ECONOMETRÍA II
Prof.: Begoña Álvarez
2007-2008

TEMA 1

INTRODUCCIÓN

Estimación por máxima verosimilitud y conceptos de teoría asintótica

1. ESTIMACIÓN POR MÁXIMA VEROSIMILITUD (*MAXIMUM LIKELIHOOD*)

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

La intuición del principio de MV es la siguiente:

1. Dado el supuesto sobre la distribución de las Yi, construimos la verosimilitud (probabilidad conjunta) de observar la muestra que tenemos. Esa función de probabilidad conjunta es una función de una serie de parámetros desconocidos.

2. Elegimos como estimadores MV aquellos valores de los parámetros desconocidos que hacen máxima esa verosimilitud.

1. ESTIMACIÓN POR MÁXIMA VEROSIMILITUD (*MAXIMUM LIKELIHOOD*)

Se trata de construir la función de probabilidad conjunta (o función de verosimilitud) de y1, y2, …. Yn. Suponemos que las observaciones son independientes y están idénticamente distribuidas (i.i.d.)

yi ~ f (yi; θ)

$$L(\theta) = f(y_1..y_n;\theta) \overset{\text{independence}}{=} \prod_{i=1}^{n} f(y_i;\theta)$$

- Si, para un determinado valor de θ, la verosimilitud es **PEQUEÑA**, es poco probable que ese θ sea el valor correcto que ha generado los datos que observamos.

- Si, para un determinado valor de θ, la verosimilitud es **GRANDE**, es bastante probable que ese θ sea el valor correcto que ha generado los datos que observamos.

1. ESTIMACIÓN POR MÁXIMA VEROSIMILITUD (*MAXIMUM LIKELIHOOD*)

Por tanto tenemos que elegir el θ que maximizar L(θ). Es decir, el estimador MV será el que satisfaga la condición de primer orden:

$$\left.\frac{\partial L(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}} = 0$$

O bien:

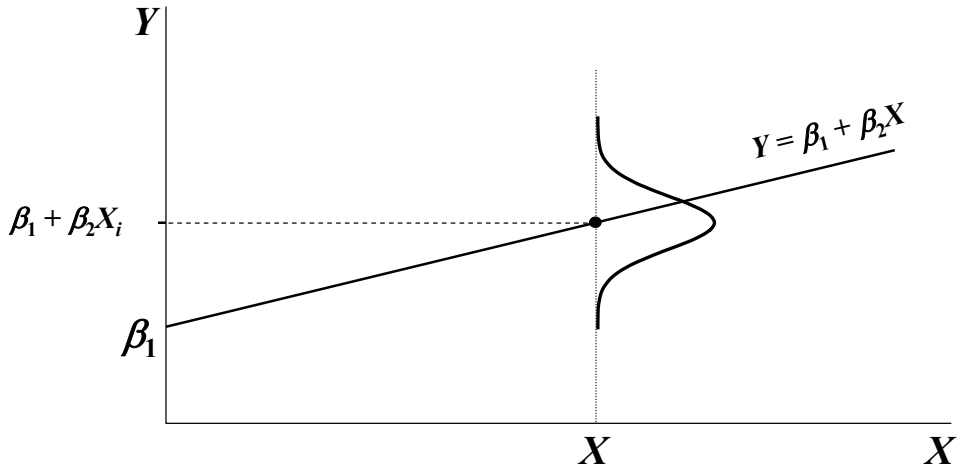
$$\left.\frac{\partial \ln L(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}} = 0$$

Generalmente, trabajamos con el logaritmo de la verosimilitud por razones prácticas.

EJERCICIO:
Estimador MV en el modelo de regresión lineal bajo el supuesto de normalidad.

$$y_i = x_i'\beta + u_i \qquad u_i \sim i.i.d.\ N(0, \sigma^2),\ i = 1, \ldots, n$$

$$f(y_i \mid x_i, \beta, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y_i - x_i'\beta}{\sigma}\right)^2}$$



EJERCICIO:
Estimador MV en el modelo de regresión lineal bajo el supuesto de normalidad.

1. Construimos la función de verosimilitud

$$L(\beta, \sigma^2; y) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - x_i'\beta)^2\right\}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right\}$$

2. Calculamos el logaritmo de la función de verosimilitud

$$\ln L(\beta, \sigma^2; y) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)$$

EJERCICIO:

Estimador MV en el modelo de regresión lineal bajo el supuesto de normalidad.

