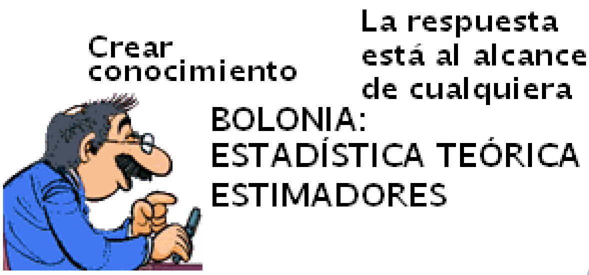
Gestión Aeronáutica: Estadística Teórica
Facultad Ciencias Económicas y Empresariales
Departamento de Economía Aplicada
Profesor: Santiago de la Fuente Fernández

ESTADÍSTICA TEÓRICA: ESTIMADORES



ESTIMADORES

Un *estimador* es un estadístico (una función de la muestra) utilizado para estimar un parámetro desconocido de la población.

Por ejemplo, si se desea conocer el precio medio poblacional de un artículo (parámetro desconocido) se recogen observaciones del precio de dicho artículo en diversos establecimientos (muestra) pudiendo utilizarse la media aritmética de las observaciones para estimar el precio medio poblacional.

Para *cada parámetro* pueden existir varios *estimadores diferentes*. En general, se elige el estimador que posea mejores propiedades que los restantes, como insesgadez, eficiencia, convergencia y robustez (consistencia).

El valor de un estimador proporciona una estimación puntual del valor del parámetro en estudio. En general, se realiza la *estimación* mediante un intervalo, es decir, se obtiene un *intervalo* $\left[\text{estadístico muestral} \pm \text{error estimación}\right]$ dentro del cual se espera se encuentre el valor poblacional dentro de un cierto nivel de confianza. El nivel de confianza es la probabilidad de que a priori el valor poblacional se encuentre contenido en el intervalo.

PROPIEDADES DE LA ESPERANZA Y VARIANZA

a) $E\left[aX_1 \pm bX_2\right] = E\left[aX_1\right] \pm E\left[bX_2\right] = a\,E\left[X_1\right] \pm b\,E\left[X_2\right]$

b) $\text{Var}\left[aX_1 \pm bX_2\right] = \text{Var}\left[aX_1\right] + \text{Var}\left[bX_2\right] = a^2\,\text{Var}\left[X_1\right] + b^2\,\text{Var}\left[X_2\right]$

SESGO

Se denomina *sesgo* de un estimador a la diferencia entre la esperanza (valor esperado) del estimador y el verdadero valor del parámetro a estimar. Es deseable que un estimador sea *insesgado o centrado*, esto es, que el sesgo sea nulo para que la esperanza del estimador sea igual al valor del parámetro que se desea estimar.

Por ejemplo, si se desea estimar la media de una población, la media aritmética de la muestra es un estimador insesgado de la misma, ya que la esperanza (valor esperado) es igual a la media poblacional.

Si una muestra $X=(x_1, x_2, \cdots, x_n)$ procede de una población de media μ:

$E[x_i]=\mu$ para $i=(1, 2, \cdots, n)$

- La media aritmética muestral es un estimador insesgado de la media poblacional:

$$E[\overline{x}]=E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right]=\frac{1}{n}E\left[\sum_{i=1}^{n} x_i\right]=\frac{1}{n}\sum_{i=1}^{n} E[x_i]=\frac{1}{n}\left(E[x_1]+E[x_2]+\cdots+E[x_n]\right)=$$

$$=\frac{1}{n}n\mu=\mu$$

- La varianza de una muestra aleatoria simple es un estimador sesgado de la varianza poblacional, siendo su esperanza:

La varianza muestral $\sigma_x^2=\dfrac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}$. Para calcular su esperanza matemática se realizan previamente algunos cálculos sumando y restando la esperanza de la variable aleatoria poblacional.

$$\sigma_x^2=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}=\frac{\sum_{i=1}^{n}(x_i-\overline{x}+\mu-\mu)^2}{n}=\frac{1}{n}\sum_{i=1}^{n}\left[(x_i-\mu)-(\overline{x}-\mu)\right]^2$$

Desarrollando el cuadrado:

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}\left[(x_i-\mu)^2 + (\bar{x}-\mu)^2 - 2(x_i-\mu)(\bar{x}-\mu)\right] =$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n}(x_i-\mu)^2 + n(\bar{x}-\mu)^2 - 2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\mu)\right) =$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n}(x_i-\mu)^2 + n(\bar{x}-\mu)^2 - 2\,(\bar{x}-\mu)\,(n\bar{x}-n\mu)\right) =$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n}(x_i-\mu)^2 + n\bar{x}^2 + n\mu^2 - \boxed{2n\bar{x}\mu} - 2n\bar{x}^2 + \boxed{2n\bar{x}\mu} + 2n\bar{x}\mu - 2n\mu^2\right) =$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n}(x_i-\mu)^2 - n(\bar{x}-\mu)^2\right)$$

Calculando la esperanza matemática de la varianza muestral σ_x^2 :

$$E\left[\sigma_x^2\right] = \frac{1}{n}E\left(\sum_{i=1}^{n}(x_i-\mu)^2 - n(\bar{x}-\mu)^2\right) = \underbrace{\frac{1}{n}\sum_{i=1}^{n}E\left[(x_i-\mu)^2\right]}_{\text{varianza poblacional}} - \underbrace{E\left[(\bar{x}-\mu)^2\right]}_{\text{varianza media muestral}}$$

En el segundo miembro aparecen dos esperanzas, la primera $E(x_i-\mu)^2$ coincide con la varianza poblacional σ^2 al tratarse de una muestra aleatoria simple, la segunda esperanza $E(\bar{x}-\mu)^2$ es la varianza de la media muestral $\dfrac{\sigma^2}{n}$

Por tanto, $E\left[\sigma_x^2\right] = \sigma^2 - \dfrac{\sigma^2}{n} = \dfrac{n-1}{n}\sigma^2$

- **La cuasivarianza de una muestra aleatoria simple es un estimador insesgado de la varianza poblacional:** $\quad s_x^2 = \dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$

Relación entre varianza y cuasivarianza:

$$n\sigma_x^2 = (n-1)\, s_x^2 \quad \mapsto \quad s_x^2 = \frac{n}{n-1}\,\sigma_x^2$$

La esperanza de la cuasivarianza s_x^2 es la varianza poblacional σ^2:

$$E\left[s_x^2\right] = E\left[\frac{n}{n-1}\,\sigma_x^2\right] = \frac{n}{n-1}\,.\,E\left[\sigma_x^2\right] = \frac{n}{n-1}\,.\,\frac{n-1}{n}\,.\,\sigma^2 = \sigma^2$$

📖 **Un estimador $\hat{\theta}$ es *insesgado* (*centrado*) cuando $E(\hat{\theta}) = \theta$**

📖 **Un estimador $\hat{\theta}$ es *sesgado* sí $E(\hat{\theta}) = \theta + \underbrace{b(\hat{\theta})}_{\text{sesgo}} \Rightarrow b(\hat{\theta}) = E(\hat{\theta}) - \theta$**

📖 **Un estimador $\hat{\theta}$ es *asintóticamente insesgado* si su posible sesgo tiende a cero al aumentar el tamaño muestral que se calcula: $\lim\limits_{n\to\infty} b(\hat{\theta}) = 0$**

Sea el estimador $\hat{\theta} = \dfrac{1}{n+1}\displaystyle\sum_{i=1}^{n} x_i$

$$E(\hat{\theta}) = E\left[\frac{1}{n+1}\sum_{i=1}^{n} x_i\right] = \frac{1}{n+1}\,E\left[\sum_{i=1}^{n} x_i\right] = \frac{1}{n+1}\sum_{i=1}^{n} E(x_i) = \frac{1}{n+1}\,(n\mu) = \frac{n\mu}{n+1}$$

$$b(\hat{\theta}) = E(\hat{\theta}) - \mu = \frac{n\mu}{n+1} - \mu = \frac{n\mu - n\mu - \mu}{n+1} = \underbrace{-\frac{\mu}{n+1}}_{\text{insesgado asintóticamente}} \xrightarrow{\;n \longrightarrow \infty\;} 0$$

ERROR CUADRÁTICO MEDIO DE LOS ESTIMADORES (ECM)

La utilización de la estimación puntual como si fuera el verdadero valor del parámetro conduce a que se pueda cometer un error más o menos grande.

El Error Cuadrático Medio (ECM) de un estimador $\hat{\theta}$ viene definido:

$$\mathrm{ECM}(\hat{\theta}) = \mathrm{E}(\hat{\theta}-\theta)^2 = \mathrm{V}(\hat{\theta}) + \Big[\underbrace{\mathrm{E}(\hat{\theta})-\theta}_{\text{sesgo}}\Big]^2 \qquad \text{sesgo } \ b(\hat{\theta}) = \mathrm{E}(\hat{\theta})-\theta$$

Cuando el estimador es centrado, el sesgo $b(\hat{\theta}) = 0 \quad \to \quad \mathrm{ECM}(\hat{\theta}) = \mathrm{V}(\hat{\theta})$

Un error cuadrático medio pequeño indicará que en media el estimador $\hat{\theta}$ no se encuentra lejos del parámetro θ.

CONSISTENCIA

Si no es posible emplear estimadores de mínima varianza, el requisito mínimo deseable para un estimador es que a medida que el tamaño de la muestra crece, el valor del estimador tienda a ser el valor del parámetro poblacional, propiedad que se denomina *consistencia*.

Un estimador $\hat{\theta}$ *consistente* es un estimador asintóticamente insesgado cuya varianza tiende a cero al aumentar el tamaño muestral.

El estimador $\hat{\theta}$ es *consistente* cuando $\lim\limits_{n\to\infty} \mathrm{E}(\hat{\theta}) = \theta \quad$ y $\quad \lim\limits_{n\to\infty} \mathrm{V}(\hat{\theta}) = 0$

EFICIENCIA

Un estimador es más *eficiente* o más *preciso* que otro estimador, si la varianza del primero es menor que la del segundo.

Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados, se dice que $\hat{\theta}_1$ es *más eficiente* que $\hat{\theta}_2$ si se verifica que $\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$

La *eficiencia relativa* se mide por la ratio: $\dfrac{\mathrm{Var}(\hat{\theta}_1)}{\mathrm{Var}(\hat{\theta}_2)}$

La *eficiencia* de los estimadores está limitada por las características de la distribución de probabilidad de la muestra de la que proceden.

Un estimador es *eficiente* cuando verifica: $\begin{cases} \textbf{Es insesgado} \\ \textbf{Posee varianza mínima} \end{cases}$

La cuestión de tener varianza mínima queda resuelta mediante la Cota de Cramér-Rao.

La varianza de un estimador verifica siempre la *Cota de Cramér-Rao*: $V(\hat{\theta}) \geq CCR$**. Un estimador será *eficiente* cuando** $V(\hat{\theta}) = CCR$

y la cota resulta $$\boxed{V(\hat{\theta}) \geq CCR = \frac{\left[1 + b(\hat{\theta})\right]^2}{nE\left[\dfrac{\vartheta \ln L(x,\theta)}{\vartheta\theta}\right]^2}}$$

Si el estimador es insesgado $b(\hat{\theta}) = 0$: $$V(\hat{\theta}) \geq CCR = \frac{1}{nE\left[\dfrac{\vartheta \ln L(x,\theta)}{\vartheta\theta}\right]^2}$$

En muestras aleatorias simples: $$V(\hat{\theta}) \geq CCR = \frac{\left[1 + b(\hat{\theta})\right]^2}{nE\left[\dfrac{\vartheta \ln L(x,\theta)}{\vartheta\theta}\right]^2}$$

Es preciso destacar que la *Cota de Cramér-Rao* (CCR) no tiene por qué tomar siempre un valor muy pequeño (cercano a cero).

Un estimador es *asintóticamente eficiente* sí: $\lim\limits_{x \to \infty} V(\hat{\theta}) = CCR$

El denominador de la Cota de Cramér-Rao es la *cantidad de información de Fisher* en una muestra:

$$I(\theta) = E\left[\frac{\vartheta \ln L(X,\theta)}{\vartheta\theta}\right]^2 \quad \text{o bien} \quad i(\theta) = E\left[\frac{\vartheta \ln L(x,\theta)}{\vartheta\theta}\right]^2 \quad \text{donde} \quad I(\theta) = n.i(\theta)$$

A la función $\ln L(X,\theta)$ **se llama *soporte* o *Log-verosimilitud.***

La cantidad de información de Fisher $I(\theta)$ puede simplificarse en una muestra aleatoria simple (m.a.s.), según sea el caso discreto o continuo, obteniendo la expresión:

$$\begin{cases} \text{Discreto:} \quad E\left[\dfrac{\vartheta \ln P(x_1, x_2, \cdots, x_n\ ;\ \theta)}{\vartheta\theta}\right]^2 = n.E\left[\dfrac{\vartheta \ln P(x\ ;\ \theta)}{\vartheta\theta}\right]^2 \\ \\ \text{Continuo:} \quad E\left[\dfrac{\vartheta \ln f(x_1, x_2, \cdots, x_n\ ;\ \theta)}{\vartheta\theta}\right]^2 = n.E\left[\dfrac{\vartheta \ln f(x\ ;\ \theta)}{\vartheta\theta}\right]^2 \end{cases}$$

MÉTODO DE MÁXIMA VEROSIMILITUD (EMV)

La estimación por máxima verosimilitud es un método de optimización que supone que la distribución de probabilidad de las observaciones es conocida.

Sea $(x_1, \cdots, x_n)$ una muestra aleatoria (no necesariamente simple) de una población X con función de masa P_θ (o función de densidad f_θ) donde $\theta = (\theta_1, \cdots, \theta_n)$.

