Statistics 523, Problem Set 7 Solutions

Wellner; 5/21/2010

1. PfS, Exercise 13.1.4, page 353: Verify the claims made in example 1.10: For all $t \in \mathbb{R}$ let

$$\mathbb{N}_t \equiv 1_{[X \le t]}, \qquad \text{and} \qquad \mathcal{A}_t \equiv \sigma[\mathbb{N}(r): \ r \le t].$$

(a) Show that the class $\mathcal{C}_s \equiv \{[X > r] : -\infty \le r \le s\}$ is a $\overline{\pi}-$system that generates $\mathcal{A}_s$. Thus any two finite measures that agree on $\mathcal{C}_s$ also agree on $\mathcal{A}_s$ by the Dynkin $\pi - \lambda$ theorem.
(b) Show that for all $-\infty < s < t < \infty$

$$E\{\mathbb{N}_t | \mathcal{A}_s\} =_{a.s.} 1_{[X \le s]} + 1_{[X > s]} \frac{F(s,t]}{1 - F(s)}.$$

where $F(s,t] \equiv F(t) - F(s)$.
(c) Similarly, verify that for all $-\infty < s < u \le t < \infty$

$$E\{1_{[X \ge u]} | \mathcal{A}_s\} =_{a.s.} 1_{[X > s]} \frac{1 - F(u-)}{1 - F(s)}.$$

(d) Show that the process

$$\mathbb{M}_t \equiv \mathbb{N}_t - \int_{(-\infty,t]} 1_{[X \ge r]} d\Lambda(r) \equiv \mathbb{N}_t - \mathbb{A}_t$$

is a martingale on $\mathbb{R}$ adapted to the $\mathcal{A}_t$'s. To do this, show that

$$E\{1_A(\mathbb{N}_t - \mathbb{N}_s)\} = E\{1_A(\mathbb{A}_t - \mathbb{A}_s)\} \tag{1}$$

for all sets $A \in \mathcal{C}_s$.

Solution: (a) The class of sets $\mathcal{C}_s = \{[X > r] : -\infty \le r \le s]\}$ is a $\overline{\pi}-$system: it is closed under finite intersections since for $r, r' \in [-\infty, s]$ we have $[X > r] \cap [X > r'] = [X > r \wedge r'] \in \mathcal{C}_s$; and it contains Ω since $[X > -\infty] = \Omega$. Furthermore, $\mathcal{C}_s$ generates $\mathcal{A}_s$: by the proof of Dynkin's $\pi - \lambda$ theorem, since $\mathcal{C}_s$ is a $\pi-$ system, it follows that $\sigma[\mathcal{C}_s] = \lambda[\mathcal{C}_s]$ where $\lambda[\mathcal{C}_s]$ is the minimal $\lambda-$system generated by $\mathcal{C}_s$.

But $\lambda[\mathcal{C}_s] \subset \mathcal{A}_s = \sigma[[X \le r] : \; r \le s]$ since $\mathcal{A}_s$ is a $\lambda-$system which contains $\mathcal{C}_s$: $[X > r] = [X \le r]^c \in \mathcal{A}_s$ for all $-\infty \le r \le s$. Furthermore $\mathcal{A}_s$ is the minimal such $\lambda-$system containing $\mathcal{C}_s$, and hence the claim holds.

(b) The claimed equality holds if and only if

$$E\left\{1_A\left(1_{[X\le s]} + 1_{[X>s]}\frac{F(s,t]}{1-F(s)}\right)\right\} = E\{1_A \mathbb{N}_t\}$$

for all $A \in \mathcal{A}_s$. But by (a) it suffices to show this identity for $A \in \mathcal{C}_s$. Thus let $A = [X > r]$ for some $r \le s$. Then the right side is

$$E\{1_{[X>r]}1_{[X\le t]}\} = E\{1_{[r<X\le t]}\} = F(t) - F(r),$$

and the left side is

$$\begin{aligned} & E\left\{1_{[X>r]}1_{[X\le s]} + 1_{[X>r]}1_{[X>x]}\frac{F(s,t]}{1-F(s)}\right\} \\ = \;& E1_{[r<X\le s]} + E\{1_{[X>s]}\}\frac{F(s,t]}{1-F(s)} = F(s) - F(r) + (1-F(s))\frac{F(s,t]}{1-F(s)} \\ = \;& F(s) - F(r) + F(t) - F(s) = F(t) - F(r). \end{aligned}$$

Thus the required identity holds and the conditional expectation is equal to the right side a.s. as claimed.

