

## 8- Estimación puntual

### 8.1 – Introducción

Supongamos la siguiente situación: en una fábrica se producen artículos, el interés está en la producción de un día, específicamente, de todos los artículos producidos en un día nos interesa una característica determinada, si el artículo es o no defectuoso. Sea  $p$  la proporción de artículos defectuosos en la **población**, es decir en la producción de un día.

Tomamos una **muestra** de 25 artículos, podemos definir la v.a.  $X$ : “número de artículos defectuosos en la muestra”, y podemos asumir que  $X \sim B(25, p)$ .

En **Probabilidades** se **conocían todos los datos sobre la v.a.  $X$** , es decir conocíamos  $p$ . De esa forma podíamos responder preguntas como: ¿cuál es la probabilidad que entre los 25 artículos halla 5 defectuosos?. Si, por ejemplo,  $p = 0.1$  entonces calculábamos  $P(X = 5)$  donde  $X \sim B(25, 0.1)$ .

En **Estadística desconocemos las características de  $X$**  total o parcialmente, y a partir de la muestra de 25 artículos tratamos de **inferir** información sobre la distribución de  $X$ , o dicho de otra forma tratamos de inferir información sobre la **población**.

Por ejemplo, en estadística sabremos que  $X$  tiene distribución binomial pero **desconocemos  $p$** , y a partir de la muestra de 25 artículos trataremos de hallar información sobre  $p$ .

En Estadística nos haremos preguntas tales como: si en la muestra de 25 artículos se encontraron 5 defectuosos, ¿ese hecho me permite inferir que el verdadero  $p$  es 0.1?.

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. Estos métodos utilizan la información contenida en una **muestra** de la población para obtener conclusiones.

La inferencia estadística puede dividirse en dos grandes áreas: **estimación de parámetros y pruebas de hipótesis**.

### 8.2 – Muestreo aleatorio

En muchos problemas estadísticos es necesario utilizar una muestra de observaciones tomadas de la población de interés con objeto de obtener conclusiones sobre ella. A continuación se presenta la definición de algunos términos

Una **población** está formada por la totalidad de las observaciones en las cuales se tiene cierto interés

En muchos problemas de inferencia estadística es poco práctico o imposible, observar toda la población, en ese caso se toma una parte o subconjunto de la población

Una **muestra** es un subconjunto de observaciones seleccionada de una población

Para que las inferencias sean válidas, la muestra debe ser representativa de la población. Se selecciona una muestra aleatoria como el resultado de un mecanismo aleatorio. En consecuencia, la selección de una muestra es un experimento aleatorio, y cada observación de la muestra es el valor observado de una variable aleatoria. Las observaciones en la población determinan la distribución de probabilidad de la variable aleatoria.

Para definir muestra aleatoria, sea  $X$  la v.a. que representa el resultado de tomar una observación de la población. Sea  $f(x)$  la *f.d.p.* de la v.a.  $X$ . supongamos que cada observación en la muestra se ob-

tiene de manera independiente, bajo las mismas condiciones. Es decir, las observaciones de la muestra se obtienen al observar  $X$  de manera independiente bajo condiciones que no cambian, digamos  $n$  veces.

Sea  $X_i$  la variable aleatoria que representa la  $i$ -ésima observación. Entonces  $X_1, X_2, \dots, X_n$  constituyen una muestra aleatoria, donde los valores numéricos obtenidos son  $x_1, x_2, \dots, x_n$ . Las variables aleatorias en una muestra aleatoria son independientes, con la misma distribución de probabilidad  $f(x)$  debido a que cada observación se obtiene bajo las mismas condiciones. Es decir las funciones de densidad marginales de  $X_1, X_2, \dots, X_n$  son todas iguales a  $f(x)$  y por independencia, la distribución de probabilidad conjunta de la muestra aleatoria es el producto de las marginales  $f(x_1)f(x_2)\dots f(x_n)$

Las variables aleatorias  $(X_1, X_2, \dots, X_n)$  constituyen una **muestra aleatoria** de tamaño  $n$  de una v.a.  $X$  si  $X_1, X_2, \dots, X_n$  son independientes idénticamente distribuidas

El propósito de tomar una muestra aleatoria es obtener información sobre los parámetros desconocidos de la población. Por ejemplo, se desea alcanzar una conclusión acerca de la proporción de artículos defectuosos en la producción diaria de una fábrica. Sea  $p$  la proporción de artículos defectuosos en la población, para hacer una inferencia con respecto a  $p$ , se selecciona una muestra aleatoria (de un tamaño apropiado) y se utiliza la proporción observada de artículos defectuosos en la muestra para estimar  $p$ .

