MAT-206: Intervalos y regiones de confianza

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

El objetivo de esta sección es abordar el problema $\boldsymbol{\theta} \in C$, donde $C \subseteq \Theta$, $C = C(\boldsymbol{X})$ es un conjunto determinado por los datos observados $\boldsymbol{X} = \boldsymbol{x}$.[1]

Definición 1:

Una estimación intervalar de un parámetro real-valuado θ es cualquier par de funciones $L(x_1, \dots, x_n)$ y $U(x_1, \dots, x_n)$ que satisfacen

$$L(\boldsymbol{x}) \leq U(\boldsymbol{x}), \qquad \forall\, \boldsymbol{x} \in \mathcal{X}.$$

Para $\boldsymbol{X} = \boldsymbol{x}$ tenemos $L(\boldsymbol{x}) \leq \theta \leq U(\boldsymbol{x})$, mientras que $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ es un intervalo aleatorio.

[1] Si θ es real-valuado, entonces C corresponde a un intervalo.

Ejemplo:

Considere X_1, X_2, X_3, X_4 una muestra aleatoria desde $\mathsf{N}(\mu, 1)$. Un estimador intervalar de μ es $[\overline{X} - 1, \overline{X} + 1]$, es decir

$$\mu \in [\overline{X} - 1, \overline{X} + 1]$$

Note que $\overline{X} \sim \mathsf{N}(\mu, 1/4)$, pero

$$\mathsf{P}(\overline{X} = \mu) = 0.$$

Mientras que,

$$\begin{aligned}
\mathsf{P}(\mu \in [\overline{X} - 1, \overline{X} + 1]) &= \mathsf{P}(\overline{X} - 1 \leq \mu \leq \overline{X} + 1) = \mathsf{P}(-1 \leq \mu - \overline{X} \leq 1) \\
&= \mathsf{P}(-1 \leq \overline{X} - \mu \leq 1) = \mathsf{P}\Big(- \frac{1}{\sqrt{1/4}} \leq \frac{\overline{X} - \mu}{\sqrt{1/4}} \leq \frac{1}{\sqrt{1/4}}\Big) \\
&= \mathsf{P}\Big(- 2 \leq \frac{\overline{X} - \mu}{\sqrt{1/4}} \leq 2\Big) = \mathsf{P}(-2 \leq Z \leq 2) = 0.9544
\end{aligned}$$

pues $Z = (\overline{X} - \mu)/\sqrt{1/4} \sim \mathsf{N}(0, 1)$.

Interpretación:

Tenemos un 95% de chances de cubrir el parámetro verdadero (desconocido) con nuestro estimador intervalar.

Observación:

En este contexto $\mathsf{P}_\theta(\theta \in [L(\boldsymbol{x}), U(\boldsymbol{x})])$ se denomina probabilidad de cobertura

Definición 2:

El coeficiente de confianza de $[L(\boldsymbol{x}), U(\boldsymbol{x})]$ es el ínfimo de las probabilidades de cobertura

$$\inf_\theta \; \mathsf{P}_\theta(\theta \in [L(\boldsymbol{x}), U(\boldsymbol{x})])$$

Observación:

Estimadores intervalares en conjunto con una medida de confianza (coeficiente de confianza) son conocidos como intervalos de confianza.

Definición 3:

Una variable aleatoria $Q(\boldsymbol{X};\boldsymbol{\theta}) = Q(X_1,\dots,X_n;\theta)$ es una cantidad pivotal o pivote si la distribución de $Q(\boldsymbol{X};\theta)$ no depende de θ. Esto es, si $\boldsymbol{X} \sim F(\boldsymbol{x};\theta)$, entonces $Q(\boldsymbol{X};\theta)$ tiene la misma distribución para todo valor de θ.

Observación:

La técnica confía en la habilidad de hallar un pivote y un conjunto A tal que el conjunto $\{\theta : Q(\boldsymbol{X};\theta) \in A\}$ sea una estimación intervalar para θ.

Ejemplo:

Si $X_1, \ldots, X_n$ es una muestra aleatoria de tamaño n desde $\mathsf{N}(\mu, \sigma^2)$, entonces

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathsf{N}(0, 1),$$

y por tanto es un pivote para μ (siempre que σ^2 sea conocido). Para cualquier constante a sigue que:

$$\begin{aligned}
\mathsf{P}\Big(-a \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq a\Big) &= \mathsf{P}\Big(-a\frac{\sigma}{\sqrt{n}} \leq \overline{X} - \mu \leq a\frac{\sigma}{\sqrt{n}}\Big) \\
&= \mathsf{P}\Big(\overline{X} - a\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + a\frac{\sigma}{\sqrt{n}}\Big),
\end{aligned}$$

es decir obtenemos el intervalo de confianza

$$\Big[\overline{X} - a\frac{\sigma}{\sqrt{n}}; \overline{X} + a\frac{\sigma}{\sqrt{n}}\Big].$$

Podemos escribir también,

$$\left\{\mu : \overline{X} - a\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + a\frac{\sigma}{\sqrt{n}}\right\}.$$

Además suponga que $a = z_{1-\alpha/2}$ para un valor de α dado. Entonces, es fácil notar que

$$\mathsf{P}\left(\mu \in \left[\overline{X} - z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}; \overline{X} + z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]\right) = 1 - \alpha,$$

corresponde a un intervalo de confianza del $100(1-\alpha)\%$ para μ.

