Let $z_{11}^*, z_{12}^*, \ldots, z_{1n}^*$, $T_1^* = T(z_{11}^*, \ldots, z_{1n}^*)$

$z_{21}^*, z_{22}^*, \ldots, z_{2n}^*$, $T_2^* = T(z_{21}^*, \ldots, z_{2n}^*)$

$\vdots$

$z_{B1}^*, z_{B2}^*, \ldots, z_{Bn}^*$, $T_B^* = T(z_{B1}^*, \ldots, z_{Bn}^*)$,

Bth bootstrap dataset

ex] know for Quiz 11, final

data $z_1, \ldots, z_7 = 1, 2, 3, 4, 5, 6, 7$

$T = \text{median}(z_1, \ldots, z_7) = 4$

Let $B = 2$ (mean is common too)

(2, 2, 2, 3, 3, 5, 6) ← ordered

1st: 3, 2, 3, 2, 5, 2, 6 $\quad T_1^* = 3$

2nd: 3, 5, 3, 4, 3, 5, 7 $\quad T_2^* = 4$

(3, 3, 3, 4, 5, 5, 7 ← ordered)

$$\bar{T}^* = \frac{1}{B}\sum_{i=1}^{B} T_i^* = \frac{3+4}{2} = 3.5$$

is the bagging estimator.

See old quiz extra #2 on the web page for another ex.

26) Let $T_{1n}, \ldots, T_{Bn}$ be iid with the same dist as statistic $T_n = \hat{\underline{\theta}}$

often $T_n = \hat{\underline{\beta}}$ or $T_n = A\hat{\underline{\beta}}$ with $\underline{\theta} = \underline{\beta}$ or $A\underline{\beta}$.

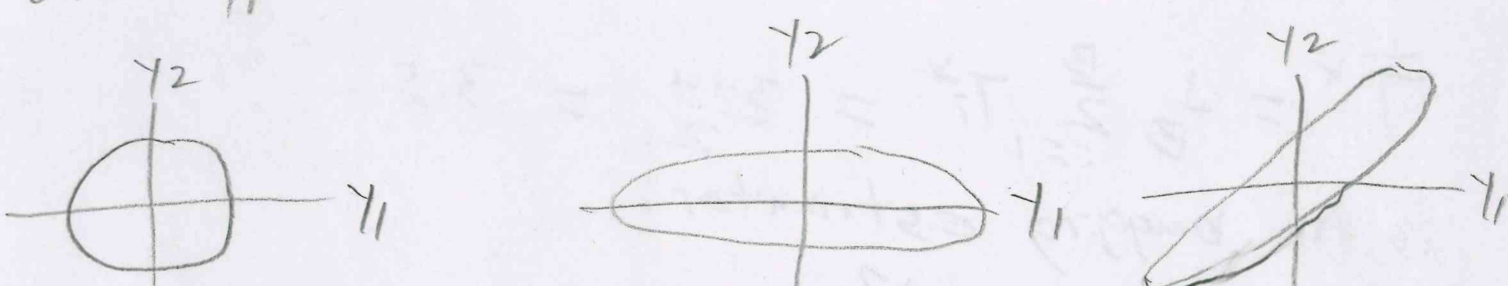
$\sqrt{n}(T_1^* - T_n), \ldots, \sqrt{n}(T_B^* - T_n)$ are often pseudodata for $\sqrt{n}(T_{1n} - \underline{\theta}), \ldots, \sqrt{n}(T_{Bn} - \underline{\theta})$.

27) $\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_p \end{pmatrix} \sim N_p(\underline{\mu}, \Sigma)$, a multivariate normal distribution with $E(\underline{Y}) = \underline{\mu} = \begin{pmatrix} E(Y_1) \\ \vdots \\ E(Y_p) \end{pmatrix}$

and covariance matrix $\underset{p \times p}{\Sigma} = (\sigma_{ij})$ where

$\sigma_{ij} = cov(Y_i, Y_j)$, Then $\sigma_{ii} = \sigma_i^2$. $Y_i \sim N(\mu_i, \sigma_i^2)$.

28) The highest density regions for $N_p(\underline{\mu}, \Sigma)$ are hyperellipoids



$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ $\sigma_1^2 = \sigma_2^2$ $\quad \Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$, $\sigma_1^2 > \sigma_2^2$ $\quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$ Y_1 and Y_2 correlated

29] often $\sqrt{n}(T_n - \underline{\theta}) \xrightarrow{D} N_p(\underline{0}, V)$.

ex] multivariate central limit theorem:

Let $\underline{x}_1, \ldots, \underline{x}_n$ be iid $p \times 1$ random vectors

with $E(\underline{x}_i) = \underline{\mu}$ and $\text{cov}(\underline{x}_i) = \Sigma$.

