Statistics 581, Midterm Exam

Wellner; 11/04/2009

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

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 $Y = (Y_1, \ldots, Y_n)$.

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.

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?

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

Do **either** problem 5 **or** problem 6.

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) $n^{1/4}(\overline{X}_n - \mu) \to_p 0$?
(vi) $\cos(S_n) = O_p(1)$?

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 five 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.

7. (30 points) Suppose that $X, X_1, X_2, \ldots, X_n$ are independent Geometric(p) random variables: $P(X = k) = p(1-p)^{k-1}$, $k = 1, 2, 3, \ldots$. 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 $(a, b, c)' \equiv \underline{y}$ where (a, b, c) depends on p. Give (a, b, c) 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 $\tilde{p}$ as an estimator of p.