Tema 4

ESTIMACIÓN PUNTUAL. INSESGADEZ Y MÍNIMA VARIANZA.

4.1. Planteamiento del problema

En el tema anterior se trataron aspectos relacionados con la reducción de los datos de una muestra en términos de un estadístico. En este tema se aborda el problema de estimación del parámetro de un modelo estadístico paramétrico (o de una función de dicho parámetro) mediante un estadístico conveniente.

En general se tiene una v.a. X con distribución en una familia de distribuciones paramétricas (es decir conocidas salvo por un parámetro), $X \rightsquigarrow F \in \{F_\theta,\ \theta \in \Theta\}$. El objetivo es determinar la distribución de la v.a. que se estudia, que en el caso paramétrico se reduce a conocer el parámetro, es decir, inferir el verdadero valor de θ, θ_0.

Para ello lo que se hace es coger una muestra, $X_1, \ldots, X_n$, m.a.s. de X, y en base a la información que proporciona la muestra se aproxima el valor de θ. Por lo tanto, el problema es escoger estadísticos, $T(X_1, \ldots, X_n)$, que para valores concretos de la muestra proporcionen buenas aproximaciones del parámetro θ.

Con dicho fin se van a escoger unos estadísticos particulares, que se denominan *estimadores*, de forma que cuando se sustituya la muestra aleatoria, $X_1, \ldots, X_n$, por sus observaciones, $T(x_1, \ldots, x_n)$ proporcionen una buena aproximación del parámetro desconocido.

En ocasiones en vez de estimar el parámetro θ, interesará estimar una *función paramétrica*, es decir, una transformación del parámetro, $g : \Theta \to \mathbb{R}$. En ese caso se buscará $g(\theta_0)$, el verdadero valor de $g(\theta)$, en vez de θ_0, el verdadero valor de θ.

Ejemplo: Sea $X \rightsquigarrow B(k_0, p)$ y $p \in (0, 1)$. A partir de una m.a.s. de X indicar algún estadístico que se pueda usar para inferir p y alguna función paramétrica que pueda ser de interés.

A la aproximación que se obtiene del parámetro a través de un estimador se la denomina estimación puntual porque lo que se obtiene al aplicarla es una valor concreto para el parámetro desconocido.

El parámetro desconocido puede ser unidimensional o multidimensional. En general se van a estudiar casos unidimensionales y, si es fácil, se generalizará al caso multidimensional.

4.1.1. Estimador

Sea X una v.a. con función de distribución en una familia de distribuciones paramétricas, $F \in \{F_\theta : \theta \in \Theta\}$ y $(X_1, \ldots, X_n)$ una m.a.s. de X.

Definición: Un estimador de θ es un estadístico, $T(X_1, \ldots, X_n)$, que toma valores en Θ.

$$T : \mathcal{X}^n \to \Theta.$$

Por tanto la diferencia entre un estimador y un estadístico es que el estimador es un estadístico que en lugar de estar definido sobre $\mathbb{R}^k$, se le exige que tome valores en el espacio paramétrico Θ, es decir, donde toma los valores el parámetro desconocido.

Si el estimador esta definido con espacio de llegada $g(\Theta)$, es decir, $T : \mathcal{X}^n \to g(\Theta)$, T es un estimador de la función paramétrica $g(\theta)$.

Para valores concretos de la muestra, $x_1, \ldots, x_n$, $T(x_1, \ldots, x_n)$ es una estimación puntual de θ o de $g(\theta)$, según el caso.

Ejemplos:

1. Sea $X_1, \ldots, X_n$ una m.a.s. de $X \rightsquigarrow \{P(\lambda), \lambda > 0\}$. $\bar{X}$ es un estimador de λ. Es más, cualquier función medible de la muestra que sea independiente del parámetro λ y tome valores positivos es un estimador del parámetro.

2. Sea $X_1, \ldots, X_n$ una m.a.s. de $X \rightsquigarrow \{B(1, p), p \in (0, 1)\}$. $\bar{X}$ es un estimador de p. Es más, cualquier función medible de la muestra que sea independiente del parámetro p y tome valores en el intervalo $(0, 1)$ es un estimador del parámetro.

