

## Statistics 583, Problem Set 8

Wellner; 5/15/2006

**Reading:** Chapter 8, sections 8.4 - 8.5, pages 18 - 27;

Wasserman, Chapter 3, pages 13-41; read Chapter 4

**Due:** Wednesday, May 24, 2006

1. Read Wasserman, section 3.4, pages 32 - 24. 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.
2. Form a 95% confidence interval for the skewness parameter assuming that the nerve data can be modeled by a Weibull distribution with parameters  $(\alpha, \beta)$ . (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?
3. Wasserman, problem 12, page 41.
4. (Bootstrapping a linear regression model a simple way.) Consider bootstrapping a linear regression model

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

where the  $\epsilon_i$  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  $\hat{\epsilon}^T = (\hat{\epsilon}_1, \dots, \hat{\epsilon}_n)$  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  $\hat{\epsilon}_1^*, \dots, \hat{\epsilon}_n^*$  is a bootstrap sample (with replacement) from  $\{\hat{\epsilon}_1, \dots, \hat{\epsilon}_n\}$ . Show that

$$E_*(n^{1/2}(\hat{\beta}^* - \hat{\beta})) = \left(\frac{1}{n} \mathbf{X}^T \mathbf{X}\right)^{-1} \left(\frac{1}{n} \mathbf{X}^T \mathbf{1}\right) Z_n$$

where  $Z_n = n^{-1/2} \sum_{i=1}^n \hat{\epsilon}_i$ .

(c) Show that if  $\max_{1 \leq i \leq n} h_{ii} \rightarrow 0$ , and  $n^{-1} \mathbf{X}^T \mathbf{X} \rightarrow V$ , a positive definite matrix, then

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

[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)  $n^{-1}\mathbf{X}^T\mathbf{X} \rightarrow V$ , a positive definite matrix;

(ii)  $\mathbf{X}^T\mathbf{1}/n \rightarrow \mathbf{h}$  with  $\mathbf{h}^TV^{-1}\mathbf{h} < 1$ ;

(iii)  $\max_{1 \leq i \leq n} h_{ii} \rightarrow 0$  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  $\max_{1 \leq i \leq n} |c_{ni}| \rightarrow 0$  where  $\mathbf{c} = n^{-1/2}(I - H)\mathbf{1}$ .

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