MAXIMUM LIKELIHOOD ESTIMATOR
MLE

consider a linear model

$$Y = X\beta + e$$

under assumption: $e \sim N(0, \sigma^2 I)$, X NONRANDOM

Given that we have a random sample of Y's, they are **i.i.d.** and also Normally distributed: $Y \sim N(X\beta, \sigma^2 I)$

the density of any y_t:

$$f(y_t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y_t - x_t\beta)^2\right\}$$

the density of the whole sample:

$$f(y_1, y_2 \dots y_T) = f(y_1) \cdot f(y_2) \cdot \dots f(y_T) =$$

$$\left(\frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{(Y_1 - x_1\beta)^2}{2\sigma^2}\right)\right) \cdot \left(\frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{(Y_2 - x_2\beta)^2}{2\sigma^2}\right)\right) \dots$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^T \left(\frac{1}{\sigma}\right)^T \exp\left(-\frac{(Y_1 - x_1\beta)^2}{2\sigma^2}\right) \cdot \exp\left(\frac{-(Y_2 - x_2\beta)^2}{2\sigma^2}\right) \dots$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^T \left(\frac{1}{\sigma}\right)^T \exp\left\{-\frac{\sum_{t=1}^{T}(Y_t - X_t\beta)^2}{2\sigma^2}\right\}$$

For the linear model, we denote by $\theta = (\beta', \sigma^2)'$ the vector of parameters

the joint density integrates to one for the true value θ_0 of the unknown parameter θ

The likelihood function:

$$L(y;\theta) = l(y_1;\theta) \cdot l(y_2;\theta) \cdots \cdot l(y_T,\theta)$$

The likelihood function's value depends on θ, i.e. is a function of θ, while y is considered given.

Let $\tilde{\theta}(y_1, \dots, y_T)$ denote the parameter vector that maximizes the likelihood function

$$L(y;\tilde{\theta}) \geq L(y;\theta^*)$$

where θ^* is some other estimator

intuition: $\tilde{\theta}$ is the parameter estimate that yield the largest likelihood of the sample at hand

$\tilde{\theta}(y_1 \dots y_T)$ is a function of random y's and is therefore random too.

We call MLE (maximum likelihood estimator) the value $\tilde{\theta}$ that maximizes the likelihood function L.

in fact we rather work with $L^* = \ln L$

$$L^* = \ln L(y, X; \beta, \sigma^2) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2}\frac{(y - X\beta)'(y - X\beta)}{\sigma^2}$$

First order conditions:

$$\frac{\partial L^*(.)}{\partial \beta} = -\frac{1}{\tilde{\sigma}^2}\left(-X'y + X'X\tilde{\beta}\right) = 0$$

$$\frac{\partial L^*(.)}{\partial \sigma^2} = -\frac{T}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4}\left(y - X\tilde{\beta}\right)'\left(y - X\tilde{\beta}\right) = 0$$

we have K+1 equations for K+1 unknowns $\tilde{\beta}$ and $\tilde{\sigma}^2$ yielding: MLE (maximum likelihood estimators:)

$$\tilde{\beta} = (X'X)^{-1}X'y$$

$$\tilde{\sigma}^2 = \frac{\hat{e}'\hat{e}}{T} = \frac{T-K}{T}\hat{\sigma}^2$$

biased in small samples.

SMALL SAMPLE PROPERTIES.

the $\tilde{\beta}$ are **Normally distributed**

$$\tilde{\beta} \sim N(\beta, \sigma^2(X'X)^{-1})$$

(ident to OLS, hence unbiased)

$T\frac{\tilde{\sigma}^2}{\sigma^2} = (T-K)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{T-K}$ biased in small samples.

We say that the MLE are most efficient ones. Indeed the satisfy the Cramer-Rao lower bound for variances of estimators. hence the variance of MLE's is the lowest possible among all estimators.

Cramer-Rao:

let $y_1, y_2 \cdots y_T$ be a random sample and $L(y;\theta) = \prod_{t=1}^{T} l_t$ the likelihood function. DEFINE THE INFORMATION MATRIX to be:

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

with (i,j) element

$$I_{ij} = -E\left[\frac{\partial^2 \ln L(y,\theta)}{\partial\theta_i\,\partial\theta_j}\right], \quad i,j = 1,2\ldots K.$$

let $\hat{\theta}$ be any unbiased estimator of θ with variance Σ.
Then the matrix $\Sigma - [I(\theta)]^{-1}$ is positive semidefinite

$\Rightarrow$ i.e. $\Sigma - [I(\theta)]^{-1} \geq 0$, so that:

$$\Sigma \geq [I(\theta)]^{-1}$$

it is impossible for an estimator to have var lower than $[I(\theta)]^{-1}$.

