tema5teorIE.pdf

(Provisional) Apuntes tema 5

3º Inferencia Estadística

Grado en Matemáticas

Facultad de Ciencias
Universidad de Granada

Tema 5: Estimación de máxima verosimilitud y otros métodos. [1]

Índice

En este tema se completan los métodos de estimación puntual vistos en el tema 4. Por un lado, se introduce la propiedad de máxima verosimilitud y se estudia su relación con las propiedades de suficiencia y eficiencia. Por otro, se añaden dos métodos de estimación específicos por el tipo de parámetros que estiman o por el tipo de problemas en que se emplean.

[1] Si encuentras algún error en el documento, escribe un correo electrónico a *melferizq@gmail.com*.

1. Estimación de máxima verosimilitud.

El método de estimación de máxima verosimilitud es un método de cálculo de estimadores basado en el llamado *principio de la máxima verosimilitud*:

Principio de la máxima verosimilitud.

Si al observar una muestra aleatoria simple de una variable aleatoria X con distribución teórica F_θ se obtienen los valores $(x_1, \dots, x_n)$ cuya probabilidad es $P_\theta(x_1, \dots, x_n)$, debe estimarse θ mediante el valor que maximiza dicha probabilidad.

Ejemplo 1. *Una urna contiene 6 bolas, entre blancas y negras, no todas del mismo color, pero se ignora cuántas hay de cada uno. Decidir cuál es el estimador máximo verosímil para la proporción de bolas blancas en la urna a partir de la extracción de dos bolas con reemplazamiento.*

Solución. La proporción total de bolas blancas puede ser $p = i/6$, con $i = 1, \dots, 5$. Para tratar de saber la composición de la urna se extraen dos bolas con reemplazamiento. De lo que se trata en este caso, al estar ante un problema de máxima verosimilitud, es de maximizar la probabilidad de que se obtenga cada uno de los posibles valores de la variable aleatoria T = *número de bolas blancas extraídas*, con $T = 0, 1, 2$. Así, calculamos la probabilidad de cada resultado en función del valor del parámetro p.

$$P[T=0] = \begin{cases} \frac{5}{6}\cdot\frac{5}{6} = 0{,}694 & \text{si } p = \frac{1}{6} \\ \frac{4}{6}\cdot\frac{4}{6} = 0{,}444 & \text{si } p = \frac{2}{6} \\ \frac{3}{6}\cdot\frac{3}{6} = 0{,}25 & \text{si } p = \frac{3}{6} \\ \frac{2}{6}\cdot\frac{2}{6} = 0{,}111 & \text{si } p = \frac{4}{6} \\ \frac{1}{6}\cdot\frac{1}{6} = 0{,}027 & \text{si } p = \frac{5}{6} \end{cases} \qquad P[T=1] = \begin{cases} 2\cdot\frac{1}{6}\cdot\frac{5}{6} = 0{,}277 & \text{si } p = \frac{1}{6} \\ 2\cdot\frac{2}{6}\cdot\frac{4}{6} = 0{,}444 & \text{si } p = \frac{2}{6} \\ 2\cdot\frac{3}{6}\cdot\frac{3}{6} = 0{,}5 & \text{si } p = \frac{3}{6} \\ 2\cdot\frac{4}{6}\cdot\frac{2}{6} = 0{,}444 & \text{si } p = \frac{4}{6} \\ 2\cdot\frac{5}{6}\cdot\frac{1}{6} = 0{,}277 & \text{si } p = \frac{5}{6} \end{cases}$$

$$P[T=2] = \begin{cases} \frac{1}{6}\cdot\frac{1}{6} = 0{,}027 & \text{si } p = \frac{1}{6} \\ \frac{2}{6}\cdot\frac{2}{6} = 0{,}111 & \text{si } p = \frac{2}{6} \\ \frac{3}{6}\cdot\frac{3}{6} = 0{,}25 & \text{si } p = \frac{3}{6} \\ \frac{4}{6}\cdot\frac{4}{6} = 0{,}444 & \text{si } p = \frac{4}{6} \\ \frac{5}{6}\cdot\frac{5}{6} = 0{,}694 & \text{si } p = \frac{5}{6} \end{cases}$$

Si se obtienen $T = 0$ bolas blancas, podría ser que la proporción de bolas blancas en la urna fuese $p = 4/6$ y que se haya producido un suceso de probabilidad 0,111, pero es más verosímil que se tenga $p = 2/6$ y que haya tenido lugar un suceso de probabilidad 0,444, y todavía es más verosímil que sea $p = 1/6$ y que se haya producido un suceso de probabilidad 0,694. Por tanto, si $T = 0$, el p que hace que el suceso tenga una máxima verosimilitud es $p = 1/6$. Por otro lado, si se obtiene $T = 1$, es decir, hay una bola blanca en la muestra, entonces el valor de p que da una mayor probabilidad al suceso observado es $p = 1/2$, de la misma forma que si $T = 2$, el valor de p que le otorga una mayor verosimilitud al suceso es $p = 5/6$. Así, el estimador máximo verosímil para la proporción de bolas blancas en la urna obtenido a partir de la extracción de dos bolas con reemplazamiento es:

$$p(T) = \begin{cases} 1/6 & \text{si } T = 0 \\ 1/2 & \text{si } T = 1 \\ 5/6 & \text{si } T = 2 \end{cases}$$

Ejemplo 2. *Sea X una variable aleatoria con distribución $B(n,p)$, con $n \in \{2,3\}, p \in \{1/2, 1/3\}$. Basándose en la observación de un valor de la variable, decidir cuáles de estos valores corresponden a la variable bajo estudio.*

Solución. De nuevo, al igual que en el ejemplo anterior, estudiamos para qué valores de n y p es más probable que se obtengan los valores $X = 0$, $X = 1$, $X = 2$, $X = 3$.

X	$(2,1/2)$	$(2,1/3)$	$(3,1/2)$	$(3,1/3)$
$P[X=0]$	$1/4$	$4/9$	$1/8$	$8/27$
$P[X=1]$	$1/2$	$4/9$	$3/8$	$12/27$
$P[X=2]$	$1/4$	$1/9$	$3/8$	$6/27$
$P[X=3]$	0	0	$1/8$	$1/27$

Así, el estimador máximo verosímil para los parámetros n y p en este caso es:

$$(n,p)(X) = \begin{cases} (2,1/3) & \text{si } X = 0 \\ (2,1/2) & \text{si } X = 1 \\ (3,1/2) & \text{si } X = 2, X = 3 \end{cases}$$

Para hacer óptimo el cálculo del estimador máximo verosímil, debemos llegar hasta su definición y obtener un método que nos permita calcularlo. Así, es necesario plantearnos cómo se puede medir la probabilidad de que se obtenga cada una de las posibles realizaciones muestrales de una muestra aleatoria simple y, en tal caso, cómo esta podría maximizarse.

1.1. Estimadores máximo verosímiles.

El principio de máxima verosimilitud dice que debemos maximizar la probabilidad de que se tenga una realización muestral concreta, y la probabilidad de que se obtenga cada una de ellas de entre todas las posibles viene dada por la función masa de probabilidad muestral o por la función de densidad muestral, según el caso. Así, una función que nos diga la verosimilitud de cada realización se deberá definir a partir de la función de probabilidad muestral de una variable aleatoria X, a la que llamamos $f_\theta^n(x_1, \dots, x_n)$.

En la función de probabilidad muestral $f_\theta^n(x_1, \dots, x_n)$ con la que hemos trabajado hasta ahora, el valor de θ siempre ha sido fijo y se obtenía su estimación a partir de variaciones de las realizaciones muestrales $(x_1, \dots, x_n)$ a las que se les aplicaba un cierto estadístico T del cual se obtenía un valor concreto para θ. Ahora se desea obtener el valor de θ haciéndolo variar en Θ hasta llegar al máximo valor de $f_\theta^n(x_1, \dots, x_n)$ para una realización muestral $(x_1, \dots, x_n)$ concreta. Por tanto, podemos definir la función de verosimilitud como la función de probabilidad muestral con la realización muestral $(x_1, \dots, x_n)$ fija y θ variable.

Función de verosimilitud.

Definición 1. *Sea X una variable aleatoria con distribución en una familia paramétrica de distribuciones, $\{F_\theta, \theta \in \Theta\}$. Sea $f_\theta(x)$ la f.m.p. o la f.d.d. de X, según el caso. Se considera $(X_1, \dots, X_n)$ una muestra aleatoria simple de X, y sea $f_\theta^n(x_1, \dots, x_n)$ su f.m.p. o su f.d.d. conjunta, con $\theta \in \Theta$. Para cada realización muestral $(x_1, \dots, x_n)$ se define la función de verosimilitud asociada a dichos valores de la muestra como una función de θ de la siguiente manera:*

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

Estimador de máxima verosimilitud (EMV).

