Stat 515 Final, Penn State Statistics

Sketch of Solutions

May 4, 2015.

1. (a) A certain town never has two sunny days in a row. Each day is classified as either being sunny, cloudy (but dry), or rainy. If it is sunny one day, then it is equally likely to be either cloudy or rainy the next day. If it is rainy one day, then there is a 50% chance it will be the same the next day and there is a 25% chance each that it will be cloudy or sunny the next day. If it is cloudy one day, then there is a 50% chance that it will be the same the next day and there is a 25% chance it will be rainy or sunny the next day. In the long run, what proportion of days are sunny and what proportion of days are cloudy? [3pts]

   (b) Consider a Markov chain with transition matrix with $p_1, p_2 \in (0, 1)$:

$$P = \begin{bmatrix} 0 & p_1 & p_1 \\ p_2 & p_1 & p_2 \\ p_2 & p_2 & p_1 \end{bmatrix}$$

   Does this Markov chain have a limiting distribution (you do not have to find this limiting distribution)? If so, carefully show how the conditions for this theoretical result are satisfied. If not, state any conditions that are violated. [4pts]

Soln (a) The transition probability matrix associated with the Markov chain $\{X_n\}$ where X_n is the weather on the nth day and $X_n = 0, 1, 2$ when the day is sunny, cloudy and rainy respectively is:

$$P = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$$

The equations for the long run proportions are:

$$\begin{aligned} \pi_0 &= \frac{1}{4}\pi_1 + \frac{1}{4}\pi_2 \\ \pi_1 &= \frac{1}{2}\pi_0 + \frac{1}{2}\pi_1 + \frac{1}{4}\pi_2 \\ \pi_2 &= \frac{1}{2}\pi_0 + \frac{1}{4}\pi_1 + \frac{1}{2}\pi_2 \\ &\pi_0 + \pi_1 + \pi_2 = 1 \text{ (redundant equation).} \end{aligned}$$

Solution to the above yields: $\pi_0 = 1/5, \pi_1 = 2/5, \pi_2 = 2/5$. Hence long run proportion of sunny and cloudy days are $1/5$ and $2/5$ respectively.

(b) Yes, it has a limiting distribution. Since 0 and 1 communicate ($P_{01} > 0, P_{10} > 0$) and 0 and 2 communicate ($P_{02} > 0, P_{20} > 0$), all 3 states communicate. The chain is therefore *irreducible.*

Since it is irreducible and the state spate space is finite, all states are *positive recurrent.*
$P_{11} > 0$ so the period of state 1 is 1. Since all states are in the same class, they all have the same period, and the chain is *aperiodic.*
Thus, the chain is irreducible, positive recurrent and aperiodic, which means it has a limiting distribution. Note that this Markov chain is a generalization of the one from part (a).

2. Suppose n lightbulbs are placed in a room and switched on at time 0. Assume the lifetime of each bulb is independently distributed according to Exponential(θ) with $\theta > 0$, i.e., they have expected lifetimes of θ. Suppose you walk into the room at some time $\tau > 0$ and the only information you have is the number of bulbs still working, W.

   (a) Suppose the time you walk into the room is random, with $\tau \sim U(a, b)$, $b > a > 0$. What is the expected value of W ? [3pts]

   (b) Now assume τ is fixed (but the rest of the description of the experiment stays the same as above.) What is the expected *total* lifetime of all the bulbs that are still working at time τ? [3pts]

   Soln: (a) $W \mid \tau \sim$Binomial$(n, e^{-\tau/\theta})$, so $E(W) = E(E(W \mid \tau)) = nE(e^{-\tau/\theta})$. Using the pdf of τ, we get $E(W) = n\theta(e^{-a/\theta} - e^{-b/\theta})$.
   (b) Since τ is fixed, a bulb still working at time τ has lifetime $= \tau + X$ where $X \sim$ Exp(θ) by memoryless property of exponential random variables. Hence a bulb still working at time τ has expected lifetime $= \tau + \theta$. Let total lifetime of still working bulbs be T. Expected total lifetime of bulbs still working at time $\tau = E(E(T \mid W)) = E(W(\tau + \theta)) = ne^{-\tau/\theta}(\tau + \theta)$.

3. Consider a (time-homogeneous) Poisson process $\{N(t),\ t > 0\}$ with $\lambda = 5$.

   (a) What is the probability of seeing 10 events in the time interval (0,2) *and* 5 events in the time interval (1,2) ? [3pts]

   (b) Suppose you know 10 events occurred in the interval (0,2). What is the distribution of the occurrence of the very first event? [3pts]

   (c) Show that this process satisfies the Markov assumption for continuous-time discrete state space Markov processes. Hint: You can use the connection between Poisson processes and exponential random variables (and properties of exponential random variables) to show this. [3pts]

