# Capítulo 5

# Estimación e Intervalos de Confianza.

Con la ayuda de muestras aleatorias simples y con el uso de los estadísticos estamos ya en disposición de describir aunque de manera somera, dos de los métodos fundamentales de la Inferencia Estadística. En la primera sección utilizamos los estadísticos como estimadores de aquellos parámetros en los que estamos interesados. Se considera el riesgo o error cometido como una variable aleatoria y por tanto tratamos de asegurar que el valor esperado de dicha variable sea lo más pequeño posible. En cuanto a Intervalos de confianza diremos que la filosofía es más o menos la misma, quizás la única diferencia esté en que para controlar el riesgo lo que se hace el minimizar la longitud del intervalo - al cual pertenece el parámetro a estimar -.

## 5.1. Estimadores y riesgo.

Tratamos de perfilar algunas de las propiedades más interesantes de los estimadores así como de presentar los métodos más importantes de construcción de dichos estimadores.

### 5.1.1. Estimación y riesgo.

Al estudiar cierto fenómeno se plantea normalmente un modelo en el que intervendrá una o varias variables. En general se supondra que interviene una sólamente. Digamos que $\xi$, cuya función de distribución es $F_{\xi}(x)$, función que se supone depende de un parámetro desconocido $\theta$ (que a su vez puede ser unidimensional o multidimensional). Acerca de $\theta$ sólo sabemos que pertenece a cierto conjunto $\Theta$ llamado espacio paramétrico. Para recordar esta dependencia entre $F_\xi$ y $\theta$ denotaremos por $F_\theta(x)$ a dicha distribución.

A continuación realizamos una m.a.s. $\xi_1, ...\xi_n$ de distribución $F_\theta(x)$ y a partir de esta pretendemos inferir o conjeturar sobre cual es el valor de $\theta$. Para realizar esto se construyen los estimadores puntuales, éstos se construyen de los estadísticos y como ya hemos dicho estimarán o aproximarán al parámetro $\theta$. Es frecuente el que $\theta$ coincida con la esperanza o la varianza poblacional del modelo (ver ejemplos más abajo).

Hemos hablado de una v.a. $\xi$ que modeliza al fenómeno, con $\xi \sim F_\theta$; se toma una m.a.s. $\xi_1, ...\xi_n$ con la misma ley y fabricamos un estadístico $\theta^* = \theta^*(\xi_1, ...\xi_n)$. $\theta^*$ es una nueva v.a. que depende de la muestra y que elegido apropiadamente será el estimador de $\theta$. Para realizar tal estimación se ha de realizar previamente la muestra y sustituir en $\theta^*$: dada $\xi_1, ...\xi_n$, ésta se realiza y se obtienen unos valores (numéricos, ya no son v.a.) $x_1, ...x_n$ que son llevados a $\theta^*$ para dar $\theta^*(x_1, ...x_n)$, aproximación o estimación del parámetro $\theta$. A $\theta^*(x_1, ...x_n)$, resultante de sustituir los valores de la realización de la m.a.s. $x_1, ...x_n$ en el estadístico $\theta^* = \theta^*(\xi_1, ...\xi_n)$, se le llama realización del estadísitico $\theta^*$.

Hemos comentado el elegir apropiadamente el estadístico, esto significa que la diferencia

$$|\theta^*(x_1, ...x_n) - \theta|$$

se lo más pequeña posible. Debemos decir que en la mayoría de las estimaciones, en lugar de considerar esta diferencia se toma $|\theta^*(x_1, ...x_n) - \theta|^2$. En todo caso ambas sirven - teóricamente - para medir el error. Hemos dicho teóricamente ya que $\theta$ no se conoce. No obstante, si lo que queremos es minimizar esta diferencia, el error, podemos considerar la v.a. $Y = |\theta^*(\xi_1, ...\xi_n) - \theta|^2$, la cual representa el error aleatorio - nótese que para cada realización de la m.a.s. se obtiene una aproximación del parámetro $\theta$ -. Bien, pues minimizar el error puede conseguirse en sentido probabilístico, haciendo que el valor esperado (la media) de la v.a. $Y$ sea lo más pequeño posible. En relación con esto veremos seguidamente algunas propiedades deseables para los estadísticos con los que vamos a realizar la estimación.

DEFINICIÓN 5.1 *Diremos que $\theta^*$ es un estimador* insesgado o centrado *de* *parámetro* $\theta$ *si* $E[\theta^*] = \theta$. *Se dirá que* $\theta^*$ *es sesgado si no es insesgado y en este caso se define el* sesgo *de* $\theta^*$ *como* $b(\theta) = E[\theta^*] - \theta$.

DEFINICIÓN 5.2 *Sea* $\theta^* = \theta^*(\xi_1, ...\xi_n)$ *el estimador y supongamos que* $E[\theta^*(\xi_1, ...\xi_n)] = \theta + b(\theta)$. *Se define el error cometido o riesgo de* $\theta^*$ *como*

$$R_{\theta^*}(\theta) = E\left[|\theta - \theta^*|^2\right].$$

Si tenemos en cuenta (??) vemos que

$$\begin{aligned} R_{\theta^*}(\theta) &= E\left[|\theta - \theta^*|^2\right] = E\left[|\theta - E[\theta^*] + E[\theta^*] - \theta^*|^2\right] \\ &= E\left[|\theta^* - E[\theta^*]|^2\right] + E\left[|E[\theta^*] - \theta|^2\right] \\ &= V[\theta^*] + E[b(\theta)], \end{aligned}$$

lo que significa que el riesgo coincide con el error cometido por una desviación cuadrática más el debido a un error de tipo sistemático. Por tanto tenemos:

PROPOSICIÓN 5.1 *Si* $\theta^*$ *es insesgado entonces*

$$R_{\theta^*}(\theta) = V[\theta^*].$$

Tenemos más aún, el hecho de ser insesgado implica que se produzca un riesgo menor al eliminar el error sistemático. Esto nos conduce al concepto de eficiencia:

DEFINICIÓN 5.3 *Diremos que* $\theta^*$ *es un estimador* eficiente *si*

1. $E[\theta^*] = \theta$ *(es insesgado), y*

2. $\theta^*$ *posee mínima varianza.*

La condición de eficiencia equivale a que la media del estimador coincide con el parámetro a estimar y su desviación cuadrática es mínima (mínimo error), i.e.

$$\begin{cases} E[\theta^*] = \theta \\ V[\theta^*] \text{ mínima entre todos los estimadores centrados.} \end{cases}$$

La segunda afirmación quiere dice que si $T$ es cualquier otro estadístico del conjunto de estadísticos

$$U_\theta = \{T = T(\xi_1, ...\xi_n) : E[T] = \theta\},$$

entonces

$$V[T] \geq V[\theta^*],$$

cualquiera que sea $\theta$ perteneciente al conjunto de sus posibles valores, conjunto que denotamos como $\Theta$.

