STAT 8004 Lecture 4

Thursday, February 12, 2015 5:33 PM

Review: $\underset{n\times 1}{Y} = \underset{n\times p}{X}\underset{p\times 1}{\beta} + \underset{n\times 1}{\epsilon}$ $\quad E(\epsilon)=0$

$E(Y)=X\beta \iff E(Y) \in C(X)$

OLS $\quad \underset{\hat{Y}\in C(X)}{\text{minimize}} \ \|Y-\hat{Y}\|_2^2$

$\iff \underset{b}{\text{minimize}} \quad Q(b) = (Y-Xb)^T(Y-Xb)$

$0 = \frac{\partial Q}{\partial b} = -X^T(Y-Xb)$

$\iff -X^TY + X^TXb = 0$

(normal equation) $\iff (X^TX)b = X^TY \quad = 0 \quad (1)$

If (X^TX) is not of full rank, there are infinity many b satisfying the normal equation.

Now: $\quad b = \underline{(X^TX)^- X^TY}$ is a solution satisfying (1).

\- G is generalized inverse of A, if $AGA = A$.

Optional work, verify $b = (X^TX)^-X^TY$ satisfying (1).

\- But, $\hat{Y} = Xb = \underbrace{X(X^TX)^-X^T}_{= P_{n\times n}}Y \triangleq PY$ is always unique.

Now, $a^T\hat{Y}$ is always estimable

and $E(a^T\hat{Y}) = a^TPX\beta$

$= a^TX\beta$

Def: $c^T\beta$ is estimable iff $c^T = a^TX$

$\iff c \in C(X^T)$

OLS $\quad \widehat{c^T\beta} = c^T(X^TX)^-X^TY$

estimator " $\widehat{a^T Y}$ " $\sim$ $\hookrightarrow$ linear.
$= a^T P Y$

—x—

- Optimality: $var(\epsilon) = \sigma^2 I$ (GM)
  then $\widehat{c^T\beta}$ is BLUE for $c^T\beta$.

  If $var(\epsilon) = \sigma^2 V$ (V is known)
  then $U = V^{-1/2} Y$, $W = V^{-1/2} X$, $\epsilon^* = V^{-1/2}\epsilon$
  then $\widehat{c^T\beta}$ based on $U = W\beta + \epsilon^*$
  is BLUE for $c^T\beta$.

- If V is unknown, one would need "likelihood" theory for asymptotic optimality.

—x—

Normal theory inference. { classical, useful, reasonably robust.

- Building blocks: linear model + multivariate normal distribution

$$Y = X\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I)$$

- R&S Chapter 4 & 5

- Def: The χ^2_n distribution is the distribution of
  $Z^T Z = \sum_{i=1}^n z_i^2$ for $Z \sim MVN_n(0, I)$.
  (iid sum of standard normal r.vs)
  chisq

- Def: If $Y \sim MVN_n(\mu, I)$, then $Y^T Y$ is called the noncentral χ^2_n with noncentrality parameter

$$\lambda = \delta^2 = \mu^T \mu \quad (\text{squared length of } \mu).$$

$$(\text{R\&S uses } \delta^2 = \tfrac{1}{2}\mu^T\mu \qquad)$$

Remark: every μ with the same $\mu^T\mu$ produces the same distribution of Y^TY.

e.g. $$(Z_1+\delta)^2 + Z_2^2 + \cdots Z_n^2 \sim \chi^2_n(\lambda)$$

Theorem: Suppose $\underset{n\times n}{A}$ symmetric with $\underline{\text{rank}(A) = k}$.

$Y \sim MVN_n(\mu, \Sigma)$ for positive definite Σ.

If $A\Sigma$ is $\underline{\text{idempotent}}$ $(A\Sigma A\Sigma = A\Sigma)$.

then $$Y^TAY \sim \underline{\chi^2_k(\mu^TA\mu)}.$$

So, if $A\mu = 0$, then $Y^TAY \sim \chi^2_k$.

(R&S, Theorem 5.5.).

