Statistics 581, Midterm Exam Solutions

Wellner; 11/06/2009

1. (24 points) **Define** any **three** of the following six terms.
(a) Convergence in probability of a sequence of random variables to a limit random variable X.
(b) Almost sure convergence of a sequence of random variables to a limit random variable X.
(c) The inverse or quantile function F^{-1} of a distribution function F.
(d) The event $\{A_n \text{ infinitely often}\} = \{A_n \text{ i.o.}\}$ for a sequence of events $\{A_n\}_{n\ge 1}$.
(e) The total variation distance between two probability measures P and Q.
(f) A normal random vector $Y = (Y_1, \ldots, Y_n)$.

Solution: See Chapters 1 and 2 of the course notes and Chapter 1 (Basic Probability Theory) of Ferguson, *A Course in Large Sample Theory*.

2. (36 points). **State** any **three** of the following:
(a) The Cramér - Wold theorem (or device).
(b) The Helly-Bray theorem.
(c) Vitale's theorem.
(d) Liapunov's CLT for a row-independent triangular array.
(e) The Glivenko-Cantelli theorem.

Solution: See Chapters 1 and 2 of the course notes and Chapter 1 (Basic Probability Theory) of Ferguson, *A Course in Large Sample Theory*.

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

3. (30 points).
A. Suppose that $X \sim N_n(\mu, I)$ where $\mu = (\mu_1, \ldots, \mu_n)' \in R^n$ and I is the $n \times n$ identity matrix. Describe the distribution of $Y \equiv X'X = |X|^2$ in terms of ordinary chi-square distributions and a Poisson random variable K, and give the distribution's name.
B. Use the description in A to compute the mean and variance of Y.
C. What is the role of the distribution of Y in a statistical problem we have discussed in class?

Solution: A. If $X \sim N_n(\mu, I)$, then $Y \equiv |X|^2 \sim \chi_n^2(\delta)$ where $\delta = \mu'\mu$; moreover $(Y|K = k) \sim \chi^2_{2k+n}$ where $K \sim \text{Poisson}(\delta/2)$.
B. By the conditional characterization of the distribution of Y we have

$$E(Y) = E(E(Y|K)) = E(2K + n) = 2(\delta/2) + n = n + \delta\,,$$

and

$$\begin{aligned} Var(Y) &= E\{Var(Y|K)\} + Var[E(Y|K)] = E\{2(2K+n)\} + Var[2K+n] \\ &= 2(2(\delta/2)+n)) + 4\delta/2 = 2n + 4\delta \end{aligned}$$

just as in problem 1 of problem set # 4.
C. In testing $H_0 : p = p_0$ versus $K_0 : p \neq p_0$ the limiting distribution of Q_n under local alternatives of the form $p_n = p_0 + cn^{-1/2}$ is $\chi^2_{k-1}(\delta)$ with $\delta = \sum_1^k c_j^2/p_{j0}$.

4. (30 points) Let X be a random variable with $E(X^2) < \infty$. Show that $Var(|X|) \leq Var(X)$. When does equality occur? (Hint for the second part: use $X = X^+ - X^-$ and $|X| = X^+ + X^-$ where $X^+ \geq 0$, $X^- \geq 0$.)

   Solution: Now

$$Var(|X|) = E(|X|^2) - \{E(|X|)\}^2 = E(X^2) - \{E(|X|)\}^2$$

   while

$$Var(X) = E(X^2) - \{E(X)\}^2,$$

   so

$$\begin{aligned} Var(X) - Var(|X|) &= E(X^2) - \{E(X)\}^2 - (E(X^2) - \{E(|X|)\}^2) \\ &= \{E(|X|)\}^2 - \{E(X)\}^2 \\ &\geq 0 \end{aligned}$$

   since, for the convex function $g(x) \equiv |x|$,

$$|E(X)| = g(E(X)) \leq Eg(X) = E(|X|)$$

   by Jensen's inequality. Equality occurs if $P(X = 0) = 1$, or if $P(X \geq 0) = 1$, or if $P(X \leq 0) = 1$. On the other hand if equality occurs, then $E(|X|) - |E(X)| = 0$, or $EX^+ + EX^- - |E(X^+) - EX^-| = 0$. Suppose first that $E(X^+) > E(X^-)$; then the last equality becomes

$$0 = E(X^+) + E(X^-) - (E(X^+) - E(X^-)) = 2EX^-.$$

   But this implies that $X^- = 0$ almost surely. Similarly, if $E(X^+) < E(X^-)$ then $0 = 2E(X^+)$ implies $X^+ = 0$ almost surely.

