Estimación Puntual

Introducción

- En **Probabilidades** se conocían todos los datos sobre la v.a.

  Ejemplo: X ~ B(25, 0.1) y podíamos calcular P(X=5).

- En **Estadística** desconocemos las características de X total o parcialmente.

  Ejemplo: sabemos que X tiene distribución binomial pero ***desconocemos p*** y a partir de una muestra de 25 artículos trataremos de hallar información sobre p.

Inferencia Estadística

- El campo de la inferencia estadística está formado por los métodos utilizados para tomar decisiones o para obtener conclusiones sobre el o los parámetros de una población a partir de la información contenida en una muestra.
- La inferencia estadística puede dividirse en dos grandes áreas:
  - Estimación de Parámetros.
  - Pruebas de hipótesis.

Estimación Puntual

- Dado un parámetro de interés, como una media poblacional μ o una proporción p de la población, el objetivo de la estimación puntual es usar una muestra para calcular un número que representa, en cierto sentido, una buena estimación del valor real del parámetro.
- El número resultante se llama estimación puntual.
- La estimación se realiza a través de algún ***estadístico*** (ej: media muestral).

Ejemplo 7.1

Suponga que el parámetro de interés es μ, el tiempo de vida útil promedio de baterías de cierto tipo.

Ej: muestra de n=3 baterías. Para c/u se indica el tiempo de vida útil observado (horas)

$x_1=5.0$, $x_2 = 6.4$ y $x_3 = 5.9$. El valor calculado del tiempo de vida útil promedio de la muestra es

$\bar{x} = 5.77$ ⟵

Podría considerarse este valor como la mejor estimación de μ a partir de la información de la muestra

---

Estimador Puntual

- Cuando un estadístico se utiliza para estimar un parámetro desconocido se lo llama **estimador puntual**.
- Es habitual simbolizar en forma genérica a un parámetro con la letra θ y al estadístico que se utiliza como estimador puntual de θ, simbolizarlo con $\widehat{\Theta}$.
- Al medir la muestra aleatoria se obtienen $x_1, x_2, \ldots, x_n$ *y entonces* ***el valor que toma*** $\widehat{\Theta}$ ***es***

$$\hat{\theta} = h(x_1, x_2, \ldots, x_n)$$

y se denomina ***estimación puntual de*** $\boldsymbol{\theta}$

Estimadores usuales

- Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una v.a. X donde $E(X) = \mu$ y $V(X) = \sigma^2$.
- Si se desconoce μ, un estadístico que se utiliza para estimar ese parámetro es la **media o promedio muestral**

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

- Si se desconoce σ^2, un estadístico que se utiliza para estimar ese parámetro es la **varianza muestral**

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

Luego veremos porque utiliza (n-1) en lugar de n

- Si se busca estimar la **proporción** p de objetos de una población que cumplen una determinada característica, el estimador puntual sería

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

donde X_i vale 1 si la i-ésima observación tiene la característica de interés y 0 si no, para i=1,2,…, n.

Más de un estimador

Puede ocurrir que se tenga más de un estimador para un parámetro.

Ejemplo: Para estimar el valor de μ puede utilizarse la media muestral o también la semisuma entre X_1 y X_n

$$\hat{\mu} = \frac{X_1 + X_n}{2}$$

En estos casos necesitamos de algún **criterio para decidir cuál es mejor estimador** de μ

Criterios para evaluar estimadores

- En una situación ideal, se podría encontrar un estimador $\widehat{\Theta}$ para el cual siempre se cumpla que $\widehat{\Theta} = \theta$
- Sin embargo, $\widehat{\Theta}$ es una función de las X_i muestrales y por lo tanto, una v.a.

$$\therefore \quad \hat{\theta} = \theta + \underbrace{error\ de\ estimacion}$$

Lo que se necesita es un estimador que tenga las propiedades de *insesgamiento* y *varianza mínima*.

