

“Introducción a la Estadística, Probabilidad e Inferencia”

Maestría en Estadística Aplicada
Facultad de Ciencias Económicas y Estadística
UNR

Introducción

En la primera parte de esta unidad consideramos algunos criterios a partir de los cuales podemos evaluar estimadores. Esto permite determinar si un estimador que se propone para estimar un parámetro desconocido es insesgado y consistente, conocer su variancia y compararla con la mínima variancia posible o con la de otro estimador.

Lo ideal sería poder encontrar estimadores insesgados, consistentes y con mínima variancia para todo parámetro. Desafortunadamente no hay garantía que de existan estimadores con tales propiedades para todos los parámetros de interés. Pero si existen algunos métodos que derivan estimadores con propiedades deseables para tamaños de muestra grandes y que con frecuencia pueden ser modificados para que resulten insesgados y de mínima variancia.

En esta clase veremos algunos de estos métodos. Específicamente los métodos de: Máxima Verosimilitud; Momentos; Mínimos Cuadrados.

Función de densidad conjunta de la muestra

Primero recordemos ... Sea X una variable aleatoria con una distribución de probabilidad que depende de un parámetro desconocido θ , $f(x, \theta)$. Sea X_1, X_2, \dots, X_n una muestra aleatoria de X . La distribución conjunta de la muestra está dada por:

$$f(x_1, x_2, \dots, x_n, \theta) = f(x_1, \theta)f(x_2, \theta) \dots f(x_n, \theta) = \prod_{i=1}^n f(x_i, \theta)$$

Esta distribución da la verosimilitud relativa de que las variables aleatorias tomen un valor particular. Si se conoce θ , el conjunto de valores particulares de las variables aleatorias más verosímiles son aquellos en los que $f(x_1, x_2, \dots, x_n, \theta)$ es máximo. Si θ es desconocido y queremos estimarlo a partir de una muestra aleatoria de valores, podríamos intentar determinar de que densidad es más verosímil que proceda este conjunto particular de valores. En otras palabras, podríamos hallar el valor de θ que hace máxima la función de $f(x_1, x_2, \dots, x_n, \theta)$.

Método de Máxima Verosimilitud

Consideremos un ejemplo. Sea X una variable aleatoria con distribución Bernoulli. Se selecciona una muestra de tamaño $n = 10$. Supongamos que entre las 10 observaciones, se observaron 6 éxitos. De que valor de θ es más probable que esta muestra provenga? La verosimilitud de observar exactamente 6 éxitos en una muestra de tamaño 10 es:

$$f(x, \theta) = \theta^6(1 - \theta)^{10-6} \quad 0 < \theta < 1$$

Esta función, que depende solamente de θ , es llamada precisamente *función de verosimilitud*. Calculemos la verosimilitud para distintos valores θ :

Los valores de $f(6, \theta)$ crecen hasta llegar al máximo en $\theta = 0.6$ y luego decrecen. Es decir, entre estos 9 valores, razonablemente se podría elegir a 0,6 como estimación de θ ya que este es el valor que maximiza (entre los listados!) la probabilidad de obtener los 6 éxitos observados. Este es el estimador máximo verosímil de θ

θ	$f(6, \theta)$
0.1	0,000006656
0.2	0,000026200
0.3	0, 000175000
0.4	0,000531000
0.5	0,000976000
0.6	0,003320000
0.7	0,003010000
0.8	0,000419000
0.9	0,000053000

Método de Máxima Verosimilitud

Formalmente...

