

# TEMA 7 ESTIMACIÓN PUNTUAL Y POR INTERVALOS

## ESTIMACIÓN PUNTUAL

Pablo Buenestado

Curso 2019-2020 Primavera

Departamento de Matemáticas (UPC)

# Índice

## 1 ESTIMACIÓN PUNTUAL

- Introducción
- Algunos conceptos generales de estimación puntual
  - Estimadores insesgados
  - Estimadores con varianza mínima
  - Reporte de una estimación puntual: el error estándar
- Métodos de estimación puntual
  - El método de momentos
  - Estimación de máxima verosimilitud
- Ejercicios
- Problemas

# Esquema

## 1 ESTIMACIÓN PUNTUAL

- Introducción
- Algunos conceptos generales de estimación puntual
  - Estimadores insesgados
  - Estimadores con varianza mínima
  - Reporte de una estimación puntual: el error estándar
- Métodos de estimación puntual
  - El método de momentos
  - Estimación de máxima verosimilitud
- Ejercicios
- Problemas

Dado un parámetro de interés, tal como la media  $\mu$  o la proporción  $p$  de una población, el objetivo de la estimación puntual es utilizar una muestra para calcular un número que representa en cierto sentido una buena suposición del valor verdadero del parámetro.

El número resultante se llama estimación puntual.

Empezaremos presentando algunos conceptos generales de estimación puntual. Y, posteriormente, describiremos e ilustraremos dos métodos importantes de obtener estimaciones puntuales: el método de momentos y el método de máxima verosimilitud.

# Esquema

## 1 ESTIMACIÓN PUNTUAL

- Introducción
- Algunos conceptos generales de estimación puntual
  - Estimadores insesgados
  - Estimadores con varianza mínima
  - Reporte de una estimación puntual: el error estándar
- Métodos de estimación puntual
  - El método de momentos
  - Estimación de máxima verosimilitud
- Ejercicios
- Problemas

El objetivo de la inferencia estadística casi siempre es sacar algún tipo de conclusión sobre uno o más parámetros (características de la población).

Para hacer eso un investigador tiene que obtener datos muestrales de cada una de las poblaciones estudiadas.

Las conclusiones pueden entonces basarse en los valores calculados de varias cantidades muestrales.

Por ejemplo, sea  $\mu$  (un parámetro) la resistencia a la ruptura promedio verdadera de conexiones alámbricas utilizadas en la unión de obleas semiconductoras.

Se podría tomar una muestra aleatoria de  $n = 10$  conexiones y determinar la resistencia a la ruptura de cada una y se tendrían las resistencias observadas  $x_1, x_2, \dots, x_{10}$ .

La resistencia a la ruptura media muestral  $\bar{x}$  se utilizaría entonces para sacar una conclusión con respecto al valor de  $\mu$ .

Asimismo, si  $\sigma^2$  es la varianza de la distribución de la resistencia a la ruptura (varianza de la población, otro parámetro), el valor de la varianza muestral  $\hat{s}^2$  se utiliza para inferir algo sobre  $\sigma^2$ .

Cuando se discuten los métodos y conceptos generales de inferencia, es conveniente disponer de un símbolo genérico para el parámetro de interés.

Utilizamos la letra griega  $\theta$  para este propósito.

El objetivo de la estimación puntual es seleccionar un solo número, con base en los datos muestrales, que represente un valor sensible de  $\theta$ .

Supongamos, por ejemplo, que el parámetro de interés es  $\mu$ , la vida útil promedio verdadera de baterías de un tipo.

Una muestra aleatoria de  $n = 3$  baterías podría dar las vidas útiles (horas) observadas  $x_1 = 5.0$ ,  $x_2 = 6.4$ ,  $x_3 = 5.9$ .

El valor calculado de la vida útil media muestral es  $\bar{x} = 5.77$  y es razonable considerar 5.77 como un valor muy factible de  $\mu$ , la "mejor suposición" del valor de  $\mu$  basado en la información muestral disponible.



Supongamos que se desea estimar un parámetro de una población (p. ej.,  $\mu$  o  $\sigma$ ) con una muestra aleatoria de tamaño  $n$ .

Recordemos por el tema previo de que antes que los datos estén disponibles, las observaciones muestrales deben ser consideradas como variables aleatorias  $X_1, X_2, \dots, X_n$ .

Se deduce que cualquier función de las  $X_i$ , es decir, cualquier estadístico, tal como la media muestral  $\bar{X}$  o la desviación estándar muestral  $\hat{S}$  también es una variable aleatoria.

### Definición

Una **estimación puntual** de un parámetro  $\theta$  es un número único que puede ser considerado como un valor sensible de  $\theta$ .

Se obtiene una estimación puntual seleccionando un estadístico apropiado y calculando su valor con los datos muestrales dados.

El estadístico seleccionado se llama **estimador puntual** de  $\theta$ .

En el ejemplo anterior sobre la batería, el estimador utilizado para obtener la estimación puntual de  $\mu$  fue  $\bar{X}$  y la estimación puntual de  $\mu$  fue 5.77.

Si las tres vidas útiles hubieran sido  $x_1 = 5.6$ ,  $x_2 = 4.5$ ,  $x_3 = 6.1$ , el uso del estimador  $\bar{X}$  habría dado por resultado la estimación  $\bar{x} = (5.6 + 4.5 + 6.1)/3 = 5.40$ .

El símbolo  $\hat{\theta}$  ("teta testada") se utiliza comúnmente para denotar tanto la estimación de  $\theta$  como la estimación puntual que resulta de una muestra dada.

Por tanto,  $\hat{\mu} = \bar{X}$  se lee como "el estimador puntual de  $\mu$  es la media muestral  $\bar{X}$ ".

La proposición "la estimación puntual de  $\mu$  es 5.77" se escribe concisamente como  $\hat{\mu} = 5.77$ .

Observemos que cuando se escribe  $\hat{\theta} = 72.5$ , no hay ninguna indicación de cómo se obtuvo esta estimación puntual (qué estadístico se utilizó). Se recomienda reportar tanto el estimador como la estimación resultante.

**Ejemplo:** Un fabricante automotriz ha producido un nuevo tipo de defensa, la que se presume absorbe impactos con menos daño que las defensas previas.

El fabricante ha utilizado esta defensa en una secuencia de 25 choques controlados con un muro, cada uno a 10 mph, utilizando uno de sus modelos de coche compacto.

Sea  $X$  = el número de choques que no provocaron daños visibles al automóvil.

