$[L_n(\underline{Y}), U_n(\underline{Y})]$ is a large sample $100(1-\delta)\%$ CI for θ if $P(L_n \le \theta \le U_n) = P(L_n(\underline{Y}) \le \theta \le U_n(\underline{Y}))$ is eventually bounded below by $1-\delta$ as $n \to \infty$.

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

CIs are a special case.

3] A statistic is a function of the data that does not depend on any unknown parameters $\underline{\theta}$.

4] P231 The quantity $R(\underline{Y}|\underline{\theta})$ is a pivot or pivotal quantity if the distribution of $R(\underline{Y}|\underline{\theta})$ does not depend on θ. $R(\underline{Y}, \theta)$ is an asymptotic pivot or asymptotic pivotal quantity if the limiting dist of $R(\underline{Y}, \underline{\theta})$ does not depend on $\underline{\theta}$.

ex) a) $\sqrt{n}(T_n - \underline{\theta}) \xrightarrow{D} N_p(\underline{0}, \Sigma)$

does not depend on $\underline{\theta}$ if Σ doesn't depend on $\underline{\theta}$ (nuire)

$\sqrt{n}\,\hat{\Sigma}^{-\frac{1}{2}}(T_n - \underline{\theta}) \xrightarrow{D} N_p(\underline{0}, I)$

useful asymptotic pivot

then $n(T_n - \underline{\theta})^T \hat{\Sigma}^{-1}(T_n - \underline{\theta}) \xrightarrow{D} \chi^2_p$

b) $Y_1, \ldots, Y_n$ iid $E(Y_i) = \mu$, $V(Y_i) = \sigma^2$

$\sqrt{n}\,\frac{(\bar{Y} - \mu)}{\sigma} \xrightarrow{D} N(0,1)$

$\sqrt{n}(\bar{Y} - \mu) \xrightarrow{D} N(0, \sigma^2)$

} need σ known to be a useful pivot

$\sqrt{n}\left(\frac{\bar{Y} - \mu}{\sigma}\right) \sim N(0,1)$, Y_i iid $N(0, \sigma^2)$

$\sqrt{n}\left(\frac{\bar{Y} - \mu}{S}\right) \sim t_{n-1}$ Y_i iid $N(\mu, \sigma^2)$

$\xrightarrow{D} N(0,1)$ Y_i iid as above

useful asymptotic pivot

C) Linear Models

$Y_i = \underline{x}_i^T \underline{\beta} + e_i$ $\quad e_i$ iid

$E(e_i) = 0$, $V(e_i) = \sigma^2$

$\underline{x}_i = (1, x_{2i}, \ldots, x_{pi})^T$ so a constant is in the model, $i = 1, \ldots, n$.

H_0 $A\underline{\beta} = \underline{0}$ $\quad H_A$ $A\underline{\beta} \neq \underline{0}$

$r \times p$ $\quad A$ full rank r

$$F_0 = \frac{1}{r\,MSE} (A\hat{\underline{\beta}})^T (A (\underline{X}^T \underline{X}) A^T)^{-1} (A\hat{\underline{\beta}})$$

$\underline{X}$ $n \times p$ with ith row $\underline{x}_i^T$.

If $e_i \sim N(0, \sigma^2)$, $\quad F_0 \sim F_{r, n-p}$ Ho true

and $r F_0 \xrightarrow{D} \chi^2_r$ under mild conditions if e_i iid.

$r F_{r,n-p} \xrightarrow{D} \chi^2_r$, so the F tests are large sample tests.

5) Consider a statistic T_n where $\sqrt{n}(T_n - \underline{\theta}) \xrightarrow{D} N_p(\underline{0}, \Sigma_T)$.

A bootstrap data set,

$Y_1^*, \ldots, Y_n^*$ or $(Y_1^*, \underline{x}_1^*), \ldots, (Y_n^* \underline{x}_n^*)$

etc, is used to compute a bootstrap statistic T_n^*.

Repeat B times to get a bootstrap sample $T_1^*, \ldots, T_B^*$.

6) A case consists of the measurements taken on a person, object or thing to get a data set.

eg $Y_1, \ldots, Y_n$ height Y_i = ith case

eg $(Y_1, \underline{x}_1^T)^T, \ldots, (Y_n, x_n^T)^T$ Y_i = height x_{i1} = weight x_{i2} = gender of ith person

$(Y_i, x_i^T)^T$ is the ith case.

7) Let the data set have n cases.

The nonparametric bootstrap creates a bootstrap data set by drawing a sample of size n with replacement from the n cases. Then T^* is computed from the bootstrap data set.

If $\underline{x}_1, \ldots, \underline{x}_n$ is the data set, then the empirical distribution is a discrete distribution where the $\underline{x}_i$ are equally likely. If $\underline{W}$ is a random vector from the empirical dist, then $P(\underline{W} = \underline{x}_i) = \frac{1}{n}$, $i = 1, \ldots, n$.

