10 Mar 01

Parametric bootstrap.

Simplest example

$Y_1, \ldots, Y_n \quad : \quad N(\mu, \sigma^2)$

$\hat{\theta} = \bar{Y}$

$Y_1^*, \ldots, Y_n^*$ from $N(\bar{Y}, s^2)$ (1)

$\hat{\theta}_1^*, \ldots, \hat{\theta}_B^*$

$\hat{\theta}_b^* = \bar{Y} + \frac{s}{\sqrt{n}} z_b \quad b = 1, \ldots, B$ representation

$$\frac{1}{B-1} \sum_b (\hat{\theta}_b^* - \bar{\hat{\theta}}_b^*)^2 = \frac{s^2}{n} \frac{1}{B-1} \sum_1^B (z_b - \bar{z})^2$$

cdly $\quad \Big| \underset{\sim}{Y} \xrightarrow{P} \frac{s^2}{n}$ as $B \to \infty$

For estimating $\operatorname{var} \bar{Y}$.

<u>For estimating distribution</u> of $\bar{Y}$. 10 Nov 01

Conditional distribution, given the data, of

(*) $\sqrt{n}(\underbrace{\hat{\theta}^*}_{\text{from (1)}} - \hat{\theta}) = s Z \xrightarrow{d} \sigma Z$ as $n \to \infty$

(**) $\sqrt{n}(\hat{\theta} - \theta) = \sigma Z'$

Can use distⁿ of (*) to approximate that of (**)

Want s near σ.

<u>Note</u> on $s \xrightarrow{?} \sigma$, given the data

<u>SLLN</u>. $\bar{Y} \to \mu$ prob 1

10 Mar 01

Example(s) where bootstrap doesn't work.

mle of $U(0, \theta)$

Going outside $\sqrt{n}$ range

Which to use - depends on skills

(1) 18 Nov. 01

Bootstrap. Efron & Tibshirani, Chapt 6.

Sample $(x_1, \ldots, x_n)$

Parameter $\theta = t(F)$ (functional)

Estimate $\hat{\theta} = s(x)$, eg. plug in estimator $t(\hat{F})$

$$\boxed{\text{How to estimate } se_F(\hat{\theta}) \quad ?}$$

Suppose $se_F(\hat{\theta}) \left(\overset{eg.}{=} \frac{\sigma_F}{n} = \frac{1}{n}\sqrt{\int x^2 dF(x) - \left(\int x dF(x)\right)^2} = h(F)\right)$

Could use $h(\hat{F}) \left(= \frac{1}{n} \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} \right)$. But $h(.)$ may be ~~messy~~ messy or unknown.

$\star$ (Ideal) bootstrap estimate

$$se_{\hat{F}}(\hat{\theta}(\underset{\sim}{x}^*)) = se_{\hat{F}}(\theta^*)$$

"the s.e. of $\hat{\theta}$ for data sets of size n randomly sampled from $\hat{F}$"

i.e. s.e. of $\hat{\theta}$ for data sets of size n randomly sampled from $\hat{F}$.

There are $m = \binom{2n-1}{n}$ distinct bootstrap samples (if data are distinct) and sampling with replacement. Eg.: 1 2 3

111, 222, 333

112, 113, 221, 223, 331, 332

123 $m = 10$

Basic idea: using $se_{\hat{F}}(\hat{\theta})$ to estimate $se_F(\hat{\theta})$

Denote these by $z^1, z^2, \ldots, z^m$

Let $\text{Prob}\{z = z^j\} = w_j \qquad j = 1, \ldots, m$

Then

$$se_{\hat{F}}(\hat{\theta}^*) = \left[\sum_{j=1}^{m} w_j \{ s(z^j) - s(\cdot) \}^2 \right]^{1/2} \qquad (*)$$

where

$$s(\cdot) = \sum_{j=1}^{m} w_j \, s(z^j)$$

a mess. (m is very large)