El parámetro que tiene que ser estimado es  $p$  = la proporción de todos los choques que no provocaron daños [alternativamente,  $p = P(\text{ningún daño en un choque})$ ].

Si se observa que  $X$  es  $x = 15$ , el estimador y estimación más razonables son

$$\text{estimador } \hat{p} = \frac{X}{n}, \quad \text{estimación} = \frac{x}{n} = \frac{15}{25} = 0.60$$

Si por cada parámetro de interés hubiera sólo un estimador puntual razonable, no habría mucho que discutir sobre la estimación puntual.

En la mayoría de los problemas, sin embargo, habrá más de un estimador razonable.

**Ejemplo:** Consideremos 20 observaciones correspondientes al voltaje de ruptura, que es una variable aleatoria normal con  $\mu$  desconocida:

24.46, 25.61, 26.25, 26.42, 26.66, 27.15, 27.31, 27.54, 27.74, 27.94  
27.98, 28.04, 28.28, 28.49, 28.50, 28.87, 29.11, 29.13, 29.50, 30.88

Como las distribuciones normales son simétricas,  $\mu$  también es la mediana de la distribución.

Se supone entonces que las observaciones dadas son el resultado de una muestra aleatoria  $X_1, X_2, \dots, X_{20}$  de esta distribución normal.

Consideremos los siguientes estimadores:

- Ⓐ Estimador =  $\bar{X}$ , estimación =  $\bar{x} = \sum x_i/n = 27.793$
- Ⓑ Estimador =  $\tilde{X}$  (mediana),  
estimación =  $\tilde{x} = (27.94 + 27.98)/2 = 27.960$
- Ⓒ Estimador =  $[\min(X_i) + \max(X_i)]/2$ ,  
estimación =  $[\min(x_i) + \max(x_i)]/2 = (24.46 + 30.88)/2 = 27.670$

### Ejemplo (continuación):

Cada uno de los estimadores a), b) y c) utiliza una medición diferente del centro de la muestra para estimar  $\mu$ .

¿Cuál de las estimaciones se acerca más al valor verdadero?

No se puede responder esta pregunta sin conocer el valor verdadero.

Una pregunta que se puede hacer es:

¿Qué estimador, cuando se utiliza en otras muestras de  $X_i$ , tiende a producir estimaciones cercanas al valor verdadero?

En breve se considerará este tipo de pregunta.

En el mejor de todos los mundos posibles, se podría hallar un estimador  $\hat{\theta}$  con el cual  $\hat{\theta} = \theta$  siempre.

Sin embargo,  $\hat{\theta}$  es una función de las  $X_i$  muestrales, así que es una variable aleatoria.

Con algunas muestras,  $\hat{\theta}$  dará un valor más grande que  $\theta$ , mientras que con otras muestras  $\hat{\theta}$  subestimaré  $\theta$ .



Si se escribe

$$\hat{\theta} = \theta + \text{error de estimación}$$

entonces un estimador preciso sería uno que produzca errores de estimación pequeños, así que los valores estimados se aproximarán al valor verdadero.

Una forma sensible de cuantificar la idea de  $\hat{\theta}$  cercano a  $\theta$  es considerar el error al cuadrado  $(\hat{\theta} - \theta)^2$ .

Con algunas muestras,  $\hat{\theta}$  se acercará bastante a  $\theta$  y el error al cuadrado se aproximará a 0.

Otras muestras pueden dar valores de  $\hat{\theta}$  alejados de  $\theta$ , correspondientes a errores al cuadrado muy grandes.

Una medida general de precisión es el error cuadrático medio:

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

Si un primer estimador tiene una media del error al cuadrado más pequeña que un segundo, es natural decir que el primer estimador es el mejor.

Sin embargo, el error cuadrático medio en general dependerá del valor de  $\theta$ .

Lo que a menudo sucede es que un estimador tendrá una media del error al cuadrado más pequeña con algunos valores de  $\theta$  y una media del error al cuadrado más grande con otros valores.

En general no es posible determinar un estimador con el error cuadrático medio mínimo.

Una forma de librarse de este dilema es limitar la atención sólo en estimadores que tengan una propiedad deseable específica y luego determinar el mejor estimador en este grupo limitado.

Una propiedad popular de esta clase en la comunidad estadística es el **insesgamiento**.

## Estimadores insesgados

Supongamos que se tienen dos instrumentos de medición: uno ha sido calibrado con precisión, pero el otro sistemáticamente da lecturas más pequeñas que el valor verdadero que se está midiendo.

Cuando cada uno de los instrumentos se utiliza repetidamente en el mismo objeto, debido al error de medición, las mediciones observadas no serán idénticas.

Sin embargo, las mediciones producidas por el primer instrumento se distribuirán en torno al valor verdadero de tal modo que en promedio este instrumento mide lo que se propone medir, por lo que este instrumento se conoce como instrumento insesgado.

El segundo instrumento proporciona observaciones que tienen un componente de error o sesgo sistemático.

## Definición

Se dice que un estimador puntual  $\hat{\theta}$  es un **estimador insesgado** de  $\theta$  si  $E(\hat{\theta}) = \theta$  con todo valor posible de  $\theta$ .

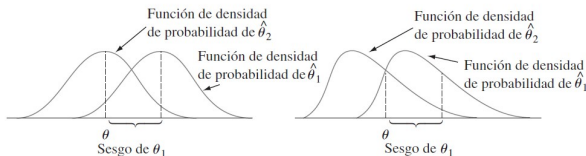
Si  $\hat{\theta}$  no es insesgado, la diferencia  $E(\hat{\theta}) - \theta$  se conoce como el **sesgo** de  $\hat{\theta}$ .

Es decir,  $\hat{\theta}$  es insesgado si su distribución de probabilidad (es decir, muestreo) siempre está "centrada" en el valor verdadero del parámetro.

Supongamos que  $\hat{\theta}$  es un estimador insesgado; entonces  
si  $\theta = 100$ , la distribución muestral  $\hat{\theta}$  está centrada en 100;  
si  $\theta = 27.5$ , en ese caso la distribución muestral  $\hat{\theta}$  está centrada en 27.5,  
y así sucesivamente.

La figura siguiente ilustra la distribución de varios estimadores sesgados e insesgados.

Observemos que "centrada" en este caso significa que el valor esperado, no la mediana de la distribución de  $\hat{\theta}$  es igual a  $\theta$ .



Parece como si fuera necesario conocer el valor de  $\theta$  (en cuyo caso la estimación es innecesaria) para ver si  $\hat{\theta}$  es insesgado.

