47) variable selection changes the dist of $\underline{U}_{jn}$ to $\underline{W}_{in}$, say.

48) $\hat{\underline{\beta}}_{VS} = \hat{\underline{\beta}}_{I_{k},0}$ with prob π_{kn}. Let

$\hat{\underline{\beta}}_{MIX} = \hat{\underline{\beta}}_{I_{k}0}$ with prob π_{kn} but

$\hat{\underline{\beta}}_{MIX}$ uses random selection instead of variable selection (can't compute since π_{kn} are unknown).

49) By large sample theory, if $S \subseteq I_j$, then $\sqrt{n}(\hat{\underline{\beta}}_{I_j} - \underline{\beta}_j) \xrightarrow{D} N_{a_j}(\underline{0}, V_j)$.

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

V_{j0} adds rows and columns of 0s corresponding to x_i not in I_j. Thus V_{j0} is singular unless I_j is the full model!

ex) $\underline{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)^T = (\beta_1, \beta_2, 0, 0)^T$ with

$\underline{\beta}_S = \begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}$, $\underline{\beta}_E = \begin{pmatrix}\beta_3\\ \beta_4\end{pmatrix} = \begin{pmatrix}0\\ 0\end{pmatrix}$.

Suppose $\sqrt{n}\,(\hat{\underline{\beta}}_S - \underline{\beta}_S) \xrightarrow{D} N_2\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}V_{11} & V_{12}\\ V_{21} & V_{22}\end{pmatrix}\right]$.

So $\sqrt{n}\,(\hat{\underline{\beta}}_{S0} - \underline{\beta}) = \sqrt{n}\left[\begin{pmatrix}\hat{\beta}_{1S}\\ \hat{\beta}_{2S}\\ 0\\ 0\end{pmatrix} - \begin{pmatrix}\beta_1\\ \beta_2\\ 0\\ 0\end{pmatrix}\right]$

$\xrightarrow{D} N_4\left[\begin{pmatrix}0\\0\\0\\0\end{pmatrix}, \begin{pmatrix}V_{11} & V_{12} & 0 & 0\\ V_{21} & V_{22} & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}\right]$.

50) $P(S \subseteq I_{min}) \to 1$ as $n \to \infty$ is a necessary condition for $\hat{\underline{\beta}}_{VS}$ to be a consistent estimator of $\underline{\beta}$ and the condition tends to hold if AIC is used

51) $\hat{\underline{\beta}}_{MIX}$ CLT: Assume $P(S \subseteq I_{min}) \to 1$ as $n \to \infty$. Let $\hat{\underline{\beta}}_{MIX} = \hat{\underline{\beta}}_{I_K 0}$ with probs π_{kn} where $\pi_{kn} \to \pi_k$ as $n \to \infty$. Denote the positive π_k by π_j. Assume $\underline{u}_{jn} = \sqrt{n}\,(\hat{\underline{\beta}}_{I_j 0} - \underline{\beta}) \xrightarrow{D} \underline{u}_j \sim N_p(\underline{0}, V_{j0})$.

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

$\underline{u}$ has a mixture dist of the $\underline{u}_j$ with probs π_j, $E(\underline{u}) = \underline{0}$ and

$\text{Cov}(\underline{u}) = \Sigma_u = \sum_j \pi_j V_{j0}$.

b) Let A be a full rank $g \times p$ constant 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

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

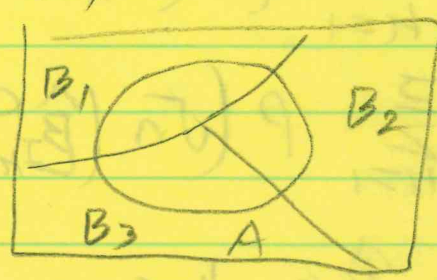
with probs π_j.

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

d) If $\pi_d = 1$, then $\sqrt{n}\,(\hat{\underline{\beta}}_{SEL} - \underline{\beta}) \xrightarrow{D}$

$\underline{u} \sim N_p(\underline{0}, V_{d0})$ where SEL is VS or MIX

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



Law of total Prob:

$P(A) = \sum_{i=1}^{K} P(A \cap B_i) =$

$\sum_{i=1}^{K} P(A \mid B_i)\, P(B_i)$.

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

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

53} Let $\hat{\underline{B}}^C_{I_k 0} \sim \underbrace{\underline{B}_{I_k 0} \,|\, (\hat{\underline{B}}_{VS} = \hat{\underline{B}}_{I_k 0})}_{\text{conditional dist is a dist}}$.

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

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

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

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

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

$$= \sum_{k=1}^{J} P(\sqrt{n}\,(\hat{\underline{B}}^C_{I_k 0} - \underline{B}) \le \underline{t}) \, \pi_{kn} = \sum_{j=1}^{J} F_{\underline{w}_{kn}}(\underline{t}) \pi_{kn}.$$

So $\hat{\underline{B}}_{VS}$ has a mixture dist of $\hat{\underline{B}}^C_{I_k 0}$ with probs π_{kn}

and " $\underline{w}_n$ $\qquad\qquad$ $\underline{w}_{kn}$."

54) $\sqrt{n}(\hat{\beta}_{MIX} - \beta)$ and $\sqrt{n}(\hat{\beta}_{VS} - \beta)$

are selecting from $\underline{u}_{kn} = \sqrt{n}(\hat{\beta}_{I_{k0}} - \beta)$ and asymptotically from $\underline{u}_k$ (where $\underline{u}_k \sim N_p(0, V_{k0})$ if $S \subseteq I_k$).

