Statistics 581, Midterm Exam Solution

Wellner; 11/8/2000

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).

   Solution: See Chapters 0,1, 2.

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.

   Solution: See Chapters 0,1, 2.

   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

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

   Solution: See Chapter 2, section x.

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 $g(x) = 1/(1+e^{-x})$.

Solution: (a) Now since F is the distribution function of a non-negative random variable, for $r \geq 1$,

$$\begin{aligned} E|X|^r &= E(X^r) = \int_0^\infty r x^{r-1}(1-F(x))\,dx \\ &= \int_0^1 r t^{r-1} dt + \int_1^\infty r x^{r-1} x^{-6}\,dx = 1 + \int_1^\infty r x^{r-7}\,dx \\ &= 1 + \frac{r}{r-6} x^{r-6} \Big|_{x=1}^{\infty} \\ &= \begin{cases} \frac{6}{6-r} & \text{if } r < 6 \\ \infty & \text{if } r \geq 6\,. \end{cases} \end{aligned}$$

Thus we have $E(X) = 6/5$, and $E(X^2) = 6/4 = 3/2$, so $Var(X) = E(X^2) - (E(X))^2 = 3/2 - (6/5)^2 = 3/50$.
(b) (i) False. By the WLLN, $\sum_{i=1}^{n} X_i = O_p(n)$.
(ii) True. $n^{1/3}(\overline{X}_n - \mu) = n^{-1/6} n^{1/2}(\overline{X}_n - \mu) = o(1)O_p(1) = o_p(1)$.
(iii) False. $n^{2/3}(\overline{X}_n - \mu) = n^{1/6}\sqrt{n}(\overline{X}_n - \mu) = n^{1/6}O_p(1) = O_p(n^{1/6})$.
(iv) True. $Y_n = \sqrt{n}(\overline{X}_n - \mu) \to_d Y \sim N(0,\sigma^2)$ and $g(x) = \tan(\pi x)$ is continuous except at $x = \pm k/2$, k odd, and this (countable) set has P_Y-measure 0. Thus by the continuous mapping theorem $g(Y_n) \to_d g(Y)$, and this implies $g(Y_n) = O_p(1)$.
(v) True. This follows easily since g is uniformly bounded between 0 and 1. In fact, since $n^{1/3}(\overline{X}_n - \mu) \to_p 0$ from (ii), $g(n^{1/3}(\overline{X}_n - \mu)) \to_p g(0) = 1/2$ by the continuous mapping theorem.

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 $P(\limsup A_n^c) \leq \limsup P(A_n^c)$ true or false? If it is true, explain why. If false, give a counterexample.

Solution: (a)

$$\limsup A_n = [A_n\ i.o.] = \cap_{n=1}^{\infty} \cup_{k=n'}^{\infty} A_k\,, \qquad \liminf A_n = [A_n\ a.a.] = \cup_{n=1}^{\infty} \cap_{k=n'}^{\infty} A_k\,.$$

(b) Fatou's lemma says that

$$E(\liminf X_n) \leq \liminf E(X_n)\,.$$

But $E(X_n) = E(1_{A_n}) = P(A_n)$ while $E(\liminf X_n) = E(\liminf 1_{A_n}) = E(1_{\liminf A_n}) = P(\liminf A_n)$. Hence Fatou's inequality implies that

$$P(\liminf A_n) \le \liminf P(A_n)\,.$$

(c) False. By taking complements in the conclusion of (b),

$$P(\limsup A_n^c) \ge \limsup P(A_n^c)\,.$$

A counterexample is provided by the "dancing functions" as follows: For integers $m \ge 1$ and $j = 1, \ldots, 2^m$, set $Y_{mj} = 1_{((j-1)/2^m, j/2^m]}(U)$ where $U \sim \text{Uniform}(0,1)$, and define $Z_1 = Y_{1,1}$, $Z_2 = Y_{1,2}$, $Z_3 = Y_{2,1}$, $Z_4 = Y_{2,2}$, $Z_5 = Y_{2,3}$, $Z_6 = Y_{2,4}$, ... Then, with $A_n = [Z_n \le 1/2]$, $A_n^c = [Z_n > 1/2]$, and we have $P(A_n^c) \to 0$ as $n \to \infty$, but $P(A_n^c\ i.o.) = P(\limsup A_n) = 1$.

