

# **MAT-206: Sesión 9, Propiedades de estimadores puntuales**

**Felipe Osorio**

[fosorios.mat.utfsm.cl](mailto:fosorios.mat.utfsm.cl)

Departamento de Matemática, UTFSM



## Definición 1 (Error Cuadrático Medio):

Sea  $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$  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

$$\text{MSE}(T, \theta) = E_\theta\{(T - g(\theta))^2\}.$$

### *Observación:*

Es fácil notar que

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



## Definición 2 (Sesgo):

El sesgo de un estimador  $T$  es definido como:

$$\text{bias}(T, \theta) = E_{\theta}(T) - g(\theta).$$

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

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

## Definición 3 (Insesgamiento):

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

$$E_{\theta}(T) = g(\theta), \quad \forall \theta \in \Theta,$$

o equivalentemente,

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



### 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:

$$\hat{\sigma}_{\text{MM}}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Además,

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2 = \mathbf{X}^\top \left( \mathbf{I} - \frac{1}{n} \mathbf{1}\mathbf{1}^\top \right) \mathbf{X} = \mathbf{X}^\top \mathbf{C} \mathbf{X},$$

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

$$\mathbb{E}(\mathbf{X}) = \mu \mathbf{1}_n, \quad \text{Cov}(\mathbf{X}) = \sigma^2 \mathbf{I}_n.$$



## Propiedades de estimadores puntuales

De este modo,<sup>1</sup>

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

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

$$E(\hat{\sigma}_{\text{MM}}^2) = E\left(\frac{1}{n}\sum_{i=1}^n(X_i - \bar{X})^2\right) = \left(\frac{n-1}{n}\right)\sigma^2.$$

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

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

Aunque

$$\lim_{n \rightarrow \infty} \operatorname{bias}(\hat{\sigma}_{\text{MM}}^2, \sigma^2) = 0.$$

---

<sup>1</sup>Si  $E(\mathbf{X}) = \boldsymbol{\delta}$  y  $\operatorname{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$ , tenemos que  $E(\mathbf{X}^\top \mathbf{A} \mathbf{X}) = \operatorname{tr} \mathbf{A} \boldsymbol{\Sigma} + \boldsymbol{\delta}^\top \mathbf{A} \boldsymbol{\delta}$ .



## Propiedades de estimadores puntuales

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

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

Suponga adicionalmente que  $X_i \sim N(\mu, \sigma^2)$ , para  $i = 1, \dots, n$ . Entonces,

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

y sabemos que

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

Podemos escribir

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

luego, tenemos

$$\text{var}(\hat{\sigma}_{\text{MM}}^2) = \frac{\sigma^4}{n^2} \text{var}(U).$$



## Propiedades de estimadores puntuales

De este modo,

$$\text{MSE}(\hat{\sigma}_{\text{MM}}^2, \sigma^2) = \left(-\frac{\sigma^2}{n}\right)^2 + \frac{\sigma^4}{n^2} \text{var}(U) = \frac{\sigma^4}{n^2} + \frac{\sigma^4}{n^2} 2(n-1) = \sigma^4 \left(\frac{2n-1}{n^2}\right).$$

Mientras que

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

Así,

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

es decir,

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

Finalmente,

$$\text{MSE}(S^2, \sigma^2) > \text{MSE}(\hat{\sigma}_{\text{MM}}^2, \sigma^2), \quad n > 1.$$



## Propiedades de estimadores puntuales

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

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

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

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





## Definición 4:

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

$$\mathbf{u}^\top (\text{MSE}(T_*, \boldsymbol{\theta}) - \text{MSE}(T, \boldsymbol{\theta})) \mathbf{u} \leq 0, \quad \forall \mathbf{u} \in \mathbb{R}^m,$$

y escribimos

$$\text{MSE}(T_*, \boldsymbol{\theta}) \leq \text{MSE}(T, \boldsymbol{\theta}).$$

## 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  $T_*$  es **T-óptimo**, si

$$\text{tr MSE}(T_*, \theta) \leq \text{tr MSE}(T, \theta), \quad (1)$$

mientras que  $T_*$  se dice **D-óptimo**, si satisface

$$\det \text{MSE}(T_*, \theta) \leq \det \text{MSE}(T, \theta). \quad (2)$$

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

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

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

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

## 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  $\gamma$ , entonces

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



## Demostración:

Considere

$$\begin{aligned}\text{Cov}_\theta(T(\mathbf{X}), U(\theta; \mathbf{X})) &= \mathbb{E}_\theta(T(\mathbf{X}) U(\theta; \mathbf{X})) = \int_A T(\mathbf{x}) \frac{\partial}{\partial \theta} \log f(\mathbf{x}; \theta) f(\mathbf{x}; \theta) d\mathbf{x} \\ &= \int_A T(\mathbf{x}) \frac{\partial}{\partial \theta} f(\mathbf{x}; \theta) d\mathbf{x}.\end{aligned}$$

Como  $T$  es regular y es insesgada, sigue que

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

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

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

es decir,

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



## 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,

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

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



### Ejemplo:

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

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

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

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



## Propiedades de estimadores puntuales

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

$$u^\top (\text{Cov}_\theta(T_*) - \text{Cov}_\theta(T))u \leq 0, \quad \forall u \in \mathbb{R}^m,$$

en cuyo caso escribimos

$$\text{Cov}_\theta(T_*) \leq \text{Cov}_\theta(T).$$

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:

$$\text{Cov}_\theta(T) \geq \left( \frac{\partial g(\theta)}{\partial \theta^\top} \right) \mathcal{F}_X^{-1}(\theta) \left( \frac{\partial g(\theta)}{\partial \theta^\top} \right)^\top, \quad \forall \theta \in \Theta.$$





### Ejemplo:

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

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

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

$$\text{var}_\theta(\bar{X}) = \frac{\sigma^2}{n}, \quad \text{var}_\theta(S^2) = \frac{2\sigma^4}{n-1}.$$

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



## 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  $\tilde{\theta}$ , entonces  $\tilde{\theta} = \hat{\theta}_{ML}$ .

### *Demostración:*

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

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

para  $T$  suficiente.



## 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  $\tilde{\theta}$ , entonces  $\tilde{\theta} = \hat{\theta}_{ML}$ .

## Demostración:

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

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

para  $\mathbf{T}$  suficiente.



## Propiedades de estimadores puntuales

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

$$\tilde{\theta}(\mathbf{x}) - \theta = \frac{U(\theta; \mathbf{x})}{\mathcal{F}_X(\theta)}, \quad \forall \theta \in \Theta,$$

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

$$\tilde{\theta}(\mathbf{x}) - \hat{\theta}_{\text{ML}} = \frac{U(\hat{\theta}_{\text{ML}}; \mathbf{x})}{\mathcal{F}_X(\hat{\theta}_{\text{ML}})}.$$

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

$$\tilde{\theta}(\mathbf{x}) - \hat{\theta}_{\text{ML}} = 0.$$