Random selection does not change the dist of $\underline{u}_{kn}$ and $\underline{u}_k$, but variable selection changes the dist of selected $\underline{u}_{kn}$ to $\underline{w}_{kn}$ and selection bias changes the dist of $\underline{u}_j$ to $\underline{w}_j$.

55) <u>Variable selection CLT</u>: Assume $P(S \subseteq I_{min}) \to 1$ as $n \to \infty$ and let $\hat{\beta}_{VS} = \hat{\beta}_{I_{k0}}$ with probs $\pi_{kn} \to \pi_k$ as $n \to \infty$. Denote the positive π_k by π_j. Assume $\underline{w}_{jn} = \sqrt{n}(\hat{\beta}^C_{I_{j0}} - \beta) \xrightarrow{D} \underline{w}_j$. Then

$\underline{w}_n = \sqrt{n}(\hat{\beta}_{VS} - \beta) \xrightarrow{D} \underline{w}$ where the cdf of $\underline{w}$ is $F_{\underline{w}}(\underline{t}) = \sum_j \pi_j F_{\underline{w}_j}(\underline{t})$.

Thus $\underline{w}$ is a mixture dist of the $\underline{w}_j$ with probs π_j.

56] $\hat{\underline{\beta}}_{S,MIX}$ seems to be a good approx for $\hat{\underline{\beta}}_{S,VS}$ unless the predictors are highly correlated. Most of the selection bias is due to predictors in E which make selection of S almost random. $\hat{\underline{\beta}}_{E,MIX}$ and $\hat{\underline{\beta}}_{E,VS}$ tend to differ, but both use 0 padding.

57] Let the bootstrap sample $T_1^*, \ldots, T_B^* =$

$$\underbrace{T_{1,1}^*, \ldots, T_{B_{1n},1}^*}_{\text{1st bootstrap component}}, \ldots, \underbrace{T_{1,J}^*, \ldots, T_{B_{Jn},J}^*}_{\text{Jth bootstrap component}}$$

1st bootstrap component $\hat{S}$ selects model I_1 Jth bootstrap component $\hat{S}$ selects model I_J

Denote $T_{1,j}^*, \ldots, T_{B_{jn},j}^*$ as the jth bootstrap component with sample mean $\overline{T}_j^*$ and sample covariance matrix $S_{T,j}^*$. Define the jth component of the iid sample $T_1, \ldots, T_B$ to have sample mean $\overline{T}_j$ and sample covariance matrix $S_{T,j}$.

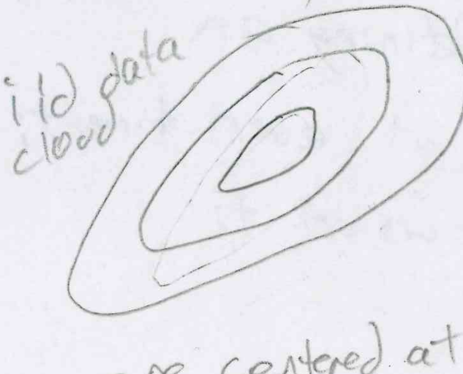
58] Under regularity conditions, if $S \subseteq I_j$, then $\sqrt{n}(\hat{\underline{\beta}}_{I_j,0} - \underline{\beta}) \xrightarrow{D} N_p(0, V_{j,0})$ and

$$\sqrt{n}(\hat{\underline{\beta}}_{I_j,0}^* - \hat{\underline{\beta}}_{I_j,0}) \xrightarrow{D} N_p(0, V_{j,0}).$$

Then the bootstrap and iid component

clouds have the same variability, asymptotically. The iid data component clouds are centered at $\underline{\beta}$. The Geometric argument holds for iid cloud with $\hat{\underline{\beta}}_{MIX}$. If the bootstrap component clouds were centered at $\underset{\sim}{\underline{\beta}}$, the bootstrap confidence regions would work.

Instead, the bootstrap component clouds are slightly shifted from a common center $\overline{T}^*$, and are each centered at $\hat{\underline{\beta}}_{I_j,0}$. So the T_i^* are "further" from $\overline{T}^*$ than the T_i are from $\overline{T}$. Geometrically, shifting the bootstrap component clouds from a common center makes the bootstrap data cloud more variable than the iid data cloud, asymptotically, resulting in slightly higher asymptotic coverage.



iid data component clouds have a common center $\overline{T} \approx \underline{\beta}$

Separating so the component clouds are centered at $\hat{\underline{\beta}}_{I_j,0}$ makes the overall bootstrap data cloud have greater variability than the iid data cloud. Hence expect coverage ≥ nominal by the Geometric argument.

More on checking x_i for the PH model:

1) A poisson regression model has

$Z|\underline{x} \sim \text{poisson}\left[\exp(\underline{\beta}'\underline{x})\right]$.

A poisson generalized additive model GAM

has $Z|\underline{x} \sim \text{poisson}\left(\exp\left[\alpha + \sum_{i=1}^{p} S_i(x_i)\right]\right)$

where the functions $S_i(x_i)$ are usually

unknown, but you can set $S_i(x_i) = \beta_i x_i$.

2) The cox PH model $Y|\underline{x}$ has

$h(\underline{x}) = e^{\underline{x}'\underline{\beta}} h_0(t)$.

3) The cox PH model can be fit

with a poisson regression by defining an

artificial variable Z_{ij} for Y_j at event times t_i

where $Z_{ij} = \begin{cases} 1 & \text{if } Y_j \text{ event occurs at } t_i \\ 0 & \text{else.} \end{cases}$

The GAM analog can also be fit to

get $\hat{\alpha}$ and $\hat{S}_i$.