tema4teorIE.pdf

(Provisional) Apuntes tema 4

3° Inferencia Estadística

Grado en Matemáticas

Facultad de Ciencias
Universidad de Granada

Tema 4: Estimación puntual. Insesgadez y mínima varianza. [1]

Índice

En este tema se introducen los métodos de estimación puntual de los parámetros de un modelo estadístico paramétrico mediante el uso de estimadores cumpliendo propiedades como la insesgadez, mínima varianza, regularidad y eficiencia.

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

1. Planteamiento del problema. Conceptos básicos.

1.1. Introducción.

Sea $F \in \mathcal{F} = \{F_\theta, \theta \in \Theta\}$ una familia de distribuciones paramétricas. Nuestro objetivo es evaluar el parámetro θ desconocido con el fin de identificar la distribución de una población dentro de $\mathcal{F}$. En concreto, queremos determinar la distribución $F \in \mathcal{F}$ de una variable aleatoria X, y como esta depende de θ que, a priori, no es conocido, debemos desarrollar métodos para obtener su valor concreto, o un valor lo más aproximado posible a su valor real, al que llamaremos θ_0.

En muchas ocasiones será necesario seleccionar un único valor de θ que constituya un pronóstico individual sobre el parámetro. Se habla entonces de *estimación puntual* (temas 4 y 5). En otras, bastará con dar un intervalo en el que, con cierta seguridad, podemos afirmar que varía θ, y así obtendremos los llamados *intervalos de confianza* (tema 6), mientras que otras veces tendremos que comprobar o desmentir alguna hipótesis sobre θ, llegando así al *contraste de hipótesis* (tema 7).

A lo largo de los temas 4 y 5 vamos a estudiar métodos con los que poder hacer una estimación puntual sobre el parámetro desconocido θ que determina la distribución de una variable aleatoria.

Tengamos en cuenta que el valor de θ que se obtenga en la estimación no será más que una aproximación de su valor real θ_0. Por ejemplo, si tras lanzar un dado 100 veces se da 0,17 como estimación puntual de la probabilidad de obtener 4, no está excluido que el valor real pueda ser 0,165, 0,175 o 1/6, puesto que 0,17 debe entenderse como resultado del compromiso de suministrar una única conjetura estadística acerca de dicha probabilidad.

Sin embargo, pese a que el hecho de dar un único valor para el parámetro θ pueda ser atrevido, se podrá asegurar (bajo ciertas condiciones) que las conclusiones que se obtengan a partir de tal aproximación son lo más acertadas posible (similares a las que se tendrían si se conociera el verdadero valor de θ). Por tanto, debemos preocuparnos por obtener buenas estimaciones puntuales, para lo cual definiremos propiedades que aseguren que el error cometido al utilizarlas es mínimo.

Sea X una variable aleatoria, y sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X. La manera de proceder en un problema de estimación puntual del valor paramétrico real θ_0 consiste en seleccionar un estadístico $T(X_1, \ldots, X_n)$ adecuado, función dependiente solo de las observaciones muestrales, y tomar como estimación θ de θ_0 el valor de T calculado a partir de la muestra obtenida. Ello supone aceptar exclusivamente que la estimación puntual del parámetro debe ser función de la muestra observada.

De manera genérica, los estadísticos $T(X_1, \ldots, X_n)$, independientes del parámetro θ_0, cuyos valores se utilizan para obtener su estimación puntual, se llamarán estimadores. Su recorrido estará en el espacio paramétrico Θ, es decir, no proporcionarán estimaciones fuera del rango del parámetro.

En ocasiones, en lugar 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 el valor de $g(\theta_0)$, el verdadero valor de $g(\theta)$, en vez de θ_0, el verdadero valor de θ.

Ejemplo 1. *Sea $X \rightsquigarrow B(n,p)$ con n conocido y $p \in (0,1)$. A partir de una muestra aleatoria simple de X, indicar algún estadístico que se pueda usar para inferir p y alguna función paramétrica que pueda ser de interés.*

Solución. Como queremos inferir el verdadero valor de $p \in (0,1)$, a falta de otras herramientas que nos digan lo buena que es la aproximación obtenida debemos conformarnos con utilizar para la estimación un estadístico que tome valores dentro del espacio paramétrico $\Theta = (0,1)$. Por tanto, nos sirve cualquier estadístico que tome valores en $(0,1)$. Por ejemplo, $X_{(1)}$, $X_{(n)}$ y $\bar{X}$ no son estadísticos convenientes puesto que no toman valores entre 0 y 1.

Por otro lado, como $0 \leq X_i \leq n, \forall i$, entonces $0 \leq \frac{X_i}{n} \leq 1, \forall i$, luego $\frac{X_i}{n}, \forall i$ y, en particular, $\frac{X_{(n)}}{n}$ (una función de un estadístico ya conocido) serán estadísticos adecuados (a pesar de que en las estimaciones se puede obtener alguna vez 0 o 1, valores que no se encuentran estrictamente contenidos en el espacio paramétrico pero pueden ser aceptados como parámetro de una distribución binomial).

1.2. Estimadores. Función de pérdida y función de riesgo.

Sea X una variable aleatoria con función de distribución en una familia de distribuciones paramétricas, $F \in \{F_\theta, \theta \in \Theta\}$, y $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X.

Estimador.

Definición 1. *Un estimador de θ es un estadístico $T(X_1, \ldots, X_n)$ que toma valores en Θ:*

$$T : \chi^n \to \Theta$$

Si el estimador está definido con espacio de llegada $g(\Theta)$, es decir, $T : \chi^n \to g(\Theta)$, 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.*

Ejemplo 2.

- *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de $X \rightsquigarrow \{P(\lambda), \lambda > 0\}$. Entonces, $\bar{X}$ es un estimador de λ, pues obviamente será la media de valores positivos y λ también lo es. De hecho, cualquier función medible de la muestra que sea independiente del parámetro λ y que tome valores positivos es un estimador de λ.*
- *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de $X \rightsquigarrow \{B(1,p), p \in (0,1)\}$. Entonces, $\bar{X}$ será la media de n valores que, o bien son 0, o bien son 1. Por tanto, $\bar{X} \in (0,1)$ (aunque puede darse la casualidad de que valga 0 o 1, hecho que también damos por válido). Así, $\bar{X}$ es un estimador de p. De hecho, 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, de entre todos, el mejor de ellos, es decir, aquel estimador con el que se comete el menor error posible a la hora de aproximar el verdadero valor de θ (al que llamamos θ_0). Para ello, damos algunos criterios de selección.

Función de pérdida.

Definición 2. *Una función $L : \Theta \times \Theta \to \mathbb{R}$ se llama función de pérdida si verifica 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.

Observación 1. *Intuitivamente, la función de pérdida indica la información que se pierde (o el error que se comete) al estimar el parámetro por el valor del estadístico* $T = t$ *si su verdadero valor es* θ.

- *i) y ii) nos dicen que el error que se comete siempre debe ser positivo, o 0 si la estimación se corresponde con el valor real del parámetro.*
- *Por otro lado, iii) nos dice que si un valor* T *nos da una mejor aproximación de* θ *que otro valor* T' *(un estadístico nos da una mejor estimación que otro), entonces el error asociado a* T *es menor que el error asociado a* T'.

Ejemplo 3. *Es sencillo comprobar que las siguientes funciones son funciones de pérdida (para ello, basta comprobar las tres condiciones de la Definición 2:*

- Error absoluto de estimación: $L(\theta, T) = |\theta - T|$.

  Asociada a esta función de pérdida se define otra, también frecuente, que penaliza con un coste $c \geq 0$ *los errores mayores que un cierto* $\epsilon \to 0$:

$$L(\theta, T) = \begin{cases} c & si\ |\theta - t| > \epsilon \\ 0 & si\ |\theta - t| \leq \epsilon \end{cases}$$

- Error cuadrático de estimación: $L(\theta, T) = (\theta - T)^2$.
- Error relativo de estimación: $L(\theta, T) = \left|\dfrac{\theta - T}{\theta}\right|$.

Dado un estimador $T(X_1, \ldots, X_n)$ de θ, impondremos que la función de pérdida $L(\theta, T(X_1, \ldots, X_n))$ sea Borel-Medible para todo $\theta \in \Theta$. Por tanto, por ser T una variable aleatoria y ser L medible, se tiene que $L(\theta, T)$ también es otra variable aleatoria, luego podremos calcularle su esperanza.

Función riesgo.

Definición 3. *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*, es decir:*

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

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

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

Observación 2. *El cálculo de $R_T^L(\theta)$ viene dado por:*

$$R_T^L(\theta) = E_\theta[L(\theta, T(X_1, \dots, X_n))] = \int_{\mathbb{R}} L(\theta, t) F_\theta(t) dt$$

donde $F_\theta(t)$ representa la distribución en el muestreo del estadístico T.

El concepto de riesgo proporciona un criterio para la compararación de estimadores en el sentido de que podemos saber si utilizar un estimador T frente a otro T' si, con la misma función de pérdida, el error medio que se comete al emplear T es menor que el que se comete al emplear T'. Por tanto, parece lógico buscar aquel estimador que minimice la función riesgo.

Estimador óptimo.

Definición 4. *Se dice que un estimador $(T(X_1, \dots, X_n))$ es óptimo bajo una función de pérdida $L(\theta, T)$ si dicho estimador minimiza uniformemente la función de riesgo $R_T(\theta)$, es decir, un estimador T tal que, para cualquier otro estimador T', verifica:*

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

Ejemplo 4. *Consideremos una variable aleatoria $X \rightsquigarrow U(0, \theta), \theta > 0$. Ya se demostró que un estadístico suficiente y completo para estimar el valor de θ es $T = X_{(n)}$ (Tema 3, Ejemplo 7), y de hecho, parece lógico que así lo sea pues el propio θ es el mayor de los valores observables de la variable. Aun así, es posible que T no sea el estimador óptimo, y de hecho en la mayoría de ejemplos que veamos, si T es un estimador suficiente y completo, se tendrá que el óptimo será un múltiplo de T. Por tanto, en este caso, para buscar al estimador óptimo de θ, consideraremos como candidatos a T y a sus múltiplos, $T_k = kT$ (luego $X_{(n)} = T_1$).*

La distribución en el muestreo de T_1 y su función de densidad vienen dadas por:

$$F_{T_1}(x) = \left(\frac{x}{\theta}\right)^n, \ x \in (0, \theta) \Rightarrow f_{T_1}(x) = \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1}, \ x \in (0, \theta)$$

Si consideramos la función pérdida $L(\theta, T) = \dfrac{(\theta - T)^2}{\theta^2}$, entonces:

$$R_{T_k}^L(\theta) = E_\theta\left[\frac{(\theta - T_k)^2}{\theta^2}\right] = E_\theta\left[\frac{\cancel{\theta^2}}{\cancel{\theta^2}} - \frac{2\cancel{\theta}T_k}{\theta^{\cancel{2}}} + \frac{T_k^2}{\theta^2}\right] = 1 - \frac{2}{\theta}E_\theta[T_k] + \frac{1}{\theta^2}E_\theta[T_k^2]$$

$$= 1 - \frac{2}{\theta}E_\theta[kT_1] + \frac{1}{\theta^2}E_\theta[k^2T_1^2] = 1 - \frac{2k}{\theta}E[T_1] + \frac{k^2}{\theta^2}E[T_1^2]$$

Calculemos $E_\theta[T_1]$ y $E_\theta[T_1^2]$.

$$E_\theta[T_1] = \int_0^\theta t \cdot f_{T_1}(t) dt = \int_0^\theta t \cdot \frac{n}{\theta}\left(\frac{t}{\theta}\right)^{n-1} dt = \int_0^\theta n \cdot \frac{t^n}{\theta^n} dt = \frac{n}{\theta^n}\int_0^\theta t^n dt = \frac{n}{\theta^n}\left[\frac{t^{n+1}}{n+1}\right]_0^\theta = \frac{n}{n+1}\theta$$

$$E_\theta[T_1^2] = \int_0^\theta t^2 \cdot f_{T_1}(t) dt = \int_0^\theta t^2 \cdot \frac{n}{\theta}\left(\frac{t}{\theta}\right)^{n-1} dt = \frac{n}{\theta^n}\int_0^\theta t^{n+1} dt = \frac{n}{\theta^n}\left[\frac{t^{n+2}}{n+2}\right]_0^\theta = \frac{n}{n+2}\theta^2$$

Por tanto:

$$R_{T_k}^L = 1 - \frac{2k}{\cancel{\theta}}\frac{n}{n+1}\cancel{\theta} + \frac{k^2}{\cancel{\theta^2}}\frac{n}{n+2}\cancel{\theta^2} = 1 - \frac{2kn}{n+1} + \frac{k^2 n}{n+2}$$

En este caso, el riesgo no depende de θ. Por tanto, $R^L_{T_k}(\theta) = R^L_{T_k}(k)$. Así, el menor valor del riesgo vendrá dado por la solución de la ecuación $(R^L_{T_k})'(k) = 0$, donde la derivada la entendemos respecto a k.

$$(R^L_{T_k})'(k) = -\frac{2n}{n+1} + \frac{2kn}{n+2} = 0 \Leftrightarrow \frac{2kn}{n+2} = \frac{2n}{n+1} \Leftrightarrow k = \frac{\cancel{2n}(n+2)}{\cancel{2n}(n+1)} = \frac{n+2}{n+1}$$

Como $(R^L_{T_k})''(k) = \dfrac{2n}{n+2} > 0, \forall k$, entonces $k = \dfrac{n+2}{n+1}$ es un mínimo de $R^L_{T_k}(\theta)$.

