2 Verossimilhança para um parâmetro

2.3 Aproximações e Assintótica de verossimilhança

Os exemplos na seção 2.1 mostram que mesmo para problemas simples de se obter, é difícil um entendimento das caracteristicas da amostra que influenciam no intervalo de confiança. Além do mais nós não temos, ainda idéia sobre qual valor de c^* é requerido para fazer intervalo de confiança baseado na verossimilhança dado por

$$\{\theta \in \Omega : D(\theta) < c^*\}$$

Seja 95%, ou 99%, intervalos de confiança. Nesta seção usamos aproximações da função de verossimilhança que ocorrem como $n \to \infty$ para direcionar ambos requisitos. Todos os resultados surgem da seguinte expansão da série de Taylor em torno do EMV

$$l(\theta) = l(\hat{\theta}) + (\theta - \hat{\theta})\ l'(\hat{\theta}) + \frac{1}{2}\ (\theta - \hat{\theta})^2\ l''(\theta^*) \ \text{ para } |\theta^* - \theta| \leq |\hat{\theta} - \theta|$$

$$l'(\theta) = l'(\hat{\theta}) + (\theta - \hat{\theta})\ \ l''(\theta^+) \ \text{ para } |\theta^+ - \theta|\ \leq |\hat{\theta} - \theta|$$

Para simplificar a notação usaremos $U(\theta) = l'(\theta)$, indicando a função escore, e usaremos $I_O(\theta) = -l''(\theta)$ indicando a informação observada, e

$$I_E(\theta) = E\{I_O(\theta)\} = E\left\{-\frac{\delta^2}{\delta\theta^2} logL(\theta; X)\right\}$$

indicando a informação esperada, ou informação de Fisher. Observe que aqui nós olhamos para as propriedades de U e I_O, termos sob considerção das propriedades da amostra ,i,e substitui-se x por X. Estes dois retmos são vitais para determinar o comportamento da expanção de verossimilhança **??** e **??** . Assim a função de verossimilhança e a função e log-verissimilhança são funções aleatórias com variaveis aleatórias determinando a posição do máximo, $\hat{\theta}$, o gradiente da função, U(θ), e a curvatura $I_O(\theta)$.

2.3.1 Resutados Práticos e Definições

Definição: se $T = T(X$ possuir a propriedade $E(T) = g(\theta)$, para qualquer g de θ ,então T ,é considerado um estimador não viciado de $g(\theta)$.

Lemma 1: $E\{U(\theta)\}$ e $Var\{U(\theta)\} = I_E(\theta)$
Provado em M232.

Lemma 2: Se T = T (X) é algum estimador não viciado para θ i.e E(T)= θ, então

$$Var(T) \geq [I_E(\theta)]^{-1}$$

Indica a limitr inferior de Cramer-Rao .
Provado em M232.

Lemma 3: Se T= T(X) for um estimador não viciado para $g(\theta)$, então

$$Var(T) \geq [g'(\theta)]^2[I_E(\theta)]^{-1}$$

Prova:(Não examinada) Seja $\phi = g(\theta)$, então pelo Lemma 2 o limite inferior de Cramér-Rao para estimadores não viciados de ϕ é $[I_E(\phi)]$ onde

$$I_E(\phi) = E\left\{\frac{\delta^2}{\delta\phi^2} l(\phi)\right\}$$

,

entretanto

$$\frac{\delta\theta}{\delta\phi} = \frac{\delta\theta}{\delta\phi}\frac{\delta l}{\delta\theta}$$

então

$$\frac{\delta^2 l}{\delta\phi^2} = \frac{\delta^2\theta}{\delta\phi^2}\frac{\delta l}{\delta\theta} + \left(\frac{\delta\theta}{\delta\phi}\right)^2\frac{\delta^2 l}{\delta\theta^2}$$

$$= \frac{\delta^2\theta}{\delta\phi}U(\theta) + \left(\frac{1}{g(\theta)}\right)^2 l''(\theta)$$

assim,

$$E\left\{-\frac{\delta^2}{\delta\phi}l(\phi)\right\} = 0 + \frac{I_E(\theta)}{[g'(\theta)]^2}$$

Aqui o ultimo passo é baseado no Lema 1 e na difinição de informação esperada. O resultado é obtido derivando a expressão **??**

2.3.2 Resultados Principais

Embora os resultados clássicos da teoria de verossimilhança que são dados nesta seção, sejam especialmente apreciaveis eles não valem universalmente, nós devemos assumir a influência das seguintes condições de regularidade.