Éste casi nunca es el caso, puesto que el insesgamiento es una propiedad general del estimador muestral, donde se centra, y generalmente no depende de cualquier valor particular del parámetro.

En el ejemplo de defensa del fabricante automotriz, se utilizó la proporción muestral  $X/n$  como estimador de  $p$ ,  $\hat{p} = X/n$ , donde  $X$ , el número de éxitos muestrales, tenía una distribución binomial con parámetros  $n$  y  $p$ .

Por lo tanto,

$$E(\hat{p}) = E(X/n) = \frac{1}{n}E(X) = \frac{1}{n}n \cdot p = p$$

### Proposición

Cuando  $X$  es una variable aleatoria binomial con parámetros  $n$  y  $p$ , la proporción muestral  $\hat{p} = \frac{X}{n}$  es un estimador insesgado de  $p$ .

No importa cuál sea el valor verdadero de  $p$ , la distribución del estimador  $\hat{p}$  estará centrada en el valor verdadero.

## Principio de estimación insesgada

A la hora de seleccionar entre varios estimadores diferentes de  $\theta$ , se elige uno insesgado.

## Proposición

Sean  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una distribución con media  $\mu$  y varianza  $\sigma^2$ . Entonces el estimador

$$\hat{\sigma}^2 = \hat{S}^2 = \frac{\sum (X_i - \bar{X})^2}{n - 1}$$

es un estimador insesgado de  $\sigma^2$ .

Recordatorio (Tema 6):  $E(\hat{S}^2) = \sigma^2$

Se pueden proponer varios estimadores diferentes de la media  $\mu$  de una distribución normal.

Si hubiera un estimador insesgado único para  $\mu$ , el problema de estimación se resolvería utilizando dicho estimador.

Desafortunadamente, éste no es el caso.

### Proposición

Si  $X_1, X_2, \dots, X_n$  es una muestra aleatoria tomada de una distribución con media  $\mu$ , entonces  $\bar{X}$  es un estimador insesgado de  $\mu$ .

Si además la distribución es continua y simétrica, entonces  $\tilde{X}$  (mediana) y cualquier media recortada también son estimadores insesgados de  $\mu$ .

El hecho de que  $\bar{X}$  sea insesgado es simplemente un replanteamiento de una de las reglas de valor esperado:  $E(\bar{X}) = \mu$  con cada valor posible de  $\mu$  (para distribuciones discretas y continuas).

El insesgamiento de los demás estimadores es más difícil de verificar.



Tanto en la situación que implica estimar una media de población normal como en otros casos, el principio de insesgamiento (prefiere un estimador insesgado a uno sesgado) no puede ser invocado para seleccionar un estimador.

Lo que ahora se requiere es un criterio para elegir entre estimadores insesgados.

## Estimadores con varianza mínima

Supongamos que  $\hat{\theta}_1$  y  $\hat{\theta}_2$  son dos estimadores de  $\theta$  insesgados.

Entonces, aunque la distribución de cada estimador esté centrada en el valor verdadero de  $\theta$ , las dispersiones de las distribuciones en torno al valor verdadero pueden ser diferentes.

### Principio de estimación insesgada con varianza mínima

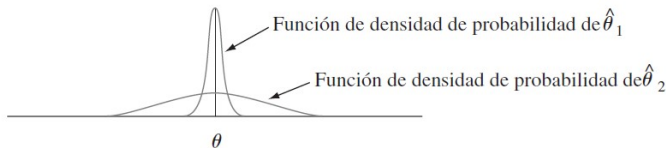
Entre todos los estimadores de  $\theta$  insesgados, se selecciona el de varianza mínima.

El  $\hat{\theta}$  resultante se llama **estimador insesgado con varianza mínima (EIVM)** de  $\theta$ .

La figura siguiente ilustra las funciones de densidad de probabilidad de los dos estimadores insesgados, donde  $\hat{\theta}_1$  tiene una varianza más pequeña que  $\hat{\theta}_2$ .

Entonces es más probable que  $\hat{\theta}_1$  produzca una estimación próxima al valor verdadero  $\theta$  que  $\hat{\theta}_2$ .

El estimador insesgado con varianza mínima es, en cierto sentido, el que tiene más probabilidades entre todos los estimadores insesgados de producir una estimación cercana al verdadero  $\theta$ .



Uno de los triunfos de la estadística matemática ha sido el desarrollo de una metodología para identificar el estimador insesgado con varianza mínima en una amplia variedad de situaciones.

El resultado más importante de este tipo para nuestros propósitos tiene que ver con la estimación de la media  $\mu$  de una distribución normal.

### Teorema

Sean  $X_1, X_2, \dots, X_n$  una muestra aleatoria tomada de una distribución normal con parámetros  $\mu$  y  $\sigma$ .

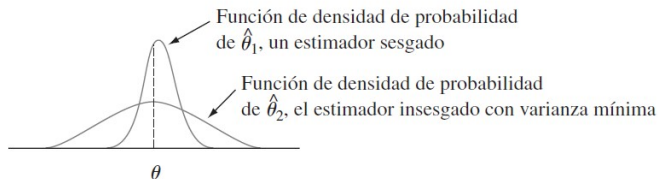
Entonces el estimador  $\hat{\mu} = \bar{X}$  es el estimador insesgado con varianza mínima para  $\mu$ .

Siempre que exista la seguridad de que la población que se está muestreando es normal, el resultado dice que  $\bar{X}$  debería usarse para estimar  $\mu$ .

En algunas situaciones, es posible obtener un estimador con sesgo pequeño que se preferiría al mejor estimador insesgado.

Esto se ilustra en la siguiente figura.

Sin embargo, los estimadores insesgados con varianza mínima a menudo son más fáciles de obtener que el tipo de estimador sesgado cuya distribución se ilustra.



## Reporte de una estimación puntual: el error estándar

Además de reportar el valor de una estimación puntual, se debe dar alguna indicación de su precisión.

La medición usual de precisión es el error estándar del estimador usado.

### Definición

El **error estándar** de un estimador  $\hat{\theta}$  es su desviación estándar

$$\sigma_{\hat{\theta}} = \sqrt{\text{Var}(\hat{\theta})}.$$

Si el error estándar implica parámetros desconocidos cuyos valores pueden ser estimados, la sustitución de estas estimaciones en  $\sigma_{\hat{\theta}}$  da el error estándar estimado (desviación estándar estimada) del estimador.

