Stat 6329 Syllabus

Stat 6329.001
Applied Probability and Stochastic Processess
Course Information

Course number/section	Stat 6329-001
Instructor	Dr. Larry P. Ammann
Email	ammann@utdallas.edu
Office	FO 2.602D
Office hours	Wed, 2:00-3:30 pm, others by appt.
Phone	(972) 883-2164
Required text	Introduction to Probability Models, 10^{th} or 11^{th} ed.
Author	Sheldon M. Ross
Suggested course materials	None

Tentative Schedule

Topics	Chapters
Review of probability and random variables	1.1-1.6; 2.1-2.3
Expectation	2.4
Joint distributions and independence	2.5-2.8
Conditioning and conditional expectation	3.1-3.7
Markov chains	4.1-6
Poisson and related processes	5.1-5.4
Renewal processes	7.1-7.3
Introduction to queuing models	8.1-3
Introduction to Brownian motion	10.1-10.6

Grading Policy

Final course grade will be based on 3 exams and homework. You may bring one sheet of notebook paper with notes for the exams.

Student Learning Objectives

1. Understand the fundamentals of probability theory and random variables
2. Understand the basic probability tools for stochastic processes: conditioinal probability and conditional expectation

3. Understand the basic properties and application of special classes of stochastic processes including Markov chains and Poisson and related processes

The complete syllabus is available here:
`http://www.utdallas.edu/~ammann/stat6329_syllabus.pdf`

Class Notes

Review of Probability and Random Variables

Probability provides a mathematical model for an *Experiment*: a process whose outcome is uncertain. A probability model contains three components.

1. Ω, the *Sample space* defined to be the set of all possible outcomes of the experiment
2. $\mathcal{A}$, a collection of events (subsets of Ω) whose probabilities we wish to model. For reasons of mathematical consistency this collection must contain Ω, complements of any set in $\mathcal{A}$ (which implies $\emptyset \in \mathcal{A}$), along with countable unions of sets in $\mathcal{A}$. We refer to such collections as sigma-algebras and events in such a collection are referred to as measureable sets. The collection of Borel subsets of $\Re$ is an example. If Ω is finite or uncountable infinite, then $\mathcal{A}$ can be the set of all subsets of Ω. However, if Ω is uncountable, then we need these restrictions.
3. A probability function $P : \mathcal{A} \to \Re$ that satisfies
   (a) $0 \le P(A) \le 1, \forall A \in \mathcal{A}$
   (b) $P(\Omega) = 1$
   (c) if $\{E_i\}$ is a countable collection of disjoint events in $\mathcal{A}$, then
   $$P\left(\cup E_i\right) = \sum_i P(E_i).$$

Note that these components, in particular the probability function, are defined in terms of the underlying experiment.

Properties

1. $P(\emptyset) = 0$. **Proof**: let

$$E_i = \emptyset,\ i \ge 1.$$

Then $\{E_i\}$ are disjoint and

$$\cup E_i = \emptyset.$$

Hence,

$$P(\emptyset) = P\left(\cup E_i\right) = \sum_{i=1}^{\infty} P(E_i) = \sum_{i=1}^{\infty} P(\emptyset).$$

The only real number that satisfies this equation is $P(\emptyset) = 0$.

2. If $E_1, \cdots E_n$ is a finite collection of disjoint events in $\mathcal{A}$, then

$$P\left(\cup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i).$$

Proof: let $E_k = \emptyset$, $k \geq n+1$. Then $\{E_i,\ i \geq 1\}$ is a countable collection of disjoint events in $\mathcal{A}$ and

$$\cup_{i=1}^n E_i = \cup_{i=1}^\infty E_i$$

Hence,

$$P\left(\cup_{i=1}^n E_i\right) = P\left(\cup_{i=1}^\infty E_i\right) = \sum_{i=1}^\infty P(E_i) = \sum_{i=1}^n P(E_i).$$

3. If $A, B \in \mathcal{A}$, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Proof: write

$$A \cup B = A \cup (B \cap A^C)$$

and note that this is a disjoint union. Therefore

$$P(A \cup B) = P(A) + P(B \cap A^C).$$

Similarly,

$$P(B) = P(A \cap B) + P(A^C \cap B).$$

Hence,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Note that this implies the corollary: if A,B are measureable sets with $A \subset B$, then $P(A) \leq P(B)$. So probability is a monotone function. More generally, let $E_1, \cdots E_n$ be an arbitrary collection of measureable sets. Then

$$P\left(\cup_{i=1}^n E_i\right) = \sum_{r=1}^n (-1)^{r+1} \sum_{1 \leq i_1 < \cdots < i_r \leq n} P\left(E_{i_1} \cap \cdots \cap E_{i_r}\right).$$

Proof: by induction. This is referred to as the *inclusion-exclusion* identity. For example,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC).$$

4. Experiments with equally likely outcomes. Suppose Ω contains n outcomes, $O_1, \cdots O_n$ and each outcome has the same probability,

$$P(O_j) = p,\ 1 \leq j \leq n.$$

Then

$$1 = P(\Omega) = P(\cup_i O_i) = \sum_{i=1}^n P(O_j) = np,$$

which implies that $p = 1/n$. Furthermore, since any event in $\mathcal{A}$ can be expressed as a union of individual outcomes, then

$$P(E) = \frac{\#\{E\}}{n}.$$

Example. What is the probability at least two people in a group have the same birthdate? Ignore leap-years and assume that 365 possible dates in a year are equally likely. Suppose the group size is $n < 365$ and no twins, triplets, etc., are in the group. Then the total number of outcomes for the n birthdates is 365^n. The number of outcomes with different dates is

$$(365)(364)\cdots(365 - n + 1),$$

so

$$P(\text{no match}) = \frac{365!}{(365-n)!365^n} = \begin{cases} 0.49, & n = 23 \\ 0.29, & n = 30 \\ 0.03, & n = 50 \end{cases}$$

Conditional Probability

We often use the language of percentages when describing probabilities. For example, suppose a large company has 10 positions to fill and has received 100 qualified applicants for those positions. Suppose a review of those applicants gave the following:

	M	F
Hired	8	2
not	52	38

Assume a probability model in which each qualified applicant is equally likely to be hired. Then we can say
1) 10% of qualified applicants were hired, $P(\text{Hired}) = 10/100$
2) 60% of qualified applicants were male, $P(M) = 60/100$
3) 8% of qualified applicants were male and hired, $P(M \cap \text{Hired}) = 8/100$
But what about the statements:
4) 80% of those hired were male, 8/10
5) 5% of qualified female applicants were hired, 2/40.

The last two statements require a different type of probability. Note the English structure of the first statement:
(10%) (of qualified applicants) (were hired)
(subject) (adjective phrase modifying subject) (verb plus object)
Note that the adjective phrase defines the reference group for the percentage and so specifies the denominator. The verb plus object defines the outcomes contained in the event. The translation into the language of probability is then

$$0.1 = \frac{\#\{\text{were hired}\}}{\#\{\text{qualified applicants}\}} = \frac{10}{100}.$$

This is ordinary probability because the reference group is the entire sample space. Now consider statement 4) above.
(80%) (of those hired) (were male)
The reference group is the set of those who were hired. Since this is not the entire sample space, then this is not an ordinary probability. Also, the requirement (were male) must be counted from within the reference group. Therefore, we can translate this statement as:

$$0.8 = \frac{\#\{\text{were male and hired}\}}{\#\{\text{were hired}\}} = \frac{P(M \cap H)}{P(H)}.$$

Similarly, 5) is equivalent to

$$0.05 = \frac{P(F \cap H)}{P(F)}.$$

We refer to these as conditional probabilities, defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \; P(B) > 0.$$

Note that conditional probability is only defined for conditions with positive probabilities. The translations of 4) and 5) are:

$$\begin{aligned} P(\text{male}|\text{hired}) &= \frac{8/100}{10/100} = 0.8 \\ P(\text{hired}|\text{female}) &= \frac{2/100}{40/100} = 0.05. \end{aligned}$$

Note that the reference group comes after the seperator — in these conditional probabilities.
Properties of Conditional Probability

1. $$P(AB) = P(A|B)P(B) = P(B|A)P(A).$$

   Note that if $P(A) = 0$ or $P(B) = 0$ then $P(AB) = 0$ from the monotonicity property of probability.

2. Let B denote an event with $P(B) > 0$ and define a function P_B on $\mathcal{A}$ by:

   $$P_B(E) = P(E|B), \; E \in \mathcal{A}.$$

   The P_B satisfies the axioms of probability and therefore it also satisfies all the properties derived from those axioms.

3. Let $\{B_i\}$ denote a finite or countably infinite *partition* of Ω, that is, these events are disjoint and their union is Ω. Assume without loss of generality that $P(B_i) > 0, \; \forall i$. Then

   $$P(A) = P\left(\cup_i (AB_i)\right) = \sum_i P(AB_i) = \sum_i P(A|B_i)P(B_i).$$

This is referred to as the *Theorem of Total Probability.* This can be extended to give Bayes Theorem: if $P(A) > 0$ then

$$P(B_k|A) = \frac{P(AB_k)}{P(A)} = \frac{P(A|B_k)P(B_k)}{\sum_i P(A|B_i)P(B_i)}.$$

In this context $P(B_i)$ is referred to as a prior probability and $P(B_k|A)$ is referred to as a posterior probability. The posterior probability of B_k represents an update to the likelihood that event B_k will occur given that we have observed an outcome in A.

