Mathematics 504
Simulation Modeling and Analysis
CSU Fullerton, Fullerton
Spring 2008

Nasser M. Abbasi

spring 2008 Compiled on October 14, 2025 at 5:13pm [public]

Contents

Chapter 1

Introduction

1.1 Course description

This course part of my Masters degree in Applied Mathematics at California State University, Fullerton

Course description (from CSUF catalogue)

```
MATH 504A    Simulation Modeling and: Prerequisites:
Math 501A,B; 502A,B; 503A,B. Corequisite: Math 504B.
Advanced techniques of simulation modeling, including
the design of Monte Carlo, discrete event, and continuous
simulations. Topics may include output data analysis,
comparing alternative system configurations,
variance-reduction techniques, and experimental design
and optimization.Units: (3)

MATH 504B Applications of Simulation Modeling
Techniques

Description: Prerequisites: Math 501A,B; 502A,B; 503A,B.
Corequisite: Math 504A. Introduction to a modern
simulation language, and its application to simulation
modeling. Topics will include development of computer
models to demonstrate the techniques of simulation modeling,
model verification, model validation, and methods
of error analysis.Units: (3)
```

1.2 Instructor

Professor Gearhart, W. B. CSUF Math department.

1.3 Class description handout/flyer

Hand out 1/22/2008.

Math 504: Simulation Modelling and Analysis

Text The course is based on notes written by the instructor. However, many of the course topics are covered in the text entitled Introduction to Probability Models, by S. Ross and published by Academic Press. This text is an excellent reference in applied probability.

Math. 438 8th edition.

Instructor	W. B. Gearhart
Office	MH 182F
Phone	714-278-3184
Email	wgearhart@fullerton.edu

Office Hours MW 4-5 pm, MWThF 2-3 pm. If you wish to see me any other time, just let me know, and we will arrange a meeting.

Course Description The course concerns the development and analysis of models of stochastic systems. There are three phases to the course. The first provides an introduction to the theory of stochastic processes. The second concerns modelling discrete event systems using simulation. The software Extend will be used to provide an introduction to the structure and use of a simulation environment. The third and last phase is devoted to further topics in stochastic modelling, and may include statistical aspects of simulation modelling, Brownian motion, signal processing, and Kalman filtering.

Exams every 5-6 weeks apart.

1

Grading There will be two exams, scheduled approximately every five to six weeks. Also, there will be a comprehensive final exam. Homework will be assigned and graded. The course grade will be based on the weighted average of the homework (5%), the average of the two exams (60%), and the final exam (35%). In cases when a student's calculated percentage is borderline, the instructor may raise the grade based on class participation and attendance, or any other evidence of a strong effort to do the course work.

Grade Scale A: 90-100 B: 80-89 C: 70-79 D: 60-69 F: 0-59

Attendance Class attendance is required. Please arrive on time. If you happen to miss a class, it is your responsibility to obtain from your classmates any missed lecture notes and assignments. However, see the instructor concerning class handouts.

Class Participation In addition to attending class, you are expected to actively participate in your own learning. In particular, you should come to class prepared, having studied the assigned readings and problems, and be ready to ask questions and participate in the class discussion.

Homework Homework due dates will be specified well in advance. Late papers will not be accepted.

Exam Make-up Policy No make-up exams will be given, unless you have a medical emergency or death in the family. These emergencies require valid documentation, and the instructor must be notified within 24 hours of the exam. The grade for a missed exam is zero.

Academic Dishonesty Academic dishonesty is obtaining or attempting to obtain credit for work by the use of any dishonest, deceptive, fraudulent, or unauthorized means. Academic dishonesty also includes helping someone commit an act of academic dishonesty. Examples of academic dishonesty include, but are not limited to:

1. Unacceptable examination behavior - communicating with fellow students, copying material from another student's exam or allowing another student to copy from an exam, possessing or using unauthorized materials, or any behavior that defeats the intent of an exam.

2. Plagiarism - taking the work of another and offering it as one's own without giving credit to that source, whether that material is paraphrased or copied in verbatim or near-verbatim form.

3. Unauthorized collaboration on a project, homework or other assignment where an instructor expressly forbids such collaboration.

4. Documentary falsification including forgery, altering of campus documents or records, tampering with grading procedures, fabricating lab assignments, or altering medical excuses.

Students who violate university standards of academic honesty are subject to disciplinary sanctions, including failure in the course, and suspension from the university. Since dishonesty in any form harms the individual, other students, and the university, policies on academic dishonesty are strictly enforced.

Emergency Information In the event of an emergency such as an earthquake or fire:

1. Take all your personal belongings and leave the classroom (or lab). Use the stairways located at the east, west, or center of the building.

2. Do not use the elevator. They may not be working once the alarm sounds.

3. Go to the lawn area towards Nutwood Avenue. Stay with class members for further instruction.

4. For additional information on exits, fire alarms and telephones, Building Evacuation Maps are located near each elevator.

5. Anyone who may have difficulty evacuating the building, please see the instructor.

Comments

1. Retain this course description and refer to it as needed during the semester.

2. All personal electronic devices, in particular cell phones, must be turned off during class.

3. Keep in mind that grades are not given, they are earned.

4. You are responsible for managing your outside responsibilities (work, family, and social) in order to allow sufficient time to meet the course requirements.

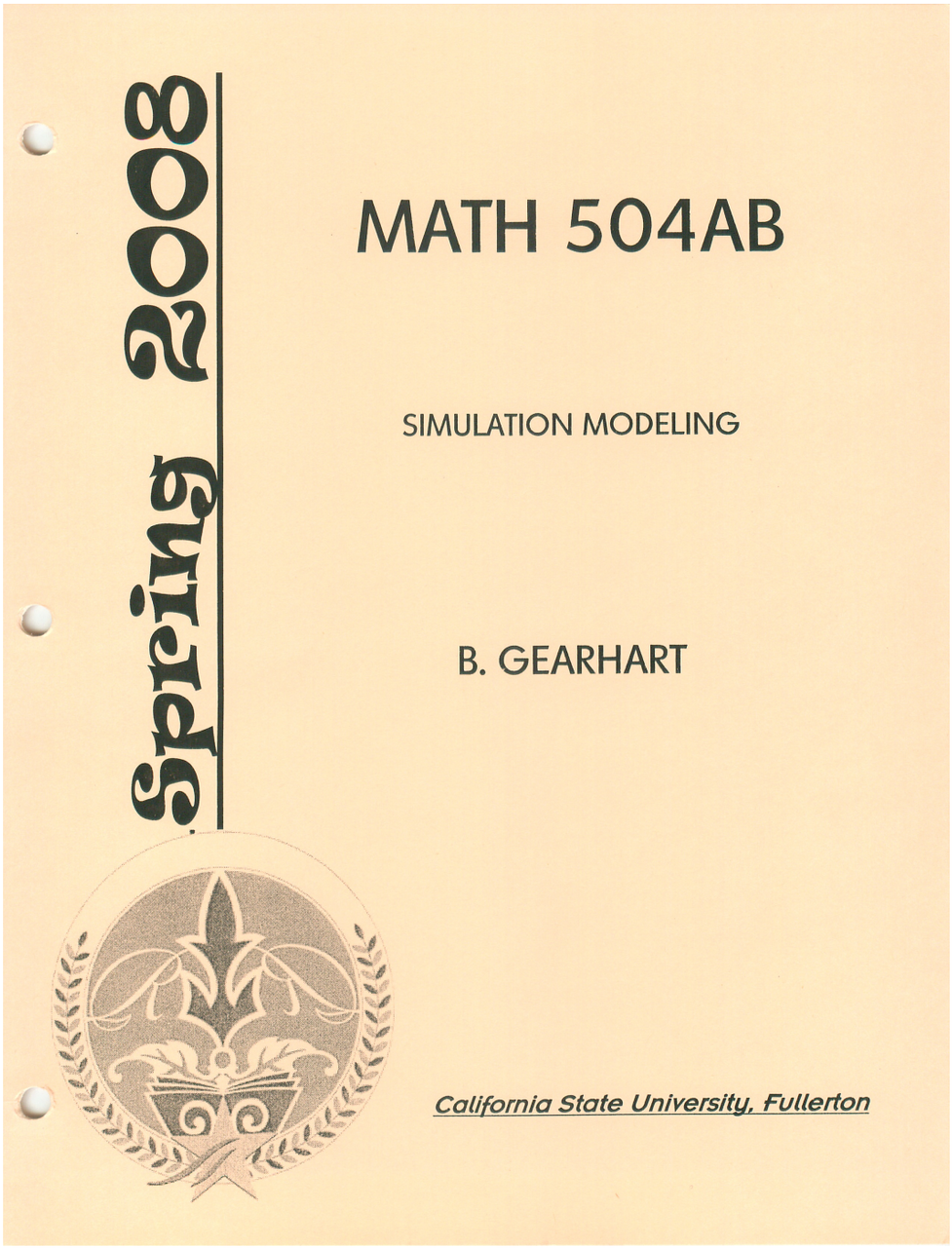
Chapter 2

Handouts given during the course

Local contents

2.1 instructor class notes

We followed mostly the instructor class notes



Introduction to Analysis and Simulation of Probability Models

William B. Gearhart
Department of Mathematics
California State University, Fullerton

Introduction to Analysis and Simulation of Probability Models

Table of Contents

Unit 1

Probability Theory

This unit covers basic topics from probability theory that are needed in the text. Much of the material is review and the presentation is concise. However, some of the topics in Chapter 3 may be new, and you will need to read this material more carefully. Chapter 3 concerns conditional expectation which is especially important in the study of stochastic processes.

Chapter 1

Basic Probability Theory

1.1 Introduction Probability models are based on the concept of a statistical experiment. An **experiment** is an action by which an observation is made. The individual observations which can result when the experiment is performed are called **outcomes.** For example, the experiment might be: Toss a coin and observe whether heads or tails appears . The individual outcomes are then heads or tails. A set of outcomes is called an **event.** The set of all possible outcomes is called the **sample space.** In these notes, we will typically denote the sample space by S.

Example 1.1.1 Suppose the experiment is to toss a coin three times and observe the sequence of heads and tails. Let H indicate that a head appears on a toss, and let T indicate that tails appears on a toss. Then we can describe the sample space as $S = \{HHH, HHT,\ HTH, HTT, THH, THT, TTH, TTT\}$. If A denotes the event that exactly one head appears in the sequence, then $A = \{HTT, THT, TTH\}$. □

1.2 The probability function For an event A of an experiment, we assign a number $P(A)$ called the probability of the event. This number is intended to reflect the **frequency interpretation** of probability; namely, that if the experiment is performed a large number of times, then $P(A)$ is the fraction of times that the event A occurs. To form this function, and stay consistent with the frequency interpretation, we require the following axioms :

a. $P(A) \geq 0$.

b. $P(S) = 1$.

c. For mutually exclusive (disjoint) events A and B, $P(A \cup B) = P(A) + P(B)$.

A probability model is developed by making assumptions about the experiment. Then, using the axioms, probabilities of events can be determined.

1.2 *Unit 1: Probability Theory*

Example 1.2.1 For the experiment of tossing a coin, we have the two outcomes H (heads appears) and T (tails appears). Thus $S = \{H, T\}$. Suppose we assume the coin is fair, which means that $P(H) = P(T)$. Then the axioms tell us that

$$1 = P(S) = P(H) + P(T) = 2P(H), \text{ and so } P(H) = P(T) = 1/2. \quad \square$$

More generally, using the same reasoning as in this example, it can be shown that if a sample space consists of a finite number of outcomes, each with the same chance of occurrence, the probability of an event A is simply

$$P(A) = \frac{n(A)}{n(S)},$$

where the symbol $n(\cdot)$ denotes the number of outcomes in the specified event.

Several useful formulas and relationships can be derived from the axioms. These include

a. If $A \subset B$, then $P(A) \leq P(B)$.

b. For any events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

c. $P(\overline{A}) = 1 - P(A)$, where $\overline{A}$ denotes the set of all outcomes not in A. This set is often called the compliment of A.

d. If A is an event consisting of a finite number of outcomes, then $P(A)$ is the sum of the probabilities of the outcomes in A.

Example 1.2.2 Consider again the experiment of Example 1.1. Recall that the sample space is $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$. If we assume the coin is fair, then we may assume that each of these 8 outcomes has the same chance of occurring, and so each has a probability of $1/8$. Let A be the event that at least one heads appears. Then A consists of 7 outcomes, and so by the formula in (d), $P(A) = 7/8$. Alternatively, we could observe that $\overline{A} = \{TTT\}$, and so, by the formula in (c), we conclude that $P(A) = 1 - P(\overline{A}) = 1 - 1/8 = 7/8$. $\square$

1.3 Conditional probability The concept of conditional probability is the cornerstone of much analysis in probability theory. For events A and B, the **conditional probability of A given B** is defined by

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad \text{provided} \quad P(B) \neq 0.$$

We think of $P(A \mid B)$ as the chance that the event A occurs, given that B has occurred, and can interpret it as the long-run fraction of times that A occurs among those times that B occurs. Often, conditional probabilities are found directly without using this formula. In fact, a common use of the formula is to find $P(A \cap B)$, as we have $P(A \cap B) = P(A \mid B)P(B)$.

Example 1.3.1 A bag contains 20 apples, of which 4 are rotten. We select one apple at random from the bag, and then select at random a second apple from the remaining 19 apples. Let A be the event that the second apple is rotten, and let B be the event that the first apple is rotten. Then we find directly that $P(A \mid B) = 3/19$. Indeed, given that B has occurred (the first apple is rotten), we are faced at the second selection with 19 apples of which 3 are rotten. To find the probability that both apples are rotten; that is, to find $P(A \cap B)$, note first that $P(B) = 4/20$, and so

$$P(A \cap B) = P(A \mid B)P(B) = (3/19)(4/20) = 3/95. \qquad \square$$

Events A and B are said to be **statistically independent** if $P(A \mid B) = P(A)$. In other words, events A and B are independent if the chance of A occurring is the same whether or not the event B has occurred. When events A and B are independent, it follows from the definition of conditional probability that $P(A \cap B) = P(A)P(B)$. Often, statistical independence can be determined from the nature of the experiment. For example, in the previous Example 3.1, the events A and B are evidently not independent since the probability of selecting a rotten apple on the second selection depends on the outcome of the first selection.

Example 1.3.2 Suppose 60% of registered voters will vote yes on a certain proposition. Then the probability is 0.6 that a voter selected at random will vote yes for the proposition. If two voters are selected at random and independently, the probability that both will vote yes is $(0.6)(0.6) = 0.36$. $\square$

1.4 *Unit 1: Probability Theory*

Exercises

1.1 The number of hurricanes reaching the east coast of the United States during a year is observed to have the following pattern.

number of hurricanes	0	1	2	3	4	5	6 or more
frequency	0.15	0.28	0.27	0.17	0.08	0.03	0.02

What is the probability that the number of hurricanes reaching the east coast of the United States is (a) at most 2, (b) at least 3, (c) either 0, 1, or 2 .

1.2 Under controlled conditions, a white rat will contract disease A with probability 0.6 , will contract disease B with probability 0.7 , and will contract both diseases with probability 0.5. Suppose a white rat is exposed to both disease A and disease B. (a) What is the probability of contracting at least one these diseases? (b) What is the probability of contracting neither?

1.3 A card is drawn at random from an ordinary deck of 52 cards. (a) What is the probability of selecting a queen or a heart? (b) What is the probability of selecting neither a queen nor a heart?

1.4 A fair die is rolled twice. (a) What is the probability that at least one die shows four dots? (b) What is the probability that at least one die shows four dots, or the total number of dots is ten?

1.5 A system consists of two components. The first component works with probability 0.9 and the second works with probability 0.8 . In order for the system to work, both components must work. Assume the components function independently of each other. (a) What is the probability the system works? (b) What is the probability the system fails?

1.6 A system consists of two components, in which the second serves only as a backup for the first. The first component works with probability 0.9 and the second works with probability 0.6 . Assume the components function independently of each other. (a) What is the probability the system works? (b) What is the probability the system fails?

1.7 A fair die is rolled twice. Let A be the event that the sum of the dots showing is 7, and let B be the event that the amount on the first die is even. Are these events independent?

1.8 Referring to Example 1.1.1, suppose the coin is not fair, and instead shows heads with probability 0.7 . Assume the three tosses are made independently. (a) What is the probability of the outcome THT? (b) What is the probability of the event exactly one head?

Chapter 2

Random Variables

2.1 Introduction A **random variable** is a function defined on a sample space. In these notes, we will consider only random variables that are real-valued.

Example 2.1.1 Suppose the experiment is toss a coin three times and observe the sequence of heads and tails. Let X be the random variable equal to the number of heads in a sequence. Then, for example, $X(\{HTH\}) = 2$ and $X(\{HTT\}) = 1$. □

Typically, uppercase letters such as X or Y are used to denote random variables, and lower case letters are used to denote the possible values of random variables.

2.2 Discrete random variables A random variable X is said to be **discrete** if the set of its possible values is countable. To describe the probabilistic structure a discrete random variable we use a **probability distribution function** which is defined for each possible value x of X by $P(X = x)$.

Example 2.2.1 For the experiment of Example 2.1.1, assume the coin is fair. Define X as the number of heads in an outcome. Then the possible values of X are $0, 1, 2$, and 3. The event $\{X = 2\}$, for example, consists of the outcomes HHT, HTH, and THH. As these outcomes are equally likely, we conclude that $P(X = 2) = 3/8$. Similarly one shows that $P(X = 0) = 1/8$, $P(X = 1) = 3/8$, and $P(X = 3) = 1/8$. □

Example 2.2.2 For a specified probability p, the Bernoulli random variable is defined as the random variable X which has value 1 with probability p, and value 0 with probability $q = 1 - p$. Thus, the probability distribution function is given by $P(X = 1) = p$, and $P(X = 0) = q$. For example, in the experiment of tossing a fair coin, if we define the random variable X by $X(H) = 1$ and $X(T) = 0$, then X is

Bernoulli with $p = 1/2$. The Bernoulli random variable is a basic building block for several other random variables, such as the binomial random variable and the geometric random variable. □

Example 2.2.3 An important example of a discrete random variable is the binomial random variable, which is based on the Bernoulli random variable mentioned in Example 2.2.2. For a specified probability p, consider a sequence of n independent Bernoulli trails, in each of which the value 1 occurs with probability p, and value 0 occurs with probability $q = 1 - p$. The binomial random variable X is defined as the number of 1's that appear in the sequence. The possible values of X are $0, 1, 2, \cdots, n$, and it can be shown that

$$P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad \text{for } x = 0, 1, 2, \cdots, n, \quad \text{where } \binom{n}{x} = \frac{n!}{x!(n-x)!} .$$

The term $\binom{n}{x}$ equals the number of subsets, consisting of x items, which can be formed from a set of n items. □

2.3 Continuous random variables A random variable X is said to be **continuous** if there is a function $f(x)$ defined on the entire real line, such that for any numbers $a < b$,

$$P(a < X \leq b) = \int_a^b f(x)dx .$$

For example, the experiment might be to select a person at random and measure their height. The random variable in this case assigns the observed height to the person selected. This random variable would be viewed as continuous. Another example, fundamental in the field of life insurance, is to select a person at random, and observe how long they live. Lifetimes are modeled as continuous random variables.

The function $f(x)$ that is used to describe the random variable is called the **density function** of X. A density function is required to satisfy the two conditions:

$$f(x) \geq 0 \text{ for all } x, \text{ and } \int_{-\infty}^{\infty} f(x)dx = 1 .$$

Example 2.3.1 Perhaps the most widely used density function in applications is the normal density. This density is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} , \quad \text{for } -\infty < x < \infty .$$

The parameters μ and $\sigma > 0$ need to be specified, and can be estimated from sample values of the random variable. □

Example 2.3.2 Another widely used density function is the gamma density. This density is given by

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} , \quad \text{for } x > 0 .$$

The density is equal to zero for $x < 0$. The parameters $\alpha > 0$ and $\beta > 0$ need to be specified, and can be estimated from sample values of the random variable. The term $\Gamma(\alpha)$ is known as the gamma function. When the argument α is a positive integer, we have $\Gamma(\alpha) = (\alpha - 1)!$. An important special case is $\alpha = 1$. The density is then called the **exponential density**, and is given by

$$f(x) = \frac{1}{\beta} e^{-x/\beta} , \quad \text{for } x > 0 .$$ □

To specify a density function, we usually indicate the form of the function only over the domain on which the density is positive.

2.4 Expected value and variance Distribution functions and densities give a complete probabilistic description of a random variable. However, it is helpful to summarize the quantitative nature of a random variable in terms of just a few numbers. Foremost among such numbers are the expected value and the variance. The **expected value**, also called the **mean**, is defined in the discrete case by

$$E(X) = \sum_x x\, P(X = x) ,$$

where the sum is taken over all possible values of X. In the continuous case, it is defined by

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx .$$

2.4 *Unit 1: Probability Theory*

The expected value of a random variable can be interpreted as the long-run average of the values of the random variable. Indeed, consider the discrete case and suppose the experiment associated with X is performed a large number of times, say N. Of the N observations, the fraction $P(X = x)$ of them would be equal to x, and so we would count $NP(X = x)$ observations with the value x. Thus, when we sum all the observed values of X and divide by N to get the usual average, we would get approximately,

$$\frac{1}{N}\sum_x x\,(NP(X=x))\,, \quad \text{which equals } E(X)\,.$$

A similar argument can be made in the continuous case.

The expected value is a measure of location or central tendency. Roughly, we could view the values of a random variable X as forming a set of points along the real line. The mean $E(X)$ is then a measure of the location or center of this set of numbers. On the other hand, to measure how far the values of the random variable deviate away from the mean, we commonly use the variance. For convenience, denote the mean $E(X)$ by μ. Then the **variance** of X is defined in the discrete case by

$$Var(X) = \sum_x (x-\mu)^2\,P(X=x)\quad,$$

and in the continuous case by

$$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx\,.$$

Note that the variance is the expected value of the random variable $(X-\mu)^2$, which is the square of the deviation of X from the mean μ. It can be shown that $Var(X) = E(X^2) - \mu^2$. This formula is often more convenient for calculations.

The variance is usually denoted by σ^2, and the square root of the variance, or σ, is known as the **standard deviation** of X. For many random variables, the bounds $\mu \pm 2\sigma$ include most of the possible values of the random variable. For example, this estimate covers about 95% of the values of a normally distributed random variable. In fact, for any random variable, these bounds will contain at least 75% of the values.

Example 2.4.1 Consider the experiment of Example 2.2.1, in which a fair coin is tossed three times, and the sequence of heads and tails observed. As shown in that example, for the random variable X equal to the number of heads in a sequence, the probability distribution is $P(X=0) = 1/8$, $P(X=1) = 3/8$, $P(X=2) = 3/8$, and

$P(X = 3) = 1/8$. Thus, the expected value of X is

$$E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{3}{2} \cdot$$

Since the coin is fair, we would anticipate the expected value to be in the center of the possible values of X, which are the integers from 0 to 3. To find the variance, we have

$$E(X^2) = 0^2 \cdot \frac{1}{8} + 1^2 \cdot \frac{3}{8} + 2^2 \cdot \frac{3}{8} + 3^2 \cdot \frac{1}{8} = 3 \ ,$$

and so $Var(X) = 3 - (3/2)^2 = 3/4$. The standard deviation of X is then $\sqrt{3/4}$. Finally, using the rule for estimating bounds on the possible values of X, we find the interval to be $\frac{3}{2} \pm 2\sqrt{3/4}$, or 1.5 ± 1.7, which is a fair estimate of the range of this random variable. □

It can be shown for the normal distribution that the expected value is equal to the parameter μ, and the variance is equal to the parameter σ^2. For the gamma distribution, the expected value is $\mu = \alpha\beta$ and the variance is $\sigma^2 = \alpha\beta^2$. For the binomial random variable with parameters n and p (recall Example 2.2.3), the expected value is $\mu = np$, and the variance is $\sigma^2 = npq$, where $q = 1 - p$.

2.5 Functions of random variables Often in applications, random variables are formed as functions of other random variables. Given a random variable X and a real-valued function h, a new random variable Y can be formed as $Y = h(X)$. From the distribution of X, the distribution of Y can be determined, and then the parameters of Y, such as the mean and variance, can be found. However, there is a more direct and usually simpler way to find the parameters of Y. Indeed, it can be shown that

$$E(Y) = \sum_x h(x)\, P(X = x) \ ,$$

when X is a discrete random variable, and

$$E(Y) = \int_{-\infty}^{\infty} h(x) f(x)\, dx \ ,$$

when X is a continuous random variable and $f(x)$ is its density function.

2.6 *Unit 1: Probability and Statistics*

Example 2.5.1 Suppose X is a continuous random variable with exponential density

$$f(x) = \frac{1}{2}e^{-x/2} \text{ for } x > 0 \ .$$

Form a new random variable Y as $Y = X^2$. Then

$$E(Y) = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_0^{\infty} x^2 \frac{1}{2} e^{-x/2} dx = 8 \ .$$

The last integral was evaluated using integration by parts twice. □

Exercises

2.1 For the experiment in Example 1.1, assume the probability of heads is 0.6. (a) What is the probability distribution of X? (b) What are the mean and variance of X.

2.2 A fair die is rolled. Let X denote the random variable equal to the number of dots showing. Find the expected value and the standard deviation of X.

2.3 Find the mean μ and the standard deviation σ of a random variable X that follows the exponential distribution with parameter β. What is the probability that the random variable lies in the interval $\mu \pm 2\sigma$?

2.4 A distribution of fundamental importance is the uniform distribution on the interval (a, b). The density function $f(x)$ is equal to $1/(b-a)$ for $a < x < b$, and is equal to 0 otherwise. Suppose X is a uniform random variable on (a, b).

(a) Show that for any interval (c, d) which is contained in (a, b), the probability $P(c < X \leq d)$ is equal to $(d-c)/(b-a)$.

(b) What are the mean μ and variance σ^2 of X. Compare the interval $\mu \pm 2\sigma$ with the actual range of X.

2.5 Suppose a random variable X follows the gamma distribution with parameters $\alpha = 2$ and $\beta = 4$. What are the expected value and standard deviation of X.

2.6 A sum of k independent and identically distributed exponential random variables follows the gamma distribution with mean $k\mu$ and $\beta = \mu$. Write a MATLAB program which will generate 1000 independent observations of a gamma random variable with mean 6 and $\beta = 2$. Tabulate a relative frequency histogram over the range $[0, 20]$ with intervals of length two. A relative frequency histogram is a table which shows the fraction of observations that fall in each of the intervals. Print a table listing each interval midpoint and the relative frequency for that interval.

2.7 For the previous problem, plot the histogram as a bar chart, and compare the shape of this graph with the graph of the density function. Why would you expect these two shapes to be similar? To address this question, develop the following line of reasoning. Suppose X is a continuous random variable with density $f(x)$. Assume a large number of random and independent observations of X are obtained, and the results summarized in a relative frequency histogram with intervals (x_{i-1}, x_i) for $i = 1, 2, \cdots, n$.

(a) Find an approximate expression for the fraction of observations that will fall in the interval (x_{i-1}, x_i). Write this expression as an integral of the density function $f(x)$. Next assume that x_{i-1} and x_i are very close, and approximate this expression in terms of $x_i - x_{i-1}$, and the value of the density function $f(x)$ at the midpoint of the interval.

(b) Use the result of part (b) to argue that the relative frequency histogram and the density function $f(x)$ of X should be approximately scaled versions of each other, and estimate the scale factor.

Expand the program written for the previous problem by including a table of the values of the gamma density at each interval midpoint. Compare this table with the relative frequency chart to estimate the scale factor, and then compare this estimate with the value predicted by your analysis above.

Chapter 3

Conditional Expectation

Review

3.1 Introduction Conditional expectation is fundamental in probability theory and is especially useful in the analysis of stochastic processes. Recall that for events A and B, the conditional probability of A, given that B has occurred, is defined by $P(A|B) = P(A \cap B)/P(B)$, whenever $P(B) \neq 0$. For a given event B, we can think of the sample space as being restricted to those outcomes in the event B, with associated probability function $P(\cdot \,|\, B)$. The function $P(\cdot \,|\, B)$ has all the properties of a probability function.

Example 3.1.1 Let X and Y be independent binomial random variables, each with parameters n and p. Then, with $q = 1 - p$,

$$P(X = k \,|\, X + Y = m) = \frac{P(X = k, Y = m - k)}{P(X + Y = m)}$$

$$= \frac{\binom{n}{k} p^k q^{n-k} \binom{n}{m-k} p^{m-k} q^{n-(m-k)}}{\binom{2n}{m} p^m q^{2n-m}} = \frac{\binom{n}{k}\binom{n}{m-k}}{\binom{2n}{m}},$$

for $k = 0, 1, \cdots, m$. Note that the hypergeometric distribution is obtained. Since X and Y are independent and identically distributed, this result might be expected. Indeed, we could view the event $X = k$, given $X + Y = m$, as selecting a subset of size m from the $2n$ trials for X and Y, and asking for the probability that k of them are among those belonging to the trials for the X variable. □

3.2 Conditional Expectation For jointly continuous random variables X and Y, the conditional expectation of Y given $X = x$ is defined by

$$E(Y \,|\, X = x) = \int_{-\infty}^{\infty} y f(y \,|\, x)\, dy$$

3.2 *Unit 1: Probability Theory*

where $f(y \mid x)$ is the conditional distribution of Y given $X = x$. In the discrete case, a similar formula holds with the integral replaced by a sums.

