Lecture 19

6.04.2021.

Econometrics-1

MAXIMUM LIKELIHOOD ESTIMATION (MLE)

MLRM: $$\underset{N\times 1}{\underline{y}} = \underset{N\times K}{X}\underset{K\times 1}{\underline{\beta}} + \underset{N\times 1}{\underline{u}}$$

$$\underset{K\times 1}{\underline{\hat{\beta}}_{OLS}} = \underbrace{\underbrace{(X'X)^{-1}}_{K\times K}\ \underbrace{X'\underline{y}}_{K\times 1}}_{K\times 1}$$

$$\underline{\hat{\beta}}_{MLE} = ?$$

→ Ultimate objective for this course.

Steps in MLE

Step 1: Assume a particular form of distribution to describe a population.

e.g. (a) $X \sim N(\mu, \sigma^2)$

(b) $Z \sim \underbrace{B(n, p)}$

Binomial distribution:
· Probability of attaining $Z = k$ successes in n independent trials, where
$P(\text{success in each trial}) = p$.

Step 2 : Given the distribution assumed in Step 1, Calculate the probabity / likelihood of Observing the "given Sample".

Step 3 : $\boxed{\text{Choose the parameters}}$ of distribution (specified in step 1) in a manner that maximizes the probability / likelihood of Observing the given sample (as calculated in step 2).

This Choice leads to MLE.

Example 1 :

- A coin is tossed 50 times (n trials) and we observe 30 heads.
(z = 30 successes)

- In each independent trial:

$$\Pr(\text{heads}) = p \quad \in [0,1]$$

$$\Pr(\text{tails}) = 1-p \quad \in [0,1]$$

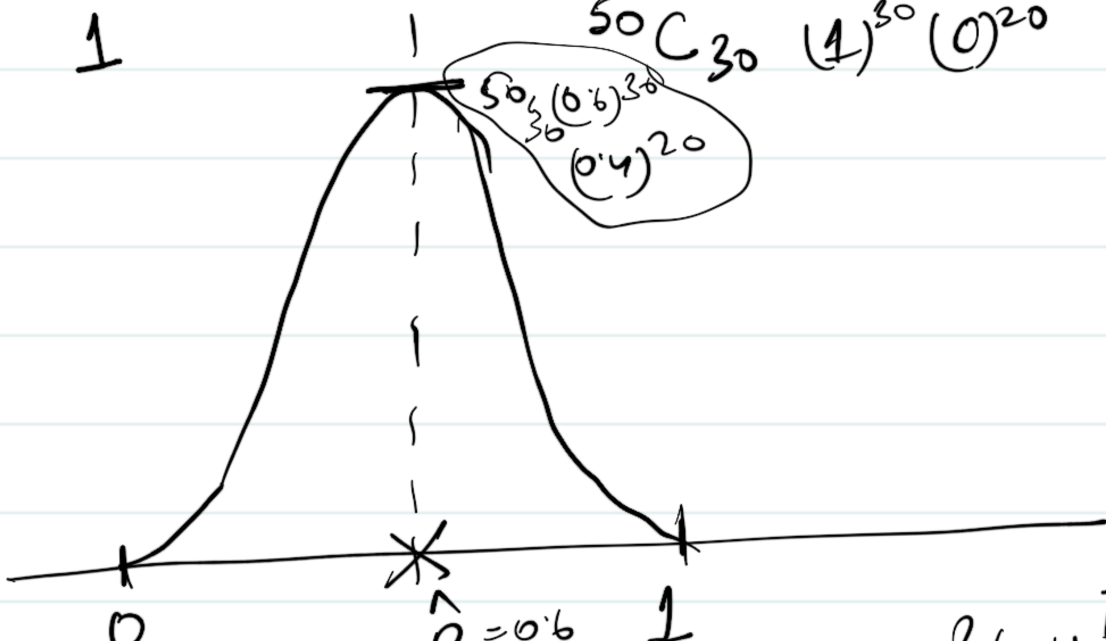
After 50 trials, using the binomial distribution definition, we have

$$\Pr(30 \text{ heads in } 50 \text{ trials}) = {}^{50}C_{30}\,(p)^{30}\,(1-p)^{\overbrace{50-30}^{=20}}$$

$$= {}^{50}C_{30}\,(p)^{30}\,(1-p)^{20}$$

$$Pr(30 \text{ heads in } 50 \text{ trials}) = {}^{50}C_{30}\, p^{30} (1-p)^{20}$$

p	$Pr(30H \text{ in } 50 \text{ trials})$
0	${}^{50}C_{30} (0)^{30} (1)^{20} = 0$
0.01	${}^{50}C_{30} (0.01)^{30} (0.99)^{20}$
0.02	${}^{50}C_{30} (0.02)^{30} (0.98)^{20}$
$\vdots$	
$\hat{p}_{ML} = 0.6$	${}^{50}C_{30} (0.6)^{30} (0.4)^{20}$
$\vdots$	
0.99	${}^{50}C_{30} (0.99)^{30} (0.01)^{20}$
1	${}^{50}C_{30} (1)^{30} (0)^{20} = 0$



$\hat{p}_{ML} = 0.6$ that maximize $Pr(30H \text{ in } 50 \text{ trials})$

Basically, we want to

$$\max \Pr(30H \text{ in } 50 \text{ trials})$$

$$= \max_{\{p\}} L = \overbrace{{}^{50}C_{30}\,(p)^{30}\,(1-p)^{20}}^{\text{likelihood of observing the given sample}}$$

L: Likelihood function.