La no unicidad del estimador de un parámetro plantea el problema de encontrar el mejor estimador. Para ello hay que establecer criterios de selección entre los estimadores para encontrar el mejor en algún sentido. Una opción es seleccionar el estimador en base a una función denominada de pérdida.

4.1.2. Función de pérdida y función de riesgo

Definición: A cualquier función $L : \Theta \times \Theta \to \mathbb{R}$, que verifique las siguientes propiedades:

(i) $L(\theta, T) \geq 0$, $\forall \theta \in \Theta, T \in \Theta$.

(ii) $L(\theta, T) = 0$, si $T = \theta$.

(iii) $L(\theta, T) \leq L(\theta, T')$ si la distancia de T a θ es menor que la distancia de T' a θ.

se la denomina *función de pérdida*.

$L(\theta, t)$ sería la pérdida que conlleva estimar el parámetro por el valor t si su verdadero valor es θ usando la función de pérdida L.

Ejemplos:

1. $L(\theta, T) = |\theta - T|$ (error absoluto de estimación).
2. $L(\theta, T) = (\theta - T)^2$ (error cuadrático de estimación).
3. $L(\theta, T) = \left| \dfrac{\theta - T}{\theta} \right|$ (error relativo de estimación).

Dado un estimador $T(X_1, \ldots, X_n)$ de θ, la función $L(\theta, T(X_1, \ldots, X_n))$, para cada $\theta \in \Theta$, es una variable aleatoria, siempre que L sea Borel-medible.

Definición: Se define la pérdida media o *función riesgo* de un estimador como la función (del parámetro θ) que asigna a cada valor del parámetro la pérdida media asociada al estimador bajo la función de pérdida L.

$$R_T^L(\theta) = E_\theta[L(\theta, T)].$$

En particular, la función riesgo de una función paramétrica $g(\theta)$ se definiría como:

$$R_{g,T}^L(\theta) = E_\theta[L(g(\theta), T)].$$

Definición: Se dice que un estimador $T(X_1, \ldots, X_n)$ es *óptimo bajo una función de pérdida* $L(\theta, T)$ si dicho estimador minimiza uniformemente la función de riesgo $R_T(\theta)$; esto es, un estimador, T, tal que, para cualquier otro, T' se tiene

$$R_T^L(\theta) \leq R_{T'}^L(\theta), \quad \forall \theta \in \Theta.$$

En general, el estimador óptimo no tiene porqué existir. Al no tener asegurada la existencia del estimador óptimo basado en la función de riesgo, el problema de estimación se puede reconsiderar mediante una de las dos siguientes vías:

1. Restringir la clase de estimadores imponiendo propiedades deseables de los mismos.
2. Introducir una nueva medida de la función de riesgo que permita ordenar totalmente la clase de todos los estimadores.

Estimación de menor error cuadrático medio (ECM)

Un criterio de comparación usual en múltiples ámbitos y aplicaciones de la estadística es el llamado criterio de menor *error cuadrático medio*. Se considera como función de pérdida:

$$L(\theta, T) = (\theta - T)^2$$

y como función de riesgo:

$$R_T(\theta) = E_\theta\left[(\theta - T)^2\right] = ECM_T(\theta).$$

Propiedades:

- El criterio del ECM tiene ventajas desde el punto de vista del manejo analítico, frente a otras funciones de riesgo.
- El ECM se interpreta como el grado de dispersión del estimador en torno al verdadero valor del parámetro, θ.
- El ECM puede descomponer en términos de la varianza y una función denominada sesgo:

  $$ECM_T(\theta) = Var_\theta(T) + B_T^2(\theta)$$

  donde $B_T(\theta)$ es la función denominada sesgo que se define como:

  $$B_T(\theta) = E_\theta[T] - \theta.$$

- Si el estimador considerado verifica la propiedad de insesgadez, es decir, $E_\theta[T] = \theta$ o $B_T(\theta) = 0$, se verifica que el ECM coincide con la varianza del estimador.

  $$ECM_T(\theta) = E_\theta\left[(\theta - T)^2\right] = Var_\theta(T).$$

4.2. Estimación insesgada de mínima varianza

Como se ha visto en la sección anterior, existe una relación sencilla e intuitiva que liga al ECM, la varianza y el sesgo de un estimador. En el problema de la búsqueda de un estimador óptimo, en algún sentido, se explota dicha relación dentro de la clase de estimadores que verifican ciertas propiedades, como la propiedad de insesgadez.

