Estimación puntual de parámetros

X: variable aleatoria cuya distribución de probabilidad depende de (y/o está determinada) por un cierto parámetro θ **que desconocemos** y que nos **interesa estimar**.

Sólo sabemos que X pertenece a una cierta familia de distribuciones de probabilidad, pero no sabemos qué distribución concreta.

El objetivo es **estimar** el valor de θ a partir de una muestra $x_1, x_2, \ldots, x_n$ que se supone que se ha obtenido aleatoria e independientemente, es decir, a partir de una realización de $(X_1, X_2, \ldots, X_n)$.

Ejemplos

- $X \sim \text{BER}(p)$. El parámetro es $\theta = p$.

  Ejemplo: p es porcentaje de gente que va a votar "no" en un referendo. Queremos estimar p a partir de muestra de tamaño n de respuestas.

- $X \sim \mathcal{N}(0, \sigma^2)$. El parámetro aquí sería $\theta = \sigma$ o $\theta = \sigma^2$.

  Ejemplo: X registra errores en mediciones con un cierto aparato (calibrado de manera que el error medio es nulo).

- $X \sim \text{EXP}(\lambda)$. Podría ser $\theta = \lambda$ o quizás $\theta = 1/\lambda$.

  Ejemplo: tiempo de espera hasta siguiente mensaje, o tiempo de espera hasta que un cliente es atendido en cola de IKEA.

- $X \sim \text{POISS}(\lambda)$. Podría ser $\theta = \lambda$ o quizás $\theta = e^{-\lambda}$.

  Ejemplo: analizamos el número de ocurrencias de un fenómeno relativamente raro en un intervalo de tiempo relativamente corto.

Generalidades sobre estimadores

- **Muestra empírica**. Disponemos de una muestra $x_1, x_2, \ldots, x_n$ de valores concretos de X (obtenidos de forma aleatoria e independiente).

Calculamos con cierta función $h : \mathbb{R}^n \to \mathbb{R}$ una **estimación** de θ:

$$h(x_1, x_2, \ldots, x_n) \quad \Longleftarrow \quad \text{ESTIMACIÓN DE } \theta$$

La función h podría ser la media aritmética, el máximo, etc.

- **Contrapartida teórica.** Tenemos $(X_1, X_2, \ldots, X_n)$ clones independientes de X. La distribución de probabilidad del estadístico

$$T = h(X_1, X_2, \ldots, X_n)\,,$$

recoge todos las posibles estimaciones de θ y sus respectivas probabilidades de ocurrencia. T es un **estadístico estimador** de θ o simplemente un **estimador** de θ.

Sesgo de un estimador

La primera propiedad deseable de un estimador T es que al menos en media no se equivoque.

Decimos que $T = h(X_1, X_2, \ldots, X_n)$ es un **estimador insesgado** del parámetro θ si

$$\boxed{\mathbf{E}(T) = \theta}$$

Si $\mathbf{E}(T) \neq \theta$, entonces T es un estimador sesgado de θ.

La diferencia $\mathbf{E}(T) - \theta$ es el **sesgo** (al alza si es positivo, a la baja si es negativo).

Ejemplos

EJEMPLO 1. X variable aleatoria. Queremos estimar $\mu = \mathbf{E}(X)$.

Usamos la media muestral $\overline{X}$, que siempre es estimador insesgado de μ, pues

$$\mathbf{E}(\overline{X}) = \mu\,.$$

EJEMPLO 2. X variable aleatoria. Queremos estimar $\sigma^2 = \mathbf{V}(X)$.

Usamos la cuasivarianza muestral S^2, que siempre es estimador insesgado de σ^2, pues

$$\mathbf{E}(S^2) = \sigma^2\,.$$

¡Atención!, la cuasidesviación típica muestral S es un estimador sesgado de σ, pues $\mathbf{E}(S)^2 < \mathbf{E}(S^2)$.

EJEMPLO 3. $X \sim \text{BER}(p)$. Queremos estimar p.

• Como $\mathbf{E}(X) = p$, podemos usar $\overline{X}$ (que aquí es simplemente proporción de unos) como estimador insesgado de p.