There exist estimator with var. equal to $[I(\theta)]^{-1}$ and

one of them is MLE.

derivation:

$$\ln L = -\frac{T}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)$$

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

$$= \frac{1}{\sigma^2}(X'Y - X'X\beta) = \frac{1}{\sigma^2}X'(Y - X\beta)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = \frac{-T}{2\sigma^2} + \frac{1}{2\sigma^4}(Y - X\beta)'(Y - X\beta)$$

$$E\left[\frac{\partial \ln L}{\partial \beta}\right] = 0 \quad \text{and} \quad E\left[\frac{\partial \ln L}{\partial \sigma^2}\right] = 0$$

The Cramer-Rao lower bounds for the variances of $\hat{\beta}$ and $\hat{\sigma}^2$ are: $\sigma^2(X'X)^{-1}$ and $2\sigma^4/T$ respectively.

$$\frac{\partial^2 \ln L}{\partial\beta\partial\beta'} = -\frac{1}{\sigma^2}X'X \Rightarrow -E\left[\frac{\partial^2 \ln L}{\partial\beta\partial\beta'}\right] = \frac{X'X}{\sigma^2}$$

$$\frac{\partial^2 \ln L}{\partial^2\sigma^2} = \frac{T}{2\sigma^4} - \frac{1}{\sigma^6}(Y-X\beta)'(Y-X\beta) \Rightarrow -E\left[\frac{\partial^2 \ln L}{\partial^2\sigma^2}\right] = \frac{T}{2\sigma^4}$$

$$\frac{\partial^2 \ln L}{\partial\beta\partial\sigma^2} = -\frac{1}{\sigma^4}X'(Y-X\beta); \quad -E\left[\frac{\partial^2 \ln L}{\partial\beta\partial\sigma^2}\right] = 0$$

Therefore the information matrix is:

$$I(\beta,\sigma^2) = \begin{bmatrix} \frac{X'X}{\sigma^2} & 0 \\ 0 & \frac{T}{2\sigma^4} \end{bmatrix}$$

1) MLE of all $\tilde{\beta}$'s are uncorrelated with $\tilde{\sigma}^2$ estimator of variance

2) $\tilde{\sigma}^2$ is BIASED IN SMALL SAMPLE AND WE DON'T CARE THAT IT IS EFFICIENT. IT IS NOT GOOD.

2) ASYMPTOTIC SETUP

WE CONSIDER A LARGE SAMPLE AND CARE ABOUT CONSISTENCY OF ESTIMATORS

$\text{plim}\ \tilde{\beta} = \beta$

$\text{plim}\ \hat{\sigma}^2 = \sigma^2$ (because if T is large it is irrelevant if we divide $\hat{e}'\hat{e}$ by T or $(T-K)$)

WE WILL SEE THAT MLE IS

- ASYMPTOTICALLY NORMALLY DISTRIBUTED

$$\sqrt{T}\,(\tilde{\theta}_T - \theta) \xrightarrow{A} N\left(0,\ \lim_{T\to\infty}\left[\frac{I(\theta)}{T}\right]^{-1}\right)$$

- ASYMPTOTICALLY MOST EFFICIENT:

indeed the asymptotic analogue of the Cramer-Rao lower bound is:

$$\frac{1}{T}\lim_{T\to\infty}\left[\frac{I(\theta)}{T}\right]^{-1}$$

asy variance of MLE satisfies the asy CRAMER RAO LOWER BOUND. IT APPROACHES THIS LIMIT AS $T\to\infty$.

$$\lim_{T\to\infty}\left[\frac{I(\beta,\sigma^2)}{T}\right] = \begin{bmatrix} \frac{\Sigma_{xx}}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{bmatrix}$$

inverse of this matrix is the var of MLE for ANY GIVEN VALUE OF T

where Σ_{xx} is $\lim \frac{X'X}{T}$

The Information matrix can be consistently estimated from:

i) $I(\tilde{\theta})$

ii) $-\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \log l(y_t;\tilde{\theta})}{\partial\theta\partial\theta'} = \hat{J}$ (Hessian)

iii) $\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \log l(y_t;\tilde{\theta})}{\partial\theta}\frac{\partial \log l(y_t;\tilde{\theta})}{\partial\theta'} = \hat{I}$

outer product of scores.

MLE estimation of nonlinear models

- Poisson
- exponential
- binomial

$$Y = X\beta + e \qquad e \sim (0, \sigma^2 I)$$

X : assumed fixed, constant

- Assumption $e_t \sim N(0, \sigma^2)$ NORMALITY

$$\Rightarrow f(e_t) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\left(-\frac{e_t^2}{2\sigma^2}\right)$$

Log-likelihood function

$$L(\beta_1, \ldots \beta_K, \sigma^2) = -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)$$

<u>maximize</u> $L(\cdot)$ w.r. to unknowns

$$\begin{bmatrix} \frac{\partial L}{\partial \beta_1} \\ \frac{\partial L}{\partial \beta_2} \\ \vdots \\ \frac{\partial L}{\partial \beta_K} \\ \frac{\partial L}{\partial \sigma^2} \end{bmatrix}$$

vector of FOC, where $\theta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \\ \sigma^2 \end{bmatrix}$
(scores)

$$\tilde{\beta}_{MLE} = b_{OLS} = (X'X)^{-1} X'Y$$

$$\tilde{\sigma}^2_{ML} \neq \hat{\sigma}^2_{OLS}$$

$$\tilde{\sigma}^2_{ML} = \frac{1}{T} \sum \hat{e}_t^2$$

2.

