Estimadores

Kevin Piterman

23 de Junio, 2020

Resumen.
Idea: conocemos parte de la distribución salvo algunos parámetros. Queremos realizar observaciones que nos permitan "estimar" el parámetro desconocido.

$$\theta? \longrightarrow \text{Buscamos una v.a. } \underbrace{\hat{\theta}(X_1,\ldots,X_n)}_{\text{aleatorio}} \text{ tal que } \underbrace{\hat{\theta}(x_1,\ldots,x_n)}_{\text{número}} \approx \theta.$$

Pregunta: ¿qué quiere decir que $\hat{\theta} \approx \theta$?

- Pedir que se parezca numéricamente es muy fuerte.
- Como estamos en el mundo de las probabilidades, nos enfocamos en ver "la probabilidad que suceda algo".

La noción más acertada en este contexto es que se parezcan bajo probabilidad.

Definición. Sea $\hat{\theta}$ un estimador de un parámetro θ de una determinada distribución.

1. $\hat{\theta}$ es un estimador **consistente** si
$$\hat{\theta}(X_1,\ldots,X_n) \xrightarrow[n\to\infty]{P} \theta.$$
2. El **sesgo** de $\hat{\theta}$ es
$$b(\hat{\theta}) = \mathbb{E}_\theta(\hat{\theta}-\theta).$$
3. $\hat{\theta}$ es un estimador **insesgado** si
$$\mathbb{E}(\hat{\theta}(X_1,\ldots,X_n)) = \theta.$$
4. $\hat{\theta}$ es un estimador **asintóticamente insesgado** si
$$\lim_{n\to\infty} \mathbb{E}(\hat{\theta}(X_1,\ldots,X_n)) = \theta.$$

Tenemos dos tipos principales de estimadores: de **Momentos (Mom)** y de **Máxima Verosimilitud (EMV)**

Estimadores de Momentos: se construye el estimador despejando el parámetro a partir de los momentos y utilizando las estimaciones conocidas de los momentos (LGN). Receta:

1. Tenemos una muestra aleatoria $X_1,\ldots,X_n$ iid.
2. Considerar todos los momentos necesarios $\mathbb{E}(X)$, $\mathbb{E}(X^2)$, ..., $\mathbb{E}(X^k)$ que nos permitan despejar nuestro parámetro θ en función de $\mathbb{E}(X),\ldots,\mathbb{E}(X^k)$. O sea, $\theta = \Theta(\mathbb{E}(X),\ldots,\mathbb{E}(X^k))$.
3. Por LGN, $\mathbb{E}(X^i)$ se estima con $\sum_{j=1}^n \frac{X_j^i}{n}$.
4. Estimamos θ con
$$\hat{\theta}(X_1,\ldots,X_n) = \Theta\left(\sum_{j=1}^n \frac{X_j}{n}, \sum_{j=1}^n \frac{X_j^2}{n}, \ldots, \sum_{j=1}^n \frac{X_j^k}{n}\right).$$

Estimadores de Máxima Verosimilitud: se construye el estimador del parámetro maximizando la probabilidad de obtener la muestra dada. Se conoce la distribución salvo algunos parámetros. Receta:

1. Tenemos una muestra aleatoria $X_1, \ldots, X_n$ iid, con densidad puntual p_θ (discreta) o f_θ (continua).
2. Las densidades dependen de θ.
3. Fijamos valores observados $x = (x_1, \ldots, x_n)$.
4. Consideramos la función de verosimilitud $L(\theta, x) : \mathbb{R} \to \mathbb{R}$ dada por:

$$\text{(Caso discreto)} \quad L(\theta, x) = P(X_1 = x_1, \ldots, X_n = x_n) \overset{\text{iid}}{=} \prod_i p_\theta(x_i)$$

$$\text{(Caso continuo)} \quad L(\theta, x) = f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) \overset{\text{iid}}{=} \prod_i f_\theta(x_i).$$

5. El EMV es $\hat{\theta}(X_1, \ldots, X_n)$ tal que (para cada x fijo!!) $\hat{\theta}(x)$ maximiza la función $\theta \mapsto L(\theta, x)$.

Notación: $\overline{X} = \sum_{i=1}^n \frac{X_i}{n}$. Si queremos hacer énfasis en el n, notaremos $\overline{X_n}$.

