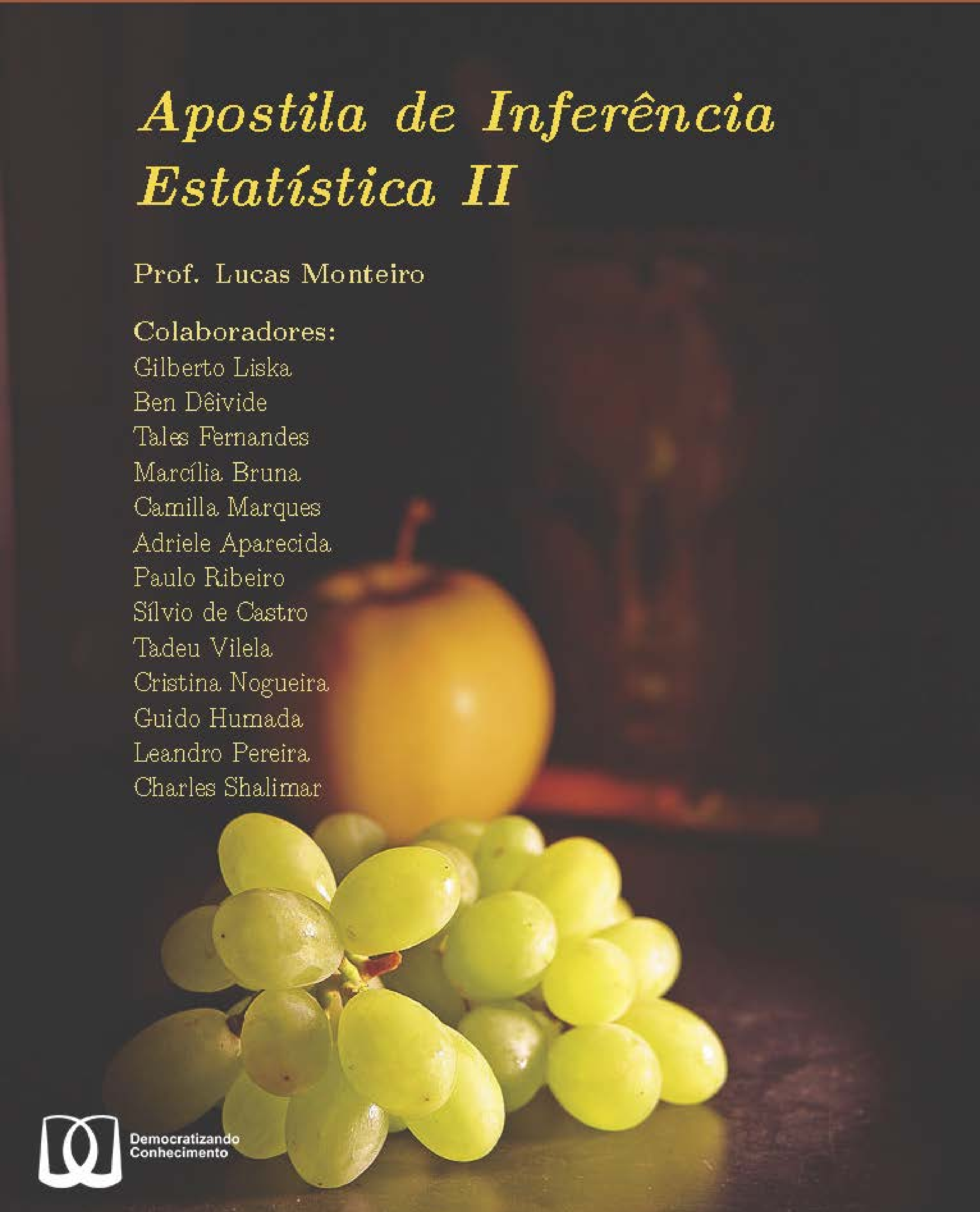


Apostila de Inferência Estatística II

Prof. Lucas Monteiro

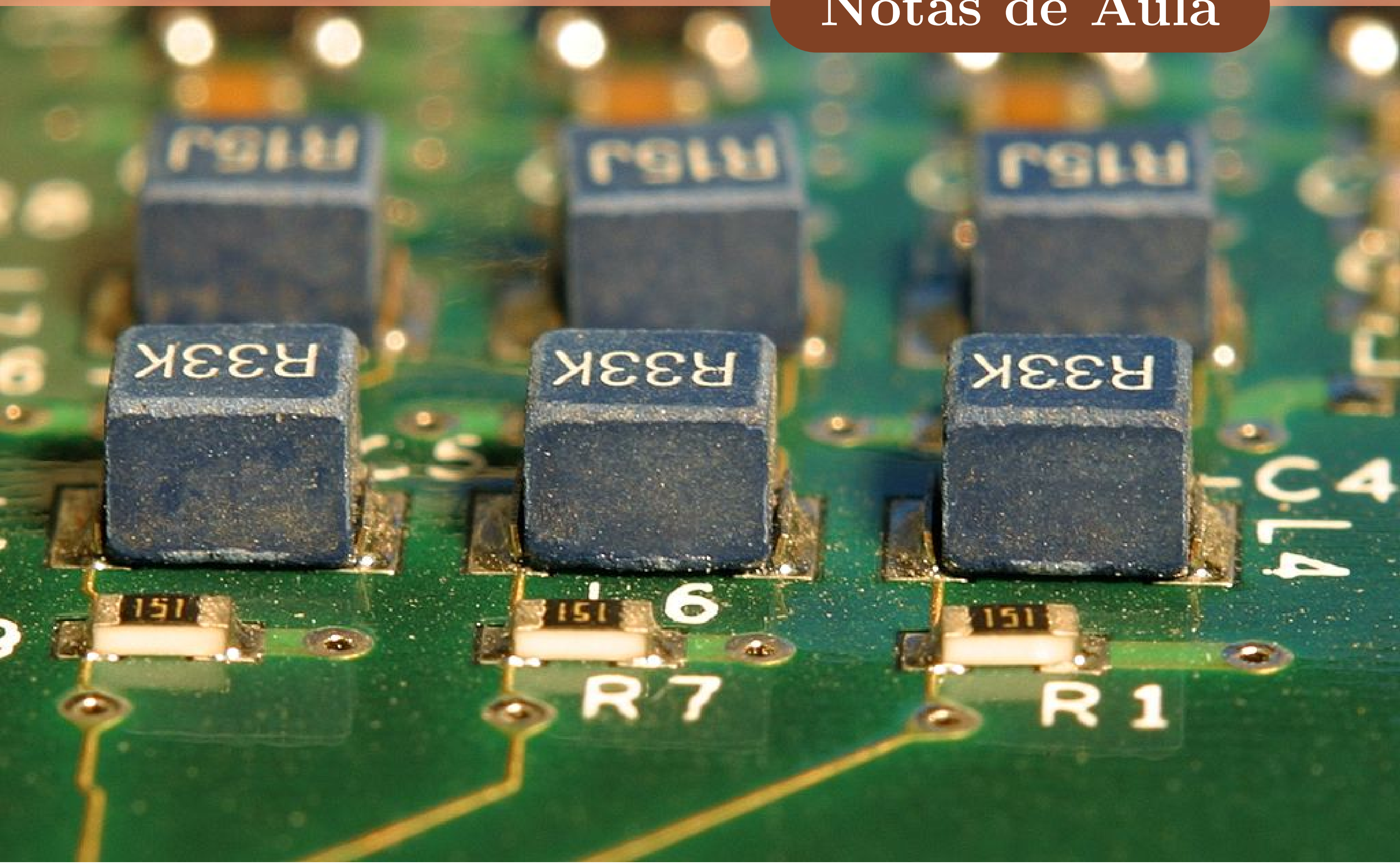
Colaboradores:
Gilberto Liska
Ben Dêivide
Tales Fernandes
Marcília Bruna
Camilla Marques
Adriele Aparecida
Paulo Ribeiro
Sílvio de Castro
Tadeu Vilela
Cristina Nogueira
Guido Humada
Leandro Pereira
Charles Shalimar

Democratizando Conhecimento

Sumário

Capítulo 1

Notas de Aula



1.1 Introdução

Até o momento, todos os critérios que consideramos na Inferência Clássica foram baseados em amostras finitas. Por outro lado, poderíamos considerar propriedades assintóticas, que descrevem o comportamento de um procedimento conforme o tamanho da amostra tende para o infinito.

O poder das avaliações assintóticas é que, quando o tamanho de uma amostra vai para o infinito, os cálculos são simplificados. Avaliações que foram impossíveis no caso de amostras finitas se tornam rotineiras. Essa simplificação também nos permite examinar outras técnicas (por exemplo *bootstrap*) que tipicamente podem ser avaliadas assintoticamente.

1.2 Estimação Pontual

Consistência

A propriedade de consistência parece ser fundamental, requerendo que o estimador convirja para o "correto" valor conforme a amostra vai para o infinito. Essa é uma propriedade fundamental, uma vez que o valor de um estimador inconsistente deve ser questionado (ou no mínimo vigorosamente investigado).

Consistência (bem como todas as outras propriedades assintóticas) consiste de uma sequência de estimadores ao invés de um único estimador, embora seja comum dizer de um estimador "consistente". Se observamos $X_1, X_2, ...$ de acordo com uma distribuição $f\left(x\,|\theta\right)$, podemos construir uma sequência de estimadores $W_n = W_n\left(X_1, ..., X_n\right)$ meramente por aplicar um mesmo procedimento em cada amostra de tamanho n. Por exemplo, $\bar{X}_1 = X_1$, $\bar{X}_2 = (X_1 + X_2)/2$, $\bar{X}_3 = (X_1 + X_2 + X_3)/3$, etc. Podemos agora definir uma sequência consistente.

Definição 1.2.1. *Uma sequência de estimadores $W_n = W_n(X_1, ..., X_n)$ é uma sequência consistente de estimadores do parâmetro θ se, para todo $\varepsilon > 0$ e todo $\theta \in \Theta$,*

$$\lim_{n\to\infty} P_\theta \left[|W_n - \theta| < \varepsilon\right] = 1 \tag{1.1}$$

Informalmente, a definição 1.2.1 diz que conforme o tamanho da amostra vai para o infinito (e a informação amostral se torna cada vez melhor), o estimador estará arbitrariamente perto do parâmetro com alta probabilidade, uma propriedade eminentemente desejável. Ou, olhando por outro lado, podemos dizer que a probabilidade de que sequência consistente de estimadores perde o verdadeiro parâmetro é pequena. Uma equivalente forma é essa: Para todo $\varepsilon > 0$ e todo $\theta \in \Theta$, uma sequência consistente W_n satisfará

$$\lim_{n\to\infty} P_\theta \left[|W_n - \theta| \geq \varepsilon\right] = 0 \tag{1.2}$$

A definição 1.2.1 deve ser comparada à definção de *convergência em probabilidade.* A definição 1.2.1 diz que uma sequência consistente de estimadores converge em probabilidade para o parâmeotro *theta* que é estimado. Contudo, a definição de convergência em probabilidade é de acordo com uma sequência de variáveis aleatórias com uma estrutura de probabilidade, a definição 1.2.1 está de acordo com uma família de estruturas de probabilidade, indexadas por *theta*. Para cada diferente valor de θ, a estrutura de probabilidade associada com a sequência W_n é diferente. E a definição diz que para cada valor de *theta*, a estrutura probabilística é tal que a sequência converge em probabilidade para o verdadeiro *theta*. Essa é a definição usual entre a definição probabilística e a definição estatística. A definição probabilística se trata de uma estrutura probabilística, mas a definição estatística está relacionada com toda família.

Definição 1.2.2. *Uma sequência de variáveis aleatórias $X_1, X_2, ...,$ converge em probabilidade para uma variável aleatória X se, para todo $\varepsilon > 0$,*

$$\lim_{n\to\infty} P\left[|X_n - X| \geq \varepsilon\right] = 0 \tag{1.3}$$

ou, equivalentemente,

$$\lim_{n\to\infty} P\left[|X_n - X| < \varepsilon\right] = 1 \tag{1.4}$$

As $X_1, X_2, ...$ na definição 1.2.2 são tipicamente variáveis aleatórias independentes e identicamente distribuídas, como em uma amostra aleatória. A distribuição de X_n muda conforme o subscrito muda, e os conceitos de convergência descrevem modos diferentes pelos quais a distribuição de X_n converge para alguma distribuição limitada conforme o subscrito se torna alto.

Exemplo 1.2.0.1. ***(Consistência de $\bar{X}$)*** *seja $X_1, X_2, ...$ uma amostra iid $N(\theta,1)$, e considere a sequência $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Mostre que $\bar{X}$ é consistente.*

SOLUÇÃO

Dado $\varepsilon > 0$. Sabe-se que $\bar{X}_n \sim N\left(\theta, \frac{1}{n}\right)$. Daí,

$$\begin{aligned}
P_\theta\left[|X_n - \theta| < \varepsilon\right] =& P_\theta\left[-\varepsilon < X_n - \theta < \varepsilon\right] \\
=& P_\theta\left[-\varepsilon + \theta < X_n < \varepsilon + \theta\right] \\
=& P_\theta\left[\theta - \varepsilon < X_n < \theta + \varepsilon\right] \\
=& \int_{\theta-\varepsilon}^{\theta+\varepsilon} \frac{1}{\sqrt{2\pi}\left(\sqrt{1/n}\right)} e^{-\frac{1}{2}\left(\frac{\bar{X}_n - \theta}{\sqrt{\frac{1}{n}}}\right)^2} d\bar{x}_n \\
=& \int_{-\varepsilon}^{\varepsilon} \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{n}{2}(y)^2} dy \\
=& \int_{-\varepsilon}^{\varepsilon} \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\sqrt{n}y\right)^2} dy
\end{aligned}$$

$$
\begin{aligned}
&= \int_{-\sqrt{n}\varepsilon}^{\sqrt{n}\varepsilon} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2} dy \\
&= P\left[-\sqrt{n}\varepsilon < Z < \sqrt{n}\varepsilon\right] \\
&\Rightarrow \lim_{n\to\infty} P\left[-\sqrt{n}\varepsilon < Z < \sqrt{n}\varepsilon\right] = 1
\end{aligned}
$$

e, portanto, $\bar{X}_n$ é uma sequência consistente de estimadores de θ.

Em geral, um cálculo detalhado, como visto no exemplo anterior, não é necessário para verificar consistência. Em algumas situações, podemos usar a Desigualdade de Chebychev, que é dada por

$$\lim_{n\to\infty} P_\theta\left[|W_n - \theta| \geq \varepsilon\right] \leq \frac{E_\theta\left[(W_n-\theta)^2\right]}{\varepsilon^2} \tag{1.5}$$

então se, para todo $\theta \in \Theta$,

$$\lim_{n\to\infty} E_\theta\left[(W_n-\theta)^2\right] = 0$$

a sequência de estimadores é consistente. Além disso,

$$E_\theta\left[(W_n-\theta)^2\right] = Var_\theta\left[W_n\right] + \left(Vies_\theta\left[W_n\right]\right)^2 \tag{1.6}$$

Definição 1.2.3. ***(Desigualdade de Chebychev)*** *Seja X uma variável aleatória e $g()$ uma função não negativa com domínio na reta real. Então*

$$P\left[g(X) \geq k\right] \leq \frac{E\left[g(X)\right]}{k^2}$$

para todo $k > 0$. Ainda, se X é uma variável aleatória com variância finita, tem-se que

$$P\left[|X-\mu| \geq r\sigma\right] = P\left[(X-\mu)^2 \geq r^2\sigma^2\right] \leq \frac{1}{r^2}, \qquad \forall r > 0 \tag{1.7}$$

alternativamente,

$$P\left[|X-\mu| < r\sigma\right] \geq 1 - \frac{1}{r^2} \tag{1.8}$$

Demonstração. As expressões 1.7 e 1.8 são obtidas considerando-se a função $\gamma(x) = (x-\mu)^2$ da seguinte forma:

$$
\begin{aligned}
P\left[g(X) > k\right] \leq & \frac{E\left[g(X)\right]}{k} \\
&\Rightarrow P\left[(X-\mu)^2 > k\right] \leq \frac{E\left[(X-\mu)^2\right]}{k} \\
&\Rightarrow P\left[(X-\mu)^2 > k\right] \leq \frac{Var\left[X\right]}{k}; \quad k = r^2 Var\left[X\right] \\
&\Rightarrow P\left[(X-\mu)^2 > r^2\sigma^2\right] \leq \frac{1}{r^2} \\
&\Rightarrow P\left[|X-\mu| > r\sigma\right] \leq \frac{1}{r^2} \\
&\Rightarrow P\left[|X-\mu| \leq r\sigma\right] \geq 1 - \frac{1}{r^2}
\end{aligned}
$$

□

Exemplo 1.2.0.2. *Seja $X \sim \exp(\lambda)$ e $g(x) = |x|$. Mostre que a desigualdade de Chebychev vale.*

SOLUÇÃO

Pela Desigualdade de Chebychev temos que

$$\begin{aligned}
P[X > k] \leq & \frac{E[X]}{k} \\
& \Rightarrow \int_k^\infty \lambda e^{-\lambda x} dx \leq \frac{\frac{1}{\lambda}}{k} \\
& \Rightarrow \int_k^\infty \lambda e^{-\lambda x} dx \leq \frac{1}{k\lambda} \\
& \Rightarrow -e^{-\lambda x}\Big|_k^\infty \leq \frac{1}{k\lambda} \\
& \Rightarrow e^{-\lambda k} \leq \frac{1}{k\lambda} \\
& \Rightarrow \log\left(e^{-\lambda k}\right) \leq \log\left(\frac{1}{k\lambda}\right) \\
& \Rightarrow -\lambda k \leq -\log(k\lambda) \\
& \Rightarrow \lambda k > \log(k\lambda)
\end{aligned}$$

Portanto, a desigualdade vale.

1.3 Estatística de Ordem

Valores amostrais como a menor, maior ou mediana observação de uma amostra aleatória podem promover informações adicionais de síntese. Por exemplo, a maior vazão de um rio ou a menor temperatura de inverno registrada durante os últimos 50 anos poderia ser um dado útil no planejamento de emergências futuras. O preço mediano de casas vendidas durante um mês anterior poderia ser útil para estimar o custo de moradia. Esses são exemplos de estatísticas de ordem.

Definição 1.3.1. *A estatística de ordem de uma amostra aleatória $X_1, \ldots, X_n$ são os valores amostrais organizados em ordem ascendente. Eles são denotados por $X_{(1)}, \ldots, X_{(n)}$.*

As estatísticas de ordem são variáveis aleatórias que satisfazem $X_{(1)} \leq \ldots \leq X_{(n)}$. Em particular

$$\begin{aligned}
X_{(1)} &= \min_{1 \leq i \leq n} \{X_i\} \\
X_{(2)} &= \text{ segundo menor } \{X_i\} \\
&\vdots \\
X_{(n)} &= \max_{1 \leq i \leq n} \{X_i\}.
\end{aligned}$$

Uma vez que elas são variáveis aleatórias, podemos discutir as probabilidades que elas assumem sobre vários valores. Para calcular essas probabilidades precisamos das fdp's ou fd's das estatísticas de ordem.

Se $X_1, \ldots, X_n$ são variáveis aleatórias iid discretas, então o cálculo das probabilidades para as estatísticas de ordem é essencialmente um exercício de contagem. Se $X_1, \ldots, X_n$ são variáveis aleatórias de uma população contínua, então expressões convenientes para a fdp de uma ou mais estatísticas de ordem podem ser obtidas. Essas podem ser usadas para obter as funções de distribuição das estatísticas de ordem.

Teorema 1.3.1. *Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma distribuição discreta com fdp $f_X(x_i) = p_i$, onde $x_1 < x_2 < \cdots$ são os possíveis valores de X em ordem ascendente. Defina*

$$P_0 = 0$$

$$\begin{aligned}
P_1 &= p_1 \\
P_2 &= p_1 + p_2 \\
&\vdots \\
P_i &= p_1 + p_2 + \ldots + p_i \\
&\vdots
\end{aligned}$$

Seja as estatísticas de ordem de uma amostra denotadas por $X_{(1)}, \ldots, X_{(n)}$. Então

$$P\left[X_{(j)} \leq x_i\right] = \sum_{k=j}^{n} \binom{n}{k} P_i^k (1 - P_i)^{n-k} \tag{1.9}$$

e

$$P\left[X_{(j)} = x_i\right] = \sum_{k=j}^{n} \binom{n}{k} \left[P_i^k (1 - P_i)^{n-k} - P_{i-1}^k (1 - P_{i-1})^{n-k}\right] \tag{1.10}$$

Se $X_1, \ldots, X_n$ é uma amostra aleatória de uma população contínua, então a situação é simplificada ligeiramente pelo fato de a probabilidade ser zero para quaisquer duas variáveis $X_j's$ iguais, livrando-nos das preocupação dos empates. Então $P\left[X_{(1)} < \ldots < X_{(n)}\right] = 1$ e o espaço amostral para $X_{(1)} < \ldots < X_{(n)}$ é $\{(x_1, \ldots, x_n) : x_1 < x_2 < \ldots < x_n\}$.

Teorema 1.3.2. *Seja $X_{(1)} < \ldots < X_{(n)}$ as estatísticas de ordem de uma amostra aleatória, $X_1, \ldots, X_n$, de uma população contínua com fd $F_X(x)$ e fdp $f_X(x)$. Então a fdp de $X_{(j)}$ é*

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!\,(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j} \tag{1.11}$$

Com as expressões 1.10 e 1.11 podemos obter as expressões para as estatísticas de ordem do máximo e mínimo de uma amostra aleatória. Assim,

$$\begin{aligned}
F_{X_{(n)}}(x) &= P\left[\max_{1 \leq i \leq n} \{X_i\}\right] = \left[F_X(x)\right]^n \\
F_{X_{(1)}}(x) &= P\left[\min_{1 \leq i \leq n} \{X_i\}\right] = \sum_{k=1}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k} \\
&= \underbrace{\sum_{k=0}^{n} \binom{n}{k} \left[F_X(x)\right]^k \left[1 - F_X(x)\right]^{n-k}}_{1} - \underbrace{\binom{n}{0}}_{1} \underbrace{\left[F_X(x)\right]^0}_{1} \left[1 - F_X(x)\right]^n \\
&= 1 - \left[1 - F_X(x)\right]^n
\end{aligned}$$

Exemplo 1.3.0.3. ***(Fdp da Estatística de Ordem Uniforme)*** *Seja $X_1, \ldots, X_n$ uma amostra iid de uma Uniforme(0,1), então $f_X(x) = 1$ para $x \in (0,1)$ e $F_X(x) = x$ para $x \in (0,1)$. Obtenha a fdp para a j-ésima estatística de ordem.*

SOLUÇÃO

Usando a expressão 1.11, temos que

$$\begin{aligned}
f_{X_{(j)}}(x) &= \frac{n!}{(j-1)!\,(n-j)!} f_X(x) \left[F_X(x)\right]^{j-1} \left[1 - F_X(x)\right]^{n-j} \\
&= \frac{n!}{(j-1)!\,(n-j)!} \left[x\right]^{j-1} \left[1 - x\right]^{n-j} \\
&= \frac{\Gamma(n+1)}{\Gamma(j)\,\Gamma(n-j+1)} x^{j-1} \left[1 - x\right]^{(n-j+1)-1}.
\end{aligned}$$

Então, a j-ésima estatística de ordem de uma amostra da Uniforme$(0{,}1)$ *tem uma distribuição Beta*$(j, n - j + 1)$*. Disso, podemos deduzir que*

$$
\begin{aligned}
E\left[X_{(j)}\right] &= \frac{j}{n+1} \\
Var\left[X_{(j)}\right] &= \frac{j\,(n-j+1)}{(n+1)^2\,(n+2)}
\end{aligned}
$$

1.4 Conceitos de Convergência

Essa seção tratará de alguns conceitos e idéias de quando o tamanho de uma amostra se aproxima do infinito e investigar o comportamento de certas quantidades amostrais e seus acontecimentos. Embora a noção de uma amostra infinita seja um artifício teórico, às vezes isso pode propiciar algumas úteis aproximações para o caso de amostra finita, uma vez que algumas expressões se tornam simplicadas quando tratadas no limite.

Vamos olhar o comportamento de $\bar{X}_n$, a média de n observações, conforme $n \to \infty$.

Convergência em Probabilidade

É um tipo de convergência considerada mais fraca que as demais e é, portanto, fácil de ser verificada.

A definição de convergência em probabilidade é dada na definição 1.2.2.

Frequentemente, estatísticos se deparam com situações o limite de uma variável aleatória é uma constante e a variável aleatória na sequência são médias amostrais. O mais famoso resultado desse tipo é o seguinte.

Teorema 1.4.1. ***(Lei Fraca dos Grandes Números - LFGN)*** *Seja* $X_1, X_2, \ldots$ *variáveis aleatórias iid com* $E[X_i] = \mu$ *e* $Var[X_i] = \sigma^2 < \infty$*. Defina* $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$*. Então, para todo* $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left[\left|\bar{X}_n - \mu\right| < \varepsilon\right] = 1 \tag{1.12}$$

ou seja, $\bar{X}_n$ *converge em probabilidade para* μ.

Demonstração. Para provar a LFGN, recorreremos à Desigualdade de Chebychev. Temos que, para todo $\varepsilon > 0$,

$$
\begin{aligned}
P[|X-\mu| \geq \varepsilon] = & P\left[(X-\mu)^2 \geq \varepsilon^2\right] \leq \frac{E\left[(X-\mu)^2\right]}{\varepsilon^2} \\
\Rightarrow & P\left[(X-\mu)^2 \geq \varepsilon^2\right] \leq \frac{E\left[(X-\mu)^2\right]}{\varepsilon^2} = \frac{Var\left[\bar{X}_n\right]}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2 n} \\
\Rightarrow & 1 - P[|X-\mu| < \varepsilon] \leq \frac{\sigma^2}{\varepsilon^2 n} \\
\Rightarrow & P[|X-\mu| < \varepsilon] \geq 1 - \frac{\sigma^2}{\varepsilon^2 n} \\
\Rightarrow & \lim_{n\to\infty} 1 - \frac{\sigma^2}{\varepsilon^2 n} = 1
\end{aligned}
$$

□

A Lei Fraca dos Grandes Números elegantemente afirma que, sob condições gerais, a média amostral se aproxima da média populacional conforme $n \to \infty$. De fato, existem versões mais gerais da LFGN, onde assumimos apenas que a média é finita. Contudo, a versão estabelecida no teorema 1.4.1 é aplicada em muitas situações práticas.

A propriedade sumarizada pela LFGN, que a sequência de mesmas quantidades amostrais se aproxima de uma constante conforme $n \to \infty$, é conhecida como *consistência*.

Exemplo 1.4.0.4. ***(Consistência de S^2)*** *Suponha que temos uma sequência $X_1, X_2, ...$ de variáveis aleatórias iid com $E[X_i] = \mu$ e $Var[X_i] = \sigma^2 < \infty$. Mostre que S_n^2 é um estimador consistente pela LFGN.*

SOLUÇÃO

Usando a desigualdade de Chebychev, temos que

$$\begin{aligned}
&P\left[\left|S_n^2 - \sigma^2\right| \geq \varepsilon\right] \leq \frac{E\left[\left(S_n^2 - \sigma^2\right)^2\right]}{\varepsilon^2} = \frac{Var\left[S_n^2\right]}{\varepsilon^2} \\
\Rightarrow &1 - P\left[\left|S_n^2 - \sigma^2\right| < \varepsilon\right] \leq \frac{Var\left[S_n^2\right]}{\varepsilon^2} \\
\Rightarrow &P\left[|X - \mu| < \varepsilon\right] \geq 1 - \frac{\frac{1}{n}\left(\mu^4 - \frac{n-3}{n-1}\sigma^4\right)}{\varepsilon^2} \\
\Rightarrow &\lim_{n\to\infty} 1 - \left[\frac{1}{n}\left(\mu^4 - \frac{n-3}{n-1}\sigma^4\right)\right]\frac{1}{\varepsilon^2} = 1
\end{aligned}$$

e portanto, uma condição suficiente que S_n^2 convirja em probabilidade para σ^2 é que $Var\left[S_n^2\right] \to 0$ conforme $n \to \infty$.

Uma extensão natural da definição 1.2.2 está relacionada a funções de variáveis aleatórias. Ou seja, se uma sequência $X_1, X_2, ...$ converge em probabilidade para uma variável aleatória X ou para uma constante a, podemos fazer qualquer conclusão sobre a sequência das variáveis aleatórias $h(X_1)$,$h(X_2)$,... para alguma função h comortada razoavelmente? O próximo teorema mostra que isso acontece.

Teorema 1.4.2. *Suponha que $X_1, X_2, ...$ convirja em probabilidade para uma variável aleatória X e h uma função contínua. Então $h(X_1)$,$h(X_2)$,... converje em probabilidade para $h(X)$.*

Exemplo 1.4.0.5. ***(Consistência de S)*** *Se S_n^2 é um estimador consistente de σ^2, então pelo teorema 1.4.2, o desvio padrão amostral $S_n = \sqrt{S_n^2} = h\left(S_n^2\right)$ é um entimador consistente de σ. Note que S_n é, de fato, um estimador viesado de σ, mas o viés desaparece assintoticamente.*

Convergência Quase-certa

Um tipo de convergência que é mais forte que a convergência em probabilidade é a convergência quase certa (algumas vezes sabidamente confundida com convergência com probabilidade 1). Esse tipo de convergência é similar à convergência pontual de uma sequência de funções, exceto que a convergência não necessariamente ocorre em um conjunto com probabilidade 0 (daí o "quase" certo).

Definição 1.4.1. *Uma Sequência de variáveis aleatórias $X_1, X_2, ...$ converge quase certo para uma variável aleatória X se, para todo $\varepsilon > 0$,*

$$P\left[\lim_{n\to\infty} |X_n - X| < \varepsilon\right] = 1 \tag{1.13}$$

Note a similaridade entre as definições 1.2.2 e 1.4.1. Embora elas pareçam similares, existem afirmações muito diferentes, sendo a definição 1.4.1 muito mais forte. Para entender a convergência quase certa, devemos relembrar conceitos a definição básica de variável aleatória. Uma variável aleatória é uma função de valor real definida sobre o espaço amostral S. Se o espaço amostral tem elementos denotados por s, $X_n(s)$ e $X(s)$ são todas funçõs definidas em S. A definição 1.4.1 afirma que X_n converge para X quase certamente se as funções $X_n(s)$ convergem para $X(s)$ para todo $s \in S$ exceto, talvez, para $s \in N$, onde $N \subset S$ e $P(N) = 0$.

Exemplo 1.4.0.6. *Seja o espaço amostral S o intervalo fechado [0,1] com distribuição de probabilidade uniforme. Defina as variáveis aleatórias $X_n(s) = s + s^n$ e $X(s) = s$. Daí*

$$\begin{aligned}\lim_{n\to\infty} |X_n(s) - X(s)| &= \lim_{n\to\infty} |s + s^n - s| \\ &= \lim_{n\to\infty} |s^n| = \begin{cases} 0, & 0 \leq s < 1 \\ 1, & s = 1 \end{cases} \\ &\Rightarrow P\left[\lim_{n\to\infty} |X_n(s) - X(s)| < \varepsilon\right] = P[0 \leq s < 1] = 1\end{aligned}$$

note que $X_n(1) = 2$ para todo n tal que $X_n(1)$ não converge para $1 = X_n(1)$. Mas, uma vez que a convergência ocorre sobre o conjunto [0,1) e $P[[0,1)] = 1$, X_n converge quase certamente para X.

Como poderia se esperar, convergência quase certa, sendo um forte critério, implica convergência em probabilidade. O contrário não é verdade. Contudo, se uma sequência converge em probabilidade, é possível encontrar uma subsequência que tem convergência quase certa.

Novamente, estatísticos são confrontados às vezes com a convergência para uma constante. Vamos afirmar agora, sem provar, uma forte analogia da LFGN, a *Lei Forte dos Grandes Números (LSGN).*

Teorema 1.4.3. ***(Lei Forte dos Grandes Números)*** *Seja $X_1, X_2, \ldots$ variáveis aleatórias iid com $E[X_i] = \mu$ e $Var[X_i] = \sigma^2 < \infty$, e defina $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$. Então, para todo $\varepsilon > 0$,*

$$P\left[\lim_{n\to\infty} \left|\bar{X}_n - \mu\right| < \varepsilon\right] = 1 \tag{1.14}$$

ou seja, $\bar{X}_n$ converge quase certo para μ

Para ambas LFGN e LSGN temos a pressuposição de variância finita. Embora essa pressuposição seja verdadeira (e desejável) em muitas aplicações, ela é, de fato, uma forte pressuposição que é necessária. Ambas as leis funcioam sem essa pressuposição. A única condição necessária no momento é que $E[|X_i|] < \infty$.

Convergência em Distribuição

A seguinte definição mostra quando uma sequência de variáveis aleatórias X_i converge em distribuição para uma v.a. X.

Definição 1.4.2. *Uma sequência de variáveis aleatórias $X_1, X_2, \ldots$, converge em distribuição para uma variável aleatória X se*

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \tag{1.15}$$

em todos os pontos x onde $F_X(x)$ é contínua.

OBS.: A convergência de $F_{X_n}(x)$ em $F_X(x)$ é apenas pontual.

Exemplo 1.4.0.7. ***(Máximo de Unformes)*** *Se $X_1, X_2, \ldots$ são iid uniformes (0,1) e $X_{(n)} = \max_{1\leq i\leq n} X_i$, vamos examinar se (e para onde) $X_{(n)}$ converge em distribuição.*

SOLUÇÃO

Conforme $n \to \infty$, esperamos que $X_{(n)}$ se aproxime de 1 e, como $X_{(n)}$ deve ser necessariamente menor que 1, temos que, para todo $\varepsilon > 0$,

$$\begin{aligned}P\left[\left|X_{(n)} - 1\right| \geq \varepsilon\right] =& P\left[X_{(n)} - 1 \geq \varepsilon\right] + P\left[X_{(n)} - 1 \leq \varepsilon\right] \\ =& \underbrace{P\left[X_{(n)} \geq 1 + \varepsilon\right]}_{=0} + P\left[X_{(n)} \leq 1 - \varepsilon\right] \\ =& 0 + P\left[X_{(n)} \leq 1 - \varepsilon\right]\end{aligned}$$

$$
\begin{aligned}
&= (1-\varepsilon)^n \\
&\Rightarrow \lim_{n\to\infty} (1-\varepsilon)^n = 0, \ \ 0 < \varepsilon < 1
\end{aligned}
$$

Portanto, $X_{(n)}$ converge em probabilidade para 1. Contudo, se tomarmos $\varepsilon = t/n$ temos que

$$
\begin{aligned}
&P\left[X_{(n)} \leq 1 - \frac{t}{n}\right] = \left(1 - \frac{t}{n}\right)^n = e^{-t} \\
\Rightarrow &P\left[X_{(n)} \leq 1 - \frac{t}{n}\right] \to e^{-t} \\
\Rightarrow &P\left[X_{(n)} > 1 - \frac{t}{n}\right] \to 1 - e^{-t} \\
\Rightarrow &P\left[X_{(n)} - 1 > -\frac{t}{n}\right] \to 1 - e^{-t} \\
\Rightarrow &P\left[-X_{(n)} + 1 < \frac{t}{n}\right] \to 1 - e^{-t} \\
\Rightarrow &P\left[n\left(1 - X_{(n)}\right) < t\right] \to 1 - e^{-t} \\
\Rightarrow &n\left(1 - X_{(n)}\right) \sim \exp(1)
\end{aligned}
$$

ou seja, a v.a. $n\left(1 - X_{(n)}\right)$ converge em distribuição para uma v.a. exponencial(1).

Notem que embora foi dito que uma sequência de variáveis aleatórias convergem em distribuição, na verdade suas funções de distribuição de probabilidade que converge, não as variáveis aleatórias. Esse modo de convergência em distribuição é um pouco diferente da convergência em probabilidade ou convergência quase certa. Porém, isso implica em outros tipos de convergência.

Teorema 1.4.4. *Se a sequência de variáveis aleatórias $X_1, X_2, \ldots$, converge em probabilidade para uma v.a. X, a sequência também converge em distribuição para X.*

Uma forma mais abrangente do teorema 1.4.4 acima é visto no seguinte teorema.

Teorema 1.4.5. *A sequência de variáveis aleatórias $X_1, X_2, \ldots$, converge em probabilidade para uma constante μ se, e somente se, a sequência também converge em distribuição para μ. Ou seja, a afirmação*

$$
P\left[\left|X_{(n)} - \mu\right| > \varepsilon\right] \to 0, \ \textit{para todo } \varepsilon > 0 \tag{1.16}
$$

é equivalente a

$$
P\left[X_n \leq x\right] \to \begin{cases} 0, \ se\, x < \mu \\ 1, \ se\, x > \mu \end{cases} \tag{1.17}
$$

A média amostral é uma estatística cujo comportamento de amostras de tamanho grande é bastante importante. Em particular, queremos investigar sua distribuição limite. Essa idéia é resumida em um dos mais importantes teoremas em estatística, o *Teorema do Limite Central (TLC)*.

Teorema 1.4.6. ***(Teorema do Limite Central)*** *Seja $X_1, X_2, \ldots$ uma sequência de v.a. iid cuja função geradora de momentos (fgm) exista e uma vizinhança de 0 (ou seja, $M_{X_i}(t)$ existe para $|t| < h$, para algum positivo h. Seja $E[X_i] = \mu$ e $Var[X_i] = \sigma^2 > 0$. (Ambos μ e σ^2 são finitos e mgf existe) Defina $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$. Seja $G_n(x)$ que denota a fda de $\sqrt{n}\left(\bar{X}_n - \mu\right)\big/\sigma$. Então, para todo x, $-\infty < x < \infty$,*

$$
\lim_{n\to\infty} G(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy
$$

ou seja, $\sqrt{n}\left(\bar{X}_n - \mu\right)\big/\sigma$ tem como distribuição limite a normal padrão.

O teorema 1.4.6 tem algumas características interessantes. De começo, não existem praticamente nenhuma suposição sobre a distribuição em que a amostra foi obtida, ou pelo menos de qual se suspeita (além da indepedência e variância finita), o resultado final será a normalidade! O ponto aqui é que a normalidade vem de somas de "pequenas" (variância finita), independentes distúrbios. A suposição de variància finita é essencialmente necessária para a convergência à normalidade. Embora isso possa ser relaxado, caso não possa ser eliminado.

Enquanto que revelamos a maravilha de TLC, é também importante refletirmos sobre suas limitações. Embora o 1.4.6 dê uma aproximação geral importante, não existe m meio automático de sabermos o quão boa é a aproximação em geral. De fato, a bondade dessa aproximação é uma função da distribuição original, e deve ser checado caso a caso. Além disso, com os correntes avanços computacionais, o uso das aproximações pelo Teorema do Limite Central fica um pouco diminuída. Contudo, desprezando suas limitações, é ainda considerado um grande resultado.

OBS.: Uma versão mais forte do TLC é omitir a hipótese de existência da mgf e manter a existência de $E[X_i] = \mu$ e $Var[X_i] = \sigma^2$.

