Estimació n puntual

Ximena Fernández

Probabilidades y Estadística (Computació n)

25 de junio de 2020

Estimadores

$X \sim F_\theta$

Supongamos que tenemos una muestra aleatoria de una distribució n F, es decir

$$X_1, \ldots, X_n \sim F_\theta \text{ i.i.d}$$

Queremos estimar θ a partir de realizaciones de la muestra (observaciones)

$$x_1, \ldots, x_n.$$

Buscamos $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ una variable aleatoria tal que

$$\hat{\theta}(x_1, ..., x_n) \simeq \theta$$

momento k es $E(X^k) \xleftarrow{P} \frac{\sum_{i=1}^{n} X_i^k}{n}$

Estimador 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 $E(X), E(X^2), \ldots, E(X^k)$ que nos permitan despejar el parámetro θ en funció n de $E(X), \ldots, E(X^k)$. O sea, $\theta = \Theta(E(X), \ldots, E(X^k))$.
3. Por LGN, $E(X^j)$ se estima como

$$E(X^j) \overset{\sim}{=} \sum_{i=1}^{n} \frac{1}{n} X_i^j = \overline{X_n^j}.$$

4. Estimamos θ como

$$\hat{\theta}(X_1, \ldots, X_n) = \Theta(\overline{X_n}, \overline{X_n^2}, \ldots, \overline{X_n^k})$$

$X \sim F_\theta$ $X_1 . X_2 \cdots X_n$

Estimador de máxima verosimilitud

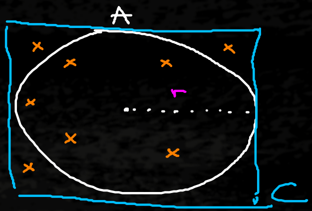
Se construye el estimador del parámetro maximizando la probabilidad de obtener la muestra dada.

Receta:

1. Tenemos una muestra aleatoria $X_1, \ldots, X_n$ iid, con densidad puntual p_θ (discreta) of f_θ (continua).
2. Fijamos valores observados $x = \underline{(x_1, \ldots, x_n)}$. fijos
3. Consideramos la funció n de verosimilitud $L(\theta, x) : \mathbb{R} \to \mathbb{R}$:
   - (Caso discreto)
     $L(\theta, x) = P(X_1 = x_1, \ldots, X_n = x_n) \overset{\text{iid}}{=} \prod_{i=1}^{n} p_\theta(x_i)$
   - (Caso continuo)
     $L(\theta, x) = f_{(X_1,\ldots,X_n)}(x_1, \ldots, x_n) \overset{\text{iid}}{=} \prod_{i=1}^{n} f_\theta(x_i)$.
4. El EMV es la variable aleatoria $\hat{\theta}(X_1, \ldots, X_n)$, que para cada realizació n $x, \hat{\theta}(x)$ maximiza la funció n

$$\theta \mapsto L(\theta, x).$$

Ejemplo



$Y \sim U_C$ tiro dardo

A = circulo de radio r

$X \sim Be(p)$ tiro adentro

$p = P(Y \in A)$

$X_1, X_2, \dots, X_n$ muestra

$$\frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow{P} E(X) = p = \frac{\text{Area } A}{\text{Area } C} = \frac{\pi \cdot r^2}{(2r)^2} = \frac{\pi}{4}$$

Queremos estimar es π.

$$\frac{1}{n}\sum_{i=1}^{n} X_i \underset{LGN}{\approx} \frac{\pi}{4} \Rightarrow \hat{\pi} = 4\,\frac{1}{n}\sum_{i=1}^{n} X_i$$

Propiedades de estimadores

Sea $\hat{\theta}$ un estimador de un parámetro θ de una determinada distribució n.

- $\hat{\theta}$ es un estimador consistente si

$$\hat{\theta}(X_1, X_2, \ldots, X_n) \xrightarrow[n\to\infty]{p} \theta$$

- $\hat{\theta}$ es un estimador insesgado si

$$E(\hat{\theta}(X_1, X_2, \ldots, X_n)) = \theta$$

- $\hat{\theta}$ es un estimador asintó ticamnte insesgado si

$$E(\hat{\theta}(X_1, X_2, \ldots, X_n)) \xrightarrow[n\to\infty]{} \theta$$

Error cuadrático medio

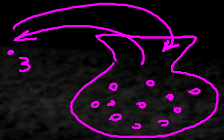
$$ECM(\hat{\theta},\theta) = E\left(\underbrace{(\hat{\theta}}_{\text{v.a}} - \underbrace{\theta}_{\text{número}})^2\right) = \underbrace{V(\hat{\theta})}_{\text{varianza}} + \Big(\underbrace{E(\hat{\theta}) - \theta}_{\text{bias}}\Big)^2$$

