Statistics 522, Final Exam

Wellner; 3/19/2008

1. (24 points). **Define** *three* of the following six terms:
   (a) An asymptotically tight sequence $\{X_n\}$ in a metric space (M, d).
   (b) Weak convergence of probability measures $\{P_n\}$ to a probability measure P on $(M, \mathcal{M})$ for a metric space (M, d).
   (c) The characteristic function of a real-valued random variable X and of a random vector $\underline{X}$ with values in $\mathbb{R}^d$.
   (d) A Brownian motion process $\mathbb{S}$ on $[0, \infty)$.
   (e) A Brownian bridge process $\mathbb{U}$ on $[0, 1]$.
   (f) The "Uniformly Asymptotically Negligible" (or UAN) property of a triangular array of random variables $\{X_{nk} : 1 \le k \le n,\ n \ge 1\}$ (in the context of the Lindeberg-Feller central limit theorem).

2. (30 points). Give careful **statements** of *any three* of the following six theorems or results:
   (a) A result connecting convergence of characteristic functions of a sequence of random variables X_n to tightness of the sequence X_n.
   (b) Donsker's theorem for the uniform empirical process $\mathbb{U}_n(t) = \sqrt{n}(\mathbb{G}_n(t) - t)$ where $\mathbb{G}_n(t) = n^{-1}\sum_{i=1}^n 1_{[0,t]}(\xi_i)$ and $\xi_1, \ldots, \xi_n$ are i.i.d. Uniform$(0, 1)$ random variables.
   (c) Any result connecting a Brownian motion process $\mathbb{S}$ on $[0, 1]$ or $[0, \infty)$ to a Brownian bridge process $\mathbb{U}$ on $[0, 1]$.
   (d) The Cramér - Lévy continuity theorem for characteristic functions.
   (e) The law of the iterated logarithm for $S_n = X_1 + \cdots + X_n$ where $X_1, \ldots, X_n$ are i.i.d. with $EX_i = 0$ and $Var(X_i) = \sigma^2$.
   (f) The Berry-esseen theorem for $S_n = X_1 + \cdots + X_n$ where $X_1, \ldots, X_n$ are i.i.d. with $E(X_1) = 0$, $Var(X_1) = \sigma^2$, and $E|X_1|^3 < \infty$.

3. (36 points) Consider the following sequences of distribution functions F_n on $\mathbb{R}$ with densities f_n with respect to Lebesgue measure:
   (a) $f_n(x) = n^{-1} 1_{[0,n]}(x)$.
   (b) $f_n(x) = ne^{-nx} 1_{[0,\infty)}(x)$.
   (c) $f_n(x) = n^{-1}e^{-x/n} 1_{[0,\infty)}(x)$.
   (i) In each of (a) - (c), describe these distributions in terms of random variables of the form $X_n = a_n Y$ where Y has a fixed distribution F_0 with density f_0; that is, find random variables of this form so that X_n has distribution function F_n with density f_n.
   (ii) Which of (a) - (c), if any, are uniformly (or asymptotically) tight sequences? In the case of a tight sequence, identify the collection of limiting distributions.

4. (25 points). Suppose that X and Y are random variables on the probability space $(\Omega, \mathcal{A}, P)$ with $X \in L_2(P)$ and $Y \in L_2(P)$ (so that $XY \in L_1(P)$), and suppose that $\mathcal{D}$ is a sub sigma-field of $\mathcal{A}$. Show that

$$Cov(X,Y) = E\left[Cov(X,Y|\mathcal{D})\right] + Cov(E(X|\mathcal{D}), E(Y|\mathcal{D}))$$

where

$$Cov(X,Y|\mathcal{D}) = E\left[(X - E(X|\mathcal{D})(Y - E(Y|\mathcal{D}))|\mathcal{D}\right].$$

(This generalizes our formula for the variance of a random variable X obtained in the midterm exam.)

