15) Let $\underset{g\times 1}{T_n}$ be a statistic, eg $\underline{T}_n = A\hat{\underline{\beta}} = \underset{g\times 1}{\hat{\underline{\theta}}}$

Suppose we want to test $H_0\ \underline{\theta} = \underline{\theta}_0$, eg $\underline{\theta} = A\underline{\beta}$. A large sample $100(1-\delta)\%$ confidence region (CR) for $\underline{\theta}$ is a set A_n such that $P[\underline{\theta} \in A_n] \to 1-\delta$ as $n \to \infty$ (or $\to 1-\tau \geq 1-\delta$ as $n \to \infty$).

Reject H_0 if $\underline{\theta}_0 \notin A_n$.

A CI is a special case of a confidence region if $g = 1$.

16) Assume $\sqrt{n}\,(T_n - \underline{\theta}) \xrightarrow{D} N_g(\underline{0}, \Sigma)$ and

$$\sqrt{n}\,(T_n^* - T_n) \xrightarrow{D} N_g(\underline{0}, \Sigma).$$

Let $\overline{T}^* = \frac{1}{B}\sum_{i=1}^{B} T_i^*$ and $S_T^* = \frac{1}{B-1}\sum_{i=1}^{B}(T_i^* - \overline{T}^*)(T_i^* - \overline{T}^*)^T$

be the sample mean and sample covariance matrix of the bootstrap sample $T_1^*, \ldots, T_B^*$.

Need $B \geq 50g$. For MLR $1 \leq g \leq p$ so need $B \geq 50p$.

see 22

17) Under 16) also assume $nS_T^* \xrightarrow{P} \Sigma$

so $[nS_T^*]^{-1} \xrightarrow{P} \Sigma^{-1}$. The usual large sample $100(1-\delta)\%$ CR for $\underline{\theta}$ is

$$\{\underline{w} : (\underline{w} - T_n)^T [S_T^*]^{-1} (\underline{w} - T_n) \leq \chi^2_{g,1-\delta}\}$$

center of hyperellipsoid

could use $d_2 F_{g,d_2,1-\delta}$

$d_2 \to \infty$ as $n \to \infty$

Since $\sqrt{n}(T_n - \underline{\theta}) \xrightarrow{D} N_g(\underline{0}, \Sigma)$,

$$\sqrt{n}(T_n - \underline{\theta})^T \Sigma^{-1} \sqrt{n}(T_n - \underline{\theta}) \xrightarrow{D} \chi^2_g$$

$$n(T_n - \underline{\theta})^T \hat{\Sigma}^{-1} (T_n - \underline{\theta}) \xrightarrow{D} \chi^2_g \quad \text{if } \hat{\Sigma} \xrightarrow{P} \Sigma$$

$$(T_n - \underline{\theta})^T [S_T^*]^{-1} (T_n - \underline{\theta}) \xrightarrow{D} \chi^2_g,$$

Reject H_0 if $(T_n - \underline{\theta}_0)^T [S_T^*]^{-1} (T_n - \underline{\theta})_0 > \chi^2_{g, 1-\delta}$.

The assumption $n[S_T^*] \xrightarrow{P} \Sigma$ is strong.

18] Modified Bickel and Ren CR

$$\{\underline{w} : (\underline{w} - T_n)^T [S_T^*]^{-1} (w - T_n) \le D^2_{(U_B, T)}\}$$

$$= \{\underline{w} : D^2_{\underline{w}}(T_n, S_T^*) \le D^2_{(U_B, T)}\}$$

$D^2_{(U_B, T)}$ is the $100 g_B$ th sample quantile

of $(T_i^* - T_n)^T (S_T^*)^{-1} (T_i^* - T_n)$, $i = 1, \ldots, B$

where $g_B \downarrow 1 - \delta$.

19] Prediction region method CR

$\{\underline{w} : (\underline{w} - \bar{T}^*)^T (S_T^*)^{-1} (\underline{w} - \bar{T}^*) \le D^2_{(U_B)}\} = \{\underline{w} : D^2_{\underline{w}}(\bar{T}^*, S_T^*) \le D^2_{(U_B)}\}$.

$D^2_{(U_B)}$ is the $100 g_B$th sample quantile of $(T_i^* - \bar{T}^*)^T (S_T^*)^{-1} (T_i^* - \bar{T}^*)$, $i = 1, \ldots, B$.

