Estimation

(1) Identifiability and Estimable:

· Consider $\underset{n\times 1}{Y} = \underset{n\times p}{X}\underset{p\times 1}{\beta} + \underset{n\times 1}{e}$. $E(e)=0$. indept with β.

where Y is observation. X is known. β unobserved.

$\Rightarrow$ We can only learn about β by $E(Y) = X\beta$.

① Def: For $E(Y) = f(\beta)$

i) β is identifiable if $f(\beta_1) = f(\beta_2) \Rightarrow \beta_1 = \beta_2$

ii) $g(\beta)$ is identifiable if $f(\beta_1) = f(\beta_2) \Rightarrow g(\beta_1) = g(\beta_2)$

Rmk: If β or $g(\beta)$ isn't identifable. Then it's impossible to know it base on $E(Y) = f(\beta)$

prop. In regression model, β is identifable

$\Leftrightarrow r(X) = p$. $(X \in M^{n\times p})$.

Pf: $(\Rightarrow)$ If $r(X) < p$. $X = P\begin{pmatrix} I & \\ & 0 \end{pmatrix}Q^T$.

$\exists\ \beta_1 \neq \beta_2$. $X\beta_1 = X\beta_2$.

$(\Leftarrow)$ $\beta_1 = (X^TX)^{-1}X^TX\beta_1 = (X^TX)^{-1}X^TX\beta_2$

$= \beta_2$. if $X\beta_1 = X\beta_2$.

Thm. $g(\beta)$ is identifable $\Leftrightarrow$ $g(\beta)$ is function of $f(\beta)$.

Pf: $g(\beta)$ is func of $f(\beta)$. if $\forall \beta_1 \neq \beta_2$.

$f(\beta_1) = f(\beta_2) \Rightarrow g(\beta_1) = g(\beta_2)$

(Set $G: f(\beta) \longmapsto g(\beta)$. well-def)

② Def: L.F of β. $\Lambda^T\beta$ is estimable if it's L.F of $f(\beta) = X\beta$. i.e. $\Lambda^T\beta = P^T X \beta$.

Rmk: P isn't unique. but MP is unique. (Orthonormal proj. on $C(X)$)

i.e. $P_1^T X = P_2^T X \Rightarrow MP_1 = MP_2$.

prop. $X\beta = \Sigma \beta_k X_k$, $X = (X_1 \cdots X_p)$ Then :

β_i isn't estimable $\Leftrightarrow \exists (\alpha_k).\ X_i = \sum_{k \neq i} \alpha_k X_k$

Cor. β_i is estimable $\Leftrightarrow$ for $\Sigma \alpha_k X_k = 0$ only holds when $\alpha_i = 0$.

Rmk: Consider: $\sum_1^p \lambda_i \beta_i$ is estimable or not

We can check each i. β_i by prop.

③ Def: $f(Y)$ is linear estimate of $\lambda^T\beta$ if $f(Y) = a_0 + a^T Y$. for $a_0 \in \mathbb{R}^1$. $a^T \in \mathbb{R}^n$.

prop. $a_0 + a^T Y$ is unbiased estimate of $\lambda^T\beta$ if $a_0 = 0$. $a^T X = \lambda^T$.

(2) Least Square:

- For $Y = X\beta + e$. $E(e) = 0$

We want to estimate: $E(Y) = X\beta \in C(X)$

So we might take vector in $C(X)$ which is closest to Y. We call $\hat{\beta}$ Least square estimate (LSE). st. $(Y - X\hat{\beta})^T(Y - X\hat{\beta}) = \min_{\beta \in \mathbb{R}^p} (Y - X\beta)^T(Y - X\beta)$

Rmk: For $\Lambda^T\beta$. LSE is $\Lambda^T\hat{\beta}$.

Thm. $\hat{\beta}$ is LSE of $\beta \Leftrightarrow MY = X\hat{\beta}$. $M = P_{C(X)|C(X)^\perp}$

Pf: $(Y - X\beta)^T(Y - X\beta) = (Y - MY)^T(Y - MY) + (MY - X\beta)^T(MY - X\beta)$

Cor. $\hat{\beta} = (X^TX)^- X^TY$ is LSE of β.

Rmk: LSE of identifiable function $f(\beta)$ is unique. Since $X\hat{\beta}_1 = MY = X\hat{\beta}_2 \Rightarrow f(\hat{\beta}_1) = f(\hat{\beta}_2)$.