3. Condiciones de primer orden

$$\frac{\partial \ln L}{\partial \beta} = \frac{-1}{2\sigma^2}(-2X'y + 2X'X\beta)$$
$$\frac{\partial \ln L}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}(y - X\beta)'(y - X\beta)$$

4. Estimadores

$$\widehat{\beta}_{MLE} = (X'X)^{-1}X'y$$
$$\widehat{\sigma}^2_{MLE} = \frac{\hat{u}'\hat{u}}{n} \text{ with } \hat{u} = y - X\hat{\beta}_{MLE}$$

En este caso, $\boxed{\widehat{\beta}_{MLE} = \widehat{\beta}_{OLS}}$ pero $\boxed{\widehat{\sigma}^2_{MLE} \neq s^2}$

1. ESTIMACIÓN POR MÁXIMA VEROSIMILITUD (*MAXIMUM LIKELIHOOD*)

VENTAJA: El estimador MV (ML=maximum likelihood) tiene propiedades asintóticas óptimas entre todos los estimadores consistentes y normales asintóticamente.

DESVENTAJA: Podemos tener problemas graves si nos equivocamos en el supuesto de la distribución. En otras palabras, el estimador ML depende de forma importante de los supuestos sobre la distribución.

Otra desventaja: El estimador MV tiene propiedades mediocres en muestras pequeñas.

2. PROPIEDADES ASINTÓTICAS

- La idea es analizar el comportamiento aproximado del estimador cuando $n \to \infty$
- En particular, nos interesa saber si los estimadores son "consistentes" y cuál es su distribución asintótica.

Estas propiedades "sustituyen" de alguna forma a otras que se obtienen en muestras pequeñas pero que no se cumplen en muchos estimadores de MV. En particular, en muchos casos, no podemos demostrar "insesgadez" $(E(\widehat{\theta}) = \theta)$ ni podemos calcular la distribución exacta del estimador (ejemplo: $\widehat{\theta} \sim N(\theta, V)$)

2. PROPIEDADES ASINTÓTICAS

1. CONSISTENCIA

Definition

Consistency. An estimator $\widehat{\theta}$ of θ is consistent if, when the sample size increases, $\widehat{\theta}$ gets "closer" to θ.

Definition

Convergence in probability (consistency). Let X_n be a random variable indexed by the size of a sample. X_n converges in probability to X, $(\text{plim}(X_n) = X \text{ or } X_n \xrightarrow{p} X)$ if

$$\lim_{n \to \infty} P\left(|X_n - X| > \varepsilon\right) = 0 \text{ for any positive } \varepsilon.$$

2. PROPIEDADES ASINTÓTICAS

1. CONSISTENCIA

1º PROCEDIMIENTO PARA DEMOSTRAR CONSISTENCIA

Sufficient condition for consistency (not necessary)

If $\lim_{n\to\infty} E(\widehat{\theta}) = \theta$ and $\lim_{n\to\infty} Var(\widehat{\theta}) = 0$

then $\operatorname{plim}(\widehat{\theta}) = \theta$ (or $\widehat{\theta} \xrightarrow{p} \theta$)

Se trata de demostrar la **Convergencia en Media Cuadrática.**

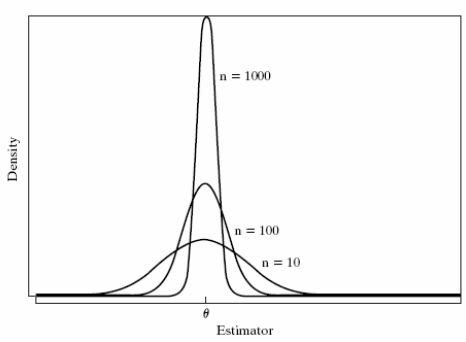


FIGURE D.1 Quadratic Convergence to a Constant, θ.

