Statistics 583, Problem Set 8

Wellner; 5/24/2000

Reading: Lecture Notes, Chapter 8, Sections 8.1-8.4
Efron and Tibshirani, Chapters 5-7 (pages 39-82), Chapter 11, pages (141 - 150).

Due: Wednesday, May 31, 2000.

1. Problem 20.4, page 295, Efron and Tibshirani: Show that for linear statistics the jackknife and bootstrap estimates of bias are zero. (Note that Efron and Tibshirani's definition (20.11) of a linear statistic corresponds to a functional $T(F)$ of the form $T(F) = \int \psi dF$ for some function ψ for a distribution function F on R, or $T(P) = \int \psi dP$ for a probability distribution P on a general sample space $\mathcal{X}$.)

2. Problem 19.4, page 281, Efron and Tibshirani:
(a) Given n distinct data items, show that the probability that a given data item does not appear in a bootstrap sample is $e_n = (1 - 1/n)^n$
(b) Show that $e_n \to e^{-1} \approx .368$ as $n \to \infty$.
(c) Hence show that the probability that each of B bootstrap samples contains an item i is $(1 - e_n)^B$. Evaluate this quantity for $n = 10, 20, 50, 100$ and $B = 10, 20, 50, 100$.
(d) Let $N_n \equiv \sum_{j=1}^n 1_{[M_j=0]}$ where $\underline{M} \equiv (M_1, \ldots, M_n) \sim$ $\text{Mult}_n(n, \underline{1}/n)$. Show that $E(n^{-1}N_n) = e_n$ as computed in (a).

3. Suppose that $\mathcal{P}_0 = \{P_\theta : \theta \in \Theta \subset R^k\}$ is a regular parametric model, and let $\dot{\mathbf{l}}_\theta = \dot{\mathbf{l}}_\theta(\cdot;\theta)$ be the vector of score functions. Usually, to find the maximum likelihood estimator $\widehat{\theta}_n$ of $\theta \in \Theta$, we assume that the true distribution $P \in \mathcal{P}_0$ and we solve

$$0 = \frac{1}{n}\sum_{i=1}^n \dot{\mathbf{l}}_\theta(X_i;\theta) = \mathbb{P}_n \dot{\mathbf{l}}_\theta(x;\theta) \tag{1}$$

for θ to find the MLE. Now suppose that the true $P \notin \mathcal{P}_0$, but we still define an estimator $\widehat{\theta}_n$ by solving (1). Then under reasonable regularity conditions we can prove that $\widehat{\theta}_n$ converges in probability to $\theta(P)$ defined as the solution to

$$0 = P(\dot{\mathbf{l}}_\theta(X;\theta(P))) = \int \dot{\mathbf{l}}_\theta(x;\theta(P))dP(x)\,. \tag{2}$$

(a) Making any regularity assumptions you want, relate $\theta(P)$ as defined in (2) to the Kullback-Leibler distance $K(P, P_\theta)$ for $\theta \in \Theta$.
(b) Find the influence function (it will be a $k-$vector of influence functions of the components) of $\theta(P)$: i.e. for $P_t \equiv (1-t)P + tQ$ and $Q = \delta_x$, compute

$$\frac{d}{dt}\theta(P_t)\big|_{t=0}\,.$$

(c) Use the result of (b) to "guess" the asymptotic variance of $\theta(\mathbb{P}_n) \equiv \widehat{\theta}_n$ for $P \notin \mathcal{P}_0$.
(d) How would you estimate the asymptotic variance you found in (c)?

4. Suppose that $T(F) = Var_F(X)$ so that $T_n \equiv T(\mathbb{F}_n) = n^{-1}\sum_{i=1}^n (X_i - \overline{X})^2$. Show that the jackknife estimate of the variance $\sigma_n^2(F) \equiv Var_F(T_n)$ is

$$\widehat{Var} = \frac{n^2}{(n-1)^3}(\widehat{\mu}_4 - \widehat{\mu}_2^2)$$

where $\widehat{\mu}_k \equiv n^{-1}\sum_{i=1}^n (X_i - \overline{X})^k$ for $k = 1, 2, \ldots$. Hence, assuming that $EX^4 < \infty$, the jackknife estimate of variance is consistent for this T:

$$n\widehat{Var} \rightarrow_p \mu_4 - \mu_2^2 = \mu_2^2\{2 + \frac{\mu^4}{\mu_2^2} - 3\} = T_2(F)(2 + \gamma_2)\,.$$