# Máxima verosimilitud

Tenemos una variable $X$ discreta con función de masa

$$f(x;\theta),$$

donde $x \in A$ y $\theta \in \Theta$. $(X_1, \ldots, X_n)$ es una muestra aleatoria de $X$.

- **A priori**, la probabilidad, conocido $\theta$, de obtener una realización específica (potencial) $(x_1, \ldots, x_n)$ de $(X_1, \ldots, X_n)$ viene dada por

$$(\star) \qquad \prod_{i=1}^{n} f(x_i;\theta)\,.$$

- En la práctica (estadística) la situación es justo la contraria: disponemos una realización concreta $(x_1, \ldots, x_n)$, pero desconocemos $\theta$.

**Idea:** para $(x_1, \dots, x_n)$ dado,

- tomamos como estimación de $\theta$ al valor $\widehat{\theta}$ que hace máxima la expresión $(\star)$;
- y entendemos que $\widehat{\theta}$ es el valor de $\theta$ más verosímil dada la realización observada de la muestra $(x_1, \dots, x_n)$.

*A priori: probabilidad de realizaciones.*
*A posteriori, verosimilitud de parámetros.*

Llamamos **función de verosimilitud** a la función $V$

$$\theta \mapsto V(\theta; x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i; \theta)\,,$$

donde $(x_1, \dots, x_n) \in A^n$ hace ahora el papel de parámetros.

El estimador de máxima verosimilitud $\mathbf{emv}_\theta(x_1, \dots, x_n)$ de $\theta$ es aquel valor $\widehat{\theta} \in \Theta$ donde $V$ se hace máxima, suponiendo que haya tal valor. Se trata de un máximo global.

Habitual: considerar

$$\mathrm{LV}(\theta; x_1, \dots, x_n) = \ln(V(\theta; x_1, \dots, x_n)) = \sum_{i=1}^{n} \ln(f(x_i; \theta)),$$

la **log-verosimilitud**. Las funciones $V$ y $LV$ alcanzan el máximo en el mismo punto $\theta$ de $\Theta$.

Si $X$ es una **variable continua** con función de densidad $f(x;\theta)$ con $\theta \in \Theta$, la función de verosimilitud se define

$$V(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i;\theta),$$

y la estimación máximo-verosímil es el valor de $\theta$ que maximiza, para una muestra $(x_1, \ldots, x_n)$ dada, la función de máximo-verosimilitud.

O quizás la función de log-verosimilitud

$$\mathrm{LV}(\theta; x_1, \ldots, x_n) = \ln(V(\theta; x_1, \ldots, x_n)) = \sum_{i=1}^{n} \ln(f(x_i;\theta)),$$

En lo que sigue solo escribiremos $V(\theta)$ y $LV(\theta)$.

Obsérvese que $\mathbf{emv}_\theta$ es función de $x_1, \ldots, x_n$, pues para cada realización, la función de verosimilitud cambia.

De manera que $\mathbf{emv}_\theta(X_1, \ldots, X_n)$ es un **estadístico estimador** del parámetro $\theta$. Que en algunos casos se puede calcular explícitamente.

# Ejemplos

Ejemplo 1. $X \sim \text{BER}(p)$. Queremos estimar $p \in [0, 1]$.

$X$ toma sólo los valores 0 y 1. De una realización $(x_1, \ldots, x_n)$ de $(X_1, \ldots, X_n)$ sólo necesitamos saber el número $m$ de unos que hay en ella (el número de ceros es $n - m$).

La función de verosimilitud $V$ viene dada por

$$p \in [0, 1] \mapsto V(p) = p^m \, (1 - p)^{n-m}$$

Queremos hallar el valor de $p$ que hace máxima esta expresión.

La función $V$ es

- continua y no negativa;
- se anula en $p = 0$ y en $p = 1$;
- y tiene un máximo en $(0, 1)$.

¿Dónde?

Obsérvese que

$$\begin{aligned}\partial_p V(p) &= m\, p^{m-1}\,(1-p)^{n-m} + p^m\,(n-m)\,(1-p)^{n-m-1} \\ &= \big[m\,(1-p) - (n-m)\,p\big]\; p^{m-1}\,(1-p)^{n-m-1}\,,\end{aligned}$$

que se anula cuando $m(1-p) = (n-m)p$. Es decir

$$\widehat{p} = \frac{m}{n}\,.$$

La estimación $\widehat{p}$ es, en efecto, función de la realización de la muestra (a través de $m$).

