Statistics 581, Midterm Exam

Wellner; 11/03/2008

This exam is to be taken without any books or notes.

1. (24 points) **Define** any **three** of the following five terms.
   (a) A *Brownian bridge process* $\mathbb{U}$.
   (b) *Convergence in distribution* of a sequence of random vectors X_n in $\mathbb{R}^k$.
   (c) *A normal random vector* $Y = (Y_1, \ldots, Y_n)$.
   (d) A *non-central chi-square distribution* with m degrees of freedom and non-centrality parameter δ.
   (e) The *Hellinger distance* between two probability measures P and Q on a measurable space $(\mathcal{X}, \mathcal{A})$.

2. (36 points). **State** any **three** of the following:
   (a) The Lindeberg-Feller central limit theorem.
   (b) The inverse transformation theorem.
   (c) The multivariate delta method or g' theorem.
   (d) The Mann-Wald or continuous mapping theorem.
   (e) The Glivenko-Cantelli theorem.
   (f) A limit theorem about the joint limiting distribution of $(\sqrt{n}(\mathbb{F}_n^{-1}(s) - F^{-1}(s)), \sqrt{n}(\mathbb{F}_n^{-1}(t) - F^{-1}(t)))$ where $0 < s < t < 1$ are fixed.

   Do **either** problem 3 **or** problem 4.

3. (40 points) Suppose that $\underline{N} = (N_1, \ldots, N_k) \sim \text{Mult}_k(n, \underline{p})$ where $\underline{p} = (p_1, \ldots, p_k)$. In class and homework problems we have discussed the chi-square statistic Q_n and the Hellinger distance statistic $4nH_n^2$ as test statistics for testing $H : \underline{p} = \underline{p_0}$ versus $K : \underline{p} \neq \underline{p}_0$. An alternative statistic for testing H versus K is the likelihood ratio statistic $2 \log \lambda_n$ where

$$\lambda_n \equiv \frac{\sup_{\underline{p}} L_n(\underline{p})}{L_n(\underline{p}_0)} = \frac{\prod_{j=1}^k \widehat{p}_j^{N_j}}{\prod_{j=1}^k p_{0j}^{N_j}} = \prod_{j=1}^k \left\{ \frac{\widehat{p}_j}{p_{0j}} \right\}^{N_j} .$$

   (a) Show that

$$2 \log \lambda_n = 2n \sum_{j=1}^k \widehat{p}_j \log \left(\frac{\widehat{p}_j}{p_{0j}} \right) .$$

   (b) If the alternative hypothesis K is true, so $\underline{p} \neq \underline{p}_0$, show that

$$n^{-1} 2 \log \lambda_n = g(\underline{\widehat{p}}) \to_p g(\underline{p}) ,$$

and identify $g(\underline{p})$ as a function of $\underline{p}$ and $\underline{p_0}$.
(c) If the alternative hypothesis K is true, so $\underline{p} \neq \underline{p}_0$, show that

$$\sqrt{n}(2n^{-1}\log\lambda_n - g(\underline{p})) = \sqrt{n}(g(\widehat{\underline{p}}) - g(\underline{p})) \to_d N(0, V^2(\underline{p}))\,,$$

and compute $V^2(\underline{p})$. Could you use this to approximate the power of the likelihood-ratio test? How?

4. (40 points). Suppose that $X_1, \ldots, X_n$ are i.i.d. with distribution function F having a continuous density function f. Let $\mathbb{F}_n$ be the empirical distribution function of the X_i's, and suppose that b_n is a sequence of positive numbers, and let

$$\hat{f}_n(x) = \frac{\mathbb{F}_n(x+b_n) - \mathbb{F}_n(x-b_n)}{2b_n}.$$

(a) Show that $E\hat{f}_n(x) \to f(x)$ if $b_n \to 0$.
(b) Show that $Var(\hat{f}_n(x)) \to 0$ if $b_n \to 0$ and $nb_n \to \infty$.
(c) Use some appropriate central limit theorem to show that (perhaps under some suitable further conditions that you might need to specify)

$$\sqrt{2nb_n}(\hat{f}_n(x) - E\hat{f}_n(x)) \to_d N(0, f(x)).$$

Hint: Write $\hat{f}_n(x)$ in terms of some Bernoulli random variables and identify $p = p_n$.
Note: This estimator $\hat{f}_n$ is a *kernel density estimator* based on the uniform kernel $k(x) = 1_{[-1,1]}(x)/2$, and can be rewritten as

$$\hat{f}_n(x) = \int_{-\infty}^{\infty} \frac{1}{b_n} k\left(\frac{x-y}{b_n}\right) d\mathbb{F}_n(y);$$

other kernel density estimators result when the uniform kernel is replaced by some other density function.

5. (35 points). Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. Uniform$(0,\theta)$. Consider $\hat{\theta}_n = \max_{1\le i\le n} X_i = X_{(n)}$ as an estimator of θ.
(a) Compute $E(\hat{\theta}_n)$.
(b) Show that $Y_n = n(\theta - \hat{\theta}_n) \to_d Y$ and find the distribution of Y.
(c) Show that $\hat{\theta}_n \to_p \theta$.
(d) Consider the function $g(x) = (1-x)^{-2}$. Does $g(Y_n) \to_d$ something? If the answer is yes, what is the limit (expressed in terms of the random variable Y)?
(e) Consider the function $g(x) = \log x$. Find the limiting distribution of $n(g(\hat{\theta}_n) - g(\theta))$.

6. (36 points).

Suppose that $X, X_1, \ldots, X_n$ are i.i.d. with distribution function F given by $P(X > x) = 1 - F(x) = 1/(1+x)^3$, $x \geq 0$.

(a) For what values of $r > 0$ is $E|X|^r < \infty$? If they are finite compute $\mu = E(X)$ and $\sigma^2 = Var(X)$.

(b) Compute $F^{-1}(t) = Q(t)$, the quantile function corresponding to F, and $f(x)$, the density function of F at x.

(c) Which of the following are true? (Briefly indicate why or why not.)

(i) $\sum_{i=1}^n X_i = O_p(n^{3/4})$.

(ii) $n^{1/5}(\overline{X}_n - \mu) = o_p(1)$.

(iii) $n^{2/3}(\overline{X}_n - \mu) = O_p(1)$.

(iv) $g(n^{1/3}(\overline{X}_n - \mu)) \rightarrow_p 1$ where $g(x) = \exp(3x)$.

(v) $h(n^{1/2}(\overline{X}_n - \mu)) = O_p(1)$ with $h(x) = 1/\cos(x)$.

(vi) $\sqrt{n}(\mathbb{F}_n^{-1}(3/4) - F^{-1}(3/4)) \rightarrow_d N(0, 3^{-1}(1/4)^{2/3})$.