(c) Similarly, the claimed equality holds for $-\infty < s < u \le t < \infty$ if and only if

$$E\{1_A 1_{[X\ge u]}\} = E\left\{1_A 1_{[X>s]}\frac{1-F(u-)}{1-F(s)}\right\}$$

for all $A \in \mathcal{A}_s$. But by (a) it suffices to show this identity for $A \in \mathcal{C}_s$. Thus let $A = [X > r]$ for some $r \le s$. Then since $r \le s < u$ the left side is

$$E\{1_{[X>r]}1_{[X\ge u]}\} = E1_{[X\ge u]} = 1 - F(u-),$$

and the right side is

$$\begin{aligned} E\left\{1_{[X>r}1_{[X>s]}\frac{1-F(u-)}{1-F(s)}\right\} &= E\{1_{[X>s]}\}\frac{1-F(u-)}{1-F(s)} \\ &= (1-F(s))\frac{1-F(u-)}{1-F(s)} = 1 - F(u-). \end{aligned}$$

Thus the required identity holds and the conditional expectation is equal to the right side a.s. as claimed.
(d) To show that $\mathbb{M}_t$ is a martingale as claimed, it suffices to prove the identity (1) for sets $A \in \mathcal{C}_s$. Thus let $A = [X > r]$ for some $r \in [-\infty, s]$. Then the left side of (1) is given by

$$\begin{aligned} E\{1_{[X>r]}(\mathbb{N}_t - \mathbb{N}_s)\} &= E\{1_{[r<X\le t]} - 1_{[r<X\le s]}\} \\ &= F(t) - F(r) - (F(s) - F(r)) = F(t) - F(s). \end{aligned}$$

On the other hand, the left side is

$$\begin{aligned} & E\{1_{[X>r]}(\mathbb{A}_t - \mathbb{A}_s)\} \\ &= E\left\{1_{[X>r]} \int_{(s,t]} 1_{[X\ge v]} d\Lambda(v)\right\} \\ &= \int_{(s,t]} \left\{E(1_{[X>r]}1_{[X\ge v]})\right\} d\Lambda(v) \\ &= \int_{(s,t]} E1_{[X\ge v]} d\Lambda(v) = \int_{(s,t]} (1 - F(v-)) \frac{1}{1 - F(v-)} dF(v) \\ &= \int_{(s,t]} dF(v) = F(t) - F(s). \end{aligned}$$

Thus the required identity holds and $\{\mathbb{M}_t, \mathcal{A}_t\}$ is a martingale.

2. PfS, Exercise 13.1.5, page 353: Verify the claims made in example 1.11: Suppose that the rv's $\xi_1, \xi_2, \ldots$ are i.i.d. Uniform$(0,1)$, and let $\mathbb{N}_n(t) \equiv n\mathbb{G}_n(t) \equiv$ (the number of ξ_i's $\le t$). Show that

$$\begin{aligned} \mathbb{M}_n(t) &\equiv \mathbb{N}_n(t) - \int_0^t \frac{n(1 - \mathbb{G}_n(r-))}{1 - r} dr \\ &= \sqrt{n}\left\{\mathbb{U}_n(t) + \int_0^t \frac{\mathbb{U}_n(r-)}{1 - r} dr\right\} \end{aligned}$$

is a martingale with covariance $n(s \wedge t)$.

Solution: Let $\mathcal{A}_t^n \equiv \sigma\{1_{[\xi_i \le s]},\ 0 \le s \le t,\ i = 1, \ldots, n\}$. Now note that

$$\mathbb{M}_n(t) = \sum_{i=1}^n \left\{1_{[\xi_i \le t]} - \int_0^t \frac{1_{[\xi_i \ge r]}}{1 - r} dr\right\}$$

is a sum of the basic one-jump counting process martingales treated in the previous problem in the special case of $F(x) = x$ for $0 \le x \le 1$, and hence with hazard rate function $\lambda(x) = 1/(1-x)$ and cumulative hazard function $\Lambda(t) = \int_0^t (1-r)^{-1} dr$. Since the sum of martingales is also a martingale, the first equality of the claim follows. Alternatively, for $0 \le s < t < 1$,

