



# Estimación puntual y por intervalo.

---

## Esquema

- [1. Estimadores insesgados.](#)
  - [2. Desigualdad de Tchebysheff.](#)
  - [3. Estimadores insesgados de varianza mínima.](#)
  - [4. Estadísticos suficientes.](#)
  - [5. Métodos de estimación puntual.](#)
    - [5.1 Estimación por máxima verosimilitud.](#)
    - [5.2 Método de los momentos](#)
    - [5.3 Estimación de máxima verosimilitud para muestras censuradas.](#)
- 

Sean  $X_1, X_2, \dots, X_n$  una muestra aleatoria de tamaño  $n$  proveniente de una distribución con función de densidad conocida  $f(x; \theta)$ , y sea  $T = u(X_1, X_2, \dots, X_n)$  cualquier estadístico. Queremos encontrar una función  $u$  que proporcione la “mejor” estimación de  $\theta$ .

### **Definición.**

Se define el error cuadrático medio:

$$ECM(T) = E[(T - \theta)^2]$$

Por tanto, se tiene:

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

El error cuadrático medio es la suma de dos cantidades positivas, una es la varianza del estimador y la otra el cuadrado del sesgo del estimador.

Visto de una manera simple, el “mejor” estimador será aquel que minimice el ECM. Sin embargo, aún en el caso en el que determinemos el ECM para un gran número de estimadores, para la mayor parte de las funciones de densidad  $f(x; \theta)$  no existe un estimador que minimice el error cuadrático medio para todos los posibles valores de  $\theta$ . Por esta razón hay que añadir criterios adicionales para la selección de los estimadores de  $\theta$ .

## Estimadores insesgados.

En el ECM el término  $[\theta - E(T)]$  recibe el nombre de sesgo del estimador.

### Definición

Se dice que  $T = u(X_1, X_2, \dots, X_n)$  es un estimador insesgado del parámetro  $\theta$ , si  $E(T) = \theta$  para cualquier  $\theta$ .

De esta forma, para cualquier estimador insesgado de  $\theta$ , la distribución de muestreo de  $T$  se encuentra centrada alrededor de  $\theta$  y  $ECM(T) = \text{var}(T)$ .

Puede demostrarse que si:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \text{ y } \forall i, X_i \rightarrow N(\mu, \sigma) \text{ entonces } E(S^2) = \sigma^2$$

Demostración

$$\frac{(n-1)S^2}{\sigma^2} \rightarrow \chi_{n-1}^2 \text{ Sea } Y = \frac{(n-1)S^2}{\sigma^2}; S^2 = \frac{\sigma^2 Y}{n-1} \quad E(S^2) = \frac{\sigma^2}{n-1} E(Y)$$

Como  $Y$  sigue una Chi-cuadrado con  $n-1$  grado de libertad;  $E(Y) = n-1$ . Luego

$$E(S^2) = \frac{\sigma^2}{n-1} E(Y) = \frac{\sigma^2}{n-1} (n-1) = \sigma^2$$

Este resultado se puede extender sin importar cuál sea la distribución de probabilidad de la población de interés.

## Teorema 1

Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de alguna población con función de densidad no especificada de manera que  $E(X_i) = \mu$  y  $\text{Var}(X_i) = \sigma^2$  para todo  $i$ . Entonces si

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

$S^2$  es un estimador insesgado de  $\sigma^2$ , es decir sólo cuando se divide por  $n-1$ , se obtiene un estimador insesgado.

Estimadores consistentes.

La idea es que conforme se va aumentando el tamaño de la muestra, el estimador se va concentrando cada vez mas alrededor de  $\theta$ .

### Definición

Sea  $T$  un estimador del parámetro  $\theta$ , y sean  $T_1, T_2, \dots, T_n$  una secuencia de estimadores que representan a  $T$  en base a muestras de tamaño  $1, 2, \dots, n$ ; respectivamente. Se dice que  $T$  es un estimador consistente de  $\theta$  si:

$$\lim_{n \rightarrow \infty} P(|T_n - \theta| \leq \varepsilon) = 1 \text{ para todos los valores de } \theta, \varepsilon > 0$$

El requisito

$$\lim_{n \rightarrow \infty} P(|T_n - \theta| \leq \varepsilon) = 1 \forall \text{ los valores de } \theta, \varepsilon > 0$$

constituye lo que se llama convergencia en probabilidad.

