Statistics 581, Problem Set 6

Wellner; 11/6/2002

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 13, 2002.

1. 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 $P_{\mu,\sigma}$?

2. Suppose that $X, X_1, X_2, \ldots, X_n$ are i.i.d. with exponential$(1/\theta)$ distribution given by $1 - F_\theta(x) = \exp(-x/\theta)$, $x \geq 0$. Thus the density of X is $f(x;\theta) = (1/\theta)\exp(-x/\theta)1_{[0,\infty)}(x)$.
   (a) Let $M_n = \mathbb{F}_n^{-1}(1/2)$ be the sample median. Find a constant c so that $cM_n \rightarrow_p \theta$, and then show that $\sqrt{n}(cM_n - \theta) \rightarrow_d N(0,\sigma^2)$ for some σ^2 and find σ^2 in terms of θ.
   (b) Do the same with M_n replaced by the $p-$th sample quantile $\mathbb{F}_n^{-1}(p)$ where $0 < p < 1$. Find p which minimizes the asymptotic variance.

3. Suppose that $X_1, \ldots, X_n$ are i.i.d. with distribution function F which has positive density f at its quartiles $F^{-1}(1/4)$ and $F^{-1}(3/4)$ and at its median $F^{-1}(1/2)$.
   (a) Let $Q_n = (X_{(3n/4)} + X_{(n/4)})/2 = (\mathbb{F}_n^{-1}(3/4) + \mathbb{F}_n^{-1}(1/4))$, the mid-quartile range. Find the asymptotic distribution of Q_n as an estimator of the population mid-quartile range $Q = Q(F) = (F^{-1}(3/4) + F^{-1}(1/4))/2$. That is, prove that
   $$\sqrt{n}(Q_n - Q) \rightarrow_d \text{“}\textit{something}\text{”}$$
   and find “something”.
   (b) Assuming that the underlying distribution is Cauchy(μ,σ) ($X = \sigma Y + \mu$ where $Y \sim$ Cauchy$(0,1)$), compare the variances of the mid-quartile range Q_n and the median M_n as estimators of μ.

4. 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

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

here δ_x is the measure which puts mass 1 as x:

$$\delta_x(A) = 1_A(x) = \begin{cases} 1 & \text{if} \quad x \in A \\ 0 & \text{if} \quad x \notin A \, . \end{cases}$$

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

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

5. **Optional bonus problem**.
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 $\{\tilde{\lambda}_n^{(k)}\}_{k \geq 1}$ 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.)

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