$$\begin{aligned}
E\{\mathbb{M}_n(t) | \mathcal{A}_s^n\} &= \sum_{i=1}^n E\left\{ 1_{[0,t]}(\xi_i) - \int_0^t \frac{1_{[\xi_i \ge r]}}{1-r} dr \middle| \mathcal{A}_s^n \right\} \\
&= \sum_{i=1}^n \left\{ 1_{[0,s]}(\xi_i) - \int_0^s \frac{1_{[\xi_i \ge r]}}{1-r} dr \right\} \\
&\qquad \text{a.s., by the result of the previous problem} \\
&= \mathbb{M}_n(s).
\end{aligned}$$

The second equality of the claim follows since

$$\begin{aligned}
&\sqrt{n}\left\{ \mathbb{U}_n(t) + \int_0^t \frac{\mathbb{U}_n(r-)}{1-r} dr \right\} \\
&= n\left\{ (\mathbb{G}_n(t) - t) - \int_0^t \frac{1 - \mathbb{G}_n(r-) - (1-r)}{1-r} dr \right\} \\
&= \mathbb{N}_n(t) - \int_0^t \frac{n(1-\mathbb{G}_n(r-))}{1-r} dr - nt + n\int_0^t \frac{1-r}{1-r} dr \\
&= \mathbb{N}_n(t) - \int_0^t \frac{n(1-\mathbb{G}_n(r-))}{1-r} dr.
\end{aligned}$$

The last claim concerning the covariance follows from the calculation of the covariance of the basic process

$$\mathbb{M}_1(t) \equiv 1_{[\xi_1 \le t]} - \int_0^t 1_{[\xi_1 \ge r]} (1-r)^{-1} dr.$$

Since $\mathbb{M}_1$ is a martingale, for $0 \le s \le t < 1$,

$$\begin{aligned} E\{\mathbb{M}_1(t)\mathbb{M}_1(s)\} &= EE\{(\mathbb{M}_1(t) - \mathbb{M}_1(s) + \mathbb{M}_1(s))\mathbb{M}_1(s)|\mathcal{A}_s^n\} \\ &= E\{\mathbb{M}_1(s)E\{\mathbb{M}_1(t) - \mathbb{M}_1(s)|\mathcal{A}_s^n\}\} + E\{\mathbb{M}_1(s)^2\} \\ &= 0 + E\{\mathbb{M}_1(s)^2\} = E\{\langle\mathbb{M}_1\rangle(s)\} \\ &= E\left\{\int_0^s \frac{1_{[\xi_1 \ge r]}}{1-r} dr\right\} \\ &= \int_0^s \frac{1-r}{1-r} dr = s \end{aligned}$$

where we have used the fact that for a (one-jump) counting process

$$\langle\mathbb{M}_1\rangle_t = \int_{[0,t]} 1_{[X_1 \ge r]}(1 - \Delta\Lambda(r))d\Lambda(r).$$

Thus $Cov[\mathbb{M}_1(t), \mathbb{M}_1(s)] = s \wedge t$, and this yields, using the fact that $\mathbb{M}_n$ is the sum of n independent copies of $\mathbb{M}_1$,

$$Cov[\mathbb{M}_n(t), \mathbb{M}_n(s)] = n(s \wedge t).$$

3. PfS, Exercise 12.8.1, page 328: Let $X_0 \equiv 0$, and $X_1, \ldots$ be i.i.d. $(0, \sigma^2)$. Define $S_k \equiv X_1 + \cdots + X_k$ for each integer $k \ge 0$.
(a) Find the asymptotic distribution of $(S_1 + S_2 + \cdots + S_n)/c_n$ for an appropriate c_n.
(b) Determine a representation for the asymptotic distribution of the "absolute area" under the partial sum process, as given by $(|S_1| + \cdots + |S_n|)/c_n$.