Estimador insesgado

- Se dice que un estimador puntual $\widehat{\Theta}$ es **estimador insesgado** del parámetro θ si

$$E(\hat{\Theta}) = \theta$$

cualquiera sea el valor verdadero de θ.

- **Sesgo del estimador** : $E(\hat{\Theta}) - \theta$
- Note que si el estimador es insesgado, su sesgo es cero.

Ejemplo 7.2

Sea $X_1, X_2, \ \dots \ , X_n$ una muestra aleatoria de una v.a. X donde $E(X)=\mu$ y $V(X)=\sigma^2$.

a) Veamos si la media muestral

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

es un estimador insesgado de μ. Debemos ver que $E(\overline{X}) = \mu$

Usando la prop.de linealidad de la esperanza

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$

Por tratarse de componentes de una muestra aleatoria

$$E(X_i) = E(X) = \mu \qquad \forall i = 1,2,\dots,n$$

Luego

$$E(\overline{X}) = \frac{1}{n} n\mu = \mu$$

b) Veamos si la varianza muestral $\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

es un estimador insesgado de σ^2. Debemos ver que $E(S^2) = \sigma^2$

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

Reescribamos la suma sumando y restando μ y desarrollando el cuadrado

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n}(X_i - \mu + \mu - \overline{X})^2 = \sum_{i=1}^{n}([X_i - \mu] + [\mu - \overline{X}])^2 =$$

$$= \sum_{i=1}^{n}([X_i - \mu]^2 + 2[X_i - \mu][\mu - \overline{X}] + [\mu - \overline{X}]^2) =$$

$$= \sum_{i=1}^{n}[X_i - \mu]^2 + \sum_{i=1}^{n} 2[X_i - \mu][\mu - \overline{X}] + \sum_{i=1}^{n}[\mu - \overline{X}]^2) =$$

$$= \sum_{i=1}^{n}[X_i - \mu]^2 + 2[\mu - \overline{X}]\underbrace{\sum_{i=1}^{n}[X_i - \mu]} + n[\mu - \overline{X}]^2) =$$

$$\sum_{i=1}^{n}[X_i - \mu] = \sum_{i=1}^{n} X_i - n\mu = n(\overline{X} - \mu)$$

$$= \sum_{i=1}^{n}[X_i - \mu]^2 + 2[\mu - \overline{X}]n[\overline{X} - \mu] + n[\mu - \overline{X}]^2) =$$

$$= \sum_{i=1}^{n}[X_i - \mu]^2 - 2n[\mu - \overline{X}]^2 + n[\mu - \overline{X}]^2) = \sum_{i=1}^{n}[X_i - \mu]^2 - n[\mu - \overline{X}]^2$$

$$E(S^2)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}(X_i-\overline{X})^2\right)=\frac{1}{n-1}E\left(\sum_{i=1}^{n}[X_i-\mu]^2-n[\mu-\overline{X}]^2\right)$$

$$=\frac{1}{n-1}\left(\sum_{i=1}^{n}E[X_i-\mu]^2-nE[\mu-\overline{X}]^2\right)=\frac{1}{n-1}\left(\sum_{i=1}^{n}V(X_i)-nE[\mu-\overline{X}]^2\right)$$

$$=\frac{1}{n-1}\left(\sum_{i=1}^{n}V(X_i)-nE[\overline{X}-\mu]^2\right)=\frac{1}{n-1}\left(\sum_{i=1}^{n}V(X_i)-nE[\overline{X}-E(\overline{X})]^2\right)$$

$$=\frac{1}{n-1}\left(\sum_{i=1}^{n}V(X_i)-nV(\overline{X})\right)$$

Usando $V(X_i)=V(X)=\sigma^2 \quad \forall i=1,2,\ldots,n$ y que $V(\overline{X})=\frac{\sigma^2}{n}$

$$E(S^2)=\frac{1}{n-1}\left(n\sigma^2-n\frac{\sigma^2}{n}\right) \Rightarrow E(S^2)=\sigma^2$$

