

## T.3 ESTIMACIÓN PUNTUAL

1. INTRODUCCIÓN: ESTIMACIÓN Y ESTIMADOR
  2. PROPIEDADES DE LOS ESTIMADORES
  3. MÉTODOS DE ESTIMACIÓN. EJEMPLO 1, EJEMPLO 2
- 

### 1. Introducción: Estimación y Estimador

En este tema se analizan las formas adecuadas para el establecimiento del conocimiento numérico o abstracto de un parámetro de una población, y que evidentemente nos es desconocido, partiendo, claro está, de la información suministrada por la muestra.

LA ESTIMACION, como proceso, consiste en que dada una población que siga una distribución de cierto tipo con función de probabilidad (de cuantía o de densidad)  $f(X, \theta)$  dependiente de un parámetro o varios desconocido(s) " $\theta$ ", aventurar en base a los datos muestrales el valor que toma o puede tomar el parámetro o parámetros.

UN ESTIMADOR  $\hat{\theta}(x)$  es una función de una muestra genérica  $x=[x_1, x_2, \dots, x_n]$ , es decir, un estadístico que utilizaremos para estimar el valor del parámetro. Por tanto es una variable aleatoria y será necesario para la estimación conocer la distribución muestral del estimador.

UNA ESTIMACION  $\hat{\theta}$  será el valor concreto que tomará el estimador al aplicarlo a la muestra concreta obtenida y será, por tanto, la solución concreta de nuestro problema. (Cuando no haya lugar para la confusión designaremos al estimador, a veces, como  $\hat{\theta}$  en vez de  $\hat{\theta}(x)$ )

LA TEORIA DE LA ESTIMACION se ocupará, dentro del marco de la perspectiva clásica, de estudiar las características deseables de los estimadores permitiéndonos escoger aquel estimador que reúna más propiedades ventajosas para que realicemos buenas estimaciones.

### 2. Propiedades de los Estimadores

**INSESGADEZ:** un estimador es insesgado o centrado cuando verifica que

$E(\hat{\theta}) = \theta$ . (Obsérvese que deberíamos usar  $\hat{\theta}(x)$  y no  $\hat{\theta}$ , pues hablamos de estimadores y no de estimaciones pero como no cabe la confusión, para simplificar,

aquí, y en lo sucesivo usaremos  $\hat{\theta}$ ). En caso contrario se dice que el estimador es sesgado.

Se llama sesgo a  $B(\hat{\theta}) = E(\hat{\theta}) - \theta$  [se designa con B de BIAS, sesgo en inglés]

Como ejemplo podemos decir que: la media muestral es un estimador insesgado de la media de la población (y lo es sea cual fuere la distribución de la población) ya que:

si el parámetro a estimar es  $\theta = \mu$  y establecemos como estimador de  $\mu$ :

$\hat{\mu} = \hat{\theta}(x) = \bar{x}$  tendremos que

$$E[\hat{\mu}(x)] = E[\bar{x}] = \mu$$

luego la media muestral es un estimador insesgado de la media poblacional.

En cambio la varianza muestral es un estimador sesgado de la varianza de la población, ya que: si utilizamos como estimador de  $\sigma^2$  la varianza muestral  $S^2$  es decir:

tendremos que

$$E[S^2] = E[S^2] = \sigma^2 (n-1) / n = \sigma^2 (1 - \frac{1}{n}) \neq \sigma^2$$

que es el parámetro a estimar. Existe pues un sesgo que será:

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

Dado que la varianza muestral no es un estimador de la varianza poblacional con propiedades de insesgadez, conviene establecer uno que si las tenga; este estimador no es otro que la cuasivarianza muestral, de ahí su importancia; así

la cuasivarianza es en función de la varianza

$$S^{*2} = \frac{n S^2}{n-1}$$

y tomada como estimador tendríamos que

$$E[S^{*2}] = E\left[\frac{n S^2}{n-1}\right] = \frac{n}{n-1} E[S^2] = \left(\frac{n}{n-1}\right) \left(\frac{n-1}{n}\right) \sigma^2 = \sigma^2$$

dado que la esperanza del estimador coincide con el parámetro a estimar podemos decir que la cuasivarianza muestral es un estimador insesgado de la varianza de la población.