Observación:

Note que este intervalo de confianza es simétrico.

Ejemplo:

Si $X_1, \ldots, X_n$ es una muestra aleatoria de tamaño n desde $\mathsf{N}(\mu, \sigma^2)$ con σ^2 desconocido. Para este caso, podemos usar el pivote

$$T = \frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t(n-1),$$

es decir,

$$\mathsf{P}(-a \leq T \leq a) = \mathsf{P}\Big(- a \leq \frac{\overline{X} - \mu}{s/\sqrt{n}} \leq a \Big)$$

que lleva al intervalo de confianza

$$\Big\{\mu : \overline{X} - t_{1-\alpha/2}(n-1)\frac{s}{\sqrt{n}} \leq \mu \leq \overline{X} + t_{1-\alpha/2}(n-1)\frac{s}{\sqrt{n}}\Big\}.$$

Ejemplo:

Considere ahora,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

que es cantidad pivotal y elija a y b, satisfaciendo que

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

desde donde obtenemos

$$\left\{\sigma^2 : \frac{(n-1)S^2}{b} \leq \sigma^2 \leq \frac{(n-1)S^2}{a}\right\}.$$

Las elecciones de a y b que producen el intervalo con el coeficiente de confianza requerido son $a = \chi^2_{1-\alpha/2}(n-1)$ y $b = \chi^2_{\alpha/2}(n-1)$.

Definición 4 (Intervalo de confianza asintótico):

Considere $\mathsf{SE} = \sqrt{\mathsf{var}(\widehat{\theta}_n)}$. Entonces $\widehat{\mathsf{SE}} = \sqrt{1/\mathcal{F}_n(\widehat{\theta}_n)}$, luego un intervalo del confianza asintótico del $100(1-\alpha)\%$ para θ es dado por:[2]

$$IC_n(\theta) = [\widehat{\theta}_n - z_{1-\alpha/2}\widehat{\mathsf{SE}}, \widehat{\theta}_n + z_{1-\alpha/2}\widehat{SE}].$$

Este procedimiento está basado en la "cantidad pivotal"

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathsf{D}} \mathsf{N}_1(0, \mathcal{F}_1^{-1}(\theta)),$$

es decir,

$$\frac{\widehat{\theta}_n - \theta}{\sqrt{\mathsf{var}(\widehat{\theta}_n)}} \xrightarrow{\mathsf{D}} \mathsf{N}_1(0, 1).$$

[2]En efecto, $\mathsf{P}_\theta(\theta \in IC_n(\theta)) \to 1 - \alpha$ para $n \to \infty$.

Ejemplo:

Sea $X_1,\dots,X_n$ muestra aleatoria desde Ber(p). Sabemos que el MLE de p es $\widehat{p}_n = \frac{1}{n}\sum_{i=1}^n x_i$, y

$$\log f(x;p) = x\log p + (1-x)\log(1-p),$$

así

$$U(x;p) = \frac{x}{p} - \frac{1-x}{1-p}, \qquad U'(x;p) = \frac{x}{p^2} + \frac{1-x}{(1-p)^2}.$$

De este modo,

$$\mathcal{F}_1(p) = \mathsf{E}\{-U'(X;p)\} = \frac{p}{p^2} + \frac{1-p}{(1-p)^2} = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)},$$

de ahí que

$$\widehat{\mathsf{SE}} = \frac{1}{\sqrt{\mathcal{F}_n(\widehat{p}_n)}} = \frac{1}{\sqrt{n\mathcal{F}_1(\widehat{p}_n)}} = \sqrt{\frac{\widehat{p}_n(1-\widehat{p}_n)}{n}},$$

luego, un intervalo de confianza del $100(1-\alpha)\%$ para p es dado por

$$\widehat{p}_n \mp z_{1-\alpha/2}\sqrt{\frac{\widehat{p}_n(1-\widehat{p}_n)}{n}}.$$

Observación:

Considere $\lambda = g(\theta)$. Sabemos que el estimador ML de λ es dado por $\widehat{\lambda}_n = g(\widehat{\theta}_n)$. Además, usando el método Delta, sigue que

$$\frac{\widehat{\lambda}_n - \lambda}{\widehat{\mathsf{SE}}(\widehat{\lambda}_n)} \xrightarrow{\mathsf{D}} \mathsf{N}(0,1),$$

donde

$$\widehat{\mathsf{SE}}(\widehat{\lambda}_n) = |g'(\widehat{\theta}_n)|\,\widehat{\mathsf{SE}}(\widehat{\theta}_n).$$

Lo que lleva al intervalo de confianza asintótico

$$IC_n(\lambda) = [\widehat{\lambda}_n - z_{1-\alpha/2}\,\widehat{\mathsf{SE}}(\widehat{\lambda}_n), \widehat{\lambda}_n + z_{1-\alpha/2}\,\widehat{\mathsf{SE}}(\widehat{\lambda}_n)].$$

Ejemplo:

Considere $X_1, \ldots, X_n$ variables aleatorias IID desde una FE 1-paramétrica, y sea $\phi = \eta(\theta)$, $\gamma(\phi) = b(\theta)$. De este modo,

$$f(x; \phi) = \exp[\phi T(x) - \gamma(\phi)] h(x),$$

es decir, la log-verosimilitud para una única observación es dada por

$$\log f(x; \phi) = \phi T(x) - \gamma(\phi) + \log h(x).$$

Lo que lleva a,

$$U(x; \phi) = T(x) - \gamma'(\phi), \qquad U'(x; \phi) = -\gamma''(\phi),$$

de ahí que, la información de Fisher es dada por

$$\mathcal{F}_1(\phi) = \mathsf{E}\{-U'(X; \phi)\} = \gamma''(\phi) = \mathsf{var}(T(X)).$$

Es decir, el error estándar adopta la forma

$$\widehat{\mathsf{SE}} = \frac{1}{\sqrt{\mathcal{F}_n(\widehat{\phi}_n)}} = \frac{1}{\sqrt{n\mathcal{F}_1(\widehat{\phi}_n)}} = \frac{1}{\sqrt{n\gamma''(\widehat{\phi}_n)}},$$

donde $\widehat{\phi}_n$ denota el MLE de ϕ. Finalmente, un intervalo de confianza del $100(1-\alpha)\%$ para ϕ es dado por

$$IC_n(\phi) = \left[\widehat{\phi}_n \mp z_{1-\alpha/2}\,\frac{1}{\sqrt{n\gamma''(\widehat{\phi}_n)}}\right].$$

Evidentemente, también podemos considerar un intervalo de confianza asintótico para $\theta = g(\phi) = \eta^{-1}(\phi)$ usando el método Delta.

Resultado 1:

Sea $\{\boldsymbol{T}_n\}$ una secuencia de vectores aleatorios k-dimensionales tal que

$$\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{\theta}) \xrightarrow{\mathsf{D}} \mathsf{N}_k(\boldsymbol{0}, \boldsymbol{\Sigma}),$$

y sea $\{\boldsymbol{A}_n\}$ una secuencia de matrices aleatorias tal que $\boldsymbol{A}_n \xrightarrow{\mathsf{P}} \boldsymbol{A}$, donde $\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A} = \boldsymbol{A}$. Entonces

$$Q_n = n(\boldsymbol{T}_n - \boldsymbol{\theta})^\top \boldsymbol{A}_n(\boldsymbol{T}_n - \boldsymbol{\theta}) \xrightarrow{\mathsf{D}} \chi^2(k).$$

Demostración:

Ver detalles en Sen y Singer (1993), página 137.

Basado en la normalidad asintótica de los MLEs

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathsf{D}} \mathsf{N}_k(\mathbf{0}, \boldsymbol{\mathcal{F}}_1^{-1}(\boldsymbol{\theta})),$$

y notando que $\boldsymbol{\mathcal{F}}_1(\widehat{\boldsymbol{\theta}}_n) \xrightarrow{\mathsf{P}} \boldsymbol{\mathcal{F}}_1(\boldsymbol{\theta})$, tenemos

$$n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^\top \boldsymbol{\mathcal{F}}_1(\widehat{\boldsymbol{\theta}}_n)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathsf{D}} \chi^2(k).$$

Es decir, podemos construir una región de confianza del $100(1-\alpha)\%$ para $\boldsymbol{\theta}$ como:

$$RC_n(\boldsymbol{\theta}) = \{\boldsymbol{\theta} : n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^\top \boldsymbol{\mathcal{F}}_1(\widehat{\boldsymbol{\theta}}_n)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \leq \chi^2_{1-\alpha}(k)\},$$

donde $\chi^2_{1-\alpha}(k)$ denota un valor cuantil $1-\alpha$ de la distribución chi-cuadrado con k grados de libertad.