Definición 2. *Sea X una variable aleatoria con distribución en una familia paramétrica de distribuciones, $\{F_\theta, \theta \in \Theta\}$. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X. Un estimador $\widehat{\theta}(X_1, \ldots, X_n)$ de θ es estimador de máxima verosimilitud (EMV) de θ si:*

$$\forall (x_1, \ldots, x_n) \in \chi^n, \quad L_{x_1,\ldots,x_n}(\widehat{\theta}(x_1, \ldots, x_n)) = \max_{\theta \in \Theta} L_{x_1,\ldots,x_n}(\theta)$$

Por tanto, el método de obtención de estimadores máximo verosímiles consiste en obtener un estimador que maximice la función de verosimilitud (o una transformación suya conveniente).

En efecto, si tal función es derivable, el estimador máximo verosímil se calcula resolviendo las ecuaciones de verosimilitud, que se obtienen al derivarla con respecto a cada uno de los parámetros e igualando las expresiones resultantes a cero. Sin embargo, lo más habitual es considerar su logaritmo neperiano, es decir, $\ln L_{x_1,\ldots,x_n}(\theta)$, ya que muchas distribuciones tienen exponenciales en sus expresiones, y como el logaritmo neperiano es una función creciente, no afecta al cálculo del máximo.

Así, si $\ln L_{x_1,\ldots,x_n}(\theta)$ es derivable con respecto a θ, 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 $\ln L_{x_1,\ldots,x_n}(\theta)$ y, por tanto, los posibles extremos de $L_{x_1,\ldots,x_n}(\theta)$, que pueden ser máximos o no. Si la solución es única y es un máximo, entonces es el EMV de θ y lo llamamos $\widehat{\theta}$. Si existen varias soluciones se puede tomar el máximo absoluto entre ellas como $\widehat{\theta}$. Una vez obtenida la solución se debe comprobar que, en efecto, se trata de un estimador.

Si la función de verosimilitud no es derivable, entonces hay que recurrir a otro tipo de métodos, como la representación de $L_{x_1,\ldots,x_n}(\theta)$ o, incluso, métodos numéricos, para obtener el máximo.

Ejemplo 3. *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de $X \rightsquigarrow \{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$. Encontrar el EMV para μ y σ^2 en el caso de un parámetro conocido y cuando ámbos parámetros son desconocidos.*

Solución.

- Suponemos μ desconocido y σ_0^2 conocido. Así:

$$f_\mu^n(x_1, \ldots, x_n) = \left(\frac{1}{\sqrt{2\pi\sigma_0^2}}\right)^n \exp\left(\frac{-\sum_{i=1}^n (x_i - \mu)^2}{2\sigma_0^2}\right) = L_{x_1,\ldots,x_n}(\mu)$$

Por tanto:

$$\ln L_{x_1,\ldots,x_n}(\mu) = \frac{-n}{2} \ln(2\pi\sigma_0^2) - \frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma_0^2}$$

Así:

$$\frac{\partial \ln L_{x_1,\ldots,x_n}(\mu)}{\partial \mu} = \sum_{i=1}^n (x_i - \mu) = 0 \Leftrightarrow \sum_{i=1}^n x_i - n\mu \Leftrightarrow \mu = \frac{1}{n}\sum_{i=1}^n x_i = \bar{X}$$

Se deduce entonces que $\bar{X}$ es candidato a EMV.

  - $\bar{X}$ es estimador de μ puesto que, igual que μ, puede tomar cualquier valor real.

- Puesto que $(L_{x_1,\dots,x_n}(\mu))'' = -n < 0$, entonces $\bar{X}$ es un máximo.

Por tanto, $\widehat{\mu} = \bar{X}$ es EMV de μ.

- Suponemos μ_0 conocido y σ^2 desconocido. En este caso:

$$\ln L_{x_1,\dots,x_n}(\sigma^2) = \frac{-n}{2}\ln(2\pi\sigma^2) - \frac{\sum_{i=1}^n (x_i-\mu_0)^2}{2\sigma^2}$$

Así:

$$\frac{\partial \ln L_{x_1,\dots,x_n}(\sigma^2)}{\partial\sigma^2} = \frac{(-n/2)\cancel{2\pi}}{\cancel{2\pi}\sigma^2} + \frac{\sum_{i=1}^n (x_i-\mu_0)^2\cdot 2}{4\sigma^4} = \frac{-n\sigma^2 + \sum_{i=1}^n (x_i-\mu_0)^2}{2\sigma^4} = 0$$
$$\Leftrightarrow -n\sigma^2 + \sum_{i=1}^n (x_i-\mu_0)^2 = 0 \Leftrightarrow \sum_{i=1}^n (x_i-\mu_0)^2 = n\sigma^2 \Leftrightarrow \sigma^2 = \frac{\sum_{i=1}^n (x_i-\mu_0)^2}{n}$$

Por tanto, $\widehat{\sigma^2} = \dfrac{\sum_{i=1}^n (X_i-\mu_0)^2}{n}$ es candidato a EMV de σ^2.

- $\widehat{\sigma^2}$ es estimador de σ^2 pues ambos solo pueden tomar valores positivos.
- Para hallar la segunda derivada de $L_{x_1,\dots,x_n}(\sigma^2)$ hacemos el cambio $\sigma^2 = x$. Así:

$$\frac{\partial^2 \ln L_{x_1,\dots,x_n}(x)}{\partial x^2} = \frac{-n\cdot 2x^2 - (nx + \sum_{i=1}^n (x_i-\mu_0)^2)\cdot 4x}{4x^4} = \frac{-1}{x^2} - \frac{nx + \sum_{i=1}^n (x_i-\mu_0)^2}{x^3}$$
$$= \underbrace{\frac{-1}{\sigma^6}}_{<0} \underbrace{-\frac{n\sigma^2 + \sum_{i=1}^n (x_i-\mu_0)^2}{\sigma^6}}_{<0} < 0$$

Por tanto, hemos demostrado que $\widehat{\sigma^2} = \dfrac{\sum_{i=1}^n (X_i-\mu_0)^2}{n}$ es EMV de σ^2.

- Para el caso en que μ y σ^2 son desconocidos, con un proceso similar se llega a que los estimadores máximo verosímiles para μ y σ^2 son, respectivamente:

$$\widehat{\mu} = \bar{X}, \qquad \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$$

Ejemplo 4. *Sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de X una variable aleatoria discreta en N puntos. Encontrar el EMV de N.*

Solución. Si X es una variable aleatoria discreta en N puntos, entonces su función masa de probabilidad viene dada por

$$P[X=x] = \frac{1}{N}, x = 1,\dots,N$$

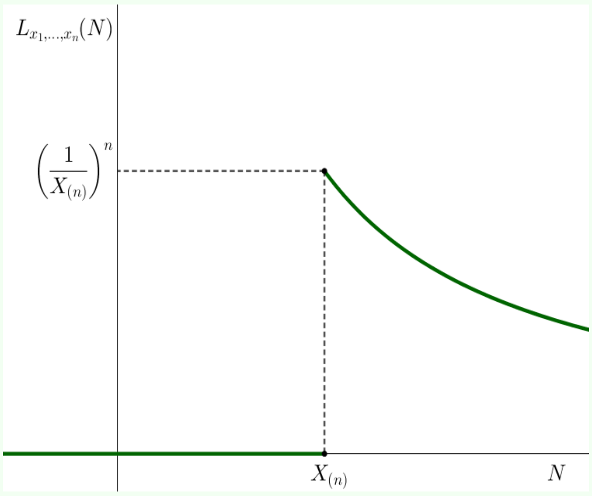
Si consideramos una muestra aleatoria simple $(X_1,\dots,X_n)$ de X, la función masa de probabilidad muestral será no nula si todos los valores X_i se encuentran entre 1 y N, es decir, si $I_{[X_{(1)}\geq 1]}$ y si $I_{[X_{(n)}\leq N]}$. Asumiendo que la condición $I_{[X_{(1)}\geq 1]}$ se cumple siempre, la función masa de probabilidad muestral de X viene dada por

$$f_N^n(x_1,\ldots,x_n) = \frac{1}{N^n} I_{[X_{(n)}\leq N]} = \frac{1}{N^n} I_{[N\geq X_{(n)}]} = L_{x_1,\ldots,x_n}(N)$$

donde en la función de verosimilitud escribimos el parámetro desconocido N en función del estadístico. En este, caso, $L_{x_1,\ldots,x_n}(N)$ no es derivable puesto que es el producto de dos funciones, una de ellas la función indicadora, que no es derivable. Por tanto, para hallar su máximo, una buena opción será representarla.