20) Hybrid CR

$\{\underline{w} : (\underline{w} - T_n)^T (S_T^*)^{-1} (\underline{w} - T_n) \le D^2_{(U_B)}\}$

shifts CR 19} to be centered at T_n

or replaces CR 18} cutoff $D^2_{(U_B,T)}$ with $D^2_{(U_B)}$.

If $\underset{\sim}{U} \sim N_g(\underline{0}, \Sigma)$ these regularity conditions often hold.

21) Suppose $\sqrt{n}(T_n - \underline{\theta}) \xrightarrow{D} \underset{\sim}{U}$, $\sqrt{n}(T_i^* - T_n) \xrightarrow{D} \underline{U}$

$\sqrt{n}(\overline{T}^* - T_n) \xrightarrow{P} \underline{0}$, $\sqrt{n}(\overline{T}^* - \underline{\theta}) \xrightarrow{D} \underline{U}$ and $(n S_T^*)^{-1}$ is not too ill conditioned for large enough n and B. Then

$D_1^2 = D^2_{T_i^*}(\overline{T}^*, S_T^*) = \sqrt{n}(T_i^* - \overline{T}^*)^T (n S_T^*)^{-1} \sqrt{n}(T_i^* - \overline{T}^*)$

$D_2^2 = D^2_{\underline{\theta}}(T_n, S_T^*) = \sqrt{n}(T_n - \underline{\theta})^T (n S_T^*)^{-1} \sqrt{n}(T_n - \underline{\theta})$

$D_3^2 = D^2_{\underline{\theta}}(\overline{T}^*, S_T^*) = \sqrt{n}(\overline{T}^* - \underline{\theta})^T (n S_T^*)^{-1} \sqrt{n}(\overline{T}^* - \underline{\theta})$

$D_4^2 = D^2_{T_i^*}(T_n, S_T^*) = \sqrt{n}(T_i^* - T_n)^T (n S_T^*)^{-1} \sqrt{n}(T_i^* - T_n)$.

So $D_j^2 \overset{D}{\approx} \underline{U}^T (n S_T^*)^{-1} \underline{U}$ for large n, B (percentiles are approx equal)

and CRs 16), 17), and 18) will have coverage near $1-\delta$.

22] Suppose $T_n = \underset{\sim}{\overline{X}} = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i$ where the $\underline{x}_i$ are iid

with $E(\underline{x}_i) = \underline{\mu} = \underline{\theta}$ and $\mathrm{cov}(\underline{x}_i) = \Sigma$. Then $\mathrm{cov}(T_n) =$

$\mathrm{cov}(\overline{x}) = \frac{\Sigma}{n}$. So $n S_T^*$ estimates Σ and $(n S_T^*)^{-1} = \frac{(S_T^*)^{-1}}{n}$

estimates Σ^{-1}.

23) Let the MLR model $\underset{\sim}{Y} = X\underline{\beta} + \underset{\sim}{\epsilon}$, $E(\underset{\sim}{\epsilon}) = \underset{\sim}{0}$, $\underline{Cov}(\underset{\sim}{\epsilon}) = \sigma^2 I$ hold.

Do <u>not</u> assume that $\underset{\sim}{\epsilon} \sim N_n(\underline{0}, \sigma^2 I)$.

Assume $\frac{X'X}{n} \xrightarrow{P} V^{-1}$ so $n(X'X)^{-1} \xrightarrow{P} V$.

<u>parametric bootstrap for MLR</u>

$\underset{\sim}{Y}_j^* = (Y_i^*)$ where $Y_i^* | \underline{x}_i \sim N_1(\underline{x}_i'\hat{\underline{\beta}}, MSE = \hat{\sigma}_n^2)$

or $\underset{\sim}{Y}_j^* \overset{iid}{\sim} N_n(X\hat{\underline{\beta}}, \hat{\sigma}_n^2 I)$ for $j = 1, \ldots, B$.

Then compute OLS on $\underset{\sim}{Y}_j^*, X$

$\hat{\underline{\beta}}_j^* = (X'X)^{-1} X' \underline{Y}_j^* \sim N_p(\hat{\underline{\beta}}, \hat{\sigma}_n^2 (X'X)^{-1})$

$\sqrt{n}(\hat{\underline{\beta}}_j^* - \hat{\underline{\beta}}) \sim N_p(\underline{0}, \hat{\sigma}_n^2\, n(X'X)^{-1}) \xrightarrow{D} N_p(\underline{0}, \sigma^2 V)$

$\sqrt{n}(\hat{\underline{\beta}} - \underline{\beta}) \xrightarrow{D} N_p(\underline{0}, \sigma^2 V)$

by OLS CLT.

