EECS 126 Probability and Random Processes University of California, Berkeley: Spring 2015
Abhay Parekh April 7, 2015

Midterm Exam

Last name	First name	SID

Rules.

- You have 80 mins (5:10pm - 6:30pm) to complete this exam.
- The exam is not open book, but you are allowed one sheet of handwritten notes; calculators will be allowed.
- No form of collaboration between the students is allowed. If you are caught cheating, you may fail the course and face disciplinary consequences.

Please read the following remarks carefully.

- Show all work to get any partial credit.
- Take into account the points that may be earned for each problem when splitting your time between the problems.

Problem	Points earned	out of
Problem 1		25
Problem 2		20
Problem 3		25
Problem 4		30
Total		100

Problem 1[25]

(a) [6] The following questions pertain to a finite state MC that consists of a single periodic class (i.e, period > 1): explain your answer — a T/F answer won't get credit.

(i) [2] Do all the states have to be recurrent?
Solution: The Markov chain is finite, so there has to be at least one recurrent state (i.e. a state which the Markov chain visits infinitely often). As the Markov chain has a single class, this means that all the states must be recurrent.

(ii) [2] Is it possible for any of the states to have self transitions?
Solution: If any state has self transitions, then the period of the Markov chain would be 1. Since the period is given to be > 1, there can be no self transitions.

(iii) [2] What does the solution to the balance equations tell you about the states of the MC?
Solution: The solution to the balance equations tells us what fraction of time the Markov chain spends in each state on an average.

(b) [7] A four state Discrete time MC with states $0, 1, 2, 3$ has the transition matrix

$$P = \begin{pmatrix} 0.2 & 0.3 & 0.5 & 0 \\ 0.2 & 0.1 & 0.3 & 0.4 \\ 0.4 & 0.6 & 0 & 0 \\ 0.1 & 0.9 & 0 & 0 \end{pmatrix}$$

Argue that this MC consists of a single recurrent class. Assume that it is in steady state after running for $n > 500$ steps. Find the following in terms of π_i's. (You don't have calculate the steady state distribution!)

(i) [3] $P(X_{1002} = 2 | X_{1000} = 0)$
Solution: Note that from state 1 there is a positive probability of jumping to all other states, and there is a positive probability of jumping back to state 1 from all other states. Thus the Markov chain has a single recurrent class.
Let p_{ij} be the (i, j)-th element in the transition matrix.

$$\begin{aligned} P(X_{1002} = 2 | X_{1000} = 0) &= p_{00}p_{02} + p_{01}p_{12} + p_{02}p_{22} + p_{03}p_{32} \\ &= (0.2 \times 0.5) + (0.3 \times 0.3) + (0.5 \times 0) + (0 \times 0) \\ &= 0.19 \end{aligned}$$

(ii) [4] $P(X_{1000} = 0 | X_{1002} = 2)$
Solution:

$$\begin{aligned} P(X_{1000} = 0 | X_{1002} = 2) &= \frac{P(X_{1000} = 0, X_{1002} = 2)}{P(X_{1002} = 2)} \\ &\approx \frac{\pi(0) \times P(X_{1002} = 2 | X_{1000} = 0)}{\pi(2)} \\ &= \frac{0.19\pi(0)}{\pi(2)}. \end{aligned}$$

(c) [6] The freeway system consists of a very long entry ramp and a road. A pacing light regulates the entry of cars from ramp to road. There are multiple exits.

(i) [3] Cars enter a freeway system at a rate of 100 cars/minute according to some unknown distribution, and the average time a car is in the freeway system is 22 mins. On average how many cars are either in the entry ramp or on the road?
Solution: Using Little's law for the entire freeway system, the average number of cars in the freeway system is simply $22 \times 100 = 2200$.

(ii) [3] On average a car waits for 2mins in the ramp to enter the freeway. How many cars on average are waiting on the ramp?
Solution: Using Little's law for the freeway ramp, the average number of cars on the ramp is $2 \times 100 = 200$.

(d) [6] $X_1, X_2, ...$ are i.i.d. Bernouli random variables. $P(X_i = 1) = \delta$. Let $S_n = X_1 + X_2 + ... + X_n$, and let m be some fixed positive integer. What is

$$\lim_{n\to\infty} \sum_{i=n\delta-m}^{n\delta+m} P(S_n = i)$$

Explain your answer.
Solution: We first note that

$$\sum_{i=n\delta-m}^{n\delta+m} P(S_n = i) = P(|S_n - n\delta| \leq m).$$

By the Central Limit Theorem, the distribution of $\frac{S_n - n\delta}{\sqrt{n\delta(1-\delta)}}$ tends to $\mathcal{N}(0,1)$. Let $Z \sim \mathcal{N}(0,1)$. We can re-write the required probability as

$$P(|S_n - n\delta| \leq m) = P\left(\frac{|S_n - n\delta|}{\sqrt{n\delta(1-\delta)}} \leq \frac{m}{\sqrt{n\delta(1-\delta)}}\right).$$

Taking the limit as n tends to infinity,

$$\lim_{n\to\infty} P\left(\frac{|S_n - n\delta|}{\sqrt{n\delta(1-\delta)}} \leq \frac{m}{\sqrt{n\delta(1-\delta)}}\right) = P(Z = 0) = 0.$$

Thus, the required limit is 0.