Por tanto, el estimador óptimo de θ bajo la función $L = \dfrac{(\theta - T)^2}{\theta^2}$ es $T = \dfrac{n+2}{n+1}X_{(n)}$.

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

1. Restringir la clase de los estimadores considerados a aquellos que cumplan alguna propiedad adicional de interés, tratando con ello de eliminar los estimadores indeseables y examinando si el criterio del mínimo riesgo selecciona entre ellos uno preferible a todos los demás.

   Ya tenemos algunas propiedades que, de algún modo, permiten hacer una estimación adecuada del parámetro desconocido. En efecto, en el tema anterior ya vimos los conceptos de suficiencia y completitud. Ahora, añadiremos a esa lista de propiedades la insesgadez, la mínima varianza y la eficiencia, y pasaremos a buscar no solo estadísticos suficientes y completos, sino estadísticos suficientes y completos que además sean estimadores insesgados, de mínima varianza, y eficientes (cuando se pueda asegurar su existencia).

2. Reforzar el criterio de preferencia de estimadores mediante el procedimiento de reducir toda la función de riesgo $R^L_T(\theta)$ a un único número representativo r_T, que permita ordenar linealmente todos los estimadores, para tratar de seleccionar el mejor.

   En esta línea destacan fundamentalmente los criterios Bayes y minimax, que no serán estudiados en este curso.

Observación 3. *Cabría en principio una tercera posibilidad: hallar el subconjunto de todos los estimadores* admisibles *(aquellos para los cuales no existe otro estimador preferible a él) y seleccionar de entre ellos aquel que tuviese propiedades adicionales deseables. Sin embargo, este proceso no es operativo debido a la dificultad de enumerar todos los estimadores admisibles.*

Error cuadrático medio (ECM).

Definición 5. *El criterio de menor error cuadrático medio (ECM) es aquel en el que se considera como función de pérdida a la función*

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

y como función de riesgo a

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

Dos razones intuitivas que hacen preferible utilizar el criterio del error cuadrático medio frente a otros son las siguientes:

- Tiene ventajas desde el punto de vista analítico frente a otras funciones de riesgo. Por ejemplo, hallar el mínimo de la función de riesgo implica hallar el mínimo de una función cuadrática, mientras que por ejemplo si consideramos el error absoluto de estimación deberemos hallar el mínimo de un valor absoluto, que implica su derivación, que puede ser compleja.
- Tiene una interpretación sencilla, pues en este caso mide el grado de dispersión del estimador en torno al verdadero valor del parámetro θ y no considera para ello distinción entre estimaciones por encima o por debajo del parámetro (no tiene en cuenta los signos de los errores en la estimación).

Proposición 1.

I) El ECM puede descomponerse 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 sesgo que se define como

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

II) Si el estimador considerado verifica $E_\theta[T] = \theta$ (lo que llamaremos estimador insesgado), o $B_T(\theta) = 0$, se verifica que el ECM coincide con la varianza del estimador:

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

Demostración. Si demostramos I), haciendo $B_T^2(\theta) = 0$ se tendrá demostrado II). Probemos entonces que se cumple I).

$$\begin{aligned}
ECM_T(\theta) &= \mathrm{E}_\theta[(\theta - T)^2] = \mathrm{E}_\theta[[(\theta - \mathrm{E}_\theta[T]) + (\mathrm{E}_\theta[T] - T)]^2] \\
&= \underbrace{\mathrm{E}_\theta[(\theta - \mathrm{E}_\theta[T])^2]}_{(1)} + \underbrace{\mathrm{E}_\theta[(\mathrm{E}_\theta[T] - T)^2]}_{(2)} + 2\underbrace{\mathrm{E}_\theta[(\theta - \mathrm{E}_\theta[T])(\mathrm{E}_\theta[T] - T)]}_{(3)} = (A)
\end{aligned}$$

$$\begin{aligned}
(1) &= \mathrm{E}_\theta[\theta^2 - 2\theta\mathrm{E}_\theta[T] + (\mathrm{E}_\theta[T])^2] = \mathrm{E}_\theta[\theta^2] - 2\theta\mathrm{E}_\theta(\mathrm{E}_\theta[T]) + \mathrm{E}[(\mathrm{E}_\theta[T])^2] \\
&= \theta^2 - 2\theta\mathrm{E}_\theta[T] + (\mathrm{E}_\theta[T])^2 = (\mathrm{E}_\theta[T] - \theta)^2 \Rightarrow (1) = B_T^2(\theta), \text{ con } B_T(\theta) = \mathrm{E}_\theta[T] - \theta. \\
(2) &= \mathrm{E}_\theta[(\mathrm{E}_\theta[T])^2 - 2\mathrm{E}_\theta[T]T + T^2] = (\mathrm{E}_\theta[T])^2 - 2\mathrm{E}_\theta[T]\mathrm{E}_\theta[T] + \mathrm{E}_\theta[T^2] = \mathrm{E}_\theta[T^2] - (\mathrm{E}_\theta[T])^2 \\
&= \mathrm{Var}_\theta(T) \Rightarrow (2) = \mathrm{Var}_\theta(T) \\
(3) &= (\theta - \mathrm{E}_\theta[T])\mathrm{E}_\theta[\mathrm{E}_\theta[T] - T] = (\theta - \mathrm{E}_\theta[T])(\mathrm{E}_\theta[T] - \mathrm{E}_\theta[T]) = 0
\end{aligned}$$

Por tanto, en efecto:

$$(A) = (1) + (2) = ECM = \mathrm{Var}_\theta(T) + B_T^2(\theta)$$

■

2. Estimación insesgada de mínima varianza.

Acabamos de ver que existe una relación sencilla e intuitiva que liga el error cuadrático medio, 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, pilar de otras propiedades como la mínima varianza, la regularidad y la eficiencia.

2.1. Insesgadez

De nuevo, sea X una variable aleatoria con función de distribución en una familia de distribuciones paramétricas, $F \in \{F_\theta, \theta \in \Theta\}$. Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de X.

Estimador insesgado.

Definición 6. *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 5. *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.*

Solución. Este ejemplo ya se vio en la Proposición 4 del Tema 1. Debido a su importancia, lo repetimos ahora. Sean $\bar{X}$ y S^2 la media y la cuasivarianza muestral, respectivamente, definidas como:

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

Sean μ y σ^2, respectivamente, la media y la varianza poblacional. Es evidente que S^2 es estimador de σ^2 puesto que ambos toman valores en $\mathbb{R}^+$. Por otro lado, como $x_1 \le \mu \le x_k$ (en una población donde la variable X presenta k modalidades), entonces $x_1 \le X_1 \le \bar{X} \le X_n \le x_k$ para cada muestra aleatoria simple que se seleccione, luego $\bar{X}$ es estimador de μ.

Demostremos que $\mathrm{E}[\bar{X}] = \mu$ y $\mathrm{E}[S^2] = \sigma^2$.

$$\mathrm{E}[\bar{X}] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[X_i] \underbrace{=}_{\text{i.d}} \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[X] = \frac{\mu\cancel{n}}{\cancel{n}} = \mu$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n}{n-1}\cdot\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n}{n-1}B_2 \Rightarrow \mathrm{E}[S^2] = \frac{n}{n-1}\mathrm{E}[B_2]$$

donde $B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$ es el momento centrado de orden k.

El teorema de König para la varianza poblacional tendrá su versión para la varianza muestral, afirmándose que

$$B_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

Por tanto:

$$\mathrm{E}[B_2] = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[X_i^2] - \mathrm{E}[\bar{X}^2] \underbrace{=}_{i.d} \frac{1}{\cancel{n}}\cancel{n}\mathrm{E}[X^2] - \mathrm{E}[\bar{X}^2] = \mathrm{E}[X^2] - (\mathrm{E}[\bar{X}])^2 = (1)$$

Utilizando que

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 \Rightarrow \mathrm{E}[X^2] = \mathrm{Var}[X] + (\mathrm{E}[X])^2$$
$$\mathrm{Var}[\bar{X}] = \mathrm{E}[\bar{X}^2] - (\mathrm{E}[\bar{X}])^2 \Rightarrow \mathrm{E}[\bar{X}^2] = \mathrm{Var}[\bar{X}] + (\mathrm{E}[\bar{X}])^2$$

y que $\mathrm{Var}[\bar{X}] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] \underbrace{=}_{i.i.d} \frac{1}{n^{\cancel{2}}}\cancel{n}\mathrm{Var}[X] = \frac{1}{n}\mathrm{Var}[X]$, se llega a:

$$(1) = \mathrm{Var}[X] + (\mathrm{E}[X])^2 - \frac{\mathrm{Var}[X]}{n} - \underbrace{(\mathrm{E}[\bar{X}])^2}_{(\mathrm{E}[X])^2} = \frac{(n-1)\mathrm{Var}[X]}{n} = \frac{(n-1)\sigma^2}{n}$$

Así, se tiene que:

$$\mathrm{E}[S^2] = \frac{\cancel{n}}{\cancel{n-1}} \cdot \frac{\cancel{(n-1)}\sigma^2}{\cancel{n}} = \sigma^2$$

Observación 4.

- *Puesto que la esperanza matemática de un vector aleatorio se define como un vector cuyas componentes son las esperanzas matemáticas de sus componentes aleatorias, si $\theta = (\theta_1, \dots, \theta_k)$ es un parámetro k-dimensional, entonces 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, es deicr: un estimador $T = (T_1, \dots, T_k)$ es insesgado en θ si se verifica*

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

- *Por lo visto en la Proposición 1, se tiene que para un estimador insesgado se verifica:*

$$ECM_T(\theta) = Var_\theta(T(X_1, \dots, X_n))$$

- *Si T es un estimador insesgado de θ, entonces en general, $h(T)$ no es un estimador insesgado de $h(\theta)$, siendo h cualquier función. Sin embargo, si h es una función lineal, dicha implicación sí se cumple; es decir, la insesgadez no se mantiene bajo transformaciones en general, pero sí se mantiene si la transformación es lineal.*
- *No tiene por qué existir algún estimador insesgado de un parámetro.*
- *Un estimador insesgado no tiene porqué ser único. De hecho, cualquier combinación convexa de estimadores insesgados es otro estimador insesgado: 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.*

Ejemplo 6. *Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de alguna población. Probar que, en general, la cuasidesviación típica (raíz cuadrada de la cuasivarianza) no es un estimador insesgado para $\sigma_\theta = \sqrt{Var_\theta(X)}$.*

Solución. Recordemos que la función Gamma viene definida por:

$$\Gamma(p) = \int_0^{+\infty} x^{p-1}e^{-x}dx, \quad p > 0$$

Vamos a demostrar que $\mathrm{E}[S] \neq \sigma$. Para ello, necesitamos un contraejemplo. Así, sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$. Consideramos entonces la variable aleatoria

$$Y = \frac{(n-1)S^2}{\sigma^2} \rightsquigarrow \chi^2(n-1) \Rightarrow \mathrm{E}[Y] = \mathrm{E}\left[\frac{(n-1)S^2}{\sigma^2}\right]$$

donde S^2 es la cuasivarianza de X. Como la función de densidad de una $\chi^2(n)$ es

$$f(x) = \frac{1}{\Gamma\left(\frac{n}{2}\right)\cdot 2^{n/2}}x^{n/2-1}e^{-x/2}, \qquad x > 0$$

entonces se tiene que la función de densidad de Y es

$$f(y) = \frac{1}{\Gamma\left(\frac{n-1}{2}\right)\cdot 2^{(n-1)/2}}y^{(n-1)/2}e^{-y/2}, \qquad y > 0$$

Para calcular la esperanza de S necesitamos que desaparezca el cuadrado en S^2, así que calculamos $\mathrm{E}[\sqrt{Y}]$.

$$\mathrm{E}[\sqrt{Y}] = \int_0^{+\infty} \sqrt{y}\frac{1}{\Gamma\left(\frac{n-1}{2}\right)\cdot 2^{(n-1)/2}}y^{(n-1)/2}e^{-y/2} = \frac{1}{2^{(n-1)/2}\Gamma\left(\frac{n-1}{2}\right)}\int_0^{+\infty} y^{n/2-1}e^{-y/2}dy = (1)$$

Como $\displaystyle\int_0^{+\infty} \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}y^{n/2-1}e^{-y/2}dy = 1$ por estar integrando la función de densidad de una $\chi^2(n)$:

$$(1) = \frac{2^{n/2-1}\Gamma\left(\frac{n}{2}\right)}{2^{(n-1)/2}\Gamma\left(\frac{n-1}{2}\right)} \Rightarrow \mathrm{E}[\sqrt{Y}] = \frac{2^{1/2}\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}$$

Por tanto:

$$\mathrm{E}\left[\sqrt{\frac{(n-1)S^2}{\sigma^2}}\right] = \mathrm{E}\left[\frac{\sqrt{n-1}S}{\sigma}\right] = \frac{\sqrt{n-1}}{\sigma}\mathrm{E}[S] \Rightarrow \mathrm{E}[S] = \frac{\sigma}{\sqrt{n-1}}\frac{2^{1/2}\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \neq \sigma$$

Por tanto, la cuasivarianza, en general, no es un estimador insesgado de la desviación típica (una función del estimador insesgado en el parámetro no es estimador insesgado de la función del parámetro).