El estimador de máxima verosimilitud (probabilidad conjunta) de θ es el formado por los valores $(\hat{\theta}_1, \cdots, \hat{\theta}_n)$ que maximizan la *función de verosimilitud* de la muestra $(x_1, \cdots, x_n)$ obtenida:

$$L(\theta) = L(X;\theta) = L(x_1, \cdots, x_n; \theta) = \begin{cases} P(x_1, \theta) \cdots P(x_n, \theta) & \text{caso discreto} \\ \\ f_\theta(x_1) \cdots f_\theta(x_n) & \text{caso continuo} \end{cases}$$

En muchas ocasiones, es más práctico encontrar el estimador de máxima verosimilitud es considerar la función soporte o *Log-verosimilitud* $\ln L(X, \theta)$, en lugar de la función de verosimilitud $L(\theta)$, ya que es más fácil de manejar y presenta los mismos máximos y mínimos.

Se despeja $\hat{\theta}=(\hat{\theta}_1, \cdots, \hat{\theta}_n)$ de la ecuación: $\left.\dfrac{\vartheta \ln L(\theta)}{\vartheta\theta}\right|_{\theta=\hat{\theta}} = 0$ y se obtiene el estimador de máxima verosimilitud E.M.V$(\hat{\theta})$

SUFICIENCIA

Un estimador $\hat{\theta}$ es *suficiente* cuando no da lugar a una pérdida de información. Es decir, cuando la información basada en $\hat{\theta}$ es tan buena como la que hiciera uso de toda la muestra.

Para identificar estadísticos suficientes se utiliza el criterio de factorización de Fisher-Neyman, que dice que dada una muestra aleatoria $(x_1, \cdots, x_n)$ de una población X con función de masa P_θ (o función de densidad f_θ) un estadístico $\hat{\theta}$ es suficiente para θ si y sólo sí:

$$\begin{cases} P_\theta(x_1, \cdots, x_n) = g\left[\hat{\theta}(x_1, \cdots, x_n), \theta\right].h(x_1, \cdots, x_n) & \text{caso discreto} \\ \\ f_\theta(x_1, \cdots, x_n) = g\left[\hat{\theta}(x_1, \cdots, x_n), \theta\right].h(x_1, \cdots, x_n) & \text{caso continuo} \end{cases}$$

Para encontrar un estadístico suficiente $\hat{\theta}$ hay que factorizar la función de verosimilitud de la forma: $L(\theta) = g(\hat{\theta}, \theta).h(x_1, \cdots, x_n)$

MÉTODO DE LOS MOMENTOS

El procedimiento consiste en igualar momentos poblacionales respecto al origen α_r a los correspondientes momentos muestrales respecto al origen a_r, formando así tantas ecuaciones como parámetros poblacionales se pretenden estimar:

$$\begin{cases} \alpha_1 = E(X) = \mu & \rightarrow & \hat{\alpha}_1 = a_1 = \dfrac{\sum_{i=1}^{n} x_i}{n} = \overline{x} \\ \alpha_2 = E(X^2) & \rightarrow & \hat{\alpha}_2 = a_2 = \dfrac{\sum_{i=1}^{n} x_i^2}{n} \\ \cdots\cdots\cdots & \rightarrow & \cdots\cdots\cdots \\ \alpha_r = E(X^r) & \rightarrow & \hat{\alpha}_r = a_r = \dfrac{\sum_{i=1}^{n} x_i^r}{n} \end{cases}$$

EJERCICIOS DE ESTIMADORES

La variable aleatoria poblacional "renta de las familias" del municipio de Madrid se distribuye siguiendo un modelo $N(\mu, \sigma)$. Se extraen muestras aleatorias simples de tamaño 4. Como estimadores del parámetro μ, se proponen los siguientes:

$$\hat{\mu}_1 = \frac{x_1 + 2x_2 + 3x_3}{6} \qquad \hat{\mu}_2 = \frac{x_3 - 4x_2}{-3} \qquad \hat{\mu}_3 = \overline{x}$$

Se pide:

a) Comprobar si los estimadores son insesgados
b) ¿Cuál es el más eficiente?
c) Si tuviera que escoger entre ellos, ¿cuál escogería?. Razone su respuesta a partir del Error Cuadrático Medio.

Solución:

a) Un estimador $\hat{\theta}$ es insesgado (o centrado) cuando se verifica $E(\hat{\theta}) = \theta$

$$E(\hat{\mu}_1) = E\left[\frac{x_1 + 2x_2 + 3x_3}{6}\right] = \frac{1}{6} E\left[x_1 + 2x_2 + 3x_3\right] =$$

$$= \frac{1}{6}\left[E(x_1) + 2E(x_2) + 3E(x_3)\right] = \frac{1}{6}\left[6\mu\right] = \mu$$

$$E(\hat{\mu}_2) = E\left[\frac{x_1 - 4x_2}{-3}\right] = -\frac{1}{3} E\left[x_1 - 4x_2\right] = -\frac{1}{3}\left[E(x_1) - 4E(x_2)\right] =$$

$$= -\frac{1}{3}\left[-3\mu\right] = \mu$$

$$E(\hat{\mu}_3) = E\left[\frac{x_1 + x_2 + x_3 + x_4}{4}\right] = \frac{1}{4} E\left[x_1 + x_2 + x_3 + x_4\right] =$$

$$= \frac{1}{4}\left[E(x_1) + E(x_2) + E(x_3) + E(x_4)\right] = \frac{1}{4}\left[4\mu\right] = \mu$$

Los tres estimadores son insesgados o centrados.

b) El estimador *más eficiente* es el que tenga menor varianza.

$$V\left[\hat{\mu}_1\right] = V\left[\frac{x_1+2x_2+3x_3}{6}\right] = \frac{1}{36}V\left[x_1+2x_2+3x_3\right] =$$

$$= \frac{1}{36}\left[V(x_1)+4V(x_2)+9V(x_3)\right] = \frac{1}{36}\left[14\sigma^2\right] = \frac{14}{36}\sigma^2 = 0,39\sigma^2$$

$$V\left[\hat{\mu}_2\right] = V\left[\frac{x_1-4x_2}{-3}\right] = \frac{1}{9}V\left[x_1-4x_2\right] = \frac{1}{9}\left[V(x_1)+16V(x_2)\right] =$$

$$= \frac{1}{9}\left[17\sigma^2\right] = \frac{17}{9}\sigma^2 = 1,89\sigma^2$$

$$V\left[\hat{\mu}_3\right] = V\left[\frac{x_1+x_2+x_3+x_4}{4}\right] = \frac{1}{16}V\left[x_1+x_2+x_3+x_4\right] =$$

$$= \frac{1}{16}\left[V(x_1)+V(x_2)+V(x_3)+V(x_4)\right] = \frac{1}{16}\left[4\sigma^2\right] = \frac{4}{16}\sigma^2 = 0,25\sigma^2$$

El estimador $\hat{\mu}_3$ es el más eficiente.

c) $ECM(\hat{\theta}) = E(\hat{\theta}-\theta)^2 = V(\hat{\theta}) + \left[b(\hat{\theta})\right]^2$ **sesgo** $b(\hat{\theta}) = \left[E(\hat{\theta})-\theta\right]$

Al ser los tres estimadores insesgados $b(\hat{\theta}) = 0 \;\mapsto\; ECM(\hat{\theta}) = V(\hat{\theta})$

El error cuadrático medio coincide con la varianza, por tanto, se elige el estimador $\hat{\mu}_3$ por presentar menor varianza.

🖫 **La variable aleatoria X representa los gastos mensuales de una empresa, cuya función de densidad es $f(\theta, x) = \theta x^{\theta-1}$ con $\theta > 0$ y $0 < x < 1$. Se realiza una m.a.s. de tamaño 3, y se proponen tres estimadores:**

$$\hat{\theta}_1 = \bar{x} \qquad \hat{\theta}_2 = \frac{x_1^2 + 2x_2^2 + 3x_3^2}{6} \qquad \hat{\theta}_3 = \frac{x_3 - 2x_1 + 4x_2}{6}$$

a) Calcule los sesgos

b) Si la muestra que se obtiene es (0,7; 0,1; 0,3), calcule las estimaciones puntuales

c) ¿Cuáles son las funciones estimadas para las estimaciones anteriores?

Solución:

a)

- $\hat{\theta}_1 = \bar{x} \Rightarrow E(\hat{\theta}_1) = E\left[\frac{x_1 + x_2 + x_3}{3}\right] = \frac{1}{3} E\left[x_1 + x_2 + x_3\right] = \frac{1}{3}(3\mu) = \mu$

donde

$$\mu = \int_{-\infty}^{\infty} x\, f(x,\theta)\, dx = \int_0^1 x\, f(x,\theta)\, dx = \int_0^1 x\, \theta x^{\theta-1}\, dx = \int_0^1 \theta\, x^{\theta}\, dx =$$

$$= \left[\frac{\theta\, x^{\theta+1}}{\theta+1}\right]_0^1 = \frac{\theta}{\theta+1}$$

Sesgo: $E(\hat{\theta}_1) = \dfrac{\theta}{\theta+1} \quad \mapsto \quad b(\hat{\theta}_1) = E(\hat{\theta}_1) - \theta = \dfrac{\theta}{\theta+1} - \theta = -\dfrac{\theta^2}{\theta+1}$

- $E(\hat{\theta}_2) = E\left[\dfrac{x_1^2 + 2x_2^2 + 3x_3^2}{6}\right] = \dfrac{1}{6}\left[\underbrace{E(x_1^2)}_{\alpha_2} + 2\underbrace{E(x_2^2)}_{\alpha_2} + 3\underbrace{E(x_3^2)}_{\alpha_2}\right] = \dfrac{1}{6}(6\alpha_2) = \alpha_2$

$$\alpha_2 = E(x^2) = \int_{-\infty}^{\infty} x^2\, f(x,\theta)\, dx = \int_0^1 x^2\, f(x,\theta)\, dx = \int_0^1 x^2\, \theta\, x^{\theta-1}\, dx =$$

$$= \int_0^1 \theta\, x^{\theta+1}\, dx = \left[\frac{\theta\, x^{\theta+2}}{\theta+2}\right]_0^1 = \frac{\theta}{\theta+2}$$

$$E(\hat{\theta}_2) = E\left[\frac{x_1^2 + 2x_2^2 + 3x_3^2}{6}\right] = \frac{1}{6}\left[\underbrace{E(x_1^2)}_{\alpha_2} + 2\underbrace{E(x_2^2)}_{\alpha_2} + 3\underbrace{E(x_3^2)}_{\alpha_2}\right] = \alpha_2 = \frac{\theta}{\theta+2}$$

Sesgo: $b(\hat{\theta}_2) = E\left(\hat{\theta}_2\right) - \theta = \dfrac{\theta}{\theta+2} - \theta = -\dfrac{\theta^2+\theta}{\theta+2}$

- $E(\hat{\theta}_3) = E\left[\dfrac{x_3 - 2x_1 + 4x_2}{6}\right] = \dfrac{1}{6}E\left[x_3 - 2x_1 + 4x_2\right] = \dfrac{1}{6}(3\mu) = \dfrac{1}{2}\mu$

$$\mu = \int_{-\infty}^{\infty} x\, f(x,\theta)\, dx = \int_0^1 x\, f(x,\theta)\, dx = \int_0^1 x\, \theta x^{\theta-1}\, dx = \int_0^1 \theta\, x^{\theta}\, dx =$$

$$= \left[\frac{\theta\, x^{\theta+1}}{\theta+1}\right]_0^1 = \frac{\theta}{\theta+1}$$

Sesgo: $b(\hat{\theta}_3) = E(\hat{\theta}_3) - \theta = \dfrac{1}{2}\left(\dfrac{\theta}{\theta+1}\right) - \theta = -\dfrac{2\theta^2+\theta}{2(\theta+1)}$

b) Las estimaciones puntuales de la muestra (0,7 ; 0,1 ; 0,3):

$$\hat{\theta}_1 = \frac{0{,}7 + 0{,}1 + 0{,}3}{3} = 0{,}367$$

$$\hat{\theta}_2 = \frac{0{,}7^2 + 2.0{,}1^2 + 3.0{,}3^2}{6} = 0{,}13$$

$$\hat{\theta}_3 = \frac{0{,}3 - 2.0{,}7 + 4.0{,}1}{6} = -\,0{,}117 \mapsto \text{no puede ser, puesto que } \hat{\theta} > 0$$

c) Las funciones estimadas para las estimaciones de la muestra:

$$\hat{\theta}_1 \Rightarrow f(0{,}367, x) = 0{,}367\, x^{0{,}367-1} = 0{,}367\, x^{-0{,}633}$$

$$\hat{\theta}_2 \Rightarrow f(0{,}13, x) = 0{,}13\, x^{0{,}13-1} = 0{,}367\, x^{-0{,}87}$$

Sea una población con media μ de la que se extraen m.a.s. de tamaño n. Considere los siguientes estimadores de la media:

$$\hat{\mu}_1 = \overline{x} \qquad\qquad \hat{\mu}_2 = \frac{1}{n+1}\sum_{i=1}^{n} x_i$$

a) Estudiar la insesgadez, la eficiencia relativa y la consistencia de ambos estimadores

b) Elegir uno de los dos en término del error cuadrático medio

Solución:

a) **Insesgadez:**

- $$E(\hat{\mu}_1) = E(\overline{x}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}(n\mu) = \mu$$

Sesgo: $b(\hat{\mu}_1) = E(\hat{\mu}_1) - \mu = \mu - \mu = 0$

- $$E(\hat{\mu}_2) = E\left(\frac{1}{n+1}\sum_{i=1}^{n} x_i\right) = \frac{1}{n+1}E\left(\sum_{i=1}^{n} x_i\right) = \frac{1}{n+1}\sum_{i=1}^{n} E(x_i) =$$
$$= \frac{1}{n+1}(n\mu) = \frac{n\mu}{n+1}$$

