Tema 5

ESTIMACIÓN DE MÁXIMA VEROSIMILITUD Y OTROS MÉTODOS.

5.1. Estimación de máxima verosimilitud

El método de obtención de estimadores de máxima verosimilitud es el método más usado debido, en parte, a las buenas propiedades asintóticas que tienen los estimadores obtenidos con él. No se trata de un método que proporciona un criterio de selección, sino que es un método de cálculo.

Para entender la idea principal de este método se va a estudiar, primeramente, un ejemplo:

Ejemplo: Sea X una v.a. con distribución $B(n, p)$, donde $n \in \{2, 3\}$ y $p \in \{1/2, 1/3\}$. Basándose en la observación de un valor de la variable, decidir cuál de estos valores de los parámetros corresponden a la variable bajo estudio.

Antes de indicar como se aplica el método de máxima verosimilitud, se introducen una serie de conceptos:

Definición: Sea X una v.a. con distribución en una familia paramétrica de distribuciones, $\{F_\theta,\ \theta \in \Theta\}$. Sea $f_\theta(x)$ la f.m.p. (caso discreto) ó la f.d.d. (caso continuo) de X. Se considera $X_1, \ldots, X_n$ una m.a.s. de X y sea $f_\theta^n(x_1, \ldots, x_n)$ su f.m.p. ó f.d.d. conjunta con $\theta \in \Theta$. Para cada $x_1, \ldots, x_n$, realización muestral, se define la *función de verosimilitud* asociada a dichos valores de la muestra como una función de θ de la siguiente forma:

$$\begin{aligned} L_{x_1,\ldots,x_n} : \quad & \Theta \to \mathbb{R}^+ \cup \{0\} \\ & \theta \to L_{x_1,\ldots,x_n}(\theta) = f_\theta^n(x_1, \ldots, x_n). \end{aligned}$$

Definición: Sea X una v.a. con distribución en una familia paramétrica de distribuciones, $\{F_\theta,\ \theta \in \Theta\}$. Sea $X_1, \ldots, X_n$ una m.a.s. de X. Un estimador $\hat{\theta}(X_1, \ldots, X_n)$ de θ es *estimador de máximo verosímilitud* (EMV) de θ si:

$$\forall (x_1, \ldots, x_n) \in \mathcal{X}^n, \quad L_{x_1,\ldots,x_n}(\hat{\theta}(x_1, \ldots, x_n)) = \max_{\theta \in \Theta} L_{x_1,\ldots,x_n}(\theta).$$

Por tanto, el método de obtención de estimadores de máxima verosimilitud consiste en obtener un estimador que maximice la función de verosimilitud. Para determinarlo se deben tener en cuenta las propiedades analíticas de dicha función:

- Si la función de verosimilitud es derivable, el estimador máximo verosímil se calcula resolviendo las ecuaciones de verosimilitud, las cuales se obtienen de derivar la función de verosimilitud con respecto a cada uno de los parámetros e igualar las expresiones a cero. Lo más habitual es tomar el logaritmo neperiano de la función de verosimilitud, $\ln L_{x_1,\ldots,x_n}(\theta)$, debido a que muchas distribuciones tienen exponenciales en sus expresiones, y como el logaritmo neperiano es una función creciente, no afecta al cálculo del máximo.

  Por tanto si $\ln L_{x_1,\ldots,x_n}(\theta)$ es derivable, se obtienen las ecuaciones de verosimilitud:

  $$\frac{\partial \ln L_{x_1,\ldots,x_n}(\theta)}{\partial \theta_j} = 0 \quad j = 1, \ldots, k, \quad \theta = (\theta_1, \ldots, \theta_k)$$

  Las soluciones de estas ecuaciones son los posibles extremos de $\ln L_{x_1,\ldots,x_n}(\theta)$, que pueden ser máximos o no. Si la solución es única y es un máximo y es el EMV de θ, $\hat{\theta}$. Si existen varias soluciones, se puede tomar el máximo absoluto entre ellas como $\hat{\theta}$. Una vez obtenida la solución se debe comprobar que, efectivamente, se trata de un estimador.