matrix of 2nd order derivatives $\dfrac{\partial^2 L(\,)}{\partial\theta\,\partial\theta'}$:

$$
\begin{bmatrix}
\dfrac{\partial^2 L}{\partial\beta_1^2} & \dfrac{\partial^2 L}{\partial\beta_1\partial\beta_2} & \dfrac{\partial^2 L}{\partial\beta_1\partial\beta_3} & \cdots & \dfrac{\partial^2 L}{\partial\beta_1\partial\sigma^2} \\
\dfrac{\partial^2 L}{\partial\beta_1\partial\beta_2} & \dfrac{\partial^2 L}{\partial\beta_2^2} & \dfrac{\partial^2 L}{\partial\beta_2\partial\beta_3} & \cdots & \dfrac{\partial^2 L}{\partial\beta_2\partial\sigma^2} \\
 & & \ddots & & \\
\dfrac{\partial^2 L}{\partial\beta_1\partial\sigma^2} & \dfrac{\partial^2 L}{\partial\beta_2\partial\sigma^2} & \cdots & \cdots & \dfrac{\partial^2 L}{\partial(\sigma_2)^2}
\end{bmatrix}
$$

$$(K+1) \times (K+1)$$

$\swarrow$ for K betas

$\searrow$ for σ^2

ex: $y_t = \beta_1 + \beta_2 p_t + \beta_3 a_t + e_t$: $\dfrac{K=3}{+\sigma^2}$

$$
\begin{array}{cc}
 & 4 \\
4 &
\begin{bmatrix}
\dfrac{\partial^2 L}{\partial\beta_1^2} & \dfrac{\partial^2 L}{\partial\beta_1\partial\beta_2} & \dfrac{\partial^2 L}{\partial\beta_1\partial\beta_3} & \dfrac{\partial^2 L}{\partial\beta_1\partial\sigma^2} \\
\dfrac{\partial^2 L}{\partial\beta_1\partial\beta_2} & \dfrac{\partial^2 L}{\partial\beta_2^2} & \dfrac{\partial^2 L}{\partial\beta_2\partial\beta_3} & \dfrac{\partial^2 L}{\partial\beta_2\partial\sigma^2} \\
\dfrac{\partial^2 L}{\partial\beta_1\partial\beta_3} & \dfrac{\partial^2 L}{\partial\beta_2\partial\beta_3} & \dfrac{\partial^2 L}{\partial\beta_3^2} & \dfrac{\partial^2 L}{\partial\beta_3\partial\sigma^2} \\
\dfrac{\partial^2 L}{\partial\beta_1\partial\sigma^2} & \dfrac{\partial^2 L}{\partial\beta_2\partial\sigma^2} & \dfrac{\partial^2 L}{\partial\beta_3\partial\sigma^2} & \dfrac{\partial^2 L}{\partial(\sigma^2)^2}
\end{bmatrix}
\end{array}
$$

Information matrix

$$I(\theta) = E - \left[\frac{\partial^2 L}{\partial \theta \, \partial \theta'} \right]$$

where $\theta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \sigma^2 \end{bmatrix}$

INVERSE OF $I(\theta)$

$I(\theta)^{-1}$ is the lower bound on the variance of any estimator

CRAMER-RAO LOWER BOUND

$$I(\theta)^{-1} = \underbrace{\begin{bmatrix} \sigma^2 (X'X)^{-1} & \begin{matrix} 0 \\ 0 \\ \vdots \end{matrix} \\ 0 \; 0 \; \cdots & \frac{2\sigma^4}{T} \end{bmatrix}}_{K+1} \quad K+1$$

4.

Cramer-Rao lower bound on variance of

$b_{OLS} = \tilde{\beta}_{MLE} = (X'X)^{-1}X'y$

is

$$\sigma^2 (X'X)^{-1}$$

$\boxed{\text{ASYMPTOTICALLY}}$: $\sigma^2 \cdot \Sigma_{xx}^{-1}$

where Σ_{xx} is plim $\frac{X'X}{T}$

WHEN X ARE RANDOM e NOT NORMAL

ex: for $y_t = \beta_1 + \beta_2 p_t + \beta_3 a_t + e_t$

Information matrix

$$I(\theta) = \begin{bmatrix} \frac{1}{\sigma^2}\begin{bmatrix} T & \Sigma p_t & \Sigma a_t \\ \Sigma p_t & \Sigma p_t^2 & \Sigma p_t a_t \\ \Sigma a_t & \Sigma a_t p_t & \Sigma a_t^2 \end{bmatrix} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \end{matrix} & \frac{T}{2\sigma^4} \end{bmatrix}$$

$$I(\theta)^{-1} = \begin{bmatrix} \begin{bmatrix} - & & - \\ - & \sigma^2 (X'X)^{-1} & - \\ - & - & - \end{bmatrix} & \begin{matrix} 0 \\ 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 \end{matrix} & \frac{2\sigma^4}{T} \end{bmatrix}$$