El error estándar estimado puede ser denotado o por  $\hat{\sigma}_{\hat{\theta}}$  (el  $\hat{\cdot}$  sobre  $\sigma$  recalca que  $\sigma_{\hat{\theta}}$  está siendo estimada) o por  $\hat{s}_{\hat{\theta}}$ .

## Ejemplo

A partir de las 20 observaciones correspondientes al voltaje de ruptura, que es una variable aleatoria normal con  $\mu$  desconocida.

Afirmamos ahora que:  $\hat{\mu} = \bar{X}$  es la mejor estimación de  $\mu$ .

Si se sabe que el valor de  $\sigma$  es 1.5, el error estándar de  $\bar{X}$  es  $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 1.5/\sqrt{20} = 0.335$ .

Si, como casi siempre es el caso, el valor de  $\sigma$  es desconocido, la estimación  $\hat{\sigma} = \hat{s} = 1.462$  se sustituye en  $\sigma_{\bar{X}}$  para obtener el error estándar estimado  $\hat{\sigma}_{\bar{X}} = \hat{s}_{\bar{X}} = \hat{s}/\sqrt{n} = 1.462/\sqrt{20} = 0.327$ .

## Ejemplo

Del ejemplo de defensa del fabricante automotriz, el error estándar del estimador  $\hat{p} = X/n$  es

$$\sigma_{\hat{p}} = \sqrt{\text{Var}(X/n)} = \sqrt{\frac{1}{n^2} \text{Var}(X)} = \sqrt{\frac{1}{n^2} n \cdot p \cdot (1-p)} = \sqrt{\frac{p \cdot (1-p)}{n}}$$

Como  $p$  es desconocida (¿de otro modo por qué estimar?), se sustituye  $\hat{p} = x/n$  en  $\sigma_{\hat{p}}$  para obtener el error estándar estimado

$$\hat{\sigma}_{\hat{p}} = \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}} = \sqrt{\frac{0.6 \cdot 0.4}{25}} = 0.098$$

Alternativamente, como el valor más grande de  $p \cdot (1-p)$  se obtiene cuando  $p = 0.5$ , un límite superior en el error estándar es

$$\sqrt{\frac{0.5 \cdot 0.5}{n}} = \sqrt{\frac{0.25}{25}} = 0.10$$



Cuando la distribución del estimador puntual  $\hat{\theta}$  es aproximadamente normal, lo que a menudo será el caso cuando  $n$  es grande, en tal caso se puede estar confiado de manera razonable en que el valor verdadero de  $\theta$  queda dentro de aproximadamente dos errores estándar (desviaciones estándar) de  $\hat{\theta}$ . Es decir:

$$(\hat{\theta} - 2 \cdot \sigma_{\hat{\theta}}, \hat{\theta} + 2 \cdot \sigma_{\hat{\theta}})$$

De este modo si una muestra de  $n = 36$  vidas útiles de componentes da  $\hat{\mu} = \bar{x} = 28.50$  y, con varianza desconocida,  $\hat{\sigma} = \hat{s} = 3.60$ , por consiguiente,  $\hat{s}/\sqrt{n} = 0.60$  dentro de dos errores estándar estimados de  $\hat{\mu}$  se transforma en el intervalo

$$(\hat{\mu} - 2 \cdot \hat{s}_{\hat{\mu}}, \hat{\mu} + 2 \cdot \hat{s}_{\hat{\mu}})$$

$$28.50 \pm 2 \cdot 0.60 = (27.30, 29.70)$$

Si  $\hat{\theta}$  no es necesariamente normal de forma aproximada pero es insesgado, entonces se puede demostrar que la estimación se desviará de  $\theta$  hasta por cuatro errores estándar cuando mucho el 6% de las veces.

Se esperaría entonces que el valor verdadero quede dentro de cuatro errores estándar de  $\hat{\theta}$  (y ésta es una proposición muy conservadora, puesto que se aplica a cualquier  $\hat{\theta}$  insesgado).

Resumiendo, el error estándar indica de forma aproximada a qué distancia de  $\hat{\theta}$  se puede esperar que quede el valor verdadero de  $\theta$ .

La forma del estimador de  $\hat{\theta}$  puede ser suficientemente complicada de modo que la teoría estadística estándar no pueda ser aplicada para obtener una expresión para  $\sigma_{\hat{\theta}}$ .

Esto es cierto, por ejemplo, en el caso  $\theta = \sigma$ ,  $\hat{\theta} = \hat{S}$ , la desviación estándar del estadístico  $\hat{S}$ ,  $\sigma_{\hat{S}}$ , en general no puede ser determinada.

# Esquema

## 1 ESTIMACIÓN PUNTUAL

- Introducción
- Algunos conceptos generales de estimación puntual
  - Estimadores insesgados
  - Estimadores con varianza mínima
  - Reporte de una estimación puntual: el error estándar
- Métodos de estimación puntual
  - El método de momentos
  - Estimación de máxima verosimilitud
- Ejercicios
- Problemas

La definición de insesgamiento no indica en general cómo se pueden obtener los estimadores insesgados.

A continuación se discuten dos métodos "constructivos" para obtener estimadores puntuales:

- el método de momentos
- el método de máxima verosimilitud

Por constructivo se quiere dar a entender que la definición general de cada tipo de estimador sugiere explícitamente cómo obtener el estimador en cualquier problema específico.

Aun cuando se prefieren los estimadores de máxima verosimilitud a los de momento debido a ciertas propiedades de eficiencia, a menudo requieren significativamente más cálculo que los estimadores de momento.

En ocasiones estos métodos dan estimadores insesgados.

## El método de momentos

La idea básica de este método es poder igualar ciertas características muestrales, tales como la media, a los valores esperados de la población correspondiente.

Luego, resolviendo estas ecuaciones con valores de parámetros conocidos, se obtienen los estimadores.

### Definición

Si  $X_1, X_2, \dots, X_n$  constituyen una muestra aleatoria proveniente de una función masa de probabilidad o de una función de densidad de probabilidad  $f(x)$ .

Para  $k = 1, 2, 3, \dots$ :

- El  $k$ -ésimo momento de la población o el  $k$ -ésimo momento de la distribución  $f(x)$ , es  $E(X^k)$
- El  $k$ -ésimo momento muestral es  $(1/n) \sum_{i=1}^n X_i^k$

Por consiguiente, para  $k = 1$ :

- el primer momento de la población es  $E(X) = \mu$
- el primer momento muestral es  $\sum X_i/n$

para  $k = 2$ :

- el segundo momento de la población es  $E(X^2)$
- el segundo momento muestral es  $\sum X_i^2/n$

y para  $k = 3$ :

- el tercer momento de la población es  $E(X^3)$
- el tercer momento muestral es  $\sum X_i^3/n$

Los momentos de la población serán funciones de cualquier parámetro desconocido  $\theta_1, \theta_2, \dots$ .

