Capítulo 5

Estimación Puntual de Parámetros

5.1 El Problema de Estimación Puntual

En el análisis económico, las teorías suelen tomar forma de relaciones entre variables (por ejemplo, ecuaciones lineales) donde algunos coeficientes tienen interpretación específica y cuyo conocimiento es de importancia para las predicciones y validación de teorías. Por ejemplo, en una función lineal de consumo $C = c_0 + c_1 Y$, donde Y es el Ingreso disponible de una familia y C es su Consumo, la pendiente c_1 es la "propensión marginal a consumir" y c_0 es el "consumo autónomo", siendo ambos coeficientes de necesarios para cualquier pronóstico.

Cuando, por razones de muestreo, las variables económicas se convierten en aleatorias, los coeficientes de las relaciones entre ellas devienen en parámetros (estructurales o derivados) de sus distribuciones, y un problema importante es "aproximarlos" a partir de valores obtenidos en una muestra. Esta aproximación, en una primera instancia, consiste en especificar un valor posible para el parámetro y adjuntar el "margen de error" asociado al uso de una muestra. Este proceso se conoce como *Estimación de Parámetros* en Estadística, y en las líneas que siguen desarrollaremos sus bases.

Por ejemplo, en el caso de la relación entre Consumo C e Ingreso disponible Y, en una primera etapa tenemos un modelo teórico que podría ser resumido:

Modelo económico

Empezando con $C = f(Y)$, $\frac{\partial f(Y)}{\partial Y} > 0$ que expresa la idea que el Consumo aumenta con el Ingreso, para hacer análisis cuantitativo tenemos que escribir la relación funcional anterior de modo más explícito. Seleccionando un modelo lineal pasamos a

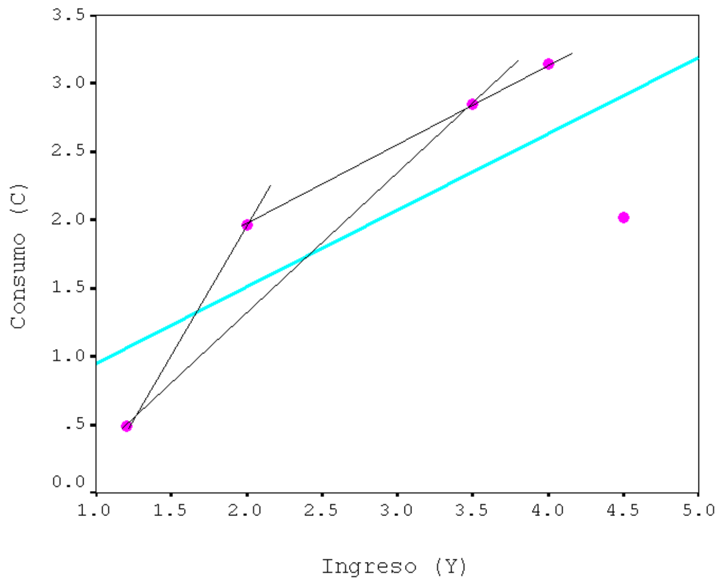
Especificación del modelo económico

$C = c_0 + c_1 Y$, $c_0 > 0$, $c_1 > 0$, donde $c_0 =$Consumo autónomo y $c_1 =$Propensión marginal a consumir. Esta especificación del modelo implica algo más que el modelo original. Así, por ejemplo, no sólo dice que el consumo es función creciente del ingreso, sino además propone una relación de proporcionalidad entre la variación del ingreso y el consumo. Adicionalmente, tiene dos parámetros característicos c_0 y c_1, con significado económico, cuyo valor es necesario conocer para poder usar plenamente el modelo, por ejemplo, en pronósticos. Si no conocemos estos valores, dado que la relación es lineal, en principio bastaría con tener dos puntos de paso de la recta para hallarlos. Asumamos que tomamos una muestra de n hogares, para las cuales hemos tomado nota de sus ingresos y consumos:

Familia	Ingreso	Consumo
1	Y_1	C_1
2	Y_2	C_2
3	Y_3	C_3
:		
n	Y_n	C_n

Si graficamos las n parejas (Y_j, C_j) en un plano, cartesiano, los más probable es que no caigan totalmente colineales y esto genera el problema de tener múltiples posibilidades de valores para los parámetros c_0 y c_1. Una explicación es que aún cuando la relación postulada por el modelo sea correcta, siempre puede haber pequeñas alteraciones de tipo aleatorio en el consumo, que originan la no colinealidad exacta y la posibilidad de muchas rectas $C = c_0 + c_1 Y$. El gráfico de abajo muestra la situación, donde la recta de línea más gruesa es la verdadera $C = c_0 + c_1 Y$.

Figura 1 Diagrama de dispersión Consumo vs Ingreso



Para darle sentido al modelo tenemos que agregar algunos supuestos de tipo probabilístico. Pasamos así a un "modelo de datos"-

Modelo de datos

Tenemos que escribir el modelo original agregando las componentes aleatorias. Una manera de hacerlo es mediante la esperanza condicional. Así tenemos $E(C\,|\,Y) = c_0 + c_1 Y$ que a su vez origina la ecuación:

$C = c_0 + c_1 Y + \varepsilon$, donde ε es variable aleatoria que representa el efecto del azar sobre el consumo.

Para completar el modelo, habrá que definir algunas características razonables para la v.a. ε. Un supuesto muy usado es asumir normalidad, con lo que nuestro modelo de datos deviene en:

$$C = c_0 + c_1 Y + \varepsilon \text{ y } \varepsilon \sim N(0, \sigma^2)$$

Finalmente y como consecuencia de todo, tenemos:

$E(C\,|\,Y) = c_0 + c_1 Y + E(\varepsilon\,|\,Y) = c_0 + c_1 Y$, pues $E(\varepsilon\,|\,Y) = 0$

$V(C\,|\,Y) = V(c_0 + c_1 Y + \varepsilon\,|\,Y) = V(c_1 Y\,|\,Y) + V(\varepsilon\,|\,Y) = \sigma^2$, pues $V(\varepsilon\,|\,Y) = V(\varepsilon) = \sigma^2$ y $V(c_1 Y\,|\,Y) = 0$

Con lo anterior, el modelo final de datos resulta $C \sim N(c_0 + c_1 Y, \sigma^2)$, que además de c_0 y c_1, tiene ahora un parámetro más σ^2. Nótese que, de los tres parámetros, dos de ellos (c_0 y c_1) son en realidad parámetros económicos, que se han convertido en parámetros estadísticos sólo por mayor conveniencia de análisis.

Estos parámetros deben ser aproximados a partir de los datos. Esta tarea es una de las más importantes de la Estadística Inferencial.

5.2 Elementos

Sea X v.a. con función de densidad o probabilidad $f_X(x,\theta)$ donde θ es un parámetro -o vector de parámetros- de interés pero de valor desconocido.

Espacio Paramétrico
Denotado Θ es el conjunto de valores posibles para θ

Espacio de Información
Denotado $\mathbf{X}$, es el conjunto de todas las muestras posibles de tamaño n, i.e. $\mathrm{X} = \{(X_1, X_2, ..., X_n) \mid X_j \in R_X\}$

Ejemplos
(1) Si $X \sim U(x;\alpha = 0, \beta = \theta) \Rightarrow \Theta =]0, +\infty[$ y $\mathrm{X} = \{(X_1, X_2, ..., X_n) \mid 0 < X_j \leq \theta\}$
(2) Si $X \sim N(x;\mu,\sigma^2) \Rightarrow \Theta = \{(\mu,\sigma^2) \mid \mu \in]-\infty,+\infty[, \sigma^2 \in]0,+\infty[\}$ y $\mathrm{X} = R^n$

5.3 El Problema de Estimación Puntual

Generalmente no se conoce el valor de θ y sólo se puede obtener una m.a. de tamaño n de la población de X, para tratar de aproximar el valor de θ. O sea, se tiene un punto $(X_1, X_2, ..., X_n)$ del Espacio de Información $\mathbf{X}$ y se desea aproximar un punto θ del Espacio Paramétrico Θ, de modo que el error de aproximación sea mínimo.

Observación
Puede haber muchas maneras de "aproximar" –estimar es el nombre técnico- un parámetro θ, y cada una tendrá un "error de aproximación". Obviamente deseamos identificar la manera con la cual se comete el menor error posible. Esto implica ponerse de acuerdo primero en qué se va a considerar un error aceptable.

5.4 Estimador y Valor Estimado

Aproximar el valor de θ a partir de una m.a. $(X_1, X_2, ..., X_n)$ pasa por resumir los *n* datos en uno solo, a partir de la aplicación de alguna "fórmula de trabajo" $H(X_1, X_2, ..., X_n)$ adecuada. Esta fórmula se usa como regla de trabajo, como una plantilla que se aplica siempre. Naturalmente, aún cuando la fórmula sea la misma, los valores que se obtengan dependen de los valores que uno encuentre al tomar efectivamente la muestra aleatoria. Para distinguir estas dos facetas del mismo proceso, creamos los conceptos de Estimador y de Valor Estimado.

Estimador
Un estimador de un parámetro θ es una estadística $H(X_1, X_2, ..., X_n)$ cuyo valor, una vez tomada la m.a., se usa como aproximación de θ. Es frecuente denotar el estimado usando la misma letra del parámetro, pero agregándole un acento circunflejo, y se escribe entonces $\hat{\theta}(X_1, X_2, ..., X_n)$ o simplemente $\hat{\theta}$.

Valor Estimado o Estimación
Es un valor particular del estimador $\hat{\theta}(X_1, X_2, ..., X_n)$, asumido ante unos datos específicos.

Error de estimación
Si $\hat{\theta}$ denota al estimador $\hat{\theta} = \hat{\theta}(X_1, X_2, ..., X_n)$, el error de estimación e se define $e := (\hat{\theta} - \theta)$

Observaciones
(1) Por lo general se escribe $\hat{\theta}$ en lugar de $\hat{\theta}(X_1, X_2, ..., X_n)$ para denotar el estimador.
(2) El valor estimado es una valor particular $\hat{\theta}_0$ de la variable aleatoria $\hat{\theta}$ que luego será usado <u>como si fuera el valor verdadero</u> del parámetro θ

(3) Naturalmente, se desea que la diferencia $|\hat{\theta}-\theta|$ sea mínima o nula, i.e. deseamos tener un "buen valor estimado". Lo anterior se logra construyendo un "buen" estimador $\hat{\theta}$, o sea, construyendo una fórmula de trabajo que, salvo por azar, funcione correctamente, que permita seleccionar un punto del Espacio Paramétrico $\boldsymbol{\Theta}$ muy próximo, o de ser posible igual, al parámetro $\boldsymbol{\theta}$. **Por lo anterior, al problema de construir un buen estimador lo llamaremos "Problema de Estimación Puntual".**

(4) La solución <u>racional</u> al problema planteado pasa por definir primero las características que consideraremos "buenas" para un estimador y luego buscar métodos que logren este tipo de estimadores.

5.5 Propiedades de un Buen Estimador

5.5.1 Insesgamiento

Un estimador $\hat{\theta}$ se dice Insesgado si $E(\hat{\theta})=\theta \quad \forall\theta\in\Theta$

Ejemplos

(1) Si $X \sim U(x;\alpha=0,\beta=\theta)$ y $(X_1,X_2,...,X_n)$ es m.a. de esta distribución y definimos $\hat{\theta}=2\overline{X}$, entonces $\hat{\theta}$ es estimador insesgado de θ .