Ejemplo 7.4

La definición de varianza estándar es un estimador sesgado de la varianza

$$E\left[\frac{(n-1)S^2}{n}\right]=\frac{n-1}{n}E(S^2)=\frac{n-1}{n}\sigma^2$$

Por ser negativo, este estimador tiende a subestimar σ^2

Resumen

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una v.a. X donde $E(X)=\mu$ y $V(X)=\sigma^2$

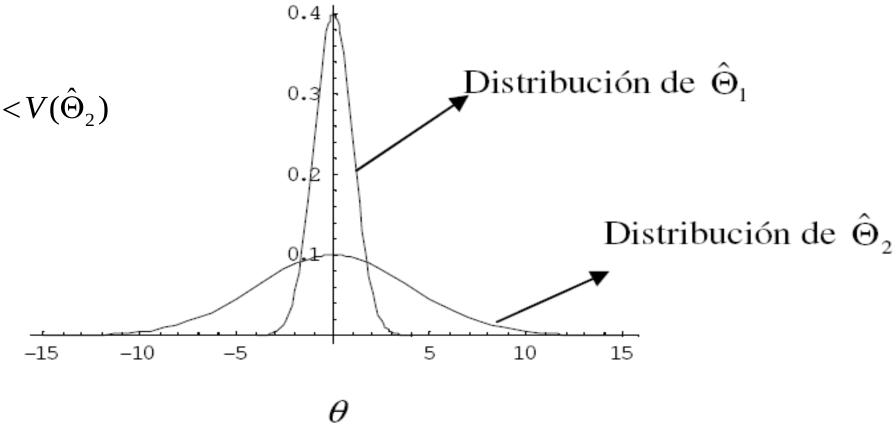
$\bar{X}$	$\frac{\sum_{i=1}^{n} x_i}{n}$	$E(\bar{X})=\mu$; $V(\bar{X})=\frac{\sigma^2}{n}$
S^2	$\frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n-1}$	$E(S^2)=\sigma^2$
S_{EMV}^2	$\frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n}$	$E(S_{EMV}^2)=\frac{(n-1)\sigma^2}{n}$

Criterios para evaluar estimadores

- En ocasiones hay más de un estimador insesgado de un parámetro θ.
- Por lo tanto necesitamos un método para seleccionar un estimador entre varios estimadores insesgados.
  - Varianza
  - Error Cuadrático medio

Varianza de un estimador

$V(\hat{\Theta}_1) < V(\hat{\Theta}_2)$



Si tenemos dos estimadores insesgados debemos elegir el de menor varianza.

- Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una v.a. X donde $E(X)=\mu$ y $V(X)=\sigma^2$
- Para estimar el parámetro μ podría utilizarse

$$\hat{\mu}_1 = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad\qquad \hat{\mu}_2 = \frac{X_1 + X_2}{2}$$

Cuál elegimos? Ya vimos que $\hat{\mu}_1$ es insesgado. Ahora veamos si $\hat{\mu}_2$ es insesgado.

$$E(\hat{\mu}_2) = E\left(\frac{X_1+X_2}{2}\right) = \frac{1}{2}(E(X_1)+E(X_2)) = \frac{1}{2}(\mu+\mu) = \frac{1}{2}2\mu = \mu$$

Por lo tanto $\hat{\mu}_2$ también es insesgado

Entonces ¿Cuál de los dos estimadores es mejor?

Calculamos la varianza de cada uno utilizando las propiedades de la varianza.

$$V(\overline{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}V\left(\sum_{i=1}^{n} X_i\right) \underset{\Uparrow}{=} \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) \underset{\Uparrow}{=} \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

Las X_i son v.a. independientes	Todas las X_i tienen la misma varianza.

$$V(\hat{\mu}_2) = V\left(\frac{X_1+X_2}{2}\right) = \frac{1}{4}(V(X_1)+V(X_2)) = \frac{1}{4}(\sigma^2+\sigma^2) = \frac{\sigma^2}{2}$$

Si n>2 el mejor estimador es $\hat{\mu}_1$ porque $V(\hat{\mu}_1) < V(\hat{\mu}_2)$

Criterios para evaluar estimadores

- Supongamos ahora que $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son dos estimadores de un parámetro θ y ***alguno de ellos no es insesgado***.
- A veces es necesario utilizar un estimador sesgado. En esos casos puede ser importante el ***error cuadrático medio del estimador***.