Ejemplo 7.

a) *Sea $(X_1, \ldots, X_n), n \geq 2$ una muestra aleatoria simple de una distribución de Bernoulli $B(1, p)$. Probar que $\dfrac{T^2 - T}{n^2 - n}$ con $T = \displaystyle\sum_{i=1}^{n} X_i$ es un estimador insesgado para $g(p) = p^2$.*

b) *Probar que para $n = 1$ no existe un estimador insesgado de p^2.*

Solución.

a) Tenemos que $T = \sum_{i=1}^{n} X_i \rightsquigarrow B(n, p)$. Por tanto:

$$\mathrm{E}[T] = np, \quad \mathrm{Var}[T] = np(1-p) = \mathrm{E}[T^2] - (\mathrm{E}[T])^2 \Rightarrow \mathrm{E}[T^2] = np - np^2 + (np)^2$$

Así, se tiene:

$$\mathrm{E}\left[\frac{T^2 - T}{n^2 - n}\right] = \frac{1}{n^2 - n}\mathrm{E}[T^2 - T] = \frac{\mathrm{E}[T^2] - \mathrm{E}[T]}{n^2 - n} = \frac{\cancel{np} - np^2 + n^2p^2 - \cancel{np}}{n^2 - n} = \frac{p^2\cancel{(n^2 - 2)}}{\cancel{n^2 - n}} = p^2$$

b) Consideramos $n = 1$. Entonces, estamos buscando un estimador h, independiente de p, de forma que $\mathrm{E}[h(X)] = p^2$. Recordemos que, si X es una variable aleatoria discreta, su esperanza viene dada por

$$\mathrm{E}[X] = \sum_x x \cdot P[X = x]$$

luego en este caso $\mathrm{E}[h(X)] = h(0) \cdot (1-p) + h(1) \cdot p$, con $h(0)$ y $h(1)$ independientes de p. Por tanto, $h(0)(1-p) + h(1)p$ es un polinomio en p de grado 1, que no puede estar igualado de ninguna forma a un polinomio de grado 2, luego $h(0)(1-p) + h(1)p \neq p^2, \forall p$.

Ejemplo 8.

a) Sea X una variable aleatoria con distribución $P(\lambda)$ y $X_1, \ldots, X_n$ una muestra aleatoria simple de X. Probar que $\bar{X}$ y S^2 son estimadores insesgados de λ.

b) Sea X una variable aleatoria con distribución $P(\lambda)$. ¿Existe algún estimador insesgado para la función paramétrica $1/\lambda$ basado en una muestra de tamaño 1?

Solución.

a) En el Ejemplo 5 ya se vio para cualquier población en general. En particular, por tanto, se cumple para una distribución de Poisson.

b) Consideramos $n = 1$. Buscamos un estimador h, independiente de λ, tal que $\mathrm{E}[h(X)] = 1/\lambda$.

$$\begin{aligned}\mathrm{E}[h(X)] &= \sum_x h(x)P[X = x] = h(0)e^{-\lambda}\frac{\lambda^0}{0!} + h(1)e^{-\lambda}\frac{\lambda^1}{1!} + h(2)e^{-\lambda}\frac{\lambda^2}{2!} + \cdots \\ &= e^{-\lambda}\left(h(0) + \lambda h(1) + \lambda^2\frac{h(2)}{2} + \cdots\right) = \frac{1}{\lambda} \Leftrightarrow h(0) + \lambda h(1) + \lambda^2\frac{h(2)}{2} + \cdots = \lambda e^{-\lambda}\end{aligned}$$

que se cumple si y solo si $h(0) = h(2) = \cdots = 0$ y $h(1) = e^{-\lambda}$, lo cual no puede darse pues h debe ser independiente de λ. Por tanto, no existe estimador insesgado de $1/\lambda$ con $n = 1$.

2.2. Estimador insesgado uniformemente de mínima varianza (UMVUE).

Nos planteamos 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 (es decir, los que tengan una menor dispersión con respecto a su media).

UMVUE.

Definición 7. *Sea X una variable aleatoria con función de distribución en una familia de distribuciones paramétricas, $F \in \{F_\theta, \theta \in \Theta\}$, y $(X_1, \ldots, X_n)$ una muestra aleatoria simple 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:*

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

Propiedades del UMVUE.

Proposición 2.

- *Si existe, el UMVUE es único: si hay dos UMVUE, estos son iguales con probabilidad 1.*
- *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:*
  - *λT_1 es UMVUE para $\lambda g_1(\theta), \forall \lambda \in \mathbb{R}$, y λT_2 es UMVUE para $\lambda g_2(\theta), \forall \lambda \in \mathbb{R}$.*
  - *$T_1 + T_2$ es UMVUE para $g_1(\theta) + g_2(\theta)$.*

Demostración. Demostramos solo el primer punto, pues es el que se suele confundir en los exámenes (se suele pensar que si dos estimadores son UMVUE, entonces su combinación lineal también lo es, y no es así por varias razones; la primera, porque el UMVUE es único, y la segunda, porque la linealidad del UMVUE solo se cumple cuando se tienen UMVUE T_1 y T_2 de funciones paramétricas $g_1(\theta)$ y $g_2(\theta)$, donde $g_1 \neq g_2$).

Si T es UMVUE para $g(\theta)$ y T' es otro estimador cualquiera y T' es otro estimador para $g(\theta)$, entonces por la Observación 4 se tiene que $T + \lambda(T' - T)$ también será un estimador insesgado en $g(\theta)$. Sin embargo, por ser T UMVUE, se cumple:

$$\mathrm{Var}_\theta[T] \leq \mathrm{Var}_\theta[T + \lambda(T' - T)] = \mathrm{Var}_\theta[T] + \lambda^2 \mathrm{Var}_\theta[T - T'] + 2\lambda \mathrm{E}_\theta[T(T' - T)], \quad \forall \lambda \in \mathbb{R}$$

Por tanto, debe cumplirse

$$\mathrm{E}_\theta[T(T' - T)] = 0 \text{ o bien } \mathrm{E}_\theta[TT'] = \mathrm{E}_\theta[T^2]$$

Luego si T' fuese también UMVUE para $g(\theta)$, resultaría

$$\mathrm{E}_\theta[(T - T')^2] = \mathrm{E}_\theta[T^2] + \mathrm{E}_\theta[T'^2] - 2\mathrm{E}_\theta[TT'] = 2\mathrm{E}_\theta[T^2] - 2\mathrm{E}_\theta[TT'] = 0$$

con lo que $T = T'$ con probabilidad 1, para cualquier θ. ■

La definición de UMVUE no es constructiva puesto que lo único que hace es, de entre todos los estimadores admisibles, darle un nombre a aquellos que tienen varianza mínima. Por tanto, es necesario encontrar métodos que permitan obtener el UMVUE para un parámetro.

Teorema de Rao-Blackwell.

Teorema 1. *Sea X una variable aleatoria con función de distribución en una familia de distribuciones paramétricas, $F \in \mathcal{F} = \{F_\theta, \theta \in \Theta\}$, y $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X. Si $T(X_1, \ldots, X_n)$ es un* ***estadístico suficiente*** *para la familia $\mathcal{F}$ y $S(X_1, \ldots, X_n)$ es un* ***estimador insesgado*** *de $g(\theta)$ con* ***momento de segundo orden finito****, entonces:*

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

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

Demostración. En primer lugar, debemos razonar que $\mathrm{E}[S/T]$ es independiente de θ, pero esto es evidente puesto que por ser T suficiente, entonces la distribución de la muestra $(X_1, \ldots, X_n)$ condicionada a T es independiente de θ, luego lo mismo ocurrirá con la distribución de S/T (por la definición de variables aleatorias condicionadas, si la variable condicionante es independiente de un parámetro, cualquier variable condicionada a ella también lo será) y con su esperanza.

a) Por las propiedades de la esperanza condicionada:

$$\mathrm{E}[\mathrm{E}[S/T]] = \mathrm{E}[S] = g(\theta)$$

puesto que S es un estimador insesgado de $g(\theta)$. Por otro lado, como ya sabemos que se cumple $\mathrm{Var}[S/T] = \mathrm{E}[S^2/T] - (\mathrm{E}[S/T])^2 \geq 0$, entonces se tiene que $(\mathrm{E}[S/T])^2 \leq \mathrm{E}[S^2/T]$. Por la linealidad de la esperanza:

$$\mathrm{E}[(\mathrm{E}[S/T])^2] \leq \mathrm{E}[\mathrm{E}[S^2/T]] = \mathrm{E}[S^2] < \infty$$

puesto que ya sabemos que S es un estimador con momento de segundo orden finito.

b) Como $\mathrm{E}[\mathrm{E}[S/T]] = g(\theta)$, entonces:

$$\underbrace{\mathrm{E}[(\mathrm{E}[S/T])^2] - \underbrace{g(\theta)^2}_{(\mathrm{E}[\mathrm{E}[S/T]])^2}}_{\mathrm{Var}[\mathrm{E}[S/T]]} \leq \underbrace{\mathrm{E}[S^2] - \underbrace{g(\theta)^2}_{(\mathrm{E}[S])^2}}_{\mathrm{Var}[S]} \Rightarrow \mathrm{Var}_\theta[\mathrm{E}[S/T]] \leq \mathrm{Var}_\theta[S]$$

■

Teorema de Lehmann-Scheffé.

Teorema 2. *Sea $T(X_1, \ldots, X_n)$ es un estadístico* ***suficiente y completo*** *para la familia de distribuciones $\mathcal{F}$ consideradas. Si $g(\theta)$ admite un estimador insesgado de segundo orden finito $S(X_1, \ldots, X_n)$, entonces existe el UMVUE de $g(\theta)$, que viene dado por*

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

Demostración. Por el Teorema de Rao-Blackwell, si T es un estadístico suficiente y S y S' son estimadores insesgados de $g(\theta)$ de segundo orden finito, entonces $\mathrm{E}[S/T]$ y $\mathrm{E}[S'/T]$ son estimadores insesgados de $g(\theta)$ con momento de segundo orden finito.

Definimos la función $g(T) = \mathrm{E}[S/T] - \mathrm{E}[S'/T]$. Calculamos su esperanza.

$$\mathrm{E}[g(T)] = \mathrm{E}[\mathrm{E}[S/T] - \mathrm{E}[S'/T]] = g(\theta) - g(\theta) = 0$$

Por ser T completo, entonces

$$P\{g(T) = 0\} = 1 \Rightarrow P\{\mathrm{E}[S/T] - \mathrm{E}[S'/T] = 0\} = 1 \Rightarrow P\{\mathrm{E}[S/T] = \mathrm{E}[S'/T]\} = 1$$

Así, existe un único estimador insesgado de $g(\theta)$ función del estadístico suficiente y completo. También por el Teorema de Rao-Blackwell se cumple $\mathrm{Var}[\mathrm{E}[S/T]] \leq \mathrm{Var[S]}$ y $\mathrm{Var}[\mathrm{E}[S'/T]] \leq$ Var[S'], luego:

$$\mathrm{Var}[\mathrm{E}[S'/T]] = \mathrm{Var}[\mathrm{E}[S/T]] \leq \mathrm{Var}[\mathrm{E}[S']]$$

Luego $\mathrm{E}[S/T]$ es UMVUE de $g(\theta)$. ■

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, $\mathrm{E}[S(x_1, \ldots, X_n)/T(X_1, \ldots, X_n)]$ es el UMVUE.
- Buscar una función $h(T)$ con $\mathrm{E}[h(T)] = g(\theta), \forall \theta \in \Theta$ (que sea insesgada en $g(\theta)$), que sea un estimador y tenga momento de segundo orden finito. Entonces, $\mathrm{E}[h(T(X_1, \ldots, X_n))/T(X_1, \ldots, X_n)] = h(T(X_1, \ldots, X_n))$ es el UMVUE.

En cualquier caso, el UMVUE será una función del estadístico suficiente y completo. La diferencia radica en que, para construir tal función, en el primer método se debe encontrar previamente otro estimador insesgado del parámetro con momento de segundo orden finito S, mientras que en el segundo método la función del estadístico se construirá a partir del cálculo de su esperanza, sin tener que buscar tal S.

Ejemplo 9. *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable X con distribución $B(1, p), p \in (0, 1)$. Encontrar el UMVUE para p.*

Solución. En el Ejemplo 6 del Tema 3 se vio que para una distribución $B(1, p)$ un estadístico suficiente y completo es $T = \sum_{i=1}^n X_i$. Por tanto, el UMVUE será una función de T.

$$\mathrm{E}[T] = np \Rightarrow h(T) = \frac{T}{n} \text{ cumple } \mathrm{E}\left[\frac{T}{n}\right] = \frac{np}{n} = p$$

Así, $h(T) = \dfrac{T}{n}$ es candidato a UMVUE.

- Se tiene que $h(T)$ es un estimador de p puesto que $0 \leq \sum_{i=1}^n X_i \leq n \Rightarrow 0 \leq \dfrac{T}{n} \leq 1$.
- $\mathrm{E}\left[\left(\dfrac{T}{n}\right)^2\right] = \dfrac{1}{n^2}\mathrm{E}[T^2] < \infty$.

Por tanto, $h(T) = \dfrac{T}{n}$ es UMVUE para p.

