Teoria-tema-1.pdf

Apuntes completos tema 1

Tema 1: Introducción a la Inferencia Estadística. Estadísticos muestrales. [1]

Índice

Introducimos en este tema algunos conceptos básicos de Inferencia Estadística. En particular, estudiaremos los elementos principales de un problema de Inferencia Paramétrica tales como muestra aleatoria simple y estadístico muestral.

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

1. Introducción.

El Cálculo de Probabilidades proporciona una teoría matemática que permite analizar propiedades de fenómenos (o experimentos) en los que interviene el azar. En general, se trabajará en un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, construido a partir del conjunto de los posibles resultados del fenómeno (espacio muestral), sobre el que se define una σ-álgebra de sucesos, y una función de probabilidad adecuada. Así, una vez el espacio probabilístico se encuentra totalmente especificado, se podrá estudiar la ocurrencia de los sucesos incluidos en Ω.

Los espacios muestrales asociados a fenómenos aleatorios podrán ser cualitativos o cuantitativos, siendo evidente que el tratamiento matemático de una variable cualitativa puede ser complejo. Por ello, con frecuencia en Probabilidad se trabajará con variables aleatorias puesto que facilitan el tratamiento numérico de los datos obtenidos en el estudio de cualquier experimento aleatorio.

Sin embargo, para poder construir una variable aleatoria que modele una situación real será necesario tener construido previamente su espacio de probabilidad asociado, y se podrá dar el caso en el que, por cualquier razón, se disponga de un conjunto de observaciones acerca del fenómeno considerado en lugar de un espacio probabilístico bien definido. Al tener algún desconocimiento sobre las variables de estudio o algunas de sus características, no podremos construir completamente la variable aleatoria necesaria para el cálculo de probabilidades, y por tanto no podremos obtener conclusiones sobre los datos a analizar.

Con el fin de obtener conclusiones a partir de unos pocos datos obtenidos en el estudio de un fenómeno aleatorio cuya variable aleatoria modelo es, de algún modo, desconocida, se desarrollan métodos que se engloban dentro de lo que se llama *Inferencia Estadística* o *Estadística Matemática.*

Se plantea el problema de cómo obtener información acerca de la ley de probabilidad de un fenómeno a partir de una observación no exhaustiva del mismo. La Inferencia Estadística trata de analizar e interpretar las observaciones de las que se dispone como método para obtener conclusiones sobre la ley de probabilidad del fenómeno de estudio.

Toda variable aleatoria X sigue una *distribución*, que es una función que indica la probabilidad de que X tome o bien un valor concreto, o bien un valor dentro de un intervalo (o un conjunto de intervalos). La falta del conocimiento total de X se presenta cuando no se conoce su distribución, bien porque no se sabe nada acerca de su expresión matemática, bien porque aunque se sepa algo, esta depende de un parámetro desconocido.

Ejemplos de la primera situación ocurren cuando, a pesar de no saber nada sobre la expresión matemática de la distribución, se saben algunas de sus propiedades, como por ejemplo que es simétrica, o continua, o absolutamente continua, o que existen sus momentos centrados hasta orden 3. En este caso, conocer la distribución de una variable a partir de algunas de sus propiedades será un problema de *Inferencia no paramétrica* (tema 9).

Para la segunda situación, lo que se hace es obtener una aproximación del parámetro que se desconoce. De este tipo de problemas se encarga la *Inferencia Paramétrica* (temas 1-8). Dependiendo de la forma en que queremos obtener el valor de tal parámetro, podemos dividir la Inferencia Paramétrica en:

- *Estimación puntual.* Se busca obtener un pronóstico numérico único acerca del parámetro.
- *Intervalos de confianza.* Se da un margen de variación sobre la acotación del parámetro; en concreto, se precisa un intervalo numérico en el que pueda afirmarse, razonablemente, que este varía.

- *Contrastes de hipótesis.* Se busca corroborar o invalidar una determinada afirmación acerca del parámetro desconocido.

1.1. Introducción a la Inferencia Paramétrica.

De forma genérica, la distribución desconocida F de la variable aleatoria X involucrada en un problema de Inferencia Estadística recibe el nombre de *distribución teórica*. El mayor o menor grado de desconocimiento acerca de la distribución teórica F se refleja mediante la familia $\mathcal{F}$ de distribuciones, candidatas a ser realmente la distribución de la población.

La situación que estudiamos en primer lugar es aquella en que la familia $\mathcal{F}$ está compuesta por distribuciones que se expresan a partir de una función fija y conocida, dependientes de un parámetro θ, de una o más dimensiones, que varía dentro de un subconjunto Θ de $\mathbb{R}^n$, denominado espacio paramétrico:

$$\mathcal{F} = \{F_\theta, \theta \in \Theta \subset \mathbb{R}^n\}$$

2. Muestra aleatoria simple.

Nos planteamos de qué forma debemos obtener observaciones a partir de las cuales se ha de intentar disminuir el desconocimiento de la distribución teórica F de la variable aleatoria X en estudio. En principio, tales observaciones pueden obtenerse llevando a cabo repeticiones del experimento aleatorio que da lugar a X y anotando los valores de la variable en cada una de ellas. Se tendrá así un conjunto de valores numéricos $(x_1, \ldots, x_n)$ que se denomina una *muestra aleatoria* de X. El número n de repeticiones efectuadas, y de observaciones obtenidas, se denomina *tamaño de la muestra.*

Observación 1. *Los procedimientos posibles para obtener una muestra aleatoria pueden ser diversos, dependiendo de las condiciones en que se efectúen las repeticiones del experimento aleatorio. Todo problema de Inferencia Estadística, para estar bien formulado, tiene que incluir la definición precisa del procedimiento de muestreo con el que se obtienen las observaciones.*

Cuando ya se ha obtenido una muestra aleatoria de tamaño n, $(x_1, \ldots, x_n)$, se dispone simplemente de un conjunto de n variables numéricas con las cuales tratar de cumplir el objetivo de precisar la distribución teórica. Sin embargo, cuando se está planificando la obtención de la muestra aleatoria y diseñando el procedimiento que se empleará para intentar conocer tal distribución, todavía no se sabe qué valores numéricos $(x_1, \ldots, x_n)$ resultarán, de manera que deben considerarse como n variables aleatorias, $(X_1, \ldots, X_n)$. Así, la muestra aleatoria, como conjunto de n variables aleatorias, tiene una distribución n-dimensional (supuesto que las observaciones individuales son unidimensionales), que determina las probabilidades de aparición de cada muestra efectiva.

Observación 2.

- *Al conjunto de valores $(x_1, \ldots, x_n)$ obtenidos en la selección de una muestra particular se le llama* realización muestral.

  A efectos prácticos, en un problema real se estudiará la muestra aleatoria $(X_1, \ldots, X_n)$ a partir de una realización muestral suya, es decir, a partir de valores concretos $(x_1, \ldots, x_n)$, donde cada valor x_i se obtiene al observar la variable aleatoria X_i correspondiente.

- *Al conjunto de todas las realizaciones muestrales, es decir, el conjunto de los posibles valores del vector $(x_1, \ldots, x_n)$, se le denomina* espacio muestral *o conjunto de posibles valores de la muestra, y se le denota por χ^n.*

En el caso en que se efectúen repeticiones independientes del experimento aleatorio (es decir, con reemplazamiento; una vez observado un individuo lo devolvemos a la población), cabe destacar que lo que se está haciendo es observar de forma independiente, para la obtención de cada valor x_i, la variable X en estudio, cuya distribución teórica es F, luego cada uno de los valores que se obtengan, al ser extraídos con repetición, tendrán la misma distribución.

Muestra aleatoria simple.

Definición 1. *Una muestra aleatoria simple, de tamaño n, de una variable aleatoria X con distribución teórica F, es un vector $(X_1, \ldots, X_n)$ formado por n variables aleatorias independientes e idénticamente distribuidas, con distribución común F.*

Al ser las variables aleatorias $X_1, \ldots, X_n$ independientes, se tiene que la función de distribución conjunta del vector aleatorio formado por dichas variables será igual al producto de las distribuciones marginales de cada una de ellas. Es decir, la función de distribución conjunta de una muestra aleatoria simple $(X_1, \ldots, X_n)$ correspondiente a una distribución de la población F es:

$$F_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n)$$

- Para el caso de variables aleatorias discretas, se cumple que la función masa de probabilidad conjunta es el producto de las funciones masa de probabilidad marginales:

$$P[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n] = P[X_1 = x_1] \cdot P[X_2 = x_2] \cdot \cdots \cdot P[X_n = x_n]$$

- Para el caso de variables aleatorias continuas, se cumple que la función de densidad de probabilidad conjunta es el producto de las funciones de densidad marginales:

$$f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n)$$

Ejemplo 1.

a) Calcular la función masa de probabilidad conjunta de una muestra aleatoria simple de tamaño arbitrario de una variable $X \rightsquigarrow B(k_0, p)$.

b) Calcular la función de densidad de probabilidad conjunta de una muestra aleatoria simple de tamaño arbitrario de una variable $X \rightsquigarrow U(a, b)$.

Solución.

1. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de $X \rightsquigarrow B(k_0, p)$.

$$P[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n] \underbrace{=}_{\text{independientes}} P[X_1 = x_1]P[X_2 = x_2]\cdots P[X_n = x_n]$$

$$\underbrace{=}_{\text{identicamente distribuidas a } X} P[X = x_1]P[X = x_2]\cdots P[X = x_n]$$

$$= \binom{k_0}{x_1} p^{x_1}(1-p)^{k_0-x_1} \binom{k_0}{x_2} p^{x_2}(1-p)^{k_0-x_2} \cdots \binom{k_0}{x_n} p^{x_n}(1-p)^{k_0-x_n}$$

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

2. Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de $X \rightsquigarrow U(a,b)$. Por el mismo razonamiento que antes, pero ahora con la función de densidad, se tiene:

$$f_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n) = f_{X_1}(x_1)\cdots f_{X_n}(x_n) = f_X(x_1)\cdots f_X(x_n) = \frac{1}{(b-a)^n}, a < x_i < b, \forall i$$

3. Función de distribución muestral.

En los problemas de Inferencia Estadística, en general se usará la función de densidad muestral para variables aleatorias continuas y la función masa de probabilidad muestral para variables aleatorias discretas. Sin embargo, asociada a la muestra también existirá una función de distribución, que tendrá propiedades de convergencia que aseguran que su uso es conveniente para aproximar la función de distribución de la variable aleatoria de partida. Así, vamos a asociar a cada muestra aleatoria simple $(X_1, \ldots, X_n)$ de una variable aleatoria X con función de distribución F una distribución muestral $F^*_{X_1,\ldots,X_n}$ que emule a F a partir, únicamente, de la información contenida en la muestra.

Función de distribución muestral.

Definición 2. *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria X con función de distribución F. Se define la función de distribución muestral de X asociada a la muestra $(X_1, \ldots, X_n)$ como*

$$F^*_{X_1,\ldots,X_n}(x) = \frac{n^o \text{ de variables } X_i \leq x}{n}, \forall x \in \mathbb{R}$$

Observación 3. *La función de distribución muestral se puede escribir como*

$$F^*_{X_1,\ldots,X_n}(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(X_i)$$

donde el sumando $I_{(-\infty,x]}(X_i)$ es la variable aleatoria definida como

$$I_{(-\infty,x]}(X_i) = \begin{cases} 1 & si \quad X_i \leq x \\ 0 & si \quad X_i > x \end{cases}$$

Esta puede verse como una variable donde con probabilidad 1 se tiene el suceso éxito *(donde el éxito se resume en $X_i \leq x$) y con probabilidad 0 se tiene el suceso* fracaso *(es decir, $X_i > x$). Por tanto, es una Bernoulli, y sabemos que la suma de n Bernoullis es una binomial. Además, se verifica:*

$$P[I_{(-\infty,x]}(X_i) = 1] = P[X_i \leq x] = F(x)$$

$$P[I_{(-\infty,x]}(X_i) = 0] = P[X_i > x] = 1 - F(x)$$

Propiedades de la función de distribución muestral.

Proposición 1.

i) Para cada realización muestral, $(x_1, \ldots, x_n) \in \chi^n$, $F^*_{X_1,\ldots,X_n}(x)$ es una función de distribución en $\mathbb{R}$. En particular, es una función a saltos, con saltos de amplitud $1/n$ en los sucesivos *valores muestrales ordenados de menor a mayor, supuestos que sean distintos, y de saltos múltiples en el caso de que varios valores muestrales coincidieran.*

ii) $\forall x \in \mathbb{R}, F^*_{X_1,\dots,X_n}(x)$ *es una variable aleatoria tal que* $nF^*_{X_1,\dots,X_n}(x) \rightsquigarrow B(n, F(x))$*:*

$$E[F^*_{X_1,\dots,X_n}(x)] = F(x), \qquad Var[F^*_{X_1,\dots,X_n}(x)] = \frac{F(x)(1-F(x))}{n}$$

iii) Para valores grandes de n*, en virtud del Teorema Central del Límite:*

$$F^*_{X_1,\dots,X_n}(x) \rightsquigarrow \mathcal{N}\left(F(x), \frac{F(x)(1-F(x))}{n}\right)$$

Demostración.

i) Asumiendo $x_1 < x_2 < \cdots < x_n$ (en otro caso, basta con que la función que se construye tenga saltos múltiples), se tiene:

$$F^*_{X_1,\dots,X_n}(x) = \begin{cases} 0/n & x < x_1 \\ 1/n & x_1 \leq x < x_2 \\ \dots & \dots \\ 1 & x \geq x_n \end{cases}$$

Así, F^* es una función de distribución en $\mathbb{R}$ pues tiene límite 0 a la izquierda, límite 1 a la derecha, es monótona no decreciente y continua a la derecha.

ii) Por la Observación 3, se tiene:

$$I_{(-\infty,x]}(X_i) \rightsquigarrow B(1, F(x)) \Rightarrow \sum_{i=1}^{n} I_{(-\infty,x]}(X_i) \rightsquigarrow B(n, F(x)) \Rightarrow nF^*_{X_1,\dots,X_n}(x) \rightsquigarrow B(n, F(x))$$

Teniendo en cuenta la linealidad de la esperanza:

$$\mathrm{E}[nF^*_{X_1,\dots,X_n}(x)] = nF(x) \Rightarrow n\mathrm{E}[F^*_{X_1,\dots,X_n}(x)] = nF(x) \Rightarrow \mathrm{E}[F^*_{X_1,\dots,X_n}(x)] = F(x)$$

Por otro lado, puesto que $\mathrm{Var}[nX] = n^2\mathrm{Var}[X]$:

$$\mathrm{Var}[nF^*_{X_1,\dots,X_n}(x)] = n^2\mathrm{Var}[F^*_{X_1,\dots,X_n}(x)] = nF(x)(1-F(x))$$
$$\Rightarrow \mathrm{Var}[F^*_{X_1,\dots,X_n}(x)] = \frac{F(x)(1-F(x))}{n}$$

iii) Como $nF^*_{X_1,\dots,X_n}(x)$ es suma de variables aleatorias independientes e idénticamente distribuidas (ind. e i.d), el Teorema Central del Límite permite afirmar que, cuando $n \to +\infty$,

$$\frac{nF^*_{X_1,\dots,X_n}(x) - nF(x)}{\sqrt{nF(x)(1-F(x))}} = \sqrt{n}\frac{F^*_{X_1,\dots,X_n}(x) - F(x)}{\sqrt{F(x)(1-F(x))}} \rightsquigarrow \mathcal{N}(0,1)$$

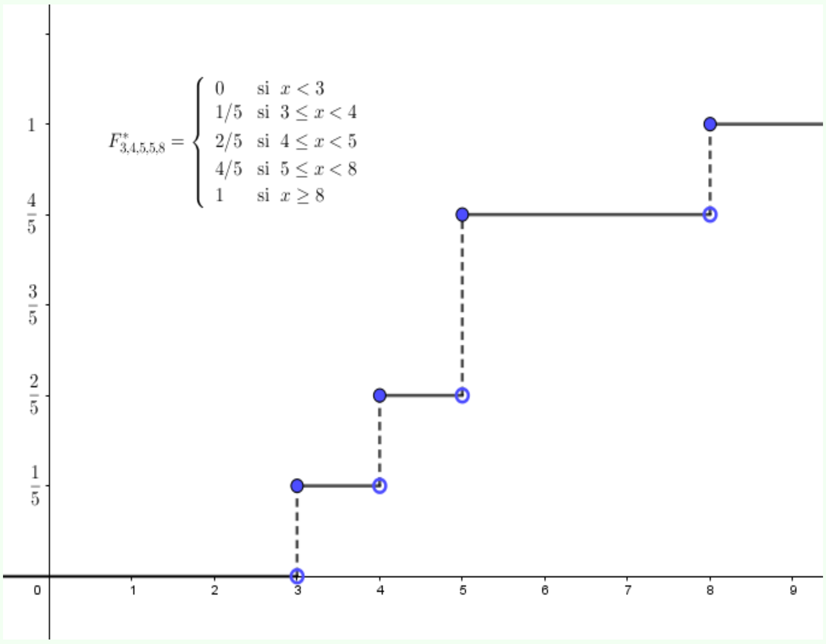
de manera que, cuando el tamaño muestral es grande, se tiene

$$F^*_{X_1,\dots,X_n}(x) \rightsquigarrow \mathcal{N}\left(F(x), \frac{F(x)(1-F(x))}{n}\right)$$

■

Ejemplo 2. *Dada una muestra aleatoria simple formada por las observaciones* $(3, 8, 5, 4, 5)$*, obtener su función de distribución muestral y realizar la representación gráfica.*

Solución. Ordenamos los valores muestrales: $3 < 4 < 5 \leq 5 < 8$. Por tanto, en $x = 5$, la función de distribución muestral tendrá un salto de amplitud $2/5$, puesto que al haber cinco observaciones muestrales, la amplitud de cada salto en caso de valores no repetidos es $1/5$.



En el punto iii) de la Proposición 1 se ha visto que, a medida que n crece (aumenta el tamaño muestral), la distribución muestral $F^*_{X_1,\ldots,X_n}(x)$ se encuentra cada vez más concentrada alrededor del valor de $F(x)$, es decir, con un tamaño muestral grande, es probable que la muestra seleccionada proporcione un valor de $F^*_{X_1,\ldots,X_n}(x)$ muy próximo al valor desconocido de $F(x)$.

Nos interesa por tanto poder asegurar que dicha aproximación se produce simultáneamente para todos los valores de x, lo cual no es inmediato, pues aunque el suceso relativo a x se produzca con probabilidad 1, el número de valores de x no es numerable. En ese sentido, el siguiente teorema prueba que $F^*_{X_1,\ldots,X_n}(x)$ converge a $F(x)$, uniformemente en x, con probabilidad 1.

Teorema de Glivenko-Cantelli.

Teorema 1. *Sea* $\{X_n\}_{n\in\mathbb{N}}$ *una sucesión de variables aleatorias independientes e idénticamente distribuidas con función de distribución común* F*. Si* $F^*_{X_1,\ldots,X_n}(x)$ *es la función de distribución muestral asociada a la muestra aleatoria simple* $(X_1, \ldots, X_n)$ *de* $X \rightsquigarrow F$*, se verifica que* $F^*_{X_1,\ldots,X_n}(x)$ *converge casi seguramente y uniformemente a la función de distribución de* X*,* F*.*

$$P\left[\lim_{n\to+\infty} \sup_{x\in\mathbb{R}} |F^*_{X_1,\ldots,X_n}(x) - F(x)| = 0\right] = 1$$

Es decir, con probabilidad 1, al tomar sucesivas observaciones independientes de la variable y considerar las correspondientes funciones de distribución muestrales, se verifica:

$$\forall \epsilon > 0, \exists n_\epsilon / n > n_\epsilon \Rightarrow F(x) \in (F^*_{X_1,\ldots,X_n}(x) - \epsilon, F^*_{X_1,\ldots,X_n}(x) + \epsilon)$$

4. Estadístico muestral.

Para obtener información acerca de la función de distribución de la variable X en estudio (que, a priori, depende de ciertos parámetros que no se conocen), se toma una muestra aleatoria simple de la misma, $(X_1, \dots, X_n)$, y se observan los valores que esta toma. Puede resultar, en ocasiones, más cómodo analizar, en lugar de todos los valores del vector muestral, una función suya, como por ejemplo la suma de los mismos. A dicha función de la muestra aleatoria simple, siempre que sea independiente de los parámetros desconocidos, le llamaremos *estadístico*, y la ventaja de su uso reside en que, en lugar de trabajar con la muestra, que es un vector aleatorio de dimensión n, se trabaja con una variable aleatoria unidimensional o con un vector de menor dimensión (en general, los estadísticos buscan reducir la dimensión con la que estamos trabajando.)

Estadístico.

Definición 3. *Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de una variable aleatoria X. Un estadístico muestral asociado a X es una función $T : (\mathbb{R}^n, \mathcal{B}^n) \longrightarrow (\mathbb{R}^k, \mathcal{B}^k)$ medible, es decir, cumple $T^{-1}(B) \in \mathcal{B}^n, \forall B \in \mathcal{B}^k$, e independiente de cualquier parámetro desconocido. Lo denotamos por $T(X_1, \dots, X_n)$, o cuando se sobreentienda que es un estadístico, simplemente T o $T(X)$.*

Observación 4. *Como T es una función medible, el estadístico muestral $T(X_1, \dots, X_n)$ es una variable aleatoria de dimensión k. Por tanto, la función T basta que esté definida en el espacio muestral χ^n y no en todo $\mathbb{R}^n$.*

Ejemplo 3. *Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de una variable aleatoria X. Algunos de los estadísticos muestrales más frecuentes en su uso son los siguientes:*

- *Media muestral:* $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.
- *Varianza muestral:* $Var[X] = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2$.
- *Cuasivarianza muestral:* $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2$.
- *Estadísticos ordenados:* $X_{(1)} = \text{mín}(X_1, \dots, X_n); X_{(n)} = \text{máx}(X_1, \dots, X_n)$.