La proporción de la muestra  $\hat{p}$  se calcula dividiendo el número de artículos defectuosos en la muestra por el número total de artículos de la muestra. Entonces  $\hat{p}$  es una función de los valores observados en la muestra aleatoria. Como es posible obtener muchas muestras aleatorias de una población, el valor de  $\hat{p}$  cambiará de una a otra. Es decir  **$\hat{p}$  es una variable aleatoria**. Esta variable aleatoria se conoce como **estadístico**.

Un **estadístico** es cualquier función de la muestra aleatoria

### Estadísticos usuales

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

Si desconocemos  $\mu$  un estadístico que se utiliza para estimar ese parámetro es la **media o promedio muestral**

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

Análogamente si se desconoce  $\sigma^2$  un estadístico usado para tener alguna información sobre ese parámetro es la **varianza muestral** que se define como

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Otro estadístico es la **desviación estándar muestral**

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$$

Como un estadístico es una variable aleatoria, éste tiene una distribución de probabilidad, esperanza y varianza.

Una aplicación de los estadísticos es obtener **estimaciones puntuales** de los parámetros desconocidos de una distribución. Por ejemplo como se dijo antes se suelen estimar la media y la varianza de una población.

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  $\theta$  y al estadístico que se utiliza como estimador puntual de  $\theta$ , simbolizarlo con  $\hat{\Theta}$ .

Por lo tanto  $\hat{\Theta}$  es una función de la muestra aleatoria:  $\hat{\Theta} = h(X_1, X_2, \dots, X_n)$

Al medir la muestra aleatoria se obtienen  $x_1, x_2, \dots, x_n$ , y entonces **el valor que toma  $\hat{\Theta}$**  es  $\hat{\theta} = h(x_1, x_2, \dots, x_n)$  y se denomina **estimación puntual de  $\theta$**

El objetivo de la estimación puntual es seleccionar un número, a partir de los valores de la muestra, que sea el valor más probable de  $\theta$ .

Por ejemplo, supongamos que  $X_1, X_2, X_3, X_4$  es una muestra aleatoria de una v.a.  $X$ . Sabemos que  $X$  tiene distribución normal pero desconocemos  $\mu$ .

Tomamos como **estimador** de  $\mu$  al promedio muestral  $\bar{X}$ , es decir  $\hat{\mu} = \bar{X}$

Tomamos la muestra (medimos  $X_1, X_2, X_3, X_4$ ) y obtenemos  $x_1 = 24$ ,  $x_2 = 30$ ,  $x_3 = 27$ ,  $x_4 = 32$

Entonces la **estimación puntual** de  $\mu$  es  $\bar{x} = \frac{24 + 30 + 27 + 32}{4} = 28.25$

Si la varianza  $\sigma^2$  de  $X$  también es desconocida, un estimador puntual usual de  $\sigma^2$  es la varianza muestral, es decir  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ , para la muestra dada la estimación de  $\sigma^2$  es 12.25.

Otro parámetro que a menudo es necesario estimar es la proporción  $p$  de objetos de una población que cumplen una determinada característica.

En este caso el estimador puntual de  $p$  sería  $\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$  donde

$$X_i = \begin{cases} 1 & \text{si la } i\text{-ésima observación tiene la característica de interés} \\ 0 & \text{caso contrario} \end{cases} \quad i = 1, 2, \dots, n$$

Por lo tanto  $\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$  es la **proporción de objetos en la muestra** cumplen la característica de interés

Puede ocurrir que se tenga más de un estimador para un parámetro, por ejemplo para estimar la media muestral se pueden considerar el promedio muestral, o también la semisuma entre  $X_1$  y  $X_n$ , es decir  $\hat{\mu} = \frac{X_1 + X_n}{2}$ . En estos casos necesitamos de algún **criterio para decidir cuál es mejor estimador** de  $\mu$ .

### 8.3 – Criterios para evaluar estimadores puntuales

Lo que se desea de un estimador puntual es **que tome valores “próximos” al verdadero parámetro**.

Se dice que el estimador puntual  $\hat{\Theta}$  es un **estimador insesgado** del parámetro  $\theta$  si  $E(\hat{\Theta}) = \theta$  cualquiera sea el valor verdadero de  $\theta$

Podemos exigir que el estimador  $\hat{\Theta}$  tenga una distribución cuya media sea  $\theta$ .