Ejemplo 10. *Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de una variable aleatoria X con distribución $U(0,\theta), \theta \in \mathbb{R}^+$. Encontrar el UMVUE para θ y $1/\theta$.*

Solución. En el Ejemplo 7 del Tema 3 se vio que para una distribución $U(0,\theta)$ un estadístico suficiente y completo es $T = X_{(n)}$. Puesto que la función de densidad de $X_{(n)}$ es

$$f_{X_{(n)}}(x) = \frac{nx^{n-1}}{\theta^n}, 0 < x < \theta$$

se tiene entonces:

$$\mathrm{E}[T] = \int_0^\theta tnt^{n-1}\theta^{-n}dt = \frac{n}{\theta^n}\int_0^\theta t^n dt = \frac{n}{\theta}\left[\frac{t^{n+1}}{n+1}\right]_0^\theta = \frac{n\theta^{n+1}}{\theta^n(n+1)} = \frac{n\theta}{n+1}$$

$\boxed{\theta}$ Tomo $h(T) = \frac{n+1}{n}T \Rightarrow \mathrm{E}[h(T)] = \theta \Rightarrow h(T) = \frac{n+1}{n}T$ candidato a UMVUE.

- Como $X_{(n)} \in \mathbb{R}$, entonces $\frac{n+1}{n}X_{(n)} \in \mathbb{R}$, luego $h(T)$ es estimador.
- $\mathrm{E}[h(T)^2] = \frac{(n+1)^2}{n^2}\mathrm{E}[T^2] = \frac{(n+1)^2}{n^2}\int_0^\theta t^2\frac{nt^{n-1}}{\theta^n} < \infty$ puesto que se está integrando un monomio de grado positivo en un espacio acotado.

Por tanto, $h(T) = \frac{n+1}{n}T$ es UMVUE para θ.

$\boxed{1/\theta}$ En este caso no podemos encontrar la función h de la misma forma que en el apartado anterior pues obtendríamos que h debe depender de θ. Calculamos entonces la esperanza de la transformación $h(T)$ imponiendo que su valor sea $1/\theta$.

$$\mathrm{E}[h(T)] = \int_0^\theta h(t)n\frac{t^{n-1}}{\theta^n}dt = \frac{1}{\theta} \Rightarrow \int_0^\theta h(t)nt^{n-1}dt = \frac{\theta^n}{\theta} \Rightarrow \int_0^\theta h(t)t^{n-1}dt = \frac{\theta^{n-1}}{n}$$

Aplicando el Teorema Fundamental del Cálculo, se tiene que existe H primitiva de $h(t)t^{n-1}$, y se cumple

$$\int_0^\theta h(t)t^{n-1} = H(\theta) - H(0) \underbrace{\Rightarrow}_{\text{Derivo en } \theta} h(\theta)\theta^{n-1} = (n-1)\frac{\theta^{n-2}}{n} \Rightarrow h(\theta) = \frac{n-1}{n\theta}$$

Así, $h(T) = \frac{n-1}{nT}$ es candidato a UMVUE.

- Si $n = 1$, entonces $h(T) = 0 \notin \mathbb{R}^+$. En cualquier otro caso, $\frac{n-1}{nT} \in \mathbb{R}^+$ puesto que $T, \frac{n-1}{n} \in \mathbb{R}^+$. Por tanto, $h(T)$ es estimador para $n > 1$.
- $\mathrm{E}[h(T)^2] = \int_0^\theta \frac{(n-1)^2}{n^2t^2}n\frac{t^{n-1}}{\theta^n}dt = \frac{(n-1)^2}{n\theta^n}\int_0^\theta t^{n-3}dt < \infty \Leftrightarrow n \geq 3$. Por tanto, $h(T)$ tiene momento de orden finito si y solo si $n = 1$ o $n \geq 3$.

Por tanto, $h(T) = \frac{n-1}{nT}$ es UMVUE para $\frac{1}{\theta}$ solo cuando $n \geq 3$.

Ejercicio tipo examen.

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

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

a) Justificar si la distribución de X *es o no exponencial uniparamétrica.*

b) Encontrar un estadístico suficiente y completo.

c) Hallar el UMVUE para θ.

Solución.

a) X no es exponencial uniparamétrica pues el conjunto $\chi = \{x/f_\theta(x) > 0\} = (\theta, +\infty)$ no es independiente del parámetro desconocido θ.

b) Calculamos la función de densidad de la muestra.

$$f_\theta^n = f_\theta(x_1)\cdots f_\theta(x_n) = \frac{1}{2}e^{-(x_1-\theta)/2}\cdots\frac{1}{2}e^{-(x_n-\theta)/2} = [x_i > \theta]$$
$$= \frac{1}{2^n}\exp\left\{\frac{1}{2}\left(\sum_{i=1}^n -(x_i-\theta)\right)\right\} I_{[X_{(1)}>\theta]} = \frac{1}{2^n}\exp\left\{\frac{-1}{2}\left(\sum_{i=1}^n x_i - n\theta\right)\right\} I_{[X_{(1)}>\theta]}$$

Podemos pensar que el estadístico que estamos buscando es $\sum X_i$, pero en este caso, el problema de escoger ese es que vamos a necesitar su distribución que, a priori, no es conocida. Por tanto, el estadístico que se busca en general es aquel del que se puede conocer su distribución. Así, parece que el estadístico que debemos escoger es $X_{(1)}$. Elegimos

$$h(x_1, \dots, x_n) = \frac{1}{2^n}\exp\left(\frac{-1}{2}\sum_{i=1}^n X_i\right), \quad g_\theta(T(X_1, \dots, X_n)) = \exp\left(\frac{-1}{2}(-n\theta)\right) I_{[X_{(1)}>\theta]}$$

Por el teorema de Factorización, $T(X_1, \dots, X_n) = X_{(1)}$ es un estadístico suficiente. Veamos si es completo. Hallamos primero su distribución.

$$f_\theta(x) = \frac{1}{2}e^{-(x-\theta)/2} \Rightarrow F_\theta(x) = \int_\theta^x \frac{1}{2}e^{-(s-\theta)/2}ds = \frac{1}{2}e^{\theta/2}\int_\theta^x e^{-s/2}ds$$
$$= \frac{1}{\cancel{2}}e^{\theta/2}\left[\frac{e^{-s/2}}{-1/\cancel{2}}\right]_\theta^x = e^{\theta/2}(-e^{-x/2} + e^{-\theta/2}) = -e^{\theta/2}e^{-x/2} + 1 = -e^{(\theta-x)/2} + 1$$
$$\Rightarrow F_{X_{(1)}}(x) = 1 - \left(1 + e^{(\theta-x)/2} - 1\right)^n = 1 - \left(e^{(\theta-x)/2}\right)^n$$
$$\Rightarrow f_{X_{(1)}}(x) = 0 + n\left(e^{(\theta-x)/2}\right)^{n-1} e^{(\theta-x)/2}\cdot\frac{1}{2} = \frac{n}{2}e^{n(\theta-x)/2}$$

Sea g una función medible con $\mathrm{E}[g(T)] = 0$. Entonces:

$$\mathrm{E}[g(T)] = \int_0^\infty g(t)\frac{n}{2}e^{n(\theta-t)/2}dt = 0,\ \forall\theta \Leftrightarrow \int_0^\infty g(t)\frac{n}{2}e^{(n\theta-nt)/2}dt = 0,\ \forall\theta$$
$$\Leftrightarrow \underbrace{\frac{n}{2}e^{n\theta/2}}_{\neq 0\ \forall\theta}\int_\theta^\infty g(t)e^{-nt/2}dt = 0,\ \forall\theta \Leftrightarrow \int_\theta^\infty g(t)e^{-nt/2}dt = 0,\ \forall\theta$$

Por el Teorema Fundamental del Cálculo, existe G primitiva de $g(t)e^{-nt/2}$, y se cumple

$$\lim_{x\to\infty} G(x) - G(\theta) = 0\ \forall\theta \underbrace{\Rightarrow}_{\text{Derivo en } \theta} -g(\theta)\underbrace{e^{-n\theta/2}}_{\neq 0\ \forall\theta} \Rightarrow g(\theta) = 0\ \forall\theta \Rightarrow g(T) = 0$$

Se tiene entonces $\{g(T) = 0\} \supseteq \{T \in \mathbb{R}\}$, luego

$$1 \geq P\{g(T) = 0\} \geq P\{T \in \mathbb{R}\} = 1 \Rightarrow P\{g(T) = 0\} = 1$$

Por tanto, $T = X_{(1)}$ es, además, un estadístico completo.

c) Para hallar el UMVUE, tenemos que encontrar un estimador insesgado con momento de segundo orden finito. Sea $h(T)$ tal UMVUE (que debe ser función del estadístico suficiente y completo).

- Imponemos $\mathrm{E}[h(T)] = \theta$.

$$\int_\theta^\infty h(t)\frac{n}{2}e^{n(\theta-t)/2}dt = \theta,\ \forall\theta \Rightarrow \frac{n}{2}e^{n\theta/2}\int_\theta^\infty h(t)e^{-nt/2}dt = \theta,\ \forall\theta$$
$$\Rightarrow \int_\theta^\infty h(t)e^{-nt/2}dt = \frac{2\theta}{ne^{n\theta/2}},\ \forall\theta$$

Por el Teorema Fundamental del Cálculo, existe primitiva H de $h(t)e^{-nt/2}$, y se cumple

$$\lim_{x\to\infty} H(x) - H(\theta) = \frac{2\theta}{ne^{n\theta/2}} = \frac{2}{n}(\theta e^{-n\theta/2})$$

Derivando con respecto a θ se tiene entonces:

$$-h(\theta)\cancel{e^{-n\theta/2}} = \frac{2}{n}\left(\cancel{e^{-n\theta/2}} - \theta\cancel{e^{-n\theta/2}}\frac{-n}{2}\right) \Rightarrow -h(\theta) = \frac{2}{n} + \theta \Rightarrow h(\theta) = -\frac{n\theta+2}{n}$$

Por tanto, $h(T) = -\dfrac{n\theta+2}{n}$ es candidato a UMVUE.

- Como $h(T) \in \mathbb{R}$, entonces $h(T)$ es estimador.
- Se tiene que $\mathrm{E}[h(T)^2] < \infty$ por ser la integral de una exponencial negativa.

Por tanto, $h(T) = -\dfrac{nT+2}{n}$ es UMVUE para θ.

3. Eficiencia.

Ya sabemos que, para que la estimación puntual sea adecuada, debemos encontrar estimadores insesgados de segundo orden finito que minimicen su varianza uniformemente. Nos planteamos ahora cómo acotar la varianza de tales estimadores.

En primer lugar haremos una criba entre las familias de distribuciones teóricas de una variable aleatoria, estando interesados ahora solo en aquellas que cumplen las condiciones de regularidad de Fréchet-Cramér-Rao. Después, definiremos la función de información de Fisher, que nos dará una forma de medir la cantidad de información que una variable aleatoria observable X lleva sobre un parámetro desconocido θ de una distribución regular. Introduciremos el concepto de estadístico regular y, con la función de información de Fisher, obtendremos una cota (de Fréchet-Cramér-Rao) para la varianza de estimadores regulares con momento de segundo orden finito e insesgados en una función paramétrica $g(\theta)$. Finalmente, daremos el concepto de estimador eficiente, que será aquel que es regular, insesgado y cuya varianza alcanza la cota de Fréchet-Cramér-Rao.

3.1. Condiciones de regularidad de Fréchet-Crámer-Rao.

Condiciones de regularidad de Fréchet-Cramér-Rao.

Definición 8. *Sea X una variable aleatoria con distribución de probabilidad en la familia de distribuciones (unidimensional) $\{F_\theta, \theta \in \Theta\}$. Sea $f_\theta(x)$ la función de densidad o la función masa de probabilidad, según el caso, para $\theta \in \Theta \subseteq \mathbb{R}$. 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, $\chi = \{x / f_\theta(x) > 0\}$, *es independiente de* θ.

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

$$\int_\chi \frac{\partial}{\partial\theta} f_\theta(x)dx = \frac{\partial}{\partial\theta}\int_\chi f_\theta(x)dx = 0 \qquad \left(\sum_\chi \frac{\partial}{\partial\theta} f_\theta(x) = \frac{\partial}{\partial\theta}\sum_\chi f_\theta(x) = 0\right), \quad \forall\theta \in \Theta$$

Proposición 3. *Si $X \rightsquigarrow F \in \{F_\theta, \theta \in \Theta\}$ cumple las condiciones de regularidad, entonces la familia de distribuciones asociadas a la muestra aleatoria simple $(X_1, \ldots, X_n)$ de X también las cumple.*

Demostración. Los puntos *i)* y *ii)* también los cumplen las familias de distribuciones exponenciales paramétricas, luego la demostración de estos puntos es similar a la demostración de los puntos *i)* y *ii)* del Teorema 2 del Tema 3.