Example 3.2.1 Let X and Y be jointly continuous random variables with density

$$f(x,y) = 4x(y-x)e^{-(x+y)} \quad , \text{ for } \quad 0 < x < y, \ 0 < y < \infty \ .$$

Then the marginal distribution of X is

$$f(x) = \int_x^\infty f(x,y)\,dy = \int_x^\infty 4x(y-x)e^{-(x+y)}\,dy = 4xe^{-2x} \ , \quad \text{for } 0 < x < \infty,$$

where the substitution $u = y - x$ and integration by parts were used to find the integral. Thus,

$$f(y \mid x) = \frac{f(x,y)}{f(x)} = (y-x)e^{-(y-x)} \quad , \text{for} \quad y > x \ .$$

It follows that

$$E(Y \mid X = x) = \int_x^\infty y(y-x)e^{-(y-x)}dy = x + 2 \ , \quad \text{for} \quad x > 0 \ . \qquad \square$$

A formula of especial importance is $E(Y) = E(E(Y \mid X))$. For jointly continuous random variables, a brief demonstration of this result is obtained by first setting

$$\phi(x) = E(Y \mid X = x) = \int_{-\infty}^\infty y\frac{f(x,y)}{f(x)}dy \ ,$$

and then noting that

$$E(E(Y \mid X)) = E(\phi(X)) = \int_{-\infty}^\infty \left(\int_{-\infty}^\infty y\frac{f(x,y)}{f(x)}dy\right) f(x)\,dx = E(Y) \ .$$

Example 3.2.2 Consider a Bernoulli experiment with probability p of success on each trial. Let X denote the number of trials until the first success occurs. To find $E(X)$, condition on the outcome of the first toss. Suppose the random variable Y is defined as 1 if the first toss is a success, and defined as 0 if the first toss is a failure. Then we have $E(X|Y = 1) = 1$, and $E(X|Y = 0) = 1 + E(X)$. Thus,

$$E(X) = pE(X|Y = 1) + (1-p))E(X|Y = 0) = p + (1-p)(1 + E(X)) \ .$$

Solving for $E(X)$ gives us $E(X) = 1/p$. In the same way, one can find the variance of X, and this calculation is left as an exercise. □

Example 3.2.3 Consider a Bernoulli experiment with probability p of success on each trial. Let X_k denote the number of trials until k consecutive successes occur. To find $E(X_k)$, we first determine the conditional expectation $\phi(n) = E(X_k | X_{k-1} = n)$. By conditioning on the result of the next toss, we obtain

$$\phi(n) = E(X_k | X_{k-1} = n) = p(n+1) + (1-p)[(n+1) + E(X_k)] .$$

Thus, $\phi(n) = n + 1 + (1-p)E(X_k)$, and since $E(X_k) = E(\phi(X_{k-1}))$, it follows that $E(X_k) = E(X_{k-1}) + 1 + (1-p)E(X_k)$. Thus, solving for $E(X_k)$ we get

$$E(X_k) = \frac{1}{p} + \frac{1}{p} E(X_{k-1}) , \quad \text{for} \quad k \geq 1 .$$

These equations can be solved recursively, starting with $E(X_1) = 1/p$, to get

$$E(X_k) = \frac{1}{p} + \frac{1}{p^2} + \cdots + \frac{1}{p^k} \quad \text{for} \quad k \geq 1 .$$

Using the formula for the sum of a geometric series, this result can be written a little more simply as

$$E(X_k) = \frac{1}{q}\left(p^{-k} - 1\right) , \quad \text{for} \quad k \geq 1 . \qquad \square$$

The formula $E(Y) = E(E(Y \mid X))$ provides not only a convenient way to calculate expected values, but yields also the law of total probability. Indeed, if A is an event, define the random variable Y by $Y = 1$ if A occurs, and $Y = 0$ otherwise. Then $P(A) = E(Y)$ and also $P(A \mid X = x) = E(Y \mid X = x)$. Hence, again assuming X is a continuous random variable, we have

$$P(A) = E(Y) = \int_{-\infty}^{\infty} E(Y \mid X = x) f(x) dx = \int_{-\infty}^{\infty} P(A \mid X = x) f(x) dx .$$

Thus we have a formula for $P(A)$ which is obtained by conditioning on the random variable X. In the discrete case, the integral is replaced by a sum.

Example 3.2.4 Suppose the number N of accidents per year at a certain industrial site follows the Poisson distribution with parameter λ. Denote by p the probability that an accident will result in a claim over \$10,000. Let Y denote the number of accidents per year which result in a claim over \$10,000. Then, setting $q = 1 - p$,

$$P(Y = k \mid N = n) = \binom{n}{k} p^k q^{n-k} \quad , \quad \text{for} \quad n \geq k \ ,$$

and $P(Y = k \mid N = n) = 0$ if $n < k$. To find $P(Y = k)$, we have

$$P(Y = k) = \sum_{n=k}^{\infty} \binom{n}{k} p^k q^{n-k} e^{-\lambda} \frac{\lambda^n}{n!} \qquad \equiv \sum P(Y|N)\, P(N)$$

$$= \frac{e^{-\lambda}(\lambda p)^k}{k!} \sum_{n=k}^{\infty} \frac{(q\lambda)^{n-k}}{(n-k)!} = \frac{e^{-\lambda}(\lambda p)^k}{k!} e^{q\lambda} = e^{-\lambda p} \frac{(\lambda p)^k}{k!} \ .$$

Thus, the unconditional distribution of Y is Poisson with parameter λp. □

Exercises

3.1 Find $E(X \mid Y = y)$, for random variables X and Y with joint density

$$f(x, y) = \frac{e^{-x/y} e^{-y}}{y} \quad , \quad 0 < x < \infty, \ 0 < y < \infty$$

3.2 Let X be a random variable with range $0, 1, 2, \cdots$. Show that

$$E(X) = \sum_{n=1}^{\infty} P(X \geq n) = \sum_{n=0}^{\infty} P(X > n) \ .$$

3.3 Let X_i, $i = 1, 2, \cdots$ be independent uniform $(0, 1)$ random variables. Define N as the smallest integer n such that $X_n < X_{n-1}$, where $X_0 = x$. Let $f(x) = E(N)$. (a) Derive an integral equation for $f(x)$ by conditioning on X_1. Solve this equation by differentiating both sides with respect to x. (b) As a second approach, determine $P(N > n)$ directly, and then use this result to find $E(N)$. Help: Note that $P(N > n) = P(x < X_1 < X_2 < \cdots < X_n)$. Find this probability by integration. By the way, you could find this probability by conditioning on X_1, but this way leads to steps similar to those of part (a).

3.4 Here is another approach to the problem in Example 3.2.3. Let N denote the number of trials until the first failure occurs. Then N follows that geometric distribution and $P(N = n) = p^{n-1}q$, for $n = 1, 2, \cdots$. Argue that $E(X_k \mid N = n) = n + E(X_k)$, if $1 \leq n \leq k$, and that $E(X_k \mid N = n) = k$, if $n > k$. Use this result to find $E(X_k)$.

Remark: The random variable N gives a handle on the probability distribution of X_k. Indeed, for $j > k$, $P(X_k = j \mid N = n) = 0$, if $n > j - k$, and if $1 \leq n \leq j - k$, then $P(X_k = j \mid N = n) = P(X_k = j - n)$. Now, conditioning on N yields a formula for $P(X_k = j)$ in terms of $P(X_k = j - 1), \cdots, P(X_k = k)$. This formula can then be used to find the values of $P(X_k = j)$, starting with $P(X_k = j) = 0$ for $j < k$, and $P(X_k = k) = p^k$.

3.5 Let X_i, $i = 1, 2, \cdots$, be independent uniform $(0, 1)$ random variables. Fix a : $0 < a \leq 1$, and let N denote the smallest value of n such that $X_1 + X_2 + \cdots + X_n \geq a$. Show that $P(N > n) = a^n/n!$, $n = 1, 2, \cdots$, and hence find $E(N)$. Help : Use induction on n, and then condition on X_n in the formula for $P(N > n)$.

3.6 Consider a sequence of independent trials, on each of which any one of m outcomes is equally likely to occur. Let N denote the number of trials until the same outcome occurs k consecutive times. Show that $E(N) = 1 + m + m^2 + \cdots + m^{k-1}$. Help: Let N_i denote the number of trials until the same outcome occurs i consecutive times. Find $E(N_i \mid N_{i-1})$, and then use $E(N_i) = E(E(N_i \mid N_{i-1}))$ to find a recursive formula for the $E(N_i)$.

3.7 Let X_i, $i = 1, 2, \cdots$ be independent uniform $(0, 1)$ random variables. Given $\lambda > 0$, let N denote the smallest value n such that

$$\prod_{i=1}^{n} X_i > e^{-\lambda} \geq \prod_{i=1}^{n+1} X_i, \quad \text{where we define } \prod_{i=1}^{0} X_i = 1.$$

Show that N follows the Poisson distribution with parameter λ. Help: Use induction on n, and find $P(N = n)$ by conditioning on X_1, keeping in mind that $P(N = n \mid X_1 = x) = 0$ if $x < e^{-\lambda}$.

3.8 An urn contains w white balls and b black balls. At each trial, a ball is drawn at random. If the ball is white, it is returned, while if it is black, it is replaced by a white ball. Find the expected number of white balls in the urn after n trials. Help: Let N_n denote the number of white balls in the urn after n trials. Find a recurrence formula for $E(N_n)$ by conditioning on N_{n-1}.

3.9 Let $X_1, X_2, \cdots$ be a sequence of independent, identically distributed, continuous random variables. A record is said to occur at time $n \geq 2$ if $X_n > max\{X_1, X_2, \cdots, X_{n-1}\}$. The value of the first variable X_1 is considered a record at time $n = 1$. Define the random variable I_i to be 1 if a record occurs at time i, and to be zero otherwise. (a) Show that $P(I_i = 1) = 1/i$, for $i \geq 1$. Help: For $i \geq 2$, $P(I_i = 1)$ is equal to $P(X_1 < X_i, X_2 < X_i, \cdots, X_{i-1} < X_i)$. Condition on X_i to solve this equation. (b) Let N_n equal the number of records that occur up to time n. Find the expected value and variance of N_n. Help: Note that $N_n = I_1 + \cdots + I_n$, and assume without proof that the variables I_i are independent. (c) Let T be the first time greater than 1 at which a record occurs. Find the probability distribution of T, and show that

3.6 *Unit 1: Probability Theory*

$P(T < \infty) = 1$, while $E(T) = \infty$. Help: Note that $P(T > n) = P(X_2 < X_1, X_3 < X_1, \cdots X_n < X_1)$, and condition on X_1.

3.10 Recall the multinomial distribution

$$P(X_1 = x_1, \cdots, X_n = x_n) = \frac{n!}{x_1! x_2! \cdots x_n!} p_1^{x_1} p_2^{x_2} \cdots p_n^{x_n} ,$$

for non negative integers $x_1, x_2, \cdots, x_n$ which sum to n. Find $cov(X_i, X_j)$. Help: Write the covariance in terms of expectations, and condition on one of the variables, say X_j. Then recall that X_i by itself is a binomial random variable.

Unit 2

Markov Chains

This unit provides an introduction to the fundamental ideas of Markov chains. The chapters cover standard topics found in most texts. Of particular interest for this course, however, is the material in Chapter 8 which concerns the recent applications of Markov chains to problems in statistical mechanics and combinatorial optimization.

Chapter 4

Markov Chains

4.1 Introduction A **stochastic process** is a family of random variables $\{X(t) \mid t \in T\}$. In applications, the elements of the index set T often relate to time. The set of possible values the random variables $X(t)$ is called the **state space**. In these notes, the index set T and the state space will be a subsets of the real line. A stochastic process is said to have the **Markov property** if, whenever $t_0 < t_1 < \cdots < t_n$ are in T, we have

$$P(X(t_n) \leq x_n \mid X(t_{n-1}) = x_{n-1}, \cdots, X(t_0) = x_0)$$

$$= P(X(t_n) \leq x_n \mid X(t_{n-1}) = x_{n-1}) .$$

In other words, probabilities concerning the chain at time t, conditioned on a finite set of past values, depend only on the latest conditioned state.

Definition 4.1.1 A stochastic process $\{X_n \mid n = 0, 1, \cdots\}$ is called a **Markov chain** if (a) the process has the Markov property, (b) the state space is countable, and (c) the conditional probabilities $P(X_{n+1} = j \mid X_n = i)$ do not depend on n. □

In some settings, property (b) is not required, in which case the chain is said to be **non-denumerable**. Also, property (c) is sometimes not required, in which case the chain is said to be **non-homogeneous**.

4.2 Transition probabilities Let I denote the state space of a Markov chain. For states $i, j \in I$, set $p_{ij} = P(X_{n+1} = j \mid X_n = i)$. These probabilities are called **transition probabilities**. Observe that

$$\sum_{j \in I} p_{ij} = 1 , \quad \text{for any state } i \in I .$$

4.2 *Unit 2: Markov Chains*

For each time $n = 0, 1, \cdots$ and state i, we say the chain is in state i at time n if $X_n = i$. Set $\pi_i^{(n)} = P(X_n = i)$. The probabilities $\{\pi_i^{(0)} \mid i \in I\}$ describe the initial distribution of the chain over the states. For any state $j \in I$,

$$\pi_j^{(n+1)} = P(X_{n+1} = j) = \sum_{i \in I} P(X_{n+1} = j,\, X_n = i)$$

$$= \sum_{i \in I} \overbrace{P(X_{n+1} = j \mid X_n = i)}^{p_{ij}} P(X_n = i) = \sum_{i \in I} \pi_i^{(n)} p_{ij} \,.$$

These equations can be expressed in matrix form. Let P denote the (possibly infinite) matrix whose (i,j)-th entry is p_{ij}, and let $\pi^{(n)}$ denote the row vector whose i-th component is $\pi_i^{(n)}$. Then the previous equations can be written

$$\pi^{(n+1)} = \pi^{(n)} P \ , \text{ for } n = 0, 1, 2, \cdots . \tag{4.2.1}$$

The matrix P is called the **probability transition matrix**. Using this formula, and starting with the initial distribution over the states, the state probabilities at any future time are determined:

$$\pi^{(n)} = \pi^{(0)} P^n \,, \quad \text{for } n = 0, 1, 2, \cdots .$$

Thus, the evolution of a Markov chain is completely determined by the initial distribution and the transition probability matrix.

Example 4.2.1 Consider a Markov chain with state space $I = \{1, 2\}$. Suppose the transition probabilities are $p_{11} = 0.2$, $p_{12} = 0.8$, $p_{21} = 0.6$, and $p_{22} = 0.4$. Then the probability transition matrix is

$$P = \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{bmatrix}$$

□

The ***n*-step transition probabilities** are defined by

$$p_{ij}^{(n)} = P(X_{\nu+n} = j \mid X_\nu = i) \,, \quad \text{for each state } i, j \in I \text{ and time } n \geq 0 \,.$$

The 1-step transition probabilities are simply the transition probabilities introduced above. In general, the multistep transition probabilities are related by the **Chapman-Kolmogorov equations**: For any times n and m,

$$p_{ij}^{(n+m)} = \sum_{\nu \in I} p_{i\nu}^{(n)} \, p_{\nu j}^{(m)} \,.$$

Indeed,

$$\begin{aligned} p_{ij}^{(n+m)} &= P(X_{n+m} = j \mid X_0 = i) = \frac{P(X_{n+m} = j, X_0 = i)}{P(X_0 = i)} \\ &= \frac{1}{P(X_0 = i)} \sum_{\nu \in I} P(X_{n+m} = j, X_0 = i, X_n = \nu) \\ &= \frac{1}{P(X_0 = i)} \sum_{\nu \in I} P(X_{n+m} = j \mid X_0 = i, X_n = \nu)\, P(X_0 = i, X_n = \nu) \\ &= \sum_{\nu \in I} P(X_{n+m} = j \mid X_n = \nu) P(X_n = \nu \mid X_0 = i) = \sum_{\nu \in I} p_{i\nu}^{(n)}\, p_{\nu j}^{(m)} . \end{aligned}$$

For each time n, let $P^{(n)}$ denote the (possibly infinite) matrix whose (i,j)-th entry is $p_{ij}^{(n)}$. Then the Chapman-Kolmogorov equation can be written

$$P^{(n+m)} = P^{(n)} P^{(m)} , \quad \text{for } n, m \geq 0 .$$

In particular, for any time $n \geq 0$, we have $P^{(n)} = P^n$.

4.3 Examples of Markov chains This section presents a few well known examples of Markov chains.

Example 4.3.1 – A random walk model. Let $\{S_i \mid i = 1, 2, \cdots\}$ be independent and identically distributed random variables with possible values $\{0, \pm 1, \pm 2, \cdots\}$. Set

$$X_n = \sum_{i=1}^{n} S_i , \quad \text{for } n = 1, 2, \cdots , \text{ with } X_0 = 0 .$$

Then $\{X_n \mid n = 0, 1, 2, \cdots\}$ is a Markov chain. One could think of this chain as modeling the motion of a particle moving along the x-axis. The motion starts at the origin, and takes a step of amount S_{n+1} at time n. Thus, X_n is the position of the particle at time n.□

do Markov chain. HW

Example 4.3.2 – A simple inventory model. A store sells a certain item for which the weekly demands are assumed to be independent and identically distributed random variables with probability distribution $\{d_j \mid j \geq 0\}$. The inventory is managed using an (s, S) policy: If at the beginning of the week the amount in inventory is s or higher, then no order is placed; otherwise, an order is placed to bring the amount in inventory up to S. It is assumed that orders are placed only at the beginning of each week, and there is no lag-time, meaning that orders are filled immediately. Also, it is assumed that demand which cannot be met immediately is lost. Let X_n denote the amount in inventory at the end of the n-th week, and let D_n denote the demand during the n-th week. Then for

$n = 0, 1, 2, \cdots$, we have $X_0 = S$, while $X_{n+1} = max\{0, X_n - D_{n+1}\}$ if $X_n \geq s$, and $X_{n+1} = max\{0,\ S - \ D_{n+1}\}$ if $X_n < s$. The state space is $\{0, 1, \cdots, S\}$. The transition probabilities are given as follows. For $j = 0$,

$$p_{i0} = \sum_{\nu=S}^{\infty} d_\nu\,, \text{ if } i < s\,, \text{ and } p_{i0} = \sum_{\nu=i}^{\infty} d_\nu\,, \text{ if } i \geq s\ .$$

Next, for $i < s$ and $0 < j \leq S$, we have $p_{ij} = d_{S-j}$, while for $i \geq s$ and $0 < j \leq i$, we have $p_{ij} = d_{i-j}$. The remaining transition probabilities are equal to zero. □

Example 4.3.3 – The $M/G/1$ queue. Consider a single server queue in which an arriving customer is served immediately if the server is idle, and otherwise the customer joins a line to wait for service. When a customer has completed service, the first person in line is served next. Assume the service times of successive customers are independent and identically distributed random variables with common density $f(t)$. Assume further that customers arrive according to a Poisson process, which means the following. First, customers arrive at a fixed arrival rate $\lambda > 0$, so that during any time interval of length t, the number of arriving customers N_t follows a Poisson distribution with parameter λt. Thus

$$P(N_t = j) = e^{-\lambda t}\,\frac{(\lambda t)^j}{j!}\,, \quad \text{for } j = 0, 1, 2, \cdots\ .$$

Secondly, during disjoint intervals of time, the numbers of arriving customers are independent random variables. Let X_n *denote the number of customers in the system at the end of the n-th service, and let* Y_n *denote the number of customers that arrive during the service of the* $(n+1)$*-st customer.* Then, for any $n \geq 1$, if $X_n = 0$, then $X_{n+1} = Y_n$, while if $X_n > 0$, then $X_{n+1} = X_n - 1 + Y_n$. Conditioning on service time yields

$$w_j = P(Y_n = j) = \int_0^\infty e^{-\lambda t}\,\frac{(\lambda t)^j}{j!}\,f(t)dt\,, \quad \text{for } j = 0, 1, 2, \cdots\ .$$

The assumption of Poisson arrivals tells us that the Y_n are independent random variables, and therefore $\{X_n \mid n = 1, 2, \cdots\}$ is a Markov chain with transition probabilities:

$$p_{0j} = w_j\,,\ j \geq 0\,;\ p_{ij} = w_{j-i+1}\,,\ j \geq i - 1 \geq 0\,;\ \text{and } p_{ij} = 0, \text{ otherwise}\,. \quad \square$$

Exercises

4.1 Consider Example 4.3.1, and suppose the probability distribution of the step sizes is $\{b_k \mid k = 0, \pm 1, \cdots, \pm m\}$. Thus, b_k is the probability that a step of k units is taken, and the possible values of k are $0, \pm 1, \cdots, \pm m$. Determine the transition probabilities in terms of these probabilities.

4.2 Verify the expressions for the transition probabilities in Example 4.3.3.

4.3 In Example 4.3.3, let ρ denote the average number of arrivals during a service period. Thus,

$$\rho = \sum_{j=0}^{\infty} j w_j \ .$$

This average ρ is a measure of traffic intensity for the system. Let S be a random variable denoting service time of a customer, and having density $f(t)$. (a) Show that $\rho = \lambda E(S)$. (b) Suppose the service time S of a customer has the exponential density

$$f(t) = \mu e^{-\mu t} \quad \text{for} \quad t > 0 \ ,$$

where $\mu > 0$ is the service rate. Find ρ, and find also the transition probabilities. (c) If ρ is not too large, the state probability vectors, $\pi^{(n)}$, converge as $n \to \infty$ to a probability distributon π, which is the same for any choice of initial state probability distribution $\pi^{(0)}$. Assume this convergence holds, and that the exponential density of part (b) is in effect. Taking the limit in (4.2.1) as $n \to \infty$, it follows that the limiting distribution π satisfies $\pi = \pi P$. Show that when $\rho < 1$, a unique solution π of this equation exists and that $\pi_i = c(\lambda/\mu)^i$, $i = 0, 1, \cdots$, for some constant c. Find the constant c. Help : Starting with the first equation in the set of equations $\pi = \pi P$, show that *any* solution must have this form. Then find the constant c.

4.5 (a) Model the experiment in Example 3.2.3 of the last chapter as a Markov chain. (b) Do the same for the experiment in problem 3.6. (c) Do the same for problem 3.8.

4.6 Consider a Bernoulli experiment with probability p of success on a trial. We wish to study the number of independent trials needed until the pattern $SFFSFS$ appears. Model this problem as a Markov chain, and specify the probability transition matrix. Help: Let S denote the state that the last trial was a success, with the previous state, if any, being a success. Define SF to be the state that the last two results were a success followed by failure, and the third previous result, if any, was a success. Similarly, define the states SFF, $SFFS$, $SFFSF$, and $SFFSFS$.

4.7 For a Markov chain, define N_i to be the number of time steps that the process, having just entered state i, stays in state i before making a jump to another state. (a) Find the probability distribution of N_i. (b) Define $q_{ij} = P(X_{n+1} = j \mid X_n = i, X_{n+1} \neq i)$. These probabilities might be called jump probabilities, as they are transitions

probabilities under the condition that a jump to a different state has occurred. Determine expressions for the q_{ij} in terms of the one-step transition probabilities.

Chapter 5

Classification of States

5.1 Introduction To study the behavior of a Markov chain, it is important to understand the nature of the states in the chain. We say a state j is **accessible** from state i if $p_{ij}^{(n)} > 0$ for some n. Two states are then said to **communicate** if each state is accessible from the other. It can be shown that communication is an equivalence relation, and this relation partitions the state space into a collection of equivalence classes. In this chapter we study the properties of the states in these classes.

5.2 Recurrent and transient states Let T_{ij} denote the first time a chain enters state j, given that it starts in state i. The time T_{ij} is called a **first entrance time**. Set

$$f_{ij}^{(n)} = P(T_{ij} = n)\,, \quad n = 1, 2, \cdots, \quad \text{and} \quad f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)} = P(T_{ij} < \infty)\,.$$

State i is called **recurrent** if $f_{ii} = 1$, and is called **transient** $f_{ii} < 1$.

Theorem 5.2.1 Suppose a Markov chain starts in a transient state i. Let N denote the number of visits to this state. Then N follows a geometric distribution with $P(N = n) = f_{ii}^{n-1}(1 - f_{ii})$, for $n \geq 1$. Hence, the expected number of visits to state i is $1/(1 - f_{ii})$. On the other hand, if state i is recurrent, then the chain returns to this state infinitely often.

Proof Suppose state i is transient. Define a Bernoulli trial as follows. The random variable T_{ii} is sampled and if it is finite (with probability f_{ii}), we imagine returning to state i, while if it is infinite (with probability $1 - f_{ii}$), we imagine leaving state i forever. Define success on a trial to mean T_{ii} is infinite. Then the event $\{N = n\}$ occurs when we observe a success for the first time on trial n. The trials are independent because of the Markov property, and so it follows that N is a geometric random variable with probability of success $1 - f_{ii}$. Thus, the results for a transient state follow. If the starting

state i is recurrent, however, the chain will necessarily return to that state. When this happens, the Markov property implies that the process starts over probabilistically, and the chain will again return to state i. Thus, the process will return to this state infinitely often. □

Theorem 5.2.2 State i is

$$\text{transient if } \sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty, \quad \text{and} \quad \text{recurrent if } \sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty.$$

Proof Define the indicator variable I_n to be 1 if $X_n = i$, and to be zero otherwise. Then the sum $I_0 + I_1 + \cdots$ equals the number of time periods the chain is in state i. However,

$$E\Big(\sum_{n=0}^{\infty} I_n \mid X_0 = i\Big) = \sum_{n=0}^{\infty} E\Big(I_n \mid X_0 = i\Big) = \sum_{n=0}^{\infty} P\Big(X_n = i \mid X_0 = i\Big) = \sum_{n=0}^{\infty} p_{ii}^{(n)}.$$

But the expected number of time periods a chain spends in a state is finite for a transient state, and infinite for a recurrent state, the conclusion of the theorem follows. □

One application of this theorem is to show that states which communicate must be of the same type.

Theorem 5.2.3 If two states communicate, then they are either both transient or both recurrent.

Proof Suppose states i and j communicate. Then there are integers n and m such that $p_{ij}^{(n)} > 0$ and $p_{ji}^{(m)} > 0$. For any integer $k \geq 0$, we have

$$p_{ii}^{(n+k+m)} \geq p_{ij}^{(n)} p_{jj}^{(k)} p_{ji}^{(m)}, \quad \text{and} \quad p_{jj}^{(n+k+m)} \geq p_{ji}^{(n)} p_{ii}^{(k)} p_{ij}^{(m)}.$$

The first inequality holds because the term on the left is the probability of going from state i to state i in $n + m + k$ steps, while the term on the right is the probability of the same transition, but along a certain path going through state j. A similar argument holds for the second inequality. These inequalities show that the sums

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} \quad \text{and} \quad \sum_{n=1}^{\infty} p_{jj}^{(n)},$$

converge or diverge together. The conclusion of the theorem follows. □

If a state is transient, then we might expect that as time goes on, visits to this state would become less likely. Indeed, the following result holds.

Theorem 5.2.4 Assume state j is transient. For any state i, $p_{ij}^{(n)} \to 0$ as $n \to \infty$.

Proof Use a proof similar to that of Theorem 5.2.2 to show that the sum

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} \quad \text{is convergent .}$$

The conclusion of then follows. □

5.3 Structure of the state space A set of states S is said to be **closed** if $p_{ij} = 0$ whenever $i \in S$ and $j \overline{\in} S$. By the Chapman-Kolmogorov equations, it follows that for a closed set S, we have $p_{ij}^{(n)} = 0$, for any $n \geq 0$, whenever $i \in S$ and $j \overline{\in} S$. Indeed, denote by I the state space,

$$p_{ij}^{(2)} = \sum_{\nu \in I} p_{i\nu} p_{\nu j} = \sum_{\nu \in S} p_{i\nu} p_{\nu j} + \sum_{\nu \overline{\in} S} p_{i\nu} p_{\nu j} = \sum_{\nu \in S} p_{i\nu} \cdot 0 + \sum_{\nu \overline{\in} S} 0 \cdot p_{\nu j} = 0 \ .$$

Continuing by induction completes the proof. It follows now that if a set of states is closed, then upon entering that set, it is not possible to leave it.

Theorem 5.3.1 In a Markov chain, the recurrent states can be partitioned uniquely into a collection of closed sets such that all states in a set communicate with each other. In addition, the chain may also contain a set of transient states from which it is possible to reach states in the closed sets, but not vice-versa.

Proof Suppose i is a recurrent state. Let C_1 be the set of states that communicate with this state. Then C_1 is a closed set of recurrent states which communicate with each other. Next, suppose state j is recurrent but not in C_1. Let C_2 be the set of all states that communicate with this state. Then C_2 is a closed set of recurrent states that communicate with each other, and this set is disjoint from the set C_1. Continuing in this fashion yields the collection of closed sets consisting of recurrent states. Any remaining states will be transient states. □

A Markov chain is said to be **irreducible** if its state space contains no proper subset that is closed.

Theorem 5.3.2 In an irreducible Markov chain, each state can be reached from any other state. Also, all states are of the same type; that is, either all states are recurrent or all states are transient.

Proof The set of all states that can be reached from a given state is a closed set. Thus the first part of theorem holds. The second part follows from Theorem 5.2.3 . □

Markov chains with a finite number of states are of special interest. In view of Theorem 5.2.1, we would expect the following result.

Theorem 5.3.3 In a Markov chain with a finite number of states, not all states can be transient.

Proof Let $\{1, 2, \cdots, m\}$ denote the state space. Then

$$\sum_{j=1}^{m} p_{ij}^{(n)} = 1 \,, \quad \text{for any} \quad n \geq 1 \,.$$

But Theorem 5.2.4 tells us that if j is a transient state, then $p_{ij}^{(n)} \to 0$, as $n \to \infty$. Thus, not all states can be transient. □

It follows now that in an irreducible Markov chain with a finite number of states, all states must be recurrent.

5.4 First entrance times In general it is difficult to get a good picture of the probability distributions for the various first entrance times. However, there are some formulas which allow us to compute these probabilities. For example, we have

$$p_{ij}^{(n)} = f_{ij}^{(n)} + f_{ij}^{(n-1)} p_{jj}^{(1)} + f_{ij}^{(n-2)} p_{jj}^{(2)} + \cdots + f_{ij}^{(1)} p_{jj}^{(n-1)} \,. \tag{5.4.1}$$

Indeed, the event that a transition from state i to state j is made in n steps will occur in exactly of n ways : either state j is entered for the first time in n steps, and this event occurs with probability $f_{ij}^{(n)}$, or state j is entered for the first time in $n-1$ steps, and this event occurs with probability $f_{ij}^{(n-1)} p_{jj}^{(1)}$, and so on. Starting with $f_{ij}^{(1)} = p_{ij}$ for a given pair of states i and j, formula (5.4.1) can be used to calculate recursively the probabilities $f_{ij}^{(n)} = P(T_{ij} = n)$, for $n = 1, 2, \cdots$.

Consider next the calculation of the probabilities $f_{ij} = P(T_{ij} < \infty)$. Recall that f_{ij} is the probability that there is finite first entrance time into state j, given that the process starts in state i. Equivalently, we can interpret f_{ij} as the probability that the process

eventually reaches state j, given that it started in state i. In the following, let I denote the state space of the Markov chain, let R denote the set of recurrent states, and let T denote the set of transient states. Two cases can be found by direct reasoning. If both $i \in R$ and $j \in R$, then f_{ij} is either 1 or 0 depending on whether i and j communicate, while if $i \in R$ and $j \in T$, then $f_{ij} = 0$ by Theorem 5.3.1. For the case $i \in T$ and $j \in T$, it is generally of more interest to find the expected number of visits to state j, given that the process started in state i. We will consider this computation later in the next chapter.

For now therefore, we consider only the case $i \in T$ and $j \in R$. Conditioning on the next state gives us

$$f_{ij} = P(T_{ij} < \infty) = \sum_{k \in I} P(T_{ij} < \infty \,|\, X_1 = k, X_0 = i)\, P(X_1 = k \,|\, X_0 = i)\,.$$

However, from the Markov property, $P(T_{ij} < \infty \,|\, X_1 = k, X_0 = i) = P(T_{kj} < \infty)$. Therefore, recalling that $P(X_1 = k \,|\, X_0 = i) = p_{ik}$ yields

$$f_{ij} = \sum_{k \in I} P(T_{kj} < \infty) p_{ik} = \sum_{k \in I} p_{ik} f_{kj}\ .$$

Let $C(j)$ denote the set of states which communicate with state j, and denote by B the set of states which are not in T and which also are not in $C(j)$. Then

$$f_{ij} = \sum_{k \in T} p_{ik}\, f_{kj} + \sum_{k \in C(j)} p_{ik}\, f_{kj} + \sum_{k \in B} p_{ik}\, f_{kj}\ .$$

However, when $k \in B$, we have $f_{kj} = 0$. Indeed, since k is not a transient state, it must be a recurrent state, and therefore as state k does not communicate with state j, it is not possible to reach state j from state k. Next observe that when $k \in C(j)$, then k is a recurrent state which communicates with state j, and therefore it is certain that the first entrance time into state j from state k is finite. Thus, $f_{kj} = 1$. We have therefore

$$f_{ij} = \sum_{k \in T} p_{ik}\, f_{kj} + \sum_{k \in C(j)} p_{ik}\ .$$

These equations, for each fixed j, listed for each $i \in T$, yield a system of equations which can be solved, in principle at least, for the f_{ij}.

