**Due:** Wednesday, June 4, 1997

**Reading:**
Lecture Notes, Chapter 8, sections 4-5 and 7.
Efron and Tibshirani, *Bootstrap*, Chapters 8-9 and 21.

1. Problem 9.6, Efron & Tibshirani, page 122. [You may assume that the design matrix $C$ contains a column of 1's -- so that the model includes an intercept term.]

2. (Bootstrapping a linear regression model a simple way.)
Consider bootstrapping a linear regression model

$$Y_i \equiv \mathbf{x}_i^T \beta + \epsilon, \quad i = 1, \cdots, n$$

where the $\epsilon_i$ are i.i.d. mean 0, finite variance, and the $\mathbf{x}_i$ are given $p-$dimensional vectors, such that there is no constant term in the regression.
A. Show that the estimated residuals $\hat{\epsilon}^T = (\hat{\epsilon}_1, \cdots, \hat{\epsilon}_n)$ satisfy $\hat{\epsilon} - \epsilon = -H\epsilon$ where $H = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is the "hat matrix" (i.e. the projection matrix onto the column space of $\mathbf{X}$).
B. Suppose that $\hat{\epsilon}_1^*, \cdots, \hat{\epsilon}_n^*$ is a bootstrap sample (with replacement) from $\{\hat{\epsilon}_1, \cdots, \hat{\epsilon}_n\}$. Show that

$$E_*(n^{1/2}(\hat{\beta}^* - \hat{\beta})) = (\frac{1}{n}\mathbf{X}^T\mathbf{X})^{-1}(\frac{1}{n}\mathbf{X}^T\mathbf{1})Z_n$$

where $Z_n \equiv n^{-1/2}\sum_{i=1}^n \hat{\epsilon}_i$.
C. Show that if $\max_{1 \le i \le n} h_{ii} \to 0$ and $n^{-1}\mathbf{X}^T\mathbf{X} \to V$, a positive definite matrix, then

$$\sqrt{n}(\hat{\beta} - \beta) \to_d N_p(0, \sigma^2 V^{-1})$$

[This is a variant of the result we established in 581 via the Lindeberg - Feller CLT.]
D. Find the mean and variance of $Z_n$.
E. Suppose that:
(i) $n^{-1}\mathbf{X}^T\mathbf{X} \to V$, a positive definite matrix,
(ii) $\mathbf{X}^T\mathbf{1}/n \to \mathbf{h}$ with $\mathbf{h}^T V^{-1}\mathbf{h} < 1$,
(iii) $\max_{1 \le i \le n} |h_{ii}| \to 0$ where $h_{ii}$, $i = 1, \cdots, n$ are the diagonal elements of the hat matrix $H$.
Show that if (i) - (iii) hold, then the bootstrap fails in the sense that the random variable $Z_n$ in B converges in distribution to a proper random

variable rather than to zero.
[Hint: show that (iii) implies that $\max_{1 \leq i \leq n} |c_{ni}| \to 0$ where $\underline{c}_n \equiv n^{-1/2}(I - H)\underline{1}$.]

3. Suppose now that the bootstrap residuals are drawn from the collection of *centered* residuals $\underline{\hat{\epsilon}} - \underline{1}(\underline{1}^T \underline{\hat{\epsilon}}/n)$.

   A. Compute $E_*(\sqrt{n}(\hat{\beta}^* - \hat{\beta}))$ and $E_*(\sqrt{n}(\hat{\beta}^* - \hat{\beta}))^{\otimes 2}$ for this bootstrap resampling scheme.

4. **Optional bonus problem:** Recast the results in sections 21.6 and 21.8 of Efron and Tibshirani, pages 307 - 309, and 310-311, in terms of the notation we used in Chapters 3 and 4 of Statistics 581. [For example replace the parameter of interest $\theta$ by $\nu$, replace $\underline{\eta}$ by $\underline{\theta}$, replace $h$ by $q$, ... .]