STAT 210B HWK #6 SOLUTIONS (DUE MAY 8)

DAVID ROSENBERG, JING LEI

(1) Consider the artificial data set consisting of the 8 numbers

$$1.2, 3.5, 4.7, 7.3, 8.6, 12.4, 13.8, 18.1$$

Let $\hat{\theta}$ be the 25% trimmed mean.

(a) Calculate bootstrap estimates of the standard error of $\hat{\theta}$, using bootstrap samples of size $B = 25, 100, 200, 500, 1000, 2000$. Use these to form a numerical estimate of the ideal bootstrap estimate (that would be obtained by taking $B \to \infty$).

(b) Repeat part (a) using ten different random number seeds and hence assess the variability in the estimates. How large should we take B to provide satisfactory accuracy?

(c) Calculate the ideal bootstrap estimate of standard error directly by summing over all possible bootstrap samples. Compare the answer to that obtained in part (a).

(2) While the bootstrap is based on resampling from the data with replacement, the jackknife is based on samples that leave out one observation at a time. There are n such samples. Let us denote the estimates based on these samples by $\hat{\theta}_{(i)}$.

Define a *linear statistic* as one that can be written in the form:

$$\hat{\theta}(x_1, \ldots, x_n) = \mu + \frac{1}{n} \sum_{i=1}^{n} \alpha(x_i)$$

where μ is a constant and α is a function.

Date: May 8, 2008 .

Show that for linear statistics the jackknife and the bootstrap are equivalent in the following sense: Given the values $\hat{\theta}_{(i)}$ for $i = 1, \ldots, n$, the value of $\hat{\theta}^*$ can be deduced for an arbitrary bootstrap sample $x_1^*, \ldots, x_n^*$.

We want to find:

$$\hat{\theta}^*(x_1^*, \ldots, x_n^*) \quad = \quad \mu + \frac{1}{n}\sum_{i=1}^{n}\alpha(x_i^*)$$

in terms of the leave-one-out estimates $\hat{\theta}_{(i)}$. It's sufficient to write $\alpha(x_i)$ in terms of these estimators.

We have

$$\hat{\theta}_{(j)}(x_1, \ldots, x_n) \quad = \quad \mu + \frac{1}{n-1}\left[-\alpha(x_j) + \sum_{i=1}^{n}\alpha(x_i)\right]$$

and

$$\begin{aligned} \sum_{j=1}^{n}\hat{\theta}_{(j)} \quad &= \quad n\mu + \frac{1}{n-1}\sum_{j=1}^{n}\left[-\alpha(x_j) + \sum_{i=1}^{n}\alpha(x_i)\right] \\ &= \quad n\mu - \frac{1}{n-1}\sum_{j=1}^{n}\alpha(x_j) + \frac{n}{n-1}\sum_{i=1}^{n}\alpha(x_i) \\ &= \quad n\mu + \sum_{i=1}^{n}\alpha(x_i) \end{aligned}$$

So

$$\sum_{j=1}^{n}\hat{\theta}_{(j)} - (n-1)\hat{\theta}_{(k)} \quad = \quad \mu + \alpha(x_j)$$

Rearranging:

$$\alpha(x_j) = -\mu + \sum_{J=1}^{n}\hat{\theta}_{(j)} - (n-1)\hat{\theta}_{(k)}$$

(3) The jackknife estimate of standard error is

$$\left[\frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}\right)^2\right]^{1/2}$$

where $\hat{\theta}_{(\cdot)} = \sum_{i=1}^n \hat{\theta}_{(i)}/n$. Show that for linear statistics the bootstrap estimate of standard error is

$$\left[\sum_{i=1}^n \left(\alpha(x_i) - \bar{\alpha}\right)^2 / n^2\right]^{1/2}$$

where $\bar{\alpha} = \sum \alpha(x_i)/n$, and the jackknife estimate of standard error is

$$\left[\sum_{i=1}^n \left(\alpha(x_i) - \bar{\alpha}\right)^2 / \{(n-1)n\}\right]^{1/2}$$

SE of Bootstrap

Let $x_1, \ldots, x_n$ be our observed iid sample from P, and let P_n be the corresponding empirical distribution. Write $X^*, X_1^*, \ldots, X_n^*$ for iid samples from P_n. Then the bootstrap estimate of variance is

$$\begin{aligned}
\mathrm{Var}_{P_n}\hat{\theta}(X_1^*, \ldots, X_n^*) &= \mathrm{Var}_{P_n}\left(\mu + \frac{1}{n}\sum_{i=1}^n \alpha(X_i^*)\right) \\
&= \frac{1}{n}\mathrm{Var}_{P_n}\alpha(X^*) \\
&= \frac{1}{n}\left(E_{P_n}\left[\alpha(X^*)\right]^2 - \left[E_{P_n}\alpha(X^*)\right]^2\right) \\
&= \frac{1}{n}\left(\frac{1}{n}\sum_{j=1}^n \alpha(X_j)^2 - \left[\frac{1}{n}\sum_{j=1}^n \alpha(X_j)\right]^2\right) \\
&= \frac{1}{n}\left(\frac{1}{n}\sum_{j=1}^n \left(\alpha(X_j) - \bar{\alpha}\right)^2\right)
\end{aligned}$$

So the BOOTSTRAP SE is just

$$\hat{SE}_{\text{bootstrap}} = \sqrt{\sum_{i=1}^n \left(\alpha(X_i) - \bar{\alpha}\right)^2 / n^2}$$

JACKKNIFE STANDARD ERROR:

From Problem 3, we had

$$\hat{\theta}_{(j)}(x_1, \ldots, x_n) = \mu + \frac{1}{n-1}\left[-\alpha(x_j) + \sum_{i=1}^n \alpha(x_i)\right]$$

and

$$\hat{\theta}_{(\cdot)}(x_1, \ldots, x_n) := \frac{1}{n}\sum_{j=1}^{n} \hat{\theta}_{(j)} = \mu + \frac{1}{n}\sum_{i=1}^{n} \alpha(x_i)$$

So

$$\begin{aligned}
\hat{\theta}_{(j)} - \hat{\theta}_{(\cdot)} &= -\frac{1}{n-1}\alpha(x_j) + \frac{1}{n-1}\sum_{i=1}^{n}\alpha(x_i) - \frac{1}{n}\sum_{i=1}^{n}\alpha(x_i) \\
&= -\frac{1}{n-1}\alpha(x_j) + \frac{1}{n(n-1)}\sum_{i=1}^{n}\alpha(x_i) \\
&= -\frac{1}{n-1}\alpha(x_j) + \frac{1}{n-1}\bar{\alpha} \\
&= \frac{1}{n-1}\left(\bar{\alpha} - \alpha(x_j)\right)
\end{aligned}$$

and the jackknife estimate of standard error is

$$\begin{aligned}
\sqrt{\frac{n-1}{n}\sum_{k=1}^{n}\left(\hat{\theta}_{(k)} - \hat{\theta}_{(\cdot)}\right)^2} &= \sqrt{\frac{n-1}{n}\sum_{k=1}^{n}\left[\frac{1}{n-1}\left(\bar{\alpha} - \alpha(x_j)\right)\right]^2} \\
&= \sqrt{\frac{1}{n(n-1)}\sum_{k=1}^{n}\left(\alpha(x_j) - \bar{\alpha}\right)^2}
\end{aligned}$$