Statistics 523, Midterm Exam Solutions

Wellner; 5/13/2013

1. (30 points). **Define** *three* of the following five terms:
   (a) An infinitely divisible random variable X
   (or distribution function F).
   (b) A stable distribution F (or stable random variable) X.
   (c) The domain of attraction of a stable distribution G.
   (d) The *strong Markov property* of a process $\{X(t) : 0 \le t < \infty\}$.
   (e) A standard Brownian motion process $\mathbb{S}$ on $[0, \infty)$.

   Solution: See Shorack, PfS, Course Notes.

2. (36 points). Give careful **statements** of *three* of the following five theorems or results:
   (a) Donsker's theorem for the partial sum process $\{\mathbb{S}_n(t) : \ 0 \le t \le 1\}$.
   (b) Donsker's theorem for the uniform empirical process $\{\mathbb{U}_n(t) : \ 0 \le t \le 1\}$.
   (c) Four properties of Brownian motion $\mathbb{S}$ on $[0, \infty)$.
   (d) A result concerning embedding of a random variable with $E(X) = 0$ and variance $Var(X) = 1$ in Brownian motion.
   (e) The strong Markov property of Brownian motion.

   Solution: See Shorack, PfS, Course Notes.

3. (24 points). Suppose that $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ are i.i.d. Uniform$(0, 1)$ random variables, and let $0 \equiv \xi_{n:0} \le \xi_{n:1} \le \xi_{n:2} \le \cdots \le \xi_{n:n} \le \xi_{n:n+1} \equiv 1$ denote the order statistics of the ξ_i's. From problem set 5 we know that

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

where $S_j = \sum_{i=1}^{j} Y_i$ where the $Y_1, Y_2, \ldots$ are i.i.d. Exponential(1). It follows that the spacings $\delta_{n,k} \equiv \xi_{n:k} - \xi_{n:k-1}$, $k = 1, \ldots, n+1$ satisfy

$$(\delta_{n:1}, \ldots, \delta_{n:n+1}) \stackrel{d}{=} \left(\frac{Y_1}{S_{n+1}}, \frac{Y_2}{S_{n+1}}, \ldots, \frac{Y_{n+1}}{S_{n+1}} \right).$$

Use the representation of the spacings $\delta_{n,k}$ in the display above to show that for each fixed k we have

$$n(\delta_{n:1}, \ldots, \delta_{n:k}) \to_d (Y_1, Y_2, \ldots, Y_k) \quad \text{as} \quad n \to \infty.$$

Solution: Note that $EY_1 = 1 < \infty$, so the strong law of large numbers yields

$$\frac{S_{n+1}}{n} = \frac{n+1}{n} \frac{S_{n+1}}{n+1} \to_{a.s.} 1 \cdot 1 = 1.$$

Thus we have

$$\begin{aligned} n(\delta_{n:1}, \ldots, \delta_{n:k}) &\stackrel{d}{=} \left(\frac{Y_1}{n^{-1}S_{n+1}}, \frac{Y_2}{n^{-1}S_{n+1}}, \ldots, \frac{Y_k}{n^{-1}S_{n+1}} \right) \\ &\to_{a.s.} (Y_1, Y_2, \ldots, Y_k), \end{aligned}$$

and this implies that $n(\delta_{n:1}, \ldots, \delta_{n:k}) \to_d (Y_1, \ldots, Y_k)$. [This can be extended to $k = k_n$ growing with n: if $k_n/n \to 0$, then, $d_{TV}(P_{n,k_n}, Q_{k_n}) \to 0$ as $n \to \infty$ where $P_{n,k}$ denotes the joint distribution of $n(\delta_{n:1}, \ldots, \delta_{n:k_n})$ and Q_{k_n} denotes the distribution of k_n independent exponential rv's; see Runnenburg, J. Th.; Vervaat, W. (1969). Asymptotical independence of the lengths of subintervals of a randomly partitioned interval; a sample from S. Ikeda's work. *Statistica Neerlandica* **23**, 67–77 .