- $$b(\hat{\mu}_2) = E(\hat{\mu}_2) - \mu = \frac{n\mu}{n+1} - \mu = \frac{n\mu - n\mu - \mu}{n+1} = \overbrace{\underbrace{-\frac{\mu}{n+1}}_{\text{sesgado asintoticamente}}}^{\to 0 \text{ cuando } n \to \infty}$$

Eficiencia:

Sean dos estimadores insesgados $\hat{\theta}_1$ y $\hat{\theta}_2$ de un parámetro desconocido θ. Se dice $\hat{\theta}_1$ es más eficiente que $\hat{\theta}_2$ si se verifica:

$\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$. La eficiencia relativa se mide por la ratio $\dfrac{\mathrm{Var}(\hat{\theta}_1)}{\mathrm{Var}(\hat{\theta}_2)}$

$$V(\hat{\mu}_1) = V(\overline{x}) = V\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}$$

$$V(\hat{\mu}_2) = V\left(\frac{1}{n+1}\sum_{i=1}^{n} x_i\right) = \frac{1}{(n+1)^2}\sum_{i=1}^{n} V(x_i) = \frac{1}{(n+1)^2}(n\sigma^2) = \frac{n}{(n+1)^2}\sigma^2$$

Eficiencia relativa:

$$\frac{Var(\hat{\mu}_1)}{Var(\hat{\mu}_2)} = \frac{\sigma^2/n}{n\sigma^2/(n+1)^2} = \frac{(n+1)^2}{n^2} > 1 \mapsto Var(\hat{\mu}_1) > Var(\hat{\mu}_2)$$

El estimador $\hat{\mu}_2$ tiene menor varianza, por lo que es *más eficiente* que $\hat{\mu}_1$

Consistencia:

Un estimador $\hat{\theta}$ es *consistente* cuando $\lim\limits_{n\to\infty} E(\hat{\theta}) = \theta$ y $\lim\limits_{n\to\infty} V(\hat{\theta}) = 0$

$$\hat{\mu}_1 \equiv \begin{cases} \lim\limits_{n\to\infty} E(\hat{\mu}_1) = \lim\limits_{n\to\infty} E(\overline{x}) = \mu \\ \lim\limits_{n\to\infty} V(\hat{\mu}_1) = \lim\limits_{n\to\infty} \dfrac{\sigma^2}{n} = 0 \end{cases} \qquad \text{es consistente}$$

$$\hat{\mu}_2 \equiv \begin{cases} \lim\limits_{n\to\infty} E(\hat{\mu}_2) = \lim\limits_{n\to\infty} \left(\mu - \dfrac{1}{n+1}\mu\right) = \mu \\ \lim\limits_{n\to\infty} V(\hat{\mu}_2) = \lim\limits_{n\to\infty} \left[\dfrac{n}{(n+1)^2}\sigma^2\right] = 0 \end{cases} \qquad \text{es consistente}$$

b) $ECM(\hat{\theta}) = E(\hat{\theta}-\theta)^2 = V(\hat{\theta}) + \left[b(\hat{\theta})\right]^2 \qquad$ **sesgo** $b(\hat{\theta}) = \left[E(\hat{\theta}) - \theta\right]$

$$ECM(\hat{\mu}_1) = V(\hat{\mu}_1) + \left[b(\hat{\mu}_1)\right]^2 = \frac{\sigma^2}{n} + 0 = \frac{\sigma^2}{n}$$

$$ECM(\hat{\mu}_2) = V(\hat{\mu}_2) + \left[b(\hat{\mu}_2)\right]^2 = \frac{n}{(n+1)^2}\sigma^2 + \left(\frac{1}{n+1}\mu\right)^2 = \frac{n\sigma^2 + \mu^2}{(n+1)^2}$$

El estimador $\hat{\mu}_1$ será elegido sí $ECM(\hat{\mu}_1) \leq ECM(\hat{\mu}_2)$ $\mapsto \dfrac{\sigma^2}{n} \leq \dfrac{n\sigma^2 + \mu^2}{(n+1)^2}$

$$\frac{n\sigma^2 + \mu^2}{(n+1)^2} = \frac{n\sigma^2}{(n+1)^2} + \frac{\mu^2}{(n+1)^2} \quad \mapsto \quad \frac{\sigma^2}{n} \leq \frac{n\sigma^2}{(n+1)^2} + \frac{\mu^2}{(n+1)^2}$$

$$\frac{\sigma^2}{n} - \frac{n\sigma^2}{(n+1)^2} \leq \frac{\mu^2}{(n+1)^2} \Rightarrow \frac{(n+1)^2\sigma^2 - n^2\sigma^2}{n(n+1)^2} \leq \frac{\mu^2}{(n+1)^2} \Rightarrow$$

$$\Rightarrow \frac{(2n+1)\sigma^2}{n(n+1)^2} \leq \frac{\mu^2}{(n+1)^2} \Rightarrow \frac{(2n+1)}{n} \leq \frac{\mu^2}{\sigma^2}$$

$$\begin{cases} \text{Si } \dfrac{2n+1}{n} \leq \dfrac{\mu^2}{\sigma^2} \mapsto \hat{\mu}_1 \text{ se elige antes que } \hat{\mu}_2 \\ \\ \text{Si } \dfrac{2n+1}{n} \geq \dfrac{\mu^2}{\sigma^2} \mapsto \hat{\mu}_2 \text{ se elige antes que } \hat{\mu}_1 \end{cases}$$

Sea $(x_1, x_2, \cdots, x_n)$ una muestra aleatoria simple de una variable aleatoria X con $E(X) = \mu$ y $Var(X) = \sigma^2$
Calcular el error cuadrático medio para los siguientes estimadores de μ:

$$\hat{\mu}_1 = x_1 \qquad \hat{\mu}_2 = \frac{3x_1 - 2x_2 + x_3}{6}$$

Solución:

- **Insesgadez**

$$E(\hat{\mu}_1) = E(x_1) = \mu \mapsto b(\hat{\mu}_1) = 0$$

$$\begin{cases} E(\hat{\mu}_2) = E\left[\dfrac{3x_1 - 2x_2 + x_3}{6}\right] = \dfrac{1}{6}\left[3E(x_1) - 2E(x_2) + E(x_3)\right] = \\ \qquad = \dfrac{1}{6}(3\mu - 2\mu + \mu) = \dfrac{2}{6}\mu = \dfrac{1}{3}\mu \\ b(\hat{\mu}_2) = E(\hat{\mu}_2) - \mu = \dfrac{1}{3}\mu - \mu = -\dfrac{2}{3}\mu \text{ sesgo} \end{cases}$$

Respecto al sesgo es mejor el primer estimador $\hat{\mu}_1$ que es insesgado o centrado.

- **Varianza**

$$Var(\mu_1) = \sigma^2$$

$$\mathrm{Var}(\mu_2) = \mathrm{Var}\left(\frac{3x_1 - 2x_2 + x_3}{6}\right) = \frac{1}{36}\mathrm{Var}\left(3x_1 - 2x_2 + x_3\right) =$$

$$= \frac{1}{36}\left[9\,\mathrm{Var}(x_1) + 4\,\mathrm{Var}(x_2) + \mathrm{Var}(x_3)\right] = \frac{1}{36}\left[9\sigma^2 + 4\sigma^2 + \sigma^2\right] =$$

$$= \frac{14}{36}\sigma^2 = \frac{7}{18}\sigma^2$$

Respecto a la varianza es mejor el segundo estimador $\hat{\mu}_2$ por ser

$$\frac{7}{18}\sigma^2 < \sigma^2$$

- **El mejor estimador será el que presente menor Error Cuadrático Medio**

$$\mathrm{ECM}(\hat{\mu}_1) = V(\hat{\mu}_1) + \left[b(\hat{\mu}_1)\right]^2 = \sigma^2 + 0 = \sigma^2$$

$$\mathrm{ECM}(\hat{\mu}_2) = V(\hat{\mu}_2) + \left[b(\hat{\mu}_2)\right]^2 = \frac{7}{18}\sigma^2 + \left(-\frac{2}{3}\mu\right)^2 = \frac{7}{18}\sigma^2 + \frac{4}{9}\mu^2$$

El primer estimador $\hat{\mu}_1$ será mejor sí

$$\sigma^2 < \frac{7}{18}\sigma^2 + \frac{4}{9}\mu^2 \quad \mapsto \quad \frac{11}{18}\sigma^2 < \frac{4}{9}\mu^2 \quad \mapsto \quad \sigma^2 < \frac{4 \times 18}{9 \times 11}\mu^2 \quad \mapsto \quad \sigma^2 < \frac{8}{11}\mu^2$$

La distribución del peso de las manzanas de una determinada cosecha sigue una distribución normal, cuyo peso medio es desconocido y cuya desviación típica es 7 gramos. Se pide:

a) Analizar cuál de los estimadores $\hat{\mu}_1$, $\hat{\mu}_2$ del peso medio es mejor respecto del sesgo y de la eficiencia, para una muestra aleatoria simple de tamaño cinco.

b) Si $\hat{\mu}_1 = \dfrac{\sum_{i=1}^{5} x_i}{5}$ y $\hat{\mu}_2 = x_1 + 2x_2 + 3x_3 - 4x_4 - x_5$, obtener los pesos medios estimados a partir de la muestra (125, 135, 130, 137, 142).

Solución:

a) El peso de las manzanas sigue una distribución $N(\mu, 7)$

Se calculan las esperanzas para analizar el sesgo de los estimadores.

$$E(\hat{\mu}_1) = E\left[\sum_{i=1}^{5} x_i / 5\right] = \frac{1}{5} E\left[\sum_{i=1}^{5} x_i\right] = \frac{1}{5}\sum_{i=1}^{5} E\left[x_i\right] = \frac{1}{5}(5\mu) = \mu$$

$$E(\hat{\mu}_2) = E(x_1 + 2x_2 + 3x_3 - 4x_4 - x_5) = E(x_1) + 2E(x_2) + 3E(x_3) - 4E(x_4) - E(x_5) =$$
$$= \mu + 2\mu + 3\mu - 4\mu - \mu = \mu$$

Los estimadores $\hat{\mu}_1$, $\hat{\mu}_2$ son insesgados (centrados).

b) Para analizar la eficiencia de los estimadores se obtienen las varianzas:

$$V(\hat{\mu}_1) = V\left[\sum_{i=1}^{5} x_i / 5\right] = \frac{1}{25} V\left[\sum_{i=1}^{5} x_i\right] = \frac{1}{25}\sum_{i=1}^{5} V\left[x_i\right] = \frac{1}{25}(5.49) = \frac{49}{5}$$

$$V(\hat{\mu}_2) = V(x_1 + 2x_2 + 3x_3 - 4x_4 - x_5) =$$
$$= V(x_1) + 4V(x_2) + 9V(x_3) + 16V(x_4) + V(x_5) =$$
$$= (49) + 4\,(49) + 9\,(49) + 16\,(49) + (49) = 31\,(49) = 1519$$

Como los dos estimadores son insesgados y $V(\hat{\mu}_1) < V(\hat{\mu}_2)$ se elige como mejor el estimador $\hat{\mu}_1$, el peso medio de la muestra de las cinco manzanas.

Sea $(x_1, x_2, x_3, x_4, x_5)$ una muestra aleatoria simple de una variable aleatoria X con distribución normal con media $(\mu - 5)$ y varianza σ^2. Se proponen los siguientes estimadores:

$$\hat{\mu}_1 = \sum_{i=1}^{5} x_i \qquad \hat{\mu}_2 = 8x_2 - 3x_5$$

Determinar cuál es el mejor estimador para μ. Justificar la respuesta.