Esto implica que la varianza de un estimador consistente  $T_n$  disminuye conforme  $n$  crece. Una condición necesaria, aunque no suficiente para que  $T_n$  sea un estimador consistente de  $\theta$  es que  $\text{Var}(T_n)$  tienda a cero cuando  $n$  tiende a infinito.

Por ejemplo: la media muestral y la varianza muestral son estimadores consistentes de  $\mu$  y  $\sigma^2$ . para demostrarlo veamos primero el importante teorema siguiente:

[\[Volver al principio\]](#)

## Desigualdad de Tchebysheff.

Sea  $X$  una variable aleatoria con función de densidad  $f(x)$  y tal que  $E(X)=\mu$  y  $\text{Var}(X)=\sigma^2$  son finitas. Entonces:

$$P(|X - \mu| \leq k\sigma) \geq 1 - \frac{1}{k^2}$$

o de otra manera

$$P(|X - \mu| > k\sigma) \leq \frac{1}{k^2}$$

### Teorema.

Sean  $X_1, X_2, \dots, X_n$   $n$  variables aleatorias IID, tales que  $E(X_i)=\mu$  y  $\text{Var}(X_i)=\sigma^2$  y son finitas. Entonces

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

es un estimador consistente de  $\mu$ . Es decir:

$$\lim_{n \rightarrow \infty} P\left(|\bar{X}_n - \mu| \leq \varepsilon\right) = 1$$

Este importante teorema se conoce con el nombre de Ley de los grandes números

Demostración

Dado que  $E(X_n) = \mu$  y  $\text{Var}(X_n) = \sigma^2 / n$ , aplicando la desigualdad de Tchebycheff, se tiene:

$$P\left(|\bar{X} - \mu| > \frac{k\sigma}{\sqrt{n}}\right) \leq \frac{1}{k^2}$$

Tomando

$$k = \frac{\varepsilon\sqrt{n}}{\sigma}$$

se tiene :

$$P\left(|\bar{X} - \mu| > \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2 \cdot n}$$

Tomando límites cuando  $n$  tiende a  $\infty$ .

$$\lim_{n \rightarrow \infty} P\left(|\bar{X} - \mu| > \varepsilon\right) = 0$$

y por tanto:

$$\lim_{n \rightarrow \infty} P\left(|\bar{X} - \mu| \leq \varepsilon\right) = 1$$

### Ejercicio.

Considérese una muestra de tamaño  $n$  de una distribución que tiene varianza conocida  $\sigma^2=10$  y media desconocida  $\mu$ . ¿Cuál debe ser el tamaño de la muestra para que la media

$\bar{X}_n$  se encuentre dentro de un intervalo igual a 2 unidades de la media poblacional con una probabilidad de al menos 0,9?.

$$P\left(\left|\bar{X} - \mu\right| \leq \frac{k\sigma}{\sqrt{n}}\right) \geq 1 - \frac{1}{k^2}$$

Por tanto:

$$1 - \frac{1}{k^2} = 0,9; \quad 0,1 = \frac{1}{k^2}; \quad k^2 = 10 \quad y \quad k = \sqrt{10}$$

De otro lado

$$\frac{k\sigma}{\sqrt{n}} = 2; \quad k^2\sigma^2 = 4n; \quad 10 \cdot 10 = 4n; \quad n = 25$$

[\[Volver al principio\]](#)

## Estimadores insesgados de varianza mínima.

Para un parámetro dado  $\theta$ , considérese la clase formada por todos los estimadores insesgados de  $\theta$ . Si  $T$  es un estadístico perteneciente a esta clase, entonces  $E(T)=\theta$  y  $ECM(T) = \text{Var}(T)$ . Debe buscarse una clase de estimadores insesgados, si es que existe, que tenga una varianza mínima para todos los valores de  $\theta$ . Este estimador recibe el nombre de estimador insesgado de varianza mínima uniforme (VMU) de  $\theta$ .

### Definición

Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria cuya función de probabilidad es  $f(x; \theta)$ . Sea el estadístico  $T=u(X_1, X_2, \dots, X_n)$  un estimador de  $\theta$  tal que  $E(T)=\theta$  y  $\text{Var}(T)$  es menor que la varianza de cualquier otro estimador insesgado de  $\theta$ . Se dice que  $T$  es un estimador insesgado de varianza mínima.