  El resto de los estadísticos ordenados, $X_{(2)}, \dots, X_{(n-1)}$, son los estadísticos muestrales que se quedan con el valor que ocupa la posición que indica su subíndice después de ordenar de menor a mayor todos los valores, es decir, por ejemplo $X_{(2)}$ dará como resultado el segundo valor más pequeño de $(X_1, \dots, X_n)$.

Ejemplo 4. *Sea X una variable aleatoria con distribución $B(1,p)$ con $p \in (0,1)$. Se toma una muestra aleatoria simple de tamaño 5, $(X_1, X_2, X_3, X_4, X_5)$, y se obtiene la siguiente observación: $(0,1,1,0,0)$. Determina el valor de los estadísticos estudiados en el Ejemplo 3.*

Solución. Basta con sustituir en las fórmulas del Ejemplo 3 para obtener los siguientes valores:

$$\bar{X} = \frac{2}{5} \qquad \text{Var}[X] = \frac{6}{25} \qquad S^2 = 0{,}3 \qquad X_{(1)} = 0 \qquad X_{(n)} = 1$$

4.1. Distribución en el muestreo y características de un estadístico.

Por ser $T(X_1, \ldots, X_n)$ una varable aleatoria, es entonces evidente que tendrá una función de distribución y unas características numéricas asociadas tales como esperanza, varianza o existencia de momentos. Así, como en el caso de una variable aleatoria cualquiera X, tener completamente determinado al estadístico T implica conocer todas sus características.

Distribución en el muestro.

Definición 4. *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria X. La distribución en el muestreo de un estadístico T definido en el espacio muestral $(\chi^n, \mathcal{B}^n)$ es la distribución de la variable aleatoria $T(X_1, \ldots, X_n)$.*

Proposición 2. *La función generatriz de momentos del estadístico media muestral viene dada por:*

$$M_{\bar{X}}(t) = (M_x(t/n))^n$$

Demostración. Recordemos que la función generatriz de momentos de una variable aleatoria es aditiva-multiplicativa en el sentido de que la función generatriz de momentos de la suma de n variables aleatorias independientes coincide con el producto de las n funciones generatrices de momentos.

Sea X una variable aleatoria, y $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X de tamaño n.

$$M_{\bar{X}}(t) = \mathrm{E}\left[e^{t\bar{X}}\right] = \mathrm{E}\left[e^{t\frac{\sum X_i}{n}}\right] = \mathrm{E}\left[e^{\frac{\sum X_i t}{n}}\right] = \mathrm{E}\left[\prod e^{\frac{tX_i}{n}}\right]$$
$$\underbrace{=}_{\text{ind}} \prod_{i=1}^{n} \mathrm{E}\left[e^{\frac{t}{n}X_i}\right] = \prod_{i=1}^{n} M_{X_i}(t/n) \underbrace{=}_{\text{i.d}} (M_X(t/n))^n$$

■

Importante para un tipo test.

Ejemplo 5. *Obtener la distribución muestral de $\bar{X}$ para $(X_1, \ldots, X_n)$ una muestra aleatoria simple de $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$.*

Solución. Aplicando la Proposición 2, se tiene:

$$M_{\bar{X}}(t) = (M_X(t/n))^n = \left(\exp\left(\frac{t}{n}\mu + \left(\frac{t}{n}\right)^2 \frac{\sigma^2}{2}\right)\right)^n = \exp\left(\cancel{n}\left(\frac{t}{\cancel{n}}\mu + \frac{t^2}{n^{\cancel{2}}}\frac{\sigma^2}{2}\right)\right) \Rightarrow \bar{X} \rightsquigarrow \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Consideremos los estadísticos ordenados, definidos como:

$$X_{(r)} = \begin{cases} \text{mín}(X_1, \ldots, X_n) & r = 1; \\ \text{mín}(\{X_1, \ldots, X_n\} - \{X_{(1),\ldots,X_{(j-1)}}\}) & r = j; \\ \text{máx}(X_1, \ldots, X_n) & r = n. \end{cases}$$

Proposición 3. *La distribución muestral general de los estadísticos de orden, según la variable aleatoria sea discreta o continua, es:*

- *X discreta:*

$$\begin{aligned} P[X_{(r)} \leq x] &= P[\text{al menos } r \text{ elementos muestrales sean } \leq x] \\ &= \sum_{i=r}^{n} \binom{n}{i} (P[X \leq x])^i (P[X > x])^{n-i} \end{aligned}$$

- *X continua:*

$$g_r(x_{(r)}) = \frac{n!}{(r-1)!(n-r)!} [F(x_{(r)})]^{r-1} [1 - F(x_{(r)})]^{n-r} f(x_{(r)})$$

Observación 5. *Las fórmulas obtenidas en la Proposición 3 cuentan de cuántas formas se pueden seleccionar permutaciones de r elementos menores o iguales que un cierto x.*

Importante para un tipo test.

Ejemplo 6.

a) Obtener las distribuciones muestrales de $X_{(1)}$ y $X_{(n)}$ para X una variable aleatoria cualquiera y $(X_1, \ldots, X_n)$ una muestra aleatoria simple suya.

b) Obtener las distribuciones muestrales de $X_{(1)}$ y $X_{(n)}$ para $X \rightsquigarrow U(a, b)$.

Solución.

a) Hallamos las distribuciones muestrales de $X_{(1)}$ y $X_{(n)}$ para una muestra aleatoria simple de tamaño n. de una variable aleatoria X con función de distribución F_X.

- Para la distribución del mínimo $X_{(1)}$, tengamos en cuenta que si el mínimo de los valores es mayor que un cierto x, entonces el resto de valores también serán mayores que x. Así:

$$\begin{aligned} &F_{X_{(1)}}(x) = P[X_{(1)} \leq x] = 1 - P[X_{(1)} > x] = 1 - P[X_1 > x, \ldots, X_n > x] \\ &\underbrace{=}_{\text{ind}} 1 - P[X_1 > X] \cdots P[X_n > x] \underbrace{=}_{\text{i.d}} 1 - (P[X > x])^n = 1 - (1 - P[X \leq x])^n \end{aligned}$$

$$\boxed{F_{X_{(1)}}(x) = 1 - (1 - F_X(x))^n} \tag{1}$$

- Para la distribución del máximo $X_{(n)}$, se tiene que si el máximo de los valores es menor que un cierto x, entonces el resto de valores son también menores que x. Por tanto:

$$F_{X_{(n)}}(x) = P[X_{(n)} \leq x] = P[X_1 \leq x, \ldots, X_n \leq x] \underbrace{=}_{\text{ind}} P[X_1 \leq x] \cdots P[X_n \leq x]$$

$$\underbrace{=}_{\text{i.d}} (P[X \leq x])^n = (F_X(x))^n$$

$$\boxed{F_{X_{(n)}}(x) = (F_X(x))^n} \tag{2}$$

b) Para obtener las funciones de distribución muestrales de $X_{(1)}$ y $X_{(n)}$ para el caso de una variable continua uniforme, basta con sustituir su función de distribución en las fórmulas (1) y (2) y su función de densidad en ambas derivadas. En este caso, nos centraremos en calcular las funciones de densidad muestrales, que se obtienen derivando las funciones de distribución muestrales:

$$F_{X_{(1)}}(x) = 1 - (1 - F_X(x))^n \Rightarrow f_{X_{(1)}}(x) = n(1 - F_X(x))^{n-1} f_X(x)$$

$$F_{X_{(n)}}(x) = (F_X(x))^n \Rightarrow f_{X_{(n)}}(x) = n(F_x(x))^{n-1} f_X(x)$$

Así, teniendo en cuenta que las funciones de distribución y de densidad de una variable $U(a,b)$ vienen dadas, respectivamente, por

$$F_X(x) = \frac{x-a}{b-a}; \quad f_X(x) = \frac{1}{b-a}, \ a < x < b$$

se tiene entonces:

$$f_{X_{(1)}}(x) = n\left(1 - \frac{x-a}{b-a}\right)^{n-1} \frac{1}{b-a}; \quad f_{X_{(n)}}(x) = n\left(\frac{x-a}{b-a}\right)^{n-1} \frac{1}{b-a}, \quad a < x < b$$

Otros ejemplos de estadísticos menos frecuentes (de los cuales ya hemos mencionado algunos casos particulares) son los siguientes.

Momentos muestrales centrados y no centrados.

Definición 5. *Se definen los momentos centrados y no centrados de una muestra aleatoria simple de tamaño n como:*

- *Momento no centrado de orden $k \in \mathbb{N}$: $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$.*

  En particular, se tiene $A_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} =$ media muestral.

- *Momento centrado de orden $k \in \mathbb{N}$: $B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$.*

  En particular, se tiene $B_1 = 0$ (por las propiedades de la media de una distribución), y definimos $B_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = Var[X] =$ Varianza muestral.

Proposición 4. *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria X.*

i) Para los momentos no centrados, se tiene:

$$E[A_k] = E[X^k] \qquad Var[A_k] = \frac{1}{n}(E[X^{2k}] - E[X^k]^2)$$

En particular, $E[A_1] = \mu$, $Var[A_1] = \frac{\sigma^2}{n}$, donde μ y σ^2 son, respectivamente, la media y la varianza poblacional.

ii) Para el momento centrado de orden 2, se tiene

$$E[B_2] = \frac{(n-1)\sigma^2}{n}$$

En particular, se verifica $E[S^2] = \sigma^2$, donde $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

A S^2 así definida le llamamos cuasivarianza muestral.

Demostración.

i)

$$\mathrm{E}[A_k] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^k\right] = \frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n} X_i^k\right] = \frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n} X^k\right] = \frac{1}{n}\mathrm{E}[nX^k] = \mathrm{E}[X^k]$$

En particular, $\mathrm{E}[A_1] = \mathrm{E}[X] = \mu$.

$$\mathrm{Var}[A_k] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^k\right] = \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} X_i^k\right] \underbrace{=}_{\text{ind}} \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_i^k]$$
$$= \frac{1}{n}\mathrm{Var}[X^k] = \frac{1}{n}(\mathrm{E}[X^{2k}] - \mathrm{E}[X^k]^2)$$

En particular, $\mathrm{Var}[A_1] = \frac{1}{n}(\mathrm{E}[X^2] - (\mathrm{E}[X])^2) = \frac{\sigma^2}{n}$.

ii) El teorema de König tiene su versión para la varianza poblacional, cumpliéndose 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 $\mathrm{Var}[\bar{X}] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] \underbrace{=}_{\text{i.i.d}} \frac{1}{n^{\cancel{2}}}\cancel{n}\mathrm{Var}[X] = \frac{1}{n}\mathrm{Var}[X]$, se llega a:

$$(1) = \mathrm{Var}[X] + (\mathrm{E}[X])^2 - \frac{\mathrm{Var}[X]}{n} - \underbrace{(\mathrm{E}[\bar{X}])^2}_{(\mathrm{E}[X])^2} = \frac{(n-1)\mathrm{Var}[X]}{n} = \frac{(n-1)\sigma^2}{n} = \mathrm{E}[B_2]$$

En particular, podemos reescribir S^2 como

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

Por tanto:

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

■

Cuantil muestral.

Definición 6. *Para cada $p \in (0,1)$, el cuantil de orden p, c_p, es un valor real tal que*

$$F_n^*(c_p) \geq p \text{ y } F_n^*(c_p^-) \leq p$$

Se puede expresar de la siguiente forma en función de los elementos de la muestra ordenada:

- *Si $np \in \mathbb{N}$, $c_p = \dfrac{X_{(np)} + X_{(np+1)}}{2}$.*
- *En otro caso, sea $[np]$ la parte entera de np, entonces $c_p = X_{([np]+1)}$.*

Función generatriz de momentos muestral.

Definición 7. *Se define la función generatriz de momentos muestral como*

$$M^*(t) = \frac{1}{n}\sum_{i=1}^{n} e^{tX_i}$$

Proposición 5. *La función generatriz de momentos muestral se utiliza para obtener los momentos no centrados, pues se verifica*

$$\left.\frac{\partial^k M^*(t)}{\partial t^k}\right|_{t=0} = A_k$$

tema2teorIE.pdf
(Definitivo) Apuntes tema 2

Tema 2: Distribuciones en el muestreo de poblaciones normales. [1]

Índice

En este tema se estudian resultados fundamentales relativos al muestreo de poblaciones normales tales como la distribución exacta (para muestras de cualquier tamaño) de estadísticos que surgen de forma natural en problemas concretos de Inferencia Estadística.

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

1. Introducción.

Entre las distribuciones de probabilidad continuas que ya han sido estudiadas destacan, por sus propiedades, algunas que serán utilizadas con mayor frecuencia. Por ejemplo, la distribución exponencial se emplea en problemas donde se estudia el tiempo de espera hasta que ocurre un suceso concreto bajo condiciones de falta de memoria (recordemos que la distribución exponencial es la única distribución continua con esta propiedad), mientras que la distribución normal puede aparecer al estudiar, entre otras, situaciones que se modelan con una distribución binomial de la que queremos conocer su comportamiento a largo plazo.

En este último caso, esto es así puesto que, casi siempre que un determinado efecto sea producido por un gran número de causas independientes, cuyas repercusiones individuales son despreciables (y de varianza finita), pero que se acumulan para dar lugar al resultado final, la distribución de este será, aproximadamente, normal (aunque cuidado: no siempre debemos generalizar y suponer que cualquier situación se modelará con una normal; debemos hacer las comprobaciones que sean necesarias).

Por tanto, podemos suponer que, en la mayoría de poblaciones que se vayan a estudiar, el modelo que mejor se adapta a ellas es normal, así que es razonable estudiar la distribución en el muestreo de los principales estadísticos de utilidad (vistos en el primer tema), correspondientes a una distribución de tipo normal.

En una primera aproximación, los problemas estadísticos relativos a una población con distribución teórica $\mathcal{N}(\mu, \sigma^2)$ solo pueden venir del desconocimiento de μ, de σ o de ambos parámetros simultáneamente. Parece lógico entonces que en su solución intervengan la media muestral, $\bar{X}$, y la cuasivarianza muestral, S^2, y que el principal interés se centre en conocer la distribución en el muestreo de dichos estadísticos, o de alguna función de ambos.

Nuestro objetivo de conocer tales distribuciones se vuelve más ambicioso al intentar efectuar comparaciones entre dos poblaciones normales independientes con distribuciones $\mathcal{N}(\mu_1, \sigma_1^2)$ y $\mathcal{N}(\mu_2, \sigma_2^2)$. En tal caso, deberemos estudiar nuevos estadísticos necesarios para ello, así como sus distribuciones en el muestreo.

2. Distribuciones χ^2 de Pearson, t de Student y F de Snedecor.

Consideramos una población descrita por una variable aleatoria unidimensional X, con distribución teórica $\mathcal{N}(\mu, \sigma^2)$, de la cual se obtiene una muestra aleatoria simple de tamaño n, $(X_1, \ldots, X_n)$. Recordemos que, en el primer tema, se demostró el siguiente resultado:

Distribución en el muestreo de la media muestral.

Proposición 1. *Si $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$, y $(X_1, \ldots, X_n)$ es una m.a.s. de X, entonces*

$$\bar{X} \rightsquigarrow \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

2.1. Distribución χ^2 de Pearson.

Distribución χ^2 de Pearson.

Definición 1. *Se dice que una variable aleatoria X tiene una distribución χ^2 de Pearson con n grados de libertad ($n \in \mathbb{N}$), y se denota $X \rightsquigarrow \chi^2(n)$, si su función de densidad de probabilidad es*

$$f_X(x) = \frac{1}{\Gamma(n/2) \cdot 2^{n/2}} x^{n/2-1} e^{-x/2}, \quad x > 0$$

Observación 1. *La distribución χ^2 de Pearson es un caso particular de la distribución Gamma, $\Gamma(p, a)$, cuya función de densidad es*

$$f(x) = \frac{a^p}{\Gamma(p)} x^{p-1} e^{-ax}, \quad x > 0$$

De hecho, si comparamos ambas funciones de densidad, se obtiene que

$$X \rightsquigarrow \chi^2(n) \Leftrightarrow X \rightsquigarrow \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$$

Propiedades de la distribución χ^2 de Pearson.

Proposición 2.

i) La función generatriz de momentos de una variable aleatoria $X \rightsquigarrow \chi^2(n)$ es

$$M_X(t) = \frac{1}{(1-2t)^{n/2}}, \quad t < \frac{1}{2}$$

ii) Los momentos no centrados de la distribución $\chi^2(n)$ vienen dados por

$$E[X^k] = 2^k \frac{\Gamma\left(\frac{n}{2} + k\right)}{\Gamma\left(\frac{n}{2}\right)}$$

En particular, $E[X] = n, E[X^2] = n^2 + 2n \Rightarrow Var[X] = E[X^2] - (E[X])^2 = 2n$.

iii) ***Reproductividad.*** *Si $X_1, \ldots, X_n$ son variables aleatorias independientes con $X_i \rightsquigarrow \chi^2(k_i)$, entonces*

$$\sum_{i=1}^{n} X_i \rightsquigarrow \chi^2\left(\sum_{i=1}^{n} k_i\right)$$

iv) ***Relación entre las distribuciones χ^2 y $\mathcal{N}(0,1)$.*** *Si $X_1, \ldots, X_n$ son variables aleatorias independientes e idénticamente distribuidas con distribución común $\mathcal{N}(0,1)$, entonces:*

$$\sum_{i=1}^{n} X_i^2 \rightsquigarrow \chi^2(n)$$

Demostración. Las propiedades *i)*, *ii)* y *iii)* se pueden demostrar teniendo en cuenta que la distribución χ^2 es un caso particular de la distribución Gamma y, por tanto, se cumplirán sus mismas propiedades para los parámetros concretos considerados. Demostramos entonces la propiedad *iv)*.

Vamos a probar que, si $X \rightsquigarrow \mathcal{N}(0,1)$, entonces $X^2 \rightsquigarrow \chi^2(1)$. Así, utilizando la reproductividad de la distribución χ^2, se tendrá que la suma de los cuadrados de n variables aleatorias normales tipificadas será igual a la suma de n distribuciones χ^2 con un grado de libertad, es decir, una distribución $\chi^2(n)$. Recordemos que si $X \rightsquigarrow \mathcal{N}(0,1)$, entonces su función de densidad viene dada por

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, x \in \mathbb{R}$$

Aplicamos el teorema de cambio de variable unidimensional. Si $Y = X^2$, entonces se tiene la transformación $y = x^2$, luego $x = \pm\sqrt{y}$, así que consideramos las funciones $h_1(y) = \sqrt{y}, h_2(y) = -\sqrt{y}$. Así, por el teorema, se tiene que:

$$\begin{aligned} g_Y(y) &= f(h_1(y)) \left| \frac{\partial h_1(y)}{\partial y} \right| + f(h_2(y)) \left| \frac{\partial h_2(y)}{\partial y} \right| \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(\sqrt{y})^2}{2} \right) \left| \frac{1}{2\sqrt{y}} \right| + \frac{1}{2\pi} \exp\left(\frac{-(-\sqrt{y})^2}{2} \right) \left| \frac{-1}{2\sqrt{y}} \right| \\ &= \frac{1}{2\sqrt{y}} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-y^2}{2} \right) + \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-y^2}{2} \right) \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{1}{\sqrt{y}}, \quad y > 0 \end{aligned}$$

Ahora, comparando la función $g_Y(y)$ con la función de densidad de la χ^2 (o con la función de densidad de la correspondiente distribución Gamma), se llega a que $Y = X^2 \rightsquigarrow \chi^2(1)$. Así, como dijimos antes, por la reproductividad de la distribución χ^2 se obtiene el resultado. ■

Observación 2. *Como evaluar la función de densidad de la distribución $\chi^2(n)$ es complejo (debido a las propiedades de la función Γ, recordemos que es una integral), se dan tablas donde χ^2 está tabulada para valores de n pequeños. Para valores de n grandes, se utiliza que su distribución se puede aproximar por una $\mathcal{N}(n, 2n)$.*

Ejemplo 1. *Calcula el valor de k o la probabilidad indicada:*

1. $P[\chi^2(10) \geq k] = 0{,}005 \Rightarrow k = 25{,}1881$.
2. $P[\chi^2(45) \leq k] = 0{,}005 = 1 - P[\chi^2(45) \geq k] \Rightarrow P[\chi^2(45) \geq k] = 0{,}995 \Rightarrow k = 24{,}3110$.
3. $P[\chi^2(14) \geq 21{,}06] = 0{,}1$.
4. $P[\chi^2(20) \leq 12{,}44] = 1 - P[\chi^2(20) \geq 14{,}44] = 0{,}1$.

En la tabla se muestra la probabilidad de que la χ^2 deje valores por encima (a la derecha) de un número dado.

Gráfica de la función de densidad de $\chi^2(n)$.

Proposición 3. *La función de densidad de la distribución $\chi^2(n)$ cumple las siguientes propiedades:*

- *Es asimétrica a la derecha y unimodal.*
- *Para $n = 1$, $\lim_{x \to 0} f(x) = +\infty$, y la función de densidad es estrictamente decreciente.*

- *Para $n = 2$, $f(0) = 1/2$, y la función de densidad es estrictamente decreciente.*
- *Para $n = 3$, $f(0) = 0$, crece hasta la moda, y luego decrece.*

2.2. Distribución t de Student.

Distribución t de Student.

Definición 2. *Sean X e Y variables aleatorias independientes con distribuciones $X \rightsquigarrow \mathcal{N}(0,1)$ e $Y \rightsquigarrow \chi^2(n)$. Entonces, la variable aleatoria*

$$T = \frac{X}{\sqrt{Y/n}}$$

se dice que tiene una distribución t de Student con n grados de libertad, y se denota $T \rightsquigarrow t(n)$.

Propiedades de la distribución t de Student.

Proposición 4.

i) La función de densidad de probabilidad de una distribución $t(n)$ viene dada por

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad t \in \mathbb{R}$$

ii) Sea T una variable aleatoria con distribución $t(n)$, con $n > 1$. Entonces, se tiene que existen los momentos no centrados $E[T^r]$ para $r < n$, y se verifica que:

- *Si r es impar, entonces $E[T^r] = 0$.*
- *Si r es par, entonces*

$$E[T^r] = \frac{\Gamma\left(\frac{r+1}{2}\right)\Gamma\left(\frac{n-r}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}$$

En particular, para $n > 2$, existen los momentos de primer y segundo orden, y se tiene $E[T] = 0, E[T^2] = Var[T] = \dfrac{n}{n-2}$.

