Estimación

X variable aleatoria

conocemos la distribución de X a menos de un parámetro Θ.

$X_1, X_2, \dots, X_n$ es una muestra aleatoria simple (MAS) si son iid $\sim X$

un estimador de Θ es una sucesión de variables aleatorias $\hat{\Theta}_n$ donde $\hat{\Theta}_n$ depende únicamente de $X_1, \dots, X_n$

$\hat{\Theta}_n$ es un estimador consistente de Θ si $\hat{\Theta}_n \xrightarrow{cs} \Theta$

ejemplo: $\overline{X_n}$ es un estimador consistente de $E(X)$

* $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ depende únicamente de $X_1, \dots, X_n$
* LFGN: $\overline{X}_n \xrightarrow{cs} E(X)$

podemos enunciar la LFGN como: el promedio muestral es un estimador consistente de la esperanza.

Ejercicio 1
Sean $(X_n)_{n\in\mathbb{N}}$ *iid* tales que $\mathbf{E}(X_1) = \mu$ y $\mathbf{Var}(X_1) = \sigma^2 < \infty$ $(\sigma > 0)$.

1. Demostrar que $\overline{X}_n$ es un estimador consistente de μ, esto es que $\overline{X}_n \xrightarrow[n]{c.s.} \mu$.

2. Demostrar que si $\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ y $s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$ entonces

$$\sigma_n^2 \xrightarrow[n]{c.s.} \sigma^2 \qquad s_n^2 \xrightarrow[n]{c.s.} \sigma^2 \qquad \sigma_n \xrightarrow[n]{c.s.} \sigma \qquad s_n \xrightarrow[n]{c.s.} \sigma$$

$Var(X_1) = \sigma^2$

$Var(X_1) = E(X_1^2) - E(X_1)^2$

$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$

a) queremos ver que $\sigma_n^2 \xrightarrow{cs} \sigma^2$

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2)$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 - \sum_{i=1}^{n} 2X_i\overline{X}_n + \sum_{i=1}^{n}\overline{X}_n^2\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\sum_{i=1}^{n} 2X_i\overline{X}_n + \frac{1}{n}\sum_{i=1}^{n}\overline{X}_n^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}_n \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{\overline{X}_n} + \frac{1}{\cancel{n}}\cancel{n}\overline{X}_n^2$$

$X \xrightarrow{cs} a$

g función continua

$g(X) \xrightarrow{cs} g(a)$

$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}_n^2 + \overline{X}_n^2$

$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2 \xrightarrow{cs} \sigma^2 = E(X_1^2) - E(X)^2$

* LFGN: $\overline{X}_n \xrightarrow{cs} E(X_1)$

entonces $\overline{X}_n^2 \xrightarrow{cs} E(X)^2$ porque $g(x) = x^2$ es continua

* sea $Y_i = X_i^2$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{cs} E(Y_1) = E(X_1^2)$$

entonces tenemos:

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2 \xrightarrow{cs} E(X_1^2) - E(X_1)^2 = \sigma^2$$

b) queremos ver que $\sigma_n \xrightarrow{cs} \sigma$

$$\left.\begin{array}{l} \sigma_n^2 \xrightarrow{cs} \sigma^2 \\ \sqrt{\ } \text{ es una función continua} \end{array}\right\} \Rightarrow \sqrt{\sigma_n^2} \xrightarrow{cs} \sqrt{\sigma^2}$$

$$\Rightarrow \sigma_n \xrightarrow{cs} \sigma$$

c) $\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \qquad \sigma_n^2 \xrightarrow{cs} \sigma^2$

$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$

queremos ver que $s_n^2 \xrightarrow{cs} \sigma^2$

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{n}{n-1}\,\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{n}{n-1}\sigma_n^2$$

$$s_n^2 = \underbrace{\frac{n}{n-1}}_{\downarrow\, 1}\,\underbrace{\sigma_n^2}_{\xrightarrow{cs}\sigma^2} \xrightarrow{cs} \sigma^2$$

$$\frac{n}{n-1} = \frac{1}{1-1/n} \xrightarrow[n\to\infty]{} 1$$

Método de los momentos

$X_1, X_2, \dots, X_n$ MAS de X

LFGN: $\overline{X_n} \xrightarrow{cs} E(X)$

$\overline{X_n^2} \xrightarrow{cs} E(X^2)$

$\overline{X_n^3} \xrightarrow{cs} E(X^3)$

$\vdots$

$\overline{X_n^k} \xrightarrow{cs} E(X^k)$

$\overline{X_n}^2 \xrightarrow{cs} E(X)^2$

$\overline{X_n^2} \xrightarrow{cs} E(X^2)$

$\overline{X_n^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{cs} E(Y_1) = E(X_1^2)$