To consider a special case, suppose the state space I is finite. Let R denote the set of recurrent states. Assume there are m recurrent states, numbered $1, 2, \cdots, m$, and that there are n transient states, numbered $m+1, m+2, \cdots, m+n$. Let Q be the $n \times n$ matrix consisting of the p_{ik} for the transient states, and let F be the $n \times m$ matrix consisting of the f_{ij}, where i is a transient state, and j is a recurrent state. Finally, let Z be the $n \times m$ matrix whose (i,j)-th entry is the second sum on the right side of the equations above. Then these equations can be written in matrix form as $F = QF + Z$, or

equivalently $(I - Q)F = Z$. We will show later that the matrix $I - Q$ is invertible. Thus, we can write $F = (I - Q)^{-1}Z$.

Example 5.4.1 Consider the Markov chain with probability transition matrix

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & .2 & .8 & 0 & 0 \\ 0 & .7 & .3 & 0 & 0 \\ .1 & 0 & .1 & .4 & .4 \\ 0 & .1 & .3 & .2 & .4 \end{bmatrix} .$$

Then $R = \{1, 2, 3\}$, $T = \{4, 5\}$, and

$$Q = \begin{bmatrix} .4 & .4 \\ .2 & .4 \end{bmatrix} , \quad F = \begin{bmatrix} f_{41} & f_{42} & f_{43} \\ f_{51} & f_{52} & f_{53} \end{bmatrix} .$$

To form the matriz Z, note first that $C(1) = \{1\}$, $C(2) = \{2, 3\}$, and $C(3) = \{2, 3\}$. It follows that

$$Z = \begin{bmatrix} .1 & .1 & .1 \\ 0 & .4 & .4 \end{bmatrix} .$$

The matrix F can now be determined, and is given by $F = (I - Q)^{-1}Z$. Performing this calculation yields

$$\begin{bmatrix} 0.2143 & 0.7857 & 0.7857 \\ 0.0714 & 0.9286 & 0.9286 \end{bmatrix} .$$

Note that columns 2 and 3 are identical, as they must be. Indeed, since states 2 and 3 are recurrent and communicate, the first entrance time into state 2 is finite if and only if the first entrance time into state 3 is finite. Thus, the probabilities of these two events must be equal. Note also that in each row, the entry in the first column, and the entry in either the second or third column, sum to one. Indeed, starting in either transient state 4 or 5, the process must enter in finite time either the closed set $\{1\}$, or the closed set $\{2, 3\}$.

5.5 Stationary distributions Let $\pi_i^{(n)} = P(X_n = i)$. Thus, $\pi_i^{(n)}$ is the probability that a Markov chain is in state i at time n. Write the state space as $I = \{0, 1, 2, \cdots\}$. Then,

$$\pi_j^{(n+1)} = \sum_{i=0}^{\infty} \pi_i^{(n)} p_{ij} , \quad \text{for any } n \geq 0 .$$

What can we say about these state probabilities $\pi_j^{(n)}$ as $n \to \infty$? Suppose it is the case that for each state j, we have $\pi_j^{(n)} \to \pi_j$ as $n \to \infty$. Then, taking the limit (at least formally) in the last equation, we get

$$\pi_j = \sum_{i=0}^{\infty} \pi_i \, p_{ij} \, .$$

A vector $\pi = (\pi_0, \pi_1, \cdots)$ which satisfies this equation and which has nonnegative components summing to one, is called a **stationary distribution** for the chain. Note that if the initial state probability distribution $\pi^{(0)} = (\pi_0^{(0)}, \pi_1^{(0)}, \cdots)$ is a stationary distribution, then all subsequent state probability distributions will be the same as this initial distribution. Hence the term stationary is used.

It is of interest to know when a Markov chain has a stationary distribution. Toward this end, we introduce a few definitions. Recall that a Markov chain returns to a recurrent state infinitely often. A recurrent state is called **positive recurrent** if the expected value of the time to return to that state is finite, and is called **null recurrent** if this expected value is infinite. The **period** of state i is the greatest common divisor of the integers $n \geq 1$ such that $p_{ii}^{(n)} > 0$. If $p_{ii}^{(n)} = 0$ for all $n \geq 1$, the period is defined to be 0. A state will be called **periodic** if its period is at least 2, and it will be called **aperiodic** if its period is 1. States that are positive recurrent and aperiodic are called **ergodic**.

To derive conditions under which a Markov chain has a stationary distribution is lengthy and involved. However, we will highlight steps which lead to a central result in the theory. This result (Theorem 5.5.3 below) states that for an irreducible and aperiodic Markov chain, a stationary distribution exists if and only if the chain is ergodic. We begin by recording, without proofs, certain fundamental results needed later.

Theorem 5.5.1 For a Markov chain with n-step transition probabilities $p_{ij}^{(n)}$, we have $p_{ij}^{(n)} \to 0$ as $n \to \infty$ whenever state j is transient or null recurrent, while if state j is positive recurrent and aperiodic, then $p_{ij}^{(n)} \to f_{ij}/\mu_j$ as $n \to \infty$, where f_{ij} is the probability that the first entrance time into state j from state i is finite, and μ_j is the expected time of return to state j (which is finite since j is positive recurrent).

The case in which state j is transient was shown in Section 5.2. The other two cases have somewhat technical and lengthy proofs. It was shown also in Section 5.2 that for a irreducible chain, either all states are transient or all states are recurrent. It is possible to say more, as the next result shows.

Theorem 5.5.2 In an irreducible Markov chain, either (a) all states are transient, (b) all states are positive recurrent, or (c) all states are null recurrent. Further, all states have the same period.

Theorem 5.5.3 Consider an irreducible and aperiodic Markov chain. Then a stationary distribution exists if and only if all states are positive recurrent. Moreover, when all states are positive recurrent, the stationary vector is unique, and given by $\pi = (\pi_0, \pi_1, \cdots)$, where π_j is the limit of $p_{ij}^{(n)}$ as $n \to \infty$ for each j. In addition, the stationary vector π is the limit of the state probability distributions $\pi^{(n)}$ as $n \to \infty$, for any initial state probability distribution $\pi^{(0)}$.

Proof Suppose first that the chain has a stationary distribution π. Then

$$\pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij}^{(n)} \ , \quad \text{for any } n \geq 1 \ .$$

Since the chain is irreducible the states must be either all transient, all positive recurrent, or all null recurrent. But it can not be the case that all states are transient, or that all states are null recurrent, for then we would have $p_{ij}^{(n)} \to 0$ as $n \to \infty$, and taking the limit as $n \to \infty$ in this equation would show that each π_j is zero, which is not possible. Thus, each state is positive recurrent.

Conversely, suppose each state is positive recurrent. By assumption each state is also aperiodic. Let π_j be the limit of $p_{ij}^{(n)}$ as $n \to \infty$. From Theorem 5.1, each $\pi_j = 1/\mu_j > 0$. Using the Chapman-Kolmogorov equations,

$$p_{\nu j}^{(n+1)} = \sum_{i=0}^{\infty} p_{\nu i}^{(n)} p_{ij} \ , \quad \text{for any state } \nu \ . \tag{5.5.1}$$

Taking the limit as $n \to \infty$ yields

$$\pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij} \ .$$

Hence $\pi = (\pi_0, \pi_1, \cdots)$ satisfies the stationary equations. The solution is unique, for if there is another stationary vector, say v, then for each state j,

$$v_j = \sum_{i=0}^{\infty} v_i p_{ij}^{(n)} \ , \quad \text{for any } n \geq 1 \ .$$

But taking the limit as $n \to \infty$, the right side converges to π_j, showing that $v_j = \pi_j$.

Finally, for an arbitrary initial state probability distributions $\pi^{(0)}$, we have

$$(5.5.2) \qquad \pi_j^{(n)} = \sum_{i=0}^{\infty} \pi_i^{(0)} p_{ij}^{(n)} \to \left(\sum_{i=0}^{\infty} \pi_i^{(0)} \right) \pi_j = \pi_j \,, \quad \text{as } n \to \infty \,,$$

which proves that last statement of the theorem. □

Theorem 5.5.3 gives us a way to identify the irreducible and aperiodic chains for which all states are positive recurrent. Namely, these are the chains for which a stationary distributions exists. A similar chacterization is possible to identify the irreducible chains that have all transient states.

Theorem 5.5.4 In an irreducible Markov chain, with state space $I = \{0, 1, \cdots\}$, all states are transient if and only if the system of equations

$$(5.5.3) \qquad x_i = \sum_{\nu=1}^{\infty} p_{i\nu} x_\nu \,, \quad i = 1, 2, \cdots,$$

has a nonzero, bounded solution.

Proof Suppose first that all states are transient. For each $i = 1, 2, \cdots$, set $x_i^{(n)} = P(T_{i0} > n)$, for $n = 1, 2, \cdots$. Then

$$(5.5.4) \qquad x_i^{(1)} = \sum_{\nu=1}^{\infty} p_{i\nu} \,, \quad \text{and} \quad x_i^{(n+1)} = \sum_{\nu=1}^{\infty} p_{i\nu} x_\nu^{(n)} \,, \quad n = 1, 2, \cdots.$$

These equations follow by conditioning on the next state. For each i, the limit

$$x_i = \lim_{n \to \infty} x_i^{(n)} \,,$$

exists, and by taking the limit as $n \to \infty$ in (5.5.4), it follows that the vector $x = (x_1, x_2, \cdots)$ satisfies (5.5.3). This solution x is bounded. To show that it is nonzero, note that for each i,

$$x_i = \lim_{n \to \infty} x_i^{(n)} = P(T_{i0} = \infty) = 1 - f_{i0} \,.$$

But if all states are transient, then $f_{i0} < 1$ for at least one $i \in \{1, 2, \cdots\}$.

5.10 *Unit 2: Markov Chains*

Conversely, suppose that (5.5.3) has a nonzero bounded solution $y = (y_1, y_2, \cdots)$. We may assume that each component is bounded by one. Now, for each i,

$$|y_i| \leq \sum_{\nu=1}^{\infty} p_{i\nu}|y_\nu| \leq \sum_{\nu=1}^{\infty} p_{i\nu} = x_i^{(1)} .$$

Therefore,

$$|y_i| \leq \sum_{\nu=1}^{\infty} p_{i\nu}|y_\nu| \leq \sum_{\nu=1}^{\infty} p_{i\nu}x_\nu^{(1)} = x_i^{(2)} .$$

Continuing in this fashion yields $|y_i| \leq x_i^{(n)}$ for each i, and any $n \geq 1$. Thus, in the limit as $n \to \infty$, we have $|y_i| \leq 1 - f_{i0}$, for each i. However, as the solution y is nonzero, $f_{i0} < 1$ for at least one i, and therefore state 0 is transient. But the chain is irreducible, and so all states are transient. □

Example 5.5.1 Consider the Markov chain with state space $I = \{0, 1, 2, \cdots\}$, and transition probabilities

$$p_{00} = 1/2\,, p_{01} = 1/2, \quad \text{and} \quad p_{i,i+1} = 1/2,\ p_{i,i-1} = 1/2\,, \quad \text{for } i \geq 1.$$

This chain is irreducible and aperiodic. Attempting to solve the stationary equations shows that all components of a solution must be equal, which is not possible as the components must sum to one. Therefore no stationary solution exists, and so by Theorem 5.5.2, either all states are null recurrent, or all states are transient. Applying Theorem 5.5.4 above and attempting to find a solution of (5.5.3), say $x = (x_1, x_2, \cdots)$, shows that for each $i \geq 2$, we have $x_i = ix_1$. Therefore no nonzero, bounded solution exists. It follows that the chain does not consist of all transient states. The only possibility that remains is that all states are null recurrent. □

Exercises

5.1 (a) Show that communication is an equivalence relation. (b) Show that if the state i is recurrent and does not communicate with state j, then $p_{ij} = 0$. (c) Suppose state j is accessible from a recurrent state i. Show that $f_{ji} = 1$.

5.2 Complete the proof of Theorem 5.2.4 .

5.3 (a) Show that the sets constructed in the proof of Theorem 5.3.1 are closed. (b) Show that the set of all states that can be reached from a given state forms a closed set.

5.4 Show that for a transient state j, and any initial state i,

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} = \frac{f_{ij}}{(1 - f_{jj})} .$$

Help: For $i = j$, use the idea of the proof of Theorem 2.2, and the result of Theorem 5.2.1 . For $i \neq j$, assume the chain starts in state i, and let V denote the number of visits to state j. Relate the expected value $E(V)$ to the sum, using the idea in the proof of Theorem 2.2. Next express $E(V)$ using conditional expectation, conditioning on the event that the first entrance time from state i to state j is finite. Then use the result of Theorem 5.2.1.

5.5 Verify the limits in (5.5.1) and (5.5.2) in the proof of Theorem 5.5.3. Help : For (5.5.1), first bound the sum on the right below by the sum as i ranges from 0 to an arbitrary $m \geq 1$. Then take the limit to show that each π_j is bounded below by $\pi_0 p_{0j} + \pi_1 p_{1j} + \cdots$. Finally, if strict inequality holds for some j, then summing j yields a contradiction. For (5.5.2), note that for any fixed j, given $\epsilon > 0$, the sum on the right can be approximated to within ϵ by a finite sum.

5.6 For an irreducible, ergodic Markov chain, let μ_i denote the mean time to return to state i. Give a heuristic argument that the stationary probabilities satisfy $\pi_i = 1/\mu_i$.

HW→ 5.7 Consider an irreducible Markov chain with a finite number of states $\{0, 1, 2, \cdots, m\}$. Let $Q_i = P(\text{visit state } m \text{ before state } 0 \mid \text{start in state } i)$. Then $Q_0 = 0$ and $Q_m = 1$. (a) Find a system of $m - 1$ linear equations that is satisfied by $Q_1, Q_2, \cdots, Q_{m-1}$. Help: Condition on the next state. (b) Show that the matrix which arises in part (a) is nonsingular. Help: Assume this matrix, say A, is singular so that there exists a vector $v \neq 0$ such that $Av = 0$. Normalize v so that one component is 1, and the rest are ≤ 1. You will need the irreducible property.

5.8 Consider an r-state Markov chain for which the probability distribution of the states at time n, $\pi^{(n)}$, converges to a probability distribution π as $n \to \infty$. Suppose that whenever the process enters state i, a reward of amount $R(i)$ is earned. Then the total reward earned up to time n is $\sum_{s=0}^{n} R(X_s)$, where X_s denotes the state at time s. Show that

$$E\left(\frac{1}{n+1} \sum_{s=0}^{n} R(X_s)\right) \to \sum_{i=1}^{r} \pi_i R(i) , \quad \text{as } n \to \infty .$$

Help: If a sequence $\{x_s\}_{s \geq 0}$ converges to x as $s \to \infty$, then

$$\frac{1}{n+1} \sum_{s=0}^{n} x_s \to x , \quad \text{as } n \to \infty .$$

5.12 *Unit 2: Markov Chains*

5.9 Consider a Markov chain for which the probability distribution of the states at time n, $\pi^{(n)}$, converges to a probability distribution $\pi = (\pi_1, \pi_2, \cdots)$ as $n \to \infty$. (a) Explain why π_i, in addition to being the long run fraction of time the process is in state i, is also the fraction of transitions into state i, as well as the fraction of transitions from state i. Help: Suppose you observe the process for a large number of transitions. Then the number of occasions the process is in state i, the number of transitions into state i, and the number of transitions from state i, differ by at most one. (b) Give an interpretation of $\pi_i p_{ij}$ as the fraction of transitions of a certain type. Use this interpretation to interpret $\sum_i \pi_i p_{ij}$. Finally, use these interpretations, and the result of part (a) to explain the formula $\pi_j = \sum_i \pi_i p_{ij}$. (c) Let A denote a set of states, and let $\overline{A}$ denote the compliment of A. Give an interpretation of the quantity

$$\sum_{i \in A} \sum_{j \in \overline{A}} \pi_i p_{ij} \, .$$

Use this intrepretation to explain the identity

$$\sum_{i \in \overline{A}} \sum_{j \in A} \pi_i p_{ij} = \sum_{i \in A} \sum_{j \in \overline{A}} \pi_i p_{ij} \, .$$

5.10 For the Markov chain in Example 5.4.1, find $f_{42}^{(n)}$ for $n = 1, 2,$ and 3. It will help to use computing software, such as MATLAB. Also calculate these probabilities by direct reasoning. For this part you will not need computing software.

5.11 Consider the Markov chain with probability transition matrix

$$\begin{bmatrix} 0.7 & 0 & 0 & 0 & 0 & 0.3 \\ 0.1 & 0 & 0 & 0.4 & 0.5 & 0 \\ 0 & 0 & 0.2 & 0 & 0.8 & 0 \\ 0 & 0.7 & 0.3 & 0 & 0 & 0 \\ 0 & 0 & 0.8 & 0 & 0.2 & 0 \\ 0.4 & 0 & 0 & 0 & 0 & 0.6 \end{bmatrix}$$

Use the graph of the chain to partition the state space into closed sets of communicating recurrent states, and a set of transient states. Use this partition to re-number the states so that the probability transition matrix is block diagonal, with blocks also along the lower part of the matrix. In other words, the matrix has the form

$$\begin{bmatrix} D_1 & 0 & 0 & 0 \\ 0 & D_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ S_1 & S_2 & \cdots & Q \end{bmatrix},$$

where the entries above are themselves matrices.

5.12 For the Markov chain of the previous problem, find the probabilities f_{ij} for each transient state i and recurrent state j. As in the previous problem, re-number the states first.

5.13 Consider the queueing model of Example 4.3.3. (a) Show that if $\rho > 1$, then all states are transient. Help: Use Theorem 5.5.4. Show that when $\rho > 1$, the system of equations has a solution $x_i = r^i$, $i = 1, 2, \cdots$, for some $r : 0 < r < 1$. (b) Show that if $\rho < 1$, then all states are positive recurrent. Help: Use Theorem 5.5.3, and apply the z-transform to show that the stationary equations have a solution. (c) Show that if $\rho = 1$, then all states are null recurrent. Help: Recall first that when $\rho = 1$, a stationary solution does not exist. Next, use a proof by contradiction to show that when $\rho = 1$, the system of equations in Theorem 5.5.4 has no solution.

5.14 Show that in Markov chain with a finite number of states, there can be no null recurrent states. Help: First note that for a null recurrent state, the set of a states that communicate with it form a closed set. Then use Theorem 5.5.1, and an argument similar to the one used to prove Theorem 5.3.3.

Chapter 6

Finite Markov Chains

6.1 Introduction A finite Markov chain is a Markov chain with a finite number of states. This case turns out to be much simpler than that of an infinite state space. There are a few special types of finite Markov chains which are of interest, and this chapter concerns two of them.

6.2 Absorbing Markov chains A state i is said to be **absorbing** if the transition probability $p_{ii} = 1$. A finite Markov chain is called an **absorbing chain** if there is at least one absorbing state, and from each state that is not absorbing, it is possible to reach a state that is absorbing. In other words, if i is a state that is not absorbing, then there is an absorbing state j such that $p_{ij}^{(n)} > 0$ for some n. This definition implies that the absorbing states are the only recurrent states, and all other states are transient. The transition matrix for an absorbing Markov chain can be written, possibly after reordering the states, in the form

$$P = \begin{bmatrix} I & 0 \\ R & Q \end{bmatrix}, \tag{6.2.1}$$

where I is an identity matrix. The first set of rows correspond to the absorbing states, and the last set of rows correspond to the transient states. The n-step probability transition matrix is then

$$P^{(n)} = P^n = \begin{bmatrix} I & 0 \\ R_n & Q^n \end{bmatrix}, \tag{6.2.2}$$

for some matrix R_n.

Example 6.2.1 Consider a simplified model of a charge account system in which an account is said to be in state n if its oldest unpaid debt is n months old as of the billing

6.2 *Unit 2: Markov Chains*

date for that debt. When a payment is made, the dollars go to reduce the oldest debts in the account. Suppose there are five states $\{u, b, 0, 1, 2\}$, where u designates paid up, and b desiginates bad debt. States u and b are viewed as absorbing states. Referring to (6.2.1), suppose the transition matrix, with this ordering of the states, has

$$R = \begin{bmatrix} .3 & 0 \\ .2 & 0 \\ .1 & .1 \end{bmatrix}, \text{ and } Q = \begin{bmatrix} .6 & .1 & 0 \\ .2 & .5 & .1 \\ .2 & .2 & .4 \end{bmatrix}.$$

Thus, the first column of the matrix R corresponds to the paid-up state, and the second column corresponds to the bad debt state. The states $\{0, 1, 2\}$ are transient, and from any transient state it is possible to reach an absorbing state. □

Theorem 6.2.1 For each transient state j, we have $p_{ij}^{(n)} \to 0$ as $n \to \infty$, for any initial state i.

This result was shown in the previous chapter, and holds for any Markov chain. For the case at hand, it tells us that $Q^n \to 0$ as $n \to \infty$. This property will be needed later.

Theorem 6.2.2 Let $A^{(n)} = Q^n R$. Then the (i, j) element of $A^{(n)}$ is equal to the probability that the process reaches the absorbing state j for the first time in $n + 1$ steps, given that the process started in transient state i.

Proof From the Chapman-Kolmogorov equations, we see that the desired probability, namely, that the process reaches the absorbing state j for the first time in $n + 1$ steps, given that the process started in state i, is equal to

$$\sum_{k \in T} q_{ik}^{(n)} r_{kj},$$

where $q_{ik}^{(n)}$ is the (i, k) element of the matrix Q^n, and T denotes the set of transient states. By matrix multiplication, this sum is seen to be the (i, j) element of the matrix $A^{(n)}$. □

Consider now the matrix $R + QR + Q^2R + \cdots$. Using Theorem 6.2.2, it follows that entry (i, j) of this matrix is the probability that the process eventually reaches absorbing state j, given that it starts in transient state i. But using Theorem 6.2.1, it can be shown that the matrix $I - Q$ is invertible, and that the inverse is

$$N = (I - Q)^{-1} = I + Q + Q^2 + \cdots.$$

Thus

$$R + QR + Q^2R + \cdots = (I + Q + Q^2 + \cdots)R = NR.$$

This result is summarized in the next theorem.

Theorem 6.2.3 The probability that the process eventually reaches absorbing state j, given that it starts in transient state i, is given by the (i, j) entry of the matrix NR, where $N = (I - Q)^{-1}$.

Example 6.2.2 Consider the simplified model of a charge account system given in Example 6.2.1. Then we have

$$NR = (I - Q)^{-1}R = \begin{bmatrix} .99 & .01 \\ .96 & .04 \\ .82 & .18 \end{bmatrix}.$$

From the second column of this matrix, for instance, we find that 1% of the new debts (0 months old) will eventually become bad debts, about 4% of the one month old debts will eventually become bad debts, and about 18% of the two month old debts will eventually become bad debts. □

6.3 Regular Markov chains A finite Markov chain is said to be **regular** if for some integer m, all entries in the m-step probability matrix $P^{(m)}$ are positive. Equivalently, a finite chain is regular if it is irreducible and aperiodic. The inventory model considered earlier in the introductory notes on Markov chains is an example of a regular chain.

Example 6.3.1 Suppose the probability transition matrix is given by

$$P = \begin{bmatrix} .60 & .40 \\ .13 & .87 \end{bmatrix}.$$

The probabilities, by the way, are based on rainfall data in Tel-Aviv over the years 1923 to 1970. The first row and column correspond to the wet-day state, and the second row and column correspond to the dry-day state. A day is considered wet if total rainfall is at least 0.1 mm during that day. □

Theorem 6.3.1 Let P be the probability transition matrix for a regular finite Markov chain with r states. Let e denote the column r-vector, each of whose components is one. Then for any column r-vector x, there is a scalar λ_x such that $P^n x \to \lambda_x e$, as $n \to \infty$.

Proof Assume first that the result is true when $P > 0$. Suppose then that $P^m > 0$. For integer $n \geq m$, set $n = km + l$, for nonnegative integers k and l. Then

$$P^n x = P^l P^{km} x = P^l(P^{km}x - \lambda_x e) + P^l(\lambda_x e) = P^l(P^{km}x - \lambda_x e) + \lambda_x e\,.$$

Note that the last result holds because the sum of the elements in any row of P^l is one (when $l = 0$, P^l is the identity matrix). Now, since the first term on the right converges to zero as $k \to \infty$, the conclusion follows. Thus, we now suppose that $P > 0$.

Let d be the magnitude the smallest entry in P. Then $0 < d \leq 1/2$. Consider the sequence $\{x^{(n)}\}$ for $n = 0, 1, \cdots$, where $x^{(0)} = x$, and $x^{(n+1)} = Px^{(n)}$ for $n \geq 0$. For each n, denote by a_n and b_n the smallest and largest entry in the vector $x^{(n)}$ respectively. Let ν be the index of the smallest entry in $x^{(n)}$. Then, for each component k,

$$x_k^{(n+1)} = \sum_{i=1}^{r} p_{ki}x_i^{(n)} = p_{k\nu}a_n + \sum\nolimits_{i\neq\nu} p_{ki}x_i^{(n)} \leq p_{k\nu}a_n + (1 - p_{k\nu})b_n\,.$$

But the term on the right is equal to $b_n - p_{k\nu}(b_n - a_n)$ which is bounded above by $b_n - d(b_n - a_n)$. It follows that

$$b_{n+1} \leq b_n - d(b_n - a_n)\,.$$

Using analogous reasoning, it can be shown that

$$a_{n+1} \geq a_n + d(b_n - a_n)\,. \tag{6.3.1}$$

These two equations give us

$$b_{n+1} - a_{n+1} \leq [b_n - d(b_n - a_n)] - [a_n + d(b_n - a_n)] \leq (1 - 2d)(b_n - a_n)\,.$$

But $0 \leq 1 - 2d < 1$. Therefore, $b_n - a_n \to 0$ as $n \to \infty$, which completes the proof. □

Some consequences of this theorem are immediate. Let e_i be the r-column vector with 1 in the i-th component, and zeros elsewhere. Then there is a scalar w_i such that $P^n e_i \to w_i e$ as $n \to \infty$. Let w be the r-row vector whose i-th component is w_i, and let W be the $r \times r$ matrix whose i-th row is the vector w. Then (a) Let u be any row probability r-vector (the components of u are nonnegative and sum to 1). Then $uP^n \to w$, as $n \to \infty$, (b) $P^n \to W$ as $n \to \infty$, (c) $PW = WP = W$, and in particular $wP = w$, (d) $w > 0$, (e) w is the unique solution, up a scalar multiple, of the equation $xP = x$, where x is a row r-vector.

Denote the state probability vector at time n by $\pi^{(n)}$. Thus, $\pi_i^{(n)} = P(X_n = i)$, and at any time n, $\pi^{(n)} = \pi^{(0)}P^n$, where $\pi^{(0)}$ is the initial state probability vector. For a

regular Markov chain, it follows that $\pi^{(n)} \to w$ as $n \to \infty$. Hence, we can interpret w as the long-run state probability vector for the process.

Example 6.3.2 Consider the weather model of Example 6.3.1. Solving the equation $wP = w$, subject to the components of w summing to 1, yields the approximate stationary solution $w = (0.25\,,\, 0.75)$. Thus, in the long-run, there is a 25% chance of having a wet day in Tel-Aviv, and a 75% chance of having a dry day. □

Exercises

6.1 (*a*) Verify that the matrix $I - Q$ is invertible. Help: Show there is no nonzero vector x such that $(I - Q)x = 0$. (*b*) Set $S_n = I + Q + \cdots + Q^n$. Show that the limit of S_n as $n \to \infty$ exists, and that this limit is the inverse of $I - Q$. Help: consider the matrix $(I - Q)S_n$. (c) Find an expression for the matrix R_n in the n-step transition matrix (6.2.2). Show that the limit of R_n as $n \to \infty$ exists and find an expression the limit.

6.2 Consider an absorbing Markov chain with probability transition matrix (2.1) for which

$$R = \begin{bmatrix} .1 & .2 \\ .3 & .1 \end{bmatrix}, \quad \text{and} \quad Q = \begin{bmatrix} .4 & .3 \\ .5 & .1 \end{bmatrix}.$$

Find the probability that the process will eventually reach absorbing state 1, given that it started in transient state 4.

6.3 For an absorbing Markov chain, let V_{ij} denote the number of visits made to transient state j before absorption, given that the process starts in transient state i. Let B be the matrix whose (i, j)-*th* entry is $b_{ij} = E(V_{ij})$. (*a*) Show that $B = N = (I - Q)^{-1}$. Help: Condition on the next state X_1. Write out the resulting equations for each (i, j), and then appeal to matrix multiplication to get the result. For $i = j$, include this initial condition as one of the visits to state j, among possibly others before absorption. (*b*) Given that the process starts in transient state i, give a formula for the expected number of steps until absorption.

6.4 Regarding the proof of Theorem 6.3.1, derive equation (6.3.1) and show that $0 < d \leq 1/2$. Further, verify the results listed in the paragraph after the statement of Theorem 6.3.1.

6.5 Consider a regular Markov chain, and denote by T_{ij} the first entrance time into state j, given that the process starts in state i. Set $m_{ij} = E(T_{ij})$. (*a*) Show that $m_{ij} = 1 + \sum_{k \neq j} p_{ik} m_{kj}$. Help: Use conditional expectation, and condition on the next state. (*b*) Let $(w_1, w_2, \cdots, w_r)$ be the stationary probability vector for the process. Show

that $m_{jj} = 1/w_j$, for each state j. Help: Use the result of (a). Multiply the i-th equation by w_i, and then sum over i. (c) Give a heuristic argument to justify the result of part (b).

6.6 Show that a finite chain is regular if and only if it is irreducible and aperiodic. Help: To show the if-part, use Theorem 5.5.1 and Theorem 5.5.2 from Chapter 5.

6.7 Let us say that a finite Markov chain with one-step transition probability matrix P is *asymptotic* if the limit of P^n as $n \to \infty$ exists. For irreducible finite Markov chains, show that a chain is asymptotic if and only if it is regular. Using the result of problem 6.6, what is the only way in which an irreducible finite chain could not be asymptotic?

The Pallet Problem

Several breweries in Canada cooperate in maintaining a collection of pallets for distribution of their product. Pallets are used for shipment to retail outlets and for the return of empty bottles. The use of common bottles and pallets by breweries allowed the return of empties to any brewery. New bottles, when shipped to the breweries from the glass manufacturers, were shipped on new pallets. The bill for new pallets was included with the bill for the new bottles. Also, additional pallets could be ordered by the individual breweries. Records of purchases were kept, and at the end of the year, the costs would be split proportionately among the breweries. The portion of the total expenses paid by a particular brewery was determined by the number of new purchases times the market share of that brewery. The pallet pool consisted of about 200, 000 pallets.

The policy for maintaining pallets was to repair them as needed. However, increasing labor costs raised concern that this policy may become too expensive, and so a study of possible maintanence policies was begun.

Information and data

Records were available which showed the average damage rates for pallets. These are summarized in the following table. The table indicates the pallets according to condition. The entries show the percentage of pallets that are damaged for the first time in the specified year.