Solución:

- Insesgadez

$$E(\hat{\mu}_1) = E\left(\sum_{i=1}^{5} x_i\right) = E\left(x_1 + x_2 + x_3 + x_4 + x_5\right) = 5(\mu - 5)$$

$$E(\hat{\mu}_2) = E\left[8x_2 - 3x_5\right] = 8E(x_2) - 3E(x_5) = 8(\mu - 5) - 3(\mu - 5) = 5(\mu - 5)$$

Ambos estimadores son sesgados, con idéntico sesgo:

$$b(\hat{\mu}_1) = b(\hat{\mu}_2) = 5(\mu - 5) - \mu = 4\mu - 25$$

- Varianza

$$Var(\mu_1) = Var\left(\sum_{i=1}^{5} x_i\right) = 5\,Var(x) = 5\sigma^2$$

$$Var(\mu_2) = Var\left(8x_2 - 3x_5\right) = 8^2\,Var(x_2) + 3^2\,Var(x_5) = 64\sigma^2 + 9\sigma^2 = 73\sigma^2$$

Dado que los dos estimadores tienen el mismo sesgo y el primer estimador μ_1 tiene menor varianza, será el estimador óptimo.
Puede observarse que presenta el menor Error Cuadrático Medio:

$$ECM(\hat{\mu}_1) = V(\hat{\mu}_1) + \left[b(\hat{\mu}_1)\right]^2 = 5\sigma^2 + \left(4\mu - 25\right)^2$$

$$ECM(\hat{\mu}_2) = V(\hat{\mu}_2) + \left[b(\hat{\mu}_2)\right]^2 = 73\sigma^2 + \left(4\mu - 25\right)^2$$

🖫 **El peso en kilos de los jamones vendidos por una empresa sigue una distribución normal con varianza 4 y peso medio desconocido. Se conoce que el peso medio de los jamones vendidos es superior a 5 kg, y se toman m.a.s. de tamaño 4 para estimar θ. ¿Cuál de los dos estimadores sería el mejor respondiendo a la insesgadez y eficiencia?**

$$\hat{\theta}_1 = \frac{x_1 + x_2 + x_3}{4} \qquad \hat{\theta}_2 = \frac{x_1 + x_2}{2}$$

Solución:

La v.a $X_i =$ 'peso en kg de los jamones' sigue una distribución normal de varianza $\sigma^2 = 4$

Para estudiar la insesgadez de los estimadores se calculan sus esperanzas:

- $E(\hat{\theta}_1) = E\left[\dfrac{x_1 + x_2 + x_3}{4}\right] = \dfrac{1}{4}\left[E(x_1) + E(x_2) + E(x_3)\right] = \dfrac{3}{4}\theta$

El sesgo del estimador $\hat{\theta}_1$ será: $b(\hat{\theta}_1) = E(\hat{\theta}_1) - \theta = \dfrac{3}{4}\theta - \theta = -\dfrac{1}{4}\theta$

- $E(\hat{\theta}_2) = E\left[\dfrac{x_1 + x_2}{2}\right] = \dfrac{1}{2}\left[E(x_1) + E(x_2)\right] = \dfrac{2}{2}\theta = \theta$

El estimador $\hat{\theta}_2$ es insesgado, $b(\hat{\theta}_2) = 0$

Atendiendo al sesgo se elige $\hat{\theta}_2$

Para analizar la eficiencia relativa de los dos estimadores se calculan las respectivas varianzas

- $$V(\hat{\theta}_1) = V\left[\frac{x_1 + x_2 + x_3}{4}\right] = \frac{1}{16}\underbrace{\left[V(x_1 + x_2 + x_3)\right]}_{\text{Observaciones independientes}} =$$

$$= \frac{1}{16}\left[V(x_1) + V(x_2) + V(X_3)\right] \underset{V(x_i) = 4}{=} \frac{1}{16}12 = \frac{12}{16} = \frac{3}{4}$$

- $V(\hat{\theta}_2) = V\left[\dfrac{x_1 + x_2}{2}\right] = \dfrac{1}{4}\left[V(x_1 + x_2)\right] = \dfrac{1}{4}\left[V(x_1) + V(x_2)\right] = \dfrac{1}{4}8 = 2$

Respecto a la varianza se elige el estimador $\hat{\theta}_1$ por ser el de menor varianza.

Aparecen decisiones contrapuestas, de modo que el estimador se elige en base al error cuadrático medio: $\text{ECM} = \text{Varianza} + (\text{sesgo})^2$

$$\text{ECM}(\hat{\theta}_1) = \frac{3}{4} + \left(-\frac{\theta}{4}\right)^2 = \frac{\theta^2 + 12}{16} \qquad \text{ECM}(\hat{\theta}_2) = 2 + 0 = 2$$

Se analiza cuando es mayor el ECM el primer estimador $\hat{\theta}_1$:

$$\text{ECM}(\hat{\theta}_1) > \text{ECM}(\hat{\theta}_2) \quad \mapsto \quad \frac{\theta^2 + 12}{16} > 2 \quad \Rightarrow \quad \theta^2 > 20 \quad \Rightarrow \quad |\theta| > \sqrt{20} \approx 4{,}47$$

Si θ es en valor absoluto mayor que 4,47, el error cuadrático medio de $\hat{\theta}_1$ es mayor, con lo que se elige el estimador $\hat{\theta}_2$.

Se conoce que el peso medio de los jamones es superior a 5 kg, no queda duda que el estimador a elegir (con menor error cuadrático medio) es $\hat{\theta}_2$

Supongamos que la distribución de ingresos de una cierta población es una variable aleatoria con media μ desconocida y varianza σ^2 también desconocida. Si queremos estimar el ingreso medio de la población mediante una m.a.s. de tamaño n, respecto de la insesgadez y de la eficiencia. ¿Cuál de los dos estimadores elegiríamos?. ¿Son consistentes?

$$\hat{\mu}_1 = \frac{\sum_{i=1}^{n} x_i}{n-1} \qquad \hat{\mu}_2 = \frac{\sum_{i=1}^{n} x_i}{n}$$

Solución:

La v.a $x_i = \text{'ingresos de cierta población'}$ sigue una distribución normal $N(\mu, \sigma)$

Para analizar el sesgo de los estimadores, se calcula la esperanza:

- $$E(\hat{\mu}_1)=E\left[\sum_{i=1}^{n}\frac{x_i}{(n-1)}\right]=\frac{1}{n-1}E\left[\sum_{i=1}^{n}x_i\right]=\frac{1}{n-1}\sum_{i=1}^{n}E(x_i)=$$
$$=\frac{1}{n-1}(n\mu)=\frac{n}{n-1}\mu$$

El sesgo del estimador $\hat{\mu}_1$ será: $b(\hat{\mu}_1)=E(\hat{\mu}_1)-\mu=\frac{n}{n-1}\mu-\mu=\frac{1}{n-1}\mu$

- $$E(\hat{\mu}_2)=E\left[\sum_{i=1}^{n}\frac{x_i}{n}\right]=\frac{1}{n}E\left[\sum_{i=1}^{n}x_i\right]=\frac{1}{n}\sum_{i=1}^{n}E(x_i)=\frac{1}{n}(n\mu)=\mu$$

El estimador $\hat{\mu}_2$, que es la media muestral, es insesgado (centrado) y el estimador que se elige.

- **La eficiencia de los estimadores se analiza a través de su varianza:**

- $$V(\hat{\mu}_1)=V\left[\sum_{i=1}^{n}\frac{x_i}{n-1}\right]=\frac{1}{(n-1)^2}V\left[\sum_{i=1}^{n}x_i\right]=\frac{1}{(n-1)^2}\sum_{i=1}^{n}V(x_i)=$$
$$=\frac{1}{(n-1)^2}(n\sigma^2)=\frac{n\sigma^2}{(n-1)^2}$$

- $$V(\hat{\mu}_2)=V\left[\sum_{i=1}^{n}\frac{x_i}{n}\right]=\frac{1}{n^2}V\left[\sum_{i=1}^{n}x_i\right]=\frac{1}{n}\sum_{i=1}^{n}V(x_i)=\frac{1}{n^2}(n\sigma^2)=\frac{\sigma^2}{n}$$

El estimador más eficiente será el de menor varianza. Comparando las varianzas de los estimadores:

$$V(\hat{\mu}_2)=\frac{\sigma^2}{n}<\frac{n\sigma^2}{(n-1)^2}=V(\hat{\mu}_1)\quad \text{puesto que } (n-1)^2<n^2$$

El estimador $\hat{\mu}_2$, que es la media muestral, es el mejor tanto al sesgo como a la eficiencia.

Los dos estimadores son consistentes:

$$\begin{cases} \lim\limits_{n\to\infty} E(\hat{\mu}_1) = \mu \quad y \quad \lim\limits_{n\to\infty} V(\hat{\mu}_1) = \lim\limits_{n\to\infty} \dfrac{n\sigma^2}{(n-1)^2} = 0 \\ \lim\limits_{n\to\infty} E(\hat{\mu}_2) = \mu \quad y \quad \lim\limits_{n\to\infty} V(\hat{\mu}_2) = \lim\limits_{n\to\infty} \dfrac{\sigma^2}{n} = 0 \end{cases}$$

Nota: El estimador de máxima verosimilitud (probabilidad conjunta) de θ es el formado por los valores $(\hat{\theta}_1, \cdots, \hat{\theta}_n)$ que maximizan la *función de verosimilitud* de la muestra $(x_1, \cdots, x_n)$ obtenida:

$$L(\theta) = L(X;\theta) = L(x_1, \cdots, x_n; \theta) = P(x_1, \theta) \cdots P(x_n, \theta) \quad \text{caso discreto}$$

En muchas ocasiones, es más práctico encontrar el estimador de máxima verosimilitud al considerar la función *Soporte* o *Log-verosimilitud* $\ln L(X, \theta)$, en lugar de la función de verosimilitud $L(\theta)$, ya que es más fácil de manejar y presenta los mismos máximos y mínimos.

Se despeja $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_n)$ de la ecuación: $\left.\dfrac{\vartheta \ln L(\theta)}{\vartheta\theta}\right|_{\theta=\hat{\theta}} = 0$ y se obtiene el estimador de máxima verosimilitud E.M.V$(\hat{\theta})$

Un atleta olímpico de salto de altura se enfrenta a un listón de 2,3 metros. Su entrenador desea estudiar el comportamiento del saltador. Sabe que el número de saltos fallidos por hora es una variable aleatoria distribuida como una Poisson de parámetro λ.

a) Calcular el estimador máximo verosímil del parámetro λ.

b) Analizar sus propiedades.

Solución:

a) Sea la v.a. X = 'número de saltos fallidos por hora'

En la distribución de Poisson: $P(X=x) = \dfrac{\lambda^x}{x!} e^{-\lambda} \quad \begin{cases} E(x) = \lambda \\ V(x) = \lambda \end{cases}$

La función de verosimilitud $L(X, \lambda)$ en una muestra aleatoria simple de tamaño n:

$$L(\lambda) = L(X,\lambda) = \prod_{i=1}^{n} P(x_i,\lambda) = \frac{\lambda^{x_1}}{x_1!}e^{-\lambda}\cdots\frac{\lambda^{x_n}}{x_n!}e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}e^{-n\lambda}$$

$$L(X,\lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}e^{-n\lambda} \Rightarrow \ln L(X,\lambda) = \ln\left[\frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}e^{-n\lambda}\right] =$$

$$= \ln(\lambda^{\sum_{i=1}^{n} x_i}) - \ln(\prod_{i=1}^{n} x_i!) + \ln(e^{-n\lambda}) = \sum_{i=1}^{n} x_i \ln\lambda - \sum_{i=1}^{n} \ln(x_i!) - n\lambda$$

$$\ln L(X,\lambda) = \sum_{i=1}^{n} x_i \ln\lambda - \sum_{i=1}^{n} \ln(x_i!) - n\lambda$$

Para obtener el estimador de máxima verosimilitud $EMV(\hat{\lambda})$, se deriva la expresión anterior respecto del parámetro para obtener, sustituyendo λ por $\hat{\lambda}$

$$\left.\frac{\vartheta \ln L(X,\lambda)}{\vartheta\lambda}\right|_{\lambda=\hat{\lambda}} = 0 \equiv \left.\sum_{i=1}^{n} x_i \frac{1}{\lambda} - n\right|_{\lambda=\hat{\lambda}} = 0 \Rightarrow \hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}$$

El estimador de máxima verosimilitud (estimador que ofrece mayor credibilidad) viene dado por la media muestral $EMV(\hat{\lambda}) = \overline{x}$

b) Propiedades de Insesgadez, Consistencia y Eficiencia

- **Insesgadez**

El estimador $\hat{\lambda}$ es insesgado (centrado) sí $E(\hat{\lambda}) = \lambda$

$$E(\hat{\lambda}) = E\left[\frac{\sum_{i=1}^{n} x_i}{n}\right] = \frac{1}{n}E(\sum_{i=1}^{n} x_i) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}(n\lambda) = \lambda$$

$$V(\hat{\lambda}) = V(\bar{x}) = V\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) = \frac{1}{n^2}(n\lambda) = \frac{\lambda}{n}$$

- **Consistencia**

Cuando no es posible emplear estimadores de máxima verosimilitud, el requisito mínimo deseable para un estimador es que sea consistente.

El estimador $\hat{\lambda}$ es *consistente* cuando $\lim_{n\to\infty} E(\hat{\lambda}) = \lambda$ y $\lim_{n\to\infty} V(\hat{\lambda}) = 0$

$$\lim_{n\to\infty} E(\hat{\lambda}) = \lim_{n\to\infty} \lambda = \lambda \quad \text{y} \quad \lim_{n\to\infty} V(\hat{\lambda}) = \lim_{n\to\infty} \frac{\lambda}{n} = 0$$

El estimador $\hat{\lambda}$ es consistente

Para que un estimador sea *eficiente* tiene que ser *centrado* y de *varianza mínima*. La varianza mínima se analiza en virtud de la acotación de Cramer-Rao:

$$V(\hat{\lambda}) \geq CCR = \frac{\left[1 + b(\hat{\lambda})\right]^2}{E\left[\dfrac{\vartheta \ln P(x_1, x_2, \cdots, x_n ; \lambda)}{\vartheta\lambda}\right]^2}$$

La cantidad de información de Fisher $I(\lambda)$ puede simplificarse en una muestra aleatoria simple (m.a.s.), obteniendo la expresión:

$$I(\lambda) = E\left[\frac{\vartheta \ln P(x_1, x_2, \cdots, x_n ; \lambda)}{\vartheta\lambda}\right]^2 = n.E\left[\frac{\vartheta \ln P(x ; \lambda)}{\vartheta\lambda}\right]$$

$$V(\hat{\lambda}) \geq CCR = \frac{\left[1 + b(\hat{\lambda})\right]^2}{n.E\left[\dfrac{\vartheta \ln P(x ; \lambda)}{\vartheta\lambda}\right]^2} \overset{b(\hat{\lambda})=0}{\mapsto} V(\hat{\lambda}) \geq CCR = \frac{1}{n.E\left[\dfrac{\vartheta \ln P(x ; \lambda)}{\vartheta\lambda}\right]^2}$$

El estimador es eficiente cuando $V(\hat{\lambda}) = CCR$

- **Eficiencia**

Siendo, $P(X = x) = \dfrac{\lambda^x}{x!} e^{-\lambda}$

$$\ln P(x,\lambda) = \ln\left[\frac{\lambda^x}{x!}e^{-\lambda}\right] = x\ln\lambda - \ln(x!) - \lambda$$

$$\frac{\vartheta \ln P(x,\lambda)}{\vartheta\lambda} = \frac{x}{\lambda} - 1 = \frac{x-\lambda}{\lambda}$$

$$E\left[\frac{\vartheta \ln P(x,\lambda)}{\vartheta\lambda}\right]^2 = E\left[\frac{x-\lambda}{\lambda}\right]^2 = \frac{1}{\lambda^2}E(x-\lambda)^2 = \frac{1}{\lambda^2}E(x-\overline{x})^2 = \frac{1}{\lambda^2}V(x) =$$
$$= \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

En consecuencia, $V(\hat{\lambda}) \geq \dfrac{1}{n\dfrac{1}{\lambda}} = \dfrac{\lambda}{n}$

El menor valor de la varianza del estimador será $\dfrac{\lambda}{n}$

Se sabe que $V(\hat{\lambda}) = V(\overline{x}) = \dfrac{\lambda}{n}$, lo que muestra $V(\hat{\lambda}) = CCR$, el estimador empleado es *eficiente*.