Example. Suppose a particular type of cancer with a high mortality rate is very costly and/or risky to diagnose with standard tools such as surgical biopsy. Suppose also that the set of symptoms experienced by patients with this cancer is relatively common so that only 5% of all patients with those symptoms actually have this type of cancer. Then this cancer is just one possible diagnosis for such patients, but it may not the most likely diagnosis. Now suppose a relatively inexpensive, low risk screening test has been developed to help diagnose this type of cancer for the target population, the set of patients who exhibit that set of symptoms. Properties of the screening test are obtained by applying it to patients in the target population, some of whom are known to have this type of cancer, the rest are known not to have this cancer. Suppose the following results were obtained: 98% of those in the target population known to have this cancer get a positive response to the screening test and 95% of those in the target population known not to have this cancer get a negative response. Is this a good screening test? We will answer this question from the perspective of a doctor who has a patient identified as belonging to the target population. Suppose this patient gets a positive response to the screening test. What does the doctor tell the patient?

First we construct a probability model for the experiment in which a patient is randomly selected from the target population. Then the translations of the results into this probability model are:

$$\begin{aligned}
P(\text{has cancer}) &= 0.05, \\
P(\text{Positive}|\text{has cancer}) &= 0.98 \\
P(\text{Negative}|\text{not cancer}) &= 0.95
\end{aligned}$$

This gives

$$\begin{aligned}
P(\text{Positive} \cap \text{has cancer}) &= P(\text{Positive}|\text{has cancer})P(\text{has cancer}) \\
&= (0.98)(0.05) = 0.0490 \\
P(\text{Negative} \cap \text{not cancer}) &= P(\text{Negative}|\text{not cancer})P(\text{not cancer}) \\
&= (0.95)(0.95) = 0.9025
\end{aligned}$$

We can put these into a probability table:

	has cancer	not cancer	total
Positive	.0490		
Negative		.9025	
total	.0500		1.0000

Note that this table only contains ordinary probabilities, not conditional probabilities. The blank cells can be filled in using the theorem of total probability:

	has cancer	not cancer	total
Positive	.0490	.0475	.0965
Negative	.0010	.9025	.9035
total	.0500	.9500	1.0000

We can obtain the answer to the question from this table:

$$P(\text{has cancer}|\text{Positive}) = \frac{.0490}{.0965} = 0.508,$$

not much better than a coin flip! How can that be a good test? But note that before the screening test result is known, the probability the patient has this type of cancer is 5%; after a positive test result, this probability increases to 50.8% from 5%. This increased risk would warrant applying the more costly definitive diagnostic tool to the patient. On the other hand, if the screening test is negative, then

$$P(\text{has cancer}|\text{Negative}) = \frac{.0010}{.9035} = 0.0011.$$

In this case the doctor would test for other diagnoses. Of course, if it turns out that none of these other diagnoses is effective, the doctor likely would then apply the definitive diagnostic test for this type of cancer given its high mortality rate.

Example: the Monte Hall problem. This problem is named after the host of a TV game show called *Let's Make a Deal*. A contestant is presented with 3 doors behind one of which is a valuable prize but the other two doors contain nothing of value. Monte asks the contestant to choose a door and the contestant will win whatever is behind the door. However, before opening the selected door, Monte opens one of the other doors he knows does not contain the valuable prize and offers the contestant the opportunity to switch his or her choice to the remaining unopened door. Should the contestant switch or stay with the original choice?

This problem first appeared in 1975 in the *American Statistician*. Quite a few people who answered based on their intuition and said the remaining unopened doors are equally likely to have the valuable prize got it wrong. It is interesting to note that the audience members who were fans of the show got it right - they always encouraged the contestant to switch! To see why, we will construct a probability model for this problem. Assume the prize is equally likely to be behind any of the doors:

$$P(A) = P(B) = P(C) = 1/3.$$

Suppose the contestant chooses door A. If the prize is behind door A, assume Monte mentally flips a coin to select which of the other doors to open, but if the prize is not behind A then

Monte must open the door he knows does not have the prize. Suppose Monte opens door C and then asks the contestant if he or she would like to switch from A to B. We need to find the conditional probability,

$$P(A|\text{opens C}) = \frac{P(A \cap \text{opens C})}{P(\text{opens C})}.$$

Now,

$$P(A \cap \text{opens C}) = P(\text{opens C}|A)P(A) = (1/2)(1/3) = 1/6.$$

Also,

$$\begin{aligned} P(\text{opens C}|A)P(A) &= (1/2)(1/3) = 1/6 \\ P(\text{opens C}|B)P(B) &= (1)(1/3) = 1/3 \\ P(\text{opens C}|C)P(C) &= (0)(1/3) = 0 \end{aligned}$$

Therefore,

$$\begin{aligned} P(\text{opens C}) &= P(\text{opens C}|A)P(A) + P(\text{opens C}|B)P(B) + P(\text{opens C}|C)P(C) \\ &= (1/2)(1/3) + (1)(1/3) + (0)(1/3) \\ &= 1/2 \end{aligned}$$

And so,

$$\begin{aligned} P(A|\text{opens C}) &= \frac{1/6}{1/2} = \frac{1}{3} \\ P(B|\text{opens C}) &= \frac{1/3}{1/2} = \frac{2}{3} \end{aligned}$$

Switch!

Independence

If $P(A|B) = P(A)$, then knowledge that the outcome of the experiment is in B does not change the likelihood that A occurs. In this case we say that events A and B are statistically independent. Note that this would imply that

$$P(AB) = P(A|B)P(B) = P(A)P(B)$$

Since $P(A|B)$ is not defined when $P(B) = 0$, we define independence more generally by **Definition**: events A and B are said to be *statistically independent* iff $P(AB) = P(A)P(B)$. This definition is equivalent to $P(A|B) = P(A)$ when $P(B) > 0$.

Random Variables

We use r.v.'s to model numeric data. A r.v. assigns a numeric value to each outcome of an experiment. Formmally, a r.v. X is a function, $X : \Omega \to \Re$ that satisfies

$$\{\omega \in \Omega : \; X(\omega) \leq a\} \in \mathcal{A}, \; \forall a \in \Re.$$

This requirement implies that events in which the random variable takes values in an interval will be measureable,

$$P(\omega \in \Omega : \; a < X(\omega) \leq b).$$

Note that ordinarily we suppress the notation that X is a function and write this probability as

$$P(a < X \leq b),$$

but we always should remember that this is just shorthand notation and X is actually a function on the sample space.

Suppose for example we toss a fair coin until a 6 appears. Assume the tosses are independent and let S denote the event that 6 appears on a toss and let F denote the complement of that event. Then the sample space of this experiment is

$$\Omega = \{S, FS, FFS, FFFS, \cdots\}.$$

Now let N denote the number of tosses required for a 6 to appear. Then $N(S) = 1$, $N(FS) = 2$, etc, and so for $n \geq 1$,

$$\begin{aligned} P(\omega \in \Omega : N(\omega) = n) &= P(N = n) \\ &= P(F \cdots FS) = [P(F)]^{n-1} P(S) \\ &= \left(\frac{5}{6}\right)^{n-1} \frac{1}{6}. \end{aligned}$$

The set of all possible values of a r.v. is called its sample space. In this example the sample space of N is the positive integers. Also,

$$\sum_{k=1}^{\infty} P(N = n) = \sum_{k=1}^{\infty} \left(\frac{5}{6}\right)^{n-1} \frac{1}{6} = \frac{1/6}{1 - 5/6} = 1.$$

R.v.'s with a finite or countably infinite sample space are referred to as discrete r.v.'s. For such r.v.'s we can derive the probability of any event from its *probability mass function*, defined by

$$p(x) = P(X = x).$$

In particular,

$$P(E) = \sum_{x \in E} p(x).$$

Probability mass functions satisfy

1. $0 \leq p(x) \leq 1, \; \forall x \in \Re$
2. $p(x) > 0$ for at most countable values x

3. $\sum_x p(x) = 1$

Any function that satisfies these condtions is the p.m.f of some r.v.

If the sample space of a r.v. is uncountable, then we refer to such r.v.'s as continuous r.v.'s. In this case the event

$$\{\omega \in \Omega : X(\omega) = x\}$$

leads to mathematical inconsistencies and so the p.m.f. is not defined for continuous r.v.'s. However, the function

$$F(x) = P(\omega \in \Omega : X(\omega) \leq x) = P(X \leq x)$$

is defined for all r.v.'s and probabilities of events associated with a r.v. can be derived from this function, referred to as the cumulative distribution function. This function satisfies the following conditions

1. $F(x)$ is monotone, nondecreasing, and right continuous
2. $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$

Anuy function that satisfies these conditions is the c.d.f. of some r.v. In the discrete case,

$$\begin{aligned} P(X = x) &= F(x) - \lim_{\epsilon \searrow 0} F(x - \epsilon) \\ &= F(x) - F(x-). \end{aligned}$$

Note that if $p(x) > 0$ then F has a discontinuity point at x. The conditions on cdf's imply that a cdf has at most countably infinite many points of discontinuity.

Bernoulli and related r.v.'s

The simplest non-trivial r.v. is defined on an experiment, referred to as a Bernoulli trial, with 2 possible outcomes. Label these outcomes as S,F and let $P(S) = p$. Then $P(F) = 1-p$. A Bernoulli r.v. is defined by $X(S) = 1$, $X(F) = 0$. Its p.m.f. is

$$p(x) = \begin{cases} 1-p, & x = 0 \\ p, & x = 1 \\ 0, & \text{otherwise} \end{cases}$$

Binomial. Suppose n independent Bernoulli trials are performed each with the same success probability p. Then the sample space consists of strings of n characters, each of which is either S or F. Let N denote the number of times S appears in the string, that is, N is the number of successes among n independent Bernoulli trials. Then the sample space is $0, 1, \cdots, n$ and the p.m.f. is

$$P(N = k) = \binom{n}{k} p^k (1-p)^{n-k}, \ 0 \leq k \leq n$$

This r.v. is named the Binomial r.v. because its pmf is a term in a binomial series. This r.v. provides a probability model for random sampling with replacement from a population with two types of individuals. In practice we usually perform sampling without replacement, but as long as the sample size is small compared to the population size, the binomial distribution provides a reasonable approximation.