Pallet condition	yr. one	yr. two	yr. three	yr. four
New pallet	22	45	33	–
Pallet repaired in year one	–	47	48	5
Pallet repaired in year two	–	–	83	17

Thus, for example, of the pallets that were damaged and repaired in year one, 47 percent were next damaged in year two, 48 percent were next damaged in year three, and 5 percent were next damaged four. The third row of the table refers to pallets that were repaired in year two, but which may also have been repaired in year one.

The foreman in charge of maintanence estimated that approximately 10 percent of damaged pallets were unrepairable. Further, pallets over two years old were not

considered worth repairing and thus were scrapped. Pallets last only four years, and then are scraped. In our analysis we will assume a pallet is damaged only once in a year.

Based on recent records, it was estimated that the average cost of a new pallet was $ 4.47, and that the average slavage value was $ 0.55. However, salvage value depènded on condition, and had been as much as $ 0.75. The foreman thought that a pallet in good condition could be salvaged for as much as $ 1.50. The average cost of reparing a pallet was $ 2.07. Records showed that during the last year, 59,000 new pallets were purchased, and 32, 000 were sold as scrape.

A Markov Chain Model

To model the effect of the current maintenance policy, the use of a regular Markov chain is suggested. The discussion below shows how to form such a model. Your task will be to complete the analysis, and then develop similar models for the other specified maintenance policies listed below.

The states of the Markov chain model, indicated below, are based on the age and condition of a pallet.

1. Zero years old.
2. One year old and undamaged
3. Two years old and undamaged
4. One year old, and repaired in year one
5. Two years old and repaired in year two (and possibly in year one)
6. Two years old and repaired in year one only
7. Three years old.

The transition probabilities are estimated using the data in the table above.

Transitions from state 1 : The only possible next states are 1, 2 and 4. A transition from state 1 to 1 means that a new pallet was damaged in year one and had to be scraped. Thus, $p_{11} = (0.22)(0.10) = 0.022$, since there is a 22 percent chance that a new pallet is damaged year one, and given that it is damaged, a 10 percent chance that it needs to be scraped. Next, $p_{12} = 1 - 0.22 = 0.78$, since 22 percent of new pallets are damaged in year one. Finally, $p_{14} = (0.22)(0.90) = 0.198$, since there is a 22 percent chance that a new pallet is damaged in year one, and given that it is damaged, a 90 percent chance that it is repairable.

Transitions from state 2 : The possible next states are 1, 3 and 5. A transition from state 2 to 1 means that a new pallet was damaged for the first time in year two, and had to be scraped. Thus, $p_{21} = (0.45/0.78)(0.10) = 0.0577$. The first factor is the probability that a new pallet is damaged in year two given that it was not damaged in year one. The second factor accounts for the 10 percent of damaged pallets that need to be scraped. Next, $p_{23} = 1 - (0.45/0.78) = 0.423$, since the term in parenthesis is the probability that a new pallet is damaged in year two, given that it was not damaged in year one. Finally, $p_{25} = (0.45/0.78)(0.90) = 0.519$, using the same reasoning as above to justify p_{21}.

6.8 *Unit 2: Markov Chains*

Transitions from state 3 : The first row of the table above tells that such a pallet will be damaged in year three, and being over two years old, it will be scraped. Thus, the only next state is 1, and so $p_{31} = 1$.

Transitions from state 4 : The possible next states are 1, 5 and 6. A transition from state 4 to 1 means that a pallet damaged and repaired in year one, was damaged in year two and had to be scraped. Thus, $p_{41} = (0.47)(0.10) = 0.047$, since from the table, a pallet damaged in year one has a 47 percent chance of being damaged in year two and there is a 10 percent chance that a damage pallet needs to be scraped. Next, by similar reasoning, $p_{45} = (0.47)(0.90) = 0.423$. Finally, $p_{46} = 1 - (0.47) = 0.530$, since a pallet is damaged in year one has a 47 percent chance of being damaged in year two.

Transitions from state 5 : The possible next states are 1 and 7. A transition from state 5 to 1 to mean that a pallet repaired in year two was damaged in year three. Thus, from the table, $p_{51} = 0.830$, where the factor of 10 percent is not used, since damaged pallets over two years old are scraped. Next, again directly for the table, we have $p_{57} = 0.17$.

Transitions from state 6 : The possible next states are 1 and 7. A transition from state 6 to 1 means that a pallet repaired in year one only was damaged in year three. Thus, from the table, $p_{61} = (0.48)/(0.53) = 0.906$, which is the conditional probability that a pallet, damaged in year, is damage in year three, given that it was not damaged in year two. Again, the factor of 10 percent is not used, since damaged pallets over two years old are scraped. Finally, $p_{67} = (0.05)/(0.53) = 0.094$, which is the conditional probability that a pallet, damaged in year one, is not damage in year three, given that it was not damaged in year two.

Transitions from state 7 : The only next state is 1. Thus, $p_{71} = 1$.

Problem Analysis

As mentioned earlier, your task is to complete the analysis above, and then develop similar models for the other specified maintenance policies listed below. Use the Markov chain model to determine the long-run state probabilities. From these long-run probabilities, form the following cost function to evaluate the policy :

$$C = c_N f_N + c_R f_R - c_S f_S \ ,$$

where f_N is the long-run fraction of new pallets per year, f_R is the long-run fraction of pallets repaired per year, and f_S is the long-run fraction of pallets that are scraped per year. Also, c_N is the average cost of a new pallet, c_R is the average cost to repair a pallet, and c_S is the average salvage value. Use the long-run state probability distribution to estimate the fractions f_N, f_R, and f_S.

Note that in the Markov chain model, it is assumed that a scraped pallet is replaced by a new one. The records indicate however that 59,000 new pallets were introduced into

the system and only 32,000 were salvaged. We shall interpert this to mean that not all scraped pallets were salvagable. Thus, although we might otherwise take $f_S = f_N$, we can use the data to set $f_S = \alpha f_N$, where α is a fraction around $32,000/59,000$.

Possible maintenance policies In your analysis, consider at least the following policies for maintaining the pallets.

1. Repair as needed. This is the current policy.
2. Repair only once.
3. Repair only in the first year.
4. Do not repair.

Chapter 7

A Random Walk Model

7.1 Introduction Consider a Markov chain with state space the integers, and one-step transition probabilities given by $p_{i,i-1} = q_i$, $p_{i,i} = r_i$, $p_{i,i+1} = p_i$, and $p_{ij} = 0$ otherwise. Thus, for each state i, q_i, p_i, and r_i are nonnegative numbers that sum to one. We shall refer to such a Markov chain as a **random walk**. This chapter considers two examples in which this type of chain arises.

7.2 A gambler's ruin problem One way to think of the random walk model is to imagine that the states represent a gambler's fortune, and state 0 represents ruin. Fix a state $n > 1$, and for each state i, $0 < i < n$, consider the probability of reaching state 0 before reaching state n. Denote this probability by P_i. Using a conditional probability argument, it follows that

$$P_i = q_i P_{i-1} + r_i P_i + p_i P_{i+1} \ , \tag{7.2.1}$$

with boundary conditions $P_0 = 1$ and $P_n = 0$. Since $1 - r_i = q_i + p_i$, this difference equation can be written as $p_i d_i = q_i d_{i-1}$, where $d_i = P_{i+1} - P_i$, for $i = 1, 2, \cdots, n-1$. It follows that $d_i = a_i d_0$, where

$$a_i = \frac{q_i q_{i-1} \cdots q_1}{p_i p_{i-1} \cdots p_1} \ , \quad \text{for} \quad i = 1, 2, \cdots, n-1 \ .$$

Since $d_0 + d_1 + \cdots + d_{n-1} = P_n - P_0 = -1$, we can solve for d_0 to obtain

$$d_0 = \frac{-1}{1 + a_1 + a_2 + \cdots + a_{n-1}} \ .$$

Finally, using $P_i = 1 + d_0 + d_1 + \cdots + d_{i-1}$, give us

$$P_i = \frac{a_i + \cdots + a_{n-1}}{1 + a_1 + a_2 + \cdots + a_{n-1}} \ , \quad \text{for} \quad i = 1, 2, \cdots, n-1 \ .$$

This expression for the probability of ruin is not easily analysed. However, we can see that it is decreasing with i, as would be expected. Indeed, the greater our initial funds, the less chance there is of ruin. Under special assumptions concerning the transition probabilities, simpler expressions are possible. For example, suppose p and q are positive numbers such that $p+q=1$, and assume that for each state $i \geq 1$, $q_i = q$, $r_i = 0$, and $p_i = p$, while $q_0 = 0$, $r_0 = q$, and $p_0 = p$. Set $\rho = q/p$. Then, for $\rho \neq 1$, we have

$$P_i = \frac{\rho^n - \rho^i}{\rho^n - 1}, \quad \text{for} \quad i = 0, 1, 2, \cdots, n, \tag{7.2.2}$$

and for $\rho = 1$,

$$P_i = 1 - \frac{i}{n}, \quad \text{for} \quad i = 0, 1, 2, \cdots, n. \tag{7.2.3}$$

7.3 A queueing model Consider a single server queue that is modeled as follows. Time as divided into intervals of some specified length h. It is assumed that (a) at most one customer arrives during a time interval, (b) at most one service completion occurs in a time interval, (c) arrivals in separate time intervals occur independently, as do service completions. Also, arrivals and service completions occur independently of each other. Finally, (d) the probability of a simultaneous arrival and departure is negligible.

For this model, we shall suppose that if a customer arrives and finds there are m customers in the system, waiting or being served, then that customer does not join the system. Denote by p the probability of an arrival during an interval of time h, and denote by r the probability of a service completion, given that service is in progress. Define X_n as the number of customers in the system at the end of the n-th time interval. Then we have a random walk with states $0, 1, 2, \cdots, m$. The rows of the transition probability matrix P are

$$i = 0: \qquad r_0 = 1 - p, \qquad p_0 = p$$

$$1 \leq i \leq m-1: \quad q_i = (1-p)r, \quad r_i = 1 - (1-r)p - (1-p)r, \quad p_i = (1-r)p$$

$$i = m: \qquad q_m = r, \qquad r_m = 1 - r$$

Assuming that $0 < r < 1$ and $0 < p < 1$, we have a regular Markov chain. The stationary vector for this chain is the solution of the system of equations $wP = w$, where

$w = (w_0, w_1, \cdots, w_m)$. Suppose $m > 1$. Starting with the first equation, and working with each successive equation, one can show that

$$w_1 = \frac{s}{1-r}\, w_0, \quad w_i = s w_{i-1}, \text{ for } 2 \le i \le m-1, \quad \text{and} \quad w_m = s(1-p) w_{m-1},$$

where $s = (1-r)p/(1-p)r$. In the following, we will consider only the case $s \neq 1$. Since the components of w must sum to one, it follows that

$$w = k\left(1-r, s, s^2, \cdots, s^{m-1}, (1-p)s^m\right), \text{ where } k = \frac{1-\rho}{(1-r)-(1-p)\rho s^m},$$

and where $\rho = p/r$. This parameter ρ can be viewed as a measure of traffic intensity. The expression for w does not lend itself easily to analysis, but some observations are possible. For example, if $\rho < 1$ and m is very large, then $k \approx (1-\rho)/(1-r)$, and so $w_0 \approx 1-\rho$. Thus, when traffic intensity is not too large ($\rho < 1$), and arriving customers stay even though many customers are there already (m large), the fraction of time the system is empty is approximately $1-\rho$. On the other hand, if $\rho > 1$ and m is large, then $w_m \approx 1-(1/\rho)$. Thus, when the system tends to grow ($\rho > 1$), and arriving customers will stay even though many customers are there already (m large), the fraction of time the system attains maximum capacity is approximately $1-(1/\rho)$.

An approximate expression for w is obtained if we assume that the length of the time interval h is very small, and that $p = \lambda h + o(h)$ and $r = \mu h + o(h)$, for some fixed parameters λ and μ. Then s and ρ are each approximately λ/μ, and if $\rho \neq 1$,

$$w \approx k\left(1, \rho, \rho^2, \cdots, \rho^m\right), \quad \text{where } k = \frac{1-\rho}{1-\rho^{m+1}}.$$

This expression is more tractable. For example, denoting by N the number of customers in the system, then we can find

$$E(N) \approx k(\,0 \cdot 1 + 1 \cdot \rho + 2\rho^2 + \cdots + m\rho^m\,),$$

which can be evaluated in closed form. Using this expression, it can be shown, as indicated in an exercise below, that for large m, $E(N) \approx \rho/(1-\rho)$ when $\rho < 1$.

Exercises

7.1 Determine the probability of ruin for the model in Section 2 when $p_i = q_i$ for each $i = 1, 2, \cdots, n-1$.

7.4 *Unit 2: Markov Chains*

7.2 Verify the basic recurrence formula (7.2.1) for the probability of ruin. Also, verify formulas (7.2.2) and (7.2.3) for the probability of ruin in the special case considered.

7.3 Let Q_i denote the probability that state n is reached before state 0, given that we start in state i. Use the conditional probability argument of Example 7.1 to form a recurrence formula for Q_i, and thus find expressions for these quantities. Verify that $P_i + Q_i = 1$ for each i.

7.4 For the single server queue model of Section 7.3, take $\rho \neq 1$, and use the simplifying assumption of a small time interval h, as indicated at the end of the section. (a) Verify the given approximate expression for $E(N)$ when m is large, and $\rho < 1$. (b) Let L denote the number of waiting customers. Find a formula for $E(L)$, the expected value of L, under the steady state conditions. Help: For both parts (a) and (b), form the derivative of $f(x) = 1 + x + \cdots + x^m$ to get an expression for the sum involved. For part (b), note that $E(L) = 1 \cdot w_2 + 2 \cdot w_3 + \cdots + (m-1)w_m$.

7.5 Consider the special case of the gamblers ruin model presented at the end of Section 7.2. The states are $0, 1, 2, \cdots$, and for positive numbers p and q such that $p + q = 1$, the transition probabilities are : for $i \geq 1$, $q_i = q$, $r_i = 0$, and $p_i = p$, while $q_0 = 0$, $r_0 = q$, and $p_0 = p$. Set $\rho = q/p$. Following the steps below, show that this Markov chain consists of all transient states if $\rho < 1$, but consists of all recurrent states when $\rho \geq 1$. (a) Recall that T_{ij} denotes the first entrance time into state j, given that the process starts in state i. Suppose the process starts in state 0. Condition on the next state to show that

$$P(T_{00} < \infty) = q + pP(T_{10} < \infty) \ .$$

(b) The probability P_i defined in Secton 2 is equal to $P(T_{i0} < T_{in})$. But $T_{in} \geq n - i$, and so it is reasonable to take $P(T_{i0} < \infty)$ as the limit of P_i as $n \to \infty$. Assume this result and find $P(T_{10} < \infty)$. (c) Combine the results of parts (a) and (b) to determine when $P(T_{00} < \infty)$ is less than 1, and when it is equal to 1. Thus find the conditions under which state 0 is transient and under which it is recurrent. Finally, note that the chain is irreducible, and so the states are either all transient or all recurrent.

Chapter 8

Markov Chain Monte Carlo Methods

8.1 Introduction This chapter introduces some of the basic ideas used in Markov chain Monte Carlo methods. These techniques provide a way to generate random observations from a specified probability distribution function. The idea is to construct a Markov chain whose limiting distribution is the specified distribution, and then to simulate the Markov chain to obtain the random observations. In these notes, we consider only finite chains. However, finite chains occur in many applications of Markov chain Monte Carlo methods.

8.2 Time reversible Markov chains An important role is played in Markov chain Monte Carlo methods by time reversible Markov chains, and so we begin with a review of this topic. Consider an irreducible, positive recurrent Markov chain $\{X_n \mid n = 0, 1, \cdots\}$, with transition probabilities $p(x, y)$ and stationary probabilities $\pi(x)$. Following notation that is common in the literature, we will denotes states by letters such as x and y, and drop the subscript notation for the transition probabilities and state distributions.

Suppose the chain has been in operation a long time, so that the stationary distribution describes the state probabilities. Suppose we follow the states backward in time; that is, we consider the process $\cdots X_n, X_{n-1}, X_{n-2}, \cdots$. This sequence of states forms a Markov chain with transition probabilities $r(x, y)$ given by

$$r(x,y) = P(X_m = y \mid X_{m+1} = x) = \frac{P(X_m = y,\, X_{m+1} = x)}{P(X_{m+1} = x)}$$

$$= \frac{P(X_m = y)\, P(X_{m+1} = x \mid X_m = y)}{P(X_{m+1} = x)} = \frac{\pi(y)p(y,x)}{\pi(x)} .$$

$\pi(x)\, r(x|y) = \pi(y)\, p(y,x)$

8.2 *Unit 2: Markov Chains*

To show that the reverse chain has the Markov property, let B denote any event involving states X_{m+2}, $X_{n+3}, \cdots$. Then we must show that

$$P(X_m = y \mid X_{m+1} = x, B) = P(X_m = y \mid X_{m+1} = x) \ .$$

However,

$$P(X_m = y \mid X_{m+1} = x, B) = \frac{P(X_m = y, X_{m+1} = x, B)}{P(X_{m+1} = x, B)} \ .$$

In the numerator we have

$$P(X_m = y, X_{m+1} = x, B) = P(B \mid X_m = y, X_{m+1} = x) P(X_m = y, X_{m+1} = x) \, .$$

But the forward chain has the Markov property, and so

$$P(X_m = y, X_{m+1} = x, B) = P(B \mid X_{m+1} = x) P(X_m = y \mid X_{m+1} = x) P(X_{m+1} = x).$$

Next, in the denominator we have

$$P(X_{m+1} = x, B) = P(B \mid X_{m+1} = x) P(X_{m+1} = x) \ .$$

Returning to the ratio with the last two expressions gives the desired result.

The reverse chain is therefore a Markov chain with transition probabilities $r(x, y)$. If $r(x, y) = p(x, y)$, then the chain is said to be **time reversible** with respect to the distribution π. Thus, from the formula above for $r(x, y)$, the requirement for time reversibility can be expressed as $\pi(x)p(x, y) = \pi(y)p(y, x)$, for all x, y. These equations are sometimes called the **balance equations**. Indeed, they say that for a time reversible chain, the rate at which the process moves from state x to state y is the same as the rate at which the process moves from state y to x.

To find the stationary distribution of a time reversible chain, the following result is useful. Suppose we can find nonnegative numbers $u(x)$ which sum to one, and satisfy the equations $u(x)p(x, y) = u(y)p(y, x)$, for all x, y. Then summing over x, we find that

$$\sum_x u(x)p(x, y) = u(y) \sum_x p(y, x) = u(y) \, .$$

Therefore, by uniqueness of the solution to the stationary equations, it follows that $u(x) = \pi(x)$ for all x. Solving the system of equations $u(x)p(x, y) = u(y)p(y, x)$, for all x and y, is often a convenient way to determine the stationary solution.

Suppose the transition matrix of an irreducible and aperiodic chain is symmetric. Then the stationary distribution is the uniform distribution (see Exercise 8.2), and it then follows that the chain is time reversible. The next example uses this observation.

Example 8.2.1 Let S denote the set of all $N \times N$ matrices with entries 0 or 1. For a given matrix, define the **nearest neighbors** of entry (i, j) as the entries in positions $(i, j-1)$, $(i, j+1)$, $(i-1, j)$ and $(i+1, j)$. Denote by T the subset of S consisting of matrices such that if an entry is 1, then the nearest neighbors are each 0. Suppose we wish to select a matrix in T at random, or, equivalently, we wish to place a uniform distribution on T. If we could conveniently list the elements of T, this task would not pose a problem. But for even moderate values of N, an enumeration of T is not easy. It is not even known in general how many elements are in T. However, use of a Markov chain Monte Carlo method allows us to do the selection without having to know this number. The Markov chain, and the simulation of the chain, are implemented as follows. We start with an element $M \in T$. Next, an entry of M is chosen at random, and we consider changing that entry. Thus, an entry of 0 would be changed to 1, and vice-versa. If the resulting matrix is still in T, then we accept the new matrix. Otherwise, we do not change the matrix. In this way, we are simulating a Markov chain with state space T. The entries in the probability transition matrix are as following : If M and L are in T, and differ in exactly one entry, then $p(M, L) = N^{-2}$, and if they differ by more than one entry, then $p(M, L) = 0$. Finally, $p(M, M)$ is determined so that the row sum is one. It is evident that the transition matrix is symmetric. Further, the chain is irreducible since all states communicate. Also, since $p(M, M) > 0$ for some $M \in T$, the chain is aperiodic. Finally, since the chain is finite, all states are positive recurrent. It follows that the chain is time reversible and the stationary distribution is the uniform distribution on T. □

8.3 Markov chain Monte Carlo methods Let X be a discrete random variable, perhaps vector-valued, and let S denote the set of possible values of X. We are concerned with situations in which the probability distribution $\pi(x) = P(X = x)$, for $x \in S$, is difficult to use analytically, and a direct simulation of the random random variable X is not practical. It is assumed however that the probability distribution π is known up to the normalizing constant that is needed so that the probabilities sum to one. One general formulation of Markov chain Monte Carlo methods is the Hastings-Metropolis algorithm. This algorithm constructs a Markov chain whose stationary distribution is the probability distribution π. A key feature of the construction is that it is

done without having to know the normalizing constant for $\pi(x)$, which may be difficult to find in practice.

The algorithm begins with selection of an irreducible Markov chain whose state space is S. In practice, and as the examples below indicate, the choice of this chain is usually motivated by the particular application. Let $q(x,y)$ denote the transition probabilities for this Markov chain. Then a time reversible Markov chain $\{X_n \,|\, n = 0,\ 1, \cdots\}$ is constructed as follows. Suppose the process is in state x. Generate a random observation, say y, from the probability distribution $\{q(x,z) \,|\, z \in S\}$. Then, with a certain probability $\beta(x,y)$, move to state y, and otherwise stay at state x. The probabilities $\beta(x,y)$ are determined so that Markov chain $\{X_n \,|\, n = 0, 1, \cdots\}$ is time reversible with respect to π. To find them, note first that the transition probabilities for the chain $\{X_n \,|\, n = 0, 1, \cdots\}$ are

$$p(x,y) = q(x,y)\,\beta(x,y)\,, \quad \text{for } y \neq x\,, \quad \text{and} \quad p(x,x) = 1 - \sum\nolimits_{y \neq x} p(x,y)\ .$$

Thus, we need $\pi(x)\,p(x,y) = \pi(y)\,p(y,x)$, for $y \neq x$, or equivalently

$$\pi(x)\,q(x,y)\,\beta(x,y) = \pi(y)\,q(y,x)\,\beta(y,x)\,, \quad \text{for} \quad y \neq x\,.$$

When $q(x,y) = 0$, we do not specify $\beta(x,y)$ as $p(x,y)$ will be zero anyway. Since $\pi(x) > 0$ for all $x \in S$, we can set $\tau(x,y) = \pi(y)\,q(y,x)/\pi(x)\,q(x,y)$. Then $\beta(x,y) = \tau(x,y)\,\beta(y,x)$. Now, if $\tau(x,y) \leq 1$, then a solution is to take $\beta(y,x) = 1$ and $\beta(x,y) = \tau(x,y)$. Otherwise, take $\beta(y,x) = 1/\tau(x,y)$ and $\beta(x,y) = 1$. In summary,

$$\beta(x,y) = min\left\{1, \tau(x,y)\right\} = min\left\{1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right\}\ .$$

With this choice of β, if the chain $\{X_n \,|\, n = 0, 1, \cdots\}$ is irreducible and aperiodic, or in other words regular, then it also time-reversible, and its stationary distribution is π. Thus, in this case, this stationary distribution will be the limiting distribution.

In general, one needs to check that the final chain is irreducible and aperiodic. A sufficient condition for irreducibility is that $q(y,x) > 0$ whenever $q(x,y) > 0$. In this case, $\beta(x,y) > 0$ whenever $q(x,y) > 0$, and hence $p(x,y) > 0$ whenever $q(x,y) > 0$. Therefore, the irreduciblity of the original chain implies the same for the final chain. That the new chain will be aperiodic is likely, since it may well be that $p(x,x) > 0$ for some x. Moreover, if the original chain is regular, and $\beta(x,y) > 0$ whenever $q(x,y) > 0$, then the new chain will also be regular.

Example 8.3.1 Here, we consider first a simple example, just to illustrate the algorithm. Suppose we wish to generate a random observation from the distribution $\pi = \{1/6, 1/3, 1/2\}$. Introduce the irreducible Markov chain with state space $\{1, 2, 3\}$, and probability transition matrix

$$q = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix} .$$

Then the probabilities $\beta(x, y)$, for $x \neq y$, are determined from the formula

$$\beta(x, y) = min\left\{1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}\right\} .$$

For example,

$$\beta(1, 2) = min\left\{1, \frac{\pi(2)(1/2)}{\pi(1)(1/2)}\right\} = min\left\{1, \frac{1/3}{1/6}\right\} = 1 .$$

In the same way, we find that $\beta(1, 3) = 1$, $\beta(2, 1) = 1/2$, $\beta(2, 3) = 1$, $\beta(3, 1) = 1/3$, $\beta(3, 2) = 2/3$. Thus, the probability transition matrix for the Markov chain used in the simulation is

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

This chain is irreducible, time reversible, and has stationary distribution π. Further, because the chain is regular, the stationary distribution is also a limiting distribution, and thus simulation can be used to generate a random observation from π. Note, by the way, that had we started the algorithm with the irreducible chain whose probability transition matrix is

$$q = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} ,$$

the algorithm would fail. The resulting Markov chain would have a probability transition matrix equal to the identity, and thus the chain would not be irreducible as required. □

A more interesting application of the Hasting-Metropolis method is the following.

8.6 *Unit 2: Markov Chains*

Example 8.3.2 – The Ising model Let S be the set of $N \times N$ matrices with entries 1 or -1. Recall that matrix entries at (i,j) and (s,t) are nearest neighbors if $|i-s| + |j-t| = 1$. Define the energy of $M \in S$ by

$$E(M) = -\sum M(i,j)\, M(s,t)\,,$$

where the sum is taken over all nearest neighbors. The value $M(i,j)$ is called the spin at site (i,j) and the energy is minimized when all spins are the same. The Ising model specifies the following probability distribution on S which weighs matrices of lowest energy highest. Given $r > 0$, define for $M \in S$,

$$\pi_r(M) = \frac{exp(-rE(M))}{\sum_{B \in S} exp(-rE(B))}\,.$$

Note that calculation of the denominator in this expression requires some effort. However, this calculation is unnecessary with the Markov chain Monte Carlo approach. The Markov chain is traversed as follows. Start with $M \in S$, select an entry at random and consider reversing its value. Thus, the underlying irreducible Markov chain is specified by the transitions probabilities

$$q(R,L) = \frac{1}{N^2} \quad \text{for} \quad R \sim L\,,$$

and $q(R,L) = 0$ otherwise, ~~except for the diagonal entry~~. The notation $R \sim L$ means that the matrices differ in exactly one entry. Now let B denote the matrix which would result if the randomly chosen value of M is reversed. Then we accept B with probability

$$\beta(M,B) = min\left\{1, \frac{\pi_r(B)(1/N^2)}{\pi_r(M)(1/N^2)}\right\} = min\left\{1, \frac{\pi_r(B)}{\pi_r(M)}\right\}.$$

Thus, the transition probabilities for the Markov chain are defined by

$$p(M,B) = \frac{1}{N^2}\, min\left\{1, \frac{\pi_r(B)}{\pi_r(M)}\right\}, \quad M \sim B\,,$$

and $P(M,M)$ is defined so that the row sum is one.

The original chain in this case is irreducible, but it is periodic with period two. However, we have $\beta(M,B) > 0$ whenever $q(M,B) > 0$. Hence, the final chain is also irreducible. Further, for at least one $M \in S$, we will have $0 < \beta(M,B) < 1$ for some B such that $M \sim B$. Therefore, $P(M,M) > 0$ for this M, and hence the final chain is aperiodic. Thus, the final Markov chain is regular, and so it is time-reversible with stationary distribution π_r. The simulation is intuitively reasonable. It encourages a move

from matrix M to a matrix B which has lower energy than M, or equivalently, for which $\pi_T(B)$ is larger than $\pi_T(M)$. □

8.4 Gibbs sampling A setting for Gibbs sampling is the following. Assume X is a vector-valued random variable, $X = (X_1, X_2, \cdots, X_n)$, with probability distribution $f(x)$ which may be specified only up to a constant multiplier. Let $\mathcal{A}$ denote a given subset of the possible values of X. Then we wish to generate a random vector whose distribution is the conditional distribution of X, given that $X \in \mathcal{A}$. This conditional distribution is

$$\pi(x) = P(X = x \mid X \in \mathcal{A}) = \frac{f(x)}{P(X \in \mathcal{A})} , \quad \text{for} \quad x \in \mathcal{A} .$$

The approach is to form a Markov chain with state space $\mathcal{A}$, which is time reversible with respect to distribution $\pi(x)$. It is assumed that we are able to generate observations from the conditional distribution

$$c_i(z) = P(X_i = z \mid X_j = x_j , \text{ for } j \neq i) .$$

The transitions are determined as follows. Suppose we are in state x. First, a coordinate index i is chosen at random from among the indices $1, 2, \cdots, n$. Next, a random observation is generated from this conditional distribution $c_i(z) = P(X_i = z \mid X_j = x_j ,$ for $j \neq i)$. Suppose the observed value is u. Then the candidate for the next state is taken to be $y = (x_1, \cdots, x_{i-1}, u, x_{i+1}, \cdots, x_n)$. If $y \in \mathcal{A}$, then y is accepted as the next state. Otherwise the process stays in state x.

The transition probabilities for the resulting Markov chain are

$$p(x, y) = \frac{1}{n} P(X_i = u \mid X_j = x_j , \text{ for } j \neq i) = \frac{1}{n} \frac{f(y)}{P(X_j = x_j, \text{ for } j \neq i)} ,$$

when x and y belong to $\mathcal{A}$ and differ in only the i-th component. Otherwise, $p(x, y) = 0$, while $p(x, x)$ is determined so that the row sum is one. To verify the balance equations, consider states x and y in $\mathcal{A}$ which differ only in the i-th component, and have common components $x_1, x_2, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n$. Then

$$\pi(x)p(x, y) = \frac{1}{n} \frac{f(x)f(y)}{P(X_j = x_j, \text{ for } j \neq i)P(\mathcal{A})} .$$

8.8 *Unit 2: Markov Chains*

However, because x and y have in common the components $x_1, x_2, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n$, it follows that

$$p(y,x) = \frac{1}{n} \frac{f(x)}{P(X_j = x_j, \text{ for } j \neq i)} .$$

Therefore

$$\pi(y)p(y,x) = \frac{1}{n} \frac{f(y)f(x)}{P(X_j = x_j, \text{ for } j \neq i)P(\mathcal{A})} .$$

Hence the balance equations hold. For the chain to be irreducible, however, the set $\mathcal{A}$ must be connected in the sense that it is possible to travel from any one element in $\mathcal{A}$ to any other by making changes in only one component at a time.

