TEMA 2:
ESTIMACIÓN PUNTUAL

Estimación I

Grado en Estadística Aplicada
Curso 2019-2020

Recordemos que nos encontramos en el contexto de la inferencia paramétrica, por lo que asumimos que la v.a. población X se distribuye según una **familia paramétrica** de distribuciones conocida $\mathcal{F}$, pero cuyo parámetro (o vector de parámetros) θ es **desconocido**: $\mathcal{F} = \{f_{X,\theta}(x), \theta \in \Theta\}$

DEFINICIÓN: ESTIMACIÓN PUNTUAL

La estimación puntual tiene como objetivo **proporcionar un valor "aproximado"** de θ (denominado estimación) basado en una muestra.

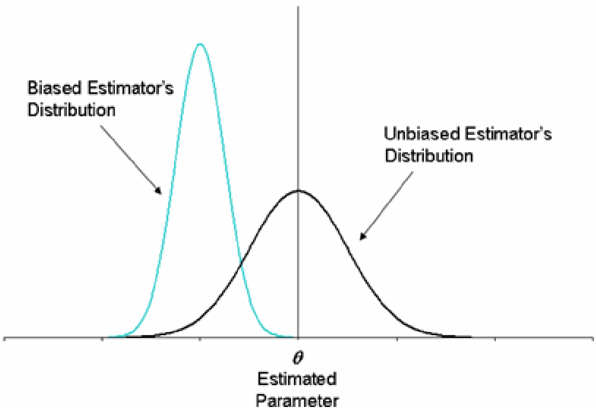
Para ello, haremos uso de **estimadores** (funciones de la muestra que no dependen del valor del parámetro), lo que nos permitirá obtener un valor concreto $\hat{\theta}$ como pronóstico de θ una vez obtenida una **realización de la muestra**.

Dado lo poco exigente que es la definición de estimador, para un problema de estimación concreto, existirán **muchos posibles estimadores** (unos con mejores propiedades que otros). Por ello, la teoría de la estimación puntual se dirige a determinar **criterios de estimación razonables** y buscar, para cada uno de ellos, cuál sería el **estimador óptimo**.

Consideremos una muestra X_1, X_2 de una distribución teórica $B(1, p)$, con $\Theta = [0, 1]$. Se consideran como posibles estimadores para p: $T_1 = \bar{X}$ y $T_2 = \frac{1}{3}(X_1 + 2X_2)$. ¿Cuál de los estimadores será mejor?

DEFINICIÓN: ESTIMADORES INSESGADOS

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim f_{X,\theta}$, con θ desconocido. Se dice que un estimador $\hat{\theta} = T(X_1, \ldots, X_n)$ es **insesgado** si $E[\hat{\theta}] = \theta$.



¿Es siempre preferible un estimador insesgado a uno sesgado?

Recordemos que, si $X_1, \ldots, X_n$ es una m.a.s. de X, tal que $E[X] = \mu$ y $Var[X] = \sigma^2$, entonces:

$$E[\bar{X}] = \mu \qquad E[S^2] = \frac{n-1}{n}\sigma^2 \qquad E[S_c^2] = \sigma^2$$

Comprueba si la media muestral es un estimador insesgado para θ en una m.a.s. de $X \sim B(m, \theta)$ y $X \sim P(\theta)$.

Ejercicio 1

Comprueba si la media muestral es un estimador insesgado para θ en una m.a.s. de $X \sim U(0, \theta)$ y $X \sim \Gamma(\theta, \lambda_0)$, con λ_0 conocido.

Como hemos visto, un estadístico insesgado no siempre es preferible a uno sesgado, pues si, en media, toma valores cercanos al parámetro, pero dichos valores son muy variados, las estimaciones pueden no ser muy buenas.