	Momentos	EMV
Media	$\hat{\mu} = \sum_{i=1}^n \frac{X_i}{n} = \overline{X}$	depende de la distribución
Varianza	$\sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \to \hat{\sigma}^2 = \left(\sum_i \frac{X_i^2}{n}\right) - \left(\overline{X}\right)^2$	depende de la distribución
λ, $\mathcal{E}(\lambda)$	$\mathbb{E}(\mathcal{E}(\lambda)) = \frac{1}{\lambda} \to \hat{\lambda} = \left(\overline{X}\right)^{-1}$	$\hat{\lambda} = \left(\overline{X}\right)^{-1}$
μ, $N(\mu, \sigma^2)$	$\hat{\mu} = \overline{X}$	$\hat{\mu} = \overline{X}$
σ, $N(\mu, \sigma^2)$	$\hat{\sigma} = \sqrt{\sum_i \frac{X_i^2}{n} - \overline{X}^2}$	$\hat{\sigma} = \sqrt{\frac{\sum_i (X_i - \overline{X})^2}{n}}$
θ, $U(0, \theta)$	$\mathbb{E}(U(0, \theta)) = \frac{\theta}{2} \to \hat{\theta} = 2\overline{X}$	$\hat{\theta} = \max\limits_{i=1,\ldots,n} X_i.$
λ, $\mathcal{P}(\lambda)$	$\hat{\lambda} = \hat{\mu} = \overline{X}$	$\hat{\lambda} = \overline{X}$

Ejemplo 1. Hallar el estimador de momentos y el EMV para la probabilidad de éxito p de una distribución Binomial $Bi(N,p)$, con N conocido. Analizar consistencia y sesgo.

Solución. Calculemos primero el estimador de momentos para el parámetro p.

$$\mathbb{E}(Bi(N,p)) = Np.$$

Entonces despejamos p de la ecuación anterior y lo que nos queda es

$$\hat{p} = \frac{\overline{X}}{N}.$$

Por LGN, $\overline{X}$ tiende en probabilidad a la esperanza que es Np. Luego $\hat{p} = \frac{\overline{X}}{N}$ tiende en probabilidad a $\frac{Np}{N} = p$. Analizamos el sesgo de este estimador:

$$\mathbb{E}(\hat{p}) = \mathbb{E}\left(\frac{\overline{X}}{N}\right) = \frac{1}{N}\mathbb{E}(\overline{X}) = \frac{1}{N}\mathbb{E}(Bi(N,p)) = \frac{1}{N}Np = p.$$

Por lo tanto, este estimador es insesgado y consistente.

Buscamos el EMV. La función de verosimilitud en este caso es:

$$L(p,x) = \prod_i p_{Bi(N,p)}(x_i) = \prod_i \binom{N}{x_i} p^{x_i}(1-p)^{N-x_i} = g(N,x)p^{\sum_i x_i}(1-p)^{\sum_i N-x_i}.$$

Como $g(N,x)$ es constante (positiva) para p y queremos maximizar $L(p,x)$, esto es lo mismo que maximizar $\ln\left(\frac{L(p,x)}{g(N,x)}\right)$. Sea $y = \sum_i x_i$. Entonces nos queda que tenemos que maximizar la siguiente función:

$$h(p) = \ln\left(\frac{L(p,x)}{g(N,x)}\right) = \ln(p^y(1-p)^{nN-y}) = y\ln(p) + (nN-y)\ln(1-p).$$

Acá, n es el tamaño de la muestra. Derivamos la función $h(p)$ y buscamos ceros:

$$h'(p) = y\frac{1}{p} - (nN-y)\frac{1}{1-p} = 0$$

Pasamos sumando y multiplicamos todo por $1-p$.

$$y\frac{1-p}{p} = (nN-y).$$

$$y\left(\frac{1}{p} - 1\right) = nN - y$$

$$\left(\frac{y}{p} - y\right) = nN - y$$

$$\frac{y}{p} = nN$$

Por lo tanto, $p = \frac{y}{nN}$.

Tarea: chequear que para $p = 1, 0$ obtenemos un mínimo para $L(p,x)$, y no un máximo, por lo que podemos excluir estos valores de nuestro análisis.

En consecuencia, el EMV para p es:

$$\hat{p} = \frac{\sum_i X_i}{nN} = \frac{\overline{X}}{N}.$$

Notar que es el mismo estimador que el calculado con los momentos, por lo que comparte las mismas propiedades de consistencia y sesgo.