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?

Solution: (a) We compute

$$\begin{aligned} 2\log\lambda_n &= 2\log\prod_{j=1}^k \left\{\frac{\widehat{p}_j}{p_{0j}}\right\}^{N_j} \\ &= 2\sum_{j=1}^k \log\left\{\left(\frac{\widehat{p}_j}{p_{0j}}\right)^{N_j}\right\} \\ &= 2\sum_{j=1}^k N_j \log\left(\frac{\widehat{p}_j}{p_{0j}}\right) \\ &= 2n\sum_{j=1}^k \widehat{p}_j \log\left(\frac{\widehat{p}_j}{p_{0j}}\right) \end{aligned}$$

where the last line follows from $\widehat{p}_j = N_j/n$.
(b) From (a) we deduce that

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

by the continuous mapping theorem since

$$g(x) = 2 \sum_{j=1}^{k} x_j \log\left(\frac{x_j}{p_{0j}}\right)$$

is continuous and $\underline{\widehat{p}} \rightarrow_p \underline{p}$.
(c) From (b) we have

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

where, with $\Sigma = \mathrm{diag}(\underline{p}) - \underline{p}\underline{p}^T$,

$$\sqrt{n}(\underline{\widehat{p}} - \underline{p}) \rightarrow_d \underline{Z} \sim N_k(0, \Sigma)\,,$$

and g is differentiable at $\underline{p}$ with gradient (vector of partial derivatives) $\nabla g(\underline{p}) \equiv \underline{d}$ given by

$$d_j = \frac{\partial}{\partial x_j} g(\underline{x}) = 2\left\{\log\left(\frac{x_j}{p_{0j}}\right) + \frac{x_j}{x_j}\right\}\Big|_{x_j = p_j} = 2\left\{1 + \log\left(\frac{p_j}{p_{0j}}\right)\right\}.$$

Hence by the delta method,

$$\sqrt{n}(g(\underline{\widehat{p}}) - g(\underline{p})) \rightarrow_d \nabla g(\underline{p})^T \underline{Z} \sim N(0, \underline{d}^T \Sigma \underline{d})\,.$$

Since both $\underline{1}'\underline{\widehat{p}} = 1$ and $\underline{1}'\underline{p} = 1$, we have $\underline{1}'\underline{Z} = 0$, and hence we can add or subtract $c\underline{1}$ from $\nabla g(\underline{p})$ (for any constant c) without changing the distribution:

$$\nabla g(\underline{p})^T \underline{Z} = (\nabla g(\underline{p})' - c\underline{1}')\underline{Z} = (\underline{d}' - c\underline{1}')\underline{Z}\,.$$

A nice choice of c is $k^{-1}\sum_{j=1}^{k} d_j$. Then $\underline{d} - \overline{d}\underline{1} \equiv \underline{c}$ where

$$\begin{aligned} c_j/2 &= \log\left(\frac{p_j}{p_{0j}}\right) - \frac{1}{k}\sum_{j'=1}^{k} \log\left(\frac{p_{j'}}{p_{0j'}}\right) \\ &= \log\left(\frac{p_j}{p_{0j}}\right) - \log\left\{\prod_{j'=1}^{k} \frac{p_{j'}}{p_{0j'}}\right\}^{1/k} \\ &= \log\left(\frac{p_j/p_{0j}}{\left\{\prod_{j'=1}^{k}(p_{j'}/p_{0j'})\right\}^{1/k}}\right). \end{aligned}$$