R1: Ω é de dimensão finita e o verdadeiro valor de θ está dentro de Ω.

R2: Dois diferentes valores de θ não resultam na mesma $f(x;\theta)$

R3: As 3 primeiras derivadas de $l(\theta)$ existem em uma vizinhança do verdadeiro valor de θ.

Estas condições em assencia apontam para o fato que a dimensão da distribuição não dependem do parametro e assegure que com $n \to \infty$ a log-verossimilhança converjiu para uma função quadrática em torno do EMV e, e assim dependa apenas da posição do EMV e a curvatura da verossimilhança no EMV.

Teorema 1: Para um problema regular de estimação, no limite $n \to \infty$, se θ é o verdadeiro valor do parâmetro então

$$\sqrt{I_E(\theta)}(\hat{\theta} - \theta) \sim N(0;1)$$

i,e $\hat{\theta} \sim N(\theta, [I_E(\theta)]^{-1})$, nós denominamos isso de distirbuição assintótica de θ

Prova: Este resultado tem base em torno da expansão **??**, veja MATH232 para maiores detalhes. Em sintese é expressado como a seguir

$$U(\theta) = -(\theta - \hat{\theta})I_O(\theta^+)$$

Desde que $\hat{\theta} \to \theta$ com $n \to \infty$ então $\theta^+ \to \theta$ também. Nós podemos aproximar $I_O(\theta)$ por $I_E(\theta)$ desde que

$$\frac{I_E(\theta)}{I_O(\theta)} \to 1$$

Desde que $U(\theta)$ seja a soma de variáveis aleatórias é então aplicado o teorema central do limite, pelo Lema 1.

$$\frac{U(\theta)}{[I_E(\theta)]^{\frac{1}{2}}} \to Z \quad \text{onde} \quad Z \sim N(0,1)$$

.

desde que

$$\begin{aligned}\sqrt{I_E(\theta)}(\hat{\theta}-\theta) &= \frac{U(\theta)}{[I_E(\theta)]^{\frac{1}{2}}}.\frac{I_E(\theta)}{I_O(\theta)} \\ &\sim Zx1 = Z\end{aligned}$$

Corolário 1: Se $\phi = g(\theta)$ então $\phi \sim N(\phi, [g'(\theta)]^2[I_E(\theta)]^{-1})$ com $n \to \infty$.
Prova: Isso segue os Lemas 2 e 3.

A consequência importante do corolário 1 é que

$$Var(\phi) = [g'(\theta)]^2 Var(\hat{\theta})$$

.

Colorário 2: Qualquer termo assintoticamente equivalente a $I_E(\theta)$ pode ser usado no lugar do Teorema 1, então

$$\begin{aligned}\sqrt{I_E(\hat{\theta})}(\hat{\theta}-\theta) &\sim N(0,1) \\ \sqrt{I_O(\theta)}(\hat{\theta}-\theta) &\sim N(0,1) \\ \sqrt{I_O(\hat{\theta})}(\hat{\theta}-\theta) &\sim N(0,1)\end{aligned}$$

Mas as propriedades de convergência para cada um destes será diferente.

Teorema 2: Para um problema regular no limite com $n \to \infty$, se θ for o verdadeiro valor do parâmetro então

$$D(\theta) = 2[l(\hat{\theta}) - l(\theta)] \sim \chi_1^2$$

.