Ejemplo:

Considere $X_1, \dots, X_n$ una muestra aleatoria desde $\mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Es fácil notar que

$$\sqrt{n}(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}) \xrightarrow{\mathsf{D}} \mathsf{N}_p(\boldsymbol{0}, \boldsymbol{\Sigma}),$$

y

$$\boldsymbol{S}_n = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n)(\boldsymbol{X}_i - \overline{\boldsymbol{X}}_n)^\top \xrightarrow{\mathsf{P}} \boldsymbol{\Sigma}.$$

De este modo, la estadística T^2 de Hotelling, satisface que

$$T_n^2 = n(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu})^\top \boldsymbol{S}_n^{-1} (\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}) \xrightarrow{\mathsf{D}} \chi^2(p).$$

Luego, una región de confianza asintótica del $100(1-\alpha)\%$ para $\boldsymbol{\mu}$ es dada por:

$$RC_n(\boldsymbol{\mu}) = \{\boldsymbol{\mu} : n(\overline{\boldsymbol{X}}_n - \boldsymbol{\mu})^\top \boldsymbol{S}_n^{-1} (\overline{\boldsymbol{X}}_n - \boldsymbol{\mu}) \leq \chi^2_{1-\alpha}(p)\}.$$

Considere la siguiente expansión de Taylor,

$$\ell_n(\boldsymbol{\theta}) \overset{\mathsf{a}}{=} \ell_n(\widehat{\boldsymbol{\theta}}_n) + \Big(\frac{\partial \ell_n(\widehat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}\Big)^\top (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n) + \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)^\top \frac{\partial^2 \ell_n(\widehat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n).$$

Usando la condición de primer orden, sigue que

$$\ell_n(\widehat{\boldsymbol{\theta}}_n) - \ell_n(\boldsymbol{\theta}) \overset{\mathsf{a}}{=} \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)^\top \Big\{ - \frac{\partial^2 \ell_n(\widehat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} \Big\} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n).$$

Asumiendo independencia, tenemos que

$$\begin{aligned} \frac{1}{n}\Big\{ - \frac{\partial^2 \ell_n(\widehat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} \Big\} &= \frac{1}{n} \sum_{i=1}^n - \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \\ &\xrightarrow{\mathsf{P}} \mathsf{E}\Big\{ - \frac{\partial^2 \ell_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} \Big\} = \boldsymbol{\mathcal{F}}_1(\boldsymbol{\theta}) \end{aligned}$$

Por los razonamientos anteriores,

$$\ell_n(\widehat{\boldsymbol{\theta}}_n) - \ell_n(\boldsymbol{\theta}) \overset{\mathsf{a}}{=} \frac{n}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n)^\top \boldsymbol{\mathcal{F}}_1(\boldsymbol{\theta})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_n),$$

es decir,

$$2\{\ell_n(\widehat{\boldsymbol{\theta}}_n) - \ell_n(\boldsymbol{\theta})\} \xrightarrow{\mathsf{D}} \chi^2(k),$$

lo que lleva a la región de confianza asintótica:

$$RC_n^*(\boldsymbol{\theta}) = \{\boldsymbol{\theta} : 2(\ell_n(\widehat{\boldsymbol{\theta}}_n) - \ell_n(\boldsymbol{\theta})) \leq \chi^2_{1-\alpha}(k)\}.$$

Regiones de confianza

En el contexto de extremum estimators, este decir $\widehat{\boldsymbol{\theta}}_n$ obtenido como solución del problema $\max\limits_{\theta} Q_n(\boldsymbol{\theta})$,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathsf{D}} \mathsf{N}_k(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}).$$

De este modo, obtenemos la región de confianza

$$RC_n^{\star}(\boldsymbol{\theta}) = \{\boldsymbol{\theta} : n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^\top \boldsymbol{A}\boldsymbol{B}^{-1}\boldsymbol{A}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \leq \chi^2_{1-\alpha}(k)\}.$$

Mientras que para funciones de inferencia del tipo $\boldsymbol{g}(\boldsymbol{\theta}) = \mathbf{0}$, sabemos que

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathsf{D}} \mathsf{N}_k(\mathbf{0}, \boldsymbol{G}^{-1}(\boldsymbol{\theta})),$$

esto lleva a la región de confianza

$$RC_n^{\dagger}(\boldsymbol{\theta}) = \{\boldsymbol{\theta} : n(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^\top \boldsymbol{S}^\top(\widehat{\boldsymbol{\theta}}_n)\boldsymbol{V}^{-1}(\widehat{\boldsymbol{\theta}}_n)\boldsymbol{S}(\widehat{\boldsymbol{\theta}}_n)(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \leq \chi^2_{1-\alpha}(k)\}.$$