Demostramos el punto *iii)*. Para ello, debemos tener en cuenta que si $(X_1, \ldots, X_n)$ es una muestra aleatoria simple de X, entonces $X_1, \ldots, X_n$ son variables aleatorias independientes e idénticamente distribuidas a X, que tiene su distribución dentro de una familia de distribuciones regular.

$$\int_{\chi^n} \frac{\partial}{\partial\theta} f_\theta(x_1, \ldots, x_n)dx_1\cdots dx_n = \int_{\chi^n} \underbrace{\frac{\partial}{\partial\theta}\prod_{i=1}^n f_\theta(x_i)}_{\sum_{i=1}^n \frac{\partial f_\theta(x_i)}{\partial\theta}\prod_{j\neq i} f_\theta(x_j)} dx_1\cdots dx_n$$

$$= \sum_{i=1}^n \int_{\chi^n} \frac{\partial f_\theta(x_i)}{\partial\theta} f_\theta(x_1)\cdots f_\theta(x_{i-1})f_\theta(x_{i+1})\cdots f_\theta(x_n)dx_1\cdots dx_n$$

$$= \sum_{i=1}^n \int_\chi \frac{\partial f_\theta(x_i)}{\partial\theta}\left(\int_\chi f_\theta(x_1)\cdots\int_\chi f_\theta(x_{i-1})\int_\chi f_\theta(x_{i+1})\cdots\int_\chi f_\theta(x_n)\right)dx_n\cdots dx_{i+1}dx_{i-1}\cdots dx_1 dx_i$$

$$\underbrace{=}_{*} \sum_{i=1}^n \int_\chi \frac{\partial f_\theta(x_i)}{\partial\theta}dx_i \underbrace{=}_{\text{i.d}} \sum_{i=1}^n 0 = 0 \Rightarrow (X_1, \ldots, X_n) \text{ es regular.}$$

En $*$ hemos usado que f_θ es una función de densidad que integra 1, luego

$$\int_\chi f_\theta(x_1)dx_1 = \cdots = \int_\chi f_\theta(x_{i-1})dx_{i-1} = \int_\chi f_\theta(x_{i+1})dx_{i+1} = \cdots = \int_\chi f_\theta(x_n)dx_n = 1$$
$$\Rightarrow \int_\chi f_\theta(x_1)\cdots \int_\chi f_\theta(x_{i-1}) \int_\chi f_\theta(x_{i+1}) \cdots \int_\chi f_\theta(x_n)dx_n \cdots dx_{i+1}dx_{i-1}\cdots dx_1 = 1$$

■

3.2. Función de información de Fisher.

Función de información de Fisher.

Definición 9. *Sea $X \rightsquigarrow F \in \{F_\theta, \theta \in \Theta\}$, cuya familia de distribuciones es **regular**. Se definen las funciones*

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

Observación 5. *Tengamos en cuenta que la función de información de Fisher solo se define para familias de distribuciones regulares, no para cualquier familia.*

Propiedades de la función de información de Fisher.

Proposición 4. *La función de información de Fisher cumple las siguientes propiedades:*

i) $I_X \geq 0$.

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

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

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

Demostración.

i) Es evidente por ser la esperanza de una variable aleatoria positiva.

ii) $\forall \theta \in \Theta$ se tiene:

$$\mathrm{E}_\theta\left[\frac{\partial \ln f_\theta(X)}{\partial \theta}\right] = \int_\chi \frac{\partial \ln f_\theta(X)}{\partial \theta} f_\theta(x)dx = \int_\chi \frac{1}{\cancel{f_\theta(x)}} \frac{\partial f_\theta(x)}{\partial \theta} \cancel{f_\theta(x)}dx = \int_\chi \frac{\partial f_\theta(x)}{\partial \theta}dx \underbrace{=}_{\text{Regular}} 0$$

$$\mathrm{Var}_\theta\left[\frac{\partial \ln f_\theta(X)}{\partial \theta}\right] = \mathrm{E}_\theta\left[\left(\frac{\partial \ln f_\theta(X)}{\partial \theta}\right)^2\right] - \underbrace{\left(\cancel{\mathrm{E}_\theta\left[\frac{\partial \ln f_\theta(X)}{\partial \theta}\right]}\right)^2}_{=0} \underbrace{=}_{\text{Def.}} I_X(\theta)$$

iii) Teniendo en cuenta que el logaritmo neperiano de un producto es la suma de los logaritmos neperianos, y que la derivada de una suma es la suma de las derivadas, se llega a la demostración de la propiedad que, además, es la extensión de la propiedad anterior (se utiliza la misma propiedad que utilizamos en el siguiente punto).

iv) Teniendo en cuenta que $f_\theta^n(x_1,\dots,x_n) = \prod_{i=1}^{n} f_\theta(x_i)$

$$I_{X_1,\dots,X_n}(\theta) = \mathrm{E}_\theta\left[\left(\frac{\partial \ln f_\theta^n(x_1,\dots,x_n)}{\partial\theta}\right)^2\right] = \mathrm{E}_\theta\left[\left(\frac{\partial \sum_{i=1}^n \ln f_\theta(x_i)}{\partial\theta}\right)^2\right]$$

$$= \mathrm{E}_\theta\left[\sum_{i=1}^n\left(\frac{\partial \ln f_\theta(x_i)}{\partial\theta}\right)^2 + \sum_{i\neq j}\frac{\partial \ln f_\theta(x_i)}{\partial\theta}\cdot\frac{\partial \ln f_\theta(x_j)}{\partial\theta}\right]$$

$$= \sum_{i=1}^n \mathrm{E}_\theta\left[\left(\frac{\partial \ln f_\theta(x_i)}{\partial\theta}\right)^2\right] + \sum_{i\neq j}\mathrm{E}_\theta\left[\frac{\partial \ln f_\theta(x_i)}{\partial\theta}\cdot\frac{\partial \ln f_\theta(x_j)}{\partial\theta}\right] = \sum_{i=1}^n \mathrm{E}_\theta\left[\left(\frac{\partial \ln f_\theta(x_i)}{\partial\theta}\right)^2\right]$$

donde en la última igualdad hemos utilizado que como X_i, X_j son independientes, entonces

$$\sum_{i\neq j}\mathrm{E}_\theta\left[\frac{\partial \ln f_\theta(x_i)}{\partial\theta}\cdot\frac{\partial \ln f_\theta(x_j)}{\partial\theta}\right] = \sum_{i\neq j}\mathrm{E}_\theta\left[\frac{\partial \ln f_\theta(x_i)}{\partial\theta}\right]\mathrm{E}_\theta\left[\frac{\partial \ln f_\theta(x_j)}{\partial\theta}\right] \underbrace{=}_{ii)} 0$$

Como $\mathrm{E}_\theta\left[\left(\frac{\theta \ln f_\theta(x_i)}{\partial\theta}\right)^2\right] = I_X(\theta)$ y todas las distribuciones están idénticamente distribuidas:

$$I_{X_1,\dots,X_n} = \sum_{i=1}^n \mathrm{E}_\theta\left[\left(\frac{\partial \ln f_\theta(x_i)}{\partial\theta}\right)^2\right] = \sum_{i=1}^n I_X(\theta) = nI_X(\theta)$$

■

3.3. Desigualdad de Fréchet-Cramér-Rao.

Estadístico regular.

Definición 10. *Sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de $X \rightsquigarrow F \in \{F_\theta, \theta\in\Theta\}$, cuya familia de distribuciones es* ***regular****, y sea $(x_1,\dots,x_n) = \underline{x}$ una realización muestral (luego $f_\theta^n(x_1,\dots,x_n) = f_\theta^n(\underline{x})$). Un estadístico $T(X_1,\dots,X_n)$ se dice que es regular en el sentido de Fréchet-Cramér-Rao si verifica:*

- *Para el caso discreto:*

$$\frac{\partial}{\partial\theta}E_\theta[T(\underline{x})] = \frac{\partial}{\partial\theta}\sum_{\underline{x}\in\chi^n} T(\underline{x})f_\theta^n(\underline{x}) = \sum_{\underline{x}\in\chi^n} T(\underline{x})\frac{\partial}{\partial\theta}f_\theta^n(\underline{x})$$

- *Para el caso continuo:*

$$\frac{\partial}{\partial\theta}E_\theta[T(\underline{x})] = \frac{\partial}{\partial\theta}\int_{\chi^n} T(\underline{x})f_\theta^n(\underline{x})dx_1\cdots dx_n = \int_{\chi^n} T(\underline{x})\frac{\partial}{\partial\theta}f_\theta^n(\underline{x})dx_1\cdots dx_n$$

Desigualdad de Cauchy-Schwarz.

Lema 1. *Si X e Y son variables aleatorias sobre el mismo espacio de probabilidad, cada una con momento de segundo orden finito. Entonces:*

I) $(E[XY])^2 \leq E[X^2]E[Y^2]$.

II) Si una de las dos variables es degenerada en 0, o las dos variables son degeneradas, se da la igualdad en I).

III) Si ninguna de las dos variables es degenerada, entonces

$$(E[XY])^2 = E[X^2]E[Y^2] \Leftrightarrow \exists a \neq 0, \exists b \neq 0 \ : \ P(aX + bY = 0) = 1$$

Cota de Fréchet-Cramér-Rao.

Teorema 3. *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple 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 finito 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 solo si $\exists a(\theta_0) \neq 0$ tal que

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

Demostración. Definimos:

$$I_{X_1,\ldots,X_n}(\theta) = \mathrm{E}\left[\left(\frac{\partial \ln f_\theta^n(x_1, \ldots, x_n)}{\partial\theta}\right)^2\right] = \mathrm{E}[Y^2], \ \text{con } Y = \frac{\partial \ln f_\theta^n(x_1, \ldots, x_n)}{\partial\theta}$$

$$\mathrm{Var}[T] = \mathrm{E}[(T - \mathrm{E}[T])^2] = \mathrm{E}[X^2], \ \text{con } X = T - g(\theta)$$

i) Por la Desigualdad de Cauchy-Schwarz, $\mathrm{E}[X^2] \cdot \mathrm{E}[Y^2] \geq (\mathrm{E}[XY])^2$, es decir:

$$I_{X_1,\ldots,X_n}(\theta) \cdot \mathrm{Var}_\theta T(X_1, \ldots, X_n) \geq (\mathrm{E}[XY])^2 \Rightarrow \mathrm{Var}_\theta T(X_1, \ldots, X_n) \geq \frac{(\mathrm{E}[XY])^2}{I_{X_1,\ldots,X_n}(\theta)} \qquad (*)$$

Calculamos $\mathrm{E}[XY]$, y después solo habrá que sustituir su valor en $(*)$.

$$\begin{aligned}
\mathrm{E}[XY] &= \mathrm{E}\left[(T - g(\theta))\frac{\partial \ln f_\theta^n(x_1, \ldots, x_n)}{\partial\theta}\right] \\
&= \mathrm{E}\left[T\frac{\partial \ln f_\theta^n(x_1, \ldots, x_n)}{\partial\theta}\right] - \underbrace{\mathrm{E}\left[g(\theta)\frac{\partial \ln f_\theta^n(x_1, \ldots, x_n)}{\partial\theta}\right]}_{g(\theta)\mathrm{E}\left[\cancel{\frac{\partial \ln f_\theta^n(x_1,\ldots,x_n)}{\partial\theta}}\right]=0} \\
&= \int_{\chi^n} T\frac{\partial \ln f_\theta^n(x_1, \ldots, x_n)}{\partial\theta} f_\theta^n(x_1, \ldots, x_n)dx_1 \cdots dx_n \underbrace{=}_{\text{Regular}} \frac{\partial}{\partial\theta}\underbrace{\int_{\chi^n} T f_\theta^n(x_1, \ldots, x_n)}_{\mathrm{E}[T]=g(\theta)} = g'(\theta)
\end{aligned}$$

ii) La igualdad en la desigualdad (∗), según el lema de Cauchy Schwarz, se da cuando alguna de las dos variables es degenerada, o cuando $\exists a, b \in \mathbb{R}, a, b \neq 0$, tales que $P[aX + bY = 0] = 1$.

- Y no puede ser degenerada, pues

$$Y = \frac{\partial \ln f_\theta^n(x_1, \ldots, x_n)}{\partial \theta} \Rightarrow \text{Var}_\theta \left[\frac{\partial \ln f_\theta^n(x_1, \ldots, x_n)}{\partial \theta} \right] = I_{X_1, \ldots, X_n}(\theta) > 0 \text{ por hipótesis.}$$

- Comprobemos que X tampoco puede ser degenerada.

$$\text{Var}_\theta[X] = \text{Var}_\theta[T - g(\theta)] = \text{Var}_\theta[T] \geq \frac{(g'(\theta))^2 \ (> 0)}{I_{X_1, \ldots, X_n}(\theta) \ (> 0)} > 0$$

- Por tanto, debe darse la otra opción:

$$\exists a, b \in \mathbb{R}, a, b \neq 0 \text{ tales que } P\left[a(T - g(\theta)) + b \frac{\partial \ln f_\theta(x_1 n \ldots, x_n)}{\partial \theta} = 0 \right] = 1$$

$$\Rightarrow \underbrace{\exists a, b \in \mathbb{R}, a, b \neq 0}_{\exists a(\theta_0) \neq 0} \quad \text{tal que} \quad \underbrace{P\left[\frac{\partial \ln f_\theta^n(x_1, \ldots, x_n)}{\partial \theta} = \frac{-a}{b}(T - g(\theta)) \right] = 1}_{P_{\theta_0}\left[\left. \frac{\partial \ln f_\theta(X_1, \ldots, X_n)}{\partial \theta} \right|_{\theta = \theta_0} = a(\theta_0)[T(X_1, \ldots, X_n) - g(\theta_0)] \right] = 1}$$

donde la última igualdad se da puesto que el valor de $\dfrac{-a}{b}$ depende del valor concreto de θ_0 en el que se evalúe la derivada.

■

3.4. Estimadores eficientes.

Estimador eficiente.

Definición 11. *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,*

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

Lema 2. *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:*

i) $g(\theta)$ es constante y, en tal caso, $T(X_1, \ldots, X_n)$ es degenerado.

ii) $g(\theta)$ es estrictamente monótona: $g'(\theta) > 0\,, \forall \theta \in \Theta$ o $g'(\theta) < 0$, $\forall \theta \in \Theta$.

