Statistics 581, Midterm Exam Solutions

Wellner; 11/10/2005

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 $\mathbb{U}$.

Solution: See chapter 1 and 2 notes.

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 $\mathbb{U}_n$.

Solution: See chapter 1 and 2 notes.

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) $\sqrt{n}(\mathbb{F}_n^{-1}(1/2) - F^{-1}(1/2)) \rightarrow_d N(0, (1/4)/[7(1/2)^{8/7}]^2)$.

Solution: (a) $E|X|^r = EX^r = \int_1^\infty x^r 7x^{-8} dx = 7\int_1^\infty x^{r-8} dx = 7/(7-r) < \infty$ if $r < 7$. If $r \geq 7$, then $EX^r = \infty$. Thus taking $r = 1$ yields $EX = 7/6$ and taking $r = 2$ yields $EX^2 = 7/5$. Hence $Var(X) = E(X^2) - (EX)^2 = (7/5) - (7/6)^2 = 7/(5 \cdot 36) = 7/180$.

(b) Now $F(x) = 1 - x^{-7}$ for $x \ge 1$, so solving $t = F(x) = 1 - x^{-7}$ for x gives $x = F^{-1}(t) = (1-t)^{-1/7}$.
(c) (i) False; since $n^{-1}\sum_1^n X_i = \overline{X}_n \to_p E(X) = 7/6$, $\sqrt{n}\overline{X} = n^{-1/2}\sum_1^n X_i \to_p \infty$.
(ii) True; since $\sqrt{n}(\overline{X}_n - \mu) \to_d N(0,\sigma^2)$, it follows that $n^{1/4}(\overline{X}_n - \mu) = n^{-1/4}\sqrt{n}(\overline{X}_n - \mu) \to_d 0 \cdot N(0,\sigma^2) = 0$ by Slutsky's theorem, and hence also this holds in probability.
(iii) False; since $n^{2/3}(\overline{X}_n - \mu) = n^{1/6}\sqrt{n}(\overline{X}_n - \mu) = n^{1/6}Z_n$ where $Z_n \to_d Z \sim N(0,\sigma^2)$, this is not $O_p(1)$.
(iv) True; since $n^{1/4}(\overline{X}_n - \mu) \to_p 0$ by (ii), the continuous mapping theorem yields $g(n^{1/4}(\overline{X}_n - \mu)) \to_p g(0) = 1/2$.
(v) True; since $Z_n \equiv \sqrt{n}(\overline{X}_n - \mu) \to_d Z \sim N(0,\sigma^2)$ and h is continuous, $h(Z_n) \to_d h(Z)$, and this implies that $h(Z_n) = O_p(1)$.
(vi) True; $F^{-1}(1/2) = (1/2)^{-1/7}$ and $f(x) = 7x^{-8}$ for $x \ge 1$. Thus $f(F^{-1}(1/2)) = 7(1/2)^{8/7}$.

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 $K : \underline{p} \ne \underline{p}_0$. 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 have

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

using $\widehat{p}_j = n^{-1} N_j$ in the last line.
(b) It follows immediately from (a) that $2n^{-1} \log \lambda_n = g(\widehat{p})$ where $g : [0,1]^k \to \mathbb{R}$ is given by

$$g(u) = 2 \sum_{j=1}^{k} u_j \log(u_j / p_{0,j}).$$

(c) By the continuous mapping theorem applied to $g(\widehat{p}_n)$ where $\widehat{p}_n \to_p p$, it follows immediately that $2n^{-1} \log \lambda_n \to_p g(p) = 2K(P, P_0)$ where $K(P, P_0)$ is the Kullback-Leibler divergence between P and P_0 determined by p and p_0 respectively.
(d) By (b) we can write

$$\begin{aligned} \sqrt{n}(2n^{-1} \log \lambda_n - g(p)) &= \sqrt{n}(g(\widehat{p}_n) - g(p)) \\ &\to_d g'(p) Z \sim N(0, g'(p) \Sigma g'(p)^T) \end{aligned}$$

by the delta-method or $g'-$ theorem where the components of the vector of derivatives $g'(p)$ are given by

$$\frac{\partial}{\partial u_j} g(u) \Big|_{u=p} = 2(\log(p_j / p_{0,j}) + 1) \equiv d_j$$

for $j = 1, \ldots, k$, and $\Sigma = \text{diag}(p) - pp^T$. Thus $V^2 = d \Sigma d^T = \sum_{j=1}^{k} (d_j^2 p_j) - (d^T p)^2 = Var_p(D)$ where $P(D = d_j) = p_j$, $j = 1, \ldots, k$. This can be used to approximate the power of the likelihood ratio test:

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

A non-central chi-square approximation of the power based on local alternatives is also possible.

