– Métodos de estimación puntual

- ***Método de los momentos***
- ***Método de máxima verosimilitud***

Método de los momentos

Definimos los ***momentos de orden k de una variable aleatoria*** como:

$$\mu_k = E\left(X^k\right) = \sum_{x_i \in R_X} x_i^{\,k}\, p(x_i) \qquad (k = 0,1,2,\ldots) \qquad \text{Si } X \text{ es discreta}$$

$$\mu_k = E\left(X^k\right) = \int_{-\infty}^{+\infty} x^k f(x)\,dx \qquad (k = 0,1,2,\ldots) \qquad \text{Si } X \text{ es continua,}$$

y definimos los correspondientes momentos ***muestrales de orden k*** como:

$$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \qquad (k = 0,1,2,\ldots),$$

Entonces la ley débil de los grandes números se puede generalizar:

$$\lim_{n\to\infty} P\left(\left|M_k - \mu_k\right| \geq \varepsilon\right) = 0 \qquad (k = 0,1,2,\ldots).$$

De acuerdo con esto parece razonable estimar los momentos poblacionales de orden k mediante los momentos muestrales de orden k: $\mu_k \sim M_k \quad (k = 0,1,2,\ldots)$.

Supongamos, entonces, una variable aleatoria X y supongamos que la distribución de X depende de r parámetros $\theta_1, \theta_2, ..., \theta_r$, esto es la *fdp* poblacional es $p(x_i, \theta_1, \theta_2, ..., \theta_r)$ si X es discreta o $f(x, \theta_1, \theta_2, ..., \theta_r)$ si es continua. Sean $\mu_1, \mu_2, ..., \mu_r$ los primeros r momentos poblacionales:

$$\mu_k = E(X^k) = \sum_{x_i \in R_X} x_i^k \, p(x_i, \theta_1, \theta_2, ..., \theta_r) \qquad (k = 1,2,...,r) \qquad \text{Si } X \text{ es discreta}$$

$$\mu_k = E(X^k) = \int_{-\infty}^{+\infty} x^k f(x, \theta_1, \theta_2, ..., \theta_r) dx \qquad (k = 1,2,...,r) \qquad \text{Si } X \text{ es continua,}$$

y sean

$M_k = \dfrac{1}{n} \sum_{i=1}^{n} X_i^k \quad (k = 1,2,...,r)$ los r primeros momentos muestrales para una muestra de tamaño n. Entonces ***el método de los momentos consiste en plantear el sistema de ecuaciones***:

$$\begin{cases} \mu_1 = M_1 \\ \mu_2 = M_2 \\ \vdots \quad \vdots \vdots \\ \mu_r = M_r \end{cases}$$

Es decir

$$\begin{cases} \displaystyle\sum_{x_i \in R_X} x_i \, p(x_i, \theta_1, \theta_2, ..., \theta_r) = \frac{1}{n} \sum_{i=1}^{n} X_i^1 \\ \displaystyle\sum_{x_i \in R_X} x_i^2 \, p(x_i, \theta_1, \theta_2, ..., \theta_r) = \frac{1}{n} \sum_{i=1}^{n} X_i^2 \\ \qquad \vdots \qquad\qquad \vdots \quad \vdots \\ \displaystyle\sum_{x_i \in R_X} x_i^r \, p(x_i, \theta_1, \theta_2, ..., \theta_r) = \frac{1}{n} \sum_{i=1}^{n} X_i^r \end{cases} \qquad \text{Si } X \text{ es discreta,}$$

o

$$\begin{cases} \displaystyle\int_{-\infty}^{+\infty} x f(x, \theta_1, \theta_2, ..., \theta_r) dx = \frac{1}{n} \sum_{i=1}^{n} X_i^1 \\ \displaystyle\int_{-\infty}^{+\infty} x^2 f(x, \theta_1, \theta_2, ..., \theta_r) dx = \frac{1}{n} \sum_{i=1}^{n} X_i^2 \\ \qquad \vdots \qquad\qquad \vdots \quad \vdots \\ \displaystyle\int_{-\infty}^{+\infty} x^r f(x, \theta_1, \theta_2, ..., \theta_r) dx = \frac{1}{n} \sum_{i=1}^{n} X_i^r \end{cases} \qquad \text{Si } X \text{ es continua.}$$

