**Statistics 583, Problem Set 9 Solutions**
Wellner; 6/2/2016

1. 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,

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

Thus we have, with $\gamma_2 \equiv \mu_4/\mu_2^2 - 3$ (the excess of kurtosis),

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

Note that if $X \sim N(\mu, \sigma^2)$, then $\gamma_2 = 0$.

2. (a) Wasserman, problem 3.8.9, page 40: Let $X_1, \ldots, X_n$ be $n$ distinct observations (no ties). Let $X_1^*, \ldots, X_n^*$ denote a bootstrap sample (from the empirical d.f. $\mathbb{F}_n$ of the $X_i$'s), and let $\overline{X}_n^* = n^{-1}\sum_{i=1}^n X_i^*$. Find: $E\{\overline{X}_n^*|X_1, \ldots, X_n\}$, $Var(\overline{X}_n^*|X_1, \ldots, X_n)$, and $Var(\overline{X}_n^*)$.
(b) Wasserman, problem 3.8.13, page 41: Let $X_1, \ldots, X_n$ be $n$ distinct observations (no ties). Let $X_1^*, \ldots, X_n^*$ denote a bootstrap sample (from the empirical d.f. $\mathbb{F}_n$ of the $X_i$'s). Let $G$ denote the marginal distribution of $X_i^*$. Note that $G(x) = P(X_i^* \le x) = E\{P(X_i^* \le x|X_1, \ldots, X_n) = E\{\mathbb{F}_n(x)\} = F(x)$. So it appears that $X_i^*$ and $X_i$ have the same distribution. But in (a) we showed that $Var(\overline{X}_n) \ne Var(\overline{X}_n^*)$. Explain.

**Solution:** (a) First, as we computed in class,

$$E(\overline{X}_n^*|X_1, \ldots, X_n) = n^{-1}\sum_{i=1}^n E(X_i^*|X_1, \ldots, X_n) = n^{-1}\sum_{i=1}^n \overline{X}_n = \overline{X}_n.$$

Similarly,

$$\begin{aligned} Var(\overline{X}_n^*|X_1, \ldots, X_n) &= n^{-2}\sum_{i=1}^n Var(X_i^*|X_1, \ldots, X_n) = n^{-2}\sum_{i=1}^n n^{-1}\sum_{j=1}^n (X_i - \overline{X}_n)^2 \\ &= n^{-1}S_n^2 \end{aligned}$$

where $S_n^2 = n^{-1}\sum_1^n (X_i - \overline{X}_n)^2$. Then it follows that

$$E(\overline{X}_n^*) = E\{E(\overline{X}_n^*|X_1, \ldots, X_n)\} = E\{\overline{X}_n\} = \mu(F),$$

and

$$\begin{aligned} Var(\overline{X}_n^*) &= E\{Var(\overline{X}_n^*|X_1, \ldots, X_n)\} + Var(E(\overline{X}_n^*|X_1, \ldots, X_n)) \\ &= E\{n^{-1}S_n^2\} + Var(\overline{X}_n) \\ &= n^{-1}\frac{n-1}{n}\sigma^2(F) + n^{-1}\sigma^2(F) \\ &= n^{-1}\sigma^2(F)\left\{\frac{n-1}{n} + 1\right\} \\ &= n^{-1}\sigma^2(F)\frac{2n-1}{n} = \frac{2n-1}{n}Var(\overline{X}_n) \approx 2Var(\overline{X}_n). \end{aligned}$$

Another way to organize this calculation is as follows:

$$\begin{aligned}
Var(\overline{X}_n^*) &= Var(\overline{X}_n^* - \overline{X}_n + \overline{X}_n - \mu + \mu) \\
&= Var(\overline{X}_n^* - \overline{X}_n) + Var(\overline{X}_n - \mu) + 2Cov(\overline{X}_n^* - \overline{X}_n, \overline{X}_n - \mu) \\
&= E\{Var(\overline{X}_n^* - \overline{X}_n)|X_1, \ldots, X_n)\} + Var\{E(\overline{X}_n^* - \overline{X}_n)|X_1, \ldots, X_n)\} \\
&\qquad + n^{-1}\sigma^2(F) \\
&\qquad + 2E\{(\overline{X}_n^* - \overline{X}_n)(\overline{X}_n - \mu)\} \\
&= E\{n^{-1}S_n^2\} + 0 + n^{-1}\sigma^2(F) + 0 \\
&= \frac{n-1}{n} n^{-1}\sigma^2(F) + n^{-1}\sigma^2(F),
\end{aligned}$$