Then $\sqrt{n}(\overline{\underline{x}} - \underline{\mu}) \xrightarrow{D} N_p(\underline{0}, \Sigma)$.

The univariate CLT is a special case with $p=1$:

$\sqrt{n}(\overline{x} - \mu) \xrightarrow{D} N(0, \sigma^2)$.

ex] $\sqrt{n}(\hat{\underline{\beta}} - \underline{\beta}) \xrightarrow{D} N_p(\underline{0}, V)$ for

PH, AFT, SPH, WPH, GCR etc

where $\hat{\underline{\beta}}$, $\underline{\beta}$, and V depend on the method.

This CLT allows hypothesis testing and CIs.

30] Let A be a constant matrix and $\underline{d}$ a constant vector.
If $\underline{z}_n \xrightarrow{D} N_p(\underline{\mu}, \Sigma)$, then

$$\underset{g \times p}{A} \underset{p \times 1}{\underline{z}_n} \xrightarrow{D} N_g(A\underline{\mu}, A \Sigma A^T),$$

$$A\underline{z}_n + \underset{g \times 1}{\underline{d}} \xrightarrow{D} N_g(A\underline{\mu} + \underline{d}, A \Sigma A^T).$$

31} $p=1$: $T^*_{(1)}, \ldots, T^*_{(B)}$

a) $100(1-\delta)\%$ percentile CI for $E(T)=\theta$ discards the smallest and largest $100\frac{\delta}{2}\%$ of $T^*_{(i)}$

b) shorth$(\lceil B(1-\delta)\rceil)$ CI of $T^*_{(i)}$

$$\left[T^*_{(\lceil B\frac{\delta}{2}\rceil)}, T^*_{(\lceil B(1-\frac{\delta}{2})\rceil)}\right]$$ eg $1-\delta=0.9$, $\delta=0.1$, $\frac{\delta}{2}=0.05$

$\left[T^*_{(\lceil B(0.05)\rceil)}, T^*_{(\lceil B\,0.975\rceil)}\right]$ is a 95% CI for θ.

Reject $H_0\ \theta=\theta_0$ if θ_0 is outside the CI.

32} Sample indices $1, \ldots, n$ with replacement

ex $n=6$ $\underset{\sim}{i}_1 = 3, 2, 3, 2, 5, 6$ so use

$(T_3, \underline{x}_3, \delta_3), (T_2, \underline{x}_2, \delta_2), (T_3, \underline{x}_3, \delta_3), (T_2, \underline{x}_2, \delta_2), (T_5, \underline{x}_5, \delta_5), (T_6, \underline{x}_6, \delta_6)$

in the 1st bootstrap data set for survival regression.

This is called the <u>nonparametric bootstrap</u> (empirical, naive, rowwise, pairs). Works for PH with right censored response T_i.

33] Get bootstrap data, do variable selection to get $\hat{\underline{\beta}}^*_{I_{min},i}$ $i=1,\dots,B$.

$\hat{\underline{\beta}}_{VS} = \hat{\underline{\beta}}_{I_{min},0}$ and $\hat{\underline{\beta}}^*_{VS,i} = \hat{\underline{\beta}}_{I_{min},0,i}$

Q11

34] $\boxed{\text{know for final}}$ Given $\hat{\underline{\beta}}_{I_{min}}$, get $\hat{\underline{\beta}}_{VS}$.

ex] $p=5$ $\hat{\underline{\beta}}_{I_{min}} = \begin{pmatrix} 1.1 \\ 3.7 \\ 4.3 \end{pmatrix}$ uses x_1, x_3, and x_5

$$\hat{\beta}_{VS} = \begin{pmatrix} 1.1 \\ 0 \\ 3.7 \\ 0 \\ 4.3 \end{pmatrix}$$

35] Compute $\underbrace{\hat{\underline{\beta}}^*_1, \dots, \hat{\underline{\beta}}^*_B}_{\text{bootstrap sample}}$ with $\hat{\underline{\beta}}_i = \hat{\underline{\beta}}_{VS,i}$. $T = \hat{\underline{\beta}} = \hat{\underline{\beta}}_{VS}$

$$\overline{T}^* = \overline{\hat{\beta}^*} = \frac{1}{B} \sum_{i=1}^{B} \hat{\underline{\beta}}_i \text{ and } \widehat{Cov}(\hat{\underline{\beta}}_{VS}) = Cov(\hat{\underline{\beta}}^*) = S^*_T$$

$$= \frac{1}{B-1} \sum_{i=1}^{B} (\hat{\underline{\beta}}^*_i - \overline{\hat{\underline{\beta}}^*})(\hat{\underline{\beta}}_i - \overline{\hat{\underline{\beta}}^*})^T$$ are the

sample mean and sample covariance matrix of the bootstrap sample

36) A squared Mahalanobis distance

$$D^2_{\underline{w}}(T, \hat{\Sigma}) = (\underline{w} - T)^T \hat{\Sigma}^{-1} (\underline{w} - T)$$