$Y_i = X_i^2$

$\overline{X_n}^2 \neq \overline{X_n^2}$

idea: para estimar θ lo despejamos a partir de $E(X), E(X^2), \dots, E(X^k)$

Ejercicio 2

Sean $X_1, X_2, \dots X_n \, iid \sim F$. Encontrar estimadores para los siguientes parámetros por el método de los momentos:

1. p si la distribución es Ber(p)
2. λ si la distribución es $\mathcal{P}(\lambda)$
3. p si la distribución es Geo(p)
4. μ y σ^2 si la distribución es $N(\mu, \sigma^2)$
5. a y b si la distribución es $\mathcal{U}[a, b]$.

2. estimador de λ para $P(\lambda)$ $\quad X \sim P(\lambda) \quad E(X) = \lambda$

$X_1, \dots, X_n$ MAS de X

$\overline{X_n} \xrightarrow{cs} E(X) = \lambda$

$\overline{X_n} \xrightarrow{cs} \lambda$

el estimador de λ es $\hat{\lambda}_n = \overline{X_n}$

3. $X \sim Geo(p) \quad E(X) = \frac{1}{p}$

$X_1, \dots, X_n$ MAS de X $\quad$ estimador de p?

$\overline{X_n} \xrightarrow{cs} \underbrace{E(X)}_{=1/p}$

$\overline{X_n} \xrightarrow{cs} \frac{1}{p}$

$\frac{1}{\overline{X_n}} \xrightarrow{cs} p$

estimador de p: $\hat{p}_n = \frac{1}{\overline{X_n}}$

4. $X \sim \mathcal{N}(\mu, \sigma^2)$ $\quad E(X) = \mu$, $Var(X) = \sigma^2$

$X_1, \dots, X_n$ MAS de X

$$\overline{X_n} \xrightarrow{cs} \underbrace{E(X)}_{=\mu}$$

estimador de μ: $\hat{\mu}_n = \overline{X_n}$

$$\overline{X_n} \xrightarrow{cs} E(X)$$

$$\overline{X_n^2} \xrightarrow{cs} E(X^2)$$

$$\sigma^2 = Var(X) = E(X^2) - E(X)^2$$

estimador de σ^2: $\hat{\sigma}_n^2 = \overline{X_n^2} - \overline{X_n}^2$

5. $X \sim \mathcal{U}[a,b]$ $\quad E(X) = \frac{a+b}{2}$ $\quad Var(X) = \frac{(b-a)^2}{12}$

$X_1, \dots, X_n$ MAS de X $\quad \hookrightarrow Var(X) = \underbrace{E(X^2)} - E(X)^2$

estimadores de a y b

$$\overline{X_n} \xrightarrow{cs} E(X) = \frac{a+b}{2}$$

$$\overline{X_n^2} \xrightarrow{cs} E(X^2) = Var(X) + E(X)^2$$

$$= \frac{(b-a)^2}{12} + \frac{(a+b)^2}{4}$$

$$= \frac{(b-a)^2 + 3(a+b)^2}{12}$$

$$= \frac{b^2 - 2ab + a^2 + 3a^2 + 6ab + 3b^2}{12}$$

$$= \frac{4b^2 + 4a^2 + 4ab}{12}$$

$$= \frac{b^2 + a^2 + ab}{3}$$

queremos escribir a y b en función de $E(X)$ y $E(X^2)$:

$$E(X) = \frac{a+b}{2} \Rightarrow a = 2E(X) - b$$

$$E(X^2) = \frac{b^2+a^2+ab}{3} \Rightarrow 3E(X^2) = b^2 + a^2 + ab$$

$$= b^2 + (2E(X)-b)^2 + (2E(X)-b)b$$

$$\Rightarrow 0 = b^2 + 4E(X)^2 - 4bE(X) + \cancel{b^2} + 2bE(X) - \cancel{b^2} - 3E(X^2)$$

$$0 = b^2 - 2bE(X) + 4E(X)^2 - 3E(X^2)$$

obtendriemos estimadores de a y b resolviendo:

$$\begin{cases} \hat{a} = 2\overline{X}_n - \hat{b} \\ \hat{b}^2 - 2\overline{X}_n\hat{b} + 4\overline{X}_n^2 - 3\overline{X_n^2} = 0 \end{cases}$$

