Clase 29/05/2023.

En la clase anterior vimos la definición de estadístico y estimador. Podemos entonces repasarla:
Sea $X_1, X_2, \dots, X_n$ una muestra aleatoria simple de una v.a. X con distribución F_X. Esta expresión m.a.s. significa que las v.a. $\{X_j\}_{j=1}^{n}$ son independientes y poseen la misma distribución F_X.
Definimos entonces un estadístico a una función de $(X_1, X_2, \dots, X_n)$: $T(X_1, \dots, X_n)$ es decir una función de la m.a.s.
Un estimador es un estadístico que nos otorga una información de un parámetro asociado a la distribución F_X.

De esta manera nuestro modelo será una función de distribución que dependerá de un parámetro θ que puede ser escalar o vectorial. Cuando el parámetro es escalar diremos que $F_X(\cdot, \theta)$ es una familia uniparamétrica.
Si $\vec{\theta} \in \mathbb{R}^d$ esto es $\vec{\theta} = (\theta_1, \theta_2, \dots, \theta_d)$ en este caso el modelo será $F_X(\cdot, \vec{\theta})$ multiparamétrica.
Ejemplos de estas situaciones son

1.- $B_1, B_2, \dots, B_n$ son Bernoullis con parámetro p y sabemos que cada $B_j \overset{d}{\sim}$ Bernoulli (p)

2.- También podemos tener $P_1, P_2, \dots, P_n$ una

Una m.a.s. de una Poisson con parámetro λ; es decir la función de probabilidad

$$\mathbb{P}(P=k)=e^{-\lambda}\frac{\lambda^k}{k!}$$

3.- Podemos tener $U_1, U_2, \ldots, U_n$ una muestra aleatoria simple de una Uniforme en el intervalo $[0,\theta]$ y la densidad en este caso es

$$U_\theta(x)=\begin{cases}\frac{1}{\theta} & 0\le x\le\theta\\ 0 & \text{en otra parte.}\end{cases}$$

El parámetro aquí es el número positivo. θ.

4.- También podemos tener una muestra de una exponencial con parámetro λ. Esto es $E_1, E_2, \ldots, E_n$ son todas exponenciales con función de densidad

$$e(x)=\begin{cases}\lambda e^{-\lambda x} & x\ge 0\\ 0 & \text{en otra parte}\end{cases}$$

y el parámetro es el valor $\lambda>0$.

5.- Por último podemos considerar el modelo de una $N(\mu,\sigma^2)$. Esto es $Z_1, Z_2, \ldots, Z_n$ es una m.a.s. cada una de las Z tiene densidad

$$g_Z(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

En este caso tenemos dos parámetros $\underline{(\mu,\sigma^2)}$.

En la clase pasada diseñamos un primer método de estimación que llamamos el "Método de los momentos" y que se basaba en la LGN. Sabemos que si $X_1, X_2, \ldots, X_n$ es una muestra aleatoria simple de una X con distribución F_X y sabemos que $E(|X|) < \infty$ entonces

$$\boxed{\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} \xrightarrow{P} E(X) = \mu.}$$

Además si $E(|X|^k) < +\infty$ entonces también se tiene que

$$\bar{m}_r = \frac{X_1^r + \cdots + X_n^r}{n} \xrightarrow{P} E(X^r) = m_r$$

El método de los momentos consiste en igualar los momentos empíricos a los verdaderos momentos y al despejar encontrar $\hat{\theta}$ el estimador del parámetro.

Ejemplo: Si $P_1, P_2, \ldots, P_n$ es una m.a.s de una Poisson de parámetro λ. Entonces

$$E(P) = e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{h=0}^{\infty} \frac{\lambda^h}{h!}$$

$$= \lambda$$

luego el estimador de λ por el método de momentos es

$$\hat{\lambda}_n = \frac{P_1 + P_2 + \cdots + P_n}{n}$$

y $\hat{\lambda}_n \xrightarrow{P} \lambda$ (Decimos que el estimador es consistente).

En esta clase introduciremos un nuevo método de estimación que llamaremos "Método de máxima verosimilitud".