Beyond the exam question: A question not addressed by the exam problem concerns the asymptotic distribution of $2 \log \lambda_n$ under the null hypothesis. Here we briefly sketch a proof of $2 \log \lambda_n \to_d \chi^2_{k-1}$. Note that by Taylor expansion we can write

$$g(u) = g(p_0) + g'(p_0)^T (u - p_0) + \frac{1}{2} (u - p_0)^T g''(u^*) (u - p_0)$$

where $|p^* - p_0| \le |u - p_0|$; here g'' is a $k \times k$ matrix. Now $g(p_0) = 0$, and from the calculation in (d) above, $g'(p_0) = 2\underline{1} = 2(1, \ldots, 1)^T$, and $g''(u) = \text{diag}(2/u)$. Thus we find, using $\underline{1}^T \widehat{p}_n = 1 = \underline{1}^T p_0$,

$$\begin{aligned} g(\widehat{p}_n) &= 2\underline{1}^T (\widehat{p}_n - p_0) + (\widehat{p}_n - p_0)^T \text{diag}(1/p_n^*)(\widehat{p}_n - p_0) \\ &= (\widehat{p}_n - p_0)^T \text{diag}(1/p_n^*)(\widehat{p}_n - p_0) \end{aligned}$$

and hence

$$2 \log \lambda_n = n g(\widehat{p}_n) = n(\widehat{p}_n - p_0)^T \text{diag}(1/p_n^*)(\widehat{p}_n - p_0)$$

where p_n^* satisfies $|p_n^* - p_0| \le |\widehat{p}_n - p_0| \to 0$. Thus $2 \log \lambda_n = Q_n + o_p(1) \to_d \chi^2_{k-1}$ under the null hypothesis $p = p_0$.

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

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

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 $f_n(u,v) \rightarrow f(u,v) = \exp(-v)1\{0 \leq u \leq v < \infty\}$, 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.

Solution: (a) By the strong (weak) law of large numbers it follows that

$$n^{-1}S_{n+1} = \frac{n+1}{n}(n+1)^{-1}S_{n+1} \rightarrow_{a.s.(p)} 1.$$

From the representation given for the joint distribution of the uniform order statistics in terms of partial sums of exponential random variables,

$$\begin{aligned}(n\xi_{n:1}, \ldots, n\xi_{n:k}) &\stackrel{d}{=} \left(\frac{S_1}{n^{-1}S_{n+1}}, \ldots, \frac{S_k}{n^{-1}S_{n+1}}\right) \\ &\rightarrow_{a.s.(p)} \left(\frac{S_1}{1}, \ldots, \frac{S_k}{1}\right) \\ &= (S_1, \ldots, S_k).\end{aligned}$$

Since almost sure convergence and convergence in probability imply convergence in distribution, we conclude that

$$(n\xi_{n:1}, \ldots, n\xi_{n:k}) \rightarrow_d (S_1, S_2, \ldots, S_k).$$

(b) The joint density of $(\xi_{n:1}, \xi_{n:2})$ is given by

$$g(u,v) = n(n-1)(1-v)^{n-2}1\{0 < u < v < 1\}.$$

Hence the joint density of $(n\xi_{n:1}, n\xi_{n:2})$ is

$$f_n(u,v) = \frac{n(n-1)}{n^2}(1 - \frac{v}{n})^{n-2}1\{0 < u < v < n\}.$$

(c) It follows immediately from the form of f_n derived in (b) that the joint density f_n of $(n\xi_{n:1}, n\xi_{n:2})$ satisfies

$$f_n(u,v) \rightarrow e^{-v}1\{0 < u < v < \infty\} \equiv f(u,v)$$

where $f(u,v)$ is the joint density of (S_1, S_2).
(d) By Scheffé's theorem the pointwise convergence of densities proved in (c) implies that if P_n denotes the distribution of $(n\xi_{n:1}, n\xi_{n:2})$ and P denotes the distribution of (S_1, S_2), then

$$d_{TV}(P_n, P) = \frac{1}{2}\int |f_n - f| d\lambda_2 = \int\int |f_n(u,v) - f(u,v)| du dv \to 0. \qquad (2)$$

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

$$P_F(\sqrt{n}\|\mathbb{F}_n - F\|_\infty > t) \le C\exp(-2t^2) \qquad (3)$$

for all F, all n, and all $t > 0$ where, by Massart (1990) $C = 2$ works.
(a) Use the inequality (3) to give a conservative $1-\alpha$ confidence band for F with the dependence on n and α made explicit.
(b) Show that (3) 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.