Resolviendo estos sistema de ecuaciones para los parámetros desconocidos $\theta_1, \theta_2, ..., \theta_r$ en función de la muestra aleatoria obtenemos los estimadores:

$$\begin{cases} \hat{\Theta}_1 = H_1(X_1, X_2, ..., X_n) \\ \hat{\Theta}_2 = H_2(X_1, X_2, ..., X_n) \\ \vdots \\ \hat{\Theta}_r = H_r(X_1, X_2, ..., X_n) \end{cases}$$

Ejemplos:

1- Sea *X* una variable aleatoria. Supongamos que *X* tiene ***distribución gama con parámetros*** σ ***y*** λ : $X \sim \Gamma(\sigma, \lambda)$, es decir su *fdp* está dada por:

$$f(x) = \begin{cases} \dfrac{1}{\sigma \Gamma(\lambda)} \left(\dfrac{x}{\sigma} \right)^{\lambda - 1} e^{-\frac{x}{\sigma}} & x > 0 \\ 0 & demás \quad valores \end{cases}$$

con $\sigma > 0$; $\lambda > 0$ y $\Gamma(\lambda) = \int_0^\infty x^{\lambda-1} e^{-x} dx$.

Sea $(X_1, X_2, ..., X_n)$ una muestra aleatoria de tamaño *n*. Deseamos calcular los estimadores de σ y λ dados por el método de los momentos.

Solución:

Como tenemos dos parámetros desconocidos a estimar, planteamos el sistema de ecuaciones:

$$\begin{cases} \mu_1 = M_1 \\ \mu_2 = M_2 \end{cases}$$

Se puede probar que

$$\mu_1 = \lambda.\sigma$$
$$\mu_2 = \lambda^2.\sigma^2 + \lambda.\sigma^2$$

Tenemos, entonces, el sistema de ecuaciones

$$\begin{cases} \lambda.\sigma = \dfrac{1}{n} \sum_{i=1}^{n} X_i \\ \lambda^2.\sigma^2 + \lambda.\sigma^2 = \dfrac{1}{n} \sum_{i=1}^{n} X_i^2 \end{cases} \Rightarrow \begin{cases} \lambda.\sigma = \overline{X} \\ \lambda^2.\sigma^2 + \lambda.\sigma^2 = \dfrac{1}{n} \sum_{i=1}^{n} X_i^2 \end{cases}$$

Reemplazando en la segunda ecuación: $\overline{X}^2+\sigma\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i^{\,2} \quad \Rightarrow \quad \sigma=\frac{\frac{1}{n}\sum_{i=1}^{n}X_i^{\,2}-\overline{X}^2}{\overline{X}}$

Y despejando λ de la primera ecuación y reemplazando la expresión hallada para σ

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

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

Solución:
Planteamos la ecuación: $\mu_1=M_1$
Sabemos que $\mu_1=E(X)=\frac{0+\theta}{2}=\frac{\theta}{2}$. Entonces $\frac{\theta}{2}=\overline{X} \quad \Rightarrow \quad \hat{\Theta}=2\overline{X}$
Observación: notar que el estimador $\hat{\Theta}=2\overline{X}$ es un estimador consistente de θ, pues
$E\left(\hat{\Theta}\right)=E\left(2\overline{X}\right)=2E\left(\overline{X}\right)=2\frac{\theta}{2}=\theta$ y $V\left(\hat{\Theta}\right)=V\left(2\overline{X}\right)=4V\left(\overline{X}\right)=4\frac{(\theta-0)^2}{12n}=\frac{\theta^2}{3n}\underset{n\to\infty}{\longrightarrow}0$

3- Sea una muestra aleatoria de una v.a. $X\sim N(\mu,\sigma^2)$.
Encuentra los estimadores de μ y σ por el método de momentos.