Observación 3. *Evaluar la función de densidad de la distribución $t(n)$ puede ser demasiado complicado, así que se dan tablas donde está tabulada para valores de n pequeños. Para valores de n grandes, se utiliza que su distribución se puede aproximar por una $\mathcal{N}(0,1)$.*

Ejemplo 2. *Calcula el valor de k o la probabilidad indicada:*

1. $P[t(26) \geq k] = 0{,}05 \Rightarrow k = 1{,}7056$.
2. $P[t(20) \leq k] = 0{,}25 = 1 - P[t(20) \geq k] = 0{,}75 \Rightarrow P[t(20) \geq k] = 0{,}75 \Rightarrow k = -0{,}687$.
3. $P[t(26) \geq k] = 0{,}9 \Rightarrow k = -1{,}315$.

4. $P[t(21) \geq 1{,}721] = 0{,}05$.
5. $P[t(11) \leq 0{,}697] = 1 - P[t(11) \geq 0{,}697] = 1 - 0{,}25 = 0{,}75$.
6. $P[t(8) \leq -2{,}306] = P[t(8) \geq 2{,}306] = 0{,}025$.

En la tabla se muestra la probabilidad de que la distribución $t(n)$ deje valores por encima (a la derecha) de un número dado.

Gráfica de la función de densidad de $t(n)$.

Proposición 5. *La función de densidad de la distribución $t(n)$ de Student cumple las siguientes propiedades:*

- *Es simétrica alrededor del 0 y unimodal.*
- *Para $n \to +\infty$, se aproxima a la gráfica de $\mathcal{N}(0,1)$.*
- *Tiene colas mayores que las de la normal, que van reduciéndose y aproximándose a las de la normal cuando n crece.*
- *Es más aplastada que la de la normal, es decir, es platicúrtica, y su zona central va creciendo y aproximándose a la de la normal conforme n crece.*

2.3. Distribución F de Snedecor.

Distribución F de Snedecor.

Definición 3. *Sean X e Y variables aleatorias independientes con distribuciones $X \rightsquigarrow \chi^2(m)$ e $Y \rightsquigarrow \chi^2(n)$. Entonces, la variable aleatoria*

$$F = \frac{X/m}{Y/n}$$

se dice que tiene una distribución F de Snedecor con (m,n) grados de libertad, y se denota por $F \rightsquigarrow F(m,n)$.

Propiedades de la distribución F de Snedecor.

Proposición 6.

i) La función de densidad de probabilidad de una distribución $F(m,n)$ viene dada por

$$g(f) = \frac{\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)} \left(\frac{m}{n}\right)^{m/2} f^{m/2-1}\left(1+\frac{m}{n}f\right)^{-\frac{m+n}{2}}, \quad f > 0$$

ii) Sea F una variable aleatoria con distribución $F(m,n)$. Entonces, se verifica que

$$E[F^r] = \left(\frac{n}{m}\right)^r \frac{\Gamma\left(\dfrac{m}{2}+r\right)\Gamma\left(\dfrac{n}{2}-r\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)}, \quad \text{para } 0 < r < \frac{n}{2}$$

En particular, $E[F] = \dfrac{n}{n-2}$ *si* $n > 2$, $E[F^2] = \dfrac{n^2(m+2)}{m(n-4)(n-2)}$ *si* $n > 4$, *luego se tiene que* $Var[F] = \dfrac{n^2(2m+2n-4)}{m(n-2)^2(n-4)}$ *si* $n > 4$.

iii) Se verifican las siguientes propiedades:

1. $F \rightsquigarrow F(m,n) \Leftrightarrow F^{-1} \rightsquigarrow F(n,m)$.
2. $T \rightsquigarrow t(n) \Leftrightarrow T^2 \rightsquigarrow F(1,n)$.

Demostración. Por su utilidad, solo vamos a demostrar los dos puntos del apartado *iii)*.

1. Si $F \rightsquigarrow F(m,n)$, entonces $F = \frac{X/m}{Y/n}$, donde $X \rightsquigarrow \chi^2(m), Y \rightsquigarrow \chi^2(n)$, con X e Y independientes. Así, se tiene que $F^{-1} = \frac{Y/n}{X/m}$, de nuevo el cociente de dos distribuciones χ^2 independientes, por tanto, una F de Snedecor, pero con los grados de libertad en orden inverso.
2. Si $T \rightsquigarrow t(n)$, entonces $T = \frac{X}{\sqrt{Y/n}}$, donde $X \rightsquigarrow \mathcal{N}(0,1)$ e $Y \rightsquigarrow \chi^2(n)$, con X e Y independientes. Por tanto, se tiene que $T^2 = \frac{X^2}{Y/n}$, y como ya demostramos que si $X \rightsquigarrow \mathcal{N}(0,1)$ entonces $X^2 \rightsquigarrow \chi^2(1)$, se tiene que T^2 es el cociente de una $\chi^2(1)$ dividida entre su grado de libertad (1) y una $\chi^2(n)$ dividida entre su grado de libertad (n), ambas independientes, y por tanto, una $T^2 \rightsquigarrow F(1,n)$. El recíproco se demuestra siguiendo la misma cadena de pasos en sentido opuesto.

■

Observación 4. *Se dan tablas donde* $F(m,n)$ *está tabulada para valores de* m *y* n *grandes y pequeños (no se estudiará ningún resultado de convergencia de la distribución* F *de Snedecor).*

Ejemplo 3. *Calcula el valor de* k *o la probabilidad indicada:*

1. $P[F(7,3) \leq k] = 0{,}95 \Rightarrow k = 8{,}89$.
2. $P[F(8,4) \geq k] = 0{,}01 = 1 - P[F(8,4) \leq k] \Rightarrow P[F(8,4) \leq k] = 0{,}99 \Rightarrow k = 14{,}8$.
3. $P[F(2,2) \leq 19] = 0{,}95$.
4. $P[F(3,5) \geq 12{,}1] = 1 - P[F(3,5) \leq 12{,}1] = 1 - 0{,}99 = 0{,}01$.
5. $P[F(60,40) \leq k] = 0{,}05 \Rightarrow k = 0{,}05$.

En la tabla se muestra la probabilidad de que la F *deje valores por debajo (a la izquierda) de un número dado.*

Gráfica de la función de densidad de $F(m;n)$.

Proposición 7. *La función de densidad de la distribución* $F(m,n)$ *de Snedecor cumple las siguientes propiedades:*

- *Es asimétrica a la derecha y unimodal.*
- *Para* $m = 1$, $\lim_{f\to 0} g(x) = +\infty$, *y la función de densidad es estrictamente decreciente.*
- *Para* $m = 2, g(0) = 1$ *y la función de densidad es estrictamente decreciente.*
- *Para* $m \geq 3, g(0) = 0$, *crece hasta la moda y luego decrece.*

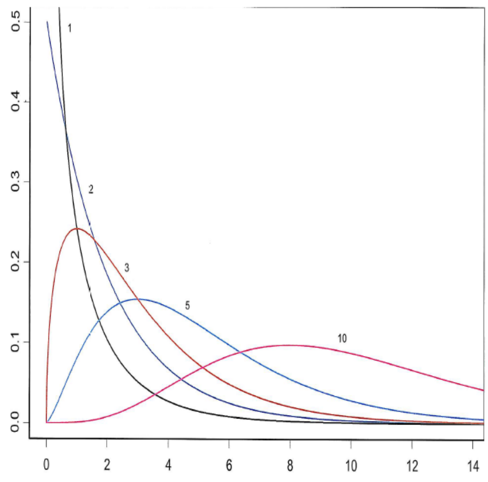


Figura 1: Funciones de densidad de casos particulares de $\chi^2(n)$.

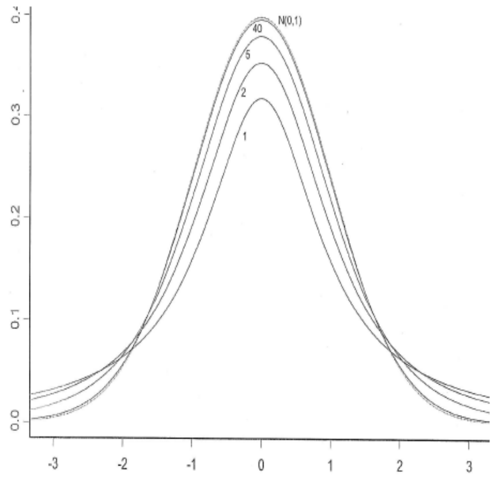


Figura 2: Funciones de densidad de casos particulares de $t(n)$.

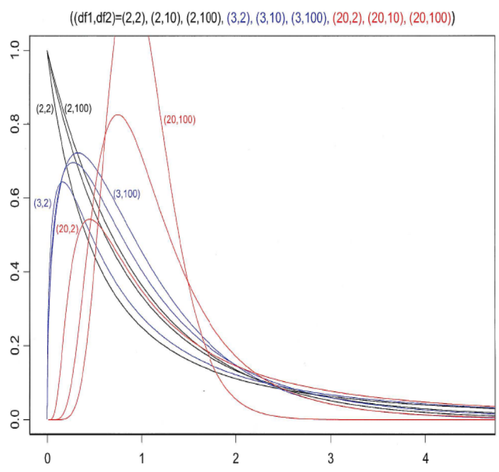


Figura 3: Funciones de densidad de casos particulares de $F(m, n)$.

3. Muestreo en poblaciones normales.

Vamos a estudiar, asociados a poblaciones que siguen una distribución normal, los estadísticos media muestral y cuasivarianza muestral y otros relacionados con ellos. En concreto, obtendremos la distribución en el muestreo de tales estadísticos, con el fin de poder hacer inferencia para una muestra tanto en una población normal unidimensional como en dos poblaciones normales unidimensionales.

3.1. Muestreo en una población normal unidimensional.

Teorema 1. *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$. Entonces, $\bar{X}$ y $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ son independientes.*

Demostración. Recordemos que $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2) \Rightarrow M_X(t) = \exp\left(t\mu + \frac{\sigma^2}{2}t^2\right)$.

Si $\bar{X}$ y $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$ fueran independientes, entonces su función generatriz de momentos conjunta sería el producto de las funciones generatrices de momentos marginales, luego debemos demostrar que se cumple la siguiente igualdad:

$$M_{(\bar{X}, X_1-\bar{X}, \ldots, X_n-\bar{X})}(t, t_1, \ldots, t_n) \overbrace{=}^{¿?} M_{\bar{X}}(t) M_{(X_1-\bar{X}, \ldots, X_n-\bar{X})}(t_1, \ldots, t_n)$$

Por la definición de función generatriz de momentos, se tiene que

$$M_{(\bar{X}, X_1-\bar{X}, \ldots, X_n-\bar{X})}(t, t_1, \ldots, t_n) = \mathrm{E}\left[\exp\left\{t\bar{X} + \sum_{i=1}^{N} t_i(X_i - \bar{X})\right\}\right] = (1)$$

Como $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ y $\sum_{i=1}^n t_i(X_i - \bar{X}) = \sum_{i=1}^n t_i X_i - \bar{X}\sum_{i=1}^n t_i$, entonces:

$$(1) = \mathrm{E}\left[\exp\left\{\frac{t}{n}\sum_{i=1}^n X_i + \sum_{i=1}^n t_i X_i - \frac{\sum_{i=1}^n X_i}{n}\sum_{i=1}^n t_i\right\}\right] = \mathrm{E}\left[\exp\left\{\sum_{i=1}^n \left(\frac{X_i t}{n} + t_i X_i - X_i \bar{t}\right)\right\}\right]$$
$$= \mathrm{E}\left[\prod_{i=1}^n \exp\left(\frac{X_i t}{n} + t_i X_i - X_i\bar{t}\right)\right] = \mathrm{E}\left[\prod_{i=1}^n \exp\left\{X_i\left(\frac{t}{n} + t_i - \bar{t}\right)\right\}\right] = (2)$$

Como todas las variables aleatorias que componen la muestra aleatoria simple son independientes (ind) e idénticamente distribuidas (i.d), se tiene entonces:

$$(2) \underbrace{=}_{\text{ind}} \prod_{i=1}^n \mathrm{E}\left[\exp\left\{X_i\left(\frac{t}{n} + t_i - \bar{t}\right)\right\}\right] = \prod_{i=1}^n M_{X_i}\left(\frac{t}{n} + t_i - \bar{t}\right) \underbrace{=}_{\text{i.d}} \prod_{i=1}^n M_X\left(\frac{t}{n} + t_i - \bar{t}\right)$$
$$= \prod_{i=1}^n \exp\left\{\mu\left(\frac{t}{n} + t_i - \bar{t}\right) + \frac{\sigma^2}{2}\left(\frac{t}{n} + t - \bar{t}\right)^2\right\} = \exp\left\{\sum_{i=1}^n\left[\mu\left(\frac{t}{n} + t_i - \bar{t}\right) + \frac{\sigma^2}{2}\left(\frac{t}{n} + t - \bar{t}\right)^2\right]\right\}$$
$$= \exp\left\{\sum_{i=1}^n \mu\frac{t}{n} + \sum_{i=1}^n t_i\mu - \sum_{i=1}^n \bar{t}\mu + \frac{\sigma^2}{2}\sum_{i=1}^n\left(\frac{t}{n} + t_i - \bar{t}\right)^2\right\} = (3)$$

Como $\sum_{i=1}^{n} t_i\mu - \sum_{i=1}^{n} \bar{t}\mu = \mu\left(\sum_{i=1}^{n} t_i - \sum_{i=1}^{n} \bar{t}\right) = \mu\sum_{i=1}^{n}(t_i - \bar{t}) = \mu \cdot 0 = 0$, entonces:

$$(3) = \exp\left\{\sum_{i=1}^{n} \mu\frac{t}{n} + \frac{\sigma^2}{2}\sum_{i=1}^{n}\left(\frac{t}{n} + t_i - \bar{t}\right)^2\right\} = \exp\left\{\mu t + \frac{\sigma^2}{2}\left(\frac{t^2}{n} + \sum_{i=1}^{n}(t_i-\bar{t})^2 + 2\sum_{i=1}^{n}\frac{t}{n}(t_i - \bar{t})\right)\right\}$$
$$= \exp\left\{\left(\mu t + \frac{\sigma^2}{2}\frac{t^2}{n}\right) + \left(\frac{\sigma^2}{2}\sum_{i=1}^{n}(t_i-\bar{t})^2\right)\right\} = \exp\left\{\mu t + \frac{\sigma^2}{2}\frac{t^2}{n}\right\}\exp\left\{\frac{\sigma^2}{2}\sum_{i=1}^{n}(t_i-\bar{t})^2\right\}$$

donde $\exp\left\{\mu t + \frac{\sigma^2}{2}\frac{t^2}{n}\right\} = M_{\bar{X}}(t)$ y $\exp\left\{\frac{\sigma^2}{2}\sum_{i=1}^{n}(t_i-\bar{t})^2\right\} = M_{(X_1-\bar{X},\dots,X_n-\bar{X})}(t_1,\dots,t_n)$. ■

Corolario 1. *Bajo las mismas hipótesis del Teorema 1, se tiene:*

i) $\bar{X} \rightsquigarrow \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

ii) ***Lema de Fisher (L.F).*** *$\bar{X}$ y S^2 son independientes.*

iii) $\frac{(n-1)S^2}{\sigma^2} \rightsquigarrow \chi^2(n-1)$.

iv) $\frac{\bar{X}-\mu}{S/\sqrt{n}} \rightsquigarrow t(n-1)$.

Demostración.

i) Aunque este resultado ya ha sido demostrado en el tema 1, volvemos a probarlo ahora pues, en este caso, el Teorema 1 hace evidente su demostración, ya que si calculamos la función generatriz de momentos de $\bar{X}$, se tiene:

$$M_{\bar{X}}(t) = M_{(\bar{X},X_1-\bar{X},\dots,X_n-\bar{X})}(t,0,\dots,0) = \exp\left(t\mu + \frac{\sigma^2}{2n}t^2\right)$$

que es la función generatriz de momentos de una $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

ii) Por el Teorema 1, se tiene que $\forall i = 1,\dots,n$, las diferencias $X_i - \bar{X}$ son independientes de $\bar{X}$, luego cualquier combinación lineal de esas diferencias, en particular, la cuasivarianza muestral S^2, también lo será.

iii) Si $X \rightsquigarrow \mathcal{N}(\mu,\sigma^2)$, sabemos que $\frac{X-\mu}{\sigma} \rightsquigarrow \mathcal{N}(0,1)$, y por la relación entre las distribuciones χ^2 y $\mathcal{N}(0,1)$ se tiene que $\left(\frac{X-\mu}{\sigma}\right)^2 \rightsquigarrow \chi^2(1)$. Así, $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \rightsquigarrow \chi^2(n)$.

Por otro lado, $S^2 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1} \Rightarrow \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}$.

Calculemos la función generatriz de momentos de $\dfrac{(n-1)S^2}{\sigma^2}$.

$$\begin{aligned}\frac{\sum_{i=1}^n (X_i-\mu)^2}{\sigma^2} &= \frac{\sum_{i=1}^n (X_i-\bar X+\bar X-\mu)^2}{\sigma^2}\\ &= \frac{\sum_{i=1}^n (X_i-\bar X)^2}{\sigma^2}+\frac{\sum_{i=1}^n (\bar X-\mu)^2}{\sigma^2}+\frac{2}{\sigma^2}(\bar X-\mu)\underbrace{\sum_{i=1}^n (X_i-\bar X)}_{=0}\end{aligned}$$

Como $\dfrac{\sum_{i=1}^n (\bar X-\mu)^2}{\sigma^2}=\dfrac{n(\bar X-\mu)^2}{\sigma^2}=\dfrac{(\bar X-\mu)^2}{\sigma^2/n}=\dfrac{(\bar X-\mu)}{\sigma/\sqrt{n}}\rightsquigarrow \mathcal{N}(0,1)$, ya que se tiene $\bar X\rightsquigarrow \mathcal{N}\left(\mu,\dfrac{\sigma^2}{n}\right)$, entonces $\left(\dfrac{\bar X-\mu}{\sigma/\sqrt{n}}\right)^2\rightsquigarrow \chi^2(1)$. Se tiene entonces:

$$\underbrace{\frac{\sum_{i=1}^n (X_i-\mu)^2}{\sigma^2}}_{Y_1}=\frac{\sum_{i=1}^n (X_i-\bar X^2)}{\sigma^2}+\left(\frac{\bar X-\mu}{\sigma/\sqrt{n}}\right)^2=\underbrace{\frac{(n-1)S^2}{\sigma^2}}_{Y_2}+\underbrace{\left(\frac{\bar X-\mu}{\sigma/\sqrt{n}}\right)^2}_{Y_3}$$

Por tanto, $M_{Y_1}(t)=\mathrm{E}[e^{tY_1}]=\mathrm{E}[e^{t(Y_2+Y_3)}]=\mathrm{E}[e^{tY_2}e^{tY_3}]\overbrace{=}^{(1)}\mathrm{E}[e^{tY_2}]\mathrm{E}[e^{tY_3}]=M_{Y_2}(t)M_{Y_2}(t)$, donde en la igualdad (1) hemos aplicado el Lema de Fisher, que nos asegura que como $\bar X$ y S^2 son independientes, entonces Y_2 e Y_3 también lo son, luego la esperanza de su producto (y de cualquier producto construido a partir de transformaciones de Y_2 e Y_3) es el producto de sus esperanzas.

Como $Y_1\rightsquigarrow\chi^2(n)$ por ser la suma de n variables aleatorias normales tipificadas, y ya demostramos que $Y_3\rightsquigarrow\chi^2(1)$, entonces se tiene:

$$(1-2t)^{-n/2}=M_{Y_2}(t)\cdot(1-2t)^{-1/2}\underbrace{\Rightarrow}_{Divido}M_{Y_2}(t)=(1-2t)^{-(n-1)/2}$$

donde $(1-2t)^{-(n-1)/2}$ es la función generatriz de momentos de una $\chi^2(n-1)$. Por tanto, $Y_2=\dfrac{(n-1)S^2}{\sigma^2}\rightsquigarrow\chi^2(n-1)$.

iv) Recordemos que si $X\rightsquigarrow\mathcal{N}(0,1)$ e $Y\rightsquigarrow\chi^2(n)$ son independientes, entonces $\dfrac{X}{\sqrt{Y/n}}\rightsquigarrow t(n)$.

Como $\dfrac{\bar X-\mu}{\sigma/\sqrt{n}}\rightsquigarrow\rightsquigarrow\mathcal{N}(0,1)$, $\dfrac{(n-1)S^2}{\sigma^2}\rightsquigarrow\chi^2(n-1)$, y ambas son independependientes (L.F):

$$\frac{\dfrac{\bar X-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{\cancel{(n-1)}S^2}{\sigma^2\cancel{(n-1)}}}}=\frac{\bar X-\mu}{\sigma/\sqrt{n}}:\frac{S}{\sigma}=\frac{\bar X-\mu}{\cancel{\sigma}/\sqrt{n}}\cdot\frac{\cancel{\sigma}}{S}=\frac{\bar X-\mu}{S/\sqrt{n}}\rightsquigarrow t(n-1)$$

■

Así, hemos obtenido una serie de estadísticos que usaremos para hacer inferencias sobre los parámetros de una población normal unidimensional en diferentes situaciones:

- Inferencia sobre μ:

σ_0^2 conocida:	$\dfrac{\bar{X}-\mu}{\sigma_0/\sqrt{n}} \rightsquigarrow \mathcal{N}(0,1)$
σ^2 desconocida:	$\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \rightsquigarrow t(n-1)$

- Inferencia sobre σ^2:

μ_0 conocida:	$\dfrac{\sum_{i=1}^n (X_i-\mu_0)^2}{\sigma^2} \rightsquigarrow \chi^2(n)$
μ desconocida:	$\dfrac{(n-1)S^2}{\sigma^2} \rightsquigarrow \chi^2(n-1)$

3.2. Muestreo en dos poblaciones normales unidimensionales.

Sean $X \rightsquigarrow \mathcal{N}(\mu_1,\sigma_1^2)$, $Y \rightsquigarrow \mathcal{N}(\mu_2,\sigma_2^2)$ dos variables aleatorias. Consideramos dos muestras aleatorias simples $(X_1,\ldots,X_{n_1})$, $(Y_1,\ldots,Y_{n_2})$ de X e Y, respectivamente. Sean $\bar{X},\bar{Y},S_1^2,S_2^2$ sus medias muestrales y sus cuasivarianzas.

