Statistics 581, Problem Set 6

Wellner; 11/7/2001

Reading: Lecture Notes Chapter 3; Ferguson, ACILST, chapter 19, pages 126 - 132; Lehmann and Casella, TPE, Sections 2.5 and 2.6, pages 113 - 129; and Section 6.2, pages 437 - 443.

Due: Wednesday, November 15, 2001.

1. 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
$$\sqrt{n} \begin{pmatrix} \hat{\alpha}_n - \alpha \\ \hat{\beta}_n - \beta \end{pmatrix} \to_d N_2(0, \Sigma)$$
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 $\hat{\beta}_n$?

2. Suppose that $X, X_1, X_2, \ldots, X_n$ are independent Poisson(λ) random variables:
$$P(X = k) \equiv p_k = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \ldots.$$
Note that
$$\frac{p_k}{p_{k-1}} = \frac{\lambda}{k},$$
and hence whole family of alternative estimators $\{\tilde{\lambda}_n^{(k)}\}_{k \ge 1}$ is given by
$$\tilde{\lambda}_n^{(k)} = k \frac{\hat{p}_n(k)}{\hat{p}_n(k-1)}$$
where $\hat{p}_n(k) \equiv n^{-1} \sum_{i=1}^n 1_{[X_i = k]}$.
(a) Show that $\tilde{\lambda}_n \to_p \lambda$ for each $k = 1, 2, \ldots$.
(b) Show that
$$\sqrt{n}(\tilde{\lambda}_n^{(k)} - \lambda) \to_d N(0, \sigma_k^2(\lambda)) \quad \text{as} \quad n \to \infty$$
and compute $\sigma_k^2(\lambda)$ explicitly as a function of k and λ.
(c) What is the asymptotic relative efficiency of $\tilde{\lambda}_n^{(k)}$ to $\widehat{\lambda}_n = \overline{X}_n$ for $k > 1$? (The ARE of $\tilde{\lambda}_n^{(1)}$ with respect to $\widehat{\lambda}_n$ was computed in the Midterm Exam Solutions.)

3. Suppose that Suppose that $X_1, \ldots, X_n, \ldots$ are i.i.d. random vectors in R^k with common distribution function F and corresponding probability measure P on $(R^k, \mathcal{B}_k)$. Let $\mathbb{P}_n$ be the empirical measure defined by

$$\mathbb{P}_n = n^{-1} \sum_{i=1}^{n} \delta_{X_i} \,,$$

and consider $\mathbb{P}_n$ and the empirical process $\mathbb{G}_n$ as indexed by a class of sets $\mathcal{C} \subset \mathcal{B}_k$:

$$\{\mathbb{P}_n(C) : \; C \in \mathcal{C}\}, \qquad \{\mathbb{G}_n(C) : \; C \in \mathcal{C}\} \,,$$

where

$$\mathbb{G}_n \equiv \sqrt{n}(\mathbb{P}_n - P) \,.$$

(a) Show that $\mathbb{G}_n \to_{f.d.} \mathbb{G}_P$ where $\mathbb{G}_P$ is a P-Brownian bridge process indexed by $\mathcal{C}$: i.e. show that for any integer m and sets $C_1, \ldots, C_m \in \mathcal{C}$,

$$(\mathbb{G}_n(C_1), \ldots, \mathbb{G}_n(C_m)) \to_d (\mathbb{G}_P(C_1), \ldots, \mathbb{G}_P(C_m)) \sim N_m(0, \Sigma)$$

where $\Sigma = (\sigma_{jj'})$ is given by

$$\sigma_{jj'} = P(C_j \cap C_{j'}) - P(C_j)P(C_{j'}) \,.$$

(b) When $\mathcal{C} = \mathcal{O} \equiv \{(-\infty, x] : \; x \in R^k\}$ specialize the result in (a) and show that it gives the finite-dimensional convergence of the empirical distribution function $\mathbb{F}_n$: i.e.

(i) show that $\mathbb{P}_n((-\infty, x]) = \mathbb{F}_n(x)$;

(ii) show that $P((-\infty, x]) = F(x)$;

(iii) show that $\mathbb{Y}(x) \equiv \mathbb{G}_P((-\infty, x])$ has mean zero and covariance

$$E\{\mathbb{Y}(x)\mathbb{Y}(y)\} = F(x \wedge y) - F(x)F(y), \qquad x, y \in R^k \,.$$

4. **Optional bonus problem:** Consider the *empirical process* $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$ as a process indexed by $\mathcal{F}$. Thus

$$\mathbb{G}_n(f) = \sqrt{n}(\mathbb{P}_n(f) - P(f)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f(X_i) - P(f)) \qquad \text{for all} \qquad f \in \mathcal{F}.$$

Show that

$$\mathbb{G}_n \to_{f.d.} \mathbb{G}_P$$

where $\mathbb{G}_P$ is a $P-$Brownian bridge process indexed by $\mathcal{F} \subset L_2(P)$ [so $\mathbb{G}_P$ is mean - zero Gaussian with covariance $Cov(\mathbb{G}_P(f), G_P(g)) = P(fg) - P(f)P(g)$, $f, g \in \mathcal{F}$]; i.e. show that for any integer k and $f_1, \ldots, f_k \in \mathcal{F}$

$$(\mathbb{G}_n(f_1), \ldots, \mathbb{G}_n(f_k)) \to_d (\mathbb{G}_P(f_1), \ldots, \mathbb{G}_P(f_k)) \sim N_k(0, \Sigma)$$

where $\Sigma = (\sigma_{ij})$ and $\sigma_{ij} = P(f_i f_j) - P(f_i)P(f_j)$.
[Note that problem 4 is the special case of the optional problem with $\mathcal{F} = \{1_C : C \in \mathcal{D}\}$, and problem 2, problem set #6 is the special case of problem this optional problem with $\mathcal{X} = R$ and the class of sets $\mathcal{D} = \{(-\infty, x] : x \in R\}$.]