OLS Estimators

- 3 approaches to parameter estimation (Wooldridge Appendix C.4)

① method of moment (MM)

idea : theoretical moments = sample moments

$E(X^n)$ (or other expected values) $\qquad \frac{1}{n}\sum_{i=1}^{n} X_i^n$

② maximum likelihood (ML)

idea : $\max \mathcal{L}(\theta; \underbrace{x_1 \cdots x_n}_{\text{observations}})$ — θ: parameters

$\mathcal{L}$: likelihood function (pmf, pdf)

③ least square (LS)

idea : $\min \sum_{i=1}^{n} (\text{error})^2$

- Illustrative (simple) example

Ber(p) $\qquad$ V X V X X X V X X X

$\qquad\qquad$ 1 0 1 0 0 0 1 0 0 0 $\qquad \hat{p} = \frac{3}{10}$?

① MM $\qquad E(X) = \bar{X}$ (match), $E(X) = p$ $\qquad \underbrace{\hat{p}}_{\text{estimator}} = \bar{X} = \frac{3}{10}$

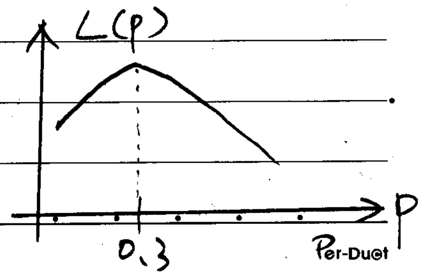
② ML $\qquad L(p) = p(1-p)p(1-p)^3 p(1-p)^3 = p^3(1-p)^7$

(or $L(p; x_1 \cdots x_{10})$)

$$\max \log L(p) = \max \left[3 \log p + 7 \log(1-p) \right]$$

$$\frac{d\, L(p)}{dp} = \frac{3}{p} - \frac{7}{1-p} = 0$$

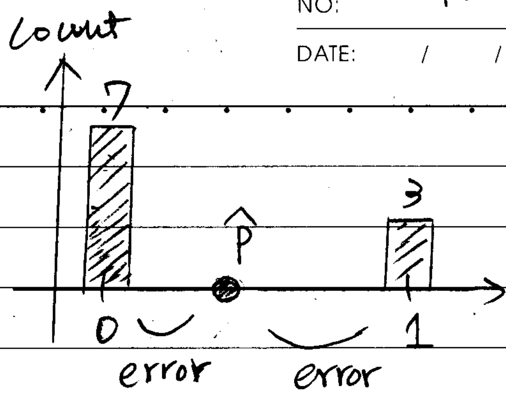
$$\Rightarrow \hat{p} = \frac{3}{10}$$



③ LS

squared error

$$\min Q(p) = \sum_{i=1}^{10} \underbrace{(x_i - p)}_{\text{error}}^2$$



$$\frac{\partial Q}{\partial p} = 3 \cdot 2(1-p) + 7 \cdot 2(-p) = 0$$

$$\hat{p} = 0.3$$

Notes:

① MM : moments can be $E(X^n)$, or any kind of expected values

ex: $X \sim \Gamma(\alpha, \beta)$, $E(X) = \frac{\alpha}{\beta}$, $Var(X) = \frac{\alpha}{\beta^2}$

Given sample mean M, sample variance V

we want to find $\hat{\alpha}, \hat{\beta}$

$$\frac{\alpha}{\beta} = M, \quad \frac{\alpha}{\beta^2} = V \Rightarrow \hat{\alpha} = \frac{M^2}{V}, \quad \hat{\beta} = \frac{M}{V}$$

ex: $y = \beta_0 + \beta_1 x + u$

$$\left.\begin{array}{l} E(u) = 0 \\ E(x \cdot u) = 0 \end{array}\right\} \Rightarrow \hat{\beta}_0, \hat{\beta}_1$$

② ML : why likelihood function $\approx$ pmf, pdf ?

ex: $X \sim N(\mu, \sigma^2)$, $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