Cor. ℓ^TMY is the unique LSE of $\ell^TX\beta$.

Rmk: Sometimes choose $\ell \in C(X)$. Then $\ell^TMY = \ell^TY$. which can reduce calculation.

Thm. $\lambda^T = \ell^TX \Leftrightarrow \lambda^T\hat{\beta}_1 = \lambda^T\hat{\beta}_2$ if $X\hat{\beta}_1 = MY = X\hat{\beta}_2$

Pf: ($\Leftarrow$) $\lambda = X^T\ell_1 + (I - N)\ell_2$. $N = P_{C(X^T)|C(X^T)^\perp}$

From $\lambda^T(\hat{\beta}_1 - \hat{\beta}_2) = 0$ and $\ell_1^TX\hat{\beta}_1 = \ell_1^TX\hat{\beta}_2$

So $e_2^T(I-N)(\hat{\beta}_1-\hat{\beta}_2)=0$

Since $\hat{\beta}_2=\hat{\beta}_1+v.\ v\in C^{\perp}(X^T)$. is still LSE.

$\therefore\ e_2^T(I-N)v = e_2^T v=0.\quad e_2\in C(X^T)$

$\therefore\ \lambda = X^T e_1$, i.e. $\lambda^T\beta$ is estimable.

Rmk: i) $\lambda^T\beta$ has unique LSE $\Leftrightarrow$ It's estimable.

ii) $e^T MY$ is unbiased linear estimator of $\lambda^T\beta$. $\lambda^T=e^T M$.

Under $Cov(e)=\sigma^2 I$. We consider estimation on σ^2.

Note that
$$\begin{cases}\hat{e} = Y - X\hat{\beta} = (I-M)Y \\ (I-M)Y = (I-M)e\end{cases}$$

It's reasonable to estimate σ^2 by $(I-M)Y$

Thm. $r(X)=r$. $Cov(e)=\sigma^2 I$. $\Rightarrow Y^T(I-M)Y/(n-r)$ is unbiased estimate of σ^2.

Pf: $E(Y^T(I-M)Y) = \sigma^2 tr(I-M) = \sigma^2(n-r)$

Rmk: $(I-M)Y$ is residual vector. $\frac{Y^T(I-M)Y}{n-r}$ is called mean square error (MSE).

(3) Best linear Estimator:

Def: a^TY is best linear unbiased estimator (BLUE) of $\lambda^T\beta$. if $E(a^TY)=\lambda^T\beta$. $Var(a^TY)\le Var(b^TY)$ for $\forall b\in\mathbb{R}^n$. $E(b^TY)=\lambda^T\beta$.

Thm. (Gauss - Markov)

For $Y = X\beta + e$. $E(e) = 0$. $Var(e) = \sigma^2 I$.

If $\lambda^T\beta$ is estimable. Then LSE of $\lambda^T\beta$ is BLUE of $\lambda^T\beta$.

Cor. If $\sigma^2 > 0$. Then there exists unique BLUE for any estimable func. $\lambda^T\beta$.

Pf. i) $Var(a^T Y) = Var(e^T M Y) + Var(a^T Y - e^T M Y)$

ii) if $a^T Y$ is BLUE of $\lambda^T\beta$. then:

$$Var(a^T - e^T M)Y = \sigma^2 (a^T - e^T M)(a - Me) = 0$$

(4) UMVUE:

① Maximum Likelihood Estimation:

Consider: $Y = X\beta + \varepsilon$. $\varepsilon \sim N_n(0, \sigma^2 I)$. $r(X) = r$.

$\Rightarrow Y \sim N_n(X\beta, \sigma^2 I)$. We have:

$$L(\beta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{(Y - X\beta)^T (Y - X\beta^T)}{2\sigma^2}}$$. Let $l = \log L$.

We obtain: $\begin{cases} \hat{\beta} \text{ satisfies } MY = X\hat{\beta}. \text{ (LSE)} \\ \hat{\sigma}^2 = \dfrac{Y^T (I - M) Y}{n} \text{ (asympotic unbiased)} \end{cases}$

② UMVUE:

We have proved $e^T X\hat{\beta}$ is unique BLUE

Next. if under $Y = X\beta + \varepsilon$. $\varepsilon \sim N_n(0, \sigma^2 I)$.

Then $\ell^T M Y$ is UMVUE particularly.