4.2.1. Estimador insesgado

Sea, como siempre a lo largo de este tema, X una v.a. con función de distribución en una familia de distribuciones paramétricas, $F \in \{F_\theta : \theta \in \Theta\}$ y $(X_1, \dots, X_n)$ una m.a.s. de X.

Definición: Un estimador $T(X_1, \dots, X_n)$ de θ es *insesgado* o centrado en el parámetro θ si su sesgo asociado es idénticamente nulo o, equivalentemente, si:

$$E_\theta[T(X_1, \dots, X_n)] = \theta, \quad \forall \theta \in \Theta.$$

Si el estimador $T^*(X_1, \dots, X_n)$ es de una función paramétrica de θ, $g(\theta)$, se dice que es insesgado en $g(\theta)$ si:

$$E_\theta[T^*(X_1, \dots, X_n)] = g(\theta), \quad \forall \theta \in \Theta.$$

Ejemplo: Sea $X_1, \dots, X_n$ una m.a.s. de alguna población. Probar que si existe la media de la población, $E_\theta X$, la media muestral es un estimador insesgado de la media poblacional, y si existe la varianza de la población, $Var_\theta X$, la cuasivarianza muestral es un estimador insesgado de ella.

Notas:

- Si $\theta = (\theta_1, \dots, \theta_k)$, un estimador insesgado de dicho vector de parámetros es un vector donde cada elemento del mismo es un estimador insesgado para cada parámetro θ_i componente de θ, es decir: Un estimador $T = (T_1, \dots, T_k)$ es insesgado en θ si se verifica

$$E_\theta[T_i] = \theta_i, \ \forall i = 1, \dots, k.$$

- Para un estimador insesgado se verifica que

$$ECM_T(\theta) = Var_\theta(T(X_1, \dots, X_n)), \quad \forall \theta \in \Theta$$

- Si T es un estimador insesgado de $\theta \not\Rightarrow h(T)$ es estimador insesgado de $h(\theta)$, siendo h cualquier función.

  Ejemplo: Sea $X_1, \dots, X_n$ una m.a.s. de alguna población. Probar que, en general, la cuasidesviación típica no es un estimador insesgado para $\sigma_\theta = \sqrt{Var_\theta X}$.

  Si h es una función lineal dicha implicación si se cumple, es decir, la insesgadez no se mantiene bajo transformaciones, en general, pero si se mantiene si la transformación es lineal.

- No tiene porque existir algún estimador insesgado de un parámetro.

  Ejemplos:

1. Sea $X_1, \ldots, X_n$, con $n \geq 2$, una m.a.s. de una distribución binomial $B(1, p)$. Probar que $\dfrac{T^2 - T}{n^2 - n}$ con $T = \sum_{i=1}^n X_i$ es un estimador insesgado para la función paramétrica $g(p) = p^2$. Probar, además, que para $n = 1$ no existe un estimador insesgado de p^2.

2. Sea X una v.a. con distribución $\mathcal{P}(\lambda)$, ¿existe algún estimador insesgado para la función paramétrica $1/\lambda$ basado en una muestra de tamaño 1?

- Un estimador insesgado no tiene porque ser único.

  Ejemplo: Sea X una v.a. con distribución $\mathcal{P}(\lambda)$ y $X_1, \ldots, X_n$ una m.a.s. de X. Probar que $\bar{X}$ y S^2 son estimadores insesgados de λ.

  Es más, si T_1 y T_2 son estimadores insesgados de θ, entonces $\alpha T_1 + (1 - \alpha)T_2$ con $0 \leq \alpha \leq 1$ es un estimador insesgado de θ.

