Statistics 581, Problem Set 6

Wellner; 11/6/2009

Reading: Lecture Notes Chapter 3, sections 1-2;
Ferguson, ACILST, chapters 19-20, pages 126 - 139;
Lehmann and Casella, TPE, Sections 2.5 and 2.6, pages 113 - 129;
and Section 6.2, pages 437 - 443.

Due: Friday, November 13, 2009.

1. Chapter 2, Exercise 5.3, page 25. [Hint: One approach uses the fact that $\mathbb{S}_n(t_j) - \mathbb{S}_n(t_{j-1}) = n^{-1/2}\sum_{i=[nt_{j-1}]+1}^{[nt_j]} X_i$, $j = 1, \ldots, k$ with $t_0 \equiv 0$ are independent random variables.]

2. Suppose that $X_1, \ldots, X_n$ are i.i.d. with the Weibull distribution F_θ given by

$$1 - F_\theta(x) = \exp(-(x/\alpha)^\beta), \qquad x \ge 0$$

where $\theta = (\alpha, \beta) \in (0, \infty) \times (0, \infty)$.
(a) Find the inverse (or quantile function) $F_\theta^{-1}(u)$ corresponding to F_θ in terms of α, β, and $u \in (0, 1)$, and show that

$$\log F_\theta^{-1}(u) = \log \alpha + \frac{1}{\beta} \log\log\left(\frac{1}{1-u}\right).$$

(b) Fix $r \in (0, 1/2)$ and $s \in (1/2, 1)$ Use the $r-$th and $s-$th quantiles of the X_i's, namely $\mathbb{F}_n^{-1}(r)$ and $\mathbb{F}_n^{-1}(s)$, to obtain simple consistent estimators $\hat{\alpha}_n$ and $\hat{\beta}_n$ of α and β. Prove that your estimators are consistent.
(c) Prove that your estimators $\hat{\alpha}_n$ and $\hat{\beta}_n$ satisfy

$$\sqrt{n}\begin{pmatrix} \hat{\alpha}_n - \alpha \\ \hat{\beta}_n - \beta \end{pmatrix} \rightarrow_d N_2(0, \Sigma)$$

and identify Σ as a function of α, β, and t.
(d) How would you choose r and s to minimize the asymptotic variance of $\hat{\beta}_n$?

3. Ferguson, ACILST, problem 6, page 93, plus the following:
(d) Construct a family of estimators $\tilde{\theta}_n$ of θ based on the sample quantile function $\mathbb{F}_n^{-1}(t)$. Show that your estimators are consistent and asymptotically normal. Give a formula for the asymptotic variance of your estimators.

4. 1. A. Let $\mathbb{U}_X$ and $\mathbb{U}_Y$ be two independent Brownian bridge processes on $[0, 1]$, and let $\lambda \in [0, 1]$. Show that the process $\mathbb{U}$ defined by $\mathbb{U} = \sqrt{1-\lambda}\,\mathbb{U}_X - \sqrt{\lambda}\,\mathbb{U}_Y$ is also a Brownian bridge process.
B. Suppose that $X_1, \ldots, X_m$ are i.i.d. F and $Y_1, \ldots, Y_n$ are i.i.d. G, with the X's and Y's independent. Let $\mathbb{F}_m$ and $\mathbb{G}_n$ denote the empirical df's of the X's and Y's respectively. Suppose that $\lambda_N \equiv m/N \to \lambda \in (0, 1)$ where $N \equiv m + n$. Show that

$$\begin{aligned} \mathbb{X}_{m,n} &\equiv \sqrt{\frac{mn}{N}}(\mathbb{F}_m - F) - \sqrt{\frac{mn}{N}}(\mathbb{G}_n - G) \\ &\Rightarrow \sqrt{1-\lambda}\,\mathbb{U}_X(F) - \sqrt{\lambda}\,\mathbb{U}_Y(G) \equiv \mathbb{X}. \end{aligned}$$

C. A distribution function F is said to be *stochastically smaller than another distribution function* G, and we write $F <_s G$, if $F(x) \geq G(x)$ for all $x \in \mathbb{R}$ with strict inequality for some $x \in \mathbb{R}$. Note that this means $F^{-1}(u) \leq G^{-1}(u)$ for all $0 < u < 1$ so that the random variables resulting from a Skorokhod construction with one uniform random variable ξ satisfy satisfy $X^* \equiv F^{-1}(\xi) \leq G^{-1}(\xi) \equiv Y^*$. Consider testing $H_0 : \ F = G$ continuous versus $H_1 : \ F <_s G$ based on the one-sided Kolmogorov-Smirnov statistic

$$D_{m,n}^+ \equiv \sqrt{\frac{mn}{N}} \|(\mathbb{F}_m - \mathbb{G}_n)^+\|_\infty = \sqrt{\frac{mn}{N}} \sup_{x \in \mathbb{R}} (\mathbb{F}_m(x) - \mathbb{G}_n(x));$$

here the notation f^+ is the *positive part* of the function f: $f^+(x) \equiv \max\{f(x), 0\}$. Use the result of B to show that under H_0 it follows that

$$D_{m,n}^+ \to_d \|\mathbb{U}^+\|_\infty = \sup_{0 \leq t \leq 1} \mathbb{U}(t).$$

D. To test the effectiveness of vitamin B_1 in stimulating growth in mushrooms, vitamin B_1 was applied to 13 mushrooms selected at random from a group of 24, while the remaining 11 did not receive this treatment. The weights of the mushrooms at the end of the period of observation were:

$\underline{X} = (18, 14.5, 13.5, 12.5, 23, 24, 21, 17, 18.5, 9.5, 14)$, $m = 11$;
$\underline{Y} = (27, 34, 20.5, 29.5, 20, 28, 20, 26.5, 22, 24.5, 34, 35.5, 19)$, $\quad n = 13$.

Plot the two empirical df's and compute $D_{m,n}^+$. What is the approximate P - value for testing H_0 versus $H_1 : \ F <_s G$? You may use your favorite tables of the distribution of $D_{m,n}^+$, or the asymptotic distribution.

5. **Optional bonus problem:** Course notes, Chapter 2, Exercise 5.6, page 27.