Since x is available, (*) can be estimated by simulation

$$\hat{se}_B^2 = \sum_{b=1}^{B} [\hat{\theta}^*(b) - \bar{\hat{\theta}}^*]^2 / B-1$$

$$\lim_{B \to \infty} \hat{se}_B^2 = se_{\hat{F}}(\hat{\theta}^*)^2 = se_{\hat{F}}$$

$$E_F(\hat{se}_B^2) = E_F(se_{\hat{F}}(\theta^*)^2)$$

$$E_{\hat{F}}(\hat{se}_B^2) = se_{\hat{F}}(\hat{\theta})^2 = \hat{se}_\infty^2 \qquad \text{p. 5?}$$

Write $se_{\hat{F}}(\hat{\theta}^*)$ rather than $se_{\hat{F}}(\hat{\theta})$ to avoid confusion between $\hat{\theta}$, the value of $s(x)$ based on the observed data and $\hat{\theta}^* = s(x^*)$ thought of as a random variable based on the bootstrap sample.

18 Nov. 05

Real World | **Bootstrap World**

$$\begin{array}{ccc}
\text{Unknown cdf} & & \text{Observed sample} \\
F & \longrightarrow & \underset{\sim}{x} = (x_1, \ldots, x_n) \\
& & \downarrow \\
& & \hat{\theta} = s(\underset{\sim}{x}) \\
& & \text{Statistic of interest}
\end{array}
\quad \Longrightarrow \quad
\begin{array}{ccc}
\text{Empirical cdf} & & \text{Bootstrap sample} \\
\hat{F} & \longrightarrow & \underset{\sim}{x}^* = (x_1^*, \ldots, x_n^*) \\
& & \downarrow \\
& & \hat{\theta}^* = s(\underset{\sim}{x}^*) \\
& & \text{Bootstrap replication}
\end{array}$$

In more complicated case replace F by P, the unknown probability model.

$$\begin{array}{ccc}
& \text{Population } F & \\
& \downarrow & \\
& \text{Data } \underset{\sim}{x} = (x_1, \ldots, x_n) & \\
\swarrow & \downarrow & \searrow \\
\underset{\sim}{x}^{*1} & \underset{\sim}{x}^{*2} & \underset{\sim}{x}^{*B} \\
\downarrow & \downarrow & \downarrow \\
s(\underset{\sim}{x}^{*1}) & s(\underset{\sim}{x}^{*2}) & s(\underset{\sim}{x}^{*B})
\end{array}$$

} sampling variability

} resampling variability

bootstrap samples

18 Nov. 04

Doesn't work for $\hat{\theta}$ mle of $U(0,\theta)$.
$\hat{F}$ is not a good estimate of F in the extreme tail.

How large to choose B?

Problems of bias.

p. 129 "The bootstrap t-intervals have good theoretical coverage properties, but tend to be erratic in practice."

p. 394 "... bootstrap ideas have been least successful in hypothesis testing problems, ..."

Programs in Efron & Tibshirani.

Problem 19.1 $s(\underset{\sim}{x}) = t(\hat{F})$ based on i.i.d. Suppose $\hat{\gamma}_B$ is a bootstrap ~~estimate~~ of one feature of the distribution of $s(\underset{\sim}{x})$. Let $\gamma(\hat{F}) = \lim_{B\to\infty} \hat{\gamma}_B$. Then

$$E\{\hat{\gamma}_B\} \approx \gamma(F)$$

Proof. $E(\cdot) = E_{\underset{\sim}{x}}(E(\cdot \mid \underset{\sim}{x}))$

19 Nov. 01

Bootstrapping and Regression.

Freedman (1981) Ann Statist. 1218–

Regression model.

$$\underset{n\times1}{Y} = \underset{n\times p}{X}\,\underset{p\times1}{\beta} + \underset{n\times1}{\epsilon}$$

$\epsilon_1, \epsilon_2, \ldots$ from F

$X^T X$ non singular, first col X is 1

$$\hat\beta = (X^T X)^{-1} X^T Y$$

$$E\hat\beta = \beta, \quad \operatorname{Var}\hat\beta = \sigma^2 (X^T X)^{-1}$$

If $X^T X/n \to \Sigma$

$$\boxed{\sqrt{n}(\hat\beta - \beta) \xrightarrow{d} N_p(0, \sigma^2 \Sigma^{-1})} \qquad (1)$$

$$(X^T X)^{-1/2}(\hat\beta - \beta)/\hat\sigma \xrightarrow{d} N_p(0, I)$$

$$\hat\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} \hat\epsilon_i^2$$

19. Nov. 01.

Bootstrapping.