There is another version of Gibbs sampling which is perhaps simpler. Suppose we have n variables, each of which can take on values $\{v_1, v_2, \cdots, v_m\}$. Let $\mathcal{G}$ denote the set of m^n possible n-tuples of values that the n variables can achieve. Let g be a positive function defined on $\mathcal{G}$ and suppose we wish to sample from the distribution

$$\pi(x_1, x_2, \cdots, x_n) = \frac{g(x_1, x_2, \cdots, x_n)}{\sum_{z \in \mathcal{G}} g(z_1, z_2, \cdots, z_n)} . \tag{8.4.1}$$

Starting in state $x = (x_1, x_2, \cdots, x_n)$, the algorithm selects at random an index $i \in \{1, 2, \cdots, n\}$, and the component x_i is then changed to a value u generated from the conditional distribution

$$h_i(z) = \frac{g(x_1, \cdots, x_{i-1}, z, x_{i+1}, \cdots, x_n)}{\sum_{s=1}^{m} g(x_1, \cdots, x_{i-1}, v_s, x_{i+1}, \cdots, x_n)} . \tag{8.4.2}$$

The transition matrix for the Markov chain is therefore

$$p(x,y) = \frac{1}{n} \frac{g(x_1, \cdots, x_{i-1}, u, x_{i+1}, \cdots, x_n)}{\sum_{s=1}^{m} g(x_1, \cdots, x_{i-1}, v_s, x_{i+1}, \cdots, x_n)} ,$$

when $x = (x_1, \cdots, x_n)$, and $y = (x_1, \cdots, x_{i-1}, u, x_{i+1}, \cdots, x_n)$, and $u \neq x_i$. Showing that the balance equations hold follows the same steps shown in the previous version of Gibbs sampling. Also, we see that the chain is irreducible since it is possible to travel from any one n-tuple in $\mathcal{G}$ to another by changing only one component at a time.

The Ising model considered earlier in Example 8.3.2 can be treated using this Gibbs sampling approach. For $\mathcal{G}$ we take the set of $N \times N$ matrices with entries 1 or -1. This set was previously denoted by S in Example 8.3.2. Thus, the variables refer to the N^2 entries in an $N \times N$ matrix, so that $n = N^2$. The v-values are 1 and -1, so that $m = 2$.

Finally, g is taken to be $g(M) = exp(-rE(M))$, for $M \in \mathcal{G}$. This Gibbs sampling approach, however, yields a different Markov chain than the one obtained in Example 8.3.2. Here we get instead the transition probabilities

$$p(M,B) = \frac{1}{N^2}\frac{g(B)}{g(M)+g(B)}, \quad \text{when} \quad M \sim B.$$

where the matrices M and B differ in at most one entry.

8.5 Simulated annealing We consider now the application of Markov chain Monte Carlo methods to optimization. Let S be a finite set, and let f be positive function defined on S. The elements of S are viewed as states of the system. Each state $x \in S$ has specified a set of adjacent or neighbor states, denoted by $S(x)$. We can imagine a graph in which the states are nodes and an edge connects each state to its neighbors. Assume this graph is **strongly connected** in the sense that it is possible to travel from any one state to another, and if $y \in S(x)$ then $x \in S(y)$, for any two states x and y.

We seek a point in S at which f achieves its minimum value. Define a probability distribution on S by

$$\pi(x) = \frac{1}{C}exp(-rf(x)), \quad \text{where} \quad C = \sum_{x\in S} exp(-rf(x)).$$

and where $r > 0$ is a constant. The basic idea is to form a regular Markov chain whose stationary distribution is π.

The Markov chain The Markov chain is traversed as follows. Given that the process is in state x, a state y in $S(x)$ is selected at random. Then, this state y is accepted as the next state with probability $\beta(x,y)$, where the probability $\beta(x,y)$ is determined by the Hastings-Metropolis criterion :

$$\beta(x,y) = min\left\{1, \frac{\pi(y)(1/|S(y)|)}{\pi(x)(1/|S(x)|)}\right\} = min\left\{1, \frac{\pi(y)|S(x)|}{\pi(x)|S(y)|}\right\}.$$

The symbol $|S(x)|$ denotes the number of elements in the set $S(x)$. The one-step transition probabilities for the resulting Markov chain are then

$$p(x,y) = \frac{1}{|S(x)|}\beta(x,y), \quad \text{for } y \in S(x), \quad \text{and} \quad p(x,x) = 1 - \sum_{y\neq x} p(x,y).$$

8.10 *Unit 2: Markov Chains*

Note that in the special case in which each neighborhood set $S(x)$ has the same number of elements, then

$$\beta(x,y) = min\left\{1, \frac{\pi(y)}{\pi(x)}\right\} = min\left\{1, exp(r[f(x) - f(y)])\right\} .$$

The process will then move from state x to y if $f(x) > f(y)$, and otherwise will accept y as the next point with probability

$$exp\left(r[f(x) - f(y)]\right) .$$

Observe that in this case, if the function f is not constant on S, the resulting Markov chain is irreducible and aperiodic.

In general now, assuming that the Markov chain is regular, or equivalently, irreducible and aperiodic, it follows that for large n,

$$P(X_n = x) \approx \pi(x) = \frac{1}{C} exp(-rf(x)) , \text{ where } C = \sum_{x \in S} exp(-rf(x)) .$$

Suppose the function f achieves its minimum value on S, say v, on a set of points $O \subset S$. In other words, $v < f(x)$ for all $x \notin O$, and $v = f(x)$ for all $x \in O$. Then for large n,

$$P(X_n \in O) \approx \frac{\sum_{x \in O} exp(-rf(x))}{C} = \left(1 + \frac{1}{|O|} \sum_{x \notin O} exp\left(r[v - f(x)]\right)\right)^{-1} .$$

But for a large value of $r > 0$, the sum on the right will be small. Therefore, the algorithm has a good chance of converging to the optimal solution in this case.

The optimization algorithm In theory, one would now simulate the Markov chain, and after a suitably large number of steps, accept the current state as the optimal solution. Thus, considering the special case in which all adjacent states have the same number of elements, the algorithm would have the form

```
Initialize: xstart, r
k := 0;
x := xstart;
do until stopping criterion:
  generate: y from S(x)
  if f(y) < f(x), then x:=y
  else
  if rand < exp(r(f(x)-f(y)), then x:=y
  end
end
```

In this code, "rand" denotes a uniform $(0, 1)$ random variable.

In practice however, the implementation takes on a modified form, designed to improve convergence. The modification concerns the parameter r. In principle, a large value of r is desirable, but may slow convergence in the initial stages. Therefore, a sequence of r-values is used. The following ideas seem to work well.

1. Initially, $r = r_0$ is chosen small enough so that the acceptance criterion holds for most of the transitions generated.

2. Then a sequence of r-values is generated, often using a simple method such as $r_{k+1} = \alpha r_k$, for $k = 0, 1, 2, \cdots$, where the factor α is around 1.01 to 1.25. It seems however that these r-values can not increase too rapidly, and that a bound of the form $r_k \leq r\, log(k + 2)$, for constant r, is needed.

3. For each r_k value, the above simulation is carried out for N_k steps. The value of N_k is required to be long enough so that a certain minimum number of accepted transitions is achieved, but it is also bounded above so that extremely long chains are not generated.

4. A stopping criterion is based on the number of consecutive steps for which no improvement in the objective is obtained. Also, an upper bound for the r_k-values is specified.

Using these ideas the code above would then be expanded as follows. Basic Hastings

```
Initialize: xstart, r(0), N(0)
k := 0;
x := xstart;
do until stopping criterion:

  for n = 1 to N(k)
    generate: y from S(x)
    if f(y) < f(x), then x:=y
    else
    if rand < exp(r(f(x)-f(y)), then x:=y
    end
  end

  k := k+1;
  calculate length  N(k)
  calculate control r(k)

end
```

In general, it is difficult to offer more specific rules for the implementation of these special steps. Procedures need to be worked out on a case by case basis.

Example 8.5.1 The traveling salesperson problem Suppose we are given an $n \times n$ matrix $D = (d_{ij})$ of positive numbers, where d_{ij} represents the distance between city i and city j. A **tour** is a closed path that passes through each city exactly once. We can represent a tour as a permutation p of the integers $\{1, 2, \cdots, n\}$, where $p(i)$ is the city visited immediately after city i. There are $(n-1)!$ tours. However, if we are concerned only the length of a tour, then we have only $(n-1)!/2$ tours to consider, since a given tour can be traversed in either of two directions. We shall refer to these tours, when the order of traversal is ignored, as **routes**. The traveling salesperson problem is to find a route that has least distance.

In terms of the formalism above, take the set S to consist of all possible routes. Thus, S consists of $(n-1)!/2$ elements. Each route then is a state of the system. For the neighborhood structure, take $S(x)$ to be the set of routes that can be obtained from the route x by altering exactly 2 edges. The function to be minimized assigns to each route in S its total distance. Thus, f is defined by

$$f(x) = \sum_{i=1}^{n} d_{i,p(i)} \ ,$$

where the route $x \in S$ is defined by the permutation p.

Exercises

8.1 Consider the Markov chain with transition matrix

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

where $0 < p < 1$. (a) Determine the transition matrix of the reverse chain. (b) Is this process time reversible? Explain.

8.2 A key result that was needed in Example 8.2.1 was that if the transition matrix of an irreducible and aperiodic chain is symmetric, then the stationary distribution is the

uniform distribution. Show this result. Help: Show directly that the row vector e with all ones satisfies $e = eP$, where P is the one-step transition matrix.

8.3 Verify the calculations of Example 8.3.1. Also verify that the matrix mentioned at the end of the example fails to work.

8.4 Let $G = (V, E)$ be an undirected, connected graph, with the property that each vertex is connected to at most r vertices. Let f be a positive function defined on V, and let π denote the probability distribution

$$\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)} .$$

If $(x, y) \in E$, define the transition probability

$$p(x, y) = \frac{1}{r} min\left\{1, \frac{f(y)}{f(x)}\right\} ,$$

with $p(x, y) = 0$ otherwise, except that $p(x, x)$ is determined so that the rows sum to one. (i) Show that the Markov chain determined by p is irreducible. (ii) Determine conditions under which the chain is regular. (iii) Show the chain is time reversible with respect to π.

8.5 Suppose $G = (V, E)$ is an undirected, connected graph. For each vertex $v \in V$, let $edge(v)$ denote the number of edges that are connected to v. Let f be a positive function defined on V, and let π denote the probability distribution

$$\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)} .$$

(a) Implement the Hastings-Metropolis method to find a regular Markov chain whose limiting distribution is π. Start with the initial irreducible chain defined by

$$q(x, y) = \frac{1}{edge(x)} , \quad \text{whenever } (x, y) \in E .$$

Note that the Markov chain with this one-step transition matrix is traversed by selecting at random one of the edges connected to x, and then making the transition to the corresponding node. $(a1)$ Show the Markov chain is irreducible. $(a2)$ Determine conditions under which the chain is regular. $(a3)$ Show the chain is time reversible with respect to π. (b) Write a MATLAB program that determines the one-step probability matrix resulting from this method. The input to this program is the function f, and the graph, represented by an adjacency matrix. An adjacency matrix is an $n \times n$ matrix, where n is the number of nodes in the graph, and where entry (i, j) is one if there is an edge connecting nodes i and j, and is zero otherwise. Use this adjacency matrix to compute the function $edge(v)$ at each node. Apply your program to the graph $G = (V, E)$, where $V = \{1, 2, 3, 4\}$, and $E = \{(1, 2), (1, 3), (2, 3), (2, 4), (3, 4)\}$,

and where $f(1) = 2$, $f(2) = 8$, $f(3) = 6$, and $f(4) = 4$. Verify (using MATLAB) that the resulting chain is regular and has the required limiting state probability distribution.

8.6 Verify that the Markov chain constructed in Example 8.3.2 of the Ising model is irreducible and that the balance equations hold with respect to the stationary distribution π_r.

8.7 Verify that the function $h(z)$ given in equation (8.4.2) is a conditional probability, and in fact that

$$h(z) = P(X_i = z \mid X_1 = x_1, \cdots, X_{i-1} = x_{i-1}, X_{i+1} = x_{i+1} \cdots, X_n = x_n) \ .$$

where X_1, X_2, $\cdots$, X_n denote the n variables, and the underlying probability function on $\mathcal{G}$ is π, as given by (8.4.1).

8.8 Show that any two-state regular Markov chain is time reversible with respect to its stationary distribution.

8.9 Referring to the method of simulated annealing, verify that if f is not a constant function, and each neighborhood set $S(x)$ has the same number of elements, then the Markov chain is irreducible and aperiodic. In other words, the chain is regular in this case.

8.10 For the traveling salesperson problem, verify that for a given route x, if two edges are removed, and they do not have node in common, then only one new route can be formed. How many new routes can be formed if the two removed edges are adjacent; that is, they have a node in common. For a given route x, how many elements are in the neighborhood structrue $S(x)$?

Unit 3

Continuous-Time Markov Chains

The notion of a Markov chain is extended in this unit to the case in which time is treated as a continuous variable. This extension permits us to model a number of discrete-event systems. In particular, we will consider waiting line models.

Chapter 9

The Poisson Process

9.1 Introduction A stochastic process $\{N(t) \mid t \geq 0\}$ which counts the number of events that have occurred up to time t is called a **counting process**. For example, a counting process might count the number of customers that arrive at a service facility, or the number of machine break-downs at a repair shop. Observe that for times $s < t$, the random variable $N(t) - N(s)$ is equal to the number of events that have occurred in the interval $(s, t]$. A counting process is said to have **independent increments** if the number of events in disjoint intervals of time are independent random variables. The process is said to have **stationary increments** if the probability distribution of the number of events that occur in an interval of time depends only on the length of the interval. Thus, the probability distribution of $N(s+t) - N(s)$ depends only on t.

9.2 The Poisson process A counting process $\{N(t) \mid t \geq 0\}$ is said to be a Poisson process with rate constant $\lambda > 0$ if (a) $N(0) = 0$, (b) the process has independent increments, and (c) the probability distribution of the number of events in an interval of length t is Poisson with parameter λt. In other words, for $s, t \geq 0$,

$$P\big(N(s+t) - N(s) = n\big) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \quad n = 0, 1, 2, \cdots.$$

Note that $E(N(t)) = \lambda t$. The parameter λ can be interpreted as the average rate per unit time, at which events occur.

There is an equivalent definition of the Poisson process that can be helpful when property (c) is not evident. Under this definition, a counting process $\{N(t) \mid t \geq 0\}$ is said to be a Poisson process with rate constant $\lambda > 0$ if (a) $N(0) = 0$, (b) the process has independent and stationary increments, (c) $P(N(h) = 1) = \lambda h + o(h)$ as $h \to 0^+$ and (d) $P(N(h) \geq 2) = o(h)$ as $h \to 0^+$. To see that this definition implies the previous one, suppose the interval $(0, t]$ is divided into a large number of intervals, say m, each of

9.2 *Unit 3: Continuous-Time Markov Chains*

length $h > 0$. Thus $mh = t$, and we have

$$P(N(t) = n) = \lim_{h \downarrow 0} \binom{m}{n} (\lambda h + o(h))^n (1 - \lambda h + o(h))^{m-n} .$$

It follows that

$$P(N(t) = n) = \lim_{h \to 0} \frac{m(m-1)\cdots(m-n+1)}{n!} (\lambda h + o(h))^n (1 - \lambda h + o(h))^{m-n}$$

$$= \lim_{h \downarrow 0} \frac{[mh][(m-1)h]\cdots[(m-n+1)h]}{n!} \frac{(\lambda + \mathrm{O}(h))^n}{(1 - \lambda h + o(h))^n} (1 - \lambda h + o(h))^m .$$

Using $t = mh$, this expression is approximately, for small $h > 0$,

$$\frac{t^n}{n!} \lambda^n (1 - \lambda h)^{t/h} , \quad \text{which converges to} \quad \frac{(\lambda t)^n}{n!} e^{-\lambda t} , \quad \text{as } h \to 0^+ .$$

9.3 Interarrival times and waiting times For a Poisson process, let T_i denote the time between the occurrence of the $(i-1)$-st and the i-th events. Then we have

$$P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t} .$$

Thus, T_1 follows the exponential distribution with mean $1/\lambda$. Next, by conditioning on T_1, we can write

$$P(T_2 > t) = \int_0^\infty P(T_2 > t \mid T_1 = s) (\lambda e^{-\lambda s}) \, ds .$$

But the term $P(T_2 > t \mid T_1 = s)$ is the probability that no events occur in the interval of time $(s, s+t)$, and is thus equal to $e^{-\lambda t}$. It follows that T_2 is independent of T_1, and performing the integration gives us $P(T_2 > t) = e^{-\lambda t}$, so that T_2 is seen to follow the exponential distribution with mean $1/\lambda$. Continuing by induction, suppose that T_1, T_2, $\cdots$, T_n are independent and each follow the exponential distribution with mean $1/\lambda$. By conditioning, we can write $P(T_{n+1} > t)$ as

$$\int_0^\infty \cdots \int_0^\infty P(T_{n+1} > t \mid T_1 = s_1, T_2 = s_2, \cdots, T_n = s_n) \, f_n(s) \, ds_1 \cdots ds_n ,$$

where f_n denotes the joint probability density of $\{T_1, T_2, \cdots, T_n\}$. Using the induction hypothesis, this joint density is

$$f_n(s_1, s_2, \cdots, s_n) = \lambda^n e^{-\lambda s} , \quad \text{where} \quad s = s_1 + s_2 + \cdots + s_n .$$

Further, the term $P(T_{n+1} > t \mid T_1 = s_1, T_2 = s_2, \cdots, T_n = s_n)$ is the probability that no events occur in the interval of time $(s, s+t)$, and is thus equal to $e^{-\lambda t}$. It follows that T_{n+1} is independent of $T_1, T_2, \cdots, T_n$, and performing the integration gives us $P(T_{n+1} > t) = e^{-\lambda t}$, so that T_{n+1} is seen to follow the exponential distribution with mean $1/\lambda$, completing the induction argument.

Consider next the waiting time until the occurrence of the n-th event, which is $S_n = T_1 + T_2 + \cdots + T_n$. To find its probability distribution, note that the event $\{S_n \leq t\}$ occurs if and only if $\{N(t) \geq n\}$. Thus, denoting by G_n and g_n the cdf and density of S_n respectively, we have

$$G_n(t) = P(S_n \leq t) = P(N(t) \geq n) = \sum_{\nu=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^\nu}{\nu!} .$$

Since $g_n(t) = G_n'(t)$, we find by differentiating this expression with respect to t, and doing some algebra, that

$$g_n(t) = \frac{\lambda}{(n-1)!} (\lambda t)^{n-1} e^{-\lambda t} .$$

Thus, S_n follows the gamma distribution with parameters $\alpha = n$ and $\beta = 1/\lambda$.

9.4 The $G/M/1$ Queue Consider the single server queue in which customers arrive and are served immediately if the server is idle, but otherwise they join the end of a single line to wait for service. When a service is completed, the next waiting customer is served. It will be assumed that service times of customers are independent and identically distributed random variables with common exponential density $\mu e^{-\mu t}$ for $t > 0$, where μ is a fixed parameter. It will be assumed further that customer arrive according to a general renewal process, which means that the times between successive arrivals are viewed as independent and identically distributed random variables with common density $f(t)$.

Let X_n denote the number of customers in the system after the n-th arrival, and let W_n denote the number of customers served during the time between the arrival of the n-th and the $(n+1)$-st customers. Then the sequence $\{X_n \mid n = 1, 2, \cdots\}$ is a stochastic process with state space $\{1, 2, \cdots\}$. To see that it is also a Markov chain, note first that for each $n \geq 1$,

$$X_{n+1} = X_n + 1 - W_n . \tag{9.4.1}$$

Let A denote the interarrival time between two successive customers (so, as indicated above, A is a random variable with density f). Given $A = t$, and $X_n = i$, and given any

history of the states previous to time n, it follows from the memoryless property of the exponential distribution that the probability distribution of W_n depends only on the values of t and i. Indeed, consider the event $\{W_n = k\}$, conditioned on this information. When $k < i$, this event is equivalent to the occurrence of k Poisson events (with parameter μ) in the interval $(0, t)$, and when $k = i$, this event is equivalent to the occurrence of at least i Poisson events in the interval $(0, t)$. For $k > i$, the event $\{W_n = k\}$ is not possible. Thus, the probability $P(W_n = k \mid A = t, X_n = i, X_{n-1}, \cdots, X_0)$ is given by

$$e^{-\mu t}\frac{(\mu t)^k}{k!}, \quad \text{for} \quad k < i, \quad \text{and} \quad \sum_{\nu=i}^{\infty} e^{-\mu t}\frac{(\mu t)^\nu}{\nu!}, \quad \text{for} \quad k = i,$$

and equals 0 otherwise. Thus, conditioning on A, we can write for $k < i$

$$P(W_n = k \mid X_n = i) = \int_0^\infty e^{-\mu t}\frac{(\mu t)^k}{k!} f(t)\, dt,$$

while

$$P(W_n = i \mid X_n = i) = \sum_{\nu=i}^{\infty} \int_0^\infty e^{-\mu t}\frac{(\mu t)^\nu}{\nu!} f(t)\, dt,$$

and $P(W_n = k \mid X_n = i) = 0$ for $k > i$. It follows from this reasoning and from formula (9.4.1), that $\{X_n\}$ is a Markov chain. The transition probabilities are

$$p_{i1} = P(X_{n+1} = 1 \mid X_n = i) = P(W_n = i \mid X_n = i) = \sum_{\nu=i}^{\infty} \int_0^\infty e^{-\mu t}\frac{(\mu t)^\nu}{\nu!} f(t)\, dt$$

and for $1 < j \leq i + 1$,

$$p_{ij} = P(X_{n+1} = j \mid X_n = i)$$

$$= P(W_n = i + 1 - j \mid X_n = i) = \int_0^\infty e^{-\mu t}\frac{(\mu t)^{i+1-j}}{(i+1-j)!} f(t)\, dt,$$

with $p_{ij} = 0$ otherwise.

Exercises

9.1 (a) Suppose the waiting time for an event follows that exponential distribution with parameter λ. Show that if the event has not yet occurred up to some time $t > 0$, then the probability that is does occur in the next time interval of length h is asymptotically

equal to $\lambda h + o(h)$ as $h \to 0^+$. (b) Suppose an event E occurs randomly according to the following properties: (i) There is a constant λ such that the probability E occurs during a time interval of length h is $\lambda h + o(h)$, (ii) the probability that E occurs two or more times during an interval of length h is $o(h)$, and (iii) the occurrences of E during disjoint time intervals are independent events. Let T be the waiting time for this event E to occur. Show that T follows the exponential distribution with parameter λ.

9.2 Consider two independent Poisson processes, type 1 having rate λ_1, and type 2 having rate λ_2. Show that the combined process, where events are observed without regard to type, is again Poisson and that the rate is $\lambda_1 + \lambda_2$.

9.3 Consider a Poisson process such that each time an event occurs, it is of type 1 with probability $p > 0$, or it is type 2 with probability $q = 1 - p > 0$. Assume these two types appear independently from event to event. Show that these two processes each are Poisson processes and that they are independent of each other. Help: Let $N_1(t)$ and $N_2(t)$ be the counting processes for each type of event. Find the joint distribution by conditioning on $N(t)$, where $N(t) = N_1(t) + N_2(t)$.

9.4 Let $N(t)$ be a Poisson counting process with rate constant λ. Let X_i be independent and identically distributed random variables, with common mean μ. Fix times t and s, with $t > s$, and define the random variable

$$Y = \prod_{i=N(s)+1}^{N(t)} X_i \, .$$

Assume the product equals one if $N(s) = N(t)$. Show that $E(Y) = e^{-\lambda(1-\mu)(t-s)}$.

9.5 A vehicle-controlled traffic light will stay green for τ seconds after a car passes through the intersection, and if no cars arrive during a period of time τ, then it turns red. Let X denote the number of cars that pass through the intersection following dissapation of the initial queue and until the light next turns red. Assume cars arrive according to a Poisson process with rate λ. Find the probability distribution of X and find the expected value of X. Help: Note that the event $\{X \geq n\}$ occurs if the interarrival times of the next n arriving cars are each no more than τ.

9.6 An investor must decide on just one investment to make during a fixed period of time T. The opportunities are of two types : those of profit L and those of profit H, where $L < H$. Opportunities appear at random according to a Poisson process of rate λ. Each time an opportunity appears it is worth L with probability p, and worth H with independent probability $q = 1 - p$. Consider the following strategy : For a given time $\tau < T$, we invest only if an opportunity of profit H occurs. After this time, we invest in the next opportunity that appears, if any.

9.6 *Unit 3: Continuous-Time Markov Chains*

(a) Show that the expected profit is

$$G(\tau) = H(1 - e^{-\lambda q \tau}) + (pL + qH)(e^{-\lambda \tau} - e^{-\lambda T}) .$$

(b) Find the optimal value of τ. Further, what is the probability under the optimal rule that the profit is zero. How might we modify τ_{opt} so as to increase the probability of some profit.

9.7 Consider the $G/M/1$ model.

(a) Is the chain irreducible? Is the chain aperiodic? Help: Write the transition probability matrix with a $+$, say, for each entry that is positive, and a 0 otherwise.

(b) Find the transition probabilities when the interarrival time distribution is exponential with parameter λ.

(c) Under the assumption of part (b), find the stationary distribution?

Chapter 10

Continuous-Time Markov Chains

here, once frog jumps out of state it has to go to a different state

10.1 Introduction A stochastic process $\{X(t) \mid t \geq 0\}$ is called a **continuous-time Markov chain** if the state space is the nonnegative integers, and if the process has the property that

we still have a discrete # of states

$$P(X(s+t) = j \mid X(s) = i,\ X(u) \text{ specified for any } u < s)$$

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

The process is said to be **homogeneous** or **stationary** if the transition probabilities $P(X(s+t) = j \mid X(s) = i)$ depend only on t. In this case, define

$$p_{ij}(t) = P(X(s+t) = j \mid X(s) = i)\,.$$

$P_{ij}^{(n)} \equiv P_{ij}(t)$

A continuous-time stochastic process, with discrete state space, can be visualized as a process that moves from state to state over time, and once having entered a particular state, it stays there for a random amount of time, and then moves to another state. For a stationary continuous-time Markov process, however, the length of time spent in a state must follow the exponential distribution. To see this property, let T_i be a random variable denoting the amount of time the process stays in state i before making a transition to another state. Consider the probability $P(T_i > s+t \mid T_i > s)$, in which it is given that the process has been in state i for a period of time s. By the Markov property and stationarity, the possible evolution of the process after this time, depends only on this state i, and not how long the process has been there. Thus, the probability that the process stays in state i for another t units of time is the same as if the process had just entered state i. In other words, $P(T_i > s+t \mid T_i > s) = P(T_i > t)$. Therefore, T_i has the memoryless property, and hence follows the exponential distribution. Using similar reasoning, one can show that the time the process spends in a given state, and the time it spends in the next state, are independent random variables.

time frog remains in same state is exp. distributed

10.2 *Unit 3: Continuous-Time Markov Chains*

To determine the probability distributions that govern jumps to the next state, we will assume that the transition probabilities can be written

$$p_{ii}(h) = 1 - v_i h + o(h)\,, \quad \text{and for } j \neq i, \quad p_{ij}(h) = q_{ij} h + o(h)\,, \text{ as } h \to 0^+\,,$$

for some constants v_i and q_{ij} for $i \neq j$. A stationary continuous-time Markov chain satisfying these assumptions is sometimes called a **pure jump process**. The first of these equations implies that once the process enters state i, it will stay in that state for a length of time that is exponentially distributed with rate v_i. At the end of that time, the process jumps to a different state j, according to certain jump probabilities, say Q_{ij}, for $j \neq i$. These jump probabilities are specified by the requirement that asymptotically as $h \to 0$,

$$P(X(t) = j \mid X(t) \neq i,\, X(t-h) = i) = Q_{ij} + O(h)\,.$$

For $j \neq i$,

$$P(X(t) = j \mid X(t) \neq i,\, X(t-h) = i) = \frac{P(X(t) = j,\, X(t) \neq i,\, X(t-h) = i)}{P(X(t) \neq i,\, X(t-h) = i)}$$

$$= \frac{P(X(t) = j,\, X(t-h) = i)}{P(X(t) \neq i,\, X(t-h) = i)}$$

$$= \frac{P(X(t) = j \mid X(t-h) = i)\, P(X(t-h) = i)}{P(X(t) \neq i \mid X(t-h) = i)\, P(X(t-h) = i)}$$

$$= \frac{P(X(t) = j \mid X(t-h) = i)}{P(X(t) \neq i \mid X(t-h) = i)}$$

$$= \frac{q_{ij}h + o(h)}{1 - p_{ii}(h)} = \frac{q_{ij}h + o(h)}{v_i h + o(h)} \;\to\; \frac{q_{ij}}{v_i}\,, \text{ as } h \to 0^+$$

Thus, the jump transition probabilities are given by

$$Q_{ij} = \frac{q_{ij}}{v_i}\,, \quad \text{for } j \neq i\,.$$

Another way to obtain these transition probabilities is to imagine a large number of entities which move through the states according to the stochastic process. During a small time interval of length h, of those entities that now are in state i, the fraction $1 - p_{ii}(h)$ will leave the state i, and of these, the fraction Q_{ij} will go to state j. Thus, the fraction leaving state i and entering state j is $Q_{ij}(1 - p_{ii}(h))$. However, this fraction is also equal

to $p_{ij}(h)$. Thus, for small h, we have approximately $Q_{ij}(1 - p_{ii}(h)) \approx p_{ij}(h)$, and in the limit as $h \to 0$, we get $Q_{ij}v_i = q_{ij}$, or $Q_{ij} = q_{ij}/v_i$.

10.2 A differential equation for the pure jump process Consider a stationary Markov process. During a time interval of length h, we can write

$$p_{ij}(t+h) = p_{ij}(t)p_{jj}(h) + \sum_{\nu \neq j} p_{i\nu}(t)p_{\nu j}(h) \ .$$

For the pure jump process, this equation becomes

$$p_{ij}(t+h) = p_{ij}(t)(1 - v_j h) + \sum_{\nu \neq j} p_{i\nu}(t)(q_{\nu j}h) + o(h) \ .$$

The quantities lumped into the $o(h)$ term depend in general on ν, however, we will assume the resulting sums are such that the limits, taken below, as $h \to 0$ are allowable. Such limits would be allowable, for example, if the state space is finite. Now, dividing through by h, we get

$$\frac{p_{ij}(t+h) - p_{ij}(t)}{h} = -p_{ij}(t)v_j + \sum_{\nu \neq j} p_{i\nu}(t)q_{\nu j} + o(h)/h \ ,$$

and taking the limit as $h \to 0$ gives us

$$\frac{d}{dt}p_{ij}(t) = -p_{ij}(t)v_j + \sum_{\nu \neq j} p_{i\nu}(t)q_{\nu j} \ .$$

system of first order linear differential eqn

Let $P(t)$ be the matrix whose (i,j)-th entry is $p_{ij}(t)$, and let B be the matrix whose (i,j)-th entry is q_{ij} for $i \neq j$, and along the diagonal are $-v_1, -v_2, \cdots$. In matrix form, the last equation is

$$P'(t) = P(t)B \ .$$

For $t = 0$, we have $P(0) = I$. When the state space is finite, we can write the solution of this initial value problem as

$$P(t) = e^{tB} \ .$$

In principle, the analysis of the long-run behavior of the process can be done using this solution. However, the analysis involves the eigenvalues and eigenvectors of B.