Geometric. Now suppose we perform a series of independent Bernoulli trials with the same success probability and continue until we observe an S. Let N denote the number of trials required to obtain S. Then the pmf of this r.v. is

$$P(N = k) = (1-p)^{k-1}p,\ k \geq 1.$$

Note that this pmf represents terms in a geometric series. An alternotive form of the geometric r.v. is to count the number of failures before the first success. Let Y denote this r.v. Then $Y = N - 1$ and so its pmf is

$$P(Y = j) = (1-p)^{j}p,\ j \geq 0.$$

A generalization of the geometric is the negative binomial r.v., the number of trials required to obtain r successes, where r is a fixed positive integer. Then the sample space for N_r is $r, r+1, \cdots$ and $\{N_r = k\}$ is the event that there are exactly $r-1$ successes among the first $k-1$ trials and trial k is an S. Therefore

$$P(N_r = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r},\ k \geq r.$$

Let M_r denote the number of failures before the r^{th} success, $M_r = N_r - r$. Then

$$P(M_r = j) = P(N_r = r + j) = \binom{r+j-1}{j} p^r (1-p)^j,\ j \geq 0.$$

Poisson. A commonly used model for count data is the Poisson r.v. with pmf

$$P(N = n) = \frac{\lambda^n}{n!} e^{-\lambda},\ n \geq 0,$$

where λ is a positive real number. Before calculators, computation of binomial probabilities was very tedious for large n. This led to one of the first limit theorems in probability: if $p_n \to 0$ in such a way that $np_n \to \lambda > 0$, then

$$\lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{\lambda^n}{n!} e^{-\lambda},\ k \geq 0.$$

Markov chain examples

Example 1. Random walk with partial reflecting barrier at 0. Let

$$\begin{aligned} P_{0,0} &= \frac{1}{2},\ P_{0,1} = \frac{1}{2},\ P_{0,i} = 0,\ i > 1, \\ P_{i,i-1} &= q,\ P_{i,i+1} = p,\ i \geq 1,\ \text{where } p + q = 1, \\ P_{i,j} &= 0,\ i \geq 1,\ j \neq i-1, i+1. \end{aligned}$$

Then TPM is

$$P = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 & \cdots \\ q & 0 & p & 0 & 0 & \cdots \\ 0 & q & 0 & p & 0 & \cdots \\ 0 & 0 & q & 0 & p & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

All states communicate so this is an aperiodic, irreducible MC. Intuitively, it seems that if q¿p then states should be recurrent. To determine whether or not states are recurrent we must consider the system of equations $\mathbf{v} = \mathbf{vP}$.

$$\begin{aligned} v_0 &= \frac{1}{2}v_0 + qv_1 \ \Rightarrow\ v_1 = \frac{1}{2q}v_0 \\ v_1 &= \frac{1}{2}v_0 + qv_2 \ \Rightarrow\ v_2 = \frac{p}{q}v_1 \end{aligned}$$

The remaining equations have the form

$$v_k = pv_{k-1} + qv_{k+1},\ k \geq 2.$$

Therefore,

$$p(v_{k-1} - v_k) = q(v_k - v_{k+1}),\ k \geq 2,$$

Let

$$\Delta_k = v_{k-1} - v_k,\ k \geq 2.$$

Then for $p \neq q$, we have the recursive equations,

$$\Delta_{k+1} = \frac{p}{q}\Delta_k \ \Rightarrow\ \Delta_k = \left(\frac{p}{q}\right)^{k-2}\Delta_2,\ k \geq 2.$$

Also,

$$\Delta_2 = v_1 - v_2 = \frac{v_0}{2q}(1 - \frac{p}{q}).$$

Since

$$\sum_{k=2}^{m} \Delta_k = v_1 - v_m, \; m \geq 2,$$

then

$$\begin{aligned} v_m &= v_1 - \sum_{k=2}^{m} \Delta_k = v_1 - \sum_{k=2}^{m} \left(\frac{p}{q}\right)^{k-2} \Delta_2 \\ &= v_1 - \frac{1-(p/q)^{m-1}}{1-p/q}\Delta_2 \\ &= \frac{v_0}{2q} - \frac{1-(p/q)^{m-1}}{1-p/q}\Delta_2 \\ &= \frac{v_0}{2q} - \frac{1-(p/q)^{m-1}}{1-p/q}\frac{v_0}{2q}(1-\frac{p}{q}) \\ &= \frac{v_0}{2q}\left(\frac{p}{q}\right)^{m-1}. \end{aligned}$$

Therefore,

$$\sum_{m=0}^{\infty} v_m$$

is convergent when $p¡q$. In that case,

$$\sum_{m=0}^{\infty} v_m = v_0 \frac{3-4p}{2-4p}.$$

Setting

$$\pi_k = \frac{v_k}{\sum v_m}$$

gives the stationary distribution,

$$\begin{aligned} \pi_0 &= \frac{2(1-2p)}{3-4p} \\ \pi_1 &= \frac{1-2p}{(1-p)(3-4p)} \\ \pi_k &= \frac{1-2p}{(1-p)(3-4p)}\left(\frac{p}{1-p}\right)^{k-1} \quad k \geq 2. \end{aligned}$$

Note that the mean time to return to state 0 is

$$\frac{1}{\pi_0} = \frac{3-4p}{2(1-2p)}.$$

It can be shown that this MC is null recurrent if $p = 0.5$ and all states are transient if $p > 0.5$.

Example 2. TPM is given by:

$$P = \begin{bmatrix} 1/4 & 3/4 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/6 & 1/6 \\ 0 & 1/4 & 1/2 & 1/4 & 0 \end{bmatrix}$$

States are classified as follows: *(1,2)* is a recurrent class, *3* is an absorbing state (recurrent class with one member), and states *(4,5)* are transient. Then

$$Q = \begin{bmatrix} 1/6 & 1/6 \\ 1/4 & 0 \end{bmatrix}$$

$$R = \begin{bmatrix} 1/3+0 & 1/3 \\ 0+1/4 & 1/2 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/3 \\ 1/4 & 1/2 \end{bmatrix}$$

$$I - Q = \begin{bmatrix} 5/6 & -1/6 \\ -1/4 & 1 \end{bmatrix}$$

$$S = (I-Q)^{-1} = \frac{20}{19}\begin{bmatrix} 1 & 1/6 \\ 1/4 & 5/6 \end{bmatrix} = \begin{bmatrix} 24/19 & 4/19 \\ 6/19 & 20/19 \end{bmatrix}$$

S contains mean number of visits to transient states. So if this MC starts in state 4, then mean number of visits to state 4 is 24/19, mean number of visits to state 5 is 4/19, and mean time to absorption is 28/19. Likewise, if the MC starts in state 5, then mean number of visits to state 4 is 6/19, mean number of visits to state 5 is 20/19, and mean time to absorption is 26/19. Absorption probabilities are

$$SR = \begin{bmatrix} 24/19 & 4/19 \\ 6/19 & 20/19 \end{bmatrix}\begin{bmatrix} 1/3 & 1/3 \\ 1/4 & 1/2 \end{bmatrix} = \begin{bmatrix} 9/19 & 10/19 \\ 7/19 & 12/19 \end{bmatrix}$$

Stationary distribution for recurrent class *(1,2)* is solution to

$$[\pi_1, \pi_2] = [\pi_1, \pi_2]\begin{bmatrix} 1/4 & 3/4 \\ 1/2 & 1/2 \end{bmatrix}, \quad \pi_1 + \pi_2 = 1.$$

Solution is

$$\pi_1 = \frac{2}{5}, \ \pi_2 = \frac{3}{5}.$$

Therefore, mean times between visits to *1, 2* are *5/2, 5/3*, respectively.

Example 3. Branching processes. Suppose each individual in a population is able to produce offspring and let N denote the total number of offspring by an individual. Assume N is independent of the numbers produced by other individuals and has pmf

$$p_j, \; j \geq 0.$$

Let X_0 denote the total number of individuals at time 0, referred to as generation 0, and let X_1 denote the total number of all offspring from generation 0. These offspring are referred to as generation 1. In general, let X_n denote the total number of all offspring from generation *n-1*. Since X_{n+1} depends only on X_n and not on the sizes of previous generations, then X_k, $k \geq 0$, is a MC.

Since $P_{0,0} = 1$ then *0* is an absorbing state. Since offspring from individuals are i.i.d. then

$$P_{i,0} = p_o^i.$$

This implies that if $p_0 > 0$ then there is a positive probability this MC will be absorbed into state 0 and so states $i \geq 1$ are transient. This means that the MC will either be absorbed into state 0 or go to $+\infty$.

Let $Z_i^{(n-1)}$ denote number of offspring of i^{th} individual in generation *n-1* and let μ, σ^2 denote its mean and variance,

$$\begin{aligned} \mu &= E\left(Z_i^{(n-1)}\right) = \sum_{j=0}^{\infty} j p_j, \\ \sigma^2 &= Var\left(Z_i^{(n-1)}\right) = \sum_{j=0}^{\infty} (j-\mu)^2 p_j. \end{aligned}$$

Suppose $X_0 = 1$ and find $E(X_n)$ and $Var(X_n)$. First note that

$$\begin{aligned} E(X_n|X_{n-1} = r) &= E\left[\sum_{i=1}^{X_{n-1}} Z_i^{(n-1)} | X_{n-1} = r\right] \\ &= \sum_{i=1}^{r} E[Z_i^{(n-1)}|X_{n-1} = r] = r\mu. \end{aligned}$$

Hence, $E(X_n|X_{n-1}) = \mu X_{n-1}$ and so

$$E(X_n) = E[E(X_n|X_{n-1})] = \mu E(X_{n-1}), \; n \geq 1.$$

Since $X_0 = 1$, then $E(X_n) = \mu^n$.

