Statistics 583, Problem Set 8 Solutions

Wellner; 6/2/2000

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 $\mathcal{X}$.)

   Solution: First note that for a linear statistic, assuming that $E_F|\psi(X)| = \int |\psi(x)| dF(x) < \infty$, we have

$$E_F T(\mathbb{F}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n}\psi(X_i)\right) = E\psi(X_1) = \int \psi dF = T(F)\,,$$

   so, with $T_n \equiv T(\mathbb{F}_n)$,

$$\text{bias}(F) \equiv E_F(T_n) - T(F) = 0\,.$$

   The bootstrap estimator of bias(F) is

$$\begin{aligned}
\text{bias}(\mathbb{F}_n) &= E_{\mathbb{F}_n}(T_n^*) - T(\mathbb{F}_n) \\
&= E_{\mathbb{F}_n}\left(\frac{1}{n}\sum_{i=1}^{n}\psi(X_i^*)\right) - T(\mathbb{F}_n) \\
&= \frac{1}{n}\sum_{i=1}^{n} E_{\mathbb{F}_n}\{\psi(X_i^*)\} - T(\mathbb{F}_n) \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\sum_{j=1}^{n}\psi(X_j) - T(\mathbb{F}_n) \\
&= T(\mathbb{F}_n) - T(\mathbb{F}_n) = 0\,,
\end{aligned}$$

   so the bootstrap estimator of bias is correct. For the jackknife, we first note that the pseudo-values, $T_{n,i}^*$ are given by

$$T_{n,i}^* \equiv nT_n - (n-1)T_{n,i} = \psi(X_i)$$

   and hence

$$\overline{T}_n^* \equiv n^{-1}\sum_{i=1}^{n} T_{n,i}^* = n^{-1}\sum_{i=1}^{n}\psi(X_i) = T_n \equiv T(\mathbb{F}_n)\,.$$

   Thus the jackknife estimate of bias is

$$\widehat{\text{bias}}_n \equiv T_n - \overline{T}_n^* = 0\,.$$

   Thus the jackknife estimate of bias is also correct.

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 $\underline{M} \equiv (M_1, \ldots, M_n) \sim$ Mult$_n(n, \underline{1}/n)$. Show that $E(n^{-1}N_n) = e_n$ as computed in (a).

Solution: (a) The probability that X_i does not appear in a bootstrap sample $X_1^*, \ldots, X_n^*$ from $\mathbb{F}_n$ is just $e_n = P(M_i = 0)$ where $M_i \sim$ Binomial$(n, 1/n)$. Thus we have $e_n = P(M_i = 0) = \binom{n}{0}(1/n)^0(1-1/n)^n = (1-1/n)^n$.
(b) Since $(1 + x/n)^n \to e^x$ for any x, it follows immediately that $e_n \to e^{-1} \approx .368$.
(c) The probability that each of B bootstrap samples contains X_i is clearly $(1-e_n)^B$. The following table gives values of this for $n = 10, 20, 50, 100$ and $B = 10, 20, 50, 100$.

B/n	10	20	50	100
10	.0137	.0118	.0108	.0105
20	.000189	.000139	.000117	.000110
50	4.89×10^{-10}	2.29×10^{-10}	1.47×10^{-10}	1.27×10^{-10}
100	2.39×10^{-19}	5.26×10^{-20}	2.16×10^{-20}	1.61×10^{-20}

(d) N_n/n is the proportion of the original sample not appearing in the bootstrap sample. Since $N_n \equiv \sum_{j=1}^n 1_{[M_j=0]}$ where each M_j is marginally Binomial$(n, 1/n)$, it follows immediately that

$$E(N_n/n) = P(M_1 = 0) = (1-1/n)^n \to e^{-1}.$$

Furthermore, from occupancy theory for urn models,

$$\sqrt{n}(n^{-1}N_n - (1-1/n)^n) \to_d N(0, e^{-1}(1-2e^{-1}));$$

see e.g. Johnson and Kotz (1977), page 317.

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

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

(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)?

Solution: (a) The Kullback-Leibler distance is

$$\begin{aligned} K(P, P_\theta) &= \int \log \frac{p(x)}{p(x;\theta)}\, dP(x) \\ &= \int \log p(x) dP(x) - \int \log p(x;\theta) dP(x)\,. \end{aligned}$$

To minimize this as a function of θ we differentiate with respect to θ_j for $j = 1, \ldots, k$. Under the assumption that the interchange of integration and differentiation is permissible,

$$\frac{\partial}{\partial \theta_j} K(P, P_\theta) = -\int \dot{\mathbf{l}}_{\theta_j}(x;\theta) dP(x)\,, \tag{3}$$

so that

$$\nabla_\theta K(P, P_\theta) = (\frac{\partial}{\partial\theta_1}, \ldots, \frac{\partial}{\partial \theta_k})^T K(P, P_\theta) = -\int \dot{\mathbf{l}}_\theta(x;\theta) dP(x)$$

while the Hessian (matrix of second derivatives) is, assuming a second interchange of integration and differentiation is permissible,

$$\left(\frac{\partial^2}{\partial\theta_i \partial\theta_j} K(P, P_\theta)\right) = \left(-\int \ddot{\mathbf{l}}_{ij}(x;\theta) dP(x)\right)\,.$$