$$\hat{\epsilon} = Y - X\hat{\beta}$$

$\frac{1}{n}\sum \hat{\epsilon}_i = 0$ since first col of X is 1

Let $\hat{F}_n$ be empirical distribution function of $\hat{\epsilon}$

$$\int x \, d\hat{F}_n = 0$$

Given Y, let $\epsilon_1^*, \ldots, \epsilon_n^*$ be conditionally independent with common distribution $\hat{F}_n$.
(Resampling the residuals.)
Let

$$Y^* = X\hat{\beta} + \epsilon^*$$

Can compute new ~~estimate~~ from these values

$$\hat{\beta}^* = (X^T X)^{-1} X^T Y^*$$

Bootstrap principle. The distribution of $\sqrt{n}(\hat{\beta}^* - \hat{\beta})$ approximates that of $\sqrt{n}(\hat{\beta} - \beta)^{(1)}$ and the former may be computed directly from the data.

Approx good if n large, $\sigma^2 p \,\mathrm{trace}(X^T X)^{-1}$ small

19 Nov. 01

Let m denote the resample size.

Consider adding one row at a time to X $\ni$

$X^T X / n \longrightarrow \Sigma$, $\epsilon_1, \epsilon_2, \ldots$ iid F

The original X is the first n of the growing sequence of rows. Need to be able to keep generating Y's

Now

$$\underset{m\times 1}{Y^*} = \underset{m\times p}{X}\ \underset{p\times 1}{\hat{\beta}} + \underset{m\times 1}{\epsilon^*}$$

$\epsilon_1^*, \ldots, \epsilon_m^*$ iid from $\hat{F}_n$

$$\underset{p\times 1}{\hat{\beta}^*} = \underset{p\times p}{(X^T X)^{-1}} X^T Y^*$$

$$\hat{\sigma}^{*2} = \frac{1}{m} \sum_{i=1}^{m} \hat{\epsilon}_i^{*2}$$

Depend on m and n

19 Nov 01

Theorem. For almost all sample sequences, given $Y_1, \ldots, Y_n$

a) $\sqrt{m}\,(\hat{\beta}^* - \hat{\beta}) \xrightarrow{d} N(0, \sigma^2 \Sigma^{-1})$ i.e. (1)

b) $\hat{\sigma}^* \xrightarrow{d} \sigma$

$m, n \to \infty$

Bootstrapping and linear regression

21 Nov. 01

$$Y_j = \underset{\sim}{x}_j' \underset{\sim}{\beta} + \epsilon_j \qquad j = 1, \ldots, n$$

Residuals $\hat{\epsilon}_j = y_j - \underset{\sim}{x}_j' \hat{\beta}$, $\boxed{r_j = \hat{\epsilon}_j / \sqrt{1 - h_j^2}}$

Bootstrap I (model-based)

h_j : leverage

For $b = 1, \ldots, B$

1. For $j = 1, \ldots, n$
   a) set $\underset{\sim}{x}_j^* = \underset{\sim}{x}_j$
   b) Randomly sample ϵ_j^* from $r_j - \bar{r}$ sample( ) with replacement
   c) set $y_j^* = \underset{\sim}{x}_j^{*\prime} \underset{\sim}{\hat{\beta}} + \epsilon_j^*$
2. OLS of $\begin{bmatrix} \underset{\sim}{x}_j^* \\ y_j^* \end{bmatrix}$ to get $\hat{\beta}_b^*$, $\hat{\sigma}_b^*$

Justification

$$\underset{\sim}{\hat{\beta}}^* = (\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}' \underset{\sim}{y}^*$$

non-singular case

$$= (\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}' (\underset{\sim}{X} \underset{\sim}{\hat{\beta}} + \underset{\sim}{\epsilon}^*)$$

$$= \underset{\sim}{\hat{\beta}} + (\underset{\sim}{X}'\underset{\sim}{X})^{-1} \underset{\sim}{X}' \underset{\sim}{\epsilon}^*$$