Definición: Error cuadrático medio de un estimador

Sea $X_1, \dots, X_n$ una m.a.s. de $X \sim f_{X,\theta}$, con θ desconocido y $\hat{\theta}$ un estimador para dicho parámetro. Llamamos **error cuadrático medio** de $\hat{\theta}$ a:

$$ECM(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

El ECM mide como de **cerca del parámetro** se encuentran las estimaciones, en media. Por lo tanto, un estimador será mejor, cuanto **menor** sea su ECM.

$$ECM(\hat{\theta}) = E[(\hat{\theta} - E[\hat{\theta}])^2] + (E[\hat{\theta}] - \theta)^2 = Var[\hat{\theta}] + (b_{\hat{\theta}}(\theta))^2,$$

donde $b_{\hat{\theta}}(\theta)$ recibe el nombre de **sesgo del estimador**, que es igual a 0 cuando el estimador es insesgado. Por lo tanto, si $\hat{\theta}$ es insesgado,

$$ECM(\hat{\theta}) = Var[\hat{\theta}]$$

Consideremos una m.a.s. de tamaño n de una distribución $N(\mu, \sigma)$. Determina si es preferible utilizar la media o la mediana muestral como estimador de μ, suponiendo que σ es conocido. Dado que necesitarás información sobre la mediana muestral M, ten en cuenta que $M \sim N\left(\mu, \frac{\sigma\sqrt{2\pi}}{2\sqrt{n}}\right)$.

EJERCICIO 2

Consideremos una m.a.s. $X_1, \ldots, X_n$ de una distribución $U(0, \theta)$. Determina si es preferible utilizar el estadístico $T = \max\{X_1, \ldots, X_n\}$ o el estadístico $U = 2\bar{X}$ para estimar θ.

Generalmente es preferible un estimador $\hat{\theta}_1$ a otro $\hat{\theta}_2$ si $ECM(\hat{\theta}_1) < ECM(\hat{\theta}_2)$, independientemente de si son sesgados o no. No obstante, en la práctica, se suele restringir la búsqueda de estimadores al subconjunto de aquellos que sí lo son.

DEFINICIÓN: ESTIMADORES EFICIENTES

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim f_{X,\theta}$, con θ desconocido y $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores **insesgados** para dicho parámetro. Diremos que $\hat{\theta}_1$ es **más eficiente** que $\hat{\theta}_2$ si $Var[\hat{\theta}_1] < Var[\hat{\theta}_2]$.

Dado que estamos asumiendo que los dos estimadores son insesgados, que $\hat{\theta}_1$ sea más eficiente que $\hat{\theta}_2$ implica que $ECM(\hat{\theta}_1) < ECM(\hat{\theta}_2)$ y, por lo tanto, $\hat{\theta}_1$ es **preferible** a $\hat{\theta}_2$.

Consideremos una m.a.s. de tamaño n de una distribución $N(\mu, \sigma)$. De cara a estimar μ, ¿qué es más eficiente: la media o la mediana muestral?

Si, dada una población $X \sim f_{X,\theta}$ y un parámetro desconocido θ, existiera, de entre todos los **estimadores insesgados**, uno que tuviera **menor varianza** que el resto, este sería el **óptimo**, desde el punto de vista del sesgo y la eficiencia.

En ese sentido, bajo ciertas condiciones es posible determinar la **varianza mínima posible** para los estimadores insesgados de cierto parámetro, por lo que, conocida dicha *cota*, si encontramos un estimador insesgado cuya **varianza coincida** con dicho valor, habremos encontrado el estimador insesgado de mínima varianza.

COTA DE CRAMER-RAO

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim f_{X,\theta}$, con θ desconocido y soporte $S = \{x / f_{X,\theta}(x) > 0\}$. Entonces, si S no depende de θ, se tiene que:

$$Var[\hat{\theta}] \geq \frac{1}{nE\left[\left(\frac{\partial}{\partial\theta} ln(f_{X,\theta})\right)^2\right]}, \quad \forall\ \hat{\theta} \text{ insesgado}$$

Sea $\hat{\theta}$ un estimador insesgado para θ. Diremos que $\hat{\theta}$ es **eficiente** si:

$$Var[\hat{\theta}] = \frac{1}{nE\left[\left(\frac{\partial}{\partial\theta} ln(f_{X,\theta})\right)^2\right]}$$

