1
1.7 Let
$$b = h^{-1}(j)$$
 and $a_r = h^{-1}(i_r)$. Then,
 $P(Y_{n+1} = j | Y_r = i_r \text{ for } 0 \le r \le n) = P(X_{n+1} = b | X_r = a_r \text{ for } 0 \le r \le n)$
Preserving the
Markov Property

The result follows since X_n is Markovian. If h is not one-to-one, it is not necessarily true that Y_n is a Markov Chain. Suppose $\{X_n\}$ is Markovian on \mathbb{Z} for $n \ge 1$ and $X_0 = 0$. Then, $Y_n = X_n + X_{n-1}$ is a counterexample (see Exercise 6.1.11). One can see that Y_n is not Markovian by observing that Y_3 depends on Y_2 and Y_1 . So, Y_3 is not indepedent from Y_1 conditional on Y_2 .

- **1.8** Not necessarily true. See $Y_n = X_n + X_{n-1}$. Exercise 6.1.3 gives another counterexample in the sum $S_n + Y_n = M_n$.
- **1.9** Part (a) can be easily shown using Markov property of X_n . For $Y_n = X_{2m}$, let {even} = { $X_{2r} = i_{2r} : 0 \le r \le m$ }, {odd} = { $X_{2r+1} = i_{2r+1} : 0 \le r \le m-1$ }, and \mathcal{I} contains all odd indices *i*. Then,

$$\mathsf{P}(X_{2m+2} \mid \text{even}) = \sum_{x_i: i \in \mathcal{I}} \frac{\mathsf{P}(X_{2m+2} = k, X_{2m+1} = x_{2m+1}, \text{even}, \text{ odd})}{\mathsf{P}(\text{even})}$$
$$= \sum_{x_i: i \in \mathcal{I}} \frac{\mathsf{P}(X_{2m+2} = k, X_{2m+1} = x_{2m+1} \mid X_{2m} = x_{2m})\mathsf{P}(\text{even}, \text{ odd})}{\mathsf{P}(\text{even})}$$

$$= \mathsf{P}(X_{2m+2} = k \mid X_{2m} = i_{2m})$$

For
$$Y_n = (X_n, X_{n+1}),$$

$$\mathsf{P}(Y_{n+1} = (k, \ell) \mid Y_0 = (i_0, i_1), \dots, Y_n = (i_n, k)) =$$
$$\mathsf{P}(Y_{n+1} = (k, \ell) \mid X_{n+1} = k) = \mathsf{P}(Y_{n+1} = (k, \ell) \mid Y_n = (i_n, k))$$

by Markov property of X_n .

2

Card Trick

This phenomenon is called *Kruskal Count*. This is also the "coupling game" on page 235-236. One intuition behind the card trick is that if at any point of the card trick, once the magician's (your) key card "clicks" with the audience's (my) keycard, our keycards would remain the same for the remainder of the trick, How do we come up with a Markov Chain to represent the "clicking" and that our keycards staying the same once they click? The latter appears to suggest an absorbing state or class (once we have the same keycard, we will always have the same keycard). One way to represent the "clicking" is to define states of the markov chain as the absolute difference between value of your keycard, $k_{n'}$ and my most recent keycard k_n . Suppose $k_n, k_{n'} \in$ $\{1,\ldots,9\}$ for $n,n' \geq 0$. Then, the states $\mathcal{S} = \{0,\ldots,8\}$. Using our definition of the states, state 0 represents the "clicking" and is the absorbing state of this finite chain. Therefore, once the chain reaches state 0, it stays there. Since the finite chain has only one absorbing state and 0 can be reached from the remaining states, the remaining states are transient. In the context of the card trick, this means that given an deck of infinite number of cards, our cards will eventually be the same at some point during the trick by the definition of transience in a finite Markov Chain. Transience property in this finite chain implies that all states other than 0 will eventually not be visited in the long run. This is confirmed by your simulations in which the probability of performing your card trick with success increases with the deck size. Another interesting discovery some of you may have found is that the probability also increases as the value assigned to face cards decreases. The intuition behind this is that by assigning low values to face cards allows the cards to run out slower. In other words, the magician receives more chances to have the same as the audience's. Thus, the probability of having the same keycard goes up.