2. PROPIEDADES ASINTÓTICAS

1. CONSISTENCIA

2º PROCEDIMIENTO PARA DEMOSTRAR CONSISTENCIA

- To prove consistency **Laws of Large Numbers** are helpful
  - Suppose that X_i, $i = 1, n$ is an i.i.d. sequence of random variables with $E|X_i| < \infty$. Then by the (weak) law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} E(X_i).$$

- The probability limits operator exhibits some nice *intuitive properties*: If X_n and Y_n are random variables with $\operatorname{plim} X_n = a$ and $\operatorname{plim} Y_n = b$ then
  - $\operatorname{plim}(X_n + Y_n) = a + b$
  - $\operatorname{plim}(X_n Y_n) = ab$
  - $\operatorname{plim}(X_n / Y_n) = a/b$ provided $b \neq 0$

Theorem

Slutsky Theorem. *For a continuous function $g(X_n)$ that is not a function of n*

$$plim\ g(X_n) = g(plim\ X_n).$$

2. PROPIEDADES ASINTÓTICAS

1. CONSISTENCIA

Los estimadores pueden ser inconsistentes por dos razones:

(1) Convergen a una constante que no coincide con el parámetro que pretendemos estimar. Es decir, son estimadores consistentes de otro parámetro, pero no del que nos interesa.

(1) No convergen a una constante. En ese caso no son estimadores consistentes de nada.

EJERCICIO:

Consistencia del estimador OLS /MV en el modelo de regresión lineal

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sum_{i=1}^{n}(x_i-\overline{x})^2} = \beta + \frac{\sum_{i=1}^{n}(x_i-\overline{x})(\varepsilon_i-\overline{\varepsilon})}{\sum_{i=1}^{n}(x_i-\overline{x})^2}.$$

1º PROCEDIMIENTO

- $E(\widehat{\beta}) = \beta$
- $Var(\widehat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2} = \frac{\frac{1}{n}\sigma^2}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2} \to 0 \text{ as } n \to \infty.$

2º PROCEDIMIENTO

- $\text{plim}(\widehat{\beta}) = \text{plim}(\beta + \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(\varepsilon_i-\overline{\varepsilon})}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2})$
- $\underset{\text{Slutsky Theorem}}{=} \beta + \frac{\text{plim}\,\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})(\varepsilon_i-\overline{\varepsilon})}{\text{plim}\,\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2}$
- $\underset{\text{Law of Large numbers}}{=} \beta + \frac{Cov(x_i,\varepsilon_i)}{Var(x_i)} = \beta$

2. DISTRIBUCIÓN ASINTÓTICA

Cuando desconocemos la distribución exacta de un estimador, podemos preguntarnos si en grandes muestras el estimador sigue alguna distribución conocida. Esto nos permitirá realizar inferencia estadística (contrastes de hipótesis) cuyos resultados serán válidos en muestras grandes.

Definition

(Convergence in Distribution): The sequence Z_n with cumulative distribution functions $F_{Z_n}(z)$, converges in distribution to a random variable Z with cumulative distribution function $F_Z(z)$ if $\lim_{n\to\infty} |F_{Z_n}(z) - F_Z(z)| = 0$, at all points of continuity of $F_Z(z)$.

La intuición de la convergencia en distribución es que la distribución de Z_n se va pareciendo cada vez más a la distribución de Z conforme aumenta el tamaño muestral

2. PROPIEDADES ASINTÓTICAS

2. DISTRIBUCIÓN ASINTÓTICA

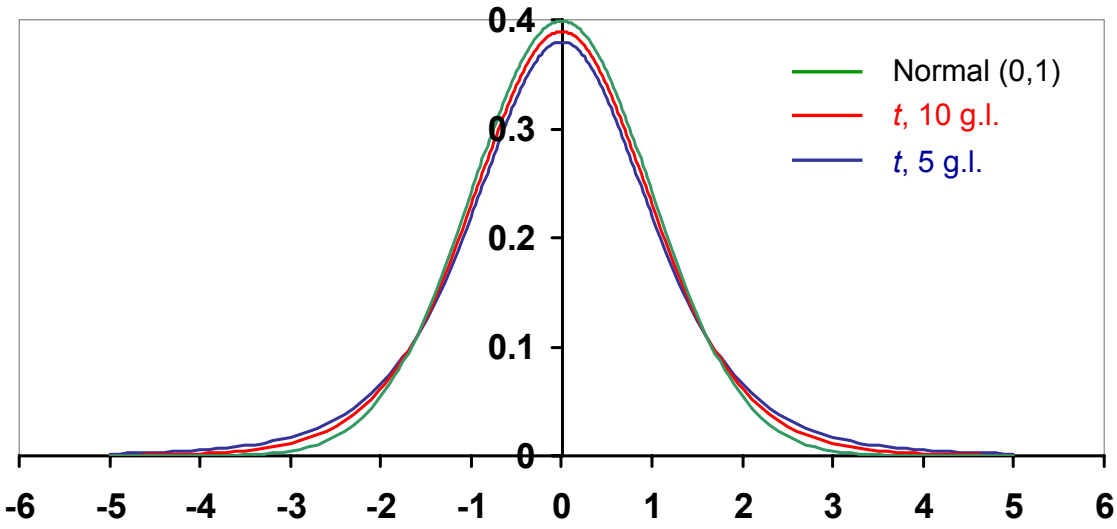
Ejemplo:

El estadístico t tiene una distribución t de Student con n-k grados de libertad. Pero, conforme $n \to \infty$ se comporta como una distribución Normal estándar. Esta es, por tanto, su distribución asintótica

Estadístico t $\sim$ t-Student

Estadístico t $\overset{a}{\sim}$ N(0,1) *o bien* Estadístico t $\overset{d}{\to}$ N(0,1)

2. DISTRIBUCIÓN ASINTÓTICA



2. PROPIEDADES ASINTÓTICAS

2. DISTRIBUCIÓN ASINTÓTICA

Una herramienta útil para derivar distribuciones asintóticas es el **Teorema Central del Límite**

If $X_1, \ldots, X_n$ represents a i.i.d. **random** sample from any probability distribution with **finite mean** μ **and finite variance** σ^2 then

$$\sqrt{n}(\overline{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$$

"$\sqrt{n}(\overline{X} - \mu)$ has a $N(0, \sigma^2)$ limiting distribution"

Igual que ocurre con la Ley de los Grandes Números, existen diferentes Teoremas Centrales del Límite cuando las Xi no son i.i.d. (se suelen exigir condiciones de diferentes para que se cumpla).

2. PROPIEDADES ASINTÓTICAS

2. DISTRIBUCIÓN ASINTÓTICA

¿Por qué es útil el Teorema Central del Límite?

Porque nos ayuda a demostrar la validez de los contrastes de hipótesis en muestras grandes (basados en los estadísticos de contraste que conocemos) incluso si desconocemos cuál es la verdadera distribución de los términos de perturbación del modelo.

En el caso del estimador MCO (*OLS*):

Un Teorema Central del Límite nos dice que aunque las perturbaciones aleatorias o términos de error del modelo de regresión no sigan una distribución Normal, si tienen media 0 y varianza finita e igual a σ^2

- $\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2(\text{plim}\,\frac{1}{n}X'X)^{-1})$, or
- $\widehat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$ asymptotically

3. PROPIEDADES DEL ESTIMADOR MÁXIMO VEROSÍMIL (*MAXIMUM LIKELIHOOD ESTIMATOR*)

Las principales propiedades el estimador MV son propiedades asintóticas (o en muestras grandes). Se cumplen bajo condiciones bastante generales (condiciones de regularidad).

- Consistencia
- Distribución asintótica Normal
- Eficiencia asintótica
- Invarianza

3. PROPIEDADES DEL ESTIMADOR MÁXIMO VEROSÍMIL (*MAXIMUM LIKELIHOOD ESTIMATOR*)

- Consistency
  - As our sample size increases $\widehat{\theta}_{MLE}$ gets 'closer' to θ. It may, however, be biased. For example, the MLE of σ^2, $\widehat{\sigma}^2_{MLE} = \frac{\widehat{u}'\widehat{u}}{n}$, is biased.

- Asymptotic Normality
  - $\sqrt{n}(\widehat{\theta}_{MLE} - \theta) \xrightarrow{d} N\left(0, \left(\lim_{n\to\infty} \frac{1}{n} I(\theta)\right)^{-1}\right)$, where $I(\theta)$ is the Information matrix, which is defined in two equivalent ways
    $$I(\theta) = -E\left(\frac{\partial^2 \ln L(\theta; y)}{\partial\theta\partial\theta'}\right) = E\left(\frac{\partial \ln L(\theta; y)}{\partial\theta}\frac{\partial \ln L(\theta; y)}{\partial\theta'}\right)$$
    - We also write: $\widehat{\theta}_{MLE} \sim N(\theta, I(\theta)^{-1})$ asymptotically

3. PROPIEDADES DEL ESTIMADOR MÁXIMO VEROSÍMIL (*MAXIMUM LIKELIHOOD ESTIMATOR*)

- Asymptotically Efficient
  - The inverse of the information matrix, $I(\theta)^{-1}$, provides a lower bound on the asymptotic covariance matrix for any consistent asymptotically normal estimator for θ.
    - This bound is often referred to as the **Cramèr-Rao lower bound**.
    - No other consistent and asymptotically normal distributed estimator has a smaller asymptotic covariance matrix.

