Tópico 6 - Estimação por intervalos: Conceituação, interpretação e construção; Intervalo de confiança para grandes amostras

Ben Dêivide

6 de outubro de 2021

1 Conceitos iniciais

A ideia básica da estimação é observar uma amostra $X_1, X_2, \ldots, X_n$ de uma população com $f_X(x;\theta)$, sendo θ desconhecido, e a partir da amostra retirar informações sobre θ.

Na estimação pontual não temos a ideia da margem do erro que é cometido ao estimarmos o parâmetro. A estimação intervalar visa preencher esta lacuna criando um intervalo de possíveis valores para o parâmetro θ com margem de erro conhecido.

Definição 1 (Intervalo de Confiança). *Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ de uma população com $f_X(x;\theta)$, parametrizada por θ. Considere ainda duas estatísticas $T_1 = t_1(X_1, X_2, \ldots, X_n)$ e $T_2 = t_2(X_1, X_2, \ldots, X_n)$ satisfazendo $T_1 \leq T_2$, tal que $P_\theta(T_1 \leq \theta \leq T_2) = \gamma$, em que γ não depende de θ. Então o intervalo aleatório (T_1, T_2) é o estimador intervalar de confiança 100γ%(ou o intervalo de confiança) de θ, sendo γ o coeficiente de confiança, e T_1 e T_2 são os limites inferior e superior, respectivamente.* □

Exemplo 1 (Estimador intervalar). *Para uma amostra X_1, X_2, X_3, X_4 de uma população normal com média μ e variância 1, isto é, $X \sim N(\mu, 1)$. Um estimador intervalar possível de μ é $[\bar{X} - 1, \bar{X} + 1]$. Isto significa que o intervalo contém o valor do parâmetro μ com uma certa confiança.* □

Qual a vantagem de ter usado um estimador com menor precisão do que um estimador pontual? Pontualmente, estimamos μ por $\bar{X}$. E agora estimamos por $[\bar{X} - 1, \bar{X} + 1]$. Apesar de abrirmos mão de uma certa precisão, ganhamos alguma confiança, isto é, uma garantia de que nossa asserção está correta.

Outra importante observação é que o intervalo $[T_1, T_2]$, Definição 1, é a quantidade aleatória e não o parâmetro θ. Assim, não podemos afirmar que θ está dentro do intervalo $[T_1, T_2]$ com uma confiança $100\gamma\%$, e sim, que o intervalo $[T_1, T_2]$ contém o parâmetro θ com uma confiança $100\gamma\%$.

Diversos métodos são encontrados na literatura para obter um intervalo de confiança, tais como:

- inverter uma estatística de teste [Ver Casella (2001, port. pág. 376;)];
- intervalos bayesianos [Casela (2001, port. pág.390); Mood (1974, pág. 396)];

- método da quantidade pivotal.

Nos restringiremos ao último método.

Definição 2 (Quantidade pivotal). *Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ de uma população com $f_X(x;\theta)$, parametrizada por θ. Considere ainda uma medida $Q = q(X_1, X_2, \ldots, X_n;\theta)$ uma função de $X_1, X_2, \ldots, X_n$ e θ. Se Q tem distribuição independente de θ, então Q é chamado de quantidade pivotal.* $\square$

Exemplo 2. *Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ de uma população $X \sim N(\theta, 9)$. Assim, podemos afirmar que $Q = \bar{X} - \theta$ é uma quantidade pivotal, pois sua distribuição é normal, tal que $Q_1 \sim N(0, 9/n)$. Agora $Q_2 = \bar{X}/\theta$ não é uma quantidade pivotal, pois $Q_2 \sim N(0, 9/(n\theta^2))$ que depende de θ.* $\square$