$$\Rightarrow \hat{b} = \frac{2\overline{X}_n \pm \sqrt{4\overline{X}_n^2 - 4(4\overline{X}_n^2 - 3\overline{X_n^2})}}{2}$$

$$= \overline{X}_n \pm \sqrt{\overline{X}_n^2 - (4\overline{X}_n^2 - 3\overline{X_n^2})}$$

$$= \overline{X}_n \pm \sqrt{3\overline{X_n^2} - 3\overline{X}_n^2}$$

$$\hat{b} = \overline{X}_n \pm \sqrt{3}\sqrt{\overline{X_n^2} - \overline{X}_n^2}$$

$$\hat{a} = 2\overline{X}_n - \hat{b}$$

$$= 2\overline{X}_n - (\overline{X}_n \pm \sqrt{3}\sqrt{\overline{X_n^2} - \overline{X}_n^2})$$

$$= \overline{X}_n \mp \sqrt{3}\sqrt{\overline{X_n^2} - \overline{X}_n^2}$$

como $a < b$: $\hat{b}_n = \overline{X}_n + \sqrt{3}\sqrt{\overline{X_n^2} - \overline{X}_n^2}$

$$\hat{a}_n = \overline{X}_n - \sqrt{3}\sqrt{\overline{X_n^2} - \overline{X}_n^2}$$

Ejercicio 3

Una pieza de una máquina se verifica al final de cada hora de producción y se cambia por una nueva en caso de encontrarse rota. El tiempo de vida en horas de la pieza se puede modelar con una variable aleatoria T con distribución exponencial de parámetro λ ($T \sim \exp(\lambda)$), por lo tanto el tiempo en horas que transcurre hasta el recambio de la pieza se puede modelar con una variable aleatoria $X = [T] + 1$, donde $[T]$ es la parte entera de T (esto es, $X = n$ si y sólo si $n - 1 \leq T < n$).

1. Hallar la función de probabilidad de la variable aleatoria X y probar que tiene distribución geométrica de parámetro $1 - e^{-\lambda}$ ($X \sim \text{Geo}(1 - e^{-\lambda})$).

2. A partir de los tiempos en los que se realiza el recambio de las piezas se desea estimar el parámetro λ del tiempo de vida de dichas piezas.

   a) Calcular λ en función de μ siendo $\mu = \mathbf{E}(X)$.

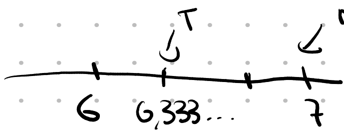
   b) ¿Cómo estimaría μ a partir de las observaciones $X_1, X_2, \ldots, X_n$ de los tiempos de recambio de las piezas?

   c) Construir un estimador consistente para λ en función de las observaciones $X_1, X_2, \ldots, X_n$ de los tiempos de recambio de las piezas.

tiempo de vida de la pieza: $T \sim \exp(\lambda)$ $\quad F_T(x) = P(T \leq x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si no} \end{cases}$

tiempo en horas que transcurre hasta el recambio:

$$X = [T] + 1$$



nos damos cuenta que está rota

$X = n$ sii $n-1 \leq T < n$

① $p_X(k) = P(X = k) = P(k-1 \leq T < k)$

$k \geq 0$

$$= F_T(k) - F_T(k-1)$$
$$= 1 - e^{-\lambda k} - (1 - e^{-\lambda(k-1)})$$
$$= -e^{-\lambda k} + e^{-\lambda(k-1)}$$
$$= e^{-\lambda(k-1)} - \underbrace{e^{-\lambda k}}_{= e^{-\lambda(k-1)} \cdot e^{-\lambda}}$$
$$= e^{-\lambda(k-1)}(1 - e^{-\lambda})$$
$$= (1 - (1 - e^{-\lambda}))^{k-1}(1 - e^{-\lambda})$$

queremos ver que

$X \sim \text{Geo}(1 - e^{-\lambda})$

$Y \sim \text{Geo}(p)$

$p_Y(k) = (1-p)^{k-1} p$

$\to p_X(k) = \underbrace{(1 - (1 - e^{-\lambda}))^{k-1}}_{(e^{-\lambda})^{k-1} = e^{-\lambda(k-1)}} (1 - e^{-\lambda})$

$\Rightarrow X \sim \text{Geo}(1 - e^{-\lambda})$