4. (40 points) (a) Define what is meant by a *Uniformly Asymptotically Negligible* (UAN) array of random variables $\{X_{n,k} : 1 \le k \le n,\ n \ge 1\}$. (b) State a theorem involving convergence of sums of independent random variables which involves the UAN condition as a key hypothesis. (c) Give a condition equivalent to the UAN condition (either in terms of characteristic functions or truncated moments).

   Solution: (a) An array $\{X_{n,k} : 1 \le k \le n\}$ of row-independent rv's is said to be UAN if $\max_{1 \le k \le n} P(|X_{n,k}| > \epsilon) \to 0$ for every $\epsilon > 0$.
   (b) The general theorem for convergence of a UAN array to an infinitely divisible law says that $S_n \equiv \sum_{i=1}^n X_{n,i} \to_d S$ where the infinitely divisible law S has characteristic function

$$\phi_S(t) = Ee^{itS} = \exp\left\{ i\beta t + \int_{\mathbb{R}} (e^{itx} - 1 - it\beta) \frac{1}{\alpha(x)} dH(x) \right\}$$

   if and only if

$$\beta_n \to \beta \quad \text{and} \quad H_n \to_d H$$

where, with $b_{nk} \equiv E(X_{nk} 1_{(-\delta,\delta)}(X_{n,k})$ for some $\delta > 0$,

$$\beta_n \equiv \sum_{k=1}^{n} \{b_{nk} + \int_{\mathbb{R}} \beta(x) dF_{nk}(x + b_{nk})\}$$

and

$$H_n(\cdot) = \sum_{k=1}^{n} \int_{-\infty}^{\cdot} \alpha(y) dF_{n,k}(y + b_{nk}).$$

(c) The UAN condition is equivalent to: (i) $\max_{1 \le k \le n} |\phi_{nk}(t) - 1| \to 0$; and also to $\max_{1 \le k \le n} E(X_{n,k}^2 \wedge 1) \to 0$ as $n \to \infty$.

5. (40 points). Suppose that $\underline{X}_n = \sum_{i=1}^{n} \underline{Y}_i$ where $\underline{Y}_1, \ldots, \underline{Y}_n$ are i.i.d. Multinomial$_k(1, \underline{p})$ random vectors; i.e. $P(\underline{Y} = \underline{y}) = \prod_{j=1}^{k} p_j^{y_j}$ for $\underline{y} = (y_1, \ldots, y_k)$ with each $y_j \in \{0, 1\}$ and $\sum_{j=1}^{k} y_j = 1$.
(a) What is the distribution of $\underline{X}_n$?
(b) Suppose that $N_\lambda \sim$ Poisson(λ).
What is the distribution of $\underline{X}_{N_\lambda} = \sum_{i=1}^{N_\lambda} \underline{Y}_i$?
(c) Now suppose that $\xi_1, \ldots, \xi_n$ are i.i.d. Uniform$(0,1)$ random variables and let $\mathbb{N}_n(t) \equiv n\mathbb{G}_n(t) \equiv \sum_{i=1}^{n} 1_{[0,t]}(\xi_i)$. Suppose that $0 < t_1 < t_2 < \cdots < t_k < 1$. What is the joint distribution of $(\mathbb{N}_n(t_1), \mathbb{N}_n(t_2) - \mathbb{N}_n(t_1), \ldots, \mathbb{N}_n(1) - \mathbb{N}_n(t_k))$?
(d) Use (b) and (c) to show that $\mathbb{N}_{N_\lambda}$ is a Poisson process with rate λ.

Solution: (a) $\underline{X}_n \sim$ Mult$_k(n, \underline{p})$; i.e.

$$P(\underline{X}_n = \underline{x}) = \binom{n}{\underline{x}} \prod_{j=1}^{k} p_j^{x_j} \quad \text{if} \;\; x_i \in \mathbb{N} \;\; \text{with} \; \sum_{j=1}^{k} x_j = n,$$

and 0 otherwise, with $\binom{n}{\underline{x}} = n! / \prod_{j=1}^{k} x_j!$.

