

de onde segue que

$$Var(S(\underline{X})) = Var(T^*(\underline{X})) + \mathbb{E}\left[Var(S(\underline{X})|T(\underline{X}))\right] \geq Var(T^*(\underline{X}))$$

Assim, $T^*(\underline{X})$ é uniformemente melhor do que $S(\underline{X})$. Além disso, como $S(\underline{X})$ é função somente da amostra e $T(\underline{X})$ é uma estatística suficiente, segue que a distribuição de $S(\underline{X})|T(\underline{X})$ independe de θ. Portanto, $T^*(\underline{X})$ é, de fato, um estimador e, além disso, é não viciado e uniformemente melhor para $\tau(\theta)$. Em outras palavras, condicionar qualquer estimador não-viciado de $\tau(\theta)$ a uma estatística suficiente para θ resultará em um melhor estimador, de forma que na busca de estimadores não-viciados ótimos, somente consideramos estatísticas que são funções de uma estatística suficiente. No entanto, o Teorema de Rao-Blackwell não fornece o ENVMV.

O teorema de Rao-Blackwell mostra como melhorar estimadores não viciados ao condicionar a estatísticas suficientes, porém ainda não sabemos se o resultado é ENVMV. Precisamos introduzir o conceito de completude de uma família de densidades.

3.5.3 Família completa de densidades

Definição 3.36 *(**Família completa de densidades**) Seja $f(t|\theta)$ uma família de funções de probabilidade ou funções densidade de probabilidade para uma estatística $T(\underline{X})$. A família de distribuições de probabilidade é chamada completa se $\mathbb{E}(g(T)) = 0$ para todo θ implicar que $\mathbb{P}(g(T) = 0) = 1$ para todo θ. De forma equivalente, $T(\underline{X})$ é uma estatística completa.*

A definição acima diz que T é completa se e somente se o único estimador não viciado de 0 que é função de T é a estatística identicamente igual a 0 com probabilidade 1.

Exemplo 3.37 *uponha que T tenha distribuição Binomial(n,p) com $0 < p < 1$ e seja g uma função tal que $\mathbb{E}(g(T)) = 0$. Então*

$$0 = \mathbb{E}(g(T)) = \sum_{t=0}^{n} g(t) \left(\begin{array}{c} n \\ t \end{array} \right) p^t (1-p)^{n-t} = (1-p)^n \sum_{t=0}^{n} g(t) \left(\begin{array}{c} n \\ t \end{array} \right) \left(\frac{p}{1-p}\right)^t$$

para $0 < p < 1$. Como $(1-p)^n$ é não nulo para qualquer p, segue que

$$0 = \sum_{t=0}^{n} g(t) \begin{pmatrix} n \\ t \end{pmatrix} \left(\frac{p}{1-p}\right)^t = \sum_{t=0}^{n} g(t) \begin{pmatrix} n \\ t \end{pmatrix} r^t$$

para todo r com $0 < r < \infty$. Uma vez que nenhum dos termos $\begin{pmatrix} n \\ t \end{pmatrix}$ é nulo, seque que $\begin{pmatrix} n \\ t \end{pmatrix}$ é não nulo e então, $g(t) = 0$ para todo $t = 0, 1, \ldots, n$. Como, por hipótese, T assume os valores $0, 1, \ldots, n$ com probabilidade 1, segue que $\mathbb{P}(g(T) = 0) = 1$ para todo p e, desta forma, T é uma estatística completa.

De forma geral, testar a completude é tarefa difícil e cheia de truques matemáticos. Segue um teorema que fornece completude de estatísticas.

Teorema 3.38 *Sejam $X_1, \ldots, X_n$ observações independentes e igualmente distribuídas de uma família exponencial com função de probabilidade ou função densidade de probabilidade da forma*

$$f(x|\theta_1, \ldots, \theta_n) = a(\theta_1, \ldots, \theta_n) b(x) exp\{\sum_{j=1}^{k} c_j(\theta_1, \ldots, \theta_n) d_j(x)\}$$

Então a estatística

$$T(\boldsymbol{X}) = \left(\sum_{i=1}^{n} d_1(X_i), \sum_{i=1}^{n} d_2(X_i), \ldots, \sum_{i=1}^{n} d_k(X_i) \right)$$

é completa, contanto que o espaço paramétrico Θ contenha um conjunto aberto em $\mathbb{R}^k$.

Finalmente, podemos apresentar um teorema que permite construir ENVMV.

Teorema 3.39 ***(Lehmann-Scheffé)*** *Seja $X_1, \ldots, X_n$ uma amostra aleatória da variável aleatória X com função de probabilidade ou função densidade de probabilidade $f(x|\theta)$, $T(\underline{X})$ uma estatística suficiente e completa para θ e $S(\underline{X})$ um estimador não-viciado de $\tau(\theta)$. Então, $T^*(\underline{X}) = \mathbb{E}(S(\underline{X})|T(\underline{X}))$ é um ENVMV para $\tau(\theta)$.*

Demonstração 3.40 *Como $T^*(\underline{X}) = \mathbb{E}(S(\underline{X})|T(\underline{X}))$ temos que*

$$\mathbb{E}(T^*(\underline{X})) = \mathbb{E}\left[\mathbb{E}(S(\underline{X})|T(\underline{X}))\right] = \mathbb{E}(S(\underline{X})) = \tau(\theta).$$

Logo, $T^*(\underline{X})$ é um estimador não-viciado de $\tau(\theta)$. Agora, na procura de ENVMV's para $\tau(\theta)$, basta procurar entre os que são função de $T(\underline{X})$ (pois os que não são podem ser melhorados através do Teorema de Rao-Blackwell). Portanto, basta provar que há um único estimador não-viciado de $\tau(\theta)$ que é função de $T(\underline{X})$. Para isto, suponha que existem $T_1(\underline{X})$ e $T_2(\underline{X})$, ambos função de T, tais que

$$\mathbb{E}(T_1(\underline{X})) = \mathbb{E}(T_2(\underline{X})) = \tau(\theta).$$

Mas então, $\mathbb{E}(T_1(\underline{X}) - T_2(\underline{X})) = 0$ e, como T é completa, implica que $T_1(\underline{X}) - T_2(\underline{X}) = 0$, logo $T_1(\underline{X}) = T_2(\underline{X})$ com probabilidade 1.

Exemplo 3.41 *Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma população com distribuição de Bernoulli de parâmetro θ, sendo $0 < \theta < 1$. Verificar que $\frac{n}{n+1}\overline{X}(1 - \overline{X})$ é um ENVMV para $\theta(1 - \theta)$. Observe que, se $f(\boldsymbol{x}|\theta)$ pertence a família exponencial. De fato, temos que*

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

e então

$$f(\underline{x}|\theta) = \exp\left\{ \sum_{i=1}^{n} x_i \log(\theta) + \left(n - \sum_{i=1}^{n} x_i\right) \log(1-\theta) \right\} = \exp\left\{ n\bar{x} \log\left(\frac{\theta}{1-\theta}\right) + n \log(1-\theta) \right\},$$

de onde concluímos que $T(\underline{X}) = \overline{X}$ *é uma estatística suficiente para* θ*. Além disso, como* $0 < \theta < 1$*, e então, o espaço paramétrico* Θ *contém um conjunto aberto de* $\mathbb{R}$ *e, portanto, segue do Teorema* **??***, que* $T(\underline{X})$ *também é uma estatística completa para* θ*. Considere agora a estatística* $S(\underline{X})$ *dada por*

$$S(\underline{X}) = \frac{n}{n-1}\overline{X}(1-\overline{X})$$

que é função de $T(\underline{X})$*. Temos que*

$$\mathbb{E}(S(\underline{X})) = \frac{n}{n-1}\mathbb{E}(\overline{X} - \overline{X}^2) = \frac{n}{n-1}\left[\mathbb{E}(\overline{X}) - Var(\overline{X}) - \mathbb{E}^2(\overline{X})\right]$$

e então

$$\mathbb{E}(S(\underline{X})) = \frac{n}{n-1}\left[\theta - \frac{\theta(1-\theta)}{n} - \theta^2\right] = \frac{n}{n-1}\left[\frac{n-1}{n}(\theta(1-\theta))\right] = \theta(1-\theta).$$

de onde concluímos que $S(\underline{X}) = \frac{n}{n-1}\overline{X}(1-\overline{X})$ *é um estimador não viciado de* $\tau(\theta) = \theta(1-\theta)$ *e, portanto, segue do Teorema de Lehmann-Scheffé que* $S(\underline{X})$ *é o único ENVMV para* $\theta(1-\theta)$*.*

Capítulo 4

Testes de hipóteses

4.1 Definições básicas

No capítulo anterior, estudamos um método de inferência denominado estimação pontual. Agora passamos para outro método de inferência que é o teste de uma hipótese. Refletindo a necessidade de construir e de avaliar testes de hipóteses.

Vamos supor uma situação em que um fabricante quer saber se um determinado tipo de barra produzido por sua fábrica atende a exigência de ter um comprimento médio de 70 cm. Na verdade, o que o fabricante está fazendo é levantar hipóteses sobre uma característica (parâmetro), neste caso a média μ (média do comprimento), de sua produção (população) de barras. Eventualmente ele poderia fazer conjecturas a respeito da distribuição da variável aleatória que representa o comprimento de barra, ou ainda, poderia estar questionando a proporção de defeituosas, etc. Essas conjecturas ou suposições são chamadas de hipóteses estatísticas. De maneira genérica, podemos enunciar:

Definição 4.1 ***(Hipótese estatística)*** *Uma hipótese estatística é uma afirmação ou conjectura sobre a distribuição de uma ou mais variáveis aleatórias.*

A definição de uma hipótese é mais genérica, mas o aspecto importante é que uma hipótese faz uma declaração sobre a população. O objetivo de um teste de hipótese é decidir, com base em uma amostra da população, qual de duas hipóteses complementares é verdadeira.

Definição 4.2 *As duas hipóteses complementares em um problema envolvendo um teste de hipóteses são chamadas* ***hipótese nula*** *e* ***hipótese alternativa****. Elas são denotadas por H_0 e H_1, respectivamente.*

Além da definição de hipóteses complementares, uma hipótese estatística também é classificada em **simples**, se especifica completamente a distribuição dos dados ou **composta**, caso contrário. Veja o exemplo a seguir.

Exemplo 4.3 *exemplo introdutório do fabricante de barras. O tal fabricante está interessado em decidir se as barras tem uma média igual a 70 cm ou diferente de 70 cm. Nesse caso, as hipóteses seriam:*

$$\begin{cases} H_0 : \mu = 70 \\ H_1 : \mu \neq 70 \end{cases}$$

Ainda com relação a esse exemplo poderíamos ter hipóteses do tipo:

$a)\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$ $\quad b)\begin{cases} H_0 : p_1 \leq p_2 \\ H_1 : p_1 > p_2 \end{cases}$ $\quad c)\begin{cases} H_0 : p_1 \geq p_2 \\ H_1 : p_1 < p_2 \end{cases}$

Em todas as situações a hipótese nula H_0 inclui a igualdade, podendo ser do tipo "simples", enquanto que H_1 é do tipo "composta"em (a), (b), (c).

Em um problema envolvendo um teste de hipótese, depois de observar a amostra, o experimentador deve decidir se aceita H_0 como verdadeira ou se a rejeita, o que implicaria que H_1 é verdadeira.

Definição 4.4 ***(Teste de hipótese (Υ))*** *Um procedimento para testar uma hipótese, ou um teste de hipótese, é uma regra específica para decidir se rejeitamos a referida hipótese.*

O subconjunto do espaço amostral para o qual H_0 será rejeitada é chamado de **região de rejeição**, ou **região crítica**. O complemento da região de rejeição é chamado de **região de aceitação**.

Um teste pode ser aleatório ou não aleatório. Por exemplo, se rejeitarmos uma hipótese nula com base no resultado do lançamento de uma moeda não viciada, estamos diante de um teste aleatório.

Definição 4.5 ***(Teste aleatório)*** *Um teste Υ de uma hipótese H é chamado de aleatório se Υ é definido pela função $\Psi_\Upsilon(x_1,\ldots,x_n) = P[\text{rejeitar } H|\text{ foi observado } (x_1,\ldots,x_n)]$. Essa função é chamada de função crítica do teste Υ.*

Definição 4.6 ***(Teste não aleatório)*** *Seja um hipótese H com teste Υ. Definir a função crítica como:*

$$\Psi_\Upsilon(x_1,\ldots,x_n) = \begin{cases} 1, & se\ (x_1,\ldots,x_n) \in \mathbb{C}_\Upsilon \\ 0, & c.c., \end{cases}$$

onde $\mathbb{C}_\Upsilon$ é a região crítica.

O desenvolvimento de um teste não aleatório é imediato: observamos uma amostra aleatória $(x_1,\ldots,x_n)$, verificamos se ela cai na região crítica e rejeitamos H quando for necessário. Por outro lado, para realizar um teste aleatório Υ de H, primeiro observamos uma amostra aleatória $(x_1,\ldots,x_n)$, depois avaliamos $\Psi_\Upsilon(x_1,\ldots,x_n)$ e finalmente observamos o resultado de algum experimento de Bernoulli auxiliar que tem $\Psi_\Upsilon(x_1,\ldots,x_n)$ como sua probabilidade de sucesso, se o experimento resulta em sucesso então, H é rejeitada.

Se a distribuição que fornecer a amostra está parametrizada por θ onde $\theta \in \Theta$, então associamos a qualquer teste uma função poder.

Definição 4.7 ***(Função poder)*** *Seja Υ um teste para H_0. A função poder do teste Υ é a probabilidade de rejeitarmos H_0 dado o valor de θ. Neste caso, temos que*

$$\pi_\Upsilon(\theta) = P[\text{rejeitar } H_0|\theta],$$

para todo valor de θ.

Claro que uma função poder ideal é uma função que é 0 (zero) para aqueles θ que pertencem a H_0 e é 1 para aqueles θ da hipótese alternativa.

Exemplo 4.8 *Seja $X_1,\ldots,X_n$ uma amostra aleatória de uma Normal $(\theta, 25)$ e considere $H_0 : \theta \le 17$ e o teste Υ : rejeitar H_0 se e somente se $\overline{X} > 17 + 5/\sqrt{n}$. Ache a função poder do teste.*

Solução:

$$\begin{aligned}\pi_\Upsilon(\theta) &= P[\overline{X} > 17 + 5/\sqrt{n}|\theta] = 1 - P[\overline{X} \le 17 + 5/\sqrt{n}|\theta] \\ &= 1 - P\left[Z \le \frac{17 + 5/\sqrt{n} - \theta}{5/\sqrt{n}}\right].\end{aligned}$$

A função poder é útil para dizer que tão bom é um teste em particular, no exemplo anterior, se $\theta > 20 \Rightarrow \pi_\Upsilon(\theta) \to 1$ e rejeitamos quase com certeza H_0.

Definição 4.9 ***(Tamanho do teste)*** *Seja Υ um teste da hipótese $H_0 : \theta \in \Theta_0$ onde $\Theta_0 \in \Theta$. O tamanho do teste Υ de H_0 é definido como:*

$$\sup_{\theta \in \Theta_0} [\pi_\Upsilon(\theta)]$$

No exemplo anterior, o tamanho do teste seria dado por:

$$\sup_{\theta \in \Theta_0} [\pi_\Upsilon(\theta)] = \sup_{\theta \le 17} \left[1 - P\left(Z \le \frac{17 + 5/\sqrt{n} - \theta}{5/\sqrt{n}}\right)\right] = 1 - P(Z \le 1) \approx 0,159.$$

Observação:

- Para o teste não aleatório:

$$\pi_\Upsilon(\theta) = P((x_1, \ldots, x_n) \in \mathbb{C}_\Upsilon | \theta).$$

- Para o teste aleatório:

$$\begin{aligned} \pi_\Upsilon(\theta) &= P[\text{rejeitar } H_0|\theta] = \int \ldots \int P[\text{rejeitar } H_0|(x_1, \ldots, x_n) f(x_1, \ldots, x_n) dx_1 \ldots dx_n] \\ &= E[\Psi_\Upsilon(X_1, \ldots, X_n)]. \end{aligned}$$

Na prática, ao realizar testes de hipóteses, podemos estar suscetíveis a erros, mais especificamente são eles:

1. **Erro tipo I:** rejeitar H_0, quando ela é verdadeira.
2. **Erro tipo II:** Não rejeitar H_0, quando ela é falsa.

A tabela a seguir resume as situações acima.

	Não rejeitar H_0	Rejeitar H_0
H_0 verdadeira	Decisão correta	Erro do tipo I
H_0 falsa	Erro do tipo II	Decisão correta

A probabilidade de cometer os erros tipo I e tipo II, ao qual chamamos de tamanho do erro, é denotada por α e β, respectivamente. Assim temos,

$$\alpha = P(\text{Erro do tipo I}) = P(\text{rejeitar } H_0 \text{ dado } H_0 \text{ verdadeira});$$

$$\beta = P(\text{Erro do tipo II}) = P(\text{aceitar } H_0 \text{ dado } H_0 \text{ falsa}).$$

A primeira probabilidade, α, é chamada *nível de significância*, enquanto que $1-\beta$ é conhecida como *poder do teste* e representada por π.

A situação ideal é aquela em que ambas as probabilidades, α e β, são próximas de zero. Entretanto, a medida que diminuímos α, β aumenta.

O poder do teste tem como objetivo conhecer o quanto o teste de hipóteses controla um erro do tipo II, ou qual a probabilidade de rejeitar a hipótese nula se realmente for falsa. Na prática, é importante que se tenham testes com nível de significância próximos do nível de significância nominal e que o poder seja alto, mesmo em situações de amostras pequenas.

O poder de um teste de hipóteses é afetado por três fatores:

1. Tamanho da amostra: Mantendo todos os outros parâmetros iguais, quanto maior o tamanho da amostra, maior o poder do teste.