5. (30 points) A sequence of random variables X_n is "bounded in probability", which we express in symbols as $X_n = O_p(1)$, if for every $\epsilon > 0$ there exist M and n_0 such that $P(|X_n| > M) < \epsilon$ for all $n > n_0$; i.e. if

$$\lim_{M\to\infty} \limsup_{n\to\infty} P(|X_n| > M) = 0\,.$$

We write $X_n = O_p(b_n)$ for a sequence of positive real numbers b_n if $X_n/b_n = O_p(1)$.
(a) Show that if $X_n \to_d X$, then $X_n = O_p(1)$.
Now suppose that $X_1, X_2, X_3, \ldots$ are i.i.d. with mean $\mu \neq 0$ and variance $\sigma^2 < \infty$ (so $E(X^2) < \infty$. Let $S_n = X_1 + \ldots + X_n$ and $\overline{X}_n = S_n/n$.
(b) Is it true that:

(i) $S_n = O_p(1)$? (ii) $S_n = O_p(n^{1/2})$? (iii) $\overline{X}_n = O_p(n^{-1/2})$?
(iv) $n^{1/2}(\overline{X}_n - \mu) = O_p(1)$? (v) $n^{1/4}(\overline{X}_n - \mu) \to_p 0$?
(v) $\cos(S_n) = O_p(1)$?

Solution: (a) Let F be the distribution function of X and let $\epsilon > 0$. Fix M large with $\pm M \in C_F = \{x \in R : F \text{ is continuous at } x\}$ and both $F(-M) < \epsilon/2$ and $1 - F(M) < \epsilon/2$. This is possible since the discontinuity points of F are at most countable and $F(-M) \to 0$, $1 - F(M) \to 0$ as $M \to \infty$. Then

$$\begin{aligned} P(|X_n| > M) &= P([X_n > M] \cup [X_n < -M]) \\ &\leq P(X_n > M) + P(X_n \leq -M) \\ &= 1 - F_n(M) + F_n(-M) \\ &\to 1 - F(M) + F(-M) \\ &< \epsilon/2 + \epsilon/2 = \epsilon\,. \end{aligned}$$

Thus for this choice of M we have

$$\limsup_{n\to\infty} P(|X_n| > M) \leq \epsilon\,.$$

Hence $X_n = O_p(1)$.
(b) (i) No; $S_n = O_p(n)$ in general (if $\mu \neq 0$), and $S_n = O_p(n^{1/2})$ if $\mu = 0$; $S_n = O_p(1)$ fails in either case.
(b) (ii) As noted in (i), this is true if $\mu = 0$, since then it follows that $n^{-1/2}S_n \to_d N(0, \sigma^2)$ so that $n^{-1/2}S_n = O_p(1)$ by the result of (a). If $\mu \neq 0$ then $S_n = O_p(n)$, but it is not $O_p(n^{1/2})$.
(b) (iii) This has the same answer as (b)(ii) since $\sqrt{n}\overline{X}_n = S_n/\sqrt{n}$.
(b) (iv) Since $n^{1/2}(\overline{X}_n - \mu) \to_d N(0, \sigma^2)$, it follows from (a) that $n^{1/2}(\overline{X}_n - \mu) = O_p(1)$.
(b) (v) Yes. Since $n^{1/2}(\overline{X}_n - \mu) = O_p(1)$ by (iv), it follows that

$$n^{1/4}(\overline{X}_n - \mu) = n^{-1/4}n^{1/2}(\overline{X}_n - \mu) = o(1)O_p(1) = o_p(1).$$

(b) (v) Yes. Since $|\cos(x)| \leq 1$ taking $M = 1$ we have $P(|\cos(S_n)| > 1) = 0$ for every $n \geq 1$, and hence $\cos(S_n)$ is trivially $O_p(1)$.