Solution: (a) With $c_n = n^{3/2}$ we have

$$T_n \equiv n^{-3/2} \sum_{j=1}^n S_j = \frac{1}{n} \sum_{j=1}^n \mathbb{S}_n(j/n) = \int_0^1 \mathbb{S}_n(t)dt$$

where $\mathbb{S}_n(t) \equiv n^{-1/2} \sum_{i=1}^{[nt]} X_i$. Now for the Skorokhod construction $\|\mathbb{S}_n - \sigma\mathbb{S}\| \to_p 0$, so for T_n of the Skorokhod construction

$$T_n \to_p \int_0^1 \sigma\mathbb{S}(t)dt \sim N(0, \sigma^2/3)$$

since

$$\begin{aligned}
E\left(\int_0^1 \mathbb{S}(t)dt\right)^2 &= E\left(\int_0^1 \mathbb{S}(t)dt \int_0^1 \mathbb{S}(t')dt'\right) \\
&= \int_0^1 \int_0^1 E(\mathbb{S}(t)\mathbb{S}(t'))dtdt' \\
&= \int_0^1 \int_0^1 t \wedge t' dtdt' = 2\int_0^1 \left(\int_0^t t'dt'\right) dt = 2\int_0^1 \frac{1}{2}t^2 dt \\
&= \frac{1}{3}.
\end{aligned}$$

Thus $T_n \to_d N(0, \sigma^2/3)$ for T_n formed from any X_i's with mean 0 and variance σ^2.
(b) Again with $c_n = n^{3/2}$,

$$A_n \equiv n^{-3/2} \sum_{j=1}^n |\mathbb{S}_n(j/n)| = \int_0^1 |\mathbb{S}_n(t)|dt,$$

and hence for A_n of the Skorokhod construction,

$$A_n \to_p \int_0^1 |\sigma \mathbb{S}(t)|dt = \sigma \int_0^1 |\mathbb{S}(t)|dt,$$

and for any X_i's with mean 0 and variance σ^2 we have $A_n \to_d \sigma \int_0^1 |\mathbb{S}(t)|dt$. The distribution of $\int_0^1 |\mathbb{S}(t)|dt$ has been determined by Takács (1993). Computation of its Laplace transform(s) was carried out by Kac (1946) and Perman and Wellner (1996). See Janson (2007) for a survey of various Brownian area distributions and their applications.

4. (a) PfS, Exercise 12.10.1, page 338. Here is a rephrasing: suppose that $Y_1, \ldots, Y_{n+1}$ are i.i.d. exponential(1) random variables, and set $T_k = Y_1 + \cdots + Y_k$ for $k = 1, 2, \ldots, n+1$. Suppose that $\xi_1, \xi_2, \ldots$ are i.i.d. Uniform$(0,1)$ and let $0 \le \xi_{n:1} \le \xi_{n:1} \le \ldots \le \xi_{n:n} \le 1$ be the order statistics of $\xi_1, \ldots, \xi_n$. Show that

$$\underline{U}_n \equiv \left(\frac{T_1}{T_{n+1}}, \ldots, \frac{T_n}{T_{n+1}}\right) \stackrel{d}{=} (\xi_{n:1}, \ldots, \xi_{n:n}) \equiv \underline{\xi}_{(n)}.$$

(b) Show that the construction in (a) for a fixed n is not correct jointly in n: i.e. show that

$$(\underline{U}_n, \underline{U}_{n+1}) \stackrel{d}{\neq} (\underline{\xi}_{(n)}, \underline{\xi}_{(n+1)}).$$

Hint: consider $n = 1$.

Solution: (a) The joint density of $Y_1, \ldots, Y_{n+1}$ is given by

$$p_{\underline{Y}}(\underline{y}) = \exp(-\sum_1^{n+1} y_j) 1_{[0,\infty)}(y_{(1)}).$$

Thus the joint density of the partial sums $\underline{T} = (T_1, T_2, \ldots, T_{n+1})$ is, by an easy calculation,

$$\begin{aligned} p_{\underline{T}}(\underline{t}) &= p_{\underline{Y}}(\underline{y}(\underline{t}))\left|\frac{\partial y}{\partial t}\right| \\ &= \exp(-t_{n+1}) 1_{[0 \le t_1 \le t_2 \le \cdots \le t_{n+1}]} \end{aligned}$$