🖫 Sea la distribución $N(\mu,\sigma)$, con la media y varianza desconocidas. Calcular los estimadores máximo-verosímiles de μ y σ^2

Solución:

Función de verosimilitud:

$L(X;\,\mu,\sigma^2) =$

$$= \left[\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_1-\mu)^2}{2\sigma^2}}\right]\left[\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_2-\mu)^2}{2\sigma^2}}\right]\cdots\left[\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}\right] =$$

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

Función soporte o Log-verosimilitud:

$$\ln\left[L(X;\,\mu,\sigma^2)\right]=\ln\left[\frac{1}{(2\pi)^{\frac{n}{2}}\,(\sigma^2)^{\frac{n}{2}}}\,e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\,\sigma^2}}\right]=$$

$$=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln(\sigma^2)-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\,\sigma^2}$$

$$\ln\left[L(X;\,\mu,\sigma^2)\right]=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln(\sigma^2)-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\,\sigma^2}$$

Para obtener los estimadores de máxima verosimilitud de $\hat{\mu}$ y $\hat{\sigma}^2$ se deriva la expresión anterior respecto de los parámetros, sustituyendo μ por $\hat{\mu}$ e σ^2 por $\hat{\sigma}^2$, e igualando a cero:

$$\left[\frac{\vartheta \ln L(X;\,\mu,\sigma^2)}{\vartheta\mu}\right]_{\mu=\hat{\mu}}=\frac{\sum_{i=1}^{n}(x_i-\hat{\mu})}{\sigma^2}=0 \quad\mapsto\quad \hat{\mu}=\frac{\sum_{i=1}^{n}x_i}{n}=\overline{x}$$

$$\left[\frac{\vartheta \ln L(X;\,\mu,\sigma^2)}{\vartheta\sigma^2}\right]_{\sigma^2=\hat{\sigma}^2}=-\frac{n}{\hat{\sigma}}+\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\hat{\sigma}^3}=0 \quad\mapsto\quad \hat{\sigma}^2=\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$$

En conclusión, $\quad \hat{\mu}=\overline{x}$ **y** $\quad \hat{\sigma}^2=\dfrac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n}=\sigma_x^2$

La condición de máximo se verifica, pues: $\left[\dfrac{\vartheta^2 \ln L(X;\,\mu)}{\vartheta\mu^2}\right]_{\mu=\hat{\mu}}=-\dfrac{n}{\sigma^2}<0$

Los estimadores máximo-verosímiles de μ y σ^2 son la media y la varianza muestrales.

🖫 En una distribución $N(\mu,\sigma)$, se estima la media poblacional μ mediante la media de una muestra aleatoria simple $(x_1, x_2, \cdots, x_n)$. El estimador es insesgado y su varianza $V(\overline{x}) = \dfrac{\sigma^2}{n}$.

Demostrar que la media muestral es un estimador eficiente.

Solución:

La varianza de un estimador verifica siempre la *Cota de Cramer-Rao*: $V(\hat{\theta}) \geq CCR$. Un estimador es eficiente cuando $V(\hat{\theta}) = CCR$

Para obtener la cota de Cramer-Rao se parte de la función de densidad poblacional:

$$CCR = \frac{1}{nE\left[\dfrac{\vartheta \ln f(x,\mu)}{\vartheta\mu}\right]^2}$$

Función de densidad poblacional: $f(x,\mu) = \dfrac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Tomando logaritmos neperianos, se tiene:

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

Derivando respecto a μ:

$$\frac{\vartheta \ln f(x,\mu)}{\vartheta\mu} = -\frac{1}{2\sigma^2}\left[-2(x-\mu)\right] = \frac{x-\mu}{\sigma^2} \quad \mapsto \quad \left(\frac{\vartheta \ln f(x,\mu)}{\vartheta\mu}\right)^2 = \frac{(x-\mu)^2}{\sigma^4}$$

La esperanza matemática:

$$E\left[\frac{\vartheta \ln f(x,\mu)}{\vartheta\mu}\right]^2 = E\left[\frac{(x-\mu)^2}{\sigma^4}\right] = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}$$

Sustituyendo el valor de la esperanza matemática en la expresión de la cota para estimadores insesgados:

$$CCR=\frac{1}{nE\left[\frac{\vartheta \ln f(x,\mu)}{\vartheta\mu}\right]^2}=\frac{1}{n\,\frac{1}{\sigma^2}}=\frac{\sigma^2}{n}$$

La varianza del estimador coincide con la cota de Cramer-Rao, $V(\overline{x})=CCR$, concluyendo que la media muestral es un estimador eficiente de la media poblacional en la distribución normal.

Calcular el estimador máximo verosímil del parámetro 'a' de las funciones:

a) $f(x,a)=a^2e^{-ax}$ siendo $x\geq 0$ en muestras aleatorias simples de tamaño n

b) $f(x,a)=ae^{-ax}$ para $x\geq 0,\ a>0$ en muestras aleatorias simples de tamaño 2

Solución:

a) $f(x,a)=a^2e^{-ax}$ $f(x,a)=a^2e^{-ax}$ donde $x\geq 0$ en m.a.s. de tamaño n

Función de verosimilitud:

$$L(X,a)=L(x_1,x_2,\cdots,x_n;a)=(a^2e^{-ax_1}).(a^2e^{-ax_2})\cdots(a^2e^{-ax_n})=a^{2n}e^{-a\sum_{i=1}^{n}x_i}$$

Función soporte o Log-verosimilitud:

$$\ln L(X,a)=\ln\left(a^{2n}e^{-a\sum_{i=1}^{n}x_i}\right)=2n\ln a-a\sum_{i=1}^{n}x_i$$

Para obtener el estimador de máxima verosimilitud $EMV(\hat{a})$, se deriva la expresión anterior respecto del parámetro para obtener, sustituyendo a por $\hat{a}$ e igualando a cero:

$$\left.\frac{\vartheta \ln L(X,a)}{\vartheta a}\right|_{a=\hat{a}}=\left.\frac{2n}{a}-\sum_{i=1}^{n}x_i\right|_{a=\hat{a}}=0\ \Rightarrow\ \hat{a}=\frac{2n}{\sum_{i=1}^{n}x_i}=\frac{2}{\overline{x}}\quad\mapsto\quad\hat{a}=\frac{2}{\overline{x}}$$

b) Sea $f(x,a)=ae^{-ax}$ para $x\geq 0,\ a>0$ en m.a.s. de tamaño 2

Función de verosimilitud:

$$L(X,a)=L(x_1,x_2;a)=(ae^{-ax_1}).(ae^{-ax_2})=a^2e^{-a(x_1+x_2)}$$

Función soporte o Log-verosimilitud:

$$\ln L(X,a)=\ln\left[a^2e^{-a(x_1+x_2)}\right]=2\ln a-a(x_1+x_2)$$

Derivando respecto de a, igualando a cero y sustituyendo a por $\hat{a}$:

$$\left.\frac{\vartheta \ln L(X,a)}{\vartheta a}\right|_{a=\hat{a}}=\left.\frac{2}{a}-(x_1+x_2)\right|_{a=\hat{a}}=0 \Rightarrow \hat{a}=\frac{2}{x_1+x_2}=\frac{1}{\bar{x}}$$

En una distribución $N(\mu,\sigma)$ se estima la media poblacional μ en una muestra aleatoria simple de tamaño n $(x_1,x_2,\cdots,x_n)$ por medio de la función muestral:

$$\hat{\mu}=\frac{\sum_{i=1}^{n} i x_i}{n}$$

Estúdiese la eficiencia del estimador.

Solución:

Un estimador es eficiente cuando $V(\hat{\theta})=CCR$

Para hallar la cota de Cramer-Rao se necesita saber en primer lugar si el estimador es insesgado o no, se calcula para ello la esperanza matemática:

$$E(\hat{\mu})=E\left[\frac{\sum_{i=1}^{n} i x_i}{n}\right]=\frac{1}{n}\sum_{i=1}^{n}E(ix_i)=$$

$$=\frac{1}{n}\sum_{i=1}^{n} i\,E(x_i)=\frac{1}{n}\left[E(x_1)+2E(x_2)+3E(x_3)+\cdots+nE(x_n)\right]=$$

$$=\frac{1}{n}\left[\mu+2\mu+3\mu+\cdots+n\mu\right]=\frac{\mu}{n}\left[1+2+3+\cdots+n\right]=$$

$$=\frac{\mu}{\cancel{n}}\,\frac{(n+1)\,\cancel{n}}{2}=\frac{(n+1)}{2}\mu$$

$$E(\hat{\mu})=\mu+b(\hat{\mu}) \;\mapsto\; \text{sesgo } b(\hat{\mu})=E(\hat{\mu})-\mu=\frac{n+1}{2}\mu-\mu=\frac{n-1}{2}\mu$$

Varianza del estimador:

$$V(\hat{\mu})=V\left[\frac{\sum_{i=1}^{n} i x_i}{n}\right]=\frac{1}{n^2}\sum_{i=1}^{n}V(i x_i)=\frac{1}{n^2}\sum_{i=1}^{n} i^2\,V(x_i)=$$

$$=\frac{1}{n^2}\left[1^2V(x_1)+2^2V(x_2)+3^2V(x_3)+\cdots+n^2V(x_n)\right]=$$

$$=\frac{1}{n^2}\left[1^2\sigma^2+2^2\sigma^2+3^2\sigma^2+\cdots+n^2\sigma^2\right]=\frac{\sigma^2}{n^2}\left[1^2+2^2+3^2+\cdots+n^2\right]=$$

$$=\frac{\sigma^2}{n^{\cancel{2}}}\,\frac{\cancel{n}\,(n+1)(2n+1)}{6} \;\mapsto\; V(\hat{\mu})=\frac{\sigma^2\,(n+1)(2n+1)}{6n}$$

Cota de Cramer-Rao $CCR=\dfrac{\left[1+b(\hat{\mu})\right]^2}{nE\left[\dfrac{\vartheta\ln L(x,\mu)}{\vartheta\mu}\right]^2}$

$$\left[1+b(\hat{\mu})\right]^2=\left[1+\frac{n-1}{2}\mu\right]^2=\frac{(n+1)^2}{4}$$

La Información de Fisher $I(\mu)=n\,I_0(\mu)$ **de la muestra:**

$$I_0(\mu)=E\left[\frac{\vartheta\ln L(X,\mu)}{\vartheta\mu}\right]^2=nE\left[\frac{\vartheta\ln L(x,\mu)}{\vartheta\mu}\right]^2$$

se obtiene a partir de la función de densidad poblacional

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

Tomando logaritmos neperianos, se tiene:

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

Derivando respecto a μ:

$$\frac{\vartheta \ln f(x,\mu)}{\vartheta\mu}=-\frac{1}{2\sigma^2}\big[-2(x-\mu)\big]=\frac{x-\mu}{\sigma^2}\ \mapsto\ \left(\frac{\vartheta \ln f(x,\mu)}{\vartheta\mu}\right)^2=\frac{(x-\mu)^2}{\sigma^4}$$

La esperanza matemática: $E\left[\dfrac{\vartheta \ln f(x,\mu)}{\vartheta\mu}\right]^2=E\left[\dfrac{(x-\mu)^2}{\sigma^4}\right]=\dfrac{\sigma^2}{\sigma^4}=\dfrac{1}{\sigma^2}$

Atendiendo a la aditividad de esta medida: $I(\mu)=n\,I_0(\mu)=\dfrac{n}{\sigma^2}$

$$CCR=\frac{\big[1+b(\hat{\mu})\big]^2}{n\,E\left[\dfrac{\vartheta \ln L(x,\mu)}{\vartheta\mu}\right]^2}=\frac{\dfrac{(n+1)^2}{4}}{\dfrac{n}{\sigma^2}}=\frac{(n+1)^2\,\sigma^2}{4n}\ \mapsto\ V(\hat{\mu})=\frac{\sigma^2\,(n+1)(2n+1)}{6n}$$

Al ser la varianza del estimador mayor que la cota de Cramer-Rao:

$$V(\hat{\mu})=\frac{\sigma^2\,(n+1)(2n+1)}{6n}>\frac{(n+1)^2\,\sigma^2}{4n}=CCR$$

El estimador *no es eficiente*, siempre que $n>1$

Para $n=1$: $V(\hat{\mu})=CCR=\dfrac{\sigma^2}{n}$ el estimador es eficiente.

Para que un estimador sea *eficiente* tiene que ser *centrado* y de *varianza mínima*. La varianza mínima se analiza en virtud de la acotación de Cramer-Rao (conocida también como desigualdad de la información).
En el caso continuo:

$$V(\hat{\theta}) \geq CCR = \frac{1}{I(\theta)} = \frac{1}{E\left[\dfrac{\vartheta \ln L(x\,;\,\theta)}{\vartheta\theta}\right]^2} = \frac{1}{n\,E\left[\dfrac{\vartheta \ln f(x\,;\,\theta)}{\vartheta\theta}\right]^2}$$

La información de Fisher $I(\theta)$ es una cota inferior para la varianza de un estimador insesgado del parámetro.

