Clase 8: Inferencia Estadística

Justo Andrés Manrique Urbina

10 de octubre de 2019

1. Métodos de Estimación

1.1. Método de momentos

$$m_k = E(X^k) : \text{k-ésimo momento poblacional.}$$

$$M_k = \bar{X^K} = \frac{\sum_{j=1}^{n} X_j}{n} : \text{k-ésimo momento muestral.}$$

Asumiendo $(X_1, X_2, \ldots, X_n)$ como una muestra aleatoria de X.

1.1.1. Ejemplo

Sea $X \sim P(\lambda)$

$$m_1 = E(X) = \lambda.$$

$$M_1 = \bar{X}.$$

El estimador de λ, por el método de los momentos, es la solución del sistema de ecuaciones (con incógnita λ):

$$m_1 = M_1.$$

Esto equivale a decir que:

$$\lambda = \bar{X}.$$

$$\hat{\lambda} = \bar{X}.$$

1.1.2. Ejemplo

Sea $X \sim N(\mu, \sigma^2)$

$$m_1 = E(X) = \mu.$$

$$m_2 = E(X^2) = \sigma^2 + \mu^2.$$

$$M_1 = \bar{X}.$$

$$M_2 = \bar{X^2}.$$

Los estimadores de μ y σ^2, por el método de momentos, corresponden a la solución del sistema de ecuaciones (con incógnita μ y σ^2) es:

$$m_1 = M_1.$$

$$\hat{\mu} = \bar{x}.$$

$$m_2 = M_2.$$

$$\sigma^2 + \mu^2 = \bar{X^2}.$$

$$\hat{\sigma^2} = \bar{X^2} - \bar{X}^2.$$

1.1.3. Ejemplo

Sea $X \sim N(0, \sigma^2)$

$$m_1 = E(X) = 0 : \text{no depende } \sigma^2.$$

Pasamos a $m_2 \to$ se descarta $m_1 = M_1$

$$m_2 = E(X^2) = \sigma^2.$$

Resolvemos la ecuación:

$$m_2 = M_2.$$

$$\to \sigma^2 = \bar{X}^2 \to \hat{\sigma^2} = \bar{X}^2.$$

2. Método de los mínimos cuadrados

Sea

$$Y = g(X, \theta) + \varepsilon.$$

$g(X, \theta)$ no es variable aleatoria, ε sí. Una muestra conjunta de X e Y es de la forma $(Y_1, X_1), \dots, (Y_n, X_n)$ dónde $Y_i = g(X_{i,\theta}) + \varepsilon_i$.

La forma de $g(\cdot)$ es conocida y th es un parámetro desconocido. Según este método el estimador de θ es el valor de θ que minimiza la función

$$Q(\theta) = \sum_{j=1}^{n} (Y_i - g(X_i, \theta))^2.$$

Observación: Sea $A = (Y_1, Y_2, \dots, Y_n)$ un vector de variables aleatorias de Y, y $B = (g(X_1), g(X_2), \dots, g(X_n))$ un vector de variables aleatorias de X. Entonces:

$$Q(\theta) = d^2(A, B).$$

El estimado de θ minimiza la distancia $d(A, B)$.

2.1. Regresión lineal simple sin intercepto

$$Y = \theta X + \varepsilon.$$

$$Y_i = \theta x_i + \varepsilon_i; i =, 1, \ldots, n.$$

$$Q(\theta) = \sum_{j=1}^{n} (Y_i - \theta X_i)^2.$$

$$\frac{\partial}{\partial \theta} Q = \sum_{j=1}^{n} -2(Y_j - \theta X_j) X_j.$$

$$\frac{\partial}{\partial \theta} = 0 \to \theta = \frac{\sum X_j Y_j}{\sum X_j^2}.$$

$$\frac{\partial^2}{\partial \theta^2} Q = \sum_{j=1}^{n} X_j^2 > 0.$$

Por lo tanto $Q(\cdot)$ es mínima. Y el estimador es:

$$\hat{\theta} = \frac{\sum_{j=1}^{n} X_j Y_j}{\sum_{j=1}^{n} X_j^2}.$$

3. Propiedad de Gauss-Markov

Si $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ son independientes, con $E(\varepsilon_j) = 0$ y $V(\varepsilon_j) = \sigma^2, j = 1, \ldots, n$. Los estimadores de mínimos cuadrados son los mejores estimadores insesgados que sean funciones lineales de $(Y_1, Y_2, \ldots, Y_n)$.

Ejercicio: Sea $\hat{Y}_j = \hat{\theta}_1 + \hat{\theta}_2 X_j; j = 1, \ldots, n$. Hallar $E(Y_j)$ y $V(Y_j)$.

Sugerencia:

$$E(Y_j) = E(\hat{\theta}_1) + X_j E(\hat{\theta}_2).$$

$$V(\hat{\theta}_1 + \hat{\theta}_2 X_j) = V(\hat{\theta}_1) + X_j^2 V(\hat{\theta}_2) + 2X_j Cov(\hat{\theta}_1, \hat{\theta}_2).$$

$$Cov(\hat{\theta}_1, \hat{\theta}_2) = Cov(\sum_{1}^{n} a_j Y_j, \sum_{1}^{n} b_j Y_j).$$

$$Cov(\hat{\theta}_1, \hat{\theta}_2) = a_1 b_1 V(Y_1) + \ldots + a_n b_n V(Y_n) = \sigma^2 \sum_{j=1}^{n} a_j b_j.$$

3.1. Método de Máxima Verosimilitud

Definition 1. *Dada la muestra registrada (u observada)* $(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$*, y* l_n *como función de verosimilitud asociada, se denota por* $L(\theta)$*, y se define como:*

$$L(\theta) = f_{(X_1,X_2,\ldots,X_n)}((X_1, X_2, \ldots, X_n)).$$

Ejemplo: Sea $X \sim P(\lambda)$, dada la muestra registrada:

$$X_1 = 0; X_2 = 1; X_3 = 0; X_4 = 2.$$

$$L(\theta) = f_{(X_1,X_2,\ldots,X_4)}(0; 1; 0; 2).$$

$$= f_{X_1}(0)f_{X_2}(1)f_{X_3}(0)f_{X_4}(2).$$

$$L(\lambda) = \frac{e^{-4\lambda}\lambda^3}{2}, \lambda > 0.$$

Es decir, $L(\theta)$ es la probabilidad de obtener justamente la muestra registrada. En general (caso continuo), $L(\theta)$ mide cuán verosímil resulta cada valor de θ. La estimación de θ, por máxima verosimilitud, es el valor de $\theta \in \Theta$ que maximiza $L(\theta)$.

En general, la estimación de θ es $g(x_1, \ldots, x_n)$. El estimado de θ se define como $\hat{\theta} = g(x_1, \ldots, x_n)$.

Propiedad: Si existe una estadística suficiente para θ, entonces el estimador de máxima verosimilitud es una función de ella.

Explicación:

$$L(\theta) = f_{X_1,\ldots,X_n}(X_1, \ldots, X_n) = h(X_1, \ldots, X_n)l(T(X_1, \ldots, X_n), \theta).$$

Maximizar $L(\theta)$ es igual a maximizar $l(T(X_1, \ldots, X_n), \theta)$.

Propiedad: Si un estimador es insesgado, entonces su varianza es:

$$V(\hat{\theta}) \leq \frac{1}{nI(\theta)}.$$

Para máxima verosimilitud, debemos conocer la distribución de la muestra.