Exemplo 1.4.0.8. ***(Aproximação Normal para a Binomial Negativa)*** *Seja $X_1, X_2, ..., X_n$ uma amostra aleatória de um distribuição binomial negativa(r,p). Lembramos que*

$$E[X] = \frac{r(1-p)}{p}, \; Var[X] = \frac{r(1-p)}{p^2}$$

e o TLC nos diz que

$$\frac{\bar{X}_n - \mu}{\sigma} = \frac{\bar{X}_n - \frac{r(1-p)}{p}}{\sqrt{\frac{\frac{r(1-p)}{p^2}}{n}}} = \sqrt{n}\frac{\bar{X}_n - \frac{r(1-p)}{p}}{\sqrt{\frac{r(1-p)}{p^2}}} \approx N(0{,}1)$$

O cálculo aproximado de probabilidade é mais fácil do que o cálculo exato. Por exemplo se $r = 10$, $p = \frac{1}{2}$ e $n = 30$, o cálculo exato seria

$$\begin{aligned}
P\left[\bar{X} \leq 11\right] =& P\left[\underbrace{\sum_{i=1}^{30} X_i}_{Y=\sum X_i \sim BN(nr,p)} \leq 330\right] \\
=& \sum_{i=0}^{330} P[Y = i] \\
=& \sum_{i=0}^{330} \binom{300 + x - 1}{x} \left(\frac{1}{2}\right)^{300} \left(\frac{1}{2}\right)^{x} = 0{,}8916
\end{aligned}$$

pelo TLC temos a seguinte aproximação considerando que $\mu = E[X] = 10$ e $\sigma = \sqrt{Var[X]} = \sqrt{20}$,

$$\begin{aligned}
P\left[\bar{X} \leq 11\right] =& P\left[\left(\bar{X}_{10} - 10\right) \leq 11 - 10\right] \\
=& P\left[\left(\frac{\bar{X}_{30} - 10}{\sqrt{20}}\right) \leq \frac{11 - 10}{\sqrt{20}}\right] \\
=& P\left[\sqrt{30}\left(\frac{\bar{X}_{30} - 10}{\sqrt{20}}\right) \leq \sqrt{30}\frac{11 - 10}{\sqrt{20}}\right] \\
=& P[Z \leq 1{,}22] = 0{,}8888
\end{aligned}$$

Uma aproximação que pode ser usada conjuntamente com o TLC é conhecido como Teorema de Slutsky, apresentado a seguir.

Teorema 1.4.7. ***(Teorema de Slutsky)*** *Se $X_n \to X$ em distribuição e $Y_n \to a$ em probabilidade, sendo a uma constante, então*

- $Y_n X_n \to aX$ *em distribuição.*
- $X_n + Y_n \to X + a$ *em distribuição.*

Uma aplicação comum do teorema 1.4.7 é ilustrada no seguinte exemplo.

Exemplo 1.4.0.9. ***(Aproximação Normal com variância estimada)*** *Sabemos pelo TLC que*

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \to N\left(0{,}1\right)$$

mas o valor de σ é desconhecido. Sabemos também que

$$S_n^2 \to \sigma^2 \Rightarrow \sqrt{S_n^2} = S_n \to \sqrt{\sigma^2} = \sigma \Rightarrow \frac{S_n}{\sigma} \to 1$$

ou seja, a variância amostral converge em probabilidade para a variância populacional. Portanto, pelo teorema 1.4.7 temos que

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{S_n} = \underbrace{\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma}}_{\to N(0{,}1)} \underbrace{\frac{\sigma}{S_n}}_{\to 1} \to N\left(0{,}1\right)$$

Logo, assintoticamente $\frac{\sqrt{n}(\bar{X}_n-\mu)}{S_n}$ é uma grandeza pivota e pode-se obter o intervalo de confiança aproximado. Daí

$$\begin{aligned}
&\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{S_n} \approx Z \\
\Rightarrow &P\left[-z_{0{,}975} < \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{S_n} < z_{0{,}975}\right] \approx 0{,}95 \\
\Rightarrow &P\left[\bar{X}_n - \frac{S_n \times 1{,}96}{\sqrt{n}} < \mu < \bar{X}_n + \frac{S_n \times 1{,}96}{\sqrt{n}}\right] \\
\Rightarrow &IC\left(\mu\right)_{aproximado} = \left[\bar{X}_n - \frac{1{,}96 \times S_n}{\sqrt{n}}; \bar{X}_n + \frac{1{,}96 \times S_n}{\sqrt{n}}\right]
\end{aligned}$$

em que $\bar{X}_n$ não precisa ser de uma distribuição normal.

O Método Delta

Nas seções anteriores foram vistas condições sobre as quais uma variável aleatória tem no limite distribuição normal. Muitas vezes, contudo, não estamos interessados na distribuição de uma determinada v.a., mas em alguma função da variável aleatória.

Por exemplo, suponha que foram observadas $X_1, X_2, ..., X_n$ v.a. independentes de uma Bernoulli(p). Em muitas situações o parâmetro de interesse é p, que é a probabilidade de sucesso, e é estimado por $\hat{p} = \left(\frac{1}{n}\right) \sum_{i=1}^{n} X_i$, que é o estimador de máxima verossimilhança de p. Em medicina, por exemplo, uma estimativa de p fornece a probabilidade de um paciente obter a cura sobre um determinado tratamento e suponha que essa probabilidade seja $\hat{p} = 0{,}9$, ou seja, a probabilidade de ele ser curado é de 90%. Considere a seguinte medida: $\frac{p}{1-p}$. Se $\hat{p} = 0{,}9$, então $\frac{\hat{p}}{1-\hat{p}} = 9$. Essa medida é chamada de chance (*odds*) e representa a chance de cura de um paciente em relação a ele não ser curado. No caso, seria de 9:1, ou seja, a cada 10 pacientes, 9 são curados.

Embora o parâmetro $\frac{p}{1-p}$ seja uma medida interessante e muito utilizada, não se conhece a distribuição desse parâmetro e fazer inferências sobre esse parâmetro são difíceis de serem realizadas.

Para situações do tipo, o método Delta poderá ser utilizado e permitirá obter aproximações razoáveis para responder questões como do exemplo acima.

O método Delta utiliza aproximação por séries de Taylor, que permitirá aproximar a média e a variância de funções de variáveis aleatórias. Faz-se, então, necessário uma breve revisão de séries de Taylor.

Definição 1.4.3. *Se uma função $g(x)$ tem derivadas de ordem r, ou seja, $g^{(r)}(x) = \frac{d^r}{dx^r} g(x)$ existe, então para qualquer constante a, o polinômio de Taylor de ordem r sobre a é*

$$T_r(x) = \sum_{i=0}^{r} \frac{g^{(i)}(a)}{i!} (x-a)^i$$

e a aproximação de $g(x)$ por $T_r(x)$ é dada por

$$g(x) \approx \underbrace{g(x_0) + \sum_{i=1}^{r} \frac{g^{(i)}(a)}{i!} (x-a)^i}_{T_r(x)}$$

Da definição 1.4.3 acima, chega-se ao maior teorema de Taylor, que diz que o resto da aproximação, $g(x) - T_r(x)$, sempre tende a zero mais rápido que os termos de ordem superior, ou seja,

$$\lim_{x \to a} \frac{g(x) - T_r(x)}{(x-a)^r} = 0$$

Em muitas aplicações estatísticas do teorema de Taylor, as séries de Taylor de primeira ordem, ou seja, aproximações usando a primeira derivada (tomando $r = 1$ nas fórmulas acima), são muito utilizadas.

Podemos estar interessados em uma versão multivariada das séries de Taylor. Seja $X_1, ..., X_k$ uma sequência de variáveis aleatórias com médias $a_1, ..., a_k$, e defina $\mathbf{X} = X_1, ..., X_k$ e $\mathbf{a} = a_1, ..., a_k$. Suponha que existe uma função diferenciável $g(\mathbf{X})$ (um estimador de algum parâmetro) para o qual queremos uma estimativa da variância. Defina

$$g_i'(\mathbf{a}) = \frac{\partial}{\partial x_i} g(\mathbf{x}) |_{x_1=a_1, \ldots, x_k=a_k}$$

A expansão de primeira ordem da série de Taylor de g sobre $\mathbf{a}$ é

$$g(\mathbf{x}) = g(\mathbf{a}) + \sum_{i=1}^{r} g_i'(\mathbf{a})(x_i - a_i) + \text{restante} \tag{1.18}$$

Como estamos interessados em uma aproximação de primeira ordem, deixaremos de lado os termos restantes, daí reescrevemos a equação 1.18 por

$$g(\mathbf{x}) \approx g(\mathbf{a}) + \sum_{i=1}^{r} g_i'(\mathbf{a})(x_i - a_i) \tag{1.19}$$

Agora, aplicando e esperança em ambos os lados da expressão 1.19 temos

$$E_a[g(\mathbf{X})] \approx g(\mathbf{a}) + \sum_{i=1}^{k} g_i'(\mathbf{a})(X_i - a_i)$$
$$\Rightarrow E_a[g(\mathbf{X})] \approx g(\mathbf{a})$$

Podemos agora aproximar a variância de $g(\mathbf{X})$ por

$$Var_a[g(\mathbf{X})] \approx E_a\left[(g(\mathbf{X}) - g(\mathbf{a}))^2\right]$$

$$\begin{aligned}
Var_a\left[g\left(\mathbf{X}\right)\right] \approx & E_a\left[\left(\sum_{i=1}^{k} g_i'\left(\mathbf{a}\right)\left(X_i - a_i\right)\right)^2\right] \\
Var_a\left[g\left(\mathbf{X}\right)\right] \approx & E_a\left[\left(\sum_{i=1}^{k} \frac{\partial}{\partial x_i} g\left(\mathbf{a}\right)\left(X_i - a_i\right)\right)^2\right] \\
\approx & E_a\left[\sum_{i=1}^{k}\left(\frac{\partial}{\partial x_i} g\left(\mathbf{a}\right)\right)^2\left(X_i - a_i\right)^2 + 2\sum_{I<j}\frac{\partial}{\partial x_i} g\left(\mathbf{a}\right)\frac{\partial}{\partial x_j} g\left(\mathbf{a}\right)\left(X_i - a_i\right)\left(X_j - a_j\right)\right] \\
\approx & \sum_{i=1}^{k}\left(\frac{\partial}{\partial x_i} g\left(\mathbf{a}\right)\right)^2 Var_a\left[X_i\right] + 2\sum_{I<j}\frac{\partial}{\partial x_i} g\left(\mathbf{a}\right)\frac{\partial}{\partial x_j} g\left(\mathbf{a}\right) E\left[\left(X_i - a_i\right)\left(X_j - a_j\right)\right] \\
\approx & \sum_{i=1}^{k}\left(\frac{\partial}{\partial x_i} g\left(\mathbf{a}\right)\right)^2 Var_a\left[X_i\right] + 2\sum_{I<j}\frac{\partial}{\partial x_i} g\left(\mathbf{a}\right)\frac{\partial}{\partial x_j} g\left(\mathbf{a}\right) COV\left[X_i, X_j\right]
\end{aligned}$$

em que a última linha da expressão acima vem da expansão dos quadrados e usando a definição de variância e covariância. Essa aproximação é muito útil porque os dá uma fórmula para a variância de uma função geral, usando apenas as simples variâncias e covariâncias.

Como X é uma variável aleatória e queremos obter uma aproximação para sua média, temos que $E\left[X\right] = \mu$, pois $X \sim f\left(x;\mu\right)$. Podemos, em analogia às expressões anteriores, obter as expressões para a média e variância de X. Assim, a média é aproximada por

$$\begin{aligned}
g\left(X\right) \approx & g\left(\mu\right) + g'\left(\mu\right)\left(X - \mu\right) \\
E_\mu\left[g\left(X\right)\right] \approx & E\left[g\left(\mu\right) + g'\left(\mu\right)\left(X - \mu\right)\right] \\
\approx & g\left(\mu\right) + g'\left(\mu\right)\left(E\left[X\right] - \mu\right) \\
\approx & g\left(E_\mu\left[X\right]\right)
\end{aligned}$$

e a variância por

$$\begin{aligned}
Var_\mu\left[g\left(X\right)\right] \approx & Var_\mu\left[g\left(\mu\right) + g'\left(\mu\right)\left(X - \mu\right)\right] \\
\approx & E_\mu\left[\left(g\left(X\right) - E\left[g\left(X\right)\right]\right)^2\right] \\
\approx & E_\mu\left[\left(g\left(\mu\right) + g'\left(\mu\right)\left(X - \mu\right) - E\left[g\left(\mu\right) + g'\left(\mu\right)\left(X - \mu\right)\right]\right)^2\right] \\
\approx & E_\mu\left[\left(g'\left(\mu\right)\right)^2\left(X - \mu\right)^2\right] - \left(g'\left(\mu\right)\right)^2 E\left[\left(X - \mu\right)^2\right] \\
\approx & \left(g'\left(\mu\right)\right)^2 Var_\mu\left[X\right]
\end{aligned}$$

Com isso, vamos retornar ao exemplo do início dessa seção.

Exemplo 1.4.0.10. *Estamos interesados nas propriedades de* $\frac{\hat{p}}{1-\hat{p}}$ *como uma estimativa de* $\frac{p}{1-p}$*, onde* p *é a uma probabilidade de sucesso. Na notação utilizada,* $g\left(p\right) = \frac{p}{1-p}$ *então* $g'\left(p\right) = \frac{p}{(1-p)^2}$*. Daí*

$$\begin{aligned}
E_p\left[g\left(\hat{p}\right)\right] &\approx g\left(E_p\left[\hat{p}\right]\right) = g\left(p\right) = \frac{p}{1-p} \\
Var_p\left[g\left(\hat{p}\right)\right] &\approx \left(g'\left(\hat{p}\right)\right)^2 Var_\mu\left[\hat{p}\right] = \frac{1}{\left(1-p\right)^4}\frac{p\left(1-p\right)}{n} = \frac{p}{n\left(1-p\right)^3}
\end{aligned}$$

que nos dá uma aproximação para a variância de nosso estimador.

A seguir mais um exemplo da utilidade das aproximações

Exemplo 1.4.0.11. *Considere duas variáveis aleatórias com distribuição exponencial, ou seja,* $X \sim \exp(\alpha)$ *e* $Y \sim \exp(\beta)$*, e independentes. Considere uma outra variável aleatória* $Z = g(X,Y) = \frac{X}{Y}$*. Qual* $E[Z]$ *e* $Var[Z]$*?*

SOLUÇÃO

Utilizando as aproximações para a média e variância temos que

$$\begin{aligned} E[g(X,Y)] \approx & g(E[X], E[Y]) \\ \approx & \frac{E[X]}{E[Y]} = \frac{\frac{1}{\alpha}}{\frac{1}{\beta}} = \frac{\beta}{\alpha} \end{aligned}$$

e para a variância,

$$Var[Z] \approx \left(\frac{\partial g(X,Y)}{\partial x}\right)^2 Var[X] + \left(\frac{\partial g(X,Y)}{\partial y}\right)^2 Var[Y]$$

cujas derivadas parciais são

$$\begin{aligned} \frac{\partial g(X,Y)}{\partial x} &= \frac{1}{y} \Rightarrow \frac{\partial}{\partial x} g\left(\frac{1}{\alpha}, \frac{1}{\beta}\right) = \beta \\ \frac{\partial g(X,Y)}{\partial y} &= -\frac{x}{y^2} \Rightarrow \frac{\partial}{\partial y} g\left(\frac{1}{\alpha}, \frac{1}{\beta}\right) = -\frac{\frac{1}{\alpha}}{\left(\frac{1}{\beta}\right)^2} = -\frac{\beta^2}{\alpha} \end{aligned}$$

logo,

$$Var[Z] \approx (\beta)^2 \frac{1}{\alpha^2} + \left(-\frac{\beta^2}{\alpha}\right)^2 \frac{1}{\beta^2} = 2\left(\frac{\beta}{\alpha}\right)^2$$

Após a exposição das idéias do método Delta e dos exemplos, vamos apresentar de maneira formal o seguinte teorema do método Delta.

Teorema 1.4.8. ***(Método Delta)*** *Seja* Y_n *uma sequência de variáveis aleatórias que satisfaz* $\sqrt{n}(Y_n - \theta) \to N(0, \sigma^2)$ *em distribuição. Para uma dada função* g *e um específico valor de* θ*, suponha que* $g'(\theta)$ *exista e não é 0. Então*

$$\sqrt{n}(g(Y_n) - g(\theta)) \to N\left(0, \sigma^2 (g'(\theta))^2\right) \tag{1.20}$$

em distribuição.

Demonstração. A prova do teorema 1.4.8 utiliza todas as considerações já feitas sobre a aproximação da média e variância e das séries de Taylor. A demonstração pode ser colocada da seguinte forma: pela expansão de Taylor de primeira ordem de $g(Y_n)$ em torno de $Y_n = \theta$ temos que

$$\begin{aligned} & g(Y_n) \approx g(\theta) + g'(\theta)(Y_n - \mu) \\ & \Rightarrow g(Y_n) - g(\theta) \approx g'(\theta)(Y_n - \mu) \\ & \Rightarrow [g(Y_n) - g(\theta)]^2 \approx [g'(\theta)(Y_n - \mu)]^2 \\ & \Rightarrow E\left[[g(Y_n) - g(\theta)]^2\right] \approx E\left[[g'(\theta)(Y_n - \mu)]^2\right] \\ & \Rightarrow E\left[[g(Y_n) - g(\theta)]^2\right] \approx (g'(\theta))^2 Var[Y_n] \end{aligned}$$

e para a média

$$g(Y_n) \approx g(\theta) + g'(\theta)(Y_n - \mu)$$

$$\Rightarrow g\left(Y_n\right) - g\left(\theta\right) \approx g'\left(\theta\right)\left(Y_n - \mu\right)$$
$$\Rightarrow E\left[g\left(Y_n\right) - g\left(\theta\right)\right] \approx E\left[g'\left(\theta\right)\left(Y_n - \mu\right)\right]$$
$$\Rightarrow E\left[g\left(Y_n\right) - g\left(\theta\right)\right] \approx g'\left(\theta\right)\underbrace{E\left[\left(Y_n - \mu\right)\right]}_{=0}$$
$$\Rightarrow E\left[g\left(Y_n\right) - g\left(\theta\right)\right] \approx 0$$
$$\Rightarrow E\left[g\left(Y_n\right)\right] \approx g\left(\theta\right)$$

Logo, considerando os resultados acima e o Teorema de Slutsky (teorema 1.4.7) o reultado do teorema 1.4.8 vale. □

OBS.: Se $g'\left(\theta\right) = 0$ o método Delta não funciona.

Neste caso vamos utilizar a expansão de Taylor de ordem 2, que será mostrado a seguir

$$g\left(Y_n\right) = g\left(\theta\right) + g'\left(\theta\right)\left(Y_n - \theta\right) + \frac{g''\left(\theta\right)}{2}\left(Y_n - \theta\right)^2 + \text{restante}$$
$$\Rightarrow g\left(Y_n\right) - g\left(\theta\right) \approx \frac{g''\left(\theta\right)}{2}\left(Y_n - \theta\right)^2$$

que será utilizado no método Delta de segunda ordem.

Teorema 1.4.9. ***(Método Delta de segunda ordem)*** *Seja Y_n uma sequência de variáveis aleatórias que satisfaz $\sqrt{n}\left(Y_n - \theta\right) \to N\left(0{,}\sigma^2\right)$ em distribuição. Para uma dada função g e um específico valor de θ, suponha que $g'\left(\theta\right) = 0$ e $g''\left(\theta\right)$ exista e não é 0. Então*

$$\sqrt{n}\left(g\left(Y_n\right) - g\left(\theta\right)\right) \to \sigma^2 \frac{g''\left(\theta\right)}{2}\chi^2_1 \tag{1.21}$$

em distribuição.

Demonstração. A prova do teorema 1.4.9 utiliza todas as considerações já feitas sobre a aproximação da média e variância e das séries de Taylor. A demonstração pode ser colocada da seguinte forma: pela expansão de Taylor de segunda ordem de $g(Y_n)$ em torno de $Y_n = \theta$ temos que

$$g\left(Y_n\right) - g\left(\theta\right) \approx \frac{g''\left(\theta\right)}{2}\left(Y_n - \theta\right)^2$$
$$\Rightarrow \frac{n\left(g\left(Y_n\right) - g\left(\theta\right)\right)}{\sigma^2} \approx \frac{g''\left(\theta\right)}{2}\left(\frac{\sqrt{n}\left(Y_n - \theta\right)}{\sigma}\right)^2$$

como $\frac{\sqrt{n}(Y_n - \theta)}{\sigma} \to N\left(0{,}1\right)$ em distribuição e $Z \sim N\left(0{,}1\right) \Rightarrow Z^2 \sim \chi^2_1$ temos que

$$\Rightarrow \frac{n\left(g\left(Y_n\right) - g\left(\theta\right)\right)}{\sigma^2 \frac{g''(\theta)}{2}} \to \chi^2_1$$
$$\Rightarrow n\left(g\left(Y_n\right) - g\left(\theta\right)\right) \to \sigma^2 \frac{g''\left(\theta\right)}{2}\chi^2_1$$

que converge em distribuição para o fator $\sigma^2 \frac{g''(\theta)}{2}\chi^2_1$, que nada mais é do que uma χ^2_1 escalonada por uma fator $\sigma^2 \frac{g''(\theta)}{2}$. □

1.5 Métodos para avaliar estimadores

Existem várias técnicas para encontrar estimadores pontuais. Uma dificuldade que existe é que pode-se aplicar mais de um método de estimação em situações particulares e devemos fazer uma escolha de um entre os estimadores utilizados. É possível que diferentes métodos de encontrar estimadores levem a mesmas respostas, e alguns com cálculos mais fáceis, mas, em muitos casos, métodos diferentes levam a resultados diferentes.

Será discutido a seguir alguns métodos para avaliar estimadores para auxiliar na escolha entre vários estimadores.

Erro Quadrático Médio

Considere uma amostra finita. Uma primeira medida de qualidade de um estimador é o erro quadrático médio, apresentado a seguir.

Definição 1.5.1. *O erro quadrático médio (MSE) de um estimador W para um parâmetro θ é a função de θ definida por $E_\theta\left[(W-\theta)^2\right]$.*

Note que MSE mede a diferença quadrática média entre o estimador W e o parâmetro θ e é uma medida razável de performance para um estimador pontual. Em geral, qualquer incremento na função da distância absoluta $|W-\theta|$ serviria como uma medida de qualidade de um estimador (erro médio absoluto, $E\left[|W-\theta|\right]$, é uma alternativa razoável), mas o MSE tem pelo menos duas vantagens sobre outras medidas de distância: primeiro, ela é tratável analiticamente e, segundo, tem interpretação

$$E_\theta\left[(W-\theta)^2\right] = Var_\theta\left[W\right] + E\left[W-\theta\right]^2 = Var_\theta\left[W\right] + \left(Vies_\theta\left[W\right]\right)^2$$

onde o viés de um estimador tem a seguinte definição

Definição 1.5.2. *O viés de um estimador pontual W para um parâmetro θ é a diferença entre o valor esperado de W e θ, ou seja, $Vies_\theta\left[W\right] = E\left[W-\theta\right]$. Um estimador cujo viés é identicamente (em θ) igual a zero é chamado não viesado e satisfaz $E\left[W\right]=\theta$ para todo θ.*

Portanto, o MSE incorpora duas componentes, uma medindo a variabilidade do estimador (precisão) e a outra medindo seu viés (acurácia). Um estimador que tem bom MSE tem uma combinação de valor pequeno de variância e viés.

Para um estimador não viesado temos que

$$E_\theta\left[(W-\theta)^2\right] = Var_\theta\left[W\right]$$

logo, se um estimador é não viesado, seu MSE é igual a sua variância.

Exemplo 1.5.0.12. ***(MSE da Normal)*** *Seja $X_1,...,X_n$ uma amostra iid de $X \sim N\left(\mu,\sigma^2\right)$. As estatísticas $\overline{X}$ e S^2 são ambas estimadores não viesados, uma vez que $E\left[\overline{X}\right]=\mu$ e $E\left[S^2\right]=\sigma^2$, para todo μ e σ^2. Os MSE's desses estimadores são dados por*

$$E\left[\left(\bar{X}-\mu\right)^2\right] = Var\left[\bar{X}\right] = \frac{\sigma^2}{n}$$
$$E\left[\left(S^2-\sigma^2\right)^2\right] = Var\left[S^2\right] = \frac{2\sigma^4}{n-1}$$

o MSE de $\bar{X}$ resulta em $\frac{\sigma^2}{n}$ até mesmo se a suposição de normalidade é perdida. Contudo, a expressão acima para o MSE de S^2 não permanece a mesma se a suposição de normalidade é relaxada.

Melhor estimador não viesado

Como visto antes, uma comparação de estimadores entre estimadores baseado no MSE pode não levar a um estimador favorito diretamente. De fato, não existe um únicoestimador "melhor em MSE" dentre uma vasta classe de estimadores. Por exemplo, o estimador $\hat{\theta} = 17$ é o melhor em MSE quando $\theta = 17$, mas é o pior estimador em outros casos. Uma forma de contornar esse problema de encontrar o melhor estimador em MSR é restringir a classe de estimadores. Uma maneira popular de restringir a classe de estimadores é considerar a classe dos estimadores não viesados.

Se W_1 e W_2 são ambos estimadores não viesados para o parâmetro θ, ou seja, $E_\theta\left[W_1\right] = E_\theta\left[W_2\right] = \theta$, então seus erros quadráticos médios são iguais a suas variâncias, logo, escolheremos o estimador com menor variância. Se podemos encontrar um estimador com a menor variância (o melhor estimador não viesado) então o problema está resolvido.

Definição 1.5.3. *Um estimador W^* é o melhor estimador não viesado de $\tau(\theta)$ se satisfaz $E_\theta\left[W^*\right] = \tau(\theta)$ para todo θ e, para qualquer outro estimador W com $E_\theta[W] = \tau(\theta)$, temos $Var_\theta\left[W^*\right] \leq Var_\theta[W]$ para todo θ. W^* é também chamado de estimador não viesado uniforme de variância mínima (UMVUE) de $\tau(\theta)$.*

Encontrar um melhor estimador não viesado (se ele existe!) não é uma tarefa fácil por uma variedade de razões, duas delas serão ilustradas no seguinte exemlo.

Exemplo 1.5.0.13. *(Estimação não viesada da Poisson) Seja $X_1, ..., X_n$ uma amostra iid e $X \sim Poisson(\lambda)$, e seja $\bar{X}$ e S^2 a média e variância amostral, respectivamente. Para a distribuição Poisson, ambas média e variância são iguais a λ. Além disso, sabe-se que*

$$E_\lambda\left[\bar{X}\right] = \lambda, \forall\lambda \qquad\qquad E_\lambda\left[S^2\right] = \lambda, \forall\lambda$$

então ambos $\bar{X}$ e S^2 são estimadores não viesados de λ.

Para determinar o melhor estimador, $\bar{X}$ ou S^2, devemos comparar suas variâncias. Sabemos que $Var_\lambda\left[\bar{X}\right] = \frac{\lambda}{n}$, mas $Var_\lambda\left[S^2\right] = \frac{\lambda}{n}$ é difícil de calcular. Esse é o primeiro problema de encontrar um melhor estimador não viesado. Os cálculos, além de difíceis, podem não serem úteis, como nesse exemplo, porque veremos que $Var_\lambda\left[\bar{X}\right] \leq Var_\lambda\left[S^2\right]$ para todo λ.

Estabeleça que $\bar{X}$ é melhor que S^2, considere a classe de estimadores

$$W_\alpha\left(\bar{X}, S^2\right) = \alpha\bar{X} + (1-\alpha) S^2$$

para cada constante α, temos que

$$E_\lambda\left[W_\alpha\left(\bar{X}, S^2\right)\right] = E_\lambda\left[\alpha\bar{X} + (1-\alpha) S^2\right] = \alpha E_\lambda\left[\bar{X}\right] + (1-\alpha) E_\lambda\left[S^2\right] = \lambda$$

ou seja, $W_\alpha\left(\bar{X}, S^2\right)$ também é um estimador não viesado para λ, então temos uma infinidade de estimadores não viesados para λ. Se $\bar{X}$ é melhor do que S^2, podemos afirmar que $\bar{X}$ é melhor do que $W_\alpha\left(\bar{X}, S^2\right)$? Além disso, como podemos garantir que não exista um outro estimador para λ melhor do que os apresentados?

O exemplo mostrou alguns problemas que poderiam ser encontrados na tentativa de obter um estimador não viesado. Suponha que, para estimar um parâmetro $\tau(\theta)$ de uma distribuição $f(x|\theta)$, possamos especificar um limite inferior, diga $g(\theta)$, sobre a variância de qualquer estimador não viesado de $\tau(\theta)$. Se pudermos encontrar um estimador não viesado W^* satisfazendo $Var_\theta\left[W^*\right] = g(\theta)$,

Teorema 1.5.1. ***(Desigualdade de Cramér-Rao)*** *Seja $X_1, ..., X_n$ uma amostra com fdp $f(\mathbf{x}|\theta)$, e seja $W(\mathbf{X}) = W(X_1, \ldots, X_n)$ qualquer estimador que satisfaça*

$$\frac{d}{d\theta} E_\theta[W(\mathbf{X})] = \int_{\aleph} \frac{\partial}{\partial\theta}[W(\mathbf{x}) f(\mathbf{x}|\theta)]\, d\mathbf{x} \tag{1.22}$$

e

$$Var_\theta\left[W\left(\mathbf{X}\right)\right] < \infty.$$

Então

$$Var_\theta\left[W\left(\mathbf{X}\right)\right] \geq \frac{\left(\frac{d}{d\theta}E_\theta\left[W\left(\mathbf{X}\right)\right]\right)^2}{E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f\left(\mathbf{X}\left|\theta\right.\right)\right)^2\right]} \tag{1.23}$$

Demonstração. A prova desse teorea é elegantemente simples e utiliza aplicação da desigualdade de Cauchy-Schwarz. Considere duas variáveis aleatórias X e Y,

$$\left[COV\left(X,Y\right)\right]^2 \leq Var\left[X\right]Var\left[Y\right] \tag{1.24}$$

rearranjando a expressão 1.24, podemos obter um limite inferior para a variância de X,

$$Var\left[X\right] \geq \frac{\left[COV\left(X,Y\right)\right]^2}{Var\left[Y\right]}$$

a chave desse teorema segue da escolha de X como sendo o estimador $W\left(\mathbf{X}\right)$ e Y como sendo a quantidade $\frac{\partial}{\partial\theta}\log f\left(\mathbf{X}\left|\theta\right.\right)$ e aplicando a desigualdade 1.24.

Primeiro, note que

$$\begin{aligned}
\frac{d}{d\theta}E_\theta\left[W\left(\mathbf{X}\right)\right] &= \frac{d}{d\theta}\left(\int\dots\int W\left(X_1,\dots,X_n\right)f_{X_1,\dots,X_n}\left(x_1,\dots,x_n\left|\theta\right.\right)dx_1,\dots,dx_n\right)\\
&= \int\dots\int W\left(X_1,\dots,X_n\right)\frac{d}{d\theta}f_{X_1,\dots,X_n}\left(x_1,\dots,x_n\left|\theta\right.\right)dx_1,\dots,dx_n\\
&= \int\dots\int W\left(\mathbf{X}\right)\left[\frac{d}{d\theta}f_{\mathbf{X}}\left(\mathbf{X}\left|\theta\right.\right)\right]\frac{f_{\mathbf{X}}\left(\mathbf{X}\left|\theta\right.\right)}{f_{\mathbf{X}}\left(\mathbf{X}\left|\theta\right.\right)}dx_1,\dots,dx_n\\
&= E_\theta\left[W\left(\mathbf{X}\right)\frac{\frac{d}{d\theta}f_{\mathbf{X}}\left(\mathbf{X}\left|\theta\right.\right)}{f_{\mathbf{X}}\left(\mathbf{X}\left|\theta\right.\right)}\right]\\
&= E_\theta\left[W\left(\mathbf{X}\right)\frac{d}{d\theta}\log\left(f_{\mathbf{X}}\left(\mathbf{X}\left|\theta\right.\right)\right)\right]
\end{aligned}$$

o qual sugere uma covariância entre $W\left(\mathbf{X}\right)$ e $\frac{d}{d\theta}\log\left(f_{\mathbf{X}}\left(\mathbf{X}\left|\theta\right.\right)\right)$. Para isso ser a covariância, é necessário subtrair o produto dos valores experados, então calculamos $E_\theta\left[\frac{d}{d\theta}\log\left(f_{\mathbf{X}}\left(\mathbf{X}\left|\theta\right.\right)\right)\right]$. Mas se aplicarmos a esperança acima com $W\left(\mathbf{X}\right)=1$, temos que

$$\begin{aligned}
\int_{-\infty}^{\infty}f'\left(x\right)dx = \int_{-\infty}^{\infty}\frac{f'\left(x\right)}{f\left(x\right)}f\left(x\right)dx &= \int_{-\infty}^{\infty}\frac{d}{dx}\log\left(f\left(x\right)\right)f\left(x\right)dx\\
&= E\left[\frac{d}{dx}\log\left(f\left(x\right)\right)\right]
\end{aligned}$$

como $f\left(x;\,\theta\right)$ é fdp,

$$\begin{aligned}
\int_{-\infty}^{\infty}\frac{d}{d\theta}f\left(x;\,\theta\right)dx = \int_{-\infty}^{\infty}\frac{\frac{d}{d\theta}f\left(x;\,\theta\right)}{f\left(x;\,\theta\right)}f\left(x;\,\theta\right)dx &= \int_{-\infty}^{\infty}\left[\frac{d}{d\theta}\log\left(f\left(x;\,\theta\right)\right)\right]f\left(x;\,\theta\right)dx\\
&= E_\theta\left[\frac{d}{d\theta}\log\left(f\left(X;\,\theta\right)\right)\right]
\end{aligned}$$

ainda, note que

$$\int_{-\infty}^{\infty} \frac{d}{d\theta} f(x;\theta)\,dx = \frac{d}{d\theta}\int_{-\infty}^{\infty} f(x;\theta)\,dx = \frac{d}{d\theta}(1) = 0$$

logo, a v.a. $\frac{d}{d\theta}\log(f(X;\theta))$ tem média zero, para qualquer que seja o parâmetro θ.

Portanto $COV\left[W(\mathbf{X}), \frac{d}{d\theta}\log(f_{\mathbf{X}}(\mathbf{X}\,|\theta))\right]$ é igual a esperança do produto, logo

$$COV\left[W(\mathbf{X}), \frac{d}{d\theta}\log(f_{\mathbf{X}}(\mathbf{X}\,|\theta))\right] = E_\theta\left[W(\mathbf{X})\frac{d}{d\theta}\log(f_{\mathbf{X}}(\mathbf{X}\,|\theta))\right] = \frac{d}{d\theta}E_\theta[W(\mathbf{X})]. \tag{1.25}$$

Também, uma vez que $E_\theta\left[\frac{d}{d\theta}\log(f_{\mathbf{X}}(\mathbf{X}\,|\theta))\right] = 0$ e do fato que $Var[X] = E[X^2] - (E[X])^2$ temos

$$Var_\theta\left[\frac{d}{d\theta}\log(f_{\mathbf{X}}(\mathbf{X}\,|\theta))\right] = E_\theta\left[\left(\frac{d}{d\theta}\log(f_{\mathbf{X}}(\mathbf{X}\,|\theta))\right)^2\right] \tag{1.26}$$

Usando a desigualdade de Cauchy-Schwarz juntamente com 1.25 e 1.26, obtemos

$$Var_\theta[W(\mathbf{X})] \geq \frac{\left(\frac{d}{d\theta}E_\theta[W(\mathbf{X})]\right)^2}{E_\theta\left[\left(\frac{d}{d\theta}\log(f_{\mathbf{X}}(\mathbf{X}\,|\theta))\right)^2\right]}$$

provando o teorema. □

Para o teorema acima, se adicionarmos a suposição de amostras independentes, então o cálculo do limite inferior é simplificado. A esperança no denominador da expressão 1.23 recai a cálculos univariados, conforme será mostrado no corolário a seguir.

Corolário 1.5.1. ***(Caso iid para a Desigualdade de Cramér-Rao)*** *Se as suposiçõs do teorema 1.5.1 são satisfeitas e, adicionalmente, se $X_1,\ldots,X_n$ são iid (independentes e identicamente distribuídas) com pdf $f(x\,|\theta)$, então*

$$Var_\theta[W(\mathbf{X})] \geq \frac{\left(\frac{d}{d\theta}E_\theta[W(\mathbf{X})]\right)^2}{nE_\theta\left[\left(\frac{d}{d\theta}\log(f_{\mathbf{X}}(X\,|\theta))\right)^2\right]}. \tag{1.27}$$

Demonstração. Precisamos mostrar apenas que

$$E_\theta\left[\left(\frac{d}{d\theta}\log(f_{\mathbf{X}}(\mathbf{X}\,|\theta))\right)^2\right] = nE_\theta\left[\left(\frac{d}{d\theta}\log(f_{\mathbf{X}}(X\,|\theta))\right)^2\right]$$

Usando o fato de a amostra $X_1,\ldots,X_n$ ser independente, temos que

$$\begin{aligned}\frac{d}{d\theta}\log(f_{X_1,\ldots,X_n}(X_1,\ldots,X_n;\theta)) &= \frac{d}{d\theta}\log\left(\prod_{i=1}^n f(X_i;\theta)\right)\\ &= \frac{d}{d\theta}\sum_{i=1}^n \log(f(X_i;\theta))\end{aligned}$$

elevando ambos os membros ao quadrado

$$\begin{aligned}\left[\frac{d}{d\theta}\log(f_{X_1,\ldots,X_n}(X_1,\ldots,X_n;\theta))\right]^2 &= \left[\frac{d}{d\theta}\sum_{i=1}^n \log(f(X_i;\theta))\right]^2\\ &= \sum_{i=1}^n\left[\frac{d}{d\theta}\log(f(X_i;\theta))\right]^2\\ &+ 2\sum_{i<j}\frac{d}{d\theta}\log(f(X_i;\theta))\frac{d}{d\theta}\log(f(X_j;\theta))\end{aligned}$$

note que a passagem da primeira expressão para a segunda se deve ao fato de aparecer somas de termos quadráticos e somas de termos com produtos cruzados. Aplicando a esperança em ambos os lados temos que

$$E_\theta\left[\left[\frac{d}{d\theta}\log\left(f_{X_1,\ldots,X_n}\left(X_1,\,\ldots,X_n;\,\theta\right)\right)\right]^2\right]=E_\theta\left[\left[\frac{d}{d\theta}\sum_{i=1}^{n}\log\left(f\left(X_i;\,\theta\right)\right)\right]^2\right]$$
$$=E_\theta\left[\sum_{i=1}^{n}\left[\frac{d}{d\theta}\log\left(f\left(X_i;\,\theta\right)\right)\right]^2\right]+2E_\theta\left[\sum_{i<j}\frac{d}{d\theta}\log\left(f\left(X_i;\,\theta\right)\right)\frac{d}{d\theta}\log\left(f\left(X_j;\,\theta\right)\right)\right]$$
$$=\sum_{i=1}^{n}E_\theta\left[\left[\frac{d}{d\theta}\log\left(f\left(X_i;\,\theta\right)\right)\right]^2\right]+2\sum_{i<j}E_\theta\left[\frac{d}{d\theta}\log\left(f\left(X_i;\,\theta\right)\right)\right]E_\theta\left[\frac{d}{d\theta}\log\left(f\left(X_j;\,\theta\right)\right)\right]$$

Note que da expressão acima o produto das esperanças é zero, uma vez que as variáveis aleatórias i e j são independentes. Note ainda que o termo $\sum_{i=1}^{n}E_\theta\left[\left[\frac{d}{d\theta}\log\left(f\left(X_i;\,\theta\right)\right)\right]^2\right]$ nada mais é do que a soma da esperança de uma mesma variável aleatória. Portanto,

$$\sum_{i=1}^{n}E_\theta\left[\left[\frac{d}{d\theta}\log\left(f\left(X_i;\,\theta\right)\right)\right]^2\right]=nE_\theta\left[\left[\frac{d}{d\theta}\log\left(f\left(X;\,\theta\right)\right)\right]^2\right]$$

finalizando a prova do corolário. □

A cota de Cramér-Rao foi apresentada para variáveis contínuas, mas também é aplicada à variáveis aleatórias discretas.

A quantidade $I\left(\theta\right)=E_\theta\left[\left[\frac{d}{d\theta}\log\left(f\left(X;\,\theta\right)\right)\right]^2\right]$ é chamada de *matriz de informação de Fisher* ou *número de informação de Fisher* da amostra. Essa terminologia reflete o fato de que o número de informação fornece um limite para a variância do melhor estimador não viesado de θ. Conforme o número de informação se torna maior e temos mais informação sobre θ, temos um menor limite para a variância do melhor estimador não viesado.