Proposición 8. $\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}} \rightsquigarrow \mathcal{N}(0,1).$

Demostración. Basta con tipificar la variable $\bar{X}-\bar{Y} \rightsquigarrow \mathcal{N}\left(\mu_1-\mu_2, \dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}\right)$. ■

Lema extendido de Fisher.

Teorema 2. *Los vectores $(\bar{X},\bar{Y})$ y (S_1^2,S_2^2) son independientes.*

Demostración. Como X e Y son independientes, entonces:

$$M_{(\bar{X},S_1^2)}M_{(\bar{Y},S_2^2)} \underbrace{=}_{\text{L.F}} M_{\bar{X}}(t_1)M_{S_1^2}(t_3)M_{\bar{Y}}(t_2)M_{S_2^2}(t_4)$$

Por tanto:

$$M_{(\bar{X},\bar{Y},S_1^2,S_2^2)}(t_1,t_2,t_3,t_4) = M_{(\bar{X},\bar{Y})}(t_1,t_2)M_{(S_1^2,S_2^2)}(t_3,t_4)$$

luego $(\bar{X},\bar{Y})$ y (S_1^2,S_2^2) son independientes. ■

Corolario 2.

i) $\dfrac{n_2}{n_1}\dfrac{\sigma_2^2}{\sigma_1^2}\dfrac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2} \rightsquigarrow F(n_1,n_2)$.

ii) $\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \rightsquigarrow F(n_1-1,n_2-1)$.

En particular, si $\sigma_1=\sigma_2$, *se tiene que* $\dfrac{S_1^2}{S_2^2} \rightsquigarrow F(n_1-1,n_2-1)$.

iii) $\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{(n-1)S_1^2}{\sigma_1^2}+\dfrac{(n_2-1)S_2^2}{\sigma_2^2}}}\sqrt{\dfrac{n_1+n_2-2}{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}} \rightsquigarrow t(n_1+n_2-2)$.

En particular, si $\sigma_1=\sigma_2$, *entonces* $\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{(n_1-1)S_1^2+(n_2-1)S_2^2}}\sqrt{\dfrac{n_1+n_2-2}{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \rightsquigarrow t(n_1+n_2-2)$.

Demostración.

i) Como $\dfrac{\sum_{i=1}^{n}(X_i-\mu_1)^2}{\sigma_1^2} \rightsquigarrow \chi^2(n_1)$ y $\dfrac{\sum_{i=1}^{n}(Y_i-\mu_2)^2}{\sigma_2^2} \rightsquigarrow \chi^2(n_2)$, y son independientes por el lema extendido de Fisher, entonces:

$$\frac{\dfrac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sigma_1^2\cdot n_1}}{\dfrac{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2}{\sigma_2^2\cdot n_2}} = \frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sigma_1^2\cdot n_1}\cdot\frac{\sigma_2^2\cdot n_2}{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2} = \frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sum_{i=1}^{n_2}(Y_i-\mu_2)^2}\cdot\frac{\sigma_2^2}{\sigma_1^2}\cdot\frac{n_2}{n_1} \rightsquigarrow F(n_1,n_2)$$

ii) Como $\dfrac{(n_1-1)S_1^2}{\sigma_1^2} \rightsquigarrow \chi^2(n_1-1)$ y $\dfrac{(n_2-1)S_2^2}{\sigma_2^2} \rightsquigarrow \chi^2(n_2-1)$, y son independientes por el lema extendido de Fisher, entonces:

$$\frac{\dfrac{\cancel{(n_1-1)}S_1^2}{\sigma_1^2\cdot\cancel{(n_1-1)}}}{\dfrac{\cancel{(n_2-1)}S_2^2}{\sigma_1^2\cdot\cancel{(n_2-1)}}} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \rightsquigarrow F(n_1-1,n_2-2)$$

iii) Para construir una t de Student necesitamos una variable normal tipificada y una χ^2 que sean independientes. De la Proposición 8 obtenemos la que será el numerador de la variable aleatoria, y para el denominador necesitaremos una $\chi^2(n_1+n_2-2)$. Del punto *iii)* del Corolario 1 se deduce que:

$$\frac{(n_1-1)S_1^2}{\sigma_1^2} \rightsquigarrow \chi^2(n_1-1), \frac{(n_2-1)S_2^2}{\sigma_2^2} \rightsquigarrow \chi^2(n_2-1) \Rightarrow \frac{(n_1-1)S_1^2}{\sigma_1^2}+\frac{(n_2-1)S_2^2}{\sigma_2^2} \rightsquigarrow \chi^2(n_1-n_2-1)$$

Ahora, dividiendo la distribución $\mathcal{N}(0,1)$ que hemos considerado entre la χ^2 recién obtenida (dividida por sus grados de libertad) y simplificando el resultado, se termina la demostración puesto que son independientes, por el lema extendido de Fisher. ∎

Así, hemos obtenido los estadísticos que usaremos para hacer inferencias sobre los parámetros de dos poblaciones normales unidimensionales en los siguientes casos:

- Inferencia sobre $\mu_1 - \mu_2$ (comparación de medias):

σ_1^2, σ_2^2 conocidas:	$\dfrac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \rightsquigarrow \mathcal{N}(0,1)$
$\sigma_1^2 = \sigma_2^2 = \sigma^2$ desconocidas:	$\dfrac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{(n_1-1)S_1^2 + (n_2-1)S_2^2}} \sqrt{\dfrac{n_1+n_2-2}{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \rightsquigarrow t(n_1+n_2-2)$

- Inferencia sobre $\dfrac{\sigma_2^2}{\sigma_1^2}$ (comparación de varianzas):

μ_1, μ_2 conocidas:	$\dfrac{n_2}{n_1} \dfrac{\sigma_2^2}{\sigma_1^2} \dfrac{\sum_{i=1}^{n_1} (X_i - \mu_1)^2}{\sum_{i=1}^{n_2} (Y_i - \mu_2)^2} \rightsquigarrow F(n_1, n_2)$
μ_1, μ_2 desconocidas:	$\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \rightsquigarrow F(n_1-1, n_2-1)$

tema3teorIE.pdf
(Provisional) Apuntes tema 3

Tema 3: Suficiencia y completitud. [1]

Índice

En este tema se estudian algunas propiedades que deben cumplir los estadísticos para poder simplificar los datos de una muestra aleatoria simple sin que se pierda información relevante, así como familias de distribuciones para las que es seguro que existen estadísticos con tales propiedades.

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

1. Introducción.

Aunque el objetivo de seleccionar una muestra aleatoria simple de una población es reducir el número de datos poblacionales a estudiar, es posible que incluso una muestra que sea representativa de la población no pueda ser estudiada de forma exhaustiva. Por ello, será también necesario encontrar métodos que nos permitan sacar conclusiones sobre la muestra sin que ello implique un esfuerzo inasequible.

Nos centraremos en estudiar lo que, en el tema 1, definimos como *estadístico*, que es una función construida a partir de los datos muestrales. Así, el proceso de investigación se simplifica y ya no hay que observar cada elemento de la muestra sino solo los estadísticos asociados a ella. Por tanto, aunque el uso de cualquier estadístico (a los que denotamos $T(X_1, \dots, X_n)$, o simplemente, $T(X)$) implica una reducción en los datos muestrales, también supone una cierta pérdida de la información sobre la muestra.

Por ejemplo, supongamos que se lanza una moneda al aire 100 veces, y tomamos dos muestras aleatorias simples de los resultados de tales lanzamientos. Sean $(X_1, \dots, X_5)$ e $(Y_1, \dots, Y_5)$ dichas muestras. Si llamamos 1 al suceso *salir cara* y 0 al suceso *salir cruz* y se tiene que $\bar{X} = \bar{Y} = 0{,}4$, entonces sabemos que en cada muestra deberá haber dos 1 y tres 0. Sin embargo, no tenemos forma de saber si las muestras son $(1, 1, 0, 0, 0)$ o $(0, 1, 0, 0, 1)$, luego a pesar de que es más sencillo estudiar y obtener conclusiones a partir de las medias muestrales, hemos perdido la información sobre la forma en que se distribuyen los datos de tales muestras.

Se plantea el problema de encontrar estadísticos T tales que la información que se pierde al utilizarlos sea irrelevante para los fines con los que los estamos empleando. Para ello, definimos las propiedades de suficiencia y completitud.

Demostrar que un estadístico es suficiente y completo en ocasiones será complejo. Por ello, se estudiarán familias de distribuciones para las cuales la existencia de un estadístico que sea a la vez suficiente y completo está asegurada, incluso habiendo métodos para su obtención directa.

2. Estadísticos suficientes.

Dada una variable aleatoria X con distribución en una familia de distribuciones $\{F_\theta, \theta \in \Theta\}$ dependientes de un parámetro θ desconocido. Recordemos que nuestro objetivo es inferir el valor de θ a partir de una muestra aleatoria simple de X.

Estadístico.

Definición 1. *Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple de una variable aleatoria X. Un estadístico muestral asociado a X es una función $T : (\mathbb{R}^n, \mathcal{B}^n) \longrightarrow (\mathbb{R}^k, \mathcal{B}^k)$ medible, es decir, cumple $T^{-1}(B) \in \mathcal{B}^n, \forall B \in \mathcal{B}^k$, e independiente de cualquier parámetro desconocido. Lo denotamos por $T(X_1, \dots, X_n)$, o cuando se sobreentienda que es un estadístico, simplemente T o $T(X)$.*

Así, cualquier estadístico $T(X_1, \dots, X_n)$ de una variable aleatoria X resume la información contenida en una muestra aleatoria simple asociada, y las inferencias que hagamos sobre θ se harán a partir de los valores de tales estadísticos. Sin embargo, con el fin de perder la menor cantidad de información de la muestra que sea posible, se utilizarán estadísticos que cumplan ciertas propiedades. La primera de ellas será la de *suficiencia*.

Fisher (1922) dijo que *un estadístico es suficiente cuando contiene toda la información contenida en la muestra sobre el parámetro que se está considerando.* Sin embargo, esta definición es demasiado general y poco rigurosa, aunque sí que puede ser útil si la reescribimos con cuidado. En efecto, si un estadístico es suficiente al contener toda la información que puede dar la muestra sobre el parámetro desconocido eso significa que cualquier información adicional que la muestra pueda aportar no proporciona información relevante sobre θ.

Es evidente que la información más importante que nos puede dar la muestra es el propio valor del estadístico. Sin embargo, su valor cambia para cada muestra considerada, y debemos entender que un estadístico es suficiente cuando, al cambiar la muestra seleccionada, su validez para inferir el parámetro θ no cambia. Es decir, sea cual sea el valor de T que se obtenga, este sigue siendo igual de válido a la hora de hacer inferencias sobre θ.

Pero, ¿cómo puede ser un estadístico igual de válido para cualquier valor que se obtenga? Esto solo sería posible si, para cualquier muestra, el valor del estadístico fuese aproximadamente el mismo. Así, buscamos estadísticos cuyos valores no difieran demasiado dependiendo de la muestra utilizada, es decir, buscamos estadísticos cuyos posibles valores están *igualmente distribuidos*. Por ejemplo, si demostramos que una variable aleatoria toma siempre valores entre 0 y 5, no nos equivocamos al suponer que puede estar distribuida según una binomial $B(5, 0'4)$. Por tanto, si un estadístico toma siempre los mismos valores (aproximadamente), para cualquier muestra considerada, podemos suponer que su distribución es siempre la misma.

Si tal suposición se convierte en una imposición, es evidente que estaremos ante estadísticos que siempre tienen la misma distribución, independientemente de la muestra que se considere y del valor concreto del estadístico que esta pueda proporcionar. Esta es la definición de estadístico suficiente.

Estadístico suficiente.

Definición 2. *Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de una variable aleatoria $X \rightsquigarrow F \in \{F_\theta, \theta \in \Theta\}$. Un estadístico $T(X_1, \ldots, X_n)$ es suficiente para la familia de distribuciones considerada (o suficiente para θ) si la distribución de la muestra condicionada a cualquier valor del estadístico, $T(X_1, \ldots, X_n) = t$, es independiente de θ.*

Observación 1. *Para cualquier variable aleatoria y para cualquier muestra aleatoria simple, siempre existe un estadístico suficiente asociado. En efecto, si $T(X_1, \ldots, X_n) = (X_1, \ldots, X_n)$ es evidente que no se pierde información. A la propia muestra le llamaremos estadístico suficiente trivial.*

Ejemplo 1. *Sea X_1, X_2 variables aleatorias independientes con distribución de Poisson de parámetro λ. Probar que $X_1 + X_2$ es suficiente para λ, y que $X_1 + 2X_2$ no lo es.*

Solución.

- Por reproductividad, si $X_1, X_2 \rightsquigarrow P(\lambda) \Rightarrow T = X_1 + X_2 \rightsquigarrow P(2\lambda)$. Probemos que T es suficiente para λ. Para ver que lo es, demostramos que la probabilidad de cualquier valor de la muestra condicionada a cualquier valor del estadístico es independiente del parámetro λ.

$$P_\lambda[X_1 = x_1, X_2 = x_2 \,/\, T(X_1, X_2) = X_1 + X_2 = t] = \frac{P_\lambda[X_1 = x_1, X_2 = x_2, T = t]}{P_\lambda[T = t]}$$

lo que nos da dos posibles resultados:

i) 0 si $x_1 + x_2 \neq t$ pues el numerador se anula por ser una intersección de sucesos que dan lugar a un suceso imposible y, por tanto, de probabilidad 0 (no es posible que $X_1 = x_1, X_2 = x_2$ y $T = X_1 + X_2 \neq x_1 + x_2$). En este caso, la probabilidad (siempre 0) es independiente de λ.

ii) Desarrollamos el caso $x_1 + x_2 = t$:

$$\frac{P_\lambda[X_1 = x_1, X_2 = x_2, T = t]}{P_\lambda[T = t]} \underbrace{=}_{\text{suceso seguro}} \frac{P_\lambda[X_1 = x_1, X_2 = x_2]}{P_\lambda[T = t]}$$

$$\underbrace{=}_{\text{ind}} \frac{P_\lambda[X_1 = x_1] \cdot P_\lambda[X_2 = x_2]}{P_\lambda[T = t]} = \frac{\cancel{e^{-\lambda}}\dfrac{\cancel{\lambda}^{x_1}}{x_1!} \cdot \cancel{e^{-\lambda}}\dfrac{\cancel{\lambda}^{x_2}}{x_2!}}{\cancel{e^{-2\lambda}}\dfrac{(2\cancel{\lambda})^t}{t!}} = [t = x_1 + x_2] = \frac{(x_1 + x_2)!}{2^{x_1 + x_2} x_1! x_2!}$$

En este caso, tenemos que la probabilidad es independiente de λ.

Por tanto, T es un estadístico suficiente.

- Para demostrar que un estadístico T no es suficiente basta con encontrar un valor concreto de T para el cual su distribución condicionada a que tome tal valor no es independiente del parámetro. Como en este caso $T = X_1 + 2X_2$, tomando $X_1 = 0, X_2 = 1$ se tiene $T = 2$.

$$\frac{P_\lambda[X_1 = 0, X_2 = 1 \,/\, T = 2]}{P_\lambda[T = 2]} = \frac{P_\lambda[X_1 = 0, X_2 = 1, T = 2]}{P_\lambda[T = 2]} \underbrace{=}_{\text{suceso seguro}} \frac{P_\lambda[X_1 = 0, X_2 = 1]}{P_\lambda[T = 2]}$$

$$\underbrace{=}_{\text{ind}} \frac{P_\lambda[X_1 = 0] \cdot P_\lambda[X_2 = 1]}{P_\lambda[X_1 = 2, X_2 = 0] + P_\lambda[X_1 = 0, X_2 = 1]} = \frac{e^{-\lambda}\dfrac{\lambda^0}{0!} \cdot e^{-\lambda}\dfrac{\lambda^1}{1!}}{e^{-\lambda}\dfrac{\lambda^2}{2!} \cdot e^{-\lambda}\dfrac{\lambda^0}{0!} + e^{-\lambda}\dfrac{\lambda^0}{0!} \cdot e^{-\lambda}\dfrac{\lambda^1}{1!}}$$

$$= \frac{\cancel{e^{-2\lambda}} \cdot \cancel{\lambda}}{\cancel{e^{-2\lambda}}\left(\dfrac{\lambda^{\cancel{2}}}{2} + \cancel{\lambda}\right)} = \frac{1}{\dfrac{\lambda}{2} + 1} = \frac{2}{\lambda + 2} \quad \text{dependiente del parámetro } \lambda.$$

Ejemplo 2. *Demostrar que, en n lanzamientos de una moneda, el número de caras es suficiente para el parámetro p.*

Solución. El lanzamiento de una moneda se puede modelar con la variable aleatoria $X \rightsquigarrow B(1, p)$, donde

$$X = \begin{cases} 1 & \text{si se obtiene cara} \\ 0 & \text{si se obtiene cruz} \end{cases}$$

Sea $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X, demostrar que el número de caras es suficiente es equivalente a demostrar que el estadístico $T = \sum_{i=1}^n X_i \rightsquigarrow B(n, p)$ es suficiente. Para ver que lo es, demostramos que la probabilidad de cualquier valor de la muestra condicionada a cualquier valor del estadístico es independiente del parámetro p.

$$P_p[X_1 = x_1, \ldots, X_n = x_n \,/\, T = t] = \frac{P_p[X_1 = x_1, \ldots, X_n = x_n, T = t]}{P_p[T = t]}$$

lo que nos da dos posibles resultados:

i) 0 si $x_1 + \ldots + x_n \neq t$ pues el numerador se anula por ser una intersección de sucesos que dan lugar a un suceso imposible y, por tanto, de probabilidad 0. Así que, en este caso, la probabilidad (siempre 0) es independiente de p.

ii) Desarrollamos el caso $x_1 + \cdots + x_n = t$:

$$\frac{P_p[X_1 = x_1, \ldots, X_n = x_n, T = t]}{P_p[T = t]} \underbrace{=}_{\text{suceso seguro}} \frac{P_p[X_1 = x_1, \ldots, X_n = x_n]}{P_p[T = t]}$$

$$= \frac{P_p[X_1 = x_1] \cdots P_p[X_n = x_n]}{P_p[T = t]} = \frac{p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n}}{\binom{n}{t} p^t (1-p)^{n-t}}$$

$$[t = \sum x_i] = \frac{\cancel{p^t}\cancel{(1-p)^{n-t}}}{\binom{n}{t}\cancel{p^t}\cancel{(1-p)^{n-t}}} = \frac{1}{\binom{n}{t}} = \frac{1}{\binom{n}{\sum x_i}} \rightsquigarrow \text{Uniforme discreta}$$

En este caso, tenemos que la probabilidad es independiente de p.

Por tanto, T es un estadístico suficiente.

Es posible que haya ocasiones en las que se disponga del estadístico asociado a la muestra, pero no la muestra en sí. A partir de un estadístico suficiente, será posible reconstruirla.

Sea X una variable aleatoria con distribución $\{F_\theta, \theta \in \Theta\}$ y $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X. Se realiza un experimento que consiste en observar la muestra y se obtienen unos valores concretos (observaciones) $(x_1, \ldots, x_n)$. Asociada a la variable X se tiene un estadístico $T(X_1, \ldots, X_n)$ que, al aplicarlo a los datos observados, nos da un valor concreto de T, es decir, $T(X_1, \ldots, X_n) = t$.

Si se pierden los datos de la m.a.s. pero se conoce el valor del estadístico y no se puede volver a observar la variable para obtener datos equivalentes a los anteriores, para reconstruir la muestra se sigue el siguiente proceso:

- Se considera una variable aleatoria X^* que tendrá la distribución de la variable aleatoria $X/T = t$, de la que se toma una muestra aleatoria simple $(X_1^*, \ldots, X_n^*)$ (por tanto, tal muestra tiene la misma distribución que $(X_1, \ldots, X_n)/T = t$).
- Si la distribución $(X_1, \ldots, X_n)/T = t$ es independiente del parámetro θ, por tanto es conocida, se pueden observar las variables $(X_1^*, \ldots, X_n^*)$, que es una muestra equivalente a $(X_1, \ldots, X_n)$ ya que las dos muestras dan lugar al mismo valor del estadístico:

$$T(X_1, \ldots, X_n) = T(X_1^*, \ldots, X_n^*)$$

Reconstrucción de una muestra.

Ejemplo 3. *Se supone que se ha lanzado una moneda 100 veces, y se sabe que se han obtenido 60 caras pero se han perdido los datos originales de si salió cara o cruz en cada tirada. Reconstruir la muestra.*

Solución. La variable aleatoria que modela si en una tirada se ha obtenido cara o cruz es la misma X del Ejemplo 2, luego de nuevo, el número de caras vendrá dada por el valor del estadístico

$T = \sum X_i$, con $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X.

Ya demostramos que T es un estadístico suficiente, y sabemos que se distribuye según una uniforme discreta:

$$T \rightsquigarrow U\left(\binom{n}{t}\right)$$

Así, podemos seleccionar cualquier muestra en la que aparezcan sesenta 1 y cuarenta 0, puesto que esta será una de las $\binom{n}{t}$ posibles muestras que nos darán el mismo valor para el estadístico, $T = t$, y así obtener una muestra equivalente a la original.