Error cuadrático medio (ECM)

El ECM es un criterio para comparar estimadores

- El error cuadrático medio de un estimador de un parámetro θ está definido como

$$ECM(\hat{\Theta}) = E\left[\left(\hat{\Theta} - \theta\right)^2\right]$$

- También puede escribirse como $ECM(\hat{\Theta}) = V(\hat{\Theta}) + \left(b(\hat{\Theta})\right)^2$ siendo $b(\hat{\Theta}) = E(\hat{\Theta}) - \theta$

Eficiencia relativa

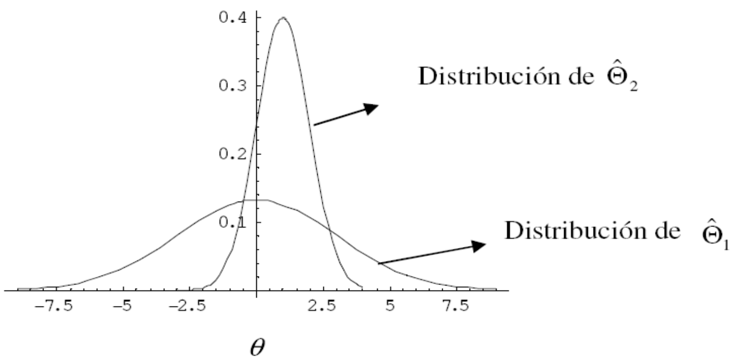
- Si $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son dos estimadores de un parámetro θ, la eficiencia relativa de $\hat{\Theta}_2$ con respecto a $\hat{\Theta}_1$ se define como $\dfrac{ECM(\hat{\Theta}_1)}{ECM(\hat{\Theta}_2)}$
- Si es un valor menor que 1, $\hat{\Theta}_1$ tiene menor ECM que $\hat{\Theta}_2$ y **por lo tanto** $\hat{\Theta}_1$ **es más eficiente que** $\hat{\Theta}_2$

Criterios para evaluar estimadores

- Si $\hat{\Theta}$ es un estimador insesgado de θ entonces $ECM(\hat{\Theta}) = V(\hat{\Theta})$
- En el ECM se consideran la varianza y el sesgo del estimador.
- Si $\hat{\Theta}_1$ y $\hat{\Theta}_2$ son dos estimadores de θ tales que

  $E(\hat{\Theta}_1) = \theta$; $E(\hat{\Theta}_2) \neq \theta$ y $V(\hat{\Theta}_2) < V(\hat{\Theta}_1)$

  habría que calcular el ECM de cada uno y tomar el menor.



Se puede ver en la figura anterior que aunque $\hat{\Theta}_2$ sea sesgado, al tener menor varianza podría llegar a tomar valores mas cercanos al verdadero parámetro $\hat{\Theta}_1$.

Ejemplo 7.6

Sean $\hat{\Theta}_1$, $\hat{\Theta}_2$ y $\hat{\Theta}_3$ tres estimadores de un parámetro θ tales que

$$E(\hat{\Theta}_1) = E(\hat{\Theta}_2) = \theta;\ E(\hat{\Theta}_3) \neq \theta$$

$$V(\hat{\Theta}_1) = 10,\ V(\hat{\Theta}_2) = 6,\ E((\hat{\Theta}_3 - \theta)^2) = 4$$

Haga una comparación de estos estimadores ¿Cuál prefiere y por qué?