Saber si un estadísitico es centrado o no, es relativamente sencillo, la cuestión más delicada es saber si su varianza es mínima o no. Para esto tenemos el siguiente criterio.

PROPOSICIÓN 5.2 *Si $\theta^*$ es insesgado (esto es, $\theta^* \in U_\theta$) entonces*

$$V[\theta^*] \geq \frac{1}{n\, I_\xi(\theta)},$$

*donde $\xi$ es la distribución que sigue el modelo,*

$$I_\xi(\theta) = E\left[\left[\frac{\partial}{\partial\theta}(\log f_\theta(\xi))\right]^2\right],$$

*y $f_\theta$ es la densidad asociada a $\xi$.*

Por ende si la varianza de un estimador insesgado coincide con $\frac{1}{n\, I_\xi(\theta)}$ entonces tiene mínima varianza y podemos decir que es eficiente.

A la esperanza $I_\xi(\theta)$ se la conoce con el nombre de *cantidad de información* de la v.a. $\xi$ y $\frac{1}{n\, I_\xi(\theta)}$ es la denominada *cota de Fisher-Cramer-Rao.*

EJEMPLO 5.1 *Véanse los ejercicios 124 y 125 -*

### 5.1.2. Obtención de estimadores.

Los métodos más frecuentes para la obtención de estimadores son dos, el *método de los momentos* y el de la *máxima verosimilitud*.

**Método de los momentos.**

Sea $\xi$ v.a. con distribución $F_\theta$ el modelo que estamos estudiando y $\xi_1, ..., \xi_n$ m.a.s. Suponemos que $\theta = (\theta_1, ..., \theta_k)$ es el parámetro ($k$-dimensional) a estimar (bastará con estimar cada una de sus componentes $\theta_j$.

El método consiste en lo siguiente: tomamos los momentos (poblacionales) $\alpha_i = E\left[\xi^i\right]$, los momentos muestrales $M_i = \frac{1}{n}\sum_{j=1}^n \xi_j^i$ y planteamos las ecuaciones

$$\{\alpha_i = M_i, \quad i = 1, ..., k \ ;$$

a continuación se tiene en cuenta que $\alpha_i = \alpha_i\left(\theta_1, ..., \theta_k\right), \ i = 1, ..., k,$ y se resuelve el sistema de $k$ ecuaciones y $k$ incógnitas, $\theta_1, ..., \theta_k$, cuya solución se expresará en la forma

$$\theta_i^* = \theta_i\left(M_1, ..., M_k\right), \quad i = 1, ..., k.$$

Según el método de los momentos, los estadísticos $\theta_i^* = \theta_i\left(M_1, ..., M_k\right)$ obtenidos así serán los estimadores de los parámetros $\theta_i$, o lo que es lo mismo, el vector aleatorio $\theta^* = \left(\theta_1\left(M_1, ..., M_k\right), ..., \theta_k^*\left(M_1, ..., M_k\right)\right)$ será el estimador del parámetro $\theta = \left(\theta_1, ..., \theta_k\right).$

Obsérvese que cuando se trata de estimar los $k$ primeros momentos poblacionales, es decir $\theta = \left(\alpha_1, ..., \alpha_k\right)$, el estimador obtenido por el método de los momentos es insesgado.

EJEMPLO 5.2 *Sea $X \sim U(0, \theta)$ y $X_1, ..., X_n$ una m.a.s.. Hallar un estimador del parámetro $\theta$ por el método de los momentos.*
*Según este método hemos de resolver*

$$\begin{aligned} M_1 &= \overline{X} = \alpha_1 \\ &= \int_0^\theta x f_X(x)\, dx \\ &= \int_0^\theta x \frac{1}{\theta} dx \\ &= \frac{\theta}{2}, \end{aligned}$$

*por tanto el estimador pedido es $\theta^* = 2\overline{X} = \frac{2}{n}\sum_{i=1}^n X_i$.*

EJEMPLO 5.3 *Sea $X \sim N(\mu,\sigma)$ y $X_1,...,X_n$ una m.a.s.. Hallar un estimador del parámetro (bidimensional) $\theta = (\mu,\sigma)$ por el método de los momentos.*

*Planteamos para ello el sistema*

$$\begin{aligned} M_1 &= \overline{X} = \alpha_1 = \mu \\ M_2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \alpha_2 = \sigma^2 + \mu^2. \end{aligned}$$

*De esto se sigue que $\mu = \overline{X}$ y $\sigma = \sqrt{M_2 - \mu^2}$, con lo cual el estimador es*

$$\begin{aligned} \theta^* &= (\mu^*, \sigma^*) \\ &= \left(\frac{1}{n}\sum_{i=1}^{n} X_i, \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\overline{X}\right)^2}\right) \\ &= \left(\overline{X}, \sqrt{\frac{n-1}{n}S^2}\right). \end{aligned}$$

En el contexto del ejemplo anterior se puede demostrar que el estimador $S^2$ es consistente en probabilidad para estimar a $\sigma^2$, es decir, que cuando elegimos $S^2$ para aproximar el valor del parámetro $\sigma^2$ se tiene que

$$\lim_{n\to\infty} P\left\{\left|S^2 - \sigma^2\right| > \epsilon\right\} = 0 \tag{5.1}$$

para todo $\epsilon$ mayor que cero. La demostración de esto consiste en utilizar la desigualdad de Markov (o de Tchevichev):

$$\begin{aligned} P\left\{\left|S^2 - \sigma^2\right| > \epsilon\right\} &= P\left\{\left(S^2 - \sigma^2\right)^2 > \epsilon^2\right\} \\ &\leq \frac{V\left[S^2\right]}{\epsilon^2} \\ &= \frac{2\sigma^4}{(n-1)\,\epsilon^2}; \end{aligned}$$

para establecer esta desigualdad hemos tenido presente que $\frac{n-1}{\sigma^2}S^2 \sim \mathcal{X}_{n-1}^2$, luego $E\left[\frac{n-1}{\sigma^2}S^2\right] = n-1$ y $V\left[\frac{n-1}{\sigma^2}S^2\right] = 2(n-1)$, y de estas dos igualdades se deduce que $E\left[S^2\right] = \sigma^2$ y $V\left[S^2\right] = \dfrac{2\sigma^4}{(n-1)}$.