2. Nível de Significância: Quanto maior o nível de significância, maior o poder do teste. Se você aumenta o nível de significância, você reduz a região de aceitação. Como resultado, você tem maior chance de rejeitar a hipótese nula. Isto significa que você tem menos chance de aceitar a hipótese nula quando ela é falsa, isto é, menor chance de cometer um erro do tipo II. Então, o poder do teste aumenta.

3. O verdadeiro valor do parâmetro a ser testado: Quanto maior a diferença entre o "verdadeiro"valor do parâmetro e o valor especificado pela hipótese nula, maior o poder do teste.

4.2 Testes para hipóteses simples

Definição 4.10 ***(Teste da razão de verossimilhança simples)*** *Seja $X_1, \ldots, X_n$ uma amostra aleatória de $f(x|\theta_0)$ ou $f(x|\theta_1)$, um teste Υ para $H_0 : X_i \sim f(x|\theta_0)$ contra $H_1 : X_i \sim f(x|\theta_1)$ é chamado de teste da razão de verossimilhança simples se Υ for definido como:*

- *Rejeitar H_0 se $\lambda < k$;*
- *Não rejeitar H_0 se $\lambda > k$;*
- *Não rejeitar, rejeitar ou aleatorizar H_0 se $\lambda = k$, onde k é constante não negativa e*

$$\lambda = \lambda(\underset{\sim}{x}) = \frac{L(\theta_0|\underset{\sim}{x})}{L(\theta_1|\underset{\sim}{x})}.$$

Existem diversas classes de testes de hipótese. Algumas dessas classes controlam a probabilidade de um erro do tipo I; por exemplo, os testes de nível α têm probabilidades de erro do tipo I de, no máximo, α para todo $\theta \in \Theta_0$. Um bom teste nesta classe também teria uma pequena probabilidade de erro do tipo II menor que a de outros testes nessa classe, certamente este seria um forte concorrente para o melhor teste na classe, uma noção que é formalizada na próxima definição.

Definição 4.11 ***(Teste mais poderoso)*** *Um teste Υ^* de $H_0 : \theta = \theta_0$ contra $H_1 : \theta = \theta_1$ é chamado teste mais poderoso de tamanho α $(0 < \alpha < 1)$ se, e somente se:*

1. $\pi_{\Upsilon^*}(\theta_0) = \alpha$
2. $\pi_{\Upsilon^*}(\theta_1) \geq \pi_{\Upsilon}(\theta_1)$, *para qualquer outro Υ tal que $\pi_{\Upsilon}(\theta_0) \leq \alpha$.*

O seguinte teorema é útil para achar um teste não aleatório mais poderoso de tamanho α.

Teorema 4.12 ***(Lema de Neyman-Pearson)*** *Suponha que as hipóteses $H_0 : \theta = \theta_0$ e $H_1 : \theta = \theta_1$ (hipóteses simples), com função de probabilidade ou função densidade de probabilidade correspondente a θ_i, $f(\boldsymbol{x}|\theta_i)$, $i = 0, 1$. Considere um teste com região crítica RC satisfazendo*

$$\underline{x} \in R_C \text{ se } f(\underline{x}|\theta_1) > kf(\underline{x}|\theta_0) \text{ e } \underline{x} \notin R_C \text{ se } f(\underline{x}|\theta_1) < kf(\underline{x}|\theta_0)$$

De forma análoga,

$$\lambda(\underline{x}) = \frac{f(\underline{x}|\theta_1)}{f(\underline{x}|\theta_0)} > k$$

para algum $k \geq 0$ e $\alpha = \mathbb{P}(\underline{x} \in R_C|H_0$ é verdadeira). Então o teste é o mais poderoso.

Exemplo 4.13 *Seja $X_1, \ldots, X_n$ uma amostra aleatória de tamanho n da variável aleatória $X \sim N(\mu, \sigma^2)$ com μ conhecido. Vamos encontrar o teste mais poderoso para testar $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 = \sigma_1^2$ ($\sigma_1^2 > \sigma_0^2$)*

Pelo Lema de Neyman-Pearson, temos que o teste mais poderoso rejeita H_0 quando $\frac{L_1(\boldsymbol{x})}{L_0(\boldsymbol{x})} > k$. Então:

$$\frac{L_1(\boldsymbol{x})}{L_0(\boldsymbol{x})} = \frac{(1/\sqrt{2\pi\sigma_1^2}) \exp[-\sum_{i=1}^n (x_i - \mu)^2 / 2\sigma_1^2]}{(1/\sqrt{2\pi\sigma_0^2}) \exp[-\sum_{i=1}^n (x_i - \mu)^2 / 2\sigma_0^2]}$$

que é equivalente a

$$\sum_{i=1}^n (x_i - \mu)^2 > \frac{\log\left(k \left(\frac{\sigma_1}{\sigma_0}\right)^n\right)}{\frac{1}{2}\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma_1^2}\right)} = c.$$

Portanto, a região crítica do teste mais poderoso é dada por

$$R_C^* = \left\{ \boldsymbol{x}; \sum_{i=1}^n (x_i - \mu)^2 > c \right\}.$$

Agora, fixando α, determinamos o valor de c pela solução da equação

$$\alpha = \mathbb{P}\,(\boldsymbol{X} \in \ R_C^* | H_0 \text{ é verdadeira}) \Rightarrow \alpha = \mathbb{P}\left(\sum_{i=1}^n (X_i - \mu)^2 > c\right) = \mathbb{P}\left(\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma_0^2} > \frac{c}{\sigma_0^2}\right).$$

Mas, sob H_0, temos que

$$\frac{X_i - \mu}{\sigma_0} \sim N(0,1) \Rightarrow \frac{(X_i - \mu)^2}{\sigma_0^2} \sim \chi_1^2 \Rightarrow \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma_0^2} \sim \chi_n^2$$

e então, fixando n, α e σ_0^2 podemos obter o valor c a partir de uma distribuição qui-quadrado com n graus de liberdade. Este raciocínio será válido se o valor de μ for conhecido.

4.3 Testes para hipóteses compostas

Consideramos testes de hipóteses simples. Agora precisamos resolver um problema de teste de hipóteses mais geral. Para isso, definimos o **espaço paramétrico**, denotado por Θ, como sendo o conjunto de todos os valores possíveis do parâmetro de interesse.

Assumimos que temos uma amostra aleatória de $f(x|\theta)$, $\theta \in \Theta$ e queremos testar $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1$, onde $\Theta_0 \subset \Theta$ e $\Theta_1 \subset \Theta$ e são disjuntos.

Definição 4.14 ***(Teste da razão de verossimilhança generalizada)*** *A estatística do teste da razão de verossimilhança generalizada para testar $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1$ é dada por*

$$\lambda(\underline{x}) = \frac{\sup_{\theta \in \Theta_0} L(\theta|\underline{x})}{\sup_{\theta \in \Theta} L(\theta|\underline{x})}$$

em que Θ é o espaço paramétrico irrestrito e Θ_0 é o espaço paramétrico restrito à hipótese nula. O teste da razão de verossimilhança generalizada tem uma região crítica da forma $\{\boldsymbol{x}; \lambda(\underline{x}) \leq c\}$ com $0 \leq c \leq 1$. Denotando $\hat{\theta}$ como o estimador de máxima verossimilhança de θ sobre o espaço Θ e $\hat{\theta}_0$ o valor de θ que maximiza a função de verossimilhança no espaço Θ_0, a estatística do teste da razão de verossimilhança generaliza pode ser escrito como

$$\lambda(\underline{x}) = \frac{L(\hat{\theta}_0|\underline{x})}{L(\hat{\theta}|\underline{x})}.$$

Exemplo 4.15 *Seja $X_1, \ldots, X_n$ uma amostra aleatória da população com distribuição normal $N(\theta, 1)$. Considere as hipóteses $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Como há apenas um valor de θ especificado por H_0, a hipótese é simples e o numerador de $\lambda(\underset{\sim}{x})$ é $L(\theta_0|\underset{\sim}{x})$. Agora, sabemos que $\hat{\theta} = \bar{X}$ é o estimador de máxima verossimilhança de θ sobre o espaço paramétrico irrestrito Θ. Logo, o denominador de $\lambda(\underset{\sim}{x})$ é $L(\bar{x}|\underset{\sim}{x})$ e a estatística do teste da razão de verossimilhança generalizada é*

$$\lambda(\underset{\sim}{x}) = \frac{L(\theta_0|\underset{\sim}{x})}{L(\bar{x}|\underset{\sim}{x})} = \frac{\exp\left[-\sum_{i=1}^n (x_i - \theta_0)^2/2\right]}{\exp\left[-\sum_{i=1}^n (x_i - \bar{x})^2/2\right]} = \exp\left\{\left(-\sum_{i=1}^n (x_i - \theta_0)^2 + \sum_{i=1}^n (x_i - \bar{x})^2\right)/2\right\}$$

e como $\sum_{i=1}^n (x_i - \theta_0)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \theta_0)^2$, a estatística do teste da razão de verossimilhanças é dada por

$$\lambda(\underset{\sim}{x}) = \exp\left\{\frac{-n(\bar{x} - \theta_0)^2}{2}\right\}.$$

Como a região crítica é dada por $RC = \{x : \lambda(\underset{\sim}{x}) \leq c\}$, segue que

$$\lambda(\underset{\sim}{x}) \leq c \Leftrightarrow \exp\left\{\frac{-n(\bar{x} - \theta_0)^2}{2}\right\} \leq c \Leftrightarrow |\bar{x} - \theta_0| \geq \sqrt{-2\log(c)/n}.$$

Portanto, a região crítica é dada por $RC = \{\underset{\sim}{x}; |\bar{x} - \theta_0| \geq \sqrt{-2\log(c)/n}\}$ e, como $0 \leq c \leq 1$, temos que $0 \leq \sqrt{-2\log(c)/n} < \infty$. Fixando o nível de significância α, podemos determinar o valor de c. Supondo que $\alpha = 0,05$, temos que

$$P\left(\underset{\sim}{x} \in RC|H_0 \text{ é verdadeira}\right) = \alpha \Rightarrow P\left(|\bar{X} - \theta_0| \geq \sqrt{-2\log(c)/n}\right) = 0,05$$

de onde segue que

$$\Rightarrow P\left(\left|\frac{\bar{X} - \theta_0}{1/\sqrt{n}}\right| \geq \frac{\sqrt{-2\log(c)/n}}{1/\sqrt{n}}\right) = 0,05 \Rightarrow P\left(\sqrt{n}|\bar{X} - \theta_0| \geq 1,96\right) = 0,05$$

onde consideramos que $z_{\alpha/2} = 1,96$ e, portanto, a região crítica do teste é

$$RC = \{\underset{\sim}{x}; \sqrt{n}|\bar{x} - \theta_0| \geq 1,96\}.$$

Definição 4.16 ***(Teste uniformemente mais poderoso)*** *Um teste Υ^* de $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1$ é chamado teste uniformemente mais poderoso de tamanho α ($0 < \alpha < 1$) se, e somente se:*

1. $\sup\limits_{\theta \in \Theta_0} \pi_{\Upsilon^*}(\theta_0) = \alpha$
2. $\pi_{\Upsilon^*}(\theta) \geq \pi_{\Upsilon}(\theta)$, $\forall \theta \in (\Theta - \Theta_0)$ *e para qualquer outro* Υ *tal que* $\pi_{\Upsilon}(\theta) \leq \alpha$.

Um teste uniformemente mais poderoso (UMP) não existe para todos os problemas de teste, mas quando existe um, podemos ver que, entre os que tem tamanho menor ou igual a α, ele tem maior chance de rejeitar H_0 quando for necessário.

O resultado a seguir estabelece condições para que se tenha o teste uniformemente mais poderoso para testar as hipóteses $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1$.

Teorema 4.17 *Se $X_1, \ldots, X_n$ segue uma distribuição da família exponencial, então o teste uniformemente mais poderoso para testar $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$ é também mais poderoso para testar $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. Temos também que o teste uniformemente mais poderoso para testar $H_0 : \theta = \theta_0$ versus $H_1 : \theta < \theta_0$ é uniformemente mais poderoso para testar $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$.*

Adicionalmente, se uma família de distribuições pertence à família exponencial, ou seja,

$$f(\underset{\sim}{x}|\theta) = a(\theta)b(\underset{\sim}{x})\exp\{d(\underset{\sim}{x})c(\theta)\}$$

então $T = \sum d(\underset{\sim}{X})$ é uma estatística suficiente.

1. *Se $c(\theta)$ é uma função monótona crescente em θ e se existe k^* tal que $P(t(X_1, \ldots, X_n) > k^*) = \alpha$, então o teste Υ^* com região crítica $\mathbb{C}_{\Upsilon^*} = [\{(x_1, \ldots, x_n)|t(x_1, \ldots, x_n) > k^*\}$ é um teste uniformemente mais poderoso de tamanho α para testar $H_0 : \theta \leq \theta_0$ contra $H_1 : \theta > \theta_0$ ou $H_0 : \theta = \theta_0$ contra $H_1 : \theta > \theta_0$.*

2. *Se $c(\theta)$ é uma função monótona decrescente em θ e se existe k^* tal que $P(t(X_1, \ldots, X_n) < k^*) = \alpha$, então o teste Υ^* com região crítica $\mathbb{C}_{\Upsilon^*} = [\{(x_1, \ldots, x_n)|t(x_1, \ldots, x_n) < k^*\}$ é um teste uniformemente mais poderoso de tamanho α para testar $H_0 : \theta \leq \theta_0$ contra $H_1 : \theta > \theta_0$ ou $H_0 : \theta = \theta_0$ contra $H_1 : \theta > \theta_0$.*

Exemplo 4.18 *Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma população que segue uma distribuição da família exponencial de parâmetro θ, $\theta > 0$. Testar $H_0 : \theta \leq \theta_0$ contra $H_1 : \theta > \theta_0$. Como, claramente, a distribuição exponencial pertence a família exponencial com $c(\theta) = -\theta$ e $d(x) = x$, então a distribuição exponencial θ tem razão de verossimilhança monótona decrescente em $t(x_1, \ldots, x_n) = \sum x_i$. De acordo com o Teorema anterior, o teste mais poderoso de tamanho α será dado por: Υ^* : rejeitar H_0 se $\sum x_i < k^*$, onde k^* é a solução de*

$$\alpha = P(\sum x_i < k^*|\theta_0) = \int_0^{k^*} \frac{1}{\Gamma(n)} \theta_0^n u^{n-1} e^{-\theta_0 u} du$$

.

Definição 4.19 ***(Razão de verossimilhança monótona)*** *Uma família de densidades $\{f(x|\theta)/\theta \in \Theta\}$ tem razão de verossimilhança monótona se existe uma estatística $T = t(X_1, \ldots, X_n)$ de forma que a razão $\frac{L(\theta'|x_1,\ldots,x_n)}{L(\theta''|x_1,\ldots,x_n)}$ é função monótona não-decrescente de T para $\theta' < \theta''$ ou é função monótona não-crescente de T para $\theta' < \theta''$.*

Exemplo 4.20 *Considerando o contexto do Exemplo **??**, determine a razão de verossimilhança.*

Solução: *Como $X \sim Exp(\theta)$ então, $L(\theta|x_1, \ldots, x_n) = \theta^n e^{-\theta \sum x_i} \prod I_{(0,\infty)}(x_i)$. Veja que*

$$\frac{L(\theta'|x_1, \ldots, x_n)}{\theta''|(x_1, \ldots, x_n)} = \frac{\theta'^n e^{-\theta' \sum x_i} \prod I_{(0,\infty)}(x_i)}{\theta''^n e^{-\theta'' \sum x_i} \prod I_{(0,\infty)}(x_i)} = \left(\frac{\theta'}{\theta''}\right)^n e^{-(\theta'-\theta'')t}$$

Dessa forma, para $\theta' < \theta''$, temos que a razão $\frac{L(\theta'|x_1,\ldots,x_n)}{L(\theta''|x_1,\ldots,x_n)}$ é função monótona não-decrescente de T.

Teorema 4.21 ***(Teorema Karlim-Rubin)*** *Seja $X_1, \ldots, X_n$ uma amostra aleatória de $f(x|\theta)$, assumir que a família de densidades $f(x|\theta)$ tal que $\{f(x|\theta)/\theta \in \Theta\}$ tem razão de verossimilhança monótona na estatística T.*

1. *Se a razão de verossimilhança monótona é uma função não-decrescente em T e se existe k^* tal que $P(t(X_1, \ldots, X_n) < k^*|\theta_0) = \alpha$, então o teste Υ com região crítica $\mathbb{C}_\Upsilon = [\{(x_1, \ldots, x_n)|t(x_1, \ldots, x_n) < k^*\}$ é um teste uniformemente mais poderoso de tamanho α para testar $H_0 : \theta \leq \theta_0$ contra $H_1 : \theta > \theta_0$.*

2. *Se a razão de verossimilhança monótona é uma função não-crescente em T e se existe k^* tal que $P(t(X_1,\ldots,X_n) > k^*|\theta_0) = \alpha$, então o teste Υ com região crítica $\mathbb{C}_\Upsilon = [\{(x_1,\ldots,x_n)|t(x_1,\ldots,x_n) > k^*\}$ é um teste uniformemente mais poderoso de tamanho α para testar $H_0 : \theta \leq \theta_0$ contra $H_1 : \theta > \theta_0$.*

Exemplo 4.22 *Seja $X_1,\ldots,X_n$ uma amostra aleatória de $f(x|\theta) = \frac{1}{\theta}x^{(1-\theta)/\theta}I_{(0,1)}(x)$. Para uma amostra de tamanho n achar um teste UMP de tamanho α para testar $H_0 : \theta \leq \theta_0$ contra $H_1 : \theta > \theta_0$.*

Primeiramente precisamos verificar se a razão de verossimilhança é monótona não-decrescente ou monótona não-crescente. Sendo assim,

$$\frac{L(\theta'|x_1,\ldots,x_n)}{\theta''|(x_1,\ldots,x_n)} = \frac{\theta'^{-n}e^{(1-\theta')/\theta'\sum lnx_i}\prod I_{(0,1)}(x_i)}{\theta''^{-n}e^{(1-\theta'')/\theta''\sum lnx_i}\prod I_{(0,1)}(x_i)} = \left(\frac{\theta''}{\theta'}\right)^n e^{(1/\theta'-1/\theta'')\sum lnx_i}$$

. Dessa forma, para $\theta' < \theta''$, temos que a razão $\frac{L(\theta'|x_1,\ldots,x_n)}{\theta''|(x_1,\ldots,x_n)}$ é função monótona não-decrescente de $T = \sum lnX_i$. Pelo Teorema de Karlim-Rubin, o teste uniformemente mais poderoso de tamanho α será Υ^* : rejeitar H_0 se $\sum lnx_i < k$, isto é, Υ^* : rejeitar H_0 se $-\sum lnx_i > k^*$, onde k^* é a solução de

$$\alpha = P(-\sum lnX_i > k^*|\theta_0)$$

.