Si  $T_1$  y  $T_2$  son dos estimadores insesgados de  $\theta$ ,  $T_1$  es más eficiente que  $T_2$  si  $\text{Var}(T_1) \leq \text{Var}(T_2)$  cumpliéndose la desigualdad en sentido estricto para algún valor de  $\theta$ .

Es muy común utilizar el cociente  $\text{Var}(T_1)/\text{Var}(T_2)$  para medir la eficiencia. La búsqueda de un estimador VMU se facilita bastante con la ayuda de un resultado que recibe el nombre de cota inferior de Cramer-Rao.

### Teorema

Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de función de densidad de probabilidad  $f(x;\theta)$ . Si  $T$  es un estimador insesgado de  $\theta$ , entonces la varianza de  $T$  debe de satisfacer la siguiente desigualdad.

$$\text{Var}(T) \geq \frac{1}{nE \left[ \left( \frac{\partial \text{Lnf}(x; \theta)}{\partial \theta} \right)^2 \right]}$$

### Definición

Si T es cualquier estimador insesgado del parámetro  $\theta$  tal que

$$\text{Var}(T) = \frac{1}{nE \left[ \left( \frac{\partial \text{Lnf}(x; \theta)}{\partial \theta} \right)^2 \right]}$$

Entonces T es un estimador eficiente de  $\theta$ .

[\[Volver al principio\]](#)

## Estadísticos suficientes.

Intuitivamente, un estadístico es suficiente para el parámetro  $\theta$  si utiliza toda la información contenida en la muestra aleatoria con respecto a  $\theta$ .

Un criterio para ver si un estadístico es suficiente viene dado por el teorema de factorización de Neyman.

### Teorema.

Sea  $X_1, X_2, \dots, X_n$  una muestra de una distribución con una función de densidad de probabilidad  $f(x; \theta)$ . Se dice que  $T=u(X_1, X_2, \dots, X_n)$  es un estadístico suficiente de  $\theta$  si y solo si la función de verosimilitud puede factorizarse de la siguiente forma:

$$L(x_1, x_2, \dots, x_n) = h(t; \theta)g(x_1, x_2, \dots, x_n)$$

Para cualquier valor  $t=u(x_1, x_2, \dots, x_n)$  de T y en donde  $g(x_1, x_2, \dots, x_n)$  no contiene al parámetro  $\theta$ .

[\[Volver al principio\]](#)

## Métodos de estimación puntual.

¿Cómo obtener estimadores que cumplan con las propiedades deseables de los estimadores?. Veamos el método de máxima verosimilitud y el de los momentos. Más adelante, en otro capítulo, se estudiará el método de mínimos cuadrados:

### Estimación por máxima verosimilitud.

En esencia el método consiste en seleccionar como estimador a aquél valor del parámetro que tiene la propiedad de maximizar el valor de la probabilidad de la muestra observada. Es decir, encontrar el valor del parámetro que maximiza la función de verosimilitud.

### Definición

Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una distribución con función de densidad de probabilidad  $f(x; \theta)$ , y sea  $L(x_1, x_2, \dots, x_n; \theta)$  la verosimilitud de la muestra como función de  $\theta$ . Si  $t = u(x_1, x_2, \dots, x_n)$  es el valor de  $\theta$  para el cual la función de verosimilitud es máxima, entonces  $T = u(X_1, X_2, \dots, X_n)$  es el estimador de máxima verosimilitud de  $\theta$ .

El método de máxima verosimilitud proporciona el estimador eficiente, si es que existe. Sin embargo, los estimadores de MV son por lo general sesgados.

Por otra parte, es más fácil, generalmente, maximizar  $\ln(L(\theta))$  que  $L(\theta)$ .

El método de máxima verosimilitud posee otra propiedad deseable conocida como propiedad de invarianza.

$$\text{Sea } \hat{\theta} = u(X_1, X_2, \dots, X_n)$$

el estimador de máxima verosimilitud de  $\theta$ . Si  $g(\theta)$  es una función univaluada de  $\theta$ , entonces el estimador de máxima verosimilitud de  $g(\theta)$  es  $g(\hat{\theta})$ .

[\[Volver al principio\]](#)

### Método de los momentos

Este método consiste en igualar los momentos apropiados de la distribución de la población con los correspondientes momentos muestrales para estimar un parámetro desconocido de la población.

### Definición

Sea  $X_1, X_2, \dots, X_n$  una muestra aleatoria con función de densidad  $f(x; \theta)$ . El  $r$ -ésimo momento alrededor de cero se define como

$$M_r' = \frac{1}{n} \sum_{i=1}^n x_i^r$$

El método de los momentos constituye una alternativa razonable cuando no pueden hallarse los estimadores de máxima verosimilitud.