Para terminar de demostrar (5.1) empleamos la conocida 'regla de sandwich', como

$$0 \underset{n\to\infty}{\longleftarrow} 0 \leq P\left\{\left|S^2-\sigma^2\right| > \epsilon\right\} \leq \frac{2\sigma^4}{(n-1)\,\epsilon^2} \underset{n\to\infty}{\longrightarrow} 0$$

entonces

$$\lim_{n\to\infty} P\left\{\left|S^2-\sigma^2\right| > \epsilon\right\} = 0.$$

EJEMPLO 5.4 *Sea $X \sim U(a,b)$ y $X_1,...,X_n$ una m.a.s.. Dar un estimador por el método de los mínimos cuadrados para el parámetro (bidimensional) $\theta = (a,b)$.*
*Sabemos que*

$$E[X] = \frac{a+b}{2}$$

*y que*

$$V[X] = \frac{(b-a)^2}{12}.$$

*Por tanto el sistema a resolver es*

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} X_i &= \frac{a+b}{2} \\ \frac{1}{n}\sum_{i=1}^{n} X_i^2 &= \frac{(b-a)^2}{12}. \end{aligned}$$

*La solución es $a = M_1 - \sqrt{3\left(M_2 - M_1^2\right)}$, $b = M_1 + \sqrt{3\left(M_2 - M_1^2\right)}$, por tanto el estimador buscado es*

$$\theta^* = \left(M_1 - \sqrt{3\left(M_2 - M_1^2\right)}, M_1 + \sqrt{3\left(M_2 - M_1^2\right)}\right).$$

EJEMPLO 5.5 *Sea $X \sim P(\theta)$ y $X_1,...,X_n$ una m.a.s.. El estimador para $\theta$ por el método de los momentos es $\theta^* = \overline{X}$.*

EJEMPLO 5.6 *Sea $X \sim B(n,p)$ y $X_1,...,X_n$ una m.a.s.. Supongamos que tanto $n$ como $p$ son parámetros desconocidos. El estimador de éstos por el método de los momentos es*

$$\begin{aligned} \theta^*\left({}_1,...,X_n\right) &= \left(n^*\left(X_1,...,X_n\right), p^*\left(X_1,...,X_n\right)\right) \\ &= \left(M_1 - \frac{M_2}{M_1} + 1, \frac{M_1^2}{M_1^2 + M_1 - M_2}\right) \end{aligned}$$

**Método de la máxima verosimilitud.**

Usado desde hace doscientos años resulta ser un método de gran aceptación. La descripción del mismo es la siguiente. Supongamos un modelo con la v.a. $\xi$ modelizando y $\xi \sim F_\theta$. Se supone que ha ocurrido el suceso $A$, podemos escribir $P_\theta(A) \doteq$ *función de verosimilitud de $\theta$ para el suceso $A$* - también se denota como $L(\theta, A)$ -; nótese que $\theta$ se entiende como una variable y $A$ es fijo. $L(\theta, A)$ indica entonces lo creíble que es cada $\theta$ en función de su verosimilitud, o de otra manera, como $A$ ha ocurrido entonces el $\theta$ que hemos de elegir como aproximación de verdadero parámetro es aquél que haga máxima la probabilidad $P_\theta(A)$.

Así pues, diremos que $\theta^*$ es *estimador de máxima verosimilitud* de $\theta$ para el suceso $A$ si

$$L(\theta^*, A) = \max_{\theta \in \Theta} L(\theta, A)$$

Obviamente se tendrá que $\theta^* = \theta^*(A)$. Veamos como se determina este estimador según sea el fenómeno de tipo discreto o continuo.

1.- Caso discreto. Al realizar una muestra $\xi_1, ..., \xi_n$ se obtiene un conjunto de números que llamamos $A$,

$$A = \{x_1, ..., x_n\}.$$

Enotnces la función de verosimilitud es

$$\begin{aligned} L(\theta, A) &= P_\theta(A) \\ &= P_\theta\{X_1 = x_1, ..., X_n = x_n\} \\ &= P_\theta\{X_1 = x_1\} ... P_\theta\{X_n = x_n\} \\ &= P_\theta^*\{x_1\} ... P_\theta^*\{x_n\}. \end{aligned}$$

En ocasiones resultará a efectos de cálculo mucho más sencillo encontrar el máximo para $\log(L(\theta, A))$ que para $L(\theta, A)$, el $\theta^*$ obtenido es el mismo para ambos.