- **Invariance**
  - E.g., if $\widehat{\theta}_{MLE}$ is the MLE of θ and if $g(\theta)$ is a continuous function of θ, the MLE of $g(\theta)$ is $g(\widehat{\theta}_{MLE})$. This invariance principle implies that we are free to reparameterize a likelihood function in any way we like, which may simplify estimation. E.g., MLE estimate of β^2 is simply $\widehat{\beta}^2_{MLE}$

4. CONCEPTOS HABITUALES EN ESTIMACIÓN ML

Definition

(Hessian): The Hessian matrix H is the $(k \times k)$ matrix of second derivatives of $\ln L(\theta; y)$ with respect to θ, $H(\theta) = \dfrac{\partial^2 \ln L(\theta; y)}{\partial\theta\partial\theta'}$

Definition

(Score): The score vector (or gradient) is a $(k \times 1)$ vector of first derivatives of $\ln L(\theta; y)$ with respect to θ, $g(\theta) = \dfrac{\partial \ln L(\theta; y)}{\partial\theta}$

4. CONCEPTOS HABITUALES EN ESTIMACIÓN ML

ESTIMACIÓN DE LA MATRIZ DE VARIANZAS Y COVARIANZAS $[-E[H]]^{-1}$

Tres métodos:

(1) **Si la expresión de $[-E[H]]^{-1}$ es conocida, podemos evaluar la matriz en el valor de los parámetros estimados (que sustituirán así a los verdaderos parámetros que aparecen en la expresión).**

(2) **Si las esperanzas de los elementos de la matriz Hessiana no son conocidos (muchas veces esos elementos son funciones no lineales para las que no es posible calcular su esperanza), entonces podemos evaluar $I=[H]^{-1}$ en los valores de los parámetros estimados.**

(3) **Como E[H] es la varianza de las primeras derivadas, podemos estimarla mediante la varianza muestral de las primeras derivadas, es decir:**

$$I(\theta) = -E\left(\frac{\partial^2 \ln L(\theta; y)}{\partial\theta\partial\theta'}\right) = E\left(\frac{\partial \ln L(\theta; y)}{\partial\theta}\frac{\partial \ln L(\theta; y)}{\partial\theta'}\right)$$

$$\widehat{I}(\theta) = \sum_{i=1}^{n} \frac{\partial \ln L(\widehat{\theta}; y)}{\partial\widehat{\theta}} \cdot \frac{\partial \ln L(\widehat{\theta}; y)}{\partial\widehat{\theta}'}$$

EJERCICIO:
Matriz de información del estimador MV en el modelo de regresión lineal

La matriz de información es:

$$I\begin{pmatrix}\beta\\ \sigma^2\end{pmatrix} = -E\begin{pmatrix}\frac{\partial^2 \ln L}{\partial\beta\partial\beta'} & \frac{\partial^2 \ln L}{\partial\beta\partial\sigma^2}\\ \frac{\partial^2 \ln L}{\partial\sigma^2\partial\beta'} & \frac{\partial^2 \ln L}{(\partial\sigma^2)^2}\end{pmatrix} = \begin{pmatrix}-E\frac{\partial^2 \ln L}{\partial\beta\partial\beta'} & -E\frac{\partial^2 \ln L}{\partial\beta\partial\sigma^2}\\ -E(\frac{\partial^2 \ln L}{\partial\beta\partial\sigma^2})' & -E\frac{\partial^2 \ln L}{(\partial\sigma^2)^2}\end{pmatrix}$$

Los gradientes o scores ya los tenemos calculados

$$\frac{\partial \ln L}{\partial\beta} = \frac{-1}{2\sigma^2}(-2X'y + 2X'X\beta)$$

$$\frac{\partial \ln L}{\partial\sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}(y - X\beta)'(y - X\beta)$$

En realidad es un vector gradiente (k x 1)

EJERCICIO:
Matriz de información del estimador MV en el modelo de regresión lineal

La matriz de información es:

$$I\begin{pmatrix}\beta\\ \sigma^2\end{pmatrix} = -E\begin{pmatrix}\frac{\partial^2 \ln L}{\partial\beta\partial\beta'} & \frac{\partial^2 \ln L}{\partial\beta\partial\sigma^2}\\ \frac{\partial^2 \ln L}{\partial\sigma^2\partial\beta'} & \frac{\partial^2 \ln L}{(\partial\sigma^2)^2}\end{pmatrix} = \begin{pmatrix}-E\frac{\partial^2 \ln L}{\partial\beta\partial\beta'} & -E\frac{\partial^2 \ln L}{\partial\beta\partial\sigma^2}\\ -E(\frac{\partial^2 \ln L}{\partial\beta\partial\sigma^2})' & -E\frac{\partial^2 \ln L}{(\partial\sigma^2)^2}\end{pmatrix}$$