Téngase en cuenta que muchas veces los parámetros son funciones de los momentos teóricos.

[\[Volver al principio\]](#)

### Estimación de máxima verosimilitud para muestras censuradas.

Una prueba típica de duración consiste en seleccionar de manera aleatoria un conjunto de artículos iguales y someterlos a un cuidados proceso hasta que el artículo falla.

Si la prueba de duración se termina sólo cuando todos los artículos han fallado, se dice que la muestra aleatoria de tiempos está completa. Sin embargo, generalmente, si la prueba termina después de un lapso determinado de tiempo  $x_0$  o después de la falla de un número determinado de unidades  $m \leq n$ . Las dos condiciones producen muestras censuradas.

Si  $X_0$  es un lapso fijo de tiempo, el número de unidades que fallan de las  $n$ , después del comienzo de la prueba hasta el tiempo  $x_0$ , es una variable aleatoria, se dice entonces que la muestra es del tipo I

Si  $m$  es fijo y el tiempo de terminación es la variable aleatoria  $X_m$ , se dice que la muestra es de tipo II.

Si no se tienen en cuenta las inferencias, existe muy poca diferencia entre ambos tipos.

Estudiamos las de tipo II.

Supongamos que el primer fallo se dio en un tiempo igual a  $x_1$  desde el comienzo y el 2º en un tiempo  $x_2$ , también desde el comienzo, y así hasta  $x_m$ , en donde  $m \leq n$ , es el número fijado de antemano para terminar la prueba. Es obvio que  $x_1 \leq x_2 \leq \dots \leq x_m$  y que  $n-m$  unidades tienen un tiempo de supervivencia  $x_m$ .

Supóngase que los tiempos de duración son variables aleatorias  $X_1, X_2, \dots, X_n$  independientes normalmente distribuidas, con función de densidad

$$f(x; \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right) \quad x > 0, \quad \theta > 0$$

El interés recae en encontrar el estimador de máxima verosimilitud de  $\theta$ .

La función de verosimilitud para un muestreo censurado del tipo II es la probabilidad de que fallen  $m$  unidades en los tiempos  $x_1, x_2, \dots, x_m$  es  $f(x_1; \theta) \cdot f(x_2; \theta) \dots f(x_m; \theta)$ . Pero ésta es una de las formas en las que pueden fallar  $m$  unidades de  $n$ . El número total de formas es:

$$V_{n,m} = \frac{n!}{(n-m)!}$$

La probabilidad de que  $n-m$  unidades sobrevivan un tiempo  $x_m$  está por la función de confiabilidad a tiempo  $x_m$ .

$$P(X > x_m) = \exp(-x_m / \theta)$$

Por tanto la función de verosimilitud es



$$L(x_1, x_2, \dots, x_n; \theta) = \frac{n!}{(n-m)!} \left\{ \frac{1}{\theta} \exp\left(-\frac{x_1}{\theta}\right) \dots \frac{1}{\theta} \exp\left(-\frac{x_m}{\theta}\right) \right\} \exp\left(-\frac{x_m}{\theta}\right) \dots \exp\left(-\frac{x_m}{\theta}\right) =$$

$$\frac{n!}{(n-m)!} \frac{1}{\theta^m} \exp\left(-\frac{\sum_{i=1}^m x_i}{\theta}\right) \cdot \exp\left(-\frac{(n-m)}{\theta} x_m\right) = \frac{n!}{(n-m)!} \frac{1}{\theta^m} \exp\left(-\frac{1}{\theta} T_m\right)$$

Siendo

$$T_m = \sum_{i=1}^m x_i + (n-m)x_m$$

Tomando logaritmos

$$Ln(L(\dots)) = Ln(n!) - Ln((n-m)!) - mLn(\theta) - \frac{1}{\theta} T_m$$

Derivando con respecto a  $\theta$ .

$$\frac{\partial Ln(L(\theta))}{\partial \theta} = -\frac{m}{\theta} + \frac{1}{\theta^2} T_m = 0$$

Se deduce que

$$\theta = \frac{T_m}{m}$$

Luego

$$\hat{\theta} = \frac{\sum_{i=1}^m x_i + (n-m)x_m}{m}$$

[\[Volver al principio\]](#)