The cdf of the empirical dist is F_n.

The empirical dist is the bootstrap dist for the bootstrap data set,

ex] Let $\underline{W} \sim F_n$ be a random vector having the empirical dist.

a) Show $E(\underline{W}) = \underline{\bar{x}} = \underline{\bar{x}}_n$.

b) Show $\text{Cov}(\underline{W}) = \frac{n-1}{n} S = \frac{n-1}{n} S_n$

↑ Sample cov matrix.

proof a) $E(\underline{W}) = \sum \underline{x}_i p_i = \sum \underline{x}_i \frac{1}{n} = \underline{\bar{x}}_n$

b) $\text{Cov}(\underline{W}) = E\left[(\underline{W} - E(\underline{W}))(\underline{W} - E(\underline{W}))^T\right]$

$= \sum (\underline{W}_i - E\underline{W})(\underline{W}_i - E(\underline{W}))^T p_i$

$= \sum (\underline{x}_i - \underline{\bar{x}})(x_i - \bar{x})^T \frac{1}{n} = \frac{n-1}{n} S_n$

□

ex) data 1, 2, 3, 4, 5, 6, 7
(4 is the sample median)

with $n=7$ and sample median $T_n = 4$.

Rootpot gives $B=2$ bootstrap data sets drawn with replacement from the data.

3 2 3 2 5 2 6 → 1st bootstrap data set

(2 2 2 3 3 5 6 $T_1^* = 3$)

3 5 3 4 3 5 7 → 2nd bootstrap dataset

(3 3 3 4 5 5 7 $T_2^* = 4$).

8) Often $\sqrt{n}(T_n - \underline{\theta}) \xrightarrow{D} \underline{U}$ and large sample theory can be used to show $\sqrt{n}(T_n^* - T_n) \xrightarrow{D} \underline{u}$.

An iid sample $T_{1n}, \ldots, T_{Bn}$ of the statistic would be useful, but we only have $T_n = T_{1n}$.

$\sqrt{n}(T_{1n}^* - T_n), \ldots, \sqrt{n}(T_{Bn}^* - T_n)$,

where $T_i^* = T_{in}^*$, is pseudodata

for $\sqrt{n}(T_{1n} - \underline{\theta}), \ldots, \sqrt{n}(T_{Bn} - \underline{\theta})$

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

9) For the bootstrap sample

$T_1^*, \ldots, T_B^*$, let

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

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

$\overline{T}^*$ is known as the bagging estimator or smoothed bootstrap estimator.

Let T_i be a random variable.

10) The bootstrap percentile method large sample $100(1-\delta)\%$ CI for θ is an interval $\left[T^*_{(k_L)}, T^*_{(k_U)} \right]$ containing $\approx \lceil B(1-\delta) \rceil$ of the T_i^*. Let

$k_1 = \lceil B\frac{\delta}{2} \rceil$ and $k_2 = \lceil B(1-\frac{\delta}{2}) \rceil$.

The usual percentile method CI is

$$\left[T^*_{(k_1)}, T^*_{(k_2)} \right].$$

11) The large sample $100(1-\delta)\%$ Shorth(c) CI for θ is

$$\left[T^*_{(s)}, T^*_{(s+c-1)} \right] \text{ where}$$

$$c = \min\left(B, \lceil B\left[1-\delta+1.12\sqrt{\frac{\delta}{B}}\right] \rceil\right).$$

12) Let T^* be a $g \times 1$ random vector. The large sample $100(1-\delta)\%$ standard bootstrap confidence region is

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

$$= \left\{ \underline{w} : D^2_{\underline{w}} (T_n, S_T^*) \leq \chi^2_{p(1-\delta)} \right\}.$$

13) The prediction region method

large sample $100(1-\delta)\%$ conf reg for $\underline{\theta}$ is

$$\left\{ \underline{w} : (\underline{w} - \bar{T}^*)^T \left[S_T^* \right]^{-1} (\underline{w} - \bar{T}^*) \leq D^2_{(U_B)} \right\}$$

$$= \left\{ \underline{w} : D^2_{\underline{w}} (\bar{T}^*, S_T^*) \leq D^2_{(U_B)} \right\} \text{ where}$$

$D^2_{(U_B)}$ is computed from

$$D_i^2 = (T_i^* - \bar{T}^*)^T (S_T^*)^{-1} (T_i^* - \bar{T}^*) \text{ for}$$

$i = 1, \ldots, n$.

14) Let $q_B = \min\left(1-\delta+0.05,\ 1-\delta+\frac{9}{B}\right)$, $\delta > 0.1$

$q_B = \min\left(1-\frac{\delta}{2},\ 1-\delta+10\delta\frac{9}{B}\right)$, $\delta \leq 0.01$.

If $1-\delta < 0.999$ and $q_B < 1-\delta+0.001$, set $q_B = 1-\delta$. Let $D^2_{(U_B)}$ be the $100\,q_B$th