### Definición

Si  $X_1, X_2, \dots, X_n$  son una muestra aleatoria de una distribución con función masa de probabilidad o función de densidad de probabilidad  $f(x; \theta_1, \dots, \theta_m)$ , donde  $\theta_1, \dots, \theta_m$  son parámetros cuyos valores son desconocidos.

Entonces los **estimadores de momento**  $\hat{\theta}_1, \dots, \hat{\theta}_m$  se obtienen igualando los primeros  $m$  momentos muestrales con los primeros  $m$  momentos de la población correspondientes y resolviendo (el sistema) para  $\theta_1, \dots, \theta_m$ .

Si, por ejemplo,  $m = 2$ ,  $E(X)$  y  $E(X^2)$  serán funciones de  $\theta_1$  y  $\theta_2$ .

El método de los momentos, para  $m = 2$ , nos pide que realicemos las siguientes 2 igualdades:

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

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

obteniendo un sistema de dos ecuaciones en función de  $\theta_1$  y  $\theta_2$ .

La solución define entonces los estimadores.

Para estimar una media  $\mu$  poblacional, el método da  $\mu = \bar{X}$ , por lo tanto el estimador es la media muestral.



## Ejemplo

Si  $X_1, X_2, \dots, X_n$  representan una muestra aleatoria de tiempos de servicio de  $n$  clientes en una instalación, donde la distribución subyacente se supone exponencial con el parámetro  $\lambda$ .

Como sólo hay un parámetro que tiene que ser estimado, el método de los momentos nos lleva a igualar el primer momento de la población con el primer momento de la muestras, es decir:

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

El estimador se obtiene igualando  $E(X)$  a  $\bar{X}$ .

Como  $E(X) = 1/\lambda$  para una distribución exponencial, el método de los momentos nos da  $1/\lambda = \bar{X}$ , o equivalentemente  $\lambda = 1/\bar{X}$ .

En conclusión, el estimador de momento de  $\lambda$  es entonces  $\hat{\lambda} = \frac{1}{\bar{X}}$

## Estimación de máxima verosimilitud

El método de máxima probabilidad lo introdujo por primera vez R. A. Fisher, genetista y estadístico en la década de 1920.

La mayoría de los estadísticos recomiendan este método, por lo menos cuando el tamaño de muestra es grande, puesto que los estimadores resultantes tienen ciertas propiedades de eficiencia deseables.

## Definición

Que  $X_1, X_2, \dots, X_n$  tengan una función masa de probabilidad o una función de densidad de probabilidad  $f(x_1, x_2, \dots, x_n; \theta_1, \dots, \theta_m)$  donde los parámetros  $\theta_1, \dots, \theta_m$  tienen valores desconocidos.

Cuando  $x_1, \dots, x_n$  son los valores muestrales observados y la función (masa) de densidad de probabilidad se considera como una función de  $\theta_1, \dots, \theta_m$  se llama **función de verosimilitud**.

Las estimaciones de máxima verosimilitud (emv)  $\hat{\theta}_1, \dots, \hat{\theta}_m$  son aquellos valores de las  $\theta_i$  que incrementan al máximo la función de probabilidad, de modo que

$$f(x_1, x_2, \dots, x_n; \hat{\theta}_1, \dots, \hat{\theta}_m) \geq f(x_1, x_2, \dots, x_n; \theta_1, \dots, \theta_m)$$

con todos los  $\theta_1, \dots, \theta_m$ .

Cuando se sustituyen las  $X_i$  en lugar de las  $x_i$ , se obtienen los **estimadores de máxima verosimilitud**.

La función de verosimilitud dice qué tan probable es que la muestra observada sea una función de los posibles valores de los parámetros.

Al incrementarse al máximo la probabilidad se obtienen los valores de los parámetros con los que la muestra observada es más probable que haya sido generada, es decir, los valores de los parámetros que "más concuerdan" con los datos observados.

## Ejemplo

Se obtuvo una muestra de diez cascos de ciclista nuevos fabricados por una compañía.

Al probarlos, se encontró que el primero, el tercero y el décimo estaban agrietados, en tanto que los demás no.

Sea  $p = P(\text{casco agrietado})$  y definiendo  $X_1, \dots, X_{10}$  como  $X_i = 1$  si el  $i$ -ésimo casco está agrietado y cero en caso contrario.

En ese caso las  $x_i$  son 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, así que la función masa de probabilidad conjunta de la muestra es

$$f(x_1, x_2, \dots, x_{10}; p) = p \cdot (1 - p) \cdot p \cdot (1 - p) \cdots (1 - p) \cdot p = p^3 \cdot (1 - p)^7$$

Ahora nos preguntamos, "¿Con qué valor de  $p$  es más probable que la muestra observada haya ocurrido?"

## Ejemplo (Continuación)

Es decir, se desea encontrar el valor de  $p$  que incrementa al máximo la función masa de probabilidad o, en forma equivalente, que incrementa al máximo el logaritmo natural de la función masa de probabilidad<sup>1</sup>.

En consecuencia, aplicamos  $\ln$  a ambos lados de la expresión anterior:

$$\ln(f(x_1, x_2, \dots, x_{10}; p)) = \ln(p^3 \cdot (1 - p)^7) = \ln(p^3) + \ln((1 - p)^7)$$

$$\ln(f(x_1, x_2, \dots, x_{10}; p)) = 3 \cdot \ln(p) + 7 \cdot \ln(1 - p)$$

la cual es una función derivable respecto de  $p$ .

Derivamos respecto de  $p$ :

$$\frac{d}{dp}[\ln(f(x_1, x_2, \dots, x_{10}; p))] = \frac{d}{dp}[3 \cdot \ln(p) + 7 \cdot \ln(1 - p)]$$

---

<sup>1</sup>Como  $\ln[g(p)]$  es una función monotónica de  $g(p)$ , determinar  $p$  para incrementar al máximo  $\ln[g(p)]$  equivale a incrementar al máximo  $g(p)$ . Es decir, el valor de  $p$  que maximice  $g(p)$  también maximiza  $\ln[g(p)]$ . En estadística, si se toma el logaritmo con frecuencia, un producto cambia a una suma, con la cual es más fácil trabajar.