24) Under the conditions of 23),

<u>residual bootstrap for MLR</u>

Let $\underset{\sim}{r}^W$ denote an $n \times 1$ random vector selected with replacement from the OLS full model residuals. Then $\underline{Y}^* = X\hat{\underline{\beta}} + \underline{r}^W$

follows a standard MLR model where elements r_i^W of $\underline{r}^W$ are iid from the empirical dist of the OLS full model residuals r_i.

Hence $E(r_i^W) = \frac{1}{n}\sum_{i=1}^n r_i = 0$, $\sigma_n^2 = V(r_i^W) = \frac{1}{n}\sum_{i=1}^n r_i^2 = \frac{n-p}{n} MSE$,

$E(\underline{r}^W) = \underline{0}$, and $Cov(\underline{Y}^*) = Cov(\underline{r}^W) = \sigma_n^2 I_n$

(Note $\sigma_n^2 = MSE$ for parametric bootstrap.)

$X\hat{\beta} = HY = PY$.

$\hat{\beta}^* = (X'X)^{-1}X'\underline{Y}^*$, $E(\hat{\underline{\beta}}^*) = \hat{\underline{\beta}}$,

$Cov(\hat{\underline{\beta}}^*) = \sigma_n^2 (X'X)^{-1}$.

It can be shown that $\sqrt{n}(\hat{\underline{\beta}}^* - \hat{\underline{\beta}}) \xrightarrow{D} N_p(\underline{0}, \sigma^2 V)$.

25] Most of the theory for the above confidence regions does not work for variable selection, so new tools are needed. Mixture distributions will be useful, $\underline{x}^T\underline{\beta} = \underline{x}_S^T\underline{\beta}_S + \underline{x}_E^T\underline{\beta}_E = \underline{x}_S^T\underline{\beta}_S$

26] A random vector $\underline{U}$ has a mixture distribution of random vectors $\underline{U}_j$ with probabilities π_j if $\underline{U}$ equals $\underline{U}_j$ with probability π_j for $j = 1, \ldots, J$. Let $\underline{U}$ and $\underline{U}_j$ be $p \times 1$ random vectors. Then the cdf of $\underline{u}$ is

$$F_{\underline{u}}(\underline{t}) = \sum_{j=1}^{J} \pi_j F_{\underline{u}_j}(\underline{t}) \text{ where}$$

$0 \le \pi_j \le 1$, $\sum_{j=1}^{J} \pi_j = 1$, $J \ge 2$ and $F_{\underline{u}_j}(\underline{t})$ is the cdf of $\underline{u}_j$. Suppose $E[h(\underline{u})]$ and $E[h(\underline{u}_j)]$ exist. Then $E[h(\underline{u})] = \sum_{j=1}^{J} \pi_j E[h(\underline{u}_j)]$ and $E[\underline{u}] = \sum_{j=1}^{J} \pi_j E(\underline{u}_j)$. Hence $\text{cov}(\underline{u}) =$

$$E(\underline{u}\,\underline{u}^T) - E(\underline{u})[E(\underline{u})]^T = \sum_{j=1}^{J} \pi_j E(\underline{u}_j \underline{u}_j^T) - E(\underline{u}) E(\underline{u})^T$$

$$= \sum_{j=1}^{J} \pi_j \,\text{cov}(\underline{u}_j) + \sum_{j=1}^{J} \pi_j \, E(\underline{u}_j)[E(\underline{u}_j)]^T - E(\underline{u})[E(\underline{u})]^T.$$

If $E[\underline{u}_j] \equiv \underline{0}$ for $j = 1, \ldots, J$, then $E(\underline{u}) = \underline{0}$ and $\text{cov}(\underline{u}) = \sum_{j=1}^{J} \pi_j \,\text{cov}(\underline{u}_j)$.

27] For a mixture distribution, the selection can't change the dist of the $\underline{u}_j$.

eg use random selection.