• Como $\mathbf{V}(X) = p(1-p)$, podemos usar S^2 como estimador insesgado de $p(1-p)$.

Problemas:

- Dos raíces para p;
- estimación sesgada de p.

EJEMPLO 4. $X \sim$ UNIF$[0, a]$. Queremos estimar a.

- Tomamos como estimador $T = \max(X_1, X_2, \ldots, X_n)$, que tiene

$$\mathbf{E}(T) = \frac{n}{n+1} a.$$

Así que T sería sesgado.

- El estimador $\frac{n+1}{n} T$ sería un estimador insesgado de a.

- Alternativamente, podríamos tomar $T' = 2\overline{X}$.

Como $\mathbf{E}(X) = a/2$, resulta que

$$\mathbf{E}(T') = 2\mathbf{E}(\overline{X}) = 2\mathbf{E}(X) = a.$$

Así que T' es un estimador insesgado del parámetro a.

EJEMPLO 5. $X \sim \text{EXP}(\lambda)$. Queremos estimar λ.

• Como $\mathbf{E}(X) = 1/\lambda$, tendríamos que $\overline{X}$ sería un estimador insesgado de $1/\lambda$.

• Pero $1/\overline{X}$ sería un estimador sesgado de λ.

• Consideremos $T = \min(X_1, X_2, \ldots, X_n)$.

Como $X \sim \text{EXP}(\lambda)$, se tiene que $T \sim \text{EXP}(n\lambda)$, pues

$$\mathbf{P}(X > t) = e^{-\lambda t} \quad \Longrightarrow \quad \mathbf{P}(T > t) = \mathbf{P}(X > t)^n = e^{-n\lambda t}.$$

Así que nT es un estimador insesgado de $1/\lambda$.

Varianza de un estimador

Para cualquier estimador es deseable que su varianza sea pequeña.

Si $\mathbf{V}(T)$ es pequeña, por Chebyshev tendremos que es poco probable que sus realizaciones se diferencien mucho de $\mathbf{E}(T)$.

Así que si además T es estimador insesgado, será muy probable que las estimaciones de θ estén próximas a θ.

- Si T_1 y T_2 son estimadores insesgados de θ, preferimos aquel que tenga menos varianza.
- ¿Y para estimadores sesgados?

Volvemos al EJEMPLO 4. $X \sim \text{UNIF}([0, a])$. Estimadores insesgados de a:

$$T_1 = \frac{n+1}{n} \max(X_1, X_2, \ldots, X_n) \quad \text{y} \quad T_2 = 2\overline{X}$$

Para T_2 tenemos que

$$\mathbf{V}(T_2) = 4\mathbf{V}(\overline{X}) = \frac{4}{n}\mathbf{V}(X) = \frac{4}{n}\frac{a^2}{12} = \frac{a^2}{3n}$$

Para T_1, recordemos que

$$f_{\max(X_1,\ldots,X_n)}(t) = n\,\frac{t^{n-1}}{a^n}, \quad \text{si } 0 \leq t \leq a.$$

Así que

$$\mathbf{E}(T_1^2) = \frac{(n+1)^2}{n(n+2)} a^2 \quad \Longrightarrow \quad \mathbf{V}(T_1) = \frac{1}{n(n+2)} a^2\,.$$

T_1 tiene menor varianza.

Error cuadrático medio

Sea $T = h(X_1, X_2, \ldots, X_n)$ un estimador del parámetro θ. Su **error cuadrático medio** es

$$\mathrm{ECM}(T) = \mathbf{E}\big((T-\theta)^2\big)$$

Ésta es la cantidad natural para medir cuán bueno es un estimador. Si $\mathrm{ECM}(T)$ es pequeño, es muy probable que las realizaciones de T estén próximas a θ.