Un estimador es eficiente cuando $V(\hat{\theta}) = CCR = I(\theta)^{-1}$

Cuando se extiende la condición de regularidad a la segunda derivada en un parámetro bidimensional $\theta = (\mu, \sigma)$, se puede utilizar una forma alternativa de la información de Fisher para obtener una nueva desigualdad de Cramer-Rao:

$$V(\hat{\theta}) \geq CCR = \frac{1}{I(\theta)} = \frac{1}{-E\left[\dfrac{\vartheta^2 \ln L(x\,;\,\theta)}{\vartheta\theta^2}\right]^2} = \frac{1}{-n\,E\left[\dfrac{\vartheta^2 \ln f(x)}{\vartheta\theta^2}\right]^2}$$

Un estimador es eficiente cuando $V(\hat{\mu}, \hat{\sigma}) = CCR = I(\mu, \sigma)^{-1}$

La matriz de información de Fisher $I(\mu, \sigma)$, matriz de segundas derivadas (Jacobiano), viene dada por la expresión:

$$I(\mu,\sigma) = -E\begin{bmatrix} \dfrac{\vartheta^2}{\vartheta\mu\,\vartheta\mu}\ln L(\mu,\sigma) & \dfrac{\vartheta^2}{\vartheta\mu\,\vartheta\sigma}\ln L(\mu,\sigma) \\ \dfrac{\vartheta^2}{\vartheta\sigma\,\vartheta\mu}\ln L(\mu,\sigma) & \dfrac{\vartheta^2}{\vartheta\sigma\,\vartheta\sigma}\ln L(\mu,\sigma) \end{bmatrix}$$

🖫 **A partir de una muestra aleatoria simple X de tamaño n, determinar la matriz de información de Fisher de una población con distribución $N(\mu,\sigma)$, respecto a los dos parámetros μ y σ.**

Solución:

Función de densidad poblacional: $f(x)=\dfrac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Función de verosimilitud:

$$L(\mu,\sigma)=L(X;\mu,\sigma)=$$

$$=\left[\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right]\left[\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right]\cdots\left[\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right]=$$

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

Función soporte o Log-verosimilitud:

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

Para determinar la información de Fisher se deriva la expresión anterior respecto de los parámetros μ y σ:

$$\ln L(\mu,\sigma)=-\frac{n}{2}\ln 2\pi-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\vartheta \ln L(\mu,\sigma)}{\vartheta\mu}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

$$\frac{\vartheta \ln L(\mu,\sigma)}{\vartheta\sigma}=-\frac{n}{\sigma}+\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^3}$$

Cálculo de las segundas derivadas $\left[\mu=\overline{x}\ ,\ \hat{\sigma}^2=\dfrac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2\right]$

$$\left.\frac{\vartheta^2 \ln L(X;\ \mu,\sigma)}{\vartheta\mu\,\vartheta\mu}\right|_{(\overline{x},\,\hat{\sigma}^2)}=\frac{\vartheta}{\vartheta\mu}\left.\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)\right)\right|_{(\overline{x},\,\hat{\sigma}^2)}=-\left.\frac{1}{\sigma^2}\right|_{(\overline{x},\,\hat{\sigma}^2)}=-\frac{n}{\hat{\sigma}^2}$$

$$\left.\frac{\vartheta^2 \ln L(X;\ \mu,\sigma)}{\vartheta\mu\,\vartheta\sigma}\right|_{(\overline{x},\,\hat{\sigma}^2)}=\left.\frac{\vartheta^2 \ln L(X;\ \mu,\sigma)}{\vartheta\sigma\,\vartheta\mu}\right|_{(\overline{x},\,\hat{\sigma}^2)}=\frac{\vartheta}{\vartheta\sigma}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)\right)_{(\overline{x},\,\hat{\sigma}^2)}=$$

$$=-\left.\frac{2\sum_{i=1}^{n}(x_i-\mu)}{\sigma^3}\right|_{(\overline{x},\,\hat{\sigma}^2)}=0$$

$$\left.\frac{\vartheta^2 \ln L(X;\ \mu,\sigma)}{\vartheta\sigma\,\vartheta\sigma}\right|_{(\overline{x},\,\hat{\sigma}^2)}=\frac{\vartheta}{\vartheta\sigma}\left(-\frac{n}{\sigma}+\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^3}\right)_{(\overline{x},\,\hat{\sigma}^2)}=$$

$$=\frac{n}{\sigma^2}+\sum_{i=1}^{n}(x_i-\mu)^2\left.\left(\frac{-3\sigma^2}{\sigma^6}\right)\right|_{(\overline{x},\,\hat{\sigma}^2)}=\frac{n}{\sigma^2}-\left.\frac{3\sum_{i=1}^{n}(x_i-\mu)^2}{\sigma^4}\right|_{(\overline{x},\,\hat{\sigma}^2)}=$$

$$=\frac{n}{\hat{\sigma}^2}-\frac{3n\hat{\sigma}^2}{\hat{\sigma}^4}=-\frac{2n}{\hat{\sigma}^2}$$

Matriz de Información de Fisher $I(\mu,\sigma)$**:**

$$I(\mu,\sigma)=-E\begin{bmatrix}\dfrac{\vartheta^2}{\vartheta\mu\vartheta\mu}\ln L(\mu,\sigma) & \dfrac{\vartheta^2}{\vartheta\mu\vartheta\sigma}\ln L(\mu,\sigma)\\[2mm] \dfrac{\vartheta^2}{\vartheta\sigma\vartheta\mu}\ln L(\mu,\sigma) & \dfrac{\vartheta^2}{\vartheta\sigma\vartheta\sigma}\ln L(\mu,\sigma)\end{bmatrix}=$$

$$=-\begin{bmatrix}-\dfrac{n}{\hat{\sigma}^2} & 0\\[2mm] 0 & -\dfrac{2n}{\hat{\sigma}^2}\end{bmatrix}=\begin{bmatrix}\dfrac{n}{\hat{\sigma}^2} & 0\\[2mm] 0 & \dfrac{2n}{\hat{\sigma}^2}\end{bmatrix}$$

Al ser x_i un elemento de una muestra aleatoria simple de una población normal $N(\mu,\sigma)$, se tiene que $E(x_i-\mu)=0$ y $E(x_i-\mu)^2=\sigma^2$

Se puede obtener la distribución de $(\hat{\mu},\hat{\sigma})$

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

Función soporte:

$$\ln f(x)=\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)=\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)+\ln\left(e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)=$$

$$=-\ln\sqrt{2\pi\sigma^2}-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]=-\ln\left(\sigma\sqrt{2\pi}\right)-\frac{1}{2}\left[\frac{(x-\mu)^2}{\sigma^2}\right]=$$

$$=-\ln\sqrt{2\pi}-\ln\sigma-\frac{1}{2}\left[\frac{(x-\mu)^2}{\sigma^2}\right]$$

Para determinar la información de Fisher se deriva la expresión anterior respecto de los parámetros μ y σ:

$$\ln f(x)=-\ln\sqrt{2\pi}-\ln\sigma-\frac{1}{2}\left[\frac{(x-\mu)^2}{\sigma^2}\right]$$

$$\frac{\vartheta}{\vartheta\mu}\ln f(x)=\frac{(x-\mu)}{\sigma^2}$$

$$\frac{\vartheta}{\vartheta\sigma}\ln f(x)=\frac{-1}{\sigma}-\frac{(x-\mu)^2}{2}\left(\frac{-2\sigma}{\sigma^4}\right)=\frac{-1}{\sigma}+\frac{(x-\mu)^2}{\sigma^3}$$

Cálculo de las segundas derivadas:

$$\frac{\vartheta^2}{\vartheta\mu\,\vartheta\mu}\,f(x)=-\frac{1}{\sigma^2}$$

$$\frac{\vartheta^2}{\vartheta\mu\,\vartheta\sigma}\,f(x)=\frac{\vartheta^2}{\vartheta\sigma\,\vartheta\mu}\,f(x)=(x-\mu)\left(\frac{-2\sigma}{\sigma^4}\right)=-\frac{2(x-\mu)}{\sigma^3}$$

$$\frac{\vartheta^2}{\vartheta\sigma\,\vartheta\sigma}\,f(x)=\frac{1}{\sigma^2}+(x-\mu)^2\left(\frac{-3\sigma^2}{\sigma^6}\right)=\frac{1}{\sigma^2}-\frac{3(x-\mu)^2}{\sigma^4}$$

Matriz de Información de Fisher $I(\mu,\sigma)$:

$$I_0(\mu,\sigma)=-E\begin{bmatrix}\dfrac{\vartheta^2}{\vartheta\mu\vartheta\mu}\,f(x) & \dfrac{\vartheta^2}{\vartheta\mu\vartheta\sigma}\,f(x)\\[2mm] \dfrac{\vartheta^2}{\vartheta\sigma\vartheta\mu}\,f(x) & \dfrac{\vartheta^2}{\vartheta\sigma\vartheta\sigma}\,f(x)\end{bmatrix}=$$

$$=-E\begin{bmatrix}-\dfrac{1}{\sigma^2} & -\dfrac{2(x-\mu)}{\sigma^3}\\[2mm] -\dfrac{2(x-\mu)}{\sigma^3} & \dfrac{1}{\sigma^2}-\dfrac{3(x-\mu)^2}{\sigma^4}\end{bmatrix}=E\begin{bmatrix}\dfrac{1}{\sigma^2} & \dfrac{2(x-\mu)}{\sigma^3}\\[2mm] \dfrac{2(x-\mu)}{\sigma^3} & -\dfrac{1}{\sigma^2}+\dfrac{3(x-\mu)^2}{\sigma^4}\end{bmatrix}=$$

$$=\begin{bmatrix}\dfrac{1}{\sigma^2} & 0\\[2mm] 0 & -\dfrac{1}{\sigma^2}+\dfrac{3\sigma^2}{\sigma^4}\end{bmatrix}=\begin{bmatrix}\dfrac{1}{\sigma^2} & 0\\[2mm] 0 & \dfrac{2}{\sigma^2}\end{bmatrix}$$

Como x un elemento de una muestra aleatoria simple de una población normal $N(\mu,\sigma)$, se tiene que $E(x-\mu)=0$ y $E(x-\mu)^2=\sigma^2$

La matriz de información de Fisher: $I_0(\mu,\sigma)=\begin{bmatrix}\dfrac{1}{\sigma^2} & 0\\[2mm] 0 & \dfrac{2}{\sigma^2}\end{bmatrix}$

Atendiendo a la aditividad de esta medida:

$$I(\mu,\sigma)=n\,I_0(\mu,\sigma)=n\begin{bmatrix}\dfrac{1}{\sigma^2} & 0\\[2mm] 0 & \dfrac{2}{\sigma^2}\end{bmatrix}=\begin{bmatrix}\dfrac{n}{\sigma^2} & 0\\[2mm] 0 & \dfrac{2n}{\sigma^2}\end{bmatrix}$$

$$\begin{vmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{vmatrix} = \frac{2}{\sigma^4} \quad \mapsto \quad I_0(\mu,\sigma)^{-1} = \frac{\sigma^4}{2}\begin{bmatrix} \frac{2}{\sigma^2} & 0 \\ 0 & \frac{1}{\sigma^2} \end{bmatrix} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \frac{\sigma^2}{2} \end{bmatrix}$$

Por el Teorema Central del Límite la distribución asintótica de $(\hat{\mu}\,,\hat{\sigma})^t$ es:

$$\sqrt{n}\left(\begin{bmatrix} \hat{\mu} \\ \hat{\sigma} \end{bmatrix} - \begin{bmatrix} \mu \\ \sigma \end{bmatrix}\right) \sim N\left(0,\ I_0(\mu,\sigma)^{-1}\right) = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix},\ \begin{bmatrix} \sigma^2 & 0 \\ 0 & \frac{\sigma^2}{2} \end{bmatrix}\right)$$

Una aplicación directa se encuentra en el Valor en Riesgo o Value at Risk (VaR), medida ampliamente utilizada en una cartera de inversiones de activos financieros.

En una cartera de inversiones, asumiendo mercados normales y que no se produce negociación en la cartera, el VaR se define como un valor límite tal que la probabilidad de que una pérdida a precios de mercados en la cartera sobre un horizonte temporal dado exceda ese valor sea el nivel de probabilidad dado.

Parámetros comunes para el VaR son probabilidades de 1% y 5% y horizontes temporales de un día y de dos semanas, aunque también se utilizan otras combinaciones.

PRÁCTICA: Una empresa que cotiza en Bolsa considera como pérdidas todos los rendimientos inferiores a 3 euros por acción, es decir, los beneficios siguen una distribución $N(\mu-3\,,\sigma)$.

Las pérdidas máximas esperadas para un nivel de significación α son $VaR = \mu - 3 - z_\alpha \sigma$.

La distribución del estimador $\widehat{VaR} = \hat{\mu} - 3 - z_\alpha \hat{\sigma}$

$$\widehat{VaR} = \hat{\mu} - 3 - z_\alpha \hat{\sigma} = \begin{pmatrix} 1 & -z_\alpha \end{pmatrix}\begin{pmatrix} \hat{\mu} \\ \hat{\sigma} \end{pmatrix} - 3$$

La distribución asintótica de $(\hat{\mu}\,,\hat{\sigma})^t$: $\sqrt{n}\,\left(\widehat{VaR} - VaR\right) \sim N\left(0,\ \sigma^2 + z_\alpha^2\,\frac{\sigma^2}{2}\right)$

MÉTODO DE LOS MOMENTOS: Procedimiento consistente en igualar momentos poblacionales respecto al origen α_r a los correspondientes momentos muestrales respecto al origen a_r, formando así tantas ecuaciones como parámetros poblacionales se pretenden estimar:

$$\alpha_1 = E(X) = \mu \Rightarrow \hat{\alpha}_1 = a_1 = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}$$

$$\alpha_2 = E(X^2) \Rightarrow \hat{\alpha}_2 = a_2 = \frac{\sum_{i=1}^{n} x_i^2}{n}$$

....

$$\alpha_r = E(X^r) \Rightarrow \hat{\alpha}_r = a_r = \frac{\sum_{i=1}^{n} x_i^r}{n}$$

Sea una población definida por: $P(X=-1) = \frac{1-\theta}{2}$, $P(X=0) = \frac{\theta+\lambda}{2}$, $P(X=1) = \frac{1-\lambda}{2}$, donde $0<\theta<1$, $0<\lambda<1$. Estimar los parámetros θ y λ por el método de los momentos, estudiando si son insesgados.