Solution: (a) Now $2e^{-2t^2} = \alpha$ if

$$t = \sqrt{\frac{1}{2}\log\left(\frac{2}{\alpha}\right)}.$$

Thus the DKW inequality yields

$$P_F\left(\|\sqrt{n}(\mathbb{F}_n - F)\|_\infty > \sqrt{\frac{1}{2}\log\left(\frac{2}{\alpha}\right)}\right) \le \alpha.$$

This implies that

$$\begin{aligned} 1-\alpha &\le P_F\left(\|\sqrt{n}(\mathbb{F}_n - F)\|_\infty \le \sqrt{\frac{1}{2}\log\left(\frac{2}{\alpha}\right)}\right) \\ &= P_F\left(\mathbb{F}_n(x) - \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)} \le F(x) \le \mathbb{F}_n(x) + \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)} \text{ for all } x\right). \end{aligned}$$

Thus $\mathbb{F}_n(x) \pm \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}$ gives a conservative $1-\alpha$ confidence band for F.
(b) If Y is a non-negative random variable, then $EY^r = \int_0^\infty r y^{r-1} P(Y \ge y) dy$. Using this formula together with the DKW inequality yields

$$E\|\sqrt{n}(\mathbb{F}_n - F)\|_\infty^r = \int_0^\infty r t^{r-1} P(\|\sqrt{n}(\mathbb{F}_n - F)\|_\infty \ge t) dt$$

$$
\begin{aligned}
&\le \int_0^\infty r t^{r-1} C \exp(-2t^2)dt = \frac{Cr}{4}\int_0^\infty t^{r-2}\exp(-2t^2)4tdt \\
&= \frac{Cr}{4}\int_0^\infty \int_0^\infty (y/2)^{(r-2)/2}\exp(-y)dy \\
&= \frac{Cr}{4} 2^{-(r-2)/2}\Gamma(r) = 2^{-r/2}\Gamma(r+1)
\end{aligned}
$$

by using $C = 2$.

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.

Solution: (a) X is the waiting time (or # of trials) until the first success (or 1) in Bernoulli(p) trials $Y_1, Y_2, \ldots$.
(b) By the weak law of large numbers applied three times,

$$\underline{\overline{V}}_n \to_p E\underline{V} = (E(X), P(X=1), P(X=2))^T = (1/p, p, qp)^T \equiv (a,b,c)^T.$$

(c) The covariance matrix of the random vector $V \equiv (X, 1\{X=1\}, 1\{X=2\})$ is given by

$$
\begin{aligned}
\Sigma &= \begin{pmatrix} Var(X) & E(X1\{X=1\}) - (1/p)p & E(X1\{X=2\}) - (1/p)qp \\ E(X1\{X=1\} - (1/p)p & Var(1\{X=1\}) & E(1\{X=1\}1\{X=2\}) - pqp \\ E(X1\{X=2\}) - (1/p)qp & E(1\{X=1\}1\{X=2\}) - pqp & Var(1\{X=2\}) \end{pmatrix} \\
&= \begin{pmatrix} q/p^2 & p-1 & 2qp-q \\ p-1 & p(1-p) & -qp^2 \\ 2qp-q & -qp^2 & qp(1-qp) \end{pmatrix}
\end{aligned}
$$

Thus the central limit theorem yields

$$\sqrt{n}(\underline{\overline{V}}_n - E(V)) \to_d N_3(0, \Sigma).$$

(d) Since $\overline{X}_n \to_p 1/p$, $\widehat{p}_1 \equiv 1/\overline{X}_n \to_p p$. Thus $\widehat{p}_1 = 1/\overline{X}_n$ is one possible estimator. Another possible estimator is $\widehat{p}_2 = n^{-1}\sum_1^n 1\{X_i = 1\} \to_p p$. Since

$$\widehat{q} \equiv n^{-1}n^{-1}\sum_1^n 1\{X_i = 2\}/n^{-1}\sum_1^n 1\{X_i=1\} \to_p qq/p = q = 1-p,$$

yet another estimator is $\widehat{p}_3 = 1 - \widehat{q}$.