Sea $X_1, \ldots, X_n$ una m.a.s. de

- $X \sim N(\mu, \sigma)$. Entonces, $S = \{x / f_{X,\theta}(x) > 0\} = \mathbb{R}$ no depende de μ ni de σ, por lo que existirá la cota.
- $X \sim Exp(\lambda)$. Entonces, $S = \{x / f_{X,\theta}(x) > 0\} = \mathbb{R}^+$ no depende de λ, por lo que existirá la cota.
- $X \sim U(0, \theta)$. Entonces, $S = \{x / f_{X,\theta}(x) > 0\} = (0, \theta)$ depende de θ, por lo que no existirá la cota.

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim B(m, p)$, con $0 \leq p \leq 1$ desconocido. Determina si existe la cota de Cramer-Rao y, de ser así, calcúlala. Comprueba si el estimador $\hat{\theta} = \frac{\bar{X}}{m}$, que es insesgado, es también eficiente.

EJERCICIO 3

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim P(\lambda)$, con $\lambda > 0$ desconocido. Determina si existe la cota de Cramer-Rao y, de ser así, calcúlala. Comprueba si el estimador $\hat{\theta} = \bar{X}$, que es insesgado, es también eficiente.

Ejercicio 4

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma_0)$, con $\mu \in \mathbb{R}$ desconocido y σ_0 conocido. Determina si existe la cota de Cramer-Rao y, de ser así, calcúlala. Comprueba si el estimador $\hat{\theta} = \bar{X}$, que es insesgado, es también eficiente.

Ciomo se ha podido comprobar en los ejemplos, a partir de la cota de Cramer-Rao es posible determinar el **estimador insesgado y eficiente**.

Sin embargo, este procedimiento **no es útil en todas las ocasiones** pues, a) existen modelos paramétricos en los que **no hay cota** de Cramer-Rao (su soporte depende de θ, como en el caso $U(0,\theta)$; y b) aún existiendo la cota, existen modelos paramétricos en los que todos los estimadores insesgados tienen una **varianza estrictamente superior** a la cota de Cramer-Rao (como en el caso exponencial, como ya se verá).

Afortunadamente, existen otros procedimientos que nos proporcionarán los **estimadores de mínima varianza** (aunque no sean eficientes). Para ello, necesitamos definir un nuevo concepto: estadístico suficiente.

Definición: estadístico suficiente

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim f_{X,\theta}$, con θ desconocido. Diremos que $T(X_1, \ldots, X_n)$ es un **estadístico suficiente** para estimar θ si contiene **toda la información** de la muestra sobre el parámetro desconocido.

En otras palabras, a efectos de estimar θ, es **igualmente útil** conocer $T(X_1, \ldots, X_n)$ que $X_1, \ldots, X_n$. Es decir, $T(X_1, \ldots, X_n)$ contiene toda la **información esencial** sobre θ desechando aquella que resulta de utilidad.

Por ejemplo, de cara a estimar el parámetro de una Bernouilli no hace falta conocer la secuencia concreta de ceros y unos, sino que nos basta con conocer el número de unos.

Dado que la definición de estadístico suficiente es **poco útil** para determinar tal estadístico, Neyman y Fisher idearon el **criterio de factorización** para establecer, en cada modelo paramétrico, cuál es el estadístico suficiente.

CRITERIO DE FACTORIZACIÓN DE NEYMAN Y FISHER

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim f_{X,\theta}$, con θ desconocido. Entonces, $\hat{\theta} = T(X_1, \ldots, X_n)$ es un **estadístico suficiente** para estimar θ sii la función de **verosimilitud muestral factoriza** de la siguiente manera:

$$f_\theta(x_1, \ldots, x_n) = g(\hat{\theta}, \theta) h(x_1, \ldots, x_n),$$

donde $g(\hat{\theta}, \theta)$ es una función del parámetro desconocido θ y de la muestra **sólo a través de $\hat{\theta}$** y $h(x_1, \ldots, x_n)$ es una función solo de la muestra y, por tanto, **no depende de θ**.