Demostración.

i) Si $g(\theta)$ es constante para todo θ, entonces $g'(\theta) = 0$ para todo θ. Así, la cota para la varianza de un estimador T de $g(\theta)$ será:

$$\text{Var}_\theta[T] \geq \frac{(g'(\theta))^2}{I_{X_1,\ldots,X_n}(\theta)} = 0$$

Si T es eficiente, entonces $\text{Var}[T] = 0$, es decir, $T(X_1, \ldots, X_n)$ toma un único valor, luego es degenerado.

ii) Si g no es estrictamente monótona, entonces existe un punto θ_0 tal que $g'(\theta_0) = 0$. En ese caso:

$$\text{Var}_{\theta_0}[T] = \frac{(g'(\theta_0))^2}{I_{X_1,\ldots,X_n}} = 0 \Rightarrow P_{\theta_0}[T = c] = 1$$

Como T es estadístico (por tanto, independiente del parámetro), se tiene entonces:

$$P[T(X_1, \ldots, X_n) = c] = 1, \ \forall \theta \Rightarrow T = c$$

y hemos llegado a una contradicción. ■

Caracterización de estimadores eficientes.

Teorema 4. *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple 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 $T(X_1, \ldots, X_n)$ un estimador de $g(\theta)$. Una condición necesaria y suficiente para que T sea eficiente es:*

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

Demostración.

$\boxed{\Rightarrow}$ Supongamos que T es eficiente.

i) Si T es eficiente, entonces se da la igualdad en la desigualdad del teorema de la Cota de Fréchet-Cramér-Rao (Teorema 3), que se da justo cuando

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

ii)

$$I_{X_1,\ldots,X_n}(\theta) = \text{Var}_\theta\left[\frac{\partial \ln f_\theta^n(x_1, \ldots, x_n)}{\partial \theta}\right] = \text{Var}_\theta\left[a(\theta)(T(X_1, \ldots, X_n) - g(\theta))\right]$$

$$= a^2(\theta)\text{Var}_\theta[T - g(\theta)] = a^2(\theta)\text{Var}_\theta[T] \underbrace{=}_{T \text{ efi}} = a^2(\theta)\frac{(g'(\theta))^2}{I_{X_1,\ldots,X_n}(\theta)}$$

Por tanto, se tiene que:

$$I^2 = a^2 \cdot (g')^2 \Rightarrow \sqrt{I^2} = \sqrt{a^2 \cdot (g')^2} \Rightarrow \underbrace{|I|}_{>0} = |a \cdot g'|$$

Como a y g' tienen el mismo signo se tiene entonces $a \cdot g' > 0$, luego $I = a(\theta) \cdot g'(\theta)$.

Para ver que T es eficiente, se tiene que demostrar que es regular, insesgado, y que alcanza la cota de FCR.

- Que alcanza la cota, de nuevo, es evidente por i) y por el Teorema 3.
- Veamos que es insesgado.

$$0 = \mathrm{E}_\theta\left[\frac{\partial \ln f_\theta^n(x_1,\dots,x_n)}{\partial\theta}\right] = \mathrm{E}_\theta[a(\theta)(T-g(\theta))] = a(\theta)\underbrace{[\mathrm{E}_\theta[T]-g(\theta)]}_{=0} \Rightarrow \mathrm{E}_\theta[T] = g(\theta)$$

- Comprobemos que es regular. Para ello, debemos comprobar que se cumple:

$$\underbrace{\frac{\partial}{\partial\theta}\int_{\chi^n} T f_\theta^n dx_1\cdots dx_n}_{=\frac{\partial}{\partial\theta}\mathrm{E}_\theta[T]=\frac{\partial}{\partial\theta}g(\theta)=g'(\theta)} = \int_{\chi^n} T\frac{\partial}{\partial\theta}f_\theta^n dx_1\cdots dx_n$$

$$\int_{\chi^n} T\cdot\frac{\partial f_\theta^n(x_1,\dots,x_n)}{\partial\theta}dx_1\cdots dx_n = \int_{\chi^n} T\cdot\underbrace{\frac{\partial f_\theta^n(x_1,\dots,x_n)}{\partial\theta}\cdot\frac{f_\theta^n(x_1,\dots,x_n)}{f_\theta^n(x_1,\dots,x_n)}}_{\frac{\partial\ln f_\theta^n}{\partial\theta}=\frac{1}{f_\theta^n}\cdot\frac{\partial f_\theta^n}{\partial\theta}}dx_1\cdots dx_n$$

$$= \int_{\chi^n} T\cdot\frac{\partial\ln f_\theta^n(x_1,\dots,x_n)}{\partial\theta}\cdot f_\theta^n(x_1,\dots,x_n)dx_1\cdots dx_n = \mathrm{E}_\theta\left[T\frac{\partial\ln f_\theta^n(x_1,\dots,x_n)}{\partial\theta}\right]$$

$$= \mathrm{E}_\theta\left[T(a(\theta)(T-g(\theta)))\right] = a(\theta)\cdot\mathrm{E}_\theta\left[T^2-Tg(\theta)\right] = a(\theta)\left[\mathrm{E}_\theta[T^2]-g(\theta)\mathrm{E}_\theta[T]\right]$$

$$= a(\theta)\underbrace{\left[\mathrm{E}_\theta[T^2]-(\mathrm{E}_\theta[T])^2\right]}_{=\mathrm{Var}_\theta[T]} = a(\theta)\mathrm{Var}_\theta[T] = a(\theta)\frac{(g'(\theta))^2}{I_{X_1,\dots,X_n}(\theta)} = \cancel{a(\theta)}\frac{(g'(\theta))^{\cancel{2}}}{a(\theta)\cancel{g'(\theta)}} = g'(\theta)$$

■

Importante para el examen.

Corolario 1.

i) *Si $T(X_1,\dots,X_n)$ 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.*

ii) *Si una función paramétrica admite dos estimadores eficientes, estos son iguales con probabilidad 1, bajo todas las distribuciones de la familia.*

iii) *Solo existen estimadores eficientes en familias de tipo exponencial paramétrico.*

Demostración.

i) Si T es eficiente para $g(\theta)$, entonces sabemos que, para cierta $a(\theta)$, se tiene

$$\begin{cases}(*) = P_\theta\left[\dfrac{\partial\ln f_\theta^n(X_1,\dots,X_n)}{\partial\theta} = a(\theta)[T(X_1,\dots,X_n)-g(\theta)]\right] = 1\\ I_{X_1,\dots,X_n}(\theta) = a(\theta)g'(\theta)\end{cases}$$

Comprobemos que se cumplen para $ag(\theta) + b$.

- Multiplicamos y dividimos primero por a, y después sumamos y restamos b, en la expresión $T - g(\theta)$, para obtener:

$$T - g(\theta) = \frac{1}{a}[aT - ag(\theta)] = \frac{1}{a}[aT + b - ag(\theta) - b] = \frac{1}{a}[aT + b - (ag(\theta) + b)]$$

Por tanto:

$$(*) = P_\theta \left[\frac{\partial \ln f_\theta^n(X_1, \ldots, X_n)}{\partial \theta} = a^*(\theta)[aT(X_1, \ldots, X_n) + b - (ag(\theta) + b)] \right] = 1$$

donde hemos elegido $a^*(\theta) = \dfrac{a(\theta)}{a}$.

- Comprobamos que $a^*(\theta)(g(\theta) + b)' = I_{X_1,\ldots,X_n}(\theta)$.

$$a^*(\theta)(ag(\theta) + b)' = \frac{a(\theta)}{\cancel{a}} \cancel{a} g(\theta) = a(\theta)g'(\theta) = I_{X_1,\ldots,X_n}(\theta)$$

Demostremos ahora que los correspondientes estimadores eficientes son $aT + b$ con probabilidad 1 bajo todas las distribuciones de la familia. Para ello, debemos ver que si dos estimadores T_1 y T_2 son eficientes para las funciones paramétricas $g_1(\theta)$ y $g_2(\theta)$, respectivamente, entonces existe una relación lineal entre T_1 y T_2.

- T_1 eficiente para $g_1(\theta) \Rightarrow P_\theta \left[\dfrac{\partial \ln f_\theta^n(X_1, \ldots, X_n)}{\partial \theta} = a_1(\theta)[T_1(X_1, \ldots, X_n) - g_1(\theta)] \right] = 1.$
- T_2 eficiente para $g_2(\theta) \Rightarrow P_\theta \left[\dfrac{\partial \ln f_\theta^n(X_1, \ldots, X_n)}{\partial \theta} = a_2(\theta)[T_2(X_1, \ldots, X_n) - g_2(\theta)] \right] = 1.$

Por tanto, se tiene que $P_\theta[a_1(\theta)T_1 - a_1(\theta)g_1(\theta) = a_2(\theta)T_2 - a_2(\theta)g_2(\theta)] = 1$ si y solo si se cumple:

$$P_\theta \left[T_1 = \frac{a_2(\theta)T_2}{a_1(\theta)} - \frac{a_2(\theta)g_2(\theta) - a_1(\theta)g_1(\theta)}{a_1(\theta)} \right] = 1$$

Así que, en efecto, existe una relación lineal entre T_1 y T_2.

ii) Por el punto i), si T_1 y T_2 son estimadores eficientes de la misma función paramétrica, entonces $P[T_1 = aT_2 + b] = 1$. Tomando $a = 1$ y $b = 0$ se llega a que $P[T_1 = T_2] = 1$, luego en efecto ambos estimadores son iguales con probabilidad 1.

iii) En primer lugar, las únicas familias de distribuciones que admiten estimadores eficientes son las regulares, que comparten con las familias exponenciales paramétricas la independencia de χ^n del parámetro desconocido y la presencia del parámetro en intervalos abiertos de $\mathbb{R}$. Por tanto, queda demostrar que si la familia de distribuciones admite estimador eficiente, entonces se cumple que existen funciones real-valuadas $Q(\theta)$ y $D(\theta)$, definidas sobre Θ, y existen funciones medibles Borel T y S, también real valuadas, tales que:

$$\forall \theta \in \Theta, \; f_\theta(x) = \exp[Q(\theta)T(x) + D(\theta) + S(x)], \; x \in \chi$$

Si la familia de distribuciones admite estimador eficiente, entonces se verifica

$$\frac{\partial \ln f_\theta^n(x_1, \ldots, x_n)}{\partial \theta} = a(\theta)(T - g(\theta)) = a(\theta)T - a(\theta)g(\theta)$$

Por tanto, tomando el estadístico T como la función Borel T de la condición para ser exponencial, y teniendo en cuenta que una función S constante no afecta a la derivada de una función, se tiene que

$$\ln f_\theta^n = T \underbrace{\int_\theta a(\theta)d\theta}_{Q(\theta)} - \underbrace{\int_\theta a(\theta)g(\theta)d\theta}_{D(\theta)} + \underbrace{S(x_1,\ldots,x_n)}_{=\text{ constante}}$$

Por tanto, en efecto la familia es exponencial paramétrica. ∎

Ejemplo 12. *Buscar la clase de funciones paramétricas que admiten estimadores eficientes para las siguientes familias de distribuciones y calcular dichos estimadores:*

a) $\{B(k_0,p), p \in (0,1)\}$.

b) $\{\mathcal{N}(\mu,\sigma_0^2), \mu \in \mathbb{R}\}$.

c) $\{\mathcal{N}(\mu_0,\sigma^2), \sigma^2 > 0\}$.

Solución. Utilizaremos el punto iii) del Corolario 1, escribiendo el logaritmo de las funciones de probabilidad muestrales como

$$\ln f_\theta^n = T \cdot Q(\theta) - D(\theta) + S(x_1,\ldots,x_n)$$

teniendo en cuenta que $Q(\theta) = \int a(\theta)d\theta$, $D(\theta) = \int a(\theta)g(\theta)d\theta$ y que $S(x_1,\ldots,x_n)$ es constante.

a) En este caso, la función masa de probabilidad muestral de una variable aleatoria $X \rightsquigarrow B(k_0,p)$ con $p \in (0,1)$ sabemos que viene dada por

$$f_p^n = \prod_{i=1}^n \binom{k_0}{x} p^{\sum x_i}(1-p)^{nk_0 - \sum x_i}$$

Por tanto:

$$\begin{aligned}
\ln f_p^n &= \underbrace{\ln\left(\prod_{i=1}^n \binom{k_0}{x}\right)}_{=S(x_1,\ldots,x_n)} + \ln p^{\sum x_i} + \ln(1-p)^{nk_0-\sum x_i} \\
&= S(x_1,\ldots,x_n) + \sum_{i=1}^n x_i \ln p + nk_0 \ln(1-p) - \sum_{i=1}^n x_i \ln(1-p) \\
&= S(x_1,\ldots,x_n) \underbrace{\sum_{i=1}^n}_{T} \underbrace{(\ln p - \ln(1-p))}_{Q(p)} + \underbrace{\underbrace{nk_0 \ln(1-p)}_{-(-nk_0\ln(1-p))}}_{D(p)}
\end{aligned}$$

Así, se tiene:

$$Q(p) = \ln p - \ln(1-p) = \int a(p)dp \Rightarrow a(p) = (\ln p - \ln(1-p))' = \frac{1}{p} + \frac{1}{1-p} \Rightarrow a(p) = \frac{1}{p(1-p)}$$

Por otra parte:

$$D(p) = -nk_0\ln(1-p) = \int a(p)g(p)dp \Rightarrow a(p)g(p) = (-nk_0\ln(1-p))' = -nk_0(\ln(1-p))'$$

$$\Rightarrow a(p)g(p) = -nk_0\left(\frac{1}{1-p}\cdot(-1)\right) = nk_0\frac{1}{1-p} = \underbrace{\frac{1}{p(1-p)}}_{a(p)}\cdot g(p) \Rightarrow g(p) = nk_0 p$$

Por tanto, la clase de funciones paramétricas que admiten estimador eficiente está compuesta por las funciones de la forma $ag(p)+b = a\cdot nk_0p+b$ con $a,b\in\mathbb{R}$, y la correspondiente familia de estimadores la completan las funciones de la forma $aT+b = a\cdot\sum_{i=1}^n X_i + b$ con $a,b\in\mathbb{R}$ (habría que comprobar que T es estimador de $g(p) = nk_0p$, pero es evidente que lo es pues ambos toman valores reales). Queda entonces demostrar que T es, en efecto, eficiente. Para ello, demostramos que $I_{X_1,\dots,X_n}(p) = a(p)g'(p)$.

$$\begin{aligned} I_{X_1,\dots,X_n}(p) &= \mathrm{Var}_p\left[\frac{\partial \ln f_p^n(x_1,\dots,x_n)}{\partial p}\right] = \mathrm{Var}_p\left[\frac{\sum_{i=1}^n x_i}{p(1-p)} - \frac{nk_0}{p(1-p)}\right] \\ &= \frac{1}{p^2(1-p)^2}\mathrm{Var}_p\left[\sum_{i=1}^n X_i\right] = \frac{nk_0\cancel{p}\cancel{(1-p)}}{p^{\cancel{2}}(1-p)^{\cancel{2}}} = \frac{nk_0}{p(1-p)} \end{aligned}$$

Por otro lado, como $g(p) = nk_0p \Rightarrow g'(p) = nk_0$ y, por tanto:

$$\Rightarrow a(p)g'(p) = \frac{1}{p(1-p)}\cdot nk_0 = \frac{nk_0}{p(1-p)} = I_{X_1,\dots,X_n}(p)$$

Así, $aT+b$ es en efecto una familia de estimadores eficientes para $ag(p)+b$.

