MAT-206: Propiedades de estimadores puntuales

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Definición 1 (Error Cuadrático Medio):

Sea $\mathcal{P} = \{\mathsf{P}_\theta : \theta \in \Theta\}$ un modelo estadístico para la variable X y sea T un estimador para $\gamma = g(\theta)$. El error cuadrático medio (MSE) de T es dado por

$$\mathsf{MSE}(T, \theta) = \mathsf{E}_\theta\{(T - g(\theta))^2\}.$$

Observación:

Es fácil notar que

$$\mathsf{MSE}(T, \theta) = \{\mathsf{E}_\theta(T) - g(\theta)\}^2 + \mathsf{var}_\theta(T).$$

Definición 2 (Sesgo):

El sesgo de un estimador T es definido como:

$$\mathsf{bias}(T,\theta) = \mathsf{E}_\theta(T) - g(\theta).$$

De este modo, usando la definición anterior, tenemos que:

$$\mathsf{MSE}(T,\theta) = \{\mathsf{bias}(T,\theta)\}^2 + \mathsf{var}_\theta(T).$$

Definición 3 (Insesgamiento):

Un estimador T para $\gamma = g(\theta)$ se dice insesgado, si

$$\mathsf{E}_\theta(T) = g(\theta), \qquad \forall\, \theta \in \Theta,$$

o equivalentemente,

$$\mathsf{bias}(T,\theta) = 0, \qquad \forall\, \theta \in \Theta.$$

Ejemplo:

Sea $X_1,\dots,X_n$ variables aleatorias IID con varianza finita. Suponga que $\gamma=\sigma^2$ es el parámetro de interés. Sabemos que el estimador MM es dado por:

$$\widehat{\sigma}^2_{\mathsf{MM}} = \frac{1}{n}\sum_{i=1}^n (X_i-\overline{X})^2.$$

Además,

$$\sum_{i=1}^n (X_i-\overline{X})^2 = \sum_{i=1}^n X_i^2 - n\overline{X}^2 = \boldsymbol{X}^\top\Big(\boldsymbol{I}-\frac{1}{n}\boldsymbol{1}\boldsymbol{1}^\top\Big)\boldsymbol{X} = \boldsymbol{X}^\top\boldsymbol{C}\boldsymbol{X},$$

donde $\boldsymbol{C}=\boldsymbol{I}-\frac{1}{n}\boldsymbol{1}\boldsymbol{1}^\top$. Como $X_1,\dots,X_n$ son IID considere

$$\mathsf{E}(\boldsymbol{X})=\mu\boldsymbol{1}_n, \qquad \mathsf{Cov}(\boldsymbol{X})=\sigma^2\boldsymbol{I}_n.$$

De este modo,[1]

$$\mathsf{E}\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right\}=\frac{1}{n}\,\mathsf{E}(\boldsymbol{X}^\top\boldsymbol{C}\boldsymbol{X})=\frac{1}{n}\{\sigma^2\operatorname{tr}\boldsymbol{C}+\mu^2\boldsymbol{1}^\top\boldsymbol{C}\boldsymbol{1}\}.$$

Como $\operatorname{tr}\boldsymbol{C}=\operatorname{tr}\boldsymbol{I}-\frac{1}{n}\operatorname{tr}\boldsymbol{1}\boldsymbol{1}^\top=n-1$ y $\boldsymbol{C}\boldsymbol{1}=\boldsymbol{0}$, sigue que

$$\mathsf{E}(\widehat{\sigma}^2_{\mathsf{MM}})=\mathsf{E}\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right\}=\Big(\frac{n-1}{n}\Big)\sigma^2.$$

Es decir, $\widehat{\sigma}^2_{\mathsf{MM}}$ es un estimador sesgado, y

$$\mathsf{bias}(\widehat{\sigma}^2_{\mathsf{MM}},\sigma^2)=\mathsf{E}(\widehat{\sigma}^2_{\mathsf{MM}})-\sigma^2=\frac{n-1}{n}\sigma^2-\sigma^2=-\frac{\sigma^2}{n}.$$

