

# Estadística

## Estimadores Puntuales: Propiedades de estimadores

Sebastián Court

### 1. Motivación

Consideremos una variable aleatoria  $X$  con ciertas características, como por ejemplo, un parámetro  $\theta$ , y una muestra aleatoria de dicha variable (digamos  $X_1, \dots, X_n$ ). Dado que no conocemos el valor del parámetro  $\theta$ , sería deseable estimarlo considerando una función  $T$  que dependa sólo de los elementos conocidos, es decir, nuestra muestra. Sin embargo, deseamos que el estimador  $T$  de  $\theta$  se comporte “aproximadamente” como  $\theta$ .

A primera vista parece razonable tener un estimador que tenga el menor error posible. Luego si tengo 2 estimadores  $T_1$  y  $T_2$  la elección natural será escoger el estimador, por ejemplo, con la menor diferencia en promedio entre  $\theta$  y el estimador, pero dado dicho promedio podría ser cero, luego nos conviene escoger la diferencia al cuadrado y promediar.

Lo anterior da paso a la siguiente definición:

**Definición 1** Sea  $T$  un estimador de un parámetro desconocido  $\theta$ . Se define el **error cuadrático medio** como el valor esperado del cuadrado de la diferencia entre  $T$  y  $\theta$ .

$$ECM(T; \theta) = \mathbb{E}[(T - \theta)^2]$$

Notemos que lo anterior, lo podemos escribir como:

$$\begin{aligned} ECM(T, \theta) &= \mathbb{E}[(T - \theta)^2] \\ &= \mathbb{E}(T) - 2\theta\mathbb{E}(T) + \theta^2 \\ &= \text{Var}(T) + [\mathbb{E}(T)]^2 - 2\theta\mathbb{E}(T) + \theta^2 \\ &= \text{Var}(T) + [\theta - \mathbb{E}(T)]^2 \end{aligned}$$

El problema es encontrar una función  $T = u(X_1, \dots, X_n)$  que sólo dependa de la muestra y que entregue la “mejor” estimación de  $\theta$ .

Para buscar dicho estimador, se utilizará el llamado error cuadrático medio. Sin embargo, dicho criterio presenta una dificultad crucial, que es que el error cuadrático medio puede variar dependiendo de los valores que tome  $\theta$ . Para ilustrar lo anterior, consideremos el siguiente ejemplo:

Sea  $X_1, \dots, X_n$  una muestra aleatoria de alguna distribución tal que  $\mathbb{E}(X_i) = \mu$  y  $\text{Var}(X_i) = \sigma^2$ ,  $i = 1, \dots, n$ . Considere los siguientes dos estimadores de  $\mu$ :

$$T_1 = \bar{X}$$

y

$$T_2 = \frac{1}{n+1} \sum_{i=1}^n X_i$$

Analizaremos los errores cuadráticos medios. Notemos que  $\mathbb{E}(T_1) = \mu$ , luego

$$ECM(T_1) = \text{Var}(T_1) + (\mathbb{E}(T_1) - \mu)^2 = \frac{\sigma^2}{n}$$

De manera analoga para  $T_2$  se obtiene que

$$ECM(T_2) = \frac{n\sigma^2 + \mu^2}{(n+1)^2}.$$

Si consideramos el caso particular  $n = 10$  y  $\sigma^2 = 100$ , entonces

$$ECM(T_1) = 10$$

y

$$ECM(T_2) = \frac{1000 + \mu^2}{121},$$

luego si  $\mu < \sqrt{210}$  se tendrá que  $ECM(T_2) < ECM(T_1)$ , pero si  $\mu > \sqrt{210}$  entonces  $ECM(T_2) > ECM(T_1)$ . ¿Cuál elegimos? En este caso podría ser útil examinar otros criterios para decidir cual de los dos estimadores es un mejor estimador para  $\mu$ .

## 2. Sesgo de un Estimador

Recordemos que el error cuadrático medio se puede escribir como:

$$ECM(T, \theta) = \text{Var}(T) + [\theta - \mathbb{E}(T)]^2$$

En la expresión anterior, la diferencia  $(\theta - \mathbb{E}(T))$  se denomina **sesgo** del estimador  $T$ .

Dado que este valor aparece en la fórmula del error cuadrático medio, es deseable que sea lo más pequeño posible en valor absoluto. Lo anterior se traduce en que, en promedio,  $T$  sea lo más cercano a  $\theta$  posible.