Pelo método jacobiano, podemos encontrar a distribuição de $Y = -lnX$. Portanto,

$$f_Y(y) = f_x(e^{-y})\left|\frac{d(e^{-y})}{dy}\right| = \frac{1}{\theta}e^{-\frac{y}{\theta}},$$

ou seja, $Y \sim Exp\left(\frac{1}{\theta}\right)$. Logo $-\sum lnX_i \sim Gama(n, 1/\theta)$ e ainda, $\frac{-2\sum lnX_i}{\theta} \sim \chi^2_{2n}$. Concluímos assim que,

$$\alpha = P(-\sum lnX_i > k^*|\theta_0) = P\left(\frac{-2\sum lnX_i}{\theta_0} > \frac{2k^*}{\theta_0}\right)$$

$$\frac{2k^*}{\theta_0} = \chi^2_{1-\alpha} \Rightarrow k^* = \frac{\chi^2_{1-\alpha}\theta_0}{2}.$$

Observação 1 (**Extensão do Teorema de Karlin-Rubin**) Se as hipóteses forem $H_0 : \theta \geq \theta_0$ contra $H_1 : \theta < \theta_0$, a desigualdade será invertida na hora de definir a região crítica no Teorema de Karlim-Rubin.

Observação 2 O teorema **??**, para a família exponencial, é um caso especial do teorema de Karlin-Rubin.

4.4 Valor p

Depois que um teste é realizado, as conclusões devem ser relatadas de algum modo estatisticamente significativo. Um método para relatar os resultados de um teste de hipótese é expor o tamanho, α, do teste utilizado e a decisão de rejeitar H_0 ou não-rejeitar H_0. O tamanho do teste fornece informações importantes. Se α for pequeno, a decisão de rejeitar H_0 é bastante convincente; mas se α for grande, a decisão de rejeitar H_0 não é muito convincente porque o teste tem uma grande probabilidade de levar, incorretamente, a esta decisão. Outro meio de relatar os resultados de um teste de hipótese é expor o valor de um determinado tipo de estatística de teste, chamado *valor p*.

Definição 4.23 ***(Valor p)*** *Seja Υ : rejeitar H_0 se $T \geq t$, se for observado $T = k$, então*

$$p = \sup_{\theta \in \Theta_0} P(T \geq k|\theta).$$

A definição quer dizer que o p-valor, também denominado nível descritivo do teste, é a probabilidade de que a estatística do teste (como variável aleatória) tenha valor extremo em relação ao valor observado (estatística) quando a hipótese H_0 é verdadeira.

Capítulo 5

Estimação Intervalar

5.1 Definições Básicas

No capítulo anterior discutimos a estimação pontual de um parâmetro θ, onde a inferência é sobre um valor único. Neste capítulo, abordaremos a estimação intervalar.

Definição 5.1 *(**Estimativa intervalar**) Uma estimativa intervalar de um parâmetro $\theta \in \mathbb{R}$ é qualquer par de funções $L(x_1, \ldots, x_n)$ e $U(x_1, \ldots, x_n)$, de uma amostra que satisfaz:*

$$L(\underset{\sim}{x}) \leq U(\underset{\sim}{x}), \forall x \in \Omega.$$

O intervalo $[L(\underset{\sim}{X}), U(\underset{\sim}{X})]$ é chamado estimador intervalar. Embora a definição utilize um intervalo fechado, algumas vezes será mais natural utilizar intervalos abertos. Por exemplo, os intervalos unilaterais $(-\infty, U(\underset{\sim}{X})]$ ou $[L(\underset{\sim}{X}), \infty)$.

Exemplo 5.2 *Para uma amostra aleatória X_1, X_2, X_3, X_4 a partir de uma $N(\mu, 1)$, um estimador intervalar de μ é $[\overline{X} - 1, \overline{X} + 1]$. Isto significa que μ está neste intervalo.*

Surge aqui uma questão natural a se perguntar: qual a vantagem de utilizar um estimador intervalar? Abrindo mão de alguma precisão em nossa estimativa, obtemos alguma confiança, ou garantia, de que nossa afirmação está correta.

Exemplo 5.3 *(Continuação do Exercício **??**) Quando estimamos μ por $\overline{X}$, a probabilidade de que estejamos estritamente corretos, isto é, $P(X = \mu)$, é 0. Entretanto, com um estimador intervalar temos uma probabilidade positiva de estarmos corretos. A probabilidade de que μ é abrangido pelo intervalo $[\overline{X} - 1, \overline{X} + 1]$ pode ser calculada como:*

$$\begin{aligned}
P(\mu \in [\overline{X} - 1, \overline{X} + 1]) &= P(\overline{X} - 1 \leq \mu \leq \overline{X} + 1) \\
&= P(-1 \leq \overline{X} - \mu \leq 1) \\
&= P\left(-2 \leq \frac{\overline{X} - \mu}{\sqrt{1/4}} \leq 2\right) \\
&= P(-2 \leq Z \leq 2) \\
&= 0,9544
\end{aligned}$$

Portanto, temos uma chance de mais de 95% de cobrir o parâmetro desconhecido com nosso estimador intervalar.

O propósito de utilizar estimador intervalar em vez de um estimador pontual é ter alguma garantia de capturar o parâmetro de interesse. A certeza dessa garantia é quantificada nas seguintes definições.

Definição 5.4 ***(Probabilidade de cobertura)*** *Para um estimador por intervalo* $[L(\underset{\sim}{X}), U(\underset{\sim}{X})]$ *do parâmetro* θ*, a probabilidade de cobertura de* $[L(\underset{\sim}{X}), U(\underset{\sim}{X})]$ *é a probabilidade deste intervalo aleatório cobrir o valor verdadeiro do parâmetro* θ*, isto é,*

$$P(\theta \in (\underset{\sim}{X}), U(\underset{\sim}{X})]|\theta).$$

Definição 5.5 ***(Coeficiente de confiança)*** *Para um estimador por intervalo* $[L(\underset{\sim}{X}), U(\underset{\sim}{X})]$ *do parâmetro* θ*, o coeficiente de confiança de* $[L(\underset{\sim}{X}), U(\underset{\sim}{X})]$ *é o ínfimo das probabilidades de cobertura:*

$$\inf_{\theta} P(\theta \in (\underset{\sim}{X}), U(\underset{\sim}{X})]|\theta).$$

Observações:

- O intervalo é a quantidade aleatória, não o parâmetro.
- Estimadores intervalares, junto com coeficientes de confiança são chamados de intervalos de confiança.
- Uma vez que não conhecemos o valor verdadeiro de θ, podemos somente garantir uma probabilidade de cobertura igual ao ínfimo, o coeficiente de confiança.

Exemplo 5.6 *Seja* $X_1, \ldots, X_n$ *uma amostra aleatória de uma população uniforme* $(0, \theta)$ *e* $Y = \text{máx}(X_1, \ldots, X_n)$*. Estamos interessados em um estimador intervalar de* θ*. Consideramos dois estimadores candidatos:* $[aY, bY]$ *para* $1 \leq a < b$ *e* $[Y + c, Y + d]$ *para* $0 \leq c < d$*, onde* a, b, c *e* d *são constantes especificadas. Para o primeiro intervalo, temos:*

$$\begin{aligned} P(\theta \in [aY, bY]|\theta) &= P(aY \leq \theta \leq bY|\theta) \\ &= P\left(\frac{1}{b} \leq \frac{Y}{\theta} \leq \frac{1}{a}\middle| \theta\right) \\ &= P(\left(\frac{1}{b} \leq T \leq \frac{1}{a}\right). \end{aligned}$$

Sabemos que $f(y) = ny^{n-1}/\theta^n$ *para* $0 \leq y \leq \theta$ *então,* $f(t) = nt^{n-1}$*, para* $0 \leq t \leq 1$*. Logo,*

$$P(\left(\frac{1}{b} \leq T \leq \frac{1}{a}\right) = \int_{1/b}^{1/a} nt^{n-1} dt = \left(\frac{1}{a}\right)^n - \left(\frac{1}{b}\right)^n.$$

A probabilidade de cobertura do primeiro intervalo é independente do valor de θ *e, assim,* $\left(\frac{1}{a}\right)^n - \left(\frac{1}{b}\right)^n$ *é o coeficiente de confiança do intervalo.*

Para o outro intervalo, temos que:

$$\begin{aligned} P(\theta \in [Y + c, Y + d]|\theta) &= P(Y + c \leq \theta \leq Y + d|\theta) \\ &= P\left(\frac{Y}{\theta} + \frac{c}{\theta} \leq \frac{Y}{\theta} + \frac{d}{\theta}\middle| \theta\right) \\ &= P(\left(1 - \frac{d}{\theta} \leq T \leq 1 - \frac{c}{\theta}\right) \\ &= \int_{1-d/\theta}^{1-c?\theta} nt^{n-1} dt = \left(1 - \frac{c}{\theta}\right)^n - \left(1 - \frac{d}{\theta}\right)^n. \end{aligned}$$

Neste caso, a probabilidade de cobertura depende de θ*. Além disso, é possível calcular diretamente que para quaisquer constantes c e d,*

$$\lim_{\theta \to \infty} \left(1 - \frac{c}{\theta}\right)^n - \left(1 - \frac{d}{\theta}\right)^n = 0,$$

mostrando que o coeficiente de confiança deste estimador intervalar é 0*.*

5.2 Métodos para encontrar estimadores

5.2.1 Invertendo uma estatística de teste

Existe uma forte correspondência entre teste de hipótese e estimação intervalar. Considere o exemplo a seguir.

Exemplo 5.7 *Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma população normal (μ, σ^2) e considere testar $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$. Para um nível α fixado, o teste da razão de verossimilhança é Υ : rejeitar H_0 se $|\overline{x} - \mu_0| > z_{\alpha/2\sigma/\sqrt{n}}$. Note que H_0 é não rejeitada para pontos amostrais com $|\overline{x} - \mu_0| \leq z_{\alpha/2\sigma/\sqrt{n}}$ ou, de modo equivalente,*

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Uma vez que o teste tem tamanho α, isto significa que $P(H_0$ é rejeitada $|\mu = \mu_0) = \alpha$ ou, de outro modo, $P(H_0$ é não rejeitada $|\mu = \mu_0) = 1 - \alpha$. Assim, podemos escrever:

$$P\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} | \mu = \mu_0\right) = 1 - \alpha$$

Mas esta declaração de probabilidade é verdadeira para cada μ_0. Deste modo, a declaração

$$P\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} | \mu = \mu_0\right) = 1 - \alpha$$

é verdadeira. O intervalo $[\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$, obtido pela inversão da região de aceitação do teste de nível α, é o intervalo de confiança $1 - \alpha$.

Podemos concluir que testes de hipóteses e estimação por intervalos fazem a mesma pergunta, mas sob perspectivas diferentes. O teste de hipótese fixa o parâmetro e pergunta quais valores amostrais são consistentes com esse valor fixo. O conjunto de confiança fixa o valor amostral e pergunta quais valores de parâmetro tornam este valor amostral mais plausível. O seguinte Teorema dá uma formalização desta correspondência.

Teorema 5.8 *Para cada $\theta_0 \in \Theta$, seja $\overline{C_\gamma}(\theta_0)$ a região de aceitação do teste de nível α de $H_0 : \theta = \theta_0$. Para cada $\underset{\sim}{x} \in \Omega$, define-se um conjunto $IC(x_1, \ldots, x_n)$ no espaço paramétrico como*

$$IC(\underset{\sim}{x}) = \{\theta_0 | \underset{\sim}{x} \in \overline{C_\gamma}(\theta_0)\}.$$

É relativamente simples construir um teste de nível α então, o método de obtenção de conjuntos de confiança invertendo regiões de confiança é útil. Todas as técnicas que temos para obter testes podem ser aplicadas à construção de conjuntos de confiança.

No Teorema se definiu $H_0 : \theta = \theta_0$, gerando um estimador bilateral, porém, na prática teremos em mente hipóteses alternativas como $H_1 : \theta \neq \theta_0$ ou $H_1 : \theta > \theta_0$. A hipótese alternativa indicará a forma da região de aceitação, $\overline{C_\gamma}(\theta_0)$, e indicará a forma do intervalo $IC(\underset{\sim}{x})$.

Exemplo 5.9 *Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma população normal (μ, σ^2). Construir um limite superior de $1 - \alpha$ de confiança para μ da forma $IC(\underset{\sim}{x}) = (-\infty, U(\underset{\sim}{x})]$. Para obter este intervalo, utilizando o Teorema da inversão para conjuntos de confiança, inverteremos testes unilaterais de $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$. Para um nível α fixado, o teste da razão de verossimilhança de H_0 versus H_1 rejeita H_0 se:*

$$\frac{\overline{X} - \mu_0}{S/\sqrt{n}} < -t_{n-1,\alpha}.$$

Asim, a região de aceitação para este teste é

$$\overline{C_\gamma}(\theta_0) = \{\underset{\sim}{x} | \overline{x} \geq \mu_0 - t_{n-1,\alpha} \frac{S}{\sqrt{n}}\}.$$

Portanto, o intervalo unilateral de $1 - \alpha$ de confiança é

$$\left(-\infty, \overline{x} + t_{n-1,\alpha} \frac{S}{\sqrt{n}}\right].$$

5.2.2 Quantidades pivotais

Considere $X_1, X_2, \ldots, X_n$ uma amostra aleatória retirada de uma população com distribuição f_θ que depende do parâmetro θ. Por exemplo, tomamos $X_1, X_2, \ldots, X_n$ uma amostra aleatória de uma população com distribuição normal com média μ desconhecida e desvio padrão conhecido $\sigma = 1$. Para propormos um intervalo de confiança para o parâmetro θ, vamos introduzir o conceito de quantidade pivotal.

Definição 5.10 *(**Quantidade pivotal**) Uma função Q da amostra $(X_1, X_2, \ldots, X_n)$ e do parâmetro θ cuja distribuição de probabilidade não depende do parâmetro θ é denominada quantidade pivotal. Desta forma, dado o nível de confiança $1 - \alpha$, tomamos*

$$1 - \alpha = P\left(a \leq Q(X_1, X_2, \ldots, X_n; \theta) \leq b\right)$$

Se a quantidade pivotal Q for inversível, podemos resolver a inequação acima em relação a θ e obter um intervalo de confiança.

Não existe uma estratégia geral para achar pivôs. Porém existem algumas regras: em geral, diferenças são quantidades pivotais de modelos de locação e razões (ou produtos) são quantidades pivotais de modelos de escala.

Exemplo 5.11 *Em modelos de locação e de escala, existem muitas quantidades pivotais. Seja $X_1, \ldots, X_n$ uma amostra aleatória de fdp's indicadas, e sejam $\overline{X}$ e S a média amostral e o desvio padrão.*

Tabela 5.1: Pivôs de locação-escala

Formas da fdp	Tipo do modelo	Quantidade pivotal
$f(x-\mu)$	Locação	$\overline{X} - \mu$
$\frac{1}{\sigma} f\left(\frac{1}{\sigma}\right)$	Escala	$\frac{\overline{X}}{\sigma}$
$\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$	Locação-escala	$\frac{\overline{X}-\mu}{\sigma}$

Exemplo 5.12 *Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição exponencial (λ). Então $T = \sum X_i$ é uma estatística suficiente para λ e $T \sim Gama(n, \lambda)$, isto é, T é um modelo de escala. Portanto, se $Q(T, \lambda) = 2T/\lambda$, então*

$$Q(T, \lambda) \sim Gama(n, \lambda(2/\lambda)) = Gama(n, 2),$$

que não depende de λ. A quantidade $Q(T, \lambda) = 2T/\lambda$ é um pivô com uma distribuição $Gama(n, 2)$ ou, equivalentemente, χ^2_{2n}.

Como visto na definição de quantidade pivotal, se $Q(\underset{\sim}{X}, \theta)$ é uma quantidade pivotal, podemos encontrar um intervalo para θ usando $Q(\underset{\sim}{X}, \theta)$.

Exemplo 5.13 *(Continuação do Exemplo* **??***) Agora, também podemos ver que se temos uma amostra $X_1, \ldots, X_n$, podemos definir $T = \sum X_i$ e $Q(T, \lambda) = 2T/\lambda \sim \chi^2_{2n}$.*

Se escolhermos constantes a e b para satisfazer $P(a \leq \chi^2_{2n} \leq b) = 1 - \alpha$, então

$$P(a \leq 2T/\lambda \leq b) = P(2T/b \leq \lambda \leq 2T/a) = 1 - \alpha.$$

.