Solución:
Planteamos las ecuaciones

$$\begin{cases}\mu_1=M_1\\ \mu_2=M_2\end{cases} \quad \Rightarrow \quad \begin{cases}\mu=\overline{X}\\ E\left(X^2\right)=\frac{1}{n}\sum_{i=1}^{n}X_i^{\,2}\end{cases}$$

pero en general es válido que $V(X)=E(X^2)-\mu^2 \quad \Rightarrow \quad E(X^2)=V(X)+\mu$
Entonces las ecuaciones quedan

$$\begin{cases}\mu=\overline{X}\\ \sigma^2+\mu^2=\frac{1}{n}\sum_{i=1}^{n}X_i^{\,2}\end{cases} \quad \Rightarrow \quad \begin{cases}\hat{\mu}=\overline{X}\\ \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}X_i^{\,2}-\overline{X}^2\end{cases}$$

4- Sea una muestra aleatoria de una v.a. $X\sim \boldsymbol{N(0,\sigma^2)}$.
Hallar un estimador por el método de los momentos de σ^2

Solución: en este caso no es conveniente plantear $\mu_1=M_1$ pues quedaría

la ecuación $0 = \overline{X}$ que no conduce a nada.
Entonces podemos plantear $\mu_2 = M_2$ es decir

$$E(X^2) = \frac{1}{n}\sum_{i=1}^{n} X_i^{\,2} \quad \Rightarrow \quad \sigma^2 + 0 = \frac{1}{n}\sum_{i=1}^{n} X_i^{\,2} \quad \Rightarrow \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^{\,2}$$

<u>Observación:</u> si $\hat{\Theta}$ es un estimador por el método de los momentos de un parámetro θ, el estimador de los momentos de $g(\theta)$ es $g(\hat{\Theta})$, si $g(x)$ es una función inyectiva.
Por ejemplo, en el ejemplo anterior un estimador de σ por el método de los momentos sería $\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^{\,2}}$. Notar que $g(x) = \sqrt{x}$ es inyectiva para los reales positivos.

<u>Método de máxima verosimilitud</u>

Uno de los mejores métodos para obtener un estimador puntual de un parámetro es el método de máxima verosimilitud.

> Supongamos que *X* es una v.a. discreta con función de distribución de probabilidad $p(x,\theta)$, 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:
>
> $$L(x_1, x_2, ..., x_n, \theta) = P(X_1 = x_1)P(X_2 = x_2)...P(X_n = x_n) = p(x_1,\theta).p(x_2,\theta).....p(x_n,\theta)$$
>
> Notar que la función de verosimilitud es una función de θ.
> El estimador de máxima verosimilitud de θ es aquel valor de θ que maximiza la función de verosimilitud

<u>La interpretación del método sería</u>: el estimador de máxima verosimilitud es aquel valor del parámetro que maximiza la probabilidad de ocurrencia de los valores muestrales

La adaptación para el caso en que X es una v.a. continua sería la siguiente

> Supongamos que *X* es una v.a. continua con función de densidad de probabilidad $f(x,\theta)$, 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:
>
> $$L(x_1, x_2, ..., x_n, \theta) = f(x_1,\theta).f(x_2,\theta).....f(x_n,\theta)$$
>
> La función de verosimilitud es una función de θ.
> El estimador de máxima verosimilitud de θ es aquel valor de θ que maximiza la función de verosimilitud

<u>Notación:</u> abreviamos estimador de máxima verosimilitud con EMV

Ejemplos:

1- Sea una muestra aleatoria de una v.a. $X \sim B(1, p)$

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

Solución:

Si $X \sim B(1, p)$ entonces $P(X = k) = \binom{1}{k} p^k (1-p)^{1-k} \quad k = 0,1$

Planteamos la función de verosimilitud

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

Esto puede escribirse:

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

Para maximizar la función de verosimilitud y ***facilitar los cálculos*** tomamos el logaritmo natural de *L*
Pues maximizar *L* es equivalente a maximizar *ln(L)* y al tomar logaritmos transformamos productos en sumas.