En efecto, $E(\hat{\theta})=E(2\overline{X})=2E(\overline{X})=2\mu_X=2\times\dfrac{\theta}{2}=\theta$.

(2) Si $X \sim U(x;0,\theta)$ y definimos $\tilde{\theta}=4\dfrac{\sum_{i=1}^{n} iX_i}{n(n+1)}$, entonces $\hat{\theta}$ es estimador insesgado de θ . Para ello basta ver que

$$E(\tilde{\theta})=E\left(4\frac{\sum_{i=1}^{n} iX_i}{n(n+1)}\right)=\frac{4}{n(n+1)}E\left(\sum_{i=1}^{n} iX_i\right)=\frac{4}{n(n+1)}\sum_{i=1}^{n} i\underbrace{E(X_i)}_{\theta/2}=\frac{4}{n(n+1)}\frac{\theta}{2}\sum_{i=1}^{n} i=\frac{4}{n(n+1)}\frac{\theta}{2}\frac{n(n+1)}{2}=\theta$$

(3) Si $X \sim U(x;0,\theta)$ y $\tilde{\tilde{\theta}}=Máx\{X_i\}$, entonces $\tilde{\tilde{\theta}}$ <u>no</u> es estimador insesgado de θ , pues $E[\tilde{\theta}]=\dfrac{n}{n+1}\theta\neq\theta$.

Veamos, usando como notación $Y=\tilde{\tilde{\theta}}$ para simplificar cálculos, tenemos $\forall X_j \quad 0\le X_j\le\theta \Rightarrow$ $0\le\underbrace{Máx\{X_j\}}_{Y}\le\theta \Rightarrow R_Y=[0,\theta]$.

Sea ahora y un elemento cualquiera de R_Y, entonces, en general:

$$F_Y(y)=P[(Y\le y)]=P[Máx\{X_j\}\le y]=P[(X_1\le y)\cap(X_2\le y)\cap...\cap(X_n\le y)]=P\left[\bigcap_{j=1}^{n}(Xj\ \le y)\right]=$$

$$=\prod_{j=1}^{n}P(X_j\ \le y)=\prod_{j=1}^{n}\int_0^y f_{X\,j}(x_j)dx_j)=\prod_{j=1}^{n}\int_0^y f_X(x_j)dx_j=\prod_{j=1}^{n}F_X(y)=[F_X(y)]^n$$. Pero por dato $X \sim U(x;0,\theta)$

luego, en este caso particular, $F_X(y)=\int_0^y\frac{1}{\theta}dx=\frac{y}{\theta}$ y $F_Y(y)=[F_X(y)]^n=\left(\frac{y}{\theta}\right)^n$, es decir, la distribución acumulativa de Y es $F_Y(y)=\left(\frac{y}{\theta}\right)^n \quad 0\le y\le\theta$. Derivando con respecto a y, obtenemos

$$f_Y(y)=\frac{d}{dy}F_Y(y)=n\left(\frac{y}{\theta}\right)^{n-1}(\frac{1}{\theta})=n\frac{y^{n-1}}{\theta^n}.$$

O sea la función de densidad de Y es $f_Y(y) = n\dfrac{y^{n\text{-}1}}{\theta^n}\quad 0 \le y \le \theta$.

Finalmente $E(Y) = \int_0^\theta y f_Y(y)dy = \int_0^\theta yn\dfrac{y^{n-1}}{\theta^n}dy = \int_0^\theta n\dfrac{y^n}{\theta^n}dy = \dfrac{n}{\theta^n}\int_0^\theta y^n dy = \dfrac{n}{\theta^n}\left[\dfrac{y^{n+1}}{n+1}\right]_0^\theta = \dfrac{n}{\theta^n}\left(\dfrac{\theta^{n+1}}{n+1}\right) = \dfrac{n}{n+1}\theta$, así que el valor esperado de Y es $E(Y) = \dfrac{n}{n+1}\theta \Leftrightarrow E[\tilde{\theta}] = \dfrac{n}{n+1}\theta \ne \theta$

5.5.2 Eficiencia

Sean $\hat{\theta}$ y $\tilde{\theta}$ dos estimadores <u>insesgados</u> del mismo parámetro θ . Diremos que $\hat{\theta}$ es más eficiente que $\tilde{\theta}$ si $V(\hat{\theta}) < V(\tilde{\theta})$

Ejemplo

Si $X \sim U(x;\alpha = 0, \beta = \theta)$ y $(X_1, X_2, ..., X_n)$ es m.a. de esta distribución y definimos $\hat{\theta} = 2\overline{X}$ y $\tilde{\theta} = (X_1 + X_n)$, entonces $\hat{\theta}$ es más eficiente que $\tilde{\theta}$, pues $E(\tilde{\theta}) = E(X_1 + X_n) = \theta$, mientras que $V(\hat{\theta}) = V(2\overline{X}) = 4V(\overline{X}) = 4\dfrac{\sigma_X^2}{n} = 4\dfrac{\theta^2}{12n} = \dfrac{\theta^2}{3n}$ y $V(\tilde{\theta}) = V(X_1 + X_2) = \dfrac{\theta^2}{12} + \dfrac{\theta^2}{12} = \dfrac{\theta^2}{6}$. Se ve que $\hat{\theta}$ es más eficiente.

5.5.3 Mejor Estimador Lineal Insesgado (MELI)

De todos los estimadores que podemos imaginar para un parámetro θ, los más simples son aquellos que tienen estructura lineal, y dentro de este grupo de estimadores, los de más interés son los insesgados. Es natural que dentro de esta última clase busquemos, si existe, aquél que sea el mejor, o sea el de menor varianza.

Definición

$\hat{\theta}$ estimador de θ se dice Mejor Estimador Lineal Insesgado (MELI o BLUE) de θ si:

(1) $\hat{\theta}$ es función **lineal** de la muestra, i.e. $\hat{\theta} = \sum_{j=1}^{n} \alpha_j X_j$

(2) $\hat{\theta}$ es **insesgado**, i.e. $E[\hat{\theta}] = \theta \quad \forall \theta \in \Theta$

(3) $\hat{\theta}$ es de **varianza mínima** en relación a cualquier otro estimador lineal e insesgado de θ

Observaciones:

- La condición (2), llamada "de insesgamiento", implica que $E[\hat{\theta}] = \theta \Leftrightarrow$

  $$E[\hat{\theta}] = E[\sum_{j=1}^{n}\alpha_j X_j] = \sum_{j=1}^{n}\alpha_j E[X_j] = \theta \ \ \forall\theta \ \varepsilon\ \Theta$$

- La condición (3), llamada "de varianza mínima" implica que $V[\hat{\theta}] = Mín \Leftrightarrow$

  $$V[\hat{\theta}] = V[\sum_{j=1}^{n}\alpha_j X_j] = \sum_{j=1}^{n}\alpha_j^2 V[X_j] = Mín$$

- Juntando (2) y (3) se ve que **construir el MELI equivale a resolver el Problema de optimización**:

  $$Mín\sum_{j=1}^{n}\alpha_j^2 V[X_j] \ \text{ s.a } \ \sum_{j=1}^{n}\alpha_j E[X_j] = \theta \ \ \forall\theta \ \varepsilon\ \Theta$$

  donde las incógnitas son las constantes $\alpha_1, \alpha_2, ..., \alpha_n$.

Ejemplo

Sea $X \sim ?(\mu,\sigma^2)$ (o sea con distribución arbitraria de media μ y varianza σ^2). Dada una m.a. $(X_1, X_2,..., X_n)$, construir el MELI de μ

Solución:

Sea $\tilde{\mu}$ **el MELI de** μ, entonces aplicando la definición:

- $\tilde{\mu}$ es función **lineal** de la muestra $\Rightarrow\ \tilde{\mu} = \sum_{j=1}^{n} \alpha_j X_j$
- $\tilde{\mu}$ debe ser **insesgado** $\Rightarrow\ E[\tilde{\mu}] = \mu \Leftrightarrow \sum_{j=1}^{n} \alpha_j E[X_j] = \mu \Leftrightarrow \sum_{j=1}^{n} \alpha_j \mu = \mu \Leftrightarrow \sum_{j=1}^{n} \alpha_j = 1$
- $\tilde{\mu}$ debe ser de **varianza mínima** $\Rightarrow V[\tilde{\mu}] = V[\sum_{j=1}^{n} \alpha_j X_j] = \sum_{j=1}^{n} \alpha_j^2 V[X_j] = \sum_{j=1}^{n} \alpha_j^2 \sigma^2 = Mín$

Luego construir $\tilde{\mu}$ **equivale a resolver el problema:**

$$Mín \quad \sigma^2 \sum_{j=1}^{n} \alpha_j^2 \quad \text{s.a} \quad \sum_{j=1}^{n} \alpha_j = 1$$

Pero como $\sigma^2 > 0$, minimizar $\sigma^2 \sum_{j=1}^{n} \alpha_j^2$ equivale a minimizar $\sum_{j=1}^{n} \alpha_j^2$, por tanto, basta con resolver:

$$Mín \quad \sum_{j=1}^{n} \alpha_j^2 \quad \text{s.a} \quad \sum_{j=1}^{n} \alpha_j = 1$$

Aplicando Multiplicadores de Lagrange, $L(\alpha_1, \alpha_2,..., \alpha_n, \lambda) = \sum_{j=1}^{n} \alpha_j^2 + \lambda(1 - \sum_{j=1}^{n} \alpha_j)$ **es el "Lagrangiano" y minimizando:**

Condiciones de primer orden $\nabla L = 0$

$$\frac{\partial L}{\partial \alpha_1} = 0 \Leftrightarrow 2\alpha_1 - \lambda = 0 \Rightarrow \alpha_1 = \frac{\lambda}{2}$$

$$\frac{\partial L}{\partial \alpha_2} = 0 \Leftrightarrow 2\alpha_2 - \lambda = 0 \Rightarrow \alpha_2 = \frac{\lambda}{2}$$

:

:

$$\frac{\partial L}{\partial \alpha_1} = 0 \Leftrightarrow 2\alpha_n - \lambda = 0 \Rightarrow \alpha_n = \frac{\lambda}{2}$$

$$\frac{\partial L}{\partial \lambda} = 0 \Leftrightarrow (1 - \sum_{j=1}^{n} \alpha_j) = 0$$

Despejando y reemplazando $\alpha_j = \frac{\lambda}{2}$ *j=1,2,...n*, en la última ecuación, se obtiene $(1 - \sum_{j=1}^{n} \frac{\lambda}{2}) = 0 \Leftrightarrow$ $1 - n\frac{\lambda}{2} = 0 \Leftrightarrow \lambda = \frac{2}{n} \Rightarrow \alpha_j = \frac{1}{n}$ *j=1,2,...n* y finalmente, asumiendo que las condiciones de segundo orden se cumplen, tenemos que el MELI de μ es $\tilde{\mu} = \sum_{j=1}^{n} \alpha_j X_j = \sum_{j=1}^{n} \frac{1}{n} X_j$; es decir, el MELI de μ es $\overline{X}$

5.5.4 Consistencia

Las propiedades de insesgamiento y eficiencia son muy deseables, pero son propiedades basadas en el valor esperado, es decir, en lo que ocurre "en promedio" con los diferentes valores del estimador, considerando todas las muestras posibles. Si un estimador va a ser usado muchas veces, (como por ejemplo cuando un diseña una metodología para estimar el la tasa de desempleo) el criterio de lo que ocurre "en el promedio" con el estimador es adecuado; pero si el estimador va a ser usado una vez (o pocas veces), realmente nos interesa que en esa vez, la discrepancia entre estimador y parámetro sea pequeña. El precio que se suele pagar es tomar una muestra de mayor tamaño. Lo anterior motiva la siguiente definición.

