Statistics 583, Problem Set 8 Solutions

Wellner; 5/25/2011

1. (a) Read Wasserman, section 3.4, pages 32 - 34. Then form confidence intervals for the skewness of the nerve data by all the methods discussed by Wasserman, section 3.4, pages 32 - 35 to see if you get results comparable to those in his table in example 3.17, page 34.
(b) Form a 95% confidence interval for the skewness parameter assuming that the nerve data can be modeled by a Weibull distribution with parameters (α, β). (That is, regard $T(P) = E_P(X - \mu(P))^3/\sigma^3(P)$ as a parametric function $g(\alpha, \beta) = T(P_{\alpha,\beta})$ for $P_{\alpha,\beta}$ a Weibull distribution on $\mathbb{R}^+$, and form a confidence interval for $g(\alpha, \beta)$ via the (parametric-) delta method. Does the resulting confidence interval include 2?

Solution (a): (i) normal bootstrap interval: based on the bootstrap standard error estimate from problem set #7 of $\hat{se}_{boot} = .163$, I get the 95% confidence interval $1.7612 \pm 1.96 * .163 = (1.442, 2.081)$.
(ii) Percentile interval: as noted by Wasserman on page 34, this is just the interval $(\theta^*_{(B\alpha/2)}, \theta^*_{B(1-\alpha/2)})$. When I compute with $B = 10^4$ I get $(\theta^*_{(B\alpha/2)}, \theta^*_{B(1-\alpha/2)}) = (1.435, 2.064)$.
(iii) Pivotal interval: as argued on Wasserman's pages 32 and 33 this interval is given by $(2\hat{\theta}_n - \theta^*_{(B(1-\alpha)/2)}, 2\hat{\theta}_n - \theta^*_{(B\alpha/2)}$. When I compute I get $(1.459, 2.088)$.
(iv) Studentized interval: this is the most computationally involved of these intervals. I proceeded as suggested in Wasserman, page 35, using the (corrected version of the) non-parametric delta method applied to the bootstrap samples. When I compute with $B = 10^4$ I get $(\hat{\theta} - z^*_{1-\alpha/2}\hat{se}_{boot}, \hat{\theta} - z^*_{\alpha/2}\hat{se}_{boot}) = (1.517, 2.281)$ where $z^*_{1-\alpha/2}$ =the α sample quantile of $Z^*_1, \ldots, Z^*_B$ and $Z^*_b \equiv (\hat{\theta}^*_b - \hat{\theta})/\hat{se}^*_b$. Here is a summary table:

method	(lower bound ,upper bound)
Normal	(1.442, 2.081)
percentile	(1.435, 2.064)
pivotal	(1.459, 2.088)
studentized pivotal	(1.517 , 2.281)

These seem to be in reasonably good agreement with the intervals obtained by Wasserman.

Solution (b): First note that

$$g(\alpha,\beta) = \frac{2\Gamma(1+1/\beta)^3 - 3\Gamma(1+1/\beta)\Gamma(1+2/\beta) + \Gamma(1+3/\beta)}{\Gamma(1+2/\beta) - \Gamma(1+1/\beta)^2}$$

is a function only of β, say $g(\beta)$. Furthermore, from our results in Stat 581,

$$\sqrt{n}(\hat{\beta}_n - \beta) \to_d N(0, \beta^2 6/\pi^2),$$

and hence, by the delta-method,

$$\sqrt{n}(g(\hat{\beta}_n) - g(\beta)) \to_d g'(\beta) N(0, \beta^2 6/\pi^2).$$

Thus a 95% parametric (model - based) confidence interval for the skewness of the nerve data is given by

$$\begin{aligned} & g(\hat{\beta}) \pm z_{.975}\sqrt{\frac{6\hat{\beta}^2[g'(\hat{\beta})]^2}{\pi^2 n}} \\ = \;& 1.77782 \pm 1.96 \cdot \frac{1.77782 \cdot 2.46058 \cdot \sqrt{6}}{\pi\sqrt{799}} \\ = \;& 1.77782 \pm 1.06 \cdot 0.0734 \\ = \;& (1.634, 1.922) \end{aligned}$$

which excludes the value 2 (the skewness of all exponential distributions), and is considerably shorter than the nonparametric CI's found in problem 1.