Lemma. $\vec{\theta} = (\theta_1 \cdots \theta_s)^T$. $\vec{Y} \sim c(\theta) h(Y) e^{[\sum_1^s \theta_i T_i(Y)]}$

Then $T(Y) = (T_1(Y) \dots T_s(Y))$ is complete.

sufficient statistics if $P(\theta)$. $D(T(Y))$ contains

a open neighbour.

If $r(X) = r$. Then $\exists Z \in M^{n \times r}$. $Z = (Z_1 \cdots Z_r)$. where

$\{Z_i\}_1^r$ is basis of $C(X)$. $X = Z \cdot A$. $A \in M^{r \times p}$.

For $\lambda^T \beta = \ell^T X \beta = \ell^T Z A \beta$. Set $\gamma = A\beta$.

Consider $Y = Z\gamma + \varepsilon$. $\varepsilon \sim N_n(0, \sigma^2 I_n)$. Still,

LSE of $\lambda^T \beta$ is: $\ell^T M Y$

Rmk: reset $\gamma = A\beta$ is for breaking the linear

restriant of X. Then exists open neighbour.

$$\Rightarrow Y \sim (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{(Y - Z\gamma)^T (Y - Z\gamma)}{2\sigma^2}}$$

$$= C_2(\gamma, \sigma^2)\, e^{-\frac{Y^T Y}{2\sigma^2} + \frac{\gamma^T Z^T Y}{\sigma^2}}$$

satisfies condition of Lemma. w.r.t $(-\frac{1}{2\sigma^2}, \frac{\gamma^T}{\sigma^2})$

and $(Y^T Y, Y^T Z)^T$. so is complete. sufficient.

$\Rightarrow \ell M Y = \ell Z (Z^T Z)^{-1} Z^T Y = f(Z^T Y)$ so UMVUE.

Rmk: $Y^T (I - M) Y = Y^T Y - Y^T Z (Z^T Z)^{-1} (Z^T Y) = f(Y^T Y, Z^T Y)$

so $Y^T (I - M) Y / n-r$ is UMVUE of σ^2.

Thm. $Y^T(I-M)Y/(n-r)$. $e^T MY$ is UMVUE of

σ^2. $e^T X\beta$ respectively. if $\varepsilon \sim N_n(0, \sigma^2 I_n)$

(5) Generalized Least Square:

Consider $Y = X\beta + \varepsilon$. $E(\varepsilon) = 0$. $Cov(\varepsilon) = \sigma^2 V$. $V > 0 \cdots (*)$

Suppose $V = aa^T$. Set $\tilde{Y} = a^{-1}Y$. $\tilde{X} = a^{-1}X$. $\tilde{\varepsilon} = a^{-1}\varepsilon$.

$\Rightarrow \tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon}$. $E(\tilde{\varepsilon}) = 0$. $Cov(\varepsilon) = \sigma^2 I \cdots (\Delta)$

$(\tilde{Y} - \tilde{X}\beta)^T(\tilde{Y} - \tilde{X}\beta) = (Y - X\beta)^T V^{-1}(Y - X\beta)$

Thm. i) $\lambda^T\beta$ is estimable in $(*) \Leftrightarrow$ in (Δ)

ii) $\hat{\beta}$ is generalized LSE if:

$X\hat{\beta} = X(X^T V^{-1} X)^- X^T V^{-1} Y$

iii) generalized LSE: $\lambda^T\hat{\beta}$ for estimable $\lambda^T\beta$ is BLUE.

iv) If $\varepsilon \sim N(0, \sigma^2 V)$. for $\lambda^T\beta$ estimable Then $\lambda^T\hat{\beta}$ (LSE) is UMVUE.

v) If $\varepsilon \sim N(0, \sigma^2 V)$. Then any generalized LSE $\hat{\beta}$ is MLE of β.

Pf. i) $\lambda^T = \ell^T X = (\ell^T a)(a^{-1}X)$.

ii) $\tilde{X}\hat{\beta} = \tilde{X}(\tilde{X}^T\tilde{X})^-\tilde{X}^T\tilde{Y}$

$\Rightarrow X\hat{\beta} = X(X^T V^{-1} X)^- X^T V^{-1} Y$.

iii). iv) is trivial since a is inevitable

For estimator (unbiased) τ. for $E(Y)$.