Visto como estadístico, sería

$$\mathbf{emv}_p(X_1, \ldots, X_n) \equiv \overline{X}.$$

Ejemplo 2. $X \sim \text{GEO}(p)$. Queremos estimar $p \in [0,1]$.

La realización $(x_1, \ldots, x_n)$ consiste de enteros positivos. La función de verosimilitud es

$$V(p) = \prod_{i=1}^{n} p\,(1-p)^{x_j - 1} = p^n\,(1-p)^{\sum_{j=1}^{n} x_j - n}\,.$$

Se obtiene que

$$\widehat{p} = \frac{n}{\sum_{j=1}^{n} x_j} = \frac{1}{\frac{1}{n}\sum_{j=1}^{n} x_j} = \frac{1}{\overline{x}}.$$

Así que

$$\mathbf{emv}_p(X_1, \ldots, X_n) \equiv \frac{1}{\overline{X}}.$$

Ejemplo 3. $X \sim \textsc{unif}[0, a]$. Queremos estimar $a > 0$.

La función de densidad de $X$ es

$$f(x; a) = \begin{cases} 1/a\,; & \text{si } 0 \le x \le a\,, \\ 0\,; & \text{resto}\,. \end{cases}$$

Dada una muestra $(x_1, \ldots, x_n)$, la función de verosimilitud es

$$V(a) = \begin{cases} 1/a^n & \text{si } \max\{x_j\} \le a\,, \\ 0 & \text{si } \max\{x_j\} > a\,. \end{cases}$$

El máximo se alcanza justamente en $a = \max\{x_1, \ldots, x_n\}$, así que

$$\mathbf{emv}_a(X_1, \ldots, X_n) = \max\{X_1, \ldots, X_n\}\,.$$

Ejemplo 4. $X \sim \text{EXP}(\lambda)$. Queremos estimar $\lambda > 0$.

Función de verosimilitud:

$$V(\lambda) = \prod_{i=1}^{n} \lambda\, e^{-\lambda\, x_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}\,.$$

Obsérvese que $V \to 0$ cuando $\lambda \downarrow 0$ y cuando $\lambda \uparrow \infty$.Tomamos logaritmos:

$$\mathrm{LV}(\lambda) = n \ln(\lambda) - \lambda n \overline{x}\,.$$

Derivando e igualando a 0:

$$\partial_\lambda \mathrm{LV}(\lambda) = \frac{n}{\lambda} - n\overline{x} = 0 \quad \Longrightarrow \quad \widehat{\lambda} = \frac{1}{\overline{x}}\,.$$

Así que

$$\mathbf{emv}_\lambda(X_1, \dots, X_n) = \frac{1}{\overline{X}}\,.$$

Ejemplo 5. $X \sim \mathcal{N}(\mu, \sigma^2)$. Hay dos parámetros. Queremos estimar $\mu \in \mathbb{R}$ y $\sigma^2 > 0$.

Función de verosimilitud:

$$V(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma^2}} e^{-(x_i-\mu)^2/(2\sigma^2)}$$
$$= \frac{1}{(2\pi)^{n/2}} \frac{1}{(\sigma^2)^{n/2}} \exp\Big( - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \Big)$$

Previo: en el semiplano $(\mu, \sigma^2)$,

- Para $\mu$ fijo, si $\sigma^2 \to +\infty$, se tiene que $V(\mu, \sigma^2) \to 0$.
- Para $\sigma^2$ fijo, si $\mu \to \pm\infty$, también se tiene que $V(\mu, \sigma^2) \to 0$.

Tomamos logaritmos:

$$LV(\mu,\sigma^2) = \ln\Big(\frac{1}{(2\pi)^{n/2}}\Big) - \frac{n}{2}\,\ln(\sigma^2) - \frac{1}{2\sigma^2}\,\sum_{i=1}^{n}(x_i-\mu)^2.$$

Y planteamos el sistema:

$$\begin{cases} \partial_\mu \; LV(\mu,\sigma^2) = 0 \\ \partial_{\sigma^2} \; LV(\mu,\sigma^2) = 0 \end{cases}$$

Solución:

$$\widehat{\mu} = \overline{x}\,, \qquad \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2.$$

Así que

$$\mathbf{emv}_\mu(X_1,\dots,X_n) = \overline{X}\,, \qquad \mathbf{emv}_{\sigma^2}(X_1,\dots,X_n) = D^2\,.$$