Tarea 1 INF 396

Rodrigo Cayazaya Marín

Rol: 201773538-4

1) Considere N el número de personas en Valparaíso, N es conocido. Hay pN personas en Valparaíso que apoyan al candidato A y (1−p)N persona que apoyan al candidato B, con $p \in [0, 1]$.
   El día de la votación cada persona decide si ir a votar o no con probabilidad 0.5 de votar. Sea N_A el número de personas que votan por el candidato A y N_B el número de personas que votan por el candidato B.

   1. Proponga una distribución de probabilidad para N_B y N_A.
   Debido a que nuestra variable aleatoria es continua, ya representa el número de personas que van a votar por N_A o por N_B, tenemos 3 opciones de distribución: Bernoulli, Binomial y Poisson.

   NA ~ B(pN,0.5)
   NB ~ B((1-p)N,0.5)

   2. ¿Cuál es su esperanza?
   $E[N_A] = pN$
   $E[N_B] = (1-p)N$

2) Estudie y explique los siguientes conceptos (conceptual y matemáticamente):

1. **Estimador insesgado**: Es aquel estimador cuyo sesgo es cero, es decir, cuya esperanza matemática coincide con el valor del parámetro que se quiere estimar. Matemáticamente se puede denotar como:

$$E[\hat{\theta}] = \theta$$

Un ejemplo sería el estimador media:

$$\bar{X} = \text{MEDIA MUESTRAL} = \frac{\sum_{i=1}^{m} X_i}{m}$$

$$E[\bar{X}] = \frac{1}{m} \sum_{i=1}^{m} E[X_i] = \frac{1}{m} \sum_{i=1}^{m} \mu = \frac{m \cdot \mu}{m} = \mu$$

$$E[\bar{X}] = \mu$$

Así podemos denotar que la esperanza de la media es la media.

2. **Estimador consistente en error cuadrático medio**: Es una herramienta que permite comprobar si un estimador es consistente. Esto significa que el sesgo del estimador se aproxima a cero cuando el tamaño de la muestra tiende a infinito. Esto cumple con la siguiente ecuación matemática:

$$lim_{n \to \infty} E[(Estimador - Valor\ verdadero\ del\ parámetro)^2] = 0$$

Lo que significa que la esperanza del error de estimación (estimador menos el valor verdadero del parámetro) debe ser cero o cerca de cero para que se considere como un estimador consistente en error cuadrático medio.

3. **Estimador eficiente**: Un estimador es eficiente cuando su varianza es menor que otro estimador.

$$\text{VAR}(\hat{\theta}_1) < \text{VAR}(\hat{\theta}_2)$$

Esta eficiencia mide la confianza con la que el estadístico obtenido en la muestra aproxime al parámetro poblacional.

4. **Estimador suficiente**: Un estimador se considera suficiente para un parámetro θ si es capaz de resumir toda la información de la muestra. Esto se puede verificar utilizando el teorema de factorización de Fisher-Neyman, el cual dice que un estimador es suficiente si y solo si la función de densidad de la muestra se puede escribir como:

$$f(x1, \ldots, xn) = h(x1, \ldots, xn) \times g(T, \theta)$$

Siendo:
$f(x1,\ldots,xn)$ la función de densidad de la muestra sobre la variable aleatoria X.
$h(x1,\ldots,xn)$ una función no negativa para los valores que conforman la muestra.
$g(T,\theta)$ una función que dependa del estadístico y del parámetro a calcular.

En otras palabras, si es posible escribir la función de densidad de la muestra de aquella forma, entonces se cumple el teorema de Fisher-Neyman, por lo que se consideraría un estimador suficiente para el parámetro θ.

3) Considere la siguiente función de densidad:

$$f(x, \alpha, \beta) = \frac{x^{\alpha} e^{-x/\beta}}{\beta^{\alpha+1}\Gamma(\alpha+1)} I_{x>0}(x),$$

donde $\alpha > 0, \beta > 0$.