El estadístico suficiente **no es único**, cualquier transformación lineal de un estadístico suficiente, proporciona otro estadístico suficiente. Trabajaremos, por tanto, con el más sencillo, que se conoce como **minimal**.

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim B(m, p)$, con $0 \leq p \leq 1$ desconocido. Determina un estadístico suficiente para estimar p.

Ejercicio 5

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim P(\lambda)$, con $\lambda > 0$ desconocido. Determina un estadístico suficiente para estimar λ.

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim U(0, \theta)$, con $\theta > 0$ desconocido. Determina un estadístico suficiente para estimar θ.

Ejercicio 6

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma)$, con $\mu \in \mathbb{R}$ y $\sigma > 0$. Determina un estadístico suficiente para estimar μ, suponiendo que $\sigma = \sigma_0$ conocido, y otro para estimar σ, suponiendo que $\mu = \mu_0$ es conocido.

Una vez estudiado el concepto de estadístico suficiente, veremos cuál es su relación con los **estadísticos de mínima varianza**.

TEOREMA DE RAO-BLACKWELL

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim f_{X,\theta}$, con θ desconocido y $\hat{\theta}_T = T(X_1, \ldots, X_n)$ un estadístico **suficiente** para estimar θ. Entonces, para cualquier $\hat{\theta}_U = U(X_1, \ldots, X_n)$ estimador **insesgado** de θ, se tiene que:

$$\varphi(\hat{\theta}_T) = E[\hat{\theta}_U | \hat{\theta}_T] = E[U(X_1, \ldots, X_n) | T(X_1, \ldots, X_n)]$$

es también un estimador de θ que satisface:

a. $\varphi(\hat{\theta}_T)$ es **insesgado** para θ.
b. $Var(\varphi(\hat{\theta}_T)) \leq Var(\hat{\theta}_U)$.
c. $Var(\varphi(\hat{\theta}_T)) = Var(\hat{\theta}_U)$ si, y sólo si, $\hat{\theta}_U$ es función de $\hat{\theta}_T$.

El teorema anterior afirma, por tanto, que es posible encontrar un estimador $\varphi(\hat{\theta}_T)$ tal que **ningún otro estimador** podrá tener una **varianza inferior** a la suya.

Dado que, como hemos visto, $\varphi(\hat{\theta}_T)$ es insesgado, se puede afirmar que $\varphi(\hat{\theta}_T)$ es el **estimador insesgado de minima varianza** (también llamado Estimador Centrado de Uniformemente Mínima Varianza (ECUMV)).

El "problema" es que su obtención es engorrosa, por lo que utilizaremos el siguiente **corolario** en su lugar.

TEOREMA DE RAO-BLACKWELL: COROLARIO

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim f_{X,\theta}$, con θ desconocido. Entonces, el estimador insesgado de mínima varianza (**EIMV**) para θ es necesariamenre **función del estadístico suficiente**.

El corolario anterior nos indica, por tanto, que, si encontramos una **función s** tal que, aplicada sobre el estadístico suficiente, de lugar a un estimador $s(\hat{\theta}_T)$ **insesgado** para θ, entonces, dicho estimador será el **insesgado de mínima varianza**, y no existirá, por tanto, un estimador insesgado mejor que él.

Nos preguntamos, por tanto, si esta función existe y si es **única**.

Según el **Teorema de Lheman-Scheffe** (que no vamos a estudiar), para todos los modelos paramétricos con los que vamos a trabajar en esta asignatura existe una **única función insesgada** del estadístico suficiente.

PROCEDIMIENTO PARA ENCONTRAR EL EIMV

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim f_{X,\theta}$, con θ desconocido. Entonces, el **EIMV** para θ se puede obtener de la siguiente manera:

1. Determina el estadístico **suficiente** minimal $\hat{\theta}_T$ para el modelo $X \sim f_{X,\theta}$.
2. Calcula la **esperanza** de $\hat{\theta}_T$.
    1. Si $E[\hat{\theta}_T] = \theta$, entonces $\hat{\theta}_T$ será directamente el EIMV.
    2. Si $E[\hat{\theta}_T] \neq \theta$, $s(\hat{\theta}_T)$ será el EIMV, con s la función que **corrige el sesgo** de $\hat{\theta}_T$.