Ejemplo 2. Sea $X_1, \ldots, X_n$ una muestra aleatoria iid con densidad

$$f(x) = \frac{3}{\theta^3} x^2 I_{(0,\theta]}(x), \quad \theta > 0.$$

1. Hallar un estimador de θ basado en el primer momento.
2. Hallar un estimador de θ basado en el segundo momento.
3. Hallar el EMV de θ.
4. Analizar la consistencia de los tres estimadores (más adelante).

Solución. Hallemos el estimador de θ utilizando el primer momento. Para ello, calculo el primer momento. Sea X variable aleatoria con densidad f.

$$\mathbb{E}(X) = \int x f(x) dx = \int x \frac{3}{\theta^3} x^2 I_{(0,\theta]}(x) dx = \int_0^\theta x \frac{3}{\theta^3} x^2 dx = \frac{3}{\theta^3} \int_0^\theta x^3 dx = \frac{3}{\theta^3} \frac{x^4}{4} |_0^\theta = \frac{3}{\theta^3} \frac{\theta^4}{4} = \frac{3}{4}\theta$$

Despejamos θ de la ecuación anterior y nos queda $\theta = \frac{4}{3}\mathbb{E}(X)$. Por lo tanto, el estimador $\hat{\theta}_1$ basado en el primer momento es

$$\hat{\theta}_1 = \frac{4}{3}\overline{X}.$$

Hallemos el estimador de θ utilizando el segundo momento. Para ello, calculo el segundo momento.

$$\begin{aligned}
\mathbb{E}(X^2) &= \int x^2 f(x) dx = \int x^2 \frac{3}{\theta^3} x^2 I_{(0,\theta]}(x) dx = \int_0^\theta x^2 \frac{3}{\theta^3} x^2 dx \\
&= \frac{3}{\theta^3} \int_0^\theta x^4 dx = \frac{3}{\theta^3} \frac{x^5}{5} |_0^\theta = \frac{3}{\theta^3} \frac{\theta^5}{5} \\
&= \frac{3}{5}\theta^2.
\end{aligned}$$

Despejamos θ en función del segundo momento, y nos queda que

$$\theta = \sqrt{\frac{5}{3}\mathbb{E}(X^2)}.$$

Luego el estimador $\hat{\theta}_2$ basado en el segundo momento es

$$\hat{\theta}_2 = \sqrt{\frac{5}{3} \sum_i \frac{X_i^2}{n}}.$$

Hallemos ahora el EMV de θ. Para ello debemos maximizar la función de verosimilitud. Fijemos un observable $x = (x_1, \ldots, x_n)$.

$$L(\theta, x) = \prod_i f(x_i) = \prod_i \frac{3}{\theta^3} x_i^2 I_{(0,\theta]}(x_i) = \left(3^n \prod_i x_i^2\right) \theta^{-3n} \prod_i I_{(0,\theta]}(x_i).$$

Notemos que si $\theta < x_i$ para algún i, entonces la expresión anterior es 0 y encontramos un mínimo (no un máximo). Por otro lado, θ^{-3n} toma valores cada vez más grandes a medida que θ se acerca a 0 (por derecha). Esto nos dice que el valor máximo de $L(\theta, x)$ se va a alcanzar en el valor más próximo a 0 que sea mayor o igual a todos los x_i. Es decir, $\theta = \max_i x_i$. Luego el EMV es $\hat{\theta} = \max_i X_i$.

Para analizar la consistencia, debemos ver un poco de álgebra de límites de convergencia en probabilidad (esperar unas clases más).

Ejemplo 3. Analizar los estimadores de momento y de máxima verosimilitud de una uniforme $U(a, b)$ para los parámetros a, b si ninguno de los dos es conocido y si alguno lo es.

Sugerencia: $\mathbb{E}(U(a, b)) = \frac{a+b}{2}$, $V(U(a, b)) = \frac{1}{12}(b-a)^2$.

