Statistics 620
Final exam, Winter 2017

1. Each of n statistics PhD students is carrying out an independent computationally intensive research project. However, the students have access to only a single shared computer. Each student works for an exponentially distributed time with mean $1/\lambda$, after which time they submit job on the computer that requires an exponentially distributed amount of computer time with mean $1/\mu$. Only one job can run at a time on the computer, and jobs run in the order they are submitted. The students do no work after submitting a job until they get the results back from the computer, at which point they start another similarly distributed cycle of work and computation. Find the long run probability that exactly k of the n students are working at any given time.

Solution: Let $X(t)$ be the number of students working at time t. We have $\nu_{k,k+1} = \mu$ for $k = 0, \ldots, n-1$, since if there is at least one job running then jobs complete at rate μ. Also, $\nu_{k,k-1} = \lambda k$, for $k = 1, \ldots, n$. This is a birth-death process, therefore time-reversible. The long run probability of k students working is equal to the stationary distribution, P_k, which satisfies the detailed balance equations,

$$\mu P_{k-1} = \lambda k P_k k, \tag{1}$$

so

$$P_k = \frac{\mu^k}{\lambda^k k!} P_0 = \frac{\mu^k}{\lambda^k k!} \left\{ \sum_{\ell=0}^{n} \frac{\mu^\ell}{\lambda^\ell \ell!} \right\}^{-1} \tag{2}$$

2. We study a simple model for public health control of an Ebola outbreak. One infected individual arrives in New York City at day zero of the outbreak. Each subsequent day, the individual either infects a new person (with probability p) or the individual becomes symptomatic and is discovered by public health officials (with probability $1-p$). Each newly infected individual behaves like the first. Find the cumulative distribution function of the time until the outbreak is discovered.

Solution: Let X_n be the number of undiscovered infected individuals on day n. Let D be the day of discovery of the outbreak. Conditional on $D > n$, $X_n = 2^n$. So, $D > n$ is exactly the event that all $\sum_{k=0}^{n} 2^k = 2^{n+1} - 1$ infected person days at time $k \leq n$ lead to infections not discoveries. Thus,

$$\mathbb{P}[D > n] = p^{2^{n-1}-1} \tag{3}$$

and D has c.d.f

$$F_D(n) = \mathbb{P}[D \leq n] = 1 - p^{2^{n-1}-1}. \tag{4}$$

3. Let S and T be stopping times for a sequence of random variables, $X_1, X_2, \ldots$. Show that the following are all stopping times:

(a) The minimum, $U = S \wedge T$.

(b) The maximum, $V = S \vee T$.

(c) The sum, $W = S + T$.

Solution: The definition of T being a stopping time for $X_0, X_1, X_2, \ldots$ is that $\{T = n\}$ is determined by $X_1, X_2, \ldots, X_n$ for all n. This is equivalent to $\{T > n\}$ being determined by $X_1, X_2, \ldots, X_n$ for all n, or to $\{T \leq n\}$ being determined by $X_1, X_2, \ldots, X_n$ for all n.

(a) We write $\{U > n\} = \{S > n\} \cap \{T > n\}$ and notice that both $\{S > n\}$ and $\{T > n\}$ are determined by $X_1, X_2, \ldots, X_n$ implying that their intersection is also.

(b) Similarly, $\{V \leq n\} = \{S \leq n\} \cap \{T \leq n\}$ and both $\{S > n\}$ and $\{T > n\}$ are determined by $X_1, X_2, \ldots, X_n$.

(c) $\{W = n\} = \bigcup_{k=0}^{n} \{S = k\} \cap \{T = n - k\}$ and, by the definition of stopping times, all terms in the union are determined by $X_1, \ldots, X_n$.

4. We use martingales to study a simple investment management model. Let Z_n be wealth at time n. At each timepoint $n \geq 1$, we allocate a fraction F of our wealth to a risky investment, modeled as

$$Z_n = \begin{cases} Z_{n-1}(1+F) & \text{with probability } p, \\ Z_{n-1}(1-F) & \text{with probability } 1-p, \end{cases}$$

for $1/2 < p < 1$. An investment strategy involves determining F as a function of $Z_0, \ldots, Z_{n-1}$. We want to choose a stragegy to maximize the expected interest rate, $\mathbb{E}[\log(Z_N/Z_0)]$ where N is a fixed integer time and Z_0 is a known constant.

(a) Find the investment strategy optimizing $E[\log(Z_n/Z_{n-1})|Z_{n-1}]$.

(b) Argue that $\log Z_n - \alpha n$ is a supermartingale for any investment strategy, where $\alpha = p\log p + (1-p)\log(1-p) + log 2$. Use part (a) to show that there is a strategy for which $\log Z_n - n\alpha$ is a martingale.

(c) Explain how (a) and (b) determine a strategy optimizing $\mathbb{E}[\log(Z_N/Z_0)]$.