Example: $$Y \sim MVN_n(X\beta, \underline{\underline{\sigma^2 I}})$$

consider $$\frac{SSE}{\sigma^2} = \frac{(Y-\hat{Y})^T(Y-\hat{Y})}{\sigma^2}$$

$$= \frac{1}{\sigma^2}((I-P)Y)^T((I-P)Y)$$

$$= \underbrace{\frac{1}{\sigma^2}\, Y^T \underbrace{(I-P)}}_{A} Y^T$$

Now, $\frac{1}{\sigma^2}(I-P)\cdot\sigma^2 I = (I-P)$ is idempotent. $(A\Sigma)$ idempotent

and. $\underbrace{(I-P)E(Y)}_{A\mu=0} = (I-P)X\beta = 0$

$$\Rightarrow \frac{SSE}{\sigma^2} \sim \chi^2_{\text{rank}(I-P)} = \chi^2_{n-\text{rank}(X)}$$

Application:

$$P\left(\begin{matrix}\text{lower } \alpha/2 \text{ quantile} \\ \text{of } \chi^2_{n-\text{rank}(X)}\end{matrix} < \frac{SSE}{\sigma^2} < \begin{matrix}\text{upper } \alpha/2 \text{ quantile} \\ \text{of } \chi^2_{n-\text{rank}(X)}\end{matrix}\right) = 1-\alpha$$

$$\Longleftrightarrow P\left(\frac{SSE}{\begin{matrix}\text{upper } \alpha/2 \text{ q}^{\text{t}} \\ \text{of } \chi^2_{n-\text{rank}(X)}\end{matrix}} < \sigma^2 < \frac{SSE}{\begin{matrix}\text{lower } \alpha/2 \text{ q}^{\text{t}} \\ \text{of } \chi^2_{n-\text{rank}(X)}\end{matrix}}\right) = 1-\alpha.$$

These give a $(1-\alpha)$-level confidence limits for σ^2.

Remark: these confidence limits are "exact", i.e. they are valid as long as the quantities are well defined.

$H_0: \sigma^2 = \nu^2$

(testing and confidence set based on SSE & $\chi^2_{n-\text{rank}(X)}$)

—x—

Theorem: suppose $Y \sim MVN_n(\mu, \sigma^2 I)$, and $BA = 0$ (B & A of appropriate size)

a) If A is symmetric, $Y^T A Y$ and BY are independent.

b) If both B and A are symmetric, then $Y^T A Y$, and $Y^T B Y$ are independent.

(weaker form of Theorems 5.6a 5.6b of R&S)

pf:

$$\begin{pmatrix} A \\ B \end{pmatrix} Y = \begin{pmatrix} AY \\ BY \end{pmatrix}$$

$$\text{var}\left(\begin{pmatrix} A \\ B \end{pmatrix} Y\right) = \begin{pmatrix} A \\ B \end{pmatrix} \underbrace{\text{var}(Y)}_{=\sigma^2 I} (A^T, B^T)$$

$$= \sigma^2 \begin{pmatrix} AA^T & AB^T \end{pmatrix}$$

$$= \sigma^2 \begin{pmatrix} AA^T & AB^T \\ BA^T & BB^T \end{pmatrix}$$

If A is symmetric $\quad BA^T = BA = 0$

$$= \sigma^2 \begin{pmatrix} AA^T & 0 \\ 0 & BB^T \end{pmatrix}$$

Fact: uncorrelated normal rvs are independent

$\Rightarrow$ AY & BY are independent

$\Rightarrow$ Any function of AY & BY are independent

$$\underbrace{Y^T A Y}_{\text{is a function of } AY} = Y^T A A^- A Y = \underbrace{(AY)^T}_{A \text{ is symmetric}} A^- (AY)$$

$\Rightarrow$ $Y^T A Y$ & BY are independent.

proof of (b) is the same if B is symmetric.

Example/Application: $\quad Y \sim MVN_n(X\beta, \sigma^2 I)$

Take $B = P$, $\quad A = (I - P) \quad (BA = 0)$

The theorem says. $\quad Y^T A Y = Y^T (I-P) Y = \underline{\underline{SSE}}$

$BY = PY = \underline{\underline{\hat{Y}}}$

$\Rightarrow$ SSE & $\hat{Y}$ are independent.

Now: If $c^T\beta$ is estimable. ($c^T = a^T X$ for some a)

and $\widehat{c^T\beta} = a^T \hat{Y}$

$\Rightarrow$ $\widehat{c^T\beta}$ and SSE are independent.

$$\widehat{c^T\beta} \sim N(c^T\beta, \ \sigma^2 c^T (X^T X)^- c)$$

was shown last time

Definition: the t-distribution with ν d.f. is the distribution of $Z/\sqrt{\frac{W}{\nu}}$ for $Z \sim N(0,1)$ and $W \sim \chi^2_\nu$. Z & W are independent.