Suppose now that $z(t)$ is the row vector representing the probability distribution of the states at time t. Thus, the i-th component of $z(t)$ is the probability

10.4 *Unit 3: Continuous-Time Markov Chains*

$z_i(t) = P(X(t) = i)$. In particular, $z(0)$ is the row vector representing the initial probability distribution of the states. Then $z(t) = z(0)P(t)$, and so

$$z'(t) = z(0)P'(t) = z(0)P(t)B = z(t)B \ .$$

Suppose that as $t \to \infty$, $z(t) \to \pi$ and $P'(t) \to 0$. Then $\pi B = 0$, and so π is a left eigenvector of B with eigenvalue 0. Writing out the equations $\pi B = 0$ componentwise, we find that

$$v_j \pi_j = \sum_{\nu \neq j} q_{\nu j} \pi_\nu \ .$$

Note that these equations could be interpreted as balance equations, in the sense that the term on the left represents flow of the process out of state j, and the term on the right represents flow into state j from the other states.

10.3 The birth-and-death process An important example of a pure jump process is the birth-and-death process. A stationary continuous-time Markov chain is called a **birth-and-death process** if for $i \geq 1$: $p_{i,i+1}(h) = \lambda_i h + o(h)$, $p_{i,i-1}(h) = \mu_i h + o(h)$, and $p_{i,i}(h) = 1 - (\lambda_i + \mu_i)h + o(h)$, and for $i = 0$: $p_{0,1}(h) = \lambda_0 h + o(h)$, and $p_{0,0}(h) = 1 - \lambda_0 h + o(h)$. For all other (i, j), $p_{ij}(h) = o(h)$. The parameters $\{\lambda_i \mid i \geq 0\}$ are called the **arrival rates**, and the parameters $\{\mu_i \mid i \geq 1\}$ are called the **departure rates**. Thus, when a birth-and-death process enters state $i \geq 1$, it will stay in that state for a period of time that is exponentially distributed with rate $\lambda_i + \mu_i$. At the end of that period of time, it will jump to state $i + 1$ with probability $Q_{i,i+1} = \lambda_i/(\lambda_i + \mu_i)$, or it will jump to state $i - 1$ with probability $Q_{i,i-1} = \mu_i/(\lambda_i + \mu_i)$. When the process enters state 0, it will stay there for a period of time that is exponentially distributed with rate λ_0, and then will necessarily jump to state 1, as $Q_{0,1} = 1$.

Example 10.3.1 – The $M/M/s$ queue. Consider a queueing model with $s \geq 1$ servers, Poisson arrivals at rate λ, independent and exponentially distributed services times with rate μ, and a single line for which arriving customers join the end of the line and the first customer in line is served by the next available server. Let $X(t)$ denote the number of customers in the system at time t. Then we have a birth-and-death model. For $i \geq s$,

$$p_{i,i+1}(h) = \big(\lambda h + o(h)\big)\big(1 - \mu h + o(h)\big)^s + o(h) = \lambda h + o(h) \ , \text{ and}$$

$$p_{i,i-1}(h) = \big(1 - \lambda h + o(h)\big)\big(s\mu h + o(h)\big) + o(h) = s\mu h \ + o(h) \ .$$

For $1 \leq i < s$,

$$p_{i,i+1}(h) = \big(\lambda h + o(h)\big)\big(1 - \mu h + o(h)\big)^i + o(h) = \lambda h + o(h) \ , \quad \text{and}$$

$$p_{i,i-1}(h) = \big(1 - \lambda h + \ o(h)\big)\big(i\mu h + o(h)\big) + o(h) = i\mu h + o(h)$$

Finally, for $i = 0$, $p_{0,1}(h) = \lambda h + o(h)$. Thus, the departure rates are $\mu_i = s\mu$ for $i \geq s$, and $\mu_i = i\mu$ for $1 \leq i < s$, while the arrival rates are $\lambda_i = \lambda$ for $i \geq 0$.

Balance equations for the birth-and-death process For a birth and death process in which arrivals occur at a slower rate than departures, one might expect that a steady-state probability distribution for the states would exist. If such is the case, denote this long-run probability distribution by $\{\pi_i \mid i = 0, 1, 2, \cdots\}$. We interpret π_i as the fraction of time in the long-run that the process spends in state i. These probabilities can be calculated by appealing to the balance equations:

$$\lambda_i \pi_i = \mu_{i+1} \pi_{i+1} \ , \quad \text{for each state } \ i = 0, 1, 2, \cdots \ .$$

These balance equations can be derived from the balance equations developed earlier in Section 10.2. Alternatively, a justification for these equations is seen by starting with state 0, and reasoning for each successive pair of states, that over a long period of time T, the average number of transitions from state i to state $i+1$ is $\lambda_i(\pi_i T)$, while the average number of transitions from state $i+1$ to state i is $\mu_{i+1}(\pi_{i+1} T)$. These flows must be equal in the steady-state situation.

From the balance equations we get

$$\pi_i = \frac{\lambda_{i-1}\lambda_{i-2}\cdots\lambda_0}{\mu_i \, \mu_{i-1} \cdots \mu_1}\pi_0 \ , \quad \text{for } i = 1, 2, \cdots \ ,$$

and using $\pi_0 + \pi_1 + \cdots = 1$, allows us to solve for all the π_i .

Example 10.3.2 – The $M/M/s$ queue. Recalling Example 10.3.1, the departure rates are $\mu_i = s\mu$ for $i \geq s$, and $\mu_i = i\mu$ for $1 \leq i < s$, while the arrival rates are $\lambda_i = \lambda$ for $i \geq 0$. It follows that

$$\pi_i = \frac{(\lambda/\mu)^i}{i!}\pi_0 \ , \quad \text{for } \ i = 1, 2, \cdots, s-1, \quad \text{and} \quad \pi_i = \frac{(\lambda/\mu)^i}{s! s^{i-s}} \, \pi_0 \ , \quad \text{for } i \geq s \ ,$$

and provided $\rho = \lambda/s\mu < 1$,

$$\frac{1}{\pi_0} = \sum_{i=0}^{s-1} \frac{(\lambda/\mu)^i}{i!} + \frac{(\lambda/\mu)^s}{s!} \frac{1}{1-(\lambda/s\mu)} .$$

10.4 First entrance times for a birth-and-death process Let T_{ij} denote the first entrance time into state j, after the process has just entered state i. For a birth-and-death process, consider the calculation of the expected value of $T_{i,i+1}$. We shall find this expected value by conditioning on the next state. Thus, let S denote the event that the next state is $i+1$. Then

$$E(T_{i,i+1}) = E(T_{i,i+1} \mid S)P(S) + E(T_{i,i+1} \mid \overline{S})P(\overline{S}) .$$

Note that $\overline{S}$ is the event that the next state is $i-1$, since it is understood that a jump has been made. Using the jump probabilities for the birth- and-death process, we have

$$E(T_{i,i+1}) = E(T_{i,i+1} \mid S)\frac{\lambda_i}{\lambda_i+\mu_i} + E(T_{i,i+1} \mid \overline{S})\frac{\mu_i}{\lambda_i+\mu_i} ,$$

where we set $\mu_0 = 0$. Now, $E(T_{i,i+1} \mid S) = 1/(\lambda_i+\mu_i)$, because the expected time to reach state $i+1$, given that the next state is $i+1$, is simply the expected time spent in state i, which is $1/(\lambda_i+\mu_i)$. Next,

$$E(T_{i,i+1} \mid \overline{S}) = \frac{1}{\lambda_i+\mu_i} + [E(T_{i-1,i}) + E(T_{i,i+1})] ,$$

for if the next state is $i-1$, then the process must spend the expected time in state i, which is $1/(\lambda_i+\mu_i)$, then spend $E(T_{i-1,i})$ to get back to state i, and then spend time $E(T_{i,i+1})$ to get from state i to state $i+1$. Substituting these results, and then solving for $E(T_{i,i+1})$, yields

$$E(T_{i,i+1}) = \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i} E(T_{i-1,i}) .$$

Starting with boundary condition $E(T_{0,1}) = 1/\lambda_0$, the remaining expected first entrance times are determined. Having found these times, it follows that for $j > i$,

$$E(T_{ij}) = E(T_{i,i+1}) + E(T_{i+1,i+2}) + \cdots + E(T_{j-1,j}) .$$

In a similar fashion, the expected first entrance times $E(T_{ij})$ for $j < i$ can be found. However, in the case $j < i$, it is generally best to determine the boundary conditions separately in each application.

Exercises

10.1 Find the expected first entrance times $T_{i,i+1}$, $i \geq 0$, for the birth-and-death process in which $\lambda_i = \lambda$ for each $i \geq 0$, and $\mu_i = \mu$ for each $i \geq 1$; that is, the $m/m/1$ queue.

10.2 Find the recurrence formula for the expected values of the first entrance times $T_{i,i-1}$ for the birth-and-death process. Remark: To use such a recurrence formula to find the expected values $E(T_{i,i-1})$, it is necessary to specify additional conditions. Generally, however, these additional conditions can only be determined on a case by case basis.

10.3 For the $M/M/1$ queue, find the expected first entrance time $E(T_{1,0})$. This time can be interpreted as the busy time for the queue. Assume $\lambda < \mu$. For the $M/M/3$ queue, find the expected values $E(T_{i,i-1})$ for all $i \geq 1$. Help: Use the recurrence formula obtained in previous problem above. Argue that for $i \geq s$, where s is the number of servers, all the values $E(T_{i,i-1})$ are the same.

10.4 Consider a set of m machines that are in continuous operation, and which fail independently of each other at an exponential rate λ. Assume there are s, $1 \leq s \leq m$, repair persons that service the machines independently and each at exponential rate μ. Let $X(t)$ denote the number of machines at time t that are *not* operational; that is, they are in the repair shop. Determine the arrival and departure rates for the birth-and-death model.

10.5 Consider a set of $m+n$ machines which fail independently of each other at an exponential rate λ. It is intended that m machines are to be in operation at any time. The remaining machines serve as spares and are called into operation when an operating machine fails. If more than n machine are in a state of failure, then all the operational machines will be in service. Suppose there are s, $1 \leq s \leq m$, repair persons that service the machines independently and each at exponential rate μ. Let $X(t)$ denote the number of machines at time t that are *not* operational; that is, they are in the repair shop. Determine the arrival and departure rates for the birth-and-death model.

10.6 Consider a sign that contains N light bulbs, each with a lifetime that follows an exponential distribution with parameter λ. Assume that the bulbs function independently of each other. Suppose it is the policy to allow bulbs to burn out until the moment the r-th bulb expires, and to then replace all burned out bulbs at that time. Define the state of the system $X(t)$ to be the number of burned out bulbs at time t. Argue that this stochastic process (a) has the Markov property, (b) is stationary, and (c) can be represented as a pure jump process. Determine the parameters v_i and q_{ij} of the jump process. (d) Determine the balance equations for this system, and find the long-run probability distribution for the states.

10.8 *Unit 3: Continuous-Time Markov Chains*

10.7 For the previous light bulb problem : (a) What is the expected time between replacements? (b) Suppose for each replacement, it costs $\alpha + \beta r$ dollars to replace the r bulbs, where α and β are constants. Develop a formula that would be a reasonable estimate of the expected cost per unit time.

10.8 Verify that the matrix B in Section 10.2 is such that the sum of the elements in any row is zero.

10.9 Consider a two-state continuous-time Markov chain in which the expected time in state 1 is a, and the expected time in state 2 is b. (a) Determine the parameters of the associated pure jump process. (b) Determine the matrix B in the differential equations for the transition probabilities. Show that this matrix has two eigenvalues, one of which is zero, and the other is negative. (c) Thus conclude that an asymptotic or steady state distribution exists, and find the limiting probability distribution.

10.10 Verify the calculations in Example 10.3.1.

10.11 For the $M/M/1$ queue, find the long-run probability distribution for the states. Use these probabilities to find $E(X)$, the expected number of customers in the system in the long-run. Help: Let $\rho = \lambda/\mu$. Then $E(X) = \rho/(1-\rho)$.

10.12 For the $M/M/2$ queue, find the long-run probability distribution for the states. Use these probabilities to find $E(X)$, the expected number of customers in the system in the long-run. Help: Let $\rho = \lambda/2\mu$. Then $E(X) = 2\rho/(1-\rho^2)$.

10.13 Compare the following queueing system in terms of the expected number in the system. (1) The $M/M/1$ queue with poisson arrivals at rate λ, and service rate μ, (2) A system consisting of two servers, each with service rate $\mu/2$, where the arrival rate is λ, however there is a separate line for each server, and customers pick a line at random, (3) The $M/M/2$ queue with poisson arrivals at rate λ, and service rate $\mu/2$. For each model, assume $\lambda < \mu$.

10.14 A military repair facility has a permanent staff of c repair-persons. However, if more than c jobs are in the system, then additional staff is brought in to handle the extra work. Thus, the facility can be modeled as if there were an infinite number of servers. Assume that the amount of time for a server to complete a job follows the exponential distribution with mean of 4 days, and that jobs arrive according to a Poisson process at rate of 1 every 2 days. Let $X(t)$ denote the number of jobs in the facility at time t. (a) Find the long-run state probability distribution. (b) Find a closed form expression for the long-run probability that more than c jobs are in the system (and thus extra staff is needed).

10.15 Consider the $M/M/1$ queue model, with the modification that when a customer finishes being served, there is a probability p that the customer will be returned to the queue to repeat the required service. (a) Assume this process follows a birth-and-death

model, and find the arrival and departure rates. (*b*) Find the long-run state probability distribution.

10.16 For the $M/M/s$ queueing model, assume that when a customer arrives and finds servers idle, the customer selects a server at random from those available. Consider a particular server, say server number 1. Show that the probability, in the long-run, that this server is idle is $1-\rho$, where $\rho = \lambda/s\mu$. Help: Condition on the number in the system in the long-run.

10.17 Write a simulation model of following queueing network. The customer requests a server for work which lasts a random amount of time following the gamma distribution with mean μ and standard deviation σ. Upon completion of service, the customer leaves the system with probability p, or returns to repeat the service with probability q.

Unit 4

Simulation Modeling

Chapters 11 and 12 of this unit introduce the idea of simulation modeling of stochastic systems. This approach is widely used today as many practical problems are far too complex to yield to mathematical analysis. Earlier chapters presented some of the mathematical techniques that are used to model stochastic systems. However, mathematical methods, at least those developed to date, often require stringent assumptions that may limit their applicability. The simulation approach is usually not limited by restrictive assumptions, but on the other hand, may not offer the power and insight generally found in an analytical model.

Developing and implementing a simulation model hinges on the application of a number statistical methods. First, the random variables that appear in the model must be identified and their distributions determined. Next, the simulation program requires the generation of random samples from these distributions. Finally, statistical analysis of the model outputs is needed to draw conclusions properly. Chapters 13-15 provide an introduction to these topics.

Chapter 11

Static Simulation Models

11.1 Introduction The term static simulation refers to a simulation model which does not involve a natural or explicit notion of time. Typically, however, static models do involve the repetition of an action or experiment, and one can associate with each repetition the passage of one unit of time. Indeed, static models in Extend have this feature, as the examples in Section 11.2 indicate. This chapter presents several examples of static simulation models, using both MATLAB and Extend.

11.2 Static simulation models We begin with a few simple examples of static simulation models. The models are presented using MATLAB.

Example 11.2.1 A coin comes up heads with probability p. This program simulates the repeated tossing of the coin until a specified number of heads, called *numheads*, appear in succession. Each toss is independent of the others. The variable *record* keeps track of the current number of heads that have appeared in a row. The outputs are the mean and standard deviation of the random variable n, which equals the number of tosses until the required streak of heads appears. In the code, the function *rand* generates a random value from the uniform distribution on $(0, 1)$.

```
% Simulates the trials until a specified number
% of heads appear in a row.
% Inputs: numheads (specified number of heads),
% p (probability of heads on a toss), sample (sample size)

meansum = 0; squaresum = 0;
for i=1:sample
  record = 0; k = 0;
  while record < numheads
    if rand < p
      record = record+1;
    else
      record = 0;
    end
    k = k+1;
```

11.2 *Unit 4: Simulation Modeling*

```
  end
  meansum = meansum + k; squaresum = squaresum + k*k;
end
mean = meansum/sample
var = (squaresum - (meansum*meansum/sample))/(sample-1)
std = sqrt(var)
```

Example 11.2.2 The program below represents a highly simplified model of a neutron passing through a shield of thickness d. The shield is assumed infinite in the vertical or y-direction, and is assumed to extend a distance d in the x-direction. A neutron enters perpendicular to the shield, and after colliding with an atom in the shield, will be reflected at an angle *theta*. The neutron then travels another unit length. It is assumed that after 6 collisions, the neutron has lost all its energy. The reflection angle *theta* is assumed to be uniformly distributed. However, since the shield is infinite in the y-direction, we can take this reflection angle to be uniform over the interval $(0, \pi)$, by reasons of symmetry.

```
% Neutron shield model
% Input: d (thickness of shield), sample (sample size)

count = 0;
for i=1:sample
  n = 1; x = 1;
  while (x>0 & x<d & n<7)
    theta = pi*rand;
    x = x + cos(theta);
    n = n+1;
  end
  if x>d
    count = count+1;
  end
end
fraction_thru = count/sample
```

Example 11.2.3 A certain component in a system (a battery for example) has a lifetime which is viewed as a random variable. The program below estimates the average number of times the component has to be replaced during a specified period of time, called *period*. The lifetime of the component, denoted by $lifetime$, is assumed to follow an Erlang distribution, consisting of $k = 2$ exponentials, each with mean *beta*. The program output a 95% confidence interval for the mean number of renewals.

```
% Renewal of a system component.
% Inputs: beta (failure rate, Erlang distribution),
% period, sample (sample size)

meansum = 0; squaresum = 0;
for i=1:sample
  n = 0; sum = 0;
```

```
  while sum < period
    x = rand(1,2);
    lifetime = -beta*(log(x(1))+log(x(2)));
    sum = sum + lifetime;
    n = n+1;
  end
  meansum = meansum + n; squaresum = squaresum + n*n;
end
mean = meansum/sample
var = (squaresum - (meansum*meansum/sample))/(sample-1)
std = sqrt(var)
lowercfd = mean - 2*std/sqrt(sample)
uppercfd = mean + 2*std/sqrt(sample)
```

11.3 Static simulation models in Extend This section introduces some of the basic ideas behind the icon-driven models of Extend. In class, further time will be taken to discuss how Extend works, and how to construct models in Extend. For now, we consider just a few simple examples. In particular, this section considers static simulation models.

The Extend program below simulates a game of tossing a coin in which the player wins a dollar with heads and loses a dollar with tails..

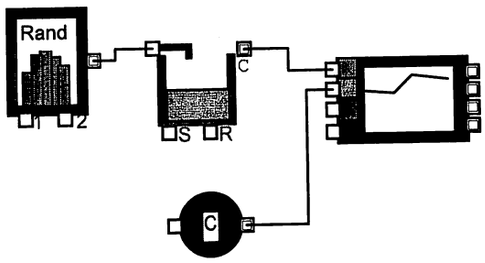


Figure 11.3.1: An Extend model for tossing a coin. With each toss, the player wins a dollar with a heads, and loses a dollar with tails.

In Extend. to run this model, one must select the option "continuous simulation" under "simulation setup" on the pull-down menu. For so-called continuous simulations, time is actually represented in the discrete units 0, 1, 2, $\cdots$. Each block generates **values** which are transmitted over the connecting line to the next block. A succeeding block takes the input value, processes it and generates an output value.

For the model above, the first block, labeled Rand, generates either a 1 or a -1, each with probability 0.5. By clicking on a block, a dialog box appears that allows the user to select these parameters. The next block, which looks like a tank, is called a **holding tank**. This block accumulates or sums the values that are sent to it. Thus, for the experiment

above, the value in the tank at any time in the simulation is the total winnings in the game. The third block over is a plotter which plots as function of time, the values sent to it. The circular block in the lower part of the figure just generates a constant value, in this case the value 0, which is then plotted also. The only purpose of this action is specify the zero level in the plotter output, in order to better compare the value of the winnings with zero.

As a more complex example, consider the one in Figure 11.3.2 which is a model of the renewal problem in Example 11.2.3. There are two holding tanks. The one at the top accumulates the successive lifetimes of the components. These lifetimes are generated by the block labeled Rand. The bottom

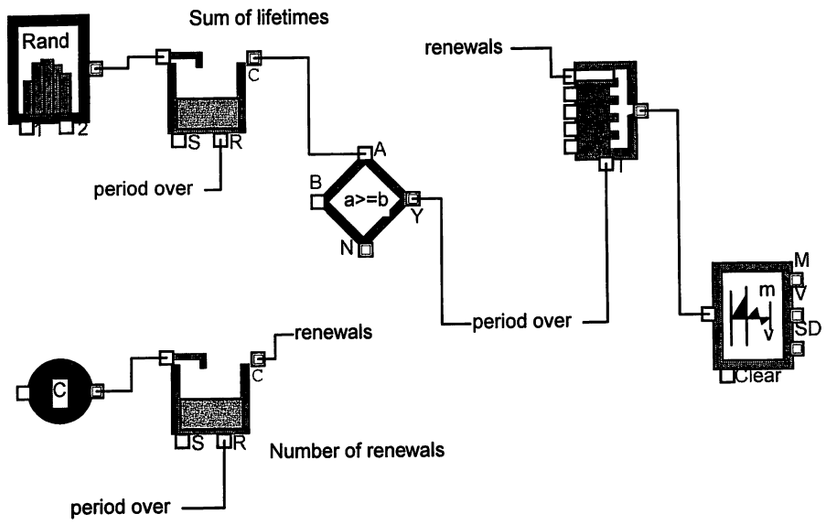


Figure 11.3.2 The renewal model of Example 11.2.3.

tank is sent the value 1 with each change in time. This lower tank, therefore, is simply holding the number of renewals at time proceeds.

When the sum of the lifetimes exceeds the period, this condition is detected by the diamond-shaped box labeled "a >=b". At this point, the text-block, entitled "period over", is sent the value 1. This action causes both holding tanks to reset there sums to zero, as done by the input to the R-connector at the bottom of each tank. Further, at the select box in the upper right corner, the value in the text-box called "renewal" is sent to

the statistics block on the far right for processing. This block compute the mean, variance, and standard deviation.

11.4 The newsstand problem The following problem, which is an introductory example from inventory theory, can be well handled using static simulation. A newsstand sells a certain newspaper each day. The number of customers who request a copy varies from day to day, and so is modeled as a random variable. Let Y denote this daily demand for the newspaper. We will assume that Y is a continuous random variable with density $f(y)$ and associated cdf $F(y)$. The question is: How many copies of the newspaper should be stocked each day?

Denote this amount to be stocked by x. Suppose the newsstand buys the newspapers for c dollars apiece, and sells each for s dollars, where $s > c$. Assume that at the end of the day, any unsold papers are discarded and have no salvage value. The daily profit Z is therefore

$$Z = s\,min(Y, x) - cx \ .$$

The optimal value of x will be determined so that the expected daily profit is maximized. To find this value, note first that

$$E(Z) = \int_0^\infty [s\,min(y, x) - cx] f(y)\,dy = s\int_0^\infty min(y, x)\,f(y)\,dy - cx \ .$$

For the integral term, we have

$$\int_0^\infty min(y, x)\,f(y)\,dy = \int_0^x yf\,(y)\,dy + x\int_x^\infty f(y)\,dy \ .$$

The second integral is equal to $1 - F(x)$. For the first integral, using integration by parts yields

$$\int_0^x yf\,(y)\,dy = xF(x) - \int_0^x F\,(y)\,dy \ .$$

Substituting into the formula for $E(Z)$, and simplifying, we get

$$E(Z) = G(x) = -s\int_0^x F(y)\,dy + (s - c)x \ .$$

The expression on the right is a function of x, and we have denoted it by $G(x)$. Since

$$G'(x) = -sF(x) + (s - c) \ , \quad \text{and} \quad G''(x) = -sf(x) \leq 0 \ ,$$

11.6 *Unit 4: Simulation Modeling*

it follows that G is concave downward, has a positive slope and is equal to zero at the origin, and goes to $-\infty$ as $x \to \infty$. Therefore, the optimal value of x, say x_{opt}, satisfies the equation $G'(x) = 0$. Thus

$$F(x_{opt}) = (s - c)/s\,, \quad \text{or} \quad x_{opt} = F^{-1}\big((s - c)/s\big)\ .$$

We shall assume the cdf F is monotonically increasing so that the inverse exists. An explicit expression for x_{opt} is possible only for simple distributions F. For example, the inverse can be found for the exponential distribution and for the Weibull distribution. However, for many reasonable choices of the demand distribution, finding the inverse explicitly is awkward at best, and usually impossible.

Further, it is not unusual in modeling stochastic systems that for certain random variables in the model, little is known about the distribution. For example, at the start of the analysis for the newspaper problem, the distribution of daily demand Y may be known only roughly. One may be able, however, to specify a range for the values of Y and perhaps a most likely value. In this case, a reasonable model for the density of Y might be the triangular distribution, which is specified by a lower bound a, an upper bound b, and a most likely value d. The shape of the density is a triangle with base on the interval $[a, b]$, and height $h = 2/(b - a)$. The density is expressed as $f(y) = h(y - a)/(d - a)$, for $a \le y \le d$, and $f(y) = h(b - y)/(b - d)$, for $d \le y \le b$. For y outside the interval $[a, b]$, the density is zero. Note that the peak of the triangle is located at $y = d$. To find the cdf F, integration gives us

$$F(y) = \frac{h}{2(d - a)}(y - a)^2\,, \quad \text{for} \quad a \le y \le d\ ,$$

and

$$F(y) = 1 - \frac{h}{2(b - d)}(b - y)^2\ , \quad \text{for} \quad d \le y \le b\ .$$

Of course, $F(y) = 0$ for $y \le a$, and $F(y) = 1$ for $y \ge b$. Note that finding an expression for F^{-1} is reasonable in this simple case.

For cases in which it is difficult or even impossible find F^{-1}, one could use graphing software to plot the graph of $G(x)$ and find the optimal value x_{opt} visually. Still, this approach may be inconvenient if we wish to perform a sensitivity analysis, particularly with regard to the choice of distribution for the demand. A simulation approach, although less accurate than an analytical solution, may be more convenient for sensitivity analyses.

Assignment Write a simulation model for the newspaper problem. Use this model to estimate x_{opt}, the optimal number of papers to be stocked each day. Take $s = 1.50$ and $c = 0.20$, and use the triangular distribution with range $a = 50$ to $b = 100$, and most likely value $d = 70$. Use a sample size of $n = 100$ days.

(a) Explain how you arrived at your estimate of x_{opt}. Corresponding to the amount you determine, find a confidence interval for the resulting mean profit $E(Z)$. Determine also a prediction interval for the daily profit, and compare this interval with the observed variations in the daily profit shown in the simulation model. State the interpretations of the confidence interval and the prediction interval you obtain.

(b) Compare the initial stock level $x = 75$ with the initial stock level $x = 85$. Based on your simulation model, which yields a larger expected daily profit? Justify your answer statistically.

(c) As a sensitivity analysis, use also for the demand distributions (i) the Poisson distribution with average $\lambda = 70$, and (ii) the gamma distribution with mean 70 and standard deviation 10.

Exercises

11.1 What does the following section of MATLAB code do to the entries in the array $\{N(1), N(2), \cdots, N(n)\}$? Justify your conclusion.

```
k = 2;
while k <= n
  j = ceil(rand*k);
  N(k) = N(j);
  N(j) = k;
  k = k + 1;
end
```

11.2 In the simulation model of a neutron passing through a shield, what is the role of the while-loop? What is the purpose of the variable `count`?

11.3 In the neutron shield model, the shield was assumed infinite in the y-direction. Suppose instead the shield has a floor at level $y = 0$, but no ceiling. How could the program be modified to handle this situation?

11.8 *Unit 4: Simulation Modeling*

11.4 Consider a best of seven match between two teams. Assume Team 1 wins a game with probability p. Write a simulation model of the best of seven series to estimate the probability Team 1 wins the series. Use p as an input to the program.

11.5 Modify the component renewal simulation model to include estimation of the standard deviation of the number of renewals, and to include a histogram of the renewal variable.

11.6 The following exercises concern the newsstand problem of Section 11.4 above.

(a) Verify the integration by parts in the derivation of the formula for the expected profit $G(x)$. Also verify the stated properties of this expected profit function $G(x)$.

(b) Suppose the distribution for demand follows the exponential density with parameter β. Find the formula for the optimal amount x_{opt} to be stocked. Determine this value x_{opt} when β, the mean of the distribution, is 70.

(c) Verify the formula for the cdf of the triangular distribution.

(d) For the triangular distribution considered in the section, one might guess that a reasonable choice for x_{opt} would be the most likely demand d. Show however that $x_{opt} < d$ when $(s-c)/s < (d-a)/(b-a)$, and that $x_{opt} > d$ when $(s-c)/s > (d-a)/(b-a)$. Show also that the value x_{opt} can range over the whole interval (a, b) depending on the relative magnitudes of $(s-c)/s$ and $(d-a)/(b-a)$.

Chapter 12

Simulation of Discrete-Event Systems

12.1 Introduction A **discrete-event system** is characterized by the property that the state of the system changes only at discrete points in time. These points in time are typically random, and the actions which result in these changes of state are called **events.** The state of the system is specified by variables which are functions of time and are called the **state variables**. For example, the system discussed below models the arrival and service of customers who enter a service facility and form a single line to await service from the next available server. Such a system is called a **single-line, multi-server queue**. We considered this model in the previous chapter, but only under the special assumptions that customer interarrival times were exponentially distributed, and that customer service times were exponentially distributed. In this chapter, these assumptions will not be required. A state variable for this model could be the number of customers in the system at any given time. This state variable takes on the values $0, 1, 2, \cdots$, and changes only at discrete points in time. There are two events for this system: a customer arrival and a service completion.

12.2 Discrete event simulation models in Extend An Extend model of a discrete event system consists of a collection blocks connected by heavy lines, as compared with the thin lines for the static or continuous simulation models discussed in a previous chapter. Each model must include the Executive block, and this block must be located to the left of the model blocks. In Figure 12.2.1 below, the Executive block is the one with the word "count" below it.

Consider for example the model below of a single-line, multi-server queue. While for continuous or static simulations, the flow along lines were values, they are now entities, such such a customers in a service system or jobs in a manufacturing system. The heavy lines indicate the flow of entities from one block to the next. As before, the thin lines represent flow of values.

12.2 *Unit 4: Simulation Modeling*

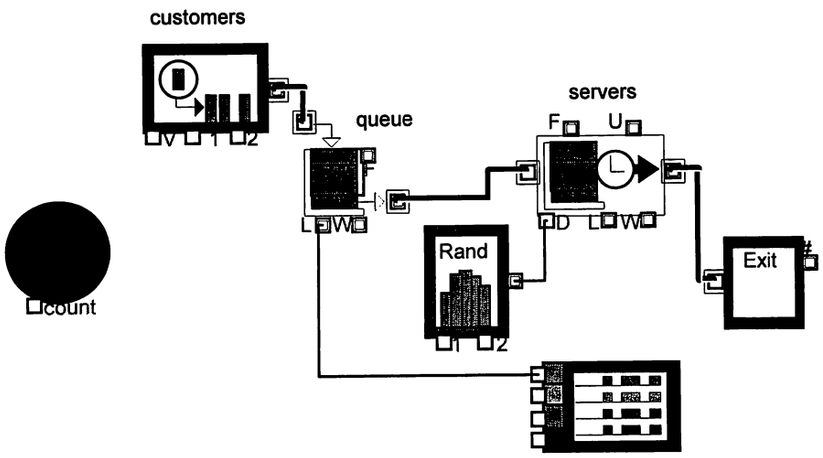


Figure 12.2.1 A single-line, multi-server queue. Each server has the same service time distribution, as specified in the Rand-block.