(b) The characteristic function of $\underline{X}_{N_\lambda}$ is given by

$$\begin{aligned} E\left\{\exp(i\underline{t}^T \underline{X}_{N_\lambda})\right\} &= E\left\{E\left\{\exp(i\underline{t}^T \underline{X}_{N_\lambda})|N_\lambda\right\}\right\} \\ &= E\left\{E\left(\exp(i\underline{t}^T \underline{Y}_1\right)^{N_\lambda}\right\} \\ &= E\left(\sum_{j=1}^k p_j e^{it_j}\right)^{N_\lambda} \\ &= \sum_{m=0}^\infty \left(\sum_{j=1}^k p_j e^{it_j}\right)^m \exp(-\lambda)\frac{\lambda^m}{m!} \\ &= \exp\left\{\lambda\left(\sum_{j=1}^k p_j(e^{it_j}-1)\right)\right\} \\ &= \prod_{j=1}^k \exp(\lambda p_j(e^{it_j}-1)) = \prod_{j=1}^k E\exp(it_j V_j) \end{aligned}$$

where $V_j \sim \text{Poisson}(\lambda p_j)$, $j = 1, \ldots, k$ are independent.
(c) It is immediate that

$$(\mathbb{N}_n(t_1), \mathbb{N}_n(t_2) - \mathbb{N}_n(t_1), \ldots, \mathbb{N}_n(1) - \mathbb{N}_n(t_k)) \sim \text{Mult}_{k+1}(n, (t_1, t_2 - t_1, \ldots, 1 - t_k)).$$

(d) By (b) and (c) we have

$$(\mathbb{N}_{N_\lambda}(t_1), \mathbb{N}_{N_\lambda}(t_2) - \mathbb{N}_{N_\lambda}(t_1), \ldots, \mathbb{N}_{N_\lambda}(1) - \mathbb{N}_{N_\lambda}(t_k)) \stackrel{d}{=} (V_1, \ldots, V_{k+1})$$

where the V_j's are independent and $V_j \sim$ Poisson$(\lambda(t_j - t_{j-1}))$. But this is just the distribution of the increments of a Poisson process with rate λ.

6. (30 points). Suppose that $X_1, X_2, \ldots$ are i.i.d. $(0,1)$ and $S_k \equiv X_1 + \cdots + X_k$ for $k = 1, 2, 3, \ldots$.
(a) What is the limiting distribution of:
   (i) $\sup_{1 \le k \le n} n^{-1/2} S_k$?
   (ii) $n^{-1} \sum_{k=1}^n 1_{[S_k > 0]}$?
   (iii) $n^{-1} \sum_{k=[an]}^n 1_{[S_k > c\sqrt{k}]}$ where $a, c > 0$?
   (iv) $n^{-1} \inf\{k \ge 1 : \ S_k \ge b\sqrt{n}\}$ for $b > 0$?

(b) Find a sequence of normalizing constants c_n so that

$$\frac{1}{c_n}\sum_{k=1}^{n} S_k^3 \to_d \text{ something as } \; n \to \infty$$

and identity "something" in terms of BM

Solution: We let $\mathbb{S}_n$ denote the partial sum process of the X_i's:

$$\mathbb{S}_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{[nt]} X_i.$$

(a) (i) Now

$$\sup_{1\le k\le n}\frac{S_k}{\sqrt{n}} = \sup_{0\le t\le 1}\mathbb{S}_n(t) \to_d \sup_{0\le t\le 1}\mathbb{S}(t).$$

where $\mathbb{S}$ is standard Brownian motion. (ii) Here we have

$$n^{-1}\sum_{i=1}^{n} 1_{[S_k>0]} = \int_0^1 1_{[\mathbb{S}_n(t)>0]}dt \to_d \int_0^1 1_{[\mathbb{S}(t)>0]}dt \sim \text{arcsin}.$$