so that the contributions of $Var(\overline{X}_n^* - \overline{X}_n)$ and $Var(\overline{X}_n - \mu)$ are approximately equal. This is important, especially since we want the first of these to estimate the second! [The marginal behavior of bootstrap estimators is largely irrelevant, since this accounts for both the deviations $T(\mathbb{F}_n^*) - T(\mathbb{F}_n)$ and $T(\mathbb{F}_n) - T(F)$ via $T(\mathbb{F}_n) - T(F) = T(\mathbb{F}_n^*) - T(\mathbb{F}_n) + T(\mathbb{F}_n) - T(F)$. What is important is that $T(\mathbb{F}_n^*) - T(\mathbb{F}_n)$ mimics (or estimates) $T(\mathbb{F}_n) - T(F)$!]

(b) The marginal distribution of the $X_i^*$'s (separately, or marginally) agrees with the marginal distribution of the (separate) $X_i$'s, but the *joint distribution* of the $X_i^*$'s is dependent. For example,

$$\begin{aligned}
G_2(x_1, x_2) &\equiv P(X_1^* \le x_1, X_2^* \le x_2) = E\{P(X_1^* \le x_1, X_2^* \le x_2 | X_1, \ldots, X_n)\} \\
&= E\{\mathbb{F}_n(x_1)\mathbb{F}_n(x_2)\} = E\{n^{-2}\sum_{i=1}^n \sum_{j=1}^n 1\{X_i \le x_1\}1\{X_j \le x_2\} \\
&= E\{n^{-2}[\sum_{i=1}^n 1\{X_i \le x_1\}1\{X_i \le x_2\} + \sum_{i \ne j} 1\{X_i \le x_1\}1\{X_i \le x_2\}]\} \\
&= n^{-2}\{nF(x_1 \wedge x_2) + n(n-1)F(x_1)F(x_2)\} \\
&= (1 - n^{-1})F(x_1)F(x_2) + n^{-1}F(x_1 \wedge x_2).
\end{aligned}$$

Thus $G_2(x_1, x_2)$ is a mixture of the independence distribution $F(x_1)F(x_2)$ and the (Fréchet bound) distribution concentrated on the diagonal, $F(x_1 \wedge x_2)$, with mixing probabilities $(1 - n^{-1})$ and $n^{-1}$. More generally,

$$\begin{aligned}
G(x_1, \ldots, x_n) &\equiv P(X_1^* \le x_1, \ldots, X_n^* \le x_n) \\
&= E\{P(X_1^* \le x_1, \ldots, X_n^* \le x_n | X_1, \ldots, X_n)\} \\
&= E\{\mathbb{F}_n(x_1) \cdots \mathbb{F}_n(x_n)\} = E\{\prod_{k=1}^n \mathbb{F}_n(x_k)\} \\
&= E\left\{ n^{-n} \sum_{j_1=1}^n \cdots \sum_{j_n=1}^n \prod_{k=1}^n 1\{X_{j_k} \le x_k\} \right\} \\
&= n^{-n} \sum_{j_1=1}^n \cdots \sum_{j_n=1}^n E\left\{ \prod_{k=1}^n 1\{X_{j_k} \le x_k\} \right\}
\end{aligned}$$

where now the computation becomes somewhat complicated. In particular, though $X_1^*$ and $X_2^*$ are correlated: from our formula (1.4.14) on page 19 of Chapter 1 (with $G(x) = H(x) = x$, $F = G_2$, and $F_X = F_Y = F$), it follows that

$$\begin{aligned}
Cov(X_1^*, X_2^*) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{G_2(x, y) - F(x)F(y)\} dx dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{(1 - n^{-1})F(x)F(y) + n^{-1}F(x \wedge y) - F(x)F(y)\} dx dy \\
&= n^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{F(x \wedge y) - F(x)F(y)\} dx dy \\
&= n^{-1} \sigma^2(F)
\end{aligned}$$

where the last equality follows from (1.14.16)-(1.14.17) on page 19, chapter 1. Thus

$$Var(\overline{X}_n^*) = n^{-2}\{n\sigma^2(F) + n(n-1)(n^{-1}\sigma^2(F))\} = n^{-1}\sigma^2(F)\{1 + \frac{n-1}{n}\}$$

in agreement with our calculation in (a).