Si calculamos el ECM de cada estimador

$$ECM(\hat{\Theta}_1) = V(\hat{\Theta}_1) = 10$$

$$ECM(\hat{\Theta}_2) = V(\hat{\Theta}_2) = 6$$

$$ECM(\hat{\Theta}_3) = E((\hat{\Theta}_3 - \Theta)^2) = 4$$

El mejor estimador es $\hat{\Theta}_3$ porque tiene menor ECM.

Consistencia de estimadores puntuales

Sea $\hat{\Theta}_n$ un estimador del parámetro θ basado en una muestra aleatoria ($X_1, X_2, \ldots, X_n$), si

$$\lim_{n\to\infty} E(\hat{\Theta}_n) = \theta \qquad \text{y} \qquad \lim_{n\to\infty} V(\hat{\Theta}_n) = 0$$

entonces $\hat{\Theta}_n$ es un **estimador consistente** de θ.

Ejemplo 7.7

La media muestral es un estimador consistente?

Sabemos que

$$a)\, E(\overline{X}) = \mu = E(X) \qquad \forall n \Longleftarrow$$

Si se cumple para todo n, también se cumple para el límite $n \to \infty$

$$b)\, V(\overline{X}) = \frac{\sigma^2}{n} = \frac{V(X)}{n} \qquad \forall n \Longleftarrow$$

Vemos que tiende a cero cuando $n \to \infty$

por lo tanto el estimador es **consistente**

Métodos de estimación puntual

Se analizan dos métodos constructivos para obtener estimadores puntuales

- Método de momentos
- Método de máxima verosimilitud: Los estimadores obtenidos de esta forma suelen ser más eficientes pero requieren más cálculo.

Método de Momentos

La idea básica es igualar ciertas características muestrales, como la media, con los valores esperados correspondientes.

Esto da lugar a un conjunto de ecuaciones cuya solución, para valores de parámetros desconocidos, producen los estimadores.

Sean $X_1, \ldots, X_n$ una muestra aleatoria de una función de probabilidad f(x).

Se definen

- El k-ésimo momento de la distribución f(x) como $\mu_k = E(X^k)$ con k=1,2,3,...
- El k-ésimo momento muestral como

$$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \qquad \text{con k=1,2,3,...}$$

Sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución con función de probabilidad $p(x_i, \theta_1, \theta_2, \ldots, \theta_r)$ si X es discreta o $f(x_i, \theta_1, \theta_2, \ldots, \theta_r)$ si X es continua, donde $\theta_1, \theta_2, \ldots, \theta_r$ son parámetros que se desconocen.

Entonces los estimadores de momento $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_r$ se obtienen igualando los primeros *r* momentos muestrales con los primeros *r* momentos de la distribución y resolviendo para $\theta_1, \theta_2, \ldots, \theta_r$

Dicho de otra forma, sean $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución cuya función de probabilidad tiene parámetros $\theta_1, \theta_2, \ldots, \theta_r$ desconocidos, deben resolverse la siguientes ecuaciones para dichos parámetros

$$E(X^k) = \frac{1}{n}\sum_{i=1}^{n} X_i^k \qquad \text{k=1, 2, ..., r}$$

siendo

$$\mu_k = E(X^k) = \begin{cases} \sum_{x_i \in R_X} x_i^k . p(x_i, \theta_1, \theta_2, ..., \theta_r) & \text{(discreta)} \\ \int_{-\infty}^{\infty} x_i^k . f(x_i, \theta_1, \theta_2, ..., \theta_r) & \text{(continua)} \end{cases}$$

Ejemplo 7.8

Sea (X_1, X_2, …, X_n) una muestra aleatoria de tamaño n de una v.a. donde $X \sim U[0,\theta]$, con θ desconocido. Hallar el estimador de θ por el método de los momentos.

