Statistics 581, Problem Set 6

Wellner; 11/2/2005

Reminder: Midterm exam: Wednesday, November 9.
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: Wednesday, November 9, 2002.

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

2. Consider a function $T : \mathcal{F} \to \mathbb{R}$ where $\mathcal{F}$ is some (sub) class of distribution functions F (examples include the mean, $T(F) = \mu(F) = \int x dF(x)$, the variance $T(F) = \sigma^2(F) = \int (x - \int y dF(y))^2 dF(x)$, the median $T(F) = F^{-1}(1/2)$, linear combinations of order statistics $T(F) = \int_0^1 F^{-1}(u)w(u)du$, the *Lorenz curve* (at $t \in (0,1)$), $T(F) = \int_0^t F^{-1}(u)du / \int_0^1 F^{-1}(u)du \equiv L(t,F)$, and so forth). [The Lorenz curve gives the percentage of "income" received by the poorest fraction t of the income distribution.] The "principle of substitution" says that $T(F)$ can be estimated by $T(\widehat{F}_n)$. for some estimator $\widehat{F}_n$ of F. If T is sufficiently "smooth", then frequently the empirical distribution function $\mathbb{F}_n$ can be taken as the estimator $\hat{F}_n$ of F.

   Give a treatment of consistency and asymptotic normality of the estimator $L(t, \mathbb{F}_n)$ of $L(t,F)$ based on our results from sections 2.4 and 2.6. You may assume that with $X \sim F$ we have $E_F|X| < \infty$ and $E_F X^2 < \infty$ (and any other additional assumptions you need).

3. Suppose that $Z \sim N(0,1)$ and, for $\mu \in R$ and $\sigma > 0$, that $X = \mu + \sigma Z \sim P_{\mu,\sigma} = N(\mu, \sigma^2)$.
   (a) Compute the likelihood ratio
$$\frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(x) = \frac{\sigma^{-1}\phi((x-\mu)/\sigma)}{\sigma^{-1}\phi(x/\sigma)} \qquad \text{and} \qquad Y \equiv \log \frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(X) .$$
   What is the distribution of Y under $P_{0,\sigma}$ and under $P_{\mu,\sigma}$?
   (b) Plot the function
$$l(\mu, \sigma; X) \equiv \log \frac{dP_{\mu,\sigma}}{dP_{0,\sigma}}(X)$$
   as a function of μ.
   (c) Find the maximum value of the function $l(\mu; X)$ in B (as a function of μ) and the value of $\mu \equiv \hat{\mu}$ which achieves the maximum.
   (d) What is the distribution of $\hat{\mu}$ under $P_{0,\sigma}$ and under $P_{\mu,\sigma}$? What is the distribution of $l(\hat{\mu}; X)$ under $P_{0,\sigma}$ and under $P_{\mu,\sigma}$?

4. 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 $k > 1$?

5. **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 $\mathcal{F} = \{1_C : C \in \mathcal{C}\}$.]