$$\underbrace{\frac{\widehat{c^T\beta} - c^T\beta}{\sqrt{\cancel{\sigma^2}\, c^T(X^TX)^- c}}}_{\sim N(0,1)} \Bigg/ \sqrt{\frac{\frac{SSE}{\cancel{\sigma^2}}}{n-\text{rank}(X)}} \qquad \frac{SSE}{\sigma^2} \sim \chi^2_{n-\text{rank}(X)}$$

$$\sim t_{n-\text{rank}(X)}$$

$$\frac{\widehat{c^T\beta} - c^T\beta}{\sqrt{MSE}\,\sqrt{c^T(X^TX)^- c}} \sim$$

Example / Application. $H_0: c^T\beta = \#$

Then we use

$$T = \frac{\widehat{c^T\beta} - \#}{\sqrt{MSE}\,\sqrt{c^T(X^TX)^- c}} \quad \text{and}$$

use $t_{n-\text{rank}(X)}$ as the null distribution, rejecting H_0, for "extreme" value of T.

Further, use $t^* =$ upper $\alpha/2$ quantile of $t_{n-\text{rank}(X)}$

then $$P\left(-t^* < \frac{\widehat{c^T\beta} - c^T\beta}{\sqrt{\ }\ \sqrt{\ }} < t^*\right) = 1-\alpha$$

$$\Longleftrightarrow P\left(\widehat{c^T\beta} - t^*\sqrt{\ }\sqrt{\ } < c^T\beta < \widehat{c^T\beta} + t^*\sqrt{\ }\sqrt{\ }\right)$$

$$\Longleftrightarrow P\Big(\underbrace{\widehat{c^T\beta} - t^*\sqrt{\ }\sqrt{\ }}\ < c^T\beta < \underbrace{\widehat{c^T\beta} + t^*\sqrt{\ }\sqrt{\ }}\Big)$$

$$= 1-\alpha.$$

A $(1-\alpha)$-level confidence limits for $c^T\beta$.

Remark: this is "exact"
$(Y = X\beta + \epsilon \quad \epsilon \sim N(0, \sigma^2 I))$

—— x ——

Further example / application = prediction

Suppose $c^T\beta$ is estimable. and y^* (independent of Y) normal with $\boxed{E(y^*) = c^T\beta \text{ and } Var(y^*) = \gamma\sigma^2}$ for known γ.

For example: if c^T is a row of X, and $\gamma = \frac{1}{m}$

y^* might be the sample mean of m future obs. at conditions given by that row of X.

or in a regression context, X is of full rank

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

then every $c^T\beta$ is estimable. one might be interested in

$$\underline{c^T\beta} = (0, x_1 - x_1', x_2 - x_2', \cdots x_k - x_k') \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

representing the difference in the mean of responses between conditions $(1, x_1, x_2 \ldots x_k)$
and $(1, x_1', x_2', \ldots x_k')$

- If $\gamma = 2$, y^* might be the the difference between the two responses at different conditions

or

$$\vec{c}^T\beta = (2, x_1+x_1', x_2+x_2', \cdots x_k+x_k') \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

then y^* might be the sum of two new observations at those two sets of conditions

Now: $$\underline{\underline{\widehat{\vec{c}^T\beta}}} - \underline{\underline{y^*}} \sim N\left(0, \sigma^2\left(\gamma + \vec{c}^T(X^TX)^{-}\vec{c}\right)\right)$$

and it is independent of SSE, so

$$\boxed{\frac{\widehat{\vec{c}^T\beta} - y^*}{\sqrt{MSE}\sqrt{\gamma + \vec{c}^T(X^TX)^{-}\vec{c}}}} \sim t_{n-\text{rank}(X)}$$

$$P(-t_* < \quad\downarrow\quad < t_*) = 1-\alpha$$

$$\Rightarrow P\left(\widehat{\vec{c}^T\beta} - t_*\sqrt{\ }\sqrt{\ } < y^* < \widehat{\vec{c}^T\beta} + t_*\sqrt{\ }\sqrt{\ }\right) = 1-\alpha.$$

The prediction limits for y^* is

$$\widehat{\vec{c}^T\beta} \pm t_*\sqrt{MSE}\sqrt{\gamma + \vec{c}^T(X^TX)^{-}\vec{c}}$$

$\uparrow$

— x —

- Summary: - model based prediction.
  - $\widehat{\vec{c}^T\beta}$ is the best.

—, —

$\boxed{\underline{\underline{Y}}, \quad \underline{\underline{X}}} \to \hat{c\beta}, \hat{\sigma}^2 \to$ hypothesis tests, confidence limits

$\boxed{c^T}(X^TX)^- X^T Y$

- More on the normal based GM model.

