Estimación

tenemos X va

no conocemos la distribución

$X_1, \ldots, X_n$ 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 $\{X_1, \ldots, X_n\}$

ej: $\overline{X}_n$ es un estimador de $E(X)$

$\hat{\theta}_n$ es un estimador consistente si $\hat{\theta}_n \xrightarrow{cs} \theta$

ej: $\overline{X}_n$ es un estimador consistente de la esperanza

LFGN: el promedio muestral es un estimador consistente de la esperanza de $E(X)$

dada X va y $X_1, \ldots, X_n$ una MAS de X tenemos que $\overline{X}_n \xrightarrow{cs} E(X)$

Ejercicio 1

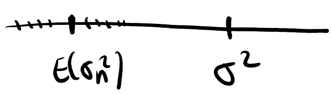
Sean $(X_n)_{n\in\mathbb{N}}$ *iid* tales que $\mathbf{E}(X_1) = \mu$ y $\mathbf{Var}(X_1) = \sigma^2 < \infty$ $(\sigma > 0)$.

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

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

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

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



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

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

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

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

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

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

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

$$LFGN: \frac{1}{n}\sum_{i=1}^{n}X_i = \overline{X}_n \xrightarrow{cs} E(X)$$

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

$\sigma^2 = E(S_n^2)$

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

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

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

Método de los momentos

$\{X_i\}_{i\in\mathbb{N}}$ sucesión de va iid $\sim X$

LFGN: $\overline{X_n} \xrightarrow{cs} E(X)$ $\qquad$ $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{cs} E(X)$

$\overline{X_n^2} \xrightarrow{cs} E(X^2)$ $\qquad$ $\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i^2}_{\overline{X_n^2}} \xrightarrow{cs} E(X^2)$

$\vdots$

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

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 $\mathcal{U}[a, b]$.

1) $\{X_i\}_{i\in\mathbb{N}}$ iid $\sim X$ $\qquad$ $X \sim Ber(p)$ $\qquad$ $E(X) = p$

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

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

3) $\{X_i\}_{i\in\mathbb{N}}$ iid $\sim X$ $\qquad$ $X \sim Geo(p)$ $\qquad$ $E(X) = \frac{1}{p} \rightsquigarrow p = \frac{1}{E(X)}$

$$\overline{X_n} \xrightarrow{cs} E(X) = \frac{1}{p}$$

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

4) $\{X_i\}_{i\in\mathbb{N}}$ iid $\sim X$ $\quad X\sim \mathcal{N}(\mu,\sigma^2)$ $\quad E(X)=\mu$

$Var(X)=\sigma^2$

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

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

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

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

estimador de σ^2: $\boxed{\sigma_n^2 = \overline{X_n^2} - \overline{X_n}^2}$

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

T= el tiempo de vida en horas de la pieza $\quad T\sim \exp(\lambda)$

X = tiempo en horas hasta que se cambia la pieza

si la pieza se rompe a las 3,5 horas $T=3,5$

nos damos cuenta que está rota a las 4 horas $\quad X=4$

$X = \lfloor T \rfloor + 1$

$X=n$ si y solamente si $n-1 \leq T < n$

$$F_T(x) = \begin{cases} 1-e^{-\lambda x} & \text{si } x\geq 0 \\ 0 & \text{si no} \end{cases}$$

1) buscamos la fpp de X

$$p_X(k) = P(X=k) = P(k-1 \leq T < k)$$
$$= F_T(k) - F_T(k-1)$$
$$= (1-e^{-\lambda k}) - (1-e^{-\lambda(k-1)})$$
$$= e^{-\lambda(k-1)} - e^{-\lambda k}$$

queremos ver que $X \sim Geo(1-e^{-\lambda})$

fpp de una $Geo(p)$: $(1-p)^{k-1} p$

fpp de $Geo(1-e^{-\lambda})$: $(1-(1-e^{-\lambda}))^{k-1}(1-e^{-\lambda}) = (e^{-\lambda})^{k-1}(1-e^{-\lambda})$

$p_X(k) = e^{-\lambda(k-1)} - e^{-\lambda k} = e^{-\lambda(k-1)}(1-e^{-\lambda})$ esto es la fpp de de $Geo(1-e^{-\lambda})$

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

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.

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

queremos escribir λ en función de $E(X)$

$$E(X) = \frac{1}{1-e^{-\lambda}} \Rightarrow \frac{1}{E(X)} = 1-e^{-\lambda}$$

$$\Rightarrow e^{-\lambda} = 1 - \frac{1}{E(X)}$$

$E(x)$

$$\Rightarrow \; -\lambda = \log\left(1 - \frac{1}{E(X)}\right)$$

$$\Rightarrow \; \lambda = -\log\left(1 - \frac{1}{E(X)}\right)$$

b) $\overline{X}_n \xrightarrow{cs} E(X)$

estimador de $E(X)$: $\overline{X}_n$

c) estimador de λ : $\hat{\lambda}_n = -\log\left(1 - \frac{1}{\overline{X}_n}\right)$

Estimación por maxima verosimilitud

$X_1, \dots, X_n$ MAS de X $\quad$ $X_1, \dots, X_n$ iid $\sim X$

observamos $(x_1, x_2, \dots, x_n)$

buscamos que maximice la probabilidad de observar los datos $(x_1, \dots, x_n)$

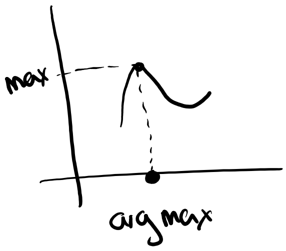
Caso discreto:

$X_1, \dots, X_n$ MAS de X $\quad$ X tiene fpp $P_X(x, \theta) = P(X = x)$

función de verosimilitud:

$L(\Theta, x_1, \dots, x_n)$ = probabilidad de observar los datos $(x_1, \dots, x_n)$

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



$$= P(X_1 = x_1) P(X_2 = x_2) \cdots P(X_n = x_n)$$

$$= \prod_{i=1}^{n} P(X_i = x_i)$$

$$= \prod_{i=1}^{n} p_X(x_i, \Theta)$$

el estimador por maxima verosimilitud de Θ es

$$\hat{\Theta} = \arg\max L(\Theta, x_1, \dots, x_n)$$

en la practica vamos a buscar donde se maximiza

$$h(\Theta, x_1, \dots, x_n) = \log L(\Theta, x_1, \dots, x_n)$$