(iii) Here for $a > 0$ and $c > 0$ we have

$$n^{-1}\sum_{k=[an]}^{n} 1_{[S_k>c\sqrt{k}]} = \int_{[an]/n}^1 1_{[\mathbb{S}_n(t)>c]}dt \to_d \int_a^1 1_{[\mathbb{S}(t)>c]}dt.$$

(iv) In this case

$$n^{-1}\inf\{k \ge 1 : \; S_k \ge b\sqrt{n}\} = \inf\{k/n \ge 1/n : \; \mathbb{S}_n(k/n) \ge b\} \to_d \inf\{t > 0 : \mathbb{S}(t) \ge b\}.$$

(b) We see that

$$\begin{aligned}\frac{1}{c_n}\sum_{k=1}^{n} S_k^3 &= \frac{n^{3/2}}{c_n}\sum_{k=1}^{n}\mathbb{S}_n(k/n)^3 = \frac{n^{3/2}\cdot n}{c_n}\frac{1}{n}\sum_{k=1}^{n}\mathbb{S}_n(k/n)^3 \\ &= \frac{1}{n}\sum_{k=1}^{n}\mathbb{S}(k/n)^3 = \int_0^1 \mathbb{S}_n(t)^3 dt \;\text{ if }\; c_n = n^{5/2} \\ &\to \int_0^1 \mathbb{S}^3(t)dt.\end{aligned}$$

7. (30 points). Suppose $\mathbb{S}$ is standard Brownian motion, and $-a < 0 < b$. Let $\tau \equiv \tau_{a,b} \equiv \inf\{t > 0 : \mathbb{S}(t) \notin (-a, b)\}$.
(a) Sketch the proof that

$$P(\mathbb{S}(\tau) = -a) = \frac{b}{b+a} \quad \text{and} \quad P(\mathbb{S}(\tau) = b) = \frac{a}{b+a}.$$

(b) What is $E\mathbb{S}^2(\tau)$? What is $E\tau$?
What is the relationship between them?
(c) How would you bound $E\tau^2$?

Solution: (a) Since $\{\mathbb{S}(t) : t \geq 0\}$ is a martingale and τ is a stopping time, the optional sampling theorem yields

$$\begin{aligned} 0 &= E\mathbb{S}(\tau) = -aP(\mathbb{S}(\tau) = -a) + bP(\mathbb{S}(\tau) = b) \\ &= -a + (a+b)P(\mathbb{S}(\tau) = b), \end{aligned}$$

and hence $P(\mathbb{S}(\tau) = b) = a/(a+b) = 1 - P(\mathbb{S}(\tau) = -a)$.
(b) Since $\{\mathbb{S}^2(t) - t : \ t \geq 0\}$ is a martingale and $\mathbb{S}$ is a stopping time

$$0 = E\{\mathbb{S}^2(\tau) - \tau\} = E(\mathbb{S}^2(\tau)) - E(\tau),$$

so that

$$E(\tau) = E(\mathbb{S}^2(\tau)) = a^2\frac{b}{a+b} + b^2\frac{a}{a+b} = ab.$$

(c) Using the martingale $Y_4(t) = \mathbb{S}^4(t) - 6t\mathbb{S}^2(t) + 3t^2$ yields

$$0 = E(\mathbb{S}^4(\tau) - 6\tau\mathbb{S}^2(\tau) + 3\tau^2),$$

and by re-arranging this we find that

$$\begin{aligned} 3E\tau^2 &= E(6\tau\mathbb{S}^2(\tau)) - E\mathbb{S}^4(\tau) \\ &\leq 6E(\tau(a \vee b)^2) \leq 6ab(a+b)^2 \end{aligned}$$

by using $E\tau = ab$ from (b). Thus $E\tau^2 \leq 2ab(a+b)^2$.