### Ejemplo (Continuación)

$$\frac{d}{dp}[\ln(f(x_1, x_2, \dots, x_{10}; p))] = \frac{d}{dp}[3 \cdot \ln(p)] + \frac{d}{dp}[7 \cdot \ln(1 - p)]$$

igualando la derivada a cero se obtiene el valor de  $p$  maximizante de  $\ln(f(x_1, x_2, \dots, x_{10}; p))$  y a su vez de  $f(x_1, x_2, \dots, x_{10}; p)$ :

$$\frac{d}{dp}[\ln(f(x_1, x_2, \dots, x_{10}; p))] = \frac{3}{p} - \frac{7}{1-p} = 0 \implies p = \frac{3}{10} = \frac{x}{n}$$

que se corresponde con  $p = \frac{3}{10} = \frac{x}{n}$ , donde  $x$  es el número de éxitos observados (cascos agrietados).

La estimación de  $p$  ahora es  $\hat{p} = \frac{3}{10}$ .

Y se llama estimación de máxima verosimilitud porque para  $x_1, \dots, x_{10}$  establecido, es el valor del parámetro que maximiza la probabilidad (función masa de probabilidad conjunta) de la muestra observada.

### Ejemplo (Continuación)

Observemos que si sólo se hubiera dicho que entre los diez cascos había tres agrietados, la expresión inicial sería reemplazada por la función masa de probabilidad binomial  $\binom{10}{3}p^3(1-p)^7$ , la cual también se incrementa al máximo con  $\hat{p} = \frac{3}{10}$ .



## Comportamiento con muestra grande del estimador de máxima verosimilitud

Aunque el principio de la estimación de máxima verosimilitud tiene un considerable atractivo intuitivo, la siguiente proposición proporciona razones adicionales fundamentales para el uso de estimadores de máxima verosimilitud.

### Proposición

En condiciones muy generales en relación con la distribución conjunta de la muestra, cuando el tamaño de la muestra  $n$  es grande, el estimador de máxima verosimilitud de cualquier parámetro  $\theta$  es aproximadamente insesgado [ $E(\hat{\theta}) \approx \theta$ ] y su varianza es casi tan pequeña como la que puede ser lograda por cualquier estimador.

Expresado de otra manera, el estimador de máxima verosimilitud  $\hat{\theta}$  es aproximadamente el estimador insesgado con varianza mínima de  $\theta$ .

Debido a este resultado y al hecho de que las técnicas basadas en el cálculo casi siempre pueden ser utilizadas para derivar los estimadores de máxima verosimilitud (aunque a veces se requieren métodos numéricos, tales como el método de Newton), la estimación de máxima verosimilitud es la técnica de estimación más ampliamente utilizada entre los estadísticos.

Muchos de los estimadores que utilizamos son estimadores de máxima verosimilitud.

La obtención de un estimador de máxima verosimilitud, sin embargo, requiere que se especifique la distribución subyacente.

# Esquema

## 1 ESTIMACIÓN PUNTUAL

- Introducción
- Algunos conceptos generales de estimación puntual
  - Estimadores insesgados
  - Estimadores con varianza mínima
  - Reporte de una estimación puntual: el error estándar
- Métodos de estimación puntual
  - El método de momentos
  - Estimación de máxima verosimilitud
- Ejercicios
- Problemas

## EJERCICIO 1

Existe interés por desarrollar aleaciones de Mg de bajo coste para varios procesos de fundición.

Es por consiguiente importante contar con formas prácticas de determinar varias propiedades mecánicas de esas aleaciones.

Un artículo publicado en Light Metal Age (octubre de 1998: 50-53) propuso un método ultrasónico para este propósito.

Consideremos la siguiente muestra de observaciones de módulo elástico (GPa) de especímenes de aleación AZ91D tomados de un proceso de fundición a troquel: 44.2, 43.9, 44.7, 44.2, 44.0, 43.8, 44.6, 43.1

Supongamos que estas observaciones son el resultado de una muestra aleatoria  $X_1, \dots, X_8$  tomada de la distribución de población de módulos elásticos en semejantes circunstancias.

Se desea estimar la varianza de la población  $\sigma^2$ .

## EJERCICIO 1

### Solución

Utilizaremos el estimador  $\hat{S}^2$  (varianza muestral corregida) por ser insesgado, es decir porque cumple que  $E(\hat{S}^2) = \sigma^2$ .

Varianza muestral corregida:  $\hat{\sigma}^2 = \hat{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  y su estimación

$$\hat{s}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{7} \sum_{i=1}^8 (x_i - 44.0625)^2 = 0.251$$

Veamos qué resultado obtendríamos al utilizar la varianza muestral:

$\hat{\sigma}^2 = S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  y su estimación

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{8} \sum_{i=1}^8 (x_i - 44.0625)^2 = 0.220$$

## EJERCICIO 2

Si  $X_1, X_2, \dots, X_n$  representan una muestra aleatoria tomada de una distribución de Rayleigh con función de densidad de probabilidad

$$f(x; \theta) = \frac{x}{\theta} e^{-\frac{x^2}{2\theta}} \quad x > 0$$

Se puede demostrar que  $E(X^2) = 2 \cdot \theta$ . Usa este hecho para construir un estimador insesgado de  $\theta$  basado en  $\sum X_i^2$  (y usa reglas de valor esperado para demostrar que es insesgado).

### Solución

Sea  $X$  una variable aleatoria con media  $\mu$  y varianza  $\sigma^2$ :

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

Para Rayleigh, dado que  $E(X^2) = 2 \cdot \theta$ , tenemos que:

$$\sigma^2 = 2 \cdot \theta - \mu^2$$

o equivalentemente:  $2 \cdot \theta = \mu^2 + \sigma^2$

## EJERCICIO 2

### Solución

El enunciado propone usar  $\sum X_i^2$  para construir nuestro estimador insesgado.

Busquemos primero la esperanza:

$$E(\sum X_i^2) = \sum E(x_i^2) = \sum(\mu^2 + \sigma^2) = n \cdot (\mu^2 + \sigma^2)$$

Y para Rayleigh nos daría:  $E(\sum X_i^2) = n \cdot 2 \cdot \theta$

Que nos da una pista para encontrar el estimador de  $\theta$ .