28] Suppose $\hat{\underline{\beta}}_{VS} = \hat{\underline{\beta}}_{I_{k},0}$ with probs π_{kn}.

Let $\hat{\underline{\beta}}_{MIX} = \hat{\underline{\beta}}_{I_k,0}$ with probs π_{kn} where random selection is used.

24) Let $\underline{x}^T\underline{\beta} = \underline{x}_S^T\underline{\beta}_S + \underline{x}_E^T\underline{\beta}_E = \underline{x}_S^T\underline{\beta}_S$.

If $S \subseteq I$, then $\underline{x}^T\underline{\beta} = \underline{x}_S^T\underline{\beta}_S = \underline{x}_I^T\underline{\beta}_S$.

For $S \subseteq I_j$ assume

$$\sqrt{n}\left(\hat{\beta}_{I_j} - \beta_{I_j}\right) \xrightarrow{D} N_{a_j}(0, V_j).$$

Then $\sqrt{n}\left(\hat{\underline{\beta}}_{I_j,0} - \underline{\beta}\right) \xrightarrow{D} N_p(0, V_{j0})$

where V_{j0} adds columns and rows of 0's corresponding to the x_i not in I_j, and V_{j0} is singular unless I_j is the full model,

$$MSE\, n(X_{I_j}' X_{I_j})^{-1} \to V_j.$$

30) $P(S \subseteq I_{min}) \to 1$ as $n \to \infty$ for C_p, AIC, BIC, lasso, elastic net. $P(S \subseteq I_{min}) \to 1$ is a necessary condition for $\hat{\beta}_{VS}$ to be a consistent est of $\underline{\beta}$ if S is unique.

---

31) Th) Assume $P(S \subseteq I_{min}) \to 1$ as $n \to \infty$ and let $\hat{\beta}_{MIX} = \hat{\underline{\beta}}_{I_k,0}$ with probs π_{kn} where $\pi_{kn} \to \pi_k$ as $n \to \infty$. Denote positive π_k by π_j. Assume $\underline{u}_{jn} = \sqrt{n}\left(\hat{\underline{\beta}}_{I_j,0} - \underline{\beta}\right) \xrightarrow{D} \underline{u}_j \sim N_p(\underline{0}, V_{j0})$.

a) Then $\underline{U}_n = \sqrt{n}(\hat{\underline{\beta}}_{MIX} - \underline{\beta}) \xrightarrow{D} \underline{U}$ where the cdf of $\underline{U}$ is $F_{\underline{U}}(\underline{t}) = \sum_j \pi_j F_{\underline{U}_j}(\underline{t})$,
a mixture dist of the $\underline{U}_j$.

b) Let A be a full rank $g \times p$ matrix with $1 \le g \le p$. Then

$$\underline{v}_n = A \underline{U}_n = \sqrt{n}(A\hat{\underline{\beta}}_{MIX} - A\underline{\beta}) \xrightarrow{D} A\underline{U} = \underline{v}$$

where $\underline{v}$ has a mixture dist of the $\underline{v}_j = A\underline{U}_j \sim N_g(\underline{0}, A V_{j0} A^T)$ with prob's π_j.

c) $\hat{\underline{\beta}}_{VS}$ is a $\sqrt{n}$ consistent estimator of $\underline{\beta}$.

d) If $\pi_d = 1$, then $\sqrt{n}(\hat{\underline{\beta}}_{SEL} - \underline{\beta}) \xrightarrow{D} \underline{U} \sim N_p(0, V_{d0})$ where SEL is VS or MIX.

32) Notation: subscripts before MIX are used for subsets of $\hat{\underline{\beta}}_{MIX} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)^T$.
Let $\hat{\beta}_{i,MIX} = \hat{\beta}_i$. If $I = \{i_1, \ldots, i_a\}$, let $\hat{\underline{\beta}}_{I,MIX} = (\hat{\beta}_{i_1}, \ldots, \hat{\beta}_{i_a})^T$. Subscrips after MIX are the ith vector from an iid sample $\hat{\underline{\beta}}_{MIX,1}, \ldots, \hat{\underline{\beta}}_{MIX,B}$. The subscript 0 is

used for zero padding.

$\hat{\underline{\beta}} = \hat{\underline{\beta}}_{FULL}$ is used for the full model.