$\{\underline{w} : D_{\underline{w}}(T, \hat{\Sigma}) \leq D^2_{(U_B)}\}$ forms a hyperellipsoid centered at T.

$$D^2_{\underline{w}}(T, \hat{\Sigma}) = D_T(\underline{w}, \hat{\Sigma}) = (T - \underline{w})^T \hat{\Sigma}^{-1} (T - \underline{w}).$$

37) Let $T_1, \ldots, T_B$ be iid. Then

$\{\underline{w} : D^2_{\underline{w}}(\overline{T}, \hat{\Sigma}) \leq D^2_{(U_B)}\}$ is a $100(1-\delta)\%$ prediction region for a future value T_f where $D^2_{(U_B)}$ is $\approx 100(1-\delta)$ percentile of $D^2_{T_i}(\overline{T}, \hat{\Sigma}) = (T_i - \overline{T})^T \hat{\Sigma}^{-1} (T_i - \overline{T})$, $i = 1, \ldots, B$.

38) A $100(1-\delta)\%$ large sample confidence region for $\underline{\theta}$ is a set $A_n \ni P(\underline{\theta} \in A_n)$ is eventually bounded below by $1-\delta$ as $n \to \infty$.

Often want $P(\underline{\theta} \in A_n) \to 1-\delta$.
A CI is a special case if θ is a scalar.

3a) Let $T_1^*, ..., T_B^*$ be the bootstrap sample
Let $\overline{T}^*$, S_T^* be the sample mean and cov matrix of the T_i^*.

a) Modified Bickel and Ren CR

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

$$\{\underline{w} : D^2_{\underline{w}}(T_n, S_T^*) \le D^2_{(U_B,T)}\}$$ where $D^2_{(U_B,T)}$ is the $100\, q_B$th sample percentile of $(T_i^* - T_n)^T (S_T^*)^{-1} (T_i^* - T_n)$, $i = 1, ..., B$

b) prediction region method CR

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

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

$D^2_{(U_B)}$ is the $100\, q_B$th sample percentile of

$$(T_i^* - \overline{T}^*)^T (S_T^*)^{-1} (T_i^* - \overline{T}^*)$$ and $q_B \downarrow 1-\delta$.

c) hybrid CR

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

CR b) to be centered at T_n or replaces

a) cutoff $D^2_{(U_B,T)}$ by $D^2_{(U_B)}$.

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

$\sqrt{n}(\overline{T}^* - T_n) \xrightarrow{P} \underline{0}$, $\sqrt{n}(\overline{T}^* - \underline{\theta}) \xrightarrow{D} \underline{U}$, and

$(n S_T^*)^{-1}$ is not too ill conditioned for large enough n and B. If $\underline{U} \sim N_g(\underline{0}, \Sigma)$, these conditions often hold. Then

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

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

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

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

So $D_j^2 \approx \underline{U}^T (n S_T^*)^{-1} \underline{U}$ for large n, B, and CRs in 39) have coverage near $1-\delta$.

ex) Let $T_n = \overline{\underline{x}} = \frac{1}{n} \sum_{i=1}^n \underline{x}_i$ where the $\underline{x}_i$ are iid with $E(\underline{x}_i) = \underline{\theta} = \underline{\mu}$ and $Cov(\underline{x}_i) = \Sigma$. Then $Cov(T_n) = Cov(\overline{\underline{x}}) = \frac{\Sigma}{n}$. So $n S_T^*$ estimates Σ and $(n S_T^*)^{-1} = \frac{(S_T^*)^{-1}}{n}$ estimates Σ^{-1}.

4] Geometric argument) Suppose there is 473 57

an iid sample $\underbrace{T_{1n}, \ldots, T_{Bn}}_{\text{problem: usually } B=1}$ of size B

of the statistic. A large sample $100(1-\delta)\%$

prediction region for T_{Fn} is

$\{w: D^2_w(\overline{T}, \hat{\Sigma}_T) \le D^2_{(U_B)}\}$.

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

$\overline{T} \to \theta$ faster than $T_i \to \theta$.

Now $D^2_{T_i}(\overline{T}, \hat{\Sigma}_T) = D_{\overline{T}}(T_i, \hat{\Sigma}_T)$

center of hyperellipsoid.