Sean X_1, X_2, \dots, X_n variables aleatorias independientes e idénticamente distribuidas con función de probabilidad $f(x, \theta)$ con $\theta \in \Omega$ y sean $\mathbf{x} = (x_1, x_2, \dots, x_n)$ los valores observados de X_1, X_2, \dots, X_n para una muestra particular. La función de verosimilitud, $L(\theta/\mathbf{x})$, está dada por:

$$L(\theta/\mathbf{x}) = \prod_{i=1}^n f(x_i, \theta)$$

El valor de θ que maximiza $L(\theta/\mathbf{x})$ es llamado **estimación máximo verosímil** de θ y lo simbolizaremos como $\hat{\theta}_{MV}$:

$$L(\hat{\theta}_{MV}/\mathbf{x}) = \max_{\theta \in \Omega} L(\theta/\mathbf{x})$$

Observación: La función de verosimilitud también es simbolizada como $L(\mathbf{x}; \theta)$.

Método de Máxima Verosimilitud

Una vez que se ha decidido utilizar este método de estimación, la búsqueda del valor específico es un problema puramente matemático: maximización de una función. Muchas de las funciones de probabilidad comúnmente utilizadas en la práctica satisfacen condiciones de regularidad que permiten obtener el estimador máximo verosímil a partir de la solución de la ecuación:

$$\frac{\partial}{\partial \theta} L(\mathbf{x}; \theta) = 0$$

Dado que $L(\mathbf{x}; \theta)$ y $\ln L(\mathbf{x}; \theta)$ (logaritmo de la verosimilitud) alcanzan su máximo en el mismo valor de θ y en general resulta más fácil hallar el máximo de $\ln L(\mathbf{x}; \theta)$, la ecuación anterior se puede reemplazar por $\frac{\partial}{\partial \theta} \ln L(\mathbf{x}; \theta) = 0$.

En muchos casos θ representará un número pero en muchos otros representará a un vector. Por ejemplo, en la distribución Normal $\theta = (\mu, \sigma^2)$.

Método de Máxima Verosimilitud

Si $\boldsymbol{\theta}$ es un vector que contiene k parámetros, los estimadores máximo verosímiles de los parámetros $\theta_1, \theta_2, \dots, \theta_k$ son los valores que hacen máxima $L(\theta_1, \theta_2, \dots, \theta_k / \mathbf{x})$, es decir la verosimilitud para una muestra dada. Si se satisfacen ciertas condiciones de regularidad, el punto en que la verosimilitud es máxima es una solución del sistema de k ecuaciones, llamadas ecuaciones máximo verosímiles:

$$\begin{aligned} \frac{\partial}{\partial \theta_1} \ln L(\theta_1, \theta_2, \dots, \theta_k / \mathbf{x}) &= 0 \\ &\vdots \\ \frac{\partial}{\partial \theta_k} \ln L(\theta_1, \theta_2, \dots, \theta_k / \mathbf{x}) &= 0 \end{aligned}$$

Se debe tener en cuenta que es necesario probar que las soluciones de las ecuaciones anteriores corresponden a máximos locales. Si la matriz de derivadas segundas es definida negativa en $\hat{\boldsymbol{\theta}}$ entonces $\hat{\boldsymbol{\theta}}$ es el estimador MV.

Método de Máxima Verosimilitud

Ejemplo: Si X_1, X_2, \dots, X_n es una muestra aleatoria de tamaño n de una distribución $N(\mu, \sigma^2)$, la función de verosimilitud es:

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

Aplicando logaritmo:

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

Las derivadas parciales de la función anterior con respecto a μ y a σ^2 son:

$$\begin{cases} \frac{\partial L(\mathbf{x}; \mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) \\ \frac{\partial L(\mathbf{x}; \mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 \end{cases}$$

Método de Máxima Verosimilitud

Igualando a cero y resolviendo las ecuaciones que resultan respecto a μ y a σ^2 , hallamos los estimadores.