Propiedad: Si $ECM(\hat{\theta},\theta) \xrightarrow[n\to\infty]{} 0$ entonces

$$\theta(X_1, X_2, \ldots, X_n) \xrightarrow[n\to\infty]{p} \theta$$

$x_1 = 3$
$x_2 = 5$
$\vdots$
$x_n = 10$

Ejercicio 1: Estimado N



$N = ?$

Una urna contiene N bolillas numeradas de 1 a N. Con el objetivo de estimar el número N de bolillas, se extrae una bolilla, se observa su número y se la repone en la urna. Se repite este experimento n veces y se denota $X_1, X_2, \cdots, X_n$ a los números observados en cada extracció n.
Hallar los estimadores de momentos y de máxma verosimilitud de N. Para cada uno, decidir si es consistente, insesgado o asintó ticamente insesgado y calcular su ECM.

X v.a discreta

X = 'número que saco'

$rg(X) = \{1, 2, \ldots, N\}$

$P(X = j) = \dfrac{1}{N}$

$X_1, X_2, \ldots, X_n$ muestra aleatoria

$$\rightarrow \quad E(X) = \sum_{i=1}^{N} i \, \underline{P(X=i)} = \sum_{i=1}^{N} i \cdot \underline{\frac{1}{N}} = \frac{1}{\cancel{N}} \frac{\cancel{N}(N+1)}{2}$$

primer momento

$$\rightarrow \quad \frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow[\text{LGN}]{P} E(X)$$

$$\shortparallel$$

$$\bar{X}_n$$

$$\boxed{\begin{aligned} E(X^2) &= \sum i^2 P(X=i) \\ &= \sum_{i=1}^{N} i^2 \frac{1}{N} = \\ &= \frac{1}{\cancel{N}} \cdot \frac{\cancel{N}(N+1)(2N+1)}{6} \end{aligned}}$$

Estimamos

$$\bar{X}_n \overset{\text{LGN}}{\approx} \frac{(N+1)}{2}$$

$$\hat{N} = 2\bar{X}_n - 1$$

Consistencia

$$\hat{N} = 2\bar{X}_n - 1 \xrightarrow[n\to+\infty]{P} 2E(X) - 1 = \cancel{2}\cdot\frac{(N+1)}{\cancel{2}} - 1 = N$$

LGN + Prop

Sesgo: $E(\hat{N}) - N = E(2\bar{X}_n - 1) - N =$

$$= 2\underbrace{E(\bar{X}_n)}_{= E(X)} - 1 - N = \cancel{2}\frac{(N+1)}{\cancel{2}} - 1 - N$$

$= 0$ ☺ $\boxed{\text{insesgado}}$

Varianza: $V(\hat{N}) = V(2\bar{X}_n - 1) = 4V(\bar{X}_n) =$

$$= 4\frac{V(X)}{n} = \frac{4}{n}\cdot\left[\frac{(N+1)(2N+1)}{6} - \left(\frac{N+1}{2}\right)^2\right]$$

$\boxed{ECM}$ = Sesgo + varianza $= 0 + 4\frac{V(X)}{n} \xrightarrow[n\to+\infty]{} 0$

$\boxed{\text{EMV}}$ $X_1 X_2 \dots X_n$ iva iid

$x = x_1, x_2 \dots x_n$ observaciones fijo

$$L_x(N) = P(X_1 = x_1, X_2 = x_2 \dots X_n = x_n) =$$

$n \leftarrow$ fijo

$$= \prod_{i=1}^{n} \underbrace{P(X_i = x_i)}_{\frac{1}{N}} = \left(\frac{1}{N}\right)^n$$

$\boxed{N \geq x_i}$ $\forall i$

$\Rightarrow N \geq \max X_i$

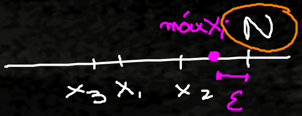
Busco N que maximiza $L_x(N)$

$$P(X_i = x_i) = \begin{cases} 1/N & \text{si } 1 \leq x_i \leq N \\ 0 & \text{si no} \end{cases}$$

$$\boxed{\hat{N} = \max(X_i)}$$

Consistencia

$$¿\ \underbrace{\max X_i}_{\hat{N}} \xrightarrow{P} N\ ?$$



Sea $\varepsilon \geq 0$ $\quad$ $\varepsilon \leq 1$

$$P(|\hat{N} - N| \geq \varepsilon) =$$

$$= P(\max X_i \leq N-1) =$$

$$= P(X_1 \leq N-1 \wedge X_2 \leq N-1 \wedge \dots \wedge X_n \leq N-1) =$$

$$\overset{iid}{=} P(X \leq N-1)^n = \left(\frac{N-1}{\underbrace{N}_{\text{número fijo}}}\right)^{n} \xrightarrow[n \to +\infty]{} 0$$

$$\boxed{\begin{array}{l} \varepsilon' > 1 \geq \varepsilon \\ 0 \leq P(|\hat{N} - N| \geq \varepsilon') \leq \\ \quad P(|\hat{N} - N| \geq \varepsilon) \end{array}}$$