Entonces

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

Y ahora podemos maximizar la función derivando e igualando a cero

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

de donde despejando *p*

$$p = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x} \qquad \text{la proporción de defectuosos en la muestra}$$

Por lo tanto se toma como estimador a $\hat{p} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

2- El tiempo de fallar T de una componente tiene una distribución exponencial con parámetro λ: $T \sim Exp(\lambda)$, es decir la *fdp* es

$$f(t; \lambda) = \begin{cases} \lambda e^{-\lambda t} & 0 \le t < \infty \\ 0 & demás \quad valores \end{cases}$$

Recordemos que la esperanza y varianza son:

$E(T) = 1/\lambda$ y $V(T) = 1/\lambda^2$, respectivamente.

Se desea calcular el estimador de máxima verosimilitud del parámetro λ para una muestra de tamaño *n*.

<u>Solución:</u>
La función de probabilidad es:
$L(t_1, t_2, ..., t_n; \lambda) = f(t_1; \lambda) f(t_2; \lambda) ... f(t_n; \lambda) = \left[\lambda e^{-\lambda t_1}\right] \times \left[\lambda e^{-\lambda t_2}\right] \times ... \times \left[\lambda e^{-\lambda t_n}\right]$,

que puede escribirse:

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

Nuevamente tomamos logaritmo natural

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

$$\frac{\partial \ln L(t_1, t_2, ..., t_n; \lambda)}{\partial \lambda} = n \frac{1}{\lambda} - \sum_{i=1}^{n} T_i = 0$$

de donde podemos despejar λ :

$\lambda = \dfrac{n}{\sum_{i=1}^{n} t_i} = \bar{t}$,   entonces el estimador de λ es $\hat{\lambda} = \dfrac{n}{\sum_{i=1}^{n} T_i}$

El método de máxima verosimilitud presenta, algunas veces, dificultades para maximizar la función de verosimilitud debido a que la ecuación obtenida a partir de $\frac{d}{d\theta} L(\theta) = 0$ no resulta fácil de resolver. O también puede ocurrir que los métodos de cálculo para maximizar $L(\theta)$ no son aplicables.

Por ejemplo:
Sea $(X_1, X_2, ..., X_n)$ una muestra aleatoria de tamaño *n* de una v.a. *X* donde $X \sim U[0, \theta]$, θ desconocido. Hallar el estimador de θ por el método máxima verosimilitud.
<u>Solución</u>:
La *f.d.p.* de *X* es

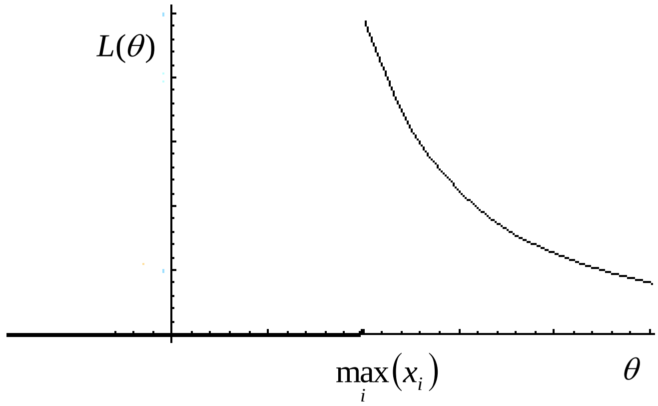
$$f(x) = \begin{cases} \dfrac{1}{\theta} & si \quad 0 < x < \theta \\ 0 & caso \quad contrario \end{cases}$$