2.- Caso continuo. En este caso el hecho conocido es el siguiente: se supone que si $\xi_1, ..., \xi_n$ es una muestra tal que $\xi_i \in (x_i - \Delta, x_i + \Delta)$, con $\Delta$ pequeño, para $i = 1, ..., n$. A la vista de esto se intentan maximizar las probabilidades $P\{\xi_i \in (x_i - \Delta, x_i + \Delta)\}$ y esto se logra maximizando la densidad conjunta del vector aleatorio $(\xi_1, ..., \xi_n)$. Por consiguiente la función de verosimilitud es

$$\begin{aligned} L(\theta, A) &= f_\theta(x_1, ..., x_n) \\ &= f_\theta(x_1) ... f_\theta(x_n). \end{aligned}$$

EJEMPLO 5.7 *Si $\xi \sim B(1,p)$, con $p$ parámetro desconocido, $\xi_1, ..., \xi_n$ es una m.a.s. y $x_1, ..., x_n$ es una realización de la misma, la función de verosimilitud es*

$$\begin{aligned} L(p,(x_1,...,x_n)) &= P^*(x_1) ... P^*(x_n) \\ &= p^{x_1}(1-p)^{1-x_1} ... p^{x_n}(1-p)^{1-x_n} \\ &= p^{\sum_{j=1}^n x_j}(1-p)^{n-\sum_{j=1}^n x_j} \\ &= p^s(1-p)^{n-s}, \end{aligned}$$

*donde $s = \sum_{j=1}^n x_j$.*

1. *Si $s = 0$ entonces $L(p,(x_1,...,x_n)) = (1-p)^n$ que es máximo cuando $p = 0$, por tanto en este caso $p^* = 0$.*

2. *Si $s = n$ $L(p,(x_1,...,x_n)) = p^n$ que es máximo cuando $p = 1$, por tanto en este caso $p^* = 1$.*

3. *Si $s \in (0,n)$*

$$\frac{\partial L}{\partial p}(p,(x_1,...,x_n)) = 0$$

*si y sólo si*

$$0 = sp^{s-1}(1-p)^{n-s} - p^s(1-p)^{n-s-1}(n-s);$$

*de esto se sigue que $p = \frac{s}{n} = \frac{\sum_{j=1}^n x_j}{n}$ (la estimación de máxima verosimilitud) y que por tanto el estimador de máxima verosimilitud es $p^* = \frac{\sum_{j=1}^n \xi_j}{n} = \overline{\xi}$.*

EJEMPLO 5.8 *Si $\xi \sim N(\mu,\sigma)$, con $\theta = (\mu,\sigma)$ parámetro desconocido, $\xi_1, ..., \xi_n$ es una m.a.s. y $x_1, ..., x_n$ es una realización de la misma, la función de verosimilitud es*

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

*En lugar de maximizar esta función lo hacemos con* $\log\left(f_\theta\left(x_1,...,x_n\right)\right)$ *i.e. con*

$$l_\theta\left(x_1,...,x_n\right)=\log\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n-\frac{\sum_{i=1}^n(x_i-\mu)^2}{2\sigma^2}.$$

*Para calcular los máximos de esta función de dos variables* $\mu\in\mathbb{R}$ *y* $\sigma\in\mathbb{R}^+$ *determinamos sus puntos críticos y vemos si estos llevados al Hessiano da lugar a una matriz definida positiva. Los puntos críticos los obtenemos resolviendo el sistema*

$$\left\{\begin{array}{l}\dfrac{\partial l_\theta}{\partial\mu}=0\\ \dfrac{\partial l_\theta}{\partial\sigma}=0\end{array}\right..$$

*Las ecuaciones son*

$$\left\{\begin{array}{l}0=\dfrac{1}{\sigma^2}\sum_{i=1}^n(x_i-\mu)\\ 0=\dfrac{-n}{\sigma}+\dfrac{\sum_{i=1}^n(x_i-\mu)^2}{\sigma^3}\end{array}\right.,$$

*y los puntos críticos resultantes son* $\mu^*=\dfrac{\sum_{i=1}^n x_i}{n}$ *y* $\sigma^*=\sqrt{\dfrac{\sum_{i=1}^n(x_i-\mu)^2}{n}}$.

*Unos cálculos más o menos sencillos aseguran que la matriz Hessiana*

$$H\left(\mu,\sigma\right)=\left(\begin{array}{cc}\dfrac{\partial^2 l_\theta}{\partial\mu^2} & \dfrac{\partial^2 l_\theta}{\partial\sigma\partial\mu}\\ \dfrac{\partial^2 l_\theta}{\partial\mu\partial\sigma} & \dfrac{\partial^2 l_\theta}{\partial\sigma^2}\end{array}\right)$$

*cuando* $(\mu,\sigma)=(\mu^*,\sigma^*)$ *es definida positiva . Probamos con esto que*

$$(\mu^*,\sigma^*)=\left(\overline{\xi},\sqrt{\frac{\sum_{i=1}^n(X_i-\mu)^2}{n}}\right)$$

*es estimador de máxima verosimilitud de* $(\mu,\sigma)$.

*Conviene observar que existen otras situaciones para este mismo modelo dependiendo de los parámetros que resulten desconocidos. Si* $\mu$ *es desconocida y* $\sigma$ *es conocida entonces el parámetro a estimar es unidimensional y segun el* *principio de máxima verosimilitud es* $\mu^* = \overline{\xi}$ *(se deja como ejercicio). Cuando* $\mu$ *sea conocida y* $\sigma$ *desconocida el estimador que resulta ser de máxima verosimilitud es*

$$\sigma^* = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n}}$$

*(también se deja como ejercicio).*

EJEMPLO 5.9 *Sea* $\xi \sim U(0,\theta)$, *con* $\theta > 0$ *parámetro desconocido y* $\xi_1, ..., \xi_n$ *es una m.a.s.. Demos un estimador máximo verosimil para* $\theta$. *La función de verosimilitud a maximizar es*

$$\begin{aligned} L_\theta(x_1,...,x_n) &= f_\theta(x_1,...,x_n) \\ &= f_\theta(x_1)...f_\theta(x_n) \\ &= \frac{1}{\theta} I_{[0,\theta]}(x_1)...\frac{1}{\theta} I_{[0,\theta]}(x_n) \\ &= \begin{cases} \dfrac{1}{\theta^n} & si\ 0 \le \text{mín}\{x_i\}, \text{máx}\{x_i\} \le \theta \\ 0 & en\ el\ resto \end{cases}, \end{aligned}$$

*con lo cual se aprecia que el máximo se alcanza en la región en la que no es cero* $L_\theta(x_1,...,x_n)$ *y en esta región vale* $\frac{1}{\theta^n}$, *que para maximizarlo hemos de elegir* $\theta$ *lo más pequeño posible. Esto se logra tomando* $\theta = \text{máx}\{x_i\}$ *y por tanto* $\theta^* = \text{máx}\{\xi_i\}$ *es el estimador de máxima verosimilitud.*

EJEMPLO 5.10 *Si* $\xi \sim B(n,p)$, *con* $\theta = (n,p)$ *parámetro desconocido. Se sabe que* $n = 2$ *ó* $3$ *y que* $p = 1/2$ *ó* $1/3$. *No obstante supondremos que sólo nos está permitido realizar una m.a.s. de tamaño 1. Se pide en estas circunstancias encontrar el estimador de máxima verosimilitud para* $\theta$.