Definição 3 (Método da quantidade pivotal). *Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ de uma população com $f_X(x;\theta)$, parametrizada por θ. Considere ainda uma quantidade pivotal $Q = q(X_1, X_2, \ldots, X_n;\theta)$ com fp ou fdp $f_Q(x)$ que não depende de θ. Fixado $0 < \gamma < 1$ existem q_1 e q_2 que dependem de γ tal que $P(q_1 < Q < q_2) = \gamma$. Se para cada amostra observada $(x_1, x_2, \ldots, x_n)$, $q_1 < q(x_1, x_2, \ldots, x_n;\theta) < q_2$ pode ser pivotado em $t_1(x_1, x_2, \ldots, x_n) < \tau(\theta) < t_2(x_1, x_2, \ldots, x_n)$ para t_1 e t_2, que não dependem de θ, então (T_1, T_2) é o intervalo de confiança $100\gamma\%$ para $\tau(\theta)$, sendo τ uma função de θ, em que $T_i = t_i(X_1, X_2, \ldots, X_n)$, $i = 1, 2$.* $\square$

Dessa definição, devemos fazer algumas observações sobre esse método:

i) q_1 e q_2 são independentes de θ, uma vez que a distribuição de Q também é;

ii) Para um valor fixado de γ, há muitos valores possíveis para q_1 e q_2 tal que $P(q_1 < Q < q_2) = \gamma$;

iii) Diferentes pares de q_1 e q_2 produzem diferentes pares de t_1 e t_2;

Dessa forma, precisamos de algum critério para escolher q_1 e q_2.

Definição 4 (Tamanho do intervalo). *Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ de uma população com $f_X(x;\theta)$, parametrizada por θ. Se (T_1, T_2) é o intervalo de confiança $100\gamma\%$ de θ, Definição 1, então o tamanho do intervalo, denotado por C, é definido por*

$$C_{T_1,T_2} = T_2 - T_1, \tag{1}$$

em que $T_2 > T_1$. $\square$

Dessa forma, devemos escolher o par q_1 e q_2 que resulte em menor comprimento C_{T_1,T_2}.

2 Intervalo de confiança (IC) para populações normais

2.1 IC para uma população normal

Considerando agora uma amostra aleatória $X_1, X_2, \ldots, X_n$ de tamanho n de uma população tenha distribuição normal, tal que, $X \sim N(\mu, \sigma^2)$, então apresentamos o seguinte Teorema

Teorema 1 (Bolfarine p. 79 no pdf, com as provas)**.** *Se $X \sim N(\mu, \sigma^2)$, então*

i) *$\bar{X}$ e S^2 são independentes;*

ii) $\dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1)$*, para σ^2 conhecido;*

iii) $\dfrac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t_{n-1}$*;*

iv) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$*;*

sendo $\bar{X} = (\sum_{i=1}^n X_i/n)$, $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$, χ^2_{n-1} é a variável aleatória de uma distribuição de qui-quadrado com $n-1$ graus de liberdade e t_{n-1} é a variável aleatória de uma distribuição de t de Student com $n-1$ graus de liberdade. □

Observamos que os itens (ii), (iii) e (iv) são todas quantidades pivotais, já que suas distribuições não dependem dos parâmetros de interesse. Dessa forma, poderemos construir intervalo de confiança para essas quantidades.

Teorema 2 (Intervalo de confiança para μ)**.** *Considerando o Teorema 1, o intervalo de confiança $100(1-\alpha)\%$ para μ é dado por:*

I) $P\left[\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$*, considerando a quantidade pivotal do item (iii), Teorema 1, para σ^2 conhecido. O quantil superior $100(\alpha/2)\%$ $z_{\frac{\alpha}{2}}$ tem distribuição normal padrão;*

II) $P\left[\bar{X} - t_{\frac{\alpha}{2},n-1}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\frac{\alpha}{2},n-1}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$*, considerando a quantidade pivotal do item (iv), Teorema 1, sendo S o desvio padrão amostral. O quantil superior $100(\alpha/2)\%$ $t_{\frac{\alpha}{2},n-1}$ tem distribuição t de Student com $n-1$ graus de liberdade.* □

Demonstração. Seja

$$Z = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1)$$

uma quantidade pivotal. Vamos agora pivotar, isto é,

$$\begin{array}{ccccc}
q_1 & \leq & \dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma} & \leq & q_2 \\
q_1\sigma & \leq & \sqrt{n}(\bar{X}-\mu) & \leq & q_2\sigma \\
q_1\dfrac{\sigma}{\sqrt{n}} & \leq & (\bar{X}-\mu) & \leq & q_2\dfrac{\sigma}{\sqrt{n}} \\
-\bar{X} + q_1\dfrac{\sigma}{\sqrt{n}} & \leq & -\mu & \leq & -\bar{X} + q_2\dfrac{\sigma}{\sqrt{n}} \\
\bar{X} - q_1\dfrac{\sigma}{\sqrt{n}} & \geq & \mu & \geq & \bar{X} - q_2\dfrac{\sigma}{\sqrt{n}} \\
T_1 = \bar{X} - q_2\dfrac{\sigma}{\sqrt{n}} & \leq & \mu & \leq & T_2 = \bar{X} - q_1\dfrac{\sigma}{\sqrt{n}}
\end{array}$$