- Si $N < X_{(n)}$, entonces $L_{x_1,\ldots,x_n}(N)$ valdrá 0 puesto que, en tal caso, $I_{[N\geq I_{X_{(n)}}]} = 0$.
- Cuando $N \geq X_{(n)}$, $L_{x_1,\ldots,x_n}(N)$ es no nula pues $I_{[N\geq I_{X_{(n)}}]}$ será siempre 1, y $\frac{1}{N^n} \neq 0$.

El máximo valor de $L_{x_1,\ldots,x_n}(N)$ se da cuando $N = X_{(n)}$ ya que es el primer valor que hace que la función de verosimilitud sea no nula y, si N sigue creciendo, el valor de $L_{x_1,\ldots,x_n}(N)$ irá decreciendo puesto que la función $\frac{1}{N^n}$ es decreciente.



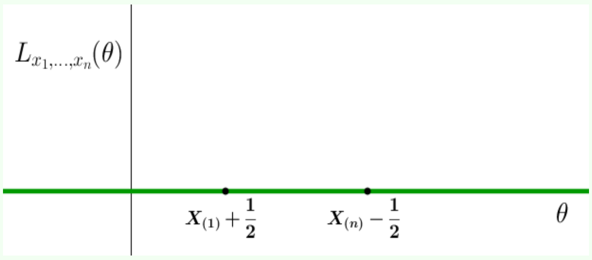
Ejemplo 5. *Sea $(X_1,\ldots,X_n)$ una muestra aleatoria simple de $X \rightsquigarrow \{U(\theta-1/2,\theta+1/2),\theta\in\mathbb{R}\}$. Encontrar el EMV para θ.*

Solución. Puesto que para una distribución uniforme continua en el intervalo $(\theta-1/2,\theta+1/2)$ su función de densidad es $f_\theta(x) = 1, \theta-1/2 < x < \theta+1/2$, entonces:

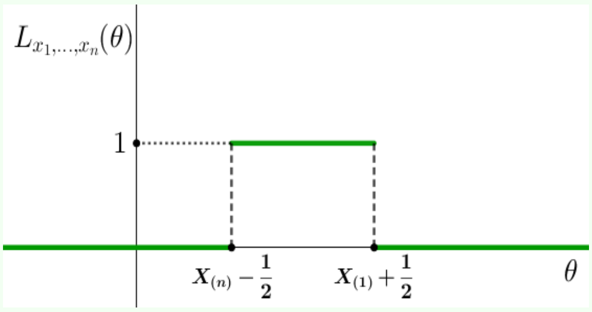
$$f_\theta^n(x_1,\ldots,x_n) = 1\cdot I_{[X_{(1)}>\theta-1/2]}\cdot I_{[X_{(n)}<\theta+1/2]} = 1\cdot I_{[\theta<X_{(1)}+1/2]}\cdot I_{[\theta>X_{(n)}-1/2]} = L_{x_1,\ldots,x_n}(\theta)$$

Así, tenemos dos formas de representar $L_{x_1,\ldots,x_n}(\theta)$, en función del orden de los valores $X_{(1)}+1/2$ y $X_{(n)}-1/2$.

- Si $X_{(1)}+\frac{1}{2} < X_{(n)}-\frac{1}{2}$, entonces si $\theta < X_{(1)}+\frac{1}{2}$ no se puede dar la condición $\theta > X_{(n)}-\frac{1}{2}$, y si se da la condición $\theta \geq X_{(n)}-\frac{1}{2}$, entonces no se cumple $\theta \leq X_{(1)}+\frac{1}{2}$. Por tanto, alguna de las funciones indicadoras $I_{[\theta<X_{(1)}+1/2]}$ o $I_{[\theta>X_{(n)}-1/2]}$ será 0. Por tanto, $L_{x_1,\ldots,x_n}(\theta) = 0$, $\forall\theta\in\Theta$.



- Si $X_{(n)} - \frac{1}{2} < X_{(1)} + \frac{1}{2}$, entonces se cumple $\theta < X_{(1)} + \frac{1}{2}$ y $\theta > X_{(n)} - \frac{1}{2}$ si y solo si $\theta \in \left(X_{(n)} - \frac{1}{2}, X_{(1)} + \frac{1}{2}\right)$, en cuyo caso las funciones indicadores valen $I_{[\theta < X_{(1)}+1/2]} = 1$ e $I_{[\theta > X_{(n)}-1/2]} = 1$ y, por tanto, $L_{x_1,\ldots,x_n}(\theta) = 1$.



Así, cualquier estadístico entre $X_{(n)} - \dfrac{1}{2}$ y $X_{(1)} + \dfrac{1}{2}$ será EMV de θ (sin incluirlos a ellos).

Observación 1. *El estimador máximo verosímil de un parámetro no tiene por qué ser único (el Ejemplo 5 lo demuestra.)*

Ejemplo 6. *Sea X una variable aleatoria con función de 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, $E[\widehat{p} - p]^2$. Comprobar que el estimador $T(X) = 1/2$ es mejor que $\widehat{p}$ en el sentido del menor error cuadrático medio.*

Solución. Si $X \rightsquigarrow B(1,p)$ se tiene $P[X = x] = p^x(1-p)^{n-x}, x = 0, 1$.

a) Como hicimos en los ejemplos 1 y 2, debemos hallar el EMV para $P[X = 0]$ y $P[X = 1]$.
   - $P[X = 0] = 1 - p$, que es una función decreciente para $p \in [1/4, 3/4]$. Por tanto, tal probabilidad es mayor cuando $p = 1/4$.
   - $P[X = 1] = p$, que es una función creciente para $p \in [1/4, 3/4]$. Así, tal probabilidad es mayor cuando $p = 3/4$.

   Por tanto, el estimador máximo verosímil para p es

$$\widehat{p}(X) = \begin{cases} 1/4 & \text{si } x = 0 \\ 3/4 & \text{si } x = 1 \end{cases}$$

b) Comprobemos que $\widehat{p}(X)$ no es insesgado.

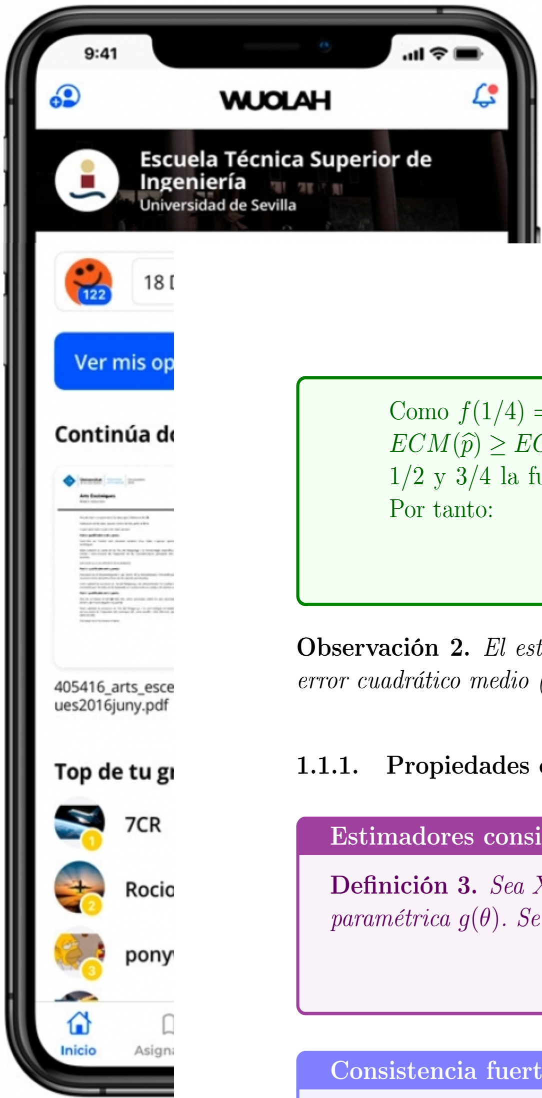
$$\mathrm{E}[\widehat{p}(X)] = \frac{1}{4}P[X = 0] + \frac{3}{4}P[X = 1] = \frac{1}{4}(1-p) + \frac{3}{4}p = \frac{1}{2}p + \frac{1}{4} \neq p \text{ (en general)}$$

c) Calculamos el error cuadrático medio de $\widehat{p}$.