Se plantea, por tanto, el problema de seleccionar un estimador óptimo dentro de la clase de estimadores insesgados. Nuestro criterio de búsqueda de los mejores estimadores será seleccionar los que tengan mínima varianza, ya que son los que están menos dispersos respecto a la media.

4.2.2. Estimador insesgado uniformemente de mínima varianza (UMVUE)

Definición: Sea X una v.a. con función de distribución en una familia de distribuciones paramétricas, $F \in \{F_\theta : \theta \in \Theta\}$ y $(X_1, \ldots, X_n)$ una m.a.s. de X. Un estimador de $g(\theta)$, $T(X_1, \ldots, X_n)$, insesgado y con momento de segundo orden finito, se dice que es un *estimador insesgado uniformemente de mínima varianza (“UMVUE”)* para $g(\theta)$ si para cualquier otro estimador de $g(\theta)$, $T'(X_1, \ldots, X_n)$, se tiene:

$$Var_\theta[T(X_1, \ldots, X_n)] \leq Var_\theta[T'(X_1, \ldots, X_n)] \quad \forall \theta \in \Theta.$$

Teorema (Unicidad del UMVUE): El UMVUE, si existe, es único, es decir, si hay dos UMVUEs son iguales con probabilidad 1.

Teorema (Linealidad del UMVUE): Si T_1 es UMVUE para una cierta función de θ, $g_1(\theta)$ y T_2 es UMVUE para otra cierta función de θ, $g_2(\theta)$ con $\theta \in \Theta$. Entonces:

1. λT_1 es UMVUE para $\lambda g_1(\theta)$ y λT_2 es UMVUE para $\lambda g_2(\theta)$,

2. $T_1 + T_2$ es UMVUE para $g_1(\theta) + g_2(\theta)$,

siendo λ cualquier valor real.

Una vez definido el UMVUE y estudiada algunas de sus propiedades, el siguiente paso es tener un método que permita su obtención de forma lo más sencilla posible. Para ello primero se estudian los Teoremas de Rao-Blackwell y Lehmann-Scheffé.

Teorema (Rao-Blackwell): Sea X una v.a. con función de distribución en una familia de distribuciones paramétricas, $F \in \{F_\theta : \theta \in \Theta\}$, y $(X_1, \ldots, X_n)$ una m.a.s. de X. Si $T(X_1, \ldots, X_n)$ un estadístico suficiente para la familia, $\{F_\theta : \theta \in \Theta\}$ y $S(X_1, \ldots, X_n)$ es un estimador insesgado de $g(\theta)$ con momento de segundo orden finito, entonces:

1. $E[S(X_1, \ldots, X_n)/T(X_1, \ldots, X_n)]$ es estimador insesgado de $g(\theta)$ y tiene momento de segundo orden finito.

2. $Var_\theta(E[S(X_1, \ldots, X_n)/T(X_1, \ldots, X_n)]) \leq Var_\theta(S(X_1, \ldots, X_n))$, $\forall \theta \in \Theta$.

Teorema (Lehmann-Scheffé): Sea $T(X_1, \ldots, X_n)$ es un estadístico suficiente y completo para la familia de distribuciones consideradas, $\{F_\theta : \theta \in \Theta\}$. Si $g(\theta)$ admite un estimador insesgado de segundo orden, $S(X_1, \ldots, X_n)$, entonces existe el UMVUE de $g(\theta)$, que viene dado por

$$E[S(X_1, \ldots, X_n)/T(X_1, \ldots, X_n)].$$

Métodos para el cálculo del UMVUE: Sea $T(X_1, \ldots, X_n)$ un estadístico suficiente y completo para $\{F_\theta : \theta \in \Theta\}$. El UMVUE para $g(\theta)$, si existe, se puede determinar mediante los dos siguientes procedimientos:

- Buscar cualquier estimador insesgado de $g(\theta)$ con momento de segundo orden finito, $S(X_1, \ldots, X_n)$. Entonces $E[S(X_1, \ldots, X_n)/T(X_1, \ldots, X_n)]$ es el UMVUE.