1. Encuentre el estimador de máxima verosimilitud para β, asumiendo α conocido.

$$Ln(\ell(\theta)) = Ln\left(\prod_{i}^{n} f(x,\alpha,\beta)\right) = \sum_{i}^{n} Ln(f(x,\alpha,\beta))$$

$$= \sum_{i}^{n} Ln\left(\frac{x_i^{\alpha} e^{-x_i/\beta}}{\beta^{\alpha+1}\Gamma(\alpha+1)}\right)$$

$$= \sum_{i}^{n} \left(Ln\left(x_i^{\alpha} e^{-x_i/\beta}\right) - Ln\left(\beta^{\alpha+1}\Gamma(\alpha+1)\right)\right)$$

$$= \sum_{i}^{n} \left(Ln\left(x_i^{\alpha}\right) - Ln\left(e^{x_i/\beta}\right) - Ln\left(\beta^{\alpha+1}\right) - Ln\left(\Gamma(\alpha+1)\right)\right)$$

$$= \alpha \sum_{i}^{n} Ln(x_i) - \frac{1}{\beta}\sum_{i}^{n} x_i - n(\alpha+1)Ln(\beta) - n Ln(\Gamma(\alpha+1))$$

$$\frac{\partial}{\partial \beta} Ln(\ell(\theta)) = \frac{1}{\beta^2}\sum_{i}^{n} x_i - \frac{n(\alpha+1)}{\beta} = 0$$

$$\frac{1}{\beta^2}\sum_{i}^{n} x_i = \frac{n(\alpha+1)}{\beta} \iff \beta = \frac{1}{n(\alpha+1)}\sum_{i}^{n} x_i$$

$$\beta = \frac{\bar{X}}{(\alpha+1)}$$

2. ¿Es el estimador insesgado?
Debido a que el estimador coincide en la media (dividido en un número conocido) y sabemos del ejercicio 2 que la media es un estimador insesgado, ya que coincide con el promedio. Entonces podemos replicar el procedimiento:

$$\beta = \frac{1}{n(\alpha+1)} \sum_{i}^{n} x_i$$

$$E[\hat{\beta}] = \frac{1}{(\alpha+1)n} \sum_{i=1}^{n} E[x_i] = \frac{1}{(\alpha+1)n} \sum_{i=1}^{n} \bar{x} = \frac{n \cdot \bar{x}}{(\alpha+1)\,n} = \frac{\bar{x}}{\alpha+1}$$

$$E[\hat{\beta}] = \frac{\bar{x}}{\alpha+1} = \beta$$

Así se concluye que **sí** es un estimador insesgado.

3. ¿Es el estimador consistente en error cuadrático medio?

Para comprobar que el estimador es consistente en error medio, se debe cumplir:

$$lim_{n\to\infty} E[(Estimador - Valor\ verdadero\ del\ parámetro)^2] = 0$$

$$\lim_{n\to\infty} E[(\hat{\beta} - \beta)^2] = \lim_{n\to\infty} E[\hat{\beta}^2 - 2\hat{\beta}\beta + \beta^2]$$

$$= \lim_{n\to\infty} E[\hat{\beta}^2] - E[2\hat{\beta}\beta] + E[\beta^2]$$

$$= \lim_{n\to\infty} \beta^2 - 2\beta\beta + \beta^2 = 0$$

Debido a que se cumplen los requerimientos, entonces **sí** es ECM.

4. ¿Es el estimador eficiente?

Una forma de comprobar su eficiencia es a través de su varianza, sabemos que:

$$\mathrm{ECM}[\hat{\theta}] = \mathbb{E}[(\hat{\theta} - \theta)^2] = \mathrm{var}[\hat{\theta}] + \mathrm{sesgo}[\hat{\theta}]^2 \, .$$

Debido a que el sesgo es 0 y el ECM es 0, entonces su varianza es 0. Por lo tanto, **sí** es eficiente.