De la primera ecuación:

$$\begin{aligned}\frac{1}{\hat{\sigma}^2} \sum_{i=1}^n (x_i - \hat{\mu}) &= 0 \Rightarrow \sum_{i=1}^n x_i - n\hat{\mu} = 0 \\ \Rightarrow \hat{\mu} &= \frac{\sum_{i=1}^n x_i}{n} = \bar{x}\end{aligned}$$

De la segunda:

$$\begin{aligned}-\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{i=1}^n (x_i - \hat{\mu})^2 &= 0 \Rightarrow -\frac{n}{\hat{\sigma}^2} + \sum_{i=1}^n (x_i - \bar{x})^2 = 0 \\ \Rightarrow \hat{\sigma}^2 &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}\end{aligned}$$

$S^2 = 1/(n-1)$ por este motivo dado que sino es insesgado sin el -1

Propiedades de los Estimadores MV

Los estimadores MV pueden ser sesgados pero muchas veces dicho sesgo puede corregirse. En el ejemplo anterior, el estimador $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$ es sesgado de σ^2 pero $\frac{n}{n-1} \hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ es insesgado.

Si la función de (densidad) de probabilidad verifica ciertas condiciones de regularidad bastante generales, los estimadores MV verifican algunas propiedades entre las que se encuentran:

1. Consistencia: $\lim_{n \rightarrow \infty} \hat{\theta}_{MV} = \theta$
2. Eficiencia asintótica: $\lim_{n \rightarrow \infty} \frac{\text{Var}(\hat{\theta}_{MV})}{1/I_n(\theta)} = 1$
3. Y son función de los estadísticos suficientes.

Propiedades de los Estimadores MV

Además, estos estimadores verifican la propiedad de invariancia: sea $\hat{\theta}_{MV}$ el estimador máximo verosímil de θ de la función de probabilidad $f(x, \theta)$. Sea $u(\theta)$ una función de θ que admite inversa. Se puede probar que el estimador máximo verosímil de $u(\theta)$ es $u(\hat{\theta}_{MV})$.

Ejemplo: En la distribución Normal, el estimador MV es

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Por la propiedad de invariancia de los estimadores MV, el estimador MV de σ es

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

Método de Máxima Verosimilitud

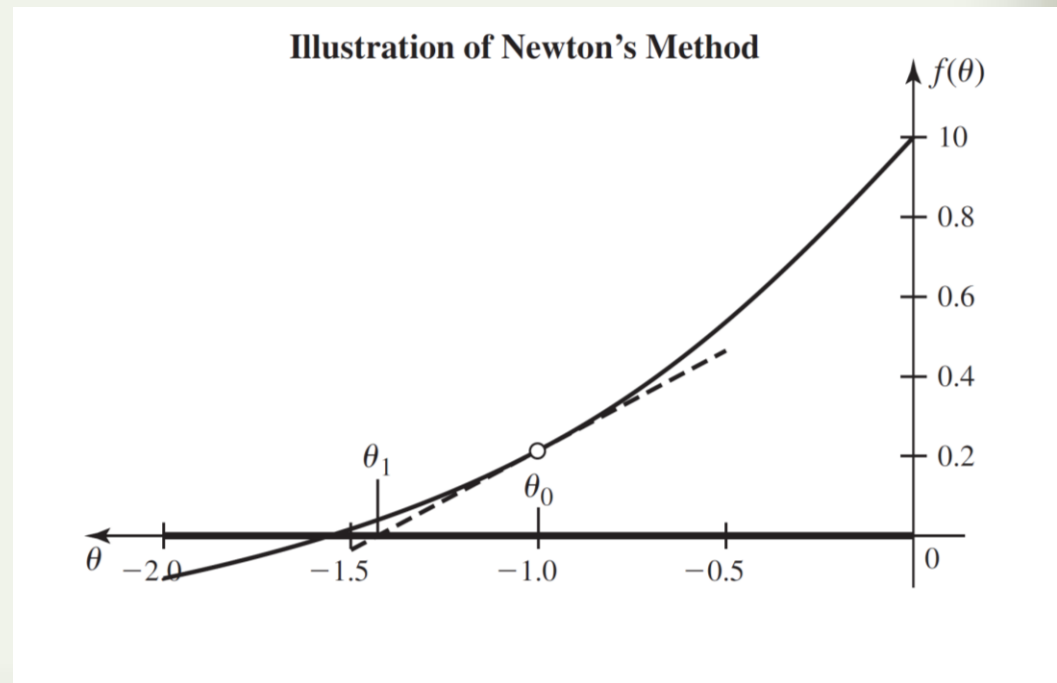
En muchos casos las ecuaciones máximo verosímiles no pueden ser resueltas explícitamente. Es decir, existen los valores $\hat{\theta}_{MV}$ para un vector de parámetros dado pero el estimador no puede ser explicitado en una expresión cerrada como función de las observaciones en la muestra. Un ejemplo de estos casos es la distribución Weibull para la que se obtiene un sistema compuesto por ecuaciones no lineales en los parámetros. En estos casos, los valores $\hat{\theta}_{MV}$ deben ser determinados empleando métodos numéricos.