Aislando  $\theta$  de la expresión anterior, el estimador insesgado será:

$$\hat{\theta} = \frac{1}{2 \cdot n} \sum_{i=1}^n X_i^2$$

Lo comprobamos calculado el valor esperado del estimador:

$$E(\hat{\theta}) = E\left(\frac{1}{2 \cdot n} \sum_{i=1}^n X_i^2\right) = \frac{1}{2 \cdot n} E\left(\sum_{i=1}^n X_i^2\right) = \frac{1}{2 \cdot n} n \cdot 2 \cdot \theta = \theta$$

### EJERCICIO 3

Sean  $X_1, \dots, X_n$  una muestra aleatoria de una distribución binomial negativa con parámetros  $r$  y  $p$ , utiliza el método de los momentos para estimarlos.

### Solución

El método de los momentos nos obliga a igualar los momentos de la población y la muestra para estimar los parámetros  $r$  y  $p$ .

En este caso, como tenemos 2 parámetros igualamos los dos primeros momentos:

- el primer momento de la población es  $E(X)$
- el primer momento muestral es  $\sum X_i/n$  o  $\bar{X}$

para  $k = 2$ :

- el segundo momento de la población es  $E(X^2)$
- el segundo momento muestral es  $\sum X_i^2/n$  o  $\overline{X^2}$



## EJERCICIO 3

### Solución (Continuación)

El método de los momentos nos lleva a las siguientes relaciones:

$$E(X) = \bar{X}$$

$$E(X^2) = \overline{X^2}$$

Para cualquier variable aleatoria  $\sigma^2 = E(X^2) - (E(X))^2$ , en consecuencia  $E(X^2) = (E(X))^2 + \sigma^2$

En el modelo binomial negativo sabemos que:  $\mu = \frac{r}{p}$  y  $\sigma^2 = \frac{r(1-p)}{p^2}$

El sistema de ecuaciones se transforma en:

$$\frac{r}{p} = \bar{X}$$

$$\left(\frac{r}{p}\right)^2 + \frac{r(1-p)}{p^2} = \overline{X^2}$$

## EJERCICIO 3

### Solución (Continuación)

Resolviendo el sistema encontramos los estimadores de los parámetros de la distribución binomial negativa por el método de los momentos:

$$\hat{r} = \frac{(\bar{X})^2}{\bar{X} + \overline{X^2} - (\bar{X})^2}$$

$$\hat{p} = \frac{\bar{X}}{\bar{X} + \overline{X^2} - (\bar{X})^2}$$

## EJERCICIO 4

Suponga que  $X_1, X_2, \dots, X_n$  es una muestra aleatoria de una distribución exponencial con parámetro  $\lambda$ . Se pide el estimador maximoverosimil de  $\lambda$ .

### Solución

Debido a la independencia, la función de verosimilitud es un producto de las funciones de densidad de probabilidad individuales:

$$f(x_1, \dots, x_n; \lambda) = (\lambda \cdot e^{-\lambda \cdot x_1}) \cdots (\lambda \cdot e^{-\lambda \cdot x_n}) = \lambda^n \cdot e^{-\lambda \sum x_i}$$

Aplicando la función logaritmo a ambos lados de la expresión anterior:

$$\ln[f(x_1, \dots, x_n; \lambda)] = \ln[\lambda^n e^{-\lambda \sum x_i}] = \ln[\lambda^n] + \ln[e^{-\lambda \sum x_i}] = n \cdot \ln[\lambda] - \lambda \sum x_i$$

Para encontrar la  $\lambda$  que maximiza la expresión anterior, calculamos la derivada respecto de  $\lambda$ :

$$\frac{d}{d\lambda}(\ln[f(x_1, \dots, x_n; \lambda)]) = \frac{d}{d\lambda}(n \cdot \ln[\lambda] - \lambda \sum x_i) = \frac{d}{d\lambda}(n \cdot \ln[\lambda]) - \frac{d}{d\lambda}(\lambda \sum x_i)$$

## EJERCICIO 4

## Solución (Continuación)

$$\frac{d}{d\lambda}(\ln[f(x_1, \dots, x_n; \lambda)]) = \frac{n}{\lambda} - \sum x_i$$

Igualamos la derivada a cero y aislamos  $\lambda$ :

$$\frac{n}{\lambda} - \sum x_i = 0 \implies \lambda = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}$$

Por consiguiente el estimador de máxima verosimilitud es  $\hat{\lambda} = \frac{1}{\bar{x}}$

## EJERCICIO 5

Sean  $X_1, X_2, \dots, X_n$  una muestra aleatoria de una distribución normal. Encuentra los estimadores maximoverosimiles de  $\mu$  y  $\sigma^2$ .

### Solución

La función de probabilidad es

$$f(x_1, \dots, x_n; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x_1 - \mu}{2\sigma^2}} \cdots \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x_n - \mu}{2\sigma^2}}$$
$$f(x_1, \dots, x_n; \mu, \sigma^2) = \left( \frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{\sum (x_i - \mu)^2}{2\sigma^2}} = \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} e^{-\frac{\sum (x_i - \mu)^2}{2\sigma^2}}$$

Aplicando logaritmos a ambos lados de la expresión anterior

$$\ln[f(x_1, \dots, x_n; \mu, \sigma^2)] = \ln \left[ \left( \frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} e^{-\frac{\sum (x_i - \mu)^2}{2\sigma^2}} \right]$$

## EJERCICIO 5

## Solución (Continuación)

$$\ln[f(x_1, \dots, x_n; \mu, \sigma^2)] = \frac{n}{2} \ln \left( \frac{1}{2\pi\sigma^2} \right) - \frac{\sum (x_i - \mu)^2}{2\sigma^2}$$

$$\ln[f(x_1, \dots, x_n; \mu, \sigma^2)] = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{\sum (x_i - \mu)^2}{2\sigma^2}$$

Para determinar los valores maximizantes de  $\mu$  y  $\sigma^2$ , se deben calcular las derivadas parciales de  $\ln(f)$  con respecto a  $\mu$  y  $\sigma^2$

$$\frac{\partial}{\partial \mu} \ln[f(x_1, \dots, x_n; \mu, \sigma^2)] = \frac{\partial}{\partial \mu} \left[ -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{\sum (x_i - \mu)^2}{2\sigma^2} \right]$$

$$\frac{\partial}{\partial \mu} \ln[f(x_1, \dots, x_n; \mu, \sigma^2)] = -\frac{-2 \sum (x_i - \mu)}{2\sigma^2} = \frac{\sum (x_i - \mu)}{\sigma^2} = \frac{\sum (x_i) - n\mu}{\sigma^2}$$