Planteamos la función de verosimilitud

$$L(x_1, x_2, ... x_n, \theta) = \begin{cases} \dfrac{1}{\theta^n} & si \quad 0 < x_i < \theta \quad \forall i \\ 0 & caso \quad contrario \end{cases} = \begin{cases} \dfrac{1}{\theta^n} & si \quad \max_i (x_i) < \theta \\ 0 & caso \quad contrario \end{cases}$$

Si derivamos con respecto a θ obtenemos $\frac{d}{d\theta}\theta^{-n} = -\frac{n}{\theta^{n+1}}$ que es siempre menor que cero. Por lo tanto la función de verosimilitud es una función decreciente para todos los $\theta > \max_i(x_i)$

Si hacemos un gráfico de la función de verosimilitud



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

Por lo tanto $\hat{\Theta} = \max_i(x_i)$

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. Específicamente si tenemos para estimar *k* parámetros $\theta_1, \theta_2, \ldots \theta_k$, entonces la función de verosimilitud es una función de *k* variables $L(x_1, x_2, \ldots, x_n, \theta_1, \theta_2, \ldots \theta_k)$ y los estimadores de máxima verosimilitud $\hat{\Theta}_1, \hat{\Theta}_2, \ldots \hat{\Theta}_k$ se obtienen al plantear (si existen las derivadas parciales) y resolver el sistema de *k* ecuaciones con *k* incógnitas $\theta_1, \theta_2, \ldots \theta_k$

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

Ejemplo:

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

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

La función de verosimilitud para una muestra aleatoria de tamaño *n* es

$$L(x_1, x_2, \ldots, x_n; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_1-\mu}{\sigma}\right)^2} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_2-\mu}{\sigma}\right)^2} \ldots \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_n-\mu}{\sigma}\right)^2} =$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

Luego

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

y el sistema de ecuaciones de verosimilitud queda:

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

Resolvemos con respecto a y σ^2 :

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

Entonces los estimadores máxima verosimilitud de y σ^2 son

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

Propiedades de los estimadores máxima verosimilitud

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***

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

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

4-Los EMV cumplen la ***propiedad de invarianza*** es decir:

si $\hat{\Theta}$ es un EMV de un parámetro θ, el EMV de $g(\theta)$ es $g\left(\hat{\Theta}\right)$, si $g(x)$ es una función inyectiva.

Ejemplos:

1- Si consideramos nuevamente la situación considerada en el Ejemplo 2, donde teníamos una v.a. T cuya distribución es una exponencial: $T \sim Exp(\lambda)$, entonces, si queremos el EMV de la varianza poblacional, podemos calcularlo recordando que $V(T) = 1/\lambda^2$, es decir, $V(T) = g(\lambda) = 1/\lambda^2$. Vimos que $\hat{\lambda} = \dfrac{n}{\sum_{i=1}^{n} T_i} = \dfrac{1}{\overline{T}}$. Por lo tanto el EMV de la varianza es $\hat{\sigma}^2 = \dfrac{1}{\hat{\lambda}^2}$.

2- Sea $X_1, X_2, \ldots\ldots\ldots, X_n$ una muestra aleatoria de una v.a. $B(1, p)$. Un EMV de *p* es $\hat{p} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Se selecciona una muestra aleatoria de *n* cascos para ciclistas fabricados por cierta compañía.
Sea *X* : " el número entre los *n* que tienen defectos" , y *p* = P(el casco tiene defecto).
Supongamos que solo se observa *X* (el número de cascos con defectos).
Si *n* = 20 y *x* = 3, es la estimación de *p* es $\hat{p} = \frac{3}{20}$
El E.M.V. de la probabilidad *(1-p)*^{5}, de que ninguno de los siguientes cinco cascos que se examinen tenga defectos será $(1-\hat{p})^5$ y su estimación en este caso $\left(1-\frac{3}{20}\right)^5$