Los EIMVs no siempre son eficientes pues, como ya hemos visto, la cota de Cramer-Rao puede no existir o no ser alcanzada por ningún estimador.

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim B(m, p)$, con $0 \leq p \leq 1$ desconocido. Determina el EIMV para estimar p.

EJERCICIO 7

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim P(\lambda)$, con $\lambda > 0$ desconocido. Determina el EIMV para estimar λ.

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim U(0, \theta)$, con $\theta > 0$ desconocido. Determina el EIMV para estimar θ.

EJERCICIO 8

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma)$, con $\mu \in \mathbb{R}$ y $\sigma > 0$. Determina el EIMV para estimar μ, suponiendo que $\sigma = \sigma_0$ conocido, y otro para estimar σ, suponiendo que $\mu = \mu_0$ es conocido.

El procedimiento anterior para determinar el EIMV no siempre se puede aplicar pues, en ocasiones, no es posible encontrar una función para corregir el sesgo.

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim \Gamma(a, \lambda_0)$, con $a > 0$ desconocido y λ_0 conocido. Determina el EIMV para estimar a.

Como ya hemos visto, un buen estimador será aquel que esté **próximo** al verdadero valor del parámetro, lo que se traduce en que sea insesgado y de mínima varianza.

No obstante, otra propiedad interesante que debería cumplir un buen estimador es que la calidad de sus pronósticos mejore cuando el **tamaño de muestra aumenta**.

DEFINICIÓN: ESTIMADOR CONSISTENTE

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim f_{X,\theta}$, con θ desconocido. Diremos que el estimador $\hat{\theta} = T(X_1, \ldots, X_n)$ es **consistente** para estimar θ si **converge en probabilidad** hacia θ:

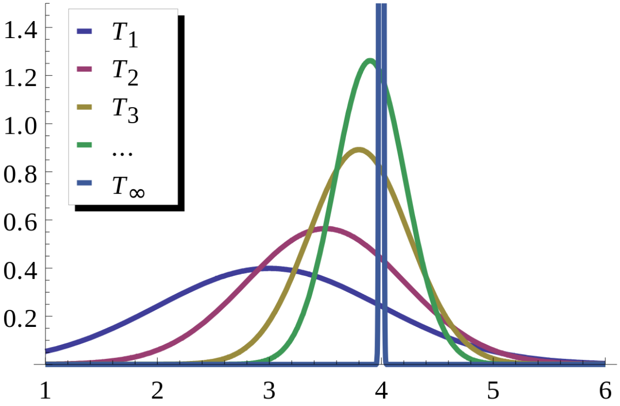
$$\hat{\theta} \xrightarrow{p} \theta \Leftrightarrow \lim_{n\to\infty} P\{\left|\hat{\theta} - \theta\right| < \varepsilon\} = 1; \quad \forall \varepsilon > 0.$$

Al utilizar un estimador consistente el pronóstico que se haga de θ estará **más próximo** a θ a medida que el tamaño muestral **aumente**.

Una forma **alternativa** de verificar si un estimador $\hat{\theta}$ es consistente para θ es:

$$\lim_{n\to\infty} E[\hat{\theta}] = \theta \qquad y \qquad \lim_{n\to\infty} Var[\hat{\theta}] = 0, \qquad \forall \theta \in \Theta$$

Ilustración de estimador consistente



EJEMPLOS DE ESTADÍSTICOS CONSISTENTES

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim f_{X,\theta}$, con θ desconocido.