$\boxed{\text{Take the natural logarithm of the likelihood function and then maximize it.}}$

WHY? → CONVENTION + CONVENIENCE

Choose p s.t.

$$\max_{\{p\}} \ln L = \ln({}^{50}C_{30}) + 30\ln(p) + 20\ln(1-p)$$

f.o.c

$$\frac{\partial \ln L}{\partial p} = 0 \Rightarrow \frac{30}{p} - \frac{20}{1-p} = 0$$

$$\Rightarrow \frac{30}{p} = \frac{20}{1-p} \Rightarrow \boxed{\hat{p}_{ML} = 0.6}$$

Example 2 :

Given: N observation of a random variable x

$$\mathcal{X} = \{x_1, x_2, x_3, \dots, x_N\}$$ → dataset / Observed sample

Assume $\mathcal{X} \overset{iid}{\sim} N(\mu, \sigma^2)$

Q: Estimate $\hat{\mu}_{ML}$ and $\hat{\sigma}^2_{ML}$

Step 2 : Given $x \overset{iid}{\sim} N(\mu, \sigma^2)$ calculate the probability of observing sample $\{x_1, x_2, \dots x_N\}$ or $\{x_i\}_{i=1}^{N}$

Use pdf of normal distribution :

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

Probability of observing $x = x_i$

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

Likelihood of observing $\{x_1, x_2, \ldots, x_N\}$ sample

$$L(x = \{x_1, \ldots, x_N\}; \mu, \sigma^2) = f(x_1) \cdot f(x_2 | x_1) \cdot f(x_3 | x_2, x_1) \cdot \ldots \cdot f(x_N | x_{N-1}, \ldots, x_1)$$

WHY ?

$\because x \overset{iid}{\sim} N(\mu, \sigma^2)$

$$= f(x_1) \cdot f(x_2) \cdot \ldots \cdot f(x_N)$$

$$= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

$$= \frac{1}{(\sqrt{2\pi\sigma^2})^N} \exp\left(\frac{-1}{2} \sum_{i=1}^{N} \left(\frac{x_i - \mu}{\sigma}\right)^2\right)$$

$$L\left(x=\{x_1, x_2, \dots, x_N\}; \mu, \sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)^2\right)$$

likelihood of observing the given sample when $x \overset{iid}{\sim} N(\mu, \sigma^2)$

Calculate the log-likelihood function:

$$l = \ln L = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

Step3: $\max\limits_{\mu, \sigma^2} \; l = \ln L$

$$\max_{\mu, \sigma^2} \left\{-\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{x_i-\mu}{\sigma}\right)^2\right\}$$

$$\max_{\{\mu, \sigma^2\}} \left\{ -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2} \sum_{i=1}^{N} \left[\frac{(x_i - \mu)^2}{\sigma^2} \right] \right.$$

F.O.C.

1) $\frac{\partial \ln L}{\partial \mu} \overset{set}{=} 0 \Rightarrow -\frac{1}{2\sigma^2} \sum_{i=1}^{N} 2(x_i - \mu) \times (-1) = 0$

$$\Rightarrow \frac{1}{\sigma^2} \sum_{i=1}^{N} (x_i - \mu) = 0 \rightarrow 1$$

2) $\frac{\partial \ln L}{\partial \sigma^2} \overset{set}{=} 0 \Rightarrow \frac{-N}{2\sigma^2} - \frac{1 \, (-1)}{2(\sigma^2)^2} \sum_{i=1}^{N} (x_i - \mu)^2 = 0$

$$\Rightarrow \frac{-N}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{N} (x_i - \mu)^2 = 0 \rightarrow 2$$

Solving 1 and 2 $\Rightarrow \hat{\mu}_{ML}$ and $\hat{\sigma}^2_{ML}$

eqn. 1 $\Rightarrow \sum_{i=1}^{N} x_i = N \hat{\mu}_{ML} \Rightarrow \boxed{\hat{\mu}_{ML} = \frac{\sum_{i=1}^{N} x_i}{N} = \bar{x}}$

Q^n. 2 $\Rightarrow$

$$\frac{-N}{2\hat{\sigma}^2_{ML}} + \frac{1}{2\left(\hat{\sigma}^2_{ML}\right)^2} \sum_{i=1}^{N} (x_i - \bar{x})^2 = 0$$

$$\Rightarrow \boxed{\hat{\sigma}^2_{ML} = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}}$$

$$\boxed{\hat{\mu}_{ML} = \bar{x}}$$

Aside: we know $s^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}$ is the unbiased estimator of population variance.

$\hat{\sigma}^2_{ML}$ is biased !! why? Because it does not account for the degrees of freedom.

ML is usually a good estimation strategy when we have large samples

$\because N \approx N-1$

MLE applied to the simple linear regression model