La diferencia  $E(\hat{\Theta}) - \theta$  se conoce como **sesgo de estimador  $\hat{\Theta}$** . Anotamos  $b(\hat{\Theta}) = E(\hat{\Theta}) - \theta$

**Notar que si un estimador es insesgado entonces su sesgo es cero**

Ejemplos:

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

Si desconocemos  $\mu$  un estadístico que se utiliza usualmente para estimar este parámetro es la **media**

**o promedio muestral**  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ . Veamos si es un estimador insesgado de  $\mu$ . Debemos ver si  $E(\bar{X}) = \mu$ .

Usamos las propiedades de la esperanza, particularmente la propiedad de linealidad.

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

Pero, tratándose de las componentes de una muestra aleatoria es:

$$E(X_i) = E(X) = \mu \quad \forall i = 1, 2, \dots, n. \quad \text{Luego:}$$

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

2- Sea  $X$  una variable aleatoria asociada con alguna característica de los individuos de una población

y sean  $E(X) = \mu$  y  $V(X) = \sigma^2$ . Sea  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  la varianza muestral (con

$\bar{X} = \left(\sum_{i=1}^n X_i\right)/n$  la esperanza muestral) para una muestra aleatoria de tamaño  $n$ ,  $(X_1, X_2, \dots, X_n)$ .

Entonces  $E(S^2) = \sigma^2$  es decir  $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  es un estimador insesgado de  $V(X) = \sigma^2$  pues:

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

Reescribiremos la suma de una forma más conveniente. Sumamos y restamos  $\mu$  y desarrollamos el cuadrado:

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n (X_i - \mu + \mu - \bar{X})^2 = \sum_{i=1}^n ([X_i - \mu] + [\mu - \bar{X}])^2 = \\ &= \sum_{i=1}^n \left\{ [X_i - \mu]^2 + 2[X_i - \mu][\mu - \bar{X}] + [\mu - \bar{X}]^2 \right\} = \sum_{i=1}^n [X_i - \mu]^2 + 2[\mu - \bar{X}] \sum_{i=1}^n [X_i - \mu] + n[\mu - \bar{X}]^2 = \\ &= \sum_{i=1}^n [X_i - \mu]^2 + 2[\mu - \bar{X}]n[\bar{X} - \mu] + n[\mu - \bar{X}]^2 = \sum_{i=1}^n [X_i - \mu]^2 - 2n[\mu - \bar{X}]^2 + n[\mu - \bar{X}]^2. \end{aligned}$$

Esto es:

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

Entonces:

$$\begin{aligned} E(S^2) &= \frac{1}{n-1} E\left(\sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{1}{n-1} E\left(\sum_{i=1}^n [X_i - \mu]^2 - n[\mu - \bar{X}]^2\right) = \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n E[X_i - \mu]^2 - nE[\bar{X} - \mu]^2\right) = \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n V(X_i) - nE[\bar{X} - E(\bar{X})]^2\right) = \frac{1}{n-1} \left(\sum_{i=1}^n V(X_i) - nV(\bar{X})\right) = \frac{1}{n-1} \left(n\sigma^2 - n\frac{\sigma^2}{n}\right), \end{aligned}$$

donde en la última igualdad tuvimos en cuenta que  $V(X_i) = V(X) = \sigma^2 \quad \forall i = 1, 2, \dots, n$  y que

$$V(\bar{X}) = \frac{\sigma^2}{n}. \text{ Luego llegamos a lo que se deseaba demostrar: } E(S^2) = \sigma^2.$$

3- Supongamos que tomamos como estimador de  $\sigma^2$  a  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$

$$\text{Entonces notar que podemos escribir } \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{n} \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{n-1}{n} S^2$$

$$\text{Por lo tanto } E(\hat{\sigma}^2) = E\left(\frac{n-1}{n} S^2\right) = \frac{n-1}{n} E(S^2) = \frac{n-1}{n} \sigma^2 \neq \sigma^2$$

Es decir  $\hat{\sigma}^2$  no es un estimador insesgado de  $\sigma^2$ , *es sesgado*, y su sesgo es