O tamanho do intervalo é:

$$\begin{aligned}
C_{T_1,T_2} &= \left(\bar{X} - q_1\frac{\sigma}{\sqrt{n}}\right) - \left(\bar{X} - q_2\frac{\sigma}{\sqrt{n}}\right) \\
&= (q_2 - q_1)\frac{\sigma}{\sqrt{n}},
\end{aligned}$$

que é o mesmo que

$$1-\alpha=\int_{q_1}^{q_2} f_Z(k)dk = F_Z(q_2)-F_Z(q_1), \tag{2}$$

em que $f_Z(z)$ é a função densidade da normal padrão. Diferenciando (2) em relação a q_1, temos

$$\frac{d}{dq_1}(F_Z(q_2)-F_Z(q_1)) = \frac{dq_2}{dq_1}f_Z(q_2)-f_Z(q_1)=0. \tag{3}$$

Isto implica em

$$\frac{dq_2}{dq_1}=\frac{f_Z(q_1)}{f_Z(q_2)}. \tag{4}$$

Para minimizar C_{T_1,T_2}, portanto fazemos $dC_{T_1,T_2}/dq_1$, isto é,

$$\frac{d}{dq_1}(q_2-q_1)\frac{\sigma}{\sqrt{n}} = \left(\frac{dq_2}{dq_1}-1\right)\frac{\sigma}{\sqrt{n}}=0. \tag{5}$$

Substituindo (4) em (5), temos

$$\left(\frac{f_Z(q_1)}{f_Z(q_2)}-1\right)\frac{\sigma}{\sqrt{n}}=0. \tag{6}$$

Isso só ocorrerá se $q_2 = q_1$. Entretanto, $\int_{q_1}^{q_1} f_Z(k)dk \neq 1-\alpha$. Ou pode ser $q_1 = -q_2$, já que $f_Z(q_1) = f_Z(-q_1)$ pela distribuição ser simétrica, daí poderemos obter $\int_{q_1}^{-q_1} f_Z(k)dk = 1-\alpha$. Portanto, para minimizar C_{T_1,T_2}, $q_1 = -q_2$. Considerando $q_1 = z_{\frac{\alpha}{2}}$, então $-q_2 = -z_{\frac{\alpha}{2}}$, e segue o resultado do item (I). O resultado do item (II) segue nos mesmos moldes, substituindo apenas σ por S. □

Teorema 3 (Intervalo de confiança para σ^2). *Considerando o Teorema 1, o intervalo de confiança $100(1-\alpha)\%$ para σ^2 é dado por:*

$$P\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right] = 1-\alpha, \tag{7}$$

sendo $\chi^2_{1-\alpha/2}$ o quantil superior $100(1-\alpha/2)\%$ e $\chi^2_{\alpha/2}$ o quantil superior $100(\alpha/2)\%$ da distribuição de qui-quadrado com $n-1$ graus de liberdade. □

Demonstração. Mood(1974, pág. 382-384); Daniel(2009, p. 332-333) □

Teorema 4 (Intervalo de confiança para p). *Pode ser abordado em Mood (1974)[pág 394-396] caso aproximado, em Daniel (2009, p. 275); Casella(2001, port pag. 446-449). Para o intervalo exato, Daniel (2009, p. 274).*

2.2 IC para mais de uma população normal

Teorema 5 (Intervalo de confiança para $\mu_1-\mu_2$). *Material antigo escrito, com base em Daniel(2009, pág. 380-383); Mood(1974, pág. 432-437)*

Teorema 6 (Intervalo de confiança para $\sigma^2-\sigma_2$). *Material antigo escrito, com base em Daniel(2009, pág. 464-470); Mood(1974, pág. 438-442)*