2. Wasserman, problem 12, page 41: Suppose that 50 people are given a placebo and 50 are given a new treatment. Thirty placebo patients show improvement, while 40 treated patients show improvement. Let $\tau = p_2 - p_1$ where p_2 is the probability of improving under treatment and p_1 is the probability of improving under placebo.
(a) Find the MLE of τ. Find the standard error and 90% confidence interval using the delta method.
(b) Find the standard error and 90% confidence interval using the bootstrap.

Solution: (a) Here $X \equiv$ number of patients showing improvement on the placebo, so $X \sim$ Binomial(m, p_1) with $m = 50$, and $Y \equiv$ number of patients showing improvement on the treatment, so $Y \sim$ Binomial(n, p_2) with $n = 50$. The MLE of $\tau = p_2 - p_1$ is simply $\hat{\tau} = \hat{p}_2 - \hat{p}_1 = Y/n - X/m = 4/5 - 3/5 = 1/5 = .20$. Furthermore, if $\lambda_N \equiv m/N \to \lambda \in [0,1]$, then

$$\begin{aligned} \sqrt{\frac{mn}{N}}(\hat{\tau} - \tau) &= \sqrt{m/N}\sqrt{n}(\hat{p}_2 - p_2) - \sqrt{n/N}\sqrt{m}(\hat{p}_1 - p_1) \\ &\to_d \sqrt{\lambda} Z_2 - \sqrt{1-\lambda} Z_1 \\ &\sim N(0, \lambda p_2 q_2 + (1-\lambda) p_1 q_1). \end{aligned}$$

Thus the standard error of $\hat{\tau}$ is

$$\begin{aligned}
\sqrt{\frac{(m/N)\hat{p}_2\hat{q}_2+(n/N)\hat{p}_1\hat{q}_1}{mn/N}} &= \sqrt{\frac{\hat{p}_2\hat{q}_2}{n}+\frac{\hat{p}_1\hat{q}_1}{m}} \\
&= \sqrt{\frac{(4/5)(1/5)}{50}+\frac{(3/5)(2/5)}{50}} \\
&= \sqrt{\frac{10}{25\cdot 50}} = \sqrt{\frac{1}{125}} = 0.0894427,
\end{aligned}$$

and a 90% confidence interval for τ is given by

$$\begin{aligned}
&\hat{\tau} \pm z_{.05}\sqrt{\frac{\hat{p}_2\hat{q}_2}{n}+\frac{\hat{p}_1\hat{q}_1}{m}} \\
&= \frac{1}{5} \pm 1.645(0.0894427) = .20 \pm 0.147 = (0.053, 0.347).
\end{aligned}$$

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

$$Y_i = \mathbf{x}_i^T\beta + \epsilon_i, \qquad i = 1, \ldots, n$$

where the ϵ_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 $\hat{\epsilon}^T = (\hat{\epsilon}_1, \ldots, \hat{\epsilon}_n)$ 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 $\hat{\epsilon}_1^*, \ldots, \hat{\epsilon}_n^*$ is a bootstrap sample (with replacement) from $\{\hat{\epsilon}_1, \ldots, \hat{\epsilon}_n\}$. 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 $Z_n = n^{-1/2}\sum_{i=1}^n \hat{\epsilon}_i$.
(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

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

[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} 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 $\max_{1\le i\le n} |c_{ni}| \to 0$ where $\mathbf{c} = n^{-1/2}(I-H)\mathbf{1}$.

Solution: (a) Now $\hat{Y} = X\hat{\beta} = X(X^TX)^{-1}X^TY = HY$, so $\hat{\underline{\epsilon}} = Y - \hat{Y} = (I-H)Y$, and

$$\begin{aligned} \hat{\underline{\epsilon}} - \underline{\epsilon} &= (I-H)Y - (Y - X\underline{\beta}) \\ &= -H(X\underline{\beta} + \underline{\epsilon}) + X\underline{\beta} = -H\underline{\epsilon} \end{aligned}$$