Otra opción sería escoger $g(p) = k_0p$, $a(p) = \frac{n}{p(1-p)}$ y $T = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}$, elección para la que habría que seguir el mismo procedimiento que hemos descrito.

Para hacer los apartados b) y c) debemos previamente estudiar $\ln f^n_{\mu,\sigma^2}$, donde f^n_{μ,σ^2} es la función de densidad muestral de una variable aleatoria $\mathcal{N}(\mu,\sigma^2)$.

La función de densidad muestral de $\mathcal{N}(\mu,\sigma^2)$ se vio en el Ejercicio 2 del Tema 1, y recordemos que viene dada por

$$f^n_{\mu,\sigma^2}(x_1,\dots,x_n) = \frac{1}{(\sigma\sqrt{n})^n}\exp\left(\frac{-1}{2}\sum_{i=1}^n\frac{(x_i-\mu)^2}{\sigma^2}\right).$$

Por tanto, su logaritmo neperiano es:

$$\begin{aligned} \ln f^n_{\mu,\sigma^2} &= \ln\left(\frac{1}{(\sigma\sqrt{2\pi})^n}\cdot\exp\left(\frac{-1}{2}\sum_{i=1}^n\frac{(x_i-\mu)^2}{\sigma^2}\right)\right) = \ln\left(\frac{1}{(\sigma\sqrt{2\pi})^n}\right) + \left(\frac{-1}{2}\sum_{i=1}^n\frac{(x_i-\mu)^2}{\sigma^2}\right) \\ &= -n\ln\sigma - n\ln(\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^n x_i - \frac{n\mu^2}{2\sigma^2} \end{aligned}$$

b) Consideramos σ_0^2 conocido. Así:

$$\ln f_{\mu,\sigma_0^2} = \underbrace{-n\ln\sigma_0 - n\ln(\sqrt{2\pi})}_{S(x_1,\dots,x_n)} - \frac{1}{2\sigma_0^2}\sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma_0^2}\sum_{i=1}^n x_i - \frac{n\mu^2}{2\sigma_0^2} = S + \underbrace{\sum_{i=1}^n x_i}_{T}\underbrace{\frac{\mu}{\sigma_0^2}}_{Q(\mu)} - \underbrace{\frac{n\mu^2}{2\sigma_0^2}}_{D(\mu)}$$

Por tanto:

$$Q(\mu) = \frac{\mu}{\sigma_0^2} = \int a(\mu)d\mu \Rightarrow a(\mu) = \left(\frac{\mu}{\sigma_0^2}\right)' \Rightarrow a(\mu) = \frac{1}{\sigma_0^2}$$

$$D(\mu)=\frac{n\mu^2}{2\sigma^2}=\int a(\mu)g(\mu)d\mu \Rightarrow a(\mu)g(\mu)=\left(\frac{n\mu^2}{2\sigma_0^2}\right)'=\frac{n}{2\sigma_0^2}(\mu^2)'=\frac{\cancel{2}\mu n}{\cancel{2}\sigma_0^2}=\frac{\mu n}{\sigma_0^2}$$

$$\Rightarrow \frac{\mu n}{\sigma_0^2}=a(\mu)g(\mu)=\frac{1}{\sigma_0^2}g(\mu)\Rightarrow g(\mu)=\mu n$$

Por tanto, la clase de funciones paramétricas que admiten estimador eficiente está compuesta por las funciones de la forma $ag(\mu)+b=a\cdot\mu n+b$ con $a,b\in\mathbb{R}$, y la correspondiente familia de estimadores la completan las funciones de la forma $aT+b=a\cdot\sum_{i=1}^n X_i+b$ con $a,b\in\mathbb{R}$ (habría que comprobar que T es estimador de $g(\mu)=\mu n$, pero es evidente que lo es pues ambos toman valores reales). Queda entonces demostrar que T es, en efecto, eficiente. Para ello demostraremos que se da la igualdad $I_{X_1,\dots,X_n}=a(\mu)g'(\mu)$.

$$a(\mu)=\frac{1}{\sigma_0^2},\; g(\mu)=\mu n\Rightarrow g'(\mu)=n\Rightarrow a(\mu)g'(\mu)=\frac{n}{\sigma_0^2}$$

Por otro lado:

$$I_{X_1,\dots,X_n}(\mu)=\mathrm{Var}_\mu\left[\frac{\partial \ln f_\mu^n(x_1,\dots,x_n)}{\partial\mu}\right]=\mathrm{Var}_\mu\left[\frac{\partial}{\partial\mu}\left(\frac{\mu}{\sigma_0^2}\sum_{i=1}^n X_i-\frac{n\mu^2}{2\sigma_0^2}\right)\right]$$

$$=\mathrm{Var}\left[\frac{\sum_{i=1}^n X_i}{\sigma_0^2}-\frac{\cancel{2}n\mu}{\cancel{2}\mu_0^2}\right]=\frac{1}{\sigma_0^4}\mathrm{Var}_\mu\left[\sum_{i=1}^n X_i-n\mu\right]=\frac{1}{\sigma_0^4}\mathrm{Var}_\mu\left[\sum_{i=1}^n X_i\right]$$

$$=\frac{1}{\sigma_0^2}\sum_{i=1}^n\mathrm{Var}_\mu[X_i]=\frac{1}{\sigma_0^4}n\sigma_0^2=\frac{n}{\sigma_0^2}=a(\mu)g'(\mu)$$

Por tanto, $aT+b$ es una familia de estimadores eficientes para $ag(\mu)+b$.

Otra opción sería escoger $g(\mu)=\mu$, $a(\mu)=\frac{n}{\sigma_0^2}$ y $T=\frac{1}{n}\sum_{i=1}^n X_i=\bar{X}$, elección para la que se debería repetir el procedimiento seguido.

c) Consideramos μ_0 conocido. Así:

$$\ln f^n_{\mu_0,\sigma^2}=-n\ln(\sqrt{2\pi})-n\ln\sigma-\frac{1}{2\sigma^2}\sum_{i=1}^n x_i^2+\frac{\mu}{\sigma^2}\sum_{i=1}^n x_i-\frac{n\mu^2}{2\sigma^2}$$

$$=\underbrace{-n\ln(\sqrt{2\pi})}_{S(x_1,\dots,x_n)}\underbrace{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n x_i^2+\frac{\mu}{\sigma_0^2}\sum_{i=1}^n x_i}_{T\cdot Q(\sigma^2)}-\underbrace{\underbrace{n\ln\sigma-\frac{n\mu^2}{2\sigma^2}}_{-(n\ln\sigma+\frac{n\mu^2}{2\sigma^2})}}_{D(\sigma^2)}$$

Por tanto, en este caso debemos considerar T y $Q(\sigma^2)$ como funciones bidimensionales y considerar su producto como un producto escalar:

$$T=\left(\sum_{i=1}^n X_i^2,\sum_{i=1}^n X_i\right),\; Q(\sigma^2)=\left(\frac{-1}{2\sigma^2},\frac{\mu}{\sigma^2}\right)$$

Así, en este caso $a(\sigma^2)$ también será una función bidimensional cuyas componentes se calculan como las derivadas respecto de σ^2 de las componentes de $Q(\sigma^2)$.

$$a_1(\sigma^2)=\left(\frac{-1}{2\sigma^2}\right)'=\frac{1}{2\sigma^4},\; a_2(\sigma^2)=\left(\frac{\mu}{\sigma^2}\right)'=\frac{-\mu}{\sigma^4}\Rightarrow a(\sigma^2)=\left(\frac{1}{2\sigma^4},\frac{-\mu}{\sigma^4}\right)$$

Por otro lado, $D(\sigma^2)$ también será bidimensional, pero en este caso es más difícil decidir cuáles son sus componentes. Puesto que en $a_1(\sigma^2)$ no está μ, entonces $D_1(\sigma^2)$ será aquel sumando donde tampoco se encuentre μ. Así:

$$D(\sigma^2) = \left(n \ln \sigma, \frac{n\mu^2}{2\sigma^2}\right) = (D_1(\sigma^2), D_2(\sigma^2))$$

Por tanto, se cumplen las igualdades $(D_1(\sigma^2))' = a_1(\sigma^2)g_1(\sigma^2)$ y $(D_2(\sigma^2))' = a_2(\sigma^2)g_12(\sigma^2)$. Hagamos, para hallar $(D_1(\sigma^2))'$, el cambio $\sigma^2 = x^2$ para que nos resulte más sencillo derivarla.

$$(D_1(x^2))' = (n \ln x^{1/2})' = \frac{n}{2}(\ln x)' = \frac{n}{2} \cdot \frac{1}{x} = \frac{n}{2\sigma^2} = (D_1(\sigma^2))'$$

$$\Rightarrow \frac{n}{2\sigma^2} = a_1(\sigma^2)g_1(\sigma^2) = \frac{1}{2\sigma^4}g_1(\sigma^2) \Rightarrow g_1(\sigma^2) = n\sigma^2$$

Por otro lado:

$$(D_2(\sigma^2))' = \frac{-n\mu^2}{2\sigma^4} = a_2(\sigma^2)g_2(\sigma^2) = \frac{-\mu}{\sigma^4}g_2(\sigma^2) \Rightarrow g_2(\sigma^2) = \frac{n\mu^{\cancel{2}}\cancel{\sigma^4}}{2\cancel{\sigma^4}\cancel{\mu}} = \frac{n\mu}{2}$$

Por tanto:

$$g(\sigma^2) = \left(n\sigma^2, \frac{n\mu}{2}\right)$$

Así, la clase de funciones paramétricas que admiten estimador eficiente (de nuevo habría que comprobar que lo son, pero es evidente hacerlo) está compuesta por las funciones de la forma

$$ag(\sigma^2) + b = a \cdot \left(n\sigma^2, \frac{n\mu}{2}\right) + b, \text{con } a, b \in \mathbb{R}$$

y la correspondiente familia de estimadores la completan las funciones de la forma

$$aT + b = a \cdot \left(\sum_{i=1}^{n} X_i^2, \sum_{i=1}^{n} X_i\right) + b, \text{con } a, b \in \mathbb{R}.$$

Así, como $g_2(\sigma^2)$ es constante y no aporta información sobre σ^2, podemos obviar a $a_2(\sigma^2)$ y a $g_2(\sigma^2)$ y centrarnos en trabajar con las primeras componentes de cada función.

Por tanto, habría que comprobar ahora que $I_{X_1,\ldots,X_n}(\sigma^2) = a_1(\sigma^2)g_1(\sigma^2)$, donde $I_{X_1,\ldots,X_n}(\sigma^2)$ se calcula como la varianza de la derivada de los sumandos de la función de densidad muestral donde no aparece μ. Sin embargo, para calcular $I_{X_1,\ldots,X_n}(\sigma^2)$ deberíamos saber cuál es la varianza de la suma de cuadrados de distribuciones normales, para lo que tendríamos que saber la varianza del cuadrado de una normal, y no la sabemos. Por tanto, en este caso no comprobamos que T sea eficiente.

Ejercicio tipo examen.

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

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

a) Sabiendo que $E_\theta[X] = \frac{2}{\theta}$ y $Var_\theta[X] = \frac{2}{\theta^2}$, comprobar que se cumplen las condiciones de regu-

laridad de FCR y calcular la función de información asociada a la muestra.

b) Dar la clase de funciones paramétricas que admiten estimador eficiente y calcular dichos estimadores.

c) Calcular la cota de la varianza de estimadores regulares, insesgados en $\frac{2}{\theta}$. ¿Es dicha cota alcanzable?