Teorema 7 (Intervalo de confiança para p_1-p_2). *Daniel(2009, pág. 439-440);*

3 Intervalos de confiança para grandes amostras

Estudamos na estimação pontual que as vezes é possível encontrar uma sequência de estimadores $W_n(X_1, X_2, \ldots, X_n)$ que assintoticamente tem distribuição normal com média θ e variância $\sigma_n^2(\theta)$, em que $\sigma_n^2(\theta)$ indica que a variância é uma função de θ e do tamanho da amostra n. Em particular, temos os estimadores de máxima verossimilhança (EMV), denotado por $\hat{\theta}_n(X_1, X_2, \ldots, X_n)$, que apresentam essa propriedade.

Teorema 8 (Teorema de Slutsky)**.** *Se* $X_n \xrightarrow{\mathrm{d}} X$ *em distribuição e* $Y_n \xrightarrow{p} a$*, uma constante, em probabilidade, então*

a) $Y_n X_n \xrightarrow{\mathrm{d}} aX$*, em distribuição;*

b) $X_n + Y_n \xrightarrow{\mathrm{d}} X + a$*, em distribuição.*

Teorema 9 (Eficiência e consistência assintótica dos EMV)**.** *Seja uma amostra aleatória* $X_1, X_2, \ldots, X_n$ *iid com fp ou fdp* $f_X(x;\theta)$*. Supondo que* $\hat{\theta}$ *denote o EMV de* θ *e que* $\tau(\theta)$ *seja uma função contínua de* θ*, sob condições de regularidade de* $f_X(x;\theta)$*, então*

$$\sqrt{n}[\tau(\hat{\theta}) - \tau(\theta)] \to N(0, \sigma_n^2(\theta)), \tag{8}$$

em que $\sigma_n^2(\theta)$ *é o limite inferior da cota de Cramer-Rao, dado por:*

$$\sigma_n^2(\theta) = \frac{\left(\frac{d}{d\theta}\tau(\theta)\right)^2}{E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_X(X;\theta)\right)^2\right]}.$$

Dizemos que $\tau(\hat{\theta})$ *é um estimador consistente e assintoticamente eficiente de* $\tau(\theta)$*.* □

Demonstração. Vamos fazer a prova considerando o EMV $\hat{\theta}$ e X uma v.a. contínua. Considerando que $\ell(\theta; X_1, X_2, \ldots, X_n) = \sum_{i=1}^n \log f_X(X_i;\theta)$ é a função log de verossimilhança, denote $\ell'(\theta, \boldsymbol{X})$ a primeira derivada da função log verossimilhança com relação a θ. Expanda essa derivada em torno do verdadeiro valor do parâmetro θ, denotado por θ_0, isto é,

$$\ell'(\theta, \boldsymbol{X}) = \ell'(\theta_0, \boldsymbol{X}) + (\theta - \theta_0)\ell''(\theta_0, \boldsymbol{X}). \tag{9}$$

Agora, substitua o EMV $\hat{\theta}$ para θ. Como $\ell'(\hat{\theta}, \boldsymbol{X}) = 0$, então

$$(\hat{\theta} - \theta_0) = \frac{-\ell'(\theta_0, \boldsymbol{X})}{\ell''(\theta_0, \boldsymbol{X})}. \tag{10}$$

Pré-multiplicando $\sqrt{n}$ em (10), em ambos os lados, temos

$$
\begin{aligned}
\sqrt{n}(\hat{\theta}-\theta_0) &= \sqrt{n}\frac{-\ell'(\theta_0,\boldsymbol{X})}{\ell''(\theta_0,\boldsymbol{X})}\\
&= \frac{\sqrt{n}\sqrt{n}}{\sqrt{n}}\frac{-\ell'(\theta_0,\boldsymbol{X})}{\ell''(\theta_0,\boldsymbol{X})}\\
&= \frac{(\sqrt{n})^2}{\sqrt{n}}\frac{-\ell'(\theta_0,\boldsymbol{X})}{\ell''(\theta_0,\boldsymbol{X})}\\
&= \frac{n}{\sqrt{n}}\frac{-\ell'(\theta_0,\boldsymbol{X})}{\ell''(\theta_0,\boldsymbol{X})}\\
&= \frac{-\frac{1}{\sqrt{n}}\ell'(\theta_0,\boldsymbol{X})}{\frac{1}{n}\ell''(\theta_0,\boldsymbol{X})} \qquad (11)\\
&= \frac{\frac{1}{\sqrt{n}}\ell'(\theta_0,\boldsymbol{X})}{-\frac{1}{n}\ell''(\theta_0,\boldsymbol{X})} \qquad (12)
\end{aligned}
$$