The following are two notable references that explore Kruskal Count through simulations and mathematical arguments. Haga and Robins (1995). J. Lagarias and E. Rains and R. Vanderbei (2001). The former paper explains the trick using a chain similar to the one we used here. The latter paper explains the trick using a coupled chain. A coupled chain is a chain that describes the joint behavior of independent copies of an original chain where each copy behaves according to the laws of the original chain. This coupling argument is used by the text on pages 235-236 albeit nowhere as rigorous as the paper's.

Note: The question didn't seem to indicate how to treat 10. Whether assigning the value of a face card or 10 to it shouldn't affect the results of your simulations by that much. The general relationships (between deck size and probability of success, etc...) should remain the same.

3

The supply of rolls on a certain day depends on the supply of rolls on previous day, amount of rolls used up, and number of rolls delivered in the evening. But the latter are iid across days and indepedent from the supply of rolls. The number of rolls used is constant across days. So the supply of rolls on a certain day only depends on supply of rolls on the previous day. Therefore, this scenario can be modeled by a Markov Chain. The transition probabilities are given by

$$P(i,j) = \begin{cases} \mathsf{p}(j) & \text{for } i = 0, 1\\ \mathsf{p}(j-i+1) & \text{for } i > 1 \text{ and } i \le j+1\\ 0 & \text{otherwise} \end{cases}$$

Let r be the price per roll, c be the extra cost of an emergency roll and s be storage cost per roll. Then, we want to minimize the expected cost, which in terms of the stationary distribution π is

$$(r+c)\pi_0 + (r+s)\sum_{k=1}^{\infty} k\pi_k$$

Also observe that P is an Upper Hessenberg matrix. Few implications result. If the expected value of number of rolls delivered is greater than 1, the supply increases indefinitely but the storage cost also increases. If it is less than 1, eventually the supply would reach 0. So the probability of having no roll in storage can be calculated. If it is equal to 1, the same probability would not be able to be calculated since the chain is not positive recurrent (stationary distribution fails to exist in this case).

Inventory Model

4

The homework has a typo: s and s_0 have to belong to the same cyclic Periodic Limits class for the equality to hold. Second note that to find

$$\lim_{n \to \infty} P^{nd}(s_0, s)$$

it suffices to find the limiting behavior of a chain with transition matrix

$$\tilde{P} = P^d$$

However, P is the transition matrix of X. Therefore, P^d is just the transition matrix of a chain which observes X every d steps. Call this chain, X^d . Then, $X^d = (X_0, X_d, X_{2d}, \ldots)$. We have seen in class that this chain is ergodic. Therefore, the limiting distribution exists, i.e.

$$\lim_{n \to \infty} \widetilde{P}^n(s_0, s) = \frac{1}{\widetilde{M}(s, s)}$$

where $\widetilde{M}(s,s)$ is the mean recurrence time of X^d . Since X moves d times more steps than X^d , intuitively speaking in terms of M(s,s),

$$\widetilde{M}(s,s) = d^{-1}M(s,s)$$

and the result follows. More mathematically: first let T_s and T'_s be the first return times to s for X^d and X respectively and observe

$$\begin{split} \widetilde{M}(s,s) &= \sum_{n=0}^{\infty} n \mathsf{P}(T_s = n) = d^{-1} \sum_{n=0}^{\infty} nd \cdot \mathsf{P}(T'_s = nd) \\ &= d^{-1} M(s,s) \end{split}$$

5

Given Exponential(λ) service time and G is the distribution of interarrival time. Chris' notes give the transition probability matrix of a GI/M/1 queue as:

$$P(j, j+1-k) = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^k}{k!} dG(t) \quad k \le j$$
$$P(j, 0) = \int_0^\infty \sum_{k=j+1}^\infty e^{-\lambda t} \frac{(\lambda t)^k}{k!} dG(t)$$

where states Q_n represents the number of people in the queue just before the n^{th} arrival. A few people asked how this was derived. So here goes: First let X_n represent the number of departures between n^{th} and $(n + 1)^{th}$ arrival. It can be seen that

$$Q_{n+1} = Q_n + 1 - X_n$$
, for $X_n < Q_n + 1$

So,

$$P(i,j) = \mathsf{P}(Q_{n+1} = j \mid Q_n = i) = \begin{cases} \mathsf{P}(X_n = i - j + 1) & \text{if } j > 0\\ \mathsf{P}(X_n \ge i + 1) & \text{if } j = 0 \end{cases}$$
$$= \begin{cases} \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^{i-j+1}}{(i-j+1)!} dG(t) & \text{if } j > 0\\ \int_0^\infty \sum_{k=i+1}^\infty e^{-\lambda t} \frac{(\lambda t)^k}{k!} dG(t) & \text{if } j = 0 \end{cases}$$

Second equality follows from to the fact that exponential services times imply the number of departures in any given interarrival time period is Poisson distributed. Using law of total probability, integrate the Poisson density over all possible interarrival times t with distribution function G(t) to obtain the result.