2.1. Teorema de factorización de Neyman-Fisher.

Comprobar que un estadístico T es suficiente aplicando la definición (como en los ejemplos 1 y 2) puede ser complejo en el caso en el que no sepamos su distribución (en los casos anteriores, sí que era sencillo saberlo aplicando reproductividad). Además, si dada una familia de distribuciones nos piden encontrar un estadístico suficiente, la definición no nos da ningún método para poder hacerlo (la definición no es constructiva). Por tanto, necesitamos un criterio que nos permita tanto probar que un estadístico es suficiente como encontrar uno que lo sea.

Teorema de factorización de Neyman-Fisher.

Teorema 1. *Sea $(X_1, \ldots, X_n)$ una m.a.s. de una v.a. $X \rightsquigarrow F \in \{F_\theta, \theta \in \Theta\}$. Sea f_θ la función masa de probabilidad o función de densidad de X bajo F_θ y sea f_θ^n la f.m.p. o f.d.d. de la muestra bajo F_θ. Un estadístico $T(X_1, \ldots, X_n)$ es suficiente si y solo si,* ***para cualquier valor*** *de $\theta \in \Theta$ se cumple*

$$f_\theta^n(x_1, \ldots, x_n) = h(x_1, \ldots, x_n) g_\theta(T(x_1, \ldots, x_n)), \quad \forall (x_1, \ldots, x_n) \in \chi^n$$

donde h es independiente de θ y g_θ depende de $(x_1, \ldots, x_n)$ solo a través de $T(x_1, \ldots, x_n)$.

Demostración. La hacemos en el caso discreto (en el caso continuo basta sustituir la función masa de probabilidad por la función de densidad)

$$\boxed{\Rightarrow}\ T(X_1, \ldots, X_n) \text{ suficiente} \overset{¿?}{\Rightarrow} \underbrace{P[X_1 = x_1, \ldots, X_n = x_n]}_{f_\theta^n} = h(x_1, \ldots, x_n) g_\theta(T(x_1, \ldots, x_n)).$$

$$\begin{aligned} P_\theta[X_1 = x_1, \ldots, X_n = x_n \ / \ T = t] &= \frac{P_\theta[X_1 = x_1, \ldots, X_n = x_n, T = t]}{P_\theta[T = t]} \\ &= \frac{P_\theta[X_1 = x_1, \ldots, X_n = x_n]}{P_\theta[T = t]} \end{aligned}$$

Por tanto, se tiene:

$$P_\theta[X_1 = x_1, \ldots, X_n = x_n] = f_\theta^n = \underbrace{P[X_1 = x_1, \ldots, X_n = x_n \ / \ T = t]}_{h(x_1, \ldots, x_n)} \underbrace{P_\theta[T = t]}_{g_\theta(T)}$$

$\boxed{\Leftarrow}$ $\underbrace{P[X_1 = x_1, \ldots, X_n = x_n]}_{f_\theta^n} = h(x_1, \ldots, x_n) g_\theta(T(x_1, \ldots, x_n))\ \forall\theta \overbrace{\Rightarrow}^{¿?} T$ suficiente.

$$P_\theta[X_1 = x_1, \ldots, X_n = x_n \,/\, T = t] = \frac{P_\theta[X_1 = x_1, \ldots, X_n = x_n, T = t]}{P_\theta[T = t]} = \begin{cases} 0 & T(x_1, \ldots, x_n) \neq t \\ \underline{*} & T(x_1, \ldots, x_n) = t \end{cases}$$

Desarrollamos $\underline{*}$, donde $P_\theta[X_1 = x_1, \ldots, X_n = x_n, T = t] = P_\theta[X_1 = x_1, \ldots, X_n = x_n]$ por ser $T = t$ un suceso seguro.

$$\underline{*} = \frac{P_\theta[(X_1, \ldots, X_n) = (x_1, \ldots, x_n)]}{P_\theta[T = t]} = \frac{h(x_1, \ldots, x_n) g_\theta(T)}{\displaystyle\sum_{(x_1,\ldots,x_n) \,/\, T(x_1,\ldots,x_n)=t} P_\theta[X_1 = x_1, \ldots, X_n = x_n]}$$

$$= \frac{h(x_1, \ldots, x_n) g_\theta(T)}{\displaystyle\sum_{(x_1,\ldots,x_n) \,/\, T(x_1,\ldots,x_n)=t} h(x_1, \ldots, x_n) g_\theta(T)} = \frac{h(x_1, \ldots, x_n)}{\displaystyle\sum_{(x_1,\ldots,x_n) \,/\, T(x_1,\ldots,x_n)=t} h(x_1, \ldots, x_n)}$$

que no depende de θ. Por tanto, el estadístico T es suficiente.

■

Propiedades de los estadísticos suficientes.

Corolario 1. *Si $T(X_1, \ldots, X_n)$ es un estadístico suficiente para $\{F_\theta, \theta \in \Theta\}$, entonces:*

i) $T(X_1, \ldots, X_n)$ es suficiente para $\{F_\theta, \theta \in \Theta'\}$ para cualquier $\Theta' \subseteq \Theta$.

ii) Si $U(X_1, \ldots, X_n)$ es un estadístico tal que $T = h(U)$, entonces U es suficiente para la misma familia.

iii) Toda transformación biunívoca de T proporciona un estadístico suficiente para la misma familia.

Demostración.

i) Es evidente, pues se cumple $f_\theta^n(x_1, \ldots, x_n) = h(x_1, \ldots, x_n) g_\theta(T(x_1, \ldots, x_n))$ para cualquier $\theta \in \Theta$ y $\theta' \in \Theta' \subseteq \Theta \Rightarrow \theta' \in \Theta$.

ii) Como T es suficiente, entonces $f_\theta^n(x_1, \ldots, x_n) = r(x_1, \ldots, x_n) g_\theta(T(x_1, \ldots, x_n))$, donde r es una función independiente de θ, y si $T = h(U)$, se tiene entonces:

$$f_\theta^n(x_1, \ldots, x_n) = r(x_1, \ldots, x_n) g_\theta(h(U)(x_1, \ldots, x_n)) = r(x_1, \ldots, x_n) \underbrace{(g_\theta \circ h)}_{h^*}(U(x_1, \ldots, x_n))$$

luego U es, en efecto, un estadístico suficiente.

iii) Sea m tal transformación biunívoca. Aplicando el teorema de factorización, se tiene:

$$\begin{aligned} f_\theta^n(x_1, \ldots, x_n) &= h(x_1, \ldots, h_n) g_\theta(T(x_1, \ldots, x_n)) = h(x_1, \ldots, x_n) g_\theta(m^{-1}(m(Tx_1, \ldots, x_n))) \\ &= h(x_1, \ldots, x_n) \underbrace{(g_\theta \circ m^{-1})}_{g^*}(m(T(x_1, \ldots, x_n))) \end{aligned}$$

Por tanto, $m(T)$ es suficiente.

Observamos que la condición *ser biunívoca* es imprescindible, pues si la función no lo es, entonces es posible que no exista su función inversa.

■

Observación 2. *Podemos recordar mejor lo que dice el corolario 1 si lo reescribimos de la siguiente forma:*

i) *Si un estadístico es suficiente para una familia de distribuciones, lo es también para cualquier subfamilia suya.*

ii) *Si tengo un estadístico suficiente U que es función de otro estadístico T, entonces T también lo es.*

iii) *Para que una función de un estadístico suficiente sea otro estadístico suficiente, tal función debe ser inyectiva.*

Importante para un tipo test.

Ejemplo 4. .

- *Si un estadístico T es suficiente para el parámetro u de una familia de distribuciones Gamma, $F_1 = \{\Gamma(u, \lambda), u, \lambda \in \mathbb{R}^+\}$, entonces T es suficiente para el parámetro n de una familia de distribuciones Erlang, $F_2 = \{\varepsilon(n, \lambda), n \in \mathbb{N}, \lambda \in \mathbb{R}^+\}$, pues $F_2 \subseteq F_1$.*
- *Si $T = 3X_{(1)}^2 + 4$ es un estadístico suficiente, entonces el estadístico $X_{(1)}$ también lo es.*
- *Si $T = X_{(1)}$ es suficiente, entonces $3T^2 + 4$ no tiene por qué serlo, ya que $y = x^2$ no es una tranformación biunívoca.*

La diferencia entre estos dos últimos apartados radica en que en el primero se parte de que una transformación es suficiente, luego el estadístico que ha sido transformado también lo es, mientras que en el segundo se tiene un estadístico suficiente y se quiere comprobar si lo es también su transformación.

- *Si $T = X_{(n)}$ es un estadístico suficiente, entonces T^3 también lo es, ya que $y = x^3$ es una función inyectiva.*

Observación 3. *El teorema de factorización también es cierto si el parámetro θ es multidimensional, en cuyo caso el estadístico también es multidimensional. En particular, aunque el parámetro sea de dimensión 1, el estadístico suficiente puede tener dimensión mayor que 1. En general, se verifica:*

$$\textit{Dimensión del estadístico suficiente} \geq \textit{Dimensión del parámetro}$$

La demostración de tal hecho es evidente, puesto que si la familia de distribuciones depende de un conjunto de parámetros desconocidos, entonces en la función de densidad o función masa de probabilidad de la muestra aparecerán todos los parámetros, luego la función g dependiente de los parámetros desconocidos a través del estadístico considerado debe contenerlos a todos.

Ejemplo 5. *Sea X una variable aleatoria con distribución en $\{\mathcal{N}(\mu, \sigma^2); \mu \in \mathbb{R}, \sigma^2 > 0\}$. Estudiar la suficiencia en los casos de σ_0^2 conocida, μ_0 conocida y μ y σ^2 desconocidas.*

Solución. Si $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$, entonces $f_{\mu,\sigma^2}(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right)$.

- Supongamos σ_0^2 conocida. Entonces, estudiamos la suficiencia para el parámetro μ.

$$\begin{aligned}
f_\mu^n(x_1,\dots,x_n) &= \frac{1}{\sqrt{2\pi\sigma_0^2}}\exp\left(-\frac{(x_1-\mu)^2}{2\sigma_0^2}\right)\cdots\frac{1}{\sqrt{2\pi\sigma_0^2}}\exp\left(-\frac{(x_n-\mu)^2}{2\sigma_0^2}\right)\\
&= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\sum(x_i-\mu)^2}{2\sigma_0^2}\right) = \left[-\sum(x_i-\mu)^2 = -\sum x_i^2 - n\mu^2 + 2\mu\sum xi\right]\\
&= \underbrace{\frac{1}{(2\pi\sigma_0^2)^{n/2}}\exp\left(-\frac{\sum x_i^2}{2\sigma_0^2}\right)}_{h(x_1,\dots,x_n)}\underbrace{\exp\left(\frac{-n\mu}{2\sigma_0^2}\right)\exp\left(\frac{\mu\sum x_i}{\sigma_0^2}\right)}_{g_\mu(T)}
\end{aligned}$$

Por el teorema de factorización, se tiene que $T=\sum_{i=1}^n X_i$ es un estadístico suficiente para μ.

- Supongamos μ_0 es conocida. Entonces, estudiamos la suficiencia para el parámetro σ^2.

$$\begin{aligned}
f_{\sigma^2}^n(x_1,\dots,x_n) &= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(\frac{-\sum(x_i-\mu_0)^2}{2\sigma^2}\right)\\
&= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(\frac{-\sum x_i^2}{2\sigma^2}\right)\exp\left(\frac{-n\sigma_0^2}{2\sigma^2}\right)\exp\left(\frac{\mu_0\sum x_i}{\sigma_0^2}\right)\\
&= \underbrace{\frac{1}{(2\pi)^{n/2}}}_{h(x_1,\dots,x_n)}\underbrace{\frac{1}{\sigma^n}\exp\left(\frac{-\sum x_i^2}{2\sigma^2}\right)\exp\left(\frac{-n\sigma_0^2}{2\sigma^2}\right)\exp\left(\frac{\mu_0\sum x_i}{\sigma_0^2}\right)}_{g_{\sigma^2}(T)}
\end{aligned}$$

Por el teorema de factorización, $T=\left(\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i\right)$ es un estadístico suficiente para σ^2.

Podríamos también haber decidido no desarrollar $f_{\sigma^2}^n$, ya que en ese caso escribirla como el producto de h y g_{σ^2} es trivial. En ese caso, habríamos obtenido que el estadístico $T=\sum(X_i-\mu_0)^2$ es suficiente, llegando a la misma conclusión pues T, así definido, es la transformación de los estadísticos $T_1=\sum X_i^2$ y $T_2=\sum X_i$. Observemos que, tal y como se dice en la observación 3, aunque el parámetro σ^2 es unidimensional, el estadístico suficiente para σ^2 es bidimensional.

- Basta repetir el razonamiento anterior con $\mu_0=\mu$, para obtener entonces:

$$h(x_1,\dots,x_n)=\frac{1}{(2\pi)^{n/2}} \qquad g_{\mu,\sigma^2}(T)=\frac{1}{\sigma^n}\exp\left(\frac{-\sum x_i^2}{2\sigma^2}\right)\exp\left(\frac{-n\sigma_0^2}{2\sigma^2}\right)\exp\left(\frac{\mu\sum x_i}{\sigma_0^2}\right)$$

Así, $T=\left(\sum_{i=1}^n X_i^2, \sum_{i=1}^n X_i\right)$ es un estadístico suficiente para (μ,σ^2).

2.1.1. Suficiencia minimal.

Aunque no vayamos a estudiarlos, es conveniente destacar que, de entre todos los estadísticos suficientes para un parámetro θ de cualquier familia de distribuciones $\{F_\theta, \theta\in\Theta\}$, siempre existirá uno que sea función de todos los demás. Diremos que tal estadístico es *suficiente minimal* (al haber sido construido a partir de otros estadísticos contiene toda su información luego es, de entre todos, el que produce una menor pérdida de información relevante).

Para distribuciones discretas y continuas el estadístico minimal suficiente siempre existe, es único, y se obtiene aplicando el teorema de factorización a cocientes entre funciones masa de probabilidad o funciones de densidad puntuales en distintos puntos.

3. Estadísticos completos.

Al contrario que el concepto de suficiencia, la *completitud* es difícil de explicar de forma intuitiva. A pesar de ello, diremos que la completitud es una propiedad que descarta la existencia de funciones unidimensionales medibles no nulas con esperanza cero bajo el parámetro θ desconocido.

Familia de distribuciones completas.

Definición 3. *Sea $\{F_\theta, \theta \in \Theta\}$ una familia de distribuciones con función de densidad o función masa de probabilidad $\{f_\theta, \theta \in \Theta\}$. Diremos que dicha familia es completa si, para cualquier función medible unidimensional g cumpliendo $E_\theta[g(X)] = 0, \forall \theta \in \Theta$ se tiene que*

$$P_\theta[g(X) = 0] = 1, \quad \forall \theta \in \Theta$$

En la Definición 3 se ha considerado una transformación g de la variable aleatoria X. Consideramos ahora una transformación de un estadístico T. Diremos que este es completo si su distribución no admite la existencia de funciones unidimensionales medibles no nulas con esperanza cero bajo el parámetro desconocido.

Estadístico completo.

Definición 4. *Un estadístico $T(X_1, \dots, X_n)$ se dice que es completo para la familia de distribuciones de X si para cualquier función medible unidimensional g se tiene:*

$$E_\theta[g(T(X_1, \dots, X_n))] = 0, \forall \theta \in \Theta \Rightarrow P_\theta[g(T(X_1, \dots, X_n)) = 0] = 1, \forall \theta \in \Theta$$

En este caso no estudiaremos ningún teorema con el que comprobar que un estadístico es completo, sino que tendremos que hacerlo siempre aplicando la definición.

Observación 4. *El concepto de estadístico completo suele ir asociado al concepto de estadístico suficiente. De hecho, si se pide dar un estadístico suficiente y completo, lo que haremos será tomar el estadístico suficiente que proporciona el teorema de factorización, y comprobar después si es completo.*

Ejemplo 6. *Sea $(X_1, \dots, X_n)$ una muestra aleatoria simple $X \rightsquigarrow B(1, p), p \in (0, 1)$. Encontrar un estadístico suficiente y completo asociado a la muestra.*

Solución. Si $X \rightsquigarrow B(1, p)$, entonces $P_p[X = x] = p^x(1-p)^{1-x}$. Así, la función masa de probabilidad de una m.a.s. de tamaño n de X viene dada por:

$$P_p[X_1=x_1,\dots,X_n=x_n]=p^{x_1}(1-p)^{1-x_1}\cdots p^{x_n}(1-p)^{1-x_n}=p^{\sum x_i}(1-p)^{n-\sum x_i}$$
$$=1\cdot\underbrace{(1-p)^n\left(\frac{p}{1-p}\right)^{\sum x_i}}_{g_p(T)}\Rightarrow T=\sum_{i=1}^n X_i \text{ es suficiente.}$$

Si $f_\theta^n=g_\theta(T)$, podemos escribir $h(x_1,\dots,x_n)=1$ para poder aplicar el teorema de factorización.

Comprobemos que T es completo, es decir, $\mathrm{E}_p[g(T)]=0,\forall p\in(0,1)\overbrace{\Rightarrow}^{¿?}P_p[g(T)=0]=1,\forall p$. Para ello, utilizamos que $T\rightsquigarrow B(n,p)$. Así:

$$E_p[g(T)]=\underbrace{\sum_{t=0}^n g(t)\binom{n}{t}p^t(1-p)^{n-t}}_{(1)}=0,\forall p\in(0,1)$$

donde, desarrollando (1) se tiene que

$$(1)=\sum_{t=0}^n g(t)\binom{n}{t}(1-p)^n\left(\frac{p}{1-p}\right)^t=0=\underbrace{(1-p)^n}_{\neq 0}\sum_{t=0}^n g(t)\binom{n}{t}\left(\frac{p}{1-p}\right)^t$$

y esto es posible si y solo si $\underbrace{\sum_{t=0}^n g(t)\binom{n}{t}\left(\frac{p}{1-p}\right)^t}_{(2)}=0$, con (2) un polinomio en la variable $\dfrac{p}{1-p}$.

$$(2)=g(0)+g(1)n\left(\frac{p}{1-p}\right)+g(2)\binom{n}{2}\left(\frac{p}{1-p}\right)^2+\cdots+g(n)\binom{n}{n}\left(\frac{p}{1-p}\right)^n=0,\quad\forall p\in(0,1)$$

Así, tenemos que el polinomio dado en (2) se anula siempre, y esto es así si y solamente si todos sus coeficientes son cero. Por tanto:

$$\begin{array}{ccccc} g(0)=0 & g(1)\binom{n}{1}=0 & g(2)\binom{n}{2}=0 & \cdots & g(n)\binom{n}{n}=0\\ \Downarrow & \Downarrow & \Downarrow & \cdots & \Downarrow\\ g(0)=0 & g(1)=0 & g(2)=0 & \cdots & g(n)=0 \end{array}$$

Por tanto, se tiene que $\forall t\in\{0,\dots,n\}, g(t)=0$. Sin embargo, como es posible que $g(t)$ también se anule para otros valores de t, entonces se cumple

$$\{t/g(t)=0\}\supseteq\{t\in\{0,\dots,n\}\}$$

Por la monotonía de la probabilidad, se tendrá entonces

$$P_p\{g(T)=0\}\geq P_p\{T\in\{0,\dots,n\}\}$$

Sin embargo, T sigue una distribución binomial cuyo primer parámetro es n, luego solo puede tomar valores entre 0 y n, así que el suceso $\{T\in\{0,\dots,n\}\}$ es seguro. Por otro lado, como cualquier suceso tiene siempre probabilidad menor o igual que 1, entonces se cumple la siguiente cadena:

$$1\geq P_p\{g(T)=0\}\geq P_p\{T\in\{0,\dots,n\}\}=1\Rightarrow P_p\{g(T)=0\}=1$$

Por tanto, T es un estadístico completo.

Ejemplo 7. *Sea* $(X_1, \ldots, X_n)$ *una muestra aleatoria simple de* $X \rightsquigarrow U(0,\theta), \theta > 0$. *Encontrar un estadístico suficiente y completo asociado a la muestra.*

Solución. Si $X \rightsquigarrow U(0,\theta)$, entonces $f_\theta(x) = \frac{1}{\theta}, 0 < x < \theta$. Por tanto, se tiene:

$$f_\theta^n = \frac{1}{\theta} \cdots \frac{1}{\theta} = \left(\frac{1}{\theta}\right)^n, 0 < x_i < \theta, \forall i = 1, \ldots, n$$

Es decir, para que la función de densidad de la muestra aleatoria simple sea no nula, todos sus valores deben estar también entre 0 y θ. Por tanto, para que esto sea así, debe cumplirse que:

- El menor valor de la m.a.s. debe ser mayor que 0, puesto que esto asegura que todos los demás valores también lo son.
- El mayor valor de la m.a.s. debe ser menor que θ, ya que esto garantiza que el resto de valores también están por debajo de θ.