Solución. Sea X una variable aleatoria con función de densidad $f_\theta(x) = \theta^2 x e^{-\theta x}$, $x > 0, \theta > 0$.

a) Comprobamos que la familia de distribuciones de X es regular.

i) $\Theta = \mathbb{R}^+$ intervalo abierto de $\mathbb{R}$.

ii) $\chi = \{x / f_\theta(x) > 0\} = \mathbb{R}^+$ independiente de θ.

iii) Hay que comprobar que se cumple

$$\frac{\partial}{\partial\theta}\int_\chi f_\theta(x)dx = 0 = \int_\chi \frac{\partial}{\partial\theta} f_\theta(x)dx$$

Sin embargo, la primera igualdad es evidente puesto que por ser $f_\theta(x)$ una función de densidad, integra 1 en su dominio y, por tanto, su derivada es 0. Así, debemos demostrar la segunda igualdad.

$$\begin{aligned}\frac{\partial}{\partial\theta} f_\theta(x) &= (\theta^2)'(xe^{-\theta x}) + \theta^2(xe^{-\theta x})' = 2\theta x e^{-\theta x} + \theta^2 x e^{-\theta x}\cdot(-x) = 2\theta x e^{-\theta x} - x\theta^2 x e^{-\theta x} \\ &= 2\cdot\frac{f_\theta(x)}{\theta} - x\cdot f_\theta(x) \Rightarrow \frac{\partial f_\theta(x)}{\partial\theta} = 2\frac{f_\theta(x)}{\theta} - x f_\theta(x) \\ &\Rightarrow \int_\chi \frac{\partial f_\theta(x)}{\partial\theta}dx = \int_\chi \left(\frac{2}{\theta}f_\theta(x) - x f_\theta(x)\right)dx = \frac{2}{\theta} - \mathrm{E}_\theta[X] = \frac{2}{\theta} - \frac{2}{\theta} = 0\end{aligned}$$

Por tanto, la familia de distribuciones de X es regular.

Calculamos ahora la función de información asociada a la muestra. En general, se podrán seguir los siguientes pasos:

I) Se halla la derivada de $\ln f_\theta(x)$ con respecto a θ.

$$\ln f_\theta(x) = 2\ln\theta + \ln x - \theta x \Rightarrow \frac{\partial \ln f_\theta(x)}{\partial\theta} = \frac{2}{\theta} - x$$

II) Se calcula la función de información de X.

$$I_X(\theta) = \mathrm{Var}_\theta\left[\frac{\partial \ln f_\theta(x)}{\partial\theta}\right] = \mathrm{Var}_\theta\left[\frac{2}{\theta} - x\right] = (-1)^2\mathrm{Var}_\theta[X] = \frac{2}{\theta^2}$$

III) Se utiliza la aditividad de la función de información para calcular la función de información asociada a la muestra.

$$I_{X_1,\dots,X_n}(\theta) = n\cdot I_X(\theta) = \frac{2n}{\theta^2}$$

b) Las funciones paramétricas que admiten estimador eficiente son de la forma $ag(\theta)+b$, con $g(\theta)$ cumpliendo

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

para cierta función $a(\theta)$.

i) Calculamos la función de densidad de la muestra.

$$f_\theta^n(x_1,\dots,x_n) = \theta^2 x_1 e^{-\theta x_1}\cdots\theta^2 x_n e^{-\theta x_n} = (\theta^2)^n \prod_{i=1}^n x_i e^{-\theta\sum_{i=1}^n x_i},\ x_i > 0$$

ii) Calculamos el logaritmo neperiano de la función de densidad muestral.

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

iii) Hallamos la derivada del logaritmo de la función de densidad muestral.

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

Así, vemos que tenemos dos opciones distintas para $a(\theta)$, T y $g(\theta)$.

1ª. $a(\theta) = -1, T = \sum_{i=1}^n X_i, g(\theta) = \frac{2n}{\theta}$.

2ª. $a(\theta) = -n, T = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}, g(\theta) = \frac{2}{\theta}$.

Ahora se tiene que comprobar que T es estimador en ambos casos, y que se cumple $I_{X_1,\dots,X_n}(\theta) = a(\theta)g'(\theta)$.

- Es evidente que en las dos opciones se tiene que T es estimador, pues en ambos casos T solo puede tomar valores positivos, al igual que θ.
- 1ª. $a(\theta)g'(\theta) = (-1)\cdot\left(\frac{2n}{\theta}\right)' = (-1)\cdot\left(\frac{-2n}{\theta^2}\right) = \frac{2n}{\theta^2} = I_{X_1,\dots,X_n}(\theta)$.

  2ª. $a(\theta)g'(\theta) = (-n)\cdot\left(\frac{2}{\theta}\right)' = (-n)\cdot\left(\frac{-2}{\theta^2}\right) = \frac{2n}{\theta^2} = I_{X_1,\dots,X_n}(\theta)$.

Por tanto, en efecto T es un estimador eficiente en cada caso. Así:

- Funciones paramétricas que admiten estimador eficiente: $ag(\theta)+b$.

$$\underset{\displaystyle a\cdot\frac{2n}{\theta}+b,\ a,b\in\mathbb{R}}{\overset{1^{\text{a}}\text{ opción}}{\downarrow}} \qquad \underset{\displaystyle a\cdot\frac{2}{\theta}+b,\ a,b\in\mathbb{R}}{\overset{2^{\text{a}}\text{ opción}}{\downarrow}}$$

- Estimadores eficientes para las funciones paramétricas:

$$\underset{\displaystyle a\cdot\sum_{i=1}^n X_i+b,\ a,b\in\mathbb{R}}{\overset{1^{\text{a}}\text{ opción}}{\downarrow}} \qquad \underset{\displaystyle a\cdot\frac{1}{n}\sum_{i=1}^n X_i+b = a\cdot\bar{X}+b,\ a,b\in\mathbb{R}}{\overset{2^{\text{a}}\text{ opción}}{\downarrow}}$$

c) Calculamos la cota de FCR y comprobamos que se alcanza solo en el segundo caso. El primero es análogo.

Por ser T estimador regular, se tiene que $\operatorname{Var}_\theta[T(X_1,\dots,X_n)] \geq \frac{(g'(\theta))^2}{I_{X_1,\dots,X_n}} = \text{cota}$.

$$g(\theta) = \frac{2}{\theta} \Rightarrow g'(\theta) = \frac{-2}{\theta^2} \Rightarrow (g'(\theta))^2 = \frac{4}{\theta^4} \Rightarrow \text{cota} = \frac{4}{\theta^4}\cdot\frac{\theta^2}{2n} = \frac{2}{n\theta^2}$$

Veamos que se alcanza la cota.

$$\mathrm{Var}_\theta[T] = \mathrm{Var}_\theta\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}_\theta[X_i] = \frac{1}{n^2}\cdot\frac{2n}{\theta^2} = \frac{2}{n\theta^2}$$

Corolario 2. *Si $T(X_1,\ldots,X_n)$ es eficiente para $g(\theta)$, entonces $T(X_1,\ldots,X_n)$ es el UMVUE de $g(\theta)$. El recíproco no es cierto: 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.*

Ejercicio tipo examen.

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

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

a) Sabiendo que esta familia de distribuciones es regular, que $E[\ln X] = \frac{1}{\theta}$ y $Var[\ln X] = \frac{1}{\theta^2}$, calcular la cota de FCR para la varianza de estimadores regulares, insesgados en θ^2, basados en una muestra aleatoria simple de tamaño n.

b) Determinar la clase de funciones paramétricas que admiten estimador eficiente basados en muestras de tamaño arbitrario, dichos estimadores y su varianza.

c) Basándose en el apartado anterior, encontrar el UMVUE para $2/\theta$.

Solución.

a) Por ser T un estimador regular e insesgado, se tiene $\mathrm{Var}_\theta[T(X_1,\ldots,X_n)] \geq \dfrac{(g'(\theta))^2}{I_{X_1,\ldots,X_n}} = \text{cota}$.

- $g(\theta) = \theta^2 \Rightarrow g'(\theta) = 2\theta \Rightarrow (g'(\theta))^2 = 4\theta^2$.
- Calculamos la función de información asociada la muestra.

I) Se halla la derivada de $\ln f_\theta(x)$ con respecto a θ.

$$f_\theta(x) = \frac{\theta}{x\cdot x^\theta} \Rightarrow \ln f_\theta(x) = \ln\theta - \ln(x\cdot x^\theta) = \ln\theta - \ln x - \theta\ln x = \frac{\partial \ln f_\theta(x)}{\partial\theta} = \frac{1}{\theta} - \ln x$$

II) Se calcula la función de información de X.

$$\mathrm{Var}_\theta\left[\frac{\partial \ln f_\theta(x)}{\partial\theta}\right] = \mathrm{Var}_\theta\left[\frac{1}{\theta} - \ln X\right] = \mathrm{Var}_\theta[\ln X] = \frac{1}{\theta^2}$$

III) Utilizamos la aditividad de la función de información para calcular la función de información asociada a la muestra.

$$I_{X_1,\ldots,X_n}(\theta) = n\cdot I_X(\theta) = \frac{n}{\theta^2}$$

Por tanto:

$$\text{cota} = 4\theta^2\cdot\frac{\theta^2}{n} = \frac{4\theta^4}{n}$$

b) Las funciones paramétricas que admiten estimador eficiente son de la forma $ag(\theta)+b$, con $g(\theta)$ cumpliendo

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

para cierta función $a(\theta)$.

i) Calculamos la función de densidad de la muestra.

$$f_\theta^n(x_1,\dots,x_n) = \frac{\theta}{x_1\cdot x_1^\theta}\cdots\frac{\theta}{x_n\cdot x_n^\theta} = \frac{\theta^n}{\left(\prod_{i=1}^n x_i\right)\left(\prod_{i=1}^n x_i^\theta\right)},\ x_i>1$$

ii) Calculamos el logaritmo neperiano de la función de densidad muestral.

$$\ln f_\theta^n(x_1,\dots,x_n) = \ln(\theta^n) - \ln\left(\left(\prod_{i=1}^n x_i\right)\left(\prod_{i=1}^n x_i^\theta\right)\right) = n\ln\theta - \ln\left(\prod_{i=1}^n x_i\right) - \theta\sum_{i=1}^n \ln x_i$$

iii) Hallamos la derivada del logaritmo de la función de densidad muestral.

$$\frac{\partial \ln f_\theta^n(x_1,\dots,x_n)}{\partial\theta} = \frac{n}{\theta} - \sum_{i=1}^n \ln x_i = (-1)\left[\sum_{i=1}^n \ln x_i - \frac{n}{\theta}\right] = (-n)\left[\frac{1}{n}\sum_{i=1}^n \ln x_i - \frac{1}{\theta}\right]$$

Así, vemos que tenemos dos opciones distintas para $a(\theta)$, T y $g(\theta)$.

1ª. $a(\theta)=-1, T=\sum_{i=1}^n \ln X_i, g(\theta)=\frac{n}{\theta}$.

2ª. $a(\theta)=-n, T=\frac{1}{n}\sum_{i=1}^n \ln X_i = \overline{\ln X}, g(\theta)=\frac{1}{\theta}$.

Ahora se tiene que comprobar que T es estimador en ambos casos, y que se cumple $I_{X_1,\dots,X_n}(\theta)=a(\theta)g'(\theta)$.

- Es evidente que en las dos opciones se tiene que T es estimador, pues en ambos casos T toma valores reales, igual que θ.
- 1ª. $a(\theta)g'(\theta) = (-1)\cdot\left(\frac{n}{\theta}\right)' = (-1)\cdot\left(\frac{-n}{\theta^2}\right) = \frac{n}{\theta^2} = I_{X_1,\dots,X_n}(\theta)$.

  2ª. $a(\theta)g'(\theta) = (-n)\cdot\left(\frac{1}{\theta}\right)' = (-n)\cdot\left(\frac{-1}{\theta^2}\right) = \frac{n}{\theta^2} = I_{X_1,\dots,X_n}(\theta)$.

Por tanto, en efecto T es un estimador eficiente en cada caso. Así:

- Funciones paramétricas que admiten estimador eficiente: $ag(\theta)+b$.

$$\underset{\substack{\downarrow\\ a\cdot\frac{n}{\theta}+b,\ a,b\in\mathbb{R}}}{\text{1ª opción}} \qquad \underset{\substack{\downarrow\\ a\cdot\frac{1}{\theta}+b,\ a,b\in\mathbb{R}}}{\text{2ª opción}}$$

- Estimadores eficientes para las funciones paramétricas:

$$\underset{\substack{\downarrow\\ a\cdot\sum_{i=1}^n \ln X_i+b,\ a,b\in\mathbb{R}}}{\text{1ª opción}} \qquad \underset{\substack{\downarrow\\ a\cdot\frac{1}{n}\sum_{i=1}^n \ln X_i+b = a\cdot\overline{\ln X}+b,\ a,b\in\mathbb{R}}}{\text{2ª opción}}$$

c) Por el corolario 2, si $T(X_1,\dots,X_n)$ es eficiente para una función paramétrica $g(\theta)$, entonces $T(X_1,\dots,X_n)$ es el UMVUE de $g(\theta)$. Por tanto, como $T=\overline{\ln X}$ es UMVUE para $1/\theta$, por la linealidad del UMVUE (Proposición 2) se tiene que $2T=2\overline{\ln X}$ es UMVUE para $2/\theta$.

Corolario 3. *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 la función*

$$Q(\theta) = \int_{\Theta} a(\theta) d\theta$$

contiene a un abierto de $\mathbb{R}$*, entonces* $T(X_1, \ldots, X_n)$ *es completo.*