A desigualdade de Cramér-Rao é muito útil na comparação do desempenho de estimadores. Para uma função diferenciável $\tau\left(\theta\right)$, temos agora um limite inferior da variânia de qualquer estimador W, tal que $E_\theta\left[W\right]=\tau\left(\theta\right)$. A cota depende apenas de $\tau\left(\theta\right)$ e $f\left(x\left|\theta\right.\right)$ e é uma cota inferior uniforme sobre a variância. Qualquer estimador candidato satisfazendo $E_\theta\left[W\right]=\tau\left(\theta\right)$ e alcançando esse limite inferior é o melhor estimador não viesado de $\tau\left(\theta\right)$.

Uma forma mais simples da desigualdade de Cramér-Rao é se o estimador W for uma identidade, ou seja, se $E_\theta\left[W\right]=\tau\left(\theta\right)=\theta$. Nesse caso a expressão do corolário 1.5.1 se reduz a

$$Var_\theta\left[W\left(\mathbf{X}\right)\right]\geq\frac{\left(\tau'\left(\theta\right)\right)^2}{nE_\theta\left[\left(\frac{d}{d\theta}\log\left(f_{\mathbf{X}}\left(X\left|\theta\right.\right)\right)\right)^2\right]}$$
$$\geq\frac{1}{nE_\theta\left[\left(\frac{d}{d\theta}\log\left(f_{\mathbf{X}}\left(X\left|\theta\right.\right)\right)\right)^2\right]}$$

que fica apenas em termos da fdp de X.

Exemplo 1.5.0.14. ***(Estimação não viesada da Poisson - Continuação)*** *Seja $X_1,\ldots,X_n$ uma amostra iid e $X\sim Poisson\left(\lambda\right)$. Sabe-se que $E_\lambda\left[\bar{X}\right]=\lambda,\,\forall\lambda$, $Var_\lambda\left[\bar{X}_n\right]=\frac{\lambda}{n}$, $E_\lambda\left[S_n^2\right]=\lambda$ e $Var_\lambda\left[S_n^2\right]$ é difícil de ser obtida. Qual é melhor estimador, considerando que conhecemos a expressão de $Var_\lambda\left[S_n^2\right]$?*

SOLUÇÃO
A cota de Cramér-Rao para $Var_\lambda\left[\bar{X}_n\right]$ é

$$
\begin{aligned}
Var_\lambda\left[\bar{X}_n\right] &\geq \frac{1}{-nE_\theta\left[\left(\frac{d^2}{d\lambda^2}\log\left(f_{\mathbf{X}}\left(X\left|\theta\right.\right)\right)\right)^2\right]} \\
&\geq \frac{1}{-nE_\lambda\left[\frac{d^2}{d\lambda^2}\log\left(\frac{\lambda^X e^{-\lambda}}{X!}\right)\right]} \\
&\geq \frac{1}{-nE_\lambda\left[\frac{d^2}{d\lambda^2}X\log\left(\lambda\right)-\lambda-\log\left(X!\right)\right]} \\
&\geq \frac{1}{-nE_\lambda\left[-\frac{X}{\lambda^2}\right]}=\frac{\lambda^2}{nE_\lambda\left[X\right]}=\frac{\lambda^2}{n\lambda}=\frac{\lambda}{n}
\end{aligned}
$$

Logo, a estatítica $\bar{X}_n$ tem variância igual a cota de Cramér-Rao e, portanto, é o melhor possível. Note que mesmo conhecendo a expressão de $Var_\lambda\left[S_n^2\right]$, podemos fazer conclusões apenas com $\bar{X}_n$.

A seguir um outro exemplo.

Exemplo 1.5.0.15. ***(Cota de Cramér-Rao para σ^2 da $N\left(\mu,\sigma^2\right)$)*** *Seja $X_1,\ldots,X_n$ uma amostra iid de uma $N\left(\mu,\sigma^2\right)$, e considere a estimação de σ^2, onde μ é desconhecido. a pdf da normal satisfaz condições do teorema 1.5.1. Calcule a cota de Cramér-Rao (CR) para $Var_{\sigma^2}\left[S^2\right]$.*
SOLUÇÃO
Vamos obter a cota de CR para $Var_{\sigma^2}\left[S^2\right]$.

$$
\begin{aligned}
Var_{\sigma^2}\left[S^2\right] &\geq \frac{1}{-nE_{\sigma^2}\left[\left(\frac{d^2}{d(\sigma^2)^2}\log\left(f_{\mathbf{X}}\left(X\left|\mu,\sigma^2\right.\right)\right)\right)^2\right]} \\
&\geq \frac{1}{-nE_{\sigma^2}\left[\frac{d^2}{d(\sigma^2)^2}\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}\right)\right]} \\
&\geq \frac{1}{-nE_{\sigma^2}\left[\frac{d^2}{d\lambda^2}\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}-\frac{1}{2}\log\left(\sigma^2\right)-\log\left(\sqrt{2\pi}\right)\right)\right]} \\
&\geq \frac{1}{-nE_{\sigma^2}\left[\frac{(x-\mu)^2}{\sigma^6}-\frac{1}{2(\sigma^2)^2}\right]}=\frac{1}{n\left(\frac{\sigma^2}{\sigma^6}-\frac{1}{2\sigma^4}\right)}=\frac{2\sigma^4}{n} \\
\Rightarrow Var_{\sigma^2}\left[S^2\right] &\geq \frac{2\sigma^4}{n}
\end{aligned}
$$

Sabe-se que a $Var_{\sigma^2}\left[S^2\right]$ é dada por

$$
\begin{aligned}
Var_{\sigma^2}\left[S^2\right] &= \frac{1}{n}\left(\mu_4-\frac{n-3}{n-1}\mu_2^2\right) \\
&= \frac{1}{n}\left(\mu_4-\frac{n-3}{n-1}\sigma^4\right) \\
&= \frac{1}{n}\left(3\sigma^4-\frac{n-3}{n-1}\sigma^4\right) \\
&= \frac{1}{n}\left(3-\frac{n-3}{n-1}\right)\sigma^4 \\
&= \frac{1}{n}\left(\frac{3n-3-n+3}{n-1}\right)\sigma^4
\end{aligned}
$$

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

Note que $\frac{2\sigma^4}{n-1}$ *é maior que* $\frac{2\sigma^4}{n}$, *ou seja, a* $Var_{\sigma^2}\left[S^2\right]$ *é maior que a cota de CR da* $Var_{\sigma^2}\left[S^2\right]$. *Portanto a variância do estimador* S^2 *de* σ^2 *não alcança a cota de CR.*

O corolário a seguir pode ser útil na procura do melhor estimador não viesado.

Corolário 1.5.2. *Seja* $X_1, \ldots, X_n$ *uma amostra iid com* $f(x|\theta)$, *onde* $f(x|\theta)$ *satisfaz as condições do teorema 1.5.1. Seja* $L(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta)$ *a função de verossimilhança. Se* $W(\mathbf{X}) = W(X_1, \ldots, X_n)$ *é qualquer estimador não viesado de* $\tau(\theta)$, *então* $W(\mathbf{X})$ *alcança o limite inferior de Cramér-Rao se, e somente se,*

$$a(\theta)\left[W(\mathbf{x}) - \tau(\theta)\right] = \frac{d}{d\theta}\log\left(L(\theta|\mathbf{x})\right) \tag{1.28}$$

para alguma função $a(\theta)$.

Exemplo 1.5.0.16. *(Aplicação do corolário 1.5.2 no exemplo anterior da Normal.)*

SOLUÇÃO

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

e então

$$\frac{d}{d\sigma^2}\log\left(L\left(\mu, \sigma^2|\mathbf{x}\right)\right) = \frac{n}{2\sigma^4}\left(\sum_{i=1}^n \frac{(x_i-\mu)^2}{n} - \sigma^2\right)$$

Portanto, tomando a $\left(\sigma^2\right) = \frac{n}{2\sigma^4}$ *mostra que o melhor estimador não viesado de* σ^2 *é* $\sum_{i=1}^n \frac{(x_i-\mu)^2}{n}$, *que é calculado somente se* μ *é conhecido. Se* μ *é desconhecido, a cota não pode ser alcançada.*

Os estimadores de máxima verossimilhança (EMV) são muito utilizados pela comunidade statística. Uma questão que pode surgir sobre esses estimadores é sobre suas propriedades. Com base em uma amostra podemos garantir apenas que a procedência dos estimadores de máxima verossimilhança de θ é boa e nada sobre suas propriedades. Se a amostra é grande, existem teoremas que garantem propriedades ótimas para os EMV. Algumas dessas propriedades serão apresentadas a seguir.

Eficiência

No espírito do limite inferior de Cramér-Rao (teorema 1.5.1), existe uma variância assintótica ótima.

Definição 1.5.4. *Uma sequência de estimadores* W_n *é assintoticamente eficiente para um parâmetro* $\tau(\theta)$ *se* $\sqrt{n}\left[W_n - \tau(\theta)\right] \to N(0, v(\theta))$ *em distribuição e*

$$v(\theta) = \frac{\left(\tau'(\theta)\right)^2}{E_\theta\left[\left(\frac{d}{d\theta}\log\left(f(X|\theta)\right)\right)^2\right]}$$

ou seja, a variância assintótica de W_n *alcança o limite inferior da cota de Cramér-Rao.*

A seguir, serão apresentados dois teoremas que afirmarão de maneira geral sobre a consistência e eficiência assintótica dos estimadores de máxima verossimilhança.

Teorema 1.5.2. ***(Consistência dos EMV)*** *Seja* $X_1, \ldots, X_n$ *uma amostra iid com* $f(x|\theta)$ *e seja* $L(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta)$ *a função de verossimilhança. Seja* $\hat{\theta}$ *o EMV de* θ. *Seja* $\tau(\theta)$ *uma função contínua de* θ. *Para todo* $\varepsilon > 0$ *e* $\theta \in \Theta$,

$$\lim_{n\to\infty} P_\theta\left(\left|\tau\left(\hat{\theta}\right) - \tau(\theta)\right| \geq \varepsilon\right) = 0 \tag{1.29}$$

ou seja, $\tau\left(\hat{\theta}\right)$ *é um estimador consistente de* $\tau(\theta)$

Teorema 1.5.3. ***(Eficiência assintótica dos EMV)*** *Seja $X_1, \ldots, X_n$ uma amostra iid com $f(x|\theta)$ e seja $L(\theta|\mathbf{x}) = \prod_{i=1}^{n} f(x_i|\theta)$ a função de verossimilhança. Seja $\hat{\theta}$ o EMV de θ. Seja $\tau(\theta)$ uma função contínua de θ. Então*

$$\sqrt{n}\left[\tau\left(\hat{\theta}\right) - \tau(\theta)\right] \to N(0, v(\theta)) \tag{1.30}$$

em que $v(\theta)$ é o limite inferior da cota de Cramér-Rao. Ou seja, $\tau\left(\hat{\theta}\right)$ é um estimador consistente e assintoticamente eficiente de $\tau(\theta)$

Demonstração. Seja $l(\theta|\mathbf{x}) = \log(L(\theta|\mathbf{x})) = \sum_{i=1}^{n} \log(f(x_i;|\theta))$. Suponha que θ_0 seja o valor real do parâmetro.

Vamos expandir $l'(\theta|\mathbf{x})$ em série de Taylor centrada em θ_0, daí

$$l'(\theta|\mathbf{x}) \approx l'(\theta_0|\mathbf{x}) + l''(\theta_0|\mathbf{x})(\theta - \theta_0) + \ldots$$

tomando $\theta = \hat{\theta}$, que é o estimador de máxima verossimilhança de θ, segue que

$$l'\left(\hat{\theta}|\mathbf{x}\right) \approx l'(\theta_0|\mathbf{x}) + l''(\theta_0|\mathbf{x})\left(\hat{\theta} - \theta_0\right) + \ldots.$$

Note que $l'\left(\hat{\theta}|\mathbf{x}\right) = 0$. Logo,

$$\Rightarrow \hat{\theta} - \theta_0 \approx \frac{-l'(\theta_0|\mathbf{x})}{l''(\theta_0|\mathbf{x})}$$

$$\Rightarrow \sqrt{n}\left(\hat{\theta} - \theta_0\right) \approx \frac{-\sqrt{n}\, l'(\theta_0|\mathbf{x})}{l''(\theta_0|\mathbf{x})}$$

$$\Rightarrow \sqrt{n}\left(\hat{\theta} - \theta_0\right) \approx \frac{\frac{1}{\sqrt{n}} l'(\theta_0|\mathbf{x})}{-\frac{1}{n} l''(\theta_0|\mathbf{x})}$$

como $l(\theta_0|\mathbf{x}) = \sum_{i=1}^{n} \log(f(x_i;|\theta))$ e $l'(\theta_0|\mathbf{x}) = \sum_{i=1}^{n} \frac{d}{d\theta} \log(f(x_i;|\theta))$, temos

$$\frac{1}{n} l'(\theta_0|\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \frac{d}{d\theta} \log(f(x_i;|\theta))$$

em que $\sum_{i=1}^{n} \frac{d}{d\theta} \log(f(x_i;|\theta))$ é uma v.a. com esperança zero e, como mostrado anteriormente, variância $I(\theta_0)$ $\left(E_{\theta_0}\left[\left(\frac{d}{d\theta} \log(f(x_i;|\theta))\right)^2\right] = I(\theta_0)\right)$.

Pelo TLC, segue que

$$\sqrt{n}\left(\frac{1}{n} l'(\theta_0|X_1, \ldots, X_n)\right) = \frac{1}{\sqrt{n}} l'(\theta_0|X_1, \ldots, X_n) \to N(0, I(\theta_0))$$

Por outro lado, considerando o termo $l''(\theta_0|\mathbf{x})$ temos que

$$l''(\theta_0|\mathbf{x}) = \sum_{i=1}^{n} \frac{d^2}{d\theta^2} \log(f(x_i;|\theta))$$

$$\frac{1}{n} l''(\theta_0|\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} \frac{d^2}{d\theta^2} \log(f(x_i;|\theta)).$$

Pela Lei Fraca dos Grandes Números ocorre que

$$-\frac{1}{n} l''(\theta_0|X_1, \ldots, X) \to -E_{\theta_0}\left[\frac{d^2}{d\theta^2} \log(f(x_i;|\theta))\right]$$

ou seja, o primeiro membro converge em probabilidade para o segundo membro. Mas

$$-E_{\theta_0}\left[\frac{d^2}{d\theta^2} \log(f(X_i;|\theta))\right] = E_{\theta_0}\left[\left(\frac{d}{d\theta} \log(f(X_i;|\theta))\right)^2\right] = I(\theta_0)$$

Portanto, retornando na expressão acima, considerando uma amostra suficientemente grande e utilizando o teorema de Slutsky temos que

$$\Rightarrow \lim_{n\to\infty} \sqrt{n}\left(\hat{\theta}-\theta_0\right) = \lim_{n\to\infty} \frac{\frac{1}{\sqrt{n}} l'\left(\theta_0\left|\mathbf{x}\right.\right)}{-\frac{1}{n} l''\left(\theta_0\left|\mathbf{x}\right.\right)} \to \frac{N\left(0, I\left(\theta_0\right)\right)}{I\left(\theta_0\right)} = N\left(0, \frac{1}{I\left(\theta_0\right)}\right)$$

ou seja, $\sqrt{n}\left(\hat{\theta}-\theta_0\right)$ converge em distribuição para uma $N\left(0, \frac{1}{I(\theta_0)}\right)$, provando o teorema. □

Cálculos e Comparações

Se um estimador de máxima verssimilhança é assintoticamente eficiente, a variância assintótica no teorema 1.5.2 é a variância do Método Delta do teorema 1.4.8 (sem o termo $\frac{1}{n}$). Então, podemos usar o limite inferior da cota de Cramér-Rao como uma aproximação para a verdadeira variância do EMV. Suponha que $X_1, \ldots, X_n$ seja uma amostra iid de $f\left(x\left|\theta\right.\right)$, $\hat{\theta}$ é uma estimativa de máxima verossimilhança de θ, e $I_n\left(\theta\right) = E_\theta\left[\left(\frac{d}{d\theta}\log\left(L\left(\theta\left|\mathbf{X}\right.\right)\right)\right)^2\right]$ é o número de informação da amostra. Assim,

$$\begin{aligned}
\sqrt{n}\left[W_n - \tau\left(\theta\right)\right] &\to N\left(0,v\left(\theta\right)\right)\\
W_n - \tau\left(\theta\right) &\to N\left(0,\frac{v\left(\theta\right)}{n}\right)\\
W_n &\to N\left(\tau\left(\theta\right),\frac{v\left(\theta\right)}{n}\right)
\end{aligned}$$

se $\hat{\theta} = \tau\left(\hat{\theta}\right)$ é uma EMV de θ, imlplica que $\hat{\theta}$ é solução de $\frac{d}{d\theta}\log\left(\prod_{i=1}^n f\left(x_i;\theta\right)\right) = 0$ e $\tau\left(\hat{\theta}\right)$ e EMV de $\tau\left(\theta\right)$. Daí,

$$\sqrt{n}\left[\tau\left(\hat{\theta}\right) - \tau\left(\theta\right)\right] \to N\left(0,v\left(\theta\right)\right)$$

converge em distribuição conforme $n \to \infty$.

Do Método Delta e da eficiência assintótica dos EMV, a variância de $\tau\left(\hat{\theta}\right)$ pode ser aproximada por

$$\begin{aligned}
Var\left(\tau\left(\hat{\theta}\right)\right) \approx \frac{v\left(\theta\right)}{n} &= \frac{\left(\tau'\left(\theta\right)\right)^2}{nE_\theta\left[\left(\frac{d}{d\theta}\log\left(f\left(\mathbf{X}\left|\theta\right.\right)\right)\right)^2\right]}\\
&= \frac{\left(\tau'\left(\theta\right)\right)^2}{-nE_\theta\left[\frac{d^2}{d\theta^2}\log\left(f\left(\mathbf{X};\left|\theta\right.\right)\right)\right]}.
\end{aligned}$$

Sabemos que

$$E_\theta\left[\left(\frac{d}{d\theta}\log\left(\prod_{i=1}^n f\left(X_i\left|\theta\right.\right)\right)\right)^2\right] = nE_\theta\left[\left(\frac{d}{d\theta}\log\left(f\left(X\left|\theta\right.\right)\right)\right)^2\right].$$

Logo,

$$\begin{aligned}
Var\left(\tau\left(\hat{\theta}\right)\right) = Var\left(\tau\left(\hat{\theta}\right)\left|\theta\right.\right) &\approx \frac{\left(\tau'\left(\theta\right)\right)^2}{E_\theta\left[\left(\frac{d}{d\theta}\log\left(\prod_{i=1}^n f\left(X_i\left|\theta\right.\right)\right)\right)^2\right]}\\
&= \frac{\left(\tau'\left(\theta\right)\right)^2}{-E_\theta\left[\frac{d^2}{d\theta^2}\log\left(\prod_{i=1}^n f\left(X_i\left|\theta\right.\right)\right)\right]}
\end{aligned}$$

cujo denominador da última expressão pode ser expresso por

$$-nE_\theta\left[\frac{d^2}{d\theta^2}\log\left(f\left(X\left|\theta\right.\right)\right)\right] = -E_\theta\left[\frac{d^2}{d\theta^2}\log\left(\prod_{i=1}^n f\left(X_i\left|\theta\right.\right)\right)\right]$$

$$= -nE_\theta\left[\frac{1}{n}\sum_{i=1}^{n}\frac{d^2}{d\theta^2}\log\left(f\left(X_i\left|\theta\right.\right)\right)\right].$$

Temos então que

$$E_\theta\left[\frac{d^2}{d\theta^2}\log\left(f\left(X\left|\theta\right.\right)\right)\right] = E_\theta\left[\frac{1}{n}\sum_{i=1}^{n}\frac{d^2}{d\theta^2}\log\left(f\left(X_i\left|\theta\right.\right)\right)\right] \approx \frac{1}{n}\sum_{i=1}^{n}\frac{d^2}{d\theta^2}\log\left(f\left(X_i\left|\theta\right.\right)\right).$$

Assim, voltando na expressão inicial acima,

$$\begin{aligned}
Var\left(\tau\left(\hat{\theta}\right)\right) &\approx \frac{\left(\tau'\left(\theta\right)\right)^2}{-nE_\theta\left[\frac{d^2}{d\theta^2}\log\left(f\left(X\left|\theta\right.\right)\right)\right]}\\
&= \frac{\left(\tau'\left(\theta\right)\right)^2}{-n\left[\frac{1}{n}\sum\limits_{i=1}^{n}\frac{d^2}{d\theta^2}\log\left(f\left(X_i\left|\theta\right.\right)\right)\right]}\\
&= \frac{\left(\tau'\left(\theta\right)\right)^2}{-\sum\limits_{i=1}^{n}\frac{d^2}{d\theta^2}\log\left(f\left(X_i\left|\theta\right.\right)\right)}\\
&= \frac{\left(\tau'\left(\theta\right)\right)^2}{-\frac{d^2}{d\theta^2}\log\left(\prod_{i=1}^{n}f\left(X_i\left|\theta\right.\right)\right)}\\
&\approx \frac{\left(\tau'\left(\theta\right)\right)^2}{-\frac{d^2}{d\theta^2}\log\left(\prod_{i=1}^{n}f\left(X_i\left|\hat{\theta}\right.\right)\right)}\\
\Rightarrow Var\left(\tau\left(\hat{\theta}\right)\right) &\approx \frac{\left(\tau'\left(\theta\right)\right)^2}{-\frac{d^2}{d\theta^2}\log\left(\prod_{i=1}^{n}f\left(X_i\left|\hat{\theta}\right.\right)\right)}.
\end{aligned}$$

Portanto, para estimar $Var\left(\tau\left(\hat{\theta}\right)\right)$, primeiro aproximamos $Var\left(\tau\left(\hat{\theta}\right)\right)$. Em seguida estimamos a aproximação resultante, usualmente substituindo $\hat{\theta}$ por θ, e o resultado pode ser denotado por $\hat{V}ar\left(\tau\left(\hat{\theta}\right)\right)$.

Segue do teorema 1.5.2 que $\frac{1}{n}\frac{d^2}{d\theta^2}\log\left(L\left(\hat{\theta}\left|\mathbf{X}\right.\right)\right)$ é um estimador consistente de $I\left(\theta\right)$, então segue que $\hat{V}ar\left(\tau\left(\hat{\theta}\right)\right)$ é um estimador consistente de $Var\left(\tau\left(\hat{\theta}\right)\right)$.

Exemplo 1.5.0.17. ***(Aproximação da variância da Binomial)*** *Seja $X_1, \ldots, X_n$ uma amostra iid de uma pop. Bernoulli(p). Sabe-se que o EMV de p é dado por $\hat{p} = \sum_{i=1}^{n} X_i/n$. Encontre a variância assintótica de $\hat{p}$ e mostre sua convergência.*

SOLUÇÃO

A variância para a média de uma população Bernoulli é dada por $Var_p\left[\bar{X}\right] = \frac{p(1-p)}{n}$. É razoável estimar $Var_p\left[\bar{X}\right]$ por

$$\hat{V}ar_p\left[\hat{p}\right] = \hat{V}ar_p\left[\bar{X}\right] = \frac{\hat{p}\left(1-\hat{p}\right)}{n} = \frac{\bar{X}\left(1-\bar{X}\right)}{n}$$

A estimativa obtida por construção assintótica é

$$\begin{aligned}
Var_p\left[\hat{p}\right] &= \frac{\left(\tau'\left(\theta\right)\right)^2}{-\frac{d^2}{d\theta^2}\log\left(\prod_{i=1}^{n}f\left(X_i\left|\hat{p}\right.\right)\right)}\\
&= \frac{1}{-\frac{d^2}{d\theta^2}\log\left(\hat{p}^{\sum_{i=1}^{n}X_i}\left(1-\hat{p}\right)^{n-\sum_{i=1}^{n}X_i}\right)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{-\frac{d^2}{d\theta^2}\sum_{i=1}^{n} X_i \log(\hat{p}) + \left(n - \sum_{i=1}^{n} X_i\right)\log(1-\hat{p})} \\
&= \frac{1}{-\left[-\frac{\sum_{i=1}^{n} X_i}{\hat{p}^2} - \frac{n-\sum_{i=1}^{n} X_i}{(1-\hat{p}^2)}\right]} \\
&= \frac{1}{n\left[-\frac{\bar{X}}{\hat{p}^2} + \frac{1-\bar{X}}{(1-\hat{p}^2)}\right]} \\
&= \frac{1}{n\left[-\frac{\bar{X}}{\bar{X}^2} + \frac{1-\bar{X}}{(1-\bar{X}^2)}\right]} = \frac{\bar{X}_n\left(1-\bar{X}_n\right)}{n}
\end{aligned}
$$

que coincide com o estimador razoável. Agora vamos verificar a eficiência assintótica. Sabemos que a quantidade

$$\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \to N(0,1)$$

converge em distribuição para uma normal padrão. Porém essa expressão está e termos da variância populacional de p. Podemos obter a variância assintótica que ficará em termos de $\hat{p}$. Pelo teorema 1.5.3 temos que

$$\sqrt{n}\,(\hat{p}-p) \to N\left(0, \underbrace{p(1-p)}_{C.R}\right)$$

converge em distribuição para uma normal. Aplicando o teorema de Slutsky (1.4.7), podemos concluir que

$$\sqrt{n}\frac{\hat{p}-p}{\hat{p}(1-\hat{p})} \to N(0,1)$$

Para complementar esse exercício, poderíamos estar interessados em encontrar a variância assintótica para a chance (odds). Em seções anteriores, encontramos a variância para a chance pelo método Delta, agora vamos fazê-lo utilizando o teorema 1.5.3. Vamos fazer um plug-in *para obter um estimador razoável da chance, ou seja, apenas substituir a versão populacional da chance por sua versão via EMV. Então, considerando $\frac{p}{1-p} = \frac{\hat{p}}{1-\hat{p}}$,*

$$
\begin{aligned}
\hat{V}ar_p\left[\frac{\hat{p}}{1-\hat{p}}\right] &= \frac{\left(\frac{d}{d\hat{p}}\left(\frac{\hat{p}}{1-\hat{p}}\right)\right)^2}{-\frac{d^2}{d\theta^2}\log\left(\hat{p}^{\sum_{i=1}^{n} X_i}(1-\hat{p})^{n-\sum_{i=1}^{n} X_i}\right)} \\
&= \frac{\hat{p}}{n(1-\hat{p})^3}
\end{aligned}
$$

Portanto, pelo teorema 1.5.3, temos que

$$\sqrt{n}\left(\frac{\hat{p}}{1-\hat{p}} - \frac{p}{1-p}\right) \to N\left(0, \frac{\hat{p}}{n(1-\hat{p})^3}\right)$$

e o estimador é assintoticamente eficiente e é igual ao obtido pelo método Delta.

A aproximação para a variância dos EMV funciona bem em muitos casos, mas não é infalível. Deve-se ter cuidado quando a função $\tau\left(\hat{\theta}\right)$ não é monótona. Em alguns casos, a derivada $\tau'(\theta)$ poderá term uma mudança de sinal, o que pode levar a uma aproximação subestimada da variância.

Bootstrap

O *bootstrap* pode ser utilizado em muitas situações. Nesta seção será abordado sua capacidade de promove meios alternativos de calcular erros padrões.

O *bootstrap* é baseado em uma simples, porém poderosa, idéia (cuja matemática pode estar bastante envolvida). Em estatística, aprendemos sobre características da população tomando amostras. Como a amostra representa a população, características análogas da amostra deve nos dar informações sobre as características da população. O *bootstrap* nos ajuda a aprender sobre essas características da amostra tomando reamostras (ou seja, tomamos amostras da amostra original) e usamos essa informação para inferir sobre a população. O *bootstrap* foi desenvolvido por Efron na década de 70.

Exemplo 1.5.0.18. ***(Variância Bootstrap)*** *Suponha que estamos interessados em calcular todas as possíveis médias de quatro números selecionados de* $\{2{,}4{,}9{,}12\}$*, onde forma tirados números com reposição. Por exemplo, possíveis retiradas são* $\{2{,}4{,}4{,}12\}$*, cuja média é 4,75, e* $\{2{,}4{,}9{,}12\}$*, cuja média é 6,5. Se estamos interessados apenas na média dos números amostrados, a ordem não é importante e, portanto, o número total de amostras distintas é obtida pela contagem das amostras com reposição não ordenadas.*

O número total de amostras distintas é $\binom{n+n-1}{n}$*. Mas agora, para a distribuição de probabilidade das médias amostrais, devemos contar os diferentes modos que uma particular média pode ocorrer.*

O valor 4,75 pode ocorrer somente se amostra contém um 2, dois 4 e um 9. O número de possíveis amostras que tem essa configuração é dado por uma permutação com elementos repetidos, ou seja, $P_4^2 = \frac{4!}{2!} = 12$.

As amostras de $\{2{,}4{,}9{,}12\}$ *com reposição, configura a forma mais simples do bootstrap, algumas vezes chamado de bootstrap não paramétrico.*

O que fizemos anteriormente foi criar uma reamostra dos possíveis valores da média amostral. Dizemos que existem $\binom{4+4-1}{4} = 35$ *possíveis valores distintos, mas esses valores não são equiprováveis (e então não podemos tratar as amostras como sendo aleatórias). As* $n^n = 4^4 = 256$ *reamostras (não distintas) são todas igualmente prováveis e podem ser tratadas como amostras aleatórias.*

Para a i-ésima reamostra, seja $\bar{x}_i^*$ *a média da reamostra. Podemos estimar a variância da média amostral* $\bar{X}$ *por*

$$Var^*\left[\bar{X}\right] = \frac{1}{n^n-1}\sum_{i=1}^{n^n}\left(\bar{x}_i^* - \bar{x}^*\right)^2 \tag{1.31}$$

em que

$$\bar{x}^* = \frac{1}{n^n}\sum_{i=1}^{n^n}\bar{x}_i^*$$

é a média da reamostra. A notação $*$ *refere-se a um valor proveniente do método bootstrap, ou reamostrado.*

Para o exemplo, a média e variância bootstrap são $\bar{x}^* = 6{,}75$ *e* $Var^*\left[\bar{X}\right] = 3{,}94$*, respectivamente.*

Vimos no exemplo anterior como calcular o erro padrão *bootstrap*, mas em um problema não seria realmente necessário. Contudo, a vantagem do método é que, assim como o método Delta, a fórmula da variância 1.31 a qualquer estimador. Portanto, para qualquer estimador $\hat{\theta}(\mathbf{x}) = \hat{\theta}$, temos que

$$Var^*\left[\hat{\theta}\right] = \frac{1}{n^n-1}\sum_{i=1}^{n^n}\left(\hat{\theta}_i^* - \bar{\theta}^*\right)^2 \tag{1.32}$$

em que $\hat{\theta}_i^*$ é o estimador calculado da i-ésima reamostra e

$$\bar{\theta}^* = \frac{1}{n^n}\sum_{i=1}^{n^n}\hat{\theta}_i^* \tag{1.33}$$

a média dos valores reamostrados.

Mas agora um problema surge. Para o exemplo anterior, com $n = 4$, existem 256 termos na soma *bootstrap*. Em muitos casos, a amostra é muito grande e o problema se torna intratável computacionalmente (com $n > 15$ o problema já começa a aparecer). Então, ainda existe um problema estatístico, que é o fato de utilizar amostras de reamostras.

Então, para uma amostra $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ e uma estimativa $\hat{\theta}(\mathbf{x}) = \hat{\theta}$, escolha B reamostras (ou *amostras bootstrap*) e calcule

$$Var_B^*\left[\hat{\theta}\right] = \frac{1}{B-1}\sum_{i=1}^{B}\left(\hat{\theta}_i^* - \bar{\theta}^*\right)^2 \tag{1.34}$$

Assim, os seguintes fatos podem ser comprovados:

- $Var_B^*\left[\hat{\theta}\right] \to Var^*\left[\hat{\theta}\right]$, conforme $B \to \infty$, ou seja, a variância *bootstrap* restrita se aproxima da variância *bootstrap*;
- $Var_n^*\left[\hat{\theta}\right] \to Var\left[\hat{\theta}\right]$, conforme $n \to \infty$, ou seja, a variância *bootstrap* se aproxima da variância do estimador (variância teórica).

O tipo de bootstrap mencionado até então é chmado de bootstrap não paramétrico, uma vez que nenhuma forma funcional para pdf ou cdf populacional foi estabelecida. Por outro lado, temos também o bootstrap paramétrico.

Suponha que temos uma amostra $X_1, \ldots, X_n$ de uma distribuição com pdf $f(x|\theta)$, cujo θ pode ser um vetor de parâmetros. Podemos estimar θ com $\hat{\theta}$, os EMV, e tirar amostras

$$X_1^*, X_2^*, \ldots, X_n^* \sim f\left(x\middle|\hat{\theta}\right).$$

Se tomarmos B como tais amostras, podemos estimar a variância de $\hat{\theta}$ usando 1.34. Note que essas amostras não são reamostradas dos dados, mas amostras aleatórias reais de $f\left(x\middle|\hat{\theta}\right)$, que algumas vezes é chamada de *distribuição plug-in*.

1.6 Teste de Hipóteses

Essa seção descreverá alguns métodos para se obter um teste em problemas complicados. Estamos pensando em problemas nos quais nenhum teste ótimo existe ou é conhecido. Em situações do tipo, a obtenção de qualquer teste razoável poderia ser útil.

Distribuição Assintótica do teste de Razão de Verossimilhança (TRV)

Um dos mais importantes métodos para modelos complicados é a construção de teste pelo método da razão de verossimilhança por causa de sua explícita definição do teste estatístico,

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

e uma explícita forma para a região de rejeição, $\{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$. Depois que os dados $\mathbf{X} = \mathbf{x}$ são observados, a função de verossimilhança, $L(\theta|\mathbf{x})$, é uma função da variável θ completamente definida. Até mesmo se os dois supremos de $L(\theta|\mathbf{x})$, sobre os conjuntos Θ_0 e Θ, não podem ser obtidos analiticamente, podem ser obtidos numericamente. Então, a estatística do teste $\lambda(\mathbf{x})$ pode ser determinada para as observações mesmo se nenhuma fórmula que defina $\lambda(\mathbf{x})$ for encontrada.

Para definir um nível α do teste, a constante c deve ser escolhida de tal forma que

$$\sup_{\theta\in\Theta_0} P_\theta\left[\lambda(\mathbf{x}) \leq c\right] \leq \alpha. \tag{1.35}$$

Se não pudermos obter uma simples fórmula para $\lambda(\mathbf{x})$, poderia parecer desesperançoso a obtenção da distribuição amostral de $\lambda(\mathbf{x})$ e então conhecer qual c deve-se ter para garantir 1.35. Contudo, se apelarmos para métodos assintóticos, podemos obter uma resposta aproximada.

Analogamente ao teorema 1.5.3, temos o seguinte resultado.

Teorema 1.6.1. ***(Distribuição Assintótica do TRV - H_0 simples)*** *Para testar $H_0 : \theta = \theta_0$ contra $H_1 : \theta \neq \theta_0$, suponha $X_1, \ldots, X_n$ são iid com $f(x|\theta)$ e $\hat{\theta}$ é o EMV de θ. Então, sob H_0, conforme $n \to \infty$,*

$$-2\log(\lambda(\mathbf{x})) \to \chi_1^2$$

em distribuição, em que χ_1^2 é uma variável aleatória χ^2 com 1 grau de liberdade.

Demonstração. Seja $\hat{\theta}$ o EMV de θ e $l(\theta|\mathbf{X}) = \log(L(\theta|\mathbf{X}))$ a log-verossimilhança. Vamos expandir $l(\theta|\mathbf{X})$ em séries de Taylor em torno de $\hat{\theta}$, daí

$$l(\theta|\mathbf{X}) \approx l\left(\hat{\theta}|\mathbf{X}\right) + \underbrace{l'\left(\hat{\theta}|\mathbf{X}\right)}_{=0,\, pois\, \hat{\theta}=\hat{\theta}_{MV}} \left(\theta - \hat{\theta}\right) + \frac{l''\left(\hat{\theta}|\mathbf{X}\right)\left(\theta - \hat{\theta}\right)^2}{2}$$

$$\begin{aligned}
l(\theta|\mathbf{X}) - l\left(\hat{\theta}|\mathbf{X}\right) &\approx \frac{l''\left(\hat{\theta}|\mathbf{X}\right)\left(\theta - \hat{\theta}\right)^2}{2} \\
\log\left(\frac{L(\theta|\mathbf{X})}{L\left(\hat{\theta}|\mathbf{X}\right)}\right) &\approx \frac{1}{2}\left(\frac{d^2}{d\theta^2}\log\left(\prod_{i=1}^n f(X_i|\theta)\right)\right)_{\theta=\hat{\theta}} \left(\theta - \hat{\theta}\right)^2 \\
&= \frac{1}{2}\left(\frac{d^2}{d\theta^2}\sum_{i=1}^n \log\left(f\left(X_i\middle|\hat{\theta}\right)\right)\right)\left(\theta - \hat{\theta}\right)^2 \\
&= \frac{1}{2}\left(\underbrace{\sum_{i=1}^n \frac{d^2}{d\theta^2}\log\left(f\left(X_i\middle|\hat{\theta}\right)\right)}_{*}\right)\left(\theta - \hat{\theta}\right)^2
\end{aligned}$$

resolvendo $*$, temos que

$$* = n\left(\frac{1}{n}\sum_{i=1}^n \frac{d^2}{d\theta^2}\log\left(f\left(X_i\middle|\hat{\theta}\right)\right)\right) = \underbrace{\frac{1}{n}\sum_{i=1}^n \frac{d^2}{d\theta^2}\log\left(f\left(X_i\middle|\hat{\theta}\right)\right)}_{\to E_\theta\left[\frac{d^2}{d\theta^2}\log\left(f\left(X_i|\hat{\theta}\right)\right)\right]}$$

Logo

$$\begin{aligned}
\log\left(\frac{L(\theta_0|\mathbf{X})}{L\left(\hat{\theta}|\mathbf{X}\right)}\right) &= \frac{1}{2}\frac{\left(\theta - \hat{\theta}\right)^2}{\frac{1}{\sum_{i=1}^n \frac{d^2}{d\theta^2}\log\left(f\left(X_i|\hat{\theta}\right)\right)}} \\
&= \frac{1}{2}\frac{\left(\theta - \hat{\theta}\right)^2}{\frac{1}{n\frac{1}{n}\sum_{i=1}^n \frac{d^2}{d\theta^2}\log\left(f\left(X_i|\hat{\theta}\right)\right)}} \\
&= \frac{1}{2}\frac{\left(\sqrt{n}\left(\hat{\theta} - \theta\right)\right)^2}{\frac{1}{\frac{1}{n}\sum_{i=1}^n \frac{d^2}{d\theta^2}\log\left(f\left(X_i|\hat{\theta}\right)\right)}}
\end{aligned}$$

Temos o teorema

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \to N(0, v(\theta_0)) = N\left(0, \frac{1}{-E_\theta\left[\frac{d^2}{d\theta^2}\log\left(f\left(X_i\middle|\hat{\theta}\right)\right)\right]}\right)$$

em distribuição,

$$\frac{\sqrt{n}\left(\hat{\theta} - \theta_0\right)}{-E_\theta\left[\frac{d^2}{d\theta^2}\log\left(f\left(X_i\middle|\hat{\theta}\right)\right)\right]} \to N(0,1)$$

$$\frac{\left(\sqrt{n}\left(\hat{\theta}-\theta_0\right)\right)^2}{-E_\theta\left[\frac{d^2}{d\theta^2}\log\left(f\left(X_i\left|\hat{\theta}\right.\right)\right)\right]} \to \chi_1^2$$

em distribuição. Então,

$$\Rightarrow \frac{1}{2}\underbrace{\frac{\left(\sqrt{n}\left(\hat{\theta}-\theta_0\right)\right)^2}{-E_\theta\left[\frac{d^2}{d\theta^2}\log\left(f\left(X_i\left|\hat{\theta}\right.\right)\right)\right]}}_{\to\chi_1^2}\frac{1}{\underbrace{\frac{\frac{1}{n}\sum_{i=1}^n\frac{d^2}{d\theta^2}\log\left(f\left(X_i\left|\hat{\theta}\right.\right)\right)}{-E_\theta\left[\frac{d^2}{d\theta^2}\log\left(f\left(X_i\left|\hat{\theta}\right.\right)\right)\right]}}_{\to-1}}$$

Portanto

$$-2\log\left(\lambda\left(\mathbf{x}\right)\right) \to \chi_1^2$$

□

Exemplo 1.6.0.19. ***(TRV para Poisson)*** *Construa o TRV para testar $H_0:\lambda=\lambda_0$ contra $H_1:\lambda\neq\lambda_0$ baseado nas observações $X_1,\ldots,X_n$ iid Poisson (λ).*

SOLUÇÃO

A verossimilhança é dada por

$$L\left(\theta_0\left|\mathbf{X}\right.\right)=\prod_{i=1}^n\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}=\frac{\lambda^{\sum_{i=1}^n x_i}e^{-n\lambda}}{\prod_{i=1}^n x_i!}$$

o TRV fica

$$\begin{aligned}\Rightarrow -2\log\left(\lambda\left(\mathbf{x}\right)\right) &= -2\log\left[\frac{\frac{\lambda_0^{\sum_{i=1}^n x_i}e^{-n\lambda_0}}{\prod_{i=1}^n x_i!}}{\frac{\hat{\lambda}^{\sum_{i=1}^n x_i}e^{-n\hat{\lambda}}}{\prod_{i=1}^n x_i!}}\right]\\
&= -2\log\left(\lambda_0^{\sum_{i=1}^n x_i}e^{-n\lambda_0}\right)+2\log\left(\hat{\lambda}^{\sum_{i=1}^n x_i}e^{-n\hat{\lambda}}\right)\\
&= -2\left[\sum_{i=1}^n x_i\log\left(\lambda_0\right)-n\lambda_0\right]+2\left[\sum_{i=1}^n x_i\log\left(\hat{\lambda}\right)-n\hat{\lambda}\right]\\
&= 2\left[\sum_{i=1}^n x_i\log\left(\frac{\hat{\lambda}}{\lambda_0}\right)+n\left(-\hat{\lambda}+\lambda_0\right)\right]\\
&= 2n\left[\hat{\lambda}\log\left(\frac{\hat{\lambda}}{\lambda_0}\right)+\left(\lambda_0-\hat{\lambda}\right)\right].\end{aligned}$$

Portanto o teste é: rejeito H_0 se

$$\begin{aligned}\lambda\left(\mathbf{x}\right)<\lambda_0 &\Leftrightarrow \log\left(\lambda\left(\mathbf{x}\right)\right)<\log\left(\lambda_0\right)\\
&\Leftrightarrow -2\log\left(\lambda\left(\mathbf{x}\right)\right)>-2\log\left(\lambda_0\right)\\
&\Leftrightarrow -2\log\left(\lambda\left(\mathbf{x}\right)\right)>\chi_{1;\alpha}^2\end{aligned}$$

em que α é o nível de significância do teste.