Then $\bar{a}^{1}\tau$ is unbiased estimator for $E(\tilde{Y})$

$Var(\hat{\beta}) \leq Var(\bar{a}^{1}\tau) \Leftrightarrow Var(a\hat{\beta}) \leq Var(\tau)$

v). $Y \sim (2\pi\sigma^2|V|)^{-\frac{n}{2}} e^{-\frac{(Y-X\beta)^T V^{-1}(Y-X\beta)}{2}}$

<u>Thm.</u> i) $A = X(X^T V^{-1} X)^- X^T V^{-1}$ is indept with choice of generalized inverse.

ii) A is proj. on $C(X)$.

<u>Pf.</u> i) set $B = V^{-\frac{1}{2}}X$. $\Rightarrow A = V^{\frac{1}{2}} B(B^T B)^- B^T V^{-\frac{1}{2}}$

ii) $V = aa^T$. consider $P = P_{C(a^{-1}X)|C^{\perp}(a^{-1}X)}$

$\therefore P = a^{-1}X(X^T V^{-1} X)^- X^T a^{T-1}$

$\therefore P a X = a X \Rightarrow AX = X$.

① <u>Thm.</u>

For $V > 0$. (under (*)). $C(VX) \subset C(X)$

$\Leftrightarrow$ LSE is BLUE. i.e. $\ell^T M Y = \ell^T A Y$.

<u>Pf.</u> <u>Lemma.</u> $C(VX) = C(X) \Leftrightarrow C(X) = C(V^{-1}X)$

which concludes $C(X)^{\perp} = C(V^{-1}X)^{\perp}$

<u>Pf.</u> $\exists B_1, B_2$. $XB_1 = VX \Rightarrow V^{-1}XB_1 = X$

$VXB_2 = X \qquad XB_2 = V^{-1}X$

$\therefore C(V^{-1}X) = C(X)$.

($\Rightarrow$) Show: $A = P_{C(X)|C(X)^{\perp}}$. i.e. $N(A) = C(X)^{\perp}$

$r(X) = r(VX) \Rightarrow C(VX) = C(X)$. So $C(X) = C(V^{-1}X)$.

Check $w \in C(X)^\perp = C(V^{-1}X)^\perp$. $AX = 0$.

$(\Leftarrow)$ $c^T MY = c^T AY \Leftrightarrow A = M \Leftrightarrow N(A) = C^\perp(X)$

prove $= N(A) = C^\perp(VX)$.

$AX = X(X^T V^{-1} X)^- X^T V^{-1} X = 0$ i.e.

$P_{C(V^{-\frac{1}{2}}X) \mid C^\perp(V^{-\frac{1}{2}}X)} (V^{-\frac{1}{2}}X) = 0$ i.e.

$V^{-\frac{1}{2}}X \perp C(V^{-\frac{1}{2}}X) \Leftrightarrow X^T V^{-1} x = 0$

$\therefore N(A) = C^\perp(VX) \Rightarrow C(VX) = C(X)$.

<u>Rmk</u>: It said generalized LSE is ordinary LSE

$\Leftrightarrow C(X) = C(VX)$ for estimable parameter.

<u>Cor.</u> If X has full rank p. $\tilde{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$

$\hat{\beta} = (X^T X)^{-1} X^T Y$. Then. $\hat{\beta} = \tilde{\beta} \Leftrightarrow C(X) = C(VX)$.

<u>Pf.</u> $(\Rightarrow)$ Set $T_1 = (X^T V^{-1} X)^{-1}$. $T_2 = (X^T X)^{-1}$. full rank

Simplify $\hat{\beta} = \tilde{\beta}$:

$\Rightarrow V^{-1} X T_1 = X T_2 \Rightarrow V^{-1} X = X T_2 T_1^{-1} =: X T_3$

$\therefore X = VXT_3$, $|T_3| \neq 0$. $\therefore C(X) = C(VX)$.

$(\Leftarrow)$ $X\hat{\beta} = MY = P_{C(X) \mid C^\perp(X)}(Y) = AY$

$= X\tilde{\beta}$ by $N(A) = C(X)^\perp \Rightarrow \hat{\beta} = \tilde{\beta}$.

<u>Lemma.</u> $(Y - X\beta)^T V^{-1} (Y - X\beta) = (Y - AY)^T V^{-1} (Y - AY) +$

$(\tilde{\beta} - \beta)^T X^T V^{-1} X (\tilde{\beta} - \beta)$

where $X\tilde{\beta} = AY$.