Usando primeiro a expressão do numerador de (11), temos que

$$
\begin{aligned}
E\left[\frac{1}{\sqrt{n}}\ell'(\theta_0,\boldsymbol{X})\right] &= \frac{1}{\sqrt{n}}E\left[\ell'(\theta_0,\boldsymbol{X})\right]\\
&= \frac{1}{\sqrt{n}}E\left[\frac{\partial}{\partial\theta_0}\sum_{i=1}^{n}\log f_X(X_i;\theta_0)\right]\\
&= \frac{1}{\sqrt{n}}E\left[\sum_{i=1}^{n}\frac{\partial}{\partial\theta_0}\log f_X(X_i;\theta_0)\right]\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\left[\frac{\partial}{\partial\theta_0}\log f_X(X_i;\theta_0)\right]\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}E\left[\frac{\partial}{\partial\theta_0}\log f_X(X_i;\theta_0)\right]\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\int_{\mathcal{X}}\frac{\partial}{\partial\theta_0}\log(f_X(t_i;\theta_0))f_X(t_i;\theta_0)dx_i\right)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\int_{\mathcal{X}}\frac{\frac{\partial}{\partial\theta_0}f_X(t_i;\theta_0)}{f_X(t_i;\theta_0)}f_X(t_i;\theta_0)dx_i\right)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\int_{\mathcal{X}}\frac{\partial}{\partial\theta_0}f_X(t_i;\theta_0)dx_i\right)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta_0}\underbrace{\int_{\mathcal{X}}f_X(t_i;\theta_0)dx_i}_{=1}\right) = 0. \qquad (13)
\end{aligned}
$$

A variância pode ser expressa da seguinte forma:

$$
\begin{aligned}
Var\left[\frac{1}{\sqrt{n}}\ell'(\theta_0,\boldsymbol{X})\right] &= E\left[\left(\frac{1}{\sqrt{n}}\ell'(\theta_0,\boldsymbol{X})\right)^2\right] - \left(E\left[\frac{1}{\sqrt{n}}\ell'(\theta_0,\boldsymbol{X})\right]\right)^2\\
&= E\left[\left(\frac{1}{\sqrt{n}}\ell'(\theta_0,\boldsymbol{X})\right)^2\right]\\
&= \frac{1}{n}E\left[\left(\ell'(\theta_0,\boldsymbol{X})\right)^2\right]\\
&= \frac{1}{n}E\left[\left(\frac{\partial}{\partial\theta_0}\log L(\theta_0;\boldsymbol{X})\right)^2\right]. \qquad (14)
\end{aligned}
$$

Existe um resultado para amostras *iid* que $E\left[\left(\frac{\partial}{\partial\theta_0}\log L(\theta_0;\boldsymbol{X})\right)^2\right] = nE\left[\left(\frac{\partial}{\partial\theta_0}\log f_X(X;\theta_0)\right)^2\right]$. Ver Casella (2001, port. p.300-301) e no material escrito de inf II. Assim,

$$
\begin{aligned}
Var\left[\frac{1}{\sqrt{n}}\ell'(\theta_0,\boldsymbol{X})\right] &= \frac{n}{n}E\left[\left(\frac{\partial}{\partial\theta_0}\log f_X(X;\theta_0)\right)^2\right]\\
&= E\left[\left(\frac{\partial}{\partial\theta_0}\log f_X(X;\theta_0)\right)^2\right] \qquad (15)\\
&= \frac{1}{\sigma_n^2(\theta_0)}. \qquad (16)
\end{aligned}
$$