$$E[(\widehat{p} - p)^2] = \left(\frac{1}{4} - p\right)^2 \underbrace{P[X = 0]}_{1-p} + \left(\frac{3}{4} - p\right)^2 \underbrace{P[X = 1]}_{p} = \frac{1}{16} = ECM(\widehat{p})$$

d) Sea $T(X) = 1/2$. Calculamos su error cuadrático medio.

$$ECM(T) = \mathrm{E}[(T(X) - p)^2] = \left(\frac{1}{2} - p\right)^2 = p^2 - p + \frac{1}{4} = f(p)$$



Como $f(1/4) = 1/16$ y entre $1/4$ y $1/2$ la función $f(p)$ es decreciente, entonces se tiene que $ECM(\widehat{p}) \geq ECM(T)$ con $p \in (1/4, 1/2)$. Igualmente, como $f(1/2) = 0$, $f(3/4) = 1/16$ y entre $1/2$ y $3/4$ la función $f(p)$ es creciente, entonces $ECM(\widehat{p}) \geq ECM(T)$ para $p \in (1/2, 3/4)$. Por tanto:

$$ECM(T) \leq ECM(\widehat{p}), \ \forall p \in \left[\frac{1}{4}, \frac{3}{4}\right]$$

Observación 2. *El estimador máximo verosímil no tiene por qué ser el mejor en el sentido del menor error cuadrático medio (el Ejemplo 6 lo demuestra).*

1.1.1. Propiedades de los EMV.

Estimadores consistentes.

Definición 3. *Sea X una variable aleatoria, y sea T_n una sucesión de estimadores de una función paramétrica $g(\theta)$. Se dice que T_n es consistente si para todo $\epsilon > 0$ y para todo $\theta \in \Theta$ se verifica*

$$P_\theta[|T_n - g(\theta)| > \epsilon] \longrightarrow 0$$

Consistencia fuerte del EMV.

Proposición 1. *Sea X una variable aleatoria con distribución en una familia de distribuciones $\{F_\theta, \theta \in \Theta\}$, con Θ un intervalo abierto de $\mathbb{R}$, en el cual $\ln L_{x_1,\ldots,x_n}(\theta)$ es derivable respecto de θ para cualquier realización muestral $(x_1, \ldots, x_n)$ de una muestra aleatoria simple $(X_1, \ldots, X_n)$ de X. Para cualquier $\theta \in \Theta$, con probabilidad $P_\theta(x_1, \ldots, x_n) = 1$, existe una sucesión $\widehat{\theta}(X_1, \ldots, X_n)$ de raíces de la ecuación de verosimilitud tal que $\widehat{\theta}(X_1, \ldots, X_n) \xrightarrow[n\to\infty]{c.s} \theta$, $\forall \theta \in \Theta$.*

Normalidad asintótica del EMV.

Proposición 2. *Sea X una variable aleatoria con parámetro en una familia de distribuciones $\{F_\theta, \theta \in \Theta\}$, con Θ un intervalo abierto de $\mathbb{R}$. Si la f.m.p. o f.d.d. $f_\theta(x)$ de X verifica:*

i) Existe $\dfrac{\partial^3 \ln f_\theta(X)}{\partial \theta^3}$ y su valor absoluto está acotado por una función $K(x)$ tal que $E_\theta[K(x)] \leq k$.

ii) $E_\theta\left[\dfrac{\partial \ln f_\theta(X)}{\partial \theta}\right] = 0$ y $E_\theta\left[\dfrac{1}{f_\theta(x)} \dfrac{\partial^2 f_\theta(X)}{\partial \theta^2}\right] = 0$.

iii) $I_X(\theta) = E_\theta\left[\left(\dfrac{\partial \ln f_\theta(X)}{\partial \theta}\right)^2\right] > 0$.

Entonces, cualquier sucesión $\widehat{\theta} = \widehat{\theta}(X_1, \ldots, X_n)$ de raíces de la ecuación de verosimilitud que sea consistente para θ verifica que

$$\sqrt{n}\left[\widehat{\theta}(X_1, \ldots, X_n) - \theta\right] \xrightarrow[n\to\infty]{L} \mathcal{N}(0, I_X^{-1}(\theta))$$

Observación 3. *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 pues, en efecto, se tiene $Var_\theta[\widehat{\theta}(X_1, \ldots, X_n)] \approx (nI_X(\theta))^{-1} = I^{-1}_{X_1,\ldots,X_n}(\theta)$, $\forall \theta \in \Theta$.*

EMV y estadísticos suficientes.

Proposición 3. *Sea X una variable aleatoria con distribución en una familia paramétrica de distribuciones, $\{F_\theta, \theta \in \Theta\}$. Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de X. Supongamos que la familia admite un estadístico suficiente $T(X_1, \dots, X_n)$. Entonces, si existe un EMV de θ, este es una función (no constante) de $T(X_1, \dots, X_n)$.*

Demostración. Si T es suficiente, aplicando el teorema de factorización se tiene

$$f_\theta^n(x_1, \dots, x_n) = h(x_1, \dots, x_n) g_\theta(T(x_1, \dots, x_n)) = L_{x_1,\dots,x_n}(\theta)$$

Como $L_{x_1,\dots,x_n}(\theta) \geq 0$, entonces $h(x_1, \dots, x_n)$ y $g_\theta(T(x_1, \dots, x_n))$ tienen el mismo signo. Como h es un número, su valor no influye en cuándo se alcanza el máximo de la función de verosimilitud (pero sí su signo):

- Si $h(x_1, \dots, x_n) \geq 0$, entonces $\widehat{\theta}$ es el valor de θ donde se alcanza $\max\limits_{\theta \in \Theta} L_{x_1,\dots,x_n}(\theta)$, es decir, el valor de θ donde se alcanza $\max\limits_{\theta \in \Theta} g_\theta(T(x_1, \dots, x_n))$.
- Si $h(x_1, \dots, x_n) \leq 0$, entonces $\widehat{\theta}$ es el valor de θ donde se alcanza $\max\limits_{\theta \in \Theta} L_{x_1,\dots,x_n}(\theta)$, es decir, el valor de θ donde se alcanza $\min\limits_{\theta \in \Theta} g_\theta(T(x_1, \dots, x_n))$.

En ambos casos, el EMV de θ, $\widehat{\theta}$, es una función de $g_\theta(T(X_1, \dots, X_n))$, que a su vez es función de $T(X_1, \dots, X_n)$. Por tanto, $\widehat{\theta}$ es función de $T(X_1, \dots, X_n)$. Además, tal función no puede ser constante, puesto que si lo fuera, entonces o el máximo o el mínimo de $g_\theta(T)$ serían constantes, es decir, $g_\theta(T)$ sería constante, y $f_\theta^n(x_1, \dots, x_n) = h(x_1, \dots, x_n)$ llegando así a contradicción con el hecho de que T es suficiente. ■

Observación 4. *Es importante destacar que la Proposición 3 no implica que el EMV tenga que ser siempre un estadístico suficiente pues, en general, una transformación de un estadístico suficiente no es suficiente.*

Ejemplo 7. *Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de $X \rightsquigarrow U[\theta, \theta+1]$. Calcular el estadístico suficiente, un EMV para θ, y comprobar que no coinciden.*

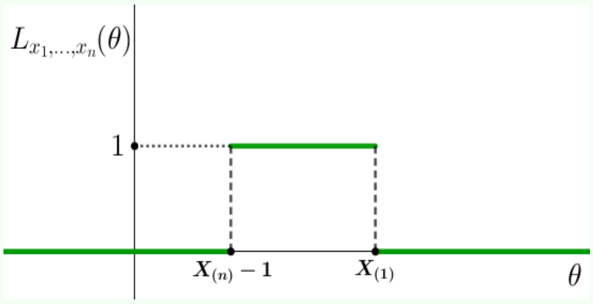
Solución. Para una distribución uniforme continua en el intervalo $[\theta, \theta+1]$, se tiene que su función de densidad es $f_\theta(x) = 1,\ \theta < x < \theta + 1$. Por tanto:

$$f_\theta^n(x_1, \dots, x_n) = 1 \cdot I_{[X_{(1)} \geq \theta]} \cdot I_{[X_{(n)} \leq \theta+1]} = 1 \cdot \underbrace{I_{[\theta \leq X_{(1)}]} \cdot I_{[\theta \geq X_{(n)}-1]}}_{g_\theta(T)}$$

Por el teorema de factorización, $T = \left(X_{(1)}, X_{(n)}\right)$ es un estadístico suficiente para θ.