No obstante, y dado que, cuando el tamaño de la muestra tiende a infinito el sesgo tiende a cero, se dice que el estimador es asintóticamente insesgado o asintóticamente centrado: podemos establecer que:

$$\lim_{n \rightarrow \infty} B(s^2) = \lim_{n \rightarrow \infty} \sigma^2/n = 0$$

Por tanto la varianza muestral es un estimador sesgado pero asintóticamente insesgado de la varianza de la población.

**CONSISTENCIA.** Un estimador es consistente si converge en probabilidad al parámetro a estimar. Esto es: si

$$\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| < \varepsilon) = 1$$

**LINEALIDAD.** Un estimador es lineal si se obtiene por combinación lineal de los elementos de la muestra; así tendríamos que un estimador lineal sería:

$$\hat{\theta} = \hat{\theta}(x) = \sum_{i=1}^n k_i x_i$$

**EFICIENCIA.** Un estimador es eficiente u óptimo cuando posee varianza mínima o bien en términos relativos cuando presenta menor varianza que otro. Quedando claro que el hecho puede plantearse también en términos más coherentes de Error Cuadrático Medio (ECM). Tendríamos que:  $ECM(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = D^2(\hat{\theta}) + (B(\hat{\theta}))^2$  por lo expresado podemos aventurar que un estimador insesgado, luego  $B(\hat{\theta}) = 0$

es el único capaz de generar eficiencia. (ir a cota de Cramer-Rao)

**SUFICIENCIA.** Un estimador es suficiente cuando  $f(x_1, x_2, \dots, x_n | \hat{\theta})$  no depende del parámetro a estimar,  $\theta$ .

En términos más simples: cuando se aprovecha toda la información muestral. (Ir a teorema de caracterización Neyman-Fisher)

### 3. Métodos de Estimación

**MÉTODO POR ANALOGÍA.** Consiste en aplicar la misma expresión formal del parámetro poblacional a la muestra, generalmente, estos estimadores son de cómoda operatividad, pero en ocasiones presentan sesgos y no resultan eficientes. Son recomendables, para muestras de tamaño grande al cumplir por ello propiedades asintóticas de consistencia.

**METODO DE LOS MOMENTOS.** Consiste en tomar como estimadores de los momentos de la población a los momentos de la muestra. Podríamos decir que es un caso particular del método de analogía. En términos operativos consiste en resolver el sistema de equivalencias entre unos adecuados momentos empíricos (muestrales) y teóricos (poblacionales).

Ejemplo:

conocemos que la media poblacional de una determinada variable  $x$  depende de un parámetro  $K$  que es el que realmente queremos conocer (estimar). Así

$$\mu = 2K + 7$$

por el método de los momentos tendríamos que

$$\hat{\theta} = \bar{x} \xleftarrow{\text{estimador}} \hat{\mu} \quad \text{de donde}$$

$$\hat{k} = (\bar{x} - 7) / 2$$

**ESTIMADORES MAXIMO-VEROSIMILES.** La verosimilitud consiste en otorgar a un estimador/estimación una determinada "credibilidad" una mayor apariencia de ser el cierto valor (estimación) o el cierto camino para conseguirlo (estimador).

En términos probabilísticos podríamos hablar de que la verosimilitud es la probabilidad de que ocurra o se dé una determinada muestra si es cierta la estimación que hemos efectuado o el estimador que hemos planteado.

Evidentemente, la máxima verosimilitud, será aquel estimador o estimación que nos arroja mayor credibilidad. En situación formal tendríamos:

Un estimador máximo-verosímil es el que se obtiene maximizando la función de verosimilitud (likelihood) de la muestra

$$L(x_1, x_2, \dots, x_n | \theta)$$

Que es la función de probabilidad (densidad o cuantía) que asigna la probabilidad de que se obtenga una muestra dependiendo del (o de los) parámetro(s) " $\theta$ " pero considerada como función de  $\theta$ . Si la distribución de la población es tal que su densidad depende de uno o más parámetros, la probabilidad (densidad) de cada realización muestral  $x_i$  (con  $i=1,2,\dots,n$ ) será:

y, a partir de aquí podremos obtener la función de verosimilitud de la muestra

$$L(x_1, x_2, \dots, x_n | \theta)$$

Si el muestreo es simple:

$$L(x_1, x_2, \dots, x_n | \theta) = f(x_1 | \theta) \cdot f(x_2 | \theta) \cdots f(x_n | \theta)$$

por ser independientes cada una de las realizaciones muestrales.