O teorema 1.6.1 pode ser extendido para casos em que a hipótese nula envolve um vetor de parâmetros. O teorema a seguir, que apenas apresentaremos sem prová-lo, nos permite verificar que 1.6.1 é verdade, pelo menos para amostras grandes.

Teorema 1.6.2. *Seja $X_1, \ldots, X_n$ uma amostra aleatória de uma fdp $f(x\,|\theta)$. Se $\theta \in \Theta_0$, então a distribuição da estatística $-2\log(\lambda(\mathbf{x}))$ converge para uma distribuição qui-quadrado conforme o tamanho da amostra $n \to \infty$. Os graus de liberdade da distribuição limite é a disferença entre o número de parâmetros livres especificado por $\theta \in \Theta_0$ e o número de parâmetros livres especificados por $\theta \in \Theta$*

A região de rejeição de H_0 é para $\theta \in \Theta_0$ com pequenos velores de $\lambda(\mathbf{x})$, que é equivalente a rejeição para grandes valores de $-2\log(\lambda(\mathbf{x}))$. Então

$$\text{rejeitar}\, H_0 \Leftrightarrow -2\log(\lambda(\mathbf{x})) \geq \chi^2_{v;\,\alpha}$$

em que v são os graus de liberdade especifiados no teorema 1.6.2. A probabililidade de erro tipo I será aproximadamente α se $\theta \in \Theta_0$ e o tamanho da amostra é grande. Desse modo, a expressão 1.35 será aproximadamente satisfeita para grandes amostras e um teste assintótico de tamanho α será definido. Note que o teorema implicará realmente que

$$\lim_{n\to\infty} P_\theta[\text{rejeitar}\, H_0] = \alpha, \text{para cada}\, \theta \in \Theta_0,$$

não que o $\sup_{\theta\in\Theta_0} P_\theta[\text{rejeitar}\, H_0]$ converge para α. Isso é tipicamente o caso para testes assintóticos de tamanho α.

O cálculo dos graus de liberdade para o teste estatístico é em geral direto. Muitas vezes, Θ pode ser representado como um subconjunto de dimensão q de um espaço Euclidiano que contém um subconjunto aberto em $\mathbb{R}^q$, e Θ_0 pode ser representado como um subconjunto de dimensão p do espaço Euclidiano que contém um subconjunto aberto em $\mathbb{R}^p$, em que $p < q$. Então $q - p = v$ são os graus de liberdade para o teste estatístico.

Exemplo 1.6.0.20. ***(TRV Multinomial)*** *Seja $\theta = (p_1,p_2,p_3,p_4,p_5)$, onde os $p_j's$ são não negativos e somam 1. Suponha que $X_1, \ldots, X_n$ são v.a. iid discretas e $P_\theta[X_i = j] = p_{ij}$, $j = 1, \ldots, 5$. Então a fdp de X_i é $f(j\,|\theta) = p_j$ e a função de verossimilhança é*

$$L(\theta\,|\mathbf{x}) = \prod_{i=1}^{n} f(x_i\,|\theta) = p_1^{y_1} p_2^{y_2} p_3^{y_3} p_4^{y_4} p_5^{y5}$$

em que y_j é igual ao número de $x_1, \ldots, x_n$ igual a j. Testar as hipóteses:

$$\begin{cases} H_0 : p_1 = p_2 = p_3 \,e\, p_4 = p_5 \\ H_1 : H_0 \, falso \end{cases}$$

SOLUÇÃO

O espaço paramétrico completo, Θ, é um conjunto real de dimensão quatro. Uma vez que $p_5 = 1 - p_1 - p_2 - p_3 - p_4$, existem somente quatro parâmetros livres. O conjunto dos parâmetros é definido por

$$\sum_{j=1}^{4} p_j \leq 1 \,e\, p_j \geq 0, j = 1, \ldots, 4,$$

um subconjunto de $\mathbb{R}^4$ contendo um subconjunto aberto de $\mathbb{R}^4$. Então $q = 4$. Existe apenas um parâmetro livre no conjunto especificado por H_0 porque, considerando que p_1, $0 \leq p_1 \leq \frac{1}{3}$, está fixado, $p_2 = p_3$ deve ser igual a p_1 e $p_4 = p_5$ deve ser igual a

$$p_5 = 1 - p_1 - p_2 - p_3 - p_4 \Rightarrow p_5 = 1 - 3p_1 - p_5 \Rightarrow p_5 = \frac{1 - 3p_1}{2}.$$

Então $p = 1$, e os graus de liberdade são $v = 4 - 1 = 3$.

Para calcular $\lambda(\mathbf{x})$, o EMV de θ sobre ambos Θ_0 e Θ devem ser determinados. Definindo

$$\frac{\partial}{\partial p_j} \log(L(\theta\,|\mathbf{x})) = 0, \text{para cada}\, j = 1, \ldots, 4,$$

e usando os fatos de que $p_5 = 1 - p_1 - p_2 - p_3 - p_4$ *e* $y_5 = n - y_1 - y_2 - y_3 - y_4$*, pode-se verificar que o EMV de* p_j *sobre* Θ *é*

$$\hat{p}_j = \frac{y_j}{n}.$$

Sob H_0*, a função de verossimilhança se reduz a*

$$L(\theta\,|\mathbf{x}) = p_1^{y_1+y_2+y_3}\left(\frac{1-3p_1}{2}\right)^{y_4+y_5}$$

Novamente, o método usual de calcular as derivadas e igualar a 0 mostram que os EMV de p_1 *sob* H_0 *é* $\hat{p}_{10} = (y_1+y_2+y_3)/(3n)$*. Portanto* $\hat{p}_{10} = \hat{p}_{20} = \hat{p}_{30}$ *e* $\hat{p}_{40} = \hat{p}_{50} = (1-3\hat{p}_{10})/2$*. Substituindo esses valores e os valores de* $\hat{p}_j$ *em* $L(\theta\,|\mathbf{x})$ *e combinando os termos com o mesmo expoente temos que*

$$\lambda(\mathbf{x}) = \left(\frac{y_1+y_2+y_3}{3y_1}\right)^{y_1}\left(\frac{y_1+y_2+y_3}{3y_2}\right)^{y_2}\left(\frac{y_1+y_2+y_3}{3y_3}\right)^{y_3}\left(\frac{y_4+y_5}{2y_4}\right)^{y4}\left(\frac{y_4+y_5}{2y_5}\right)^{y5}$$

Então o teste estatístico é

$$\begin{aligned}
&-2\log(\lambda(\mathbf{x})) = \\
&= -2\log\left[\left(\frac{y_1+y_2+y_3}{3y_1}\right)^{y_1}\left(\frac{y_1+y_2+y_3}{3y_2}\right)^{y_2}\left(\frac{y_1+y_2+y_3}{3y_3}\right)^{y_3}\left(\frac{y_4+y_5}{2y_4}\right)^{y4}\left(\frac{y_4+y_5}{2y_5}\right)^{y5}\right]\\
&= -2\left[y_1\log\left(\frac{y_1+y_2+y_3}{3y_1}\right)+y_2\log\left(\frac{y_1+y_2+y_3}{3y_2}\right)+y_3\log\left(\frac{y_1+y_2+y_3}{3y_3}\right)+y_4\log\left(\frac{y_4+y_5}{2y_4}\right)+y_5\log\left(\frac{y_4+y_5}{2y_5}\right)\right]\\
&= 2\left[y_1\log\left(\frac{3y_1}{y_1+y_2+y_3}\right)+y_2\log\left(\frac{3y_2}{y_1+y_2+y_3}\right)+y_3\log\left(\frac{3y_3}{y_1+y_2+y_3}\right)+y_4\log\left(\frac{2y_4}{y_4+y_5}\right)+y_5\log\left(\frac{2y_5}{y_4+y_5}\right)\right]\\
&= 2\sum_{i=1}^{5} y_i\log\left(\frac{y_i}{m_i}\right)
\end{aligned}$$

em que $m_1 = m_2 = m_3 = \frac{y_1+y_2+y_3}{3}$ *e* $m_4 = m_5 = \frac{y_4+y_5}{2}$*. O teste assintótico de tamanho* α *rejeita* H_0 *se* $-2\log(\lambda(\mathbf{x})) \geq \chi^2_{3,\alpha}$*. Esse exemplo é um de uma vasta classe de problemas de teste de hipóteses no qual a teoria assintótica do teste de razão de verossimilhança é extensivamente utilizado.*

Outros testes para grandes amostras

Um outro método comum de construção de um teste estatístico para grandes amostras é baseado em um estimador que tem como distribuição assintótica a normal. Suponha que desejamos testar uma hipótese sobre um parâmetro de valor real θ, e $W_n = W(X_1,\ldots,X_n)$ é um estimador pontual de θ, baseado em uma amostra de tamanho n, que foi derivado por algum método. Por exemplo, W_n poderia ser o EMV de θ. Um teste aproximado, baseado em uma aproximação pela normal, pode ser justificado do seguinte modo. Se σ_n^2 denota a variância de W_n e se podemos usar alguma forma do Teorema do Limite Central para mostrar que, conforme $n \to \infty$, $\frac{W_n-\theta}{\sigma_n}$ pode ser comparado a uma distribuição $N(0,1)$. Temos, portanto, a base de um teste aproximado.

Existem vários detalhes para serem verificados no argumento do parágrafo anterior, mas essa idéia tem aplicação em muitas situações. Por exemplo, se W_n é um EMV, o teorema 1.5.3 pode ser utilizado para validar o arqumento acima. Note que a distribuição de W_n e, talvez, o valor de σ_n depende do valor de θ. A convergência, assim, mais formalmente diz que para cada valor fixado de $\theta \in \Theta$, se usarmos a correspondente distribuição para W_n e o correspondente valor para σ_n, $\frac{W_n-\theta}{\sigma_n}$ converge para uma normal padrão. Se, para cada n, σ_n é uma constante calculável (que pode depender de θ, mas não de qualquer outro parâmetro), então um teste baseado em $\frac{W_n-\theta}{\sigma_n}$ pode ser obtido.

Em algumas situações, σ_n também depende de parâmetro desconhecido. Em tal caso, olhamos para uma estimativa S_n de σ_n com a propriedade de que $\frac{\sigma_n}{S_n}$ converge em probabilidade para 1. Então, usando o teorema de Slutsky, pode-se deduzir que $\frac{W_n-\theta}{S_n}$ também converge em distribuição para uma normal padrão. Um teste para grandes amostras pode ser baseado nesse fato.

Suponha que queremos um teste de hipóteses bilateral para testar $H_0: \theta = \theta_0$ contra $H_0: \theta \neq \theta_0$. Um teste aproximado pode ser baseado na estatística

$$Z_n = \frac{W_n-\theta_0}{S_n}$$

e rejeitaríamos H_0 se, e somente se, $Z_n < -z_{\alpha/2}$ ou $Z_n > z_{\alpha/2}$. Se H_0 é verdadeira , então $\theta = \theta_0$ e Z_n converge em distribuição para $Z \sim N(0{,}1)$. Enão, a probabilidade de erro tipo I,

$$P_{\theta_0}\left[Z_n < -z_{\frac{\alpha}{2}} \,\text{ou}\, Z_n > z_{\frac{\alpha}{2}}\right] \to P_{\theta}\left[Z_n < -z_{\frac{\alpha}{2}} \,\text{ou}\, Z_n > z_{\frac{\alpha}{2}}\right] = \alpha$$

e esse é um teste assintótico de tamanho α.

Agora considere um parâmetro alternativo de valor $\theta \neq \theta_0$. Podemos escrever.

$$Z_n = \frac{W_n - \theta_0}{S_n} = \frac{W_n - \theta}{S_n} + \frac{\theta - \theta_0}{S_n}$$

Não importa o valor de θ, o termo $\frac{W_n-\theta}{S_n} \to N(0{,}1)$. Tipicamente, é também o caso de que $\sigma_n \to 0$ conforme $n \to \infty$ (lembre que $\sigma_n = Var[W_n]$, e o estimador normalmente se torna mais preciso conforme $n \to \infty$). Então, S_n convergirá em probabilidade para 0 e o termo $\frac{\theta-\theta_0}{S_n}$ convergirá em probabilidade para $+\infty$ ou $-\infty$, dependendo se $(\theta - \theta_0)$ é positivo ou negativo. Então, Z_n convergirá para $+\infty$ ou $-\infty$ em probabilidade e

$$P_\theta\,[\text{rejeitar}\, H_0] = P_\theta\left[Z_n < -z_{\frac{\alpha}{2}} \,\text{ou}\, Z_n > z_{\frac{\alpha}{2}}\right] \to 1,\ n \to \infty.$$

Desse modo, um teste com tamanho assintótico α e poder assintótico 1 pode ser construído.

Se quiséssemos fazer um teste de hipóteses unilateral

$$\begin{cases} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{cases},$$

um teste similar poderia ser construído. Novamente, um teste estatístico $Z_n = \frac{W_n-\theta_0}{S_n}$ seria usado e o teste rejeitaria H_0 se, e somente se, $Z_n > z_\alpha$. De maneira similar ao caso acima, podemos concluir que a função poder do teste converge para 0, α ou 1, de acordo com $\theta < \theta_0$, $\theta = \theta_0$ ou $\theta > \theta_0$. Então esse teste também tem propriedades assintóticas de poder razoáveis.

Em geral, o teste de *Wald* é baseado em uma estatística da forma

$$Z_n = \frac{W_n - \theta_0}{S_n},$$

em que θ_0 é o valor na hipótese nula do parâmetro θ, W_n é um estimador de θ, e S_n é o erro padrão para W_n, uma estimativa do desvio padrão de W_n. Se W_n é o EMV de θ, então $\frac{1}{\sqrt{\hat{I}_n(W_n)}}$, em que

$$\hat{I}_n(W_n) = -\frac{\partial^2}{\partial\theta^2}\log\left(L\left(\theta\,|X\right)\right)\Bigg|_{\theta=W_n}$$

é o número de informação observado.

OBS.: S_n^2 é um estimador de $Var_\theta\left[\hat{W}_n\right]$, logo se n é grande $Var_\theta\left[\hat{W}_n\right]$ é aproximadamente a cota de Cramér-Rao e

$$Var_\theta\left[\hat{\theta}\right] \approx \frac{1}{-E_\theta\left[\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2}\log\left(f\left(X_i\,|\theta\right)\right)\right]\Big|_{\theta=\hat{\theta}}} = \frac{1}{nE_\theta\left[\frac{\partial^2}{\partial\theta^2}\log\left(f\left(X_i\,|\theta\right)\right)\right]\Big|_{\theta=\hat{\theta}}}$$

Um outro teste muito utilizado em casos de grandes amostras é o teste do *Escore* (*Score test*). A *estatística escore* é definida como

$$S(\theta) = \frac{\partial}{\partial\theta}\log\left(f\left(\mathbf{X}\,|\theta\right)\right) = \frac{\partial}{\partial\theta}\log\left(\prod_{i=1}^n f\left(X_i\,|\theta\right)\right) = \frac{\partial}{\partial\theta}\log\left(L\left(\theta\,|\mathbf{X}\right)\right).$$

Sabemos que , para todo θ, $E_\theta\,[S(\theta)] = 0$. Em particular, se estamos testando $H_0 : \theta = \theta_0$ e se H_0 é verdadeira, então $S(\theta_0)$ tem média zero. Além disso, sabemos que

$$Var_\theta\,[S(\theta)] = E_\theta\left[\left(\frac{\partial}{\partial\theta}\log\left(L\left(\theta\,|\mathbf{X}\right)\right)\right)^2\right] = -E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log\left(L\left(\theta\,|\mathbf{X}\right)\right)\right] = I_n(\theta);$$

ou seja, o número de informação é a variância da estatística *escore*. O teste estatístico para o teste do *escore* é

$$Z_S = \frac{S(\theta_0)}{\sqrt{I_n(\theta_0)}}.$$

Se H_0 é verdadeira, Z_S tem média zero e variância 1. Do teorema 1.5.3 segue que Z_S converge em distribuição para uma variável aleatória normal padrão se H_0 é verdadeira. Então o teste do escore rejeita H_0 se $|Z_S| > z_{\frac{\alpha}{2}}$ ao nível aproximado α. Se H_0 é composta, então $\hat{\theta}_0$, uma estimativa de θ assumindo que H_0 é verdadeira, substitui θ_0 em Z_S. Se $\hat{\theta}_0$ pe o EMV restrito, a maximização restrita poderia ser realizada utilizando os Multiplicadores de Lagrange. Então, o teste do *escore* é algumas vezes chamado de *teste dos Multiplicadores de Lagrange*.

A seguir, vamos demonstrar com mais detalhes a convergência da estatística do teste do escore.

Seja $S(\theta_0)$ a estatística escore definida anteriormente, ou seja,

$$\begin{aligned} S(\theta_0) &= \frac{\partial}{\partial\theta}\log\left(\prod\nolimits_{i=1}^{n} f(X_i|\theta)\right) = \sum\nolimits_{i=1}^{n}\frac{\partial}{\partial\theta}\log\left(f(X_i|\theta)\right) \\ &= n\frac{1}{n}\sum\nolimits_{i=1}^{n}\frac{\partial}{\partial\theta}\log\left(f(X_i|\theta)\right) = n\bar{Z}_n. \end{aligned}$$

Se $Z = \frac{\partial}{\partial\theta}\log(f(X_i|\theta))$, então

$$\begin{aligned} Var_{\theta_0}[S(\theta_0)] &= E_{\theta_0}\left[\left(\frac{\partial}{\partial\theta}\log\left(\prod\nolimits_{i=1}^{n} f(X_i|\theta)\right)\right)^2\right] = E_{\theta_0}\left[\left(\sum\nolimits_{i=1}^{n}\frac{\partial}{\partial\theta}\log\left(f(X_i|\theta)\right)\right)^2\right] \\ &= nE_{\theta_0}\left[\left(\frac{\partial}{\partial\theta}\log\left(f(X|\theta)\right)\right)^2\right] = \underbrace{-n\,\underbrace{E_{\theta_0}\left[\frac{\partial^2}{\partial\theta^2}\log\left(f(X|\theta)\right)\right]}_{I(\theta_0)}}_{I_n(\theta_0)} \\ &= I_n(\theta_0) = nI(\theta_0) \end{aligned}$$

Logo, temos que

$$Var\left[\bar{Z}_n\right] = Var\left[\frac{S(\theta_0)}{n}\right] = \frac{1}{n^2}Var[S(\theta_0)] = \frac{1}{n^2}nI(\theta_0) = \frac{I(\theta_0)}{n}$$

Pelo Teorema do Limite Central, segue que

$$\frac{\bar{Z}_n - E_{\theta_0}\left[\bar{Z}_n\right]}{\sqrt{\frac{I(\theta_0)}{n}}} \to N(0,1) \Rightarrow \frac{\sqrt{n}\bar{Z}_n}{\sqrt{I(\theta_0)}} \to N(0,1)$$

converge em distribuição. Como $n\bar{Z}_n = S(\theta_0)$, implica que

$$\frac{\sqrt{n}\bar{Z}_n}{\sqrt{I(\theta_0)}} = \frac{n\bar{Z}_n}{\sqrt{n}\sqrt{I(\theta_0)}} = \frac{S(\theta_0)}{\sqrt{nI(\theta_0)}} = \frac{S(\theta_0)}{\sqrt{I_n(\theta_0)}} = Z_S.$$

Portanto, $Z_S \to N(0,1)$ em distribuição.

Dessa forma, para testar as hipóteses $H_0: \theta = \theta_0$ contra $H_0: \theta \neq \theta_0$ o teste do escore é: rejeitar H_0 se $Z_S > z_{\frac{\alpha}{2}}$ ou $Z_S < -z_{\frac{\alpha}{2}}$.

1.7 Estimação Intervalar

Considere uma amostra $X_1, \ldots, X_n$ com $X_i \sim N(\mu,\sigma^2)$. Suponha que queremos encontrar um intervalo de confiança com $(1-\alpha)100\%$ de confiança para μ, considerando σ^2 desconhecido. Esse intervalo de confiança deve ser construído com base em uma quantidade pivotal. No caso acima, considere a seguinte quantidade

$$\frac{\bar{X}-\mu}{\frac{S_n}{\sqrt{n}}}, \tag{1.36}$$

daí

$$\frac{\bar{X}-\mu}{\frac{S_n}{\sqrt{n}}} = \frac{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\frac{\frac{S_n}{\sqrt{n}}}{\frac{\sigma}{\sqrt{n}}}} = \frac{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\frac{S_n}{\sigma}} = \frac{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{1}{n-1}\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}}} = \frac{Z}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}} \sim t_{n-1}$$

em que S_n é a estimativa da variància amostral. Logo, a quantidade 1.36 é uma quantidade pivotal e tem distribuição t_{n-1}. Como isso, podemos construir um intervalo de confiança para μ, pivotando a quantidade 1.36 da seguinte forma

$$\begin{aligned}
1-\alpha &= P\left[-\xi_{\frac{\alpha}{2}} < t_{n-1} < \xi_{\frac{\alpha}{2}}\right] \\
&= P\left[-\xi_{\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{\frac{S_n}{\sqrt{n}}} < \xi_{\frac{\alpha}{2}}\right] \\
&= P\left[\bar{X} - \frac{S_n}{\sqrt{n}}\xi_{\frac{\alpha}{2}} < \mu < \bar{X} + \frac{S_n}{\sqrt{n}}\xi_{\frac{\alpha}{2}}\right]
\end{aligned}$$

Portanto, o intervalo com $(1-\alpha)\,100\%$ de confiança para μ é

$$IC_{(1-\alpha)100\%}(\mu) = \left[\bar{X} - \frac{S_n}{\sqrt{n}}\xi_{\frac{\alpha}{2}};\, \bar{X} + \frac{S_n}{\sqrt{n}}\xi_{\frac{\alpha}{2}}\right]$$

Contudo, existem situações em que não é possível, ou é muito difícil, encontrar uma quantidade pivotal como em 1.36. Exploramos nas seções anteriores algumas versões aproximadas e assintóticas de conjuntos de confiança. A proposta agora é, como visto nessas seções, apresentar métodos que serão úteis em situações mais complicadas e nos proporcionarão obter alguma resposta. A resposta obtida nesse caso não é, quase que certamente, a melhor, mas também não é a pior. Em muitos casos, no entanto, são o melhor que podemos fazer.

Intervalo de Confiança aproximado utilizando EMV

Da discussão na seção 1.5 e usando o teorema 1.5.3, temos um método geral de se obter uma distribuição assintótica para o EMV. Portanto, temos um método geral para construir um intervalo de confiança. De forma análoga ao teorema 1.5.3, podemos enunciar o seguinte teorema.

Teorema 1.7.1. *Sob as mesmas condições do teorema 1.5.3, se $X_1,\ldots,X_n$ são iid $f(x\,|\theta)$ e $\hat{\theta}$ é EMV de θ, e $h\left(\hat{\theta}\right)$ é uma função do EMV de $h(\theta)$, então*

$$\sqrt{n}\left(h\left(\hat{\theta}\right) - h(\theta)\right) \to N\left(0,\frac{(h'(\theta))^2}{I(\theta)}\right) \tag{1.37}$$

em distribuição e $I(\theta) = -E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log\left(f(X\,|\theta)\right)\right]$.

Do teorema 1.7.1 podemos obter a seguinte quantidade pivotal

$$\frac{\sqrt{n}\left(h\left(\hat{\theta}_0\right)-h(\theta)\right)}{\sqrt{\frac{(h'(\theta))^2}{I(\theta)}}} = \frac{h\left(\hat{\theta}_0\right)-h(\theta)}{\sqrt{\frac{(h'(\theta))^2}{nI(\theta)}}} = \frac{h\left(\hat{\theta}_0\right)-h(\theta)}{\sqrt{\frac{(h'(\theta))^2}{I_n(\theta)}}} \sim N(0{,}1)$$

cuja variância de $h\left(\hat{\theta}\right)$ pode ser aproximada pelo método Delta, daí

$$\frac{\left(h'\left(\theta\right)\right)^2}{I_n\left(\theta\right)} = \frac{\left(h'\left(\theta\right)\right)^2}{-E_\theta\left[\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2}\log\left(f\left(X_i\left|\theta\right.\right)\right)\right]} \approx \frac{\left(h'\left(\theta\right)\right)^2}{-\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2}\log\left(f\left(X_i\left|\hat{\theta}\right.\right)\right)}$$

Logo, temos que

$$\frac{h\left(\hat{\theta}_0\right) - h\left(\theta\right)}{\sqrt{\frac{\left(h'\left(\theta\right)\right)^2}{-\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2}\log\left(f\left(X_i\left|\hat{\theta}\right.\right)\right)}}} \sim N\left(0{,}1\right) \tag{1.38}$$

é aproximadamente uma quantidade pivotal para θ.

Portanto, um intervalo de confiança com $\left(1-\alpha\right)100\%$ para $h\left(\theta\right)$ é

$$IC_{(1-\alpha)100\%}\left(h\left(\theta\right)\right) = \left[h\left(\hat{\theta}_0\right) - \sqrt{\frac{\left(h'\left(\theta\right)\right)^2}{-\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2}\log\left(f\left(X_i\left|\hat{\theta}\right.\right)\right)}}z_{\frac{\alpha}{2}};\, h\left(\hat{\theta}_0\right) + \sqrt{\frac{\left(h'\left(\theta\right)\right)^2}{-\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2}\log\left(f\left(X_i\left|\hat{\theta}\right.\right)\right)}}z_{\frac{\alpha}{2}}\right] \tag{1.39}$$

Exemplo 1.7.0.21. *Seja uma amostra aleatória $X_1,\ldots,X_n$ de uma população Bernoulli(p). Vimos que podemos estimar a razão de chances $\frac{p}{1-p}$ pelo seu EMV $\frac{\hat{p}}{1-\hat{p}}$ e que essa estimativa tem variância aproximada por*

$$\hat{V}ar\left[\frac{\hat{p}}{1-\hat{p}}\right] \approx \frac{\hat{p}}{n\left(1-\hat{p}\right)^3}.$$

Qual o intervalo de confiança assintótico de EMV para $\frac{p}{1-p}$?

SOLUÇÃO

Seja $h\left(\theta\right) = \frac{p}{1-p}$ e $h\left(\hat{\theta}_0\right) = \frac{\hat{p}}{1-\hat{p}}$. Pela expressão 1.38 temos que

$$\begin{aligned}
1-\alpha &= P\left[-z_{\frac{\alpha}{2}} < \frac{h\left(\hat{\theta}_0\right) - h\left(\theta\right)}{\sqrt{Var\left[h\left(\hat{\theta}_0\right)\right]}} < z_{\frac{\alpha}{2}}\right] \\
&= P\left[-z_{\frac{\alpha}{2}} < \frac{h\left(\hat{\theta}_0\right) - h\left(\theta\right)}{\sqrt{\frac{\left(h'\left(\theta\right)\right)^2}{I_n\left(\theta\right)}}} < z_{\frac{\alpha}{2}}\right] \\
&= P\left[-z_{\frac{\alpha}{2}} < \frac{\frac{\hat{p}}{1-\hat{p}} - \frac{p}{1-p}}{\sqrt{Var\left[\frac{\hat{p}}{1-\hat{p}}\right]}} < z_{\frac{\alpha}{2}}\right] \\
&= P\left[\frac{\hat{p}}{1-\hat{p}} - z_{\frac{\alpha}{2}}\sqrt{Var\left[\frac{\hat{p}}{1-\hat{p}}\right]} < \frac{p}{1-p} < \frac{\hat{p}}{1-\hat{p}} + z_{\frac{\alpha}{2}}\sqrt{Var\left[\frac{\hat{p}}{1-\hat{p}}\right]}\right] \\
&= P\left[\frac{\hat{p}}{1-\hat{p}} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}}{n\left(1-\hat{p}\right)^3}} < \frac{p}{1-p} < \frac{\hat{p}}{1-\hat{p}} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}}{n\left(1-\hat{p}\right)^3}}\right]
\end{aligned}$$

que é o intervalo de confiança aproximado.

Uma forma mais restritiva de aproximação por verossimilhança, mas que, quando aplicável, retorna melhores intervalos, é baseado na estatística do *escore*. A quantidade aleatória

$$Q\left(\mathbf{X}\left|\theta\right.\right) = \frac{\frac{\partial}{\partial\theta}\log\left(L\left(\theta\left|\mathbf{X}\right.\right)\right)}{\sqrt{-E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log\left(L\left(\theta\left|\mathbf{X}\right.\right)\right)\right]}} \tag{1.40}$$

tem distribuição assintoticamente $N(0{,}1)$ conforme $n \to \infty$. Então, o conjunto

$$\left\{\theta : |Q(\mathbf{X}\,|\theta)| \leq z_{\frac{\alpha}{2}}\right\} \tag{1.41}$$

é um conjunto aproximado de confiança $1-\alpha$. Note que, aplicando os resultados da seção 1.5, temos que

$$E_\theta\left[Q(\mathbf{X}\,|\theta)\right] = \frac{E_\theta\left[\frac{\partial}{\partial\theta}\log\left(L(\theta\,|\mathbf{X})\right)\right]}{\sqrt{-E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log\left(L(\theta\,|\mathbf{X})\right)\right]}} = 0$$

e

$$Var_\theta\left[Q(\mathbf{X}\,|\theta)\right] = \frac{Var_\theta\left[\frac{\partial}{\partial\theta}\log\left(L(\theta\,|\mathbf{X})\right)\right]}{-E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log\left(L(\theta\,|\mathbf{X})\right)\right]} = 1,$$

e essa aproximação corresponde aos dois primeiros momentos de uma variável aleatória normal padrão.

Exemplo 1.7.0.22. ***(Intervalo Escore da Binomial)*** *Novamente usando a binomial como exemplo, se $Y = \sum_{i=1}^{n} X_i$, onde cada X_i é uma variável aleatória independente Bernoulli(p). Obtenha o intervalo de confiança para p utilizando na estatística escore.*

SOLUÇÃO

Pela expressão 1.40 temos que

$$\begin{aligned} Q(Y\,|p) &= \frac{\frac{\partial}{\partial p}\log\left(L(p\,|Y)\right)}{\sqrt{-E_p\left[\frac{\partial^2}{\partial p^2}\log\left(L(p\,|Y)\right)\right]}} \\ &= \frac{\frac{y}{p} - \frac{n-y}{1-p}}{\sqrt{\frac{n}{p(1-p)}}} \\ &= \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \end{aligned}$$

onde $\hat{p} = \frac{y}{n}$. De 1.41, um intervalo de confiança $1-\alpha$ aproximado é dados por

$$\left\{p : \left|\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}\right| \leq z_{\frac{\alpha}{2}}\right\}.$$

Esse é o intervalo que resulta invertendo a estatística escore. Para calcular esse intervalo precisamos resolver um termo quadrático em p.

Um outro teste baseado em verossimilhança seria considerar o fato de que $-2\log(\lambda(\mathbf{x}))$ tem distribuição assintótica qui-quadrado. Vimos que se $X_1,\ldots,X_n$ são iid $f(x\,|\theta)$ e $\hat{\theta}$ é o EMV de θ, então

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta\in\Theta_0} L(\theta\,|\mathbf{x})}{\sup_{\theta\in\Theta} L(\theta\,|\mathbf{x})} = \frac{\sup_{\theta\in\Theta_0} L(\theta\,|\mathbf{x})}{L\left(\hat{\theta}\,|\mathbf{x}\right)}$$

que sob $H_0 \Rightarrow -2\log(\lambda(\mathbf{x})) \to \chi^2_v$ em distribuição se H_0 é composta. Se H_0 é simples $(H_0 : \theta = \theta_0)$ então,

$$\lambda(\mathbf{x}) = \frac{L(\theta_0\,|\mathbf{x})}{L\left(\hat{\theta}\,|\mathbf{x}\right)}$$

que sob $H_0 \Rightarrow -2\log(\lambda(\mathbf{x})) \to \chi^2_1$ em distribuição. Portanto, o conjunto de confiança

$$\left\{\theta : -2\log\left(\frac{L(\theta_0\,|\mathbf{x})}{L\left(\hat{\theta}\,|\mathbf{x}\right)}\right) < \chi^2_{1;\alpha}\right\}, \tag{1.42}$$

é um intervalo de confiança $1-\alpha$ aproximado.

Exemplo 1.7.0.23. *(Intervalo pelo TRV para Binomial) Para* $Y = \sum_{i=1}^{n} X_i$*, onde cada* X_i *é uma variável aleatória independente Bernoulli(*p*). Obtenha o conjunto de confiança* $1-\alpha$ *aproximado para* p*.*

SOLUÇÃO

Pela expressão 1.42 temos que

$$\left\{\theta : -2\log\left(\frac{L\left(\theta_0\,|\mathbf{x}\right)}{L\left(\hat{\theta}\,|\mathbf{x}\right)}\right) \leq \chi^2_{1;\alpha}\right\}$$

$$\left\{p : -2\log\left(\frac{p^y\left(1-p\right)^{n-y}}{\hat{p}^y\left(1-\hat{p}\right)^{n-y}}\right) \leq \chi^2_{1;\alpha}\right\}$$

é o conjunto de confiança $1-\alpha$ *aproximado pelo TRV.*

1.8 Teste para Tabelas de Contingência

Dados binomiais colhidos de mais de uma população são às vezes apresentados em uma tabela de contingência. Para o caso de duas populações, a tabela poderia se parecer como a seguinte:

	População		
	1	2	Total
Sucessos	S_1	S_2	$S = S_1 + S_2$
Fracassos	F_1	F_2	$F = F_1 + F_2$
Total	n_1	n_2	$n = n_1 + n_2$

em que a população 1 é binomial(n_1, p_1), com S_1 sucessos e F_1 fracassos, e a população 2 é binomial(n_2, p_2), com S_2 sucessos e F_2 fracassos. As hipótese que são normalmente de interesse são

$$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$$

Vamos mostrar que um teste pode ser baseado na estatística

$$T = \frac{\left(\hat{p}_1 - \hat{p}_2\right)^2}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\left(\hat{p}\left(1-\hat{p}\right)\right)}$$

em que $\hat{p}_1 = \frac{S_1}{n_1}$, $\hat{p}_2 = \frac{S_2}{n_2}$ e $\hat{p} = \frac{S_1+S_2}{n_1+n_2}$. Mostraremos também que conforme n_1,$n_2 \to \infty$, a distribuição de T se aproxima de uma χ^2_1. (Esse é um caso especial de um teste conhecido como *teste de independência qui-quadrado*)

Demonstração. O teste deve ser da forma: rejeito H_0 se $T > k$, para algum valor de k.

Mas para calcular a taxa de erro tipo I, é necessário conhecer a distribuição de T sob H_0, mas essa distribuição parece ser bastante complicada.

Usando a aproximação da distribuição normal para a binomial, temos que, sob H_0 $(p_1 = p_2 = p)$ $\hat{p}_1 \sim Binomial\left(n_1,p\right)$, $\hat{p}_2 \sim Binomial\left(n_2,p\right)$, $Var\left[\hat{p}_1\right] = \frac{p(1-p)}{n_1}$ e $Var\left[\hat{p}_2\right] = \frac{p(1-p)}{n_2}$.

Logo, $\hat{p}_1$ pode ser aproximado por

$$\hat{p}_1 \sim N\left(p_1, \frac{p_1\left(1-p_1\right)}{n_1}\right),$$

da mesma forma

$$\hat{p}_2 \sim N\left(p_2, \frac{p_2\left(1-p_2\right)}{n_2}\right).$$

Sob $H_0 : p_1 = p_2 = p$ temos que

$$\hat{p}_1 = \frac{S_1}{n_1} \sim N\left(p, \frac{p\left(1-p\right)}{n_1}\right)$$

$$\hat{p}_2 = \frac{S_2}{n_2} \sim N\left(p, \frac{p\,(1-p)}{n_2}\right).$$

Por suposição, $\hat{p}_1$ e $\hat{p}_2$ são independentes e intuitivamente isso é verdade. Daí, pelo teorema da Convolução, a distribuição de $\hat{p}_1 - \hat{p}_2$ é

$$\begin{aligned} \hat{p}_1 - \hat{p}_2 &\sim N\left(0, \left(\frac{1}{n_1} + \frac{1}{n_2}\right) p\,(1-p)\right) \\ \Rightarrow \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) p\,(1-p)}} &\sim N\,(0{,}1)\,. \end{aligned}$$

Portanto,

$$\begin{aligned} Z &= \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \hat{p}\,(1-\hat{p})}} \sim N\,(0{,}1) \\ &\Rightarrow Z^2 = T \sim \chi^2_1, \end{aligned}$$

ou seja, a estatística Z tem distribuição normal padrão e sabe-se que Z^2 tem distribuição qui-quadrado com 1 grau de liberdade.