Similarly,

$$Var(X_n|X_{n-1} = r) = \sigma^2 X_{n-1} = Var\left(\sum_{i=1}^{r} Z_i^{(n-1)}\right) = \sigma^2 r,$$

so

$$Var(X_n|X_{n-1}) = \sigma^2 X_{n-1}.$$

Therefore,

$$\begin{aligned}
Var(X_n) &= E[Var(X_n|X_{n-1})] + Var[E(X_n|X_{n-1})] \\
&= E[\sigma^2 X_{n-1}] + Var(\mu X_{n-1}) \\
&= \sigma^2 \mu^{n-1} + \mu^2 Var(X_{n-1}).
\end{aligned}$$

Let $a_n = Var(X_n)$. Then

$$\begin{aligned}
a_n &= \sigma^2 \mu^{n-1} + \mu^2 a_{n-1} \\
&= \sigma^2 \mu^{n-1} + \mu^2(\sigma^2 \mu^{n-2} + \mu^2 a_{n-2}) \\
&= \sigma^2 \mu^{n-1} + \sigma^2 \mu^n + \mu^4 a_{n-2} \\
&= \sigma^2 \mu^{n-1} + \sigma^2 \mu^n + \mu^4(\sigma^2 \mu^{n-3} + \mu^2 a_{n-3}) \\
&= \sigma^2 \mu^{n-1} + \sigma^2 \mu^n + \sigma^2 \mu^{n+1} + \mu^4 a_{n-3}.
\end{aligned}$$

Continuing this process gives

$$a_n = \sigma^2 \sum_{k=n-1}^{2n-2} \mu^k.$$

If $\mu = 1$, then $a_n = n\sigma^2$. If $\mu \neq 1$, then

$$a_n = \sigma^2 \frac{\mu^{n-1} - \mu^{2n-1}}{1-\mu} = \sigma^2 \mu^{n-1} \frac{1-\mu^n}{1-\mu}.$$

If $\mu < 1$, then

$$\mu^n = E(X_n) = \sum_{j=1}^{\infty} jP(X_n = j) \geq \sum_{j=1}^{\infty} P(X_n = j)P(X_n \geq 1).$$

This implies that

$$\lim_{n\to\infty} P(X_n \geq 1) = 0,$$

and so,

$$\pi_0 = \lim_{n\to\infty} P(X_n = 0) = 1.$$

It can be shown (proof is more complicated) that $\pi_0 = 1$ if $\mu = 1$. Now suppose $\mu > 1$. Then first step analysis gives

$$\pi_0 = \sum_{j=0}^{\infty} P(\text{pop. dies out}|X_1 = j)p_j.$$

Now, the population dies out iff the population generated by each offspring dies out. Since the number of offspring for an individual is independent of number of offspring for other individuals, then

$$P(\text{pop. dies out}|X_1 = j) = \pi_0^j.$$

Hence,

$$\pi_0 = \sum_{j=0}^{\infty} \pi_0^j p_j.$$

This solution may not be unique, so we take π_0 to be the smallest positive solution.

Suppose for example that

$$p_0 = 0.5,\ p_1 = 0.25,\ p_2 = 0.25.$$

Then $\mu = 0.75$ and so $\pi_0 = 1$. Now suppose

$$p_0 = 0.25,\ p_1 = 0.25,\ p_2 = 0.5.$$

Then $\mu = 1.25$ and π_0 satisfies

$$\begin{aligned} \pi_0 &= 0.25 + 0.25\pi_0 + 0.5\pi_0^2 \\ &\Rightarrow 0.5\pi_0^2 - 0.75\pi_0 + 0.25 = 0 \\ &\Rightarrow 2\pi_0^2 - 3\pi_0 + 1 = 0. \end{aligned}$$

Solutions are

$$\frac{3 \pm \sqrt{9-8}}{4} = \frac{1}{2},\ 1,$$

so $\pi_0 = 0.5$.

Example 4. List replacement algorithms. A data archive has 5 large drives arranged in a stack. When a request is made for a file stored on a particular drive, that drive is retrieved from the stack, the file is extracted, and then the drive is returned to the stack but not necessarily in the same position. Suppose drive retrieval time is proportional to the position of the drive in the stack. In that case we would like to have the most frequently requested drive at the top of the stack to minimize retrieval time. So let's consider a replacement algorithm in which a requested drive is returned to the top of the stack and drives that were above it are moved down one position. We can analyze this algorithm by tracking a particular drive, say drive A, and using the following model. Selections are independent, probability of selecting A is p and the probability of selecting each of the other items is $(1-p)/4 = q$ (others are equally likely). Let X_n denote position of drive A after the n^{th} selection where position 1 denotes the top of the stack. Note that if A is at position 1, then it will stay there if it is selected, otherwise it will move down to position 2. If A is at position k, $1 < k \leq 5$, then it will move to position 1 if it is selected, it will stay at position k if an item above it is selected, and it will move to position $k+1$ if an item below it is selected. The TPM therefore is

$$P = \begin{bmatrix} p & 1-p & 0 & 0 & 0 \\ p & q & 3q & 0 & 0 \\ p & 0 & 2q & 2q & 0 \\ p & 0 & 0 & 3q & q \\ p & 0 & 0 & 0 & 1-p \end{bmatrix}$$

The stationary distribution is given by

$$\begin{aligned} \pi_1 &= p\pi_1 + p\pi_2 + p\pi_3 + p\pi_4 + p\pi_5 = p \\ \pi_2 &= (1-p)\pi_1 + q\pi_2 \\ \pi_3 &= 3q\pi_2 + 2q\pi_3 \\ \pi_4 &= 2q\pi_3 + 3q\pi_4 \\ \pi_5 &= q\pi_4 + (1-p)\pi_5. \end{aligned}$$

We are mainly interested in π_1, the long-term proportion of times drive A is at position 1. For this algorithm, $\pi_1 = p$.

Now consider a different replacement algorithm: the requested drive is returned to position 1 if it was at 1, otherwise it is returned to 1 position above its previous position. The TPM for this algorithm is

$$P = \begin{bmatrix} 1-q & q & 0 & 0 & 0 \\ p & 1-p-q & q & 0 & 0 \\ 0 & p & 1-p-q & q & 0 \\ 0 & 0 & p & 1-p-q & q \\ 0 & 0 & 0 & p & 1-p \end{bmatrix}$$

Stationary distribution is given by

$$\begin{aligned} \pi_1 &= (1-q)\pi_1 + p\pi_2 \Rightarrow \pi_2 = \frac{q}{p}\pi_1 \\ \pi_2 &= q\pi_1 + (1-p-q)\pi_2 + p\pi_3 \Rightarrow \pi_3 = \frac{q^2}{p^2}\pi_1 \\ \pi_3 &= q\pi_2 + (1-p-q)\pi_3 + p\pi_4 \Rightarrow \pi_4 = \frac{q^3}{p^3}\pi_1 \\ \pi_5 &= q\pi_4 + (1-p)\pi_5 \Rightarrow \pi_5 = \frac{q^4}{p^4}\pi_1 \end{aligned}$$

Also, the equation $\sum \pi_k = 1$ gives

$$1 = \pi_1 \sum_{k=0}^{4} \left(\frac{q}{p}\right)^k.$$

If $p = q$ then

$$\pi_1 = \frac{1}{5} = \pi_2 = \pi_3 = \pi_4 = \pi_5.$$

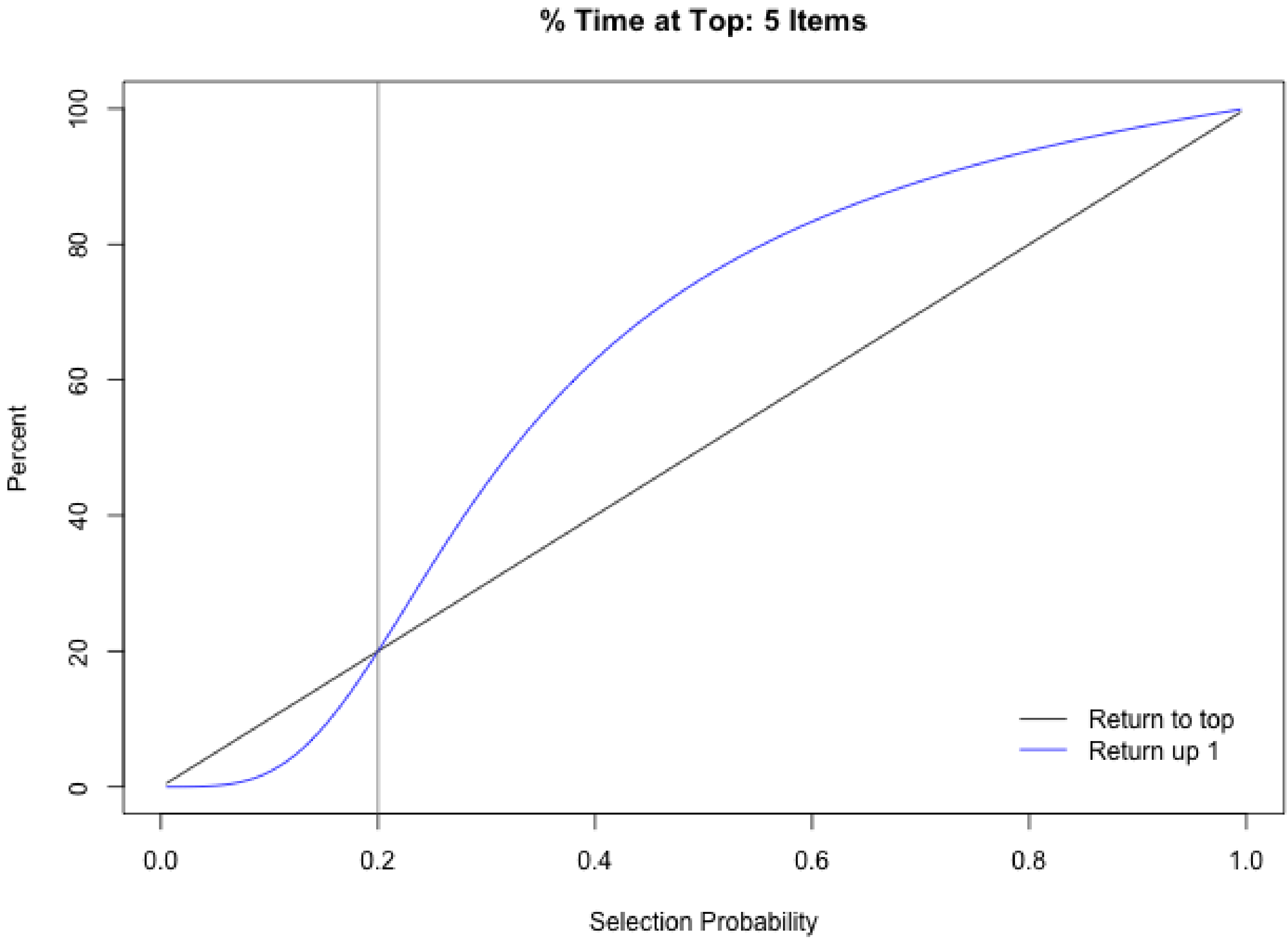
If $p \neq q$ then

$$\pi_1 = \frac{1}{\sum_{k=0}^{4} \left(\frac{q}{p}\right)^k} = \frac{1 - \frac{p}{q}}{1 - \left(\frac{q}{p}\right)^5}.$$