An less rigorous way of arriving at the same conclusion is to observe that S_n is a Binomial with mean $n\delta$ and variance $n\delta(1-\delta)$. As $n \to \infty$ the distribution is more spread out, since variance $\to \infty$ as well. Thus the amount of probability mass concentrated within m of the mean will decrease to zero.

Problem 2[20]

Bob gets three kinds of email: Critical, Not Critical and Spam. Each of these emails arrives as independent Poisson Processes with rates: λ_c, λ_{nc} and λ_s per hour respectively.

Bob's email spam filter classifies incoming emails and places each into one of two folders: Inbox and Spam. The filter attempts to place all emails that are Critical and Not Critical in the Inbox folder and to place the Spam emails in the Spam folder. Let p_c be the probability that the filter classifies a critical email correctly and places it in the inbox folder and $1-p_c$ be the probability that it is placed in the spam folder. Similarly let p_{nc} be the probability that a non-critical email is placed in the inbox, and p_s be the probability that a spam email is placed in the spam folder.

(a) [10] Suppose that at some point in time that there are 3 emails in the Inbox folder. What is the probability that all three are Spam?
Solution: Arrival of emails in Bob's Inbox folder is a merging of three independent Poisson processes with rates $\lambda_c p_c$, $\lambda_{nc} p_{nc}$, and $\lambda_s(1-p_s)$. Thus, an email in the inbox is spam with probability

$$p = \frac{\lambda_s(1-p_s)}{\lambda_c p_c + \lambda_{nc} p_{nc} + \lambda_s(1-p_s)}.$$

The probability that all three emails are spam is just p^3.

(b) [10] The probability that Bob will forward a Critical email in his Inbox to his colleagues is p_{cf}, and the probability that he will forward a Not Critical email in his Inbox to his colleagues is p_{ncf}. On average, how many emails does Bob forward in 1 hour?
Solution: The rate at which Bob receives Critical emails in his Inbox is $\lambda_c p_c$. The rate at which he forwards these is $\lambda_c p_c p_{cf}$. Similarly the rate at which Bob receives non-Critical emails in his Inbox folder is $\lambda_{nc} p_{nc}$, and the rate at which he forwards these is $\lambda_{nc} p_{nc} p_{ncf}$. Thus, the total rate at which Bob forwards emails is $\lambda_c p_c p_{cf} + \lambda_{nc} p_{nc} p_{ncf}$.

Problem 3 [25]

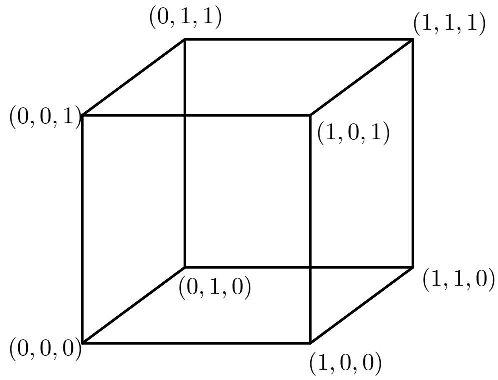

Figure 1: Unit cube

Lazy random walk on a cube: Consider a 3-dimensional unit cube as shown in Figure 1. An ant moves around on the vertices of this cube as follows: if at time n the ant is at a vertex v, then at time $n+1$ it remains on the same vertex v with probability $\frac{1}{2}$, or it jumps to one of the adjacent vertices with equal probability. Let X_n denote the position of the ant at time n. Note that X_n is a vector in $\mathbb{R}^3$.

(a) [5] Write the transition probability matrix P, and find the stationary distribution π.
Solution: Denote the states by $\{0,1,2,3,4,5,6,7\}$ based on the binary representation of these numbers (e.g state 5 is $(1,0,1)$). The transition matrix is given by

$$P = \begin{pmatrix} 1/2 & 1/6 & 1/6 & 0 & 1/6 & 0 & 0 & 0 \\ 1/6 & 1/2 & 0 & 1/6 & 0 & 1/6 & 0 & 0 \\ 1/6 & 0 & 1/2 & 1/6 & 0 & 0 & 1/6 & 0 \\ 0 & 1/6 & 1/6 & 1/2 & 0 & 0 & 0 & 1/6 \\ 1/6 & 0 & 0 & 0 & 1/2 & 1/6 & 1/6 & 0 \\ 0 & 1/6 & 0 & 0 & 1/6 & 1/2 & 0 & 1/6 \\ 0 & 0 & 1/6 & 0 & 1/6 & 0 & 1/2 & 1/6 \\ 0 & 0 & 0 & 1/6 & 0 & 1/6 & 1/6 & 1/2 \end{pmatrix}$$

Since this is a doubly stochastic matrix, the stationary distribution $\pi = \frac{1}{8}[1,1,1,1,1,1,1,1]$. Alternatively, we can plug this value of π to verify $\pi P = \pi$ and conclude that this must be the stationary distribution.