Definición

Sea $(X_1, X_2,..., X_n)$ m.a. de tamaño n y sea $\hat{\theta}$ un estimador de θ. Diremos que $\hat{\theta}$ es Consistente si

$$\lim_{n\to\infty} P(|\hat{\theta}-\theta|\leq\varepsilon)=1 \quad \forall\ \varepsilon>0$$

Observaciones:

(1) Informalmente, $\hat{\theta}$ es Consistente si cuando la muestra es de tamaño suficientemente grande, $\hat{\theta}$ y θ coinciden

(2) En sentido estricto, como para cada *n* hay un estimador $\hat{\theta}$, debiera escribirse $\hat{\theta}_n$ y decirse que "$\hat{\theta}_n=\hat{\theta}$ $(X_1,X_2,...,X_n)$ es estimador consistente si $\lim\limits_{n\to\infty} P(|\hat{\theta}_n-\theta|\leq\varepsilon)=1 \quad \forall\ \varepsilon>0$"

Ejemplos

(1) Sea $X\sim ?(\mu,\sigma^2)$ (o sea, con distribución arbitraria de media μ y varianza σ^2). ¿Es $\tilde{\mu}:=\overline{X}$ es estimador consistente de μ?

Solución

Aplicando la Ley de Grandes Números, ya vimos que dado un $\varepsilon>0$ arbitrario, se cumple que $\lim\limits_{n\to\infty} P(|\overline{X}-\mu|\leq\varepsilon)=1$, es decir, en efecto $\tilde{\mu}:=\overline{X}$ es estimador consistente de μ

(2) Sea $X\sim N(\mu,\sigma^2)$ ¿Es S^2 estimador consistente de σ^2?

Solución

Como ya vimos en el capítulo anterior, $E[S^2]=\sigma^2$ y $V[S^2]=\dfrac{2\sigma^4}{n-1}$, luego dado un $\varepsilon>0$, queremos probar que $n\to\infty \Rightarrow P(|S^2-\sigma^2|\leq\varepsilon)\to 1$. Ahora bien, aplicando la D. de Tchebychev:

$P(|S^2-\sigma^2|\leq k\sqrt{\dfrac{2\sigma^4}{n-1}})\geq 1-\dfrac{1}{k^2}$ $\forall k$, si hacemos $k\sqrt{\dfrac{2\sigma^4}{n-1}}=\varepsilon$, entonces $P(|S^2-\sigma^2|\leq\varepsilon)\geq 1-\dfrac{2\sigma^4}{\varepsilon^2 n-1}$ y $n\to\infty \Rightarrow P(|S^2-\sigma^2|\leq\varepsilon)\to 1$, así pues, S^2 estimador consistente de σ^2

(3) Sea $\hat{\theta}$ un estimador insesgado de θ y tal que $\lim\limits_{n\to\infty} V(\hat{\theta})=0$. Probar que $\hat{\theta}$ un estimador consistente de θ.

Solución

Aplicando la D. de Tchebychev: $P(|\hat{\theta}-\theta|\leq k\sqrt{V(\hat{\theta})}\geq 1-\dfrac{1}{k^2}$ $\forall k>0$. Dado un $\varepsilon>0$, tomemos $k\sqrt{V(\hat{\theta}}=\varepsilon$.

Entonces $P(|\hat{\theta}-\theta|\leq\varepsilon)\geq 1-\dfrac{V(\hat{\theta})}{\varepsilon^2}$ y si $n\to\infty \Rightarrow P(|\hat{\theta}-\theta|\leq\varepsilon)=1$ pues $\lim\limits_{n\to\infty} V(\hat{\theta})=0$.

Proposición

Sea $\hat{\theta}$ un estimador de θ tal que

(1) $\lim_{n\to\infty} E(\hat{\theta})=\theta$ ($\hat{\theta}$ es "asintóticamente insesgado") y

(2) $\lim_{n\to\infty} V(\hat{\theta})=0$ ($\hat{\theta}$ es "asintóticamente eficiente")

Entonces $\hat{\theta}$ un estimador consistente de θ.

Demostración:

Es similar al ejemplo (4): se aplica la D. de Tchebychev y luego se toma el límite.

Ejemplo

Sea $X \sim N(\mu,\sigma^2)$ y sea $\tilde{\sigma}^2 = \dfrac{\sum_{j=1}^{n}(X_j-\overline{X})^2}{n}$. Verificar que $\tilde{\sigma}^2$ es estimador consistente de σ^2

Solución

Apliquemos la proposición anterior:

(1) $E[\tilde{\sigma}^2]=E[\dfrac{\sum_{j=1}^{n}(X_j-\overline{X})^2}{n}]=E[\dfrac{(n-1)S^2}{n}]=\dfrac{(n-1)E[S^2]}{n}=\dfrac{(n-1)\sigma^2}{n}$. O sea, $\tilde{\sigma}^2$ es sesgado. Pero obviamente $\lim_{n\to\infty} E[\tilde{\sigma}^2]=\sigma^2 \Rightarrow \tilde{\sigma}^2$ **es estimador "asintóticamente insesgado" de σ^2**

(2) $V[\tilde{\sigma}^2]=V[\dfrac{(n-1)S^2}{n}]=\dfrac{(n-1)^2V[S^2]}{n^2}=\dfrac{(n-1)^2}{n^2}\dfrac{2\sigma^4}{n-1}=\dfrac{(n-1)2\sigma^4}{n^2}$ y por tanto $\lim_{n\to\infty} V(\tilde{\sigma}^2)=0$ o sea, $\tilde{\sigma}^2$ **es "asintóticamente eficiente" .**

Como se cumplen las condiciones (1) y (2), podemos decir que $\tilde{\sigma}^2$ es estimador consistente de σ^2

Definición (Límite en Probabilidad)

Sea $\{W_n\}$ una sucesión de variables aleatorias y sea C una constante. Se dice que C es el límite en probabilidad de $\{W_n\}$, lo que se denota escribiendo $Plim W_n = C$, si $\lim_{n\to\infty} P(|W_n-C|\le\varepsilon)=1\ \forall\ \varepsilon>0$

Observación

Comparando con la definición de consistencia, se ve que esta última es un caso particular de límite en probabilidad. En otras palabras: $\hat{\theta}$ es estimador Consistente de θ $\Leftrightarrow$ $Plim\,\hat{\theta}=\theta$. Esto no sería mayormente importante si no fuera porque el límite en probabilidad tiene propiedades útiles.

Propiedades:

Sean $\{W_n\}$ y $\{V_n\}$ sucesiones tales que Plim W_n = C y Plim V_n = B, entonces se cumple:

(1) Plim [$W_n \pm V_n$] = C ± B

(2) Plim [W_n][V_n] = CB

(3) Plim [W_n]/[V_n] = C/B si B ≠ 0

Proposición (Teorema de Slutsky)

Sea $\{W_n\}$ sucesión tal que Plim W_n = C y sea g(t) una función continua de t. Entonces Plim g(W_n)=g(C)

Observación:

Aunque no demostraremos estas propiedades, las usaremos libremente para demostrar la consistencia.

Ejemplo

Sea $X \sim N(\mu,\sigma^2)$, como $\sigma = g(\sigma^2) = \sqrt{\sigma^2}$ y ya vimos que S^2 es estimador consistente de σ^2, entonces podemos escribir :

- $\text{Plim}\, S^2 = \sigma^2$
- Aplicando el T. de Slutsky: $\text{Plim}\, S = \text{Plim}\sqrt{S^2} = \sqrt{\text{Plim} S^2} = \sqrt{\sigma^2} = \sigma$

De lo anterior podemos decir que S es estimador consistente de σ. Nótese que de haber usado la definición misma de consistencia, hubiera sido muy difícil probarla.

Análogamente, como Plim $Plim\,\overline{X}^2 = (Plim\,\overline{X})^2 = (\mu)^2 = \mu^2$, entonces podemos decir que $\overline{X}^2$ es estimador consistente de μ^2.

Ejemplo

Sea $\boldsymbol{X \sim U(x;0,\theta)}$ y sea $\hat{\theta} = Máx\{X_j\}$. Verificar que $\hat{\theta}$ es estimador consistente de θ

Solución

Por comodidad de notación sea $Y = Máx\{X_j\}$. En el capítulo anterior se probó que la distribución acumulativa de $Y = Máx\{X_j\}$ es $F_Y(y) = \left(\frac{y}{\theta}\right)^n \quad 0 \le y \le \theta$ y derivando obtuvimos la función de densidad del máximo como $f_Y(y) = n\frac{y^{n-1}}{\theta^n} \quad 0 \le y \le \theta$.

Aplicando la propiedad anterior tenemos $E(Y) = \int_0^\theta y f_Y(y)dy = \frac{n}{n+1}\theta$

Análogamente $E(Y^2) = \int_0^\theta y^2 f_Y(y)dy = \left(\frac{n}{n+2}\right)\theta^2$ y $V(Y) = \left(\left(\frac{n}{n+2}\right) - \left(\frac{n}{n+1}\right)^2\right)\theta^2$

o regresando a la notación del estimador:

$E(\hat{\theta}) = \frac{n}{n+1}\theta$ y $V(\hat{\theta}) = \left(\left(\frac{n}{n+2}\right) - \left(\frac{n}{n+1}\right)^2\right)\theta^2$

Tomando límite cuando n tiende al infinito se ve que se cumplen las dos condiciones de Insesgamiento asintótico y Eficiencia asintótica, así que $\hat{\theta}$ es estimador consistente de $\boldsymbol{\theta}$

5.6 Métodos de Estimación

Vistas las propiedades de un buen estimador, queda la tarea de diseñar métodos o sistemas que proporcionen la base para construir buenos estimadores. Los tres métodos básicos son: Método de Momentos, Método de Máxima Verosimilitud y Método de Mínimos Cuadrados. Se diferencian en lo que asumen como entrada del proceso de estimación:

- El Método de Momentos asume que conocemos la relación entre el o los parámetros que deseamos estimar y los valores esperados de las sucesivas potencias de X (los "momentos" de X).
- El Método de Máxima Verosimilitud asume que conocemos la forma de la función de densidad o probabilidad de X .
- El Método de Mínimos Cuadrados supone que se conoce la forma de la relación entre dos variables (en términos del valor esperado de una de ellas) y que los parámetros son coeficientes de esta relación.

5.6.1 Método de Momentos

5.6.1.1 Definiciones Básicas

Sea X una v.a. con función de densidad o probabilidad $f_X(x;\theta)$ donde θ es un vector de parámetros $\theta=(\theta_1,\theta_2,...,\theta_p)$. La forma de $f_X(x;\theta)$ no necesariamente es conocida.

