

Statistics 583, Problem Set 4

Wellner; 4/29/2015

Reading: Chapter 8, sections 8.1- 8.4, pages 1 - 18;
Wasserman, Chapters 2-3, pages 13-41;
Van der Vaart, Chapter 23, pages 323-340.

Due: Wednesday, May 13, 2015

Reminder 1: Midterm exam, Wednesday, May 6.

Reminder 2: Third (and last) make-up lecture, Wednesday, May 13, 11:30 AM, ARC G070.

- (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 $e_n \rightarrow e^{-1} \approx .368$ as $n \rightarrow \infty$.
(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 $\underline{M} \equiv (M_1, \dots, M_n) \sim \text{Mult}_n(n, \underline{1}/n)$. Show that $E(n^{-1}N_n) = e_n$ as computed in (a).
- Suppose that $T(F) = \text{Var}_F(X)$ so that $T_n \equiv T(\mathbb{F}_n) = n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Show that the jackknife estimate of the variance $\sigma_n^2(F) \equiv \text{Var}_F(T_n)$ is

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

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

$$n\widehat{\text{Var}} \rightarrow_p \mu_4 - \mu_2^2 = \mu_2^2 \left\{ 2 + \frac{\mu_4}{\mu_2^2} - 3 \right\} = T_2(F)(2 + \gamma_2).$$

- (a) Wasserman, problem 3.8.9, page 40: Let X_1, \dots, X_n be n distinct observations (no ties). Let X_1^*, \dots, X_n^* denote a bootstrap sample (from the empirical d.f. \mathbb{F}_n of the X_i 's), and let $\bar{X}_n^* = n^{-1} \sum_{i=1}^n X_i^*$. Find: $E\{\bar{X}_n^* | X_1, \dots, X_n\}$, $\text{Var}(\bar{X}_n^* | X_1, \dots, X_n)$, and $\text{Var}(\bar{X}_n^*)$.
(b) Wasserman, problem 3.8.13, page 41.

4. Wasserman, example 3.10, page 29.

(a) In the 3rd line of this example, Wasserman says that the jackknife estimate of the standard error of the skewness estimator for the nerve data is .17. Check this. (Show your code or method of calculation.)

(b) I claimed in the solution to problem set #3, problem 3 (see page 9 of the solution set), that the estimated standard error using the delta method is .163 rather than .18 as Wasserman claimed. Check this. (Show your code or method of calculation.)

(c) On page 31, Wasserman claims that the bootstrap based on $B = 10^3$ replications gives .16 as a bootstrap estimate of the standard error. Try it yourself to see what you get. (Show your code or method of calculation.)

The data is posted in two forms at:

<http://www.stat.washington.edu/jaw/COURSES/580s/583/sp15.probsets.html/nrvdat>
and

<http://www.stat.washington.edu/jaw/COURSES/580s/583/sp15.probsets.html/nerve.dat>

(d) Plot the empirical distribution function of this data together with the MLE of the distribution function assuming: (i) an exponential distribution; (ii) a Weibull distribution. What alternative methods for (a) and (b) does this suggest?

5. **Optional bonus problem:** Show that for linear statistics the jackknife and bootstrap estimates of bias are zero. (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} .)

6. **Optional bonus problem:** Consider a V -functional of order $r = 2$ given by $T(P) = \iint h(x, y) dP(x) dP(y)$ where h is permutation symmetric. Find the jackknife estimate of bias for the (V -statistic) estimator $T(\mathbb{P}_n)$ of $T(P)$.