Según la definición anterior, debe igualarse el 1er. momento muestral con el 1er. momento de la distribución.

$$E(X) = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \therefore \quad E(X) = \overline{X}$$

Por lo tanto $E(X) = \dfrac{0+\theta}{2} = \dfrac{\theta}{2} \Rightarrow \dfrac{\theta}{2} = \overline{X} \Rightarrow \hat{\Theta} = 2\overline{X}$

Verifique que este estimador es consistente (es decir que tenemos que ver que el valor esperado coincide con el parámetro y que la varianza es cero o tiende a cero a cuando el tamaño de la muestra tiende a infinito.

$$E(\hat{\Theta}) = E(2\overline{X}) = 2E(\overline{X}) = 2\frac{\theta}{2} = \theta$$

$$V(\hat{\Theta}) = V(2\overline{X}) = 4V(\overline{X}) = 4\frac{(\theta-0)^2}{12n} = \frac{\theta^2}{3n} \underset{n\to\infty}{\to} 0$$

Ejemplo 7.9

Estimar por el método de los momentos los parámetros μ y σ^2 de una distribución normal.

Como son dos parámetros utilizaremos los dos primeros momentos de la distribución normal, que están dados por:

$$\mu_1 = E(X) = \mu$$
$$\mu_2 = E(X^2) = \sigma^2 + \mu^2$$

Luego, igualando los dos primeros momentos de la distribución normal con sus respectivos momentos muestrales tenemos que:

$$\mu_1 = M_1 \Rightarrow \mu = \overline{X} \Rightarrow \hat{\mu} = \overline{X}$$

$$\mu_2 = M_2 \Rightarrow \sigma^2 + \mu^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \qquad \therefore \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$$

Método de Máxima Verosimilitud

Sea *X es una v.a. discreta con fdp p(x, θ) si es discreta o f(x,θ) si es continua , donde* θ es un parámetro desconocido.

Sean x_1, x_2, … x_n *los valores observados de una muestra aleatoria* de tamaño *n.*

Se define la ***función de verosimilitud como la función de distribución conjunta de las observaciones***

$$L(x_1, x_2,..., x_n, \theta) = P(X_1 = x_1)P(X_2 = x_2)...P(X_n = x_n) = p(x_1, \theta)p(x_2, \theta)...p(x_n, \theta)$$

o $$L(x_1, x_2,..., x_n, \theta) = f(x_1, \theta)f(x_2, \theta)...f(x_n, \theta)$$

según si X es discreta o continua respectivamente.

El **estimador de máxima verosimilitud (EMV)** de θ es aquel valor de θ que maximiza la función de verosimilitud

---

Ejemplo 7.10

Sea (X_1, X_2, …, X_n) una muestra aleatoria de tamaño n de una v.a. X ~ B(1, p).

Por ejemplo, se eligen al azar n objetos de una línea de producción, y cada uno se clasifica como defectuoso (en cuyo caso x_i=1) o no defectuoso (en cuyo caso x_i=0).

Entonces p = P(X_i=1), es decir es la verdadera proporción de objetos defectuosos en la producción total.

Queremos hallar el EMV de p.

La función de verosimilitud será

$$L(x_1, x_2,..., x_n, \theta) = p(x_1, \theta)p(x_2, \theta)...p(x_n, \theta) = \left[p^{x_1}(1-p)^{1-x_1}\right]\left[p^{x_2}(1-p)^{1-x_2}\right]..\left[p^{x_n}(1-p)^{1-x_n}\right]$$

Esto puede escribirse

$$L(x_1, x_2,..., x_n, \theta) = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

Aplicando logaritmo natural de L

$$\ln(L(x_1, x_2,..., x_n, \theta)) = \left(\sum_{i=1}^{n} x_i\right)\ln p + \left(n - \sum_{i=1}^{n} x_i\right)\ln(1-p)$$

Derivando respecto de p e igualando a cero

$$\frac{\partial \ln(L(x_1, x_2,..., x_n, \theta))}{\partial p} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p} = 0$$

$$\frac{\sum_{i=1}^{n} x_i}{p} = \frac{n - \sum_{i=1}^{n} x_i}{1-p} \Rightarrow (1-p)\sum_{i=1}^{n} x_i = p\left(n - \sum_{i=1}^{n} x_i\right)$$

$$\sum_{i=1}^{n} x_i - p\sum_{i=1}^{n} x_i = pn - p\sum_{i=1}^{n} x_i \Rightarrow \sum_{i=1}^{n} x_i = pn \quad \Longrightarrow \quad p = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

$$\hat{p} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

La proporción de defectuosos de la muestra .