Momentos Poblacionales

El k-ésimo momento poblacional, denotado m_k se define mediante $m_k = E(X^k)$ **si existe el correspondiente valor esperado**

Momentos Muestrales

Si $(X_1, X_2,..., X_n)$ es m.a. tomada de la población de X , el k-ésimo momento muestral, M_k se define mediante $M_k = \sum_{j=1}^{n} X^k / n$

Propiedad

M_k es estimador insesgado de m_k

En efecto: $E(M_k) = E\left(\sum_{j=1}^{n} X^k / n\right) = \frac{1}{n} E\left(\sum_{j=1}^{n} X^k\right) = \frac{1}{n}\sum_{j=1}^{n} \underbrace{E(X^k)}_{m_k} = \frac{1}{n} n m_k = m_k$

Observación

Se prueba además que si m_{2k} existe, entonces $V(M_k) = \dfrac{m_{2k} - m_k^2}{n}$ y como esta última tiende a cero cuando crece n, resulta que M_k es estimador <u>consistente</u> de m_k

Racionalidad del método

En general, se puede considerar que $m_k = E(X^k)$ define una ecuación que expresa al k-ésimo momento poblacional como función de los parámetros. Por ejemplo, en el caso continuo: $m_k = E(X^k) = \int_{-\infty}^{\infty} x^k f_X(x;\theta_1,\theta_2,...,\theta_p)dx = h_k(\theta_1,\theta_2,...,\theta_p) \quad k = 1,2,..., p$, ya que al integrar con respecto a X , esta variable y su f_X desaparecen, quedando todo en términos de $(\theta_1,\theta_2,...,\theta_p)$

Aplicando la idea anterior y tomando los p primeros momentos, obtenemos un sistema de p ecuaciones simultáneas y probablemente no lineales **("ecuaciones estructurales")** de la forma:

$$\begin{cases} m_1 = & E\left(X^1\right) = & h_1(\theta_1,\theta_2,\ldots,\theta_p) \\ m_2 = & E\left(X^2\right) = & h_2(\theta_1,\theta_2,\ldots,\theta_p) \\ \vdots & \vdots & \vdots \\ m_p = & E\left(X^p\right) = & h_p(\theta_1,\theta_2,\ldots,\theta_p) \end{cases}$$

Y como tenemos a mano $M_1, M_2, \ldots, M_p$ que son "buenos estimadores" de $m_1, m_2, \ldots, m_p$, o sea del lado izquierdo del sistema, podemos "despejar" los parámetros $(\theta_1,\theta_2,\ldots,\theta_p)$ y obtener estimadores de ellos.

5.6.1.2 Estimadores de Momentos

En el contexto anterior, diremos que $\hat\theta_1,\hat\theta_2,\ldots,\hat\theta_p$ estimadores de $\theta_1,\theta_2,\ldots,\theta_p$ respectivamente, son estimadores obtenidos mediante el Método de Momentos, si son solución al sistema de ecuaciones:

$$\begin{cases} M_1 = & E\left(X^1\right) = & h_1(\hat\theta_1,\hat\theta_2,\ldots,\hat\theta_p) \\ M_2 = & E\left(X^2\right) = & h_2(\hat\theta_1,\hat\theta_2,\ldots,\hat\theta_p) \\ \vdots & \vdots & \vdots \\ M_p = & E\left(X^p\right) = & h_p(\hat\theta_1,\hat\theta_2,\ldots,\hat\theta_p) \end{cases}$$

Ejemplos:

(1) Sea X ~ U(x;0,θ). Estimar θ mediante el Método de Momentos.

Solución:
Aquí tenemos p=1 parámetro y podemos escribir:

$$\mathbf{m_1 = E[X^1] = h_1(\theta) = \theta/2} \quad \textbf{("ecuación estructural")}$$

Como $M_1 = \Sigma X_j/n = \overline{X} \Rightarrow$ Reemplazando parámetros por estimadores en la ecuación estructural:

$$\mathbf{M_1 = h_1(\hat\theta) \Leftrightarrow \overline{X} = \hat\theta/2} \quad \textbf{("ecuación de estimación")}$$

La solución es $\hat\theta$ **= 2** $\overline{X}$ que es el "estimador de momentos" de **θ**

Observación:
El estimador $\hat\theta$ **es insesgado y de mínima varianza, y por lo general será un buen estimador,** pero como el método no contempla el rango de X, es posible que para ciertas muestras, el valor estimado resulte absurdo. Así, por ejemplo, si n=4 y la m.a. es $X_1=0, X_2=0, X_3=0, X_4=1$ entonces $\overline{X}$ **=0.25 y** el valor estimado de $\hat\theta$ es $\hat\theta = 2\overline{X} = 0.5$ **que es totalmente contradictorio con el valor X_4=1** (recordar que el rango de X_j es $0 \le X_j \le \theta \ \forall\ j$)

(2) Sea X ~Γ(x;α,β). Estimar α y β usando Momentos.

Solución:
En este caso hay p=2 parámetros en juego. Tomando los dos primeros momentos, llegamos a la ecuaciones estructurales:

$$\mathbf{m_1 = E[X^1] = h_1(\alpha,\beta) = \alpha\beta}$$
$$\mathbf{m_2 = E[X^2] = h_2(\alpha,\beta) = V(X) + (E[X])^2 = \alpha\beta^2 + (\alpha\beta)^2}$$

Como $M_1 = \Sigma X_j/n = \overline{X}$ y $M_2 = \Sigma X_j^2/n \Rightarrow$ Las ecuaciones de estimación son:

$$\mathbf{M_1 = h_1(\hat\alpha, \hat\beta) \Leftrightarrow \overline{X} = \hat\alpha\,\hat\beta} \qquad \text{(a)}$$
$$\mathbf{M_2 = h_2(\hat\alpha, \hat\beta) \Leftrightarrow \Sigma X_j^2/n = \hat\alpha\,\hat\beta^2 + (\hat\alpha\,\hat\beta)^2} \quad \text{(b)}$$

Resolviendo el sistema (a) y (b):

De (a) $\hat{\alpha} = \overline{X}/\hat{\beta}$ y en (b): $\Sigma X_j^2/n = (\overline{X}/\hat{\beta})\hat{\beta}^2 + [(\overline{X}/\hat{\beta})\hat{\beta}]^2 = \overline{X}\hat{\beta} + \overline{X}^2 \Rightarrow$

$\hat{\beta} = [(\Sigma X_j^2/n) - \overline{X}^2]/\overline{X} = [M_2 - \overline{X}^2]/\overline{X}$ y por tanto $\hat{\alpha} = \overline{X}^2/[M_2 - \overline{X}^2]$ son los estimadores de momentos.

Observación:

Nótese que como en este caso $X_j > 0$, se cumple que $\overline{X} > 0$, por lo que el denominador de $\hat{\beta}$ está bien definido. Por otra parte, $[M_2 - \overline{X}^2] = [(\Sigma X_j^2/n) - \overline{X}^2] = [\Sigma X_j^2 - n\overline{X}^2]/n = [\Sigma(X_j - \overline{X})^2]/n \geq 0$, de modo que los estimadores de $\hat{\alpha}$ **y** $\hat{\beta}$ nunca tomarán valores absurdos (negativos).

(3) Sea $X \sim U(x; \alpha,\beta)$. Estimar α y β usando el Método de Momentos.

Solución:

- Las ecuaciones estructurales son:

$$m_1 = E[X^1] = h_1(\alpha,\beta) = \frac{\alpha+\beta}{2}$$

$$m_2 = E[X^2] = h_2(\alpha,\beta) = V(X) + (E[X])^2 = \frac{(\beta-\alpha)^2}{12} + (\frac{\alpha+\beta}{2})^2$$

- Las ecuaciones de estimación son:

$$M_1 = h_1(\hat{\alpha},\hat{\beta}) = \frac{\hat{\alpha}+\hat{\beta}}{2} \qquad \text{(a)}$$

$$M_2 = h_2(\hat{\alpha},\hat{\beta}) = \frac{(\hat{\beta}-\hat{\alpha})^2}{12} + (\frac{\hat{\alpha}+\hat{\beta}}{2})^2 \qquad \text{(b)}$$

- Resolviendo el sistema (a) y (b):

(a) $\Rightarrow \hat{\alpha} = 2M_1 - \hat{\beta}$ y en (b) $M_2 = \dfrac{(\hat{\beta} - 2M_1 + \hat{\beta})^2}{12} + (\dfrac{2M_1 - \hat{\beta} + \hat{\beta}}{2})^2$ y por tanto,

$$M_2 = \frac{(2\hat{\beta} - 2M_1)^2}{12} + M_1^2 = \frac{4(\hat{\beta} - M_1)^2}{12} + M_1^2 = \frac{(\hat{\beta} - M_1)^2}{3} + M_1^2 \Rightarrow (\hat{\beta} - M_1)^2 = 3(M_2 - M_1^2) \Rightarrow$$

$\hat{\beta} = M_1 \pm \sqrt{3(M_2 - M_1^2)}$ y por tanto $\hat{\alpha} = M_1 \mp \sqrt{3(M_2 - M_1^2)}$. Como sabemos, en una distribución uniforme $\alpha < \beta$, por tanto:

$\hat{\alpha} = M_1 - \sqrt{3(M_2 - M_1^2)}$ **y** $\hat{\beta} = M_1 + \sqrt{3(M_2 - M_1^2)}$ **son los estimadores de momentos.**

Observación:

$(M_2 - M_1^2) = \dfrac{\sum_{j=1}^{n} X_j^2}{n} - \overline{X}^2 = \dfrac{\sum_{j=1}^{n} X_j^2 - n\overline{X}^2}{n} = \dfrac{\sum_{j=1}^{n}(X_j - \overline{X})^2}{n} \geq 0$, de modo que el radicando en la raíz de $\sqrt{3(M_2 - M_1^2)}$ nunca será negativo.

5.6.2 Método de Máxima Verosimilitud

El método de momentos (que es el más antiguo) tiene la debilidad de no tomar en cuenta la estructura del espacio paramétrico Θ; en particular, si Θ está asociado al espacio de información X, pueden presentarse casos como el del ejemplo (1), en que el método proporciona valores estimados absurdos. Un método que resuelve este problema es el de Máxima Verosimilitud.

5.6.2.1 Definiciones Básicas

Sea X una v.a. con función de densidad o probabilidad $f_X(x;\theta)$ donde $\theta \in \Theta$, es parámetro o vector de parámetros, de valor desconocido el valor de θ, la forma de $f_X(x;\theta)$ es conocida. Sea $(X_1, X_2, ..., X_n)$ una m.a.

Definición (Función de Verosimilitud).