(b) [7] Find the expected time to return to vertex $(0,0,0)$ for the first time, given that $X_0 = (0,0,0)$.
Solution: Since $\pi((0,0,0)) = \frac{1}{8}$, the expected time to return is simply $\frac{1}{\pi((0,0,0))} = 8$. Alternatively, we can also set up the one step equations as done in the next part to solve this problem.

(c) [7] Find the expected time to reach $(1,1,1)$ for the first time, given that $X_0 = (0,0,0)$.
Solution: Let α be the expected time to reach $(1,1,1)$ from $(0,0,0)$. By symmetry, note that the expected time for each $(1,1,1)$ from $(0,0,1)$ or $(0,1,0)$ or $(1,0,0)$ is the same, which we call β. Similarly, the expected time to reach $(1,1,1)$ from $(1,1,0)$ or $(1,0,1)$ or $(0,1,1)$ be γ. We can now obtain the following linear equations:

$$\begin{aligned}
\alpha &= \frac{1}{2}(1+\alpha) + \frac{1}{2}(1+\beta) \\
\beta &= \frac{1}{2}(1+\beta) + \frac{1}{6}(1+\alpha) + \frac{1}{3}(1+\gamma) \\
\gamma &= \frac{1}{2}(1+\gamma) + \frac{1}{3}(1+\beta) + \frac{1}{6}
\end{aligned}$$

We solve these to get $\alpha = 20$, $\beta = 18$, $\gamma = 14$. The answer is thus 20.

(d) [6] For $n \geq 1$, let $Y_n = X_n - X_{n-1}$. Thus Y_n is the direction the ant moved in at time n. Is $Y_1, Y_2, \ldots$ a Markov chain?
Solution: No, $Y_1, Y_2, \ldots$ is not a Markov chain. To see this, suppose the sequence of X_n's is:

$$\begin{aligned}
X_0 &= (0,0,0) \\
X_1 &= (0,0,1).
\end{aligned}$$

In this case, $Y_1 = (0,0,1)$ and Y_2 has the distribution

$$P(Y_2 = y_2) = \begin{cases} 1/2 & \text{if } y_2 = (0,0,0) \\ 1/6 & \text{if } y_2 = (0,0,-1) \\ 1/6 & \text{if } y_2 = (0,1,0) \\ 1/6 & \text{if } y_2 = (1,0,0) \\ 0 & \text{otherwise.} \end{cases}$$

Now suppose we had

$$\begin{aligned}
X_0 &= (1,1,0) \\
X_1 &= (1,1,1).
\end{aligned}$$

In this case too, we have $Y_1 = (0,0,1)$. The distribution of Y_2 is now

$$P(Y_2 = y_2) = \begin{cases} 1/2 & \text{if } y_2 = (0,0,0) \\ 1/6 & \text{if } y_2 = (0,0,1) \\ 1/6 & \text{if } y_2 = (0,1,0) \\ 1/6 & \text{if } y_2 = (1,0,0) \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the distribution of Y_2 is not dependent only on the value of Y_1 and therefore $Y_1, Y_2, \ldots$ is not a Markov chain.

Problem 4 [30] Students arrive at a pastry shop according to a Poisson Process of rate $\lambda = 30$ per hour. The students independently buy Donuts or Eclairs, each with probability $\frac{1}{2}$. Assume that the shop opens at time zero and continues to be open indefinitely.

(a) [6] Conditional on 100 customers arriving in the first 5 hours of the shop opening, what is the probability that n donuts were sold in the first 4 hours of opening?
Intuitively, each of the 100 customers comes in the first 4 hours independently with probability 0.8 and buys a donut with probability 0.5. Thus, each customer independently both comes in the first 4 hours and buys a donut with prob. 0.4. Thus:

$$P(\text{n customers buy donuts in first 4 hours } |100 \text{ in 5 hours}) = \binom{100}{n}(0.4)^n(0.6)^{100-n}.$$

(b) [8] Find the pmf of the number of eclairs sold just before the first donut is sold.

n eclair buyers arrive before the first donut customer if the with probability $\frac{1}{2}^{n+1}$. Thus this number is a geometric r.v.

(c) [8] Define the n^{th} sale reversal if the n^{th} customer buys something different from the $n-1$st customer. For example, in the sequence of sales $DDEDDEE$ the third fourth, and sixth sales are reversals. Now consider the process defined by the times of reversals. Find the expected time to the first reversal. Find the expected time between any two subsequent reversals.

The first customer cannot be a reversal, and each subsequent customer causes a sales reversal with probability $\frac{1}{2}$. Now the donut customers and eclair customers can be viewed as indpendent Poisson processes each with rate 15. Thus after the first arrival, the time to the next reversal is exponentially distributed with mean $\frac{1}{15}$.

(d) [8] Find the probability density function of the time between reversals.

Exponential with rate 15 as explained earlier.