6. (30 points) Suppose that $X, X_1, \ldots, X_n$ are i.i.d. with distribution function F given by $P(X > x) = 1 - F(x) = 1/x^5$, $x \geq 1$, $F(x) = 0$, $x \leq 1$.
(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 $F^{-1}(t)$ for $0 < t < 1$.
(c) 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) $g(n^{1/3}(\overline{X}_n - \mu)) = O_p(1)$ where $g(x) \equiv \Phi(x) = \int_{-\infty}^x \phi(z)dz$ is the standard normal distribution function.

Solution: (a) $E|X|^r = EX^r = \int_1^\infty x^r 5x^{-6}dx = 5\int_1^\infty x^{r-6}dx = 5/(5-r) < \infty$ if $r < 5$. If $r \geq 5$, then $EX^r = \infty$. Thus taking $r = 1$ yields $EX = 5/4$ and taking $r = 2$ yields $EX^2 = 5/3$. Hence $Var(X) = E(X^2) - (EX)^2 = (5/3) - (5/4)^2 = 5/(3 \cdot 16) = 5/48$.
(b) Now $F(x) = 1 - x^{-5}$ for $x \geq 1$, so solving $t = F(x) = 1 - x^{-5}$ for x gives $x = F^{-1}(t) = (1-t)^{-1/5}$.
(c) (i) False; since $n^{-1}\sum_1^n X_i = \overline{X}_n \to_p E(X) = 5/4$, $\sqrt{n}\overline{X} = n^{-1/2}\sum_1^n X_i \to_p \infty$.
(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 $Z_n \equiv n^{1/2}(\overline{X}_n - \mu) \to_d Z \sim N(0, \sigma^2)$ by (ii), the continuous mapping theorem yields $\tan(\pi n^{1/2}(\overline{X}_n - \mu)) \to_d \tan(\pi Z)$.
(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 $x \geq 1$. Thus $f(F^{-1}(1/2)) = 7(1/2)^{8/7}$.

7. (30 points) Suppose that $X, X_1, X_2, \ldots, X_n$ are independent Geometric(p) random variables:

$$P(X = k) = p(1-p)^{k-1}, \quad k = 1, 2, 3, \ldots.$$

Recall that $E(X) = 1/p$ and $Var(X) = (1-p)/p^2 \equiv q/p^2$.
(a) Use the weak law of large numbers to show that the random vector

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

converges in probability to some vector $\underline{y}$ where $\underline{y}$ depends on p. Give $\underline{y}$ explicitly in terms of p.
(b) Use the multivariate CLT to show that

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

for some covariance matrix Σ; compute Σ explicitly in terms of λ.
(c) The usual estimator of p is $\widehat{p}_n = 1/\overline{X}_n$. A friend suggests the following alternative estimator of p:

$$\tilde{p}_n = 1 - \frac{\sum_{i=1}^{n} 1_{[X_i=2]}}{\sum_{i=1}^{n} 1_{[X_i=1]}} = 1 - \frac{\overline{Y}_{3,n}}{\overline{Y}_{2,n}}.$$

Is $\tilde{p}_n$ a consistent estimator of p?
(d) If the answer to (c) is yes, find the asymptotic variance of $\tilde{p}_n$ as an estimator of p.