- Como consecuencia de la Ley débil de los grandes números, si $f_{X,\theta}$ es tal que $E[X] = \theta$, entonces $\hat{\theta} = \bar{X}$ es estimador consistente para estimar θ, pues $\bar{X} \xrightarrow{p} E[X] = \theta$.
  - Por lo tanto, si $X \sim P(\lambda)$, $\bar{X}$ es estimador consistente para estimar λ.
  - Así mismo, si $X \sim N(\mu, \sigma)$, $\bar{X}$ es estimador consistente para estimar μ y S^2 y S_c^2 lo son para estimar σ^2.
- Si $X \sim U(0, \theta)$, $\hat{\theta} = \text{máx}\{X_1, \ldots, X_n\}$ es consistente para estimar θ pues, $E[\hat{\theta}] = \frac{n}{n+1}\theta$ y $Var[\hat{\theta}] = \frac{n\theta^2}{(n+2)(n+1)^2}$ (ejercicio 4, Tema 1).

EJERCICIO 9

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim Exp(\lambda)$, con λ desconocido. Demuestra que $\hat{\theta} = 1/\bar{X}$ es un estimador consistente para estimar λ.

Hasta el momento, hemos estudiado **propiedades importantes** que han de cumplir los estimadores, algunas de las cuales nos han indicado directamente cuál es el **mejor estimador** que podemos utilizar.

No obstante, como hemos visto, esta mecánica **no siempre es útil**, por lo que sería deseable contar con algún **otro método** que nos proporcione buenos estimadores.

DEFINICIÓN: MÉTODO DE LOS MOMENTOS

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim f_{X,\theta}$, con θ desconocido. El **método de los momentos** se basa en estimar los momentos poblacionales mediante los correspondientes momentos muestrales pues, como ya hemos visto, estos últimos son **consistentes** para estimar los primeros.

Así, si deseamos estimar k parámetros $\theta_1, \ldots, \theta_k$, el procedimiento consiste en:

1. Calcular los k primeros **momentos poblaciones**.
2. **Despejar** los parámetros $\theta_1, \ldots, \theta_k$ en función de los anteriores.
3. Obtener los estimadores de $\theta_1, \ldots, \theta_k$ sin más que sustituir en la expresión anterior los momentos poblacionales por los **momentos muestrales**.

Nótese que se pueden utilizar momentos respecto al origen o respecto a la media. Lo habitual es trabajar con los momentos conocidos, para facilitar las cuentas.

- Si solo se desea estimar un parámetro θ, y $E[X] \neq 0$, se obtiene la ecuación
$$\theta = g(E[X])$$
y, por lo tanto,
$$\hat{\theta} = g(\bar{X}).$$
- Si se desean estimar dos parámetros θ_1 y θ_2, obtenemos dos ecuaciones:
$$\left.\begin{array}{l} \theta_1 = g_1(E[X], Var[X]) \\ \theta_2 = g_2(E[X], Var[X]) \end{array}\right\} \quad \text{por lo que:} \quad \begin{array}{l} \hat{\theta}_1 = g_1(\bar{X}, S^2) \\ \hat{\theta}_2 = g_2(\bar{X}, S^2) \end{array}$$

PROPIEDADES DE LOS ESTIMADORES “DE LOS MOMENTOS”

- Es un método muy **sencillo**, aunque no se obtienen estimadores con demasiadas propiedades.
- Los estimadores son **consistentes**, pues se obtienen como función de estimadores consistentes.
- Si los parámetros de la distribución poblacional coinciden con alguno de los k primeros **momentos poblaciones respecto al origen**, entonces los estimadores son **insesgados**.
- Si los parámetros de la distribución poblacional coinciden con alguno de los k primeros **momentos poblaciones respecto a la media**, entonces los estimadores son **asintóticamente insesgados**.

Calcula el estimador de los momentos para una m.a.s. $X_1, \ldots, X_n$ de

- a) $X \sim B(m, p)$, con p desconocido.
- b) $X \sim U(0, \theta)$, con θ desconocido.
- c) $X \sim N(\mu, \sigma)$, con μ y σ desconocidos.