Por tanto, f_θ^n es no nula cuando se cumple $X_{(1)} > 0$ y $X_{(n)} < \theta$. Así, considerando la función indicadora, se llega finalmente a que la función de densidad de la muestra aleatoria simple es:

$$f_\theta^n = f_\theta(x_1, \ldots, x_n) = \left(\frac{1}{\theta}\right)^n I_{[X_{(1)}>0]}(x_1, \ldots, x_n) I_{[X_{(n)}<\theta]}(x_1, \ldots, x_n)$$

Tomando $h(x_1, \ldots, x_n) = I_{[X_{(1)}>0]}(x_1, \ldots, x_n)$, $g_\theta(T) = \left(\frac{1}{\theta}\right)^n I_{[X_{(n)}<\theta]}(x_1, \ldots, x_n)$, se llega a que $T(X_1, \ldots, X_n) = X_{(n)}$ es un estadístico suficiente. Ahora, debemos calcular su función de densidad.

$$F_{X(n)}(x) = (F_X(x))^n = \left(\frac{x}{\theta}\right)^n \Rightarrow f_{X_{(n)}}(x) = n(F_X(x))^{n-1} f_\theta(x) = n\left(\frac{x}{\theta}\right)^{n-1} \cdot \frac{1}{\theta},\ 0 < x < \theta$$

Comprobemos que T es completo. Para ello, calculamos $\mathrm{E}_\theta[g(T)]$.

$$\mathrm{E}_\theta[g(T)] = \int_0^\theta g(t) f_{X_{(n)}}(t)dt = \int_0^\theta g(t) n \frac{t^{n-1}}{\theta^n} = \frac{n}{\theta^n} \int_0^\theta g(t) t^{n-1} dt$$

Como tenemos $\mathrm{E}_\theta[g(T)] = 0, \forall\theta$ para comprobar la completitud, entonces:

$$\underbrace{\frac{n}{\theta^n}}_{\neq 0} \int_0^\theta g(t) t^{n-1} dt = 0\ \forall\theta \iff \int_0^\theta g(t) t^{n-1} dt = 0\ \forall\theta$$

Por el Teorema Fundamental del Cálculo, tal integral existe, y es

$$\int_0^\theta g(t) t^{n-1} dt = G(\theta) - G(0)$$

con G una primitiva de $g(t)t^{n-1}$. Así, derivando con respecto a θ y teniendo en cuenta que $G(0)$ no depende de θ, se tiene:

$$(G(\theta) - G(0))' = G'(\theta) = g(\theta) \underbrace{\theta^{n-1}}_{\neq 0} = 0\ \forall\theta \Rightarrow g(\theta) = 0\ \forall\theta$$

Ahora, razonando como en el Ejemplo 6 se tiene:

$$\{t/g(t) = 0\} \supseteq \{t > 0\} \Rightarrow 1 \geq P_\theta\{g(T) = 0\} \geq P_\theta\{T > 0\} = 1 \Rightarrow P_\theta\{g(T) = 0\} = 1$$

Por tanto, $X_{(n)}$ es un estadístico completo.

4. Suficiencia y completitud en familias exponenciales.

Estudiamos las llamadas familias de distribuciones exponenciales paramétricas, para las cuales el proceso de obtención de estadísticos suficientes y completos puede simplificarse si se dan las circunstancias adecuadas.

Sea $\mathcal{F} = \{F_\theta, \theta \in \Theta\}$ una familia de distribuciones de probabilidad paramétricas con función de densidad o función masa de probabilidad $\{f_\theta, \theta \in \Theta\}$.

4.1. Familia exponencial uniparamétrica. Familia exponencial k-paramétrica.

Familia exponencial uniparamétrica.

Definición 5. *Se dice que la familia de distribuciones $\mathcal{F}$ es exponencial uniparamétrica si se cumplen las siguientes tres condiciones:*

i) El espacio paramétrico es un intervalo real: $\Theta \subseteq \mathbb{R}$.

ii) El conjunto de valores de la variable no depende de θ:

$$\chi = \{x / f_\theta(x) > 0\} \text{ independiente de } \theta, \ \forall \theta \in \Theta$$

iii) 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:

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

Importante para un tipo test.

Ejemplo 8. *Sea X una variable aleatoria con función de densidad de probabilidad $f_\theta(x) = e^{x-2\theta}, \theta > 0, x > \theta$. Entonces, aunque se cumplan las condiciones* i) *y* iii) *de la Definición 5, observemos que no se cumple* ii) *pues el conjunto χ está formado por valores que deben ser mayores que θ. Por tanto, χ depende de θ, luego en este caso la distribución no es exponencial uniparamétrica.*

Ejemplo 9. *Comprobar que la familia $\{P(\lambda), \lambda > 0\}$ es una familia exponencial uniparamétrica.*

Solución. Si $X \rightsquigarrow P(\lambda)$, entonces $P_\lambda[X = x] = e^{-\lambda}\dfrac{\lambda^x}{x!}, x = 0, 1, 2, \ldots$. Por tanto:

i) $\Theta = \mathbb{R}^+$ intervalo de $\mathbb{R}$, luego se cumple la primera condición.

ii) Una distribución de Poisson toma valores positivos o el 0, independientemente de cuál sea su parámetro. Por tanto, $\chi = \{x / P_\lambda[X = x] > 0\} = \{0, 1, 2, \ldots\}$ independiente de λ, luego se cumple la segunda condición.

iii) Reescribimos su función masa de probabilidad. Para ello, debemos calcular primero su exponencial y después tomar logaritmos para que no cambie:

$$P_\lambda[X = x] = e^{-\lambda}\frac{\lambda^x}{x!} = \exp\left\{\ln\left(e^{-\lambda}\frac{\lambda^x}{x!}\right)\right\} = \exp\{-\lambda + x\ln\lambda - \ln x!\}$$

y tomamos $D(\lambda) = -\lambda$, $Q(\lambda)T(x) = x\ln\lambda$, $S(x) = \ln x!$ para que se cumpla la condición.

Por tanto, es una distribución exponencial uniparamétrica.

Ejemplo 10. *Comprobar que la familia $\{B(k_0, p), p \in (0,1)\}$ es una familia exponencial uniparamétrica.*

Solución. Si $X \rightsquigarrow B(k_0, p), p \in (0,1)$, entonces $P_p[X = x] = \binom{k_0}{p} p^x (1-p)^{k_0 - x}, x = 0, 1, \dots, k_0$.
Por tanto:

i) $\Theta = (0,1)$ intervalo de $\mathbb{R}$, luego se cumple la primera condición.

ii) Una distribución Binomial de parámetros k_0 y p puede tomar valores desde el 0 hasta k_0, parámetro que no estamos considerando desconocido. Por tanto, $\chi = \{x/P_p[X = x] > 0\} = \{0, 1, 2, \dots, k_0\}$ independiente de p, luego se cumple la segunda condición.

iii)
$$\begin{aligned} P_p[X = x] &= \exp\left\{\ln \binom{k_0}{x} + x \ln p + (k_0 - x) \cdot \ln(1-p)\right\} \\ &= \exp\left\{\ln \binom{k_0}{x} + x(\ln p - \ln(1-p)) + k_0 \ln(1-p)\right\} \end{aligned}$$

y tomamos $S(x) = \ln \binom{k_0}{x}$, $Q(p)T(x) = x(\ln p - \ln(1-p))$, $D(p) = k_0 \ln(1-p)$. para que se cumpla la condición.

Por tanto, es una distribución exponencial uniparamétrica.

Familia exponencial k-paramétrica.

Definición 6. *Se dice que la familia de distribuciones $\mathcal{F}$ es exponencial k-paramétrica si se cumplen las siguientes tres condiciones:*

i) El espacio paramétrico es un intervalo real multidimensional: $\Theta \subseteq \mathbb{R}^k$.

ii) El conjunto de valores de la variable no depende de $\theta = (\theta_1, \dots, \theta_k)$:
$$\chi = \{x/f_\theta(x) > 0\}, \forall \theta \in \Theta$$

iii) Existen funciones real-valuadas $Q_h(\theta)$ y $D_h(\theta)$ con $h = 1, \dots, k$ definidas sobre Θ, y existen funciones medibles Borel T y S, también real valuadas, tales que:
$$\forall \theta \in \Theta,\ f_\theta(x) = \exp\left[\sum_{h=1}^{k} Q_h(\theta) T_h(x) + D(\theta) + S(x)\right],\ x \in \chi$$

Ejemplo 11.

- *Comprobar que la familia de distribuciones normales $\{\mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0\}$ es bi-paramétrica en los dos parámetros.*
- *Comprobar que es uniparamétrica en cada parámetro.*

Solución. Si $X \rightsquigarrow \mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma^2 > 0$, entonces $f_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right), x \in \mathbb{R}$.

- Caso biparamétrico.
  i) $\Theta = (\mathbb{R}, \mathbb{R}^+)$ intervalo de $\mathbb{R}^2$, luego se cumple la primera condición.
  ii) $\chi = \{x/f_{\mu,\sigma^2} > 0\} = \mathbb{R}$ independiente de μ y σ^2, luego se cumple la segunda condición.
  iii) Haciendo la exponencial de la función de densidad y después tomando logaritmos, se llega a:

$$f_{\mu,\sigma^2}(x) = \exp\left\{-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma^2 - \frac{(x-\mu)^2}{2\sigma^2}\right\}$$
$$= \exp\left\{-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma^2 - \frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}\right\}$$

  Para que se cumpla la condición, se toman las funciones $S(x) = -\frac{1}{2}\ln(2\pi)$, $D(\mu,\sigma^2) = -\frac{1}{2}\ln\sigma^2 - \frac{\mu^2}{2\sigma^2}$, $T_1(x) = x^2$, $Q_1(\mu,\sigma^2) = -\frac{1}{2\sigma^2}$, $T_2(x) = x$, $Q_2(\mu,\sigma^2) = \frac{\mu}{\sigma^2}$.

  Por tanto, es una distribución exponencial bi-paramétrica.

  Observemos que podríamos haber cogido una función distinta para $T_1(x)$ (podríamos haberle puesto un denominador, por ejemplo). En general, nos conviene (y se verá en el teorema siguiente) coger, para $T_h(x)$, las funciones más simples.

- Para el caso uniparamétrico, con cambios mínimos se puede comprobar a partir del punto anterior que se cumplen las condiciones *i)* y *ii)*
  - Si suponemos que μ es el parámetro desconocido, entonces para que se cumpla *iii)* escogemos las funciones:

$$S(x) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma^2 - \frac{x^2}{2\sigma^2}, \quad D(\mu) = \frac{-\mu^2}{2\sigma^2}, \quad Q(\mu)T(x) = \frac{\mu x}{\sigma^2}$$

  - Si suponemos que el parámetro que no conocemos es σ^2, entonces se toman las siguientes funciones:

$$S(x) = -\frac{1}{2}\ln(2\pi), \quad D(\sigma^2) = -\frac{1}{2}\ln\sigma^2, \quad Q(\sigma^2)T(x) = \frac{1}{\sigma^2}\left(-\frac{x^2+\mu^2}{2} + \mu x\right)$$

  En ambos casos, se tiene que la distribución normal es exponencial uniparamétrica.

4.2. Suficiencia y completitud.

Suficiencia y completitud en familias exponenciales.

Teorema 2. *Sea una variable aleatoria $X \rightsquigarrow \{F_\theta, \theta \in \Theta\}$ con familia de funciones asociadas $\{f_\theta, \theta \in \Theta\}$, siendo f.d.d. o f.m.p. según el caso. Si la familia de distribuciones es exponencial k-paramétrica, entonces la familia de distribuciones asociadas a una muestra aleatoria simple $(X_1, \ldots, X_n)$ de X es también exponencial k-paramétrica:*

$$f_\theta^n(x_1,\ldots,x_n) = \exp\left\{\sum_{h=1}^{k} Q_h(\theta)\left(\sum_{i=1}^{n} T_h(x_i)\right) + \sum_{i=1}^{n} S(x_i) + nD(\theta)\right\}, (x_1,\ldots,x_n) \in \chi^n$$

Además, se tiene:

- *El estadístico* $\left(\sum_{i=1}^{n} T_1(X_i), \dots, \sum_{i=1}^{n} T_k(X_i)\right)$ *es suficiente para* θ.
- *Si* $k \leq n$ *y el conjunto imagen de la función* $Q(\theta) = (Q_1(\theta), \dots, Q_k(\theta))$ *contiene a un abierto de* $\mathbb{R}^k$*, el estadístico* $\left(\sum_{i=1}^{n} T_1(X_i), \dots, \sum_{i=1}^{n} T_k(X_i)\right)$ *es también completo.*

Demostración. Demostramos solo la primera parte del teorema: si la familia de distribuciones es exponencial, también lo es la familia de distribuciones asociadas a una muestra aleatoria simple.

i) Θ es un intervalo de $\mathbb{R}^k$ por ser el producto de intervalos de $\mathbb{R}$.

ii) $\chi^n = \chi \times \cdots \times \chi$ es independiente de θ por ser el producto cartesiano de conjuntos independientes de θ.

iii)

$$\begin{aligned}
f_\theta(n) &= f_\theta(x_1) \cdots f_\theta(x_n) = \exp\{\cdots\} \cdots \exp\{\cdots\} \\
&= \exp\left\{\sum_{h=1}^{k} T_h(x_1) Q_h(\theta) + D(\theta) + S(x_1) + \cdots + \sum_{h=1}^{k} T_h(x_n) Q_h(\theta) + D(\theta) + S(x_n)\right\} \\
&= \exp\left\{nD(\theta) + \sum_{i=1}^{n} S(x_i) + \sum_{i=1}^{n} \sum_{h=1}^{k} T_k(x_i) Q_h(\theta)\right\} \\
&= \exp\left\{nD(\theta) + \sum_{i=1}^{n} S(x_i) + \sum_{h=1}^{k} Q_h(\theta) \sum_{i=1}^{n} T_h(x_i)\right\}
\end{aligned}$$

■

Ejemplo 12. *Sea* X *una variable aleatoria con distribución* $\mathcal{N}(\mu, \sigma^2)$ *con* $\mu \in \mathbb{R}$ *y* $\sigma^2 > 0$*. Hallar un estadístico suficiente y completo basado en una muestra aleatoria simple de tamaño* n*.*

Solución. Recordemos que, en este caso, se tenían las funciones

$$T_1(x) = x^2,\ Q_1(\mu, \sigma^2) = -\frac{1}{2\sigma^2},\ T_2(x) = x,\ Q_2(\mu, \sigma^2) = \frac{\mu}{\sigma^2}$$

Por el Teorema 2 se tiene que $\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ es un estadístico suficiente para (μ, σ^2).

Como $\text{Im}(Q_1) = \mathbb{R}^-, \text{Im}(Q_2) = \mathbb{R}$ y evidentemente $(\mathbb{R}^-, \mathbb{R})$ contiene a un abierto de $\mathbb{R}^2$ (de hecho, él mismo lo es), entonces $\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ es también completo.

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:

$$R_{T_k}^L(\theta) = E_\theta\left[\frac{(\theta - T_k)^2}{\theta^2}\right] = E_\theta\left[\frac{\cancel{\theta^2}}{\cancel{\theta^2}} - \frac{2\cancel{\theta} T_k}{\cancel{\theta^2}} + \frac{T_k^2}{\theta^2}\right] = 1 - \frac{2}{\theta}E_\theta[T_k] + \frac{1}{\theta^2}E_\theta[T_k^2]$$
$$= 1 - \frac{2}{\theta}E_\theta[kT_1] + \frac{1}{\theta^2}E_\theta[k^2T_1^2] = 1 - \frac{2k}{\theta}E[T_1] + \frac{k^2}{\theta^2}E[T_1^2]$$

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 $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X.

Estimador insesgado.

Definición 6. *Un estimador $T(X_1, \ldots, X_n)$ de θ es insesgado o centrado en el parámetro θ si su sesgo asociado es idénticamente nulo o, equivalentemente, si*

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

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

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

Ejemplo 5. *Sea $(X_1, \ldots, X_n)$ 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 $x_1 \leq \mu \leq x_k$ (en una población donde la variable X presenta k modalidades), entonces $x_1 \leq X_1 \leq \bar{X} \leq X_n \leq x_k$ 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 $\theta = (\theta_1, \ldots, \theta_k)$ 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 $T = (T_1, \ldots, T_k)$ es insesgado en θ si se verifica*

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

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

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

- *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 $(X_1, \ldots, X_n)$ 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 $X_1,\dots,X_n$ 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:

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

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

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

Por tanto, $h(T) = \dfrac{n+1}{n}T$ 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 $1/\theta$.

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

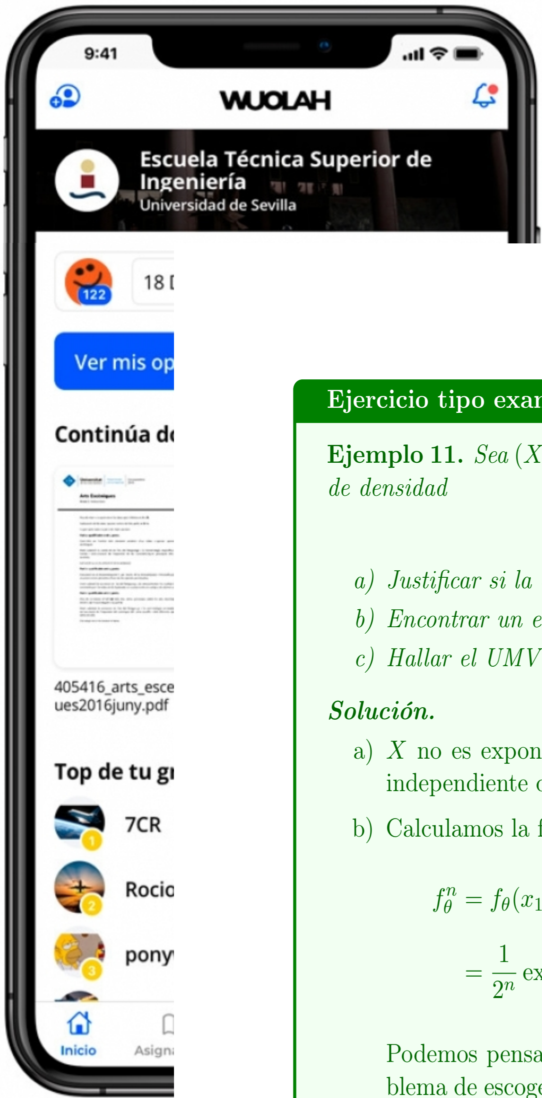
Aplicando el Teorema Fundamental del Cálculo, se tiene que existe H primitiva de $h(t)t^{n-1}$, y se cumple

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

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

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

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



Ejercicio tipo examen.

Ejemplo 11. *Sea $(X_1,\ldots,X_n)$ 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.

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

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

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

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

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

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

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

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:

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

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

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 $(*)$.

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

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

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

- Comprobemos que X tampoco puede ser degenerada.

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

- Por tanto, debe darse la otra opción:

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

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

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 $g(\theta)$, $T(X_1, \dots, X_n)$, se dice que es eficiente si es regular, insesgado, y su varianza alcanza la cota de FCR para cualquier valor del parámetro, es decir,*

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

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 $T(X_1, \dots, X_n)$ si:*

i) $g(\theta)$ es constante y, en tal caso, $T(X_1, \dots, X_n)$ 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:

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

Si T es eficiente, entonces $\text{Var}[T] = 0$, es decir, $T(X_1, \dots, X_n)$ 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:

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

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

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

y hemos llegado a una contradicción. ■

Caracterización de estimadores eficientes.

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

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

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

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

ii)

$$I_{X_1,\dots,X_n}(\theta) = \text{Var}_\theta\left[\frac{\partial \ln f_\theta^n(x_1, \dots, x_n)}{\partial\theta}\right] = \text{Var}_\theta\left[a(\theta)(T(X_1, \dots, X_n) - g(\theta))\right]$$
$$= a^2(\theta)\text{Var}_\theta[T - g(\theta)] = a^2(\theta)\text{Var}_\theta[T] \underbrace{=}_{T \text{ efi}} = a^2(\theta)\frac{(g'(\theta))^2}{I_{X_1,\dots,X_n}(\theta)}$$

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

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

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

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

teniendo en cuenta que $Q(\theta)=\displaystyle\int a(\theta)d\theta$, $D(\theta)=\displaystyle\int a(\theta)g(\theta)d\theta$ y que $S(x_1,\dots,x_n)$ 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:

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

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:

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

$$\Rightarrow a(p)g(p) = -nk_0\left(\frac{1}{1-p}\cdot(-1)\right) = nk_0\frac{1}{1-p} = \underbrace{\frac{1}{p(1-p)}}_{a(p)}\cdot g(p) \Rightarrow g(p) = nk_0p$$

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 $I_{X_1,\ldots,X_n}=a(\mu)g'(\mu)$.

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

Por otro lado:

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

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

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

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:

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

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

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.

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\begin{array}{cc}
1^{\text{a}} \text{ opción} & 2^{\text{a}} \text{ opción} \\
\downarrow & \downarrow \\
a\cdot\dfrac{2n}{\theta}+b,\ a,b\in\mathbb{R} & a\cdot\dfrac{2}{\theta}+b,\ a,b\in\mathbb{R}
\end{array}$$

- Estimadores eficientes para las funciones paramétricas:

$$\begin{array}{cc}
1^{\text{a}} \text{ opción} & 2^{\text{a}} \text{ opción} \\
\downarrow & \downarrow \\
a\cdot\sum_{i=1}^n X_i+b,\ a,b\in\mathbb{R} & a\cdot\dfrac{1}{n}\sum_{i=1}^n X_i+b=a\cdot\bar{X}+b,\ a,b\in\mathbb{R}
\end{array}$$

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

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

Veamos que se alcanza la cota.

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

Corolario 2. *Si $T(X_1,\dots,X_n)$ es eficiente para $g(\theta)$, entonces $T(X_1,\dots,X_n)$ es el UMVUE de $g(\theta)$. El recíproco no es cierto: si $T(X_1,\dots,X_n)$ es el UMVUE para $g(\theta)$ eso no implica que $T(X_1,\dots,X_n)$ 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 $\text{Var}_\theta[T(X_1,\dots,X_n)] \geq \dfrac{(g'(\theta))^2}{I_{X_1,\dots,X_n}} = \text{cota}$.

- $g(\theta) = \theta^2 \Rightarrow g'(\theta) = 2\theta \Rightarrow (g'(\theta))^2 = 4\theta^2$.
- 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.

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

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

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

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: $ag(\theta)+b$.

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

- Estimadores eficientes para las funciones paramétricas:

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

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 $T(X_1, \dots, X_n)$ es eficiente para $g(\theta)$, entonces $T(X_1, \dots, X_n)$ 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 $T(X_1, \dots, X_n)$ es completo.

tema5teorIE.pdf

(Provisional) Apuntes tema 5

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 $f_\theta^n(x_1, \ldots, x_n)$.