Representamos la función de verosimilitud, cuyos valores dependen, al igual que en el Ejemplo 5, del orden de $X_{(1)}$ y $X_{(n)} - 1$.

- Si $X_{(1)} < X_{(n)} - 1$, entonces si $\theta \leq X_{(1)}$ no se puede dar la condición $\theta \geq X_{(n)} - 1$, y si se da la condición $\theta \geq X_{(n)} - 1$, entonces no se cumple $\theta \leq X_{(1)}$. Por tanto, alguna de las funciones indicadoras $I_{[\theta \leq X_{(1)}]}$ o $I_{[\theta \geq X_{(n)}-1]}$ será 0. Por tanto, $L_{x_1,\dots,x_n}(\theta) = 0,\ \forall \theta \in \Theta$.
- Si $X_{(n)} - 1 < X_{(1)}$, entonces se cumple $\theta \leq X_{(1)}$ y $\theta \geq X_{(n)} - 1$ si y solo si $\theta \in \left[X_{(n)} - 1, X_{(1)}\right]$, en cuyo caso las funciones indicadoras valen $I_{[\theta \leq X_{(1)}]} = 1$ y $I_{[\theta \geq X_{(n)}-1]} = 1$ y, por tanto, $L_{x_1,\dots,x_n}(\theta) = 1$.



Así, cualquier estadístico entre $X_{(n)}-1$ y $X_{(1)}$ (ambos inclusive) es EMV de θ. y no coincide con el estadístico suficiente ya que este es bidimensional y el EMV es unidimensional.

EMV y estimadores eficientes.

Proposición 4. *Sea X una variable aleatoria con distribución en una familia 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 $T(X_1, \ldots, X_n)$.*

Demostración. Si T es eficiente para $g(\theta)$, entonces se cumple

$$P_\theta\left[\frac{\partial \ln f_\theta^n(x_1,\ldots,x_n)}{\partial \theta} = a(\theta)(T-\theta)\right] = 1$$

Así, $\max\limits_{\theta\in\Theta} \ln L_{x_1,\ldots,x_n}(\theta) = \max\limits_{\theta\in\Theta} \ln f_\theta^n(x_1,\ldots,x_n)$.

$$\frac{\partial \ln L_{x_1,\ldots,x_n}(\theta)}{\partial\theta} = \frac{\partial \ln f_\theta^n(x_1,\ldots,x_n)}{\partial\theta} = \underbrace{a(\theta)}_{\neq 0}(T-\theta) = 0,\ \forall\theta \Leftrightarrow T = \theta,\ \forall\theta$$

Por tanto, $\widehat{\theta} = T$ es el EMV de θ. ∎

1.2. Estimadores de máxima verosimilitud de una función paramétrica.

La principal razón por la que son tan usados los estimadores de máxima verosimilitud es su buen comportamiento a la hora de estimar funciones paramétricas $g(\theta)$ a partir de una estimación del parámetro θ. Así, sea $g : \Theta \to \Lambda$ una función paramétrica. Definimos el concepto de función de verosimilitud sobre Λ a partir de la función de verosimilitud definida sobre Θ.

Función de verosimilitud de $g(\theta)$.

Definición 4. *Para cada realización muestral $(x_1, \ldots, x_n)$ de una muestra aleatoria simple $(X_1, \ldots, X_n)$ de una variable aleatoria X, se define la función de verosimilitud de $\lambda = g(\theta)$ asociada a dicha realización como:*

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

Observación 5. *Los pasos para poner en práctica la Definición 4 son sencillos. En primer lugar, calculamos la imagen inversa de $\lambda \in g(\Theta)$, es decir, $g^{-1}(\lambda) \in \Theta' \subseteq \Theta$. Después, evaluamos la función de verosimilitud $L_{x_1,\ldots,x_n}$ en todos los $\theta \in \Theta'$. Finalmente, la función de verosimilitud $M_{x_1,\ldots,x_n}(\lambda)$ será el supremo de los valores $L_{x_1,\ldots,x_n}(\theta), \theta \in \Theta'$.*

Estimador máximo verosímil de $g(\theta)$.

Definición 5. *Un estimador $\widehat{\lambda}(X_1,\ldots,X_n)$ de $\lambda = g(\theta)$ es estimador de máxima verosimilitud (EMV) de $\lambda = g(\theta)$ si maximiza su función de verosimilitud, es decir, si cumple:*

$$\forall (x_1,\ldots,x_n) \in \chi^n, \quad M_{x_1,\ldots,x_n}(\widehat{\lambda}(x_1,\ldots,x_n)) = \max_{\lambda\in\Lambda} M_{x_1,\ldots,x_n}(\lambda)$$

Recordemos que hallar el UMVUE de una función paramétrica $g(\theta)$ en general no puede hacerse a partir del UMVUE que ya se tenga calculado para θ. Así, por ejemplo, si T es el UMVUE de un parámetro p, en general no es cierto que e^T sea el UMVUE para e^p. Sin embargo, sí se cumplirá que si $\widehat{p}$ es EMV de p, entonces $e^{\widehat{p}} = \widehat{e^p}$ es EMV de e^p.

Teorema de invariancia de Zehna.

Teorema 1. *Sea X una variable aleatoria con distribución en una familia paramétrica de distribuciones, $\{F_\theta, \theta \in \Theta\}$. Sea $(X_1,\ldots,X_n)$ una muestra aleatoria simple de X. Sea g una función medible. Si $\widehat{\theta}(X_1,\ldots,X_n)$ es EMV de θ, entonces $g(\widehat{\theta}(X_1,\ldots,X_n))$ es EMV de $g(\theta)$.*

Demostración. Tenemos que demostrar que $g(\widehat{\theta})$ es el EMV de $g(\theta)$, siendo $\widehat{\theta}$ el EMV de θ, es decir, debemos llegar a que $M_{x_1,\ldots,x_n}(g(\widehat{\theta})) \geq M_{x_1,\ldots,x_n}(\lambda)$, $\forall \lambda \in \Lambda$, $\forall (x_1,\ldots,x_n) \in \chi^n$. Para ello, utilizamos que

$$M_{x_1,\ldots,x_n}(g(\widehat{\theta})) = \sup_{\theta\in g^{-1}(g(\widehat{\theta}))} L_{x_1,\ldots,x_n}(\theta) = L_{x_1,\ldots,x_n}(\widehat{\theta}).$$

Así, se tiene que $\forall \lambda = g(\theta)$:

$$M_{x_1,\ldots,x_n}(\lambda) = \sup_{\theta\in g^{-1}(\lambda)} L_{x_1,\ldots,x_n}(\theta) \underbrace{\leq}_{g^{-1}(\lambda)\subseteq\Theta} \sup_{\theta\in\Theta} L_{x_1,\ldots,x_n}(\theta) \underbrace{=}_{\widehat{\theta}\text{ EMV}} L_{x_1,\ldots,x_n}(\widehat{\theta}) = M_{x_1,\ldots,x_n}(g(\widehat{\theta})),$$

Leyendo la desigualdad de derecha a izquierda se tiene $M_{x_1,\ldots,x_n}(g(\widehat{\theta})) \geq M_{x_1,\ldots,x_n}(\lambda)\ \forall\lambda$, es decir, $g(\widehat{\theta})$ maximiza la función de verosimilitud de λ, que es a lo que queríamos llegar. ■

Ejemplo 8. *En el muestreo de una variable aleatoria $X \rightsquigarrow \{P(\lambda, \lambda > 0)\}$ se obtiene que en n observaciones aparece y veces el valor 0. Obtener un EMV de λ a partir de esta información.*

Solución. En este caso no se tiene una muestra aleatoria simple $(X_1,\ldots,X_n)$ de X sino una observación particular, $(x_1,\ldots,x_n)$ donde aparece y veces el valor 0.

Definimos la variable aleatoria $Y = n^o$ *de veces que aparece el 0 en n observaciones*. Así definida, $Y \rightsquigarrow B(n,p)$, donde p es la probabilidad de $X = 0$, es decir, $p = P[X = 0] = e^{-\lambda}$.