μ: unknown, σ^2: known $(=1)$

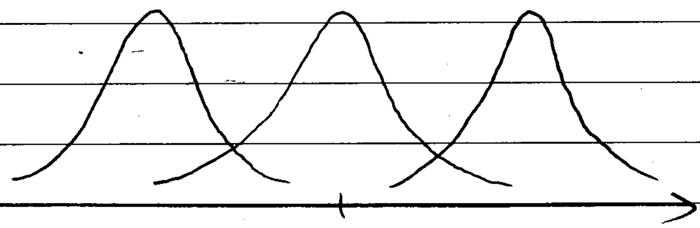
if we observe $x_i = 10$, what is the most probable value of μ ?

you would say $\hat{\mu} = 10$.

$N(5,1)$ $N(10,1)$ $N(15,1)$

Given $X_i = 10$

$N(10,1)$ gives the maximum likelihood



10

difference :
$$\begin{cases} \text{pdf} : f(x \mid \mu, \sigma^2) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ \text{likelihood} : f(\mu, \sigma^2 \mid x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \end{cases}$$

observation

③ LS :

useful in regression

$$y = \beta_0 + \beta_1 x + u$$

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} (\underbrace{y_i - (\beta_0 + \beta_1 x)}_{\text{error}})^2$$

- Simple regression (chap 2)

$$y = \beta_0 + \beta_1 x + u$$

MM :
$$E(u) = 0 \rightarrow \cancel{\frac{1}{n}} \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad ①$$
$$E(xu) = 0 \rightarrow \cancel{\frac{1}{n}} \sum x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad ②$$

by ①, $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad ③$$

plugging ③ into ②

$$\sum x_i [\, y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i \,] = 0$$

$$\sum x_i (y_i - \bar{y}) = \hat{\beta} \sum x_i (x_i - \bar{x})$$

why? $\rightarrow (x_i - \bar{x})$ $\qquad$ $(x_i - \bar{x}) \leftarrow$ why?

$$\therefore \quad \hat{\beta}_1 = \underbrace{\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}, \quad \underbrace{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}}$$

sample version

think: $y = \beta_0 + \beta_1 x + u$

$$\underbrace{\beta = \frac{Cov(X,Y)}{Var(X)}}$$ ← population version

(why? $Cov(X,Y) = Cov(X, \beta_0 + \beta_1 X + u)$

$$= \beta_1 \underbrace{Cov(X,X)}_{\substack{\parallel \\ Var(X)}} + \underbrace{Cov(X,u)}_{\substack{\parallel \\ 0}}$$
)

Note: we have a number of expressions for β

~~like hedge~~ ratio ← $\beta = \rho \frac{\sigma_y}{\sigma_x} \left(= \frac{\rho \sigma_x \sigma_y}{\sigma_x^2} = \frac{Cov(X,Y)}{Var(X)} \right)$

or $\rho = \frac{\beta \sigma_x}{\sigma_y}$

$$\Rightarrow \rho^2 = \frac{\beta^2 \sigma_x^2}{\sigma_y^2} = \frac{\text{variance explained by } x}{\text{total variance in } y}$$

like R^2

$$\left(\begin{array}{c} \underbrace{y}_{\text{total}} = \underbrace{\beta_0 + \beta_1 x}_{\text{explained}} + \underbrace{u}_{\text{unexplained}} \\ \sigma_y^2 = \beta_1^2 \sigma_x^2 + \sigma_u^2 \end{array} \right)$$

$$\Rightarrow \underbrace{\rho^2}_{\text{percentage of explained part.}} = \frac{\sigma_y^2 - \sigma_u^2}{\sigma_y^2} = 1 - \underbrace{\frac{\sigma_u^2}{\sigma_y^2}}_{\text{percentage of unexplained part.}}$$

(LS)

- OLS = ordinary least square

$$Q(b_0, b_1) = \sum (y_i - b_0 - b_1 x_i)^2$$

why b_0, b_1 not β_0, β_1 ?