Uma outra forma de construir o teste seria por meio do TRV assintótico. Para tal, temos que

$$\begin{aligned} \lambda\,(\mathbf{x}) &= \frac{\sup_{\theta\in\Theta_0} L\,(\theta\,|\mathbf{x})}{\sup_{\theta\in\Theta} L\,(\theta\,|\mathbf{x})} \\ &= \frac{\sup_{p_1=p_2} L\,(\theta\,|\mathbf{x})}{\sup_{p_1,p_2\in[0,1]} L\,(\theta\,|\mathbf{x})} \\ &= \frac{\sup_{p_1=p_2} p_1^{S_1}\,(1-p_1)^{n_1-S_1}\,p_2^{S_2}\,(1-p_2)^{n_2-S_2}}{\sup_{p_1,p_2\in[0,1]} p_1^{S_1}\,(1-p_1)^{n_1-S_1}\,p_2^{S_2}\,(1-p_2)^{n_2-S_2}} \\ &= \frac{\sup_{p_1=p_2} p^{S_1+S_2}\,(1-p)^{n_1+n_2-(S_1+S_2)}}{\sup_{p_1,p_2\in[0,1]} p_1^{S_1}\,(1-p_1)^{n_1-S_1}\,p_2^{S_2}\,(1-p_2)^{n_2-S_2}} \\ &= \frac{\left(\frac{S_1+S_2}{n_1+n_2}\right)^{S_1+S_2}\left(1-\frac{S_1+S_2}{n_1+n_2}\right)^{n_1+n_2-(S_1+S_2)}}{\left(\frac{S_1}{n_1}\right)^{S_1}\left(1-\frac{S_1}{n_1}\right)^{n_1-S_1}\left(\frac{S_2}{n_2}\right)^{S_2}\left(1-\frac{S_2}{n_2}\right)^{n_2-S_2}}. \end{aligned}$$

Portanto, sob $H_0 \Rightarrow -2\log\left(\lambda\,(\mathbf{x})\right) \sim \chi^2_1$ pelo TRV assintótico. Recusar H_0 se $-2\log\left(\lambda\,(\mathbf{x})\right) > \chi^2_{1;\alpha}$. □

1.9 Propriedades Assintóticas das Estatísticas de Ordem

Uma das principais de se utilizar o comportamento assintótico está nas estatísticas de ordem.

Pelo Teorema do Limite Central, temos que, para a média amostral,

$$\sqrt{n}\left(\bar{X} - \mu\right) \to N\left(0{,}\sigma^2\right) \tag{1.43}$$

a variável aleatória em 1.43 converge em distribuição. Será que existe alguma propriedade parecida para uma grandeza similar, como a mediana? De maneira geral, é possível se obter distribuições assintóticas para as estatísticas de ordem? Esse assunto será discutido a seguir.

Normalidade Assintótica da Mediana

Seja $X_1, \ldots, X_n$ uma amostra iid de uma população com fdp $f(\cdot)$ e fd $F(\cdot)$. Se $F(\mu) = P(X \leq \mu) = \frac{1}{2}$ então $\mu = mediana$. O argumento é baseado na distribuição binomial. Seja M_n a mediana amostral. Por exemplo, se n é ímpar,

$$M_n = X^n_{\left(\frac{n-1}{2}+1\right)}$$

em termos de estatística de ordem.

Vamos definir

$$Y_i = \begin{cases} 1, \text{se } X_i \leq \mu + \frac{a}{\sqrt{n}} \\ 0, \text{caso contrario} \end{cases}$$

em que a é uma constante. Logo, $Y_i \sim Bernoulli$ com probabilidade de sucesso

$$P[Y_i = 1] = P\left[X_i \leq \mu + \frac{a}{\sqrt{n}}\right] = F\left[\mu + \frac{a}{\sqrt{n}}\right] = p_n.$$

Novamente supondo n ímpar, o evento $\left\{M_n \leq \mu + \frac{a}{\sqrt{n}}\right\}$ significa que pelo menos metade dos X_i são maiores que $\mu + \frac{a}{\sqrt{n}}$ e, portanto, mais da metade dos Y_i são 1 e $\left\{\sum_{i=1}^n Y_i \geq \frac{n+1}{2}\right\}$. Logo,

$$\begin{aligned} P\left[\sqrt{n}\left(M_n - \mu\right) \leq a\right] &= P\left[M_n \leq \mu + \frac{a}{\sqrt{n}}\right] \\ &= P\left[\sum_{i=1}^n Y_i \geq \frac{n+1}{2}\right] \\ &= P\left[\underbrace{\sum_{i=1}^n Y_i}_{\sum_{i=1}^n Y_i \sim Binomial(n, p_n)} - np_n \geq \frac{n+1}{2} - np_n\right] \\ &= P\left[\frac{\sum_{i=1}^n Y_i - np_n}{\sqrt{np_n\left(1-p_n\right)}} \geq \frac{\frac{n+1}{2} - np_n}{\sqrt{np_n\left(1-p_n\right)}}\right], \end{aligned}$$

quando $n \to \infty$, então

$$p_n = P\left[X_i \leq \mu + \frac{a}{\sqrt{n}}\right] \to \frac{1}{2},$$

uma vez que $\lim_{n\to\infty} \frac{a}{\sqrt{n}} = 0$ e $P(X \leq \mu) = \frac{1}{2}$. Pelo Teorema do Limite Central, temos que

$$\frac{\sum_{i=1}^n Y_i - np_n}{\sqrt{np_n\left(1-p_n\right)}} = \frac{n\left(\frac{\sum_{i=1}^n Y_i}{n} - np_n\right)}{\sqrt{n}\sqrt{p_n\left(1-p_n\right)}} = \frac{\sqrt{n}\left(\bar{Y}_n - E[Y]\right)}{\sqrt{Var[Y]}} \to N(0,1)$$

Vejamos a convergência de

$$\frac{\frac{n+1}{2} - np_n}{\sqrt{np_n\left(1-p_n\right)}}$$

em que $p_n = F\left[\mu + \frac{a}{\sqrt{n}}\right]$. Daí,

$$\begin{aligned} \frac{\frac{n+1}{2} - np_n}{\sqrt{np_n\left(1-p_n\right)}} &= \frac{1}{\sqrt{p_n\left(1-p_n\right)}}\left[\frac{1}{2\sqrt{n}} + \frac{\frac{n}{2} - np_n}{\sqrt{n}}\right] \\ &= \frac{1}{\sqrt{p_n\left(1-p_n\right)}}\left[\frac{1}{2\sqrt{n}} + \left(\frac{n}{2\sqrt{n}} - \frac{np_n}{\sqrt{n}}\right)\right] \\ &= \frac{1}{\sqrt{p_n\left(1-p_n\right)}}\left[\frac{1}{2\sqrt{n}} + \left(\frac{\sqrt{n}}{2} - \frac{\sqrt{n}p_n}{1}\right)\right] \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{p_n (1-p_n)}} \left[\frac{1}{2\sqrt{n}} + \sqrt{n} \left(\frac{1}{2} - p_n \right) \right] \\
&= \frac{1}{\sqrt{p_n (1-p_n)}} \frac{1}{2\sqrt{n}} - \frac{\sqrt{n}}{\sqrt{p_n (1-p_n)}} \left(p_n - \frac{1}{2} \right)
\end{aligned}
$$

lembrando que quando $n \to \infty$, $p_n (1-p_n) \to \frac{1}{4}$ e $\frac{1}{2\sqrt{n}} \to 0$. Portanto, o único limite a ser feito é

$$
\begin{aligned}
&\Rightarrow \frac{1}{\sqrt{p_n (1-p_n)}} \frac{1}{2\sqrt{n}} - \frac{\sqrt{n}}{\sqrt{p_n (1-p_n)}} \left(p_n - \frac{1}{2} \right) = -2\sqrt{n} \left(p_n - \frac{1}{2} \right) \\
&= -2 \frac{\left[F\left(\mu + \frac{a}{\sqrt{n}} \right) - F(\mu) \right]}{\frac{1}{\sqrt{n}}} \\
&= -2a \frac{\left[F\left(\mu + \frac{a}{\sqrt{n}} \right) - F(\mu) \right]}{\frac{a}{\sqrt{n}}}.
\end{aligned}
$$

Quando $n \to \infty$, $\frac{a}{\sqrt{n}} \to 0$ e o limite acima é o mesmo que

$$
\lim_{\Delta\mu \to \infty} -2a \frac{[F(\mu + \Delta\mu) - F(\mu)]}{\Delta\mu} = -2aF'(\mu) = -2af(\mu)
$$

que é a definição de derivada de uma função $F(\mu)$. Logo

$$
P\left[\sqrt{n}(M_n - \mu) \leq a\right] \to P\left[Z \leq -2af(\mu)\right] = P\left[\frac{Z}{2f(\mu)} \leq -a \right].
$$

Considerando a fd de $\sqrt{n}(M_n - \mu)$ denotada por H, temos que

$$
\begin{aligned}
&H(a) \to 1 - \Phi_{0,\left(\frac{1}{2f(\mu)}\right)^2}(-a) \\
\Rightarrow &H'(a) \to -\Phi'_{0,\left(\frac{1}{2f(\mu)}\right)^2}(-a)(-1) = \Phi'_{0,\left(\frac{1}{2f(\mu)}\right)^2}(-a)
\end{aligned}
$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada de uma normal. Como a normal com média zero é simétrica em relação à origem temos que

$$
H'(a) = \phi_{0,\frac{1}{4f^2(\mu)}}(a)
$$

em que $\phi(\cdot)$ é a função densidade de probabilidade de uma normal. Portanto, a convergência em distribuição de $\sqrt{n}(M_n - \mu)$ é para uma distribuição normal.

Com isso, acabamos de demonstrar o seguinte teorema.

Teorema 1.9.1. *Seja μ a mediana populacional de uma amostra $X_1, \ldots, X_n$ iid de uma população $f(\cdot)$ e $F(\cdot)$ sua função de distribuição acumulada. Seja M_n a estatística de ordem relativa à mediana de uma amostra de tamanho n. Se $n \to \infty$, então a variável aleatória $\sqrt{n}(M_n - \mu)$ converge em distribuição para uma $N\left(0, \frac{1}{4f^2(\mu)}\right)$.*

No teorema acima, a posição relativa é mantida quando n cresce.

Uma outra situação é fixar uma posição $X_k^{(n)}$, k fixo e n cresce. As situações mais importantes são $X_1^{(n)}$, que é a estatística de ordem do mínimo, e $X_n^{(n)}$, que é a estatística de ordem do máximo. Ambas são estatísticas de valor extremo. Vejamos o caso $X_n^{(n)}$. Tem-se

$$
F_{X_n^{(n)}}(x) = [F_X(x)]^n.
$$

No entanto, $X_n^{(n)}$ tem tendência a concentrar seus valores em valores cada vez maiores. A idéia é normalizar $X_n^{(n)}$ na tentativa de se obter uma distribuição limite para

$$\frac{X_n^{(n)} - a_n}{b_n}.$$

(Note que $\frac{\bar{X}_n-\mu}{\frac{\sigma}{\sqrt{n}}} = \sqrt{n}\left(\frac{\bar{X}_n-\mu}{\sigma}\right)$)). A sequência de valores a_n tem a função de "puxar" os valores limite e os valores de b_n fazem uma mudança de escala na variável $X_n^{(n)}$. As sequências $\{a_n\}$ e $\{b_n\}$ devem ser obtidas.

Vejamos um exemplo com a distribuição logística.

Exemplo 1.9.0.24. *Considere a função de distribuição acumulada da distribuição logística*

$$F(x) = \frac{1}{1+e^{-x}}.$$

Como obter $\{a_n\}$ e $\{b_n\}$.

SOLUÇÃO

É razoável supor a_n seje próximo de $E\left[X_n^{(n)}\right]$. Pelo Teorema da Probabilidade Integral temos que $F(X) \sim U(0,1)$, daí diante de uma amostra $\{X_1,\ldots,X_n\}$ e aplicando o teorema da Probabilidade Integral,

$$\{F(X_1),\ldots,F(X_n)\} \Rightarrow \max\{F(X_1),\ldots,F(X_n)\} = F(\max\{X_1,\ldots,X_n\}) = F\left(X_n^{(n)}\right)$$

pois F é crescente. Assim, utilizando a esperança da uniforme temos que

$$E\left[F\left(X_n^{(n)}\right)\right] = \frac{n}{n+1}$$

e utilizando o método Delta, temos que

$$F\left(E\left[X_n^{(n)}\right]\right) \approx E\left[F\left(X_n^{(n)}\right)\right]$$

$$\frac{1}{1+e^{-E\left[X_n^{(n)}\right]}} = 1 - \frac{1}{1+e^{-E\left[X_n^{(n)}\right]}} \approx 1 - \frac{1}{1+n}$$

OBS.:

$$\frac{1}{1+e^{-x}} = \frac{1}{e^{-x}e^{x}+e^{-x}} = \frac{1}{e^{-x}(e^{x}+1)} = \frac{e^{x}}{e^{x}+1} = 1 - \frac{1}{1+e^{x}}$$

Portanto,

$$e^{E\left[X_n^{(n)}\right]} = n \Rightarrow E\left[X_n^{(n)}\right] \approx \log(n) = a_n.$$

E ainda

$$\begin{aligned}
\frac{X_n^{(n)} - a_n}{b_n} = \frac{X_n^{(n)} - \log(n)}{b_n} &\Rightarrow \lim_{n\to\infty} P\left[\frac{X_n^{(n)} - \log(n)}{b_n} \leq x\right] \\
&\Rightarrow \lim_{n\to\infty} P\left[X_n^{(n)} \leq xb_n + \log(n)\right] \\
&\Rightarrow \lim_{n\to\infty} F_{X_n^{(n)}}(xb_n + \log(n)) \\
&\Rightarrow \lim_{n\to\infty} \left[F(xb_n + \log(n))\right]^n \\
&\Rightarrow \lim_{n\to\infty} \left[\frac{1}{1+e^{-xb_n-\log(n)}}\right]^n \\
&\Rightarrow \lim_{n\to\infty} \left[1+e^{-xb_n-\log(n)}\right]^{-n} \\
&\Rightarrow \lim_{n\to\infty} \left[1+\frac{e^{-xb_n}}{n}\right]^{-n}
\end{aligned}$$

$$\Rightarrow e^{-e^{x}}$$

que é a densidade de uma variável aleatória com distribuição Gumbel considerando $b_n = 1$.

Podemos, agora, apresentar um teorema universal para a estatística de ordem de valor extremo.

Teorema 1.9.2. *Seja* $X_1, \ldots, X_n$ *uma amostra iid e* $F(\cdot)$ *a fd de* X_i. *Se* $\frac{X_n^{(n)} - a_n}{b_n}$ *converge em distribuição, então o limite é:*

- $$G_1(x;\gamma) = e^{-x^{-\gamma}}, \gamma > 0 \tag{1.44}$$

- $$G_2(x;\gamma) = e^{-|x|^{\gamma}} I_{(-\infty,0)}(x) + I_{(0,\infty)}(x), \gamma > 0 \tag{1.45}$$

- $$G_3(x;\gamma) = e^{-e^{x}} \tag{1.46}$$

em que as distribuições 1.44, 1.45 e 1.46 são amplamente conhecidas por Weibull, Fréchet e Gumbel.

Capítulo 2

Exercícios Resolvidos

2.1 Lista 1

Questão 1(Caderno)

Provar que $E[(W_n - \theta)^2] = Var_\theta(W_n) + [\text{Viés}_\theta(W_n)]^2$.
SOLUÇÃO

Sendo $E[(W_n - \theta)^2]$ o erro quadrático médio, temos

$$\begin{aligned} E[(W_n-\theta)^2] &= E_\theta[W_n - E_\theta(W_n) + E_\theta(W_n) - \theta^2], \\ &= E_\theta[(W_n - E_\theta(W_n))^2 + 2(W_n - E_\theta(W_n))(E_\theta(W_n - \theta) + \\ &\quad +(E_\theta(W_n - \theta)^2], \\ &= E_\theta[(W_n - E_\theta(W_n))^2] + 2E_\theta[(W_n - E_\theta(W_n))(E_\theta(W_n) - \theta)] + \\ &\quad +E[(E_\theta(W_n) - \theta)^2], \end{aligned}$$

Observando $2E[(W_n - E_\theta(W_n))(E_\theta(W_n) - \theta)]$, temos que $(E_\theta(W_n) - \theta)$ é uma constante, portanto, pode sair da esperança. Assim,

$$\begin{aligned} E_\theta[(W_n-\theta)^2] &= E_\theta[(W_n - E_\theta(W_n))^2] + 2(E(W_n) - \theta)E[W_n - E_\theta(W_n)] \\ &\quad (E_\theta(W_n) - \theta)^2, \end{aligned}$$

temos que $E[W_n - E_\theta(W_n)] = E_\theta[W_n] - E_\theta[W_n] = 0$. Dessa forma,

$$E_\theta[(W_n-\theta)^2] \;=\; E_\theta[(W_n - E_\theta(W_n))^2] + (E_\theta(W_n)-\theta)^2.$$

Pela definição 11(MOOD, 1974,pág.67), $Var_\theta[W_n] = E_\theta[(W_n - E_\theta(W_n))^2]$ e $E_\theta(W_n)-\theta$ é chamado de viés (MOOD, 1974, pág. 293). Portanto,

$$E[(W_n-\theta)^2] = Var_\theta(W_n) + [\text{Viés}_\theta(W_n)]^2.$$

Questão 2(Caderno)

Sejam $X_1, X_2, X_3, X_4 e X_5$ uma amostra com distribuição uniforme, isto é, $X_i \sim U(0,1)$. Obter:

a) $E[X_{(5)}]$;

b) $E[X_{(1)}]$;

c) $E[X_{(3)}]$.

SOLUÇÃO

Algumas considerações iniciais:

- A função densidade da população uniforme [0,1] é dada por:

$$\begin{aligned} f_X(x) &= \frac{1}{b-a} I_{a,b}(x), \\ &= \frac{1}{1-0} I_{0,1}(x), \\ &= 1. \end{aligned} \tag{2.1}$$

  Temos ainda que a função distribuição

$$F_X(x) \;=\; \int_0^x f(z)dz = x \tag{2.2}$$

- Seja X_1, X_2, ..., X_n, uma amostra aleatória de tamanho n de uma função de distribuição $F(.)$. Então $Y_1 \leq Y_2 \leq \ldots \leq Y_\alpha \leq \ldots \leq Y_n$, em que Y_α é a X_i considerada em ordem crescente de magnitude chamada estatística de ordem correspondente à amostra X_1, X_2, ..., X_n. Considere $y \in \mathbb{R}$ fixado, e seja $Z_i = I_{(-\infty,y]}(X_i)$, isto é, $Z_i = 1$ se $X_i \in (-\infty,y]$ e zero caso contrário. Então, a variável aleatória Z_i tem distribuição de Bernoulli de parâmetro $p = F_X(y)$. Com isto,

$$\sum_{i=1}^{n} Z_i = \text{número de } X_i{}'s \leq y,$$

  tem distribuição binomial com parâmetros n e $F_X(y)$. Assim, a função de distribuição marginal de Y_α, $\alpha = 1,2,\ldots,n$ é dado por:

$$\begin{aligned} F_{Y_\alpha}(y) &= P[Y_\alpha \leq y] = P\left[\sum_{i=1}^{n} Z_i \geq \alpha\right], \\ F_{Y_\alpha}(y) &= \sum_{j=\alpha}^{n} \binom{n}{j} [F(y)]^j [1-F(y)]^{n-j}, \end{aligned} \tag{2.3}$$

  ou seja, deve-se garantir que $Y_\alpha \leq y$, que em outras palavras, que ao menos α valores dos $X_i{}'s$ sejam menores ou iguais a y, como observado na Figura 2.1.

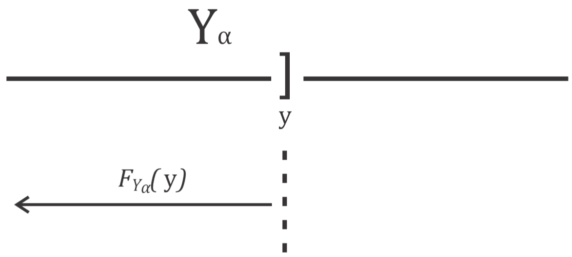


Figura 2.1: Reta da função distribuição

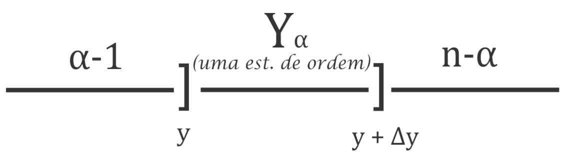


Figura 2.2: Reta da função densidade

Para se determinar a função densidade de probabilidade de Y_α, será usado o auxílio da Figura 2.2.

Nos cálculos apresentados, será utilizada a função de probabilidade multinomial, com função conjunta dada por:

$$P(k_1,k_2,\dots,k_m) = \frac{n!}{k_1!k_2!\dots k_m!}p_1^{k_1}p_2^{k_2}\dots p_m^{k_m},$$

com $\sum_{i=1}^{m} p_i = 1$ e $\sum_{i=1}^{m} k_i = n$, $k_i \in \mathbb{N}$, $0 \leq k_i \leq n$.
Assim, a probabilidade de $\alpha - 1$ $X_i's$ estarem contidos no intervalo $(-\infty,y]$ é $p_1 = F(y)$; a probabilidade de 1 X_i estar contido em $(y,y+\Delta y]$, é $p_2 = F(y+\Delta y) - F(y)$; e a probabilidade de $n-\alpha$ $X_i's$ estarem contidos no intervalo $(y+\Delta y,\infty)$ é $p_3 = 1 - F(y+\Delta y)$. Portanto,

$$\begin{aligned}
f_{Y_\alpha}(y) &= \lim_{\Delta y\to 0}\frac{F_{Y_\alpha}(y+\Delta y) - F_{Y_\alpha}(y)}{\Delta y} = \lim_{\Delta y\to 0}\frac{P[y < Y_\alpha \leq y+\Delta y]}{\Delta y},\\
&= \lim_{\Delta y\to 0} P[(\alpha-1) \text{ dos } X_i \leq y;\ 1\ X_i \text{ em } (y,y+\Delta y];\\
&\quad (n-\alpha) \text{ dos } X_i > y+\Delta y]/\Delta y,\\
&= \frac{n!}{(\alpha-1)!(n-\alpha)!}[F(y)]^{\alpha-1}[1-F(y)]^{n-\alpha}\times\\
&\quad \times \underbrace{\lim_{\Delta y\to 0}\frac{F(y+\Delta y)-F(y)}{\Delta y}}_{f(y)},\\
f_{Y_\alpha}(y) &= \frac{n!}{(\alpha-1)!(n-\alpha)!}[F(y)]^{\alpha-1}[1-F(y)]^{n-\alpha}f(y). \qquad (2.4)
\end{aligned}$$

Utilizando (2.3), pode-se obter a função de distribuição das estatísticas de ordem Y_1 e Y_n. Para Y_α com $\alpha = 1$, tem-se

$$F_{Y_1}(y) = \sum_{j=1}^{n}\binom{n}{1}[F(y)]^1[1-F(y)]^{n-1}.$$

Observando que

$$\binom{n}{0}[F(y)]^0[1-F(y)]^{n-0} + F_{Y_1}(y) = 1,$$

e que,

$$\binom{n}{0}[F(y)]^0[1-F(y)]^{n-0} = [1-F(y)]^n,$$

então, a função de distribuição $F_{Y_1}(y)$ será:

$$F_{Y_1}(y) = 1 - [1-F(y)]^n. \tag{2.5}$$

Derivando (2.5) com relação a y, encontra-se a função densidade, ou seja,

$$f_{Y_1} = \frac{\partial F_{Y_1}(y)}{\partial y} = nf(y)[1-F(y)]^{n-1} \tag{2.6}$$

Já a função de distribuição para $F_{Y_n}(y)$, utilizando (2.3), é

$$\begin{aligned}
F_{Y_n}(y) &= \sum_{j=n}^{n}\binom{n}{j}[F(y)]^j[1-F(y)]^{n-j},\\
F_{Y_n}(y) &= \binom{n}{n}[F(y)]^n[1-F(y)]^{n-n},\\
F_{Y_n}(y) &= [F(y)]^n,
\end{aligned} \tag{2.7}$$

em que derivando (2.7) com relação a y, encontra-se a função densidade,

$$f_{Y_n} = \frac{dF_{Y_n}(y)}{dy} = nf(y)[F(y)]^{n-1}. \tag{2.8}$$

(a) Esperança de $X_{(5)}$:
De acordo com (2.2) e (2.7), temos a função distribuição de $X_{(5)}$ dada por

$$F(x) = [x]^5 = x^5, \tag{2.9}$$

sendo a função densidade, usando as expressões (2.9), (2.1), $n = 5$ em (2.8), temos

$$f_{X_{(5)}}(x) = 5 \times 1 \times [x]^{5-1} = 5x^4 \qquad \text{para} \qquad 0 < x < 1. \tag{2.10}$$

Assim,

$$\begin{aligned}
E[X_{(5)}] &= \int_0^1 x f_X(5) dx,\\
&= \int_0^1 x(5x^4)dx,\\
&= \int_0^1 5x^5 dx,\\
&= 5\left[\frac{x^6}{6}\right]_0^1 = \frac{5}{6}.
\end{aligned} \tag{2.11}$$

(b) Esperança de $X_{(1)}$:
De acordo com as expressões (2.1) e (2.6), temos a função distribuição de $X_{(1)}$ dada por

$$F_{X_{(1)}}(x) = 1 - [1-x]^5, \tag{2.12}$$

sendo a função densidade de $X_{(1)}$, usando as expressões (2.6), (2.1) e $n = 5$, então

$$f_{X_{(1)}}(x) = 5 \times 1 \times [1-x]^4 \qquad \text{para} \qquad 0 < x < 1. \tag{2.13}$$

Assim,

$$
\begin{aligned}
E[X_{(1)}] &= \int_0^1 x f_{X_{(1)}}(x)dx, \\
&= \int_0^1 x \times 5[1-x]^4 dx, \\
&= 5\int_0^1 x(1-x)^4 dx,
\end{aligned}
$$

usando o binômio de Newton:

$$
\begin{aligned}
(a+b)^4 &= a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4, \\
(1-x)^4 &= 1 + 4x + 6x^2 - 4x^3 + x^4,
\end{aligned}
$$

temos

$$
\begin{aligned}
E[X_{(1)}] &= 5\int_0^1 x[1 - 4x + 6x^2 - 4x^3 + x^4]dx, \\
&= 5\int_0^1 [x - 4x^2 + 6x^3 - 4x^4 + x^5]dx, \\
&= 5\left[\frac{x^2}{x} - \frac{4x^3}{3} + \frac{6x^4}{4} - \frac{4x^5}{5} + \frac{x^6}{6}\right]_0^1, \\
&= 5\left[\frac{1}{x} - \frac{4}{3} + \frac{6}{4} - \frac{4}{5} + \frac{1}{6}\right]_0^1, \\
&= \left[\frac{1}{30}\right], \\
&= \frac{1}{6}.
\end{aligned} \tag{2.14}
$$

(b) Esperança de $X_{(3)}$:
Usando a expressão (2.4) para $\alpha = 3$, temos

$$
\begin{aligned}
f_{X_{(3)}}(x) &= \frac{5!}{(3-1)!(5-3)!} \times 1 \times [x]^{3-1}[1-x]^{5-3}, \\
&= \frac{5.4.3.\cancel{2}!}{2!\,\cancel{2}!} x^2[1-x]^2, \\
&= 30x^2[1-x]^2.
\end{aligned}
$$

Calculando a esperança para $X_{(3)}$,

$$
\begin{aligned}
E[X_{(3)}] &= \int_0^1 x f_{X_{(3)}} dx, \\
&= \int_0^1 x[30x^2(1-x)^2]dx, \\
&= \int_0^1 30x^3[1 - 2x + x^2]dx, \\
&= 30\int_0^1 (x^3 - 2x^4 + x^5)dx, \\
&= 30\left[\frac{x^4}{4} - \frac{2x^5}{5} + \frac{x^6}{6}\right]_0^1, \\
&= 30\left[\frac{1}{4} - \frac{2}{5} + \frac{1}{6}\right],
\end{aligned}
$$

$$
\begin{aligned}
&= 30 \times \frac{1}{60}, \\
&= \frac{1}{2}. \qquad (2.15)
\end{aligned}
$$

Um resultado interessante nessa questão, é que para essa situação em que a distribuição da v.a. é uma Uniforme(0,1), a equação (2.4) pode ser reescrita como

$$
f_{X_{(\alpha)}}(x) = \frac{\Gamma(n+1)}{\Gamma(\alpha)\Gamma(n-\alpha+1)} x^{\alpha-1}(1-x)^{(n-\alpha+1)-1}, \qquad (2.16)
$$

que tem distribuição beta(α,$n-\alpha+1$). A partir disso, podemos deduzir que

$$
E[X_{(\alpha)}] = \frac{\alpha}{n+1} \qquad \text{e} \qquad Var[X_{(\alpha)}] = \frac{\alpha(n-\alpha+1)}{(n+1)^2(n+2)}.
$$

Assim,

$$
\begin{aligned}
E[X_{(1)}] &= \frac{\alpha}{n+1} = \frac{1}{5+1} = \frac{1}{6}, \\
E[X_{(3)}] &= \frac{\alpha}{n+1} = \frac{3}{5+1} = \frac{1}{2}, \\
E[X_{(5)}] &= \frac{\alpha}{n+1} = \frac{5}{5+1} = \frac{5}{6}.
\end{aligned}
$$

Questão 3 (Caderno)

Qual a probabilidade de em 400 lances de uma moeda honesta sair 281 e 293 casos?

SOLUÇÃO

Sabemos que esse problema é um experimento em que a população tem distribuição Binomial. Porém, se torna não trivial se feita por esta distribuição. Assim, como auxílio do teorema 21 (MOOD, 1974, pág. 120) poderemos aproximar a distribuição binomial pela distribuição normal, do qual temos,

$$
\begin{aligned}
p\left[a \leq \frac{X-np}{\sqrt{npq}} \leq b\right] &= P[np + a\sqrt{npq} \leq X \leq np + b\sqrt{npq}], \\
&\rightarrow \Phi(b) - \Phi(a) \qquad n \rightarrow \infty. \qquad (2.17)
\end{aligned}
$$

Calculando $P[281 \leq X \leq 293]$ pela distribuição binomial, temos

$$
P[281 \leq X \leq 293] = \sum_{j=281}^{293} \binom{400}{j} \left(\frac{1}{2}\right)^j \left(\frac{1}{2}\right)^{400-j},
$$

esse resultado sem o uso do computador, demandaria algum tempo para obter o resultado.

Agora, usando a aproximação à normal,

$$
\begin{aligned}
P[281 \leq X \leq 293] &= \Phi\left(\frac{293 - 400 \times 1/2}{\sqrt{400 \times 1/2 \times 1/2}}\right) - \Phi\left(\frac{281 - 400 \times 1/2}{\sqrt{400 \times 1/2 \times 1/2}}\right), \\
&= \Phi\left(\frac{93}{10}\right) - \Phi\left(\frac{81}{10}\right), \\
&\approx \Phi(9{,}3) - \Phi(8{,}1), \\
&\approx 2{,}220446 \times 10^{-16} = 0
\end{aligned}
$$

Questão 5.22 (Casella, pág. 259)

Sejam X e Y variáveis aleatórias *iid* N(0,1), e define $Z = \min(X,Y)$. Prove que $Z^2 \sim \chi^2_1$.
SOLUÇÃO
Calculando,

$$\begin{aligned} F_{Z^2}(z) &= P[(\min(X,Y))^2 < z] = P[-\sqrt{z} \le \min(X,Y) \le \sqrt{z}], \\ &= P[\min(X,Y) < \sqrt{z}] - P[\min(X,Y) < -\sqrt{z}], \\ &= [1 - P[\min(X,Y) > \sqrt{z}]] - [1 - P[\min(X,Y) > -\sqrt{z}]], \\ &= P[\min(X,Y) > -\sqrt{z}] - P[\min(X,Y) > \sqrt{z}], \end{aligned}$$

considerando que X e Y são independentes, temos

$$F_{Z^2} = P(X > -\sqrt{z})P(Y > -\sqrt{z}) - P(X > \sqrt{z})P(Y > \sqrt{z}),$$

considerando ainda que X e Y são identicamente distribuídos, implica que essas variáveis aleatórias têm mesma distribuição, isto é, $P(X > \sqrt{z}) = P(Y > \sqrt{z})$. Assim,

$$F_{Z^2}(z) = (1 - F_X(-\sqrt{z}))^2 - (1 - F_X(\sqrt{z}))^2.$$

Percebemos ainda que $1 - F_X(\sqrt{z}) = F_X(-\sqrt{z})$, então

$$\begin{aligned} F_{Z^2}(z) &= [1 - F_X(-\sqrt{z})]^2 - [F_X(-\sqrt{z})]^2, \\ &= 1 - 2F_X(-\sqrt{z}) + [F_X(-\sqrt{z})]^2 - [F_X(-\sqrt{z})]^2, \\ &= 1 - 2F_X(-\sqrt{z}). \end{aligned}$$

Encontrando a função densidade de Z^2, temos

$$\begin{aligned} f_{Z^2}(z) &= \frac{d}{dz}F_{Z^2}(z) = \frac{d}{dz}[1 - 2F_X(-\sqrt{z})], \\ &= -2f_X(-\sqrt{z}) \times (-1/2z^{1/2-1}), \\ &= \frac{\cancel{2}}{\cancel{2}} f_X(-\sqrt{z}) \times z^{-1/2}, \end{aligned}$$

como $f_X(-\sqrt{z}) = \frac{1}{\sqrt{2\pi}} e^{-1/2(\sqrt{2})^2}$, então

$$f_{Z^2}(z) = \frac{1}{\sqrt{2\pi}} e^{-1/2z} z^{-1/2},$$

que é a distribuição de qui-quadrado com 1 grau de liberdade.

Questão 5.23 (Casella, pág. 260)

Se $\overline{X}_1$ e $\overline{X}_2$ são as médias de duas amostras de tamanhos independentes n de uma população com variância σ^2, encontre um valor para n de modo que $P[|\overline{X}_1 - \overline{X}_2| < \sigma/5] \approx 0{,}99$. Justifique seus cálculos.
SOLUÇÃO
Usando o teorema do limite central, temos que $\overline{X}_1 \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ e $\overline{X}_2 \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. Considerando que $\overline{X}_1 - \overline{X}_2 \sim N(\mu^*, \sigma^{2*})$,

$$\begin{aligned} \mu^* &= E[\overline{X}_1 - \overline{X}_2] = E[\overline{X}_1] - E[\overline{X}_2] = \mu - \mu = 0, \\ \sigma^{2*} &= Var(\overline{X}_1) + Var(\overline{X}_2) - \underbrace{2Cov(\overline{X}_1, \overline{X}_2)}_{\text{0, pois } \overline{X}_1 \text{ e } \overline{X}_2 \text{ são independentes.}}, \\ \sigma^{2*} &= \frac{\sigma^2}{n} + \frac{\sigma^2}{n} = 2\frac{\sigma^2}{n}. \end{aligned}$$

Podemos obter o valor de n da seguinte forma,

$$\begin{aligned} 0{,}99 &\approx P[|\overline{X}_1 - \overline{X}_2| < \sigma/5], \\ &= P\left[\frac{-\sigma/5}{\sigma/\sqrt{n/2}} < \frac{\overline{X}_1 - \overline{X}_2}{\sigma/\sqrt{n/2}} < \frac{\sigma/5}{\sigma/\sqrt{n/2}}\right]. \end{aligned}$$

Sabemos que agora $\frac{(\overline{X}_1 - \overline{X}_2)-0}{\sigma/\sqrt{n/2}} \sim N(0{,}1)$, portanto,

$$\begin{aligned} 0{,}99 &\approx P\left[-\frac{\cancel{\sigma}}{5} \times \frac{\sqrt{n/2}}{\cancel{\sigma}} < Z < \frac{\cancel{\sigma}}{5} \times \frac{\sqrt{n/2}}{\cancel{\sigma}}\right], \\ &\approx P\left[-\frac{1}{5}\sqrt{\frac{n}{2}} < Z < \frac{1}{5} \times \sqrt{\frac{n}{2}}\right] \end{aligned}$$

Dessa forma, podemos obter o quantil,

$$P[z > q] \approx 0{,}005,$$

sendo $q = \frac{1}{5}\sqrt{\frac{n}{2}}$, verificando numa tabela da normal padrão, ou no programa R, temos que

```
> qnorm(0.995)
[1] 2.5758929
```

podemos então igualar,

$$\begin{aligned} \frac{\sqrt{n}}{5\sqrt{2}} &= 2{,}578929, \\ \sqrt{n} &= 2{,}578929 \times 5\sqrt{2}, \\ \sqrt{n} &= 18{,}21431337, \\ n &= (18{,}21431337)^2, \\ n &= 331{,}76 \approx 332. \end{aligned}$$

Questão 5.31 (Casella, pág. 260)

Suponha que $\overline{X}$ seja a média de 100 observações de uma população com média μ e variância $\sigma^2 = 9$. Encontre o limite entre os quais $\overline{X} - \mu$ estará, com probabilidade de elo menos 0,90. Utilize a desigualdade de Chebychev e o teorema do limite central, e comente cada um deles.

SOLUÇÃO

Temos a desigualdade de Chebychev:

$$P[|\overline{X} - \mu| \geq t\sigma_{\overline{X}}] \leq \frac{1}{t^2},$$

ou

$$P[|\overline{X} - \mu| < t\sigma_{\overline{X}}] \leq 1 - \frac{1}{t^2}.$$

Temos que $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n} = \frac{9}{100} \Rightarrow \sigma_{\overline{X}} = \frac{3}{10}$, então

$$P[|\overline{X} - \mu| < \frac{3t}{10}] > 1 - \frac{1}{t^2}.$$

Queremos encontrar os limites entre os quais $\overline{X} - \mu$ esteja com probabilidade de pelo menos 0,90. Assim,

$$1 - \frac{1}{t^2} \geq 0{,}90 \Rightarrow \frac{1}{t^2} \geq 1 - 0{,}90 \Rightarrow t^2 \geq \frac{1}{0{,}10} \Rightarrow t \Rightarrow \sqrt{10}.$$

Os limites então podem ser encontrados usando:

$$P\left[|\overline{X}-\mu|<\frac{3t}{10}\right]=P\left[-\frac{3t}{10}<\overline{X}-\mu<\frac{3t}{10}\right],$$

e como $t=\sqrt{10}$, então

$$\begin{aligned}P\left[|\overline{X}-\mu|<\frac{3t}{10}\right] &= P\left[-\frac{3\sqrt{10}}{10}<\overline{X}-\mu<\frac{3\sqrt{10}}{10}\right],\\ &= P[-0{,}94868<\overline{X}-\mu<0{,}94868]\geq 0{,}90.\end{aligned}$$

Usando o teorema do limite central, temos que $\overline{X}\sim N(\mu,\sigma^2_{\overline{X}})$, sendo $\sigma^2_{\overline{X}}=\frac{\sigma^2}{n}=\frac{9}{100}\Rightarrow\sigma^2_{\overline{X}}=0{,}09$, e sabendo que $z=\left(\frac{\overline{X}-\mu}{\sqrt{0{,}09}}\right)\sim N(0{,}1)$, portanto,

$$\begin{aligned}P[\alpha_1<Z<\alpha_2]&\geq 0{,}90,\\ P[-1{,}645<Z<1{,}645]&\geq 0{,}90,\\ P[-1{,}645<\frac{\overline{X}-\mu}{\sqrt{0{,}09}}<1{,}645]&\geq 0{,}90,\\ P[-1{,}645\sqrt{0{,}09}<\overline{X}-\mu<1{,}645\sqrt{0{,}09}]&\geq 0{,}90,\\ P[-0{,}4935<\overline{X}-\mu<0{,}4935]&\geq 0{,}90.\end{aligned}$$

Questão 5.34 (Casella, pág. 261)

Sejam $X_1,X_2,\ldots,X_n$ uma amostra aleatória de uma população com média μ e variância σ^2. Mostre que

$$E\left[\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}\right]=0 \qquad \text{e} \qquad Var\left[\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}\right]=1.$$

Assim, a normalização de $\overline{X}_n$ no teorema de limite central apresenta variáveis aleatórias que têm as mesmas média e variância que a distribuição limite $N(0{,}1)$.