Es claro que si en n observaciones de X aparece y veces el valor 0, entonces se está trabajando con una única observación de la variable Y (es decir, una muestra aleatoria simple de Y de tamaño 1), y estamos interesados por tanto en calcular $P[Y = y]$, que es la función masa de probabilidad muestral de la muestra aleatoria simple considerada y, por tanto, coincide con la función de verosimilitud.

$$P_p[Y = y] = \binom{n}{y} p^y (1-p)^{n-y} = L_y(p),\ y = 0, 1, \ldots, n \Rightarrow \ln L_y(p) = \ln \binom{n}{y} + y \ln p + (n-y) \ln(1-p)$$

Por tanto:

$$\frac{\partial \ln L_y(p)}{\partial p} = \frac{y}{p} + \frac{n-y}{1-p} \cdot (-1) = 0 \Leftrightarrow \frac{y - yp - pn + yp}{p(1-p)} = 0 \Leftrightarrow y - pn = 0 \Leftrightarrow p = \frac{y}{n}$$

$$\frac{\partial^2 \ln L_y(p)}{\partial p^2} = \underbrace{\frac{-y}{p^2}}_{\leq 0} - \underbrace{\frac{n-y}{(1-p)^2}}_{\leq 0\ (n \geq y)} \leq 0 \Rightarrow \widehat{p} = \frac{y}{n} \text{ EMV para } p$$

Por el teorema de invariancia de Zehna, como $p = e^{-\lambda}$, entonces $\widehat{p} = \widehat{e^{-\lambda}}$. Así:

$$\widehat{e^{-\lambda}} = \frac{y}{n} = e^{-\widehat{\lambda}} \Rightarrow \widehat{\lambda} = -\ln\left(\frac{y}{n}\right)$$

Ejemplo 9. *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.*

Solución. Sea X = *tiempo de vida de cierto tipo de lámparas*, con $X \rightsquigarrow \exp(\theta)$. La función de densidad de X viene dada por $f_\theta(x) = \theta e^{-\theta x}$, y su función de distribución es $F_\theta(x) = 1 - e^{-\theta x}$, $x \geq 0$. Tengamos en cuenta que como observamos la variable X n veces, entonces si trabajamos con una realización muestral $(x_1, \ldots, x_n)$ de una muestra aleatoria simple $(X_1, \ldots, X_n)$ de X. Por otro lado:

$$P[X > 500] = 1 - F(500) = 1 - (1 - e^{-\theta \cdot 500}) = e^{-\theta \cdot 500}$$

Así, por el teorema de invariancia de Zehna se tiene que

$$P[\widehat{X > 500}] = e^{-\widehat{\theta} \cdot 500}$$

Hallamos entonces $\widehat{\theta}$.

$$L_{x_1, \ldots, x_n}(\theta) = \theta^n e^{-\theta \sum_{i=1}^n x_i} = f_\theta^n(x_1, \ldots, x_n)$$

$$\frac{\partial \ln L_{x_1, \ldots, x_n}(\theta)}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^n x_i = 0 \Leftrightarrow \theta = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{X}}$$

Como $\dfrac{\partial^2 \ln L_{x_1, \ldots, x_n}(\theta)}{\partial \theta^2} = \dfrac{-n}{\theta^2} < 0$, entonces $\widehat{\theta} = \dfrac{1}{\bar{X}}$ es EMV de θ. Así, $P[\widehat{X > 500}] = \exp\left\{\dfrac{-500}{\bar{X}}\right\}$.

Ejemplo 10. *Para una muestra aleatoria de tamaño n de una población $\mathcal{N}(\mu, \sigma^2)$ tal que la media y la varianza son iguales, es decir, $\mu = \theta = \sigma^2$, demostrar que el estimador máximo verosímil de θ es una raíz de la ecuación de segundo grado $\theta^2 + \theta m_1 - m_2 = 0$, donde m_1 y m_2 son los momentos muestrales no centrados de orden 1 y 2, respectivamente.*

Solución. Partimos de una variable $X \rightsquigarrow \mathcal{N}(\theta, \theta)$, donde su función de densidad viene definida por

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\theta^2}} \exp\left\{\frac{-1}{2\theta^2}(x-\theta)^2\right\}$$

Para una muestra aleatoria simple de tamaño n, la función de densidad muestral viene dada por:

$$\begin{aligned} f_\theta^n(x_1, \dots, x_n) &= \left(\frac{1}{\sqrt{2\pi\theta^2}}\right)^n \exp\left\{-\frac{1}{2\theta^2}\sum_{i=1}^n (x_i-\theta)^2\right\} \\ &= \left(\frac{1}{\sqrt{2\pi\theta^2}}\right) \exp\left\{-\frac{1}{2\theta^2}\left(\sum_{i=1}^n x_i^2 - 2\theta\sum_{i=1}^n x_i - n\theta^2\right)\right\} = L_{x_1,\dots,x_n}(\theta) \end{aligned}$$

Por tanto:

$$\ln L_{x_1,\dots,x_n}(\theta) = n\ln(\theta\sqrt{2n}) - \frac{1}{2\theta^2}\sum_{i=1}^n x_i^2 - \frac{1}{\theta}\sum_{i=1}^n x_i - \frac{n}{2}$$

Así:

$$\frac{\partial \ln L_{x_1,\dots,x_n}(\theta)}{\partial\theta} = \frac{-n}{\theta} + \frac{1}{\theta^3}\sum_{i=1}^n x_i^2 - \frac{1}{\theta^2}\sum_{i=1}^n x_i = \frac{1}{\theta^3}\left(-n\theta^2 - \theta\sum_{i=1}^n x_i + \sum_{i=1}^n x_i^2\right)$$

Así, igualando a 0 y haciendo el cambio $m_1 = \frac{1}{n}\sum_{i=1}^n x_i$, $m_2 = \frac{1}{n}\sum_{i=1}^n x_i^2$, se tiene:

$$-n\theta^2 - n\theta m_1 + nm_2 = 0 \Rightarrow n(-\theta^2 - \theta m_1 + m_2) = 0 \Rightarrow \theta^2 + \theta m_1 - m_2 = 0$$

Continuación del Ejemplo 11 del Tema 4.

Ejemplo 11. *Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de una variable aleatoria X con función de densidad*

$$f_\theta(x) = \frac{1}{2}e^{-(x-\theta)/2},\ x > \theta, \theta \in \mathbb{R}$$

d) Hallar el EMV para θ. ¿Es insesgado?

Solución. Ya se calculó la función de densidad muestral, que coincide con la función de verosimilitud. Así, se tiene que

$$L_{x_1,\dots,x_n}(\theta) = \frac{1}{2n}\exp\left\{\frac{-1}{2}\left(\sum_{i=1}^n x_i - n\theta\right)\right\} I_{[\theta < X_{(1)}]}$$

Por tanto, se tiene que la función de verosimilitud no es derivable, así que debemos estudiar su máximo estudiando previamente su crecimiento y decrecimiento.

Para $\theta > X_{(1)}$ se tiene $L_{x_1,\dots,x_n}(\theta) = 0$ puesto que $I_{[\theta < X_{(1)}]} = 0$. Por tanto, debemos estudiar su crecimiento cuando $\theta < X_{(1)}$. Así, definimos y estudiamos el signo de la función $g(\theta)$ definida como:

$$g(\theta) = \frac{1}{2n}\exp\left\{\frac{-1}{2}\sum_{i=1}^n x_i\right\}\exp\left\{\frac{n\theta}{2}\right\} \Rightarrow \frac{\partial g(\theta)}{\partial\theta} = \frac{1}{4}\underbrace{\exp\left\{\frac{-1}{2}\sum_{i=1}^n x_i\right\}}_{>0}\underbrace{\exp\left\{\frac{n\theta}{2}\right\}}_{>0} > 0$$

Así, se tiene que g crece en su dominio, luego $L_{x_1,\dots,x_n}(\theta)$ es creciente cuando $\theta < X_{(1)}$. Por tanto, el valor máximo de $L_{x_1,\dots,x_n}(\theta)$ se alcanzaría (teóricamente) cuando $\theta = X_{(1)}$. Sin embargo, esto no es posible puesto que si $\theta = X_{(1)}$, entonces $I_{[\theta<X_{(1)}]} = 0$. Por tanto, por ser la función de densidad no nula cuando $x > \theta$, entonces la función de verosimilitud no tiene máximo (sí que tiene supremo), luego no existe estimador máximo verosímil para θ.