LS : find $(b_0, b_1) = (\hat\beta_0, \hat\beta_1)$ s.t. Q is minimized

$$\left.\frac{\partial Q}{\partial b_0}\right|_{(\hat\beta_0, \hat\beta_1)} = 0 \rightarrow \sum 2(y_i - \hat\beta_0 - \hat\beta_1 x_i)(-1) = 0$$

$$\left.\frac{\partial Q}{\partial b_1}\right|_{(\hat\beta_0, \hat\beta_1)} = 0 \rightarrow \sum 2(y_i - \hat\beta_0 - \hat\beta_1 x_i)(-x_i) = 0$$

$$\Rightarrow \begin{cases} \sum (y_i - \hat\beta_0 - \hat\beta_1 x_i) = 0 \\ \sum x_i (y_i - \hat\beta_0 - \hat\beta_1 x_i) = 0 \end{cases} \Leftarrow \begin{pmatrix} E(u) = 0 \\ E(xu) = 0 \end{pmatrix}$$

exactly the same as the two equations from MM

- residual : $\hat u_i = y_i - \hat y_i = y_i - (\hat\beta_0 + \hat\beta_1 x_i)$

$$\hat y_i \downarrow \hat\beta_0 + \hat\beta_1 x_i$$

Note :

① $y = \beta_0 + \beta_1 x + u$ (true model for population)

② $E(y|x) = \beta_0 + \beta_1 x$ (PRF) not $E(x)$

average

$E(y) = \beta_0 + \beta_1 E(x)$

③ $\hat y = \hat\beta_0 + \hat\beta_1 x$ (SRF)

(not $y = \hat\beta_0 + \hat\beta_1 x + u$
$\hat y = \hat\beta_0 + \hat\beta_1 \hat x$) no such thing!

average

$\bar y = \hat\beta_0 + \hat\beta_1 \bar x$

Per-Duet

But we do have $y = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{u}$ → not u

this is how residual is defined

i.e. $\hat{u} = y - \underbrace{(\hat{\beta}_0 + \hat{\beta}_1 x)}_{=\hat{y}}$

(Note:

$$y = \beta_0 + \beta_1 x + u \quad \text{(population)}$$
$$= \hat{\beta}_0 + \hat{\beta}_1 x + \hat{u} \quad \text{(sample)}$$
)

- In fact:

$$\begin{cases} E(u) = 0 \longrightarrow \sum_{i=1}^{n} \hat{u}_i = 0 & \left(\text{not } \sum_{i=1}^{n} u_i = 0\right) \\ E(xu) = 0 \longrightarrow \sum_{i=1}^{n} x_i \hat{u}_i = 0 & \left(\text{not } \sum_{i=1}^{n} \hat{x}_i \hat{u}_i = 0\right) \end{cases}$$

do not get confused! no such thing!

- $$\sum (y_i - \bar{y})^2 = \sum (\hat{y} - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2$$

$$\underbrace{(y_i - \hat{y})}_{\hat{u}_i} + (\hat{y} - \bar{y})$$

$$\begin{array}{ccccc} y_i & & \xrightarrow{T} & & \bar{y} \\ & R\searrow & & \nearrow E & \\ & & \hat{y}_i & & \end{array}$$

Why is the cross term 0?

$$\sum \hat{u}_i (\hat{y} - \bar{y}) = \sum \hat{u}_i \hat{y}_i - \underbrace{\sum \hat{u}_i \bar{y}}_{=0}$$

$\bar{y}$ → const

$\hat{y} \to \hat{\beta}_0 + \hat{\beta}_1 x_i$, $\quad \hat{y}_i \to \hat{\beta}_0 + \hat{\beta}_1 x_i$

$$= \sum \hat{u}_i \hat{\beta}_0 + \sum \hat{u}_i \hat{\beta}_1 x_i$$

$\hat{\beta}_0$ → const $= 0$, $\quad \hat{\beta}_1$ → const $= 0$

SST = SSE + SSR

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

$$\left(\rho^2 = \frac{\beta^2 \sigma_x^2}{\sigma_y^2} = 1 - \frac{\sigma_u^2}{\sigma_y^2}\right)$$

Per-Duet