SOLUÇÃO

Sabemos que $E[\overline{X}]=E\left[\frac{\sum X}{n}\right]=\frac{1}{n}E[\sum X]=\frac{nE[X]}{n}=E[X]=\mu$ e $Var[\overline{X}]=Var\left[\frac{\sum X}{n}\right]=\frac{1}{n^2}Var[\sum X]=\frac{nVar[X]}{n^2}=\frac{Var[X]}{n}$. Assim,

$$E\left[\frac{\overline{X}-\mu}{\frac{\sigma}{n}}\right]=\frac{\sqrt{n}}{\sigma}E[\overline{X}-\mu]=\frac{\sqrt{n}}{\sigma}[E[\overline{X}]-\mu],$$

como $E[\overline{X}]=\mu$, segue que

$$E\left[\frac{\overline{X}-\mu}{\frac{\sigma}{n}}\right]=\frac{\sqrt{n}}{\sigma}[\mu-\mu]=0.$$

Para a variância, temos

$$Var\left[\frac{\overline{X}-\mu}{\frac{\sigma}{n}}\right]=\frac{n}{\sigma^2}Var[\overline{X}-\mu]=\frac{n}{\sigma^2}Var[\overline{X}],$$

como $Var[\overline{X}]=\frac{\sigma^2}{n}$, então

$$Var\left[\frac{\overline{X}-\mu}{\frac{\sigma}{n}}\right]=\frac{n}{\sigma^2}\times\frac{\sigma^2}{n}=1.$$

Questão 5.44 (Casella, pág. 263)

Sejam X_i, $i = 1{,}2, \ldots$, variáveis aleatórias de Bernoulli(p) independentes e seja $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

a) Mostre que $\sqrt{n}(Y_n - p) \to N(0{,}p(1-p))$ em distribuição;

b) Mostre que para $p \neq 1/2$, a estimativa da variância de $Y_n(1-Y_n)$, satisfaz $\sqrt{n}[Y_n(1-Y_n)-p(1-p)] \to N(0{,}(1-2p)^2 p(1-p))$ em distribuição.

SOLUÇÃO

(a) Sabemos que $X_i \sim Bernoulli(p)$, com esperança $E[X] = p$ e $Var[pq]$. A distribuição de $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, pode ser encontrado pela função geradora de momentos, que segue,

$$\begin{aligned} E\left[e^{t\sum X}\right] &= E\left[e^{t[X_1+X_2+\ldots+X_n]}\right], \\ &= E[e^{tX_1}]E[e^{tX_2}]\ldots E[e^{tX_n}]. \end{aligned}$$

A função geradora de $X \sim Bernoulli(p)$,

$$\begin{aligned} E[e^{tX}] &= e^{t0}(1-p) + e^t p, \\ &= (q + pe^t), \quad q = 1-p. \end{aligned}$$

Assim,

$$\begin{aligned} E\left[e^{t\sum X}\right] &= (q+pe^t)(q+pe^t)\ldots(q+pe^t), \\ &= (q+pe^t)^n. \end{aligned}$$

Portanto, $W = \sum X \sim Binomial(n{,}p)$, sabendo que $E[W] = np$ e $Var[W] = npq$. Dessa forma, $Y_n = \frac{W}{n}$. Como Y_n é uma média amostral, pelo teorema do limite central, esta estimativa aproximadamente $(n \to n)$ normal, com esperança

$$E\left[\frac{W}{n}\right] = \frac{E[W]}{n} = \frac{np}{n} = p,$$

e variância,

$$Var\left[\frac{W}{n}\right] = \frac{Var[W]}{n^2} = \frac{npq}{n^2} = \frac{pq}{n},$$

e portanto,

$$Y_n \sim N\left(p, \frac{pq}{n}\right).$$

Podemos então concluir que $\sqrt{n}(Y_n - p)$ é uma combinação linear da estatística Y_n. Como essa estatística tem distribuição normal padrão, a combinação linear terá também distribuição normal, com esperança

$$\begin{aligned} E[\sqrt{n}(Y_n - p)] &= E[\sqrt{n}Y_n] - \sqrt{n}p, \\ &= \sqrt{n}E[Y_n] - \sqrt{n}p, \\ &= \sqrt{n}p - \sqrt{n}p, \\ &= 0, \end{aligned}$$

e variância,

$$\begin{aligned} Var[\sqrt{n}(Y_n - p)] &= Var[\sqrt{n}Y_n - \sqrt{n}p] = Var[\sqrt{n}Y_n], \\ &= nVar[Y_n], \\ &= n\frac{pq}{n}, \\ &= np(1-p), \quad q = 1-p. \end{aligned}$$

Podemos então concluir que em distribuição,

$$\sqrt{n}(Y_n - p) \to N(0{,}p(1-p)).$$

(b) Esse exemplo pode ser resolvido fazendo analogia ao exemplo 10.1.15 (Casella, p. 425, português).

2.2 Lista 2

Questão do Caderno

Seja a variável aleatória $X \sim exp(\alpha)$. Então, calcule a distribuição de $W = \frac{X - 1/\alpha}{1/\alpha}$.

SOLUÇÃO

Podemos expressar $W = \alpha X - 1$. Como $X \sim exp(\alpha)$, ou seja,

$$f(x;\alpha) = \alpha e^{-\alpha x}, \quad x > 0,$$

então a função densidade de W pode ser expressa,

$$\begin{aligned} f(w) &= \alpha e^{-\alpha\left(\frac{w+1}{\alpha}\right)} \times \frac{1}{\alpha}, \\ &= e^{-(w+1)} I_{(-1,\infty)}(w). \end{aligned}$$

Observamos $f(w)$ tem distribuição exponencial, porém deslocando w para o intervalo de -1 e ∞.

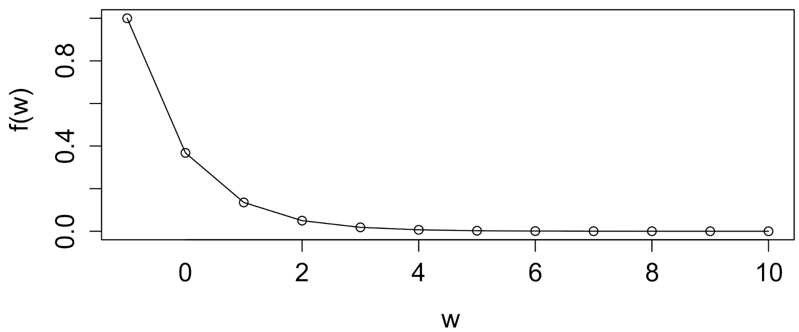


Figura 2.3: Gráfico da função densidade $f_W(w)$.

Vamos mostrar que $f(w)$ é uma função densidade de probabilidade com média 0 e variância 1, já que w é uma variável normalizada. Assim,

$$\begin{aligned} \int_{-1}^{\infty} f(w)dw &= \int_{-1}^{\infty} e^{-(w+1)}dw, \\ &= e^{-1}\int_{-1}^{\infty} e^{-w}dw, \\ &= e^{-1}\left[0 + e^{1}\right], \\ &= 1. \end{aligned}$$

Questão do Caderno

Calcular aproximadamente $f(x) = e^{-x^2}$, para $x = 0{,}25$.

SOLUÇÃO

$$f(x) = e^{-x^2}$$

$$f'(x) = -2xe^{-x^2}$$

$$f''(x) = 4x^2e^{-x^2} - 2e^{-x^2}$$

Aproximação de primeira ordem (Série de Taylor):

$$\begin{aligned} f(x) &\approx f(x_0) + f'(x_0)(x - x_0), \\ &\approx e^{-x_0^2} - 2x_0e^{-x_0^2}(x - x_0), \end{aligned}$$

para $x_0 = 0$

$$f(x = 0{,}25) \quad \approx \quad 1.$$

Aproximação de segunda ordem (Série de Taylor):

$$
\begin{aligned}
f(x) &\approx f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)(x_0 - x)^2}{2},\\
&\approx 1 - x^2,
\end{aligned}
$$

para $x_0 = 0$

$$f(x = 0{,}25) \quad \approx \quad 0{,}9375.$$

Usando o R, temos que $f(x) = 0{,}9394131$, e usando a aproximação de segunda ordem $f(x) = 0{,}9375$.

Questão do Caderno

Seja as variáveis aleatória $X \sim Exp(\alpha)$ e $Y \sim Exp(\beta)$, independentes, e $Z = \frac{X}{Y}$. Então, calcule a $E[Z]$ e $Var[Z]$.

SOLUÇÃO

Usando um resultado do Mood (1974, pág. 187), temos

$$f_Z(z) = \int_0^\infty |y| f_{X,Y}\left(uy, y\right).$$

Como X e Y são independentes, então $f_{X,Y}(x,y) = f_X(x) f_Y(y)$. Assim,

$$
\begin{aligned}
f_Z(z) &= \int_0^\infty y\alpha e^{-\alpha(zy)} \beta e^{-\beta y} dy,\\
&= \alpha\beta \int_0^\infty y e^{-y(\alpha z + \beta)} dy,
\end{aligned}
$$

calculando a integral por partes, considerando $\int u dv = uv - \int v du$, então $u = y$, $du = dy$, $dv = e^{-y(\alpha z+\beta)}dy$ e $v = -e^{-y(\alpha z+\beta)}/(\alpha z + \beta)$. Assim,

$$
\begin{aligned}
f_Z(z) &= \frac{\alpha\beta}{\alpha z + \beta}\left[-y e^{-y(\alpha z+\beta)} + \int_0^\infty e^{-y(\alpha z+\beta)} dy\right]_0^\infty,\\
&= \frac{\alpha\beta}{\alpha z + \beta}\left[-y e^{-y(\alpha z+\beta)} - \frac{e^{-y(\alpha z+\beta)}}{\alpha z + \beta}\right]_0^\infty,\\
&= \frac{\alpha\beta}{\alpha z + \beta}\left[\left(\lim_{y\to\infty} \frac{-y}{e^{y(\alpha z+\beta)}} + \lim_{y\to\infty} \frac{-1}{(\alpha z + \beta) e^{y(\alpha z+\beta)}}\right) - \left(0 - \frac{1}{(\alpha z + \beta)}\right)\right].
\end{aligned}
$$

Observando o limite $\lim_{y\to\infty} \frac{-y}{e^{y(\alpha z+\beta)}} = \frac{-\infty}{\infty}$, estamos com um problema de indeterminação. Assim, a solução será usar a regra de L'Hopital, ou seja, considerando $f(y) = -y$ e $g(y) = e^{y(\alpha z+\beta)}$, então

$$\lim_{y\to\infty} \frac{f'(y)}{g'(y)} = \lim_{y\to\infty} \frac{-1}{(\alpha z + \beta)e^{y(\alpha z+\beta)}} = 0.$$

Podemos então retornar a densidade,

$$
\begin{aligned}
f_Z(z) &= \frac{\alpha\beta}{\alpha z + \beta}\left[(0 + 0) + \frac{1}{\alpha z + \beta}\right],\\
&= \frac{\alpha\beta}{(\alpha z + \beta)^2} I_{(0,\infty)}(z).
\end{aligned}
$$

Se considerarmo $\alpha = \beta$, obteremos

$$f_Z(z) = \frac{1}{(1 + y)^2},$$

como observado no Mood (1974, pág. 208), sendo esta uma densidade com média infinita.

$$E[Z] = \alpha\beta \left(\frac{\frac{\beta}{\alpha z+\beta} + \log(\alpha z)}{\alpha^2} \right)_0^\infty = \infty$$

$$E[Z^2] = \int_0^\infty \alpha\beta \left(\frac{\frac{-\beta^2}{\alpha z+\beta} - 2\beta\log(\alpha z+\beta) + \alpha z}{\alpha^3} \right)_0^\infty = \infty$$

$$Var[Z] = \infty$$

Questão do Caderno

Seja $X_1, X_2, \ldots, X_n$ uma amostra de uma população com distribuição Poisson, isto é

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x > 0, \quad \lambda > 0.$$

Prove que $E[X] = Var[X] = \lambda$.
SOLUÇÃO

$$\begin{aligned}
E[X] &= \sum_{x=0}^{\infty} x P(X = x), \\
&= \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!}, \\
&= 0 \times \frac{\lambda^0 e^{-\lambda}}{0!} + \sum_{x=1}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!}, \\
&= \sum_{x=1}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x(x-1)!}, \\
&= \sum_{x=1}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-1)!}, \\
&= e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda}{\lambda} \frac{\lambda^x}{(x-1)!}, \\
&= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^x \lambda^{-1}}{(x-1)!}, \\
&= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!},
\end{aligned}$$

observe que $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots = \sum_{j=0}^{\infty} \frac{x^j}{j!}$, então

$$\begin{aligned}
E[X] &= \lambda e^{-\lambda} e^{\lambda}, \\
&= \lambda.
\end{aligned}$$

$$\begin{aligned}
E[X^2] &= \sum_{x=0}^{\infty} x^2 P(X = x), \\
&= \sum_{x=0}^{\infty} x^2 \frac{\lambda^x e^{-\lambda}}{x!}, \\
&= 0^2 \times \frac{\lambda^0 e^{-\lambda}}{0!} + \sum_{x=1}^{\infty} x^2 \frac{\lambda^x e^{-\lambda}}{x!},
\end{aligned}$$

$$\begin{aligned}
&= \sum_{x=1}^{\infty} x^2 \frac{\lambda^x e^{-\lambda}}{x(x-1)!},\\
&= \sum_{x=1}^{\infty} x \frac{\lambda^x e^{-\lambda}}{(x-1)!},\\
&= \sum_{x=1}^{\infty} x \frac{e^{-\lambda}\lambda^x}{(x-1)!},
\end{aligned}$$

fazendo $m = x - 1$, temos

$$\begin{aligned}
&= \sum_{m=0}^{\infty} (m+1) \frac{e^{-\lambda}\lambda^{(m+1)}\lambda^{-1}}{m!},\\
&= \left[\sum_{m=0}^{\infty} m \frac{e^{-\lambda}\lambda^{(m+1)}}{m!} + \sum_{m=0}^{\infty} \frac{e^{-\lambda}\lambda^{(m+1)}}{m!}\right],\\
&= \lambda \left[\underbrace{\sum_{m=0}^{\infty} m \frac{e^{-\lambda}\lambda^{m}}{m!}}_{E[X]} + \underbrace{\sum_{m=0}^{\infty} \frac{e^{-\lambda}\lambda^{m}}{m!}}_{=1}\right],\\
&= \lambda[\lambda+1].
\end{aligned}$$

Assim,

$$\begin{aligned}
Var[X] &= E[X^2] - (E[X])^2,\\
&= \lambda^2 + \lambda - \lambda^2,\\
&= \lambda.
\end{aligned}$$

Questão do caderno

Mostrar que se $X \sim Gama\left(\frac{1}{2},\frac{1}{2}\right)$, então $\alpha X \sim Gama\left(\frac{1}{2},\frac{1}{2c}\right)$.

SOLUÇÃO 1

Seja a distribuição gama com a parametrização r e λ,

$$f(x;r,\lambda) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, \quad x > 0.$$

Seja a transformação $Y = cX$, sendo c uma constante, então

$$\begin{aligned}
f(y) &= \frac{\lambda^r}{\Gamma(r)} \left(\frac{y}{c}\right)^{r-1} e^{-\lambda\frac{y}{c}} \left(\frac{1}{c}\right),\\
&= \frac{\lambda^r}{\Gamma(r)} \left(\frac{1}{c}\right)^{r-1} y^{r-1} e^{-\frac{\lambda}{c}y} \left(\frac{1}{c}\right),\\
&= \frac{\lambda^r}{\Gamma(r)} \frac{\left(\frac{1}{c}\right)^{r}}{\left(\frac{1}{c}\right)} y^{r-1} e^{-\frac{\lambda}{c}y} \left(\frac{1}{c}\right),\\
&= \frac{\frac{\lambda}{c}^r}{\Gamma(r)} y^{r-1} e^{-\frac{\lambda}{c}y},
\end{aligned}$$

sendo portanto, $Y \sim Gama(r,\frac{\lambda}{c})$. Se considerarmos $r = \lambda = \frac{1}{2}$, então $Y \sim Gama(\frac{1}{2},\frac{1}{2c})$.

SOLUÇÃO 2

Sabendo que $X \sim Gama(r,\lambda)$, para descobrir a função densidade de $Y = cX$, sendo c uma constante, então

$$F_Y(y) \quad = \quad F_Y(Y \leq y),$$

$$\begin{aligned} &= F_X(cX \leq y), \\ &= F_X\left(X \leq \frac{y}{c}\right), \\ &= F_X\left(\frac{y}{c}\right), \end{aligned}$$

como queremos encontrar a função densidade, basta derivar a função acumulada. Assim,

$$\begin{aligned} \frac{d}{dy}F_Y(y) &= \frac{d}{dy}F_X\left(\frac{y}{c}\right), \\ f_Y(y) &= f_X\left(\frac{y}{c}\right) \times \frac{1}{c}. \end{aligned}$$

Dessa forma, chegaremos ao resultado da **SOLUÇÃO 1**.

Questão de Caderno

Calcule a $Var[S_n^2]$, sabendo que $S_n^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}$.

SOLUÇÃO

Vamos fazer algumas definições:

$$\overline{X} = \frac{\sum_{i=1}^n X_i}{n}, \quad S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1},$$

$$\mu = \frac{\sum_{i=1}^N X_i}{N}, \quad \mu_k = \frac{\sum_{i=1}^N (a_i - \mu)^2}{N}, \quad \mu_k' = \frac{\sum_{i=1}^N a_i^k}{N}.$$

Seja a transformação $Z_i = X_i - \mu$ para $i = 1{,}2, \ldots ,n$, então $E[Z_i] = 0$. Reescrevendo S^2 em função de Z_i, chamando $X_i = Z_i + \mu$ e $\overline{X} = \frac{\sum_{i=1}^n (X_i - \mu) + \mu}{n} = \frac{\sum_{i=1}^n (Z_i) + \mu}{n} = \frac{\sum_{i=1}^n Z_i}{n} + \mu$. Podemos então redefinir:

$$\begin{aligned} S^2 &= \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}, \\ &= \frac{\sum_{i=1}^n (Z_i + \cancel{\mu} - \frac{\sum_{i=1}^n Z_i}{n} + \cancel{\mu})^2}{n-1}, \\ &= \frac{1}{n-1}\sum_{i=1}^n \left(Z_i - \frac{\sum_{i=1}^n Z_i}{n}\right)^2, \\ &= \frac{1}{n-1}\sum_{i=1}^n \left[Z_i^2 - 2Z_i\frac{\sum_{i=1}^n Z_i}{n} + \left(\frac{\sum_{i=1}^n Z_i}{n}\right)^2\right], \\ &= \frac{1}{n-1}\left[\sum_{i=1}^n Z_i^2 - 2\frac{\sum_{i=1}^n Z_i \sum_{i=1}^n Z_i}{n} + n\left(\frac{\sum_{i=1}^n Z_i}{n}\right)^2\right], \\ &= \frac{1}{n-1}\left[\sum_{i=1}^n Z_i^2 - 2\frac{(\sum_{i=1}^n Z_i)^2}{n} + \frac{(\sum_{i=1}^n Z_i)^2}{n}\right], \\ &= \frac{1}{n-1}\left[\sum_{i=1}^n Z_i^2 - \frac{(\sum_{i=1}^n Z_i)^2}{n}\right], \\ &= \frac{1}{n(n-1)}\left[n\sum_{i=1}^n Z_i^2 - \left(\sum_{i=1}^n Z_i\right)^2\right] \end{aligned}$$

Dessa forma, temos que a $Var[S^2] = E[(S^2)^2] - (E[S^2])^2 = E[S^4] - \mu_2^2$, precisando apenas encontrar $E[S^4]$. Assim,

$$E[S^4] = E[(S^2)^2],$$

$$
\begin{aligned}
&= E\left[\frac{1}{n(n-1)}\left(n\sum_{i=1}^{n} Z_i^2 - \left(\sum_{i=1}^{n} Z_i\right)^2\right)\right]^2, \\
&= \frac{1}{n^2(n-1)^2} E\left[\left(n\sum_{i=1}^{n} Z_i^2 - \left(\sum_{i=1}^{n} Z_i\right)^2\right)\right]^2, \\
&= \frac{1}{n^2(n-1)^2} E\left[\left(n\sum_{i=1}^{n} Z_i^2\right)^2 - 2n\sum_{i=1}^{n} Z_i^2 \left(\sum_{i=1}^{n} Z_i\right)^2 + \left(\sum_{i=1}^{n} Z_i\right)^4\right], \\
&= \frac{1}{n^2(n-1)^2} E\left[n^2\left(\sum_{i=1}^{n} Z_i^2\right)^2 - 2n\sum_{i=1}^{n} Z_i^2 \left(\sum_{i=1}^{n} Z_i\right)^2 + \left(\sum_{i=1}^{n} Z_i\right)^4\right].
\end{aligned}
$$

Sabendo que $Z_1, Z_2, \ldots, Z_n$ são independentes, então:

$$
\begin{aligned}
E[Z_iZ_j] &= E[Z_i]E[Z_j] = 0, \\
E[Z_i^3Z_j] &= E[Z_i^3]E[Z_j] = E[Z_i^3] \times 0 = 0, \\
E[Z_i^2Z_jZ_k] &= E[Z_i^2]E[Z_j]E[Z_k] = E[Z_i^2] \times 0 \times 0 = 0, \\
E[Z_iZ_jZ_kZ_l] &= E[Z_i]E[Z_j]E[Z_k]E[Z_l] = E[Z_i] \times 0 \times 0 \times 0 = 0, \\
E[Z_i^2Z_j^2] &= E[Z_i^2]E[Z_j^2] = \mu_2\mu_2 = \mu_2^2.
\end{aligned}
$$

Resolvendo a primeira parte do somatório dentro da esperança,

$$
\begin{aligned}
E\left[\left(\sum_{i=1}^{n} Z_i^2\right)^2\right] &= E\left[\left(Z_1^2 + Z_2^2 + \ldots + Z_n^2\right)\left(Z_1^2 + Z_2^2 + \ldots + Z_n^2\right)\right], \\
&= E\left[Z_1^2Z_1^2 + \ldots + Z_n^2Z_n^2 + Z_1^2Z_2^2 + \ldots + Z_1^2Z_n^2 + \right. \\
&\quad \left. + Z_2^2Z_1^2 + Z_2^2Z_3^2 + \ldots + Z_2^2Z_n^2 + \ldots + Z_n^2Z_{n-1}^2\right], \\
&= E\left[\sum_{i=1}^{n} Z_i^4 + \sum_{i\neq j} Z_i^2Z_j^2\right], \\
&= E\left[\sum_{i=1}^{n} Z_i^4\right] + E\left[\sum_{i\neq j} Z_i^2Z_j^2\right], \\
&= \sum_{i=1}^{n} E[Z_i^4] + E\left[\sum_{i=1}^{n} Z_i^2 \sum_{j=1}^{n-1} Z_j^2\right], \\
&= nE[Z_i^4] + \sum_{i=1}^{n}\sum_{j=1}^{n-1} E[Z_i^2Z_j^2], \\
&= n\mu_4 + \sum_{i=1}^{n}\sum_{j=1}^{n-1} E[Z_i^2]E[Z_j^2], \\
&= n\mu_4 + n(n-1)\mu_2\mu_2, \\
&= n\mu_4 + n(n-1)\mu_2^2,
\end{aligned}
$$

Resolvendo a segunda parte do somatório dentro da esperança,

$$
\begin{aligned}
E\left[\left(\sum_{i=1}^{n} Z_i^2\right)\left(\sum_{i=1}^{n} Z_i\right)^2\right] &= E\left[\left(Z_1^2 + Z_2^2 + \ldots + Z_n^2\right)(Z_1 + Z_2 + \ldots + Z_n)^2\right], \\
&= E\left[\sum_{i=1}^{n} Z_i^4\right] + E\left[\sum_{i\neq j} Z_i^2Z_j^2\right] + ,
\end{aligned}
$$

$$
\begin{aligned}
& + \underbrace{2E\left[\sum_{i\neq j} Z_i^3 Z_j\right]}_{=0,\quad \text{pois } E[Z_j]=E[Z_k]=0} + \underbrace{E\left[\sum_{i\neq j\neq k} Z_i^2 Z_j Z_k\right]}_{=0,\quad \text{pois } E[Z_j]=E[Z_k]=0}, \\
&= \sum_{i=1}^{n} E[Z_i^4] + E\left[\sum_{i=1}^{n} Z_i^2 \sum_{j=1}^{n-1} Z_j^2\right], \\
&= nE[Z_i^4] + \sum_{i=1}^{n}\sum_{j=1}^{n-1} E[Z_i^2 Z_j^2], \\
&= n\mu_4 + \sum_{i=1}^{n}\sum_{j=1}^{n-1} E[Z_i^2]E[Z_j^2], \\
&= n\mu_4 + n(n-1)\mu_2\mu_2, \\
&= n\mu_4 + n(n-1)\mu_2^2,
\end{aligned}
$$

Resolvendo a terceira parte do somatório dentro da esperança,

$$
\begin{aligned}
E\left[\left(\sum_{i=1}^{n} Z_i\right)^4\right] &= E\left[(Z_1 + Z_2 + \ldots + Z_n)^2 (Z_1 + Z_2 + \ldots + Z_n)^2\right], \\
&= E\left[\sum_{i=1}^{n} Z_i^4 + 3\sum_{i\neq j} Z_i^2 Z_j^2 + \underbrace{4\sum_{i\neq j} Z_i^3 Z_j + 6\sum_{i\neq j\neq k} Z_i Z_j Z_k + \sum_{i\neq j\neq k\neq l} Z_i Z_j Z_k Z_l}_{=0,\quad \text{pois } E[Z]=0}\right], \\
&= \sum_{i=1}^{n} E[Z_i^4] + 3E\left[\sum_{i=1}^{n} Z_i^2 \sum_{j=1}^{n-1} Z_j^2\right], \\
&= nE[Z_i^4] + 3\sum_{i=1}^{n}\sum_{j=1}^{n-1} E[Z_i^2 Z_j^2], \\
&= n\mu_4 + 3\sum_{i=1}^{n}\sum_{j=1}^{n-1} E[Z_i^2]E[Z_j^2], \\
&= n\mu_4 + 3n(n-1)\mu_2\mu_2, \\
&= n\mu_4 + 3n(n-1)\mu_2^2.
\end{aligned}
$$

Com esses resultados poderemos obter $E[S^4]$, do qual temos

$$
\begin{aligned}
E[S^4] &= \frac{1}{n^2(n-1)^2} E\left[n^2\left(\sum_{i=1}^{n} Z_i^2\right)^2 - 2n\sum_{i=1}^{n} Z_i^2 \left(\sum_{i=1}^{n} Z_i\right)^2 + \left(\sum_{i=1}^{n} Z_i\right)^4\right], \\
&= \frac{1}{n^2(n-1)^2}\left(n^2 E\left[\left(\sum_{i=1}^{n} Z_i^2\right)^2\right] - 2nE\left[\sum_{i=1}^{n} Z_i^2 \left(\sum_{i=1}^{n} Z_i\right)^2\right] + \right. \\
&\quad \left. + E\left[\left(\sum_{i=1}^{n} Z_i\right)^4\right]\right], \\
&= \frac{1}{n^2(n-1)^2}\left[n^2\left(n\mu_4 + n(n-1)\mu_2^2\right) - 2n\left(n\mu_4 + n(n-1)\mu_2^2\right) + n\mu_4 + 3n(n-1)\mu_2^2\right], \\
&= \frac{1}{n^2(n-1)^2}\left[(n^2 - 2n)\left(n\mu_4 + n(n-1)\mu_2^2\right) + n(\mu_4 + 3(n-1)\mu_2^2)\right],
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{n(n-1)^2}\left[(n-2)\left(n\mu_4+n(n-1)\mu_2^2\right)+(\mu_4+3(n-1)\mu_2^2)\right],\\
&= \frac{1}{n(n-1)^2}\left[n^2\mu_4+n^2(n-1)\mu_2^2-2n\mu_4-2n(n-1)\mu_2^2+\mu_4+3\mu_2^2 n-3\mu_2^2\right],\\
&= \frac{1}{n(n-1)^2}\left[\mu_4(n^2+2n+1)+\mu_2^2(n^2(n-1)-2n(n-1)+3n-3)\right],\\
&= \frac{1}{n(n-1)^2}\left[\mu_4[(n-1)(n-1)]+\mu_2^2(n^2-2n+3)(n-1)\right],\\
&= \frac{1}{n(n-1)}\left[(n-1)\mu_4+\mu_2^2(n^2-2n+3)\right]
\end{aligned}
$$

Dessa forma temos que a $Var[S^2]$ segue

$$
\begin{aligned}
Var[S^2] &= E[S^4]-\mu_2^2,\\
&= \frac{(n-1)\mu_4+\mu_2^2(n^2-2n+3)}{n(n-1)}-\mu_2^2,\\
&= \frac{1}{n}\left(\mu_4-\frac{n-3}{n-1}\mu_2^2\right).
\end{aligned}
$$

Questão de Caderno

Teorema fundamental da transformação de probabilidade:

Sejam U uma variável aleatória uniforme $U(0,1)$ e X uma variável aleatória com densidade f e função de distribuição F contínua e invertível, então $X = F^{-1}(U)$ possui densidade f. Sendo F^{-1} a função inversa da função de distribuição F.

Prova

Seja X uma variável aleatória com função de distribuição F e função densidade f. Se $u = F(x)$, então o jacobiano da transformação é $du/dx = F'(x) = f(x)$, em que U é uma variável aleatória uniforme $U(0,1)$, com função densidade $g(u) = 1$, para $0 < u < 1$ e $g(u) = 0$ para outros valores de u. Assim, a variável aleatória $X = F^{-1}(U)$ tem densidade f dada por:

$$f_X(x) = g(u)\left|\frac{du}{dx}\right| = g[F_X(x)]f(x) = f(x).$$

Em outras palavras a variável aleatória $X = F^{-1}(U)$ possui função densidade $f_X(x)$, estabelecendo o resultado almejado e assim, a prova fica completa.

Questão de Caderno

Prove a desigualdade de Cauchy-Schwarz: $(Cov(X,y))^2 \leq Var(X)Var(Y)$.

SOLUÇÃO

Suponha que $E[X^2]$ e $E[Y^2]$ sejam finitos. Então

$$(E[XY])^2 = |E[XY]|^2 \leq E[X^2]E[Y^2].$$

Seja a função

$$
\begin{aligned}
0 \leq h(t) &= E[(tX-Y)^2],\\
&= E[t^2X^2-2tXY+Y^2],\\
&= E[X^2]t^2-2E[XY]t+E[Y^2].
\end{aligned}
$$

Usando a fórmula de báskara, b^2-4ac de uma função do segundo grau ax^2+bx+c, então

$$
\begin{aligned}
h(t) > 0 &\Rightarrow 4(E[XY])^2-4E[X^2]E[Y^2] < 0,\\
&\Rightarrow (E[XY])^2 < E[X^2]E[Y^2].
\end{aligned}
$$

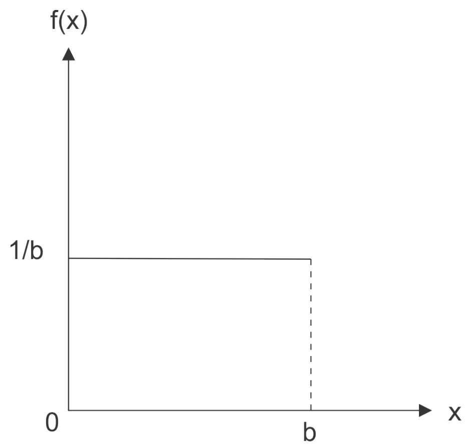
Se reescrevermos $|E[UV]| \leq \sqrt{E[U]^2E[V^2]}$ e fazendo $U = X - E[X]$ e $V = Y - E[Y]$, então

$$\begin{aligned} |E[(X - E[X])(Y - E[Y])]| &\leq \sqrt{E[(X - E[X])^2]E[(Y - E[Y])^2}, \\ |Cov(X,Y)| &\leq \sigma_X\sigma_Y. \end{aligned}$$

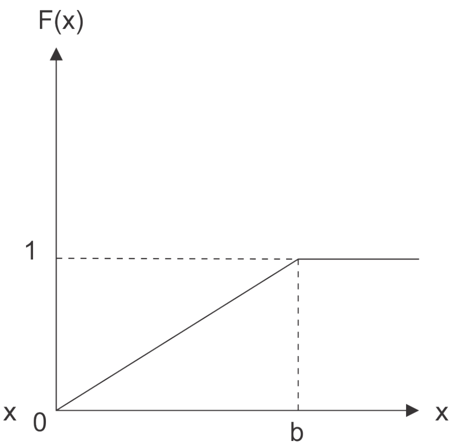
2.3 Lista 3

Questão 1

Sejam $X_1, X_2, \ldots, X_n$ $i.i.d. \sim U(0,b)$. Portanto, $f_n(x) = \frac{1}{b}$ e $F_X(x) = \frac{x}{b}$ $0 \leq x \leq b$.



(a) Função densidade.



(b) Função Distribuição.

Figura 2.4: Função densidade (2.4a) e função distribuição (2.4b).

Mostre que os estimadores $X_{(n)} = \max\{X_1,X_2,\ldots,X_n\}$ define uma sequência consistente para o parâmetro b.
SUGESTÃO : Utilizando a desigualdade de Chebychev.

$$\lim_{n\to\infty} E_b[(X_{(n)} - b)^2] = 0.$$

Ver Casella-Berger (pág. 468, ex. 10.1.2).

Como $F_{X_{(n)}}(x) = (F_X(x))^n = \left(\frac{x}{n}\right)^n \Rightarrow f_{X_{(n)}}(x) = n\left(\frac{x}{n}\right)^{n-1}$, calcule

$$E_b[(X_{(n)} - b)^2] = \int_0^b (x-b)^2 n \left(\frac{x}{b}\right)^{n-1} dx$$

SOLUÇÃO

$$\begin{aligned} E_b[(X_{(n)} - b)^2] &= \int_0^b (x-b)^2 n \left(\frac{x}{b}\right)^{n-1} dx, \\ &= n \times \left(\frac{1}{b}\right)^{n-1} \int_0^b [x^2 - 2xb + b^2]x^{n-1}dx, \\ &= \left(\frac{1}{b}\right)^{n-1} \left[\int_0^b x^2x^{n-1}dx - 2b\int_0^b xx^{n-1}dx + b^2\int_0^b x^{n-1}dx\right], \\ &= n \times \left(\frac{1}{b}\right)^{n-1} \left[\int_0^b x^{n+1}dx - 2b\int_0^b x^n dx + b^2\int_0^b x^{n-1}dx\right], \\ &= n \times \left(\frac{1}{b}\right)^{n-1} \left[\frac{x^{n+2}}{n+2} - 2b\frac{x^{n+1}}{n+1} + b^2\frac{x^n}{n}\right]_0^b, \end{aligned}$$

$$
\begin{aligned}
&= n \times \left(\frac{1}{b}\right)^{n-1}\left[\frac{b^{n+2}}{n+2} - 2b\frac{b^{n+1}}{n+1} + b^2\frac{b^n}{n}\right],\\
&= n \times \left(\frac{1}{b}\right)^{n-1}\left[\frac{b^{n+2}}{n+2} - 2\frac{b^{n+2}}{n+1} + \frac{b^{n+2}}{n}\right],\\
&= n \times \frac{1}{b^{n-1}}\left[\frac{b^{n+2}(n+1)n - 2b^{n+2}(n+2)(n) + b^{n+2}(n+2)(n+1)}{(n+2)(n+1)n}\right],\\
&= b^{1-n}\left[\frac{b^{n+2}(n+1)n - 2b^{n+2}(n+2)(n) + b^{n+2}(n+2)(n+1)}{(n+2)(n+1)}\right],\\
&= \left[\frac{b^{(1-n)+n+2}(n+1)n - 2b^{(1-n)+n+2}(n+2)(n) + b^{(1-n)+n+2}(n+2)(n+1)}{(n+2)(n+1)}\right],\\
&= \left[\frac{b^3(n+1)n - 2b^3(n+2)(n) + b^3(n+2)(n+1)}{(n+2)(n+1)}\right],\\
&= \frac{nb^3}{n+2} - 2\frac{nb^3}{n+1} + b^3.
\end{aligned}
$$

Aplicando o limite, temos

$$
\lim_{n\to\infty} E_b[(X_{(n)} - b)^2] \; = \; \lim_{n\to\infty}\frac{nb^3}{n+2} - \lim_{n\to\infty}\left(2\frac{nb^3}{n+1}\right) + b^3.
$$

Usando a regra de L'Hopital,

$$
\begin{aligned}
\lim_{n\to\infty} E_b[(X_{(n)} - b)^2] &= b^3 - 2b^3 + b^3,\\
&= 0.
\end{aligned}
$$

Portanto, a prova fica concluída, mostrando que $X_{(n)}$ define uma sequência consistente para o parâmetro b.

Questão 2

Considere uma sequência de estimadores $W_n = W_n(X_1,\ldots,X_n)$ tais que $E_\theta[W_n] = \tau(\theta)$, isto é, W_n é um estimador não viesado de $\tau(\theta)$, para todo n. Suponha também que $Var_\theta[W_n]$ atinja a cota de Cramer-Rao. Prove que W_n $n = 1,2,\ldots$ é uma sequência consistente para $\tau(\theta)$.

SOLUÇÃO

Para mostrar que Wn é um estimador consistente para $\tau(\theta)$, basta provar:

$$
\lim_{n\to\infty} P[|W_n - \tau(\theta)| < \epsilon] = 1.
$$

Seja a desigualdade de Chebychev,

$$
P[(W_n - \tau(\theta))^2 < \epsilon^2] \geq 1 - \frac{E_\theta[(W_n - \tau(\theta))^2]}{\epsilon^2},
$$

e que $E_\theta[(W_n - \tau(\theta))^2 = Var_\theta(W_n)$, já que Viés$_\theta(W_n) = 0$, então

$$
P[(W_n - \tau(\theta))^2 < \epsilon^2] \geq 1 - \frac{Var_\theta(W_n)}{\epsilon^2}.
$$

Sabemos que a $Var_\theta(W_n)$ atingiu a cota de Cramer-Rao, isto é, $Var_\theta(W_n) = \frac{[\tau'(\theta)]^2}{nI(\theta)}$. Assim,

$$
P[(W_n - \tau(\theta))^2 < \epsilon^2] \geq 1 - \frac{[\tau'(\theta)]^2}{nI(\theta)\epsilon^2}.
$$

Passando o limite, temos

$$\begin{aligned}
\lim_{n\to\infty} P[(W_n - \tau(\theta))^2 < \epsilon^2] &\geq \lim_{n\to\infty} \left[1 - \frac{[\tau'(\theta)]^2}{nI(\theta)\epsilon^2}\right], \\
&\geq 1 - \lim_{n\to\infty} \left[\frac{[\tau'(\theta)]^2}{nI(\theta)\epsilon^2}\right], \\
&= 1.
\end{aligned}$$

Portanto, provado que W_n é uma sequência consistente para $\tau(\theta)$.

Questão 3

Seja $X_1, \ldots ,X_n$ uma amostra *iid* com $X_i \sim N(\mu,\sigma^2)$. $T_{n-1} = \frac{\overline{X}_n-\mu}{S/\sqrt{n}}$ é uma quantidade pivotal $t-Student$ com $n-1$ graus d liberdade e $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$, utilizada para obter intervalos de confiança a $\gamma\%$ para a média da forma $\left[\overline{X}_n - q_2\frac{S}{\sqrt{n}}, \overline{X}_n - q_1\frac{S}{\sqrt{n}}\right]$, onde q_1 e q_2 são definidos por $P\left(q_1 < \frac{\overline{X}_n-\mu}{S/\sqrt{n}<q+2}\right) = \gamma$ (MOOD, pág. 381). Na ausência de uma tabela para a distribuição t_{n-1} obtenha um intervalo de confiança aproximado a 90% para a média. Utilize o fato abaixo:

$$T_{n-1} = \frac{\overline{X}_n - \mu}{S/\sqrt{n}} = \frac{\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}}{\sqrt{S/\sigma^2}} \xrightarrow{\text{d}} N(0,1).\text{(Magalhães, pág. 321)}$$

SOLUÇÃO

Sabendo que em distribuição,

$$T_{n-1} = \frac{\overline{X}_n - \mu}{S/\sqrt{n}} = \frac{\frac{\overline{X}_n-\mu}{\sigma/\sqrt{n}}}{\sqrt{S/\sigma^2}} \xrightarrow{\text{d}} N(0,1),$$

a um nível de significância $\alpha/2 = 5\%$, então o intervalo de confiança a 90% de probabilidade para a média será

$$\left[\overline{X}_n + q_{5\%}\frac{S}{\sqrt{n}}, \overline{X}_n - q_{5\%}\frac{S}{\sqrt{n}}\right].$$

Como $q_{5\%} = 1{,}64$ é o quantil da distribuição normal padrão, temos

$$\left[\overline{X}_n + 1{,}64\frac{S}{\sqrt{n}}, \overline{X}_n - 1{,}64\frac{S}{\sqrt{n}}\right].$$

Questão 4

Provar que a informação de Fisher $I(\theta) = E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right] = -E_\theta\left[\left(\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right)\right]$.