21 Mar 01

Write $E^*\{\ \}$ for $E\{\ |Y_1, \ldots, Y_n\}$

$$E^*\{\epsilon_j^*\} = \frac{1}{n}\sum_j (r_j - \bar{r}) = 0$$

$$\mathrm{Var}^*\{\epsilon_j^*\} = \frac{1}{n}\sum_j (\hat{r}_j - \bar{r})^2$$

So $E^*\{\hat{\underset{\sim}{\beta}}^*\} = \hat{\beta}$,

$$\mathrm{Var}^*\{\hat{\underset{\sim}{\beta}}^*\} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'[\mathrm{Var}^*\{\epsilon_j^*\}]\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$$

$$= (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\sum_j (r_j - \bar{r})^2/n$$

and

$$\frac{1}{n}\sum (r_j - \bar{r})^2 \simeq \frac{1}{n-p}\sum_j \hat{\epsilon}_j^2 = \hat{\sigma}^2$$

Estimate $\mathrm{Var}^*\{\hat{\underset{\sim}{\beta}}^*\}$ by $\frac{1}{B-1}\sum_b (\hat{\underset{\sim}{\beta}}_b^* - \bar{\hat{\underset{\sim}{\beta}}}^*)'(\hat{\underset{\sim}{\beta}}_b^* - \bar{\hat{\underset{\sim}{\beta}}}^*)$ Want B large.

The use of r_j, rather than $\hat{\epsilon}_j$, appears important.

Now can: estimate quantiles for example.
estimate prediction uncertainty

21 Mar. 01

Bootstrap II (case-based)

For $b = 1, \dots, B$

1. sample $i_1^*, \dots, i_n^*$ for $\{1, 2, \dots, n\}$ with replacement
2. For $j = 1, \dots, n$
   set $\underset{\sim}{x}_j^* = \underset{\sim}{x}_{i_j^*}$, $y_j^* = y_{i_j^*}$
3. OLS of $\begin{bmatrix} x_j^* \\ y_j^* \end{bmatrix}$ to get $\hat{\beta}_b^*$, $\hat{\sigma}_b^*$

Disadvantage - inefficient

Advantage - robust to heteroscedasticity

Bootstrapping and WLS

21 Mar 01

$$\hat{\underset{\sim}{\beta}} = (\underset{\sim}{X}'\underset{\sim}{W}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{W}\underset{\sim}{y}$$

Method I.

$$r_j = (y_j - \underset{\sim}{x}_j'\hat{\underset{\sim}{\beta}})\sqrt{w_j} \,/\, \sqrt{1-h_j}$$

ϵ_j^* sampled from $r_j - \bar{r}$

$$y_j^* = \underset{\sim}{x}_j^{*\prime}\hat{\underset{\sim}{\beta}} + \epsilon_j^* / \sqrt{w_j}$$

Method II.

$$w_j^* = w_{i_j^*}, \quad x_j^* = x_{i_j^*}, \quad y_j^* = y_{i_j^*}$$

Bootstrapping and M-estimates

21 Mar 01

Estimating equations

$$\sum_{j=1}^{n} \underset{\sim}{x}_j \psi\left(\frac{y_j - \underset{\sim}{x}_j' \hat{\underset{\sim}{\beta}}}{s}\right) = \underset{\sim}{0}$$

for some s
e.g. MAD

$$\operatorname{var}\{\hat{\underset{\sim}{\beta}}\} \sim \sigma^2 \frac{E\{\psi^2(\varepsilon/\sigma)\}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}}{[E\{\dot{\psi}(\varepsilon/\sigma)\}]^2}$$

$$\hat{\varepsilon}_j = y_j - \underset{\sim}{x}_j' \hat{\beta}$$

$$r_j = \hat{\varepsilon}_j \big/ \sqrt{1 - d h_j}$$

$$d = \frac{2 \sum (\hat{\varepsilon}_j/s)\psi(\hat{\varepsilon}_j/s)}{\sum \dot{\psi}(\hat{\varepsilon}_j/s)} - \frac{\sum \psi^2(\hat{\varepsilon}_j/s)}{\{\sum \dot{\psi}(\hat{\varepsilon}_j/s)\}^2}$$

Randomly sample from $\{r_1, \ldots, r_n\}$ bias correction later

get s^*

E.Q. $$\sum_{j=1}^{n} \underset{\sim}{x}_j \left\{\psi\left(\frac{y_j^* - \underset{\sim}{x}_j' \hat{\beta}^*}{s^*}\right) - \frac{1}{n}\sum_{k=1}^{n} \psi\left(\frac{r_k}{s^*}\right)\right\} = 0$$

21 Nov 01

Bootstrapping nonlinear regression models

$$y_j = \mu(x_j, \beta) + \epsilon_j \qquad j = 1, \ldots, n$$

Resample modified* residuals

or

resample cases

* mean adjustment + correction for bias due to nonlinearity

21 Nov 01

Bootstrapping glm's and gam's.