33) The mixture dist is asy normal if $\pi_d = 1$ or if for each π_j,

$$A\underline{v}_j \sim N_g(\underline{0}, A V_{j0} A^T) \sim N_g(\underline{0}, A \Sigma A^T).$$

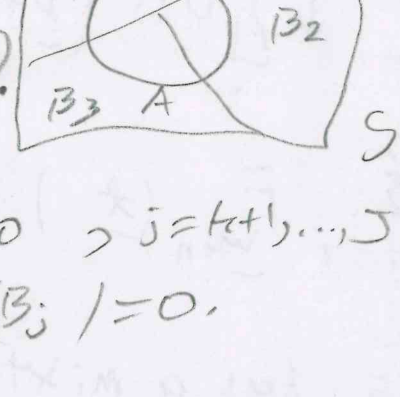
Then $\sqrt{n}\left(A\hat{\underline{\beta}}_{MIX} - A\underline{\beta}\right) \xrightarrow{D} A\underline{u} \sim N_g(\underline{0}, A\Sigma A^T)$.

This special case occurs for $\hat{\underline{\beta}}_{S,MIX}$ if $\sqrt{n}\left(\hat{\underline{\beta}} - \underline{\beta}\right) \xrightarrow{D} N_p(\underline{0}, V)$ where V is diagonal and nonsingular.

34) $B_1, \ldots, B_k$ partition sample space S if $B_i \cap B_j = \emptyset$ $i \neq j$, $P(B_i) > 0$ and $\bigcup_i B_i = S$.

law of total prob

Then $P(A) = \sum_{i=1}^{k} P(A \cap B_i) = \sum_{i=1}^{k} P(A|B_i)P(B_i)$.



Add sets $B_{k+1}, \ldots, B_J$ where $P(B_j) = 0$, $j = k+1, \ldots, J$.
by defining $P(A|B_j) = 0$ if $P(B_j) = 0$.

Then $P(A) = \sum_{i=1}^{J} P(A|B_i)P(B_i)$.

35) Let $\hat{\underline{\beta}}^{c}_{I_k,0} \sim \underbrace{\hat{\underline{\beta}}_{I_k,0} \mid (\hat{\underline{\beta}}_{VS} = \hat{\underline{\beta}}_{I_k,0})}_{\text{a conditional dist is a dist}}$

Denote $F_{\underline{z}}(\underline{t}) = P(z_1 \le t_1, \ldots, z_p \le t_p)$ by $P(\underline{z} \le \underline{t})$.

Let $\underline{w}_n = \sqrt{n}(\hat{\underline{\beta}}_{VS} - \underline{\beta})$ and $\underline{w}_{kn} = \sqrt{n}(\hat{\underline{\beta}}_{I_k,0} - \underline{\beta})$.

Then $F_{\underline{w}_n}(\underline{t}) = P[\sqrt{n}(\hat{\underline{\beta}}_{VS} - \underline{\beta}) \le \underline{t}] =$

$$\sum_{k=1}^{J} P[\underbrace{\sqrt{n}(\hat{\underline{\beta}}_{VS} - \underline{\beta}) \le \underline{t}}_{A} \mid \underbrace{\hat{\underline{\beta}}_{VS} = \hat{\underline{\beta}}_{I_k,0}}_{B_k}]\, P(\hat{\underline{\beta}}_{VS} = \hat{\underline{\beta}}_{I_k,0})$$

(the $I_k \ni P(\hat{\underline{\beta}}_{VS} = \hat{\underline{\beta}}_{I_k,0}) > 0$ form a partition)

$$= \sum_{k=1}^{J} P[\sqrt{n}(\hat{\underline{\beta}}_{I_k,0} - \underline{\beta}) \le \underline{t}) \mid \hat{\underline{\beta}}_{VS} = \hat{\underline{\beta}}_{I_k,0}]\, \pi_{kn}$$

$$= \sum_{k=1}^{J} P[\sqrt{n}(\hat{\underline{\beta}}^{c}_{I_k,0} - \underline{\beta}) \le \underline{t}]\, \pi_{kn}$$

$$= \sum_{k=1}^{J} F_{\underline{w}_{kn}}(\underline{t})\, \pi_{kn}.$$

So $\hat{\underline{\beta}}_{VS}$ has a mixture dist of the $\hat{\underline{\beta}}^{c}_{I_k,0}$ with probs π_{kn}

and " $\underline{w}_n$ " $\underline{w}_{kn}$ "!!