<u>Solution</u>:

$$\mathbb{E}[\log Z_n | Z_{n-1}, F] = \log Z_{n-1} + p\log(1+F) + (1-p)\log(1-F). \tag{5}$$

Setting the derivative with respect to F equal to zero gives,

$$\begin{aligned} \frac{p}{1+F} - \frac{(1-p)}{1-F} &= 0, \\ F &= 2p - 1. \end{aligned} \tag{6}$$

Thus, investing a fraction $2p - 1$ gives the best expected interest rate over one time step. In principle, F could depend on $Z_1, \ldots, Z_{n-1}$ but the optimal choice doesn't. For this choice, (5) gives

$$\begin{aligned} \mathbb{E}[\log Z_n | Z_{n-1}, F] &= \log Z_{n-1} + p\log(2p) + (1-p)\log(2(1-p)) \\ &= \log Z_{n-1} + p\log(p) + (1-p)\log(1-p) + \log 2 \\ &= \log Z_{n-1} + \alpha. \end{aligned} \tag{7}$$

This demonstrates that $\{\log Z_n - \alpha n\}$ is a martingale for this optimal strategy. For any other strategy, $\mathbb{E}[\log Z_n | Z_{n-1}, F]$ cannot be greater so $\{\log Z_n - \alpha n\}$ is a supermartingale. A supermartingale has non-increasing expectation, so $\mathbb{E}[\log(Z_N/Z_0)]$ is maximized for the strategy (6) that has $\{\log Z_n - \alpha n\}$ a martingale, i.e., $\mathbb{E}[\log(Z_N/Z_0)] = N\alpha$.

5. Let $\{B_k(t), 1 \leq k \leq n, t \geq 0\}$ be a collection of n independent standard Brownian motions. Let $X(t) = \sum_{k=1}^{n} [B_k(t)]^2$.

(a) Find $\lim_{h\to 0} h^{-1}\mathbb{E}[X(t+h)-X(t)|X(t)=x]$ and $\lim_{h\to 0} h^{-1}\mathrm{Var}[X(t+h)-X(t)|X(t)=x]$.

(b) Argue that $\{X(t)\}$ is a diffusion process and write down the stochastic differential equation that it solves.

(c) Let A_t be the event that $X(s)$ hits zero for some $s \geq t$. What do you think $\lim_{t\to\infty}\mathbb{P}(A_t)$ is, for each value of n? You may use results we established in class for the random walk to guide your reasoning.

Solution: (a) Define $\Delta = X(t+\delta) - X(t)$ and $\Delta_k = B_k(t+\delta) - B_k(t)$. We calculate,

$$\begin{aligned}\Delta &= \sum_{k=1}^{n} B_k(t+\delta)^2 - B_k(t)^2 \\ &= \sum_{k=1}^{n} \Big\{ \big[B_k(t) + \Delta_k\big]^2 - B_k(t)^2 \Big\} \\ &= \sum_{k=1}^{n} \Delta_k^2 + 2\sum_{k=1}^{n} \Delta_k B_k(t).\end{aligned}$$

Since $\Delta_k \sim N[0,\delta]$ and $\{\Delta_k\}$ and $\{B_k(t)\}$ are independent, we have

$$\mathbb{E}[\Delta|B_1(t),\ldots,B_k(t)] = n\delta \tag{8}$$

$$\begin{aligned}\mathrm{Var}[\Delta|B_1(t),\ldots,B_k(t)] &= \sum_{k=1}^{n}\Big[\mathrm{Var}(\Delta_k^2) + 4\mathrm{Var}\big(\Delta_k B_k(t)|B_k(t)\big) + 4\mathrm{Cov}\big(\Delta_k^2, \Delta_k B_k(t)|B_k(t)\big)\Big] \\ &= 2\delta^2 n + 4\delta\sum_{k=1}^{n} B_k^2(t).\end{aligned} \tag{9}$$

Now, using the tower property, we obtain

$$\mathbb{E}[\Delta|X(t)] = n\delta \tag{10}$$

$$\mathrm{Var}[\Delta|X(t)] = 2\delta^2 n + 4\delta X(t). \tag{11}$$

Hence, $\lim_{h\to 0} h^{-1}\mathbb{E}[X(t+h)-X(t)|X(t)=x] = n$ and $\lim_{h\to 0} h^{-1}\mathrm{Var}[X(t+h)-X(t)|X(t) = x] = 4x$.

(b) $\{X(t)\}$ evidently has continuous sample paths, but it is not immediately evident that it has the Markov property. Indeed, usually a non-invertible function of diffusions is non-Markovian. Here, it happens that the infinitesimal mean and variance depend only on the current state, so we have the Markovian property. Thus, $\{X(t)\}$ is a diffusion.

(c) The event $\{X(s)=0\}$ is equivalent to the event $B_1(s) = B_2(s) = \cdots = B_n(s) = 0$, which is a return to the origin for an n-dimensional Brownian motion. We know from class and homework that a simple random walk in $\mathbb{R}^d$ is null recurrent when $d \leq 2$ and transient when $d \geq 3$. Since Brownian motion can be constructed as a limit of simple random walks, we may expect that

$$\lim_{t\to\infty}\mathbb{P}(A_t) = \begin{cases} 1 & \text{when } n = 1,2 \\ 0 & \text{when } n \geq 3 \end{cases} \tag{12}$$