Logo, um intervalo de confiança $1 - \alpha$ paraλ é dado por:

$$IC(\underset{\sim}{x}) = \left[2\frac{\sum x_i}{b}, 2\frac{\sum x_i}{a} \right],$$

onde $a = \chi^2_{2n,1-\alpha/2}$ e $b = \chi^2_{2n,\alpha/2}$.

5.2.3 Usando a função de distribuição acumulada como quantidade pivotal

Na seção anterior, vimos que um pivô, Q, leva a um conjunto de confiança da forma,

$$IC(\underset{\sim}{x}) = \{\theta_0 | a \leq Q(\underset{\sim}{x}, \theta_0) \leq b\}.$$

Se, para cada $\underset{\sim}{x}$, a função $Q(\underset{\sim}{x}, \theta)$ é monótona de θ, então podemos garantir que o conjunto $IC(\underset{\sim}{x})$ é um intervalo.

Nesta seção trabalharemos com outro pivô, um que é totalmente geral e, com poucas suposições, vai garantir um intervalo.

Basearemos a construção do nosso intervalo de confiança para o parâmetro θ numa estatística T que assume valores reais e que tem função de distribuição acumulada (fda) $F(t|\theta)$. Assumiremos inicialmente que T é uma variável aleatória contínua para depois vermos o caso discreto.

Primeiro, lembraremos que $F(t|\theta)$ tem distribuição uniforme$(0, 1)$ e por isso é uma quantidade pivotal então, se $\alpha_1 + \alpha_2 = \alpha$, uma região de aceitação de nível α para as hipótese $H_0 : \theta = \theta_0$ é

$$\overline{C_\gamma} = \{t | \alpha_1 \leq F(t|\theta_0) \leq 1 - \alpha_2\}$$

com o intervalo associado,

$$IC(\underset{\sim}{x}) = \{\theta | \alpha_1 \leq F(t|\theta_0) \leq 1 - \alpha_2\},$$

e para garantir que $IC(\underset{\sim}{x})$ seja um intervalo, precisamos que $F(t|\theta_0)$ seja monótona em θ.

Teorema 5.14 ***(Pivotando uma fda contínua)*** *Seja T uma estatística com f.d.a. contínua $F(t|\theta)$. Seja $\alpha_1 + \alpha_2 = \alpha$, com $0 < \alpha < 1$. Suponha que para cada $t \in T$, as funções $\theta_L(t)$ e $\theta_U(t)$ possam ser definidas como segue.*

1. *Se $F(t|\theta)$ é uma função decrescente de θ para cada t, definimos $\theta_L(t)$ e $\theta_U(t)$ por:*

$$F(t|\theta_L(t)) = \int_{-\infty}^{t} f(s|\theta_L(t))ds = 1 - \int_{t}^{\infty} f(s|\theta_L(t))ds = 1 - \alpha_2;$$

$$F(t|\theta_U(t)) = \int_{-\infty}^{t} f(s|\theta_U(t))ds = \alpha_1.$$

2. Se $F(t|\theta)$ é uma função crescente de θ para cada t, definimos $\theta_L(t)$ e $\theta_U(t)$ por:

$$F(t|\theta_L(t)) = \int_{-\infty}^{t} f(s|\theta_L(t))ds = \alpha_1;$$

$$F(t|\theta_U(t)) = \int_{-\infty}^{t} f(s|\theta_U(t))ds = 1 - \int_{t}^{\infty} f(s|\theta_U(t))ds = 1 - \alpha_2.$$

Então, o intervalo aleatório $[\theta_L(t), \theta_U(t)]$ é um intervalo de coeficiente de confiança $1-\alpha$ para θ.

Exemplo 5.15 *Se $X_1, \ldots, X_n$ são i.i.d. com fdp $f(x|\theta) = 3x^2\theta^{-3}I_{[0,\theta]}(x)$, então $Y = max(X_1, \ldots, X_n)$ é suficiente para θ com função de distribuição*

$$F_Y(y|\theta) = [F_X(y|\theta)]^n = \left(\frac{y}{\theta}\right)^{3n}.$$

*Como $F_Y(y|\theta)$ é uma função decrescente de θ, pelo item 1 do Teorema **??**, se $\alpha_1 + \alpha_2 = \alpha$, $\left(\frac{y}{\theta_U(t)}\right)^{3n} = \frac{\alpha}{2}$ e $\left(\frac{y}{\theta_L(t)}\right)^{3n} = 1 - \frac{\alpha}{2}$, que nos dão as soluções:*

$$\theta_U(t) = \frac{y}{(\alpha/2)^{1/3n}},$$

$$\theta_L(t) = \frac{y}{(1-\alpha/2)^{1/3n}}.$$

Desse modo, o intervalo aleatório $IC(\underset{\sim}{x}) = \left\{\theta | \frac{y}{(1-\alpha/2)^{1/3n}} \leq \theta \leq \frac{y}{(\alpha/2)^{1/3n}}\right\}$ é um intervalo de confiança $1-\alpha$ para θ.

No teorema mesmo que não exista uma solução analítica para $\theta_L(t)$ e $\theta_U(t)$ ela pode ser numericamente adotada já que o Teorema **??** não faz exigência de uma solução analítica.

Agora, consideremos o caso discreto.

Teorema 5.16 ***(Pivotando um fda discreta)*** *Seja T uma estatística com f.d.a. discreta $F(t|\theta) = P(T \leq t|\theta)$. Seja $\alpha_1 + \alpha_2 = \alpha/2$, com $0 < \alpha < 1$. Suponha que para cada $t \in T$, as funções $\theta_L(t)$ e $\theta_U(t)$ possam ser definidas como segue.*

1. Se $F(t|\theta)$ é uma função decrescente de θ para cada t, definimos $\theta_L(t)$ e $\theta_U(t)$ por:

$$P(T \leq t|\theta_U(t)) = \alpha_1,$$

$$P(T \geq t|\theta_L(t)) = \alpha_2.$$

2. Se $F(t|\theta)$ é uma função crescente de θ para cada t, definimos $\theta_L(t)$ e $\theta_U(t)$ por:

$$P(T \geq t|\theta_U(t)) = \alpha_1,$$

$$P(T \leq t|\theta_L(t)) = \alpha_2.$$

Então, o intervalo aleatório $[\theta_L(t), \theta_U(t)]$ é um intervalo de coeficiente de confiança $1-\alpha$ para θ.

Exemplo 5.17 *Se $X_1, \ldots, X_n$ uma amostra aleatória de uma população de Poisson com parâmetro λ e definimos $Y = \sum X_i$. Y é suficiente para λ e $Y \sim Poisson(n\lambda)$. Para $\alpha_1 = \alpha_2 = \alpha/2$, se observarmos $Y = y_0$ então, precisaremos resolver as equações para λ,*

$$\sum_{k=0}^{y_0} e^{-n\lambda}\frac{(n\lambda)^k}{k!} = \alpha/2 \quad e \quad \sum_{k=y_0}^{\infty} e^{-n\lambda}\frac{(n\lambda)^k}{k!} = \alpha/2$$

Sabe-se que se $X \sim Gama(\alpha, \beta)$, para α inteiro, então $\forall x \;\; P(X \leq x) = P(Y \geq \alpha)$, onde $Y \sim Poisson(x/\beta)$. Sendo assim,

$$\alpha/2 = \sum_{k=0}^{y_0} e^{-n\lambda}\frac{(n\lambda)^k}{k!} = P(Y \leq y_0|\lambda) = P(\chi^2_{2(y_0+1)} > 2n\lambda)$$

com solução $\lambda_U(y_0) = \frac{1}{2n}\chi^2_{2(y_0+1),\alpha/2}$.

De modo similar,

$$\alpha/2 = \sum_{k=y_0}^{\infty} e^{-n\lambda}\frac{(n\lambda)^k}{k!} = P(Y \geq y_0|\lambda) = P(\chi^2_{2y_0} < 2n\lambda)$$

com solução $\lambda_L(y_0) = \frac{1}{2n}\chi^2_{2y_0,1-\alpha/2}$.

Portanto, o intervalo aleatório $IC(\underset{\sim}{x}) = \left\{\lambda | \frac{1}{2n}\chi^2_{2y_0,1-\alpha/2} \leq \lambda \leq \frac{1}{2n}\chi^2_{2(y_0+1),\alpha/2}\right\}$ é um intervalo de confiança $1-\alpha$ para λ.

5.3 Intervalos de confiança associados à distribuição normal

Consideremos uma amostra aleatória simples $X_1, X_2, \ldots, X_n$, obtida de uma população com distribuição normal, com média μ e variância σ^2 desconhecidas.

5.4 Intervalo de confiança para a média

Como neste caso a variância é desconhecida, utilizaremos a variância amostral S^2 no lugar de σ^2. Assim, temos a seguinte quantidade pivotal para μ

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

ou seja, a variável T tem distribuição t de Student com $n-1$ graus de liberdade.

Então, ao fixarmos o nível de significância α, dados a e b constantes, temos que

$$P\left(a \leq T \leq b\right) = 1 - \alpha,$$

isto é,

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

.

Agora, precisamos encontrar a e b que minimizam esse intervalo que tem amplitude expressa por $L = \frac{S}{\sqrt{n}}(b-a)$. Isto é, queremos minimizar este intervalo sujeito a restrição

$$\int_a^b f_T(t)dt = 1 - \alpha, \tag{5.1}$$

onde $f_T(t)$ é a densidade de uma distribuição t-Student com $n-1$ graus de liberdade. Em **??**, considerando $b = g(a)$, temos que

$$f_T(b)\frac{db}{da} - f_T(a) = 0 \Rightarrow \frac{db}{da} = \frac{f_T(a)}{f_T(b)}$$

.

E, para minimizar L, fazemos $dL/da = 0$ isto é,

$$\frac{dL}{da} = \frac{S}{\sqrt{n}}\left(\frac{db}{da} - 1\right) = 0 \Rightarrow \frac{S}{\sqrt{n}}\left(\frac{f_T(a)}{f_T(b)} - 1\right) = 0$$

Esta equação só é zero se $f_T(a) = f_T(b)$, o que significa que $a = b$ ou $a = -b$. Sendo assim, para $b = t_{((n-1),\alpha/2)}$, temos que

$$IC(\underset{\sim}{x}) = \left(\overline{X} - t_{\alpha/2}\frac{s}{\sqrt{n}}; \overline{X} + t_{\alpha/2}\frac{s}{\sqrt{n}}\right)$$

é um intervalo com $100(1-\alpha)\%$ de confiança para μ, com variância desconhecida.

5.4.1 Intervalo de confiança para a variância

Consideremos uma amostra aleatória $X_1, \ldots, X_n$ de tamanho n de uma população com distribuição normal com média μ e variância σ^2. Um estimador para σ^2 é a variância amostral s^2. Assim, sabemos que a quantidade pivotal

$$Q = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Seja $1-\alpha$ a probabilidade da variável Q, com $n-1$ graus de liberdade, tomar valores entre a e b, tais que $P[Q < a] = P[Q > b] = \alpha/2$. Então

$$P\left(a \leq Q \leq b\right) = 1 - \alpha,$$

isto é,

$$P\left(\frac{(n-1)S^2}{b} \leq \sigma^2 \leq \frac{(n-1)S^2}{a}\right) = 1 - \alpha$$

.

Novamente, precisamos encontrar a e b que minimizam a amplitude $L = (n-1)S^2\left(\frac{1}{a} - \frac{1}{b}\right)$, do intervalo.

Seja $f_Q(q)$ a densidade de uma distribuição χ^2_{n-1}, então

$$\int_a^b f_Q(q)dq = 1 - \alpha, \tag{5.2}$$

precisa ser derivado. Em **??**, considerando $b = g(a)$, temos que

$$f_Q(b)\frac{db}{da} - f_Q(a) = 0 \Rightarrow \frac{db}{da} = \frac{f_Q(a)}{f_Q(b)}$$

.

E, para minimizar L, fazemos $dL/da = 0$ isto é,

$$\frac{dL}{da} = (n-1)S^2\left(-\frac{1}{a^2} + \frac{1}{b^2}\frac{db}{da}\right) = 0 \Rightarrow (n-1)S^2\left(-\frac{1}{a^2} + \frac{1}{b^2}\frac{f_Q(a)}{f_Q(b)}\right) = 0$$

Portanto, $a^2 f_Q(a) = b^2 f_Q(b)$. Uma solução para a e b pode ser obtida por tentativa e erro ou integração numérica, mas, para qualquer a e b que satisfazem **??**. Em especial, tome $a = \chi^2_{\alpha/2}$ e $b = \chi^2_{1-\alpha/2}$. Logo, o intervalo com nível $100(1-\alpha)\%$ de confiança para σ^2 será dado por

$$IC(\sigma^2, 1-\alpha) = \left(\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{\alpha/2}} \right).$$

5.5 Intervalos de confiança quando $n \to \infty$

Em estimação pontual vimos que é possível achar $T_n = t_n(X_1, \ldots, X_n)$ estimador de θ que são assintoticamente normais com média θ e variância $\sigma_n^2(\theta)$. Em particular, se $\hat{\theta}$ é o EMV então,

$$\hat{\theta} \approx N(\theta, \sigma_n^2(\theta)), n \to \infty$$

onde $\sigma_n^2(\theta) = -\frac{1}{nE\left\{\frac{\partial^2}{\partial\theta^2} lnf(x|\theta)\right\}}$.

Com isso,

$$\frac{T_n - \theta}{\sigma_n^2(\theta)}$$

pode ser usado para gerar um intervalo de confiança para θ, tal que

$$P\left(-z_{\alpha/2} < \frac{T_n - \theta}{\sigma_n^2(\theta)} < z_{\alpha/2} \right) = 1 - \alpha.$$

Exemplo 5.18 *Seja $X_1, \ldots, X_n$ é uma amostra aleatória de uma distribuição exponencial (θ). Sabemos que $\hat{\theta} = 1/\overline{X}$ é EMV para θ e que $\sigma_n^2(\theta) = \frac{\theta^2}{n}$. Quando $n \to \infty$, $\hat{\theta} \approx N(\theta, \frac{\sigma^2}{n})$ e portanto,*

$$IC(\underset{\sim}{x}) = \left[\frac{1/\overline{x}}{1 + z_{\alpha_2}/\sqrt{n}}, \frac{1/\overline{x}}{1 - z_{\alpha_2}/\sqrt{n}} \right]$$

é um intervalo com nível $100(1-\alpha)\%$ de confiança para θ.

Capítulo 6

Exercícios resolvidos

6.1 Exercícios do Capítulo 1

1. Um pesquisador deseja estimar a média de uma população usando uma amostra suficientemente grande para ter uma probabilidade de 0,95 que a média amostral não seja diferente da média populacional por mais de 25 por cento do desvio padrão. Qual deve ser o tamanho dessa amostra?

   Solução: A probabilidade de interesse pode ser representada pela seguinte expressão:

   $$P(|\bar{X} - \mu| < 0,25\sigma) \geq 0,95$$

   Pela Lei Fraca dos Grandes números, segue-se que se n é um inteiro tal que $n > \frac{\sigma^2}{\epsilon^2 \delta}$ então, $P(|\bar{X} - \mu| < \epsilon) \geq 1 - \delta$. Dessa forma, basta considerar

   $$\begin{cases} \delta = 0,05 \\ \epsilon = 0,25\sigma \end{cases}$$

   Portanto, tem-se que,

   $$n > \frac{\sigma^2}{(0,25\sigma^2) \times 0,05} \ ,$$

   ou seja, $n > 320$.

2. Seja X_1, X_2 uma amostra aleatória de tamanho dois da densidade

   $$f(x) = \frac{1}{2} e^{-\frac{1}{2}x} I_{(0,\infty)}(x)$$

   Usar os resultaos das distribuições χ^2 e F para definir a distribuição de $\frac{X_1}{X_2}$.

   Solução: Temos que $f(x_i) = \frac{1}{2} e^{-\frac{1}{2}x_i} I_{(0,\infty)}(x_i)$, logo $X_i \sim Exp\left(\frac{1}{2}\right)$, $i = 1, 2$.
   Sendo assim, podemos afirmar que $X_i \sim Gama\left(1, \frac{1}{2}\right)$ e consequentemente, $X_i \sim \chi^2_2$. Defina $Y = \frac{X_1}{X_2}$, então

   $$Y = \frac{X_1}{X_2} = \frac{\frac{X_1}{2}}{\frac{X_2}{2}} \sim F(2, 2)$$

3. Seja $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, calcular $E[S^2]$ sendo que $X_1, X_2, \ldots, X_n$ é uma amostra aleatória de uma distribuição normal de média μ e variância σ^2.

   Solução:

$$
\begin{aligned}
E[S^2] &= E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] \\
&= E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu + \mu - \bar{X})^2\right] \\
&= \frac{1}{n-1}E\left[\sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2\right] \\
&= \frac{1}{n-1}\left\{E\left[\sum_{i=1}^{n}(X_i - \mu)^2\right] - E\left[n(\bar{X} - \mu)^2\right]\right\} \\
&= \frac{1}{n-1}E\{nE[(X - \mu)^2] - nE[(\bar{X} - \mu)^2]\} \\
&= \frac{1}{n-1}\left[n\sigma^2 - n\frac{\sigma^2}{n}\right] \\
&= \frac{1}{n-1}\left[\sigma^2(n-1)\right] \\
&= \sigma^2
\end{aligned}
$$

4. Seja $X_1, X_2, \ldots, X_{10}$, uma amostra aleatória de uma distribuição normal padrão. Calcular $P\left(2,56 < \sum_{i=1}^{10}\chi^2 < 18,3\right)$.

   Solução: Queremos encontrar $P(2,56 < \sum_{i=1}^{10}\chi^2 < 18,3)$. Para isso, vamos denotar $Y = \sum_{i=1}^{10}\chi^2$, como as v.as X_i, $i = 1, 2, \ldots, 10$ são $N(0,1)$, então $Y \sim \chi^2_{10}$. Sendo assim, $P(2,56 < Y < 18,3) = P(Y < 18,3) - P(Y < 2,56) = 0,95 - 0,01 = 0,94$.