SOLUÇÃO

Considere a função $l'(x;\theta) = \frac{\partial}{\partial\theta}\log f(x;\theta) = \frac{f'(X;\theta)}{f(x;\theta)}$. Então

- $\int f'(x;\theta)dx = \int \frac{\partial}{\partial\theta} f(x;\theta)dx = \frac{\partial}{\partial\theta}\underbrace{\int f(x;\theta)dx}_{=1} = 0.$
- $\int f''(x;\theta)dx = \int \frac{\partial^2}{\partial\theta^2} f(x;\theta)dx = \frac{\partial^2}{\partial\theta^2}\underbrace{\int f(x;\theta)dx}_{=1} = 0.$
- $\frac{\partial^2}{\partial\theta^2}\log f(x;\theta) = \frac{\partial}{\partial\theta}\left[\frac{f''(x;\theta)}{f(x;\theta)}\right] = \frac{f''(x;\theta)f(x;\theta)-[f'(x;\theta)]^2}{[f(x;\theta)]^2} = \frac{f''(x;\theta)}{f(x;\theta)} - [l'(x;\theta)]^2.$

Assim,

$$\begin{aligned}
-E_\theta\left[\left(\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right)\right] &= -\int\left(\frac{f''(x;\theta)}{f(x;\theta)}-[l'(x;\theta)]^2\right)f(x;\theta)dx,\\
&= -\int\left(\frac{f''(x;\theta)}{f(x;\theta)}-\left[\frac{\partial}{\partial\theta}\log f(x;\theta)\right]^2\right)f(x;\theta)dx,\\
&= -\underbrace{\int f''(x;\theta)dx}_{=0}+\int\left(\frac{\partial}{\partial\theta}\log f(x;\theta)^2\right)f(x;\theta)dx,\\
&= E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right].
\end{aligned}$$

Questão 5

Mostrar que $Var_\lambda[\overline{X}] \Leftrightarrow a(\lambda)[\overline{X}-\lambda] = \frac{\partial}{\partial\theta}\log L(\theta|x_1,\ldots,x_n)$ para o caso da distribuição poisson com parâmetro λ.

SOLUÇÃO

Seja a função densidade da distribuição Poisson:

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!},$$

sendo a esperança $E[X]=\lambda$. Então, a função de verossimilhança é dada por:

$$\begin{aligned}
L(\lambda|x_1,\ldots,x_n) &= \prod_{i=1}^n f(x;\lambda),\\
&= \prod_{i=1}^n \frac{e^{-\lambda}\lambda_i^x}{x_i!},\\
&= \frac{e^{\lambda n}\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!},
\end{aligned}$$

aplicando o logaritmo, temos

$$\begin{aligned}
L(\lambda|x_1,\ldots,x_n) &= \log\left(\frac{e^{-\lambda n}\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}\right),\\
&= \lambda n\log(e)+\sum_{i=1}^n x_i\log(\lambda)-\log\left(\prod_{i=1}^n x_i!\right),\\
&= -\lambda n+\sum_{i=1}^n x_i\log(\lambda)-\sum_{i=1}^n \log x_i!.
\end{aligned}$$

Aplicando a derivada, temos

$$\frac{\partial}{\partial\lambda}\log L(\lambda|x_1,\ldots,x_n) = \frac{1}{\lambda}\sum_{i=1}^n x_i - n.$$

Pré-multiplicando os termos por $1/n$, segue que

$$\begin{aligned}
\frac{1}{n}\frac{\partial}{\partial\lambda}\log L(\lambda|x_1,\ldots,x_n) &= \frac{1}{n}\left(\frac{1}{\lambda}\sum_{i=1}^n x_i - n\right),\\
\frac{1}{n}\frac{\partial}{\partial\lambda}\log L(\lambda|x_1,\ldots,x_n) &= \left(\frac{1}{\lambda}\frac{\sum_{i=1}^n x_i}{n}-1\right),\\
\frac{\partial}{\partial\lambda}\log L(\lambda|x_1,\ldots,x_n) &= n\left(\frac{1}{\lambda}\frac{\sum_{i=1}^n x_i}{n}-1\right),\\
\frac{\partial}{\partial\lambda}\log L(\lambda|x_1,\ldots,x_n) &= \frac{n}{\lambda}\left(\frac{\sum_{i=1}^n x_i}{n}-\lambda\right).
\end{aligned}$$

2.4 Lista 4

Questão 1

Prove que $E_\theta\left[\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^n f(x_i,\theta)\right)^2\right] = -E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log\prod_{i=1}^n f(x_i,\theta)\right]$.

SOLUÇÃO

Seja,

$$
\begin{aligned}
\frac{\partial}{\partial\theta}\prod_{i=1}^n f(x_i,\theta) &\overset{\text{iid}}{=} \frac{\partial}{\partial\theta}\sum_{i=1}^n \log f(x_i,\theta),\\
&= \sum_{i=1}^n \frac{\partial}{\partial\theta}\log f(x_i,\theta),\\
&= \sum_{i=1}^n \frac{f'(x_i,\theta)}{f(x_i,\theta)},
\end{aligned}
$$

então

$$
\begin{aligned}
\frac{\partial^2}{\partial\theta^2}\prod_{i=1}^n f(x_i,\theta) &= \frac{\partial}{\partial\theta}\left(\sum_{i=1}^n \frac{f'(x_i,\theta)}{f(x_i,\theta)}\right),\\
&= \sum_{i=1}^n \frac{\partial}{\partial\theta}\frac{f'(x_i,\theta)}{f(x_i,\theta)},\\
&= \sum_{i=1}^n \frac{f''(x_i,\theta)f(x_i,\theta) - f'(x_i,\theta)f'(x_i,\theta)}{(f(x_i,\theta))^2},\\
&= \sum_{i=1}^n \left[\frac{f''(x_i,\theta)}{f(x_i,\theta)} - \left(\frac{f'(x_i,\theta)}{f(x_i,\theta)}\right)^2\right].
\end{aligned}
$$

Logo,

$$
\begin{aligned}
-E\left[\frac{\partial^2}{\partial\theta^2}\log\prod_{i=1}^n f(x_i,\theta)\right] &= -E\left\{\sum_{i=1}^n \left[\frac{f''(x_i,\theta)}{f(x_i,\theta)} - \left(\frac{f'(x_i,\theta)}{f(x_i,\theta)}\right)^2\right]\right\},\\
&= -\sum_{i=1}^n E\left[\frac{f''(x_i,\theta)}{f(x_i,\theta)} - \left(\frac{f'(x_i,\theta)}{f(x_i,\theta)}\right)^2\right],\\
&= -\sum_{i=1}^n \left[\int\left(\frac{f''(x_i,\theta)}{f(x_i,\theta)} - \left(\frac{f'(x_i,\theta)}{f(x_i,\theta)}\right)^2\right) f(x_i,\theta)dx\right],\\
&= -\sum_{i=1}^n \left[\int f''(x_i,\theta)dx - \int\left(\frac{f'(x_i,\theta)}{f(x_i,\theta)}\right)^2 f(x_i,\theta)dx\right],\\
&= -\sum_{i=1}^n \left\{\int \frac{\partial^2}{\partial\theta^2} f(x_i,\theta)dx - E\left[\left(\frac{f'(x_i,\theta)}{f(x_i,\theta)}\right)^2\right]\right\},\\
&= -\sum_{i=1}^n \left\{\frac{\partial^2}{\partial\theta^2}\underbrace{\int f(x_i,\theta)dx}_{=1} - E\left[\left(\frac{f'(x_i,\theta)}{f(x_i,\theta)}\right)^2\right]\right\},\\
&= -\sum_{i=1}^n \left\{0 - E\left[\left(\frac{f'(x_i,\theta)}{f(x_i,\theta)}\right)^2\right]\right\},\\
&= E\left[\left(\sum_{i=1}^n \frac{f'(x_i,\theta)}{f(x_i,\theta)}\right)^2\right],
\end{aligned}
$$

$$= E\left[\left(\frac{\partial^2}{\partial\theta^2}\log\prod_{i=1}^{n} f(x_i,\theta)\right)^2\right]$$

Questão 2

Encontrar a cota de Cramer-Rao para $\hat{p}/(1-\hat{p})$.
SOLUÇÃO
Seja $\hat{\theta} = \hat{p}/(1-\hat{p})$. Então,

$$Var_\theta(\hat{\theta}) \geq \frac{\left[(p/(1-p))'\right]^2}{-E\left(\frac{\partial^2}{\partial\theta^2}\log L(p|\mathbf{x})\right)}.$$

Derivando o numerador, temos

$$\left[(p/(1-p))'\right]^2 = \left[\frac{(1-p)+p}{(1-p)^2}\right]^2 = \frac{1}{(1-p)^4}$$

Calculando o denominador, temos

$$\begin{aligned}
\frac{\partial^2}{\partial\theta^2}\log L(p|\mathbf{x}) &= \frac{\partial^2}{\partial\theta^2}\left(n\hat{p}\log(p) + n(1-\hat{p})\log(1-p)\right), \\
&= -\frac{n\hat{p}}{p^2} - \frac{n(1-\hat{p})}{(1-p)^2}.
\end{aligned}$$

$$\begin{aligned}
-E\left(\frac{\partial^2}{\partial\theta^2}\log L(p|\mathbf{x})\right) &= E\left(\frac{n\hat{p}}{p^2} + \frac{n(1-\hat{p})}{(1-p)^2}\right), \\
&= \frac{nE(\hat{p})}{p^2} + \frac{n(1-E(\hat{p}))}{(1-p)^2}, \\
&= \frac{np}{p^2} + \frac{n(1-p)}{(1-p)^2}, \\
&= \frac{n}{p} + \frac{n}{(1-p)}, \\
&= \frac{n(1-p)+np}{p(1-p)}, \\
&= \frac{n}{p(1-p)}.
\end{aligned}$$

Portanto,

$$\begin{aligned}
Var_\theta(\hat{\theta}) &\geq \frac{\frac{1}{(1-p)^4}}{\frac{n}{p(1-p)}}, \\
&\geq \frac{p}{n(1-p)^3}.
\end{aligned}$$

Dessa forma, podemos concluir que o estimador $\hat{\theta} = \hat{p}/(1-\hat{p})$, atinge a cota de Cramer-Rao, já que, assintoticamente,

$$\sqrt{n}\left(\frac{\hat{p}}{1-\hat{p}} - \frac{p}{1-p}\right) \xrightarrow{\text{d}} n\left(0, \frac{p}{(1-p)^3}\right).$$

Questão 3

Encontrar a cota de Cramer-Rao $Var\left[\widehat{p}(1-\widehat{p})\right]$.
SOLUÇÃO
Seja $\hat{\theta} = \widehat{p}(1-\widehat{p})$. Então

$$
\begin{aligned}
Var_\theta(\hat{\theta}) &\geq \frac{\left[(p(1-p))'\right]^2}{E\left(\frac{\partial^2}{\partial\theta^2}\log L(p|\mathbf{x})\right)},\\
&\geq \frac{(1-2p)^2}{\frac{n}{p(1-p)}},\\
&\geq \frac{p(1-p)(1-2p)^2}{n}.
\end{aligned}
$$

Dessa forma, podemos concluir que o estimador $\hat{\theta} = \hat{p}/(1-\hat{p})$, atinge a cota de Cramer-Rao, já que, assintoticamente,

$$\sqrt{n}\left[\hat{p}(1-\hat{p}) - p(1-p)\right] \xrightarrow{\text{d}} n\left(0,p(1-p)(1-2p)^2\right).$$

Questão 4

Fazer o estimador de máxima verossimilhança do parâmetro p_i da distribuição multinomial.
SOLUÇÃO
Seja a distribuição multinomial,

$$f(x_1,x_2,\ldots,x_m|p_1,p_2,\ldots,p_m) = \frac{n!}{\prod_{i=1}^m x_i!}\prod_{i=1}^m p_i^{x_i},$$

em que $x_1 + x_2 + \ldots + x_m = n$ e $p_1 + p_2 + \ldots + p_m = 1$.
Encontrando a verossimilhança, temos

$$l(p_1,p_2,\ldots,p_m) = \log(n!) - \sum_{i=1}^m \log(x_1!) + \sum_{i=1}^m x_i \log(p_i).$$

No entanto, não podemos maximizar, temos que levar em consideração uma restrição. Por isso, usamos um multiplicador de Lagrange, λ. Assim,

$$L(p_1,p_2,\ldots,p_m,\lambda) = \log(n!) - \sum_{i=1}^m \log(x_1!) + \sum_{i=1}^m x_i\log(p_i) + \lambda(\sum_{i=1}^m p_i - 1),$$

derivando, temos

$$
\begin{aligned}
\frac{\partial}{\partial p_i}L(p_1,p_2,\ldots,p_m,\lambda) &= \log(n!) - \sum_{i=1}^m \log(x_1!) + \sum_{i=1}^m x_i\log(p_i) + \lambda(\sum_{i=1}^m p_i - 1),\\
&= \frac{\sum_{i=1}^m x_i}{\hat{p}_i} + \lambda = 0, \quad \text{sob restrição } \lambda = -n\\
\hat{p}_i &= \frac{\sum_{i=1}^m x_i}{n}.
\end{aligned}
$$

Questão 5 - Exemplo 10.3.4, pág. 439 (Casella - port.)

Seja $\theta = (p_1,p_2,p_3,p_4,p_5)$, onde os p_j's são não negativos e somam 1. Suponha que $X_1,X_2,\ldots,X_n$ sejam variáveis aleatórias iid e discretas $P_\theta(X_j = j) = p_j$, $j = 1,\ldots,5$. Portanto, a fp de X_i é $f(j|\theta) = p_j$ e a função de verossimilhança é

$$L(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta) = p_1^{y_1}p_2^{y_2}p_3^{y_3}p_4^{y_4}p_5^{y_5},$$

onde y_j = número de $x_1, \ldots, x_n$ igual a j. Considere testar

$$H_0 : p_1 = p_2 = p_3 \quad \text{e} \quad p_4 = p_5 \quad versus \quad H_1 : H_0 \text{ não é verdadeira.}$$

O espaço paramétrico completo Θ, é realmente um conjunto quadrimensional. Uma vez que $p_5 = 1 - p_1 - p_2 - p_3 - p_4$, existem somente quatro parâmetros livres. O conjunto de parâmetros é definido por

$$\sum_{j=1}^{4} p_j \leq 1 \quad \text{e} \quad p_j \geq 0, \quad j = 1, \ldots, 4,$$

um subconjunto de $I\!R^4$ contendo um subconjunto aberto de $I\!R^4$. Portanto, $q = 4$. Existe somente um parâmetro livre no conjunto especificado por H_0 porque, uma vez que p_1, $0 \leq p_1 \leq 1/3$ é fixo, $p_2 = p_3$ deve ser igual a p_1 e $p_4 = p_5$ deve ser igual $(1 - 3p_1)/2$. Portanto, $p = 1$, e os graus de liberdade são $\nu = 4 - 1 = 3$.

Para calcular $\lambda(\mathbf{x})$, o EMV de θ de acordo com Θ_0 e Θ deve ser determinado. Definindo

$$\frac{\partial}{\partial p_j} \log L(\theta|\mathbf{x}) = 0, \quad \text{para cada } j = 1, \ldots, 4,$$

e utilizando os fatos de que $p_5 = 1 - p_1 - p_2 - p_3 - p_4$ e $y_5 = n - y_1 - y_2 - y_3 - y_4$, podemos verificar que o EMV de p_j de acordo com $\hat{\theta}_j = y_j/n$ (valor encontrado no quesito anterior). Segundo H_0, a função de verossimilhança se reduz para

$$L(\theta|\mathbf{x}) = p_1^{y_1+y_2+y_3} \left(\frac{1 - 3p_1}{2}\right)^{y_4+y_5}.$$

Aplicando o logaritmo na função de verossimilhança e posteriormente, derivando-a, temos

$$\begin{aligned}
\log(L(\theta|\mathbf{x})) &= (y_1 + y_2 + y_3)\log(p_1) + (y_4 + y_5)\log(1 - 3p_1) - (y_4 + y_5)\log(2), \\
\frac{\partial}{\partial p_1}\log(L(\theta|\mathbf{x})) &= \frac{y_1 + y_2 + y_3}{p_1} - 3\left(\frac{y_4 + y_5}{1 - 3p_1}\right),
\end{aligned}$$

igualando a zero, segue que

$$\hat{p_1} = \frac{y_1 + y_2 + y_3}{3n}.$$

Com esses resultados, a função de verossimilhança é calculada:

$$\begin{aligned}
L(\theta|\mathbf{x}) &= \left(\frac{y_1 + y_2 + y_3}{3n}\right)^{y_1+y_2+y_3} \left(\frac{1 - 3\hat{p}_{10}}{2}\right)^{y_4+y_5}, \\
&= \left(\frac{y_1 + y_2 + y_3}{3n}\right)^{y_1+y_2+y_3} \left(\frac{1 - 3\left(\frac{y_1+y_2+y_3}{3n}\right)}{2}\right)^{y_4+y_5}, \\
&= \left(\frac{y_1 + y_2 + y_3}{3n}\right)^{y_1+y_2+y_3} \left(\frac{n - (y_1 + y_2 + y_3)}{2n}\right)^{y_4+y_5}.
\end{aligned}$$

Logo,

$$\begin{aligned}
\lambda(\mathbf{x}) &= \frac{\left(\frac{y_1 + y_2 + y_3}{3n}\right)^{y_1+y_2+y_3} \left(\frac{n - (y_1 + y_2 + y_3)}{2n}\right)^{y_4+y_5}}{\hat{p}_1^{y_1}\hat{p}_2^{y_2}\hat{p}_3^{y_3}\hat{p}_4^{y_4}\hat{p}_5^{y_5}}, \\
&= \frac{\left(\frac{y_1 + y_2 + y_3}{3n}\right)^{y_1+y_2+y_3} \left(\frac{n - (y_1 + y_2 + y_3)}{2n}\right)^{y_4+y_5}}{\left(\frac{y}{n}\right)^{y_1}\left(\frac{y}{n}\right)^{y_2}\left(\frac{y}{n}\right)^{y_3}\left(\frac{y}{n}\right)^{y_4}\left(\frac{y}{n}\right)^{y_5}}, \\
&= \left(\frac{y_1 + y_2 + y_3}{3n}\right)^{y_1}\left(\frac{n}{y_1}\right)^{y_1}\left(\frac{y_1 + y_2 + y_3}{3n}\right)^{y_2}\left(\frac{n}{y_2}\right)^{y_2} \times \\
&\quad \times \left(\frac{y_1 + y_2 + y_3}{3n}\right)^{y_3}\left(\frac{n}{y_3}\right)^{y_3}\left(\frac{n - (y_1 + y_2 + y_3)}{2n}\right)^{y_4}\left(\frac{n}{y_4}\right)^{y_4} \times
\end{aligned}$$

$$
\begin{aligned}
& \left(\frac{n-(y_1+y_2+y_3)}{2n}\right)^{y_5}\left(\frac{n}{y_5}\right)^{y_5}, \\
= & \left(\frac{y_1+y_2+y_3}{3y_1}\right)^{y_1}\left(\frac{y_1+y_2+y_3}{3y_2}\right)^{y_2}\left(\frac{y_1+y_2+y_3}{3y_3}\right)^{y_3} \times \\
& \times \left(\frac{n-(y_1+y_2+y_3)}{2y_4}\right)^{y_4}\left(\frac{n-(y_1+y_2+y_3)}{2y_5}\right)^{y_5}, \\
= & \left(\frac{y_1+y_2+y_3}{3y_1}\right)^{y_1}\left(\frac{y_1+y_2+y_3}{3y_2}\right)^{y_2}\left(\frac{y_1+y_2+y_3}{3y_3}\right)^{y_3} \times \\
& \times \left(\frac{y_4+y_5}{2y_4}\right)^{y_4}\left(\frac{y_4+y_5}{2y_5}\right)^{y_5}.
\end{aligned}
$$

Portanto, a estatística do teste é

$$
\begin{aligned}
-2\log\lambda(\mathbf{x}) &= -2[y_1\log(y_1+y_2+y_3) - y_1\log(3y_1) + \\
&\quad +y_2\log(y_1+y_2+y_3) - y_1\log(3y_2) + \\
&\quad +y_3\log(y_1+y_2+y_3) - y_1\log(3y_3) + \\
&\quad +y_4\log(y_4+y_5) - y_4\log(2y_4) + \\
&\quad +y_5\log(y_4+y_5) - y_4\log(2y_5)], \\
&= 2\left[y_1\log\left(\frac{3y_1}{y_1+y_2+y_3}\right) + y_2\log\left(\frac{3y_2}{y_1+y_2+y_3}\right) + y_3\log\left(\frac{3y_3}{y_1+y_2+y_3}\right) + \right. \\
&\quad \left. +y_4\log\left(\frac{2y_4}{y_4+y_5}\right) + y_5\log\left(\frac{2y_5}{y_4+y_5}\right)\right], \\
&= 2\sum_{i=1}^{5} y_i\log\left(\frac{y_i}{m_i}\right),
\end{aligned}
$$

em que $m_1 = m_2 = m_3 = (y_1+y_2+y_3)/3$ e $m_4 = m_5 = (y_4+y_5)/2$. O teste assintótico de tamanho α rejeita H_0 se $-\log\lambda(\mathbf{x}) \geq \chi^2_{3,\alpha}$.

Questão 7

Um um sorteio com um dado foi realizado 20 vezes, sendo o resultado: Considere testar $H_0 : p_1 = p_2 =$

Número	1	2	3	4	5	6
Frequência	4	6	3	3	2	2

$p_3 \quad versus \quad H_1 : p_4 = p_5 = p_6$, sendo o teste de tamanho 90%.

SOLUÇÃO

Pelo exercício anterior, sabemos que $-2\log\lambda(\mathbf{x}) = 2\sum_{i=1}^{6} y_i \log\left(\frac{y_i}{m_i}\right)$, e

$$
\begin{cases}
m_1 = m_2 = m_3 = \frac{y_1+y_2+y_3}{3} = \frac{4+6+3}{3} = \frac{13}{3} \\
m_4 = m_5 = m_6 = \frac{y_4+y_5+y_6}{3} = \frac{3+2+2}{3} = \frac{7}{3}.
\end{cases}
$$

Logo,

$$
\begin{aligned}
-2\log\lambda(\mathbf{x}) &= 2\sum_{i=1}^{6} y_i\log\left(\frac{y_i}{m_i}\right), \\
&= 2\left[y_1\log\left(\frac{y_1}{m_1}\right) + y_2\log\left(\frac{y_2}{m_2}\right) + y_3\log\left(\frac{y_3}{m_3}\right) + \right. \\
&\quad \left. +y_4\log\left(\frac{y_4}{m_4}\right) + y_5\log\left(\frac{y_5}{m_5}\right) + y_6\log\left(\frac{y_6}{m_6}\right)\right], \\
&= 2\left[4\log\left(\frac{4}{13/3}\right) + 6\log\left(\frac{6}{13/3}\right) + 3\log\left(\frac{3}{13/3}\right) + \right.
\end{aligned}
$$

$$
\begin{aligned}
& +3\log\left(\frac{3}{7/3}\right) + 2\log\left(\frac{2}{7/3}\right) + 2\log\left(\frac{2}{7/3}\right)\Bigg], \\
= \;& 2(-0{,}3202 + 1{,}9525 - 1{,}1032 + 0{,}7539 - 0{,}3083 - 0{,}3083), \\
= \;& 1{,}333 \sim \chi^2_{4,\alpha=0{,}10} = 7{,}78.
\end{aligned}
$$

Como $1{,}333 > 7{,}78$, então na há evidências para rejeitar a hipótese H_0 com um nível de significância de 10%.

Questão 8

O que significa "escore" para a a estatística?
SOLUÇÃO
De acordo com o Wikpedia - The Free Encyclopedia, em estatística, o escore, função escore, escore eficiente ou informante, indica o quão sensível a função de verossimilhança $L(\theta,X)$ depende de seu parâmetro θ. Explicitamente, o escore para θ é o gradiente da log-verossimilhança com respeito a θ.
O escore desempenha um papel importante em vários aspectos de inferência. Por exemplo:

- na formulação de uma estatística para um teste mais poderoso;
- na aproximação do erro em uma estimativa de máxima verossimilhança;
- em demonstrações de estimativas de máxima verossimilhança assintoticamente suficientes;
- em formulações de intervalos de confiança;
- em demonstrações da desigualdade de Cramér-Rao.

A função escore também desempenha um papel em regras de estatística computacional, como por exemplo, cálculo das estimativas de máxima verossimilhança.

Exemplo 10.3.5, Casela (2001), pág. 493

Testes binomiais com grande amostras
Sejam $X_1, X_2, \ldots, X_n$ uma variável aleatória de uma população de Bernoulli (p). Considere testar $H_0 : p \leq p_0 \quad versus \quad H_1 : p > p_0$, onde $0 < p_0 < 1$ é um valor específico. O EMV de p, com base em uma amostra de tamanho n, é $\hat{p}_n = \sum_{i=1}^{n} /n$. Uma vez que $\hat{p}_n$ é apenas uma média amostral, o Teorema do Limite Central é aplicado e afirma que para qualquer p, $0 < p < 1$, $(\hat{p} - p)/\sigma_n$ converge para uma variável aleatória normal padrão. Aqui $\sigma_n = \sqrt{p(1-p)/n}$, um valor que depende do parâmetro desconhecido p. Uma estimativa razoável de σ_n é $\{S_n = \sqrt{\hat{p}_n}(1-\hat{p})/n\}$, e é possível demonstrar (veja o Exercício 5.32, p. 261) que σ_n/S_n converge em probabilidade para 1. Portanto, para qualquer p, $0 < p < 1$,

$$
\frac{\hat{p}_n - p}{\sqrt{\dfrac{\hat{p}_n(1-\hat{p}_n)}{n}}} \to n(0{,}1).
$$

A estatística de Wald Z_n é definida substituindo-se p por p_0, e o teste de Wald para grandes amostras rejeita H_0 se $Z_n > z_\alpha$. Como uma estimativa alternativa de σ_n, é fácil verificar que $1/I_n(\hat{p}) = \hat{p}_n(1-\hat{p}_n)/n$. Desse modo, a mesma estatística Z_n é obtida se utilizarmos o número de informações para derivar um erro padrão para $\hat{p}_n$.
Se houvesse interesse em testar a hipótese bilateral $H_0 : p = p_0 \quad versus H_1 : p \neq p_0$, onde $0 < p_0 < 1$ é um valor específico, mais uma vez, a estratégia acima poderia ser aplicada. Entretanto, neste caso, existe um teste aproximado alternativo. De acordo com o Teorema do Limite Central, para qualquer p, $0 < p < 1$,

$$
Z'_n = \frac{\hat{p}_n - p}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim n(0{,}1) \quad \text{(aproximadamente).}
$$

O teste de nível aproximado α rejeita H_0 se $|Z'_n| > z_{\alpha/2}$.

Exemplo 10.3.6, Casella (2001), pág. 495

Teste de escore binomial
Tenha como referência novamente o modelo de Bernoulli do Exemplo 10.3.5 e considere testar $H_0 : p = p_0 \quad versus \quad H_1 : p \neq p_0$. Cálculos diretos geram

$$S(p) = \frac{\hat{p}_n - p}{p(1-p)/n} \quad \text{e} \quad I_n(p) = \frac{n}{p(1-p)}.$$

Desse modo, a estatística de escore é

$$Z_S = \frac{S(p_0)}{\sqrt{I_n(p_0)}} = \frac{\hat{p}_n - p_0}{\sqrt{p_0(1-p_0)/n}},$$

a mesma do resultado do exercício 10.3.5.

Exemplo 10.4.5, Casella (2001), pág. 500

Intervalo de Poisson aproximado
Se $X_1, \ldots, X_n$ são Poisson(λ) iid, então sabemos que

$$\frac{\widetilde{X} - \lambda}{S/\sqrt{n}} \to n(0{,}1).$$

Contudo, isto é verdadeiro, mesmo se não obtivermos uma amostra de uma população de Poisson. Utilizando a suposição de Poisson, sabemos que $Var[X] = \lambda = E[X]$ e $\widetilde{X}$ consiste em um bom estimador de λ. Portanto, utilizando a suposição de Poisson, poderíamos também obter um intervalo de confiança aproximado, a partir do fato de que

$$\frac{\widetilde{X} - \lambda}{\sqrt{\widetilde{X}/n}} \to n(0{,}1),$$

que é o intervalo que resulta da inversão do teste de Wald. Podemos utilizar a suposição de Poisson de outra maneira. Uma vez que $Var[X] = \lambda$, segue que

$$\frac{\widetilde{X} - \lambda}{\sqrt{\lambda}/n} \to n(0{,}1),$$

resultando no intervalo correspondente ao teste de pontuação, que também é o intervalo de verossimilhança de (10.4.2) e é melhor, de acordo com Wilks (1938).

2.5 Lista 5

Questão 10.31, Casella (2001), pág. 511

Dados binomiais coletados a partir de mais de uma população frequentemente são apresentados em uma $t$abela de contingência. Para o caso de duas populações,a tabela pode se parecer do seguinte modo:

	População 1	População 2	Total
Sucessos	S_1	S_2	$S = S_1 + S_2$
Fracassos	F_1	F_2	$F = F_1 + F_2$
Total	n_1	n_2	$n = n_1 + n_2$

onde a População 1 é binomial(n_1,p_1), com S_1 sucessos e F_1 fracassos, e a População 2 é binomial(n_1,p_2), com S_2 sucessos e F_2 fracassos. Uma hipótese geralmente interessante é

$$H_0 : p_1 = p_2 \qquad \text{versus} \qquad H_1 : p_1 \neq p_2.$$

a) Mostre que um teste pode ter como base na estatística

$$T = \frac{(\hat{p}_1 - \hat{p}_2)^2}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)(\hat{p}(1-\hat{p}))}$$

onde $\hat{p}_1 = S_1/n_1$, $\hat{p}_2 = S_2/n_2$ e $\hat{p} = (S_1+S_2)/(n_1+n_2)$. Além disso, mostre que à medida que $n_1,n_2 \to \infty$, a distribuição T se aproxima de χ^2_1. (Este é um caso especial de um teste conhecido como teste de independência qui-quadrado).

SOLUÇÃO

Usando a aproximação normal para a binomial

Seja,

$$X_{pop_1} \sim Binomial(n_1,p_1), \quad E[\hat{p}_1] = p_1, \quad Var[\hat{p}_1] = \frac{p_1(1-p_1)}{n_1}$$

e

$$X_{pop_2} \sim Binomial(n_2,p_2), \quad E[\hat{p}_2] = p_2, \quad Var[\hat{p}_2] = \frac{p_2(1-p_2)}{n_2},$$

logo $\hat{p}_1$ e $\hat{p}_2$ pode ser aproximada por

$$\hat{p}_1 \sim N\left(p_1, \frac{p_1(1-p_1)}{n_1}\right)$$

e

$$\hat{p}_2 \sim N\left(p_2, \frac{p_2(1-p_2)}{n_2}\right),$$

respectivamente.

Sob $H_0 : p_1 = p_2$,

$$\hat{p}_1 \sim N\left(p, \frac{p(1-p)}{n_1}\right)$$

e

$$\hat{p}_2 \sim N\left(p, \frac{p(1-p)}{n_2}\right),$$

e considerando que $\hat{p}_1$ e $\hat{p}_2$ são independentes, então

$$\hat{p}_1 - \hat{p}_2 \sim N\left[0, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)p(1-p)\right].$$

Pelo teorema do limite central (Casela, pág. 236) e pelo teorema de Slutsky (Casela, pág. 239)

$$\frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)p(1-p)}} \overset{\text{TLC}}{\sim} N(0,1)Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\hat{p}(1-\hat{p})}} \sim N(0,1) \xrightarrow{\text{d}} T(Z^2) \sim \chi^2_1$$

Usando a razão de verossimilhança

$$\begin{aligned}
\Lambda &= \frac{\sup\limits_{p_1=p_2} L(\theta|x_1 \dots x_n)}{\sup\limits_{p_1,p_2\in[0,1]} L(\theta|x_1 \dots x_n)}, \\
&= \frac{\sup\limits_{p_1=p_2} p_1^{S_1}(1-p_1)^{n_1-S_1} p_2^{S_2}(1-p_2)^{n_2-S_2}}{\sup\limits_{p_1,p_2\in[0,1]} p_1^{S_1}(1-p_1)^{n_1-S_1} p_2^{S_2}(1-p_2)^{n_2-S_2}}, \\
&= \frac{\sup\limits_{p\in[0,1]} p^{S_1}(1-p)^{n_1-S_1} p^{S_2}(1-p)^{n_2-S_2}}{\sup\limits_{p_1,p_2\in[0,1]} p_1^{S_1}(1-p_1)^{n_1-S_1} p_2^{S_2}(1-p_2)^{n_2-S_2}},
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\left(\frac{S_1+S_2}{n_1+n_2}\right)^{S_1}\left[1-\left(\frac{S_1+S_2}{n_1+n_2}\right)\right]^{n_1-S_1}\left(\frac{S_1+S_2}{n_1+n_2}\right)^{S_2}\left[1-\left(\frac{S_1+S_2}{n_1+n_2}\right)\right]^{n_2-S_2}}{\left(\frac{S_1}{n_1}\right)^{S_1}\left[1-\frac{S_1}{n_1}\right]^{n_1-S_1}\left(\frac{S_2}{n_2}\right)^{S_2}\left[1-\frac{S_2}{n_2}\right]^{n_2-S_2}},\\
&= \frac{\left(\frac{S_1+S_2}{n_1+n_2}\right)^{S_1}\left(\frac{S_1+S_2}{n_1+n_2}\right)^{S_2}\left[1-\left(\frac{S_1+S_2}{n_1+n_2}\right)\right]^{n_1+n_2-S_1-S_2}}{\left(\frac{S_1}{n_1}\right)^{S_1}\left[1-\frac{S_1}{n_1}\right]^{n_1-S_1}\left(\frac{S_2}{n_2}\right)^{S_2}\left[1-\frac{S_2}{n_2}\right]^{n_2-S_2}}.
\end{aligned}
$$

Sob H_0, $-2\log\Lambda_n \sim \chi^2_1$.

Fazendo essas duas aproximações, vamos fazer um estudo de simulação sobre o erro tipo I, e observar a melhor aproximação, considerando tamanhos de amostras diferentes, bem como parâmetros diferentes, com os níveis de significância 1% e 5% de probabilidade.

Estudo de simulação

Para se avaliar as probabilidades de se cometer o erro tipo I , duas populações foram geradas por meio de simulação computacional utilizando o software livre R (R CORE TEAM, 2012).

Para a geração dessas populações foram consideradas a distribuição binomial. Para simulação foi considerado três fatores:

1. o tamanho das amostras das populações:
   a) amostras iguais: $n_1 = n_2$ =10, 30, 60, 100;
2. parâmetros das populações:
   a) parâmetros iguais: $p_1 = p_2$ =0.1, 0.3, 0.5, 0.7, 0.9;
3. nível de significância: α =0.01 e 0.05;

Avaliação do erro tipo I

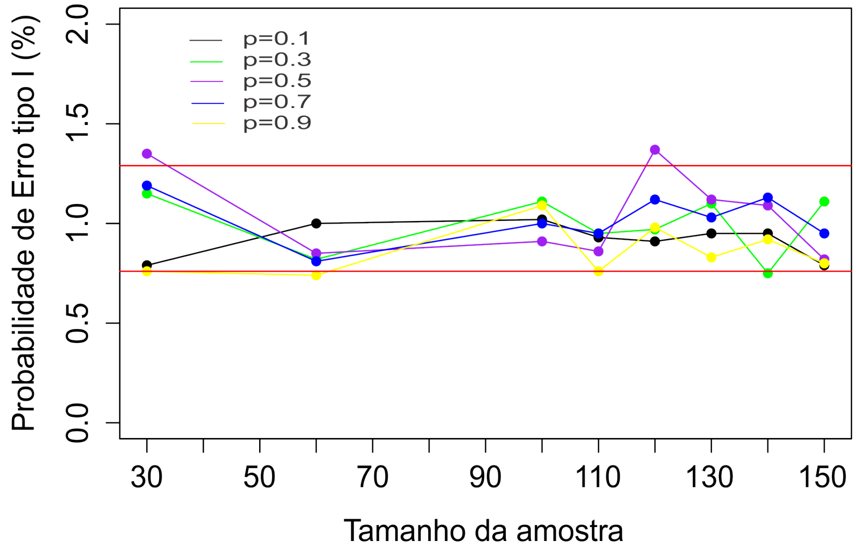
A conclusão quanto ao erro tipo I será obtida pela comparação entre os valores simulados e o intervalo de confiança para proporções(LEEMIS e TRIVEDI, 1996), cujos intervalos de 99% de confiança para o valor nominal de 1% é [0,76%;1,29%] e para 5% é [4,46%; 5,59%].

Tabela 2.1: Taxa de Erro Tipo I, para nível de significância $\alpha = 1\%$.

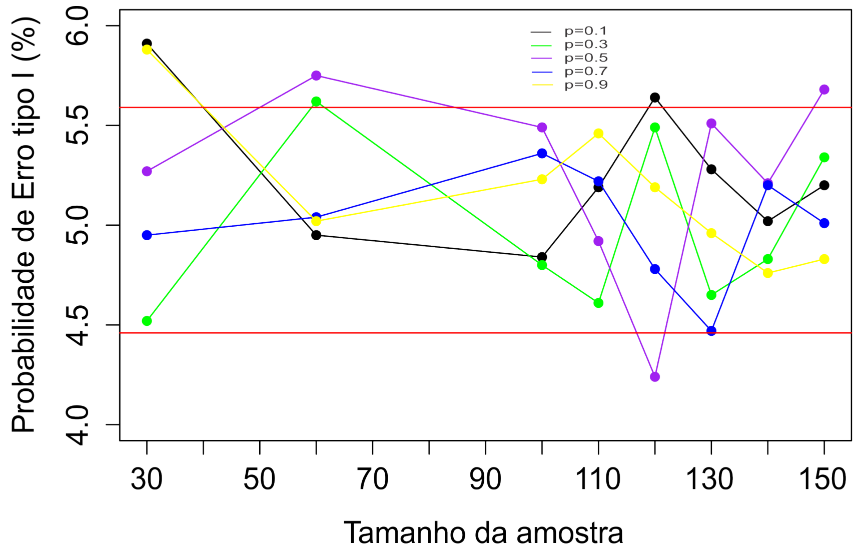
Aproximação pela Normal								
	Amostra $n_1 = n_2$							
Parâmetros $p_1 = p_2$	30	60	100	110	120	130	140	150
0.1	0,79%	1,00%	1,02%	0,93%	0,91%	0,95%	0,95%	0,79%
0.3	1,15%	0,82%	1,11%	0,95%	0,97%	1,10%	0,75%	1,11%
0.5	1,35%	0,85%	0,91%	0,86%	1,37%	1,12%	1,09%	0,82%
0.7	1,19%	0,81%	1,00%	0,95%	1,12%	1,03%	1,13%	0,95%
0.9	0,76%	0,74%	1,09%	0,76%	0,98%	0,83%	0,92%	0,80%
Razão de Verossimilhança								
0.1	8,76%	1,46%	1,16%	1,06%	0,96%	1,10%	1,14%	0,99%
0.3	1,17%	0,97%	1,13%	0,98%	0,97%	1,10%	0,81%	1,15%
0.5	1,35%	0,93%	0,93%	0,93%	1,37%	1,12%	1,09%	0,82%
0.7	1,19%	0,87%	1,02%	0,98%	1,12%	1,03%	1,16%	0,99%
0.9	8,47%	1,31%	1,17%	0,81%	1,06%	1,04%	1,13%	1,06%

Tabela 2.2: Taxa de Erro Tipo I, para nível de significância $\alpha = 5\%$.