Un método numérico que puede ser empleado es el Método de Newton-Raphson cuyo objetivo es aproximar la solución de $f(\theta) = 0$ donde $f(\theta)$ es una función real de una variable real, es decir encontrar el valor de θ en el cual el valor de $f(\theta)$ es 0.

Método de Newton-Raphson

El procedimiento comienza con un valor inicial θ_0 que actualiza en sucesivas iteraciones hasta que converge. En la i -ésima iteración, la actualización consiste en aproximar la curva $f(\theta)$ a través de la línea tangente a dicha curva que pasa por el punto $(\theta_i, f(\theta_i))$.

El nuevo valor de θ es aquel en el que la línea que aproxima la curva cruza al eje horizontal. El proceso de actualización continua hasta que el resultado se estabiliza (la diferencia entre dos valores sucesivos de θ es pequeña).



Método de los Momentos

El método de los momentos es un método intuitivo para estimar parámetros cuando otros métodos, más atractivos, pueden ser difíciles de implementar. También puede utilizarse para obtener valores iniciales para aplicar los métodos iterativos mencionados anteriormente.

Sea $f(x, \theta_1, \theta_2, \dots, \theta_k)$ una función de probabilidad que depende de k parámetros. Sea μ'_j el momento de orden j , es decir:

$$\mu'_j = E(X^j)$$

Y μ_j el momento centrado de orden j , es decir:

$$\mu_j = E[X - E(X)]^j$$

En general, μ_j será función de los k parámetros $\theta_1, \theta_2, \dots, \theta_k$.

Método de los Momentos

Sea una muestra aleatoria de la distribución $f(x, \theta_1, \theta_2, \dots, \theta_k)$. A partir de los valores muestrales x_1, x_2, \dots, x_n se pueden calcular los momentos muestrales respectivos como:

$$m'_j = \frac{1}{n} \sum_{i=1}^n x_i^j$$

Y

$$m_j = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^j$$

Los **estimadores de momentos** son los valores $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ que son solución de las k ecuaciones

$$m'_j = \mu'_j \quad j = 1, 2, \dots, k$$

Método de los Momentos

Ejemplo: Sea X una variable aleatoria con función de densidad

$$f(x, \theta) = \frac{\theta}{x^{\theta+1}} \quad x \geq 1, \theta > 1$$

El momento de primer orden es:

$$\mu'_1 = E(X) = \int_1^{\infty} x \frac{\theta}{x^{\theta+1}} dx = -\frac{\theta}{\theta + 1}$$

El respectivo momento muestral es

$$m'_1 = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

De acuerdo a este método:

$$-\frac{\hat{\theta}}{\hat{\theta} + 1} = \bar{x}$$

Propiedades de los Estimadores de Momentos

Por lo tanto:

$$\hat{\theta} = \frac{\bar{x}}{\bar{x} - 1}$$

Propiedades. Se puede demostrar que, bajo condiciones bastante generales, los estimadores deducidos por el método de los momentos son consistentes. Pero en general no son asintóticamente eficientes.