Si T_1 y T_2 son estimadores del parámetro θ, T_1 es **más eficiente** que T_2 si

$$\mathrm{ECM}(T_1) < \mathrm{ECM}(T_2)\,.$$

Lema 1

$$\mathrm{ECM}(T) = \mathbf{V}(T) + \big(\mathbf{E}(T) - \theta)\big)^2 = \mathrm{varianza}(T) + \mathrm{sesgo}^2(T)\,.$$

Demostración. Sea $\mu = \mathbf{E}(T)$.

$$\begin{aligned}
\mathrm{ECM}(T) &= \mathbf{E}\big((T-\theta)^2\big) \\
&= \mathbf{E}\Big(\big((T-\mu) + (\mu-\theta)\big)^2\Big) \\
&= \mathbf{E}\big((T-\mu)^2\big) + 2(\mu-\theta)\mathbf{E}(T-\mu) + (\mu-\theta)^2 \\
&= \mathbf{E}\big((T-\mu)^2\big) + (\mu-\theta)^2\,,
\end{aligned}$$

- El lema 1 nos dice, en particular, que puede ser preferible un estimador sesgado con poca varianza que un estimador insesgado con mucha varianza.
- Si T es un estimador de un parámetro θ, por Chebyshev

$$\mathbf{P}\big(|T-\theta| \geq \varepsilon\big) \leq \frac{\mathrm{ECM}(T)}{\varepsilon^2}\,.$$

  Es decir, que si $\mathrm{ECM}(T)$ es pequeño entonces es (relativamente) poco probable que las estimaciones de θ obtenidas con T yerren demasiado.

EJEMPLO. Sea $X \sim \mathcal{N}(\mu, \sigma^2)$.

• **Cuasivarianza muestral**. Tomamos S^2 como estimador de σ^2. Sabemos que es insesgado: $\mathbf{E}(S^2) = \sigma^2$.

Para calcular $\mathbf{V}(S^2)$,

- podemos apelar a la fórmula general (que involucra cuarto momento de X);
- o recordar que, en este caso, $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$.De manera que, si $Z \sim \chi^2_{n-1}$,

$$\mathbf{V}(S^2) = \mathbf{V}\Big(\frac{\sigma^2}{n-1}\,Z\Big) = \frac{\sigma^4}{(n-1)^2}\,\mathbf{V}(Z) = \frac{2}{n-1}\,\sigma^4.$$

Y por tanto

$$\mathrm{ECM}(S^2) = \frac{2}{n-1}\,\sigma^4\,.$$

• **Varianza muestral**. Tomamos ahora como estimador (sesgado) de σ^2 el estadístico

$$D^2 = \frac{1}{n} \sum_{j=1}^{n} (X_j - \overline{X})^2$$

Como $D^2 = S^2\,(n-1)/n$, tenemos

$$\mathbf{E}(D^2) = \frac{n-1}{n}\,\mathbf{E}(S^2) = \frac{n-1}{n}\,\sigma^2 = \sigma^2 \overbrace{-\frac{1}{n}\,\sigma^2}^{\text{sesgo}}$$

Pero

$$\mathbf{V}(D^2) = \frac{(n-1)^2}{n^2}\mathbf{V}(S^2) = \frac{(n-1)^2}{n^2}\,\frac{2\,\sigma^4}{n-1} = \frac{2(n-1)\,\sigma^4}{n^2}\,,$$

de manera que

$$\mathrm{ECM}(D^2) = \frac{\sigma^4}{n^2} + \frac{2(n-1)\,\sigma^4}{n^2} = \Big(\frac{2n-1}{n^2}\Big)\,\sigma^4\,.$$

Notación

La función de densidad (o la función de masa) de X depende de un cierto parámetro θ.

Escribiremos (tanto para densidades como para funciones de masa)

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

- El argumento x recorre los valores de X. El **soporte** de X, $\mathbf{sop}_\theta$ es el conjunto de valores x para los que $f(x; \theta) > 0$. Puede depender de θ.
- El segundo argumento recorre los posibles valores del parámetro $\theta \in \Theta$. Θ es el **espacio de parámetros**.

- $\text{UNIF}[0, b]$.

$$f(t; b) = \begin{cases} 1/b & \text{si } t \in (0, b)\,, \\ 0 & \text{si } t \notin (0, b)\,. \end{cases}$$

$b \in \Theta = (0, +\infty)$. $t \in \mathbf{sop}_b = [0, b]$.