EJERCICIO 10

Calcula el estimador de los momentos para una m.a.s. $X_1, \ldots, X_n$ de

- a) $X \sim P(\lambda)$, con λ desconocido.
- b) $X \sim U(-\theta, \theta)$, con θ desconocido.
- c) $X \sim \Gamma(\alpha, \lambda)$, con α y λ desconocidos.

Como acabamos de ver, el **método de los momentos** permite obtener estimadores de una manera **sencilla**. No obstante, las propiedades de los mismos **no son demasiado buenas**, por lo que, a continuación, presentamos un método alternativo para obtener estimadores con muy buenas propiedades.

El método se basa en el concepto de **verosimilitud**, por lo que recordamos que la verosimilitud de una muestra es la **probabilidad** de que ésta ocurra.

DEFINICIÓN: MÉTODO DE LA MÁXIMA VEROSIMILITUD

El criterio de estimación máximo verosimil se basa en aceptar el "principio de verosimilitud", que afirma que **siempre sucede lo más probable**.

Por ello, vamos a asumir que la muestra de la que disponemos ha sido **observada** puesto que era lo **más probable** y, por tanto, vamos a buscar qué valor del **parámetro** hace que la verosimilitud de la muestra sea **máxima**.

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim f_{X,\theta}$, con θ desconocido. Llamaremos **estimador máximo verosimil** de θ a $\hat{\theta}_{MV}$ tal que:

$$f_{\hat{\theta}_{MV}}(x_1, \ldots, x_n) = \max_{\theta \in \Theta} f_\theta(x_1, \ldots, x_n)$$

- El **procedimiento** para encontrar el valor de θ que máximiza la verosimilitud es el habitual: **derivar** $f_\theta(x_1, \ldots, x_n)$ con respecto a θ y buscar qué valor de θ hace que la derivada se **anule**.
- Sin embargo, la función de verosimilitud es un **producto de densidades**, lo que la hace especielmente **incómoda para la derivación**. Por dicho motivo, de cara a encontrar el estimador máximo verosimil de θ, lo habitual es recurrir al **logaritmo de la verosimilitud** (pues se trata de una función creciente y, por tanto, alcanzará el valor máximo en el mismo punto) y resolver la llamada **ecuación de verosimilitud**:

$$\left.\frac{\partial}{\partial\theta}\ln(f_\theta(x_1, \ldots, x_n))\right|_{\theta=\hat{\theta}_{MV}} = 0$$

- **No es necesario** calcular la derivada segunda y ver que es negativa pues, **a nivel teórico** está demostrado que, si existe la raiz de la ecuación de verosimilitud, es siempre un **máximo**.

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim B(m, p)$, con $0 \leq p \leq 1$ desconocido. Determina el estadístico máximo verosimil para p.

EJERCICIO 11

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim P(\lambda)$, con $\lambda > 0$ desconocido. Determina el estadístico máximo verosimil para λ.

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim U(0, \theta)$, con $\theta > 0$ desconocido. Determina el estadístico máximo verosimil para θ.

EJERCICIO 12

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma)$, con $\mu \in \mathbb{R}$ y $\sigma > 0$. Determina los estadísticos máximo verosimiles para μ y σ^2, suponiendo $\sigma = \sigma_0$ y $\mu = \mu_0$ conocidos, respectivamente.

ESTIMACIÓN MÁXIMO VEROSIMIL PARA MÁS DE UN PARÁMETRO

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim f_{X,\theta_1,\ldots,\theta_k}$, con $\theta_1, \ldots, \theta_k$ desconocidos.

Los estimadores máximo verosímiles de $\theta_1, \ldots, \theta_k$ en el caso de que $f_{\theta_1,\ldots,\theta_k}(x_1, \ldots, x_n)$ sea derivable en todos los parámetros, se obtendrán como las raices del siguiente sistema de ecuaciones de verosimilitud:

$$\begin{aligned}
\frac{\partial}{\partial \theta_1} \ln(f_{\theta_1,\ldots,\theta_k}(x_1, \ldots, x_n)) &= 0 \\
\frac{\partial}{\partial \theta_2} \ln(f_{\theta_1,\ldots,\theta_k}(x_1, \ldots, x_n)) &= 0 \\
&\vdots \\
\frac{\partial}{\partial \theta_k} \ln(f_{\theta_1,\ldots,\theta_k}(x_1, \ldots, x_n)) &= 0
\end{aligned}$$

En este caso tampoco es necesario verificar que las raices son máximos, pues está demostrado a nivel teórico, que siempre es así.