Prova: Este resultado é baseado na expansão **??** e **??**, veja M232 para maiores detalhes. Em sintese o resultado é o seguinte:

$$\begin{aligned} D(\theta) &= \{l(\hat{\theta}) - [l(\hat{\theta}) + (\theta - \hat{\theta})U(\hat{\theta}) - \frac{1}{2}(\theta - \hat{\theta})^2 I_O(\theta^*)]\} \\ &= (\hat{\theta} - \theta)^2 I_O(\theta^*) \end{aligned}$$

Como na prova do teorema 1 $\theta^* \to \theta$ então

$$\begin{aligned} D(\theta) &= [\sqrt{I_E(\theta)}(\hat{\theta} - \theta)^2]^2 \frac{I_O(\theta^*)}{I_E(\theta)} \\ &\sim Z^2 x 1 = Z^2 \end{aligned}$$

Do teorema 1, Z é $N(0,1)$, então Z^2 é χ_i^2, é resultado disso.

2.3.3 Discussão dos Resultados Principais

O Teorema 1 e os corolários associados dizem que:

- O estimador de verossimilhança de $\hat{\theta}$ de θ é assintóticamente não viciado, i.e. $E(\hat{\theta}) \sim \theta$
- $\hat{\theta}$ assintoticamente atinge o limite inferior de Cramér-Rao, i.e. $Var(\hat{\theta}) \to [I_E(\theta)]^{-1}$.
- Chamamos $[I_E(\theta)]^{-1/2}$ de erro padrão de $th\hat{e}ta$, algumas vezes denotado por $se(\hat{\theta})$.
- Podemos construir um intervalo de confiança assintoticamente valido $100(1-\alpha)\%$ para θ na forma $\hat{\theta} \pm z_{\alpha/2}[I_E(\theta)] - \frac{1}{2}$, onde $\Phi(z_{\alpha/2}) = 1 - \alpha/2$,e.g. se $\alpha = 0,05$ então $z_{\alpha/2} = 1,96$. Podemos denotar este intervalo por $(\tilde{\theta}_l, \tilde{\theta}_u) = (th\hat{e}ta - z_{\alpha/2} x se(\hat{\theta}), \hat{\theta} + z_{\alpha/2} x se(th\hat{e}ta))$.
- O estimador de máxima verissimilhança $\hat{\phi} = g(\hat{\theta}) de \phi = g(\theta)$ é aasintoticamente nhão viciado, i.e $Eg(\hat{\theta}) \to g(\theta)$.
- $\phi = g(\hat{\theta})$ assintoticamente atinge o limite inferior de Cramér-Rao, i.e. $Var(g\hat{\theta}) \to [g'(\theta)]^2 [I_E(\theta)]^{-1} = [I_E(\phi)]^{-1}$

- Chamamos $[g'(\theta)][I_E(\theta)]^{-\frac{1}{2}}$ o erro padrão de ϕ.

- Podemos construir um intervalo de confiança $100(1-\alpha)\%$. Assintóticamente válido da forma $g(\hat{\theta}) \pm Z_{\alpha/2}[g'(\theta)][I_E(\theta)]^{-\frac{1}{2}}$ onde $\Phi(Z_{\alpha/2}) = 1-\alpha/2$. Denotamos este intervalo por $(\tilde{\phi}_e, \tilde{\phi}_u) = (\hat{\phi}) - Z_{\alpha/2}\, x\, se(\hat{\phi}), \hat{\phi} + Z_{\alpha/2}\, x\, se(\hat{\phi})$.

- Em todas as declarações acima $I_E(\theta)$ pode ser substituido pelos seus termos assintoticamente equivalentes $I_E(\hat{\theta}), I_O(\hat{\theta})$ da mesma forma $g'(\theta)$pode ser substituido por$g'(\hat{\theta})$. Houve muito estudo para determinar que estes são os melhores termos para se usar na prática. E foi verificado que $I_O(\hat{\theta})$ possui as melhores propriedades. Contudo é óbvio que queremos obter a melhor aproximação da função de log-verossimilhança, então observando a curvatura observada no máximo limitado para ser o melhor do que usar a curvatura esperadapara o verdadeiro valor deθ. Além disso a informação observada na EMV é a mais fácil entre os quatro possiveis termos, para avaliar, o que é sempre muito bom.