since $y_j = t_j - t_{j-1}$ for $j = 1, \ldots, n+1$ with $t_0 \equiv 0$ so that the matrix involved in the Jacobian is lower triangular with 1's on the diagonal, -1's just below the diagonal, and 0's elsewhere, so the Jacobian is 1. It follows that the joint density of $(\underline{U}, T_{n+1})$ with $\underline{U} \equiv (T_1, \ldots, T_n)/T_{n+1}$ is given by

$$\begin{aligned} p_{\underline{U}, T_{n+1}}(\underline{u}, t_{n+1}) &= p_{\underline{T}}(u_1 t_{n+1}, \ldots, u_n t_{n+1}, t_{n+1}) \cdot t_{n+1}^n \\ &= t_{n+1}^n \exp(-t_{n+1}) 1_{[0 \le u_1 \le \ldots \le u_n \le 1,\ 0 \le t_{n+1} < \infty]}. \end{aligned}$$

Computing the marginal density of $\underline{U} = (U_1, \ldots, U_n)$ yields

$$\begin{aligned} p_{\underline{U}}(\underline{u}) &= \int_0^\infty p_{\underline{U}, T_{n+1}}(\underline{u}, t_{n+1}) dt_{n+1} \\ &= 1_{[0 \le u_1 \le \ldots \le u_n \le 1]} \int_0^\infty t_{n+1}^n \exp(-t_{n+1}) dt_{n+1} \\ &= n! 1_{[0 \le u_1 \le \ldots \le u_n \le 1]} \end{aligned}$$

since $\int_0^\infty r^n e^{-r} dr = \Gamma(n+1) = n!$. Thus $\underline{U} \stackrel{d}{=} (\xi_{n:1}, \ldots, \xi_{n:n})$.
(b) Note that for $n = 1$,

$$(\underline{U}_1, \underline{U}_2) = \left(\frac{Y_1}{Y_1 + Y_2}; \frac{Y_1}{Y_1 + Y_2 + Y_3}, \frac{Y_1 + Y_2}{Y_1 + Y_2 + Y_3} \right)$$

so that the representation of $\xi_{2:1}$ is

$$\frac{Y_1}{Y_1 + Y_2} \cdot \frac{Y_1 + Y_2}{Y_1 + Y_2 + Y_3} <_{a.s.} \frac{Y_1}{Y_1 + Y_2},$$

while the representation of $\xi_{2:2}$ is

$$\begin{aligned} & \frac{Y_1+Y_2}{Y_1+Y_2+Y_3} \\ & \quad = \frac{Y_1}{Y_1+Y_2} \cdot \frac{Y_1+Y_2}{Y_1+Y_2+Y_3} + \frac{Y_2}{Y_1+Y_2+Y_3} \\ & \quad \neq \frac{Y_1}{Y_1+Y_2}. \end{aligned}$$

On the other hand, for the order statistics of ξ_1, ξ_2 i.i.d. Uniform$(0,1)$, if $\xi_1 < \xi_2$, then at sample size $n = 1$ we have $\xi_{1:1} = \xi_1$, while at sample size $n = 2$ we have $(\xi_{2:1}, \xi_{2:2}) = (\xi_1, \xi_2)$; i.e. the value $\xi_{1:1}$ appears among the values $(\xi_{2:1}, \xi_{2:2})$; in particular as $\xi_{2:1}$ in this case.

In general in going from $n-1$ to n all the values $\xi_{n-1:i}$ appear in the list of $\xi_{n:j}$'s, but one new value gets added, namely ξ_n which divides one of the spacings intervals for sample size $n-1$ into two sub-intervals. On the other hand, the construction in terms of exponential random variables involves a shortening of all the spacing intervals at sample size $n-1$, and adding a new spacing interval from T_n/T_{n+1} to 1 of length Y_{n+1}/T_{n+1}.

References

Janson, S. (2007). Brownian excursion area, Wright's constants in graph enumeration, and other Brownian areas. *Probab. Surv.* **4** 80–145 (electronic).

Kac, M. (1946). On the average of a certain Wiener functional and a related limit theorem in calculus of probability. *Trans. Amer. Math. Soc.* **59** 401–414.

Perman, M. and Wellner, J. A. (1996). On the distribution of Brownian areas. *Ann. Appl. Probab.* **6** 1091–1111.

Takács, L. (1993). On the distribution of the integral of the absolute value of the Brownian motion. *Ann. Appl. Probab.* **3** 186–197.