Denotada $L(\theta)$, es la distribución conjunta de la muestra, vista como función del parámetro θ. Formalmente:

$$L(\theta) = \prod_{j=1}^{n} f_{X\,j}(x_j;\theta) \quad \theta \in \Theta$$

Observación:

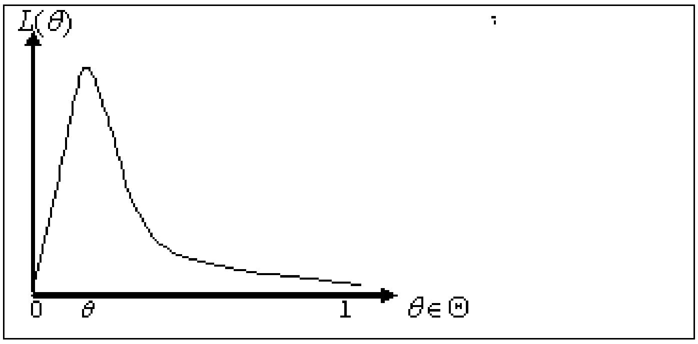
Dada la m.a., $L(\theta)$ cambia de valores conforme θ se desplaza por Θ, de modo que en algunos casos, la probabilidad asociada a la m.a. a través de $L(\theta)$ aumenta y en otros disminuye. Es natural esperar que θ sea tal que la m.a. dada sea factible o en otras palabras, sea la m.a. más probable.

Ejemplo

Si X ~ G(p=θ;x) $\Rightarrow$ la función de probabilidad de X es $P_X(x;\theta) = \theta(1-\theta)^{x-1}$ y por tanto la Función de Verosimilitud de θ es:

$$L(\theta) = \prod_{j=1}^{n} \theta(1-\theta)^{x_j-1} = \theta^n (1-\theta)^{\sum_{j=1}^{n}(x_j-1)} = \theta^n (1-\theta)^{n(\overline{X}-1)} \quad 0 < \theta < 1$$

y la gráfica de esta función $L(\theta)$ de θ es:



Para $\theta = 0$ y $\theta = 1$, ocurre $L(\theta) = 0$ y para valores intermedios la función de verosimilitud crece primero y decrece después. O sea, la probabilidad de la muestra depende del valor que demos al parámetro. Algunos valores de θ son "más probables" o "verosímiles" que otros. Naturalmente, el punto donde $L(\theta)$ tiene un máximo puede ser visto como "el más probable" o "más verosímil" y es un candidato a valor estimado.

Definición (Estimador de Máxima Verosimilitud).

El estimador de θ obtenido por el Método de Máxima Verosimilitud, es el estimador $\hat{\theta}$ que maximiza $L(\theta)$.

Formalmente, $\hat{\theta}$ es estimador Máximo verosímil de θ *si* $\hat{\theta}$ es solución al problema:

$$\underset{\theta}{Máx}\, L(\theta) \;\; s.a.\; \theta \in \Theta$$

Observaciones:

(1) El método de máxima verosimilitud escoge como valor de θ (o punto del espacio paramétrico Θ) aquél que maximiza la probabilidad asociada a la m.a. $(X_1,X_2,...,X_n)$. La idea es hacer que coincidan "lo probable" y "lo posible".

(2) La maximización de $L(\theta)$ puede hacerse con los métodos clásicos (vía diferenciación sin o con restricciones) o con métodos heurísticos (fundamentalmente gráficos).

(3) Como $L(\theta)$ y ln($L(\theta)$) tienen los mismos puntos críticos, pero ln($L(\theta)$) suele tener una estructura más simple, **es común obtener el estimador MV de θ maximizando ln($L(\theta)$) en lugar de** $L(\theta)$.

Ejemplos:

(1) Sea X ~ G(p=θ;x). Hallar el estimador máximo verosímil de θ.

Solución:

- X ~ G(p=θ;x) $\Rightarrow P_X(x;\theta) = \theta(1-\theta)^{x-1} \quad x = 1,2,3,...$
- Dada una m.a. ($X_1,X_2,...,X_n$) $\Rightarrow P_{X_j}(x_j;\theta) = P_X(x_j;\theta) = \theta(1-\theta)^{x-1} \quad x_j = 1,2,3,...$
- $L(\theta) = \prod_{j=1}^{n} P_{X_j}(x_j;\theta) = \prod_{j=1}^{n} \theta(1-\theta)^{x_j-1} = \theta^n(1-\theta)^{\sum_{j=1}^{n}(x_j-1)} = \theta^n(1-\theta)^{n(\overline{X}-1)} \quad 0 < \theta < 1$
- Como $L(\theta) = \theta^n(1-\theta)^{n(\overline{X}-1)}$ es diferenciable, podemos maximizarla usando derivadas ordinarias:

$$\frac{\partial L(\theta)}{\partial \theta} = n\theta^{(n-1)}(1-\theta)^{n(\overline{X}-1)} - n(\overline{X}-1)\theta^n(1-\theta)^{n(\overline{X}-1)-1}$$

- Condición de primer orden: $\frac{\partial L(\theta)}{\partial \theta} = 0 \Leftrightarrow n\theta^{(n-1)}(1-\theta)^{n(\overline{X}-1)} - n(\overline{X}-1)\theta^n(1-\theta)^{n(\overline{X}-1)-1} = 0 \Leftrightarrow$

  $n(1-\theta) - n(\overline{X}-1)\theta = 0 \Leftrightarrow \theta = \frac{1}{\overline{X}}$

- Asumiendo que la condición de segundo orden se cumple, tenemos que $L(\theta)$ es máxima si $\theta = \frac{1}{\overline{X}}$. Para resaltar que este no es el valor real de θ sino el "valor más probable o verosímil", escribimos $\hat{\theta} = \frac{1}{\overline{X}}$ **y decimos que $\hat{\theta}$ es el Estimador de Máxima Verosimilitud de θ.**

(2) Sea X ~ N(μ,σ^2). Hallar los estimadores MV de μ y σ^2.

Solución:

- Aquí $f_X(x;\mu,o^2) = \frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma}$ para $-\infty < x < +\infty$; $\mu \varepsilon R$ y $\sigma^2 > 0$
- Dada la m.a. ($X_1,X_2,...,X_n$) $\Rightarrow f_{X_j}(x_j;\mu,o^2) = \frac{e^{-(x_j-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma}$ y la función de verosimilitud de μ y σ^2 es:

  $L(\mu,\sigma^2) = \prod_{j=1}^{n} f_{X_j}(x_j;\mu,o^2) = \prod_{j=1}^{n}\left(\frac{e^{-(x_j-\mu)^2/2\sigma^2}}{\sqrt{2\pi}\sigma}\right) = \frac{e^{-\sum_{j=1}^{n}(x_j-\mu)^2/2\sigma^2}}{(\sqrt{2\pi})^n\sigma^n} = \frac{e^{-\sum_{j=1}^{n}(x_j-\mu)^2/2\sigma^2}}{(\sqrt{2\pi})^n(\sigma^2)^{n/2}}$ que es una función diferenciable de μ y σ^2. Para maximizar $L(\mu,\sigma^2)$, maximicemos $\ln[L(\mu,\sigma^2)]$ que es equivalente.

- $\ln[L(\mu,\sigma^2)] = \ln\left[\frac{e^{-\sum_{j=1}^{n}(x_j-\mu)^2/2\sigma^2}}{(\sqrt{2\pi})^n(\sigma^2)^{n/2}}\right] = -\sum_{j=1}^{n}(x_j-\mu)^2/2\sigma^2 - n\ln(\sqrt{2\pi}) - \frac{n}{2}\ln(\sigma^2)$ y aplicando condiciones de primer orden $\nabla\ln(L(\mu,\sigma^2) = 0$:

$$\frac{\partial}{\partial\mu}\ \ln[L(\mu,\sigma^2)]=0 \Rightarrow -\frac{-2\sum_{j=1}^{n}(x_j-\mu)}{2\sigma^2}=0 \qquad \text{(a)}$$

$$\frac{\partial}{\partial\sigma^2}\ \ln[L(\mu,\sigma^2)]=0 \Rightarrow -\frac{-\sum_{j=1}^{n}(x_j-\mu)^2}{2(\sigma^2)^2}-\frac{n}{2}\frac{1}{\sigma^2}=0 \qquad \text{(b)}$$

- Resolviendo (a) y (b):

De (a) $\Rightarrow \sum_{j=1}^{n}(x_j-\mu)=0 \Rightarrow \sum_{j=1}^{n}x_j-n\mu=0 \Rightarrow \mu=(\sum_{j=1}^{n}x_j)/n=\overline{X}$ y reemplazando en (b)

$-\frac{-\sum_{j=1}^{n}(x_j-\overline{X})^2}{2(\sigma^2)^2}-\frac{n}{2}\frac{1}{\sigma^2}=0 \Rightarrow \frac{-\sum_{j=1}^{n}(x_j-\overline{X})^2}{\sigma^2}+n=0 \Rightarrow \sigma^2=\frac{\sum_{j=1}^{n}(x_j-\overline{X})^2}{n}$. Asumiendo las condiciones de 2do. orden, los valores de μ y σ^2 que máximizan $L(\mu,\sigma^2)$ son

$$\mu=\overline{X} \text{ y } \sigma^2=\frac{\sum_{j=1}^{n}(x_j-\overline{X})^2}{n}$$

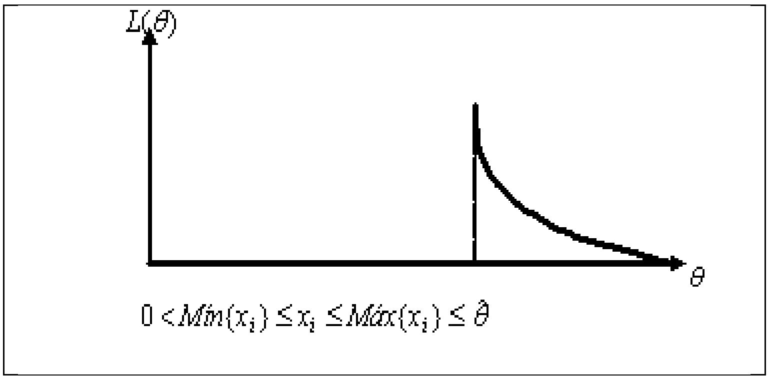
Como los valores anteriores no son los valores reales de μ y σ^2, sino los "más probables" o "más verosímiles", escribimos:

$\hat{\mu}=\overline{X}\ ;\ \hat{\sigma}^2=\frac{\sum_{j=1}^{n}(x_j-\overline{X})^2}{n}$ **y decimos que $\hat{\mu}$ y $\hat{\sigma}^2$ son los estimadores MV de μ y σ^2**

(3) Sea X ~ U(0,θ). Hallar el estimador MV de θ

Solución:

- Como $f_X(x,\theta)=\frac{1}{\theta}\quad 0<x\le\theta \Rightarrow L(\theta)=(\frac{1}{\theta})^n=\theta^{-n}$
- Derivando: $\frac{\partial}{\partial\theta}L(\theta)=-n\theta^{-n-1}\neq 0\ \forall\theta$ ¿No existe punto crítico? ¿No existe el estimador MV? No, no es este el caso. Lo que sucede es que hemos planteado la función de verosimilitud descuidadamente: $f_X(x,\theta)=\frac{1}{\theta}$ $0<x\le\theta$ implica que $f_X(x,\theta)=0$ si $x\notin\ ]0,\theta]$. De modo que lo correcto es escribir $L(\theta)$ como
  $L(\theta)=\begin{cases}\theta^{-n} & si \quad 0<x_j\le\theta \\ 0 & en\ otro\ caso\end{cases}=\begin{cases}\theta^{-n} & 0<Mín\{X_j\}\le X_j\le Máx\{X_j\}\le\theta \\ 0 & en\ otro\ caso\end{cases}$
  y se ve que aunque en la función de verosimilitud no aparece explícitamente la m.a. ($X_1,X_2,...,X_n$), ésta figura acotando inferiormente a θ.
- De lo anterior se deduce que $\frac{\partial}{\partial\theta}L(\theta)=-n\theta^{-n-1}<0$ si $0<x_j\le Máx\{x_j\}\le\theta$ y en otro caso la derivada es cero o no existe. Sin embargo, es fácil ver que $L(\theta)$ es función no creciente de θ, por lo que su mayor valor se obtiene dando a θ el menor valor posible. Como $0<x_j\le Máx\{x_j\}\le\theta$, entonces el menor valor posible para θ es $\theta=Máx\{x_j\} \Rightarrow \hat{\theta}=Máx\{x_j\}$ es el estimador MV de θ. El gráfico anexo muestra el razonamiento. Nótese que en este caso, hemos usado un método gráfico y heurístico para maximizar la función de verosimilitud, pues el método clásico de derivadas no funciona.