- $\frac{\partial^2 \ln L}{\partial\beta\partial\beta'} = -\frac{X'X}{\sigma^2}$
- $\frac{\partial^2 \ln L}{(\partial\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{(y-X\beta)'(y-X\beta)}{\sigma^6} = \frac{n}{2\sigma^4} - \frac{u'u}{\sigma^6}$
- $\frac{\partial^2 \ln L}{\partial\beta\partial\sigma^2} = -\frac{(X'y-X'X\beta)}{\sigma^4} = -\frac{X'u}{\sigma^4}\left(= (\frac{\sigma^2 \ln L}{\partial\sigma^2\partial\beta'})'\right)$

EJERCICIO:
Matriz de información del estimador MV en el modelo de regresión lineal

La matriz de información es:

$$I\begin{pmatrix}\beta\\ \sigma^2\end{pmatrix} = -E\begin{pmatrix}\frac{\partial^2 \ln L}{\partial\beta\partial\beta'} & \frac{\partial^2 \ln L}{\partial\beta\partial\sigma^2}\\ \frac{\partial^2 \ln L}{\partial\sigma^2\partial\beta'} & \frac{\partial^2 \ln L}{(\partial\sigma^2)^2}\end{pmatrix} = \begin{pmatrix}-E\frac{\partial^2 \ln L}{\partial\beta\partial\beta'} & -E\frac{\partial^2 \ln L}{\partial\beta\partial\sigma^2}\\ -E(\frac{\partial^2 \ln L}{\partial\beta\partial\sigma^2})' & -E\frac{\partial^2 \ln L}{(\partial\sigma^2)^2}\end{pmatrix}$$

- $-E(\frac{\partial^2 \ln L}{\partial\beta\partial\beta'}) = X'X/\sigma^2$
- $-E(\frac{\partial^2 \ln L}{(\partial\sigma^2)^2}) = \frac{n}{2\sigma^4} \Leftarrow E(\frac{u'u}{\sigma^6}) = \frac{\Sigma_{i=1}^n Eu_i^2}{\sigma^6} = \frac{n\sigma^2}{\sigma^6}$
- $-E(\frac{\partial^2 \ln L}{\partial\beta\partial\sigma^2}) = 0 \quad \Leftarrow E(X'u) = X'E(u) = 0$

EJERCICIO:
Matriz de información del estimador MV en el modelo de regresión lineal

Por tanto:

$$\Rightarrow I\begin{pmatrix}\beta\\ \sigma^2\end{pmatrix} = \begin{pmatrix}\frac{X'X}{\sigma^2} & 0\\ 0 & \frac{n}{2\sigma^4}\end{pmatrix}$$

$$\Rightarrow I^{-1}\begin{pmatrix}\beta\\ \sigma^2\end{pmatrix} = \begin{pmatrix}(\frac{X'X}{\sigma^2})^{-1} & 0\\ 0 & (\frac{n}{2\sigma^4})^{-1}\end{pmatrix} = \begin{pmatrix}\sigma^2(X'X)^{-1} & 0\\ 0 & 2\sigma^4/n\end{pmatrix}$$

$$\begin{pmatrix}\widehat{\beta}_{MLE}\\ \widehat{\sigma}^2_{MLE}\end{pmatrix} \sim N\left(\begin{pmatrix}\beta\\ \sigma^2\end{pmatrix}, \begin{pmatrix}\sigma^2(X'X)^{-1} & 0\\ 0 & 2\sigma^4/n\end{pmatrix}\right)$$

EJERCICIO:

Supongamos una muestra aleatoria simple $X_1, X_2, \ldots, X_n$ con función de distribución de probabilidad definida por:

$$f(x_i) = \begin{cases} \theta^{x_i}(1-\theta)^{1-x_i} & x_i = 0, 1 \\ 0 \end{cases}$$

1) Calcula $E(X_i)$ y $Var(X_i)$
2) Encuentra el estimador MV de θ
3) ¿Es Θ_{MV} consistente?
4) Encuentra su distribución asintótica