since $H(X\underline{\beta}) = X\underline{\beta}$ (since $X\underline{\beta}$ is already in the column space of X!)
(b). Let $\underline{\hat{\epsilon}_i^*}$ be a sample with replacement from $\{\hat{\epsilon}_i : \ i = 1, \ldots, n\}$, and let $Y_i^* = x_i\hat{\beta} + \hat{\epsilon}_i$, $i = 1, \ldots, n$. Thus in vector notation, $Y^* = X\hat{\underline{\beta}} + \hat{\underline{\epsilon}}^*$ and

$$\begin{aligned} \hat{\underline{\beta}}^* &= (X^TX)^{-1}X^TY^* = (X^TX)^{-1}X^T(X\hat{\underline{\beta}} + \hat{\underline{\epsilon}}^*) \\ &= \hat{\underline{\beta}} + (X^TX)^{-1}X^T\hat{\underline{\epsilon}}^*. \end{aligned}$$

Thus

$$\sqrt{n}(\hat{\underline{\beta}}^* - \hat{\underline{\beta}}) = (n^{-1}X^TX)^{-1}(n^{-1}X^T)\sqrt{n}\hat{\underline{\epsilon}}^*,$$

and, since $E_*(\hat{\epsilon}_i^*) = n^{-1}\sum_{i=1}^n \hat{\epsilon}_i$ the expected value is

$$\begin{aligned} E_*(\sqrt{n}(\hat{\underline{\beta}}^* - \hat{\underline{\beta}}) &= (n^{-1}X^TX)^{-1}(n^{-1}X^T)\underline{1}n^{-1/2}\sum_{i=1}^n \hat{\epsilon}_i \\ &= (n^{-1}X^TX)^{-1}(n^{-1}X^T)\underline{1}Z_n. \end{aligned}$$

(c). To show that

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

first write $\sqrt{n}(\hat{\underline{\beta}} - \underline{\beta}) = \sqrt{n}(X^TX)^{-1}X^T\underline{\epsilon}$ so that, for any fixed vector $\lambda \in \mathbb{R}^p$,

$$\underline{\lambda}^T\sqrt{n}(\hat{\beta} - \beta) = \sqrt{n}\underline{\lambda}^T(X^TX)^{-1}X^T\underline{\epsilon} \equiv \sum_{i=1}^n a_{ni}\epsilon_i \equiv \sum_{i=1}^n X_{n,i}$$

where the vector $a_n \equiv \sqrt{n} X (X^T X)^{-1} \underline{\lambda}$. Hence we have $EX_{ni} = 0$, $Var(X_{ni}) = a_{ni}^2 \sigma^2$, and

$$\begin{aligned} \sigma_n^2 &\equiv \sum_{i=1}^n \sigma_{ni}^2 = \sigma^2 |\underline{a}|^2 \\ &= \sigma^2 n \underline{\lambda}^T (X^T X)^{-1} \underline{\lambda} \\ &\to \sigma^2 \underline{\lambda}^T V^{-1} \underline{\lambda} > 0 \end{aligned}$$

since V is positive definite.

To check the Lindeberg condition, write

$$\begin{aligned} &\frac{1}{\sigma_n^2} \sum_{i=1}^n E\{|X_{ni}|^2 1_{[|X_{ni}| > \epsilon \sigma_n]}\} \\ &= \frac{1}{\sigma_n^2} \sum_{i=1}^n a_{ni}^2 E\{\epsilon_1^2 1_{[|\epsilon_1| > \epsilon \sigma_n / |a_{ni}|]}\} \\ &\le \frac{1}{\sigma^2} E\{\epsilon_1^2 1_{[|\epsilon_1| > \epsilon \sigma_n / \max_i |a_{ni}|]}\} \\ &\to 0 \qquad \text{by the DCT since } \; E(\epsilon^2) < \infty \end{aligned}$$

if $\max_{1 \le i \le n} |a_{ni}|^2 \to 0$. But we can write $a_{ni} = \sqrt{n} \underline{x}_i^T (X^T X)^{-1} \underline{\lambda}$, so that, by Cauchy - Schwarz,

$$\begin{aligned} \max_{1 \le i \le n} |a_{ni}|^2 &\le n \max_{1 \le i \le n} (\underline{x}_i^T (X^T X)^{-1} \underline{x}_i)(\underline{\lambda}^T (X^T X)^{-1} \underline{\lambda}) \\ &= \max_{1 \le i \le n} h_{ii} \underline{\lambda}^T (n^{-1} X^T X)^{-1} \to \; 0 \cdot \underline{\lambda}^T V^{-1} \underline{\lambda} = 0. \end{aligned}$$