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

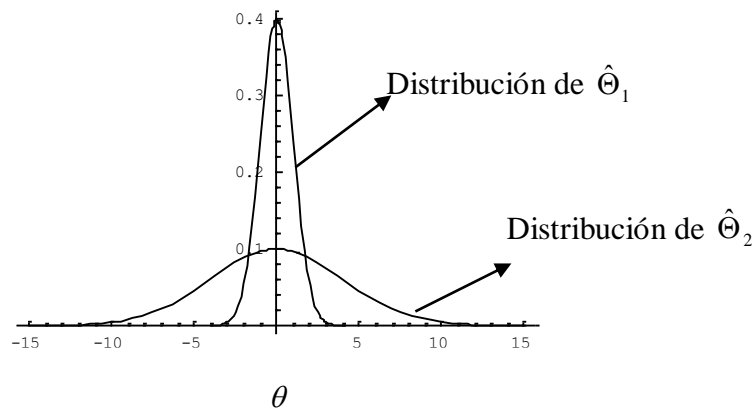
Como el sesgo es negativo el estimador tiende a subestimar el valor de verdadero parámetro

En ocasiones hay más de un estimador insesgado de un parámetro  $\theta$

Por lo tanto necesitamos un método para seleccionar un estimador entre varios estimadores insesgados.

### **Varianza y error cuadrático medio de un estimador puntual**

Supongamos que  $\hat{\theta}_1$  y  $\hat{\theta}_2$  son dos estimadores insesgados de un parámetro  $\theta$ . Esto indica que la distribución de cada estimador está centrada en el verdadero parámetro  $\theta$ . Sin embargo las varianzas de estas distribuciones pueden ser diferentes. La figura siguiente ilustra este hecho.



Como  $\hat{\theta}_1$  tiene menor varianza que  $\hat{\theta}_2$ , entonces es más probable que el estimador  $\hat{\theta}_1$  produzca una estimación más cercana al verdadero valor de  $\theta$ . Por lo tanto si tenemos dos estimadores insesgados se seleccionará aquel que tenga menor varianza.

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

Estimamos al parámetro  $\mu$  con la **media o promedio muestral**  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ . Sabemos que es un

estimador insesgado de  $\mu$ . Anotamos  $\hat{\mu}_1 = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$

Supongamos que tomamos otro estimador para  $\mu$ , lo anotamos  $\hat{\mu}_2 = \frac{X_1 + X_n}{2}$

Entonces como

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

$\hat{\mu}_2 = \frac{X_1 + X_n}{2}$  es también un estimador insesgado de  $\mu$

¿Cuál de los dos estimadores es mejor?

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

Ya sabemos cuál es la varianza de  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  (se la halló para T.C.L.):

$$V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i),$$

donde en la última igualdad hemos tenido en cuenta que, por tratarse de una muestra aleatoria, las  $X_i$  con  $i=1, 2, \dots, n$  son variables aleatorias independientes y, en consecuencia, la varianza de la suma de ellas es la suma de las varianzas. Si tenemos en cuenta que además todas tienen la misma distribución que  $X$  y por lo tanto la misma varianza:

$$V(X_i) = V(X) = \sigma^2 \quad \forall i = 1, 2, \dots, n, \text{ tenemos}$$

$$V(\bar{X}) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}.$$

Análogamente calculamos la varianza de  $\hat{\mu}_2 = \frac{X_1 + X_n}{2}$ :

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

Vemos que si  $n > 2$  entonces  $V(\hat{\mu}_1) < V(\hat{\mu}_2)$ . Por lo tanto si  $n > 2$  es mejor estimador  $\hat{\mu}_1$

Supongamos ahora que  $\hat{\Theta}_1$  y  $\hat{\Theta}_2$  son dos estimadores de un parámetro  $\theta$  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.

El **error cuadrático medio** de un estimador  $\hat{\Theta}$  de un parámetro  $\theta$  está definido como

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

El error cuadrático medio puede escribirse de la siguiente forma:

$$ECM(\hat{\Theta}) = V(\hat{\Theta}) + (b(\hat{\Theta}))^2$$

Dem.) Por definición  $ECM(\hat{\Theta}) = E\left[(\hat{\Theta} - \theta)^2\right]$ . Sumamos y restamos el número  $E(\hat{\Theta})$ :

$ECM(\hat{\Theta}) = E\left[(\hat{\Theta} - E(\hat{\Theta}) + E(\hat{\Theta}) - \theta)^2\right]$ , y desarrollamos el cuadrado:

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

Aplicamos propiedades de la esperanza:

$$= E\left[(\hat{\Theta} - E(\hat{\Theta}))^2\right] + \underbrace{(E(\hat{\Theta}) - \theta)^2}_{b(\hat{\Theta})^2} + 2(E(\hat{\Theta}) - \theta) \underbrace{E(\hat{\Theta} - E(\hat{\Theta}))}_0 = V(\hat{\Theta}) + (b(\hat{\Theta}))^2$$

El error cuadrático medio es un criterio importante para comparar estimadores.