En la función de probabilidad muestral $f_\theta^n(x_1, \ldots, x_n)$ 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 $(x_1, \ldots, x_n)$ 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 $f_\theta^n(x_1, \ldots, x_n)$ para una realización muestral $(x_1, \ldots, x_n)$ concreta. Por tanto, podemos definir la función de verosimilitud como la función de probabilidad muestral con la realización muestral $(x_1, \ldots, x_n)$ 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 $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X, y sea $f_\theta^n(x_1, \ldots, x_n)$ su f.m.p. o su f.d.d. conjunta, con $\theta \in \Theta$. Para cada realización muestral $(x_1, \ldots, x_n)$ se define la función de verosimilitud asociada a dichos valores de la muestra como una función de θ de la siguiente manera:*

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

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 $(X_1, \dots, X_n)$ una muestra aleatoria simple de X. Un estimador $\widehat{\theta}(X_1, \dots, X_n)$ de θ es estimador de máxima verosimilitud (EMV) de θ si:*

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

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, $\ln L_{x_1,\dots,x_n}(\theta)$, 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 $\ln L_{x_1,\dots,x_n}(\theta)$ es derivable con respecto a θ, se obtienen las ecuaciones de verosimilitud:

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

Las soluciones de estas ecuaciones son los posibles extremos de $\ln L_{x_1,\dots,x_n}(\theta)$ y, por tanto, los posibles extremos de $L_{x_1,\dots,x_n}(\theta)$, 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 $L_{x_1,\dots,x_n}(\theta)$ o, incluso, métodos numéricos, para obtener el máximo.

Ejemplo 3. *Sea $(X_1, \dots, X_n)$ 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:

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

Por tanto:

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

Así:

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

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

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

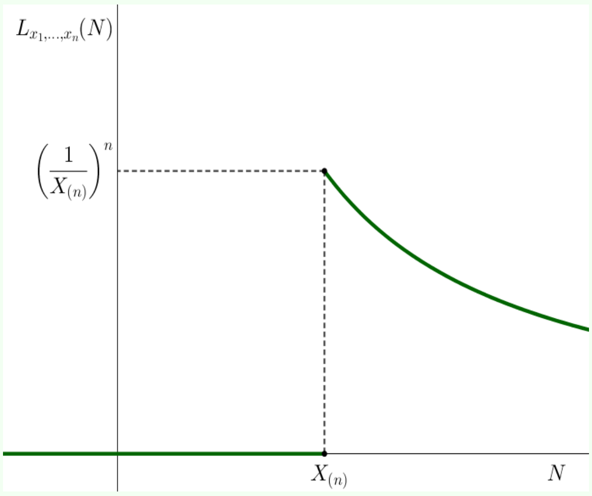
Si consideramos una muestra aleatoria simple $(X_1,\dots,X_n)$ de X, la función masa de probabilidad muestral será no nula si todos los valores X_i se encuentran entre 1 y N, es decir, si $I_{[X_{(1)}\geq 1]}$ y si $I_{[X_{(n)}\leq N]}$. Asumiendo que la condición $I_{[X_{(1)}\geq 1]}$ se cumple siempre,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

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

donde en la función de verosimilitud escribimos el parámetro desconocido N en función del estadístico. En este, caso, $L_{x_1,\dots,x_n}(N)$ 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 $L_{x_1,\dots,x_n}(N)$ valdrá 0 puesto que, en tal caso, $I_{[N\geq I_{X_{(n)}}]} = 0$.
- Cuando $N \geq X_{(n)}$, $L_{x_1,\dots,x_n}(N)$ es no nula pues $I_{[N\geq I_{X_{(n)}}]}$ será siempre 1, y $\frac{1}{N^n} \neq 0$.

El máximo valor de $L_{x_1,\dots,x_n}(N)$ 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 $L_{x_1,\dots,x_n}(N)$ irá decreciendo puesto que la función $\frac{1}{N^n}$ es decreciente.



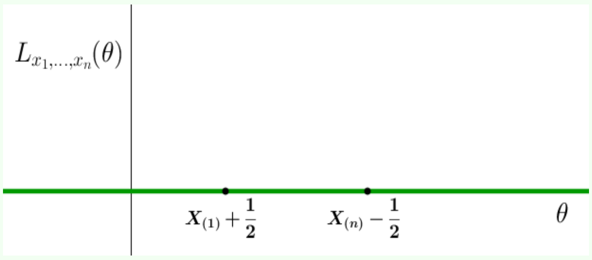
Ejemplo 5. *Sea $(X_1,\dots,X_n)$ una muestra aleatoria simple de $X \rightsquigarrow \{U(\theta-1/2,\theta+1/2),\theta\in\mathbb{R}\}$. Encontrar el EMV para θ.*

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:

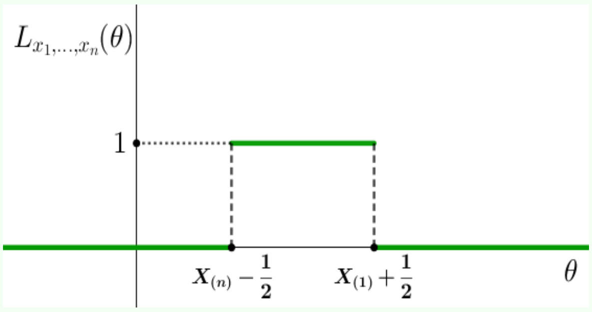
$$f_\theta^n(x_1,\dots,x_n) = 1\cdot I_{[X_{(1)}>\theta-1/2]}\cdot I_{[X_{(n)}<\theta+1/2]} = 1\cdot I_{[\theta<X_{(1)}+1/2]}\cdot I_{[\theta>X_{(n)}-1/2]} = L_{x_1,\dots,x_n}(\theta)$$

Así, tenemos dos formas de representar $L_{x_1,\dots,x_n}(\theta)$, 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, $L_{x_1,\dots,x_n}(\theta) = 0$, $\forall\theta\in\Theta$.



- 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, $L_{x_1,\dots,x_n}(\theta) = 1$.



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 $(X_1, \ldots, X_n)$ una muestra aleatoria simple de X. Supongamos que la familia admite un estadístico suficiente $T(X_1, \ldots, X_n)$. Entonces, si existe un EMV de θ, este es una función (no constante) de $T(X_1, \ldots, X_n)$.*

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

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

Como $L_{x_1,\ldots,x_n}(\theta) \geq 0$, entonces $h(x_1, \ldots, x_n)$ y $g_\theta(T(x_1, \ldots, x_n))$ 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 $h(x_1, \ldots, x_n) \geq 0$, entonces $\widehat{\theta}$ es el valor de θ donde se alcanza $\max_{\theta \in \Theta} L_{x_1,\ldots,x_n}(\theta)$, es decir, el valor de θ donde se alcanza $\max_{\theta \in \Theta} g_\theta(T(x_1, \ldots, x_n))$.
- Si $h(x_1, \ldots, x_n) \leq 0$, entonces $\widehat{\theta}$ es el valor de θ donde se alcanza $\max_{\theta \in \Theta} L_{x_1,\ldots,x_n}(\theta)$, es decir, el valor de θ donde se alcanza $\min_{\theta \in \Theta} g_\theta(T(x_1, \ldots, x_n))$.

En ambos casos, el EMV de θ, $\widehat{\theta}$, es una función de $g_\theta(T(X_1, \ldots, X_n))$, que a su vez es función de $T(X_1, \ldots, X_n)$. Por tanto, $\widehat{\theta}$ es función de $T(X_1, \ldots, X_n)$. 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 $f_\theta^n(x_1, \ldots, x_n) = h(x_1, \ldots, x_n)$ 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 $(X_1, \ldots, X_n)$ una muestra aleatoria simple de $X \rightsquigarrow U[\theta, \theta+1]$. Calcular el estadístico suficiente, un EMV para θ, y comprobar que no coinciden.*

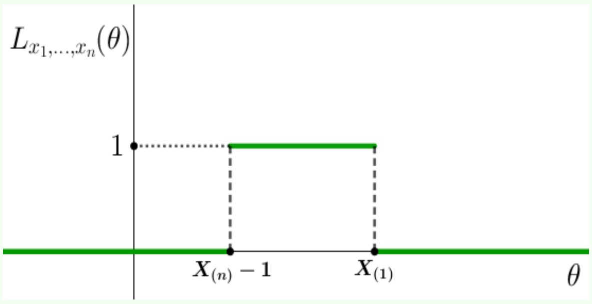
Solución. Para una distribución uniforme continua en el intervalo $[\theta, \theta+1]$, se tiene que su función de densidad es $f_\theta(x) = 1$, $\theta < x < \theta + 1$. Por tanto:

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

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, $L_{x_1,\ldots,x_n}(\theta) = 0$, $\forall \theta \in \Theta$.
- 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, $L_{x_1,\ldots,x_n}(\theta) = 1$.



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 $(X_1, \dots, X_n)$ una m.a.s. de X. Si $T(X_1, \dots, X_n)$ es un estimador eficiente para θ, entonces existe un único EMV y coincide con $T(X_1, \dots, X_n)$.*

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

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

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

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

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

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

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 $Y =$ *nº 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.

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

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 $(x_1,\dots,x_n)$ de una muestra aleatoria simple $(X_1,\dots,X_n)$ 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 $\widehat{\theta}$.

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

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

Como $\dfrac{\partial^2 \ln L_{x_1,\dots,x_n}(\theta)}{\partial\theta^2}=\dfrac{-n}{\theta^2}<0$, 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:

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

Por tanto:

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

Así:

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

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* $(X_1, \ldots, X_n)$ *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

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

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 $L_{x_1,\ldots,x_n}(\theta) = 0$ 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

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

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:

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

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 $\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$, para tener que:

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

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:

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

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

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

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 $\mathrm{E}_\theta[\epsilon] = 0$, 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:

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

TEMA 6: Estimación por intervalos de confianza

- Planteamiento del problema y conceptos básicos.
- Construcción de intervalos.
- Intervalos de confianza para los parámetros de una población normal.
- Intervalos de confianza para los parámetros de dos poblaciones normales.

Planteamiento del problema y conceptos básicos

$(X_1, \ldots, X_n)$ muestra aleatoria simple de $X \rightarrow \{P_\theta;\ \theta \in \Theta\}$

Se trata de usar los datos muestrales para construir un subconjunto del espacio paramétrico para el que podamos afirmar, con un margen de error prefijado, que contiene al verdadero valor del parámetro.

Intervalo de confianza: *Un intervalo aleatorio,* $\Big(I_1(X_1, \ldots, X_n),\ I_2(X_1, \ldots, X_n)\Big)$*, es un intervalo de confianza para* θ *al nivel de confianza* $1 - \alpha$ $(0 < \alpha < 1)$ *si*

$$\forall \theta \in \Theta, \quad P_\theta\Big(I_1(X_1, \ldots, X_n) \leq \theta \leq I_2(X_1, \ldots, X_n)\Big) \geq 1 - \alpha.$$

Interpretación: Sea cual sea el verdadero valor del parámetro, la probabilidad de que el intervalo lo contenga es, al menos, $1 - \alpha$.

Según la definición frecuentista de probabilidad, esto puede interpretarse como que el $(1 - \alpha)100\,\%$ de las veces que se observe una muestra de tamaño n, el intervalo concreto obtenido, $(I_1(x_1, \ldots, x_n),\ I_2(x_1, \ldots, x_n))$, contendrá al verdadero valor del parámetro; por tanto, cada vez que tomamos una muestra, tenemos una *confianza* del $(1 - \alpha)100\,\%$ de que así ocurre.

Intervalo de confianza de menor longitud esperada uniformemente: *Un intervalo* $\Big(I_1(X_1, \ldots, X_n),\ I_2(X_1, \ldots, X_n)\Big)$*, con nivel de confianza* $1-\alpha$*, es de menor longitud esperada uniformemente a dicho nivel si para cualquier otro intervalo* $\Big(I_1'(X_1, \ldots, X_n),\ I_2'(X_1, \ldots, X_n)\Big)$ *al mismo nivel se tiene:*

$$\forall \theta \in \Theta, \quad E_\theta\left[I_2(X_1, \ldots, X_n) - I_1(X_1, \ldots, X_n)\right] \leq E_\theta\left[I_2'(X_1, \ldots, X_n) - I_1'(X_1, \ldots, X_n)\right].$$

Métodos de construcción

$(X_1, \ldots, X_n)$ muestra aleatoria simple de $X \longrightarrow \{P_\theta;\ \theta \in \Theta\}$

Intervalos obtenidos mediante la desigualdad de Chebychev: Si $T(X_1, \ldots, X_n)$ es un estimador insesgado de θ con varianza uniformemente acotada ($E_\theta\left[T(X_1, \ldots, X_n)\right] = \theta$, y $Var_\theta\left[T(X_1, \ldots - X_n)\right] \leq c$, $\forall \theta \in \Theta$), para cualquier $k > 0$ se tiene:

$$P_\theta\Big(T(X_1, \ldots - X_n) - k < \theta < \ T(X_1, \ldots - X_n) + k\Big) \geq 1 - \frac{c}{k^2}, \quad \forall \theta \in \Theta,$$

y $(T(X_1, \ldots, X_n) - k, T(X_1, \ldots, X_n) + k)$ es un intervalo de confianza para θ al nivel de confianza $1 - c/k^2$.

Intervalos obtenidos mediante el método pivotal:

Pivote para un parámetro: Función de la muestra y del parámetro, $T(X_1, \ldots, X_n;\ \theta)$, tal que $\forall \theta \in \Theta,\ T(X_1, \ldots, X_n;\ \theta)$ es una variable aleatoria con distribución independiente de θ.

Descripción del método: Sea $T(X_1, \ldots, X_n;\ \theta)$ un pivote estrictamente monótono en θ:

$i)$ Se buscan dos valores, λ_1 y λ_2 tales que $P_\theta\Big(\lambda_1 < T(X_1, \ldots, X_n;\ \theta) < \lambda_2\Big) \geq 1 - \alpha, \quad \forall \theta \in \Theta.$

$ii)$ Se resuelven en θ las ecuaciones $T(X_1, \ldots, X_n;\ \theta) = \lambda_1$ y $T(X_1, \ldots, X_n;\ \theta) = \lambda_2$, y las soluciones, $\widehat{\theta}_1(X_1, \ldots, X_n)$ y $\widehat{\theta}_2(X_1, \ldots, X_n)$, proporcionan un intervalo de confianza para θ al nivel $1 - \alpha$:

- T creciente $\rightarrow\ P_\theta\left(\widehat{\theta}_1(X_1, \ldots, X_n) < \theta < \widehat{\theta}_2(X_1, \ldots, X_n)\right) \geq 1 - \alpha, \quad \forall \theta \in \Theta$
- T decreciente $\rightarrow\ P_\theta\left(\widehat{\theta}_2(X_1, \ldots, X_n) < \theta < \widehat{\theta}_1(X_1, \ldots, X_n)\right) \geq 1 - \alpha, \quad \forall \theta \in \Theta.$

Determinación de pivotes en distribuciones continuas:

- Si X es de tipo continuo con función de distribución F_θ, la siguiente función es un pivote para θ:
$$T(X_1, \ldots, X_n; \theta) = -2 \sum_{i=1}^{n} \ln F_\theta(X_i) \ \longrightarrow\ \chi^2(2n).$$
- Si $S(X_1, \ldots, X_n)$ es un estadístico con distribución de tipo continuo y F_θ^S es su función de distribución, la siguiente función constituye un pivote para θ:
$$T(X_1, \ldots, X_n; \theta) = F_\theta^S\big(S(X_1, \ldots, X_n)\big) \ \longrightarrow\ U(0,1).$$

En cualquiera de los casos, si los pivotes son monótonos en θ, darán lugar a intervalos de confianza para θ.

A1: Intervalos de confianza para la media de una normal con varianza conocida

$(X_1, \dots, X_n)$ muestra aleatoria simple de $X \longrightarrow \{\mathcal{N}(\mu, \sigma_0^2);\ \mu \in \mathbb{R}\}$

PIVOTE : $T(X_1, \dots, X_n;\ \mu) = \dfrac{\overline{X} - \mu}{\sigma_0/\sqrt{n}} \longrightarrow \mathcal{N}(0,1)$

$$\forall \mu \in \mathbb{R}, \quad \underbrace{P_\mu\left(\lambda_1 < \frac{\overline{X}-\mu}{\sigma_0/\sqrt{n}} < \lambda_2\right)} = P_\mu\left(\overline{X} - \lambda_2 \frac{\sigma_0}{\sqrt{n}} < \mu < \overline{X} - \lambda_1 \frac{\sigma_0}{\sqrt{n}}\right)$$

$$\downarrow$$

$$P(\lambda_1 < Z < \lambda_2) = F_Z(\lambda_2) - F_Z(\lambda_1) \quad \left(Z \to \mathcal{N}(0,1), \quad F_Z(z) = P(Z \le z)\right)$$

$$\Downarrow$$

$$\boxed{F_Z(\lambda_2) - F_Z(\lambda_1) = 1 - \alpha \longrightarrow \left(\overline{X} - \lambda_2 \frac{\sigma_0}{\sqrt{n}},\ \ \overline{X} - \lambda_1 \frac{\sigma_0}{\sqrt{n}}\right) \text{ IC para } \mu \text{ a nivel } 1-\alpha}$$

Se busca, entre estos intervalos, el que minimice uniformemente ($\forall \mu \in \mathbb{R}$) *la longitud media:*

$$E_\mu\left[\left(\overline{X} - \lambda_1 \frac{\sigma_0}{\sqrt{n}}\right) - \left(\overline{X} - \lambda_2 \frac{\sigma_0}{\sqrt{n}}\right)\right] = (\lambda_2 - \lambda_1)\frac{\sigma_0}{\sqrt{n}}.$$

Problema A1: *Minimizar la función* $\lambda_2 - \lambda_1$ *con* λ_1, λ_2 *sujetos a* $F_Z(\lambda_2) - F_Z(\lambda_1) = 1 - \alpha$.

Solución: Notando $f_Z = F_Z'$ a la función de densidad de la $\mathcal{N}(0,1)$:

$$F(\lambda_1, \lambda_2, \lambda) = \lambda_2 - \lambda_1 + \lambda[F_Z(\lambda_2) - F_Z(\lambda_1) - (1-\alpha)]$$

$$\left.\begin{array}{l} \dfrac{\partial F(\lambda_1, \lambda_2, \lambda)}{\partial \lambda_1} = -1 - \lambda f_Z(\lambda_1) = 0 \\ \dfrac{\partial F(\lambda_1, \lambda_2, \lambda)}{\partial \lambda_2} = 1 + \lambda f_Z(\lambda_2) = 0 \end{array}\right\} \Rightarrow f_Z(\lambda_1) = f_Z(\lambda_2) \Rightarrow \lambda_1 = \pm\lambda_2 \ \ (f_Z \text{ es simétrica}).$$

Dado que $\lambda_1 = \lambda_2$ no verifica $F_Z(\lambda_2) - F_Z(\lambda_1) = 1 - \alpha$, ha de ser $\lambda_1 = -\lambda_2$, y la restricción $F_Z(\lambda_2) - F_Z(-\lambda_2) = 1 - \alpha$ conduce a $\lambda_2 = z_{\alpha/2}$, siendo $P(Z > z_{\alpha/2}) = \alpha/2$.

I.C. para μ de menor longitud media uniformemente, a nivel de confianza $1-\alpha$

$$\left(\overline{X} - z_{\alpha/2} \frac{\sigma_0}{\sqrt{n}},\ \ \overline{X} + z_{\alpha/2} \frac{\sigma_0}{\sqrt{n}}\right)$$

INTERVALOS UNILATERALES

$$\left\{\begin{array}{l} \lambda_2 = +\infty \Rightarrow \lambda_1 = -z_\alpha \longrightarrow \left(-\infty,\ \overline{X} + z_\alpha \dfrac{\sigma_0}{\sqrt{n}}\right). \\[2ex] \lambda_1 = -\infty \Rightarrow \lambda_2 = z_\alpha \longrightarrow \left(\overline{X} - z_\alpha \dfrac{\sigma_0}{\sqrt{n}},\ +\infty\right). \end{array}\right.$$

A2: Intervalo de confianza para la media de una normal con varianza desconocida

$(X_1, \dots, X_n)$ muestra aleatoria simple de $X \longrightarrow \{\mathcal{N}(\mu, \sigma^2);\ \mu \in \mathbb{R},\ \sigma^2 \in \mathbb{R}^+\}$

PIVOTE : $T(X_1, \dots, X_n;\ \mu) = \dfrac{\overline{X} - \mu}{S/\sqrt{n}} \longrightarrow t(n-1)$

$$\forall \mu \in \mathbb{R},\ \sigma^2 \in \mathbb{R}^+,\ \underbrace{P_{\mu,\sigma^2}\left(\lambda_1 < \frac{\overline{X} - \mu}{S/\sqrt{n}} < \lambda_2\right)} = P_{\mu,\sigma^2}\left(\overline{X} - \lambda_2 \frac{S}{\sqrt{n}} < \mu < \overline{X} - \lambda_1 \frac{S}{\sqrt{n}}\right)$$

$$\downarrow$$

$$P(\lambda_1 < T < \lambda_2) = F_T(\lambda_2) - F_T(\lambda_1) \quad \big(T \to t(n-1),\ F_T(t) = P(T \leq t)\big)$$

$$\Downarrow$$

$$\boxed{F_T(\lambda_2) - F_T(\lambda_1) = 1 - \alpha \ \to \ \left(\overline{X} - \lambda_2 \frac{S}{\sqrt{n}},\ \ \overline{X} - \lambda_1 \frac{S}{\sqrt{n}}\right) \ \text{IC para } \mu \text{ a nivel } 1 - \alpha}$$

Se busca entre ellos el que minimice uniformemente $(\forall \mu \in \mathbb{R},\ \forall \sigma^2 \in \mathbb{R}^+)$ *la longitud media:*

$$E_{\mu,\sigma^2}\left[\left(\overline{X} - \lambda_1 \frac{S}{\sqrt{n}}\right) - \left(\overline{X} - \lambda_2 \frac{S}{\sqrt{n}}\right)\right] = (\lambda_2 - \lambda_1) E_{\mu,\sigma^2}\left[S/\sqrt{n}\right].$$

Problema A2: *Minimizar la función* $\lambda_2 - \lambda_1$ *con* λ_1, λ_2 *sujetos a* $F_T(\lambda_2) - F_T(\lambda_1) = 1 - \alpha$.