Pelo Teorema Central do limite, temos que

$$\frac{\frac{1}{\sqrt{n}}\ell'(\theta_0,\boldsymbol{X}) - 0}{\sqrt{1/\sigma_n^2(\theta_0)}} \xrightarrow{\text{d}} N(0,1), \qquad (17)$$

ou

$$\frac{1}{\sqrt{n}}\ell'(\theta_0,\boldsymbol{X}) \xrightarrow{\text{d}} N(0,1/\sigma_n^2(\theta_0)). \qquad (18)$$

Se considerarmos o denominador de (11), temos

$$
\begin{aligned}
-\frac{1}{n}\ell''(\theta_0,\boldsymbol{X}) &= -\frac{1}{n}\left(\frac{\partial^2}{\partial\theta_0^2}\sum_{i=1}^{n}\log f_X(X_i;\theta_0)\right)\\
&= -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial^2}{\partial\theta_0^2}\log f_X(X_i;\theta_0)\right)
\end{aligned} \qquad (19)
$$

Observe que $\frac{\partial^2}{\partial\theta_0^2}\log f_X(X_i;\theta_0)$ pode ser encarada como uma variável aleatória. Se denotarmos $\frac{\partial^2}{\partial\theta_0^2}\log f_X(X_i;\theta_0) = Y_i$, então

$$-\frac{1}{n}\ell''(\theta_0,\boldsymbol{X}) = -\frac{1}{n}\sum_{i=1}^{n}Y_i = -\bar{Y}. \qquad (20)$$

Pela Lei Fraca dos Grandes números,

$$-\bar{Y} \xrightarrow{p} -E\left[\frac{\partial^2}{\partial\theta_0^2}\log f_X(X_i;\theta_0)\right] = E\left[\left(\frac{\partial}{\partial\theta_0}\log L(\theta_0;X)\right)^2\right] = \frac{1}{\sigma_n^2(\theta_0)}. \tag{21}$$

Portanto, pelo Teorema de Slutsky, item (a), como

$$-\frac{1}{n}\ell''(\theta_0,\boldsymbol{X}) \xrightarrow{p} \frac{1}{\sigma_n^2(\theta_0)}$$

e

$$\frac{1}{\sqrt{n}}\ell'(\theta_0,\boldsymbol{X}) \xrightarrow{\text{d}} N(0,1/\sigma_n^2(\theta_0)),$$

então considerando que $W \sim N(0,1/\sigma_n^2(\theta_0))$, logo

$$\frac{\frac{1}{\sqrt{n}}\ell'(\theta_0,\boldsymbol{X})}{-\frac{1}{n}\ell''(\theta_0,\boldsymbol{X})} \xrightarrow{\text{d}} \sigma_n^2(\theta_0)W. \tag{22}$$

Dessa forma, $\sigma_n^2(\theta_0)W$ também tem distribuição normal com parâmetros

$$E[\sigma_n^2(\theta_0)W] = \sigma_n^2(\theta_0)E[W] = 0,$$

e

$$Var[\sigma_n^2(\theta_0)W] = \sigma_n^4(\theta_0)Var[W] = \frac{\sigma_n^4(\theta_0)}{\sigma_n^2(\theta_0)} = \sigma_n^2(\theta_0),$$

isto é, $\sigma_n^2(\theta_0)W \sim N(0,\sigma_n^2(\theta_0))$. Logo,

$$\sqrt{n}(\theta - \theta_0) \xrightarrow{\text{d}} N(0,\sigma_n^2(\theta_0)),$$

provando o Teorema. □

Exemplo 3 (Normalidade e consistência assintótica)**.** *O Teorema 9 mostra que estimadores EMV $\tau(\hat{\theta})$ de $\tau(\theta)$ são assintoticamente normal, e por consequência eficientes. Ainda mais, a normalidade assintótica implica em consistência. Suponha que*

$$\sqrt{n}\frac{W_n - \mu}{\sigma} \xrightarrow{\text{d}} Z \text{ em distribuição,}$$

em que $Z \sim N(0,1)$. Aplicando o Teorema de Slutsky, temos

$$W_n - \mu = \underbrace{\left(\frac{\sigma}{\sqrt{n}}\right)}_{\xrightarrow{p}\left(\frac{\sigma}{\sqrt{n}}\right)}\underbrace{\left(\sqrt{n}\frac{W_n - \mu}{\sigma}\right)}_{\xrightarrow{\text{d}} Z} \xrightarrow{\text{d}} \lim_{n\to\infty}\left(\frac{\sigma}{\sqrt{n}}\right)Z = 0,$$

deste modo, $W_n - \mu \to 0$ converge em distribuição. e o Teorema (Casella, port. pag. 211) mostra que a convergência em distribuição para um ponto implica em convergência em probabilidade. Logo, $W_n \xrightarrow{p} \mu$, isto é, W_n é um estimador consistente. □