Finally, express P(i, j) in terms of Chris' notation: P(j, j + 1 - k)and find they are equivalent. Now back to our main event.

Let μ be the expected value of the distribution defined by G and assume $[\lambda\mu]^{-1} < 1$. $\lambda\mu$ is called the traffic intensity. The traffic intensity is defined as the ratio of the mean of service time to mean of interarrival time. Intuitively, if it is greater than 1, the queue length would become infinitely large since the service time is on average larger than the interarrival time. So the states, Q_n , become transient. We shall see that for traffic intensity less than 1, the queue length reaches an equilibrium, i.e. limiting distribution exists and so the chain is positive recurrent (thus chain is null recurrent when it equals 1). To make notation simple, let

$$\alpha_k = e^{\lambda t} \frac{(\lambda t)^k}{k!}$$

Then our transition matrix is given by

$$P = \begin{pmatrix} 1 - \alpha_0 & \alpha_0 & 0 & 0 & \cdots & \cdots \\ 1 - \alpha_0 - \alpha_1 & \alpha_0 & \alpha_1 & 0 & \cdots & \cdots \\ 1 - \alpha_0 - \alpha_1 - \alpha_2 & \alpha_0 & \alpha_1 & \alpha_2 & 0 & \cdots \\ \vdots & & \alpha_0 & \cdots & \cdots & \alpha_3 & \cdots \\ \vdots & & & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

 $\pi = \pi P$ yields the following system of equations:

$$\pi_0 = \sum_{i=0}^{\infty} \pi_i \left(1 - \sum_{j=0}^i \alpha_j \right) \tag{1}$$

$$\pi_k = \sum_{i=0}^{\infty} \pi_{k+i-1} \alpha_i \quad \text{for } k > 0 \tag{2}$$

with

$$\sum_{i=0}^{\infty} \pi_i = 1 \tag{3}$$

Let our guess for solution be

$$\pi_k = c\gamma^k \quad \text{for } k \ge 0 \tag{4}$$

Plugging in back into (2), we have

$$\gamma = \sum_{i=1}^{\infty} \alpha_i \gamma^i$$

Now plug in the value of α_i .

$$\begin{split} \gamma &= \sum_{i=1}^{\infty} \int_{0}^{\infty} \frac{(\lambda \gamma t)^{i}}{i!} e^{-\lambda t} dG(t) \\ &= \int_{0}^{\infty} \left[\sum_{i=1}^{\infty} \frac{(\lambda \gamma t)^{i}}{i!} \right] e^{-\lambda t} dG(t) \\ &= \int_{0}^{\infty} e^{\lambda \gamma t} e^{-\lambda t} dG(t) \\ &= \mathsf{E}_{G} \left[e^{(\lambda \gamma - \lambda) t} \right] = \mathsf{M}[\lambda(\gamma - 1)] \end{split}$$

where M is the moment generating function of G. Now we proceed to find γ that solves this equality. Let $B(\gamma) = \mathsf{M}[\lambda(\gamma - 1)]$. One can verify that $B(\gamma)$ is a convex function and nondecreasing on [0, 1] by observing the derivatives of $B(\gamma)$.

Also, B(0) > 0, B(1) = 1, and $B'(1) = \mu\lambda$. By our assumption that traffic intensity is less than 1, $B'(1) = \mu\lambda > 1$. Now consider two functions, $y = \gamma$ and $y = B(\gamma)$. We know they must intersect in at least one point, at 1. Is the line $y = \gamma$ merely tangent to $B(\gamma)$? Since B'(1), the slope of the tangent line at 1, is greater than 1, it is not true γ only intersects $B(\gamma)$ at 1 but is true that $y = \gamma$ intersects $B(\gamma)$ at 2 points on [0, 1] by properties of convexity. Call ν the other intersecting point, i.e. $B(\nu) = \nu \neq 1$. Because B(0) > 0 and $B(\gamma)$ is nondecreasing on $[0, 1], 0 < \nu < 1$.