Observación:
Lo que hace a este ejemplo diferente de los anteriores, es que el parámetro θ y el rango de R_X están asociados, o sea el espacio paramétrico Θ y el espacio de información X no son independientes, a diferencia de lo que sucedía en los otros ejemplos, y al maximizar $L(\theta)$ esta asociación se convierte en una restricción.

En general, el estimador $\hat{\theta}$ es solución al problema $\underset{\theta}{Máx}\, L(\theta)$ s.a $\theta \in \Theta$, pero la condición $\theta \in \Theta$ no es una restricción real si Θ y X son independientes (en el sentido matemático, no en el probabilístico). **Cuando Θ y X están asociados, hallar el estimador $\hat{\theta}$ implica respetar y tomar en cuenta la restricción $\theta \in \Theta$**. El caso de la distribución uniforme muestra lo dicho.

(4) Sea (X,Y) v.a. con función de densidad conjunta

$$f_{XY}(\theta; x, y) = \theta^2 (1-\theta)^{x+y-2} \quad x = 1,2,3,.. \quad y = 1,2,3,... \quad 0 < \theta < 1$$

Hallar el estimador MV de θ

Solución:

- Dada una m.a. de n parejas $(X_1,Y_1),(X_2,Y_2),...,(X_n,Y_n)$, la función de verosimilitud es:

$$L(\theta) = \prod_{j=1}^{n} f_{XY}(\theta; x_j, y_j) = \prod_{j=1}^{n} \theta^2 (1-\theta)^{x_j+y_j-2} = \theta^{2n}(1-\theta)^{\sum_{j=1}^{n}(x_j+y_j-2)} = \theta^{2n}(1-\theta)^{n(\bar{X}+\bar{Y}-2)}$$

- Como θ y el rango de R_X no están asociados y $L(\theta)$ es diferenciable, apliquemos optimización clásica a ln($L(\theta)$) para simplificar:
  $\ln[L(\theta)] = 2n\ln(\theta) + n(\bar{X}+\bar{Y}-2)\ln(1-\theta)$ y derivando con respecto a θ obtenemos
  $\frac{\partial}{\partial\theta}\ln[L(\theta)] = \frac{2n}{\theta} - \frac{n(\bar{X}+\bar{Y}-2)}{(1-\theta)} = 0 \Rightarrow 2n - 2n\theta = n\theta(\bar{X}+\bar{Y}) - 2n\theta \Rightarrow \theta = \frac{2}{(\bar{X}+\bar{Y})}$ es el punto crítico de $L(\theta)$.
- Luego, generalizando, y previa verificación del máximo, se tiene que $\hat{\theta} = \frac{2}{(\bar{X}+\bar{Y})}$ es el estimador MV de θ.

5.6.3 Método de Mínimos Cuadrados

Este método se aplica cuando lo que se tiene es una relación entre dos variables y los parámetros que definen la relación deben ser estimados. Como los modelos de la economía aplicada se formulan, por lo general, como ecuaciones, esto hace del método una herramienta ideal para "ajustar" los modelos a datos empíricos y convierte a los Mínimos Cuadrados en el método más usado de la Econometría básica.

5.6.3.1 El problema y sus elementos

Supongamos a Y, X y ε tales que X es variable observable no aleatoria (lo que a veces se llama una "variable matemática"), Y es variable aleatoria observable y ε es variable aleatoria no observable. Supongamos que estas variables están relacionadas mediante la ecuación $Y = \varphi(X;\theta) + \varepsilon$, donde $\varphi(X;\theta)$ es función bien especificada (de forma conocida) y θ es un parámetro o vector de parámetros por estimar. **La función $\varphi(X;\theta)$ puede considerarse una función de "enlace" entre la componente aleatoria Y del modelo y el residuo no sistematico y aleatorio ε**. También se dice que $\varphi(X;\theta)$ es la "componente sistemática" del modelo y ε la "componente aleatoria" del mismo.

Dada una m.a. de n parejas de observaciones $(X_1,Y_1), (X_2,Y_2),\cdots,(X_n,Y_n)$ que satisfacen la relación $Y_j = \varphi(X_j;\theta) + \varepsilon_j \quad j = 1,2,3,...,n$. El problema es estimar θ

Observación:

- Si recordamos la definición de media condicional de Y dado X=x, $E[Y \mid X = x]$, sabremos que por lo general $E[Y \mid X = x]$ es una función explícita de x que se puede escribir como $E[Y \mid X = x] = \varphi(\theta;x)$, donde θ es un parámetro característico de la forma funcional que tome φ
- Si definimos $\varepsilon := Y - E[Y \mid X = x]$ no es difícil ver que $E[\varepsilon] = 0$, y podemos escribir tautológicamente $Y = E[Y \mid X = x] + \varepsilon$ o equivalentemente $Y = \varphi(x;\theta) + \varepsilon$ y si simplificamos la notación, llegamos a $Y = \varphi(X;\theta) + \varepsilon$ que es el "modelo" del problema de estimación planteado.

5.6.3.2 Los Supuestos Clásicos y el Método de Mínimos Cuadrados

Para trabajar con comodidad, debemos imponer algunas condiciones que sean simplificadoras pero plausibles. El conjunto mínimo de supuestos que con más frecuencia se emplea se conoce como de "supuestos clásicos" que, como ya se dijo, tienen por finalidad obtener soluciones iniciales al problema de estimación. En cursos más avanzados, estos supuestos se van relajando poco a poco para tratar modelos más generales y flexibles.

Supuestos clásicos

Dado el modelo $Y = \varphi(X;\theta) + \varepsilon$ y una m.a. de n parejas de observaciones $(X_1,Y_1),\ (X_2,Y_2),\cdots,(X_n,Y_n)$ que satisfacen la correspondiente relación $Y_j = \varphi(X_j;\theta) + \varepsilon_j$, asumiremos que:

(1) $E(\varepsilon_j) = 0 \quad \forall j$

(2) $V(\varepsilon_j) = \sigma^2 \quad \forall j$ **(Homogeneidad de varianzas u Homocedasticidad)**

(3) $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j$ **(No autocorrelación)**

(4) $\varphi(X;\theta)$ es una función sin error de especificación (de forma conocida y correcta)

(5) X_j es variable no aleatoria o siéndolo tiene sus valores ya dados.

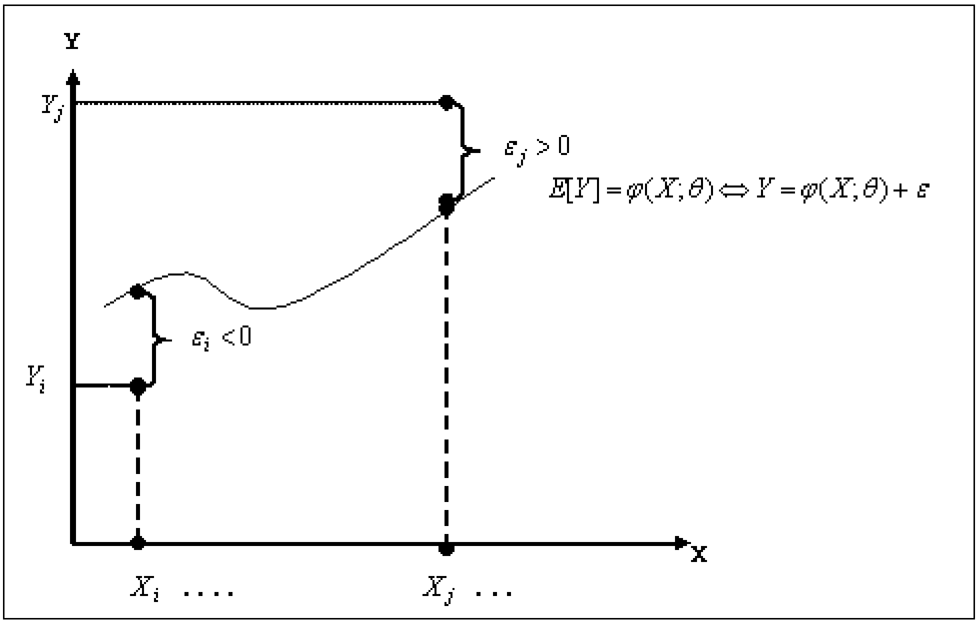
Definición (Estimador de Mínimos Cuadrados Ordinarios MCO o LS)

Bajo los supuestos (1) a (5), definimos el estimador $\hat{\theta}$ obtenido mediante el Método de Mínimos Cuadrados Ordinarios (MCO) como el valor $\hat{\theta}$ que es solución al problema:

$$\underset{\theta}{Mín}\, Q(\theta) = \sum_{j=1}^{n} \left[Y_j - \varphi(X_j;\theta) \right]^2 \quad \text{s.a } \theta \in \Theta$$

Nota: La lógica del método

Como $Y_j = \varphi(X_j;\theta) + \varepsilon_j \quad \forall j$, bajo los supuestos clásicos y tomando valor esperado tenemos que $E[Y] = E[\varphi(X;\theta)] + E[\varepsilon] = \varphi(X;\theta)$, es decir en un plano XY, la ecuación $E[Y] = \varphi(X;\theta)$ es una "curva" cuya posición depende de θ, pero que tiene la propiedad de pasar entre los distintos puntos muestrales (X_j, Y_j): algunas
veces pasará "debajo" de (X_j, Y_j) y entonces $\varepsilon_j < 0$, otras veces pasará "sobre" (X_j, Y_j) y entonces $\varepsilon_j > 0$, **pero "en promedio" pasa cerca de los puntos. Por tanto debemos tomar un valor de θ que haga pasar a $E[Y] = \varphi(X;\theta)$ lo más cerca posible de todos los puntos** (X_j, Y_j) y esto se obtiene tomado θ de modo que minimice las diferentes distancias $|Y_j - \varphi(X_j;\theta)|$. Una manera de hacer esto es minimizar los cuadrados de estas distancias y como queremos minimizar todas simultáneamente, **podemos tomar a θ de modo que $\sum_{j=1}^{n} \left[Y_j - \varphi(X_j;\theta) \right]^2$ sea mínima.**



Observaciones

- En Economía, la variable ε representa un elemento de ajuste que se incorpora a un modelo económico $E[Y] = \varphi(X;\theta)$ y que representa los efectos fortuitos de otras fuerzas no contempladas en el modelo.
- La ecuación $E[Y] = \varphi(X;\theta)$ se conoce como Modelo de Regresión de Y sobre X, por analogía con la esperanza condicional de Y dado X, y es la manera que tiene el economista de representar y explicar el efecto que tiene una variable X (llamada 'exógena' o 'independiente') sobre la tendencia promedio de otra variable Y (llamada 'endógena' o 'dependiente').