Si  $\hat{\Theta}_1$  y  $\hat{\Theta}_2$  son dos estimadores de un parámetro  $\theta$ .

La eficiencia relativa de  $\hat{\Theta}_2$  con respecto a  $\hat{\Theta}_1$  se define como  $\frac{ECM(\hat{\Theta}_1)}{ECM(\hat{\Theta}_2)}$

Si la eficiencia relativa es menor que 1 entonces  $\hat{\Theta}_1$  tiene menor error cuadrático medio que  $\hat{\Theta}_2$

**Por lo tanto  $\hat{\Theta}_1$  es más eficiente que  $\hat{\Theta}_2$**

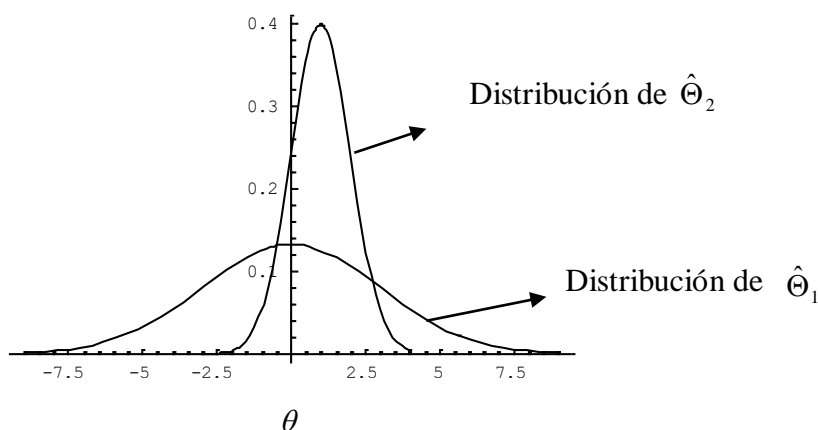
Observaciones:

1- Si  $\hat{\Theta}$  es un estimador insesgado de  $\theta$ , entonces  $ECM(\hat{\Theta}) = V(\hat{\Theta})$

2- A veces es preferible utilizar estimadores sesgados que estimadores insesgados, si es que tienen un error cuadrático medio menor.

En el error cuadrático medio se consideran tanto la varianza como el sesgo del estimador.

Si  $\hat{\theta}_1$  y  $\hat{\theta}_2$  son dos estimadores de un parámetro  $\theta$ , 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 error cuadrático medio de cada uno, y tomar el que tenga menor error cuadrático medio. Pues puede ocurrir que  $\hat{\theta}_2$ , aunque sea sesgado, al tener menor varianza tome valores mas cercanos al verdadero parámetro que  $\hat{\theta}_1$



Ejemplo:

Supóngase que  $\hat{\theta}_1$ ,  $\hat{\theta}_2$  y  $\hat{\theta}_3$  son dos estimadores de un parámetro  $\theta$ , y que  $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$ ;  $E(\hat{\theta}_3) \neq \theta$ ,  $V(\hat{\theta}_1) = 10$ ,  $V(\hat{\theta}_2) = 6$  y  $E[(\hat{\theta}_3 - \theta)^2] = 4$ . Haga una comparación de estos estimadores. ¿Cuál prefiere y por qué?

Solución: Calculamos el error cuadrático medio de cada estimador

$$ECM(\hat{\theta}_1) = V(\hat{\theta}_1) = 10 \text{ pues } \hat{\theta}_1 \text{ es insesgado}$$

$$ECM(\hat{\theta}_2) = V(\hat{\theta}_2) = 6 \text{ pues } \hat{\theta}_2 \text{ es insesgado}$$

$$ECM(\hat{\theta}_3) = E[(\hat{\theta}_3 - \theta)^2] = 4 \text{ es dato}$$

En consecuencia  $\hat{\theta}_3$  es el mejor estimador de los tres dados porque tiene menor error cuadrático medio.

### Consistencia de estimadores puntuales

Sea  $\hat{\theta}_n$  un estimador del parámetro  $\theta$ , basado en una muestra aleatoria  $(X_1, X_2, \dots, X_n)$  de tamaño  $n$ . Se dice que  $\hat{\theta}_n$  es un estimador consistente de  $\theta$  si

$$\lim_{n \rightarrow \infty} P(|\hat{\theta}_n - \theta| \geq \varepsilon) = 0 \text{ para todo } \varepsilon > 0$$

Observación:



Este tipo de convergencia, que involucra a una sucesión de variables aleatorias, se llama **convergencia en probabilidad** y es la misma que consideramos en relación a la **ley de los grandes números**.