- $\text{GEO}(q)$.

$$f(k; q) = q(1-q)^{k-1}.$$

$q \in \Theta = [0, 1]$. $k \in \mathbf{sop}_q = \mathbb{N}$.

- $\mathcal{N}(0, \sigma^2)$.

$$f(x; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/\sigma^2}\,.$$

$\sigma^2 \in \Theta = (0, +\infty)$. $x \in \mathbf{sop}_{\sigma^2} = \mathbb{R}$.

Funciones de densidad/masa de muestras aleatorias

X es variable aleatoria con función de densidad/de masa $f(x;\theta)$, para $\theta \in \Theta$ y $\mathbf{x} \in \mathbf{sop}_\theta$.

La función de densidad/masa (conjunta) $f(\mathbf{x};\theta)$ de la muestra aleatoria $\mathbb{X} = (X_1, \ldots, X_n)$ es

$$f(\mathbf{x};\theta) = f(x_1, x_2, \ldots, x_n;\theta) = \prod_{j=1}^{n} f(x_j;\theta)$$

para $\theta \in \Theta$ y $\mathbf{x} \in (\mathbf{sop}_\theta)^n$.

Fuera de $(\mathbf{sop}_\theta)^n$, se tiene $f(\mathbf{x};\theta) \equiv 0$.

Medias

Usaremos la notación $\mathbf{E}_\theta$ para indicar que calculamos esperanzas usando la función de densidad/masa $f(x;\theta)$.

Si X es continua, para una función $g(X)$ de la variable X, tenemos que

$$\mathbf{E}_\theta(g(X)) = \int_{\mathbb{R}} g(x)\, f(x;\theta)\, dx = \int_{\mathbf{sop}_\theta} g(x)\, f(x;\theta)\, dx\,,$$

Y para una función $h(\mathbb{X})$ (es decir, para un estadístico) se tiene

$$\begin{aligned}
\mathbf{E}_\theta(h(\mathbb{X})) &= \int_{\mathbb{R}^n} h(\mathbf{x}) \cdot f(\mathbf{x};\theta)\, d\mathbf{x} \\
&= \int_{\mathbb{R}^n} h(x_1, x_2, \ldots, x_n) \cdot f(x_1, x_2, \ldots, x_n;\theta)\, dx_1 dx_2 \cdots dx_n \\
&= \int_{\mathbf{sop}_\theta^n} h(x_1, x_2, \ldots, x_n) \cdot f(x_1, x_2, \ldots, x_n;\theta)\, dx_1 dx_2 \cdots dx_n\,.
\end{aligned}$$

Si X es discreta, se tiene para una función g

$$\mathbf{E}_\theta(g(X)) = \sum_{x \in \mathbf{sop}_\theta} g(x) f(x;\theta)\,.$$

Y para una función $h(\mathbb{X})$, estadístico de X, se tiene

$$\begin{aligned}\mathbf{E}_\theta(h(\mathbb{X})) &= \sum_{\mathbf{x} \in \mathbf{sop}_\theta^n} h(\mathbf{x})\ f(\mathbf{x};\theta)\\ &= \sum_{(x_1,\dots,x_n) \in \mathbf{sop}_\theta^n} h(x_1,\dots,x_n) \prod_{j=1}^{n} f(x_j;\theta)\,.\end{aligned}$$

¿Cómo “construir” estimadores?

En las ilustraciones vistas hasta ahora, los estimadores considerados han surgido de manera “natural”.

¿Algún método general? Dos ideas:

- Método de momentos.
- Máxima verosimilitud.

Método de momentos

La idea es natural y se basa en que

- la media $\overline{X}$ muestral debe parecerse a la media $\mathbf{E}_\theta(X)$
- la media $\mathbf{E}_\theta(X)$ es de hecho una función de θ.

"Por consiguiente":

- planteamos la ecuación

$$(\dagger) \quad \mathbf{E}_\theta(X) = \overline{X}\,,$$

- despejamos θ de la ecuación $(\dagger)$,
- ése es el estimador M_θ de θ.