Aunque

$$\lim_{n\to\infty}\mathsf{bias}(\widehat{\sigma}^2_{\mathsf{MM}},\sigma^2)=0.$$

[1] Si $\mathsf{E}(\boldsymbol{X})=\boldsymbol{\delta}$ y $\mathsf{Cov}(\boldsymbol{X})=\boldsymbol{\Sigma}$, tenemos que $\mathsf{E}(\boldsymbol{X}^\top\boldsymbol{A}\boldsymbol{X})=\operatorname{tr}\boldsymbol{A}\boldsymbol{\Sigma}+\boldsymbol{\delta}^\top\boldsymbol{A}\boldsymbol{\delta}$.

El "factor de corrección" $\frac{n}{n-1}$, lleva al estimador

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

Suponga adicionalmente que $X_i \overset{\text{ind}}{\sim} \mathsf{N}(\mu, \sigma^2)$, para $i = 1, \dots, n$. Entonces,

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

y sabemos que

$$\mathsf{E}(U) = n - 1, \qquad \text{y} \qquad \mathsf{var}(U) = 2(n-1).$$

Podemos escribir

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

luego, tenemos

$$\mathsf{var}(\widehat{\sigma}^2_{\mathsf{MM}}) = \mathsf{var}\left(\frac{\sigma^2}{n}\, U\right) = \frac{\sigma^4}{n^2}\, \mathsf{var}(U).$$

De este modo,

$$\mathsf{MSE}(\widehat{\sigma}^2_{\mathsf{MM}}, \sigma^2) = \Big(-\frac{\sigma^2}{n}\Big)^2 + \frac{\sigma^4}{n^2}\,\mathsf{var}(U) = \frac{\sigma^4}{n^2} + \frac{\sigma^4}{n^2}2(n-1) = \sigma^4\Big(\frac{2n-1}{n^2}\Big).$$

Mientras que

$$U = \frac{(n-1)S^2}{\sigma^2} \quad \Longrightarrow \quad S^2 = \frac{\sigma^2}{n-1}\,U.$$

Así,

$$\mathsf{MSE}(S^2, \sigma^2) = 0 + \mathsf{var}(S^2),$$

es decir,

$$\mathsf{MSE}(S^2, \sigma^2) = 0 + \frac{\sigma^4}{(n-1)^2}\,\mathsf{var}(U) = \frac{2(n-1)\sigma^4}{(n-1)^2} = \frac{2\sigma^4}{n-1}.$$

Finalmente,

$$\mathsf{MSE}(S^2, \sigma^2) > \mathsf{MSE}(\widehat{\sigma}^2_{\mathsf{MM}}, \sigma^2), \qquad n > 1.$$

Suponga $\Theta \subseteq \mathbb{R}^k$ y que el parámetro de interés $\boldsymbol{\gamma}$ es un vector k-dimensional, esto es, $\boldsymbol{g} : \Theta \to \Gamma \subseteq \mathbb{R}^m$. Entonces, $\boldsymbol{T}$ se dice un estimador insesgado, si

$$\mathsf{E}(\boldsymbol{T}) = \boldsymbol{g}(\boldsymbol{\theta}), \qquad \forall\, \boldsymbol{\theta} \in \Theta.$$

La extensión del error cuadrático medio para el caso multiparamétrico adopta la forma

$$\begin{aligned} \mathsf{MSE}(\boldsymbol{T}, \boldsymbol{\theta}) &= \mathsf{E}_\theta\{(\boldsymbol{T} - \boldsymbol{g}(\boldsymbol{\theta}))(\boldsymbol{T} - \boldsymbol{g}(\boldsymbol{\theta}))^\top\} \\ &= \mathsf{Cov}_\theta(\boldsymbol{T}) + \{\mathsf{E}_\theta(\boldsymbol{T}) - \boldsymbol{g}(\boldsymbol{\theta})\}\{\mathsf{E}_\theta(\boldsymbol{T}) - \boldsymbol{g}(\boldsymbol{\theta})\}^\top \\ &= \mathsf{Cov}_\theta(\boldsymbol{T}) + \{\mathsf{bias}(\boldsymbol{T}, \boldsymbol{\theta})\}\{\mathsf{bias}(\boldsymbol{T}, \boldsymbol{\theta})\}^\top. \end{aligned}$$