Para introducirlo consideremos una m.a.s. $X_1, X_2, \ldots, X_n$ de una distribución $F_X(\cdot, \theta)$

Es decir la familia depende del parámetro θ. Supongamos (esto no es en absoluto necesario) que X posee densidad $f(x, \theta)$ lo que implica, debido a que las v.a. son independientes, que la densidad del vector $(X_1, X_2, \ldots, X_n)$ resulta ser

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

El método consiste en reemplazar los x_j de arriba por las observaciones $(X_1, X_2, \ldots, X_n)$ y construir la función de verosimilitud

$$\boxed{L_n(\theta) = \prod_{j=1}^{n} f(X_j, \theta)}$$

y también su logaritmo

$$\boxed{\mathcal{L}_n(\theta) = \log L_n(\theta) = \sum_{j=1}^{n} \log f(X_j, \theta).}$$

El método consiste en definir el estimador de máxima verosimilitud como aquel $\hat{\theta}_n$ que

Como el valor que maximiza $L_n(\theta)$ o $\mathcal{L}_n(\theta)$ esto se escribe

$$\boxed{\hat{\theta}_n = \arg\max_{\theta} L_n(\theta)}$$

Para ejemplificar consideremos algunos ejemplos

Ejemplo 1.-

Sea $E_1, E_2, \ldots, E_n$ una muestra aleatoria simple de una exponencial con parámetro λ. Entonces la verosimilitud se escribe

$$L_n(\lambda) = \lambda^n \prod_{j=1}^{n} e^{-\lambda X_j}$$

y su logaritmo

$$\mathcal{L}_n(\lambda) = n \log \lambda - \lambda \sum_{j=1}^{n} X_j$$

para encontrar el máximo derivamos

$$\mathcal{L}_n'(\lambda) = \frac{n}{\lambda} - \sum_{j=1}^{n} X_j$$

igualamos a cero y así obtenemos

$$\boxed{\hat{\lambda}_n = \left(\frac{1}{n} \sum_{j=1}^{n} X_j\right)^{-1} = \left(\overline{X}_n\right)^{-1}}$$

y como un adicional tenemos $\overline{X}_n \xrightarrow{LGN} E(X) = 1/\lambda$, luego

$$\hat{\lambda}_n \xrightarrow{P} \lambda.$$

Otro ejemplo importante es el de la gaussiana $N(\mu, \sigma^2)$

Ejemplo 2: $Z_1, Z_2, \ldots, Z_n$ es una m.a.s de una gaussiana con densidad

$$\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Luego $$\boxed{L_n(\mu, \sigma^2) = \frac{1}{(2\pi)^{n/2}\sigma^n} \prod_{j=1}^{n} e^{-\frac{(X_j-\mu)^2}{2\sigma^2}}}$$

y su logaritmo

$$\boxed{\mathcal{L}_n(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \sum_{j=1}^{n}\frac{(X_j-\mu)^2}{2\sigma^2}}$$

Ahora el logaritmo de la verosimilitud es una función de dos variables y para conseguir los puntos que hacen el máximo debemos derivar con respecto a cada una de las variables e igualar a cero.

① $$\frac{\partial \mathcal{L}_n}{\partial \mu} = \frac{1}{\sigma^2}\sum_{j=1}^{n}(X_j-\mu) = 0$$

$$\sum_{j=1}^{n} X_j - n\mu = 0 \Rightarrow \boxed{\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} X_j}$$

② $$\frac{\partial \mathcal{L}_n}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \sum_{j=1}^{n}\frac{(X_j-\mu)^2}{2\sigma^4} = 0$$

Entonces obtenemos despejando

$$-\frac{n\sigma^2}{2}+\sum_{j=1}^{n}\frac{(X_j-\mu)^2}{2}=0$$

$$\Rightarrow \qquad \sigma_n^2=\frac{1}{n}\sum_{j=1}^{n}(X_j-\mu)^2$$

y como ya habíamos visto que $\hat{\mu}=\bar{X}_n$

$$\boxed{\hat{\sigma}_n^2=\frac{1}{n}\sum_{j=1}^{n}(X_j-\bar{X}_n)^2.}$$

Estos los estimadores fueron los mismos que obtuvimos por el método de los momentos y que fueron analizados en la clase pasada para así obtener su comportamiento.

El próximo ejemplo es importante pues muestra que para encontrar el punto que hace máxima a la verosimilitud no siempre se puede derivar la función.

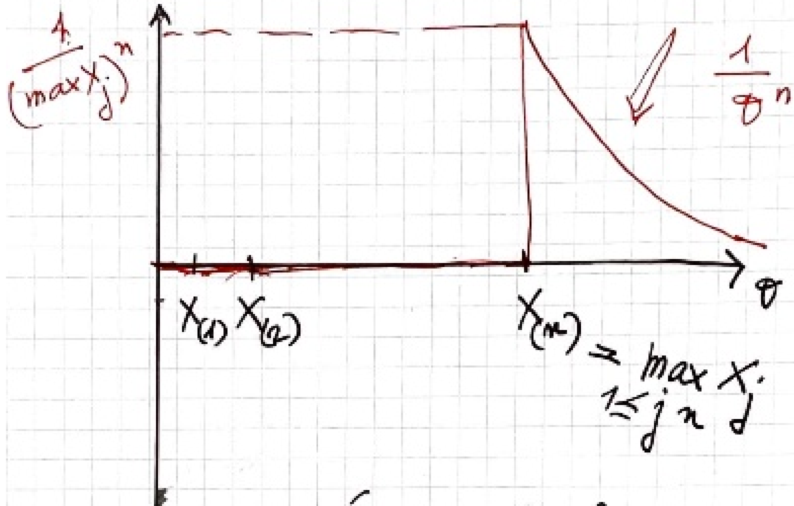
Ejemplo 3.-
Supongamos que $X_1, X_2, \ldots, X_n$ es una muestra aleatoria simple de una uniforme en el intervalo $[0,\theta]$ con $\theta>0$. Como sabemos la densidad de una tal variable es