② Estimation of σ^2:

In $\tilde{Y} = \tilde{X}\beta + \tilde{\varepsilon}$, we have unbiased estimator of σ^2 is: $\hat{\sigma}^2 = \dfrac{Y^T (Q^T)^{-1} (I - M^*) Q^{-1} Y}{n-r}$

where $M^* = Q^{-1} X (X^T V^{-1} X)^- X^T (Q^T)^{-1}$, $r = r(X)$.

By reduction: $(I - M^*) Q^{-1} = Q^{-1} (I - A)$

$\Rightarrow \hat{\sigma}^2 = \dfrac{Y^T (I-A)^T V^{-1} (I-A) Y}{n-r}$.

Thm. $V^{-1}(I-A) = (I-A)^T V^{-1} (I-A) = (I-A)^T V^{-1}$

Pf: $\Longleftrightarrow A^T V^{-1} = A^T V^{-1} A$.

$\because A$ is indept with choice of $(X^T V^{-1} X)^-$

Choose it's M-P inversed.

$\Rightarrow$ it's easy to check it!

Cor. reduce $\hat{\sigma} = \dfrac{Y^T V^{-1} (I-A) Y}{n-r}$

(b) Sampling Dist. of estimator:

① For $Y = X\beta + \varepsilon$, $\varepsilon \sim N_n(0, \sigma^2 I)$, $\ell \in M^{n \times s}$.

i) $E(\ell^T M Y) = \ell^T X \beta$

ii) $\mathrm{Cov}(\ell^T M Y) = \sigma^2 \ell^T M \ell = \sigma^2 \Lambda^T (X^T X)^- \Lambda$. $\Lambda^T = \ell^T X$.

$\therefore\ \ell^T M Y = \ell^T X\hat{\beta} = \Lambda^T\hat{\beta} \sim N_s(\ell^T X\beta,\ \sigma^2\ell^T M\ell)$

$= N_s(\Lambda^T\beta,\ \sigma^2\Lambda^T(X^TX)^-\Lambda)$

Rmk: i) If $\ell = I_n$. Then $MY = X\hat{\beta} \sim N_n(X\beta, \sigma^2 M)$

ii) $|X^TX| \neq 0 \Rightarrow \hat{\beta} \sim N_p(\beta, \sigma^2(X^TX)^{-1})$

For $\dfrac{Y^T(I-M)Y}{n-r}$:

since $Y \sim N_n(X\beta, \sigma^2 I)$. $\therefore \dfrac{Y^T(I-M)Y}{\sigma^2} \sim \chi^2(n-r, \gamma)$

$\gamma = \dfrac{\beta^T X^T(I-M)X\beta}{\sigma^2} = 0$. $\therefore Y^T(I-M)Y \sim \sigma^2\chi^2_{n-r}$.

② Consider $Y = X\beta + \varepsilon$. $\varepsilon \sim N_n(0, \sigma^2 V)$. $V > 0$. $\ell \in M^{n\times s}$

i) $E(\ell^T AY) = \ell^T X\beta$.

ii) $Cov(\ell^T AY) = \sigma^2\ell^T AVA^T\ell$

Lemma. i) $AVA^T = AV = VA^T$

ii) $A^TV^{-1}A = A^TV^{-1} = V^{-1}A$

Pf: Since A is indept with choice of $(X^TV^{-1}X)^-$

choose it's M.P inverse.

$\Rightarrow \ell^T AY = \ell^T X\hat{\beta} \sim N_s(\ell^T X\beta,\ \sigma^2\ell^T X(X^TV^{-1}X)^- X^T\ell)$

since $AVA^T = AV = X(X^TV^{-1}X)^- X^T$.

Denote $\lambda^T = \ell^T X$. $\lambda^T\hat{\beta} \sim N_s(\lambda^T\beta,\ \sigma^2\lambda^T(X^TV^{-1}X)^-\lambda)$

Rmk: To obtain confidence region of $\lambda^T\beta$:

$$\frac{\lambda^T\hat{\beta} - \lambda^T\beta}{\sqrt{MSE\,\lambda^T(X^TV^{-1}X)^-\lambda}} = \frac{\lambda^T\hat{\beta} - \lambda^T\beta}{\sqrt{\sigma^2\lambda^T(X^TV^{-1}X)^-\lambda}} \Big/ \sqrt{MSE/\sigma^2} \sim t(n-r(X))$$