5. Qual é a probabilidade de que o maior valor de uma amostra aleatória de tamanho 10 de uma variável com distribuição uniforme no intervalo (0,1) supere 0,9? Qual é a probabilidade deste valor estar abaixo de 0,5?

   Solução:

   Se $X_i \sim U(0,1)$, $i = 1, 2, \ldots, n$, então $F_{X_i}(x_i) = x_i$. Defina $Y_n = max(X_1, X_2, \ldots, X_n)$, sabemos que $F_{Y_n}(y_n) = [F_X(x)]^n$. Dessa forma,

   - $P(Y_n > 0,9) = 1 - P(Y_n \leq 0,9) = 1 - F_{Y_n}(0,9) = 1 - [F_X(0,9)]^{10} = 1 - (0,9)^{10}$
   - $P(Y_n < 0,5) = F_{Y_n}(0,5) = [F_X(0,5)]^{10} = (0,5)^{10}$

6. Nas condições do exercício anterior, qual é a probabilidade de que o menor valor seja menor a 0,5? E menor a 0,9?

Solução:

Agora, defina $Y_1 = min(X_1, X_2, \ldots, X_n)$. Então,

$$F_{Y_1}(y_1) = 1 - [1 - F_X(x)]^n$$

Portanto,

- $P(Y_1 < 0,5) = F_{Y_1}(0,5) = 1 - [1 - F_X(0,5)]^{10} = 1 - [1 - (0,5)]^{10} = 1 - [0,5]^{10}$
- $P(Y_1 < 0,9) = F_{Y_1}(0,9) = 1 - [1 - F_X(0,9)]^{10} = 1 - [1 - (0,9)]^{10} = 1 - [0,1]^{10}$

7. Assuma que seu tempo para chegar a um certo sinal na rua a caminho do seu trabalho tem distribuição uniforme no intervalo de 0 a 10 minutos, e que você faz este percurso 5 vezes por semana.

   (a) Qual é a distribuicção do seu menor tempo semanal até o sinal?

   (b) Calcule $E[Y_1]$, o valor esperado do menor tempo.

   (c)Qual é a distribuição conjunta de Y_1 e Y_n, o menor e o maior tempo?

   Solução:
   Seja X_i o tempo do i-ésimo percurso semanal. Sabendo que $X_i \sim U(0,10), i = 1, 2, \ldots, 5$, então $f_{X_i}(x_i) = \frac{1}{10} I_{(0,10)}(x_i)$ e $F_{X_i}(x_i) = \frac{x_i}{10}$. Defina $Y_1 = min(X_1, X_2, \ldots, X_5)$.

   (a) $f_{Y_1}(y_1) = n \times (1 - F_X(y_1))^{n-1} \times f_X(y_1) = 5\left(1 - \frac{y_1}{10}\right)^4 \frac{1}{10} = \frac{1}{2}\left(1 - \frac{y_1}{10}\right)^4 I_{(0,10)}(y_1)$

   (b)$E[Y_1] = \displaystyle\int_0^{10} y_1 f_{Y_1}(y_1) dy_1 = \int_0^{10} y_1 \frac{1}{2}\left(1 - \frac{y_1}{10}\right)^4 dy_1 =$

   $\begin{cases} u = 1 - \frac{y_1}{10} \Rightarrow du = -\frac{dy_1}{10} \\ 10u = 10 - y_1 \Rightarrow y_1 = 10(1-u) \end{cases}$

   $= \displaystyle\int_0^1 10(1-u)u^4 \frac{1}{2}\frac{-5}{-5} dy == \int_0^1 -50(1-u)u^4 du = -50\left(\frac{u^6}{6} - \frac{u^5}{5}\right)\Bigg|_0^1 = -50\left(\frac{1}{6} - \frac{1}{5}\right) = \frac{5}{3}$

   (c)$f_{y_1,y_5}(y_1, y_5) = \frac{5!}{3!}\left[\frac{y_5}{10} - \frac{y_1}{10}\right]^3 \frac{1}{10}\frac{1}{10} I_{(y_1,10)}(y_5) = \frac{1}{5}\left(\frac{y_1 - y_5}{10}\right)^3 I_{(y_1,10)}(y_5)$

6.2 Exercícios do Capítulo 2

1. Seja $X \sim Exp(\lambda)$. Para uma amostra aletória $X_1, X_2, \ldots, X_n$ dessa distribuição, testar se $T = \sum_{i=1}^{n} X_i$ é estatística suficiente para λ.

   Solução: Se $X \sim Exp(\lambda)$, então $f_X(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x)$. Logo sua densidade conjunta de $X_1, X_2, \ldots, X_n$ é dada por:

$$f(x_1, x_2, \ldots, x_n|\lambda) = \prod_{i-1}^{n} \lambda e^{-\lambda x_i} I_{(0,\infty)}(x) = \lambda^n exp\left\{-\lambda \sum_{i-1}^{n} x_i\right\} \prod_{i-1}^{n} I_{(0,\infty)}(x)$$

Pelo Critério de fatoração de Neyman, definindo $g(t,\lambda) = \lambda^n exp\left\{-\lambda\sum_{i-1}^{n} x_i\right\}$ e $h(x_1,\ldots,x_n) = \prod_{i-1}^{n} I_{(0,\infty)}(x)$, então $T = \sum_{i=1}^{n} X_i$ é estatística suficiente para λ.

2. Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ de uma população com distribuição gama com parâmetros α e β. Achar uma estastística suficiente bidimensional para o parâmetro (α, β).

   Solução:

   Temos que:

$$f(x_1,x_2,\ldots,x_n|\alpha,\beta) = \prod_{i-1}^{n} \frac{\beta^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i} = \left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^n exp\left\{(\alpha-1)\sum_{i=1}^{n} lnx_i - \beta\sum_{i=1}^{n} x_i\right\}$$

   Portanto, $\left(\sum_{i=1}^{n} lnX_i, \sum_{i=1}^{n} X_i\right)$ é estatística suficiente para (α, β).

3. Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ de uma população com função de densidade de probabilidade dada por:

$$f(x|\theta) = \theta x^{\theta-1},\ 0 < x < 1,\ \theta > 0$$

   .

   Encontre uma estatística suficiente para θ.

   Solução:

   Para a densidade em questão, segue-se que:

$$f(x_1,x_2,\ldots,x_n|\theta) = \prod_{i-1}^{n} \theta x_i^{\theta-1} I_{(0,1)}(x) = \prod_{i-1}^{n} \theta exp\left\{(\theta-1)lnx_i\right\} I_{(0,1)}(x) = \theta^n exp\left\{(\theta-1)\sum_{i=1}^{n} lnx_i\right\} \prod_{i-1}^{n} I$$

   .

   Pelo Critério de fatoração de Neyman, definindo $g(t,\theta) = \theta^n exp\left\{(\theta-1)\sum_{i=1}^{n} lnx_i\right\}$ e $h(x_1,\ldots,x_n) = \prod_{i-1}^{n} I_{(0,1)}(x)$, então $T = \sum_{i=1}^{n} lnX_i$ é estatística suficiente para θ.

4. Qual das seguintes famílias de massa/densidade pertencem à família exponencial? Para as que pertencem definir uma estatística suficiente.

   (a)$f(x|\theta) = 1/9, x \in \{0, 1+\theta, \ldots, 9+\theta\}$;
   (b)A família de distribuições $N(\theta, \theta^2)$;
   (c)$f(x|\theta) = 2\frac{(x+\theta)}{(1+2\theta)}, x \in (0,1), \theta > 0$;
   (d)$f(x|\theta) = \frac{\theta}{(1+x)^{1+\theta}}, x \in R^+$.

Solução:
(a) Veja que a própria função indicadora depende do parâmetro, dessa forma, não conseguimos separá-la como produto de funções que só dependem de x e só de θ. Logo, a família de densidades deste item não pertence à família exponencial.
(b)

$$\begin{aligned} f(\underset{\sim}{x}|\theta) &= \prod_{i-1}^{n} \frac{1}{\sqrt{2\pi\theta^2}} exp\left\{-\frac{1}{2\theta^2}(x_i-\theta)^2\right\} = (2\pi\theta^2)^{-\frac{n}{2}} exp\left\{-\frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i^2-2x_i\theta+\theta^2)\right\} \\ &= (2\pi\theta^2)^{-\frac{n}{2}} exp\left\{-\frac{\sum_{i=1}^{n} x_i^2}{2\theta^2} + \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{n}{2}\right\} \end{aligned}$$

Não conseguimos escrever um produto de $c(\theta)$ por $d(x)$ dentro do núcleo da família exponencial, logo, a família de densidade acima não pertence à família exponencial.

(c)$f(\underset{\sim}{x}|\theta) = \prod_{i-1}^{n} \frac{2(x_i+\theta)}{1+2\theta} = \left(\frac{2}{1+2\theta}\right)^n exp\left\{\sum_{i=1}^{n} ln(x_i+\theta)\right\}.$
Não, pois é impossível transformar soma de uma variável aleatória com um parâmetro, em um produto.

(d) $f(\underset{\sim}{x}|\theta) = \prod_{i-1}^{n} \frac{\theta}{(1+x)^{1+\theta}} = \theta^n exp\left\{-(1+\theta)\sum_{i=1}^{n} ln(x_i+1)\right\}$

Temos que, a família de densidade acima pertence à família exponencial e $T = \sum_{i=1}^{n} ln(X_i + 1)$ é estatística suficiente para θ.

5. Achar uma estatística suficiente para θ de uma distribuição uniforme no intervalo $[0, \theta]$.

Solução: Se $X \sim U[0, \theta]$, então $f_X(x) = \frac{1}{\theta} I_{(0,\theta)}(x)$. E, sua densidade conjuta é dada por:

$$f(\underset{\sim}{x}|\theta) = \left(\frac{1}{\theta}\right)^n \prod_{i-1}^{n} I_{(0,\theta)}(x_i) = \left(\frac{1}{\theta}\right)^n I_{(x_{(n)},\infty)}(\theta) \times 1 = g(x_{(n)}|\theta) \times h(x_1, x_2, \ldots, x_n)$$

Pelo Critério de Fatoração de Neyman, temos que $S(x_1, x_2, \ldots, x_n) = X_n$ é estatística suficiente para θ.

6. Seja X uma variável aleatória com função de densidade

$$f(x|\theta) = \frac{4}{\theta} x^3 exp\{-x^4/\theta\},\ x > 0,\ \theta > 0$$

Ache uma estatística suficiente para θ. Justifique sua escolha.

Solução: Claramente, $f(x|\theta)$ pertence à família exponencial, portanto, para uma amostra aleatória de tamanho n, $\sum_{i-1}^{n} d(X_i) = \sum_{i-1}^{n} X_i^4$ é estatística suficiente para θ.

6.3 Exercícios do Capítulo 3

1. Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ de uma população com distribuição Bernoulli de parâmetro θ.

   (a) Achar a cota inferior de Cramer-Rao para a variância de estimadores não viciados de $\theta(1-\theta)$.
   (b)Baseado no EMV ache um ENVMV para $\theta(1-\theta)$.

   Solução:

   (a) Temos que $\tau(\theta) = \theta(1-\theta) = \theta - \theta^2$. Seja $T = T(X_1, \ldots, X_n)$ um estimador não viciado de $\tau(\theta)$. Temos que, $Var(T)$ satisfaz,

$$Var(T) \geq \frac{\left[\frac{\partial}{\partial\theta}\tau(\theta)\right]^2}{nE\left[\left(\frac{\partial}{\partial\theta}lnf(X|\theta)\right)^2\right]} \quad (*)$$

   Veja que, $f(x|\theta) = \theta^x(1-\theta)^{1-x}I_{\{0,1\}}(x)$. Aplicando ln na função de densidade, segue-se que:

$$lnf(x|\theta) = xln\theta + (1-x)ln(1-\theta)$$

   .

   Derivando a função de log verossimilhança em relação à θ, temos:

$$\frac{\partial}{\partial\theta}lnf(x|\theta) = \frac{x}{\theta} - \frac{1-x}{1-\theta} = \frac{x-\theta}{\theta(1-\theta)}$$

   .

   Sendo assim,

$$E\left[\left(\frac{\partial}{\partial\theta}lnf(X|\theta)\right)^2\right] = E\left[\left(\frac{X-\theta}{\theta(1-\theta)}\right)^2\right] = \frac{1}{\theta^2(1-\theta)^2}E\left[(X-\theta)^2\right] = \frac{1}{\theta^2(1-\theta)^2}Var(X)$$

$$= \frac{1}{\theta^2(1-\theta)^2}\theta(1-\theta) = \frac{1}{\theta(1-\theta)} \quad (1)$$

   Por fim, $\left[\frac{\partial}{\partial\theta}\tau(\theta)\right]^2 = \left[\frac{\partial}{\partial\theta}(\theta-\theta^2)\right] = (1-2\theta)^2 \quad (2)$
   Substituindo (1) e (2) em (*), temos que:

$$Var(T) \geq \frac{(1-2\theta)^2}{n\frac{1}{\theta(1-\theta)}}$$

   .

   (b)Veja que, $\displaystyle\sum_{i=1}^{n}\frac{\partial}{\partial\theta}lnf(x|\theta) = \sum_{i=1}^{n}\left[\frac{X-\theta}{\theta(1-\theta)}\right] = \frac{n}{\theta(1-\theta)}\left[\frac{\sum_{i=1}^{n}x_i}{n} - \theta\right] = \frac{n}{\theta(1-\theta)}[\bar{x}-\theta].$

   Logo, sabemos por um lem que $\bar{X}$ é ENVMV para $\tau(\theta) = \theta$. Portanto, $\bar{X}(1-\bar{X})$ é ENVMV para $\tau(\theta) = \theta(1-\theta)$.

2. Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ de $f(x|\theta) = \theta x^{\theta-1} I_{(0,1)}(x)$, onde $\theta > 0$.

(a) Achar o EMV de $\mu = \theta/(1+\theta)$.

(b)Achar uma estatística suficiente e checar se é completa.

(c)Existe alguma função de θ para a qual existe um estimador não viciado cuja variância atinge a cota inferior de Cramer-Rao?

Solução:

(a)Primeiramente, vamos encontrar o EMV para θ.

$L(\theta|x_1,\ldots,x_n) = \prod_{i-1}^{n} \theta x^{\theta-1} I_{(0,1)}(x) = \theta^n \left[\prod_{i-1}^{n} x_i\right]^{\theta-1}.$

Aplicando o ln na função de verossimilhança, segue-se que,

$lnL(\theta|\underset{\sim}{(x)}) = nln\theta + (\theta-1)\sum_{i-1}^{n} lnx_i.$

A equação $\frac{\partial}{\partial\theta} lnL(\theta|\underset{\sim}{(x)}) = 0$ se reduz a :

$$\frac{n}{\theta} + \sum_{i-1}^{n} lnx_i = 0$$

que tem como solução, $\hat{\theta} = -\frac{n}{\sum_{i-1}^{n} lnX_i} = \frac{n}{\sum_{i-1}^{n} lnX_i^{-1}}$.

Como $\mu = \theta/(1+\theta)$ é função bijetora, então,

$$\hat{\mu} = \frac{\frac{n}{\sum_{i-1}^{n} lnX_i^{-1}}}{1 - \frac{n}{\sum_{i-1}^{n} lnX_i^{-1}}}$$

.

(b) $f(\underset{\sim}{x}|\theta) = \theta^n \left[\prod_{i-1}^{n} x_i\right]^{\theta-1} = \theta^n exp\left\{(\theta-1)\sum_{i-1}^{n} lnx_i\right\}.$

Como $f(x|\theta)$ pertence à família exponencial, podemos afirmar que $T = \sum_{i-1}^{n} lnX_i$ é estatística completa e minimal suficiente para θ.

(c)Por um lema sabemos que, se $T(X)$ é estimador não viciado de $\tau(\theta)$ então $T(X)$ atinge a cota inferior de Cramer-Rao s.s.s

$$\sum_{i=1}^{n} \frac{\partial}{\partial\theta} lnf(x|\theta) = a(\theta)\left[T(X) - \tau(\theta)\right]$$

Então, vamos tentar encontrar uma expressão análoga para esse caso,

$$\frac{\partial}{\partial\theta} lnf(x|\theta) = \frac{n}{\theta} + \sum_{i-1}^{n} lnx_i = -n\left[\frac{\sum_{i-1}^{n} lnx_i^{-1}}{n} - \frac{1}{\theta}\right]$$

Portanto, $T(X) = \dfrac{\sum_{i-1}^{n} lnx_i^{-1}}{n}$ é ENVMV para $\tau(\theta) = \frac{1}{\theta}$

3. Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ de $f(x|\theta) = \frac{ln\theta}{\theta-1}\theta^x I_{(0,1)}(x)$, onde $\theta > 0$.

   (a)Achar uma estatística suficiente.

   (b) Existe alguma função de $\tau(\theta)$ para a qual existe um estimador não viciado cuja variância atinge a cota inferior de Cramer-Rao?

   Solução:

   (a) Como $f(x|\theta) = \frac{ln\theta}{\theta-1}\theta^x$ pode ser escrito da seguinte forma:

$$f(x|\theta) = \frac{ln\theta}{\theta - 1} exp\{xln\theta\}$$

   ou seja, $f(x|\theta)$ pertence à família exponencial e consequentemente $T = \sum_{i-1}^{n} X_i$ é estatística suficiente para θ.