Solución: Es el Problema A1 con la restricción dada por la función de distribución de una $t(n-1)$; por tanto, λ_1 y λ_2 deben verificar $f_T(\lambda_1) = f_T(\lambda_2)$, donde f_T es la función de densidad de la distribución $t(n-1)$; como esta función es también simétrica, razonando como A1 se tiene $\lambda_1 = -\lambda_2$ y $\lambda_2 = t_{n-1;\ \alpha/2}$, siendo $P(T > t_{n-1;\ \alpha/2}) = \alpha/2$.

I.C. para μ de menor longitud media uniformemente, a nivel de confianza $1 - \alpha$

$$\left(\overline{X} - t_{n-1;\ \alpha/2} \frac{S}{\sqrt{n}},\ \ \overline{X} + t_{n-1;\ \alpha/2} \frac{S}{\sqrt{n}}\right)$$

INTERVALOS UNILATERALES
$$\begin{cases} \lambda_2 = +\infty \Rightarrow \lambda_1 = -t_{n-1;\ \alpha} \ \longrightarrow \ \left(-\infty,\ \overline{X} + t_{n-1;\ \alpha} \dfrac{S}{\sqrt{n}}\right). \\[2ex] \lambda_1 = -\infty \Rightarrow \lambda_2 = t_{n-1;\ \alpha} \ \longrightarrow \ \left(\overline{X} - t_{n-1;\ \alpha} \dfrac{S}{\sqrt{n}},\ +\infty\right). \end{cases}$$

B1: Intervalo de confianza para la varianza de una normal con media conocida

$(X_1, \ldots, X_n)$ muestra aleatoria simple de $X \longrightarrow \{\mathcal{N}(\mu_0, \sigma^2);\ \sigma^2 \in \mathbb{R}^+\}$

$$\text{PIVOTE}: \ T(X_1, \ldots, X_n;\ \sigma^2) = \frac{\sum_{i=1}^{n} (X_i - \mu_0)^2}{\sigma^2} \longrightarrow \chi^2(n)$$

$$\forall \sigma^2 \in \mathbb{R}^+, \quad \underbrace{P_{\sigma^2}\left(\lambda_1 < \frac{\sum_{i=1}^{n} (X_i - \mu_0)^2}{\sigma^2} < \lambda_2\right)} = P_{\sigma^2}\left(\frac{\sum_{i=1}^{n} (X_i - \mu_0)^2}{\lambda_2} < \sigma^2 < \frac{\sum_{i=1}^{n} (X_i - \mu_0)^2}{\lambda_1}\right)$$

$$P(\lambda_1 < T < \lambda_2) = F_T(\lambda_2) - F_T(\lambda_1) \quad \left(T \to \chi^2(n),\ F_T(t) = P(T \le t)\right)$$

$$\Downarrow$$

$$\boxed{F_T(\lambda_2) - F_T(\lambda_1) = 1 - \alpha \ \to \left(\frac{\sum_{i=1}^{n} (X_i - \mu_0)^2}{\lambda_2}, \ \frac{\sum_{i=1}^{n} (X_i - \mu_0)^2}{\lambda_1}\right) \ \text{IC para } \sigma^2 \text{ a nivel } 1 - \alpha}$$

$$\textit{Longitud media} \ \to E_{\sigma^2}\left[\sum_{i=1}^{n} (X_i - \mu_0)^2/\lambda_1 - \sum_{i=1}^{n} (X_i - \mu_0)^2/\lambda_2\right] = \left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2}\right) E_{\sigma^2}\left[\sum_{i=1}^{n} (X_i - \mu_0)^2\right].$$

Problema B1: *Minimizar la función* $\frac{1}{\lambda_1} - \frac{1}{\lambda_2}$ *con* λ_1, λ_2 *sujetos a* $F_T(\lambda_2) - F_T(\lambda_1) = 1 - \alpha$.

$$F(\lambda_1, \lambda_2, \lambda) = \tfrac{1}{\lambda_1} - \tfrac{1}{\lambda_2} + \lambda[F_T(\lambda_2) - F_T(\lambda_1) - (1-\alpha)] \ \to \left.\begin{array}{l} \dfrac{\partial F}{\partial \lambda_1} = -\dfrac{1}{\lambda_1^2} - \lambda f_T(\lambda_1) = 0 \\[2ex] \dfrac{\partial F}{\partial \lambda_2} = \dfrac{1}{\lambda_2^2} + \lambda f_T(\lambda_2) = 0 \end{array}\right\} \Rightarrow \frac{f_T(\lambda_1)}{f_T(\lambda_2)} = \frac{\lambda_2^2}{\lambda_1^2}.$$

En este caso, la asimetría de la distribución hace que los valores λ_1 y λ_2 no sean los de colas iguales. Sin embargo, la diferencia no es suficientemente importante (sobre todo para grandes muestras), y en la práctica se usa el intervalo de colas iguales: $\lambda_1 = \chi^2_{n;\ 1-\alpha/2}$ y $\lambda_2 = \chi^2_{n;\ \alpha/2}$, donde $P\left(\chi^2(n) > \chi^2_{n;\alpha/2}\right) = \alpha/2$.

I.C. para σ^2 *de menor longitud media uniformemente (≈), a nivel de confianza* $1 - \alpha$

$$\left(\frac{\sum_{i=1}^{n} (X_i - \mu_0)^2}{\chi^2_{n;\ \alpha/2}}, \ \frac{\sum_{i=1}^{n} (X_i - \mu_0)^2}{\chi^2_{n;\ 1-\alpha/2}}\right)$$

$$\textbf{INTERVALOS UNILATERALES}\begin{cases} \lambda_2 = +\infty \Rightarrow \lambda_1 = \chi^2_{n;\ 1-\alpha} \longrightarrow \left(0,\quad \dfrac{\sum\limits_{i=1}^{n}(X_i-\mu_0)^2}{\chi^2_{n;\ 1-\alpha}}\right) \\ \\ \lambda_1 = 0 \Rightarrow \lambda_2 = \chi^2_{n;\ \alpha} \longrightarrow \left(\dfrac{\sum\limits_{i=1}^{n}(X_i-\mu_0)^2}{\chi^2_{n;\ \alpha}},\quad +\infty\right). \end{cases}$$

B2: Intervalo de confianza para la varianza de una normal con media desconocida

$(X_1, \ldots, X_n)$ muestra aleatoria simple de $X \longrightarrow \{\mathcal{N}(\mu, \sigma^2);\ \mu \in \mathbb{R},\ \sigma^2 \in \mathbb{R}^+\}$

$$\text{PIVOTE}: \ T(X_1, \ldots, X_n;\ \sigma^2) = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \longrightarrow \chi^2(n-1)$$

$$\forall \mu \in \mathbb{R},\ \sigma^2 \in \mathbb{R}^+,\ \ \underbrace{P_{\mu,\sigma^2}\left(\lambda_1 < \frac{(n-1)S^2}{\sigma^2} < \lambda_2\right)} = P_{\mu,\sigma^2}\left(\frac{(n-1)S^2}{\lambda_2} < \sigma^2 < \frac{(n-1)S^2}{\lambda_1}\right)$$

$$\downarrow$$

$$P\left(\lambda_1 < T < \lambda_2\right) = F_T(\lambda_2) - F_T(\lambda_1) \quad \left(T \to \chi^2(n-1),\ F_T(t) = P(T \leq t)\right)$$

Se razona como en B1, sustituyendo $\sum_{i=1}^{n}(X_i - \mu_0)^2$ por $(n-1)S^2$ y la distribución $\chi^2(n)$ por $\chi^2(n-1)$.

I.C. para σ^2 de menor longitud media uniformemente (≈), a nivel de confianza $1-\alpha$

$$\left(\frac{(n-1)S^2}{\chi^2_{n-1;\ \alpha/2}},\ \ \frac{(n-1)S^2}{\chi^2_{n-1;\ 1-\alpha/2}}\right)$$

INTERVALOS UNILATERALES

$$\begin{cases} \lambda_2 = +\infty \Rightarrow \lambda_1 = \chi^2_{n-1;\ 1-\alpha} \longrightarrow \left(0,\ \ \dfrac{(n-1)S^2}{\chi^2_{n-1;\ 1-\alpha}}\right) \\[2ex] \lambda_1 = 0 \Rightarrow \lambda_2 = \chi^2_{n-1;\ \alpha} \longrightarrow \left(\dfrac{(n-1)S^2}{\chi^2_{n-1;\ \alpha}},\ \ +\infty\right). \end{cases}$$

C1: Intervalo de confianza para la diferencia de medias de dos normales (varianzas conocidas)

$(X_1, \ldots, X_{n_1})$ muestra aleatoria simple de $X \longrightarrow \{\mathcal{N}(\mu_1, \sigma_1^2);\ \mu_1 \in \mathbb{R}\}$

$(Y_1, \ldots, Y_{n_2})$ muestra aleatoria simple de $Y \longrightarrow \{\mathcal{N}(\mu_2, \sigma_2^2);\ \mu_2 \in \mathbb{R}\}$

$(X_1, \ldots, X_{n_1}),\ (Y_1, \ldots, Y_{n_2})$ independientes

$$\text{PIVOTE}: \ T(X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2};\ \mu_1 - \mu_2) = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \longrightarrow \mathcal{N}(0,1)$$

$$\forall \mu_1, \mu_2 \in \mathbb{R},\ \underbrace{P_{\mu_1,\mu_2}(\lambda_1 < T < \lambda_2)} = P_{\mu_1,\mu_2}\left(\overline{X} - \overline{Y} - \lambda_2\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < \overline{X} - \overline{Y} - \lambda_1\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

$$\downarrow$$

$$P(\lambda_1 < Z < \lambda_2) = F_Z(\lambda_2) - F_Z(\lambda_1) \quad \big(Z \to \mathcal{N}(0,1),\ \ F_Z(z) = P(Z \leq z)\big)$$

$$\Downarrow$$

$$\boxed{F_Z(\lambda_2) - F_Z(\lambda_1) = 1 - \alpha \longrightarrow \left(\overline{X} - \overline{Y} - \lambda_2\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\ \ \overline{X} - \overline{Y} - \lambda_1\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right) \ \text{IC para } \mu_1 - \mu_2 \text{ a nivel } 1 - \alpha}$$

El problema de buscar, entre estos intervalos, el que minimice uniformemente la longitud media es exactamente el Problema A1, y la solución es, por tanto, $\lambda_1 = -z_{\alpha/2},\ \lambda_2 = z_{\alpha/2}$.

I.C. para $\mu_1 - \mu_2$ de menor longitud media uniformemente, a nivel de confianza $1 - \alpha$

$$\left(\overline{X} - \overline{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\quad \overline{X} - \overline{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

INTERVALOS UNILATERALES
$$\begin{cases} \lambda_2 = +\infty \longrightarrow \left(-\infty,\ \ \overline{X} - \overline{Y} + z_\alpha\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right). \\[2ex] \lambda_1 = -\infty \longrightarrow \left(\overline{X} - \overline{Y} - z_\alpha\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},\ \ +\infty\right). \end{cases}$$

C2: Intervalo de confianza para la diferencia de medias de dos normales (varianzas desconocidas pero iguales)

$(X_1, \ldots, X_{n_1})$ muestra aleatoria simple de $X \longrightarrow \{\mathcal{N}(\mu_1, \sigma^2);\ \mu_1 \in \mathbb{R},\ \sigma^2 \in \mathbb{R}^+\}$

$(Y_1, \ldots, Y_{n_2})$ muestra aleatoria simple de $Y \longrightarrow \{\mathcal{N}(\mu_2, \sigma^2);\ \mu_2 \in \mathbb{R},\ \sigma^2 \in \mathbb{R}^+\}$

$(X_1, \ldots, X_{n_1}),\ (Y_1, \ldots, Y_{n_2})$ independientes

$$\text{PIVOTE}:\ T(X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2};\ \mu_1 - \mu_2) = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \longrightarrow t(n_1 + n_2 - 2)$$

$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$

$$\forall \mu_1, \mu_2 \in \mathbb{R},\ \sigma^2 \in \mathbb{R}^+,\ \underbrace{P_{\mu_1,\mu_2,\sigma^2}(\lambda_1 < T < \lambda_2)} = P_{\mu_1,\mu_2,\sigma^2}\left(\overline{X} - \overline{Y} - \lambda_2 S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < \overline{X} - \overline{Y} - \lambda_1 S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

$$\downarrow$$

$$P(\lambda_1 < T < \lambda_2) = F_T(\lambda_2) - F_T(\lambda_1) \quad \Big(T \to t(n_1 + n_2 - 2),\ F_T(t) = P(T \leq t)\Big)$$

$$\Downarrow$$

$$\boxed{F_T(\lambda_2) - F_T(\lambda_1) = 1 - \alpha \ \rightarrow\ \left(\overline{X} - \overline{Y} - \lambda_2 S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}},\ \ \overline{X} - \overline{Y} - \lambda_1 S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)\ \text{ IC para } \mu_1 - \mu_2 \text{ a nivel } 1 - \alpha}$$

El problema de buscar, entre estos intervalos, el que minimice uniformemente la longitud media es exactamente el Problema A2, y la solución es $\lambda_2 = t_{n_1+n_2-2;\ \alpha/2},\ \lambda_1 = -\lambda_2$.

I.C. para $\mu_1 - \mu_2$ de menor longitud media uniformemente, a nivel de confianza $1 - \alpha$

$$\left(\overline{X} - \overline{Y} - t_{n_1+n_2-2;\ \alpha/2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}},\quad \overline{X} - \overline{Y} + t_{n_1+n_2-2;\ \alpha/2} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

$$\textbf{INTERVALOS UNILATERALES} \begin{cases} \lambda_2 = +\infty \longrightarrow \left(-\infty,\ \ \overline{X} - \overline{Y} + t_{n_1+n_n-2;\ \alpha}\ S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right). \\[2ex] \lambda_1 = -\infty \longrightarrow \left(\overline{X} - \overline{Y} - t_{n_1+n_2-2;\ \alpha}\ S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}},\ \ +\infty\right). \end{cases}$$

D1: Intervalo de confianza para el cociente de varianzas de dos normales (medias conocidas)

$(X_1,\dots,X_{n_1})$ muestra aleatoria simple de $X \longrightarrow \{\mathcal{N}(\mu_1,\sigma_1^2);\ \ \sigma_1^2\in\mathbb{R}^+\}$

$(Y_1,\dots,Y_{n_2})$ muestra aleatoria simple de $Y \longrightarrow \{\mathcal{N}(\mu_2,\sigma_2^2);\ \sigma_2^2\in\mathbb{R}^+\}$

$(X_1,\dots,X_{n_1}),\ (Y_1,\dots,Y_{n_2})$ independientes

$$\text{PIVOTE}:\ T(X_1,\dots,X_{n_1},Y_1,\dots,Y_{n_2};\ \sigma_1^2/\sigma_2^2)=\frac{\sum\limits_{i=1}^{n_2}(Y_i-\mu_2)^2/n_2\sigma_2^2}{\sum\limits_{i=1}^{n_1}(X_i-\mu_1)^2/n_1\sigma_1^2}\longrightarrow F(n_2,n_1)$$

$$\forall\sigma_1^2,\sigma_2^2\in\mathbb{R}^+,\ \underbrace{P_{\sigma_1^2,\sigma_2^2}\left(\lambda_1<\frac{\sum\limits_{i=1}^{n_2}(Y_i-\mu_2)^2/n_2\sigma_2^2}{\sum\limits_{i=1}^{n_1}(X_i-\mu_1)^2/n_1\sigma_1^2}<\lambda_2\right)}_{\downarrow}=P_{\sigma_1^2,\sigma_2^2}\left(\lambda_1\frac{\sum\limits_{i=1}^{n_1}(X_i-\mu_1)^2/n_1}{\sum\limits_{i=1}^{n_2}(Y_i-\mu_2)^2/n_2}<\frac{\sigma_1^2}{\sigma_2^2}<\lambda_2\frac{\sum\limits_{i=1}^{n_1}(X_i-\mu_1)^2/n_1}{\sum\limits_{i=1}^{n_2}(Y_i-\mu_2)^2/n_2}\right)$$

$$P(\lambda_1<T<\lambda_2)=F_T(\lambda_2)-F_T(\lambda_1)\quad \Big(T\to F(n_2,n_1),\ F_T(t)=P(T\le t)\Big)$$

$$\Downarrow$$

$$\boxed{F_T(\lambda_2)-F_T(\lambda_1)=1-\alpha\ \to\ \left(\lambda_1\frac{\sum\limits_{i=1}^{n_1}(X_i-\mu_1)^2/n_1}{\sum\limits_{i=1}^{n_2}(Y_i-\mu_2)^2/n_2},\ \ \lambda_2\frac{\sum\limits_{i=1}^{n_1}(X_i-\mu_1)^2/n_1}{\sum\limits_{i=1}^{n_2}(Y_i-\mu_2)^2/n_2}\right)\ \text{IC para } \tfrac{\sigma_1^2}{\sigma_2^2} \text{ a nivel } 1-\alpha}$$

Se busca, entre todos estos intervalos, el que minimice uniformemente la longitud media, lo que se reduce a *minimizar la función* $\lambda_2-\lambda_1$ *con* λ_1,λ_2 *sujetos a* $F_T(\lambda_2)-F_T(\lambda_1)=1-\alpha$.

Es de nuevo el Problema A1, cuya solución es $f_T(\lambda_2)=f_T(\lambda_1)$, siendo ahora f_T la función de densidad de la $F(n_2,n_1)$. De nuevo, en este caso, la asimetría de la distribución hace que el intervalo de mínima longitud esperada uniformemente no sea el de colas iguales, aunque en la práctica es este último el que suele utilizarse.

I.C. para $\frac{\sigma_1^2}{\sigma_2^2}$ *de menor longitud media uniformemente* (≈), *a nivel de confianza* $1-\alpha$

$$\left(F_{n_2,n_1;\ 1-\alpha/2}\frac{\sum\limits_{i=1}^{n_1}(X_i-\mu_1)^2/n_1}{\sum\limits_{i=1}^{n_2}(Y_i-\mu_2)^2/n_2},\ \ F_{n_2,n_1;\ \alpha/2}\frac{\sum\limits_{i=1}^{n_1}(X_i-\mu_1)^2/n_1}{\sum\limits_{i=1}^{n_2}(Y_i-\mu_2)^2/n_2}\right)$$

INTERVALOS UNILATERALES
$$\begin{cases} \lambda_2 = +\infty \longrightarrow \left(F_{n_2,n_1;\,1-\alpha} \dfrac{\sum\limits_{i=1}^{n_1}(X_i-\mu_1)^2/n_1}{\sum\limits_{i=1}^{n_2}(Y_i-\mu_2)^2/n_2},\ +\infty \right). \\[2em] \lambda_1 = 0 \longrightarrow \left(0,\quad F_{n_2,n_1;\,\alpha} \dfrac{\sum\limits_{i=1}^{n_1}(X_i-\mu_1)^2/n_1}{\sum\limits_{i=1}^{n_2}(Y_i-\mu_2)^2/n_2} \right). \end{cases}$$

D2: Intervalo de confianza para el cociente de varianzas de dos normales (medias desconocidas)

$(X_1, \ldots, X_{n_1})$ muestra aleatoria simple de $X \longrightarrow \{\mathcal{N}(\mu_1, \sigma_1^2);\ \mu_1 \in \mathbb{R},\ \sigma_1^2 \in \mathbb{R}^+\}$

$(Y_1, \ldots, Y_{n_2})$ muestra aleatoria simple de $Y \longrightarrow \{\mathcal{N}(\mu_2, \sigma_2^2);\ \mu_2 \in \mathbb{R},\ \sigma_2^2 \in \mathbb{R}^+\}$

$(X_1, \ldots, X_{n_1}),\ (Y_1, \ldots, Y_{n_2})$ independientes

$$\text{PIVOTE}: \ T(X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2};\ \sigma_1^2/\sigma_2^2) = \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} \longrightarrow F(n_2-1, n_1-1)$$

$$\forall \mu_1, \mu_2 \in \mathbb{R},\ \sigma_1^2, \sigma_2^2 \in \mathbb{R}^+,\ \ \underbrace{P_{\mu_1,\mu_2,\sigma_1^2,\sigma_2^2}\left(\lambda_1 < \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} < \lambda_2\right)} = P_{\mu_1,\mu_2,\sigma_1^2,\sigma_2^2}\left(\lambda_1 \frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \lambda_2 \frac{S_1^2}{S_2^2}\right)$$

$$\downarrow$$

$$P(\lambda_1 < T < \lambda_2) = F_T(\lambda_2) - F_T(\lambda_1) \quad \left(T \to F(n_2-1, n_1-1),\ F_T(t) = P(T \leq t)\right)$$

$$\Downarrow$$

$$\boxed{F_T(\lambda_2) - F_T(\lambda_1) = 1-\alpha \ \to \ \left(\lambda_1 \frac{S_1^2}{S_2^2},\ \ \lambda_2 \frac{S_1^2}{S_2^2}\right) \ \text{IC para } \frac{\sigma_1^2}{\sigma_2^2} \text{ a nivel } 1-\alpha}$$

El problema de minimizar la longitud de los intervalos se resuelve como en D1, tomando el intervalo de colas iguales.

I.C. para $\frac{\sigma_1^2}{\sigma_2^2}$ de menor longitud media uniformemente (≈), a nivel de confianza $1-\alpha$

$$\left(F_{n_2-1,n_1-1;\ 1-\alpha/2}\ \frac{S_1^2}{S_2^2},\ \ F_{n_2-1,n_1-1;\ \alpha/2}\ \frac{S_1^2}{S_2^2}\right)$$

INTERVALOS UNILATERALES

$$\begin{cases} \lambda_2 = +\infty, \ \longrightarrow \ \left(F_{n_2-1,n_1-1;\ 1-\alpha}\ \frac{S_1^2}{S_2^2},\ +\infty\right). \\[2ex] \lambda_1 = 0 \ \longrightarrow \ \left(0,\ F_{n_2-1,n_1-1;\ \alpha}\ \frac{S_1^2}{S_2^2}\right). \end{cases}$$