3. Consider a $V-$functional of order $r = 2$ given by $T(P) = \int\int h(x, y) dP(x) dP(y)$ where $h$ is permutation symmetric. Find the jackknife estimate of bias for the ($V-$statistic) estimator $T(\mathbb{P}_n)$ of $T(P)$. Also find the jackknife estimator of $T(P)$.

**Solution:** Now we have $T(\mathbb{P}_n) = n^{-2} \sum_{i=1}^n \sum_{j=1}^n h(X_i, X_j) \equiv T_n$. Then

$$\begin{aligned}
T_{n,i} &= T(\mathbb{P}_{n-1,i}) = \frac{1}{(n-1)^2} \sum_{j \ne i} \sum_{j' \ne i} h(X_j, X_{j'}) \\
&= \frac{1}{(n-1)^2} \left\{ n^2 T_n - \sum_{j'=1}^n h(X_i, X_{j'}) - \sum_{j=1}^n h(X_j, X_i) + h(X_i, X_i) \right\}.
\end{aligned}$$

Thus

$$\begin{aligned} T_{n,\cdot} &= n^{-1}\sum_{i=1}^{n} T_{n,i} \\ &= \frac{1}{(n-1)^2}\left\{ n^2 T_n - \frac{1}{n}\sum_{i=1}^{n}\sum_{j'=1}^{n} h(X_i, X_{j'}) - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} h(X_j, X_i) + \frac{1}{n}\sum_{i=1}^{n} h(X_i, X_i)\right\}. \end{aligned}$$

Thus

$$\begin{aligned} T^*_{n,i} &= nT_n - (n-1)T_{n,i} \\ &= nT_n - \frac{1}{n-1}\left\{ n^2 T_n - \sum_{j'=1}^{n} h(X_i, X_{j'}) - \sum_{j=1}^{n} h(X_j, X_i) + h(X_i, X_i)\right\}, \end{aligned}$$

so that

$$\begin{aligned} \overline{T}^*_n &= nT_n - (n-1)T_{n,\cdot} \\ &= nT_n - \frac{1}{n-1}\left\{ n^2 T_n - 2nT_n + n^{-1}\sum_{i=1}^{n} h(X_i, X_i)\right\} \\ &= \left( n - \frac{n^2}{n-1} + \frac{2n}{n-1}\right) T_n - \frac{1}{n-1}\sum_{1}^{n} h(X_i, X_i) \\ &= \frac{n}{n-1}T_n - \frac{1}{n(n-1)}\sum_{1}^{n} h(X_i, X_i). \end{aligned}$$

It follows that the jackknife estimate of bias is given by

$$\begin{aligned} \widehat{\text{bias}}_n &= T_n - T^*_n = \left(1 - \frac{n}{n-1}\right) T_n + \frac{1}{n(n-1)}\sum_{1}^{n} h(X_i, X_i) \\ &= \frac{1}{n-1}\left\{ \frac{1}{n}\sum_{1}^{n} h(X_i, X_i) - T_n \right\}. \end{aligned}$$

Note that $T_n = V_n$ and

$$\begin{aligned} T^*_n &= \frac{n}{n-1}T_n - \frac{1}{n(n-1)}\sum_{1}^{n} h(X_i, X_i) \\ &= \frac{1}{n(n-1)}\left\{ \sum_{i=1}^{n}\sum_{j=1}^{n} h(X_i, X_j) - \sum_{i=1}^{n} h(X_i, X_i)\right\} \\ &= \frac{1}{n(n-1)}\sum_{i \neq j} h(X_i, X_j) = U_n, \end{aligned}$$

so that the jack-knife estimator of $T(P)$ we just found is exactly $U_n$.

4. Wasserman, problem 3.8.11, page 41: Let $X_1, \ldots, X_n$ be i.i.d. Uniform$(0, \theta)$. The MLE of $\theta$ is $\widehat{\theta}_n \equiv X_{(n)} = \max\{X_1, \ldots, X_n\}$.
(a) Find the distribution of $\widehat{\theta}_n$ and the exact and limiting distribution of $n(\theta - \widehat{\theta}_n)$.
(b) Compare the true and limiting distribution of $n(\theta - \widehat{\theta}_n)$ with the parametric and nonparametric bootstrap distributions when $\theta = 1$.
(c) Show that for the parametric bootstrap $P(\hat{\theta}_n^* = \hat{\theta}_n) = 0$ but for the nonparametric bootstrap $P(\hat{\theta}_n^* = \hat{\theta}_n) = 1 - (1 - 1/n)^n \to 1 - e^{-1} \approx .632\ldots$.