Hence, by the Lindeberg - Feller CLT and (f)

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

By the Cramér - Wold device, this yields (e) under the hypothesis $\max h_{ii} \to 0$.
(d) To calculate the variance, we first note that from (a) $E(\hat{\epsilon}) = (I - H) E(\epsilon) = 0$, and $E(Z_n) = 0$. Similarly,

$$\begin{aligned} Var(Z_n) &= \frac{1}{n} (\underline{1}^T (I - H)(I - H) \underline{1}) \sigma^2 \\ &= \sigma^2 \left\{ 1 - (n^{-1} \underline{1}^T X)(n^{-1} X^T X)^{-1} (n^{-1} X^T \underline{1}) \right\}. \end{aligned}$$

(e) If $n^{-1} X^T \underline{1} \to h$, $n^{-1} X^T X \to V$ with V positive definite, and $h^T V^{-1} h < 1$, then from (d)

$$\begin{aligned} Var(Z_n) &= \sigma^2 \left\{ 1 - (n^{-1} \underline{1}^T X)(n^{-1} X^T X)^{-1} (n^{-1} X^T \underline{1}) \right\} \\ &\to \sigma^2 \{1 - h^T V^{-1} h\} \equiv \sigma^2 c^2 > 0. \end{aligned}$$

To show that $Z_n \rightarrow_d$ using the Lindeberg - Feller CLT requires a bit more in the way of hypotheses and some more work: write

$$Z_n = n^{-1/2}\underline{1}^T(I-H)\underline{\epsilon} \equiv \sum_{i=1}^n c_{ni}\epsilon_i \equiv \sum_{i=1}^n X_{ni};$$

thus the vector $\underline{c}_n = n^{-1/2}(I-H)\underline{1}$. Thus $E(X_{ni}) = 0$, $\sigma_{ni}^2 = Var(X_{ni}) = c_{ni}^2\sigma^2$, and, as above,

$$\begin{aligned} \sigma_n^2 &= \sum_{i=1}^n \sigma_{ni}^2 = \sigma^2 \sum_{i=1}^n c_{ni}^2 = \sigma^2 \underline{c}^T \underline{c} \\ &= \sigma^2\{1 - n^{-1}\underline{1}^T X(X^TX)^{-1}X^T\underline{1}\} \rightarrow \sigma^2(1 - h^TV^{-1}h) \end{aligned}$$

under the above hypotheses. Finally, if $\max_{1\le i\le n}|c_{ni}| \rightarrow 0$, then, for $\epsilon > 0$,

$$\begin{aligned} &\frac{1}{\sigma_n^2}\sum_{i=1}^n E\{|X_{ni}|^2 1_{[|X_{ni}|>\epsilon\sigma_n]}\} \\ &\quad = \frac{1}{\sigma_n^2}\sum_{i=1}^n c_{ni}^2 E\{\epsilon_1^2 1_{[|\epsilon_1|>\epsilon\sigma_n/|c_{ni}|]}\} \\ &\quad \le \frac{1}{\sigma^2}E\{\epsilon_1^2 1_{[|\epsilon_1|>\epsilon\sigma_n/\max_i|c_{ni}|]}\} \\ &\quad \rightarrow 0 \qquad \text{by the DCT since } \; E(\epsilon^2) < \infty \end{aligned}$$

if $\max_{1\le i\le n}|c_{ni}|^2 \rightarrow 0$. But

$$\begin{aligned} |c_{ni}| &\le n^{-1/2} + n^{-1/2}|\sum_{j=1}^n h_{ij}| = n^{-1/2} + n^{-1/2}\underline{1}^T\underline{h} \\ &\le n^{-1/2} + n^{-1/2}\sqrt{\underline{1}^T\underline{1}}\sqrt{\underline{hh}} \\ &= n^{-1/2} + \sqrt{\sum_{j=1}^n h_{ij}^2} = n^{-1} + \sqrt{h_{ii}}, \end{aligned}$$

using $H = H^T$ and $HH = H$, so $\max_i|c_{ni}| \le n^{-1/2} + \sqrt{\max_i h_{ii}} \rightarrow 0$ by (iii). Thus the Lindeberg - Feller CLT yields $Z_n/\sigma_n \rightarrow_d N(0,1)$; combining this with (d) yields

$$\sqrt{n}E_*(\hat{\beta}^* - \hat{\beta}) \rightarrow_d V^{-1}hN(0, c^2\sigma^2).$$

We conclude from (e) and (i) that the bootstrap *fails* in the situation (at least under the additional hypothesis that $\max h_{ii} \rightarrow 0$).