Sin embargo, esta conclusión es muy estricta y se obtiene al ser excesivamente formales puesto que, intuitivamente, el EMV debería ser $X_{(1)}$ y realmente no ocurriría nada extraño si esa fuese nuestra conclusión. Para que esto sea estrictamente cierto, debemos escribir $x \geq \theta$ como condición para que la función de densidad sea no nula, y en ese caso $L_{x_1,\dots,x_n}(\theta)$ alcanza su máximo en $X_{(1)}$, $\widehat{\theta} = X_{(1)}$.

Por otra parte, como $X_{(1)}$ es un estimador función del estadístico suficiente y completo para θ con momento de segundo orden finito, entonces si fuera insesgado en θ, se tendría que $X_{(1)}$ es el UMVUE de θ, contradiciendo así la unicidad del UMVUE, que ya demostramos que era $-\frac{nT+2}{n} \neq T$ en general, con $T = X_{(1)}$.

Ejemplo 12. *Discutir la existencia del estimador máximo verosímil en la distribución uniforme $U(0,\theta)$ y en la distribución uniforme $U[0,\theta]$.*

Solución. El razonamiento para la solución de este ejemplo es similar al del Ejemplo 11. En el caso de la distribución $U(0,\theta)$ no existe EMV ya que al calcular $L_{x_1,\dots,x_n}(\theta)$ se obtiene que su máximo se alcanzaría en $\theta = X_{(n)}$, que no es posible puesto que $x < \theta$ y, por tanto, debe cumplirse $X_{(n)} < \theta$.

Por otro lado, si $x \in [0,\theta]$, entonces se cumple $\theta \leq X_{(n)}$ en cuyo caso es posible que se dé la igualdad $\theta = X_{(n)}$, valor para el cual se alcanza el máximo de $L_{x_1,\dots,x_n}(\theta)$. Así, $\widehat{\theta} = X_{(n)}$.

Continuación del Ejemplo 13 del Tema 4.

Ejemplo 13. *Sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de una variable X con función de densidad*

$$f_\theta(x) = \theta^2 x e^{-\theta x},\ x > 0, \theta > 0$$

d) Encontrar los EMV de la media y de la varianza de X.

Solución. Recordemos que $\mathrm{E}_\theta[X] = \frac{2}{\theta}$ y $\mathrm{Var}_\theta[X] = \frac{2}{\theta^2}$. Ya se demostró que $T = \bar{X}$ es eficiente para $2/\theta$. Por la Proposición 4, se tiene entonces que T también es el EMV de $2/\theta$, es decir,

$$\widehat{\frac{2}{\theta}} = \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i = \widehat{\mathrm{E}_\theta[X]}$$

Por el teorema de invariancia de Zehna, se tiene que

$$\widehat{\frac{\theta}{2}} = \frac{1}{\widehat{2/\theta}} = \frac{n}{\sum_{i=1}^n X_i} \Rightarrow \widehat{\theta} = 2 \cdot \widehat{\frac{\theta}{2}} = \frac{2n}{\sum_{i=1}^n X_i} \Rightarrow \widehat{\frac{2}{\theta^2}} = \frac{2}{\widehat{\theta}^2} = 2 \cdot \frac{\left(\sum_{i=1}^n X_i\right)^2}{2n^2} = \widehat{\mathrm{Var}_\theta[X]}$$

Para comprobarlo y recordar cómo se hace según el procedimiento teórico estudiado, maximizamos la función de verosimilitud $L_{x_1,\dots,x_n}(\theta)$.

$$L_{x_1,\dots,x_n}(\theta) = \theta^{2n}\prod_{i=1}^{n} x_i e^{-\theta\sum_{i=1}^n x_i} \Rightarrow \ln L_{x_1,\dots,x_n}(\theta) = 2n\ln\theta + \ln\left(\prod_{i=1}^{n} x_i\right) - \theta\sum_{i=1}^{n} x_i$$

$$\frac{\partial L_{x_1,\dots,x_n}(\theta)}{\partial\theta} = \frac{2n}{\theta} - \sum_{i=1}^{n} x_i = 0 \Leftrightarrow \frac{2n}{\theta} = \sum_{i=1}^{n} x_i \Leftrightarrow \theta = \frac{2n}{\sum_{i=1}^n x_i} \text{ candidato a EMV.}$$

$$\frac{\partial^2 \ln L_{x_1,\dots,x_n}(\theta)}{\partial\theta^2} = \frac{-2n}{\theta^2} < 0 \Rightarrow \frac{2n}{\sum_{i=1}^n x_i} \text{ máximo} \Rightarrow \widehat{\theta} = \frac{2n}{\sum_{i=1}^n X_i}$$

A partir de aquí, el procedimiento es el mismo que antes.

Continuación del Ejemplo 14 del Tema 4.

Ejemplo 14. *Sea X una variable aleatoria con función de densidad*

$$f_\theta(x) = \frac{\theta}{x^{1+\theta}},\ x > 1$$

d) Encontrar un EMV para θ.

Solución. Ya se demostró que $T = \overline{\ln X}$ es eficiente para $1/\theta$. Por la Proposición 4, se tiene entonces que T también es el EMV de $1/\theta$, es decir,

$$\widehat{\frac{1}{\theta}} = \overline{\ln X} = \frac{1}{n}\sum_{i=1}^{n} \ln X_i$$

Por el teorema de invariancia de Zehna, como $\theta = \frac{1}{1/\theta}$, entonces $\widehat{\theta} = \frac{1}{\widehat{1/\theta}}$. Así:

$$\widehat{\theta} = \frac{1}{\overline{\ln X}} = \frac{n}{\sum_{i=1}^n \ln X_i}$$

Para recordar cómo se haría según los pasos estudiados, maximizamos la función de verosimilitud.

$$L_{x_1,\dots,x_n}(\theta) = \frac{\theta^n}{\left(\prod_{i=1}^n x_i\right)\left(\prod_{i=1}^n x_i^\theta\right)} \Rightarrow \ln L_{x_1,\dots,x_n}(\theta) = n\ln(\theta) - \ln\left(\prod x_i\right) - \ln\left(\prod x_i^\theta\right)$$

$$\frac{\partial L_{x_1,\dots,x_n}(\theta)}{\partial\theta} = \frac{n}{\theta} - \sum_{i=1}^{n} \ln x_i = 0 \Leftrightarrow \frac{n}{\theta} = \sum_{i=1}^{n} \ln x_i \Leftrightarrow \theta = \frac{n}{\sum_{i=1}^n \ln x_i} \text{ candidato a EMV.}$$

$$\frac{\partial^2 \ln L_{x_1,\dots,x_n}(\theta)}{\partial\theta^2} = \frac{-n}{\theta^2} < 0 \Rightarrow \frac{n}{\sum_{i=1}^n \ln x_i} \text{ máximo} \Rightarrow \widehat{\theta} = \frac{n}{\sum_{i=1}^n \ln x_i}$$

Ejemplo 15. *Para una muestra aleatoria simple $(X_1,\dots,X_n)$ de una variable aleatoria X con función de densidad dada por $f_\theta(x) = \frac{2}{\theta^2}xe^{-x^2/\theta^2}$, $x > 0$. ¿Para qué función de θ es $\frac{1}{m_2}$ el EMV?*

Solución. La función de densidad muestral de X para una muestra aleatoria simple de tamaño n de X coincide con la función de verosimilitud, y viene dada por

$$f_\theta^n(x_1,\dots,x_n)=\frac{2}{\theta^2}x_1e^{-x_1^2/\theta^2}\cdots\frac{2}{\theta^2}x_ne^{-x_n^2/\theta^2}=\left(\frac{2}{\theta^2}\right)^n\prod_{i=1}^n x_i\exp\left\{-\frac{1}{\theta^2}\sum_{i=1}^n x_i^2\right\}=L_{x_1,\dots,x_n}(\theta)$$

Así, se tiene:

$$\ln L_{x_1,\dots,x_n}(\theta)=n\ln 2-n\ln\theta^2+\sum_{i=1}^n\ln x_i-\frac{1}{\theta^2}\sum_{i=1}^n x_i^2$$

Por tanto:

$$\frac{\partial \ln L_{x_1,\dots,x_n}(\theta)}{\partial\theta}=\frac{-2n}{\theta}+\frac{2\sum_{i=1}^n x_i^2}{\theta^3}=0\Leftrightarrow\theta^2=\frac{1}{n}\sum_{i=1}^n x_i^2$$

Como $\dfrac{\partial^2\ln L_{x_1,\dots,x_n}(\theta)}{\partial\theta^2}<0$, $\widehat{\theta^2}=m_2$. Por el teorema de invariancia de Zehna, $\dfrac{1}{m_2}$ es EMV de $\dfrac{1}{\theta^2}$.