- Buscar una función $h(T)$, siendo T un estadístico suficiente y completo, tal que $E_\theta[h(T)] = g(\theta)$ $\forall \theta \in \Theta$, es decir, que sea insesgada en $g(\theta)$, que sea un estimador y que tenga momento de segundo orden finito. Entonces $E[h(T(X_1, \ldots, X_n))/T(X_1, \ldots, X_n)] = h(T(X_1, \ldots, X_n))$ es el UMVUE.

Si no existiera un estadístico suficiente y completo, esto no implica que no exista UMVUE, sólo que habrá que calcularlo de otra forma.

Ejemplos

1. Sea $X_1, \ldots, X_n$ una m.a.s. con distribución $B(1, p)$, $p \in (0, 1)$. Encontrar el UMVUE para p.

2. Sea $X_1, \ldots, X_n$ una m.a.s. con distribución $\mathcal{U}(0, \theta)$, $\theta \in \mathbb{R}^+$. Encontrar el UMVUE para θ y $1/\theta$.

3. Sea $X_1, \ldots, X_n$ una m.a.s. con distribución $P(\lambda)$, $\lambda \in \mathbb{R}^+$. Encontrar, si existe, el UMVUE para la función paramétrica λ^s, $s \in \mathbb{N}$.

4.3. Estimadores eficientes

En esta sección del tema se va a estudiar el concepto de estimador eficiente con idea de, posteriormente, estudiar la relación que existe entre dichos estimadores y el UMVUE. En toda esta sección consideraremos un espacio paramétrico unidimensional para poder trabajar con el concepto de varianza en lugar del de matriz de varianzas-covarianzas.

Para dar la definición de estimador eficiente, previamente, se van a estudiar condiciones de regularidad.

4.3.1. Condiciones de reguralidad de Fréchet-Crámer-Rao

Sea X una v.a. con distribución de probabilidad en la familia de distribuciones $\{F_\theta : \theta \in \Theta\}$. Sea $f_\theta(x)$ la fdp o fmp, según el caso, para cada $\theta \in \Theta \subseteq \mathbb{R}$. Se dice que esta familia de distribuciones cumple las *condiciones de regularidad* si:

(i) Θ es un intervalo abierto de $\mathbb{R}$.

(ii) El conjunto de valores de la variable, $\{x : f_\theta(x) > 0\} = \mathcal{X}$, es independiente de θ.

(iii) $\forall x \in \chi$, $f_\theta(x)$ es derivable respecto de θ y se verifica que

$$\int_{\mathcal{X}} \frac{\partial}{\partial\theta} f_\theta(x)dx = \frac{\partial}{\partial\theta}\int_{\mathcal{X}} f_\theta(x)dx = 0 \quad \left(\sum_{\mathcal{X}} \frac{\partial}{\partial\theta} f_\theta(x) = \frac{\partial}{\partial\theta}\sum_{\mathcal{X}} f_\theta(x) = 0\right). \quad \forall\theta \in \Theta.$$

Esta tercera condición es equivalente a comprobar que

$$E_\theta\left[\frac{\partial \ln f_\theta(x)}{\partial\theta}\right] = 0.$$

Teorema: Si $X \rightsquigarrow F \in \{F_\theta : \theta \in \Theta\}$ cumple las condiciones de reguralidad, entonces la familia de distribuciones asociada a la m.a.s. $(X_1, \dots, X_n)$ de X también las cumple.

4.3.2. Función de información de Fisher

Definición: Sea $X \rightsquigarrow F \in \{F_\theta : \theta \in \Theta\}$, cuya familia de distribuciones es regular. Se definen las funciones

$$I_X(\theta) = E_\theta\left[\left(\frac{\partial \ln f_\theta(X)}{\partial\theta}\right)^2\right], \quad I_{(X_1,\dots X_n)}(\theta) = E_\theta\left[\left(\frac{\partial \ln f_\theta^n(X_1,\dots,X_n)}{\partial\theta}\right)^2\right]$$

que se denominan *función de información de Fischer* asociada a X y a la muestra, respectivamente.