and

$$\pi_k = \left(\frac{q}{p}\right)^{k-1} \frac{1 - \frac{p}{q}}{1 - \left(\frac{q}{p}\right)^5}, \quad 1 \leq k \leq 5.$$

Here is a graphical comparison of these two algorithms. It shows that the second algorithm performs better than the first. When $p < 0.2$, drive A is less likely to be selected than the other drives. In this case Algorithm 2 results in a lower proportion of time spent by A at the top. However, when $p > 0.2$, drive A is more likely to be selected and Algorithm 2 results in a higher proportion of time spent by A at the top.



Homework Assignments

Homework Assignment 1

Due date: Jan. 22, 2015.

Text, pages 15-20

1. Exercise 8
2. Exercise 10
3. Exercise 12 (note: assume experiment is repeated independently
4. Exercise 20

5. Exercise 21
6. Exercise 22
7. Exercise 29
8. Exercise 42

Solutions for Homework Assignment 1

1. Exercise 8.

$$1 \geq P(E \bigcup F) = P(E) + P(F) - P(E \bigcap F).$$

This implies that

$$P(E \bigcap F) \geq P(E) + P(F) - 1.$$

2. Exercise 10. Proof by induction. Result holds trivially for *n=1*. Now assume

$$P\left(\bigcup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} P(E_i),$$

and show result holds for *n+1*.

$$\begin{aligned}
P\left(\bigcup_{i=1}^{n+1} E_i\right) &= P\left(\bigcup_{i=1}^{n} E_i \bigcup E_{n+1}\right) \\
&= \sum_{i=1}^{n} P(E_i) + P(E_{n+1}) - P\left(\bigcup_{i=1}^{n} E_i \bigcap E_{n+1}\right) \\
&= \sum_{i=1}^{n+1} P(E_i) - P\left(\bigcup_{i=1}^{n} E_i \bigcap E_{n+1}\right) \\
&\leq \sum_{i=1}^{n+1} P(E_i).
\end{aligned}$$

3. Exercise 12 (note: assume experiment is repeated independently). Let

$$A = (E \bigcup F)^c$$

and let

$$p = P(E \bigcup F) = P(E) + P(F).$$

Then *P(A) = 1-p*. Sample space is
$E, F, AE, AF, AAE, AAF, AAAE, AAAF, ...$, so

$$P(\overbrace{A \cdots A}^{n-1} E) = [P(A)]^{n-1} P(E),$$

and so

$$\begin{aligned}
P(\text{E } occurs\ before\ F) &= \sum_{n=1}^{\infty} P(\overbrace{A \cdots A}^{n-1} E) \\
&= \sum_{n=1}^{\infty} (1-p)^{n-1} P(E) \\
&= \frac{P(E)}{p} \\
&= \frac{P(E)}{P(E)+P(F)}.
\end{aligned}$$

4. Exercise 20

$$\begin{aligned}
P(same\ on\ exactly\ 2\ of\ 3\ dice) &= \sum_{k=1}^{6} [P(kkk^c) + P(kk^ck) + P(k^ckk)] \\
&= \sum_{k=1}^{6} [3(\frac{1}{6})(\frac{1}{6})(\frac{5}{6})] \\
&= \frac{15}{36}
\end{aligned}$$

5. Exercise 21

$$P(CB|M) = 0.05, \quad P(CB|F) = 0.0025, \quad P(M) = P(F) = 0.5$$

So,

$$\begin{aligned}
P(M|CB) &= \frac{P(M \bigcap CB)}{P(CB)} \\
&= \frac{P(M \bigcap CB)}{P(CB|M)P(M) + P(CB|F)P(F)} \\
&= \frac{(0.05)(0.5)}{(0.05)(0.5) + (0.0025)(0.5)} \\
&= \frac{500}{525} = \frac{20}{21}
\end{aligned}$$

6. Exercise 22. Note that the game must end after *2n* games. This implies that after *2k* games, *k=1,2,...,n-1*, they must be tied, and that games *2n-1* and *2n* must be either AA or BB. Let

$$q = p^2 + (1-p)^2 = 1 - 2p(1-p).$$

Then

$$\begin{aligned} P(\textit{game ends at } 2n) &= [2p(1-p)]^{n-1}[p^2+(1-p)^2] \\ &= q^{n-1}(1-q), \end{aligned}$$

and so

$$\begin{aligned} P(\text{A } \textit{wins}) &= \sum_{n=1}^{\infty} P(\textit{game ends at } 2n \textit{ and } A \textit{ wins}) \\ &= \sum_{n=1}^{\infty} [2p(1-p)]^{n-1}p^2 \\ &= \frac{p^2}{p^2+(1-p)^2}. \end{aligned}$$

7. Exercise 29
a) If E,F are mutually exclusive, then

$$P(E|F) = \frac{P(E \bigcap F)}{P(F)}$$

if $P(F) > 0$, otherwise $P(E|F)$ is undefined.
b) If $E \subset F$, then

$$P(E \bigcap F) = P(E) = 0.6,$$

and

$$P(F) \geq P(E) = 0.6.$$

Hence,

$$P(E|F) = \frac{P(E)}{P(F)} = \frac{0.6}{P(F)}.$$

c) If $F \subset E$, then $P(E \bigcap F) = P(F)$ and so $P(E|F) = 1$.

8. Exercise 42

$$P(H|2H \textit{ coin}) = 1, \quad P(H|\text{f}\textit{air coin}) = 1/2, \quad P(H|\text{b}\textit{iased coin}) = 3/4,$$

so

$$\begin{aligned} P(H) &= (1)(1/3) + (1/2)(1/3) + (3/4)(1/3) \\ &= 1/3 + 1/6 + 1/4 = 3/4. \end{aligned}$$

Therefore,

$$P(2H\ coin|H) = \frac{P(H|2H\ coin)P(2H\ coin)}{P(H)} = \frac{1/3}{3/4} = 4/9.$$

$$P(\text{b}iased\ coin|H) = \frac{P(H|\text{b}iased\ coin)P(\text{b}iased\ coin)}{P(H)} = \frac{1/4}{3/4} = 1/3.$$

$$P(\text{f}air\ coin|H) = \frac{P(H|\text{f}air\ coin)P(\text{f}air\ coin)}{P(H)} = \frac{1/6}{3/4} = 2/9.$$

Homework Assignment 2

Due date: Feb. 5, 2018.

Problem set on pages 87-95, 10th edition (pages 80-90, 11th edition)

1. Text, p. 87: 11
2. Text, p. 87: 13
3. Text, p. 88: 25
4. Text, p. 88: 26
5. Text, p. 90: 37
6. Text, p. 95: 79 (in 11th Edition this is 83)
7. Let $X_1, \cdots, X_n$ be independent r.v.'s each with the same exponential distribution with rate λ. Let $Y = \min(X_1, \cdots, X_n)$. Show that Y has an exponential distribution with rate $n\lambda$.
8. Let X have a Poisson distribution with mean λ. Find $P(X \text{ is odd})$. **Hint**: express

$$h(\lambda) = P(X \text{ is odd})$$

as a series in λ, obtain

$$h'(\lambda) = \frac{\partial}{\partial \lambda} h(\lambda),$$

show that

$$\lim_{\lambda \to 0} h(\lambda) = 0,$$

and assume that $h(\lambda)$ has the form

$$h(\lambda) = ae^{b\lambda} + c,$$

Then solve for the unknowns, a, b, c.

Solutions for Homework Assignment 2

1. (P. 87, 11) The draws are independent since the balls are replaced after each draw. So number of white balls in 4 draws has Binomial(4,.5) distribution.

$$P(X = 2) = \binom{4}{2}(.5)^4 = 3/8.$$

2. (P. 87, 13) Under the assumption of no ESP, probability of correct prediction of toss is 0.5, so expected number of correct predictions would be 5. Getting 7 correct is higher than expected and so is some evidence that there may be some ESP. Getting more than 7 correct would be even stronger evidence, so the relevant event to judge the strength of evidence for ESP is the event: getting at least 7 correct. Under the assumption of no ESP, the number correct has Binomial(10,.5) distribution.

$$\begin{aligned} P(X \geq 7) &= \sum_{k=7}^{10} \binom{10}{k}(.5)^{10} \\ &= \frac{176}{1024} = \frac{11}{64} = 0.172. \end{aligned}$$

3. (P. 88, 25) Let N denote the number of games played. Then the event that 7 games are played is the event that each team wins 3 games after 6 games are played. This probability is the Binomial probability of 3 successes out of 6 trials, and so is given by

$$\binom{6}{3}p^3(1-p)^3 = 20[p(1-p)]^3.$$

The function $h(p) = p(1-p)$, $0 < p < 1$, is maximized at the solution to $1 - 2p = 0$ (solution to derivative of h equals 0), which has solution $p = 1/2$. This is the maximum since 2^{nd} derivative of h is -2.