Ejemplo 7.11

El tiempo de falla T de una componente tiene una distribución exponencial con parámetro λ. $T \sim Exp(\lambda)$, es decir la fdp es

$$f(t;\lambda) = \begin{cases} \lambda e^{-\lambda t} & t \geq 0 \\ 0 & si\ no \end{cases}$$

Recordemos que $E(T) = \frac{1}{\lambda}$ y que $V(T) = \frac{1}{\lambda^2}$

Calculemos el EMV del parámetro λ para una muestra de tamaño n.

La función de verosimilitud será

$$L(t_1,t_2,\ldots,t_n;\lambda)= f(t_1;\lambda)f(t_2;\lambda)\ldots f(t_n;\lambda)=\left[\lambda e^{-\lambda t_1}\right]\left[\lambda e^{-\lambda t_2}\right]\ldots\left[\lambda e^{-\lambda t_n}\right]$$

Esto puede escribirse

$$L(t_1,t_2,\ldots,t_n;\lambda)=(\lambda)^n e^{-\lambda\sum_{i=1}^{n} t_i}$$

Aplicando logaritmo natural de L

$$\ln(L(t_1,t_2,\ldots,t_n;\lambda))=\ln\left(\lambda^n e^{-\lambda\sum_{i=1}^{n} t_i}\right)=n\ln(\lambda)-\lambda\sum_{i=1}^{n} t_i$$

Derivando respecto de λ e igualando a cero

$$\frac{\partial\ln(L(t_1,t_2,\ldots,t_n;\lambda))}{\partial\lambda}=n\frac{1}{\lambda}-\sum_{i=1}^{n} t_i=0 \Longrightarrow \lambda=\frac{n}{\sum_{i=1}^{n} t_i}=\frac{1}{\bar{t}}$$

Por lo tanto el estimador es

$$\hat{\lambda}=\frac{n}{\sum_{i=1}^{n} T_i}=\frac{1}{\overline{T}}$$

Método de Máxima Verosimilitud

Características

- ❑ Algunas veces, resulta difícil maximizar la función de verosimilitud debido a que la ecuación obtenida a partir de

$$\frac{\partial L(\theta)}{\partial\theta}=0$$

  no resulta fácil de resolver.
- ❑ También puede ocurrir que los métodos de cálculo para maximizar *$L(\theta)$ no sean aplicables*

Ejemplo 7.12

Sea $(X_1, X_2, \ldots, X_n)$ una muestra aleatoria de tamaño n de una v.a. $X \sim U(0, \theta)$, con θ desconocido. Hallar el estimador de θ por el método de máxima verosimilitud.

La *fdp* de X es

$$f(x)=\begin{cases}\dfrac{1}{\theta} & 0<x<\theta\\ 0 & si\ no\end{cases}$$

La función de verosimilitud es

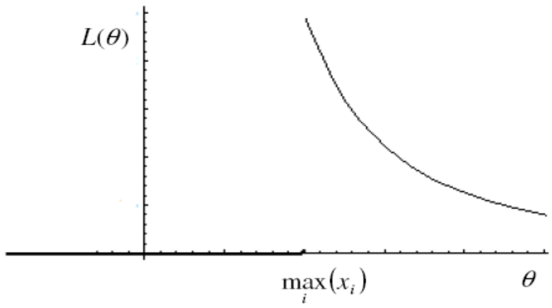
$$L(x_1,x_2,\ldots,x_n;\theta)=\begin{cases}\dfrac{1}{\theta^n} & si\quad \max_i(x_i)<\theta\\ 0 & si\ no\end{cases}$$

Si derivamos con respecto a θ obtenemos

$$\frac{d}{d\theta}\theta^{-n}=-\frac{n}{\theta^{n+1}} \Longleftarrow \boxed{\text{Siempre es menor que cero}}$$

La función de máxima verosimilitud es una función decreciente para todos los

$$\theta>\max_i(x_i)$$



Vemos que donde la función tiene el máximo hay una discontinuidad no evitable. Por lo tanto

$$\hat{\Theta} = \max_i (x_i)$$

Método de Máxima Verosimilitud

El método de máxima verosimilitud puede emplearse en el caso donde hay más de un parámetro desconocido para estimar.