Definición 4:

Sean $\boldsymbol{T}$ y $\boldsymbol{T}_*$ dos estimadores para $\boldsymbol{\gamma}$. Decimos que $\boldsymbol{T}_*$ tiene error cuadrático medio más pequeño que $\boldsymbol{T}$ si

$$\boldsymbol{u}^\top \big(\mathsf{MSE}(\boldsymbol{T}_*, \boldsymbol{\theta}) - \mathsf{MSE}(\boldsymbol{T}, \boldsymbol{\theta})\big)\boldsymbol{u} \leq 0, \qquad \forall\, \boldsymbol{u} \in \mathbb{R}^m,$$

y escribimos

$$\mathsf{MSE}(\boldsymbol{T}_*, \boldsymbol{\theta}) \leq \mathsf{MSE}(\boldsymbol{T}, \boldsymbol{\theta}).$$

Observación:

En general, evaluar la condición dada por la definición anterior puede ser difícil.

Esto ha motivado la introducción de algunos criterios más simples para comparar entre diferentes estimadores. En efecto, decimos que $\boldsymbol{T}_*$ es T-óptimo, si

$$\operatorname{tr} \mathsf{MSE}(\boldsymbol{T}_*, \boldsymbol{\theta}) \leq \operatorname{tr} \mathsf{MSE}(\boldsymbol{T}, \boldsymbol{\theta}), \tag{1}$$

mientras que $\boldsymbol{T}_*$ se dice D-óptimo, si satisface

$$\det \mathsf{MSE}(\boldsymbol{T}_*, \boldsymbol{\theta}) \leq \det \mathsf{MSE}(\boldsymbol{T}, \boldsymbol{\theta}). \tag{2}$$

El criterio dado en la Definición 4 también es conocido como M-optimalidad.

Debemos destacar que en ocasiones el error cuadrático medio es definido como

$$\begin{aligned}\mathsf{E}_\theta[\|\boldsymbol{T}-\boldsymbol{g}(\boldsymbol{\theta})\|^2] &= \|\,\mathsf{E}_\theta(\boldsymbol{T})-\boldsymbol{g}(\boldsymbol{\theta})\|^2+\sum_{j=1}^m \mathsf{var}(T_j) \\ &= \|\,\mathsf{bias}(\boldsymbol{T},\boldsymbol{\theta})\|^2+\operatorname{tr}\mathsf{Cov}_\theta(\boldsymbol{T}),\end{aligned}$$

que corresponde al criterio de T-optimalidad.

Observación:

Lamentablemente, es muy poco frecuente encontrar un estimador que siempre (es decir, para todo $\boldsymbol{\theta}\in\Theta$) sea mejor.

Definición 5 (Mejor estimador insesgado):

Para el modelo estadístico $\mathcal{P} = \{\mathsf{P}_\theta : \theta \in \Theta\}$, un estimador insesgado T_* para $\gamma = g(\theta) \in \mathbb{R}$ se dice el mejor estimador insesgado (BUE), si para cualquier otro estimador insesgado

$$\mathsf{var}_\theta(T_*) \leq \mathsf{var}_\theta(T), \qquad \forall\, \theta \in \Theta.$$

Resultado 1 (Cota de Cramér-Rao):