   (b) Façamos as contas,

$$\begin{aligned}
\sum_{i=1}^{n} \frac{\partial}{\partial\theta} lnf(x|\theta) &= \sum_{i=1}^{n}\left[\frac{\partial}{\partial\theta}\left(ln(ln\theta) - ln(\theta-1) + xln\theta\right)\right] = \sum_{i=1}^{n}\left[\frac{1}{\theta ln\theta} - \frac{1}{\theta - 1} + \frac{x}{\theta}\right] \\
&= \sum_{i=1}^{n}\left[\frac{\theta - 1 - \theta ln\theta}{(\theta-1)\theta ln\theta} + \frac{x}{\theta}\right] = n\left[\frac{\theta - 1 - \theta ln\theta}{(\theta-1)\theta ln\theta}\right] + \frac{\sum_{i=1}^{n} x_i}{n} \\
&= \frac{n}{\theta}\left[\bar{x} - \frac{\theta - 1 - \theta ln\theta}{(\theta-1)\theta ln\theta}\right].
\end{aligned}$$

   Logo, existe $\tau(\theta) = \frac{\theta-1-\theta ln\theta}{(\theta-1)\theta ln\theta}$ para o qual existe um estimador não viciado cuja variância atinge a cota inferior de Cramer-Rao, que é $\bar{X}$.

4. Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ de uma população com distribuição Bernoulli de parâmetro θ. Mostrar que a variância de $\bar{X}$ atinge a cota inferior de Cramer-Rao e que $\bar{X}$ é ENVMV de θ.

   Solução:

   $Var(\bar{X}) = Var\left(\dfrac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}(n\theta(1-\theta)) = \frac{\theta(1-\theta)}{n}$

   Lembre-se que se $X \sim Ber(\theta)$, então $Y = \sum_{i=1}^{n} X_i \sim Bin(n, \theta)$, portanto $Var(Y) = n\theta(1-\theta)$.

   Por outro lado, para $T = T(\underset{\sim}{X}) = \bar{X}$ um estimador não viciado de $\tau(\theta) = \theta$, a cota inferior de Cramer-Rao é dada por:

$$Var(T) \geq \frac{\left[\frac{\partial}{\partial\theta}\tau(\theta)\right]^2}{nE\left[\left(\frac{\partial}{\partial\theta}lnf(X|\theta)\right)^2\right]} \quad (*)$$

Veja que, $f(x|\theta) = \theta^x(1-\theta)^{1-x}I_{\{0,1\}}(x)$. Aplicando ln na função de massa de probabilidade, segue-se que:

$$lnf(x|\theta) = xln\theta + (1-x)ln(1-\theta)$$

.

Derivando a função de log verossimilhança em relação à θ, temos:

$$\frac{\partial}{\partial\theta}lnf(x|\theta) = \frac{x}{\theta} - \frac{1-x}{1-\theta} = \frac{x-\theta}{\theta(1-\theta)}$$

.

Sendo assim,

$$E\left[\left(\frac{\partial}{\partial\theta}lnf(X|\theta)\right)^2\right] = E\left[\left(\frac{X-\theta}{\theta(1-\theta)}\right)^2\right] = \frac{1}{\theta^2(1-\theta)^2}E\left[(X-\theta)^2\right] = \frac{1}{\theta^2(1-\theta)^2}Var(X)$$

$$= \frac{1}{\theta^2(1-\theta)^2}\theta(1-\theta) = \frac{1}{\theta(1-\theta)} \quad (1)$$

Por fim, $\left[\frac{\partial}{\partial\theta}\tau(\theta)\right]^2 = \left[\frac{\partial}{\partial\theta}(\theta)\right] = 1 \quad (2)$
Substituindo (1) e (2) em (*), temos que:

$$Var(T) \geq \frac{(1}{n\frac{1}{\theta(1-\theta)}} = \frac{\theta(1-\theta)}{n}$$

.

Conclui-se então que $\bar{X}$ atinge a cota inferior de Cramer-Rao.

Além disso,

$$\sum_{i=1}^{n}\frac{\partial}{\partial\theta}lnf(x|\theta) = \sum_{i=1}^{n}\left[\frac{X-\theta}{\theta(1-\theta)}\right] = \frac{n}{\theta(1-\theta)}\left[\frac{\sum_{i=1}^{n}x_i}{n} - \theta\right] = \frac{n}{\theta(1-\theta)}[\bar{x}-\theta].$$

Logo, sabemos por um lema que $\bar{X}$ é ENVMV para $\tau(\theta) = \theta$. Adicionalmente, $E[\bar{X}]\theta$, portanto, $\bar{X}$ é ENVMV de θ.

5. Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ de uma distribuição Gama de parâmetros $\alpha > 0$ e $\lambda > 0$, isto é, $f(x|\alpha,\lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}I_{(0,\infty)}(x)$ com α conhecido. Achar um ENVMV de $1/\lambda$ que atinge a cota inferior de Cramer-Rao.

Solução:

$$\begin{aligned}\sum_{i=1}^{n}\frac{\partial}{\partial\lambda}lnf(x|\alpha,\lambda) &= \sum_{i=1}^{n}\left[\frac{\partial}{\partial\lambda}\left(\alpha ln\lambda - ln\Gamma(\alpha)+(\alpha-1)lnx-\lambda x\right)\right] = \sum_{i=1}^{n}\left[\frac{\alpha}{\lambda}-x\right]\\ &= \frac{n\alpha}{\lambda}-\sum_{i=1}^{n}x_i = -n\alpha\left[\frac{\bar{x}}{\alpha}-\frac{1}{\lambda}\right]\end{aligned}$$

Dessa forma, $\frac{\bar{X}}{\alpha}$ é ENVMV para $\tau(\lambda)=\frac{1}{\lambda}$.

6. Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ de uma distribuição Binomial de parâmetro θ e m, este último conhecido.

   (a) Estimar θ pelo método dos momentos e pelo método de máxima verossimilhança.

   (b)Existe algum ENVMV de θ? Determine-o.

   Solução:

   (a) Comecemos pelo método dos momentos. Se $X \sim Bin(m,\theta)$, então $E[X]=m\theta$. Logo,

$$\frac{\sum_{i=1}^{n}x_i}{n}=m\theta$$

   Sendo assim, $\hat{\theta}=\frac{\bar{X}}{m}$.

   Seguindo, pelo método de máxima verossimilhança,

$$L(\theta|\underset{\sim}{x})=\prod_{i=1}^{n}\binom{m}{x}\theta^{x}(1-\theta)^{m-x}=\left[\prod_{i=1}^{n}\binom{m}{x}\right]\theta^{\sum_{i=1}^{n}x_i}(1-\theta)^{mn-\sum_{i=1}^{n}x_i}$$

$$\Rightarrow lnL(\theta|\underset{\sim}{x})=ln\left[\prod_{i=1}^{n}\binom{m}{x}\right]+\sum_{i=1}^{n}x_i ln\theta+(mn-\sum_{i=1}^{n}x_i)ln(1-\theta)$$

$$\Rightarrow \frac{\partial}{\partial\theta}lnL(\theta|\underset{\sim}{x})=\frac{\sum_{i=1}^{n}x_i}{\theta}-\frac{mn-\sum_{i=1}^{n}x_i}{1-\theta}=\frac{\sum_{i=1}^{n}x_i-nm\theta}{\theta(1-\theta)}$$

   Igualando a derivada à zero e isolando o parâmetro de interesse θ, encontramos $\hat{\theta}=\frac{\bar{X}}{m}$ é EMV.

   (b) Como $E\left[\frac{\bar{X}}{m}\right]=\frac{1}{m}E[X]=\frac{1}{m}m\theta=\theta$, isto é, $T(\underset{\sim}{X})=\frac{\bar{X}}{m}$ é estimador não viciado de $\tau(\theta)=\theta$. Além disso, $V\left[\frac{X}{m}\right]=\frac{1}{m^2}V[X]=\frac{m\theta(1-\theta)}{m^2}=\frac{\theta(1-\theta)}{m}$, e por outro lado,

$$Var(T)\geq\frac{\left[\frac{\partial}{\partial\theta}\tau(\theta)\right]^2}{mE\left[\left(\frac{\partial}{\partial\theta}lnf(X|\theta)\right)^2\right]}\quad(*)$$

   onde,

$$\frac{\partial}{\partial\theta}lnf(x|\theta)=\frac{x}{\theta}-\frac{1-x}{1-\theta}=\frac{x-\theta}{\theta(1-\theta)}$$

   .

Portanto,

$$E\left[\left(\frac{\partial}{\partial\theta}lnf(X|\theta)\right)^2\right] = E\left[\left(\frac{X-\theta}{\theta(1-\theta)}\right)^2\right] = \frac{1}{\theta^2(1-\theta)^2}E\left[(X-\theta)^2\right] = \frac{1}{\theta^2(1-\theta)^2}Var(X)$$

$$= \frac{1}{\theta^2(1-\theta)^2}\theta(1-\theta) = \frac{1}{\theta(1-\theta)} \quad (1)$$

Substituindo em (*), temos:

$$Var(T) \geq \frac{1}{m\frac{1}{\theta(1-\theta)}} = \frac{\theta(1-\theta)}{m},$$

.

isto é, $T(X)$ atinge a cota inferior de Cramer-Rao, então por definição, $\frac{\bar{X}}{m}$ é ENVMV para θ.

7. Para uma amostra aleatória de tamanho n de uma distribuição de Poisson:

   (a) Testar se $T(X_1,\ldots,X_n) = I_{0,1}(x_1)$ é estimador não viciado de $\tau(\lambda) = (1+\lambda)e^{-\lambda}$.

   (b) Achar a cota inferior de Cramer-Rao de estimadores não viciados de $\tau(\lambda)$.

   (c)Achar uma estatística suficiente para λ.

   (d) Achar o ENVMV de $\tau(\lambda)$.

   Solução:

   (a)$E[T] = E[I_{\{0,1\}}(x)] = P(X=0)+P(X=1) = \frac{e^{-\lambda}\lambda^0}{0!} + \frac{e^{-\lambda}\lambda^1}{1!} = e^{-\lambda} + e^{-\lambda}\lambda = e^{-\lambda}(1+\lambda) = \tau(\lambda)$

   (b) Seja $\tau(\lambda) = (1+\lambda)e^{-\lambda}$, entao

$$\frac{\partial}{\partial\lambda}\tau(\lambda) = -e^{-\lambda} + e^{-\lambda} - \lambda e^{-\lambda} = -\lambda e^{-\lambda}$$

   Seguindo, defina $T = T(X_1,\ldots,X_n)$ um estimador não viesado de $\tau(\lambda)$. Queremos encontrar a expressão ao lado direito de:

$$Var(T) \geq \frac{\left[\frac{\partial}{\partial\lambda}\tau(\lambda)\right]^2}{-nE\left[\frac{\partial^2}{\partial\lambda^2}lnf(X|\theta)\right]} \quad (*)$$

   Sendo assim,

   - $\frac{\partial}{\partial\lambda}lnf(x|\lambda) = \frac{\partial}{\partial\lambda}\left[xln\lambda - \lambda - lnx!\right] = \frac{x}{\lambda} - 1$
   - $\frac{\partial^2}{\partial\lambda^2}lnf(x|\lambda) = \frac{\partial}{\partial\lambda}\left[\frac{x}{\lambda} - 1\right] = -\frac{x}{\lambda^2}$.

   Dessa forma,

$$Var(T) \geq \frac{(-\lambda e^{-\lambda})^2}{-E\left[-\frac{x}{\lambda^2}\right]} = \frac{\lambda^2 e^{-2\lambda}}{\frac{1}{\lambda^2}E[X]} = \frac{\lambda^2 e^{-2\lambda}}{\frac{1}{\lambda^2}\lambda} = \lambda^3 e^{-2\lambda}$$

   (c)Como $f(x|\lambda) = e^{-\lambda}\frac{1}{x!}exp\{xln\lambda\}$, pertence à família exponencial, então $\sum_{i=1}^{n} X_i$ é estatística suficiente para λ.

(d) Sabemos que se $X \sim Poisson(\lambda)$, então $S = \sum_{i=1}^{n} X_i$ é estatística suficiente para λ. Defina $T' = E[T|S]$. Então,

$$\begin{aligned}
T' &= E\left[I_{\{0,1\}}(x)|\sum_{i=1}^{n} X_i = s\right] = P\left[X = 0|\sum_{i=1}^{n} X_i = s\right] + P\left[X = 1|\sum_{i=1}^{n} X_i = s\right] \\
&= \frac{P(X=0)P\left(\sum_{i=2}^{n} X_i = s\right)}{P\left(\sum_{i=1}^{n} X_i = s\right)} + \frac{P(X=1)P\left(\sum_{i=2}^{n} X_i = s-1\right)}{P\left(\sum_{i=1}^{n} X_i = s\right)} \\
&= \frac{\frac{\left(e^{-\lambda}\right)((n-1)\lambda)^s}{s!} + \frac{\lambda e^{-\lambda} e^{-(n-1)\lambda}((n-1)\lambda)^{s-1}}{(s-1)!}}{\frac{e^{-n\lambda}(n\lambda)^s}{s!}} \\
&= \frac{(n-1)^{s-1}[n-1+s]}{n^s}
\end{aligned}$$

Logo, $T' = \frac{(n-1)^{s-1}[n-1+s]}{n^s}$ é ENVMV para $\tau(\lambda)$.

8. Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ de $f(x|\theta) = \theta(1+x)^{-(1+\theta)} I_{(0,\infty)}(x),\ \theta > 0$.
   (a) Achar o EMV de $1/\theta$.
   (b) Achar uma estatística completa e minimal suficiente, se existir.
   (c)Achar a cota inferior de Cramer-Rao para estimadores não viciados de $1/\theta$.
   (d)Achar um ENVMV de $1/\theta$, se existir.

Solução:
(a) Neste caso, a função de verossimilhança é expressa por:

$$L(\theta|\underset{\sim}{x}) = \theta^n \left[\prod_{i=1}^{n}(1+x_i)\right]^{-(1+\theta)}$$

.

E, sua respectiva função de log verossimilhança,

$$lnL(\theta|\underset{\sim}{x}) = nln\theta - (1+\theta)\sum_{i=1}^{n} ln(1+x_i)$$

A sua respectiva derivada igualada á zero, resulta em:

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} ln(1+X_i)}$$

Como, $\frac{1}{\theta}$ é função bijetora de θ, então seu respectivo EMV é dado por $\frac{1}{\hat{\theta}} = \frac{\sum_{i=1}^{n} ln(1+X_i)}{n}$.
(b) $f(x|\theta) = \theta exp\{-(1+\theta)ln(1+x)\}$, ou seja, pertence à família exponencial,e portanto, $\sum_{i=1}^{n} d(X_i) = \sum_{i=1}^{n} ln(1+X_i)$ é estatística completa e minimal suficiente.
(c)Façamos as contas,

- $\frac{\partial}{\partial\theta} lnf(x|\theta) = \frac{\partial}{\partial\theta}\left(ln\theta - (1+\theta)ln(1+x)\right) = \frac{1}{\theta} - ln(1+x)$
- $\frac{\partial^2}{\partial\theta^2} lnf(x|\theta) = -\frac{1}{\theta^2}$
- $\frac{\partial}{\partial\theta}\tau(\theta) = \frac{\partial}{\partial\theta}\frac{1}{\theta} = -\frac{1}{\theta^2}$

A cota inferior de Cramer-Rao para estimadores não viciador de $\frac{1}{\theta}$ será dada por:

$$Var(T) \geq \frac{\left(-\frac{1}{\theta^2}\right)^2}{-nE\left(-\frac{1}{\theta^2}\right)} = \frac{1}{n\theta^2}$$

.

(d)

$$\begin{aligned}\sum_{i=1}^{n}\frac{\partial}{\partial\theta} lnf(x|\theta) &= \sum_{i=1}^{n}\left[\frac{1}{\theta} - ln(1+x)\right] = \frac{n}{\theta} - \sum_{i=1}^{n} ln(1+x) \\ &= -n\left[\frac{\sum_{i=1}^{n} ln(1+x_i)}{n} - \frac{1}{\theta}\right]\end{aligned}$$

Concluímos que $T = \frac{\sum_{i=1}^{n} ln(1+X_i)}{n}$ é ENVMV.

6.4 Exercícios do Capítulo 4

1. Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição normal de média desconhecida μ e variância conhecida σ_0^2. Provar que o teste mais poderoso de tamanho α para testar $H_0 : \mu = \mu_0$ contra $H_1 : \mu = \mu_1$, $\mu_0 < \mu_1$ é γ: rejeitar H_0 se $\bar{x} \geq \mu_0 + \frac{\sigma_0^2}{\sqrt{n}} z_{1-\alpha}$.

   Solução: Primeiramente, veja que:

   $L(\theta|\underset{\sim}{x}) = (2\pi)^{-\frac{n}{2}}(\sigma_0^2)^{-\frac{n}{2}} exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}.$

   Seja $\lambda = \frac{L_0}{L_1}$, isto é,

   $\lambda = \frac{L(\mu_0|\underset{\sim}{x})}{L(\mu_1|\underset{\sim}{x})} = exp\left\{-\frac{1}{2\sigma_0^2}\left[\sum_{i=1}^{n}(x_i-\mu_0)^2 - \sum_{i=1}^{n}(x_i-\mu_1)^2\right]\right\} = exp\left\{-\frac{1}{2\sigma_0^2}\left[n(\mu_0^2-\mu_1^2) - 2\sum_{i=1}^{n}x_i(\mu_0-\mu_1\right.\right.$

   $= exp\left\{\frac{-n(\mu_0^2-\mu_1^2)}{2\sigma_0^2}\right\} exp\left\{\frac{2\sum_{i=1}^{n}x_i(\mu_0-\mu_1)}{2\sigma_0^2}\right\}.$

   Por Neyman-Pearson, temos o seguinte teste:

   γ : rejeitar H_0 se $\lambda \leq K$.

$$exp\left\{\frac{-n(\mu_0^2-\mu_1^2)}{2\sigma_0^2}\right\} exp\left\{\frac{2\sum_{i=1}^{n}x_i(\mu_0-\mu_1)}{2\sigma_0^2}\right\} \leq K$$

$$exp\left\{\frac{2\sum_{i=1}^{n}x_i(\mu_0-\mu_1)}{2\sigma_0^2}\right\}\leq K_1$$

$$\frac{(\mu_0-\mu_1)\sum_{i=1}^{n}x_i}{\sigma_0^2}\leq K_2$$

.