Soln: (a) (b) Occurrence of time of first event has the same distribution as that of the minimum of 10 i.i.d. Unif(0,2) random variables.
(c) First recall that the Markov assumption for a continuous-time discrete state space Markov process $\{X(t)\}$ with $i, j \in \Omega$ is:

$$P(X(t+s) = j \mid X(s) = i, X(u) = x(u), u \in [0, s)) = P(X(t+s) = j \mid X(s) = i).$$

Note that if the process is time homogeneous,

$$P(X(t + s) = j \mid X(s) = i) = P(X(t) = j \mid X(0) = i).$$

Now define T_i as the time the process leaves state i given that it was at i at time 0. (Note: for the Poisson process described here, it can only move from i to $j = i + 1$). By memoryless property of exponentials, we know that for $s, t > 0$,

$$P(T_i > s + t \mid T_i > s) = P(T_i > t). \tag{1}$$

But,

$$P(T_i > t) = P(N(u) = i, u \in (0, t] \mid N(0) = i), \tag{2}$$

and

$$P(T_i > s + t \mid T_i > s) = P(N(u) = i, s < u \leq s + t \mid N(u) = i, u \in [0, s]). \tag{3}$$

Hence, from (1),(2),(3), it is clear that

$$P(N(u) = i, s < u \leq s+t \mid N(u) = i, u \in [0, s]) = P(N(u) = i, u \in (0, t] \mid N(0) = 0),$$

thereby satisfying the Markovian assumption.
(Students who stated the Markov assumption for continuous-time discrete state space Markov processes received partial credit. An answer that informally stated that the exponential waiting times/memoryless property led to the Markov property received most of the credit. A more detailed argument along the lines above resulted in full credit.)

4. Helicopters land at a small hangar (a place for storing aircraft) at the Poisson rate of 20 per day. However, they will only land if there are either 0 or 1 helicopters (including the one being worked on) at the hangar. That is, they will not land if there are 2 or more helicopters at the hangar. The hangar can only work on one helicopter at a time. Assume that the amount of time required to service a helicopter is exponentially distributed with a mean of 2 hours.

   (a) What is the generator matrix (following class notation this is denoted by G) for this continuous-time Markov chain? [3pts]

   (b) Does this process satisfy the detailed balance condition ? Justify your answer. [3pts]

   (c) Every hour without any helicopters results in a loss of $\$1,000$ for the hangar, while every hour with at least one helicopter results in a profit of $\$5,000$. In the long run, how much profit can the hangar expect to make per hour? (Note: you do not have to simplify your final answer.) [3pts]

Soln: (a) G, for this process.

$$G = \begin{bmatrix} -20 & 20 & 0 \\ 12 & -32 & 20 \\ 0 & 12 & -12 \end{bmatrix},$$

(b) We can find π such that detailed balance is satisfied, so:

$$\pi_1 = \frac{5}{3}\pi_0, \;\; \pi_2 = \frac{5}{3}\pi_1 = \left(\frac{5}{3}\right)^2 \pi_0.$$

Since $\sum_{i=0}^{2} \pi_i = 1$, $\pi_0 = (1 + \frac{5}{3} + \left(\frac{5}{3}\right)^2)^{-1} = \frac{9}{49}$, and $\pi_1 = \frac{15}{49}, \pi_2 = \frac{25}{49}$, so answer $= \pi_1 + \pi_2 = \frac{40}{49}$. Equivalently, we could solve part (c) and plug π in to show detailed balance holds.

(c) To get long run proportion of times spent in each state, let $\pi = (\pi_0, \pi_1, \pi_2)$. π that satisfies detailed balance also solves $\pi G = 0$ (or, equivalently, one could directly solve $\pi G = 0$). Hence, long run proportion of time spent in state 0 (no helicopters) $= \frac{9}{49}$, and long run proportion of time spent in states 1 or 2 is $\frac{40}{49}$. So expected profit per hour = -1000($\frac{9}{49}$) + 5000$\frac{40}{49}$ = \$3897.959.

5. Let $\{X(t), t \geq 0\}$ be standard Brownian motion. That is, $X(0) = 0$, every increment $X(s+t) - X(s)$ is $\mathrm{N}(0, t)$, and for every set of n disjoint time intervals, the increments are independent random variates.

   (a) Show that $\{X(t), t \geq 0\}$ is an example of a continuous-time continuous-state-space Markov process. [4pts]

   (b) Show that the stochastic process $\{Y(t), t \geq 0\}$ with $Y(t) = X^2(t) - t$ is a martingale. [4pts]

   (c) Define $\{Z(t), t \geq 0\}$ as $Z(t) = \exp(\lambda X(t) - \lambda^2 t/2)$, λ is a constant. $\{Z(t), t \geq 0\}$ is known to be a martingale (you do not have to prove this). Let T be the first time that $X(t)$ reaches 2-4t, that is, $T = \min\{t : X(t) = 2 - 4t\}$. What is $E(T)$? Fully justify your answer. [4pts]