Suponga que se satisfacen las condiciones A1-A4, que la información de Fisher es tal que $0 < \mathcal{F}_X(\theta) < \infty$ y sea $\gamma = g(\theta)$, donde g es continua y diferenciable con $g' \neq 0$. Si T es estimador insesgado para γ, entonces

$$\mathsf{var}_\theta(T) \geq \frac{\{g'(\theta)\}^2}{\mathcal{F}_X(\theta)}, \qquad \forall\, \theta \in \Theta.$$

Demostración:

Considere

$$\mathsf{Cov}_\theta(T(\boldsymbol{X}), U(\theta;\boldsymbol{X})) = \mathsf{E}_\theta(T(\boldsymbol{X})\,U(\theta;\boldsymbol{X})) = \int_A T(\boldsymbol{x})\frac{\partial}{\partial\theta}\log f(\boldsymbol{x};\theta)\,f(\boldsymbol{x};\theta)\,\mathrm{d}\boldsymbol{x}$$
$$= \int_A T(\boldsymbol{x})\frac{\partial}{\partial\theta}\,f(\boldsymbol{x};\theta)\,\mathrm{d}\boldsymbol{x}.$$

Como T es regular y es insesgada, sigue que

$$\mathsf{Cov}_\theta(T(\boldsymbol{X}), U(\theta;\boldsymbol{X})) = \frac{\partial}{\partial\theta}\int_A T(\boldsymbol{x})\,f(\boldsymbol{x};\theta)\,\mathrm{d}\boldsymbol{x} = \frac{\partial}{\partial\theta}\,\mathsf{E}_\theta(T(\boldsymbol{X})) = g'(\theta).$$

Haciendo $h(\theta) = g'(\theta)/\mathcal{F}_X(\theta)$, tenemos

$$\begin{aligned} 0 &\leq \mathsf{var}_\theta(T(\boldsymbol{X}) - h(\theta)U(\theta;\boldsymbol{X})) \\ &= \mathsf{var}_\theta(T(\boldsymbol{X})) + h^2(\theta)\,\mathsf{var}_\theta(U(\theta;\boldsymbol{X})) - 2h(\theta)\,\mathsf{Cov}_\theta(T(\boldsymbol{X}), U(\theta;\boldsymbol{X})) \\ &= \mathsf{var}_\theta(T(\boldsymbol{X})) + h^2(\theta)\mathcal{F}_X(\theta) - 2h(\theta)g'(\theta), \end{aligned}$$

es decir,

$$0 \leq \mathsf{var}_\theta(T(\boldsymbol{X})) - \{g'(\theta)\}^2/\mathcal{F}_X(\theta).$$

Considere el caso en que $\Theta \subset \mathbb{R}^k$ y que el parámetro de interés es $\boldsymbol{\gamma} \in \mathbb{R}^m$ con $\boldsymbol{\gamma} = \boldsymbol{g}(\boldsymbol{\theta})$. Sea $\boldsymbol{T}$ y $\boldsymbol{T}_*$ dos estimadores insesgados de $\boldsymbol{\gamma}$. Decimos que $\boldsymbol{T}_*$ tiene covarianza más pequeña que $\boldsymbol{T}$, si

$$\boldsymbol{u}^\top(\mathsf{Cov}_\theta(\boldsymbol{T}_*) - \mathsf{Cov}_\theta(\boldsymbol{T}))\boldsymbol{u} \leq 0, \qquad \forall\, \boldsymbol{u} \in \mathbb{R}^m,$$

en cuyo caso escribimos

$$\mathsf{Cov}_\theta(\boldsymbol{T}_*) \leq \mathsf{Cov}_\theta(\boldsymbol{T}).$$

Suponga las condiciones A1-A4 y que la matriz de información de Fisher es no singular. Entonces la cota de Cramér-Rao asume la forma:

$$\mathsf{Cov}_\theta(\boldsymbol{T}) \geq \Big(\frac{\partial \boldsymbol{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\Big)\boldsymbol{\mathcal{F}}_X^{-1}(\boldsymbol{\theta})\Big(\frac{\partial \boldsymbol{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top}\Big)^\top, \qquad \forall\, \boldsymbol{\theta} \in \Theta.$$