*Para poder aplicar el criterio de máxima verosimilitud hemos de tener en cuenta la siguiente tabla de las probabilidades* $P^*(x_1)$ :

| $x$ | $(2,1/2)$ | $(2,1/3)$ | $(3,1/2)$ | $(3,1/3)$ | *prob. máx.* |
|---|---|---|---|---|---|
| 0 | 1/4 | 4/9 | 1/8 | 8/27 | 4/9 |
| 1 | 1/2 | 4/9 | 3/8 | 12/27 | 1/2 |
| 2 | 1/4 | 1/9 | 3/8 | 6/27 | 3/8 |
| 3 | 0 | 0 | 1/8 | 1/27 | 1/8 |

,

*así, si ocurre que la realización de la m.a.s.* $\xi_1$ *es* $x = 0$*, la máxima verosimilitud se alcanza con el par* $(n,p) = (2, 1/2)$*, para* $x = 1$ *la elección será* $(n,p) = (2, 1/2), ...$

## 5.2. Intervalos de confianza.

En lugar de dar una estimación directa del parámetro desconocido lo que haremos será contruir un intervalo, $I = [\theta_1^*(x_1,..,x_n), \theta_2^*(x_1,..,x_n)]$, cuyos extremos son las realizaciones de dos estadísticos, $\theta_1^*(\xi_1,..,\xi_n)$ y $\theta_2^*(\xi_1,..,\xi_n)$, y tal que, con una probabilidad alta tengamos la certeza de que el valor real de dicho parámetro esté en este intervalo y que la longitud del mismo sea lo más pequeña posible.

DEFINICIÓN 5.4 *Sean* $\theta_1^*(\xi_1,..,\xi_n)$ *y* $\theta_2^*(\xi_1,..,\xi_n)$ *dos estimadores de* $\theta$*. Diremos que* $I = [\theta_1^*(\xi_1,..,\xi_n), \theta_2^*(\xi_1,..,\xi_n)]$ *es un intervalo de confianza* $1-\alpha$ *para* $\theta$ *si*

1. $\theta_1^*(x_1,..,x_n) \leq \theta_2^*(x_1,..,x_n)$ *para toda realización* $x_1,..,x_n$ *de la muestra, y*

2. $P\{\theta_1^*(\xi_1,..,\xi_n) \leq \theta \leq \theta_2^*(\xi_1,..,\xi_n)\} = 1-\alpha$ *para todo* $\theta \in \Theta$.

Como ya se ha comentado nuestro objetivo es que $\alpha$ sea pequeño con un intervalo no muy grande.

Observemos que una vez que tenemos el intervalo

$$I = [\theta_1^*(\xi_1,..,\xi_n), \theta_2^*(\xi_1,..,\xi_n)]$$

y se realiza la muestra $x_1,..,x_n$ entonces se genera un intervalo realizando $\theta_1^*$ y $\theta_2^*$,

$$i = [\theta_1^*(x_1,..,x_n), \theta_2^*(x_1,..,x_n)].$$

Si éste es de confianza $1-\alpha$ entonces indicará que para 100 intervalos realizados (para lo cual necesitaremos realizar 100 muestras) en un $1-\alpha\,\%$ de ocasiones el valor real del parámetro estará dentro de dichos intervalos, mientras que sobre el resto no sabremos nada.

Veamos ahora un método bastante eficaz en la construcción de intervalos de confianza.

*Método de la cantidad pivotal.*

DEFINICIÓN 5.5 *Dada una m.a.s.* $\xi_1, ..., \xi_n$ *con distribución* $F_\theta$*, diremos que el estadístico* $Q(\theta, \xi_1, ..., \xi_n)$ *, que depende de* $\xi_1, ..., \xi_n$ *y de* $\theta$*, es una* cantidad pivotal *si tiene una función de distribución asociada que es independiente de* $\theta$*.*

El método para construir el intervalo es el siguiente: damos $q_1$ y $q_2$, dos números independientes de $\theta$ y $Q(\theta, \xi_1, ..., \xi_n)$ cantidad pivotal tal que

$$P\{q_1 \leq Q(\theta, \xi_1, ..., \xi_n) \leq q_2\} = 1 - \alpha,$$

y de esta igualdad hemos de pasar a considerar una igualdad con la forma de la dada en (ii) de la Definición 5.3,

$$P\{\theta_1^*(\xi_1, ..., \xi_n) \leq \theta \leq \theta_2^*(\xi_1, ..., \xi_n)\} = 1 - \alpha.$$

EJEMPLO 5.11 *Si* $\xi \sim N(\mu, \sigma)$*, con* $\sigma$ *conocido y* $\mu$ *parámetro a estimar. Buscamos para éste un intervalo de confianza. Hemos de hallar una cantidad pivotal en la que esté involucrada una m.a.s. de tipo* $N(\mu, \sigma)$ *y el parámetro* $\mu$*. Para este caso tomamos*

$$Q(\mu, \xi_1, ..., \xi_n) = \frac{\overline{\xi} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

*Busquemos ahora dos números* $q_1$ *y* $q_2$ *tales que*

$$P\left\{q_1 \leq \frac{\overline{\xi} - \mu}{\sigma/\sqrt{n}} \leq q_2\right\} = 1 - \alpha,$$

*es decir*

$$P\left\{\overline{\xi} - q_2\left(\sigma/\sqrt{n}\right) \leq \mu \leq \overline{\xi} - q_1\left(\sigma/\sqrt{n}\right)\right\} = 1 - \alpha,$$

*con lo que a falta de saber quienes son* $q_1$ *y* $q_2$ *el intervalo será*

$$I = \left[\overline{\xi} - \left(\sigma/\sqrt{n}\right) q_2, \overline{\xi} - \left(\sigma/\sqrt{n}\right) q_1\right].$$