Siguiendo la lógica anterior, tenemos que:

**Definición 2** Un estimador  $T = u(x_1, \dots, X_n)$  se dice **insesgado** si  $\mathbb{E}(T) = \theta$ .

Para ejemplificar la situación anterior, consideremos una muestra aleatoria  $X_1, \dots, X_n$  con  $\mathbb{E}(X_i) = \mu$  y  $\text{Var}(X_i) = \sigma^2$ , y  $T_1 = \bar{X}$  y  $T_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  como estimadores de  $\mu$  y  $\sigma^2$  respectivamente.

Notemos que

$$\begin{aligned} \mathbb{E}(T_1) &= \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n} \mathbb{E}\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) \\ &= \frac{1}{n} \sum_{i=1}^n \mu \\ &= \mu \end{aligned}$$

Por lo tanto,  $\bar{X}$  es un estimador **insesgado** para  $\mu$ .

Sin embargo, notemos que para  $T_2$  tenemos que:

$$\begin{aligned} \mathbb{E}(T_2) &= \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) \\ &= \frac{1}{n} \mathbb{E}\left(\sum_{i=1}^n [(X_i - \mu) - (\bar{X} - \mu)]^2\right) \\ &= \frac{1}{n} \mathbb{E}\left(\sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2\right) \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}((X_i - \mu)^2) - n\mathbb{E}((\bar{X} - \mu)^2) \\ &= \frac{1}{n} \left[ \sum_{i=1}^n \sigma^2 - n \frac{\sigma^2}{n} \right] \quad \text{pues } \text{Var}(X_i) = \sigma^2, \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \\ &= \frac{n-1}{n} \sigma^2 \end{aligned}$$

Por lo tanto  $T_2$  es un estimador **sesgado** para  $\sigma^2$ , pero lo anterior se puede solucionar considerando como estimador de  $\sigma^2$  a  $T_3 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ , ya que en ese caso  $\mathbb{E}(T_3) = \sigma^2$ .

Es por esta razón que generalmente se ve calcula la varianza muestral dividiendo por  $n - 1$  en vez de por  $n$ .

### 3. Estimadores Consistentes

Intuitivamente, otra característica deseable para un estimador de un parámetro  $\theta$  es que, nos proporcione una mejor estimación si nos basamos en 30 observaciones que si lo hacemos en sólo 5.

Formalmente, si tenemos una sucesión de estimadores  $T_n$  (donde  $n$  es el tamaño de la muestra), deseamos que se “acerquen”, en algún sentido, a  $\theta$  a medida que  $n$  crece.

Bajo esta idea podemos definir la noción de consistencia.

**Definición 3** Sea  $T$  una estimador para  $\theta$  y sean  $T_1, T_2, \dots, T_n, \dots$  una secuencia de estimadores de  $\theta$  que representan a  $T$  en base a muestras de tamaño  $1, 2, \dots, n, \dots$  respectivamente.

Se dice que  $T$  es **consistente** para  $\theta$  si  $\forall \varepsilon > 0$ :

$$\lim_{n \rightarrow \infty} \mathbb{P}(|T_n - \theta| \leq \varepsilon) = 1$$

Sin embargo, verificar la condición de consistencia es algo complicado, y por esta razón se verifican condiciones más fuertes como convergencia en media cuadrática, es decir que:

$$\lim_{n \rightarrow \infty} \mathbb{E}[(T_n - \theta)^2] = 0$$

Lo anterior da origen al siguiente resultado:

**Proposición 1** Si  $T_n$  es un estimador tal que:

$$\lim_{n \rightarrow \infty} \mathbb{E}(T_n) = \theta$$

$$\lim_{n \rightarrow \infty} \text{Var}(T_n) = 0$$

entonces  $T_n$  es un estimador **consistente** para  $\theta$ .

Un estimador que cumpla la primera igualdad enunciada en la proposición (i.e.  $\lim_{n \rightarrow \infty} \mathbb{E}(T_n) = \theta$ ) es llamado **asintóticamente insesgado**.

Notar que la proposición anterior sólo entrega una condición suficiente, y por lo tanto no nos sirve para probar que un estimador NO es consistente. Sin embargo, la siguiente proposición nos da un criterio para este caso.

**Proposición 2** Si  $T_n$  es un estimador consistente de  $\theta$  y  $\mathbb{E}(T_n)$  es finito entonces  $T_n$  es **asintóticamente insesgado**.