Note that at least one component c_j of $\underline{c}$ is non-zero under the alternative K; if all the c_j's are zero, then

$$\frac{p_j}{p_{0j}} = \left\{\prod_{j'=1}^{k}(p_{j'}/p_{0j'})\right\}^{1/k} \qquad \text{for } j = 1, \ldots, k\,,$$

which implies $p_j = p_{0j}$ for all j. Finally, we can use this limiting distribution to approximate the power of the likelihood ratio test of H versus K. Since $2\log\lambda_n \to_d \chi^2_{k-1}$ under H, the power of the test "reject H if $2\log\lambda_n > \chi^2_{k-1,\alpha}$" is given by

$$\begin{aligned} P_p(2\log\lambda_n > \chi^2_{k-1,\alpha}) &= P_p\left(\sqrt{n}(2n^{-1}\log\lambda_n - g(\underline{p})) > \sqrt{n}(n^{-1}\chi^2_{k-1,\alpha} - g(\underline{p})\right) \\ &\doteq P(N(0, c^T\Sigma c) > \sqrt{n}(n^{-1}\chi^2_{k-1,\alpha} - g(\underline{p}))) . \end{aligned}$$

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 $K : F \neq G$. 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?

Solution: (a) & (b): By the inverse transformation,

$$\mathbb{F}_m =_d \mathbb{K}_m(F), \qquad \text{and} \qquad \mathbb{G}_n = \mathbb{L}_n(G) ,$$

and, furthermore,

$$\sqrt{m}(\mathbb{F}_m - F) =_d \mathbb{U}_m(F), \qquad \text{and} \qquad \sqrt{n}(\mathbb{G}_n - G) =_d \mathbb{V}_n(G) .$$

Since $F = G$ under the null hypothesis H, and F is continuous, we have

$$\begin{aligned}
D_{m,n} &= \sqrt{\frac{mn}{N}} \|\mathbb{F}_m - \mathbb{G}_n\|_\infty \\
&=_d \sqrt{\frac{mn}{N}} \|(\mathbb{K}_m(F) - F) - (\mathbb{L}_n(F) - F)\|_\infty \\
&= \sup_{0 \le t \le 1} |\sqrt{1 - n/N}\sqrt{m}(\mathbb{K}_m(t) - t) - \sqrt{m/N}\sqrt{n}(\mathbb{L}_n(t) - t)| \\
&= \sup_{0 \le t \le 1} |\sqrt{1 - \lambda_N}\mathbb{U}_m(t) - \sqrt{\lambda_N}\mathbb{V}_n(t)| \\
&\to_d \sup_{0 \le t \le 1} |\sqrt{1 - \lambda}\mathbb{U}(t) - \sqrt{\lambda}\mathbb{V}(t)| \,.
\end{aligned}$$

But the process $\mathbb{Z}(t) \equiv \sqrt{1 - \lambda}\mathbb{U}(t) - \sqrt{\lambda}\mathbb{V}(t)$ is a mean zero Gaussian process with

$$E[\mathbb{Z}(s)\mathbb{Z}(t)] = (1 - \lambda)(s \wedge t - st) + \lambda(s \wedge t - st) = s \wedge t - st \,.$$

Thus $\mathbb{Z}$ is a Brownian bridge process on $[0, 1]$, and $\mathbb{Z} =_d \mathbb{U}$. Hence we conclude that

$$D_{m,n} \to_d \sup_{0 \le t \le 1} |\mathbb{U}(t)| \,.$$

(c) By the Glivenko-Cantelli theorem, under K we have

$$\frac{D_{m,n}}{\sqrt{mn/N}} = \|\mathbb{F}_m - \mathbb{G}_n\|_\infty \to_{a.s.} \|F - G\|_\infty > 0 \,.$$

(d) Another natural test statistic is the two-sample Cramér - von Mises statistic

$$C^2_{m,n} = \int_{-\infty}^{\infty} \left\{ \sqrt{\frac{mn}{N}} \left(\mathbb{F}_m(x) - \mathbb{G}_n(x)\right) \right\}^2 d\mathbb{H}_N(x)$$

where $\mathbb{H}_N = (m/N)\mathbb{F}_n + (n/N)\mathbb{G}_n$ is the *pooled empirical distribution function.*