Ejemplos

Ejemplo 1. $X \sim \text{Ber}(p)$.

Como $\mathbf{E}_p(X) = p$, la ecuación es

$$p = \overline{X}.$$

Así que el estimador es $M_p = \overline{X}$.

- El estimador es el estadístico que registra la proporción de unos en la muestra $(X_1, X_2, \dots, X_n)$;
- la estimación, dada la lista $(x_1, x_2, \dots, x_n)$ de ceros y unos, viene dada por la proporción de unos en esa lista.

Ejemplo 2. $X \sim \text{EXP}(\lambda)$.

Como $\mathbf{E}_\lambda(X) = 1/\lambda$, la ecuación es

$$\frac{1}{\lambda} = \overline{X}.$$

Así que el estimador es $M_\lambda = 1/\overline{X}$.

Ejemplo 3. $X \sim \text{UNIF}[0, a]$.

Como $\mathbf{E}_a(X) = a/2$, la ecuación es

$$\frac{a}{2} = \overline{X}.$$

Así que el estimador es $M_a = 2\overline{X}$.

Cuando hay varios parámetros, como en las variables normales, recurriremos a más momentos. Además de a $\mathbf{E}_\theta(X)$, podemos apelar a $\mathbf{E}_\theta(X^2)$ o a $\mathbf{V}_\theta(X)$ (o, incluso, si hiciera falta, a $\mathbf{E}_\theta(X^k)$ con $k \geq 3$).

Ejemplo 4. $\mathcal{N}(\mu, \sigma^2)$.

Como $\mathbf{E}_{\mu,\sigma^2}(X) = \mu$ y $\mathbf{E}_{\mu,\sigma^2}(X^2) = \mathbf{V}_{\mu,\sigma^2}(X) + \mathbf{E}_{\mu,\sigma^2}(X)^2 = \sigma^2 + \mu^2$, las ecuaciones son

$$\mu = \overline{X} \qquad \text{y} \qquad \sigma^2 + \mu^2 = \overline{X^2}.$$

Los estimadores por momentos son

$$M_\mu = \overline{X} \qquad \text{y} \qquad M_{\sigma^2} = D^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X})^2$$

Obsérvese que M_{σ^2} no es S^2.

A veces el primer momento no da información.

Ejemplo 5. Consideremos la siguiente función de densidad (triangular, simétrica con moda y media en 0): para $\theta \in \Theta = (0, +\infty)$,

$$f(x;\theta) = \begin{cases} \frac{1}{\theta}(1 - |x|/\theta)\,, & \text{si } |x| < \theta \\ 0\,, & \text{en otro caso} \end{cases}$$

En este caso $\mathbf{E}_\theta(X) = 0$, para todo $\theta \in (0, +\infty)$. De hecho, por simetría, todos los momentos impares son cero. Los momentos pares son

$$\mathbf{E}_\theta(X^{2k}) = \frac{1}{(2k+1)(k+1)}\theta^{2k}\,; \quad \text{en particular,} \quad \mathbf{E}_\theta(X^2) = \frac{\theta^2}{6}\,.$$

Por consiguiente,

$$M_\theta = \sqrt{6\overline{X^2}}\,.$$

Usando distintos momentos se pueden obtener estimadores alternativos.

Ejemplo 6. $X \sim \textsc{ray}(\theta)$

La función de densidad de $X \sim \textsc{Ray}(\theta)$, donde $\theta > 0$, viene dada por

$$f(x;\theta) = \begin{cases} \frac{x}{\theta^2} e^{-x^2/2\theta^2}, & \text{si } x > 0\,, \\ 0, & \text{si } x \leq 0\,. \end{cases}$$

Se tiene que

$$\mathbf{E}_\theta(X) = \sqrt{\pi/2}\,\theta\,, \qquad \mathbf{E}_\theta(X^2) = 2\theta^2$$

Esto nos da dos estimadores:

$$M_\theta = \sqrt{\frac{2}{\pi}}\,\overline{X} \qquad \text{y} \qquad M_\theta^\star = \sqrt{\frac{1}{2}\,\overline{X^2}}\,.$$