### 4. Estimadores Eficientes

En lo que sigue nos restringuiremos al caso de estimadores insesgados. En este caso tendremos que  $\text{ECM}(T, \theta) = \text{Var}(T)$  y buscaremos un estimador que tenga la menor varianza posible entre todos los estimadores.

Para formalizar lo anterior, definamos la noción de **eficiencia**.

**Definición 4**  $T_1$  es un estimador **más eficiente** de  $\theta$  que  $T_2$  si  $\text{Var}(T_1) \leq \text{Var}(T_2)$ , cumpliéndose con desigualdad estricta para algún valor de  $\theta$ .

**Nota:** Cuando los estimadores son sesgados, se emplean los errores cuadráticos medios para determinar eficiencias relativas.

El problema de como encontrar un estimador de mínima varianza se facilita bastante utilizando la llamada cota de **Cramer-Rao**.

Antes de enunciar dicho teorema, es importante definir algunas funciones importantes:

**Definición 5** Se define la **cantidad de información de Fischer** dada por la variable aleatoria  $X$  sobre el parámetro  $\theta$  como la función:

$$I(\theta) = \mathbb{E} \left[ \left( \frac{\partial \ln f(X; \theta)}{\partial \theta} \right)^2 \right]$$

De la misma forma podemos definir una cantidad similar para una muestra  $X_1, \dots, X_n$ .

**Definición 6** Se define la cantidad de información de Fischer dada la muestra aleatoria  $X_1, \dots, X_n$  sobre el parámetro  $\theta$  como la función:

$$I_n(\theta) = \mathbb{E} \left[ \left( \frac{\partial \ln f_n(X; \theta)}{\partial \theta} \right)^2 \right]$$

Es importante notar que las últimas dos definiciones se pueden relacionar mediante la fórmula

$$I_n(\theta) = nI(\theta) \quad (\text{cuando el dominio de } X \text{ no dependa de } \theta)$$

**Nota:** Cada vez que aparezca la hipótesis “que el dominio de  $X$  no dependa de  $\theta$ ” esta se puede reemplazar por la siguiente hipótesis:

$$\frac{\partial}{\partial \theta} \int_D f(x, \theta) dx = \int_D \frac{\partial f(x, \theta)}{\partial \theta} dx.$$

En otras palabras, la información de Fischer en una muestra aleatoria de  $n$  observaciones es simplemente  $n$  veces la información de Fischer en una sola observación.

Otras propiedades de la información de Fischer son:

**Propiedad 1** Si el dominio de  $X$  no depende de  $\theta$  se tiene que:

$$I_n(\theta) = \text{Var} \left( \frac{\partial \ln f_n(X; \theta)}{\partial \theta} \right)$$

$$I_n(\theta) = -\mathbb{E} \left( \frac{\partial^2 \ln f_n(X; \theta)}{\partial \theta^2} \right)$$

Ahora, estamos en condiciones de enunciar el teorema que nos ayudará a encontrar estimadores eficientes para un parámetro.

**Teorema 1** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una distribución con densidad  $f(x; \theta)$  tal que el dominio de  $X$  no dependa de  $\theta$ . Si  $T$  es un estimador insesgado de  $\theta$ , entonces la varianza de  $T$  debe satisfacer

$$\text{Var}(T) \geq \frac{1}{nI(\theta)}$$

Notemos que el teorema anterior establece una cota inferior para la varianza de un estimador de  $\theta$ . Sin embargo, el teorema anterior no dice que el estimador de mínima varianza sea igual a la cota inferior de Cramer-Rao. En otras palabras, es posible encontrar un estimador insesgado de  $\theta$  que tenga la varianza más pequeña posible de entre todos los estimadores insesgados de  $\theta$ , pero cuya varianza es más grande que el límite inferior de Cramer-Rao.

En el caso especial en que se tiene la igualdad con la cota de Cramer-Rao, se tiene la siguiente definición:

**Definición 7** Si  $T$  es un estimador insesgado del parámetro  $\theta$  tal que

$$\text{Var}(T) = \frac{1}{nI(\theta)}$$

entonces se dice que  $T$  es un estimador **eficiente** de  $\theta$ .

Con lo anterior, tenemos que un estimador insesgado será eficiente si es el de mínima varianza y además es igual al límite inferior de Cramer-Rao.

Dicho estimador de  $\theta$  será el mejor estimador insesgado de  $\theta$ .