El estimador que maximice

$$L(x_1, x_2, \dots, x_n | \theta)$$

será el estimador máximo-verosímil (E.M.V.) Y será aquel valor/expresión para el que se verifique la derivada se iguale a cero:

$$\frac{\partial L(x_1, x_2, \dots, x_n | \theta)}{\partial \theta} = 0$$

Si lo planteado fuera EMV de varios parámetros

$$\theta = (\theta_1, \theta_2, \dots, \theta_n)$$

las expresiones serían.:

$$\left\{ \begin{array}{l} \frac{\partial L(X | \theta)}{\partial \theta_1} = 0 \\ \frac{\partial L(X | \theta)}{\partial \theta_2} = 0 \\ \vdots \\ \frac{\partial L(X | \theta)}{\partial \theta_n} = 0 \end{array} \right.$$

Debido a que la función de verosimilitud es a fin de cuentas una función de probabilidad, será una función definida no negativa y por lo tanto alcanzará su máximo en los mismos puntos que su logaritmo. Por esta razón suele maximizarse

$$\ln L(x_1, x_2, \dots, x_n | \theta)$$

en lugar de la propia función de verosimilitud . Suele hacerse esto en todos aquellos casos en los que la función de verosimilitud depende de funciones exponenciales.

### **ejemplo 1**

Obtener el E.M.V. del parámetro  $\lambda$  de una distribución de Poisson

Para una muestra de tamaño  $n$  tendremos que la función de de verosimilitud será:

$$L(x_1, x_2, \dots, x_n | \theta = \lambda) = \frac{e^{-\lambda} \cdot \lambda^{x_1}}{x_1!} \cdot \frac{e^{-\lambda} \cdot \lambda^{x_2}}{x_2!} \cdot \dots \cdot \frac{e^{-\lambda} \cdot \lambda^{x_n}}{x_n!}$$

maximizar  $L$  será equivalente a maximizar el numerador de  $L$  ; si llamamos  $L'$  a dicho numerador y tomamos logaritmos tendremos que es

$$L' = e^{-n\lambda} \cdot \lambda^{\sum x_i}$$

tomando logaritmos

$$\ln L' = -n\lambda + \sum x_i \ln \lambda$$

maximizando dicho logaritmo :

$$\frac{\partial \ln L'}{\partial \lambda} = -n + \sum x_i \frac{1}{\lambda} = 0 \Rightarrow \lambda = \frac{\sum x_i}{n} = \bar{x}$$

luego el estimador máximo verosímil de

### **ejemplo 2**

En una población normal estimar por el método de máxima verosimilitud los parámetros .

la función de verosimilitud será:

$$L(x_1, x_2, \dots, x_n | \theta = \mu, \theta_2 = \sigma^2) = \frac{e^{-\frac{1}{2} \left( \frac{x_1 - \mu}{\sigma} \right)^2}}{\sigma \cdot \sqrt{2\pi}} \cdot \frac{e^{-\frac{1}{2} \left( \frac{x_2 - \mu}{\sigma} \right)^2}}{\sigma \cdot \sqrt{2\pi}} \cdot \dots$$

$$\frac{e^{-\frac{1}{2} \left( \frac{x_n - \mu}{\sigma} \right)^2}}{\sigma \cdot \sqrt{2\pi}} = \frac{e^{-\frac{1}{2} \sum \left( \frac{x_i - \mu}{\sigma} \right)^2}}{\sigma^n \cdot 2\pi^{\frac{n}{2}}}$$

maximizar  $L$  es equivalente a maximizar  $\ln L$ .

$$\ln L = -\frac{1}{2 \cdot \sigma^2} \sum (x_i - \mu)^2 - n \ln(\sigma \sqrt{2\pi})$$

Maximizando:

$$\frac{\partial \ln L}{\partial \mu} = -\frac{1}{2 \sigma^2} 2 \sum (x_i - \mu) (-1) = 0$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{1}{2} \sum (x_i - \mu)^2 (-2) \cdot \frac{1}{\sigma^3} - \frac{n \sqrt{2\pi}}{\sigma \sqrt{2\pi}} = 0$$

despejando: de la primera :

$$n \mu = \sum x_i \rightarrow \mu = \bar{x}$$

de la segunda :

$$\sum (x_i - \mu)^2 - n \sigma^2 = 0 \rightarrow \sigma^2 = s^2$$

el estimador (MV) de la media poblacional será la media de la población mientras que el de la varianza poblacional corresponderá a su homónima muestral. Ambos estimadores reúnen casi-todas las propiedades salvo en el caso de la varianza muestral que es asintóticamente insesgada.