If we assume that the latter matrix $B = B(\theta)$ is positive definite at $\theta = \theta(P)$ solving $0 = \int \dot{\mathbf{l}}_\theta(x;\theta) dP(x)$, then $\theta(P)$ minimizes $K(P, P_\theta)$ over $\theta \in \Theta$.
(b) For $P_t = (1-t)P + tQ$,

$$\begin{aligned} 0 &= \int \dot{\mathbf{l}}_\theta(x; \theta(P_t)) dP_t(x) \\ &= \int \dot{\mathbf{l}}_\theta(x;\theta(P_t)) dP(x) + t\int \dot{\mathbf{l}}_\theta(x;\theta(P_t)) d(Q-P)(x)\,. \end{aligned}$$

Differentiating across this identity with respect to t, and setting $t = 0$ yields

$$0 = \int \ddot{\mathbf{l}}_{\theta\theta}(x;\theta(P)) \left\{\frac{d}{dt}\theta(P_t)|_{t=0}\right\} dP(x) + \int \dot{\mathbf{l}}_\theta(x;\theta(P)) d(Q-P)(x) + 0\,, \tag{4}$$

or

$$\int \ddot{\mathbf{l}}_{\theta\theta}(x;\theta(P)) dP(x) \left\{\frac{d}{dt}\theta(P_t)|_{t=0}\right\} = -\int \dot{\mathbf{l}}_\theta(x;\theta(P)) d(Q-P)(x)\,,$$

or, assuming that

$$B \equiv -\int \ddot{\mathbf{l}}_{\theta\theta}(x;\theta(P))dP(x)$$

is positive definite,

$$\left\{\frac{d}{dt}\theta(P_t)|_{t=0}\right\} = B^{-1}\int \dot{\mathbf{l}}_\theta(x;\theta(P))d(Q-P)(x)\,.$$

To find the influence function we take $Q=\delta_{x_0}$ to obtain

$$IC(x_0;\theta,P) = B^{-1}\dot{\mathbf{l}}_\theta(x_0;\theta(P))\,.$$

(c) Hence a reasonable guess at the asymptotic variance-covariance matrix of $\theta(\mathbb{P}_n)$ is:

(5) $$E_P[IC(X;\theta,P)^{\otimes 2}] = B^{-1}E_P[\dot{\mathbf{l}}_\theta(X;\theta(P))^{\otimes 2}]B^{-1} \equiv \Sigma\,.$$

Note that this is a "sandwich" form for the asymptotic variance-covariance matrix, and when $P \notin \mathcal{P}$ we will in general have $B \neq E_P[\dot{\mathbf{l}}_\theta(X;\theta(P))^{\otimes 2}] \equiv A$.
(d) A natural estimator of Σ in (5) is given by $\hat{\Sigma} = \hat{B}^{-1}\hat{A}\hat{B}^{-1}$ where

$$\hat{B} \equiv -\mathbb{P}_n\left(\ddot{\mathbf{l}}_{\theta\theta}(X;\theta(\mathbb{P}_n))\right) = -n^{-1}\sum_{i=1}^n \ddot{\mathbf{l}}_{\theta\theta}(X_i;\theta(\mathbb{P}_n))$$

and

$$\hat{A} \equiv \mathbb{P}_n[\dot{\mathbf{l}}_\theta(X;\theta(\mathbb{P}_n))^{\otimes 2}] = n^{-1}\sum_{i=1}^n [\dot{\mathbf{l}}_\theta(X_i;\theta(\mathbb{P}_n))^{\otimes 2}]\,.$$

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:

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

Solution: If $T_n = n^{-1}\sum_{j=1}^n (X_j - \overline{X})^2$, then some algebra yields

$$T_{n,i}^* = nT_n - (n-1)T_{n,i} = \frac{n}{n-1}(X_i - \overline{X})^2$$

and hence

$$\overline{T}_n^* = \frac{n}{n-1}\widehat{\mu}_2\,.$$

Furthermore,

$$
\begin{aligned}
\widehat{\mathrm{Var}}_n &= \frac{1}{n(n-1)}\sum_{i=1}^{n}(T^*_{n,i} - \overline{T}^*_n)^2 \\
&= \frac{1}{n(n-1)}\sum_{i=1}^{n} T^{*2}_{n,i} - \frac{1}{n-1}(\overline{T}^*_n)^2 \\
&= \frac{1}{n(n-1)}\sum_{i=1}^{n}(\frac{n}{n-1}(X_i - \overline{X})^2)^2 - \frac{1}{n-1}(\frac{n}{n-1}\widehat{\mu}_2)^2 \\
&= \frac{1}{n-1}\frac{n^2}{(n-1)^2}\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^4 - \frac{1}{n-1}\frac{n^2}{(n-1)^2}\widehat{\mu}_2^2 \\
&= \frac{n^2}{(n-1)^3}(\widehat{\mu}_4 - \widehat{\mu}_2^2)\,.
\end{aligned}
$$

Thus we have

$$
n\widehat{\mathrm{Var}}_n = \frac{n^3}{(n-1)^3}(\widehat{\mu}_4 - \widehat{\mu}_2^2)\,. \to_p \mu_4 - \mu_2^2 = \mu_2^2\{2 + \frac{\mu_4}{\mu_2^2} - 3\} = T_2(F)(2+\gamma_2)\,.
$$

where the (excess of) kurtosis is

$$
\gamma_2 \equiv \frac{\mu_4}{\mu_2^2} - 3\,.
$$