- Si la función de verosimilitud no es derivable, entonces hay que recurrir a otro tipo de métodos, incluso métodos numéricos, para obtener el máximo.

Ejemplos:

1. Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim \{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$. Encontrar el EMV para μ y σ^2, en el caso de un parámetro conocidos y cuando ambos parámetros son desconocidos.

2. Sea $X_1, \ldots, X_n$ una m.a.s. de X cuya f.m.p es: $P_N[X = x] = \frac{1}{N},\ x = 1, \ldots, N$ (uniforme discreta en N puntos). Calcular el EMV de N.

3. Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim \{U(\theta - 1/2, \theta + 1/2), \theta \in \mathbb{R}\}$. Calcular un EMV para θ. (El EMV no tiene que ser único).

4. Sea X una v.a. con distribución $B(1,p)$ con $p \in [1/4, 3/4]$. Calcular el EMV de p, para una muestra de tamaño 1, ver que no es insesgado y calcular su error cuadrático medio: $E[\hat{p} - p]^2$. Comprobar que el estimador $T(X) = 1/2$ es mejor que $\hat{p}$ en el sentido del ECM. (El EMV no tiene porque ser el mejor en el sentido del menor error cuadrático medio).

5.1.1. Propiedades de los EMV

Aunque estos ejemplos muestran que los EMV no tienen por qué ser únicos, ni insesgados, ni minimizar el error cuadrático medio, sí que tienen ciertas propiedades.

Teorema 1: (Propiedades asintóticas)

Bajo condiciones bastantes generales, que incluyen las condiciones de regularidad de Fréchet-Cramér-Rao, si las ecuaciones de verosimilitud tienen solución única, $\hat{\theta}(X_1, \ldots, X_n)$, esta solución satisface:

- El EMV es fuertemente consistente.

$$\hat{\theta}(X_1, \ldots, X_n) \xrightarrow[n\to+\infty]{c.s.} \theta, \quad \forall \theta \in \Theta.$$

- El EMV es asintóticamente normal.

$$\hat{\theta}(X_1, \ldots, X_n) \xrightarrow[n\to+\infty]{L} \mathcal{N}(\theta, 1/(nI_X(\theta))), \quad \forall \theta \in \Theta.$$

La normalidad asintótica implica que, para muestras grandes, la distribución del EMV es aproximadamente normal, de media θ, y su varianza alcanza la cota de FCR ($Var_\theta[\hat{\theta}(X_1, \ldots, X_n)] \approx 1/(nI_X(\theta)) = 1/I_{X_1,\ldots,X_n}(\theta)$).

Teorema 2: (Relación entre EMV y estadístico suficiente)

Sea X una v.a. con distribución en una familia paramétrica de distribuciones, $\{F_\theta, \theta \in \Theta\}$. Sea $X_1, \ldots, X_n$ una m.a.s. de X. Supongamos que la familia admite un estadístico suficiente $T(X_1, \ldots, X_n)$. Entonces, si existe un EMV de θ, es una función (no constante) de $T(X_1, \ldots, X_n)$.

Este resultado no implica que el EMV tenga que ser un estadístico suficiente.

Ejemplo: Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim U(\theta, \theta+1)$. Calcular el estadístico suficiente, un EMV para θ, y comprobar que no coinciden.

Teorema 3: (Relación entre EMV y estimador eficiente)

Sea X una v.a. con distribución en una familia paramétrica de distribuciones, $\{F_\theta, \theta \in \Theta\}$, regular con $0 < I_X(\theta) < \infty$. Sea $X_1, \ldots, X_n$ una m.a.s. de X. Si $T(X_1, \ldots, X_n)$ es un estimador eficiente para θ, entonces existe un único EMV y coincide con $T(X_1, \ldots, X_n)$.