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma)$, con μ y σ desconocidos. Determina los estadísticos máximo verosimiles para μ y σ^2.

PROPIEDADES DE LOS ESTIMADORES MÁXIMO VEROSÍMILES

- El estimador máximo verosimil siempre es **función del estadístico suficiente**. Por lo tanto,
  - si es **insesgado**, es directamente el EIMV. Además, en ese caso, siempre se alcanza la cota de Cramer-Rao y, por tanto, será también **eficiente**.
  - si no insesgado, se puede **corregir el sesgo** y obtener así el **EIMV**. No obstante, en este caso no se alcanzará la cota.

- Ejemplos del primer caso: $X_1, \ldots, X_n$ m.a.s. de $X \sim B(m, p)$ ó $X \sim P(\lambda)$.
- Ejemplo del segundo caso: $X_1, \ldots, X_n$ m.a.s. de $X \sim U(0, \theta)$.

- Si $\hat{\theta}_{MV}$ es el estimador máximo verosimil de θ y $g(\theta)$ es una función **continua** de θ entonces,

$$\widehat{(g(\theta))}_{MV} = g(\hat{\theta}_{MV})$$

Obtén el estadístico máximo verosimil para el percentil 95 de una distribución teórica normal a partir de una m.a.s. de dicha distribución.

Ejercicio 13

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim P(\lambda)$. Obten el estadístico máximo verosimil de $p(X = 0)$ sabiendo que los resultados de una realización muestral fueron: 1, 3, 5, 5, 2, 8, 6, 4.

PROPIEDADES DE LOS ESTIMADORES MÁXIMO VEROSÍMILES

- El estimador máximo verosimil es **asintóticamente insesgado**, es decir,

$$\lim_{n\to\infty} E[\hat{\theta}_{MV}] = \theta$$

- El estimador máximo verosimil es **asintóticamente eficiente**, es decir,

$$\lim_{n\to\infty} Var[\hat{\theta}_{MV}] = \frac{1}{nE\left[\left(\frac{\partial}{\partial\theta} ln(f_{X,\theta})\right)^2\right]}$$

- Si el estimador máximo verosimil se obtiene como raíz de la ecuación de verosimilitud, entonces es **consistente** y **asintóticamente normal** (CAN):

$$\hat{\theta}_{MV} \overset{n\to\infty}{\sim} N(\theta, \frac{1}{\sqrt{nE\left[\left(\frac{\partial}{\partial\theta} ln(f_{X,\theta})\right)^2\right]}})$$

- Sea $X_1, \dots, X_n$ m.a.s. de $X \sim B(m, p)$. En este caso, $\hat{p}_{MV} = \frac{\bar{X}}{m}$ se obtiene como raiz de la ecuación de verosimilitud y, por tanto, es CAN y, por tanto, $\bar{X} \overset{n\to\infty}{\sim} N(p, \sqrt{\frac{p(1-p)}{nm}})$.
- Sea $X_1, \dots, X_n$ m.a.s. de $X \sim U(0, \theta)$. En este caso, $\hat{\theta}_{MV} = \max\{X_1, \dots, X_n\}$ no se obtiene como raiz de la ecuación de verosimilitud y, por tanto, no es CAN (aunque sí es consistente).

EJERCICIO 14

Sea $X_1, \ldots, X_n$ una m.a.s. de $X \sim f_{X,\theta}(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & \text{si } x > 0 \\ 0 & e.o.c \end{cases}$

a) Determina el estimador máximo verosimil de θ.

b) Determina el EIMV.

c) ¿Coinciden? ¿Alcanza alguno la cota de Cramer-Rao?