*La longitud del intervalo es* $l = \frac{\sigma}{\sqrt{n}}(q_2 - q_1)$*, y se hace mínima cuando tomamos* $q_1 = -q_2$ *(ver Figura 1). Al número que con una* $N(0,1)$ *deja a su izquierda una probabilidad* $1 - \alpha$ *(o a la derecha una probabilidad* $\alpha$*) lo denotamos por* $Z_\alpha$*. Con esta nomenclatura tendremos que* $q_2 = Z_{\alpha/2}$ *y por tanto el intervalo de confianza* $1 - \alpha$ *para* $\mu$ *es*

$$I = \left[\overline{\xi} - Z_{\alpha/2}\left(\sigma/\sqrt{n}\right), \overline{\xi} + Z_{\alpha/2}\left(\sigma/\sqrt{n}\right)\right].$$

*por ejemplo, para* $\alpha = 0{,}00270$ *se tiene que* $Z_\alpha = -3$ *y con esta elección de* $\alpha$ *encontraríamos un intervalo de confianza* $0{,}9973$*.*

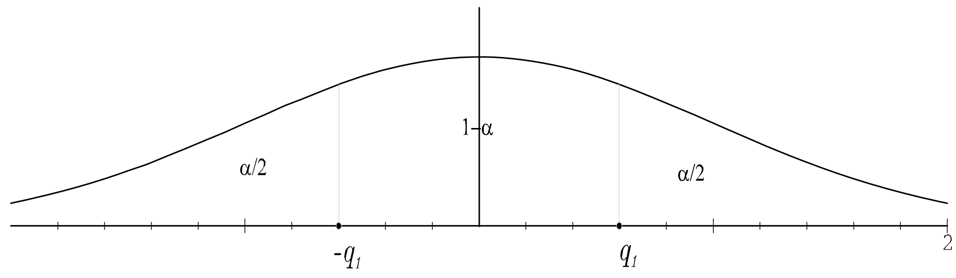


Figura 5.1: Elección de los puntos $q_i$

EJEMPLO 5.12 *Sea $\xi \sim N(\mu,\sigma)$, con $\sigma$ y $\mu$ desconocidos, pero siendo $\mu$ el parámetro a estimar. Damos para $\mu$ un intervalo de confianza. En esta ocasión no podemos utilizar la cantidad pivotal del ejemplo anterior ya que $\sigma$ es un parámetro desconocido. Ahora tomaremos como cantidad pivotal a*

$$Q\left(\mu,\xi_1,...,\xi_n\right)=\frac{\overline{\xi}-\mu}{S/\sqrt{n}}\sim \mathsf{t}_{n-1}.$$

*Busquemos ahora dos números $q_1$ y $q_2$ tales que*

$$P\left\{q_1\leq \frac{\overline{\xi}-\mu}{S/\sqrt{n}}\leq q_2\right\}=1-\alpha.$$

*Se repite el proceso del ejemplo anterior, incluso en la determinación de los puntos $q_1$ y $q_2$ ya que la 'naturaleza' de las densidades de una $N(0,1)$ y de una $\mathsf{t}_{n-1}$ es la misma. Al número que con una $\mathsf{t}_{n-1}$ deja a su izquierda una probabilidad $1-\alpha$ ( o la derecha una probabilidad $\alpha$) lo denotamos por $\mathsf{t}_{n-1,\alpha}$. Con esta nomenclatura tendremos que $q_2=\mathsf{t}_{n-1,\alpha/2}$*

$$P\left\{\overline{\xi}-\mathsf{t}_{n-1,\alpha/2}\left(S/\sqrt{n}\right)\leq\mu\leq\overline{\xi}+\mathsf{t}_{n-1,\alpha/2}\left(S/\sqrt{n}\right)\right\}=1-\alpha.$$

EJEMPLO 5.13 *Si $\xi\sim N(\mu,\sigma)$, con $\mu$ desconocido pero siendo $\sigma^2$ el parámetro a estimar. Para este terecer caso tomamos como cantidad a*

$$Q\left(\mu,\xi_1,...,\xi_n\right)=\frac{(n-1)\,S^2}{\sigma^2}\sim \mathcal{X}^2_{n-1}.$$

*Los puntos $q_i$ son:*

$$q_1=\mathcal{X}^2_{n-1,1-\frac{\alpha}{2}}$$

*y*

$$q_2=\mathcal{X}^2_{n-1,\frac{\alpha}{2}},$$

*resultando ser*

$$I = \left[ \frac{(n-1)\, S^2}{\mathcal{X}^2_{n-1,\frac{\alpha}{2}}}, \frac{(n-1)\, S^2}{\mathcal{X}^2_{n-1,1-\frac{\alpha}{2}}} \right]$$

*el intervalo de confianza* $1-\alpha$ *para* $\sigma^2$.

EJEMPLO 5.14 *Puede darse la circunstancia de no encontrar una cantidad pivotal al querer estimar cierto parámetro* $\theta$, *correspondiente a cierto modelo. Una alternativa consiste en considerar un estimador insesgado* $\theta^*$ *de* $\theta$ *y buscar un* $\epsilon > 0$ *de forma que*

$$P\left\{|\theta^* - \theta| < \epsilon\right\} = 1-\alpha.$$

*Para ajustar el* $\epsilon$ *se procede así: en lugar de tomar la igualdad tomamos un mayor o igual (mayor confianza) y usamos la desigualdad de Markov:*

$$\begin{aligned} P\left\{|\theta^* - \theta| < \epsilon\right\} &= 1 - P\left\{|\theta^* - \theta| \geq \epsilon\right\} \\ &\geq 1 - \frac{V\left[\theta^*\right]}{\epsilon^2} = 1-\alpha, \end{aligned}$$

*por tanto bastará con tomar* $\epsilon$ *tal que* $1 - \frac{V[\theta^*]}{\epsilon^2} = 1-\alpha$ *para dar con un intervalo de confianza* $\geq 1-\alpha$, *aunque este no sea de longitud mínima.*