5.1.2. Estimadores de máxima verosimilitud de una una función paramétrica

Sea $g : \Theta \to \Lambda$ una función paramétrica. Se puede definir el concepto de función de verosimilitud sobre Λ a partir de la función de verosimilitud definida sobre Θ de la siguiente forma:

Definición: Para cada $x_1, \ldots, x_n$, realización muestral, se define la *función de verosimilitud* de $\lambda = g(\theta)$ asociada a dicha realización como:

$$\begin{aligned} M_{x_1,\ldots,x_n} :\ & \Lambda \to \mathbb{R}^+ \cup \{0\} \\ & \lambda \to M_{x_1,\ldots,x_n}(\lambda) = \sup_{\theta \in g^{-1}(\lambda)} L_{x_1,\ldots,x_n(\theta)}. \end{aligned}$$

Por otro lado, de forma análoga a como se definió el EMV de θ se puede definir el de λ.

Definición: Un estimador $\hat{\lambda}(X_1, \ldots, X_n)$ de λ es *estimador de máximo verosímilitud* (EMV) de λ si:

$$\forall (x_1, \ldots, x_n) \in \mathcal{X}^n, \quad M_{x_1,\ldots,x_n}(\hat{\lambda}(x_1, \ldots, x_n)) = \max_{\lambda \in \Lambda} M_{x_1,\ldots,x_n}(\lambda).$$

Teorema de invarianza de Zehna: Sea X una v.a. con distribución en una familia paramétrica de distribuciones, $\{F_\theta,\, \theta \in \Theta\}$. Sea $X_1, \ldots, X_n$ una m.a.s. de X. Sea g una función medible. Si $\hat{\theta}(X_1, \ldots, X_n)$ es EMV de θ, entonces $g\left(\hat{\theta}(X_1, \ldots, X_n)\right)$ es EMV de $g(\theta)$.

Ejemplos:

1. En el muestreo de una v.a. $X \sim \{\mathcal{P}(\lambda),\ \lambda > 0\}$, se obtiene que en n observaciones aparece y veces el valor 0. Obtener un EMV de λ a partir de esta información.

2. La duración de cierto tipo de lámparas es exponencial de media θ, desconocida. Después de observar el tiempo de vida de n lámparas, estimar por máxima verosimilitud la probabilidad de que la duración de una lámpara sea superior a 500 horas.

5.2. Otros métodos de estimación puntual: método de los momentos y de mínimos cuadrados

5.2.1. Método de los momentos

El método de los momentos, el cual fue introducido por K. Pearson, es el método más antiguo y sencillo para obtener estimadores de los parámetros poblacionales.

Este método consiste en estima cualquier función medible de los momentos poblacionales por la misma función de los momentos muestrales. En particular se igualan tantos momentos muestrales como parámetros haya que estimar, a los correspondientes momentos poblacionales, que son funciones de los parámetros desconocidos, y se resuelve el sistema de ecuaciones resultante obteniendose estimadores de los parámetros.

La idea es la siguiente: Sea X una v.a. con distribución en una familia paramétrica de distribuciones, $\{F_\theta,\ \theta \in \Theta \subseteq \mathbb{R}^k\}$, es decir, $\theta = (\theta_1, \ldots, \theta_k)$. Sea $X_1, \ldots, X_n$ una m.a.s. de X. Sean $m_1, \ldots, m_k$ los k primeros momentos no centrados de X,

$$m_j(\theta_1, \ldots, \theta_k) = E_{(\theta_1,\ldots,\theta_k)}[X^j] = \begin{cases} \sum x_i^j P_{\theta_1,\ldots,\theta_k}[X = x_i] & \text{caso discreto} \\ \int x^j f_{\theta_1,\ldots,\theta_k}(x)\, dx & \text{caso continuo} \end{cases}$$

En general m_j será una función de los k parámetros $\theta_1, \ldots, \theta_k$.