Solución:

Puesto que hay que estimar dos parámetros θ y λ hay que calcular los dos primeros momentos.

$$\overbrace{\alpha_1 = \mu = E(X) = \sum_i x_i P(X=x_i)}^{\text{momentos poblacionales}} =$$

$$= (-1)\left(\frac{1-\theta}{2}\right) + (0)\left(\frac{\theta+\lambda}{2}\right) + (1)\left(\frac{1-\lambda}{2}\right) = \frac{\theta-\lambda}{2}$$

$$\alpha_2 = E(X^2) = \sum_i x_i^2 \, P(X=x_i) =$$

$$= (-1)^2\left(\frac{1-\theta}{2}\right) + (0)^2\left(\frac{\theta+\lambda}{2}\right) + (1)^2\left(\frac{1-\lambda}{2}\right) = \frac{2-\theta-\lambda}{2}$$

$$\overbrace{a_1 = \overline{x} = \frac{\sum_i x_i}{n} \qquad a_2 = \frac{\sum_i x_i^2}{n}}^{\text{momentos muestrales}}$$

$$\begin{cases} \alpha_1 = a_1 \Rightarrow \dfrac{\theta-\lambda}{2} = \overline{x} \Rightarrow \theta - \lambda = 2\overline{x} \\ \alpha_2 = a_2 \Rightarrow \dfrac{2-\theta-\lambda}{2} = a_2 \Rightarrow -\theta - \lambda = 2a_2 - 2 \end{cases} \mapsto$$

$$\mapsto \begin{cases} \theta - \lambda = 2\overline{x} \\ -\theta - \lambda = 2a_2 - 2 \end{cases} \mapsto \boxed{\begin{matrix} \hat{\lambda} = 1 - a_2 - \overline{x} \\ \hat{\theta} = 1 - a_2 + \overline{x} \end{matrix}}$$

- **Insesgadez**

$$E(\hat{\theta}) = E(1 - a_2 + \overline{x}) =$$

$$= 1 - E(a_2) + E(\overline{x}) = 1 - \alpha_2 + \mu = 1 - \left(\frac{2-\theta-\lambda}{2}\right) + \left(\frac{\theta-\lambda}{2}\right) = \theta$$

$$E(\hat{\lambda}) = E(1 - a_2 - \overline{x}) = 1 - E(a_2) - E(\overline{x}) =$$

$$= 1 - \alpha_2 - \mu = 1 - \left(\frac{2-\theta-\lambda}{2}\right) - \left(\frac{\theta-\lambda}{2}\right) = \lambda$$

Los estimadores θ y λ son insesgados.

Un estimador $\hat{\theta}$ es suficiente cuando no da lugar a una pérdida de información. Es decir, cuando la información basada en $\hat{\theta}$ es tan buena como la que hiciera uso de toda la muestra.

Para identificar estadísticos suficientes se utiliza el Teorema de Factorización, que establece que dada una muestra aleatoria $(x_1, \cdots, x_n)$ de una población X con función de densidad f_θ (caso continuo) un estadístico $\hat{\theta}$ es suficiente para θ sí y sólo sí:

$$f_\theta(x_1, \cdots, x_n) = g\left[\hat{\theta}(x_1, \cdots, x_n), \theta\right] . h(x_1, \cdots, x_n)$$

Para encontrar un estadístico suficiente $\hat{\theta}$ hay que factorizar la función de verosimilitud de la forma: $L(\theta) = g(\hat{\theta}, \theta) . h(x_1, \cdots, x_n)$

Una muestra aleatoria $(x_1, \cdots, x_n)$ de la población tiene como función de densidad $f_\theta(x) = \begin{cases} \theta x^{\theta-1} & \text{si } x \in (0,1) \\ 0 & \text{en el resto} \end{cases} \quad \theta > 0$

a) Hallar un estadístico suficiente
b) Estimador de máxima verosimilitud de θ
c) Estimador de θ por el método de los momentos

Solución:

a) Para encontrar un estadístico suficiente $\hat{\theta}$ hay que factorizar la función de verosimilitud de la forma: $L(\theta) = g(\hat{\theta}, \theta) . h(x_1, \cdots, x_n)$

Función de verosimilitud:

$$L(\theta) = f_\theta(x_1)\, f_\theta(x_2) \cdots f_\theta(x_n) = (\theta x_1^{\theta-1})\,(\theta x_2^{\theta-1}) \cdots (\theta x_n^{\theta-1}) = \theta^n (x_1 \cdots x_n)^{\theta-1}$$

Por tanto, $\hat{\theta} = \left(x_1, \cdots, x_n\right)$ es un estadístico suficiente.

b) $L(\theta) = \theta^n (x_1, \cdots, x_n)^{\theta-1}$

Función soporte o Log-verosimilitud:

$$\ln L(\theta) = \ln\left[\theta^n (x_1, \cdots, x_n)^{\theta-1}\right] = \ln\theta^n + \ln\prod_{i=1}^{n} x_i^{\theta-1} = \ln\theta^n + \sum_{i=1}^{n} \ln(x_i^{\theta-1})$$

$$\ln L(\theta) = n\ln\theta + [\theta-1]\sum_{i=1}^{n}\ln(x_i) \Rightarrow \frac{\partial \ln L(\theta)}{\partial\theta} = \frac{n}{\theta} + \sum_{i=1}^{n}\ln(x_i) = 0$$

$$\mapsto \hat{\theta} = -\frac{n}{\sum_{i=1}^{n}\ln(x_i)}$$

c) Se plantea la ecuación $E[X] = \overline{x}$

$$\overline{x} = E(X) = \int_0^1 x\, f_\theta(x)\,dx = \int_0^1 x\,\theta x^{\theta-1}\,dx = \int_0^1 \theta x^{\theta}\,dx = \theta\left[\frac{x^{\theta+1}}{\theta+1}\right]_0^1 = \frac{\theta}{\theta+1}$$

$$\overline{x}\,(\theta+1) = \theta \Rightarrow \hat{\theta} = \frac{\overline{x}}{1-\overline{x}}$$

Una muestra aleatoria $(x_1, \cdots, x_n)$ de la población tiene como función de densidad

$$f_\theta(x) = \begin{cases} e^{-x+\theta} & \text{si } x > 0 \\ 0 & \text{en el resto} \end{cases}$$

a) Hallar un estimador por el método de los momentos de θ
b) Estudiar si el estimador encontrado en el apartado anterior es insesgado para estimar el parámetro θ

Solución:

a) Se plantea la ecuación: $E[X] = \overline{x}$

$$\overline{x} = E[X] = \int_\theta^\infty x\, f_\theta(x)\,dx = \overbrace{\int_\theta^\infty x\, e^{-x+\theta}\,dx}^{\text{integración por partes}} = \theta + 1 \Rightarrow \hat{\theta} = \overline{x} - 1$$

$$\int \underbrace{x}_{u}\, \underbrace{e^{-x+\theta}dx}_{dv} = \underbrace{x}_{u}\underbrace{(-e^{-x+\theta})}_{v} - \int \underbrace{-e^{-x+\theta}}_{v}\,\underbrace{dx}_{du} = -x e^{-x+\theta} - e^{-x+\theta} =$$

$$= -(1+x)\,e^{-x+\theta} = -e^{-\theta}\left(\frac{1+x}{e^{x}}\right)$$

$$\int_{\theta}^{\infty} x\,e^{-x+\theta}\,dx = -e^{-\theta}\left(\frac{1+x}{e^{x}}\right)_{\theta}^{\infty} = 1+\theta$$

b) Un estimador es insesgado o centrado cuando su valor probable coincide con el valor del parámetro a estimar. Es decir, $E\left[\hat{\theta}\right]=\theta$

$$E\left[\hat{\theta}\right] = E(\overline{x}-1) = E(\overline{x})-1 = (\theta+1)-1 = \theta$$

Una muestra aleatoria $(x_1, \cdots, x_n)$ de la población tiene como función de densidad $f_{\theta}(x)=\begin{cases}\theta^2 x e^{-\theta x} & \text{si } x>0\\ 0 & \text{en el resto}\end{cases}$

Hallar el estimador de máxima verosimilitud de θ

Solución:

La función de verosimilitud $L(\theta)$

$$L(\theta) = f_{\theta}(x_1)\,f_{\theta}(x_2)\cdots f_{\theta}(x_n) = \left[\theta^2 x_1 e^{-\theta x_1}\right]\left[\theta^2 x_2 e^{-\theta x_2}\right]\cdots\left[\theta^2 x_n e^{-\theta x_n}\right] =$$

$$= \theta^{2n}(x_1\cdots x_n)\,e^{-(\theta x_1+\theta x_2+\cdots+\theta x_n)} = \theta^{2n}(x_1\cdots x_n)\,e^{-\theta\sum_{i=1}^{n} x_i}$$

$$L(\theta) = \theta^{2n}(x_1\cdots x_n)\,e^{-\theta\sum_{i=1}^{n} x_i} \Rightarrow \ln L(\theta) = \left(\ln\theta^{2n}(x_1\cdots x_n)\,e^{-\theta\sum_{i=1}^{n} x_i}\right)$$

$$\ln L(\theta) = (2n)\ln\theta + \ln\prod_{i=1}^{n} x_i - \theta\sum_{i=1}^{n} x_i \Rightarrow$$

$$\Rightarrow \ln L(\theta) = (2n)\ln\theta + \sum_{i=1}^{n}\ln x_i - \theta\sum_{i=1}^{n} x_i$$

$$\frac{\partial \ln L(\theta)}{\partial\theta} = \frac{2n}{\theta} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \hat{\theta} = \frac{2n}{\sum_{i=1}^{n} x_i}$$

El coseno X del ángulo con el que se emiten los electrones en un proceso radioactivo es una variable aleatoria con función de densidad

$$f_\theta(x) = \begin{cases} (1+\theta x)/2 & -1 \le x \le 1 \ , \ -1 \le \theta \le 1 \\ 0 & \text{en el resto} \end{cases}$$

Se considera una muestra aleatoria $(X_1, \cdots, X_n)$ de la variable aleatoria

a) Obtener el estimador θ por el método de los momentos

b) Calcular la varianza de este estimador y demostrar que es consistente

Solución:

a) Se plantea la ecuación $E[X] = \overline{x}$

$$\overline{x} = E[X] = \int_{-1}^{1} x \left(\frac{1+\theta x}{2}\right) dx = \left[\frac{x^2}{2} + \frac{\theta x^3}{6}\right]_{-1}^{1} = \frac{\theta}{3} \quad \Rightarrow \quad \hat{\theta} = 3\overline{x}$$

b) $$V(\hat{\theta}) = V(3\overline{x}) = 9V(\overline{x}) = 9\,\frac{V(X)}{n} = \frac{9}{n} V(X)$$

$$V(X) = E(X^2) - [E(X)]^2 = \int_{-1}^{1} x^2 \left(\frac{1+\theta x}{2}\right) dx - \left[\frac{\theta}{3}\right]^2 =$$

$$= \left[\frac{x^3}{6} + \frac{\theta x^4}{8}\right]_{-1}^{1} - \left[\frac{\theta}{3}\right]^2 = \frac{3-\theta^2}{9}$$

de donde, $$V(\hat{\theta}) = \frac{9}{n} V(X) = \frac{9}{n}\left[\frac{3-\theta^2}{9}\right] = \frac{3-\theta^2}{n}$$

Consistencia o robustez del estimador $\hat{\theta}$:

$$\lim_{n\to\infty} E(\hat{\theta}) = \lim_{n\to\infty} E(3\overline{x}) = \lim_{n\to\infty} 3E(\overline{x}) = 3E(X) = 3\,\frac{\theta}{3} = \theta$$

$$\lim_{n\to\infty} V(\hat{\theta}) = \lim_{n\to\infty} V(3\overline{x}) = \lim_{n\to\infty} \frac{3-\theta^2}{n} = 0$$

Queda probado que el estimador $\hat{\theta}$ es consistente

En un estacionamiento el número de veces que se abre la barrera en un intervalo de 10 minutos, para que pasen vehículos en un sector de seguridad, se considera una variable aleatoria con distribución de Poisson de parámetro λ desconocido.

a) En una muestra aleatoria de 8 intervalos de 10 minutos cada uno, elegidos de forma independiente, se registra para cada intervalo el valor que toma la variable en estudio.

3	5	8	7	4	5	6	2

Encontrar la estimación máximo verosímil de λ

b) Sea $(x_1, \cdots, x_n)$ una muestra aleatoria de tamaño n que sigue una distribución de Poisson. Si $\hat{\lambda}_1 = \sum_{i=1}^{n} \frac{x_i}{n}$, $\hat{\lambda}_2 = \frac{x_1 + 3x_n}{4}$ son estimadores.