EJEMPLO 5.15 *En este ejemplo mostramos como con la aproximación que proporciona el Teroema Central del Límite (TCL) se puede construir un intervalo de confianza. Analicemos esto con el Ejercicio 123. Se trata de un modelo descrito con la ayuda de una v.a.* $\xi \sim B(1,p)$, $p$ *parámetro a estimar. Aplicamos el TCL para asegurar que*

$$\overline{\xi} \sim N\left(\mu, \sigma/\sqrt{n}\right)$$

*donde* $\mu = E\left[\xi\right] = p$ *y* $\sigma^2 = V\left[\xi\right] = p(1-p)$. *La cantidad pivotal a usar es*

$$Q = \frac{\overline{\xi} - \mu}{\sigma/\sqrt{n}} \sim N(0,1),$$

*pero el problema es que no conocemos a* $\sigma$. *Para seguir hemos de dar una estimación de* $\sigma$, *bien calculando media muestral (que en este problema es 0.2) y haciendo* $\sigma^2 \approx (0{,}2)\,(0{,}8)$, *o bien realizando la cuasivarianza muestral (realizando* $S^2$*),* $\sigma \approx \sqrt{S^2}$. *Con la primera estimación llegamos a que*

$$I = \left[\overline{\xi} - Z_{\alpha/2}\left(\sigma/\sqrt{n}\right), \overline{\xi} + Z_{\alpha/2}\left(\sigma/\sqrt{n}\right)\right]$$

*es un intervalo de confianza* $1-\alpha$ *para* $p$[1].

EJEMPLO 5.16 *En este ejemplo buscamos un intervalo de confianza para* $\theta = \mu_1 - \mu_2$, *siendo* $X \sim N(\mu_1, \sigma_1)$, $Y \sim N(\mu_2, \sigma_2)$, $X_1, ..., X_n$ *m.a.s. de* $X$ *e* $Y_1, ..., Y_m$ *m.a.s. de* $Y$ *(ambas independientes). Se supone que las varianzas* $\sigma_i^2$ *son conocidas. La cantidad pivotal a usar es*

$$Q = \frac{\left(\overline{X} - \overline{Y}\right) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim N(0,1).$$

*Los detalles se dejan para el lector.*

## 5.3. Problemas del Capítulo V

1. La distancia $\xi$ entre un árbol cualquiera y el árbol más próximo a él en un bosque sigue una distribución con función de densidad de probabilidad

$$f_\xi(x) = \begin{cases} 2\theta x \exp\left(-\theta x^2\right) & \text{si } x \geq 0 \\ 0 & \text{en el resto} \end{cases}$$

   ($\theta > 0$). Obtener el estimador de máxima verosimilitud de $\theta$, supuesto que se realiza una m. a. s. de tamaño $n$.

2. Idem para

$$f_\xi(x) = \begin{cases} \frac{x^2}{2\theta^3} \exp\left(-x/\theta\right) & \text{si } x \geq 0 \\ 0 & \text{en el resto} \end{cases}$$

3. El coseno $\xi$ del ángulo con el que se emiten los electrones en un proceso radiactivo es una variable aleatoria con densidad de probabilidad

$$f_\xi(x) = \begin{cases} \frac{1+\theta x}{2} & \text{si } x \in [-1,1] \\ 0 & \text{en el resto} \end{cases}$$

   donde $\theta \in [-1,1]$. Dada una m.a.s. $\xi_1, ..., \xi_n$, encontrar el estimador de $\theta$ por el método de los momentos.

[1] También podría intentarse despejar $p$ de $P\left\{q_1 \leq \frac{\overline{\xi}-p}{\sqrt{p(1-p)/n}} \leq q_2\right\} = 1-\alpha$.

4. Una variable aleatoria $\xi$ tiene definida su densidad según las igualdades

$$P_{\xi}^{*}(-1) = \frac{1-\theta_1}{2},\ P_{\xi}^{*}(0) = \frac{\theta_1+\theta_2}{2},\ P_{\xi}^{*}(1) = \frac{1-\theta_2}{2},$$

con $\theta_1,\ \theta_2 \in (0,1)$. Estimar el parámetro $\theta = (\theta_1, \theta_2)$ por el método de los momentos.

5. Idem para

$$f_{\xi}(x) = \begin{cases} \theta x^{\theta-1} & \text{si } x \in [0,1] \\ 0 & \text{en el resto} \end{cases}$$

con $\theta > 0$.

6. Sea $\xi_1, ..., \xi_n$ una m.a.s. de tamaño n y cuya distribución viene definida como

$$P_{\xi_i}^{*}(x) = p(1-p)^{x-1}, \quad x = 0, 1, 2, ...$$

para $i = 1, 2, ..., n$ $(0 < p \leq 1)$. Se pide

   *a*) Estimador de $p$ por el método de los momentos.

   *b*) Estimador de $p$ por el método de la máxima verosimilitud. (Ayuda: $E[\xi_i] = \frac{1}{p}$.)

7. Idem con el caso en que $\xi_1, ..., \xi_n$ una m.a.s. de tamaño n y cuya distribución viene definida por la fucnión de densidad

$$f_{\xi}(x) = \begin{cases} e^{\theta-x} & \text{si } x \in [\theta, +\infty] \\ 0 & \text{en el resto} \end{cases}$$

(se supone $\theta > 0$).

8. Sea la variable aleatoria $\xi$ cuya función de densidad de probabilidad es

$$f_{\xi}(x) = \begin{cases} \dfrac{\theta}{x^{\theta+1}} & \text{si } x \geq 1 \\ 0 & \text{en el resto,} \end{cases}$$

donde se supone que $\theta > 1$. Dada una muestra aleatoria simple de $\xi$ y tamaño $n$, $\xi_1, ...\xi_n$, se pide:

a) Determinar un estimador del parámetro $\theta$ por el método de los momentos.

b) Obtener un estimador para $\theta$ por el método de máxima verosimilitud.