Aproximação pela Normal								
	Amostra $n_1 = n_2$							
Parâmetros $p_1 = p_2$	30	60	100	110	120	130	140	150
0.1	5,91%	4,95%	4,84%	5,19%	5,64%	5,28%	5,02%	5,20%
0.3	4,52%	5,62%	4,80%	4,61%	5,49%	4,65%	4,83%	5,34%
0.5	5,27%	5,75%	5,49%	4,92%	4,24%	5,51%	5,21%	5,68%
0.7	4,95%	5,04%	5,36%	5,22%	4,78%	4,47%	5,20%	5,01%
0.9	5,88%	5,02%	5,23%	5,46%	5,19%	4,96%	4,76%	4,83%
Razão de Verossimilhança								
0.1	11,09%	5,97%	4,88%	5,20%	5,79%	5,38%	5,38%	5,37%
0.3	5,50%	5,62%	4,92%	4,81%	5,65%	4,71%	4,83%	5,28%
0.5	5,27%	5,75%	5,49%	4,92%	4,24%	5,51%	5,21%	5,68%
0.7	9,80%	5,04%	5,40%	5,42%	4,95%	4,63%	5,21%	5,05%
0.9	11,14%	5,90%	5,28%	5,49%	5,32%	5,22%	5,18%	4,97%

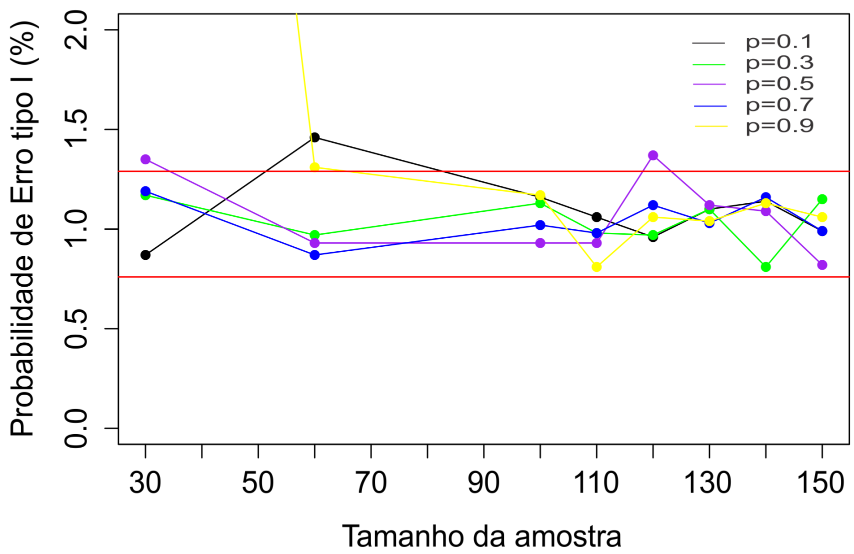


(a) Normal, $\alpha = 0{,}01$.

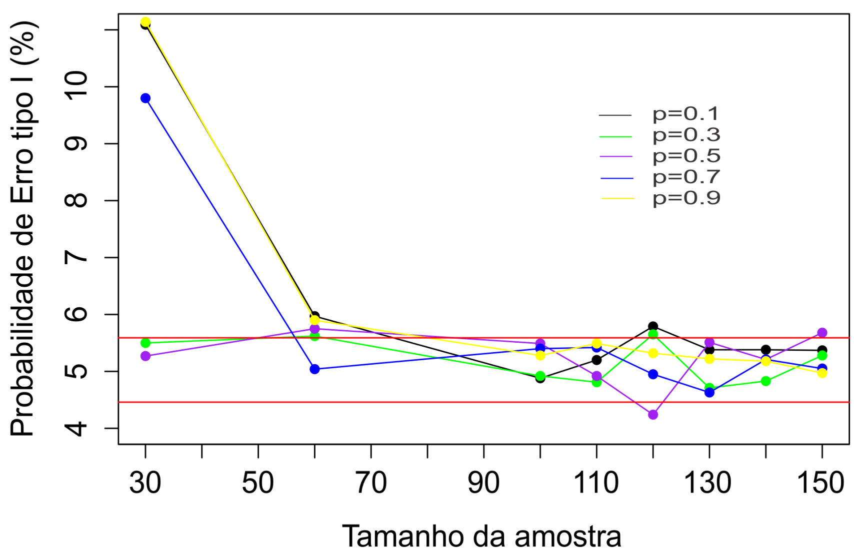


(b) Normal, $\alpha = 0{,}05$.

Figura 2.5: Gráficos representando a simulação do erro tipo I, usando a aproximação normal.



(a) Razão de Verossimilhança $\alpha = 0{,}01$.



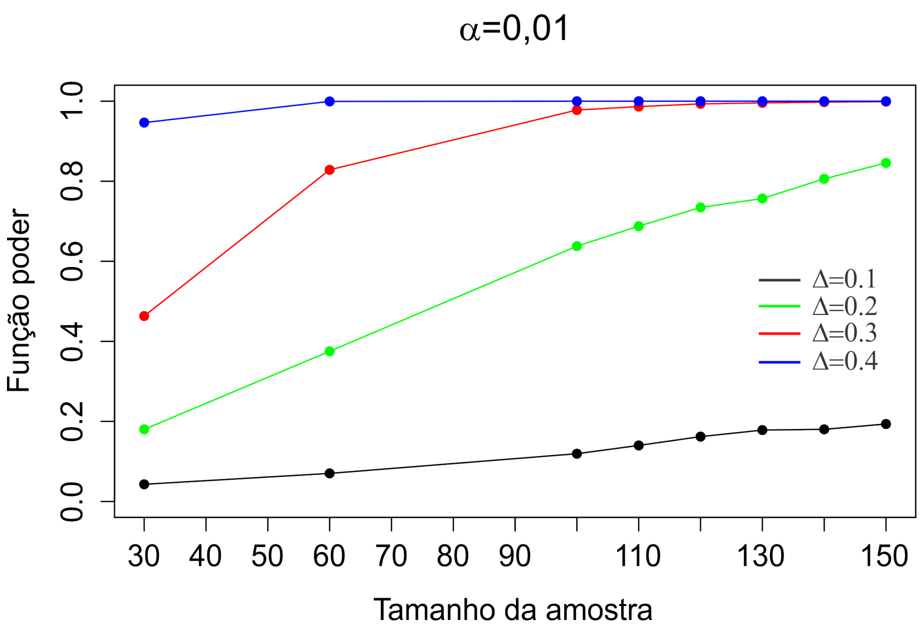
(b) Razão de Verossimilhança $\alpha = 0{,}05$.

Figura 2.6: Gráficos representando a simulação do erro tipo I, usando a aproximação da razão de verossimilhança.

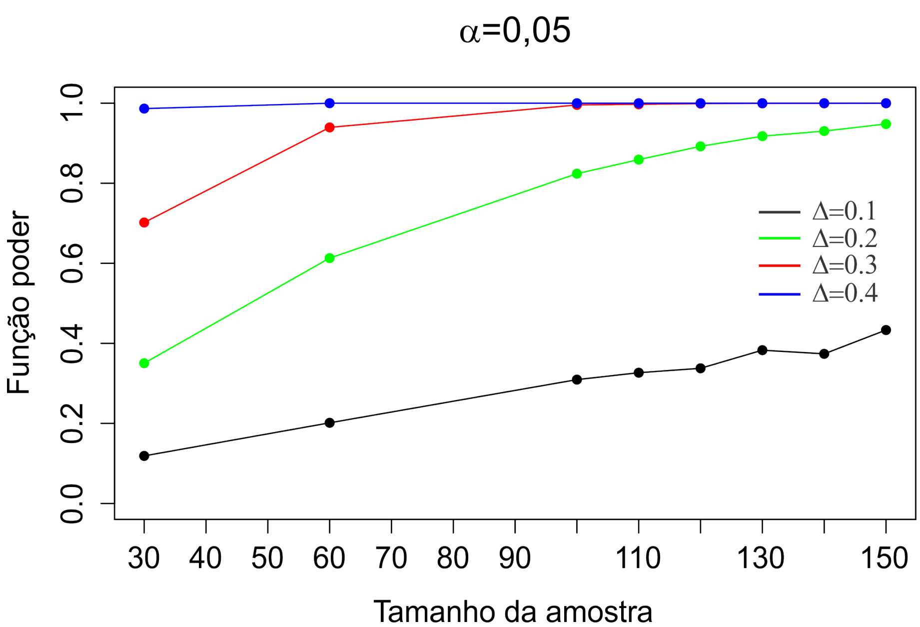
Avaliação do Poder

Tabela 2.3: Função poder dos testes baseados na aproximações normal e razão de verossimilhança.

$\Delta : p_1 \neq p_2$	**Aproximação**	α	**Função poder**							
			Tamanho da amostra (n)							
			30	**60**	**100**	**110**	**120**	**130**	**140**	**150**
0.1	Normal	1	0,0429	0,0701	0,1193	0,1400	0,1620	0,1783	0,1802	0,1935
		5	0,1188	0,2015	0,3094	0,3267	0,3377	0,3830	0,3738	0,4333
	Raz. Veros	1	0,0429	0,0735	0,1225	0,1400	0,1620	0,1785	0,1805	0,1958
		5	0,1202	0,2015	0,3094	0,3267	0,3377	0,3830	0,3743	0,4333
0.2	Normal	1	0,1801	0,3752	0,6379	0,6878	0,7346	0,7569	0,8061	0,8458
		5	0,3504	0,6130	0,8239	0,8591	0,8922	0,9176	0,9304	0,9480
	Raz. Veros	1	0,1801	0,3817	0,6417	0,6907	0,7346	0,7596	0,8105	0,8489
		5	0,3628	0,6130	0,8239	0,8591	0,8924	0,9176	0,9321	0,9483
0.3	Normal	1	0,4632	0,8286	0,9780	0,9865	0,9934	0,9958	0,9976	0,9989
		5	0,7019	0,9395	0,9957	0,9972	0,9990	0,9996	0,9997	0,9999
	Raz. Veros	1	0,4632	0,8339	0,9790	0,9870	0,9934	0,9960	0,9977	0,9989
		5	0,7222	0,9395	0,9958	0,9972	0,9990	0,9996	0,9997	0,9999
0.4	Normal	1	0,9466	0,9994	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
		5	0,9865	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
	Raz. Veros	1	0,9466	0,9995	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
		5	0,9866	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

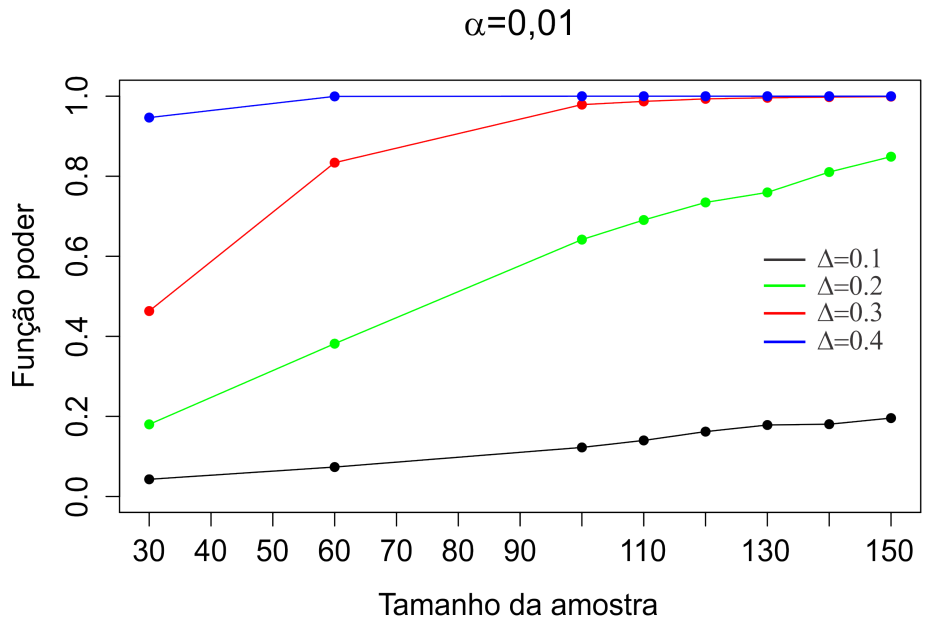


(a) Normal.

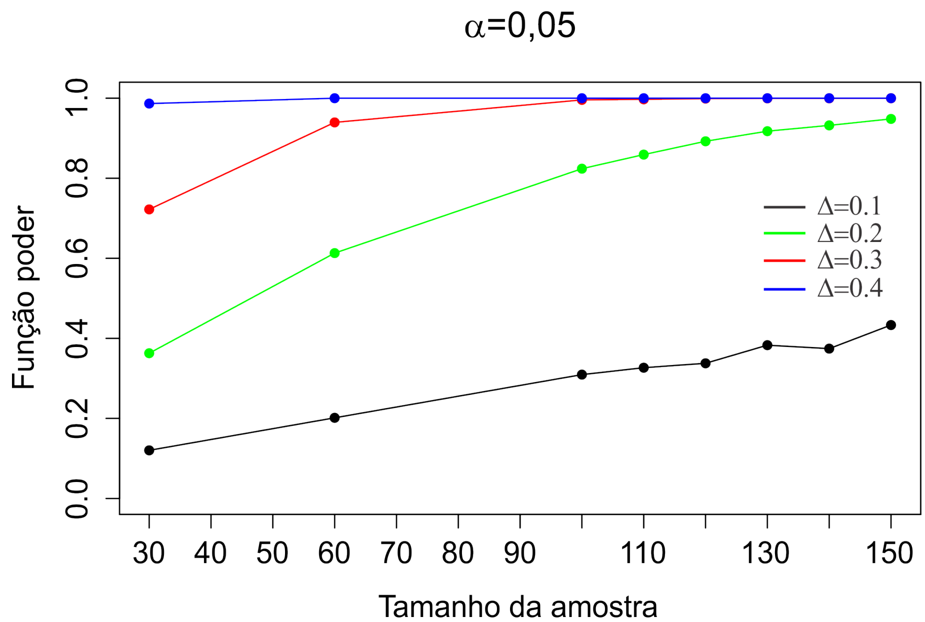


(b) Normal.

Figura 2.7: Gráficos representando a simulação poder, usando a aproximação normal.



(a) Razão de Verossimilhança.



(b) Razão de Verossimilhança.

Figura 2.8: Gráficos representando a simulação poder, usando a aproximação da razão de verossimilhança.

Rotina no R

```
###############################################
#comparison between the two methods for
#evaluating the contingency tables 2 x 2
###############################################

#          N: number of simulation
#         n1: sample size 1
#         n2: sample size 2
#p1 and p2: probability of success
#               on each trial
#      alpha: significance level

###########################################
#population with binomial distribution
###########################################

t.comp <- function(N,n1,n2,p1,p2,alpha){

exp1    <- rbinom(N,n1,p1)
exp2    <- rbinom(N,n2,p2)
est.p1 <- exp1/n1
est.p2 <- exp2/n2
est.p   <- (exp1+exp2)/(n1+n2)

#################################################
# approximation by the normal distribution:
#################################################

z    <- (est.p1-est.p2)^2/((1/n1+1/n2)*est.p*(1-est.p))
res <- 1-pchisq(z,1)

rate <- res < alpha #p-value < alpha
cont.rate <-length(rate[rate==TRUE]) #counting the number of
                                      #p-value < alpha
error <- cont.rate/N  #Type I error or Power

###################
#likelihood ratio
###################

lratio <- -2*((exp1+exp2)*log((exp1+exp2)/(n1+n2))+(n1+n2-
               (exp1+exp2))*log(1-(exp1+exp2)/(n1+n2))
              -exp1*log(exp1/n1)-(n1-exp1)*log(1-exp1/n1)
              -exp2*log(exp2/n2)-(n2-exp2)*log(1-exp2/n2))

res1 <- 1-pchisq(lratio,1)

rate1 <- res1 < alpha #p-value < alpha
cont.rate1 <-length(rate1[rate1==TRUE]) #counting the number of
                                         #p-value < alpha
error1 <- cont.rate1/N  #Type I error or Power
```

```
list(errorI.normal=error, errorI.likelihood=error1)
}

#simulation: Type I error or Power

N  <- 10000
n1 <- 60
n2 <- 60
p1 <- 0.8
p2 <- 0.3
alpha <-0.05

t.comp(N,n1,n2,p1,p2,alpha)
```

2.6 Lista 6

Questão 1 - Gibbons(2003)- pág.86 - exemplo 2.1

As maiores temperaturas durante o dia, por 10 dias foram comparadas com as médias destes dias nos anos anteriores. A e B, representam a "maior" e a "menor" temperatura, respectivamente, os dados foram descritos por: AABABBAAAB. Teste a hipótese nula de que não houve mudança a um nível de 0,05. Faça o teste exato e aproximado usando uma estatística assintótica.

SOLUÇÃO

Hipóteses:

- H_0: A sequência foi gerada de forma aleatória;
- H_1 : A sequência não foi gerada de forma aleatória.

Nível de significância: $\alpha = 0{,}05$

Dados: AABABBAAAB

Informações:

$$\begin{cases} n_A = n_2 = 6 \\ n_B = n_1 = 4 \\ R = 6 \end{cases}$$

TESTE EXATO

A distribuição de probabilidade de R é:

$$f_R(r) = \begin{cases} 2\binom{n_1-1}{r/2-1}\binom{n_2-1}{r/2-1}/\binom{n_1+n_2}{n_1}, & r \text{ par}, \\ \left[\binom{n_1-1}{(r-1)/2}\binom{n_2-1}{(r-3)/2} + \binom{n_1-1}{(r-3)/2}\binom{n_2-1}{(r-1)/2}\right]/\binom{n_1+n_2}{n_1}, & r \text{ ímpar}, \end{cases}$$
$$r = 2,3,\ldots,n_1+n_2.$$

Calculando as probabilidades:

$$f_R(2) = 2\binom{4-1}{2/2-1}\binom{6-1}{2/2-1}/\binom{4+6}{4} = 2\binom{3}{0}\binom{5}{0}/\binom{10}{4} = 0{,}01.$$

$$f_R(3) = \left[\binom{4-1}{(3-1)/2}\binom{6-1}{(3-3)/2} + \binom{4-1}{(3-3)/2}\binom{6-1}{(3-1)/2}\right]/\binom{4+6}{4}$$

$$= \left[\binom{3}{1}\binom{5}{0} + \binom{3}{0}\binom{5}{1}\right] \Big/ \binom{10}{4} = 0{,}048$$

$$f_R(9) = \left[\binom{4-1}{(9-1)/2}\binom{6-1}{(9-3)/2} + \binom{4-1}{(9-3)/2}\binom{6-1}{(9-1)/2}\right] \Big/ \binom{4+6}{4}$$

$$= \left[\underbrace{\binom{3}{4}}_{=\nexists}\binom{5}{3} + \binom{3}{3}\binom{5}{4}\right] \Big/ \binom{10}{4} = 0{,}024$$

$$f_R(10) = 2\binom{4-1}{10/2-1}\binom{6-1}{10/2-1} \Big/ \binom{4+6}{4} = 2\underbrace{\binom{3}{4}}_{=\nexists}\binom{5}{4} \Big/ \binom{10}{4} = \nexists.$$

Assim, para $P(R \leq 2) = 0{,}01$ e $P(R \geq 9) = 0{,}024$ são as respectivas maiores probabilidades que não excedem $\alpha/2 = 0{,}025$ (teste bilateral). Assim, nosso $\alpha_{real} = P(R \leq 2 \text{ ou } R \geq 9) = 0{,}01 + 0{,}024 = 0{,}034$. Se tivéssemos adotado $P(R \leq 3) = 0{,}048$ e $P(R \geq 9) = 0{,}024$, nosso $\alpha_{real} = P(R \leq 3 \text{ ou } R \geq 9) = 0{,}048 + 0{,}024 = 0{,}072$, excederia o α adotado.

Portanto, para critério de decisão, rejeitaremos a hipótese H_0 se $2 \geq R \geq 9$. Como nosso $R = 6$, não temos evidências para rejeitar a hipótese de que a amostra foi gerada de forma aleatória ao nível de significância de 5% de probabilidade.

TESTE APROXIMADO

Considere $\lambda = n_1/n$, $n = n_1 + n_2$, então R tem esperança $E[R] \approx n2\lambda(1-\lambda)$ e variância $Var[R] \approx n2\lambda^2(1-\lambda)^2$. Assim,

$$\frac{R - 2n\lambda(1-\lambda)}{2\lambda(1-\lambda)\sqrt{n}} \xrightarrow{\text{d}} n(0{,}1).$$

O teste de hipótese pode ser adotado: rejeita-se a hipótese de aleatoriedade (H_0) se

$$\left|\frac{R - 2n\lambda(1-\lambda)}{2\lambda(1-\lambda)\sqrt{n}}\right| \geq z_{\alpha/2},$$

ou fazendo a correção de continuidade,

$$\left|\frac{R - 0{,}5 - 1 - 2n_1n_2/n}{\sqrt{2n_1n_2(2n_1n_2 - n)/[n^2(n-1)]}}\right| \geq z_\alpha$$

Usando o teste com a correção de continuidade, temos:

$$\frac{6 - 0{,}5 - 1 - 2.4.6/10}{\sqrt{2.4.6(2.4..6 - 10)/[10^2(10-1)]}} \geq z_{0{,}05}$$

$$\frac{-0{,}3}{1{,}4236} = -0{,}2107 \geq 1{,}645.$$

Desse forma, o teste com aproximação assintótica também, não temos evidências para rejeitar a hipótese de que a amostra foi gerada de forma aleatória ao nível de significância de 5% de probabilidade.

Questão 2

Seja uma amostra $X_1, X_2, \ldots, X_n$ de uma população F_X com mediana Md desconhecida. Queremos testar a hipótese que $H_0 : Md = Md_0$ vs $H_1 : Md > Md_0$ e que $H_0 : Md = Md_0$ vs $H_1 : Md \neq Md_0$. Como $P[X > Md] = P[X < Md] = 1/2$, podemos pensar que para cada observação X_i, temos um ensaio de Bernoulli, isto é, X_i pode ser maior que Md_0 com probabilidade 1/2, e portanto, a amostra pode ser pensada como uma binomial(n,1/2). Para uma amostra de tamanho n defina um teste para

$H_0 : Md = Md_0$ vs $H_1 : Md > Md_0$ e para $H_0 : Md = Md_0$ vs $H_1 : Md \neq Md_0$. Obtenha uma distribuição assintótica para este teste.

SOLUÇÃO 1 - Considerando F_X conhecida de forma assintótica

Considere uma amostra $X_1, \ldots, X_n$ *iid* de uma população $f_X(x)$ com mediana Md desconhecida. Seja md_n a mediana de uma amostra de tamanho n. Considere a variável aleatória $\sqrt{n}(Md - md_n)$. Sabe-se que, conforme $n \to \infty$,

$$\sqrt{n}(Md - md_n) \xrightarrow{\text{d}} n\left(0, \frac{1}{4f^2(m)}\right),$$

em que $f(m)$ é o valor da densidade de X na mediana populacional.

Vamos propor um teste assintótico para testar $H_0 : Md = Md_0$ vs $H_1 : Md \neq Md_0$. Pelo TLC, temos:

$$Z = \frac{md_n - Md}{\frac{\sqrt{\frac{1}{4f^2(m)}}}{\sqrt{n}}} \xrightarrow{\text{d}} n(0,1).$$

Pelo teste de Wald temos que, sob a hipótese H_0 verdadeira, a estatística Z fica

$$Z_w = \frac{\sqrt{n}(md_n - Md_0)}{\frac{1}{2f(Md_0)}} \approx n(0,1).$$

Portanto, o critério de decisão será: rejeita-se H_0 se $|Z_w| > |Z_{\alpha/2}|$, sendo $Z_{\alpha/2}$ o quantil de uma normal $n(0,1)$ e α, o nível de significância do teste.

Para testarmos a hipótese $H_0 : Md = Md_0$ vs $H_1 : Md > Md_0$, podemos utilizar o teste construído anteriormente de forma análoga, baseado na estatística de Wald e o teste ser unilateral. Assim, o critério de decisão será: rejeita-se H_0 se $|Z_w| > |Z_\alpha|$, sendo $Z_{\alpha/2}$ o quantil de uma normal $n(0,1)$ e α, o nível de significância do teste.

SOLUÇÃO 2 - Considerando F_X desconhecida de forma exata e assintótica

Seja uma amostra $X_1, X_2, \ldots, X_n$ de F_X desconhecida. Na hipótese nula admite-se que a mediana populacional possui um determinado valor particular, Md_0:

$$\begin{cases} H_0 : Md = Md_0 \\ H_1 : Md \neq Md_0 \end{cases}$$

A estatística do teste é $Y = \sum_{i=1}^{n} I_{(0,Md_0]}(\text{x})$, isto é, número de observações abaixo de Md_0. Se a hipótese nula for verdadeira e a amostra aleatória, o número de observações com valor inferior a Md_0 é uma variável aleatória binomial com parâmetros n e $p = 0{,}5$. Então, o teste de hipótese é equivalente a testar:

$$\begin{cases} H_0 : p = 1/2 \\ H_1 : p \neq 1/2 \end{cases}$$

Posteriormente, calcularemos as probabilidades de Y_i com $i = 1,2,\ldots,n$, de uma $Binom(n,1/2)$. Definindo um nível de significância α, e observando a Figura 2.9, teremos que $P[Y \leq y_j \cup Y \geq y_w] \leq \alpha$. Assim, nosso critério de decisão será que rejeitaremos a hipótese H_0 se $Y \leq y_j$ ou $Y \geq y_w$ ao nível de significância α.

Se o tamanho da amostra é muito grande, o cálculo das probabilidades da binomial pode ser aproximado pela normal padrão, sendo a estatística do teste com a correção de continuidade sob H_0 verdadeiro:

$$Z = \frac{(Y + 0{,}5) - 1/2 \times n}{\sqrt{n \times 1/2(1 - 1/2)}} \sim n(0,1)$$

Para um α, os quantis da distribuição normal que estabelecem as regiões de aceitação e de rejeição da hipótese nula, $Z_{\alpha/2} = 1{,}96$, isto é, rejeita-se a hipótese H_0 se $|Z| \geq Z_{\alpha/2}$.

A outra hipótese em estudo é:

$$\begin{cases} H_0 : Md = Md_0 \\ H_1 : Md > Md_0. \end{cases}$$

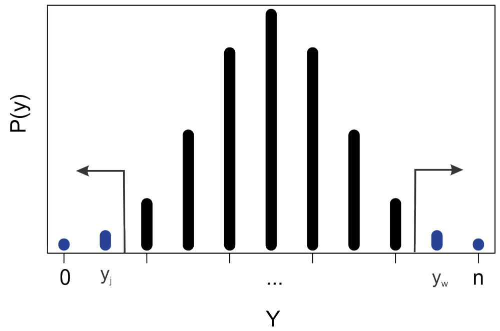


Figura 2.9: Gráfico da probabilidade $P(y)$.

De forma similar, ao teste anterior, teremos um teste equivalente:

$$\begin{cases} H_0 : p = 1/2 \\ H_1 : p > 1/2. \end{cases}$$

Posteriormente, calcularemos as probabilidades de Y_i com $i = 1,2,\ldots,n$, de uma $Binom(n,1/2)$. Definindo um nível de significância α, e observando a Figura 2.10, teremos que $P[Y \geq y_0] \leq \alpha$. Assim, nosso critério de decisão será que rejeitaremos a hipótese H_0 se $Y \geq y_o$ ao nível de significância α.

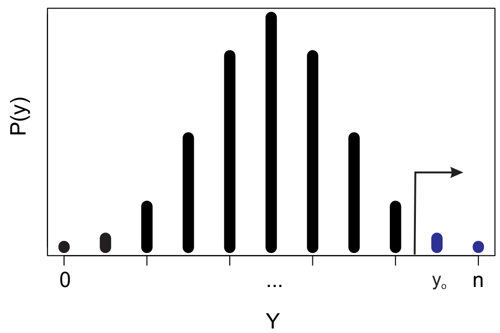


Figura 2.10: Gráfico da probabilidade $P(y)$.

Se o tamanho da amostra é muito grande, o cálculo das probabilidades da binomial pode ser aproximado pela normal padrão, sendo a estatística do teste com a correção de continuidade sob H_0 verdadeiro:

$$Z = \frac{(Y + 0{,}5) - 1/2 \times n}{\sqrt{n \times 1/2(1 - 1/2)}} \sim n(0{,}1)$$

Para um α, os quantis da distribuição normal que estabelecem as regiões de aceitação e de rejeição da hipótese nula, $Z_\alpha = 1{,}65$, isto é, rejeita-se a hipótese H_0 se $|Z| > Z_\alpha$.

Questão 3

Seja $X_1, X_2, \ldots, X_n$ uma amostra *iid* de uma uniforme $U(0{,}1)$, e $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ correspondentes estatísticas de ordem. Prova-se que $(X_{(n)} - 1)/(1/n) = n(X_{(n)} - 1) \xrightarrow{d} W = -X$.
OBS.: A f.d.p. da distribuição $W = -X$ é $f(x) = e^x$, $-\infty < x \leq 0$.

Calcule exatamente a probabilidade de que o maior($X_{(20)}$) em uma amostra de tamanho 20 seja menor que 0,92. Faça agora o cálculo aproximado utilizando a estatística assintótica.
SOLUÇÃO

Verificando o colorário pág.252 (Mood, 1974):

$$F_{X_{(n)}}(x) = [F(x)]^n,$$

sendo $F(x) = x$, já que é de uma uniforme $U(0{,}1)$. Para $n = 20$, então

$$F_{X_{(n)}}(0{,}92) = [0{,}92]^{20} = 0{,}1887$$

Utilizando a aproximação proposta pelo exercício, temos que

$$\frac{X_{(20)}-1}{1/n} = n(X_{(n)}-1) \xrightarrow{\text{d}} W = -X.$$

Daí,

$$\begin{aligned}
P[X_{(20)} < 0{,}92] &= P\left[\frac{X_{(20)}-1}{1/20} < \frac{0{,}92-1}{1/20}\right] \\
&= P\left[\underbrace{\frac{X_{(20)}-1}{1/20}}_{W} < -1{,}6\right] \\
&= \int_{-\infty}^{-1{,}6} e^{x}dx \\
&= [e^{x}]_{-\infty}^{-1{,}6} \\
&= e^{-1{,}6} - \underbrace{\lim_{x\to-\infty} e^{x}}_{=0} \\
&= e^{-1{,}6} \\
&= 0{,}2019.
\end{aligned}$$

Questão 4

O teste de aderência qui-quadrado e o teste de aderência de Kolmogorov-Smirnov podem ser aplicados em uma mesma situação. Faça alguns comentários comparando os dois.

SOLUÇÃO

TESTE DE ADERÊNCIA QUI-QUADRADO

- Dados agrupados
- Classes que possuem valores menores que cinco, devem ser agrupados em outras classes, sendo fator limitante para o uso em dados com poucas classes.
- Uma característica interessante, é que o teste de qui-quadrado considera o erro de forma cumulativa em todas as classes e no teste de Kolmogorov-Smirnov, considera somente na classe em que o erro foi o maior.

TESTE DE ADERÊNCIA KOLMOGOROV-SMIRNOV

- Pode ser usado tanto em dados agrupados, quanto para dados individuais
- Nos dados agrupados não há restrição quanto ao número nem ao valor da classe.

Questão 5

Prova-se que quando λ é grande, a distribuição de Poisson com parâmetro λ pode ser aproximada por uma normal. Se $X \sim Poisson(20)$, calcule aproximadamente a probabilidade de $3 \leq X \leq 30$.

SOLUÇÃO

Sabemos que se $X \sim Poisson(\lambda)$, para um λ grande, então

$$\frac{X-\lambda}{\sqrt{\lambda}} \sim n(0{,}1).$$

$$Z_1 = \frac{30 - 20}{\sqrt{20}} = 2{,}2360679$$

$$Z_2 = \frac{3 - 20}{\sqrt{20}} = -3{,}801325562$$

Assim, calcular $P(3 \leq X \leq 30)$, usando a padronização será

$$P(Z_2 \leq Z \leq Z_1) = 0{,}4999 + 0{,}4871 = 0{,}9870.$$

Questão 6

Calcule um intervalo de confiança aproximado a um nível de confiança 95% para θ de $f(x;\theta) = \theta e^{-\theta x}$, se for observada uma amostra de tamanho 20 e média 4. Repita o exercício para uma Bernoulli X $\theta = P[X = 1]$ se foi observada uma amostra de tamanho 20 e média 7/20.

SOLUÇÃO

POPULAÇÃO EXPONENCIAL

O estimador de máxima verossimilhança de θ é $\hat{\theta}(x_1, \ldots, x_n) = 1/\overline{x}_n$. Sabemos que

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{\text{d}} n(0,\nu(\theta)),$$

com $\nu(\theta) = \frac{1}{E_\theta\left[\left(\frac{\partial}{\partial\theta}\log\left(\theta e^{-\theta X}\right)\right)^2\right]} = \frac{1}{E_\theta[(1/\theta - X)^2]} = \frac{1}{E_\theta[1/\theta^2 - 2X/\theta + X^2]} = 1/(1/\theta^2) = \theta^2$

Portanto, $\sqrt{n}(\hat{\theta} - \theta) \sim n(0,\sigma^2) \Rightarrow \hat{\theta} \sim n(\theta,\theta^2/n)$ e $\sqrt{n}(\hat{\theta} - \theta)/\theta \sim n(0,1)$. Dessa forma, poderemos construir o intervalo de confiança,

$$\begin{aligned}
IC(\theta) &= P\left[-z_{\alpha/2} < \frac{\left(\frac{1}{\overline{X}_n} - \theta\right)}{\theta/\sqrt{n}} < z_{\alpha/2}\right] \\
&= P\left[-z_{\alpha/2} < \frac{\sqrt{n}}{\theta}\left(\frac{1}{\overline{X}_n} - \theta\right) < z_{\alpha/2}\right] \\
&= P\left[-z_{\alpha/2} < \frac{\sqrt{n}}{\theta\overline{X}_n} - \sqrt{n} < z_{\alpha/2}\right] \\
&= P\left[\sqrt{n} - z_{\alpha/2} < \frac{\sqrt{n}}{\theta\overline{X}_n} < \sqrt{n} + z_{\alpha/2}\right] \\
&= P\left[\frac{1}{\sqrt{n} + z_{\alpha/2}} < \frac{\theta\overline{X}_n}{\sqrt{n}} < \frac{1}{\sqrt{n} - z_{\alpha/2}}\right] \\
&= P\left[\frac{\sqrt{n}}{\sqrt{n} + z_{\alpha/2}} < \theta\overline{X}_n < \frac{\sqrt{n}}{\sqrt{n} - z_{\alpha/2}}\right] \\
&= P\left[\frac{\sqrt{n}}{\sqrt{n}(1 + z_{\alpha/2}/\sqrt{n})} < \theta\overline{X}_n < \frac{\sqrt{n}}{\sqrt{n}(1 - z_{\alpha/2}/\sqrt{n})}\right] \\
&= P\left[\frac{1}{(1 + z_{\alpha/2}/\sqrt{n})} < \theta\overline{X}_n < \frac{1}{(1 - z_{\alpha/2}/\sqrt{n})}\right] \\
&= P\left[\frac{1/\overline{X}_n}{(1 + z_{\alpha/2}/\sqrt{n})} < \theta < \frac{1/\overline{X}_n}{(1 - z_{\alpha/2}/\sqrt{n})}\right].
\end{aligned}$$

Assim, $\overline{X}_n = 4$ e $n = 20$, então

$$IC(\theta) = P\left[\frac{1/4}{(1 + 1{,}96/\sqrt{20})} < \theta < \frac{1/4}{(1 - 1{,}96/\sqrt{20})}\right]$$

$$= \quad P\,[0{,}1738 < \theta < 0{,}4451]\,.$$

POPULAÇÃO BERNOULLI

Para o caso $X \sim Bernoulli(\theta)$, para uma amostra $n = 20$, e o estimador de θ pelo método da máxima verossimilhança $\hat{\theta}(x_1, \ldots, x_n) = 7/20$, pelo intervalo de Wald, considerando $z_{0{,}05/2} = 1{,}96$,

$$IC(\theta) = \left[\hat{\theta} - z_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \leq \theta \leq \hat{\theta} + z_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}\right].$$

Assim,

$$\begin{aligned} IC(\theta) &= P\left[7/20 - 1{,}96\sqrt{\frac{7/20(1-7/20)}{20}} \leq \theta \leq 7/20 + 1{,}96\sqrt{\frac{7/20(1-7/20)}{20}}\right] \\ &= P\,[0{,}35 - 1{,}2090 \leq \theta \leq 0{,}35 + 1{,}2090] \\ &= P\,[0{,}141 \leq \theta \leq 0{,}559] \end{aligned}$$

Questão 7

Seja uma amostra $\{0.1,0.3,0.35,0.5,0.7,0.8,0.82\}$. Aplique o teste de aderência de Pearson e Kolmogorov-Smirnov, considerando a hipótese H_0 : amostra vem de uma uniforme $U(0{,}1)$.

SOLUÇÃO

Hipóteses:

- H_0 : Amostra vem de uma uniforme $U(0{,}1)$
- H_1 : Não segue essa distribuição.

Nível de significância: $\alpha = 0{,}20$.

TESTE DE ADERÊNCIA QUI-QUADRADO (PEARSON)

Ordenando os dados em classes: sendo que f_i é a frequência observada, e $e_i = nP_i$ é a frequência

Classes	$\mathbf{f_i}$	$\mathbf{e_i}$
$0{,}0 \vdash 0{,}2$	1	1,4
$0{,}2 \vdash 0{,}4$	2	1,4
$0{,}4 \vdash 0{,}6$	1	1,4
$0{,}6 \vdash 0{,}8$	1	1,4
$0{,}8 \vdash 1{,}0$	2	1,4
Total	7	7

esperada, em que P_i é a distribuição uniforme. Considerando que cada categoria é equiprovável, já que a amostra é *iid*, então a função acumulada de $U(0{,}1)$ é $F(x) = x$, e que portanto, $P_i = 0{,}2$. Assim, $e_i = 7 \times 0{,}2 = 1{,}4$.

Calculando a estatística Q, temos

$$\begin{aligned} Q &= \sum_{i=1}^{5} \frac{(f_i - e_i)^2}{e_i} \\ &= \frac{(1-1{,}4)^2}{1{,}4} + \frac{(2-1{,}4)^2}{1{,}4} + \frac{(1-1{,}4)^2}{1{,}4} + \frac{(1-1{,}4)^2}{1{,}4} + \frac{(2-1{,}4)^2}{1{,}4} \end{aligned}$$

$$= \quad 0{,}8571.$$

Considerando $\alpha = 0{,}20$ e $\nu = 5 - 1 = 4gl$, então $\chi^2_{0,20;4} = 5{,}99$.
CONCLUSÃO: Como $Q < \chi^2_{0,20;4}$, não temos evidência para rejeitar a hipótese de que a amostra vem de uma uniforme $U(0{,}1)$.
TESTE DE ADERÊNCIA KOLMOGOROV-SMIRNOV

Seja a acumulada empírica $F_n(x)$:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(X_i).$$

E $F(x)$ é a função acumulada da distribuição em análise.

x	**F(x)**	**Fn(X)**	**|F(x) − Fn(x)|**
0,10	0,100	0,143	0,043
0,30	0,300	0,286	0,014
0,35	0,350	0,429	0,079
0,50	0,500	0,571	0,071
0,70	0,700	0,714	0,014
0,80	0,800	0,857	0,057
0,82	0,820	1,000	**0,180**

$D_{max} = 0{,}18$, verificando na tabela de Kolmogorov-Smirnov, $D_{critico} = 0{,}381$ para $\alpha = 0{,}20$ e $n = 7$. Assim, $D_{max} < D_{critico}$, não temos evidências para rejeitar a hipótese de que a amostra vem de uma uniforme $U(0{,}1)$.