4. (P. 88, 26) For $i = 2$,

$$E(N) = 2P(N = 2) + 3P(N = 3) = 2(p^2 + (1-p)^2) + 3(2p(1-p)) = 2 + 2p(1-p).$$

For $i = 3$,

$$\begin{aligned} E(N) &= 3P(N = 3) + 4P(N = 4) + 5P(N = 5) \\ &= 3(p^3 + (1-p)^3) + 4(3p^3(1-p) + 3p(1-p)^3) + 5(6p^2(1-p)^2) \\ &= 3 + 3p(1-p) + 6p^2(1-p)^2. \end{aligned}$$

Since both are non-decreasing functions of $p(1-p)$, then they are maximized at $p = 1/2$ as shown in the previous problem.

5. (P. 90, 37) For $0 \le x \le 1$,

$$\begin{aligned} P(\max(X_1, \cdots, X_n) \le x) &= P(X_1 \le x, \cdots, X_n \le x) \\ &= \prod_{k=1}^{n} P(X_k \le x) \\ &= x^n. \end{aligned}$$

So, $F_M(x) = x^n$, $0 \le x \le 1$. Density function is the derivative,

$$f_M(x) = nx^{n-1}, \quad 0 \le x \le 1.$$

6. (P. 95, 79) We assume conditions hold under which the derivative can be passed inside the expectation. Then

$$K'(t) = \frac{d}{dt}\log(E[e^{tX}]) = \frac{E[Xe^{tX}]}{E[e^{tx}]}$$

and so

$$K'(0) = \frac{E(X)}{E(1)} = E(X).$$

Also,

$$\begin{aligned} K''(t) &= \frac{d}{dt}\frac{E[Xe^{tX}]}{E[e^{tx}]} \\ &= \frac{E[e^{tX}]E[X^2e^{tX}] - E[Xe^{tX}]E[Xe^{tX}]}{(E[e^{tX}])^2}. \end{aligned}$$

and so

$$K''(0) = \frac{E[X^2]E[1] - (E[X])^2}{(E[1])^2} = E[X^2] - (E[X])^2 = \mathrm{Var}(X).$$

7. (Extra)

$$\begin{aligned} P(Y > y) &= P(\min(X_1, \cdots, X_n) > y) \\ &= P(X_1 > y, \cdots, X_n > y) \\ &= \prod_{k=1}^{n} P(X_k > y) \\ &= e^{-ny\lambda}. \end{aligned}$$

So, $F_Y(y) = 1 - e^{-n\lambda y}$, which is d.f. of exponential distribution with rate $n\lambda$. Density is the dervative,

$$f_Y(y) = n\lambda e^{-n\lambda y}, \; y \ge 0.$$

8. (Extra 2) First note that

$$h(\lambda) = P(X \text{ is odd}) = \sum_{k=0}^{\infty} \frac{\lambda^{2k+1}}{(2k+1)!} e^{-\lambda},$$

and

$$P(X \text{ is even}) = 1 - h(\lambda) = \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} e^{-\lambda}.$$

Also,

$$\begin{aligned} h'(\lambda) &= \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} e^{-\lambda} - h(\lambda) \\ &= P(X \text{ is even}) - h(\lambda) \\ &= 1 - 2h(\lambda). \end{aligned}$$

This gives the differential equation,

$$h'(\lambda) + 2h(\lambda) = 1,$$

with boundary condition, $h(0) = 0$. Substituting the general solution,

$$h(\lambda) = ae^{b\lambda} + c,$$

gives

$$\begin{aligned} abe^{b\lambda} + 2ae^{b\lambda} + 2c &= 1 \\ a + c &= 0. \end{aligned}$$

The solution is

$$c = \frac{1}{2},\ a = -\frac{1}{2},\ b = -2.$$

So,

$$P(X \text{ is odd}) = \frac{1}{2}(1 - e^{-2\lambda}).$$

Alternatively, note that

$$\begin{aligned}
P(X \text{ is odd}) &= P(X \text{ is odd}) - P(X \text{ is even}) + P(X \text{ is even}) \\
&= \sum_{k=0}^{\infty} \frac{\lambda^{2k+1}}{(2k+1)!} e^{-\lambda} - \sum_{k=0}^{\infty} \frac{\lambda^{2k}}{(2k)!} e^{-\lambda} + 1 - P(X \text{ is odd}) \\
&= -e^{-\lambda} \sum_{k=0}^{\infty} \frac{(-\lambda)^{2k+1}}{(2k+1)!} - e^{-\lambda} \sum_{k=0}^{\infty} \frac{(-\lambda)^{2k}}{(2k)!} + 1 - P(X \text{ is odd}) \\
&= -e^{-\lambda} \sum_{n=0}^{\infty} \frac{(-\lambda)^{n}}{n!} + 1 - P(X \text{ is odd}) \\
&= -e^{-2\lambda} + 1 - P(X \text{ is odd}).
\end{aligned}$$

Therefore,

$$2P(X \text{ is odd}) = 1 - e^{-2\lambda},$$

and so,

$$P(X \text{ is odd}) = \frac{1}{2}(1 - e^{-2\lambda}).$$

Homework Assignment 3

Due date: Feb. 19, 2018

Exercises for Chapter 3, p. 163 (11th edition), p. 173 (10th edition)

1. Problem 3
2. Problem 4
3. Problem 11
4. Problem 15
5. Problem 17
6. Problem 24 (**hint**: use first step analysis, that is, condition on the outcome of the first coin flip)
7. Problem 40

Solutions for Homework Assignment 3

1. (p.173, 3).

$$\begin{aligned}
E(X|Y=1) &= \frac{1\cdot\frac{1}{9}+2\cdot\frac{1}{3}+3\cdot\frac{1}{9}}{\frac{1}{9}+\frac{1}{3}+\frac{1}{9}} \\
&= \frac{10/9}{5/9} \\
&= 2.
\end{aligned}$$

$$\begin{aligned}
E(X|Y=2) &= \frac{1\cdot\frac{1}{9}+2\cdot 0+3\cdot\frac{1}{18}}{\frac{1}{9}+0+\frac{1}{18}} \\
&= \frac{5/18}{3/18} \\
&= \frac{5}{3}.
\end{aligned}$$

$$\begin{aligned}
E(X|Y=3) &= \frac{1\cdot 0+2\cdot\frac{1}{6}+3\cdot\frac{1}{9}}{0+\frac{1}{6}+\frac{1}{9}} \\
&= \frac{12/18}{5/18} \\
&= 2.4.
\end{aligned}$$

2. (p.173, 4). Not independent because, for example, $P(X=1,Y=3)=0$, but

$$P(X=1)=2/9,\ P(Y=3)=5/18,$$

and so

$$P(X=1)P(Y=3)=5/81\neq P(X=1,Y=3).$$

3. (p.174, 11).

$$\begin{aligned}
f_Y(y) &= \int_{-y}^{y} f(x,y)dx \\
&= \frac{1}{8}e^{-y}\int_{-y}^{y}(y^2-x^2)dx \\
&= \frac{1}{8}e^{-y}[y^2x-\frac{1}{3}x^3]|_{-y}^{y} \\
&= \frac{1}{8}e^{-y}\frac{4}{3}y^3 \\
&= \frac{1}{6}y^3e^{-y}.
\end{aligned}$$

This is $Gamma(4,1)$ density. So,

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{\frac{1}{8}e^{-y}(y^2-x^2)}{\frac{1}{6}y^3e^{-y}} \\ &= \frac{3}{4}y^{-3}(y^2-x^2), \end{aligned}$$

for $|x| \leq y,\ 0 < y < \infty$. Since this conditional density is symmetric about $x = 0$ for each $0 < y < \infty$, then $E(X|Y = y) = 0$.

4. (p.174, 15).

$$f_Y(y) = \int_0^y f(x,y)dx = e^{-y},$$

$0 < y < \infty$. This is *exponential*(1) density, so

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = y^{-1},$$

$0 < x < y,\quad 0 < y < \infty$. This is Uniform$(0, y)$, and so

$$E(X^2|Y = y) = y^{-1}\int_0^y x^2dx = \frac{1}{3}y^2.$$

5. (p.175, 17). The joint density of X, Y is given by

$$f_{X,Y}(i,y) = \frac{y^i}{i!}e^{-y}Cy^{s-1}e^{-\alpha y} = \frac{Cy^i}{i!}y^{i+s-1}e^{-(1+\alpha)y}.$$

Hence,

$$\begin{aligned} P(X = i) &= \int f_{X,Y}(i,y)dy \\ &= \frac{Cy^i}{i!}\frac{\Gamma(i+s)}{(1+\alpha)^{s+i}}, \end{aligned}$$

and so,

$$\begin{aligned} f_{Y|X}(y|i) &= \frac{f_{X,Y}(i,y)}{P(X = i)} \\ &= Dy^{i+s-1}e^{-(1+\alpha)y}, \end{aligned}$$

where D is a constant that does not depend on y and is the value that makes this conditional density integrate to 1. This implies that the conditional density is $Gamma(i + s, 1 + \alpha)$.