## EJERCICIO 5

## Solución (Continuación)

$$\frac{\partial}{\partial \sigma^2} \ln[f(x_1, \dots, x_n; \mu, \sigma^2)] = \frac{\partial}{\partial \sigma^2} \left[ -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{\sum (x_i - \mu)^2}{2\sigma^2} \right]$$

$$\frac{\partial}{\partial \sigma^2} \ln[f(x_1, \dots, x_n; \mu, \sigma^2)] = -\frac{n}{2} \frac{2\pi}{2\pi\sigma^2} + \frac{\sum (x_i - \mu)^2}{2\sigma^4} = -\frac{n}{2\sigma^2} + \frac{\sum (x_i - \mu)^2}{2\sigma^4}$$

igualando a cero ambas derivadas parciales encontramos el sistema de ecuaciones que nos permite encontrar los estimadores maximoverosimiles buscados:

$$\frac{\partial}{\partial \mu} \ln[f(x_1, \dots, x_n; \mu, \sigma^2)] = \frac{\sum x_i - n\mu}{\sigma^2} = 0$$

$$\frac{\partial}{\partial \sigma^2} \ln[f(x_1, \dots, x_n; \mu, \sigma^2)] = -\frac{n}{2\sigma^2} + \frac{\sum (x_i - \mu)^2}{2\sigma^4} = 0$$

## EJERCICIO 5

### Solución (Continuación)

Al resolver obtenemos que:

$$\hat{\mu} = \bar{X}$$

$$\hat{\sigma}^2 = \frac{\sum (X_i - \bar{X})^2}{n}$$

El estimador de máxima verosimilitud de  $\sigma^2$  no es el estimador insesgado, por consiguiente dos principios diferentes de estimación (insesgamiento y máxima verosimilitud) dan dos estimadores diferentes.



# Esquema

## 1 ESTIMACIÓN PUNTUAL

- Introducción
- Algunos conceptos generales de estimación puntual
  - Estimadores insesgados
  - Estimadores con varianza mínima
  - Reporte de una estimación puntual: el error estándar
- Métodos de estimación puntual
  - El método de momentos
  - Estimación de máxima verosimilitud
- Ejercicios
- Problemas

## Problema 1

Se examina cada uno de 150 artículos recién fabricados y se anota el número de rayones por artículo (se supone que los artículos están libres de rayones) y se obtienen los siguientes datos:

Número de rayones por artículo	0	1	2	3	4	5	6	7
Frecuencia observada	18	37	42	30	13	7	2	1

Sea  $X$  = el número de rayones en un artículo seleccionado al azar y supongamos que  $X$  tiene una distribución de Poisson con parámetro  $\lambda$ .

- Determina un estimador insesgado de  $\lambda$  y calcula la estimación de los datos. [Sugerencia:  $E(X) = \lambda$  para una distribución Poisson de  $X$ , por lo tanto  $E(\bar{X}) = ?$ ]
- ¿Cuál es la desviación estándar (error estándar) de su estimador? Calcula el error estándar estimado. [Sugerencia:  $\sigma_X^2 = \lambda$  con distribución de Poisson de  $X$ .]

## Problema 2

Con una larga varilla de longitud  $\mu$ , va a trazar una curva cuadrada en la cual la longitud de cada lado es  $\mu$ . Por consiguiente el área de la curva será  $\mu^2$ . Sin embargo, no conoce el valor de  $\mu$  así que decide hacer  $n$  mediciones independientes  $X_1, X_2, \dots, X_n$  de la longitud. Supongamos que cada  $X_i$  tiene una media  $\mu$  (mediciones insesgadas) y varianza  $\sigma^2$ .

- a Demuestra que  $\bar{X}^2$  no es un estimador insesgado de  $\mu^2$ . [Sugerencia: Con cualquier variable aleatoria  $Y$ ,  $E(Y^2) = Var(Y) + [E(Y)]^2$ . Aplica ésta con  $Y = \bar{X}$ .]
- b ¿Con qué valor de  $k$  es el estimador  $\bar{X}^2 - k \cdot \hat{S}^2$  insesgado para  $\mu^2$ ? [Sugerencia: Calcula  $E(\bar{X}^2 - k \cdot \hat{S}^2)$ .]

### Problema 3

Consideremos una muestra aleatoria de  $X_1, \dots, X_n$  de la función de densidad de probabilidad

$$f(x; \theta) = 0.5(1 + \theta \cdot x) \quad -1 \leq x \leq 1$$

donde  $-1 \leq \theta \leq 1$  (esta distribución se presenta en la física de partículas).

Demuestra que  $\hat{\theta} = 3 \cdot \bar{X}$  es un estimador insesgado de  $\theta$ . [Sugerencia: Primero determina  $\mu = E(X) = E(\bar{X})$ .]

## Problema 4

Sea  $X$  la proporción de tiempo destinado que un estudiante seleccionado al azar pasa resolviendo cierta prueba de aptitud. Supongamos que la función de densidad de probabilidad de  $X$  es

$$f(x; \theta) = \begin{cases} (\theta + 1)x^\theta & 0 \leq x \leq 1 \\ 0 & \text{otros} \end{cases}$$

donde  $-1 < \theta$ . Una muestra aleatoria de diez estudiantes produce los datos  $x_1 = 0.92$ ,  $x_2 = 0.79$ ,  $x_3 = 0.90$ ,  $x_4 = 0.65$ ,  $x_5 = 0.86$ ,  $x_6 = 0.47$ ,  $x_7 = 0.73$ ,  $x_8 = 0.97$ ,  $x_9 = 0.94$ ,  $x_{10} = 0.77$ .

- a. Usa el método de momentos para obtener un estimador de  $\theta$  y luego calcula la estimación con estos datos.
- b. Obtén el estimador de máxima verosimilitud de  $\theta$  y luego calcula la estimación con los datos dados.

## Problema 5

Consideremos la muestra aleatoria  $X_1, X_2, \dots, X_n$  de la función de densidad de probabilidad exponencial desplazada

$$f(x; \lambda, \theta) = \begin{cases} \lambda \cdot e^{-\lambda(x-\theta)} & x \geq \theta \\ 0 & \text{otros} \end{cases}$$

- a. Obten los estimadores de máxima verosimilitud de  $\theta$  y  $\lambda$ .
- b. Si  $n = 10$  observaciones de tiempo entre vehículos son realizadas y se obtienen los siguientes resultados 3.11, 0.64, 2.55, 2.20, 5.44, 3.42, 10.39, 8.93, 17.82 y 1.30, calcula las estimaciones de  $\theta$  y  $\lambda$ .