2. Otros métodos de estimación puntual.

Estudiamos ahora dos métodos de estimación puntual útiles bien por el tipo de parámetros concretos que estiman, bien porque se utilizan en problemas de Inferencia específicos.

2.1. Método de los momentos.

Con el método de los momentos se podrá estimar cualquier función medible de los momentos poblacionales por la misma función de los correspondientes momentos muestrales. En particular, lo que se hace es igualar 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 obteniéndose así estimadores de los parámetros.

Método de los momentos.

Sea X una variable aleatoria con distribución en una familia paramétrica de distribuciones, $\{F_\theta,\theta\in\Theta\subseteq\mathbb{R}^k\}$, es decir, $\theta=(\theta_1,\dots,\theta_k)$. Sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de X.

Sean $m_1,\dots,m_k$ los k primeros momentos no centrados de X:

$$m_j(\theta_1,\dots,\theta_k)=\mathrm{E}_{(\theta_1,\dots,\theta_k)}[X^j]=\begin{cases}\displaystyle\sum_{i=1}^n x_i^jP_{\theta_1,\dots,\theta_k}[X=X_i] & \text{caso discreto}\\[2ex] \displaystyle\int_\chi x^jf_{\theta_1,\dots,\theta_k}(x)dx & \text{caso continuo}\end{cases}$$

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

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

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

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: $\theta_1^*, \ldots, \theta_k^*$

De la misma forma, se podrían igualar los k primeros momentos centrados poblacionales, μ_j, a los correspondientes momentos centrados muestrales, B_j, para obtener el sistema siguiente:

$$\left.\begin{array}{c} \mu_1(\theta_1, \ldots, \theta_k) = B_1 \\ \vdots \\ \mu_j(\theta_1, \ldots, \theta_k) = B_j \\ \vdots \\ \mu_k(\theta_1, \ldots, \theta_k) = B_k \end{array}\right\}$$

Ejemplo 16. *Estimar mediante el método de los momentos los parámetros de las siguientes distribuciones:*

a) $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$.

b) $X \rightsquigarrow U(a, b)$.

c) $X \rightsquigarrow P(\lambda)$.

d) $X \rightsquigarrow U(0, \theta)$.

Solución.

a) Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$. Sus momentos poblacionales son:

$$m_1 = \mathrm{E}[X], \qquad m_2 = \mathrm{E}[X^2] = \sigma^2 + \mu^2$$

Los correspondientes momentos muestrales son:

$$A_1 = \bar{X}, \qquad A_2 = \frac{\sum_{i=1}^{n} X_i^2}{n}$$

Igualando los momentos, se tiene:

$$\mu = \bar{X} \Rightarrow \mu^* = \bar{X}, \qquad \frac{\sum_{i=1}^{n} X_i^2}{n} = \sigma^2 + \mu^2 \Rightarrow (\sigma^2)^* = \frac{\sum_{i=1}^{n} X_i^2}{n} - \bar{X}$$

b) Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de $X \rightsquigarrow U(a, b)$. Calculamos sus momentos no centrados de primer orden (muestral y poblacional) y los igualamos.

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = m_1 = A_1 = \mathrm{E}[X] = \frac{a+b}{2}$$

Por otro lado, para practicar con los momentos centrados muestrales y poblacionales, calculamos los momentos centrados de segundo orden y los igualamos. Para ello, utilizamos que $\text{Var}[X] = \text{E}[X^2] - (\text{E}[X])^2$, para tener que:

$$\mu_2 = B_2 \Rightarrow \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2 = \text{E}[X^2] - (\text{E}[X])^2 = \text{Var}[X] = \frac{(b-a)^2}{12}$$

Por tanto, hemos llegado al sistema:

$$\begin{cases} \bar{X} = \dfrac{a+b}{2} & \Rightarrow a^* = \bar{X} - \sqrt{3}\sqrt{\dfrac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2} \\ \\ \dfrac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2 = \dfrac{(b-a)^2}{12} & \Rightarrow b^* = \bar{X} + \sqrt{3}\sqrt{\dfrac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2} \end{cases}$$

c) Sea $(X_1 \dots, X_n)$ una muestra aleatoria simple de $X \rightsquigarrow P(\lambda)$. Entonces:

$$\bar{X} = A_1 = m_1 = \text{E}[X] = \lambda \Rightarrow \lambda^* = \bar{X}$$

d) Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de $X \rightsquigarrow U(0, \theta)$. Entonces:

$$\bar{X} = A_1 = m_1 = \text{E}[X] = \frac{\theta}{2} \Rightarrow \theta^* = 2\bar{X}$$

Proposición 5. *Los estimadores obtenidos por el método de los momentos son consistentes. Es decir, si $\theta = h(m_1, \dots, m_k)$, con h una función continua, y $A_1^{(n)}, \dots, A_k^{(n)}$ son los momentos muestrales correspondientes a una muestra de tamaño n, entonces:*

$$\theta_n^* = h(A_1^{(n)}, \dots, A_k^{(n)}) \xrightarrow[n\to\infty]{c.s} h(m_1, \dots, m_k)$$

2.2. Método de mínimos cuadrados.

Con gran frecuencia se presentan situaciones en las que se observa una magnitud que depende de ciertas condiciones experimentales y de ciertos parámetros que se desean estimar. A esta situación la denotaremos $X = \varphi(t, \theta)$, donde t representa las diversas condiciones experimentales, $\theta \in \mathbb{R}^k$ representa los parámetros desconocidos, φ es es una función dada y X es el efecto que se observa. Si las observaciones se producen sin error, una vez realizado un número suficiente de observaciones de X, la determinación de los parámetros conduce al problema algebraico de resolver un sistema de ecuaciones en las incógnitas $\theta_1, \dots, \theta_k$.

Sin embargo, normalmente se cometerán errores aleatorios de medida de forma que, con más precisión, el resultado de la medición será $X = \varphi(t, \theta) + \epsilon$, siendo ϵ una variable aleatoria sobre la cual, al menos en una primera aproximación, puede suponerse que tiene una distribución que no depende ni de t ni de θ. La ausencia de errores sistemáticos, más fáciles de corregir en la práctica, permite suponer que los errores no tienen sesgo, es decir, que $\text{E}_\theta[\epsilon] = 0$, así como que los errores en diferentes medidas son independientes.

En estas condiciones el problema adquiere, además de su componente algebraica, una clara componente estadística. Así, después de realizar un número suficientemente grande de observaciones, X_i, en condiciones experimentales prefijadas, t_i, podríamos dar a los parámetros θ_j valores arbitrarios y echarle la culpa a los errores ϵ_i, cometidos en cada observación, de las discrepancias entre las observaciones y los valores $\varphi(t_i, \theta)$. Por supuesto, esto no es razonable. Lo sensato es buscar aquellos valores de los parámetros que hagan que el vector $(\varphi(t_1, \theta), \varphi(t_2, \theta), \ldots, \varphi(t_n, \theta))$ sea lo más próximo posible a la muestra aleatoria simple $(X_1, \ldots, X_n)$ de forma que el primero constituya la mejor previsión posible del segundo.

Método de mínimos cuadrados.

Sea $X = \varphi(t, \theta)$ una magnitud dependiente de t (condiciones experimentales) y de $\theta = (\theta_1, \ldots, \theta_k)$ un parámetro desconocido. Puesto que las observaciones de X conllevan un error de medida aleatorio, entonces se obtiene que una muestra aleatoria simple de X vendrá dada por

$$(X_1, \ldots, X_n) = (\varphi(t_1, \theta) + \epsilon_1, \ldots, \varphi(t_n, \theta) + \epsilon_n)$$

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

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

Si la función φ es derivable respecto a θ, la solución verificará:

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

Ejemplo 17. *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.*

Solución. Puesto que el espacio recorrido viene dado por $e = \frac{1}{2} a t^2$, entonces se tiene

$$X_i = \frac{1}{2} \theta t_i^2 + \epsilon_i$$

El estimador mínimo cuadrático de θ es aquel valor que minimiza $\sum_{i=1}^n \left[X_i - \frac{1}{2}\theta t_i^2\right]^2$, que se obtendrá cuando $\sum_{i=1}^n \left[X_i - \frac{1}{2}\theta t_i^2\right] t_i^2 = 0$. Quitando paréntesis se llega al estimador mínimo cuadrático:

$$\theta_* = \frac{2 \sum_{i=1}^n X_i t_i^2}{\sum_{i=1}^{t_i^4}}$$