6. (p.176, 24). Let N_{ij} denote the number of trials required to obtain at least i heads and at least j tails in a sequence of i.i.d. Bernoulli trials. Let $X_1, X_2, \cdots$ denote the Bernoulli trials. Note that N_{10} has a geometric distribution with success probability p and N_{01} has a geometric distribution with success probability $1-p$.

   a) Condition on X_1:

$$E(N_{11}) = E(N_{11}|X_1 = H)P(X_1 = H) + E(N_{11}|X_1 = T)P(X_1 = T)$$

   Note that if $X_1 = H$ then $N_{11} = 1 + N^*_{01}$ where N^*_{01} is the number of trials among $X_2, X_3, \cdots$ until the first tail appears. Likewise, if $X_1 = T$ then $N_{11} = 1 + N^*_{10}$ where N^*_{10} is the number of trials among $X_2, X_3, \cdots$ until the first head appears. Therefore N^*_{01}, N^*_{10} are each independent of X_1 and so

$$\begin{aligned} E(N_{11}) &= E(1 + N^*_{01}|X_1 = H)P(X_1 = H) + E(1 + N^*_{10}|X_1 = T)P(X_1 = T) \\ &= E(1 + N^*_{01})p + E(1 + N^*_{10})(1-p) \\ &= p(1 + 1/(1-p)) + (1-p)(1 + 1/p) \\ &= \frac{1}{p} + \frac{p}{1-p} \\ &= \frac{1 - p + p^2}{p(1-p)}. \end{aligned}$$

   b) Let M_{11} denote the number of heads during N_{11} trials. Then, as above,

$$\begin{aligned} E(M_{11}) &= E(M_{11}|X_1 = H)P(X_1 = H) + E(M_{11}|X_1 = T)P(X_1 = T) \\ &= E(1 + N_{01} - 1)p + 1(1-p) \\ &= (1-p) + \frac{p}{1-p} \\ &= \frac{1 - p + p^2}{1-p} \end{aligned}$$

   c) Let L_{11} denote the number of tails during N_{11} trials. Then $L_{11} = N_{11} - M_{11}$, and so,

$$\begin{aligned} E(L_{11}) &= E(N_{11}) - E(M_{11}) \\ &= \frac{1 - p + p^2}{p}. \end{aligned}$$

   d) First note that $E(N_{20}) = 2/p$. Similar to part a) above,

$$\begin{aligned} E(N_{21}) &= E(N_{21}|X_1 = H)P(X_1 = H) + E(N_{21}|X_1 = T)P(X_1 = T) \\ &= E(1 + N^*_{11})p + E(1 + N^*_{20})(1-p) \\ &= p + \frac{1 - p + p^2}{1-p} + (1-p) + \frac{2(1-p)}{p} \\ &= \frac{p^2}{1-p} + \frac{2}{p}. \end{aligned}$$

7. (p.179, 40).

a) Each time the prisoner returns to the cell, he is unable to determine which door he selected last time he was in the cell. So each time he returns, he is faced with the same situation as before. Let $X_k,\ k \geq 1$ denote the door selected on the k^{th} attempt. Then freedom occurs the first time the prisoner selects door 3. Let T denote the time until freedom. Note that if $X_1 = 1$, then he travels for 2 days and then is back where he started. Let T^* denote the number of days until freedom after his first return to the cell. Then

$$\begin{aligned} E(T|X_1 = 1) &= E(T^* + 2) \\ E(T|X_1 = 2) &= E(T^* + 3) \\ E(T|X_1 = 3) &= 0 \end{aligned}$$

Let $E(T) = m$. Since T and T^* have the same distribution, then

$$\begin{aligned} E(T) = m &= E(E(T|X_1)) \\ &= E(T|X_1 = 1)p_1 + E(T|X_1 = 2)p_2 + E(T|X_1 = 3)p_3 \\ &= (E(T^*) + 2)p_1 + (E(T^*) + 3)p_2 + 0 \\ &= (m + 2)p_1 + (m + 3)p_2 \end{aligned}$$

This equation is solved by

$$E(T) = m = \frac{2p_1 + 3p_2}{1 - p_1 - p_2} = \frac{2p_1 + 3p_2}{p_3}.$$

For this problem it is easier to obtain the variance directly from

$$Var(T) = E(T^2) - (E(T))^2$$

We can obtain $r = E(T^2)$ as above.

$$\begin{aligned} E(T^2) &= E(E(T^2|X_1)) \\ &= E(T^2|X_1 = 1)p_1 + E(T^2|X_1 = 2)p_2 + E(T^2|X_1 = 3)p_3 \\ &= E[(T^* + 2)^2])p_1 + E[(T^* + 3)^2]p_2 + 0 \\ &= p_1[E(T^2) + 4E(T) + 4] + p_2[E(T^2) + 6E(T) + 9] \end{aligned}$$

This gives

$$r = p_1 r + p_2 r + m(4p_1 + 6p_2) + 4p_1 + 9p_2,$$

and so

$$r = 2m^2 + \frac{4p_1 + 9p_2}{p_3},$$

Finally,

$$\mathrm{Var}(T) = r - m^2 = m^2 + \frac{4p_1 + 9p_2}{p_3}$$

For $p = (.5, .3, .2)$ we have

$$E(T) = 9.5, \quad \mathrm{Var}(T) = 113.75$$

b) Since the prisoner knows which door he selected previously, we can just consider the possible door choices that lead to freedom. They are:

$$3,\ 13,\ 23,\ 123,\ 213$$

The total times for these choices are:

$$0,\ 2,\ 3,\ 5,\ 5$$

and the corresponding probabilities are:

$$1/3,\ (1/3)(1/2),\ (1/3)(1/2),\ (1/3)(1/2)(1),\ (1/3)(1/2)(1)$$

Therefore, the p.m.f of T is

$$P(T=0) = 1/3,\ P(T=2) = 1/6,\ P(T=3) = 1/6,\ P(T=5) = 1/3,$$

and so

$$E(T) = 0 + 2/6 + 3/6 + 5/3 = 2.5,$$

$$\mathrm{Var}(T) = 0 + 4/6 + 9/6 + 25/3 - 25/4 = 51/12 = 4.25.$$

Homework Assignment 4

Due date: Feb. 28, 2018

1. Text, p. 276: 6 (11th edition: p. 261)
2. Text, p. 277: 14 (11th edition: p. 262)
3. Text, p. 277: 16 (11th edition: p. 263)
4. Text, p. 278: 20 (11th edition: p. 263)
5. Text, p. 278: 21 (11th edition: p. 263)
6. Text, p. 278: 22 **Hint, continued**. Use *modulo 13* arithmetic for the states of Y_n (11th edition: p. 263)
7. A computer program consists of a sequence of addresses that must be fetched from one of three locations, local memory (RAM), cache memory, or virtual memory (swap). A simple memory model can be expressed as follows: if the current address is in RAM, then the next address will be in RAM, cache, or swap with probabilities $0.8, 0.15, 0.05$, respectively; if the current address is in cache, then the next address will be in RAM, cache, or swap with probabilities $0.05, 0.9, 0.05$, respectively; if the current address is in swap, then the next address will be in RAM, cache, or swap with probabilities $0.2, 0.3, 0.5$, respectively. Let X_n denote the location of the n^{th} address and assume that $X_n,\ n \geq 1$ is a Markov chain. Find the proportion of time spent in each memory location over a long period of time.
8. Consider a component that begins operation with 0 damage. Suppose that at the end of a period of operation, it has accumulated damage N_1, where

   $$P(N_1 = k) = pq^k,\ k \geq 0,$$

   $0 < p < 1$, and $q = 1 - p$. If $N_1 \geq m$, where $m > 0$ is some fixed integer, then the component is replaced with an identical spare so that the damage at the beginning of the next period of operation for the component in use would be 0. Otherwise, the component begins the next period of operation with damage N_1. Damage is cumulative. That is, damage that occurs during a period of operation is added to the damage the component had at the beginning of the period, with the understanding that if the cumulative damage is m or higher at the end of a period, the component is replaced. Let N_k denote the damage that occurs during period k and assume that $\{N_k,\ k \geq 1\}$ are i.i.d. r.v.'s having the same geometric distribution given for N_1. Let $X_n,\ n \geq 1$ denote the damage of the component in use at the beginning of period n, and note that

   $$\begin{aligned} X_{k+1} &= X_k + N_k,\ \text{if } X_k + N_k \leq m - 1, \\ &\quad 0,\ \text{otherwise.} \end{aligned}$$

(a) What is the transition probability matrix of this Markov chain?

(b) Find its stationary distribution.

(c) What is the mean time between visits to state 0?

Note that the states are $0, 1, \cdots, m-1$.

Solutions for Homework Assignment 4

1. (p. 276, 6). Proof is by induction. The result holds trivially for $n = 1$. Now suppose it holds for n. From the Chapman-Kolmogorov equations,

$$
\begin{aligned}
P^{(n+1)} &= P^{(n)}P \\
&= \begin{bmatrix} \frac{1}{2}+\frac{1}{2}(2p-1)^n & \frac{1}{2}-\frac{1}{2}(2p-1)^n \\ \frac{1}{2}-\frac{1}{2}(2p-1)^n & \frac{1}{2}+\frac{1}{2}(2p-1)^n \end{bmatrix} \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix} \\
&= \begin{bmatrix} \frac{p}{2}+\frac{p}{2}(2p-1)^n+\frac{1-p}{2}-\frac{1-p}{2}(2p-1)^n & \frac{1-p}{2}+\frac{1-p}{2}(2p-1)^n+\frac{p}{2}-\frac{p}{2}(2p-1)^n \\ \frac{p}{2}-\frac{p}{2}(2p-1)^n+\frac{1-p}{2}+\frac{1-p}{2}(2p-1)^n & \frac{1-p}{2}-\frac{1-p}{2}(2p-1)^n+\frac{p}{2}+\frac{p}{2}(2p-1)^n \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{2}+\frac{2p-1}{2}(2p-1)^n & \frac{1}{2}-\frac{2p-1}{2}(2p-1)^n \\ \frac{1}{2}-\frac{2p-1}{2}(2p-1)^n & \frac{1}{2}+\frac{2p-1}{2}(2p-1)^n \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{2}+\frac{1}{2}(2p-1)^{n+1} & \frac{1}{2}-\frac{1}{2}(2p-1)^{n+1} \\ \frac{1}{2}-\frac{1}{2}(2p-1)^{n+1} & \frac{1}{2}+\frac{1}{2}(2p-1)^{n+1} \end{bmatrix}
\end{aligned}
$$

So the result holds for all n by induction.