Recall that $\gamma = 1$ also satisfies $B(\gamma) = \gamma$. But $\gamma = 1$ is useless because the solutions to $\pi = \pi P$ would not be normalizable under this case. $\gamma = \nu$ does not have this problem. So $\gamma = \nu$ is the solution to $B(\gamma) = \gamma$ we care about. Also, note that (4) satisfies (1) because we can easily derive (1) from (2) and (3). In other words, (4) satisfies (1) because of the dependence between balance equations. Using (4), we have

$$1 = \sum_{j=0}^{\infty} \pi_j = c \sum_{j=0}^{\infty} \nu^j = \frac{c}{1 - \nu}$$

 $\Rightarrow c = 1 - \nu$. Therefore, the stationary distribution is given by

$$\pi_k = (1-\nu)\nu^k$$
, for $k \ge 0$

where $M[\lambda(\nu - 1)] - \nu = 0$ for $\nu \in (0, 1)$.

6 -

The proof makes use of a number-theoretic fact that the textbook calls *postage stamp lemma* which states that: for positive and coprime integers, i_1 and i_2 , there exists N such that $\forall n \geq N$, we can find $p, q \in \mathbb{N}$ such that $n = p \cdot i_1 + q \cdot i_2$.

More generally, it states that if i_1, \ldots, i_k are positive and coprime, there exists a positive integer N such that for all $n \ge N$, we can find non-negative integers n_j such that $n = \sum_{j=1}^k n_j i_j$

Finite Recurrence

Let a finite, irreducible and aperiodic chain be given. By definition of aperiodic chains, we know that

$$\gcd\{d \ge 1 : P_{ii}^d > 0\} = 1$$

Now choose $r_1, \ldots, r_k \in \{d \ge 1 : P_{ii}^d > 0\}$. Then, by the Lemma above, we know there exists a positive integer N such that for all $n \ge N$, we can find α_j such that $n = \sum_{j=1}^k \alpha_k r_k$. Then, by Chapman Kolmogorov, for any state ℓ

$$P^{n}(\ell,\ell) \ge \prod_{j=1}^{k} [P^{r_j}(\ell,\ell)]^{\alpha_j} > 0$$

Therefore, $P^n(\ell, \ell) > 0 \quad \forall n \ge N.$

Since every state communicates with each other, we can find m such that $P^m(\gamma, \ell) > 0$ for some given pair of states γ and ℓ . It follows that

$$P^{n+m}(\gamma,\ell) \ge P^m(\gamma,\ell)P^n(\ell,\ell) > 0$$

The chain is finite, irreducible and aperiodic. The states in this chain need to be recurrent. Why? Well, if not, then all states are transient by irreducibility property. But by finiteness, transience implies that eventually we will run out of states to travel and that creates a contradiction. Also, recurrent states in a finite chain are always positive recurrent. The proof of this is given in Section 6.3 Lemma (5). Therefore, the stationary distribution always exists.

The transition probabilities are given by

$$P(i,j) = \begin{cases} p & \text{for } j = 0\\ 1-p & \text{for } j = i+1\\ 0 & \text{otherwise} \end{cases}$$

The stationary distribution, ρ can be obtained by solving

 $\rho = \rho P$

which gives

$$\rho_0 = p \sum_{j=0}^{\infty} \rho_j = p$$

Markov Flu

36-703

Homework #4 Solutions

$$\rho_k = (1-p)\rho_{k-1}$$

Now do the Plug-N'-Chug and we obtain that for each $k \ge 0$,

 $\rho_k = p(1-p)^k$

People interpreted this question two different ways, both of which are acceptable. The second interpretation has the following transition probability matrix:

$$P(i,j) = \begin{cases} p & \text{for } j = 0, i > 0\\ 1 & \text{for } i = 0, j = 1\\ 1 - p & \text{for } j = i + 1, i > 0\\ 0 & \text{otherwise} \end{cases}$$

The more important message of this question is that by treating the outbreak of the flu as arrivals (or successes in a sequence of Bernoulli trials) we define a renewal process whose interarrival times are geometrically distributed.

8 —

As long as there's no gobbledygook here, you received credit.

Murkiest Points