Por otro lado, asociados a la muestra se pueden obtener los k primeros momentos no centrados muestrales, que son:

$$A_1 = \sum_{i=1}^n \frac{X_i}{n}, \ldots, A_j = \sum_{i=1}^n \frac{X_i^j}{n}, \ldots, A_k = \sum_{i=1}^n \frac{X_i^k}{n}$$

Igualando los k primeros momentos poblacionales, m_j, a los correspondientes momentos muestrales, A_j, se obtiene un sistema de k ecuaciones con k incógnitas, $\theta_1, \ldots, \theta_k$,

$$\left.\begin{array}{c} m_1(\theta_1, \ldots, \theta_k) = A_1 \\ \vdots \\ m_j(\theta_1, \ldots, \theta_k) = A_j \\ \vdots \\ m_k(\theta_1, \ldots, \theta_k) = A_k \end{array}\right\}$$

cuyas soluciones son los estimadores de los parámetros: $\hat{\theta}_1, \ldots, \hat{\theta}_k$.

Ejemplos: Estimar mediante el método de los momentos los parámetros de las siguientes distribuciones:

1. $X \sim \mathcal{N}(\mu, \sigma^2)$.
2. $X \sim U(a, b)$.
3. $X \sim \mathcal{P}(\lambda)$.
4. $X \sim U(0, \theta)$.

La propiedad más notable de los estimadores obtenidos por el método de los momentos es la propiedad de consistencia: Si $\theta = h(m_1, \ldots, m_k)$, h continua, y $A_1^{(n)}, \ldots, A_k^{(n)}$ son los momentos muestrales correspondientes a una muestra de tamaño n, entonces

$$\hat{\theta}_n = h(A_1^{(n)}, \ldots, A_k^{(n)}) \to h(m_1, \ldots, m_k) = \theta \quad c.s. \quad (n \longrightarrow \infty)$$

5.2.2. Método de mínimos cuadrados

Sea $X = \varphi(t, \theta)$ una magnitud (de interés) que depende de ciertas condiciones experimentales (t) y de ciertos parámetros desconocidos a estimar $(\theta = (\theta_1, \ldots, \theta_k) \in \Theta \subseteq \mathbb{R}^k)$.

En principio, si fuese posible observar X bajo distintas condiciones experimentales (t_i) se podrían procurar suficientes relaciones del tipo

$$\left.\begin{array}{c} x_1 = \varphi(t_1, \theta) \\ \vdots \\ x_n = \varphi(t_n, \theta) \end{array}\right\}$$

para determinar $\theta_1, \ldots, \theta_k$ despejando en el sistema obtenido.

Sin embargo, las observaciones de X conlleva un error de medida aleatorio, ϵ, de forma que se obtendrían relaciones en términos de variables aleatorias, del tipo

$$\left.\begin{array}{c} X_1 = \varphi(t_1, \theta) + \epsilon_1 \\ \vdots \\ X_n = \varphi(t_n, \theta) + \epsilon_n \end{array}\right\}$$

Se plantea, entonces, el problema de estimar θ a partir de la muestra aleatoria $X_1, \ldots, X_n$.

El método de mínimos cuadrados consiste en elegir el θ que minimice

$$\sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} \left(X_i - \varphi(t_i, \theta)\right)^2$$

Si φ es derivable respecto a θ, la solución verificará el sistema

$$\sum_{i=1}^{n} \left(X_i - \varphi(t_i, \theta)\right) \frac{\partial}{\partial \theta_j} \varphi(t_i, \theta) = 0, \quad j = 1, \ldots, k.$$

Las propiedades de los estimadores obtenidos mediante el método de mínimos cuadrados dependen del problema particular analizado. Este método tendrá especial interés en el desarrollo del Tema 8.

Ejemplo: Para estimar la aceleración de la gravedad en una ciudad, θ, se deja caer un objeto durante tiempos $t_1, \ldots, t_n$ y se mide el espacio recorrido en cada tiempo. Si X_i representa la medida correspondiente al espacio recorrido en el tiempo t_i, con error de medida ϵ_i, estimar θ por el método de mínimos cuadrados. $(e = \frac{1}{2}\, at^2)$