Testing: $H_0: \boxed{\underline{\underline{C\beta}} = d}$ for $\left(C = \begin{pmatrix} c_1^T \\ c_2^T \\ \vdots \\ c_l^T \end{pmatrix}_{l \times k} \right)$

$c_1^T, c_2^T, \cdots c_l^T$ are corresponding to estimable functions, $\text{rank}(C) = l$

$\Leftrightarrow$ (exists an A, such that $C = \underline{\underline{A}}X$)

OLS estimator: $\widehat{C\beta} = A\hat{Y}$

Since, $\hat{Y}$ and SSE are independent,

$\underbrace{\widehat{C\beta}}$ and $\underbrace{SSE}$ are also independent.

Rationale: building a test for $H_0: C\beta = d$ with

$\underbrace{\widehat{C\beta} - d}_{l \times 1}$ & SSE

$\sim$

$MVN(\underbrace{C\beta - d}, \underbrace{\sigma^2 C(X^TX)^- C^T})$

Consider. $\underbrace{(\widehat{C\beta} - d)}^T \underbrace{(\sigma^2 C(X^TX)^- C^T)^{-1}} \underbrace{(\widehat{C\beta} - d)}$.

$\text{rank}(C) = l$.

as a discrepancy measure

$\operatorname{rank}(C) = \ell$.

as a discrepancy measure

Notice $\boxed{(\cancel{\sigma^2} \underline{C(X^TX)^- C^T})^{-1} \cancel{\sigma^2} \underline{C(X^TX)^- C^T}} = I$

is idempotent.

Then, $\frac{1}{\sigma^2} (C\hat{\beta} - d)^T \boxed{(C(X^TX)^- C^T)^{-1}} (C\hat{\beta} - d)$

SSH_0 $=$

$\sim \chi^2_\ell(\delta^2)$

$$\boxed{\delta^2 = \frac{1}{\sigma^2} (C\beta - d)^T (C(X^TX)^- C^T)^{-1} (C\beta - d)}$$

If H_0 is true, $\delta^2 = 0$, o.w. SSH_0 will be "big".

we already know $SSE/\sigma^2 \sim \chi^2_{n - \operatorname{rank}(X)}$

So, comparisons between SSH_0/σ^2 & SSE/σ^2

seems plausible for testing $H_0: C\beta = d$.

Def: If $U \sim \chi^2_{r_1}$, independent of $V \sim \chi^2_{r_2}$

then $\dfrac{U/r_1}{V/r_2}$ is called F distribution.

with d.f.s r_1, r_2. (Snedecor's F)

Def: If $U \sim \chi^2_{r_1}(\lambda)$, & $V \sim \chi^2_{r_2}$, then

$\dfrac{U/r_1}{V/r_2}$ is called the noncentral F distribution

with noncentrality parameter λ.

Define: $F = \dfrac{\frac{1}{\sigma^2} SSH_0/\ell}{\frac{1}{\sigma^2} SSE/ \ldots} = \dfrac{MSH_0}{MSE}$

Define: $\boxed{F = \dfrac{1/\sigma^2 \quad\cdots\quad /\ell}{1/\sigma^2\ SSE/n-\text{rank}(X)} = \dfrac{MSH_0}{MSE}}$

and conclude F is $F_{\ell,\, n-\text{rank}(X)}$ with

noncentrality parameter $\underline{\underline{\lambda = \delta^2}}$

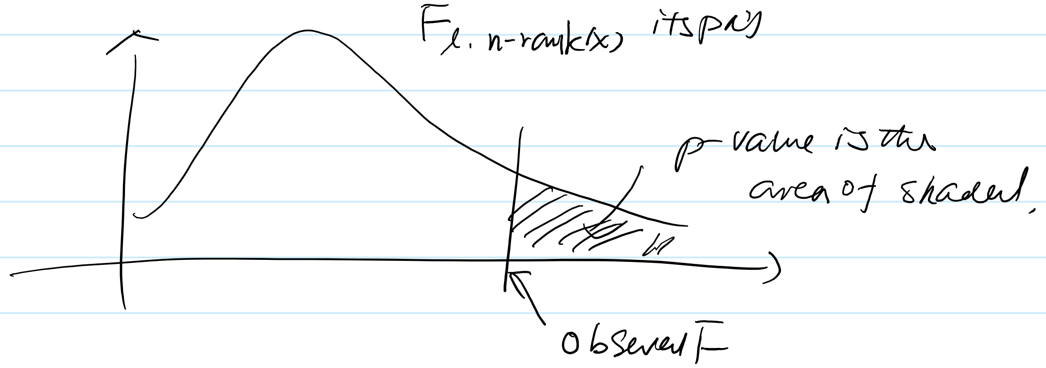
So, an α level test of $H_0: C\beta = d$ can be done by rejecting H_0 if

$$\underline{\underline{F}} > \text{upper}_\alpha \text{ quantile of } F_{\ell,\, n-\text{rank}(X)}$$

with centrality parameter 0.

or, a p-value for this hypothesis is

$\left(\begin{array}{c}\text{central F distribution} \\ \text{probability with d.f.} \\ \ell,\ n-\text{rank}(X)\end{array}\right)$ to right of the observed F



- H_0: $\boxed{C\beta = d}$ ←

- power of test.