Como $\sigma_n^2(\theta)$ depende de θ, uma aproximação (Método Delta) para a variância pode ser expresso por

$$\sigma_n^2(\hat{\theta}|\theta) = \sigma_n^2(\hat{\theta}) \approx \frac{\left(\frac{d}{d\theta}\tau(\theta)\right)^2 |_{\theta=\hat{\theta}}}{E_\theta\left[\left(\frac{\partial}{\partial\theta}\log L(\theta;\boldsymbol{X})\right)^2\right] |_{\theta=\hat{\theta}}}, \tag{23}$$

em que $L(\theta;\boldsymbol{X}) = L(\theta; X_1, X_2, \ldots, X_n) = \prod_{i=1}^n f_X(X_i;\theta)$ é a função de verossimilhança.

A quantidade,

$$E_\theta\left[\left(\frac{\partial}{\partial\theta}\log L(\theta;\boldsymbol{X})\right)^2\right] \tag{24}$$

é conhecida como número de informação ou informação de Fisher. Uma outra forma de apresentar (24) é

$$E_\theta\left[\left(\frac{\partial}{\partial\theta}\log L(\theta;\boldsymbol{X})\right)^2\right] = -E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log L(\theta;\boldsymbol{X})\right]. \tag{25}$$

Considerando uma amostra iid, a expressão (24) pode ser dada por

$$E_\theta\left[\left(\frac{\partial}{\partial\theta}\log L(\theta;\boldsymbol{X})\right)^2\right] = nE_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_X(X;\theta)\right)^2\right]. \tag{26}$$

A prova desses resultados está no material escrito de Inf II.

Na prática,

$$\begin{aligned}
\sqrt{n}[\tau(\hat{\theta}) - \tau(\theta)] &\to N(0, \sigma_n^2(\theta)) \\
\frac{\sqrt{n}[\tau(\hat{\theta}) - \tau(\theta)]}{\sqrt{\sigma_n^2(\theta)}} &\to N(0,1) \\
\tau(\hat{\theta}) - \tau(\theta) &\to N(0, \sigma_n^2(\theta)/n) \\
\tau(\hat{\theta}) &\to N(\tau(\theta), \sigma_n^2(\theta)/n)
\end{aligned}$$

Com essas informações, poderemos agora construir intervalos de confiança para grandes amostras. Usando a aproximação em (23), temos

$$\frac{\sqrt{n}[\tau(\hat{\theta}) - \tau(\theta)]}{\sqrt{\sigma_n^2(\hat{\theta})}} \xrightarrow{\text{d}} N(0,1), \tag{27}$$

pois, pelo Teorema 9 sabemos que $\sqrt{n}[\tau(\hat{\theta}) - \tau(\theta)] \xrightarrow{\text{d}} N(0,1)$. Pelo mesmo Teorema, sabemos que os estimadores de EMV são consistentes assintoticamente, e ainda sabendo pelo princípio da invariância (Mood, 1974, p. 284; Casela, 2001, port p. 285) que se $\hat{\theta}$ é um EMV de θ, então $\sigma_n^2(\hat{\theta})$ também é um EMV de $\sigma_n^2(\theta)$. Logo, $\sigma_n^2(\hat{\theta}) \xrightarrow{p} \sigma_n^2(\theta)$. Assim, pelo Teorema de Slutsky fica provado a convergência em distribuição de (27).

Assim, um intervalo de confiança aproximado é

$$\tau(\hat{\theta}) - z_{\frac{\alpha}{2}}\sqrt{\sigma_n^2(\hat{\theta})} \leq \tau(\theta) \leq \tau(\hat{\theta}) + z_{\frac{\alpha}{2}}\sqrt{\sigma_n^2(\hat{\theta})}, \tag{28}$$

sendo $z_{\frac{\alpha}{2}}$ o quantil superior $100(\alpha/2)\%$ com distribuição normal padrão.

