Statistics 581, Midterm Exam

Wellner; 11/09/2005

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

1. (24 points) **Define** *any three* of the following terms. In each case, provide an appropriate context for your definition.
   (a) Convergence in distribution of a sequence of random variables.
   (b) Convergence almost surely (of a sequence of random variables).
   (c) The inverse or quantile function F^{-1} of a distribution function F.
   (d) A chi-square distribution with m degrees of freedom and non-centrality parameter δ.
   (e) A Brownian bridge process $\mathbb{U}$.

2. (24 points) **State** *any three* of the following results:
   (a) The Lindeberg-Feller CLT.
   (b) The continuous mapping or Mann-Wald theorem.
   (c) The Glivenko-Cantelli theorem.
   (d) The inverse transformation theorem.
   (e) A result about $(Y^{(1)}|Y^{(2)})$ assuming that $Y = (Y^{(1)}, Y^{(2)}) \sim N_m((\mu^{(1)}, \mu^{(2)})^T, \Sigma)$ where
   $$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$
   and Σ_{22} is non-singular.
   (f) The elementary Skorokhod theorem.
   (g) A result connecting the quantile process $\mathbb{V}_n$ to the empirical process $\mathbb{U}_n$.

3. (30 points).
   Suppose that $X, X_1, \ldots, X_n$ are i.i.d. with distribution function F given by $P(X > x) = 1 - F(x) = 1/x^7$, $x \geq 1$, $F(x) = 0$, $x \leq 1$.
   (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.
   (c) Which of the following are true? (Briefly indicate why or why not.)
   (i) $\sum_{i=1}^n X_i = O_p(n^{1/2})$.
   (ii) $n^{1/4}(\overline{X}_n - \mu) = o_p(1)$.
   (iii) $n^{2/3}(\overline{X}_n - \mu) = O_p(1)$.
   (iv) $g(n^{1/4}(\overline{X}_n - \mu)) \rightarrow_p 1/2$ where $g(x) = 1/(1 + e^{-x})$.
   (v) $h(n^{1/2}(\overline{X}_n - \mu)) = O_p(1)$ with $h(x) = \exp(x)$.
   (vi) $\sqrt{n}(\mathbb{F}_n^{-1}(1/2) - F^{-1}(1/2)) \rightarrow_d N(0, (1/4)/[7(1/2)^{8/7}]^2)$.

Do **either** problem 4 **or** problem 5.

4. (30 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

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

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

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

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

5. (30 points) Let $Y_1, Y_2, \ldots$ be i.i.d. exponential(1) random variables and define $S_j = Y_1 + \cdots + Y_j$ for $j \geq 1$. It is well-known (and easy to prove) that

$$(1) \qquad \left(\frac{S_1}{S_{n+1}}, \ldots, \frac{S_n}{S_{n+1}} \right) \stackrel{d}{=} (\xi_{(1)}, \ldots, \xi_{(n)}) \equiv (\xi_{n:1}, \ldots, \xi_{n:n})$$

where $0 \leq \xi_{n:1} \leq \ldots \leq \xi_{n:n} \leq 1$ are the order statistics of n i.i.d. Uniform$(0,1)$ random variables $\xi_1, \ldots, \xi_n$.
(a) Use the representation (1) to prove that for any fixed $k \geq 1$ we have

$$(n\xi_{n:1}, n\xi_{n:2}, \ldots, n\xi_{n:k}) \rightarrow_d (S_1, \ldots, S_k).$$

(b) Find or state the joint density f_n of $(n\xi_{n:1}, n\xi_{n:2})$. [Hint. First find the joint density of $(\xi_{n:1}, \xi_{n:2})$: there are $n(n-1)$ ways of choosing two of the n variables to be the first two order statistics, and the remaining variables must be larger than the second order statistic. Now find the density of $n(\xi_{n:1}, \xi_{n:2})$.]

(c) Show that the density f_n you computed in (b) satisfies $f_n(u,v) \to f(u,v) = \exp(-v)1\{0 \le u \le v < \infty\}$, the joint density of (S_1, S_2).
(d) Use (c) and a result from chapter 2 to conclude that (for $k = 2$) the convergence in (a) can be strengthened to convergence in total variation distance.

Do **either** problem 6 **or** problem 7.

6. (30 points) Suppose that $X_1, \ldots, X_n$ are i.i.d. with distribution function F, and let $\mathbb{F}_n(x) = n^{-1}\sum_{i=1}^n 1_{(-\infty,x]}(X_i)$ be the empirical distribution function of the X_i's. A famous inequality due to Dvoretzky, Kiefer, and Wolfowitz (1956) yields

$$P_F(\sqrt{n}\|\mathbb{F}_n - F\|_\infty > t) \le C\exp(-2t^2) \tag{2}$$

for all F, all n, and all $t > 0$ where, by Massart (1990) $C = 2$ works.
(a) Use the inequality (2) to give a conservative $1-\alpha$ confidence band for F with the dependence on n and α made explicit.
(b) Show that (2) implies that for any $r > 0$ we have

$$\limsup_{n\to\infty} E\|\sqrt{n}(\mathbb{F}_n - F)\|_\infty^r \le C_r$$

for some constant C_r depending only on r.

7. (30 points) Suppose that $X, X_1, \ldots, X_n$ are independent Geometric(p) random variables: $P(X = k) = (1-p)^{k-1}p$ for $k = 1, 2, \ldots$. Thus $E(X) = 1/p$ and $Var(X) = q/p^2$ with $q = 1-p$.
(a) What is the meaning of X in terms of i.i.d. Bernoulli(p) random variables $Y_1, Y_2, \ldots$?
(b) Use the weak law of large numbers to show that the random vector

$$\underline{\overline{V}}_n \equiv \frac{1}{n}\sum_{i=1}^n (X_i, 1_{[X_i=1]}, 1_{[X_i=2]})^T$$

converges in probablity to some vector $(a,b,c)^T \equiv \underline{v}$ where (a,b,c) depends on p. Give (a,b,c) explicitly in terms of p.
(c) Use the multivariate CLT to show that

$$\sqrt{n}(\underline{\overline{V}}_n - \underline{v}) \to_d \underline{W} \sim N_3(0,\Sigma)$$

for some covariance matrix Σ; compute Σ explicitly in terms of p.
(d) Based on the result of (b), suggest two different estimators of p.