Determinar el mejor estimador del parámetro λ

Solución:

a) X = "número de veces que se abre la barrera en un intervalo de 10 minutos", $X \sim P(\lambda)$

$$P(X,\lambda) = \begin{cases} \dfrac{\lambda^x}{x!} e^{-\lambda} & x = 0, 1, 2, \ldots \\ 0 & \text{otro caso} \end{cases}$$

Función de verosimilitud:

$$L(\lambda) = L(X,\lambda) = \prod_{i=1}^{n} P(x_i,\lambda) = \frac{\lambda^{x_1}}{x_1!} e^{-\lambda} \cdots \frac{\lambda^{x_n}}{x_n!} e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-n\lambda}$$

$$\ln L(X,\lambda) = \ln\left(\frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} e^{-n\lambda} \right) = \ln(\lambda^{\sum_{i=1}^{n} x_i}) - \ln\left(\prod_{i=1}^{n} x_i! \right) + \ln(e^{-n\lambda}) =$$

$$= \sum_{i=1}^{n} x_i \ln\lambda - \sum_{i=1}^{n} \ln(x_i!) - n\lambda$$

$$\ln L(X,\lambda) = \sum_{i=1}^{n} x_i \ln\lambda - \sum_{i\in=1}^{n} \ln(x_i!) - n\lambda$$

Para obtener el estimador de máxima verosimilitud $EMV(\hat{\lambda})$, se deriva la expresión anterior respecto del parámetro para obtener, sustituyendo λ por $\hat{\lambda}$

$$\left.\frac{\vartheta \ln L(X,\lambda)}{\vartheta\lambda}\right|_{\lambda=\hat{\lambda}} = 0 = \left.\sum_{i=1}^{n} x_i \frac{1}{\lambda} - n\right|_{\lambda=\hat{\lambda}} = 0 \;\Rightarrow\; \hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}$$

El estimador de máxima verosimilitud viene dado por la media muestral $EMV(\hat{\lambda}) = \overline{x}$

Utilizando la muestra aleatoria de ocho intervalos de 10 minutos, se obtiene el estimador máximo verosímil:

$$\hat{\lambda} = \frac{3+5+8+7+4+5+6+2}{8} = 5$$

En consecuencia, en una muestra aleatoria de ocho intervalos de 10 minutos cada uno, elegidos de forma independiente, la estimación máxima verosímil corresponde a que la barrera se abre 5 veces.

b) Se analizan si los estimadores son o no insesgados, esto es, si la esperanza del estimador coincide con el parámetro a estimar.

$$E(\hat{\lambda}_1) = E\left(\sum_{i=1}^{n} \frac{x_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{n\lambda}{n} = \lambda$$

$$E(\hat{\lambda}_2) = E\left(\frac{x_1 + 3x_n}{4}\right) = \frac{1}{4}\left[E(x_1) + 3E(x_2)\right] = \frac{\lambda + 3\lambda}{4} = \lambda$$

Ambos estimadores son insesgados.

Varianza de cada estimador:

$$V(\hat{\lambda}_1) = V\left(\sum_{i=1}^{n} \frac{x_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) = \frac{n\lambda}{n^2} = \frac{\lambda}{n}$$

$$V(\hat{\lambda}_2) = V\left(\frac{x_1 + 3x_n}{4}\right) = \frac{1}{16}\left[V(x_1) + 9V(x_2)\right] = \frac{\lambda + 9\lambda}{16} = \frac{10\lambda}{16} = \frac{5\lambda}{8}$$

La efectividad de los estimadores depende del tamaño de la muestra:

- **Si la muestra es igual a 1 $(n=1)$ el estimador más eficiente es $\hat{\lambda}_2$**
- **Si la muestra es mayor que 1 $(n>1)$ el estimador más eficiente es $\hat{\lambda}_1$**

Sea $(x_1, \cdots, x_n)$ una muestra aleatoria de tamaño n, distribuida según $f(x,\theta)$ con θ desconocido, donde X representa el tiempo máximo necesario para determinar un proceso en segundos:

$$f(X,\theta) = \begin{cases} (\theta+1)x^{\theta} & 0 \le x \le 1\,;\ \theta > -1 \\ 0 & \text{otro caso} \end{cases}$$

a) Determinar el estimador máximo verosímil de θ
b) Determinar la estimación máximo verosímil de θ en una muestra aleatoria simple constituida por los datos:

0,7 0,9 0,6 0,8 0,9 0,7 0,9 0,8

Estimar la probabilidad del tiempo máximo necesario para terminar un proceso, que no exceda de 0,25 segundos ni supere los 0,75 segundos.

c) Determinar el estimador máximo verosímil de: (a) $\theta+1$ (b) $\dfrac{2\theta+1}{\theta-1}$

Solución:

a) Función de verosimilitud $L(\theta) = L(X,\theta) = \prod_{i=0}^{n} f(x_i,\theta)$

$$L(x_1, x_2, \cdots, x_n;\theta) = \left[(\theta+1)x_1^{\theta}\right]\left[(\theta+1)x_2^{\theta}\right]\cdots\left[(\theta+1)x_n^{\theta}\right] = (\theta+1)^n \prod_{i=0}^{n} x_i^{\theta}$$

Función soporte o Log-verosimilitud:

$$\ln L(x_1, x_2, \cdots, x_n;\theta) = \ln\left((\theta+1)^n \prod_{i=0}^{n} x_i^{\theta}\right) = \ln(\theta+1)^n + \ln\left(\prod_{i=0}^{n} x_i^{\theta}\right)$$

$$\ln L(x_1, x_2, \cdots, x_n;\theta) = n\ln(\theta+1) + \theta\sum_{i=0}^{n} \ln x_i$$

Para obtener el estimador de máxima verosimilitud $\text{EMV}(\hat{\theta})$, se deriva la expresión anterior respecto del parámetro para obtener, sustituyendo θ por $\hat{\theta}$

$$\left.\frac{\vartheta \ln L(x_1, x_2, \cdots\cdots, x_n; \theta)}{\vartheta\theta}\right|_{\theta=\hat{\theta}} = \frac{n}{\hat{\theta}+1} + \sum_{i=0}^{n} \ln x_i = 0 \quad \mapsto \quad \frac{n}{\hat{\theta}+1} = -\sum_{i=0}^{n} \ln x_i$$

$$\hat{\theta} = \text{EMV}(\theta) = -\frac{n}{\sum_{i=0}^{n} \ln x_i} - 1$$

b) El estimador máximo verosímil de θ con los datos de la muestra:

$$\hat{\theta} = -\frac{8}{\ln 0{,}7 + \ln 0{,}9 + \ln 0{,}6 + \ln 0{,}8 + \ln 0{,}9 + \ln 0{,}7 + \ln 0{,}9 + \ln 0{,}8} - 1 =$$

$$= 4{,}027 - 1 = 3{,}027$$

El estimador máximo verosímil $\text{EMV}(\theta) = \hat{\theta} = 3{,}027$

De otra parte,

$$P(0{,}25 \le X \le 0{,}75) = \int_{0{,}25}^{0{,}75} f(X, \hat{\theta})\,dx = \int_{0{,}25}^{0{,}75} (\hat{\theta}+1)\, x^{\hat{\theta}}\, dx = (\hat{\theta}+1) \left.\frac{x^{\hat{\theta}+1}}{(\hat{\theta}+1)}\right|_{0{,}25}^{0{,}75} =$$

$$= \left. x^{\hat{\theta}+1} \right|_{0{,}25}^{0{,}75} = 0{,}75^{4{,}027} - 0{,}25^{4{,}047} = 0{,}31$$

La probabilidad del tiempo máximo necesario para terminar el proceso (entre 0,25 y 0,75 segundos) es 0,31

c) Estimador máximo verosímil de $\theta + 1$ y $\dfrac{2\theta+1}{\theta-1}$

$$\text{EMV}[\theta+1] = \text{EMV}[\theta] + 1 = \hat{\theta} + 1 = 3{,}027 + 1 = 4{,}027$$

$$\text{EMV}\left(\frac{2\theta+1}{\theta-1}\right) = \frac{2\,\text{EMV}(\theta)+1}{\text{EMV}(\theta)-1} = \frac{2\hat{\theta}+1}{\hat{\theta}-1} = \frac{2 \times 3{,}027 + 1}{3{,}027 - 1} = 2{,}493$$

En la distribución $B(m, p)$ se considera como estimador del parámetro p el estadístico $\hat{p} = \dfrac{\overline{x}}{m}$, siendo $\overline{x}$ la media muestral en muestras aleatorias simples de tamaño n. Hallar la eficiencia del estimador.

Solución:

Un estimador es eficiente cuando su varianza coincide con la Cota de Cramer-Rao $V(\hat{p}) = CCR$

$$CCR = \frac{1}{n.E\left[\dfrac{\vartheta \ln P(x\,;\,p)}{\vartheta\theta}\right]^2} \qquad I(\theta) = n.E\left[\dfrac{\vartheta \ln P(x\,;\,p)}{\vartheta\theta}\right]^2$$ Información de Fisher

Insesgadez: $E(\hat{p}) = E\left(\dfrac{\overline{x}}{m}\right) = \dfrac{mp}{m} = p$

El estimador es insesgado o centrado

Para determinar la CCR se considera la función de cuantía de la distribución binomial: $P(x, p) = \begin{pmatrix} m \\ x \end{pmatrix} p^x\,(1-p)^{m-x}$

Función soporte o Log-verosimilitud:

$$\ln P(x, p) = \ln \begin{pmatrix} m \\ x \end{pmatrix} + x\ln p + (m-x)\ln(1-p)$$

$$\frac{\vartheta \ln P(x, p)}{\vartheta p} = \frac{x}{p} - \frac{(m-x)}{1-p} = \frac{(1-p)x - p(m-x)}{p(1-p)} = \frac{x - mp}{p(1-p)}$$

$$E\left[\frac{\vartheta \ln P(x, p)}{\vartheta p}\right]^2 = E\left[\frac{x-mp}{p(1-p)}\right]^2 = \frac{E(x-mp)^2}{p^2(1-p)^2}$$

En una distribución binomial: $E(X) = mp\,,\ V(X) = E(X-mp)^2 = mp(1-p)$

con lo que,

$$E\left[\frac{\vartheta \ln P(x,p)}{\vartheta p}\right]^2 = E\left[\frac{x-mp}{p(1-p)}\right]^2 = \frac{E(x-mp)^2}{p^2(1-p)^2} = \frac{m\cancel{p}\,\cancel{(1-p)}}{p^{\cancel{2}}(1-p)^{\cancel{2}}} = \frac{m}{p(1-p)}$$

$$CCR = \frac{1}{n.E\left[\frac{\vartheta \ln P(x\,;\,p)}{\vartheta\theta}\right]^2} = \frac{1}{n\,\frac{m}{p(1-p)}} = \frac{p(1-p)}{n\,m}$$

De otra parte, la varianza del estimador:

$$V(\hat{p}) = V\left(\frac{\overline{x}}{m}\right) = \frac{1}{m^2}V(\overline{x}) = \frac{1}{m^2}\frac{mp(1-p)}{n} = \frac{p(1-p)}{mn}$$

Como la varianza del estimador coincide con la cota de Cramer-Rao el estadístico $\hat{p} = \dfrac{\overline{x}}{m}$ es eficiente.

Una urna contiene bolas blancas y negras. Sea p la probabilidad de extraer una bola blanca cuando se realiza una extracción al azar. Asociado a este experimento aleatorio se encuentra la variable aleatoria X que puede tomar los valores:

X = 1 si la bola extraída es blanca
X = 0 si la bola extraída es negra

Se selecciona una muestra aleatoria con reemplazamiento de tamaño 3 (x_1, x_2, x_3), siendo x_i la variable aleatoria a la extracción i-ésima, y se supone que ha resultado la siguiente relación (B, N, B). Como el parámetro p es desconocido se pretende saber, entre los valores, $p = 0{,}65$ y $p = 0{,}73$ qué valor hace más probable la aparición de dicha extracción.

Solución:

La distribución de probabilidad será una B(1; p): $P(X = x) = p^x (1-p)^{1-x}$

Si la muestra (B, N, B) es independiente, siendo $P(B) = p \qquad P(N) = 1 - p$

$$P(B, N, B) = P(B \cap N \cap B) = P(B).P(N).P(B) = p.(1-p).p = p^2.(1-p)$$

$$\text{entonces} \begin{cases} p = 0{,}65: \quad P(B,N,B) = 0{,}65^2 . 0{,}35 = 0{,}1479 \\ \\ p = 0{,}73: \quad P(B,N,B) = 0{,}73^2 . 0{,}27 = 0{,}1439 \end{cases}$$

Resulta más probable (p = 0,65), siendo más verosímil.
Se considera la m.a.s. (x_1, x_2, x_3), siendo las variables aleatorias x_i independientes, tomando los valores 0, 1, con distribución B(1, p), la distribución de probabilidad asociada será:

$$\left.\begin{array}{l} P(x_1, p) = P(X = x_1) = p^{x_1} (1-p)^{1-x_1} \\ P(x_2, p) = P(X = x_2) = p^{x_2} (1-p)^{1-x_2} \\ P(x_3, p) = P(X = x_3) = p^{x_3} (1-p)^{1-x_3} \end{array}\right\} x_i = 1, 0 \ \text{sea bola blanca o negra}$$

Función de verosimilitud

$$L(p) = \prod_{i=1}^{3} P(x_i, p) = p^{x_1} (1-p)^{1-x_1} . p^{x_2} (1-p)^{1-x_2} . p^{x_3} (1-p)^{1-x_3} =$$

$$= p^{x_1+x_2+x_3} (1-p)^{3-(x_1+x_2+x_3)}$$

En la muestra (B, N, B) el valor que toma la función de verosimilitud será:

$$L(p) = p^{1+0+1} (1-p)^{3-(1+0+1)} = p^2 . (1-p)$$


INVESTIGACIÓN
en Cómic
Análisis
Multivariante
Análisis
Factorial
Componentes
Principales (ACP)
Análisis
Cluster
SPSS