So $\overline{T} \in \{w: D^2_w(T_i, \hat{\Sigma}_T) \le D^2_{(U_B)}\}$ iff

$T_i \in \{w: D^2_w(\overline{T}, \hat{\Sigma}_T) \le D^2_{(U_B)}\}$ which

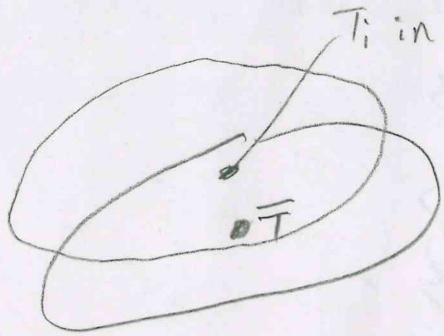
occurs $\approx 100(1-\delta)\%$ eg 95%. If $\overline{T} \to \theta$

fast enough, then $P[\theta \in \underbrace{\{w: D^2_w(T_i, \hat{\Sigma}_T) \le D^2_{(U_B)}\}}_{\text{CR for } \theta}]$

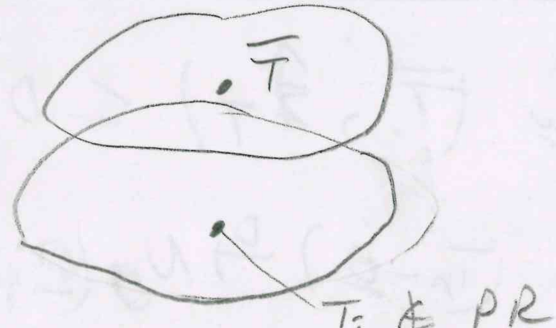
$\to 1-\delta$.

suppose $\frac{95}{100}$ T_i are in the prediction region,

contains $\simeq 95\%$ of T_i if n and B are large



occurs $\simeq 95\%$ of the time



occurs $\simeq 5\%$ of the time

42) If $\sqrt{n}(\underline{T}_n - \underline{\theta}) \xrightarrow{D} \underline{u}$ and $\sqrt{n}(\underline{T}_i^* - \underline{T}_n) \xrightarrow{D} \underline{u}$, then the bootstrap data cloud is like the iid data cloud of $T_1, \ldots, T_B$ shifted to be centered at T_n. Hence by the geometric argument, 39) gives CRs. Need $B \geq 50p$ if $\underline{\theta}$ is $p \times 1$.

43) know for final and Quiz 11 473 58

	Estimate	std err	95% shorth CI
x1	-42.48	51.2863	[-192.281, 52.492]
x2	0		[0.000, 0.268]
x3	1.17	0.0598	[0.992, 1.289]
x4	0		[0.000, 0.840]
x5	0		[0.000, 1.916]
x6	0.15	0.0368	[0.0747, 0.215]

if model selected in advance

Output for $\hat{\beta}_{Imin0}$

a) Give the CI for β_3

b) Give $\hat{\underline{\beta}}_{VS} = \hat{\underline{\beta}}_{Imin,0}$

See old quiz extra #3.

Soln a) $\boxed{[0.992, 1.289]}$

b) $\hat{\underline{\beta}}_{Imin0} = (-42.48, 0, 1.17, 0, 0, 0.15)^T$

44] variable selection theory is hard.

$$\underline{x}^T\underline{\beta} = \underline{x}_S^T\underline{\beta}_S + \underline{x}_E^T\underline{\beta}_E = \underline{x}_S^T\underline{\beta}_S$$

45] A random vector $\underline{u}$ has a <u>mixture distribution of random vectors $\underline{u}_j$ with probabilities π_j</u> if $\underline{u} = \underline{u}_j$ with probability π_j for $j = 1, \ldots, J$ (and if the selection mechanism does not change the distribution of the $\underline{U}_j$, eg use random selection). Let $\underline{U}$ and $\underline{U}_j$ be random vectors. Then the cdf of $\underline{U}$ is $F_{\underline{U}}(\underline{t}) = \sum_{j=1}^{J} \pi_j F_{\underline{U}_j}(\underline{t})$ 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 $E[\bar{h}(\underline{U})]$ and $E[\bar{h}(\underline{U}_j)]$ exist. Then

$$E[\bar{h}(\underline{U})] = \sum_{j=1}^{J} \pi_j \, E[\bar{h}(\underline{U}_j)]$$ and

$$E[\underline{U}] = \sum_{j=1}^{J} \pi_j \, E(\underline{U}_j).$$

If $E[\underline{U}_j] = \underline{\theta}$ for $j = 1, \ldots, J$, then $E(\underline{U}) = \underline{\theta}$ and

$$\operatorname{cov}(\underline{U}) = \sum_{j=1}^{J} \pi_j \operatorname{cov}(\underline{U}_j).$$

46) Random selection works: generate a Uniform $(0,1)$ RV $W \perp\!\!\!\perp \underline{U}_1, \ldots, \underline{U}_J$ and set $\underline{U} = \underline{U}_j$ if $W \in \left(\sum_{k=0}^{j-1} \pi_k, \sum_{k=0}^{j} \pi_k \right)$ with $\pi_0 = 0$