Propiedades:

(i) $I_X \geq 0$.

(ii) $E_\theta \left[\dfrac{\partial \ln f_\theta(X)}{\partial \theta} \right] = 0$ y $Var_\theta \left[\dfrac{\partial \ln f_\theta(X)}{\partial \theta} \right] = I_X(\theta), \quad \forall \theta \in \Theta$.

(iii) $E_\theta \left[\dfrac{\partial \ln f_\theta^n(X_1, \ldots, X_n)}{\partial \theta} \right] = 0$ y $Var_\theta \left[\dfrac{\partial \ln f_\theta^n(X_1, \ldots, X_n)}{\partial \theta} \right] = I_{(X_1,\ldots,X_n)}(\theta), \quad \forall \theta \in \Theta$.

(iv) $I_{X_1,\ldots,X_n}(\theta) = nI_X(\theta), \quad \forall \theta \in \Theta$. (Aditividad)

4.3.3. Desigualdad de Fréchet-Cramér-Rao

Definición: Sea $(X_1, \ldots, X_n)$ una m.a.s. de $X \rightsquigarrow F \in \{F_\theta : \theta \in \Theta\}$, cuya familia de distribuciones es regular. Un estadístico $T(X_1, \ldots, X_n)$ se dice que es *regular* en el sentido de Fréchet-Cramér-Rao si verifica:

- Caso discreto

$$\frac{\partial}{\partial \theta} \sum_{(x_1,\ldots,x_n)\in \mathcal{X}^n} T(x_1, \ldots, x_n) f_\theta^n(x_1, \ldots, x_n) = \sum_{(x_1,\ldots,x_n)\in \mathcal{X}^n} T(x_1, \ldots, x_n) \frac{\partial f_\theta^n(x_1, \ldots, x_n)}{\partial \theta}$$

- Caso continuo

$$\frac{\partial}{\partial \theta} \int_{\mathcal{X}^n} T(x_1, \ldots, x_n) f_\theta^n(x_1, \ldots, x_n) dx_1 \cdots dx_n = \int_{\mathcal{X}^n} T(x_1, \ldots, x_n) \frac{\partial f_\theta^n(x_1, \ldots, x_n)}{\partial \theta} dx_1 \cdots dx_n$$

Es decir,

$$\frac{\partial}{\partial \theta} E_\theta \left[T(X_1, \ldots, X_n) \right] = E_\theta [T(X_1, \ldots, X_n) \frac{\partial \ln f_\theta^n(X_1, \ldots, X_n)}{\partial \theta}]$$

Teorema (cota de Fréchet-Cramér-Rao): Sea $(X_1, \ldots, X_n)$ una m.a.s. de $X \rightsquigarrow F \in \{F_\theta : \theta \in \Theta\}$, cuya familia de distribuciones es regular con $0 < I_X(\theta) < +\infty \ \forall \theta \in \Theta$. Si $T(X_1, \ldots, X_n)$ es un estadístico regular, de segundo orden e insesgado en una función paramétrica derivable $g(\theta)$, entonces se tiene

(i) $Var_\theta T(X_1, \ldots, X_n) \geq \dfrac{(g'(\theta))^2}{I_{X_1,\ldots,X_n}(\theta)} \quad \forall \theta \in \Theta$.

(ii) Para los puntos θ_0 tales que $g'(\theta_0) \neq 0$, se dará la igualdad si y sólo si $\exists a(\theta_0) \neq 0$ tal que

$$P_{\theta_0} \left[\left. \frac{\partial \ln f_\theta(X_1, \ldots, X_n)}{\partial \theta} \right|_{\theta=\theta_0} = a(\theta_0) \left[T(X_1, \ldots, X_n) - g(\theta_0) \right] \right] = 1$$

4.3.4. Estimador eficiente

Definición: Sea $\{F_\theta : \theta \in \Theta\}$ regular, $0 < I_X(\theta) < +\infty$ $\forall\theta \in \Theta$ y $g(\theta)$ una función paramétrica derivable. Un estimador de $g(\theta)$, $T(X_1, \ldots, X_n)$, se dice que es eficiente si es regular, insesgado y su varianza alcanza la cota de FCR para cualquier valor del parámetro, es decir,

$$Var_\theta(T(X_1, \ldots, X_n)) = \frac{[g'(\theta)]^2}{I_{X_1,\ldots,X_n}(\theta)}, \quad \forall\theta \in \Theta.$$

El estimador eficiente no tiene porque existir.

