Statistics 581, Midterm Exam

Wellner; 11/08/2002

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

1. (24 points) **Define** *any three* of the following terms. In each case, provide an appropriate context for your definition.
(a) A normal random vector $Y = (Y_1, \ldots, Y_n)$.
(b) Convergence in rth mean (of a sequence of random variables).
(c) Convergence in probability (of a sequence of random variables).
(d) The inverse or quantile function F^{-1} of a distribution function F.
(e) A uniformly integrable sequence of random variables X_n.

2. (24 points) **State** *any three* of the following results:
(a) The dominated convergence theorem.
(b) The delta-method or g'-theorem for a differentiable function $g : R^k \to R^m$.
(c) The Cramér-Wold device.
(d) Vitale's theorem.
(e) The basic inequality and Markov's inequality.
(f) The elementary Skorokhod theorem.
(g) 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).
A. Let $c \in (0, \infty]$. Give examples of sequences of random variables X_n with $X_n \to_p 0$ but $E(X_n) \to c$.
B. Give an example of a sequence of random variables which is not uniformly integrable.
C. Show that if $|X_n| \leq Y$ where Y is integrable then $\{X_n\}$ is uniformly integrable.
D. Show that if $\{X_n\}$ is a sequence with $\limsup_{n\to\infty} E|X_n|^r < \infty$ for some $r > 1$, then $\{X_n\}$ is uniformly integrable.

5. (30 points)
Suppose that $X, X_1, \ldots, X_n$ are i.i.d. Exponential(θ) random variables so that $P_\theta(X > x) = \exp(-\theta x) = 1 - F_\theta(x)$ for $x > 0$.
A. Fix $x_0 > 0$ and let $\mathbb{F}_n(x) = n^{-1}\sum_{i=1}^n 1_{[X_i \le x]} = n^{-1}\sum_{i=1}^n 1_{(\infty,x]}(X_i)$ denote the empirical distribution function. Show that

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - 1/\theta \\ \mathbb{F}_n(x_0) - F_\theta(x_0) \end{pmatrix} \rightarrow_d Y \sim N_2(0, \Sigma)$$

and find Σ.
B. Let $g(\theta) \equiv F_\theta(x_0) = 1 - \exp(-\theta x_0)$, and consider the two estimators of $F = F_\theta$ given by $T_{n,1} \equiv g(\widehat{\theta}_n)$ and $T_{n,2} \equiv \mathbb{F}_n(x_0)$ where $\widehat{\theta}_n \equiv 1/\overline{X}_n$. Show that

$$\sqrt{n}\begin{pmatrix} T_{n,1} - F_\theta(x_0) \\ T_{n,2} - F_\theta(x_0) \end{pmatrix} \rightarrow_d \tilde{Y}$$

and find the distribution of $\tilde{Y}$.
C. What is the advantage of $T_{n,2} = \mathbb{F}_n(x_0)$ as an estimator even though it is inefficient when the exponential model holds?

Do **either** problems 6 **or** problem 7.

6. (32 points).
A. Suppose that X_{ni}, $i = 1, \ldots, n$ are independent Bernoulli(p_{ni}) random variables. Show that if $\sum_{i=1}^n p_{ni}(1 - p_{ni}) \rightarrow \infty$ then

$$\frac{\sum_{i=1}^n (X_{ni} - p_{ni})}{\sqrt{\sum_1^n p_{ni}(1 - p_{ni})}} \rightarrow_d N(0, 1)\,.$$

B. Use A to show that if $p_{ni} = p_n$ for all $i = 1, \ldots, n$ where $p_n \rightarrow p_0 > 0$, then

$$\sqrt{n}(n^{-1}\sum_1^n X_{ni} - p_n) \rightarrow_d N(0, p_0(1 - p_0))\,.$$

C. Use the result of B to show that if $x \in R$, $t \in (0, 1)$ is fixed, and F is continuous at $F^{-1}(t)$, then

$$\sqrt{n}(\mathbb{F}_n(F^{-1}(t) + xn^{-1/2}) - F(F^{-1}(t) + xn^{-1/2})) \rightarrow_d N(0, t(1 - t))\,.$$

D. Suppose that Δ_{ni}, $i = 1, \ldots, n$ are independent and identically distributed Mult$_k(1, p_n)$ random vectors where $p_n = (p_{n1}, \ldots, p_{nk}) \rightarrow (p_{0,1}, \ldots, p_{0,k})$. Show that

$$\sqrt{n}\left(n^{-1}\sum_1^n \Delta_{ni} - p_n\right) \rightarrow_d N_k(0, \text{diag}\,(p_0) - p_0 p_0^T)\,.$$

7. (32 points). Suppose that $N = (N_1, \ldots, N_k) \sim \ Mult_k(n, \underline{p})$ and consider testing $H_0 : \underline{p} = \underline{p}_0$ versus $K_0 : \underline{p} \neq \underline{p}_0$. Instead of the chi-square statistic Q_n, consider the test statistic given by

$$H_n^2 \equiv 4n \sum_{i=1}^{k} (\sqrt{\hat{p}_i} - \sqrt{p_{i0}})^2.$$

The statistic H_n^2 is $8n$ times the square of the *Hellinger distance* between $\hat{\underline{p}}$ and $\underline{p}_0$.
A. Find the limiting distribution of H_n^2 under the null hypothesis H_0.
B. Find the limit of $n^{-1} H_n^2$ under fixed alternatives $\underline{p} \neq \underline{p}_0$ in K_0, and use this to show that the test based on H_n^2 is consistent for fixed $\underline{p} \in K_0$.
C. Find the limiting distribution of H_n^2 under local alternatives $\underline{p}_n = \underline{p}_0 + \underline{c}/\sqrt{n}$ (with $\underline{c}'\underline{1} = 0$), and use this to approximate the power of this test. Compare the (local asymptotic) power of this test to the chi-square test.