Statistics 581, Midterm Exam

Wellner; 11/6/2000

1. (24 points) **Define** *any three* of the following terms. In each case, provide an appropriate context for your definition.
   (a) A probability measure P on a measurable space $(\Omega, \mathcal{A})$.
   (b) The total variation distance between two probability measures P and Q.
   (c) Absolute continuity of a measure ν with respect to a measure μ on a measure space $(\Omega, \mathcal{A})$.
   (d) Almost sure convergence (of a sequence of random variables).
   (e) Convergence in distribution (of a sequence of random variables).
   (f) The definition of a non-central chi-square random variable with n degrees of freedom and noncentrality parameter δ (in terms of central chi-square random variables and a random variable with a Poisson distribution).

2. (24 points) **State** *any three* of the following results, providing the appropriate context for your statement:
   (a) The fundamental event identity for the inverse transformation.
   (b) The dominated convergence theorem.
   (c) A (joint) central limit theorem (asymptotic normality result) for the distribution of k different sample quantile $(\mathbb{F}_n^{-1}(t_1), \ldots, \mathbb{F}_n^{-1}(t_k))$ where $0 < t_1 < \ldots < t_k < 1$.
   (d) The Cramér - Wold device.
   (e) The Lindeberg-Feller central limit theorem.
   (f) The Mann-Wald or continuous mapping theorem.

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

3. (30 points) (a) **State** the Glivenko-Cantelli theorem. Then **prove** that it holds if it holds for the case of i.i.d. Uniform$(0,1)$ random variables.
   (b) **Prove** the Glivenko-Cantelli theorem for i.i.d. Uniform$(0,1)$ random variables: if $\xi_1, \ldots, \xi_n, \ldots$ are i.i.d. Uniform$(0,1)$ with empirical distribution function

$$\mathbb{G}_n(t) = \frac{1}{n}\sum_{i=1}^{n} 1_{[0,t]}(\xi_i)\,, \qquad \text{then} \qquad \sup_{0\le t\le 1} |\mathbb{G}_n(t) - t| \to_{a.s.} 0\,.$$

4. (30 points) Suppose that $X, X_1, \ldots, X_n$ are i.i.d. with distribution function F given by $P(X > x) = 1 - F(x) = 1/x^6$, $x \ge 1$, $F(x) = 0$, $x \le 1$.
   (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) Which of the following statements are true? (Briefly indicate why or why not.)

   (i) $\sum_{i=1}^n X_i = O_p(n^{1/2})$.

   (ii) $n^{1/3}(\overline{X}_n - \mu) = o_p(1)$.

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

(iv) $\tan(\pi\sqrt{n}(\overline{X}_n - \mu)) = O_p(1)$.

(v) $g(n^{1/3}(\overline{X}_n - \mu)) = O_p(1)$ with $g(x) = 1/(1+e^{-x})$.

Do **any two** of problems 5, 6, and 7.

5. (30 points) Let $(\Omega, \mathcal{A}, P)$ be a probability space. Let $\{A_n\}$ be a sequence of events, $A_n \subset \Omega$, $A_n \in \mathcal{A}$ for $n = 1, 2, \ldots$, and let $X_n = 1_{A_n}$ be the indicator functions of the events A_n.
(a) Define the sets $\limsup A_n = [A_n \ i.o.]$ and $\liminf A_n = [A_n \ a.a.]$ in terms of the collection $\{A_n\}$.
(b) What does Fatou's lemma say about $E(\liminf X_n)$? Translate this into a statement relating $P(\liminf A_n)$ to $P(A_n)$.
(c) Based on your answer to (b), is the inequality $P(\limsup A_n^c) \le \limsup P(A_n^c)$ true or false? If it is true, explain why. If false, give a counterexample.

6. (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 $K : \underline{p} \ne \underline{p}_0$. 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 $\underline{p} \ne \underline{p}_0$, 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 $\underline{p} \ne \underline{p}_0$, show that

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

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

7. (30 points) Suppose that $X_1, \ldots, X_m$ are i.i.d. F and $Y_1, \ldots, Y_n$ are i.i.d. G and independent of the X's. Consider testing $H : F = G$ (with $F = G$ continuous) versus $K : F \ne G$. Let $\mathbb{F}_m$ denote the empirical distribution function of the X_i's, and let $\mathbb{G}_n$ denote the empirical distribution function of the Y_j's, and set $N \equiv m+n$. The *two-sample Kolmogorov statistic* for testing H versus K is

$$D_{m,n} \equiv \sqrt{\frac{mn}{N}} \|\mathbb{F}_m - \mathbb{G}_n\|_\infty = \sqrt{\frac{mn}{N}} \sup_x |\mathbb{F}_m(x) - \mathbb{G}_n(x)| .$$

(a) By introducing two independent samples $\xi_1, \ldots, \xi_m$ and $\zeta_1, \ldots, \zeta_n$ of Uniform$(0,1)$ random variables, with corresponding empirical distribution functions

$$\mathbb{K}_m(u) \equiv m^{-1} \sum_{i=1}^{m} 1_{[0,u]}(\xi_i), \qquad \text{and} \qquad \mathbb{L}_n(u) \equiv n^{-1} \sum_{i=1}^{n} 1_{[0,u]}(\zeta_i),$$

and uniform empirical processes $\mathbb{U}_m$ and $\mathbb{V}_n$ defined by

$$\mathbb{U}_m(t) = \sqrt{m}(\mathbb{K}_m(t) - t), \qquad \text{and} \qquad \mathbb{V}_n(t) = \sqrt{n}(\mathbb{L}_n(t) - t),$$

show that under the null hypothesis H

$$D_{m,n} =_d \|\sqrt{1-\lambda_N}\mathbb{U}_m - \sqrt{\lambda_N}\mathbb{V}_n\|_\infty = \sup_{0 \le t \le 1} |\sqrt{1-\lambda_N}\mathbb{U}_m(t) - \sqrt{\lambda_N}\mathbb{V}_n(t)|.$$

where $\lambda_N \equiv m/N$, $\overline{\lambda}_N = 1 - \lambda_N = n/N$.

(b) Use the result of (a) to show that when $m, n \to \infty$ with $\lambda_N \to \lambda \in [0,1]$,

$$D_{m,n} \to_d \|\sqrt{1-\lambda}\mathbb{U} - \sqrt{\lambda}\mathbb{V}\|_\infty =_d \sup_{0 \le t \le 1} |\mathbb{U}(t)|$$

where $\mathbb{U}$ and $\mathbb{V}$ denote two independent Brownian bridge processes.

(c) When the alternative hypothesis holds, so that $F \neq G$, show that $D_{m,n}/\sqrt{mn/N} \to_{a.s.}$ "something", and find "something".

(d) Based on our discussions in class and homework, can you suggest any other test statistic for testing H versus K?