**Solution:** (a) Let $F_\theta$ denote the Uniform$(0, \theta)$ distribution. The distribution function of $\widehat{\theta}_n$ is just

$$\begin{aligned} P_\theta(\widehat{\theta}_n \le x) &= P_\theta(X_{(n)} \le x) = P_\theta(X_1 \le x, \ldots, X_n \le x) \\ &= P_\theta(X_1 \le x)^n = (x/\theta)^n, \qquad 0 \le x \le \theta, \text{ so that} \\ f_{\widehat{\theta}_n}(x) &= \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1} 1_{[0,\theta]}(x). \end{aligned}$$

Thus $\widehat{\theta}_n/\theta \stackrel{d}{=} \xi_{(n)}$, the largest order statistic of a sample $\xi_1, \ldots, \xi_n$ of i.i.d. Uniform$(0,1)$ random variables, so that $n(1 - \widehat{\theta}_n/\theta) \stackrel{d}{=} n\xi_{(1)} \to_d Y$ and $n(\theta - \widehat{\theta}_n) \to_d \theta Y$ with density $\theta^{-1}\exp(-x/\theta)1_{[0,\infty)}(x)$ where $Y \sim$ exponential(1).
(b) The parametric bootstrap estimator of

$$\begin{aligned} K_n(x, F_\theta) &\equiv P_\theta(n(\theta - \widehat{\theta}_n) \ge x) = \left(1 - \frac{x/\theta}{n}\right)^n, \quad 0 \le x \le n\theta \\ &\to \exp(-x/\theta) \quad \text{as} \quad n \to \infty \end{aligned}$$

is, since $\widehat{\theta}_n \to_{a.s.} \theta$,

$$\begin{aligned} K_n(x, F_{\widehat{\theta}_n}) &\equiv P_{\widehat{\theta}_n}(n(\widehat{\theta}_n - \widehat{\theta}_n^*) \ge x) = \left(1 - \frac{x/\widehat{\theta}_n}{n}\right)^n, \quad 0 \le x \le n\widehat{\theta} \\ &\to_{a.s.} \exp(-x/\theta) \quad \text{as} \quad n \to \infty. \end{aligned}$$

The nonparametric bootstrap estimator is given by

$$\begin{aligned} K_n(x, \mathbb{F}_n) &= P_{\mathbb{F}_n}(n(\widehat{\theta}_n - \widehat{\theta}_n^*) \ge x) = P^*(n(X_{(n)} - X_{(n)}^*) \ge x) \\ &= P^*(X_1^* \le X_{(n)} - x/n, \ldots, X_n^* \le X_{(n)} - x/n) \\ &= P^*(X_1^* \le X_{(n)} - x)^n = \mathbb{F}_n(X_{(n)} - x/n)^n. \end{aligned}$$

Thus, with $x = n(X_{(n)} - X_{(n-1)}) \to_d \theta Y$ we have

$$\begin{aligned} &P_{\mathbb{F}_n}(n(\widehat{\theta}_n - \widehat{\theta}_n^*) \ge n(X_{(n)} - X_{(n-1)}) \\ &= \mathbb{F}_n(X_{(n-1)})^n = \left(\frac{n-1}{n}\right)^n = (1 - 1/n)^n \to e^{-1}. \end{aligned}$$

(c) Furthermore for the parametric bootstrap $P_{\hat{\theta}_n}(\widehat{\theta}_n^* = \widehat{\theta}_n) = 0$, but for the nonparametric bootstrap

$$\begin{aligned} P_{\mathbb{F}_n}(\widehat{\theta}_n^* = \widehat{\theta}_n) &= P_{\mathbb{F}_n}(X_{(n)}^* = X_{(n)}) = \mathbb{F}_n(X_{(n)})^n - \mathbb{F}_n(X_{(n-1)})^n \\ &= 1 - \left(\frac{n-1}{n}\right)^n = 1 - (1 - 1/n)^n \to 1 - e^{-1}. \end{aligned}$$