Most of the blocks in this model appear only in discrete-event simulation, and must be selected from the Discrete-Event Library. The block representing the server is called a "multiple activity" block. This block can represent any specified number of servers. Clicking on this block brings up a dialog box in which one specifies the number of servers. The Rand-block connected to the D-connector generates the random service time of each server.

The queue-block holds entities that are waiting for a server to become free. Note that the plotter is connected to the L-connector of the queue-block. This connector outputs the the length of the line at each point in time.

The block labeled "customers" is called a generator block, and produces the entities that flow through the system. By clicking on this block, a dialogue box appears that allows the user to specify the interarrival-time distribution.

Instead of using a multiple activity block, one could instead use several delay-blocks to model the individual servers, each with its own Rand-block to simulate the service times. This type of arrangement is shown below in Figure 7.2.2. Here, each server is represented by a separate delay-block. A Rand-block connected to the D-connector is used to simulate the service times. Thus each server has its own service-time distribution.

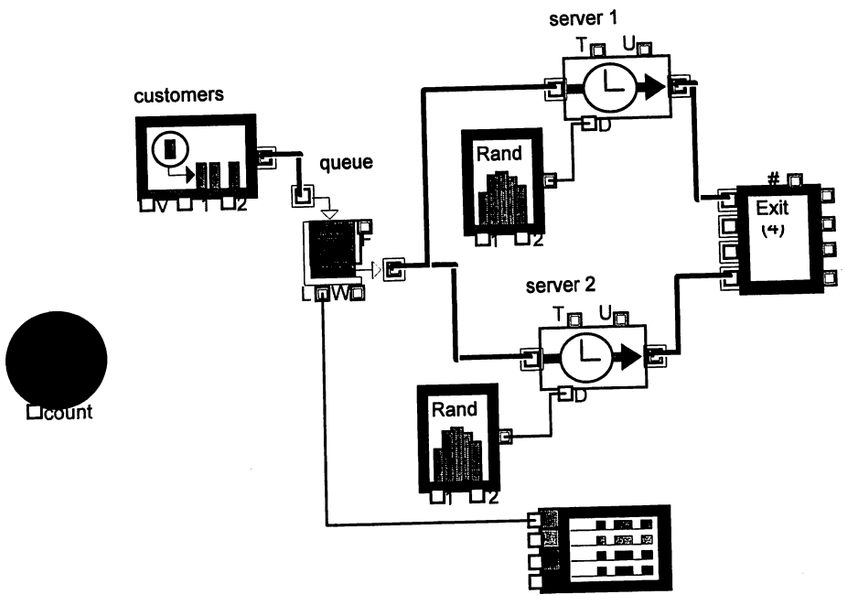


Figure 12.2.2 A two-server queue. Each server has its own service time distribution, specified by Rand-blocks connected to the D-connector of each.

The systems modeled above are examples of **open-loop** systems. The customers, or the entities, that move through the system are generated from the outside, and then enter the system. Generally these entities then leave the system after period of time. For such systems, if the entities enter the system faster than they leave, the values for the state variables could become arbitrary large. Thus, we need to be concerned about stability of open-loop systems, or in other words, whether the state variables become unbounded in time. In contrast, a **closed-loop** system is one in which the number of entities that move through the system stay fixed. In closed loop system, the values of the state variables generally reach limiting values in time.

12.3 Examples of closed-loop systems This section gives two examples of closed-loop systems. Each example involves a fixed number of entities which circulate through the system over time. To model such situations in Extend, a useful block is the Resource block which is used to start the simulation. Initially, this block contains all the entities that flow through the system.

12.4 *Unit 4: Simulation Modeling*

12.3.1 The delivery van problem This problem concerns an actual study conducted for a large transportation company based in Stuttgart, Germany. The project was undertaken to investigate the efficiency of certain aspects of their operations. At the end of this sub-section, you are asked to write a simulation model of the operations under study, and apply the model to develop recommendations.

ExpressO is an integrated transportation company. An important part of their business is the delivery van service. In each center of operation, the company runs a fleet of vans which make home deliveries for many of the large department stores and also deliver goods brought to warehouses by trucking firms. At one of their major centers, the manager was concerned that too much time was spent waiting for the drivers to report in and to receive new instructions from dispatchers. For each delivery, the driver received instructions from the dispatcher regarding particular details of the assignment, and upon return to the office, the driver reported to the dispatcher regarding the successful completion of the assignment, or reported any difficults that may have occurred. At the completion of the report, the driver would receive a new assignment. The office currently had two dispatchers, and the manager speculated that the addition of two more dispatchers would sufficiently reduce the waiting time. The following data and information is available :

1. At the time of the study, the company had 30 vans and 2 dispatchers. Delivery times averaged about 2 hours, and were estimated to follow a gamma distribution with standard deviation of 1 hour. Time with dispatchers averaged about 10 minutes, and was estimated to follow a gamma distribution with standard deviation of 3 minutes.

2. The hourly wage rate for dispatchers is 15 dollars per hour, and for drivers is 12 dollars per hour. The billing rate to customers is 30 dollars per hour for the van and driver.

3. Possible cost functions : Let s denote the number of dispatchers. Then we could seek to minimize the total cost $C(s) = 15s + wL_q$, where the cost w is either the wage of the drivers (12 dollars per hour), or the billing rate (30 dollars per hour) for van and driver. Use each of these costs and compare the results.

A remark about the cost functions By the way, given more information about operating costs we could consider a profit maximization function as the criterion for determining the optimal number of dispatchers. For example, if we also allow the number

of vans to be a variable of interest, we might use the profit function of the form

$$P(x, s) = (30 - d)(x - L_q) - (15s + [12 + c]x) ,$$

where : x = number of vans, d = operating cost of a van, and c = capital cost per hour of a van. However, as values for d and c are not made available, we will use the cost functions above.

Assignment Develop a simulation model of the delivery operations. Assume there is sufficient supply of assignments received by the company, so that the only delay is that of a van waiting for the dispatcher. A more complete study would model also the flow of assignments received by the company, the repair and maintenance of the vans, and might include the possiblity of purchasing additional vans. Run the simulation for a period of 12 hours and compute the resulting values of the cost functions. Make a number of independent runs of the simulation to analyse the cost values and to estimate the average cost during the 12 hour period.

Show in your report how you determined the number of dispatchers which results in least cost. Develop a table which shows for a given number of dispatchers, the corresponding prediction intervals for the cost values, and the confidence intervals for means, for both cost functions. The reader can then scan your table to see which number of dispatchers yields to least cost. Determine the number of dispatchers that results in the least costs, and explain how you arrived at your conclusion.

12.3.2 The machine-repair problem with spares Consider a set of $m + n$ machines which fail independently of each other. It is intended that m machines are to be in operation at any time. The remaining machines serve as spares and are called into operation when an operating machine fails. If more than n machine are in a state of failure, then all the operational machines will be in service. Suppose there are s, $1 \leq s \leq m$, repair persons who service the machines independently.

Assignment Write an Extend model of this machine-repair problem with spares. Consider the case of two servers ($s = 2$), six machines ($m = 6$), and two spares ($n = 2$). Take the service-time distribution to be gamma with a mean and standard deviation of 8 and 2.5 days respectively. Take the mean time between breakdowns to follow an exponential distribution with a mean of 4 days. Program inputs would include : Number of machines, number of repair stations, mean time between break downs, mean repair time, and duration of the simulation. Program outputs would include : Average number of machine that are not operational.

12.6 *Unit 4: Simulation Modeling*

Problem Suppose it is desired that 80% of the time there should be at least 4 machines in operation. Determine how many spares are needed in this case.

Exercises

12.1 In a queueing system, customers arrive with interarrival times that are independent and uniformly distributed on the interval 0.5 to 3 hours. Customers begin by first going to server A, and then with probability p (given) go to server B, or with probability $1-p$ go to server C, and then leave the system. Assume the work times with each server are independent and exponentially distributed with mean 2 hours. Write an Extend model for this system.

12.2 In a queueing system, customers first go to server A, and then with probability p return to server A, or with probability $1-p$ go to server B, and then leave the system. Assume the work times with each server are independent and exponentially distributed with mean 2 hours. Write an Extend model for this system.

12.3 Consider again the machine-repair problem of Section 12.3.2. Recall that there is a set of $m+n$ machines which fail independently of each other, and assume now that they fail at an exponential rate λ. Recall that m machines are to be in operation at any time, while the remaining machines serve as spares and are called into operation when an operating machine fails. If more than n machine are in a state of failure, then all the operational machines will be in service. Suppose there are s, $1 \leq s \leq m$, repair persons who service the machines independently and each at exponential rate μ. Let $X(t)$ denote the number of machines at time t that are *not* operational; that is, they are in the repair shop. Derive the arrival and departure rates for the birth-and-death model of the machine-repair problem with spares. Use these rates to develop a formula for the average number of machines that are not operational. Write a program, in MATLAB say, which computes the long-run state probabilities and implements your formula for the average number of machine that are not operational. Use the model developed in the assignment above to compare your results with theory. In particular, consider the following cases: (a) two servers, six machines, and two spares, and (b) three servers, six machines and two spares. For each case, take the mean service times to be 8 days, and the mean time between breakdowns to be 4 days for each run. Study also how the computed averages change with increasing duration of the simulation.

(a)

(b)

mean serve = 2 days
mean fail = 4 days

} with these we get average in queue = 4

Chapter 13

Preparation of Simulation Model Inputs

13.1 Introduction A simulation model will typically include a number of random variables such as interarrival times and the service times in a queuing network, or customer demand and re-order times in an inventory model. To implement a simulation model, therefore, it is necessary to determine the distributions of the random variables that comprise the model. The analysis leading to the description of such random variables generally consists three phases. These are (1) *Data collection and summary*, (2) *Fitting distributions to the data*, and finally (3) *Testing for goodness-of-fit*.

13.2 Data collection and summary We begin by noting that in developing simulation models, it is common to encounter random variables for which very little data is available. Often, at best, one may be able to obtain only estimates of extreme values (smallest and/or largest values), the mode (most frequent value), and perhaps the mean. In these situations, the triangular distribution, or the beta distribution, are often used.

Example 13.2.1 — The traingular density Suppose for a certain random variable it is known only that the distribution has support on the interval $[a, b]$ and has a mode (most frequent value) equal to m. The triangular density $f(x)$ is the continuous function which is linear on the interval $[a, m]$ with $f(a) = 0$, linear on the interval $[m, b]$ with $f(b) = 0$, and which is zero elsewhere. The height h of the density at $x = m$ must satisfy $\frac{1}{2}h(b - a) = 1$, so that the total area bounded by the graph of f and the x-axis will be one. □

When it is possible to obtain a reasonable number of independent observations of a random variable, one should first form a histogram, or frequency chart, for the data. By examining the shape of this graph, one can then make a choice for a type of distribution that may fit the data well. Indeed, a histogram can be viewed as a scaled version of the

13.2 *Unit 4: Simulation Modeling*

density function. The histogram is formed by dividing the range of observed values into subintervals $[b_0, b_1)$, $[b_1, b_2), \cdots, [b_{l-1}, b_l)$, usually of equal length $\Delta = b_i - b_{i-1}$. The number of observations n_i that lie in the interval $[b_{i-1}, b_i)$ is estimated by

$$n_i \approx \int_{b_{i-1}}^{b_i} f(x)dx \approx \Delta \cdot f(m_i)\,, \quad \text{where } m_i \text{ is some point in } [b_{i-1}, b_i)\ .$$

Thus, when Δ is small and there are a sufficiently large number of observations, the shape of the histrogram will resemble the graph of the density f. There is some art in choosing of the number of intervals so that the shape of the density can be discerned. If too many or too few intervals are used, then the histogram will not have a shape similar to that of the density. Empirical studies have suggested that using about $l = 1 + log_2 n$ intervals will often work well.

In selecting the type of distribution, it is sometimes helpful to estimate quantiles of the data, such as the octiles, in order to study the symmetry or skewness of the data. A histogram is helpful in this regard. Indeed, from the histogram one can form the empirical cdf $F(x)$ of the data, which is defined by $F(x) =$ the fraction of data values which are $\leq x$. Then, for example, the third quartile q_3 is estimated by the equation $0.75 = F(q_3)$.

13.3 Fitting the data Once the type of distribution has been selected, we can then fit the distribution to the data. Fitting a distribution is commonly done by estimating the parameters of the distribution using a method of parameter estimation, such as the method of moments (MME) or the method of maximum likelihood (MLE).

Example 13.3.1 Suppose we wish to fit a gamma distribution to sample data from which it has been calculated that the sample mean is $\overline{x}$ and the sample variance is s^2. The gamma density has the form $Kx^{\alpha-1}e^{-x/\beta}$, for an appropriate constant K. Recall that the mean of this density is $\mu = \alpha\beta$ and the variance is $\sigma^2 = \alpha\beta^2$. The method of moments determines the parameter α and β by requiring that $\overline{x} = \alpha\beta$ and $s^2 = \alpha\beta^2$. These equation can now be solved for the parameters α and β. □

Example 13.3.2 Suppose we wish to fit the exponential distribution to data $x_1, x_2, \cdots, x_n$. The exponential density is $f(x) = \frac{1}{\beta}\, e^{-x/\beta}$ for $x > 0$. Using the method

of maximum likelihood, we form the likelihood function

$$L(\beta) = \prod_{i=1}^{n} \frac{1}{\beta} e^{-x_i/\beta} = \frac{1}{\beta^n} e^{-n\overline{x}/\beta} ,$$

and then determine the value of β which maximizes this expression. Taking the logarithm of both sides and using calculus, one finds that the optimal choice of β is $\beta = \overline{x}$. □

13.4 Testing goodness-of-fit Suppose finally that we have determined a distribution to represent a certain random variable, and we wish to see how well this distribution models the data. For this purpose, the so-called χ^2 test is often used. The stated null hypothesis specifies that the random variable follows the specified distribution. To form the test, one divides the range of the n data values into k subintervals. Let n_i denote the number of data values in the i-th interval. Corresponding to the distribution specified in the null hypothesis, let p_i be the theoretical fraction of observations that belong to the i-th interval. Then the test statistic is given by

$$S = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} .$$

A large value of this statistic would be evidence for rejection of the null hypothesis since, under the null hypothesis, n_i and np_i should be close to each other.

In forming this test of hypothesis, there are two cases to consider. First, suppose the distribution specified in the null hypothesis is determined *independently* of the data. For example, we may have data which is assumed to be uniform on the interval [0, 1], and we want to test whether there is evidence to reject this hypothesis. For cases such as this, it has been shown that for large sample sizes n, the distribution of the test statistic S under the null hypothesis is approximately χ^2 with $k - 1$ degrees of freedom. Thus, using a level of significance α, we would reject the null hypothesis if $S > \chi^2_{k-1,\alpha}$, where $\chi^2_{k-1,\alpha}$ is the critical value of the χ^2 distribution with $k - 1$ degrees of freedom. Recall that by definition, a random variable which follows the χ^2 distribution with $k - 1$ degrees of freedom will exceed $\chi^2_{k-1,\alpha}$ with probability α.

The second case to consider is that in which the distribution in the null hypothesis *depends* on the data. For example, the parameters of the distribution might have been computed from the data using, say, the method of maximum likelihood. Under these circumstances, it is known that for large sample sizes, the distribution of the test statistic S under the null hypothesis is approximately χ^2 with $k - 1 - m$ degrees of freedom,

13.4 *Unit 4: Simulation Modeling*

where m is the number of parameters which were estimated using the data. Thus, in this case, we would reject the null hypothesis if $S > \chi^2_{k-1-m,\alpha}$.

Another important test for goodness-of-fit is the Kolmogorov-Smirnov test. Suppose we have obtained a random sample of values $x_1, x_2, \cdots, x_n$ of the random variable X. Assume the data is written in increasing order. Suppose a cdf $F(x)$ is given (assumed to not depend on the data), and we wish to test whether the distribution of this random variable follows this cdf . The test begins by forming the empirical cdf of X, which is defined by

$$F_n(x) = \frac{1}{n}[\text{number of } x_i \text{ values which are} \leq x] \ .$$

Thus, $F_n(x) = i/n$ for $x_i \leq x < x_{i+1}$, while $F_n(x) = 0$ for $x < x_1$ and $F_n(x) = 1$ for $x \geq x_n$. Define

$$D_n = \sup_x \mid F_n(x) - F(x) \mid \ .$$

Note that D_n is a measure of how close F_n is to F. The computation of D_n is made easier by noting that because $F_n(x)$ has the constant value i/n on the interval $[x_i, x_{i+1})$, and because F is monotonic, we can write

$$D_n = \max_i D_n(i) \ , \tag{13.4.1}$$

where

$$D_n(i) = max\left\{ \left|F(x_i) - \frac{i-1}{n}\right| , \left|F(x_i) - \frac{i}{n}\right| \right\} .$$

Now, to test the null hypothesis that X follows the distribution F, we reject at the level of significance α if $D_n > d_{n,1-\alpha}$ where $d_{n,1-\alpha}$ is a critical value determined so that the probability of rejection is α when the null hypothesis is true. These d-values have been tabulated. What is remarkable about them is that they do not depend on the distribution from which the data was drawn.

The foregoing test of hypothesis assumes that the distribution in the null hypothesis does not depend on the observed data values of the random variable X. It is also possible to use the Kolmogorov-Smirnov test for certain cases in which the cdf F depends on the data. Critical points have been determined for the normal distribution, the exponential and Weibull distributions, for example. A detailed discussion of the Kolmogorov-Smirnov test can be found in the text by Law and Kelton, *Simulation Modeling and Analysis, 2nd edition*, McGraw-Hill, 1991.

Exercises

13.1 Find the mean of the triangular density over the interval $[a, b]$ with mode m.

13.2 Find MME and the MLE for the parameter β in the density $f(x) = Kxe^{-x/\beta}$ for $x > 0$.

13.3 Test the following data for uniformity over the integers from 1 to 5.

Data : 3 4 3 2 2 1 1 1 3 3 4 4 5 3 3 2 1 2 2 4 3 .

13.4 Verify formula (13.4.1) for the Kolmogorov-Smirnov test.

13.5 Test the following data for uniformity on the interval $[0, 1]$.

.358	.997	.205	.587	.254	.563	.984
.408	.258	.429	.801	.838	.934	.356
.543	.817	.923	.248	.710	.410	.197

Use the χ^2 goodness-of-fit test.

Chapter 14

Random Number Generation

14.1 Introduction This chapter provides a brief introduction to methods for generating random samples from specified probability distributions. Two general methods are presented, along with some special techniques for the normal distribution.

14.2 Generation of uniformily distributed random numbers The ability to generate random samples from the uniform distribution on $(0,1)$ is the cornerstone of all methods for generating random samples from specified distributions. Insight into the need for this ability can be seen from the following result. Suppose F is an arbitrary cdf and U is a uniformily distributed random variable on $(0,1)$. Assume F has an inverse which is defined on $(0,1)$. Denoted the inverse by F^{-1} and set $X = F^{-1}(U)$. Then X is a random variable with cdf F. Indeed,

$$P(X \leq x) = P\left(F^{-1}(U) \leq x\right) = P\left(U \leq F(x)\right) = F(x) \ .$$

Thus, to generate a random sample from X, it suffices to generate a random sample from the uniform distribution of U, say $u_1, u_2, \cdots, u_n$, and then the random sample from X is $F^{-1}(u_1)$, $F^{-1}(u_2), \cdots, F^{-1}(u_n)$. This technique is well defined in general, but to be practical, the inverse cdf needs to be easily obtained.

Example 14.2.1 The exponential with mean β has cdf $F(x) = 1 - e^{-x/\beta}$. The inverse is $F^{-1}(x) = -\beta ln(1-x)$. Thus, if U is uniform $(0,1)$, then $X = -\beta ln(1-U)$ follows the exponential distribution with mean β. Since $1-U$ is also uniform $(0,1)$, we usually take $X = -\beta ln(U)$, to save arithmetic. □

Another example in which the inverse transform is easily obtained is the Weibull distribution for which the cdf is $F(x) = 1 - e^{-x^\gamma/\beta}$, for $x > 0$, where $\gamma > 0$, $\beta > 0$. For many important distributions, such as the gamma, beta, and normal, the inverse cdf

method is not easily implemented. Methods that apply to these distributions will be introduced in later sections.

However, whether the method used is the inverse cdf method or any another, a key ingredient is the generation of numbers that are uniformily distributed on $(0, 1)$. For this task, the technique commonly used today is based on a linear congruential iteration of the form

$$Z_{i+1} = (aZ_i + c)\ mod(m)\ , \quad \text{for given integers } a,\ c, \text{ and } m\ ,$$

and some initial integer value Z_0, called the seed. The operation $mod(m)$ means that the number $aZ_i + c$ is divided by m, and the remainder is set equal to Z_{i+1}. The Z_i are thus all integers and satisfy $0 \leq Z_i \leq m - 1$. The uniform random number stream on $(0, 1)$, $u_1, u_2, \cdots$, is formed by setting $u_i = Z_i/m$. In considering the effectiveness of such a method, there are two basic questions. One concerns the range of integers Z_i that are produced. This question seems to be well understood, and results are known that characterize the iterations of full cycle (every integer between 0 and $m - 1$ is produced), and those iterations that produce large sets of Z_i values. The second question concerns the uniformity and independence of the supposed random stream $u_1, u_2, \cdots$. This question is not as well understood as the former, but many results are known. For example, there are estimates of the correlation between u_i and u_{i+1} in terms of the constants $a, c,$ and m.

14.3 The inverse cdf method in the discrete case Suppose X is a discrete random variable with values $x_1, x_2, \cdots$, and pdf $P(X = x_i) = p_i$ for $i = 1, 2, \cdots$. Let U denote a uniform random variable on $(0, 1)$. To generate a random observation of X using the cdf method, note that the event

$$A_i\colon \quad \sum_{n=1}^{i-1} p_n < U \leq \sum_{n=1}^{i} p_n\ ,$$

occurs with probability p_i. Thus, the method is to generate a random value of U, determine the index i for which the event A_i occurs, and then select the value x_i for X.

Example 14.3.1 Let X be a Bernoulli random variable with $P(X = 1) = p$, and $P(X = 0) = 1 - p$. Then a random observation from X is obtained by generating a random observation from the uniform $(0, 1)$ distribution, say u, and then selecting the value for 1 for X if $u \leq p$, and selecting 0 otherwise. □

This inverse cdf method can applied in principle to any discrete distribution. In special cases, particularly when the discrete variable has an infinite number of possible values, it is helpful to seek simplifications.

Example 14.3.2 Consider the geometric distribution : $P(X = i) = p(1-p)^{i-1}$, for $i = 1, 2, \cdots$. In the inequalities defining the event A_i, the sums in this case can be written in closed form. Setting $\beta = 1/ln(1-p)$, it turns out that the event A_i occurs for $i = 1 + [\beta ln(1-U)]$, where $[x]$ is the largest integer less than or equal to x. Thus, we conclude that the random variable $1 + [\beta ln(U)]$ follows the geometric distribution with probability p of success on a trial, where U is uniform $(0, 1)$. Note that $1 - U$ has been replaced by U to save arithmetic.

14.4 The rejection method One of the more intriguing methods for generating random samples is the so-called rejection method of J. Von Neumann (Various techniques used in connection with random digits, U.S. National Bureau of Standards Applied Mathematics Series, No. 12, pages 36-38, 1951). Suppose f is a density function from which we wish to generate random observations, and suppose g is another density for which we have a method to generate random samples. Assume there is a constant c such that $f(x) \leq cg(x)$ for all x. This density g is sometimes called a majorizing density for f. The method proceeds as follows.

Step 1 Generate a random observation, say u, from the uniform $(0, 1)$ distribution, and independently generate a random observation, say y, from the density g.

Step 2 If $u \leq f(y)/cg(y)$, then take y as the observation from the density $f(x)$. Otherwise, return to Step 1 and continue.

The proof that this method works is instructive, and helps estimate the number of times these steps will be repeated until an observation is accepted. Let X denote the random variable whose values are generated by this method. We need to show that density of X is f. Let Y be a random variable whose density is g. For a uniform $(0, 1)$ random variable U that is independent of Y, we have

$$P(X \leq x) = P(Y \leq x \mid U \leq f(Y)/cg(Y)) = \frac{1}{p} P(Y \leq x, U \leq f(Y)/cg(Y)) ,$$

where $p = P(U \leq f(Y)/cg(Y))$. Notice that p is the probability that the method terminates on a given trial, so this number is of interest to us. Conditioning on Y, we get

$$P(X \leq x) = \frac{1}{p}\int_{-\infty}^{\infty} P(Y \leq x,\, U \leq f(Y)/cg(Y) \mid Y = y)g(y)\,dy$$

$$= \frac{1}{p}\int_{-\infty}^{x} P(\,U \leq f(y)/cg(y))\,g(y)\,dy$$

$$= \frac{1}{p}\int_{-\infty}^{x} (f(y)/cg(y))\,g(y)\,dy = \frac{1}{pc}\int_{-\infty}^{x} f(y)\,dy = \frac{1}{pc}F(x)\ ,$$

where F is the cdf corresponding to the density f. Now, allowing $x \to \infty$, we see that $pc = 1$. Thus, $P(X \leq x) = F(x)$, and so it follows that X has the density f, which completes the proof.

This argument shows that $c = 1/p$, and recalling that for the geometric distribution, the expected number of trials until the first success is $1/p$, we see that the constant c is the average number of times the steps will be repeated until a value is accepted. Thus, it is best to take c as small as possible.

Example 14.4.1 Consider the beta distribution

$$f(x) = \frac{1}{B(\alpha,\beta)}\,x^{\alpha-1}(1-x)^{\beta-1}\ , \quad \text{for } 0 < x < 1\ .$$

where $\alpha > 1$ and $\beta > 1$. The rejection method can be implemented in this case by taking g to be the uniform density on $(0,1)$. The constant c is then the maximum of f on $(0,1)$. This maximum can be computed using calculus.

Example 14.4.2 Consider the gamma$(\alpha, 1)$ distribution

$$f(x) = \frac{1}{\Gamma(\alpha)}\,x^{\alpha-1}e^{-x}\ , \quad \text{for } x > 0$$

where $0 < \alpha < 1$. The rejection method can be implemented in this case with $g(x) = Kh(x)$, where $h(x)$ equals $x^{\alpha-1}$ for $0 < x < 1$, and equals e^{-x} for $x > 1$. The constant K is $\alpha e/(\alpha + e)$, so that g is a density. Sampling from g can be done with the inverse cdf method. Because $x^{\alpha-1}e^{-x} \leq h(x)$ for $x > 0$, with equality at $x = 1$, it follows that the best constant c in the rejection method is $1/K\Gamma(\alpha)$. For example, when $\alpha = 1/2$, $c \approx 1.34$. □

If X is a random variable with gamma$(\alpha, 1)$ distribution as in Example 9.4.2, then βX has the gamma(α, β) density. Thus, we can generate observations from the gamma(α, β) distribution when $0 < \alpha < 1$ and $\beta > 0$. Finally, recall that a sum $X_1 + X_2 + \cdots + X_n$ of independent random variables, where X_i is gamma(α_i, β), is itself gamma(α, β) with $\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_n$. Thus, for the gamma$(\alpha, \beta)$ distribution when $\alpha > 1$, we can write $\alpha = n + q$, where n is an integer and $0 \leq q < 1$. This distribution can then be simulated with a sum of n exponential variables with mean β, and a gamma(q, β) variable.

14.5 The standard normal distribution There are a number of methods available for this case. One is based on the result that if a random variable X has density

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}, \quad \text{for } x > 0,$$

then the variable Z defined by $Z = X$ with probability $1/2$, and $Z = -X$ with probability $1/2$, follows the standard normal distribution. Random observations from the density f can be generated with the rejection method, using the majorizing density $g(x) = e^{-x}$. The constant c turns out to be $\sqrt{2e/\pi} \approx 1.32$.

Another approach is based on the central limit theorem. Suppose $U_1, U_2, \cdots$, are independent uniform $(0, 1)$ random variables. Set $X = U_1 + U_2 + \cdots + U_n$. Then $E(X) = n/2$ and $var(X) = n/12$. For large n, the random variable

$$Z = \frac{X - (n/2)}{\sqrt{n/12}} \quad \text{is approximately standard normal.}$$

A common choice is $n = 12$, as then Z is simply $X - 6$.

The last method we mention was probably motivated by studies of the bivariate normal distribution. Suppose X and Y are independent standard normal variables. Let (R, Φ) denote the polar coordinates of the point with cartesian coordinates (X, Y). Set $D = R^2$. The transformation relating values (x, y) of (X, Y) and values (d, ϕ) of (D, Φ) is specified by $x = d^{1/2} cos\phi$ and $y = d^{1/2} sin\phi$. Forming the Jacobian we get

$$\begin{bmatrix} \frac{\partial x}{\partial d} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial d} & \frac{\partial y}{\partial \phi} \end{bmatrix} = \begin{bmatrix} \frac{cos\phi}{2\sqrt{d}} & -\sqrt{d}\, sin\phi \\ \frac{sin\phi}{2\sqrt{d}} & \sqrt{d}\, cos\phi \end{bmatrix} = \frac{1}{2},$$

and so the joint density of (D, Φ) is found to be

$$f_{D,\Phi}(d,\phi) = f_{X,Y}(x,y) \cdot \frac{1}{2} = \frac{1}{4\pi} e^{-d/2} = \frac{1}{2\pi} \cdot \frac{1}{2} e^{-d/2} ,$$

where $d \geq 0$ and $0 \leq \phi < 2\pi$. It follows that D and Φ are independent, and that Φ is uniform on $[0, 2\pi)$, while D follows the exponential distribution with mean 2. These results tell us that if U_1 and U_2 are independent uniform $(0, 1)$ variables, then the random variables Z_1 and Z_2 defined by

$$Z_1 = (-2lnU_1)^{1/2} cos(2\pi U_2) , \quad \text{and} \quad Z_2 = (-2lnU_1)^{1/2} sin(2\pi U_2) ,$$

are independent and each follow the standard normal distribution.

Exercises

14.1 Develop the inverse cdf method for generating random observations from the Weibull distribution.

14.2 Verify the results in Example 14.3.2.

14.3 Use the result of Example 14.3.1 to determine a method for generating random observations from the binomial distribution.

14.4 Verify the results in Example 14.4.2.

14.5 Use the rejection method to develop an algorithm for generating random observations from the beta distribution $60x^2(1-x)^3$ on $(0, 1)$. Use the majorizing density $g(x) = 1$. What is the expected number of trials until acceptance of a generated value?

14.6 Determine the inverse cdf method for generating random observations from the triangular distribution on $(1, 4)$ with mode at 3.

14.7 Suppose the random variable X has density $f(x) = \left(2/\sqrt{2\pi}\right) e^{-x^2/2}$, $x > 0$. Define the random variable Y by $Y = X$ with probability $1/2$, and $Y = -X$ otherwise. Show that Y follows the standard normal distribution.