Solutions: (a) Need to show that $P(X(t) > x \mid X(u), u \in [0, s]) = P(X(t) > x \mid X(s))$. You can argue this by using the fact that Brownian motion has independent increments. The fact that it is a continuous-time and continuous state space process follows from the definition of Brownian motion.
(b) $E(Y(t) \mid X(u), u \in [0, s]) = E(X^2(t) \mid X(u), u \in [0, s]) - t = E(X^2(t) \mid X(s)) - t$ from Markov property
Now, $E(X^2(t) \mid X(s)) - t = E((X(t) - X(s) + X(s))^2 \mid X(s)) - t$, and using the fact that $X(t) - X(s)$ is independent of $X(s)$ (independent increments), it follows that $E((X(t) - X(s) + X(s))^2 \mid X(s)) - t = X^2(s) - s = Y(s)$. Thus, $Y(s)$ is a martingale.
(c) Since T is bounded and does not depend on the future, it follows that we can the optional sampling theorem to calculate this.
By optional sampling theorem, $E(X(T)) = E(X(0)) = 0$. But $X(T) = 2 - 4T$ and hence $2 - 4E(T) = 0$. That is, $\mathrm{E}(T) = 1/2$.
(This problem originally had another part which I later deleted; hence the information about $Z(t)$ ends up actually not being necessary for this problem at all. Only need $X(t)$ to be a martingale.)

6. Consider a regression of a variable Y on X where the regression model is as follows, $Y_i \sim EMG(\beta_0 + \beta_1 X, \sigma, \lambda)$, where the exponentially modified Gaussian random variable, $\text{EMG}(\mu, \sigma, \lambda)$, has pdf $f(x; \mu, \sigma, \lambda) = \frac{\lambda}{2}\exp(\frac{\lambda}{2}(2\mu + \lambda\sigma^2 - 2x))\text{erfc}\left(\frac{\mu+\lambda\sigma^2-x}{\sqrt{2}\sigma}\right)$, and erfc is the complementary error function defined as

$$\text{erfc}(x) = \frac{2}{\pi}\int_x^\infty e^{-t^2}dt.$$

Assume that σ is known. Let the independent priors for $\beta_0, \beta_1, \lambda$ be $p(\beta_0), p(\beta_1), p(\lambda)$ respectively.

(a) Provide pseudocode for a Metropolis-Hastings algorithm to construct a Markov chain with stationary distribution $\pi(\beta_0, \beta_1, \lambda \mid \mathbf{Y}, \mathbf{X})$ for a data set of size n, $(X_1, Y_1), \ldots, (X_n, Y_n)$. $\mathbf{Y}, \mathbf{X}$ are $(Y_1, \ldots, Y_n)$, $(X_1, \ldots, X_n)$ respectively. You should provide enough detail so anyone reading it should be able to write code based on your description. You do not have to provide starting values or specific tuning parameters since those will depend on the particulars of the data. You should, however, list at the beginning of the algorithm any/all tuning parameters for the algorithm that you will have to adjust in order to make it work well. [5pts]

(b) Suppose your Markov chain is $(\beta_0^{(1)}, \beta_1^{(1)}, \lambda^{(1)}), \ldots (\beta_0^{(n)}, \beta_1^{(n)}, \lambda^{(n)})$. Provide estimators for (i) $E_\pi(\lambda)$, and (ii) $E_\pi\left(\frac{1}{\beta_1+\lambda}\right)$. [2pts]

(c) State the theoretical result that justifies the use of the above estimator of $E_\pi(\lambda)$. State sufficient conditions for the theorem to hold (you do not have to prove that these conditions hold). [3pts]

(d) You are worried about the influence of the priors on the posterior so you would like to see how the posterior changes if the prior is modified to $p^*(\beta_0), p^*(\beta_1), p^*(\lambda)$, independent of each other. Describe in detail how you would use the Markov chain above (do not construct a new Metropolis-Hastings algorithm) to approximate $E_{\pi^*}(\lambda)$ where π^* is the new posterior pdf. Briefly explain when your approach is likely to work well and when it will not. [4pts]

(e) Now suppose that σ is also assumed to be unknown, and has prior $p(\sigma)$. Describe all the ways in which your MCMC algorithm from part (a) will change when you now have to approximate expectations with respect to $\pi(\beta_0, \beta_1, \lambda, \sigma \mid \mathbf{Y}, \mathbf{X})$. [3pts]

Soln: (a) For full credit, students would have to write out the posterior distribution of interest, then provide a clear description of either a variable-at-a-time or all-at-once Metropolis-Hastings algorithm. If variable-at-a-time, full conditional distributions are also required.
(b) Usual Monte Carlo estimators (sample means) using the Markov chain states.
(c) Strong Law of Large Numbers. Assumptions: Harris-ergodic Markov chain with the usual finite moment condition.
(d) Describe importance sampling. Potential problem: if the new posterior distribution is heavier-tailed than the previous posterior.
(e) You would simply add one more update to the M-H algorithm, updating the full conditional distribution of σ. The other updates would change because they would now use the most up to date value of σ in the Markov chain.