Suele escribirse también  $\hat{\Theta}_n \xrightarrow{P} \theta$ .

Este tipo de convergencia debe distinguirse de la considerada en relación al teorema central del límite. En este último caso teníamos una sucesión de distribuciones:  $F_{Z_n}(z) = P(Z_n \leq z)$  y se considera el límite  $\lim_{n \rightarrow \infty} F_{Z_n}(z) = \lim_{n \rightarrow \infty} P(Z_n \leq z) = \Phi(z)$ .

Se habla, entonces, de **convergencia en distribución** y suele indicarse  $Z_n \xrightarrow{d} Z \sim N(0,1)$ .

**Teorema.** Sea  $\hat{\Theta}_n$  un estimador del parámetro  $\theta$  basado en una muestra aleatoria  $(X_1, X_2, \dots, X_n)$ . Si  $\lim_{n \rightarrow \infty} E(\hat{\Theta}_n) = \theta$  y  $\lim_{n \rightarrow \infty} V(\hat{\Theta}_n) = 0$ , entonces  $\hat{\Theta}_n$  es un estimador consistente de  $\theta$ .

Dem.)

Utilizamos la desigualdad de Chebyshev  $\forall \varepsilon > 0$ :

$$P(|\hat{\Theta}_n - \theta| \geq \varepsilon) \leq \frac{E(\hat{\Theta}_n - \theta)^2}{\varepsilon^2} = \frac{1}{\varepsilon^2} ECM(\hat{\Theta}_n) = \frac{1}{\varepsilon^2} [V(\hat{\Theta}_n) + b(\hat{\Theta}_n)^2]$$

Entonces, al tomar el límite  $\lim_{n \rightarrow \infty}$  y teniendo presente que  $\lim_{n \rightarrow \infty} E(\hat{\Theta}_n) = \theta$  y  $\lim_{n \rightarrow \infty} V(\hat{\Theta}_n) = 0$ , vemos que  $\lim_{n \rightarrow \infty} P(|\hat{\Theta}_n - \theta| \geq \varepsilon) = 0 \quad \forall \varepsilon > 0$ , es decir  $\hat{\Theta}_n$  es un estimador convergente de  $\theta$ .

**Ejemplo:**

Sea  $X$  una variable aleatoria que describe alguna característica numérica de los individuos de una población y sean  $\mu = E(X)$  y  $\sigma^2 = V(X)$  la esperanza poblacional y la varianza poblacional, respectivamente. Sea  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  la esperanza muestral basada en una muestra aleatoria  $(X_1, X_2, \dots, X_n)$ . Entonces  $\bar{X}$  es un estimador consistente de la esperanza poblacional  $\mu = E(X)$ .

Sabemos que

$$a) E(\bar{X}) = \mu = E(X) \quad \forall n$$

$$b) V(\bar{X}) = \frac{\sigma^2}{n} = \frac{V(X)}{n} \quad \forall n$$

La propiedad a) ya me dice que  $\bar{X}$  es un estimador insesgado de  $\mu = E(X)$ .

Por otra parte si a) vale para todo  $n$ , también vale en particular en el límite  $n \rightarrow \infty$ :

$$\lim_{n \rightarrow \infty} E(\bar{X}) = \mu = E(X).$$

Además, de b) deducimos inmediatamente que

$$\lim_{n \rightarrow \infty} V(\bar{X}) = 0.$$

Por lo tanto vemos que  $\bar{X}$  es un estimador consistente de  $\mu = E(X)$ .

## 8.4 – Métodos de estimación puntual

Los criterios anteriores establecen propiedades que es deseable que sean verificadas por los estimadores. Entre dos estimadores posibles para un dado parámetro poblacional es razonable elegir aquél que cumple la mayor cantidad de criterios o alguno en particular que se considera importante para el problema que se esté analizando. Sin embargo estos criterios no nos enseñan por sí mismos a construir los estimadores. Existen una serie de métodos para construir estimadores los cuales en general se basan en principios básicos de razonabilidad. Entre éstos podemos mencionar:

- **Método de los momentos**

- **Método de máxima verosimilitud**

### Método de los momentos

Se puede probar usando la desigualdad de Chebyshev el siguiente resultado:

Ley débil de los grandes números:

Sean  $(X_1, X_2, \dots, X_n)$   $n$  variables aleatorias independientes todas las cuales tienen la misma esperanza  $\mu = E(X)$  y varianza  $\sigma^2 = V(X)$ . Sea  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ . Entonces

$$\lim_{n \rightarrow \infty} P(|\bar{X} - \mu| \geq \varepsilon) = 0$$

Decimos que  $\bar{X}$  converge a  $\mu$  en probabilidad y lo indicamos:  $\bar{X} \xrightarrow{p} \mu$ .