En ese caso la función de verosimilitud es una función de varias variables.

$$L(x_1, x_2, ..., x_n; \theta_1, \theta_2, ..., \theta_k)$$

En este caso, los estimadores de máxima verosimilitud $\hat{\Theta}_1, \hat{\Theta}_2, ..., \hat{\Theta}_k$ se obtienen al plantear (si existen las derivadas parciales) y resolver el sistema de k ecuaciones con k incógnitas $\theta_1, \theta_2, ..., \theta_k$

$$\frac{d}{d\theta_i} L(x_1, x_2, ..., x_n; \theta_1, \theta_2, ..., \theta_k) = 0 \qquad \text{i =1,2,... k}$$

Ejemplo 7.13

La variable aleatoria X tiene distribución $N(\mu, \sigma^2)$ ambos parámetros desconocidos para los cuales se desea encontrar los estimadores máxima verosimilitud.

La fdp es

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$

La función de máxima verosimilitud es

$$L(x_1, x_2, ..., x_n; \mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

Luego

$$\ln(L(x_1, x_2, ..., x_n; \mu, \sigma^2)) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

El sistema de ecuaciones de verosimilitud queda

$$\begin{cases} \dfrac{\partial \ln(L(x_1, x_2, ..., x_n; \mu, \sigma^2))}{\partial \mu} = \displaystyle\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2 = 0 \\ \dfrac{\partial \ln(L(x_1, x_2, ..., x_n; \mu, \sigma^2))}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2}\displaystyle\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^4} = 0 \end{cases}$$

Resolviendo con respecto a μ y σ^2

$$\begin{cases} \mu = \dfrac{1}{n}\sum_{i=1}^{n} x_i = \overline{x} \\ \sigma^2 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2 \end{cases}$$

Entonces, los estimadores son

$$\begin{cases} \hat{\mu} = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} \\ \hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2 \end{cases}$$

Propiedades de los Estimadores de Máxima Verosimilitud

Propiedad 1

Los EMV pueden ser ***sesgados****, pero en general si* $\hat{\Theta}$ *es el EMV de un parámetro* θ *basado en* una muestra de tamaño *n, entonces*

$$\lim_{n\to\infty} E(\hat{\Theta}) = \theta$$

es decir son ***asintóticamente insesgados***

Propiedad 2

Bajo condiciones bastantes generales se puede probar que los EMV son ***asintóticamente consistentes***

Propiedad 3

Bajo condiciones bastantes generales se puede probar que los EMV ***asintóticamente tienen varianza mínima***

Propiedad 4

Los EMV cumplen la ***propiedad de invarianza****. E*s decir que si $\hat{\Theta}$ es un EMV de un parámetro θ, el EMV de $g(\theta)$ *es* $g(\hat{\Theta})$*, si g(x) es una función inyectiva* (es decir que a cada valor de la imagen le corresponde un único valor en el dominio).

Ejemplo 7.14

En el Ejemplo 7.11 teníamos una v.a. T con distribución exponencial, *T ~ Exp(*λ *).* Si queremos el EMV de la varianza poblacional, podemos calcularlo recordando que $V(T) = 1/\lambda^2$ es decir que $V(T) = g(\lambda) = 1/\lambda^2$

Vimos que $$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} T_i} = \frac{1}{\overline{T}}$$ por lo tanto $$\sigma^2 = \frac{1}{\hat{\lambda}^2}$$