Ejemplos

(1) Modelo de Regresión Lineal sin intercepto
Sea el modelo $Y_j = \varphi(X_j;\theta) = \theta X_j + \varepsilon_j$. Estimar θ usando Mínimos Cuadrados Ordinarios.

Solución

- De $Y_j = \varphi(X_j;\theta) = \theta X_j + \varepsilon_j \;\Rightarrow\; \varepsilon_j = [Y_j - \varphi(X_j;\theta])=[Y_j - \theta X_j]$ y **la función objetivo por minimizar es** $Q(\theta) = \sum_{j=1}^{n}\left[Y_j - \theta X_j\right]^2$
- **Aplicando derivadas:**

$$\frac{\partial Q(\theta)}{\partial\theta} = 0 \Leftrightarrow \frac{\partial}{\partial\theta}\sum_{j=1}^{n}\left[Y_j - \theta X_j\right]^2 = 0 \Leftrightarrow -2\sum_{j=1}^{n}\left[Y_j - \theta X_j\right]X_j = 0 \Leftrightarrow \sum_{j=1}^{n}X_jY_j - \theta\sum_{j=1}^{n}X_j^2 = 0 \Rightarrow$$

$\theta = \dfrac{\sum_{j=1}^{n}X_jY_j}{\sum_{j=1}^{n}X_j^2}$ que es el punto crítico de la función objetivo. Verificando que las condiciones de 2do. orden se cumplen, este será el valor de θ que minimiza la suma de cuadrados de los errores. Como no es el valor real, sino sólo una aproximación óptima, lo denotaremos $\hat{\theta}$ y diremos que:

$\hat{\theta} = \dfrac{\sum_{j=1}^{n}X_jY_j}{\sum_{j=1}^{n}X_j^2}$ es el estimador de $\hat{\theta}$ obtenido mediante Mínimos Cuadrados o en la jerga econométrica, el estimador MCO de $\hat{\theta}$

(2) Modelo de Regresión Lineal con intercepto

Sea el modelo $Y_j = \varphi(X_j,\alpha,\beta) = \alpha + \beta X_j + \varepsilon_j$. Estimar α y β usando Mínimos Cuadrados Ordinarios.

Solución

- En este caso $\varepsilon_j = [Y_j - \alpha - \beta X_j]$ y **la función objetivo por minimizar es** $Q(\alpha,\beta) = \sum_{j=1}^{n}\left[Y_j - \alpha - \beta X_j\right]^2$
- **Aplicando derivadas:**

$$\nabla Q(\alpha,\beta) = 0 \Leftrightarrow \frac{\partial Q(\alpha,\beta)}{\partial\alpha} = 0 \text{ y } \frac{\partial Q(\alpha,\beta)}{\partial\beta} = 0 \Leftrightarrow$$

$$\frac{\partial Q(\alpha,\beta)}{\partial\alpha} = 0 \Rightarrow \frac{\partial}{\partial\alpha}\sum_{j=1}^{n}\left[Y_j - \alpha - \beta X_j\right]^2 = -2\sum_{j=1}^{n}\left[Y_j - \alpha - \beta X_j\right] = 0$$

$$\frac{\partial Q(\alpha,\beta)}{\partial\alpha} = 0 \Rightarrow \frac{\partial}{\partial\beta}\sum_{j=1}^{n}\left[Y_j - \alpha - \beta X_j\right]^2 = -2\sum_{j=1}^{n}\left[Y_j - \alpha - \beta X_j\right]X_j = 0$$

El sistema 2x2 resultante es:

$$\sum_{j=1}^{n} Y_j - \sum_{j=1}^{n} \alpha - \beta \sum_{j=1}^{n} X_j = 0 \Rightarrow n\overline{Y} - \alpha n - \beta n\overline{X} = 0 \Rightarrow \alpha n + \beta n\overline{X} = n\overline{Y}$$

$$\sum_{j=1}^{n} X_j Y_j - \sum_{j=1}^{n} \alpha X_j - \beta \sum_{j=1}^{n} X_j^2 = 0 \Rightarrow \alpha n\overline{X} + \beta \sum_{j=1}^{n} X_j^2 = \sum_{j=1}^{n} X_j Y_j$$

- Matricialmente el sistema anterior es: $\begin{pmatrix} n & n\overline{X} \\ n\overline{X} & \sum_{j=1}^{n} X_j^2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} n\overline{Y} \\ \sum_{j=1}^{n} X_j Y_j \end{pmatrix}$ que se puede resolver aplicando inversa o con la Regla de Cramer o por sustitución. Resolviendo y verificando el mínimo, tenemos que

$\hat{\beta} = \dfrac{\sum_{j=1}^{n} X_j Y_j - n\overline{X}\,\overline{Y}}{\sum_{j=1}^{n} X_j^2 - n\overline{X}^2}$ y $\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$ son los estimadores MCO de β y α respectivamente

Observación: (fórmulas alternativas para $\hat{\beta}$ y $\hat{\alpha}$)

- $$\hat{\beta} = \frac{\sum_{j=1}^{n} X_j Y_j - n\overline{X}\,\overline{Y}}{\sum_{j=1}^{n} X_j^2 - n\overline{X}^2} = \frac{\sum_{j=1}^{n} (X_j - \overline{X}) Y_j}{\sum_{j=1}^{n} (X_j - \overline{X})^2} = \frac{\sum_{j=1}^{n} (X_j - \overline{X})(Y_j - \overline{Y})}{\sum_{j=1}^{n} (X_j - \overline{X})^2} = \sum_{j=1}^{n} \left(\frac{(X_j - \overline{X})}{\sum_{j=1}^{n} (X_j - \overline{X})^2} \right) Y_j$$
- $$\hat{\alpha} = \frac{\overline{Y} \sum_{j=1}^{n} X^2 - \overline{X} \sum_{j=1}^{n} X_j Y_j}{\sum_{j=1}^{n} X_j^2 - n\overline{X}^2}$$
- La verificación se apoya en la identidad:

$$\sum_{j=1}^{n} (X_j - \overline{X})(Y_j - \overline{Y}) = \sum_{j=1}^{n} (X_j - \overline{X}) Y_j - \sum_{j=1}^{n} (X_j - \overline{X}) \overline{Y} = \sum_{j=1}^{n} (X_j - \overline{X}) Y_j - \overline{Y} \sum_{j=1}^{n} (X_j - \overline{X}) = \sum_{j=1}^{n} (X_j - \overline{X}) Y_j$$

- Estos estimadores MCO, bajo los supuestos clásicos son óptimos en el sentido que resultan insesgados y eficientes en relación a otros estimadores lineales e insesgados de β y α.

Proposición (Teorema de Gauss-Markov)
En el modelo $Y_j = \alpha + \beta X_j + \varepsilon_j$ y bajo los supuestos clásicos, los estimadores MCO $\hat{\alpha}$ y $\hat{\beta}$ son MELI de α y β respectivamente.

Demostración: (sólo para el caso de β)

- Primero recordemos los supuestos clásicos:

(1) $E(\varepsilon_j) = 0 \quad \forall j$

(2) $V(\varepsilon_j) = \sigma^2 \quad \forall j$ **(Homogeneidad de varianzas u Homocedasticidad)**

(3) $Cov(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j$ **(No autocorrelación)**

(4) $\varphi(X; \theta)$ es una función sin error de especificación (de forma conocida y correcta)

(5) X_j es variable no aleatoria o siéndolo tiene sus valores ya dados.

Dado el modelo $Y_j = \alpha + \beta X_j + \varepsilon_j$ y dada una m.a. de n parejas $(X_1,Y_1),(X_2,Y_2), ...,(X_n,Y_n)$, sea $\tilde{\beta}$ el MELI de β; aplicando las condiciones que definen al MELI tenemos:

- $\tilde{\beta}$ es función **lineal** de la muestra $\Rightarrow \tilde{\beta} = \sum_{j=1}^{n} c_j Y_j$
- $\tilde{\beta}$ debe ser **insesgado** $\Rightarrow E[\tilde{\beta}] = \beta \ \ \forall\beta \text{ y } \forall\alpha \Leftrightarrow \sum_{j=1}^{n} c_j E[Y_j] = \beta \ \ \forall\beta \text{ y } \forall\alpha$.

Como $E[Y_j] = E[\alpha + \beta X_j + \varepsilon_j] = E[\alpha] + E[\beta X_j] + E[\varepsilon_j] = \alpha + \beta X_j$ pues α y β son constantes, X_j es no aleatoria y $E[\varepsilon_j] = 0$ de acuerdo a los supuestos (1) y (5); por tanto $E[\tilde{\beta}] = \beta \ \ \forall\beta \text{ y } \forall\alpha \Leftrightarrow$ $\sum_{j=1}^{n} c_j [\alpha + \beta X_j] = \beta \ \ \forall\beta \text{ y } \forall\alpha \Leftrightarrow \alpha\sum_{j=1}^{n} c_j + \beta\sum_{j=1}^{n} c_j X_j = \beta \ \ \forall\beta \text{ y } \forall\alpha \Leftrightarrow \sum_{j=1}^{n} c_j = 0$ y $\sum_{j=1}^{n} c_j X_j = 1$ **que son las dos condiciones de insesgamiento**.