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

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

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

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

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

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

$$\lim_{n \rightarrow \infty} P(|M_k - \mu_k| \geq \varepsilon) = 0 \quad (k = 0, 1, 2, \dots).$$

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

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

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

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

y sean

$M_k = \frac{1}{n} \sum_{i=1}^n X_i^k \quad (k = 1, 2, \dots, r)$  los  $r$  primeros momentos muestrales para una muestra de tamaño  $n$  ( $X_1, X_2, \dots, X_n$ ). Entonces **el método de los momentos consiste en plantear el sistema de ecuaciones:**

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

Es decir

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

o

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

Resolviendo este sistema de ecuaciones para los parámetros desconocidos  $\theta_1, \theta_2, \dots, \theta_r$  en función de la muestra aleatoria  $(X_1, X_2, \dots, X_n)$  obtenemos los estimadores:

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

#### Observación:

En la forma que presentamos aquí el método necesitamos conocer la forma de la *fdp* poblacional, por lo tanto estamos frente a un caso de estimación puntual **paramétrica**.

#### Ejemplos:

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

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

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

Sea  $(X_1, X_2, \dots, X_n)$  una muestra aleatoria de tamaño  $n$ . Deseamos calcular los estimadores de  $\sigma$  y  $\lambda$  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

$$\begin{aligned} \mu_1 &= \lambda \cdot \sigma \\ \mu_2 &= \lambda^2 \cdot \sigma^2 + \lambda \cdot \sigma^2 \end{aligned}$$

Tenemos, entonces, el sistema de ecuaciones

$$\begin{cases} \lambda \cdot \sigma = \frac{1}{n} \sum_{i=1}^n X_i \\ \lambda^2 \cdot \sigma^2 + \lambda \cdot \sigma^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \end{cases} \Rightarrow \begin{cases} \lambda \cdot \sigma = \bar{X} \\ \lambda^2 \cdot \sigma^2 + \lambda \cdot \sigma^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 \end{cases}$$

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

Y despejando  $\lambda$  de la primera ecuación y reemplazando la expresión hallada para  $\sigma$

$$\begin{cases} \hat{\lambda} = \frac{n\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ \hat{\sigma} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n\bar{X}} \end{cases}$$

2- Sea  $(X_1, X_2, \dots, X_n)$  una muestra aleatoria de tamaño  $n$  de una v.a.  $X$  donde  $X \sim U[0, \theta]$ ,  $\theta$  desconocido. Hallar el estimador de  $\theta$  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  $\frac{\theta}{2} = \bar{X} \Rightarrow \hat{\Theta} = 2\bar{X}$

Observación: notar que el estimador  $\hat{\Theta} = 2\bar{X}$  es un estimador consistente de  $\theta$ , pues

$$E(\hat{\Theta}) = E(2\bar{X}) = 2E(\bar{X}) = 2\frac{\theta}{2} = \theta \quad \text{y} \quad V(\hat{\Theta}) = V(2\bar{X}) = 4V(\bar{X}) = 4\frac{(\theta-0)^2}{12n} = \frac{\theta^2}{3n} \xrightarrow{n \rightarrow \infty} 0$$

3- Sea  $(X_1, X_2, \dots, X_n)$  una muestra aleatoria de una v.a.  $X \sim N(\mu, \sigma^2)$ .

Encuentra los estimadores de  $\mu$  y  $\sigma$  por el método de momentos.

Solución:

Planteamos las ecuaciones

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

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

Entonces las ecuaciones quedan

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

4- Sea  $(X_1, X_2, \dots, X_n)$  una muestra aleatoria de una v.a.  $X \sim N(0, \sigma^2)$ .

Hallar un estimador por el método de los momentos de  $\sigma^2$

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

la ecuación  $0 = \bar{X}$  que no conduce a nada.

Entonces podemos plantear  $\mu_2 = M_2$  es decir

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

**Observación:** si  $\hat{\theta}$  es un estimador por el método de los momentos de un parámetro  $\theta$ , 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  $\sigma$  por el método de los momentos sería

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}. \text{ Notar que } g(x) = \sqrt{x} \text{ es inyectiva para los reales positivos.}$$

### **Método de máxima verosimilitud**

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  $\theta$  es un parámetro desconocido. Sean  $x_1, x_2, \dots, x_n$  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:

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

Notar que la función de verosimilitud es una función de  $\theta$ .