Solution: (a) Since $E(X_1, 1_{[X_1=1]}, 1_{[X_1=2]}) = (1/p, p, p(1-p))$, it follows by the weak law of large numbers that

$$\underline{\overline{Y}}_n \equiv \frac{1}{n}\sum_{i=1}^{n}(X_i, 1_{[X_i=1]}, 1_{[X_i=2]})' \to_p (1/p, p, p(1-p)) \equiv \underline{y}.$$

(b) Since $E(X_1^2 + 1^2_{[X_1=1]} + 1^2_{[X_1=2]}) \le ((1-p)/p^2 + 1/p^2 + 1 + 1) < \infty$, it follows from the multivariate CLT that

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

where, with $q \equiv 1-p$,

$$\Sigma = Cov(X, 1_{\{x\}}(X), 1_{\{2\}}(X)) = \begin{pmatrix} q/p^2 & -q & (2p-1)q \\ -q & pq & -p^2 q \\ (2p-1)q & -p^2 q & pq(1-pq) \end{pmatrix}.$$

Here we used

$$Var(X) = (1-p)/p^2 = q/p^2, \qquad E\{(X - 1/p)1_{\{1\}}(X)\} = p - p/p = p - 1 = -q,$$
$$E\{(X - 1/p)1_{\{2\}}(X)\} = 2pq - pq/p = (2p-1)q, \qquad Var(1_{\{1\}}(X)) = pq,$$
$$Var(1_{\{2\}}(X)) = P(X=2)(1 - P(X=2)) = pq(1-pq),$$
$$E\{1_{\{1\}}(X)1_{\{2\}}(X)\} - E\{1_{\{1\}}(X)\}E\{1_{\{2\}}(X)\} = 0 - p \cdot pq = -p^2 q.$$

(c) From the convergence in (a) and the continuous mapping theorem it follows that

$$\tilde{p}_n \to_p 1 - \frac{pq}{p} = 1 - (1-p) = p.$$

Thus $\tilde{p}_n$ is a consistent estimator of p. Moreover, with $g(u,v,w) \equiv 1 - w/v$, $g'(u,v,w) = (0, w/v^2, -1/v)$, and hence $g'(\underline{y}) = g'(1/p, p, pq) = (0, q/p, -1/p)$. Thus by the g'-theorem or delta-method

$$\begin{aligned}
\sqrt{n}(\tilde{p}_n - p) &= \sqrt{n}(g(\underline{Y}_n) - g(\underline{y})) \\
&\to_d \underline{g}'\underline{W} = \frac{1}{p}(qW_2 - W_3) \\
&\sim N(0, q(q+1)/p)
\end{aligned}$$

Since

$$\begin{aligned}
Var((qW_2 - W_3)/p) &= p^{-2}\left\{q^2 Var(W_2) + Var(W_3) - 2qCov(W_2, W_3)\right\} \\
&= p^{-2}\left\{q^2 pq + pq(1-pq) + 2qp^2 q\right\} \\
&= p^{-2}pq\left\{q^2 + 1 - pq + 2pq\right\} = \frac{q}{p}\{1 + pq + q^2\} \\
&= \frac{q}{p}\{1 + q(p+q)\} = \frac{q(q+1)}{p}.
\end{aligned}$$

The following calculations were not a part of the problem, but give a comparison of the two estimators $\tilde{p}_n$ and $\hat{p}_n$: First we find the asymptotic distribution of $\hat{p}_n$: with $h(x) = 1/x$ we find $h'(x) = -1/x^2$ so that $h'(1/p) = -p^2$; thus the delta method yields

$$\sqrt{n}(\hat{p}_n - p) = \sqrt{n}(h(\overline{X}_n) - h(1/p)) \to_d (-p^2)W_1 \sim p^2 N(0, q/p^2) = N(0, qp^2).$$

Thus the ratio of asymptotic variances, or ARE of $\tilde{p}_n$ with respect to $\hat{p}_n$ is

$$ARE(\tilde{p}_n, \hat{p}_n) = \frac{qp^2}{q(q+1)/p} = \frac{p^3}{q+1} = \frac{p^3}{2-p}.$$

Since the ARE varies from 0 to .2 as p goes from 0 to .647, we see that $\tilde{p}_n$ is considerably less efficient as an estimator of p for small values of p. When p is close to 1, the ARE is close to 1, so $\tilde{p}_n$ has only slightly larger variance than $\hat{p}_n$.