9. Sea $\xi$ una variable aleatoria que modeliza a cierto fenómeno. Se sabe que
$$P\{\xi=-1\}=\frac{1-\theta}{2},\ P\{\xi=0\}=\frac{1}{2},\ P\{\xi=1\}=\frac{\theta}{2},$$
donde $\theta$ es cierto parámetro del que sólo se sabe que $\theta\in(0,1)$.

   a) Dada una m.a.s. $\xi_1,...,\xi_n$ de la v.a. $\xi$, determinar el estimador de máxima verosimilitud del parámetro $\theta$.

   b) Si se toma $n=50$ y se realiza la muestra obteniéndose los resultados,

| | | | |
|---|---|---|---|
| $x_i:$ | $-1$ | 0 | 1 |
| $n_i:$ | 10 | 25 | 15 |

   donde $n_i$ es la frecuencia del dato $x_i$ (número de veces que aparece $x_i$), realizar una estimación del parámetro $\theta$ usando para ello el método de la máxima verosimilitud.

10. Sea una población modelizada con ayuda de la variable aleatoria $\xi$, de la que se sabe que su densidad de probabilidad es
$$f_\xi(x)=\begin{cases} x^\theta(\theta+1) & \text{si } x\in[0,1]\\ 0 & \text{en el resto,}\end{cases}$$
con $\theta$ un parámetro tal que $\theta>-1$. Dada una muestra aleatoria simple, hallar un estimador para $\theta$ empleando para ello el principio de máxima verosimilitud.

11. Deducir de manera razonada un intervalo de confianza para la varianza de una población que se distribuye normalmente y donde se supone que la esperanza es desconocida. Aplica el resultado obtenido al siguiente ejemplo: un metalúrgico ha hecho 4 determinaciones del punto de fusión del manganeso, 1269, 1271, 1263 y 1265 grados C. Tras valorar los resultados y los posibles errores decidió que la varianza habría de ser 1 ó 2. ¿ Es posible esta afirmación con los datos experimentales obtenidos, suponiendo normalidad y a un nivel de confianza del 95 % ?

12. Sobre la base de la realización de una muestra aleatoria simple de 81 observaciones, los expertos de seguridad estimaron que el tiempo de reacción de los camioneros ante una luz roja era, en media, 2 segundos, con una cuasivarianza muestral de $(0'60)^2$, esto es

$$\frac{\sum_{i=1}^{144} x_i}{81} = 2 \quad \text{y} \quad \frac{\sum_{i=1}^{144} (x_i - 2)^2}{80} = (0'60)^2 .$$

Se desea calcular un intervalo de confianza del 99 % para el tiempo medio de reacción:

*a*) Haciendo la hipótesis de normalidad y tomando como varianza poblacional la cuasivarianza muestral.

*b*) Haciendo hipótesis de normalidad y suponiendo desconocida la desviación típica poblacional.

13. Se sostiene la hipótesis de que la demanda diaria (en barriles) de petróleo por parte de la Comunidad Económica Europea (CEE) a los países árabes, posee una desviación típica de 200, mientras que la de EEUU a este mismo grupo de países tiene como desviación típica 100. Se ha realizado una muestra aleatoria simple de tamaño 125 para la demanda de la CEE, resultando que la demanda media ha sido de 300 barriles. También se ha realizado un muestreo aleatorio simple de tamaño 100 para la demanda de EEUU, la media ha sido de 250. Suponiendo que la demanda diaria de petróleo (tanto de la CEE como de EEUU) está regida por una ley de probabilidad normal, elaborar con todo detalle un intervalo de confianza del 95 % para la diferencia de las demandas medias. Nota: se supone que las demandas de petróleo por parte de EEUU y la CEE son independientes.

14. Una población está representada por una variable aleatoria $\xi$, cuya función de distribución es

$$F_\xi(x) = \begin{cases} 1 - \left(\dfrac{1}{x}\right)^\theta & \text{si } x > 1, \\ 0 & \text{en el resto,} \end{cases}$$

y donde $\theta$ es un parámetro real positivo. Sea una muestra aleatoria simple $\xi_1, ..., \xi_n$ perteneciente a dicha población. Se pide:

a) Determinar la función de densidad de probabilidad asociada a la variable aleatoria $\xi$.

b) Encontar un estimador del parámetro $\theta$ usando para ello el principio de máxima verosimilitud.

c) Idem usando el método de los momentos.

15. Sea $\xi$ una variable aleatoria cuya función de densidad es

$$f_\xi(x) = \begin{cases} \frac{2a}{1-a} x^{\frac{2a}{1-a}-1} & \text{si } x \in [0,1], \\ 0 & \text{en el resto.} \end{cases}$$

donde $a$ es un parámetro del cual se sabe que es.positivo. Sea $\xi_{1,}\xi_2, ..., \xi_n$ una muestra aleatoria simple de $\xi$. Se pide:

a) Hallar un estimador de máxima verosimilitud para el parámetro $a$.

b) Obtener por el método de los momentos un estimador para $a$.

16. Una empresa desea determinar la proporción de clientes dispuestos a adquirir uno de sus productos. Estima que dicha proporción es 0,4 ó 0,5. Decidir en base al Principio de Máxima Verosimilitud, una estimación de dicha proporción si después de realizar una muestra aleatoria simple de tamaño 15 entre sus clientes potenciales, 6 de ellos afirmaron estar dispuestos a la adquirir y los 9 restantes no estaban dispuestos a optar por dicho producto.

17. Sea la v.a. $\xi$ cuya función de densidad de probabilidad es

$$f_\xi(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & \text{si } x \geq 0 \\ 0 & \text{en el resto.} \end{cases}$$

a) Dada una m.a.s. $\xi_1, ..., \xi_n$ de la v.a. $\xi$, determinar por el método de máxima verosimilitud un estimador del parámetro $\theta$.

b) ¿Es insesgado el estimador obtenido en el apartado anterior?

c) ¿Cuál es el riesgo para este estimador?

18. Sea una v.a. $\xi$ cuya densidad de probabilidad es

$$f(x) = \begin{cases} (\theta + 1)\, x^{\theta} & \text{si } x \in (0,1) \\ 0 & \text{en el resto,} \end{cases}$$

con $\theta$ cierto parámetro desconocido del que sólo se sabe que $\theta > -1$. Dada una muestra aleatoria simple $\xi_1, \xi_2, ..., \xi_n$ se pide:

*a*) Un estimador de $\theta$ por el método de la máxima verosimilitud.

*b*) Un estimador de $\theta$ por el método de los momentos.