El estimador de máxima verosimilitud de  $\theta$  es aquel valor de  $\theta$  que maximiza la función de verosimilitud

**La interpretación del método sería:** 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  $\theta$  es un parámetro desconocido. Sean  $x_1, x_2, \dots, x_n$  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:

$$L(x_1, x_2, \dots, x_n, \theta) = f(x_1, \theta) \cdot f(x_2, \theta) \dots f(x_n, \theta)$$

La función de verosimilitud es una función de  $\theta$ .

El estimador de máxima verosimilitud de  $\theta$  es aquel valor de  $\theta$  que maximiza la función de verosimilitud

**Notación:** abreviamos estimador de máxima verosimilitud con EMV

### Ejemplos:

1- Sea  $(X_1, X_2, \dots, X_n)$  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

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

Esto puede escribirse:

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

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

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

Y ahora podemos maximizar la función derivando e igualando a cero

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

de donde despejando  $p$

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

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

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

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

Recordemos que la esperanza y varianza son:

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

Se desea calcular el estimador de máxima verosimilitud del parámetro  $\lambda$  para una muestra de tamaño  $n$ .

Solución:

La función de probabilidad es:

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

que puede escribirse:

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

Nuevamente tomamos logaritmo natural

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

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

de donde podemos despejar  $\lambda$  :

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

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  $(X_1, X_2, \dots, X_n)$  una muestra aleatoria de tamaño  $n$  de una v.a.  $X$  donde  $X \sim U[0, \theta]$ ,  $\theta$  desconocido. Hallar el estimador de  $\theta$  por el método máxima verosimilitud.

Solución:

La f.d.p. de  $X$  es

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

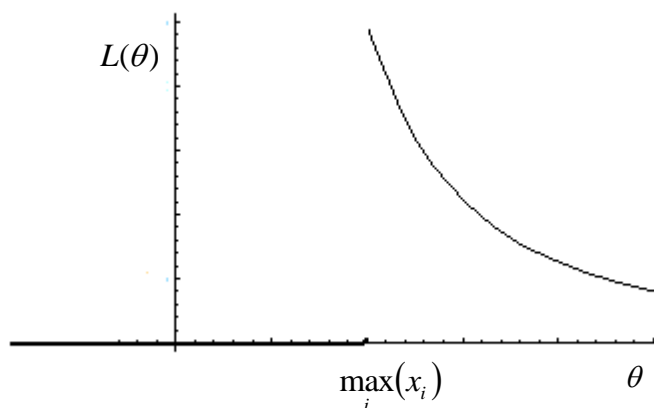
Planteamos la función de verosimilitud

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



Si derivamos con respecto a  $\theta$  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  $\theta_1, \theta_2, \dots, \theta_k$ , entonces la función de verosimilitud es una función de  $k$  variables  $L(x_1, x_2, \dots, x_n, \theta_1, \theta_2, \dots, \theta_k)$  y los estimadores de máxima verosimilitud  $\hat{\Theta}_1, \hat{\Theta}_2, \dots, \hat{\Theta}_k$  se obtienen al plantear ( si existen las derivadas parciales) y resolver el sistema de  $k$  ecuaciones con  $k$  incógnitas  $\theta_1, \theta_2, \dots, \theta_k$

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

Ejemplo:

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

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

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

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

Luego

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

y el sistema de ecuaciones de verosimilitud queda:

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

Resolvemos con respecto a  $\mu$  y  $\sigma^2$  :

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

Entonces los estimadores máxima verosimilitud de  $\mu$  y  $\sigma^2$  son

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

### Propiedades de los estimadores máxima verosimilitud

1- Los EMV pueden ser *sesgados*, pero en general si  $\hat{\Theta}$  es el EMV de un parámetro  $\theta$  basado en una muestra de tamaño  $n$ , entonces  $\lim_{n \rightarrow \infty} E(\hat{\Theta}) = \theta$ , es decir son *asintóticamente insesgados*

2- Bajo condiciones bastantes generales se puede probar que los EMV son *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  $\theta$ , el EMV de  $g(\theta)$  es  $g(\hat{\Theta})$ , 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:  $T \sim \text{Exp}(\lambda)$ , 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

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^n T_i} = \frac{1}{\bar{T}}. \text{ Por lo tanto el EMV de la varianza es } \hat{\sigma}^2 = \frac{1}{\hat{\lambda}^2}.$$

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

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  $p = P(\text{el casco tiene defecto})$ .

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  $\left(1-\frac{3}{20}\right)^5$