4. Suppose now that the bootstrap residuals are drawn from the collection of *centered* residuals $\hat{\epsilon} - \mathbf{1}(\mathbf{1}^T\hat{\epsilon}/\mathbf{n})$ Compute $E_*(\sqrt{n}(\hat{\beta}^* - \hat{\beta}))$ and $E_*(\sqrt{n}(\hat{\beta}^* - \hat{\beta}))^{\otimes 2}$ for this bootstrap resampling scheme.

   Solution: When the resampling is done from the *centered* residuals $\underline{\hat{\epsilon}} - \underline{1}(\underline{1}^T\underline{\hat{\epsilon}}/n)$, the nonzero term in $E_*(\sqrt{n}(\underline{\hat{\beta}}^* - \underline{\hat{\beta}}))$ which we investigated in problem 4 above vanishes: Since

$$\sqrt{n}(\underline{\hat{\beta}}^* - \underline{\hat{\beta}}) = (X^T X)^{-1} X^T \underline{\hat{\epsilon}}^*,$$

   where

$$E_*(\underline{\hat{\epsilon}}^*) = \underline{1}\frac{1}{n}\sum_{i=1}^{n}(\hat{\epsilon}_i - \underline{1}^T\underline{\hat{\epsilon}}/n) = \underline{0},$$

   it follows that

$$E_*\{\sqrt{n}(\underline{\hat{\beta}}^* - \underline{\hat{\beta}}))\} = (X^T X)^{-1} X^T E_*(\underline{\hat{\epsilon}}^*) = \underline{0}.$$

   Furthermore,

$$\begin{aligned} E_*\{[\sqrt{n}(\underline{\hat{\beta}}^* - \underline{\hat{\beta}})]^{\otimes 2}\} &= n(X^T X)^{-1} X^T E_*(\underline{\hat{\epsilon}}^* \underline{\hat{\epsilon}}^{*T}) X (X^T X)^{-1} \\ &= n(X^T X)^{-1} \hat{\sigma}_F^2 \end{aligned}$$

   since

$$E_*(\underline{\hat{\epsilon}}^* \underline{\hat{\epsilon}}^{*T}) = I\frac{1}{n}\sum_{i=1}^{n}(\hat{\epsilon}_i - \underline{1}\underline{\hat{\epsilon}}/n)^2 \equiv \hat{\sigma}_F^2 I.$$

   This modification of the bootstrap procedure seems appropriate when the design matrix X does not contain a column of 1's. See Freedman (1981), *Ann. Statist.* **9**, 1218 - 1228; especially the discussion on page 1220, the positive theorem on page 1223, and the discussion on page 1224 (upon which this problem is based).

5. Wasserman, problem 2, page 59.

   Solution: By easy calculation,

$$\begin{aligned} R(f(x), \widehat{f}_n(x)) &= E(f(x) - \widehat{f}_n(x))^2 \\ &= E(f(x) - E\widehat{f}_n(x) + E\widehat{f}_n(x) - \widehat{f}_n(x))^2 \\ &= E\{(f(x) - E\widehat{f}_n(x))^2 \\ &\qquad + 2(f(x) - E\widehat{f}_n(x))E(E\widehat{f}_n(x) - \widehat{f}_n(x)) + (E\widehat{f}_n(x) - \widehat{f}_n(x))^2\} \\ &= (f(x) - E\widehat{f}_n(x))^2 + 0 + E\{(E\widehat{f}_n(x) - \widehat{f}_n(x))^2\} \\ &= \text{bias}_n^2(x) + Var[\widehat{f}_n(x)]. \end{aligned}$$

   Note that this is really a general calculation for estimation with squared error loss (which we carried out in a more general setting in Chapter 5).