Solución. Comencemos primero con los estimadores de momento. Utilizo el primer y segundo momento para poder despejar a y b (como tengo dos incógnitas, debo tener al menos dos ecuaciones para poder despejarlas). Sea $\mu = \mathbb{E}(U(a, b))$ y sea $\sigma = \sqrt{V(U(a, b))}$. Como tenemos una expresión sencilla para σ y se puede escribir en términos del primer y segundo momento, vamos a utilizar la ecuación que involucra al desvío estándar para despejar a y b.

$$\mu = \frac{a+b}{2}, \quad \sigma = \frac{1}{\sqrt{12}}(b-a).$$

Ahora hacemos los despejes. De la ecuación de la media, obtenemos que $b = 2\mu - a$. En la ecuación del desvío σ, reemplazamos el valor de b y obtenemos que:

$$\sigma = \frac{1}{\sqrt{12}}(2\mu - a - a) = \frac{1}{\sqrt{12}}(2\mu - 2a) = \frac{1}{\sqrt{3}}(\mu - a)$$

Luego $a = \mu - \sqrt{3}\sigma$. Reemplazando en la expresión para b, tenemos que

$$b = 2\mu - a = 2\mu - \mu + \sqrt{3}\sigma = \mu + \sqrt{3}\sigma.$$

Recordemos que $\hat{\mu} = \overline{X}$ y que $\hat{\sigma} = \sqrt{\sum_i \frac{X_i^2}{n} - \overline{X}^2}$. Luego los estimadores de momento de a y b son:

$$\hat{a} = \overline{X} - \sqrt{3}\sqrt{\sum_i \frac{X_i^2}{n} - \overline{X}^2}$$

$$\hat{b} = \overline{X} + \sqrt{3}\sqrt{\sum_i \frac{X_i^2}{n} - \overline{X}^2}.$$

Notemos que estos estimadores funcionan tanto si no conocemos ninguno de los dos parámetros, como si conocemos alguno de ellos. En el último caso, también podríamos haber despejado el parámetro desconocido directamente del primer o del segundo momento sin necesidad de considerar las dos ecuaciones simultáneamente.

Ahora calculemos el EMV. Debemos maximizar la función de verosimilitud. Sea $x = (x_1, \ldots, x_n)$ y sea f la densidad de $U(a, b)$. Entonces

$$L(a, b, x) = \prod_i f(x_i) = \prod_i \frac{1}{b-a} I_{(a,b)}(x_i) = \frac{1}{(b-a)^n} \prod_i I_{(a,b)}(x_i).$$

Para maximizar la función anterior, necesito que la productoria de las indicadoras siempre sea 1 y que la dierencia $b - a$ sea lo más chica posible. Para la primea condición, debemos pedir que $a \leq \min_i x_i$ y que $b \geq \max_i x_i$. Luego el valor de $b - a$ se minimiza siempre que $a = \min_i x_i$ y $b = \max_i x_i$. Ver Figura 1.

Por otro lado, si conocemos a alguno de los dos parámetros (o sea conocemos a o conocemos b) entonces el estimador para el parámetro desconocido es el mismo que describimos recién, y la justificación es la misma: debemos achicar $b - a$ siempre que b y a estén en los límites correctos. Acá estoy asumiendo que cuando uno de los dos parámetros está fijo, no agarré un observable que me caiga por fuera del intervalo. Por ejemplo, si a está fijo entonces no elegí ningún x_i menor que a. Ídem con b.

En conclusión, el EMV para a y b son, respectivamente,

$$\hat{a} = \min_i X_i, \quad \hat{b} = \max_i X_i.$$

Estos estimadores siguen siendo los mismos si alguno de los parámetros es conocido.

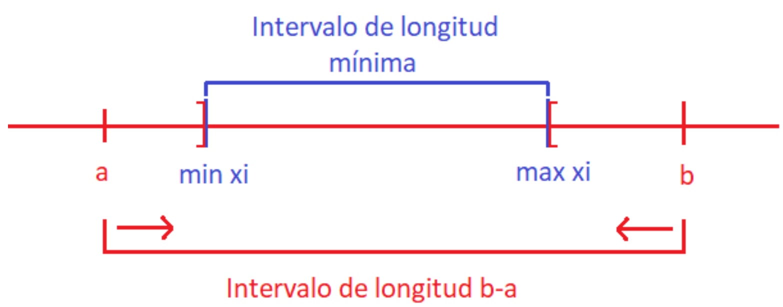

Figure 1: Esquema de la situación para minimizar $L(a, b, x)$. Notar que $b \geq \max_i x_i$ y $a \leq \min_i x_i$. Buscamos minimizar $b - a$ sujeto a estas condiciones.