Como $\mu_1 > \mu_0$, então $\mu_0 - \mu_1 < 0$, logo:

$$\sum_{i=1}^{n}x_i \geq K_3$$

$$\bar{x}\geq K_4$$

Considerando que $\alpha = P(\bar{X}\geq K_4|\mu=\mu_0) = 1-\Phi\left(\frac{K_4-\mu_0}{\frac{\sigma_0}{\sqrt{n}}}\right)$

$$\left(\frac{K_4-\mu_0}{\frac{\sigma_0}{\sqrt{n}}}\right) = z_{1-\alpha}$$

$$K_4 = \mu_0 + \frac{\sigma_0^2}{\sqrt{n}}z_{1-\alpha}$$

Temos o teste mais poderoso definido como: γ : rejeitar H_0 se $\bar{x}\geq \mu_0 + \frac{\sigma_0^2}{\sqrt{n}}z_{1-\alpha}$.

2. Seja $X_1,\ldots,X_n$ uma amostra aleatória de tamanho n de uma população com distribuição normal de média μ_0 conhecida e variância σ^2. Determine o teste mais poderoso para testar $H_0 : \sigma^2 = \sigma_0^2$ contra $H_1 : \sigma^2 = \sigma_1^2$, sendo $\sigma_0^2 < \sigma_1^2$.

   Solução:

   Com base na função de verossimilhança já definida na questão 1, temos que:

   $\lambda = \frac{L(\sigma_0^2|\underset{\sim}{x})}{L(\sigma_1^2|\underset{\sim}{x})} = \left(\frac{\sigma_1^2}{\sigma_0^2}\right)^{\frac{n}{2}} exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu_0)^2\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)\right\}$

   Por Neyman-Pearson, segue-se que,

   γ : rejeita H_0 se $\lambda \leq k$.

   Então,

$$\left(\frac{\sigma_1^2}{\sigma_0^2}\right)^{\frac{n}{2}} exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu_0)^2\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)\right\}\leq k$$

$$exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu_0)^2\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)\right\}\leq k_1$$

$$-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu_0)^2\left(\frac{1}{\sigma_0^2}-\frac{1}{\sigma_1^2}\right)\leq k_2$$

Como $\sigma_1^2 > \sigma_0^2$, então, $\frac{1}{\sigma_1^2} < \frac{1}{\sigma_0^2}$ e, consequentemente, $\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2} > 0$.

$$-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu_0)^2\leq k_3$$

$$\sum_{i=1}^{n}(x_i-\mu_0)^2\geq k_4$$

$$\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\sigma^2}\geq k_5$$

Dessa forma:

$$\alpha = P\left(\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\sigma^2}\geq k_5|\sigma^2=\sigma_0^2\right) = P\left(\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\sigma_0^2}\geq k_5\right)$$

Portanto, $k_5 = \chi^2_{(1-\alpha),n}$ e assim, γ : rejeitar H_0 se $\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{\sigma_0^2}\geq \chi^2_{(1-\alpha),n}$, isto é, rejeitar H_0 se $\sum_{i=1}^{n}(x_i-\mu_0)^2\geq \chi^2_{(1-\alpha),n}\sigma_0^2$

3. Supor que X tem distribuição de Bernoulli de parâmetro θ. Considere uma amostra aleatória de tamanho 4 de X. Calcular α e β se queremos testar $H_0 : \theta = \frac{1}{4}$ contra $H_1 : \theta = \frac{3}{4}$ e adotamos o teste γ : rejeitar H_0 se obtivermos 4 sucessos na amostra.

   Solução: Por hipótese do enunciado, temos que:

   γ : rejeitar H_0 se $\sum_{i=1}^{4}X_i = 4$. Lembrando que se $X\sim Ber(\theta)$, então $\sum_{i=1}^{4}X_i\sim(4,\theta)$.

   Vamos calcular α e β tal que:

   - $\alpha = P[\text{rejeitar } H_0|\theta=1/4] = P\left[\sum_{i=1}^{4}X_i=4|\theta=1/4\right] = \binom{4}{4}\left(\frac{1}{4}\right)^4\left(1-\frac{1}{4}\right)^{4-4} = \left(\frac{1}{4}\right)^4$
   - $\beta = P[\text{aceitar } H_0|\theta=3/4] = 1-P[\text{rejeitar } H_0|\theta=3/4] = 1-P\left[\sum_{i=1}^{4}X_i=4|\theta=3/4\right] =$ $1-\binom{4}{4}\left(\frac{3}{4}\right)^4\left(1-\frac{3}{4}\right)^{4-4} = 1-\left(\frac{3}{4}\right)^4$

4. Derive o teste mais poderoso de tamanho α para testar $H_0 : \lambda = \lambda_0$ contra $H_1 : \lambda = \lambda_1$, $\lambda_0 < \lambda_1$, baseados numa amostra aleatória de n observações de uma variável aleatória exponencial de parâmetro λ.

Solução: Se $X \sim Exp(\lambda)$, então $L(\lambda|\underset{\sim}{x}) = \lambda^n exp\left\{-\sum_{i=1}^{n} x_i\lambda\right\}$

Logo, $\lambda = \frac{L_0}{L_1} = \dfrac{\lambda_0^n exp\left\{-\sum_{i=1}^{n} x_i\lambda_0\right\}}{\lambda_1^n exp\left\{-\sum_{i=1}^{n} x_i\lambda_1\right\}} = \left(\frac{\lambda_0}{\lambda_1}\right)^n exp\left\{-\sum_{i=1}^{n} x_i(\lambda_0 - \lambda_1)\right\}$.

Por Neyman-Pearson, temos que:

γ : rejeita H_0 se $\lambda \leq k$.

$$\left(\frac{\lambda_0}{\lambda_1}\right)^n exp\left\{-\sum_{i=1}^{n} x_i(\lambda_0 - \lambda_1)\right\} \leq k$$

$$exp\left\{-\sum_{i=1}^{n} x_i(\lambda_0 - \lambda_1)\right\} \leq k_2$$

Como $\lambda_1 > \lambda_0, \lambda_0 - \lambda_1 < 0$ então -$(\lambda_0 - \lambda_1) > 0$.

$$-\sum_{i=1}^{n} x_i(\lambda_0 - \lambda_1) \leq k_3$$

$$\sum_{i=1}^{n} x_i \leq k_4$$

$$2\lambda\sum_{i=1}^{n} x_i \leq k_5$$

Então, $\alpha = P\left(2\lambda\sum_{i=1}^{n} x_i \leq k_5|\lambda = \lambda_0\right) = P\left(2\lambda_0\sum_{i=1}^{n} x_i \leq k_5\right)$

$$k_5 = \chi^2_{\alpha,2n}$$

Portanto, γ : rejeitar H_0 se $2\lambda_0\sum_{i=1}^{n} x_i \leq \chi^2_{\alpha,2n}$, isto é, γ : rejeitar H_0 se $\sum_{i=1}^{n} x_i \leq \chi^2_{\alpha,2n}2\lambda_0$

5. Seja X uma única observação de $f(x|\theta) = \theta x^{\theta-1}I_{(0,1)}(x)$,$\theta > 0$.

(a)Para testar $H_0 : \theta \leq 1$ contra $H_1 : \theta > 1$, achar a função poder e o tamanho do teste dado por γ : rejeitar H_0 se e somente se $x \geq 1/2$.

(b)Achar o teste mais poderoso de tamanho α para testar $H_0 : \theta = 2$ contra $H_1 : \theta = 1$.

Solução:

(a) A função poder é definida da seguinte forma:

$\Pi_\gamma(\theta) = P[rejeitar H_0|\theta] = P[X \geq 1/2|\theta] = \int_{1/2}^{1} \theta x^{\theta-1}dx = 1 - \left(\frac{1}{2}\right)^\theta$.

Com base na função poder encontrada,

$\alpha = sup_{\theta\leq 1}\Pi_\gamma(\theta) = 1 - \left(\frac{1}{2}\right)^1 = \frac{1}{2}$

(b) $\lambda = \frac{L_0}{L_1} = \frac{f(x|\theta=2)}{f(x|\theta=1)} = \frac{2x}{1} = 2x$

γ : rejeitar H_0 se $2x \leq k$, o equivalente a rejeitar H_0 se $x \leq k_1$.

Portanto, $\alpha = P(X \leq k_1|\theta = 2) = \displaystyle\int_0^{k_1} 2xdx = k_1^2$.

$$k_1 = \sqrt{\alpha}$$

.

Logo, γ : rejeitar H_0 se $x \leq \sqrt{\alpha}$.

6. Seja X uma única observação de $f(x|\theta) = (2\theta x + 1 - \theta)I_{(0,1)}(x)$, onde $-1 \leq \theta \leq 1$.

   (a)Achar o teste mais poderoso de tamanho α para testar $H_0 : \theta = 0$ contra $H_1 : \theta = 1$.

   (b)Para testar $H_0 : \theta \leq 0$ contra $H_1 : \theta > 0$, adotamos o seguinte procedimento γ : rejeitar H_0 se e somente se $x > 1/2$. Achar a função poder e o tamanho do teste.

   Solução:

   (a)$\lambda = \frac{L_0}{L_1} = \frac{f(x|\theta=0)}{f(x|\theta=1)} = \frac{1}{2x}$

   γ : rejeitar H_0 se $\frac{1}{2x} \leq k$, o equivalente a rejeitar H_0 se $x \geq k_1$.

   Portanto, $\alpha = P(X \geq k_1|\theta = 0) = \displaystyle\int_{k_1}^{1} dx = 1 - k_1$.

$$k_1 = 1 - \alpha$$

   .

   Logo, γ : rejeitar H_0 se $x \geq 1 - \alpha$.

   (b)$\Pi_\gamma(\theta) = P[rejeitar H_0|\theta] = P[X > 1/2|\theta] = \displaystyle\int_{1/2}^{1}(2\theta x + 1 - \theta)dx = \frac{\theta}{4} + \frac{1}{2}$.

   Com base na função poder encontrada,

   $\alpha = sup_{\theta \leq 0}\Pi_\gamma(\theta) = \Pi_\gamma(0) = \frac{1}{2}$.

7. Para uma amostra aleatória de tamanho 3 de uma Bernoulli de parâmetro θ queremos testar $H_0 : \theta \leq 0,5$ contra $H_0 : \theta > 0,5$ usando as seguintes regiões críticas:

$$C_{\gamma'} = \left\{(x_1, x_2, x_3)|\sum_{i=1}^{3} x_i = 3\right\}$$

$$C_{\gamma''} = \left\{(x_1, x_2, x_3)|\sum_{i=1}^{3} x_i \geq 2\right\}$$

   Comparar as funções poder.

   Solução: Para o primeiro caso, temos:

   $\Pi_{\gamma'}(\theta) = P\left(\displaystyle\sum_{i=1}^{3} x_i = 3|\theta\right) = \binom{3}{3}\theta^3(1-\theta)^{3-3} = \theta^3$.

   Já para $C_{\gamma''}$,

   $\Pi_{\gamma''}(\theta) = P\left(\displaystyle\sum_{i=1}^{3} x_i \geq 2|\theta\right) = 1 - P\left[\left(\displaystyle\sum_{i=1}^{3} x_i = 0\right) + \left(\displaystyle\sum_{i=1}^{3} x_i = 1\right)\right]$

$= 1 - \left[\binom{3}{0}\theta^0(1-\theta)^{3-0} + \binom{3}{1}\theta^1(1-\theta)^{3-1}\right] = 1-(1-\theta)^3 - 3\theta(1-\theta)^2.$

Como $sup_{\theta\leq 1/2}\Pi_{\gamma'}(\theta) = 0,125 < 0,5 = sup_{\theta\leq 1/2}\Pi_{\gamma''}(\theta)$, então é preferível $\Pi_{\gamma'}(\theta)$.

8. Seja $X_1,\ldots,X_n$ uma amostra aleatória de $f(x|\theta) = \frac{1}{\theta}x^{(1-\theta)/\theta}I_{(0,1)}(x)$. Achar o teste uniformemente mais poderoso de tamanho α para testar $H_0 : \theta \leq \theta_0$ contra $H_1 : \theta > \theta_0$.

   Solução:

   Veja que:

   $f(x|\theta) = \frac{1}{\theta}exp\left\{\frac{1-\theta}{\theta}lnx\right\}$.

   Como $c(\theta) = \frac{1-\theta}{\theta}$ é função monótona decrescente em θ, então, por Karlim-Rubim adaptdo para distribuições que pertencem à família exponencial, existe k tal que

$$P\left(\sum_{i=1}^n lnX_i < k|\theta_0\right) = \alpha$$

$$P\left(-\sum_{i=1}^n lnX_i > k|\theta_0\right) = \alpha$$

   Seja $Y = -lnX$, então $X = e^{-y}$. Dessa forma,

$$f_Y(y) = \left|\frac{dx}{dy}\right| f_X(e^{-y}) = \left|-e^{-y}\right| \frac{1}{\theta}e^{-y(1-\theta)/\theta} = \frac{1}{\theta}e^{-\frac{y}{\theta}}$$

   .

   Logo $Y \sim Exp\left(\frac{1}{\theta}\right) \Rightarrow -\sum_{i=1}^n lnX_i \sim Gama\left(n, \frac{1}{\theta}\right) \Rightarrow -\frac{2\sum_{i=1}^n lnX_i}{\theta} \sim \chi^2_{2n}$.

   Seguindo, temos que:

$$\alpha = P\left(-\frac{2\sum_{i=1}^n lnX_i}{\theta} > k\right)$$

$$k = \chi^2_{(1-\alpha),2n}$$

   O teste mais poderoso de tamanho α é dado por:

   γ : rejeitar H_0 se $\sum_{i=1}^n lnX_i < \frac{-\chi^2_{(1-\alpha),2n}\theta_0}{2}$.

9. Seja $X_1,\ldots,X_n$ uma amostra aleatória de $f(x|\theta) = \theta^2xe^{-\theta x}I_{(0,\infty)}(x)$.Para testar H_0 : $\theta \leq 1$ contra $H_1 : \theta > 1$, para n=1, foi usado γ : rejeitar H_0 se $x_1 \leq 1$. Achar a função poder e o tamnho do teste.

   Solução: $\Pi_\gamma(\theta) = P(X_1 \leq 1|\theta) = \int_0^1 \theta^2xe^{-\theta x}dx =$

   - $u = x \Rightarrow du = dx$

- $e^{-\theta x}dx \Rightarrow v = -\frac{e^{-\theta x}}{\theta}$
  $= \theta^2 \left[x\left(-\frac{e^{-\theta x}}{\theta}\right) + \int_0^1 \frac{e^{-\theta x}}{\theta}dx \right] = -e^{-\theta}(\theta+1)+1.$
  Além disso,
  $\alpha = sup_{\theta \leq 1}\Pi_\gamma(\theta) = \Pi_\gamma(1) = 0,2642$

6.5 Exercício do Capítulo 5

1. Seja $X_1,\ldots,X_n$ uma amostra aleatória de uma distribuição normal de média μ e variância σ^2. Construir um intervalo de $1-\alpha$ de confiança para μ da forma $IC(\underset{\sim}{x}) = \left(-\infty, U(\underset{\sim}{x})\right)$.

   Solução:

   Seja $X \sim N(\mu,\sigma^2)$, queremos encontrar $IC(\underset{\sim}{x}) = \left(-\infty, U(\underset{\sim}{x})\right)$. Dessa forma, vamos considerar as seguintes hipóteses:
   $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$
   Sendo assim, o teste será dado por:
   γ : rejeitar H_0 se $\frac{\bar{x}-\mu_0}{\frac{S}{\sqrt{n}}} < -t_{(n-1)}$, isto é,$\gamma$: rejeitar H_0 se $\mu_0 \leq \bar{x} + t_{(n-1),\alpha}\frac{S}{\sqrt{n}}$
   Portanto, o IC será expresso por:

   $$IC(\underset{\sim}{x}) = \left(-\infty, \bar{x} + t_{(n-1),\alpha}\frac{S}{\sqrt{n}}\right)$$

2. Seja $X_1,\ldots,X_n$ uma amostra aleatória de $f(x|\theta) = \theta x^{\theta-1}I_{(0,1)}(x)$, $\theta > 0$. Achar uma quantidade pivotal,usando $-lnX$, e usá-la para determinar um estimador por intervalos de θ.

   Solução:
   Vamos encontrar a distribuição de $Y = -lnX$.
   Se $y = -lnx$ então $x = e^{-y}$. Logo,

   $$f_Y(y) = \left|\frac{dx}{dy}\right| f_X(x) = \left|-e^{-y}\right| \theta(e^{-y})^{\theta-1} = \theta e^{-\theta y}$$

   ou seja, $Y \sim Exp(\theta)$. Então $2\theta Y = -2\theta lnX \sim Gama(1,1/2)$ e $T = -2\theta\sum_{i=1}^{n} lnX_i \sim Gama(n,1/2) = \chi^2_{2n}$ é quantidade pivotal.Portanto, seja $\alpha = \alpha_1 + \alpha_2$

   $$P\left(\chi^2_{2n,1-\alpha_1} \leq -2\theta\sum_{i=1}^{n} lnX_i \leq \chi^2_{2n,1-\alpha_2}\right) = 1-\alpha$$

   $$P\left(\frac{\chi^2_{2n,\alpha_1}}{-2\sum_{i=1}^{n} lnX_i} \leq \theta \leq \frac{\chi^2_{2n,1-\alpha_2}}{-2\sum_{i=1}^{n} lnX_i}\right) = 1-\alpha$$

O intervalo de confiança será dado por:

$$IC(\underset{\sim}{x}) = \left(\frac{\chi^2_{2n,\alpha_1}}{-2\sum_{i=1}^{n} lnX_i}, \frac{\chi^2_{2n,1-\alpha_2}}{-2\sum_{i=1}^{n} lnX_i} \right)$$

3. Seja $X_1, \ldots, X_n$ uma amostra aleatória de $f(x|\theta) = (\theta+1)x^\theta I_{(0,1)}(x)$, $\theta > 0$. Construa um intervalo de coeficiente de confiança $1-\alpha$ baseado em $Q(\underset{\sim}{X}, \theta) = -(\theta+1)\sum_{i=1}^{n} lnX_i$.