- $\tilde{\beta}$ debe ser de **varianza mínima** $\Rightarrow V[\tilde{\beta}] = V[\sum_{j=1}^{n} c_j Y_j] = \sum_{j=1}^{n} c_j^2 V[Y_j] = \sum_{j=1}^{n} c_j^2 \sigma^2 = Mín$

pues $V[\sum_{j=1}^{n} c_j Y_j] = \sum_{j=1}^{n} c_j^2 V[Y_j] + 2\sum_{i<j} c_i c_j Cov(Y_i, Y_j)$ y $V[Y_j] = V[\alpha + \beta X_j + \varepsilon_j] = V[\varepsilon_j] = \sigma^2$ ya que ($\alpha + \beta X_j$) es no aleatoria y $V[\varepsilon_j] = \sigma^2$ de acuerdo a los supuestos (5) y (2) respectivamente; y por otra parte $Cov(Y_i, Y_j) = E[(Y_i - E[Y_i])(Y_j - E[Y_j])] = E[\varepsilon_i \varepsilon_j] = Cov(\varepsilon_i, \varepsilon_j) = 0$ **de acuerdo al supuesto (3)**

- Hallar el MELI equivale entonces, a resolver el problema:

$$Mín \sum_{j=1}^{n} c_j^2 \ \text{ s.a. } \sum_{j=1}^{n} c_j = 0 \ \text{ y } \sum_{j=1}^{n} c_j X_j = 1$$

- Aplicando Lagrange:

$L(c_1, c_2, ...c_n, \lambda_1, \lambda_2) = \sum_{j=1}^{n} c_j^2 - \lambda_1 \sum_{j=1}^{n} c_j + \lambda_2 (1 - \sum_{j=1}^{n} c_j X_j)$ es el Lagrangiano y minimizando:

Condiciones de primer orden $\nabla L = 0$

$$\frac{\partial L}{\partial c_1} = 0 \Leftrightarrow 2c_1 - \lambda_1 - \lambda_2 X_1 = 0$$
$$\frac{\partial L}{\partial c_2} = 0 \Leftrightarrow 2c_2 - \lambda_1 - \lambda_2 X_2 = 0$$
:
:
$$\frac{\partial L}{\partial c_n} = 0 \Leftrightarrow 2c_n - \lambda_1 - \lambda_2 X_n = 0$$
$$\frac{\partial L}{\partial \lambda_1} = 0 \Leftrightarrow \sum_{j=1}^{n} c_j = 0$$
$$\frac{\partial L}{\partial \lambda_2} = 0 \Leftrightarrow 1 - \sum_{j=1}^{n} c_j X_j = 0$$

La estrategia para resolver el sistema es algo estándar: Despejar los coeficientes $\{c_j\}$ en términos de λ_1 y λ_2 y reemplazar en las ecuaciones derivadas de las restricciones:

$$c_1 = \frac{\lambda_1 + \lambda_2 X_1}{2} \; ; \; c_2 = \frac{\lambda_1 + \lambda_2 X_2}{2} \; ;\ldots c_j = \frac{\lambda_1 + \lambda_2 X_j}{2} \; ; \ldots ; \; c_n = \frac{\lambda_1 + \lambda_2 X_n}{2}$$

En la penúltima la ecuación:

$\sum_{j=1}^{n} c_j = 0 \Leftrightarrow \sum_{j=1}^{n} (\frac{\lambda_1 + \lambda_2 X_j}{2}) = 0 \Leftrightarrow \sum_{j=1}^{n} (\lambda_1 + \lambda_2 X_j) = 0 \Leftrightarrow n\lambda_1 + \lambda_2 n\overline{X} = 0 \Leftrightarrow \lambda_1 + \lambda_2 \overline{X} = 0$ y por tanto $\lambda_1 = -\lambda_2 \overline{X}$

En la última la ecuación:

$$\sum_{j=1}^{n} c_j X_j = 1 \Leftrightarrow \sum_{j=1}^{n} (\frac{\lambda_1 + \lambda_2 X_j}{2}) X_j = 1 \Leftrightarrow \sum_{j=1}^{n} (\lambda_1 X_j + \lambda_2 X_j^2) = 2 \Leftrightarrow \lambda_1 n\overline{X} + \lambda_2 \sum_{j=1}^{n} X_j^2 = 2 \Leftrightarrow$$

$$-\lambda_2 n\overline{X}^2 + \lambda_2 \sum_{j=1}^{n} X_j^2 = 2 \Rightarrow \lambda_2 = \frac{2}{\sum_{j=1}^{n} X_j^2 - n\overline{X}^2} \text{ y } \lambda_1 = -\frac{2\overline{X}}{\sum_{j=1}^{n} X_j^2 - n\overline{X}^2}$$

Reemplazando en $c_j = \frac{\lambda_1 + \lambda_2 X_j}{2} \Rightarrow c_j = \frac{-\frac{2\overline{X}}{\sum_{j=1}^{n} X_j^2 - n\overline{X}^2} + \frac{2}{\sum_{j=1}^{n} X_j^2 - n\overline{X}^2} X_j}{2} \Rightarrow c_j = \frac{X_j - \overline{X}}{\sum_{j=1}^{n} X_j^2 - n\overline{X}^2}$ y

asumiendo que se cumplen las condiciones de 2do. orden:

el MELI de β es $\tilde{\beta} = \sum_{j=1}^{n} c_j Y_j = \sum_{j=1}^{n} \left(\frac{X_j - \overline{X}}{\sum_{j=1}^{n} X_j^2 - n\overline{X}^2} \right) Y_j$

Observación: Mejor Estimador Lineal Afín

En la definición del MELI hemos usado la noción de función lineal como equivalente a combinación lineal, esto es una función de la forma $\sum_{j=1}^{n} c_j Y_j$. Pero podríamos entender función lineal como una función de la forma $c_0 + \sum_{j=1}^{n} c_j Y_j$ esto es, una "función lineal afín", en cuyo caso c_0 se convierte en una incógnita más del problema. Por lo general, al construir el MELI a partir de una función lineal afín, resulta que c_0=0 y por tanto es irrelevante la distinción entre función lineal $\sum_{j=1}^{n} c_j Y_j$ y función lineal afín $c_0 + \sum_{j=1}^{n} c_j Y_j$. Sin embargo, para ciertos modelos, pueden presentarse diferencias.

Ejemplo

En el modelo $Y_j = \alpha_0 + \beta X_j + \varepsilon_j$ donde α_0 es parámetro de valor conocido (y por tanto no hay necesidad de estimarlo) y bajo los supuestos clásicos:

- Si usamos como función lineal de la m.a. a $\sum_{j=1}^{n} c_j Y_j$, el MELI de β resulta $\tilde{\beta} =$

$$\sum_{j=1}^{n} c_j Y_j = \sum_{j=1}^{n} \left(\frac{X_j - \overline{X}}{\sum_{j=1}^{n} X_j^2 - n\overline{X}^2} \right) Y_j$$ **que es el mismo MELI del modelo** $Y_j = \alpha + \beta X_j + \varepsilon_j$ donde α es parámetro desconocido.

- Si usamos como función lineal de la m.a. a $c_0 + \sum_{j=1}^{n} c_j Y_j$, el MELI de β resulta $\tilde{\tilde{\beta}} =$

$$c_0 + \sum_{j=1}^{n} c_j Y_j = -\alpha_0 \frac{\sum_{j=1}^{n} X_j}{\sum_{j=1}^{n} X_j^2} + \frac{\sum_{j=1}^{n} X_j Y_j}{\sum_{j=1}^{n} X_j^2}$$ **que es diferente al obtenido antes. Además** $\tilde{\tilde{\beta}}$ **es más eficiente que** $\tilde{\beta}$

5.7 Propiedades adicionales

5.7.1 Invarianza

Sean $\theta \in R^p$ y $\gamma \in R^q$ $(q \leq p)$ parámetros tales que $\gamma = h(\theta)$. Sean $\tilde{\theta}$ y $\tilde{\gamma}$ estimadores de θ y γ respectivamente, obtenidos mediante un método $\mathcal{M}$. Diremos que el método $\mathcal{M}$ tiene la propiedad de invarianza si se cumple $\tilde{\gamma} = h(\tilde{\theta})$ $\tilde{\gamma} = h(\tilde{\theta})$

Propiedad 1.
En el contexto de la definición anterior, si la función $\gamma = h(\theta)$ además tiene inversa $\theta = h^{-1}(\gamma)$, los métodos de Momentos, Máxima Verosimilitud y Mínimos Cuadrados tienen la propiedad de invarianza.

Propiedad 2.
En general, el método de Máxima Verosimilitud tiene la propiedad de invarianza, esto es, si $\theta \in R^p$ y $\gamma \in R^q$ $(q \leq p)$ son parámetros tales que $\gamma = h(\theta)$ y $\tilde{\theta}_{MV}$ y $\tilde{\gamma}_{MV}$ son los estimadores máximo verosímiles de θ y γ respectivamente, entonces se cumple que $\tilde{\gamma}_{MV} = h(\tilde{\theta}_{MV})$

Ejemplo:
Si X~P(X; λ) y $\gamma = h(\lambda) = \lambda^2$, entonces como el estimador MV de λ es $\tilde{\lambda}_{MV} = \overline{X}$, aplicando la propiedad de invarianza podemos decir que $\tilde{\gamma}_{MV} = \overline{X}^2$

5.7.2 Distribución Asintótica del Estimador Máximo Verosímil

Proposición
Sea X v.a. con distribución $f_X(x;\theta)$ con $\theta \varepsilon \Theta$ y sea $\hat{\theta}$ el estimador Máximo verosímil de θ. Entonces, bajo las "condiciones de regularidad" R1 a R6,

R1: Si $\theta \neq \theta' \Rightarrow f_X(x;\theta) \neq f_X(x;\theta')$.
R2: R_X no depende de θ.
R3: El verdadero valor de θ es "punto interior" del espacio paramétrico Θ.
R4: $f_X(x;\theta)$ es función tres veces diferenciable de θ $(\exists \frac{d^{(3)} f_X(x;\theta)}{d\theta^{(3)}})$
R5: $\frac{d^{(3)} \int f_X(x;\theta)dx}{d\theta^{(3)}} = \int \frac{d^{(3)} f_X(x;\theta)}{d\theta^{(3)}} dx$

R6: $\frac{d^{(3)}f_X(x;\theta)}{d\theta^{(3)}} \leq M(x)$ donde $M(x)$ es una función definida en un entorno del verdadero valor de θ:

Si el tamaño de muestra *n* es grande, $\hat{\theta}$ *tiene distribución normal*: $\hat{\theta} \sim N(\theta, \sigma^2_{\hat{\theta}})$ donde

$$\sigma^2_{\hat{\theta}} = \left\{ nE[(\frac{\partial}{\partial\theta}\ln f_X(x;\theta))^2] \right\}^{-1}.$$

Ejemplo 1

Si $X \sim Exp(x;\theta)$, ya vimos que $\hat{\theta}_{MV} = 1/\overline{X}$.

- Como $\ln f_X(x;\theta) = \ln\theta - \theta x$, derivando tenemos: $\frac{\partial}{\partial\theta}\ln f_X(x;\theta) = \frac{1}{\theta} - x$;
- De lo anterior: $E[(\frac{\partial}{\partial\theta}\ln f_X(x;\theta))^2] = E[(\frac{1}{\theta} - X)^2] = \frac{1}{\theta^2}$ pues coincide con la definición de la varianza de una distribución exponencial.
- Por tanto, para n grande, la varianza asintótica de $\hat{\theta}_{MV} = 1/\overline{X}$ es $\sigma^2_{\hat{\theta}} = \left\{ nE[(\frac{\partial}{\partial\theta}\ln f_X(x;\theta))^2] \right\}^{-1} = \frac{\theta^2}{n}$
- Aplicando la propiedad, tenemos que: $\hat{\theta}_{MV} = 1/\overline{X} \sim N(\theta, \frac{\theta^2}{n})$

Ejemplo 2

En el ejercicio anterior, si se deseáramos que $P[|\hat{\theta}_{MV} - \theta| \leq 0.1\theta] = 0.95$, entonces asumiendo un n "grande" y aplicando la normalidad asintótica de $\hat{\theta}_{MV}$:

$$0.95 = P[|\hat{\theta}_{MV} - \theta| \leq 0.1\theta] = P[\frac{|\hat{\theta}_{MV} - \theta|}{\sigma_{\hat{\theta}}} \leq \frac{0.1\theta}{\sigma_{\hat{\theta}}}] = P[\frac{|\hat{\theta}_{MV} - \theta|}{\frac{\theta}{\sqrt{n}}} \leq \frac{0.1\theta}{\frac{\theta}{\sqrt{n}}}] = P[|Z| \leq 0.1\sqrt{n}]$$ y de la tabla Z

tenemos $0.1\sqrt{n} = 1.96$ y por tanto $n \cong 386$