Tarea: Sesgo y varianza

Probar que asintóticamente insesgado

Ejercicio 2: Seamos consistentes

a) Sea $X_1, \cdots, X_n$ una muestra aleatoria con

$$X_i \sim U(-\theta, \theta), \ \ \theta > 0.$$

Hallar el estimador de momentos de θ y probar que es consistente.

b) Sea $X_1, \cdots, X_n$ una muestra aleatoria con densidad

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

Hallar el EMV de θ y decidir si es consistente.

a) $X \sim U(-\theta, \theta)$ $\qquad X_1\ X_2 \ldots X_n$ muestra aletoria

$E(X) = 0$!

$\underset{\overline{X}_n}{\overset{\text{LGN}}{\rightsquigarrow}}$

$$E(X^2) = V(X) = \frac{(2\theta)^2}{12} = \frac{\theta^2}{3}$$

Obs:

$$V(X) = E(X^2) - \underbrace{E(X)^2}_{=0}$$

ss LGN

$$\frac{\sum_{i=1}^{n} X_i^2}{n}$$

$b = \theta$

$a = -\theta$

Estimo $\dfrac{\hat{\theta}^2}{3} = \dfrac{\sum_{i=1}^{n} X_i^2}{n} = \overline{X_n^2}$

$$\Rightarrow \boxed{\hat{\theta} = \sqrt{3\,\overline{X_n^2}}}$$

$X_1 \ldots X_n$ iid

$\Downarrow$

$X_1^2 \ldots X_n^2$ iid

Por LGN $\overline{X_n^2} \longrightarrow E(X^2)$

Consistencia

$$\hat{\theta}_n = \sqrt{3\,\overline{X_n^2}} \xrightarrow{P} \sqrt{3 \cdot \overset{\theta^2/3}{E(X^2)}} = \theta$$

EMV

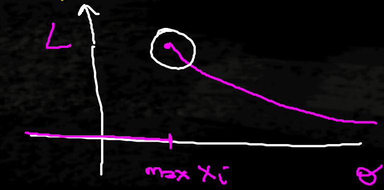
$$f_\theta(x) = \frac{3}{\theta^3} x^2 \, I_{(0,\theta]}(x) \qquad \theta > 0$$

$x = x_1, \dots, x_n$ observaciones

$$L_x(\theta) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} \frac{3}{\theta^3} x_i^2 \, I_{(0,\theta]}(x_i)$$

$$= \frac{3^n}{\theta^{3n}} \underbrace{\prod_{i=1}^{n} x_i^2}_{\text{cte}} \cdot \underbrace{\prod_{i=1}^{n} I_{(0,\theta]}(x_i)}_{=1 \text{ si } x_i \leq \theta \ \forall i}$$

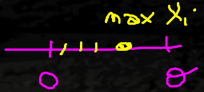
$\theta \geq \max x_i$



$$L_x'(\theta) = 3^n \cdot \underbrace{\prod_{i=1}^{n} x_i \, (-3n)\, \theta^{-3n-1}}_{<0}$$

EMV $\hat{\theta} = \text{máx}\, X_i$

$$\hat{\theta} = \max X_i \xrightarrow{P} \theta \ ?$$



$$P\Big(\big|\overbrace{\max_{1\le i\le n} X_i - \theta}^{<0}\big| \geq \varepsilon\Big) =$$

$$= P\left(\theta - \max_{1\le i\le n} X_i \geq \varepsilon\right) =$$

$$= P\left(\max X_i \leq \theta - \varepsilon\right) =$$

$$= P\left(X_1 \leq \theta-\varepsilon,\ X_2 \leq \theta - \varepsilon,\ \dots,\ X_n \leq \theta-\varepsilon\right)$$

$$= P\left(X \leq \theta - \varepsilon\right)^n = \underbrace{F_X(\theta-\varepsilon)}_{<1}{}^n$$

$$= \left(\frac{\theta-\varepsilon}{\theta}\right)^{3n} \longrightarrow 0 \Rightarrow \hat{\theta}_n \xrightarrow{P} \theta \text{ es consistente}$$

$$P(X \leq \theta - \varepsilon) = \int_{-\infty}^{\theta-\varepsilon} \frac{3}{\theta^3} x^2 I_{(0,\theta]}(x)\,dx$$

$$= \int_{0}^{\theta-\varepsilon} \frac{3}{\theta^3} x^2\,dx =$$

$\varepsilon < \theta$

$$= \frac{\cancel{3}}{\theta^3}\frac{x^3}{\cancel{3}}\Big|_0^{\theta-\varepsilon} = \left(\frac{\theta-\varepsilon}{\theta}\right)^3$$