$$\mathcal{V}(\theta)=\begin{cases}\frac{1}{\theta} & \text{para } x\in[0,\theta]\\ 0 & \text{en otro caso.}\end{cases}$$

Entonces la verosimilitud es

$$l_n(\theta)=\begin{cases}\frac{1}{\theta^n} & \text{si } X_i<\theta \text{ para todo } i\\ 0 & \text{en otro caso}\end{cases}$$

La figura siguiente nos muestra su comportamiento



De la observación de la figura se ve que el $\arg\max \mathcal{L}_n(\theta) = \max_{1\le j\le n} X_j$.

Ejemplo 4.
Supongamos que $X_1, X_2, \dots, X_n$ es una m.a.s. de una v.a. Binomial S_N de parámetro p.
Sabemos que

$$\underline{P}(S_N = k) = \binom{N}{k} p^k (1-p)^{N-k}$$

donde $k = 0, 1, 2, \dots, N$. De esta forma la función de verosimilitud se escribe:

$$L_n(p) = \prod_{j=1}^{n} \binom{N}{X_j} p^{X_j} (1-p)^{N-X_j}$$

Si tomamos el logaritmo tenemos

$$\mathcal{L}_n(p) = \sum_{j=1}^{n} \log\binom{N}{X_j} + (\log p) X_j + (N-X_j)\log(1-p)$$

Entonces al derivar

$$\frac{d\mathcal{L}_n}{dp} = \frac{1}{p}\sum_{j=1}^{n} X_j - \frac{1}{(1-p)} \sum_{j=1}^{n} (N - X_j) = 0$$

$$= \frac{1}{p}\sum_{j=1}^{n} X_j + \frac{1}{(1-p)}\sum_{j=1}^{n} X_j = \frac{1}{(1-p)} Nn$$

$$= \left(\frac{1}{p} + \frac{1}{1-p}\right)\sum_{j=1}^{n} X_j = \frac{1}{(1-p)} N n.$$

$$= \frac{1}{p\cancel{(1-p)}}\sum_{j=1}^{n} X_j = \frac{1}{\cancel{(1-p)}} Nn$$

$$\Rightarrow \quad \boxed{\frac{1}{Nn}\sum_{j=1}^{n} X_j = \hat{p}}$$

Podemos entonces calcular la esperanza de este estimador y así obtener

$$E\,\hat{p} = \frac{1}{Nn}(Np)n = p.$$

Luego la esperanza del estimador es el verdadero parámetro. Esto último nos lleva entonces a la definición siguiente.

Definición: El sesgo de un estimador $\hat{\theta}$ es definido por $\boxed{\text{Sesgo}(\hat{\theta}) = E\,\hat{\theta} - \theta}$

Un estimador es insesgado si:

$$\boxed{\text{Sesgo}(\hat{\theta}) = 0 \iff E(\hat{\theta}) = \theta.}$$

Otro concepto importante para medir la bondad de un estimador es el siguiente:
El error cuadrático medio de un estimador se define

$$\boxed{ECM(\hat{\theta}) = E[(\hat{\theta}-\theta)^2].}$$

Podemos entonces obtener la descomposición siguiente

$$E[(\hat{\theta}-\theta)^2] = E[((\hat{\theta}-E(\hat{\theta})+[E(\hat{\theta})-\theta])^2]$$

$$= E[(\hat{\theta}-E(\hat{\theta}))^2] + 2E[(\hat{\theta}-E(\hat{\theta}))(E(\hat{\theta})-\theta)]$$

$$+ (E(\hat{\theta})-\theta)^2$$

$$= E[(\hat{\theta}-E(\hat{\theta}))^2] + (E(\hat{\theta})-\theta)^2$$

$$= Var(\hat{\theta}) + [\text{Sesgo}(\hat{\theta})]^2$$

En otras palabras

$$\boxed{ECM(\hat{\theta}) = Var(\hat{\theta}) + (\text{Sesgo}(\hat{\theta}))^2}$$