Ejemplo:

Suponga que tenemos $X_1, \dots, X_n$ variables aleatorias independientes desde $\mathsf{N}(\mu, \sigma^2)$ con parámetro de interés $\boldsymbol{\theta} = (\mu, \sigma^2)^\top$. La matrix de información de Fisher está dada por

$$\boldsymbol{\mathcal{F}}_X(\boldsymbol{\theta}) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix},$$

cuya matriz inversa corresponde a la cota de Cramér-Rao. Sabemos que los estimadores insesgados $\overline{X}$ y S^2 son independientes y que

$$\mathsf{var}_\theta(\overline{X}) = \frac{\sigma^2}{n}, \qquad \mathsf{var}_\theta(S^2) = \frac{2\sigma^4}{n-1}.$$

Por tanto no se alcanza la cota inferior para la varianza.

Definición 6 (Eficiencia):

La eficiencia de un estimador insesgado T es definida como la razón de su varianza y la cota de Cramér-Rao. Esto es,

$$\mathsf{EFF}(T,\theta) = \frac{\{g'(\theta)\}^2/\mathcal{F}_X(\theta)}{\mathsf{var}_\theta(T)}.$$

Un estimador que alcanza la cota de Cramér-Rao se dice un estimador eficiente. Más aún, un estimador eficiente es BUE.

Ejemplo:

Considere $X_1, \ldots, X_n$ variables aleatorias IID desde Exp(λ) con $\lambda > 0$. Sea $\gamma = 1/\lambda$ el parámetro de interés. Un estimador insesgado para λ es $\overline{X}$ con varianza $\frac{1}{n\lambda^2}$.

Como la información de Fisher es n/λ^2 y $g'(\lambda) = -1/\lambda^2$, tenemos

$$\frac{\{-1/\lambda^2\}^2}{n/\lambda^2} = \frac{1}{n\lambda^2},$$

es decir, $\overline{X}$ es eficiente.

Resultado 2:

Bajo las condiciones A1 a A4 el estimador de máxima verosimilitud satisface las siguientes propiedades:

(a) El estimador ML depende de los datos via la estadística suficiente.

(b) Si existe un estimador insesgado y eficiente $\widetilde{\boldsymbol{\theta}}$, entonces $\widetilde{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\text{ML}}$.

Demostración:

(a) sigue notando que, por el Teorema de factorización de Fisher-Neyman

$$L(\boldsymbol{\theta}; \boldsymbol{x}) \propto g(\boldsymbol{T}(\boldsymbol{x}), \boldsymbol{\theta}),$$

para $\boldsymbol{T}$ suficiente.

Por simplicidad para la prueba de (b) sólo consideraremos el caso en que θ es real-valuado. Sabemos que $\widetilde{\theta}$ alcanza la cota de Cramér-Rao pues es un estimador eficiente. Así, tenemos que

$$\widetilde{\theta}(\boldsymbol{x}) - \theta = \frac{U(\theta; \boldsymbol{x})}{\mathcal{F}_X(\theta)}, \qquad \forall\, \theta \in \Theta,$$

que es válido en particular para $\theta = \widehat{\theta}_{\mathsf{ML}}$. Es decir,

$$\widetilde{\theta}(\boldsymbol{x}) - \widehat{\theta}_{\mathsf{ML}} = \frac{U(\widehat{\theta}_{\mathsf{ML}}; \boldsymbol{x})}{\mathcal{F}_X(\widehat{\theta}_{\mathsf{ML}})}.$$

Como $\widehat{\theta}_{\mathsf{ML}}$ maximiza $\ell(\theta; \boldsymbol{x})$ y por tanto, $U(\widehat{\theta}_{\mathsf{ML}}; \boldsymbol{x}) = 0$. De este modo,

$$\widetilde{\theta}(\boldsymbol{x}) - \widehat{\theta}_{\mathsf{ML}} = 0.$$