Correspondentemente, o teorema 2 fornece:

- Pelos argumentos na seção**??**temos que o estimador mais sensível no intervalo de confiança é da forma:

  $$\theta\ E\ \Omega : D(\theta) < c^*$$

  Para algum valor de c^*. Para resaltar que o intervalo é exatamente um intervalo de confiança a $100(1-\alpha)\%$ precisamos escolher c^4 que em repetidas amostras,a, proporção de intervalos que contém o verdadeiro valor de θ seja $1-\alpha$. Isso é genericamente impossível de se fazer sem valer-se do uso de métodos computacionais. Entretanto enquanto n se torna grande o Teorema 2 sugere que uma escolha alternativa seja $c^* = c_\alpha$, com $\mathrm{P}\{\chi_1^2 \geq c^\alpha\} = \alpha$, e.g se $\alpha = 0,05$ então $c = 3,84$. Uma simples e sensível apriximação é usar esta escolha para c^*mesmo se n for pequeno e assumir que $100(1-\alpha)$seja um intervalo aproximado. Claramente este será um intervalo de confiança baseado na verissimilhança apropriado, mas para pequenas amostras o verdadeiro valor de α pode ser ligeiramente diferente do requerido

- Como discutido na seção **??** se o intervalo de confiança baseado na verossimilhança de $100(1-\alpha)\%$ para θ está em $(\hat{\theta}_l, \hat{\theta}_u)$ então o intervalo correpondente para $\phi = g(\theta)$ é

$(\hat{\phi},\hat{\phi}_u) = g(\hat{\theta}_l, g\hat{\theta}u)$

Observe que dos Teoremas 1 e 2 temos duas aproximações para obter os intervalos de confiança, aquelas baseadas no Teorema 1 fornecem intervalos $(\sim \theta_l, \sim \theta_u)$ ao passo que para o parâmetro θ. Os intervalos de confiança baseados no Teorema 1 são simples de contribuir, mas não possuem a importante propriedade de invariância uma vez que se $\theta = g(\theta)$ então

$$\{g(\sim \theta_l), g(\sim \theta_u)\} = \left\{ g(\hat{\theta} - Z_{\alpha/2}[I_E(\theta)]^{-}\frac{1}{2}), g(\hat{\theta} + Z_{\alpha/2}[I_E(\theta)]^{-}\frac{1}{2}) \right\} \neq$$
$$\left\{ g(\hat{\theta}) - Z_{\alpha/2}[g'(\theta)][I_E(\theta)]^{-}\frac{1}{2}, g(\hat{\theta}) + Z_{\alpha/2}[g'(\theta)][I_E(\theta)]^{-}\frac{1}{2} \right\} = \{\sim \phi_l, \sim \phi_u\}$$

A menos linear que g seja linear. Intervalos de confiança construidos com o uso do Teorema 1 são intervalos de confiança de um padrão baseados em verossimilhança, mas eles não são baseados na log-verossimilhança correspondendo a aproximação quadrática para a função exata de log-verossimilhança. Entretanto podemos ver o quão bom será um intervalo de confiança $(\tilde{\theta}_l \tilde{\theta}_u)$ aproximando o máximo para estimar o intervalo $(\hat{\theta}_l, \hat{\theta}_u)$ vendo o quão próximo a log-verossimilhança quadrática é do θ. Mesmo que se a aproximação for boa deve ser lembrado que se o interesse for em algum outro parâmetro $\phi = g(\theta)$, então cuidado,as considerações precisam ser usadas para obter o intervalo de confiança para ϕ. A melhor aprocimação é usar intervalos baseados na verossimilhança $(\hat{\phi}_l, \hat{\phi}_u) = (g(\hat{\theta}_l), g(\hat{\theta}_l))$, mas mesmo que $\hat{\theta}_l$ e $\hat{\theta}_u$ não possam ser precisamente encontrados um aproximação pode ser requerida. Examinando o log-verossimilhança para ϕ em torno de EMV isso pode ser visto se $l(\phi)$ for a função mais viezada (menos quadrática) do que $l(\theta)$.

- Se $l(\phi)$ é menos viezado, a melhor aproximação é usar o intervalo aproximado $(\tilde{\theta}_l \tilde{\theta}_u)$.
- Se $l(\theta)$ é o menos viezado, a melhor aproximação é usar o intervalo aproximado $(g(\tilde{\theta}_l), g(\tilde{\theta}_u))$.