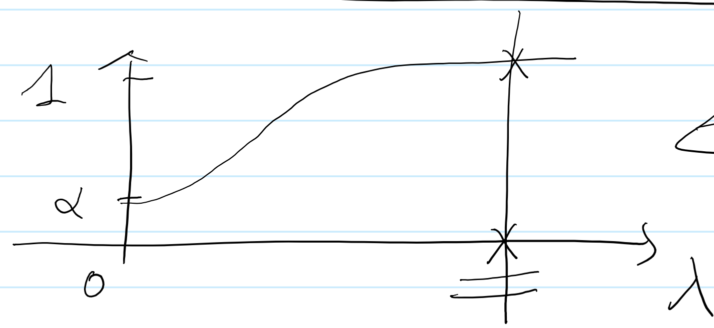
  I can use the noncentral version of the F distribution to evaluate the power of the test.

  (α-level test, $\boxed{F_{\alpha,\ell,\, n-\text{rank}(X)}(0)}$ is the critical)

($\underline{\underline{\alpha}}$- level test, $\boxed{F_{\alpha, \ell, n-\text{rank}(X)}(0)}$ is the critical value)

power = P (rejecting H_0)

$= P\left(\underline{\underline{\text{a noncentral } F_{\ell, n-\text{rank}(X)}(\lambda)}} > \downarrow \right)$



- All previous procedures are "exact"
  $Y \sim MVN(X\beta, \sigma^2 I)$
  provided all quantities are well-defined.

— ℓ —

- Normal theory and "maximum likelihood"

Def: suppose that a random vector $\underline{\underline{U}}$ has pmf or pdf $f(u \mid \theta)$. If $U = u$ is observed /realized and $f(u \mid \hat{\theta}) = \max_{\theta} \underline{\underline{f(u \mid \theta)}}$

we will call $\hat{\theta}$ a MLE. of θ.

For $Y \sim MVN_n(X\beta, \sigma^2 I)$ (GM)

$\rightarrow f(y \mid X\beta, \sigma^2) = \boxed{(2\pi)^{-n/2} \, |\sigma^2 I|^{-1/2}}$

$\exp\left(-\frac{1}{2} (Y - X\beta)^T (\sigma^2 I)^{-1} (Y - X\beta)\right)$

any

$\exp(-\frac{1}{2}\underbrace{(Y-X\beta)'(\sigma^2 I)^{-1}(Y-X\beta)}_{\|})$

for any σ^2 fixed,

$\exp(-\frac{1}{2\sigma^2}\underline{(Y-X\beta)^T(Y-X\beta)})$

maximize $f(y|X\beta)$

is equivalent to

minimize $(Y-X\beta)^T(Y-X\beta)$

or, $(Y-\hat{Y})^T(Y-\hat{Y})$ s.t. $\hat{Y}\in C(X)$,

$\Rightarrow$ MLE of $\widehat{X\beta}$ is equivalent to OLS.

Consider. $\log f(y|\hat{Y},\sigma^2)$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\underbrace{(Y-\hat{Y})^T(Y-\hat{Y})}_{\substack{\| \\ SSE}}$$

$$\frac{d}{d\sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{SSE}{2\sigma^4}$$

$$\Rightarrow \hat{\sigma}^2_{MLE} = \frac{SSE}{n} = \frac{(n-\text{rank}(X))MSE}{n}$$

$< MSE$

Because MSE is unbiased for σ^2

$\Rightarrow \hat{\sigma}^2_{MLE}$ is biased downwards.

$\boxed{\text{This fact motivates the REML for (restricted MLE) linear mixed models}}$

1

— $\boxed{\hat{Y}}$ using OLS is sound for solving the big class of problems in $Y \sim MVN_n(X\beta, \sigma^2 I)$.

$$\boxed{\begin{array}{c} \left(\sigma^2, \quad c^T\beta, \quad \boxed{C\beta} \right) \\ \chi^2 \qquad t \qquad F \\ \\ SSE = (Y-\hat{Y})^T(Y-\hat{Y}) \\ \widehat{c^T\beta} = c^T(X^TX)^{-}X^TY \end{array}}$$

— x —