Lema: Sea $\{F_\theta : \theta \in \Theta\}$ regular, $0 < I_X(\theta) < +\infty$ $\forall\theta \in \Theta$ y $g(\theta)$ una función paramétrica derivable. Entonces $g(\theta)$ admite un estimador eficiente $T(X_1, \ldots, X_n)$ si:

- $g(\theta)$ es constante y en tal caso $T(X_1, \ldots, X_n)$ es degenerado.
- $g(\theta)$ es estrictamente monótona: $g'(\theta) > 0$ $\forall\theta \in \Theta$ o $g'(\theta) < 0$ $\forall\theta \in \Theta$.

Teorema (caracterización de estimadores eficientes): Sea $(X_1, \ldots, X_n)$ una m.a.s. de $X \rightsquigarrow F \in \{F_\theta : \theta \in \Theta\}$ regular, con $0 < I_X(\theta) < \infty$ $\forall\theta \in \Theta$, $g(\theta)$ una función paramétrica derivable, con $g'(\theta) \neq 0$, $\forall\theta \in \Theta$ y $T(X_1, \ldots, X_n)$ es un estimador de $g(\theta)$. Una condición necesaria y suficiente para que T sea eficiente es:

$$\forall\theta \in \Theta, \exists a(\theta) \neq 0 \text{ tal que} \begin{cases} i)\ P_\theta \left[\dfrac{\partial \ln f_\theta^n(X_1, \ldots, X_n)}{\partial\theta} = a(\theta)[T(X_1, \ldots, X_n) - g(\theta)]\right] = 1 \\ ii)\ I_{(X_1,\ldots,X_n)}(\theta) = a(\theta)g'(\theta). \end{cases}$$

Corolario 1: Si $T(X_1, \ldots, X_n)$ es un estimador eficiente para $g(\theta)$, con $g'(\theta) \neq 0$, las únicas funciones paramétricas que admiten estimadores eficientes son las de la forma $ag(\theta) + b$ y los correspondientes estimadores eficientes son $aT + b$, con probabilidad 1, bajo todas las distribuciones de la familia.

Corolario 2: Si una función paramétrica admite dos estimadores eficientes, estos son iguales con probabilidad 1, bajo todas las distribuciones de la familia.

Corolario 3: Sólo existen estimadores eficientes en familias de tipo exponencial.

Corolario 4: Si $T(X_1, \ldots, X_n)$ es eficiente para $g(\theta)$, entonces $T(X_1, \ldots, X_n)$ es suficiente. Si además la imagen de $Q(\theta) = \int a(\theta)\, d\theta$ contiene a un abierto de $\mathbb{R}$, entonces $T(X_1, \ldots, X_n)$ es completo.

Corolario 5: Si $T(X_1, \ldots, X_n)$ es eficiente para $g(\theta)$, suficiente y completo, entonces $T(X_1, \ldots, X_n)$ es el UMVUE de $g(\theta)$.

El recíproco no es cierto, es decir, si $T(X_1,\ldots,X_n)$ es el UMVUE para $g(\theta)$ eso no implica que $T(X_1,\ldots,X_n)$ sea eficiente para dicha función paramétrica.

Ejemplo: Buscar la clase de funciones paramétricas que admiten estimadores eficientes para las siguientes familias de distribuciones y calcular dichos estimadores:

1. $\{B(k_0,p),\ p\in(0,1)\}$,
2. $\{\mathcal{N}(\mu,\sigma_0^2),\ \mu\in\mathbb{R}\}$,
3. $\{\mathcal{N}(\mu_0,\sigma^2),\ \sigma^2>0\}$.

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:

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

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,

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

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. ($e = \frac{1}{2}\, at^2$)