Davison and Hinkley (1997), Chapt. 7

Basics $\quad EY = \mu$

$$\text{Var}\, Y = \varphi V(\mu)$$

φ: overdispersion

$$\eta = \underset{\sim}{x}'\underset{\sim}{\beta}, \quad g(\mu) = \eta \quad \text{(link)}$$

E.Q. $$\sum_{j=1}^{n} \frac{y_j - \mu_j}{V(\mu_j)} \times \frac{\underset{\sim}{x}_j}{\dot{g}(\mu_j)} = \underset{\sim}{0}$$

Algorithm.

Regress $z_j = \eta_j + (y_j - \mu)\dot{g}(\mu_j)$ on $\underset{\sim}{x}_j$ with weights w_j

$$w_j^{-1} = V(\mu_j)\dot{g}(\mu_j)^2$$

Hat matrix $\quad X(X'WX)^{-1}X'W \qquad \exists$ symmetric variant

$$\hat{\varphi} = \frac{1}{n-p} \sum_{j=1}^{n} \frac{(y_j - \hat{\mu}_j)^2}{V(\hat{\mu}_j)}$$

21 Nov 01

Residuals

1. Standardized Pearson

$$r_{Pj} = \frac{y_j - \hat{\mu}_j}{\sqrt{\hat{\varphi} V(\hat{\mu}_j)(1-h_j)}}$$

2. Standardized on predictor scale

$$r_{Lj} = \frac{g(y_j) - g(\hat{\mu}_j)}{\sqrt{\hat{\varphi}\, \dot{g}(\hat{\mu}_j)^2 V(\hat{\mu}_j)(1-h_j)}}$$

3. Deviance

$$d_j = \operatorname{sgn}(y_j - \hat{\mu}_j)\sqrt{2\{ l_j(y_j) - l_j(\hat{\mu}_j)\}} \qquad \text{includes } \hat{\varphi}$$

Standardized deviance

$$r_{Dj} = d_j / \sqrt{1-h_j}$$

21 Nov 01

Bootstrapping by cases.

For $b = 1, \ldots, B$

1. sample $i_1^*, \ldots, i_n^*$ from $\{1, 2, \ldots, n\}$
2. set $\underset{\sim}{x}_j^* = \underset{\sim}{x}_{i_j^*}$, $y_j^* = y_{i_j^*}$
3. glm with $\begin{bmatrix} \underset{\sim}{x}_j^* \\ y_j \end{bmatrix}$

But be clear what a case is. (Aggregation occurs)

Via residuals.

1. $\varepsilon_1^*, \ldots, \varepsilon_n^*$ a sample from $\{r_{P_j} - \bar{r}_P\}$

$$y_j^* = \hat{\mu}_j + \sqrt{\hat{\varphi} V(\hat{\mu}_j)}\, \varepsilon_j^* \qquad (\text{Might omit } \hat{\varphi})$$

But cannot expect y_j^* to be: count, proportion, $> 0, \ldots$

2. $$y_j = g^{-1}\left[\underset{\sim}{x}^{*\prime} \underset{\sim}{\hat{\beta}} + \dot{g}(\hat{\mu}_j) \sqrt{\hat{\varphi} V(\hat{\mu}_j)}\, \varepsilon_j^*\right]$$

$\varepsilon_1^*, \ldots, \varepsilon_n^*$ from $\{r_{L_j}\}$

21 Nov 01

3. Deviance based

$$d_j = d(y_j, \hat{\mu}_j)$$

$$\epsilon_j = d(y_j, \mu_j)$$

$\epsilon_1^*, \ldots, \epsilon_r^*$ from $\{r_{Dj}\}$

y_j^* solves $d(y_j^*, \hat{\mu}_j) = \epsilon_j^*$