Método de Mínimo Cuadrados

Sea Y una variable aleatoria cuyo valor depende de otras variables X_1, X_2, \dots, X_p . Supongamos que la dependencia es de la forma:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

Donde β_j $j = 0, 1, \dots, p$ son parámetros desconocidos, los valores de las variables X_j se consideran fijos y ε es una variable aleatoria. Este modelo es llamado modelo lineal general e indica que la variable Y es un resultado aleatorio que puede descomponerse en una componente aleatoria mas $p + 1$ términos asociados a los valores de las variables X_j . El análisis de este modelo depende de las hipótesis que hagamos acerca de la variable aleatoria ε . Sean las hipótesis:

$$\begin{aligned} E(\varepsilon) &= 0 \\ V(\varepsilon) &= \sigma^2 \end{aligned}$$

Es decir, el valor esperado y la variancia de ε no dependen de los valores de X_j .

Método de Mínimo Cuadrados

Entonces:

$$\begin{aligned} E(Y) &= E(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon) \\ &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + E(\varepsilon) \end{aligned}$$

y

$$V(Y) = V(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon) = V(\varepsilon) = \sigma^2$$

Notar que nada se ha supuesto acerca de la distribución de la variable aleatoria ε ; sólo se formula una hipótesis acerca de su valor esperado y su variancia.

Supongamos que se observa la variable Y en n combinaciones distintas de valores de las variables X_j y que las observaciones son realizadas de manera independientes. Luego $(y_i; \mathbf{x}_i), (y_i; \mathbf{x}_i), \dots, (y_i; \mathbf{x}_i)$ puede ser considerada una muestra aleatoria de la variable Y para los valores de X_j

Método de Mínimo Cuadrados

Los **estimadores mínimos cuadráticos** de los parámetros $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ son los valores que minimizan:

$$S(\boldsymbol{\beta}) = \sum [Y_i - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi})]^2$$

Para minimizar $S(\boldsymbol{\beta})$ debemos resolver el sistema de $p + 1$ ecuaciones:

$$\begin{cases} \frac{\partial S(\boldsymbol{\beta})}{\partial \beta_0} = 0 \\ \frac{\partial S(\boldsymbol{\beta})}{\partial \beta_1} = 0 \\ \dots \\ \frac{\partial S(\boldsymbol{\beta})}{\partial \beta_p} = 0 \end{cases} \quad \text{Derivo e igualo a cero}$$

Algunas Propiedades de los Estimadores de Mínimos Cuadráticos

Propiedades. Los estimadores mínimo cuadráticos son insesgados. Además, si se supone que la variable aleatoria ε es Normal, entonces se puede aplicar el método de máxima verosimilitud para estimar los parámetros del modelo. En este caso, los estimadores máximo verosímiles y los estimadores mínimos cuadráticos coinciden (esto no siempre es cierto).

Ejemplo: Sea el modelo $y = \beta_0 + \beta_1 x_1 + \varepsilon$.

Los estimadores mínimo cuadráticos de β_0 y β_1 son los valores que minimizan

$$S(\boldsymbol{\beta}) = \sum [y_i - (\beta_0 + \beta_1 x_{1i})]^2$$

Estimadores de Mínimos Cuadráticos

Derivando $S(\boldsymbol{\beta})$ respecto de β_0 y β_1

$$\begin{cases} \frac{\partial S(\boldsymbol{\beta})}{\partial \beta_0} = \sum_{i=1}^n 2[y_i - (\beta_0 + \beta_1 x_{1i})](-1) \\ \frac{\partial S(\boldsymbol{\beta})}{\partial \beta_1} = \sum_{i=1}^n 2[y_i - (\beta_0 + \beta_1 x_{1i})](-x_{1i}) \end{cases}$$

Igualando a cero y despejando se obtiene

$$\begin{cases} \hat{\beta}_1 = \frac{\sum_{i=1}^n y_i (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \end{cases}$$

**Sistema de ecuaciones
Resolver B0 o B1 primero**

**En muchos casos, este metodo
coincido con Maxima Verosimilitud**