<u>Solução:</u> Se $y = -(\theta+1)lnx$ então $x = e^{-\frac{y}{\theta+1}}$. Logo,

$$f_Y(y) = \left|\frac{dx}{dy}\right| f_X(x) = \left|-\frac{1}{\theta+1}e^{-\frac{y}{\theta+1}}\right| (\theta+1)(e^{-\frac{y}{\theta+1}})^\theta = e^{-y} \Rightarrow Y \sim Exp(1)$$

Logo,

$-2(\theta+1)\sum_{i=1}^{n} lnX_i \sim Gama(n, 1/2) = \chi^2_{2n}$ é quantidade pivotal.Portanto,

$$P\left(\chi^2_{2n,\alpha_1} \leq -2(\theta+1)\sum_{i=1}^{n} lnX_i \leq \chi^2_{2n,1-\alpha_2}\right) = 1-\alpha$$

$$P\left(\frac{\chi^2_{2n,\alpha_1}}{-2\sum_{i=1}^{n} lnX_i} - 1 \leq \theta \leq \frac{\chi^2_{2n,1-\alpha_2}}{-2\sum_{i=1}^{n} lnX_i} - 1\right) = 1-\alpha$$

O intervalo de confiança será dado por:

$$IC(\underset{\sim}{x}) = \left(\frac{\chi^2_{2n,\alpha_1}}{-2\sum_{i=1}^{n} lnX_i} - 1, \frac{\chi^2_{2n,1-\alpha_2}}{-2\sum_{i=1}^{n} lnX_i} - 1 \right)$$

4. Determinar, usando uma amostra aleatória de tamanho n de $f(x|\theta) = 3x^2\theta^{-3}I_{(0,\theta)}(x)$, um intervalo de coeficiente de confiança de $1-\alpha$ para θ. Considere caudas iguais.

<u>Solução:</u>

Defina $Y = max(X_1, \ldots, X_n)$. Como $F_X(x|\theta) = \int_0^x 3t^2\theta^{-3}dt = \frac{x^3}{\theta^3}$.

Então $F_Y(y) = [F_X(y)]^n = \left(\frac{y^3}{\theta^3}\right)^n = \frac{y^{3n}}{\theta^{3n}}$ é função decrescente em θ. Sendo assim, considerando caudas iguais, segue-se que o Intervalo de confiança será dado por: $IC(\underset{\sim}{x}) = (\theta_l(y), \theta_u(y))$, onde:

- $\frac{y^{3n}}{\theta_u(y)^{3n}} = \frac{\alpha}{2} \Rightarrow \theta_u(y) = \frac{y}{(\frac{\alpha}{2})^{\frac{1}{3n}}}$

- $\frac{y^{3n}}{\theta_l(y)^{3n}} = 1 - \frac{\alpha}{2} \Rightarrow \theta_u(y) = \frac{y}{\left(1-\frac{\alpha}{2}\right)^{\frac{1}{3n}}}$

5. Para uma população com f.d.p. $f(x|\theta) = \frac{4}{\theta^2} x I_{(0,\theta/2)}(x) - \frac{4}{\theta^2}(x-\theta) I_{(\theta/2,\theta)}(x)$, $\theta > 0$ e uma única observação:

   (a)Achar T, o EMV, para θ.

   (b)Construir um intervalo aleatório de coeficiente $1-\alpha$ de confiança para θ da forma $\left(\frac{T}{2}, \lambda T\right), \lambda > 1$.

   Solução:

   (a) Observe que $L(\theta) = \frac{4}{\theta^2} x I_{(2x<\theta)}(\theta) - \frac{4}{\theta^2}(x-\theta) I_{(0<\theta<2x)}(\theta)$, pois para o primeiro intervalo em que a função está definida, $0 < x < \frac{\theta}{2} \Rightarrow \theta > 2x$. Por outro lado, $\frac{\theta}{2} < x < \theta \Rightarrow x < \theta < 2x$. Ao juntar esses dois intervalor para θ chegamos à conclusão de que $T = \hat{\theta} = 2X$ é EMV para θ, uma vez que a primeira expressão de $L(\theta)$ é decrescente em θ e a segunda é crescente em θ, tendo seu valor máximo em 2x.

   (b)
$$P\left(\frac{T}{2} \le \theta \le \lambda T\right) = 1 - \alpha$$
$$P\left(X \le \theta \le 2X\lambda\right) = 1 - \alpha$$
$$P\left(1 \le \frac{\theta}{X} \le 2\lambda\right) = 1 - \alpha$$
$$P\left(\frac{1}{2\lambda} \le \frac{X}{\theta} \le 1\right) = 1 - \alpha$$
$$P\left(\frac{\theta}{2\lambda} \le X \le \theta\right) = \int_{\frac{\theta}{2\lambda}}^{\frac{\theta}{2}} \frac{4}{\theta^2} x dx - \int_{\frac{\theta}{2}}^{\theta} \frac{4}{\theta^2}(x-\theta) dx = 1 - \frac{1}{2\lambda^2}$$

   Logo, $1 - \alpha = 1 - \frac{1}{2\lambda^2} \Rightarrow \lambda = \frac{1}{\sqrt{2\alpha}}$. O IC será expresso por:
$$IC(\underset{\sim}{x}) = \left[X, \frac{2X}{\sqrt{\alpha}}\right]$$

6. Seja X uma única observação de $f(x|\theta) = \theta x^{\theta-1} I_{(0,1)}(x)$, $\theta > 0$. Se $Y = -\frac{1}{lnX}$, mostrar que $\left(\frac{Y}{2}, Y\right)$ é um intervalo de confiança para θ. Achar seu coeficiente de confiança.

   Solução: Se $y = -\frac{1}{lnx}$ então $x = e^{-\frac{1}{y}}$. Logo,
$$f_Y(y) = \left|\frac{dx}{dy}\right| f_X(x) = \left|\frac{1}{y^2} e^{-\frac{1}{y}}\right| \theta (e^{-\frac{1}{y}})^{\theta-1} = \frac{\theta}{y^2} e^{-\frac{\theta}{y}}$$
   .

   Logo,
$$1 - \alpha = P\left(\frac{Y}{2} < \theta < Y\right) = P\left(\theta \le Y \le 2\theta\right) = \int_{\theta}^{2\theta} \frac{\theta}{y^2} e^{-\frac{\theta}{y}} dy = 0,2386$$
   .

   Portanto, $\alpha = 1 - 0,2386 = 0,7614$. Assim, $\left(\frac{Y}{2}, Y\right)$ é intervalo de coeficiente de confiança $\alpha = 0,7614$.

7. Seja X uma única observação de $f(x|\theta) = \theta e^{-x\theta} I_{(0,\infty)}(x)$, onde $\theta > 0$.

(a) $(X, 2X)$ é um intervalo de confiança para $1/\theta$? Calcular seu coeficiente de confiança.

(b)Achar outro intervalo para $1/\theta$ que tenha o mesmo coeficiente de confiança do item anterior.

Solução:

(a)

$$P\left(X < \frac{1}{\theta} < 2X\right) = P\left(1 < \frac{1}{\theta X} < 2\right) = P\left(\theta < \frac{1}{X} < 2\theta\right) = P\left(\frac{1}{2\theta} < X < \frac{1}{\theta}\right)$$

$$= \int_{\frac{1}{2\theta}}^{\frac{1}{\theta}} \theta e^{-x\theta} dx = e^{-\frac{1}{2}} - e^{-1} = 0,2386$$

. Portanto, $\alpha = 1 - 0,2386 = 0,7613$

(b)Seja $\alpha = \alpha_1 + \alpha_2$. Se $y = 2\theta X$ então $X = e^{-\frac{Y}{2\theta}}$. Logo,

$$f_Y(y) = \left|\frac{dx}{dy}\right| f_X(x) = \left|\frac{1}{2\theta}\right| \theta(e^{-\theta \frac{y}{2\theta}}) = \frac{1}{2} e^{-\frac{y}{2}} \Rightarrow Y \sim Exp(1/2) = \chi^2_2$$

.

$$P\left(\chi^2_{2,\alpha_1} \leq 2\theta X \leq \chi^2_{2,1-\alpha_2}\right) = 1 - \alpha$$

$$P\left(\frac{\chi^2_{2,\alpha_1}}{2X} \leq \theta \leq \frac{\chi^2_{2,1-\alpha_2}}{2X}\right) = 1 - \alpha$$

$$P\left(\frac{2X}{\chi^2_{2,1-\alpha_2}} \leq \frac{1}{\theta} \leq \frac{2X}{\chi^2_{2,\alpha_1}}\right) = 1 - \alpha$$

O intervalo de confiança será dado por:

$$IC(\underset{\sim}{x}) = \left(\frac{2X}{\chi^2_{2,1-\alpha_2}}, \frac{2X}{\chi^2_{2,\alpha_1}}\right)$$

8. Seja $X_1, \ldots, X_n$ uma amostra aleatória de $f(x|\theta) = \frac{1}{\theta} x^{(1-\theta)/\theta} I_{[0,1]}(x)$ com $\theta > 0$.

(a)Achar um estimador por intervalo de coeficiente de confiança $1 - \alpha$.

(b)Achar a amplitude esperada do intervalo.

Solução:

(a)Seja $\alpha = \alpha_1 + \alpha_2$, sabendo que $Y = -\frac{2}{\theta} lnX \sim Exp(1/2) = \chi^2_2$ então $-\frac{2}{\theta}\sum_{i=1}^{n} lnX_i \sim Gama(n, 1/2) = \chi^2_{2n}$. Logo,

$$P\left(\chi^2_{2n,\alpha_1} \leq -\frac{2}{\theta}\sum_{i=1}^{n} lnX_i \leq \chi^2_{2n,1-\alpha_2}\right) = 1 - \alpha$$

$$P\left(\frac{\chi^2_{2n,1-\alpha_2}}{-2\sum_{i=1}^{n} lnX_i} \leq \frac{1}{\theta} \leq \frac{\chi^2_{2n,\alpha_1}}{-2\sum_{i=1}^{n} lnX_i}\right) = 1-\alpha$$

$$P\left(\frac{-2\sum_{i=1}^{n} lnX_i}{\chi^2_{2n,\alpha_1}} \leq \theta \leq \frac{-2\sum_{i=1}^{n} lnX_i}{\chi^2_{2n,1-\alpha_2}}\right) = 1-\alpha$$

O intervalo de confiança será dado por:

$$IC(\underset{\sim}{x}) = \left(\frac{-2\sum_{i=1}^{n} lnX_i}{\chi^2_{2n,\alpha_1}}, \frac{-2\sum_{i=1}^{n} lnX_i}{\chi^2_{2n,1-\alpha_2}}\right)$$

(b) A amplitude esperada é:

$$E[L] = E\left[\frac{2\sum_{i=1}^{n} lnX_i}{\chi^2_{2n,\alpha_1}} - \frac{2\sum_{i=1}^{n} lnX_i}{\chi^2_{2n,1-\alpha_2}}\right] = \left[-\frac{2}{\chi^2_{2n,\alpha_1}} + \frac{2}{\chi^2_{2n,1-\alpha_2}}\right] E\left[-\sum_{i=1}^{n} lnX_i\right]$$

Como $-\sum_{i=1}^{n} lnX_i \sim Gama(n, 1/\theta)$, então, $E\left[-\sum_{i=1}^{n} lnX_i\right] = n\theta$.

Portanto, $E[L] = \left[-\frac{2}{\chi^2_{2n,\alpha_1}} + \frac{2}{\chi^2_{2n,1-\alpha_2}}\right] n\theta$.

9. Seja $X_1, \ldots, X_n$ uma amostra aleatória de $f(x|\theta) = \theta e^{-\theta x} I_{(0,\infty)}(x)$, onde $\theta > 0$.

   (a)Achar um IC de coeficiente $1-\alpha$ para a média da população.

   (b)Achar uma quantidade pivotal baseada em $Y = min(X_1, \ldots, X_n)$ e usá-la para achar um estimador por intervalo de θ.

   Solução:

   (a) Como $X \sim Exp(\theta)$, então $E[X] = \frac{1}{\theta}$.

   Seja $Y = 2\theta X \Rightarrow X = \frac{Y}{2\theta}$.

   $f_Y(y) = \frac{1}{2\theta}\theta exp\left\{-\theta\frac{y}{2\theta}\right\} = \frac{1}{2}e^{-\frac{y}{2}}$.

   $Y \sim Exp(1/2) \Rightarrow 2\theta\sum_{i=1}^{n} X_i \sim Gama(n, 1/2) = \chi^2_{2n}$.

   Seja $\alpha_1 + \alpha_2 = \alpha$.

$$P\left(\chi^2_{2n,\alpha_1} \leq -2\theta\sum_{i=1}^{n} X_i \leq \chi^2_{2n,1-\alpha_2}\right) = 1-\alpha$$

$$P\left(\frac{2\sum_{i=1}^{n} X_i}{\chi^2_{2n,1-\alpha_2}} \leq \frac{1}{\theta} \leq \frac{2\sum_{i=1}^{n} X_i}{\chi^2_{2n,\alpha_1}}\right) = 1-\alpha$$

Portanto, o IC será dado por:

$$IC(\underset{\sim}{x}) = \left(\frac{2\sum_{i=1}^{n} X_i}{\chi^2_{2n,1-\alpha_2}}, \frac{2\sum_{i=1}^{n} X_i}{\chi^2_{2n,\alpha_1}}\right)$$

(b) Defina $Y = min(X_1, \ldots, X_n)$. Dessa forma,

$$f_Y(y) = \theta e^{-\theta y} n \left[1-(1-e^{-\theta y})\right]^n = n\theta e^{-n\theta y} \Rightarrow Y \sim Exp(n\theta)$$

Seja $T = 2n\theta Y \Rightarrow Y = \frac{T}{2\theta n}$, e sua densidade é dada por:

$$f_T(t) = \frac{1}{2\theta n} n\theta e^{n\theta \frac{t}{2\theta n}} = \frac{1}{2} e^{-\frac{t}{2}} \Rightarrow T \sim Exp(1/2) = \chi^2_2.$$

Para $\alpha_1 + \alpha_2 = \alpha$,

$$P\left(\chi^2_{2,\alpha_1} \leq 2\theta n Y \leq \chi^2_{2,\alpha_2}\right) = 1-\alpha$$

$$P\left(\frac{\chi^2_{2,\alpha_1}}{2nY} \leq \theta \leq \frac{\chi^2_{2,\alpha_2}}{2nY}\right) = 1-\alpha$$

Logo, o IC é dado por:

$$IC((\underset{\sim}{x})) = \left(\frac{\chi^2_{2,\alpha_1}}{2nY}, \frac{\chi^2_{2,\alpha_2}}{2nY}\right)$$

10. Seja $X_1, \ldots, X_n$ uma amostra aleatória de $f(x|\theta) = \theta e^{-\theta x} I_{(0,\infty)}(x)$, onde $\theta > 0$ e n suficientemente grande. Achar um IC assintótico de coeficiente $1-\alpha$. Usar caudas iguais.

<u>Solução:</u>
Temos que $X \sim Exp(\theta)$, e sabemos que o EMV para θ é $\frac{1}{\bar{X}}$, como n é suficientemente grande, $\hat{\theta} \sim N(\theta, \frac{\theta^2}{n})$. Logo,

$$\frac{\frac{1}{\bar{X}} - \theta}{\sqrt{\frac{\theta^2}{n}}} \sim N(0,1)$$

Dessa forma,

$$IC((\underset{\sim}{x})) = \left(\frac{\frac{1}{\bar{X}}}{1+\frac{z_{\alpha/2}}{\sqrt{n}}}, \frac{\frac{1}{\bar{X}}}{1-\frac{z_{\alpha/2}}{\sqrt{n}}}\right)$$

ANEXO A - Condições de regularidade

Seja $X_1, \ldots, X_n$ uma a.a. de $f(x_1, \ldots, x_n|\theta)$. Seja $T = T(X_1, \ldots, X_n)$ um estimador não viciado de $\tau(\theta)$. Faremos as seguintes suposições chamadas condições de regularidade.

1. $\frac{\partial}{\partial\theta} ln f(x|\theta) \ \exists \ \forall x$ e $\forall\theta$;
2. $\frac{\partial}{\partial\theta} \int \ldots \int \prod_{i=1}^{n} f(x_i|\theta) d_{x_1} \ldots d_{x_n}$;
3. $\frac{\partial}{\partial\theta} \int \ldots \int T(x_1, \ldots, x_n) \prod_{i=1}^{n} f(x_i|\theta) d_{x_1} \ldots d_{x_n} = \int \ldots \int T(x_1, \ldots, x_n) \frac{\partial}{\partial\theta} \prod_{i=1}^{n} f(x_i|\theta) d_{x_1} \ldots d_{x_n}$;
4. $0 < \mathbb{E}\left[\left(\frac{\partial}{\partial\theta} ln f(X_1, \ldots, X_n)\right)^2\right] < \infty$, $\forall\theta$.