Exemplo 4 (Intervalos de confiança para grandes amostras)**.** *Seja uma amostra aleatória $X_1, X_2, \ldots, X_n$ de uma população com distribuição de Bernoulli(p). Construa um intervalo de confiança aproximado para p. Sabemos que o estimador EMV de θ é $\hat{p} = \bar{X}$ (Casella, port. p. 283). Para calcular $\sigma_n^2(p)$, usaremos a aproximação de* (23)*, isto é,*

$$
\begin{aligned}
\sigma_n^2(\hat{p}) &\approx \frac{\left(\frac{d}{dp}\tau(p)\right)^2|_{p=\hat{p}}}{E_p\left[\left(\frac{\partial}{\partial p}\log L(p;\boldsymbol{X})\right)^2\right]|_{p=\hat{p}}} \\
&\approx \frac{1}{E_p\left[\left(\frac{\partial}{\partial p}\log L(p;\boldsymbol{X})\right)^2\right]|_{p=\hat{p}}} \\
&\approx \frac{1}{-\frac{\partial^2}{\partial p^2}\log L(p;\boldsymbol{X})|_{p=\hat{p}}} \\
&\approx \frac{n}{\hat{p}(1-\hat{p})}, \ \textit{para detalhes ver Inf II (Lucas, p. 46)}
\end{aligned}
$$

Assim, um intervalo de confiança com base em (27) *é*

$$\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}. \tag{29}$$

sendo $z_{\frac{\alpha}{2}}$ o quantil superior $100(\alpha/2)\%$ com distribuição normal padrão.

Outro método de encontrar intervalos de confiança aproximado se baseiam em encontrar pivôs aproximados (ou assintóticos). Se tivermos quaisquer estatísticas W e V e um parâmetros θ de modo que à medida que $n \to \infty$,

$$\frac{W-\theta}{V} \xrightarrow{\text{d}} N(0,1), \text{Ver detalhes, Casella, port. p. 440} \tag{30}$$

então podemos formar um intervalo de confiança aproximado para θ dado por

$$W - z_{\frac{\alpha}{2}}V \leq \tau(\theta) \leq W + z_{\frac{\alpha}{2}}V, \tag{31}$$

que é essencialmente o intervalo do Tipo Wald. A aplicação direta do Teorema do Limite Central e o Teorema de Slutsky, geralmente resultará em um intervalo de confiança aproximado.

Exemplo 5. *Se $X_1, X_2, \ldots, X_n$ são iid com média μ e variância σ^2, então a partir do Teorema do Limite central, temos que*

$$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \xrightarrow{\text{d}} Z \sim N(0,1),$$

que implica também que

$$\sqrt{n}(\bar{X}-\mu) \xrightarrow{\text{d}} Z_{0,\sigma^2} \sim N(0,\sigma^2). \tag{32}$$

Assim, usando o Teorema de Slutsky,

$$\frac{\sqrt{n}(\bar{X}-\mu)}{S} \xrightarrow{\text{d}} \frac{Z_{0,\sigma^2}}{\sigma}.$$

Isso ocorre pois, há a convergência em distribuição de (32) *e a convergência em probabilidade de* $S^2 \xrightarrow{p} \sigma^2$ *(Ver Casella port pag. 208; material escrito de Inf II, pág. 58-61). Logo,*

$$\frac{Z_{0,\sigma^2}}{\sigma} \sim Z \sim N(0,1). \tag{33}$$

Outra forma de mostra isso é:

$$\underbrace{\left(\frac{\sigma}{S}\right)}_{\xrightarrow{p} 1} \underbrace{\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\right)}_{\xrightarrow{d} Z} \sim 1 \times Z \sim N(0,1), \qquad \textit{(usando o Teorema de Slutsky)},$$

Dessa forma poderemos construir um intervalo de confiança aproximado $1-\alpha$, *tal que*

$$\bar{X} - z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \tag{34}$$

sendo $z_{\frac{\alpha}{2}}$ *o quantil superior* $100(\alpha/2)\%$ *com distribuição normal padrão.* $\square$