2. (p. 277, 14).

   a) All states communicate so they are recurrent.

   b) All states communicate so they are recurrent.

   c) States 1,3 communicate so they are recurrent. States 4,5 communicate so they are recurrent. There is a positive probability of making a transition from state 2 into state 1, which is recurrent, so state 2 must be transient.

   d) States 1,2 communicate so they are recurrent. State 3 is an absorbing state. There is a positive probability of making a transition from state 4 to state 3, so state 4 is a transient state. There is a positive probability of making a transition from state 5 to state 1, so state 5 is a transient state.

3. (p. 277, 16). Let $C(i)$ denote the recurrent class to which i belongs. Suppose there exists state $j \notin C(i)$. If $i \nrightarrow j$, then

$$P^n_{ij} = 0, \ \forall n \geq 1.$$

Now suppose $i \to j$. We already have seen that if state i is recurrent and if $i \to j$, then state j must be recurrent and $i \leftrightarrow j$. This would imply $j \in C(i)$, a contradiction. Therefore, if state i and state j do not communicate, then state j cannot be accessible from state i, that is, $P^n_{ij} = 0$, $\forall n \geq 1$.

4. (p. 278, 20). Let $\underline{v}$ denote a vector of 1's and let P_k denote the k-th column of P. Then

$$\underline{v}P_k = \sum_{j=0}^{M} P_{jk} = 1 = v_k$$

since P is doubly stochastic. Therefore, $\underline{v}$ is a solution to $\underline{v} = \underline{v}P$ and so

$$\pi_k = \frac{1}{\sum_{j=0}^{m} 1} = \frac{1}{M+1}.$$

5. (p. 278, 21).

a) The TPM is

$$P = \begin{bmatrix} 1-3\alpha & \alpha & \alpha & \alpha \\ \alpha & 1-3\alpha & \alpha & \alpha \\ \alpha & \alpha & 1-3\alpha & \alpha \\ \alpha & \alpha & \alpha & 1-3\alpha \end{bmatrix}.$$

Note that for $n = 1$,

$$P_{i,i}^n = 1 - 3\alpha = \frac{1}{4} + \frac{3}{4}(1-4\alpha)^n.$$

and for $i \neq j$, $n = 1$,

$$P_{i,j}^n = \alpha = \frac{1}{4}[1 - (1-4\alpha)^n].$$

Now assume these hold for n. Then

$$\begin{aligned} P_{i,i}^{n+1} &= P^n[i,]P[,i] \\ &= [\frac{1}{4} + \frac{3}{4}(1-4\alpha)^n](1-3\alpha) + 3\alpha[1-(1-4\alpha)^n] \\ &= \frac{1}{4}(1-3\alpha) + \frac{3}{4}(1-3\alpha)(1-4\alpha)^n + \frac{1}{4}3\alpha - \frac{1}{4}3\alpha(1-4\alpha)^n \\ &= \frac{1}{4} + \frac{3}{4}(1-4\alpha)^n[1-3\alpha-\alpha] \\ &= \frac{1}{4} + \frac{3}{4}(1-4\alpha)^{n+1}. \end{aligned}$$

Also, for $i \neq j$,

$$\begin{aligned} P_{i,j}^{n+1} &= P^n[i,]P[,j] \\ &= [\frac{1}{4} + \frac{3}{4}(1-4\alpha)^n]\alpha + \frac{1}{4}[1-(1-4\alpha)^n](1-3\alpha) + 2\alpha\frac{1}{4}[1-(1-4\alpha)^n] \\ &= \frac{1}{4} + \frac{1}{4}(1-4\alpha)^n[3\alpha - 1 + 3\alpha - 2\alpha] \\ &= \frac{1}{4}[1-(1-4\alpha)^{n+1}]. \end{aligned}$$

So the result holds for all n by induction.

b) This is a doubly stochastic matrix, so from the previous problem, $\pi_i = 1/4$, $1 \leq i \leq 4$.

6. (p. 278, 22). Define a Markov chain by

$$X_n = Y_n \text{ (modulo 13)}.$$

Then the states of this MC are $0, 1, \cdots, 12$ and the event that the sum of the dice is a multiple of 13 corresponds to state 0 of this MC. Also, for $i = 0, 1, \cdots, 6$

$$P_{i,j} = P(X_{n+1} = j | X_n = i) = 1/6, \ \ j = i+1, \cdots, i+6,$$

and for $7 \leq i \leq 11$,

$$P_{i,j} = P(X_{n+1} = j | X_n = i) = 1/6, \ \ j = i+1, \cdots, 12, \ \ 0, \cdots, i-7.$$

$P_{i,j} = 0$ otherwise. Now, for each state j there are only 6 states from which j is accessible, and the corresponding transition probability is $1/6$ for each of those states. This implies that the TPM is doubly stochastic and so $\pi_0 = 1/13$.

7. (Extra 1) The TPM for this MC is

$$P = \begin{bmatrix} 0.8 & 0.15 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

To find the stationary distribution we must solve $v = vP$, $\sum v_i = 1$. This gives the system of equations,

$$\begin{aligned} v_1 &= 0.8v_1 + 0.05v_2 + 0.2v_3 \\ v_3 &= 0.05v_1 + 0.05v_2 + 0.5v_3 \\ 1 &= v_1 + v_2 + v_3 \end{aligned}$$

Simplification yields

$$\begin{aligned} 4v_1 &= v_2 + 4v_3 \\ 10v_3 &= v_1 + v_2 \end{aligned}$$

Solving for v_2 in the first equation and substituting that into the second gives

$$\begin{aligned} v_1 &= \frac{14}{5}v_3 \\ v_2 &= \frac{36}{5}v_3. \end{aligned}$$

Hence,

$$1 = \frac{14}{5}v_3 + \frac{36}{5}v_3 + v_3$$

which implies that

$$\pi_3 = \frac{5}{55} = 0.0909, \;\; \pi_2 = \frac{36}{55} = 0.6545, \;\; \pi_1 = \frac{14}{55} = 0.2545.$$

So over a long period of time about 25.5% of the addresses are in RAM, about 65.4% are in cache, and about 9.1% are in swap.

8. (Extra 2) First note that if $X_k = 0$, then $X_{k+1} = 0$ if $N_k = 0$ or $N_k \geq m$. Also,

$$P(N_k \geq m) = \sum_{r=m}^{\infty} pq^r = q^m.$$

So,

$$P(X_{k+1} = 0|X_k = 0) = p + q^m.$$

For $1 \leq i \leq m-1$,

$$\begin{aligned} P(X_{k+1} = 0|X_k = i) &= P(X_k + N_k \geq m|X_k = i) \\ &= P(N_k \geq m - i) \\ &= q^{m-i}. \end{aligned}$$

For $1 \leq j \leq m-1$,

$$P(X_{k+1} = j|X_k = 0) = P(N_k = j) = pq^j.$$

For $1 \leq i \leq j \leq m-1$,

$$P(X_{k+1} = j|X_k = i) = P(N_k = j - i) = pq^{j-i}.$$

Therefore, the TPM is

$$P = \begin{bmatrix} p + q^m & pq & pq^2 & pq^3 & \cdots & pq^{m-1} \\ q^{m-1} & p & pq & pq^2 & \cdots & pq^{m-2} \\ q^{m-2} & 0 & p & pq & \cdots & pq^{m-3} \\ q^{m-2} & 0 & 0 & p & \cdots & pq^{m-4} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ q & 0 & 0 & 0 & \cdots & p \end{bmatrix}$$

To find the stationary distribution, we must solve $\underline{\pi} = \underline{\pi} P$. Ignore the first equation and start with π_1.

$$\pi_1 = pq\pi_0 + p\pi_1 \Rightarrow q\pi_1 = pq\pi_0 \Rightarrow \pi_1 = p\pi_0.$$

$$\pi_2 = pq^2\pi_0 + pq\pi_1 + p\pi_2 \Rightarrow q\pi_2 = pq^2\pi_0 + p^2q\pi_0 \Rightarrow \pi_2 = p\pi_0.$$

We can use induction to solve the remaining equations. Suppose $\pi_i = p\pi_0$, $1 \leq i \leq k$. The equation for π_{k+1} is

$$\begin{aligned}
\pi_{k+1} &= \sum_{j=0}^{k+1} pq^{k+1-j}\pi_j \\
&= p\pi_{k+1} + pq^{k+1}\pi_0 + \sum_{j=1}^{k} p^2 q^{k+1-j}\pi_0.
\end{aligned}$$

This gives

$$q\pi_{k+1} = pq^{k+1}\pi_0 + p^2 \sum_{j=1}^{k} q^{k+1-j}\pi_0,$$

and so,

$$\begin{aligned}
\pi_{k+1} &= pq^k\pi_0 + p^2\pi_0 \sum_{j=1}^{k} q^{k-j} \\
&= pq^k\pi_0 + p^2\pi_0 \frac{1-q^k}{1-q} \\
&= pq^k\pi_0 + p\pi_0(1-q^k) \\
&= p\pi_0.
\end{aligned}$$

Now use the equation $\sum \pi_i = 1$ to obtain

$$1 = \sum_{k=0}^{m-1} \pi_k = \pi_0 + (m-1)p\pi_0 = \pi_0(1 + (m-1)p),$$

and so,

$$\pi_0 = \frac{1}{1+(m-1)p}, \quad \pi_k = \frac{p}{1+(m-1)p}, \ 1 \leq k \leq m-1.$$

The mean time to return to state 0 is

$$m_0 = \frac{1}{\pi_0} = 1 + (m-1)p.$$

Review Problems