5. (30 points). Suppose that $\xi_1, \xi_2, \ldots$ are i.i.d. Uniform$(0,1)$ random variables, and let $\mathbb{G}_n(t) = n^{-1}\sum_{i=1}^n 1_{[0,t]}(\xi_i)$ be the empirical distribution function of the ξ_i's. Define $\mathcal{F}_n(t) \equiv \sigma\{\mathbb{G}_n(r) : \ t \le r \le 1\}$ for $0 < t \le 1$.
(a) Let $0 < s < t < 1$. Show that

$$(n\mathbb{G}(s), n(\mathbb{G}_n(t) - \mathbb{G}_n(s)), n(1 - \mathbb{G}_n(t))) \sim \text{Mult}_3(n, (s, t-s, 1-t)).$$

(b) Find the conditional distribution of $n\mathbb{G}_n(s)$ conditional on $n(1 - \mathbb{G}_n(t))$.
(c) Use the result of (b) to show that

$$E\left\{\frac{\mathbb{G}_n(s)}{s}\Big|\mathbb{G}_n(t)\right\} = \frac{\mathbb{G}_n(t)}{t} \qquad \text{almost surely.}$$

Note: It is true that the conditional distribution of $\mathbb{G}_n(s)$ given the $\sigma-$field $\mathcal{F}_n(t)$ is the same as the conditional distribution of $\mathbb{G}_n(s)$ given the random variable $\mathbb{G}_n(t)$; this is the Markov property of $\mathbb{G}_n$. Thus (c) shows that

$$E\left\{\frac{\mathbb{G}_n(s)}{s}\Big|\mathcal{F}_n(t)\right\} = \frac{\mathbb{G}_n(t)}{t} \qquad \text{almost surely;}$$

that is, $\{(\mathbb{G}_n(t)/t, \mathcal{F}_n(t)) : \ 0 < t \le 1\}$ is a reverse martingale.

6. (36 points)
(a) Suppose that $\epsilon_1, \ldots, \epsilon_n, \ldots$ are i.i.d. Rademacher random variables (i.e. $P(\epsilon_i = \pm 1) = 1/2$ for $i = 1, 2, \ldots$. Let $X_{ni} = (i/n)\epsilon_i$ for $1 \le i \le n$. If $S_n \equiv \sum_{i=1}^n X_{ni}$, find constants σ_n such that $S_n/\sigma_n \to_d Z \sim N(0,1)$.
(b) Suppose we replace $\epsilon_1, \ldots, \epsilon_n$ in (a) by $Y_1, \ldots, Y_n$ i.i.d. with $E(Y_i) = 0$, $Var(Y_i) = \sigma^2$, and we replace $\{(i/n) : 1 \le i \le n\}$ in (a) with $\{a_{ni} : 1 \le i \le n\}$ with $\max_{i \le n} a_{ni}^2 / \sum_1^n a_{nj}^2 \to 0$. Let $X_{ni} \equiv a_{ni}Y_i$ for $1 \le i \le n$ and define $S_n \equiv \sum_{i=1}^n X_{ni}$. Again find constants σ_n^2 such that $S_n/\sigma_n \to_d Z \sim N(0,1)$.

Do either problem 7 or problem 8.

7. (30 points)
Suppose that $X_1, X_2, \ldots$ are i.i.d. Cauchy random variables with density function $f(x) = \pi^{-1}(1 + x^2)^{-1}$. We showed in class that the characteristic function of each of these X_i's is $\phi(t) = e^{-|t|}$.
(a) If $a_i \in \mathbb{R}$ for $i = 1, \ldots, n$, find the characteristic function of $\sum_{i=1}^n a_i X_i$.
(b) Give conditions on the numbers $\{a_i : 1 \le i \le n\}$ which imply that $\sum_1^n a_i X_i \to_d$ something.
(c) What is the limiting distribution in (b)?

8. (30 points)
(a) Suppose that X is a Rademacher random variable; i.e. $P(X = \pm 1) = 1/2$. Find the characteristic function ϕ_X of X.
(b) Let X_1, X_2 be two independent Rademacher random variables. Find the characteristic function of $Z \equiv X_1 + X_2$.
(c) Use the two independent Rademacher random variables X_1, X_2 in (b) to define two independent Bernoulli random variables Y_1, Y_2. Use the results of (b) to calculate the characteristic function of $Y_1 + Y_2$.
(d) Let $Z_1, \ldots, Z_n$ be i.i.d. each with the same distribution as $Z = X_1 + X_2$ in (b). What is the characteristic function of $S_n \equiv \sum_{i=1}^n Z_i$?
(e) Show that the characteristic function you computed in (c) converges to a function φ that is not a characteristic function.
(f) Find a sequence of numbers a_n such that the characteristic function of S_n/a_n converges to a proper characteristic function and $S_n/a_n \to_d$ something; identify "something".