14.8 Let U_1 and U_2 be independent uniform $(0, 1)$ random variables. Set $X = 2U_1 - 1$ and $Y = 2U_2 - 1$. Show that the distribution of (X, Y) is uniform on the square of side 2 and center at the origin. Show that the distribution of (X, Y), conditioned on the event $X^2 + Y^2 \leq 1$, is uniform on the unit circle with center at the origin. Explain how this result could be used to generate random points in the unit circle.

14.9 Let (X, Y) be uniformily distributed on the unit circle with center at the origin. Let $R \geq 0$ be defined by $R^2 = X^2 + Y^2$, and let (R, Φ), where $0 \leq \Phi < 2\pi$, denote the polar coordinates the point (X, Y). Show that R^2 and Φ are independent, that R^2 is uniformily distributed on $(0, 1)$, and that Φ is uniformily distributed on $[0, 2\pi)$.

14.10 For a piecewise continuous function $f(x)$ on $[0, 1]$, the integral

$$I = \int_0^1 f(x)\,dx \quad \text{can be estimated by} \quad \frac{1}{n}\sum_{i=1}^{n} f(u_i)\ ,$$

where $u_1, u_2, \cdots, u_n$ are n independent random numbers generated from the uniform $(0, 1)$ distribution. This method is an example of **Monte Carlo integration**. Show that the expected value of the estimates equals the integral, and that with probability about 95%, the error of estimation is at most $2\sigma/\sqrt{n}$, where

$$\sigma^2 = \int_0^1 [f(x)]^2\,dx - I^2\,.$$

By the way, this method is not very efficient in one dimension, but its performance becomes increasingly better as the dimension of the integral goes up.

Chapter 15

Analysis of Simulation Model Outputs

15.1 Introduction There are basically two types of simulation models : terminating and nonterminating. A **terminating simulation** model is one in which either (1) there is a specified time at which the simulated activities end, in which case we say the simulation is **time-controlled**, or (2) there is some event that causes the simulation to end, in which case we say the simulation is **event-controlled**. For example, a simulation model of a business which is open only during specified hours each day, and for which the activities can be considered independent from day to day, would be a time-controlled simulation. On the other hand, consider a simulation model of a queueing system with a waiting room of limited capacity. Suppose it is desired to estimate the first time at which the room becomes full. The simulation would end when the waiting room becomes full, and thus the model is a terminating simulation which is event-controlled. A **nonterminating simulation** model is one that has no natural end to the activities being simulated. For example, a simulation model of a maintance facility for military weapons, or a simulation model of an emergency medical service, would be examples of nonterminating simulation models. For these types of simulation, modeling is often done to determine the behavior of the system under steady state conditions. Sometimes, however, such systems do not have a steady state, in which case it is necessary to specify a way to terminate the simulation.

15.2 Statistical methods for terminating simulation models The main approach for these models is to simply make a number of independent runs of the simulation, and thus obtain a random sample of observations for each measure of performance. One can then apply standard statistical methods to estimate means and variances. Specifically, suppose we obtain n independent observations $Y_1, Y_2, \cdots, Y_n$ of a certain measure of performance.

15.2 *Unit 4: Statistical Methods in Simulation Modeling*

Then a $100(1-\alpha)\%$ confidence interval for the mean is

$$\overline{Y}(n) \pm c(n,\alpha)\frac{S_n}{\sqrt{n}} ,$$

where

$$\overline{Y}(n) = \frac{1}{n}\sum_{i=1}^{n} Y_i , \quad \text{and} \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}(n))^2 .$$

Typically, the validity of this confidence interval is based on the central limit theorem, in which case the coefficient $c(n,\alpha)$ would be $z_{\alpha/2}$, the critical point for the normal distribution. However, if we find that the data is normal or nearly so, then for $c(n,\alpha)$ we could use $t_{n-1,\alpha/2}$, the critical point for the t-distribution with $n-1$ degrees of freedom.

One concern is the choice of sample size n. A common idea is to make several initial runs of the simulation model, and then use these to obtain a rough estimate the sample variance. Denoting this estimate by S, we then choose the sample size n so that

$$c(n,\alpha)\frac{S}{\sqrt{n}} \leq \beta ,$$

where β is either a prescribed absolute error, or an error that is expressed relative to the magnitude of the mean. In the latter case, one does need to have some idea of the magnitude of the mean, which is the quantity being measured in the first place.

15.3 Statistical methods for nonterminating simulation models Consider a single run of the model which yields observations of random variables $Y_1, Y_2, \cdots$, that represent a certain measure of system performance. For example, the measure of performance might be customer waiting time, where Y_i is the waiting time of the i-th customer. To estimate the average of the performance measure, one can simply use the sample mean of the Y_i. However, to estimate the error in the sample mean, we must keep in mind that the Y_i are typically correlated, as they are in the example of waiting times. For such cases, we can not use the formula for sample variance to estimate variance. One way to deal with this difficulty is provided by the **method of batch means**.

We shall assume the process $\{Y_i\}$ is **covariance stationary** which means there are constants μ and σ such that $E(Y_i) = \mu$ and $Var(Y_i) = \sigma^2$ for each i, and $cov(Y_i, Y_{i+j})$ depends only on j. For random variables X and Y, the covariance is defined by

$$cov(X,Y) = E[(X-\overline{X})(Y-\overline{Y})] .$$

The covariance is important in the measurement of correlation between random variables. The assumption of covariance stationarity implies that the correlation between Y_i and Y_{i+j} depends only on the amount of time between the observations, but not on how long the simulation has been running. In applications, the covariance between Y_i and Y_{i+j} is expected to diminish substantially as j increases.

The method of batch means divides the observations into batches of length l, say $\{Y_1, Y_2, \cdots, Y_l\}$, $\{Y_{l+1}, Y_{l+2}, \cdots, Y_{2l}\}$, $\{Y_{2l+1}, \cdots, Y_{3l}\}$, and so on. Suppose that we have n batches, giving us a total of $m = nl$ observations. For each batch we determine the sample mean

$$X_k(l) = \frac{1}{l} \sum_i Y_i \ ,$$

where the sum for this k^{th} batch extends over the range of indices $kl - l + 1$ to kl. The overall mean for a set of n batches is then

$$\overline{X}\,(m) = \frac{1}{n} \sum_{k=1}^{n} X_k(l) = \frac{1}{m} \sum_{i=1}^{m} Y_i \ .$$

We can now argue that if l is large enough, then the random variables $\{X_k(l) \mid k = 1, 2, \cdots, n\}$ are nearly independent. We can also argue, by appealing to the central limit theorem, that for l large enough, and under certain assumptions concerning the correlations between the Y_i, each batch mean $X_k(l)$ is approximately normal. Finally, because of the covariance stationarity, we have $E(X_k(l)) = \mu$ for each k, and also $Var(X_k(l))$ does not depend on k (although it does depend on l). Thus, it can be argued that the random variables $X_1(l), X_2(l), \cdots, X_n(l)$ are nearly independent, identically distributed normal random variables. It follows that a $100(1-\alpha)\%$ confidence interval for the mean μ is

$$\overline{X}\,(m) \pm t_{n-1,\alpha/2} \frac{S_{n,l}}{\sqrt{n}} \ , \quad \text{where} \quad S_{n,l}^2 = \frac{1}{n-1} \sum_{k=1}^{n} (\, X_k(l) - \overline{X}\,(m)\,)^2 \, .$$

To implement this scheme, it is generally necessary to run the simulation for an initial warmup period so that during the remainder of the simulation run, the assumption of covariance stationarity will be reasonable. It is important that l be chosen large enough so that the batch means will be independent. Without this independence, the validity of the confidence interval is in doubt. Also, the variance estimator $S_{n,l}$ will be biased, and will likely under estimate the true variance. The normality of the batch means is not so

15.4 *Unit 4: Statistical Methods in Simulation Modeling*

critical, because the central limit can be applied when the sample size n is large enough, often $n \geq 30$ or 40 is sufficient.

As independence of the batch means is the most critical part of this method, a statistical test could be included to check for this condition. One convenient test is the following. Assume that a sequence of n observations is given. For example, consider $5, 8, 9, 2, -1, -4, 0$, and 3. Scanning from left to right, we determine for each successive pair of numbers, whether the change is an increase, in which case we record a $+$, or a decrease, in which case we record a $-$. For the sequence at hand, we get $+ + - - - + +$. We next count the number of runs in the sequence of $+$'s and $-$'s. In this example, there are 3 runs, the initial $+ +$ is the first run, followed by $- - -$, and finally $+ +$ at the end. In general it can be shown that for a random sequence of n numbers, the number R of runs satisfies

$$\mu_R = E(R) = \frac{2n-1}{3}, \quad \text{and} \quad \sigma_R^2 = Var(R) = \frac{16n-29}{90},$$

and for large n, the statistic $W = (R - \mu_R)/\sigma_R$ is approximately standard normal. Using this result, we can test for evidence of non-randomness in a sequence by checking whether value of W lies in the anticipated range of, say, -2 to 2, as would be expected under the null hypothesis of randomness.

Exercises

15.1 Suppose $X_1, X_2, \cdots$ is a covariance stationary process with mean μ and variance σ^2, and let ρ_ν denote the correlation of X_i and $X_{i+\nu}$. Let $\overline{X}_n$ denote the mean of $X_1, X_2, \cdots, X_n$. Show that

$$var(\overline{X}_n) = \frac{\sigma^2}{n}\left(1 + 2\sum_{\nu=1}^{n-1}(1 - \nu/n)\,\rho_\nu\right).$$

Hence argue that in the method of batch means, the variance $var(X_k(l))$ does not depend on k.

15.2 Suppose $X_1, X_2, \cdots$ is a covariance stationary process with mean μ and variance σ^2, and let ρ_ν denote the correlation of X_i and $X_{i+\nu}$. Denote by S_n the sample variance of $X_1, X_2, \cdots, X_n$. Show that

$$E(S_n) = \sigma^2\left(1 - \frac{2}{n-1}\sum_{\nu=1}^{n-1}(1 - \nu/n)\,\rho_\nu\right).$$

Hence argue that for positively correlated processes, the sample variance will tend to underestimate the actual variance σ^2. Thus explain why the sample variance $S_{n,l}$ in the method of batch means may tend to underestimate the true variance when the batch means are not independent.

15.3 Apply the runs up and down test to the following data : 22, 14, 24, 19, 16, 21, 23, 18, 20, 22, 17, 15, 21, 12, 11, 14, 25, 23, 22, 19, 16, 23, 25, 20, 21. Is there any evidence of nonrandomness at the 95% level of significance?

SPRING SEMESTER 2008*

DATE	DAY	EVENT
January 17	Thursday	Semester begins
January 19	Saturday	First day of classes
January 21	Monday	Martin Luther King, Jr. Day - CAMPUS CLOSED
February 12	Tuesday	Lincoln's Birthday - Campus open
February 18	Monday	Presidents' Day - CAMPUS CLOSED
March 31-April 6	Monday-Sunday	Spring recess - No classes; Campus open, EXCEPT on Monday, March 31 due to Cesar Chavez Day
March 31	Monday	Cesar Chavez Day - CAMPUS CLOSED
April 7	Monday	Classes resume
May 9	Friday	Last day of classes
May 10-16	Saturday-Friday	Semester examinations
May 17-18	Saturday-Sunday	Commencement Exercises
May 23	Friday	Semester ends; Grade reports due
May 26	Monday	Memorial Day -- CAMPUS CLOSED

* This calendar is subject to change without notice.

2.2 Monday 1/22/2008 notes

Hand out 1/22/2008.

Math 504: Simulation Modelling and Analysis

Text The course is based on notes written by the instructor. However, many of the course topics are covered in the text entitled Introduction to Probability Models, by S. Ross and published by Academic Press. This text is an excellent reference in applied probability.

Math. 438 8th edition.

Instructor	W. B. Gearhart
Office	MH 182F
Phone	714-278-3184
Email	wgearhart@fullerton.edu

Office Hours MW 4-5 pm, MWThF 2-3 pm. If you wish to see me any other time, just let me know, and we will arrange a meeting.

Course Description The course concerns the development and analysis of models of stochastic systems. There are three phases to the course. The first provides an introduction to the theory of stochastic processes. The second concerns modelling discrete event systems using simulation. The software Extend will be used to provide an introduction to the structure and use of a simulation environment. The third and last phase is devoted to further topics in stochastic modelling, and may include statistical aspects of simulation modelling, Brownian motion, signal processing, and Kalman filtering.

Exams every 5-6 weeks apart.

Grading There will be two exams, scheduled approximately every five to six weeks. Also, there will be a comprehensive final exam. Homework will be assigned and graded. The course grade will be based on the weighted average of the homework (5%), the average of the two exams (60%), and the final exam (35%). In cases when a student's calculated percentage is borderline, the instructor may raise the grade based on class participation and attendance, or any other evidence of a strong effort to do the course work.

Grade Scale A: 90-100 B: 80-89 C: 70-79 D: 60-69 F: 0-59

Attendance Class attendance is required. Please arrive on time. If you happen to miss a class, it is your responsibility to obtain from your classmates any missed lecture notes and assignments. However, see the instructor concerning class handouts.

Class Participation In addition to attending class, you are expected to actively participate in your own learning. In particular, you should come to class prepared, having studied the assigned readings and problems, and be ready to ask questions and participate in the class discussion.

Homework Homework due dates will be specified well in advance. Late papers will not be accepted.

Exam Make-up Policy No make-up exams will be given, unless you have a medical emergency or death in the family. These emergencies require valid documentation, and the instructor must be notified within 24 hours of the exam. The grade for a missed exam is zero.

Academic Dishonesty Academic dishonesty is obtaining or attempting to obtain credit for work by the use of any dishonest, deceptive, fraudulent, or unauthorized means. Academic dishonesty also includes helping someone commit an act of academic dishonesty. Examples of academic dishonesty include, but are not limited to:

1. Unacceptable examination behavior - communicating with fellow students, copying material from another student's exam or allowing another student to copy from an exam, possessing or using unauthorized materials, or any behavior that defeats the intent of an exam.

2. Plagiarism - taking the work of another and offering it as one's own without giving credit to that source, whether that material is paraphrased or copied in verbatim or near-verbatim form.

3. Unauthorized collaboration on a project, homework or other assignment where an instructor expressly forbids such collaboration.

4. Documentary falsification including forgery, altering of campus documents or records, tampering with grading procedures, fabricating lab assignments, or altering medical excuses.

Students who violate university standards of academic honesty are subject to disciplinary sanctions, including failure in the course, and suspension from the university. Since dishonesty in any form harms the individual, other students, and the university, policies on academic dishonesty are strictly enforced.

Emergency Information In the event of an emergency such as an earthquake or fire:

1. Take all your personal belongings and leave the classroom (or lab). Use the stairways located at the east, west, or center of the building.

2. Do not use the elevator. They may not be working once the alarm sounds.

3. Go to the lawn area towards Nutwood Avenue. Stay with class members for further instruction.

4. For additional information on exits, fire alarms and telephones, Building Evacuation Maps are located near each elevator.

5. Anyone who may have difficulty evacuating the building, please see the instructor.

Comments

1. Retain this course description and refer to it as needed during the semester.

2. All personal electronic devices, in particular cell phones, must be turned off during class.

3. Keep in mind that grades are not given, they are earned.

4. You are responsible for managing your outside responsibilities (work, family, and social) in order to allow sufficient time to meet the course requirements.

A problem in conditional probability (the first simulation HW, confidence interval, histogram

Handout · 1st lecture 1/22/08

A Problem in Conditional Probability

A number is chosen at random from the interval $[0, 1]$. This value is placed in a box, and twice this value is placed in a second box. One of these boxes is selected at random and opened to reveal the number inside. Given this observed value, what is the probability that this number is the smaller of the two.

1. A Solution Let the random variable X denote the observed number, and let S denote the event that the selected box contains the smaller number. We seek $P(S|X = x)$ for $0 \le x \le 2$. The quantity $P(S|X = x)$ is undefined for other values of x. We will apply Bayes Theorem which gives us

$$P(S|X = x) = \frac{f_X(x|S)P(S)}{f_X(x)} ,$$

where $f_X(x)$ is the density function of the random variable X, and $f_X(x|S)$ is the conditional density of X given the event S. If $1 < x \le 2$, then evidently we have the larger of the two numbers, and so $P(S|X = x) = 0$ when $1 < x \le 2$. Thus, we need consider only the case $0 \le x \le 1$. Since a box is selected at random, $P(S) = 1/2$. Next, the conditional density of X given the event S is just the uniform density on $[0, 1]$. Thus, $f_X(x|S) = 1$ for $0 \le x \le 1$, and $f_X(x|S) = 0$ otherwise. Finally, to determine the density function of the random variable X, we use

$$f_X(x) = f_X(x|S)P(S) + f_X(x|\bar{S})P(\bar{S}) .$$

The conditional density of X given the event $\bar{S}$, is the uniform density on $[0, 2]$. Thus, $f_X(x|\bar{S}) = 1/2$ for $0 \le x \le 2$, and $f_X(x|\bar{S}) = 0$ otherwise. Hence, for $0 \le x \le 1$,

$$f_X(x) = f_X(x|S)P(S) + f_X(x|\bar{S})P(\bar{S}) = 1 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4} ,$$

while for $1 < x \le 2$,

$$f_X(x) = f_X(x|S)P(S) + f_X(x|\bar{S})P(\bar{S}) = 0 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} ,$$

and otherwise, $f_X(x) = 0$. Returning now to the formula for $P(S|X = x)$ we have for $0 \le x \le 1$,

$$P(S|X = x) = \frac{(1)(1/2)}{3/4} = \frac{2}{3} .$$

This result shows that when we select a box at random, and observe a value between 0 and 1, there is a 2/3 chance that the observed value is the smaller of the two.

Exercise

1. Suppose that when we select a box, and observe the value, we have an opportunity to switch to the other box. The result above suggests that if we observe a value between 0 and 1, then we should switch, and otherwise, hold the value we have. Let the random variable Y denote the reward using such a strategy. Write a simulation program (in MATLAB, say) to estimate the expected value of Y. Use a 95% confidence interval, and determine the sample size so that the relative accuracy of your estimate is about one percent. In your report, explain how you determined your sample size. Also, compare theory and practise; that is, did your confidence interval include the true value.

2. **Expected Value of Y** Suppose now the strategy is to switch if the observed value is less than or equal to 1, and otherwise to hold. Let Y be the reward using this strategy. Then

$$E(Y) = E(Y \mid X \leq 1)P(X \leq 1) + E(Y \mid X > 1)P(X > 1) .$$

Consider first the events $\{X \leq 1\}$ and $\{X > 1\}$. In order for the event $\{X > 1\}$ to occur, we must select the box with the larger value, which occurs with probability 1/2, and also the original value must be in the interval $(1/2, 1)$, which occurs with probability 1/2. Since these two events are independent, it follows that $P(X > 1) = (1/2)(1/2) = 1/4$, and further, $P(X \leq 1) = 1 - 1/4 = 3/4$.

Next consider the expected value of Y given that the event $\{X > 1\}$ has occurred. Then Y is the observed value X. Given that the event $\{X > 1\}$ has occurred, the random variable X is uniformly distributed over the interval $(1, 2)$. Hence, $E(Y \mid X > 1) = 3/2$.

Consider now the expected value of Y given that the event $\{X \leq 1\}$ has occurred. Here, we will switch to the value in the other box. However, we will either (α) switch to the larger value, which occurs with probability $2/3$, or (β) switch to the smaller value, which occurs with probability $1 - 2/3 = 1/3$. In case (α), Y is the larger value, which is uniformly distributed over the interval $(0, 2)$. Hence its expected value is 1, and so $E(Y|\alpha) = 1$. In case (β), Y is the smaller value, which now, because the event $\{X \leq 1\}$ has taken place, is uniformly distributed over the interval $(0, 1/2)$. Hence, the expected value is $1/4$, and so $E(Y|\beta) = 1/4$. Thus,

$$E(Y \mid X \leq 1) = E(Y|\alpha)P(\alpha) + E(Y|\beta)P(\beta) = 1 \cdot \frac{2}{3} + \frac{1}{4} \cdot \frac{1}{3} = \frac{3}{4}\,.$$

We are ready finally to compute the expected value $E(Y)$. From the formula above, we get

$$E(Y) = E(Y|X \leq 1)P(X \leq 1) + E(Y|X > 1)P(X > 1) = \frac{3}{4} \cdot \frac{3}{4} + \frac{3}{2} \cdot \frac{1}{4} = \boxed{\frac{15}{16}}\,.$$

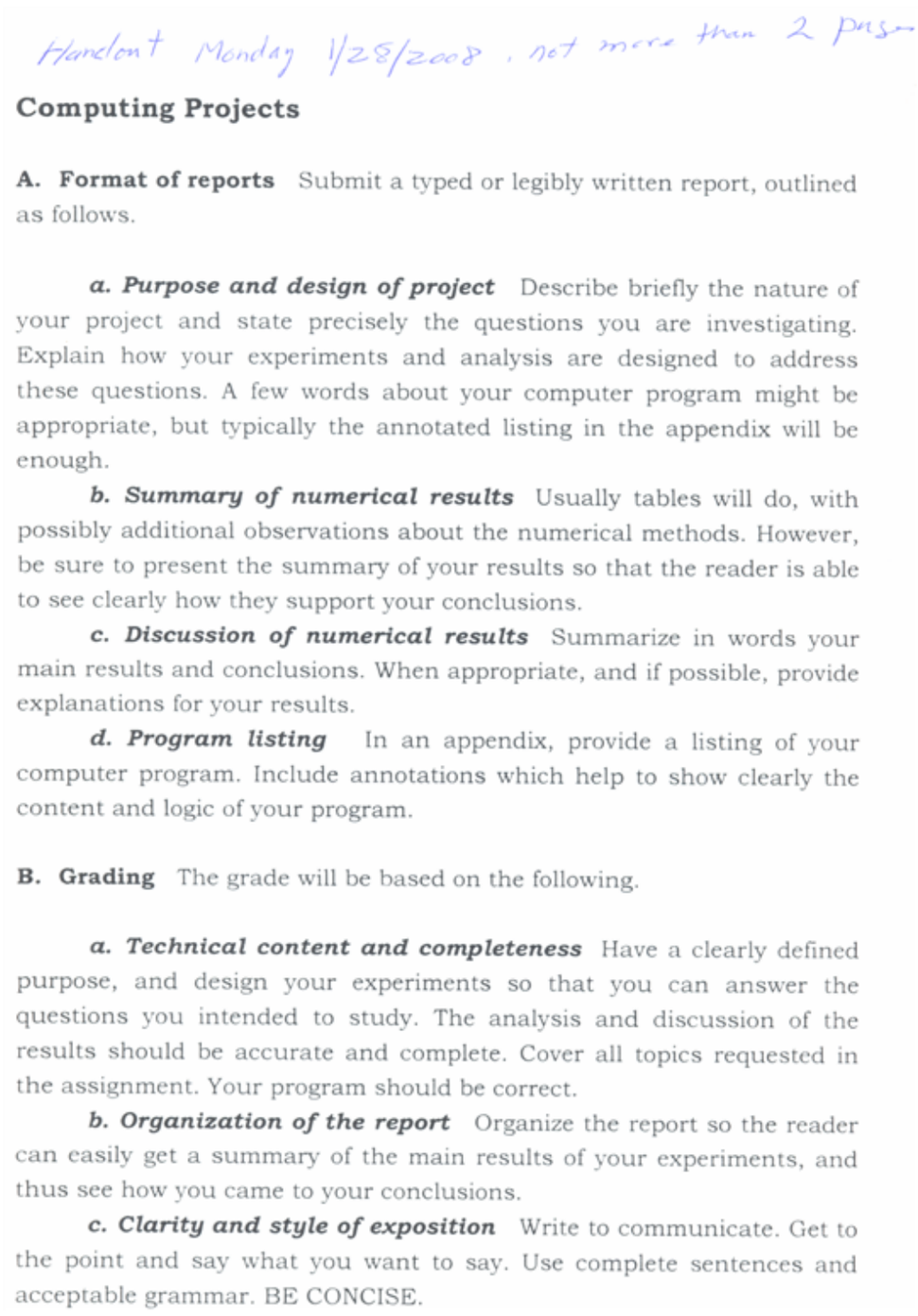
Exercise

1. Find the density function of the random variable Y.

try to do this $f_Y(y) = f_Y(y|S)P(S) + f_Y(y|\bar{S})P(\bar{S})$

2.3 Monday 1/28/08 notes

Computing project guideline



Handout Monday 1/28/2008. not more than 2 pages

Computing Projects

A. Format of reports Submit a typed or legibly written report, outlined as follows.

a. Purpose and design of project Describe briefly the nature of your project and state precisely the questions you are investigating. Explain how your experiments and analysis are designed to address these questions. A few words about your computer program might be appropriate, but typically the annotated listing in the appendix will be enough.

b. Summary of numerical results Usually tables will do, with possibly additional observations about the numerical methods. However, be sure to present the summary of your results so that the reader is able to see clearly how they support your conclusions.

c. Discussion of numerical results Summarize in words your main results and conclusions. When appropriate, and if possible, provide explanations for your results.

d. Program listing In an appendix, provide a listing of your computer program. Include annotations which help to show clearly the content and logic of your program.

B. Grading The grade will be based on the following.

a. Technical content and completeness Have a clearly defined purpose, and design your experiments so that you can answer the questions you intended to study. The analysis and discussion of the results should be accurate and complete. Cover all topics requested in the assignment. Your program should be correct.

b. Organization of the report Organize the report so the reader can easily get a summary of the main results of your experiments, and thus see how you came to your conclusions.

c. Clarity and style of exposition Write to communicate. Get to the point and say what you want to say. Use complete sentences and acceptable grammar. BE CONCISE.

2.4 Monday 1/28/08 notes

Continuous approximation to random walks

Hand out Math 504
Monday 1/28/2008

Continuous Approximations To Random Walks

1. A Simple Random Walk Consider a particle that moves along the real line in such a way that, at each point in time, it makes one step to the right with probability p, and one step to the left with probability $q = 1 - p$. Assume that distance is measured in multiples of an amount Δx, and that time is measured in multiples of Δt. Let X_n denote the position of the particle after n steps. We shall assume the particle starts at the origin.

Let $\pi_j^{(n)} = P(X_n = j\Delta x)$. Thus, $\pi_j^{(n)}$ is the probability that the particle is located at $j\Delta x$ at time $n\Delta t$. Conditioning on the next state, we can write

$$\begin{aligned} P(X_{n+1} = j\Delta x) &= P(X_{n+1} = j\Delta x \mid X_n = (j-1)\Delta x)P(X_n = (j-1)\Delta x) \\ &+ P(X_{n+1} = j\Delta x \mid X_n = (j+1)\Delta x)P(X_n = (j+1)\Delta x) \ , \end{aligned}$$

or equivalently

$$\pi_j^{(n+1)} = p\pi_{j-1}^{(n)} + q\pi_{j+1}^{(n)} \ ,$$

for $j = 0, \pm 1, \pm 2, \cdots$, and $n = 0, 1, \cdots$. Consider now a fixed position x and time t, subject to $x = j\Delta x$ and $t = n\Delta t$. Suppose that when Δx and Δt are small, we have the approximation

$$\pi_j^{(n)} \approx f(x,t)\Delta x \ ,$$

where f is some function of x and t. Note that for each fixed t, the function f is a density function that describes probabilistically the location of the particle. From the recurrence formula above, we see that for such an approximation to hold, we need approximately

$$f(x, t + \Delta t) \cong pf(x - \Delta x, t) + qf(x + \Delta x, t) \ .$$

Assuming f is twice continuously differentiable, the Taylor series expansion yields

$$\begin{aligned} f(x,t) + \Delta t\frac{\partial f}{\partial t} + O(\Delta t)^2 &= p\left[f(x,t) - \Delta x\frac{\partial f}{\partial x} + \frac{1}{2}(\Delta x)^2\frac{\partial^2 f}{\partial x^2}\right] \\ &+ q\left[f(x,t) + \Delta x\frac{\partial f}{\partial x} + \frac{1}{2}(\Delta x)^2\frac{\partial^2 f}{\partial x^2}\right] + O(\Delta x)^3 \ , \end{aligned}$$

which, upon simplification, gives us

$$\frac{\partial f}{\partial t} = -(p-q)\frac{\Delta x}{\Delta t}\frac{\partial f}{\partial x} + \frac{1}{2}\frac{(\Delta x)^2}{\Delta t}\frac{\partial^2 f}{\partial x^2} + \frac{O(\Delta x)^3}{\Delta t} + O(\Delta t) \ .$$

In order to obtain a limiting equation, assume that for small Δx and Δt, there are constants β and D such that approximately,

$$(p-q)\frac{\Delta x}{\Delta t} = \beta \quad \text{and} \quad \frac{1}{2}\frac{(\Delta x)^2}{\Delta t} = D \ .$$

Recalling that $q = 1 - p$, these approximations tell us that in the limit we need

$$p = \frac{1}{2}\left(1 + \frac{\beta\Delta x}{2D}\right) \quad \text{and} \quad q = \frac{1}{2}\left(1 - \frac{\beta\Delta x}{2D}\right) .$$

Going to the limit then, we arrive at the partial differential equation

$$\frac{\partial f}{\partial t} = -\beta\frac{\partial f}{\partial x} + D\frac{\partial^2 f}{\partial x^2} .$$

Using the Fourier transform, this equation can be solved to get

$$f(x,t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] , \text{ where } \mu = \beta t , \text{ and } \sigma = \sqrt{2Dt} . \qquad (1)$$

The steps of this method are outlined in Problem 3 below.

There is an alternative way to reach this conclusion. After n steps, the particle will have made a certain number of steps to the right, say R_n, and a certain number of steps to the left, say L_n. Then $R_n + L_n = n$ and $X_n = (R_n - L_n)\Delta x$. Hence, $X_n = (2R_n - n)\Delta x$. Note that R_n is a binomial random variable with parameters n and p. Thus, the central limit theorem tells us that the distribution of R_n, and hence of X_n, is approximately normal. Further, the mean of position is

$$E(X_n) = (2E(R_n) - n)\Delta x = (2(np) - n)\Delta x = n\Delta x(2p-1) = n\frac{\beta(\Delta x)^2}{2D} ,$$

where we have used the formula for p above. Hence, the mean of position is

$$\left(n\frac{\beta\Delta x}{2D}\right)\Delta x = (n\Delta t)\frac{\beta(\Delta x)^2}{2D\Delta t} .$$

However, in the limit as n gets larger, we have $t = n\Delta t$ and $(\Delta x)^2/\Delta t = 2D$. Thus, the mean of position is simply $\mu = \beta t$.

Next, continuing this line of reasoning, we argue that the variance of position is

$$Var(X_n) = Var((2R_n - n))\Delta x) = Var(2R_n\Delta x) = 4(\Delta x)^2 Var(R_n) .$$

But the variance of R_n is npq. Hence, the variance of position is

$$4(\Delta x)^2 npq = 4(\Delta x)^2 n\frac{1}{2}\left(1 + \frac{\beta\Delta x}{2D}\right)\frac{1}{2}\left(1 - \frac{\beta\Delta x}{2D}\right) = \frac{(\Delta x)^2}{\Delta t}(n\Delta t)\left[1 - \left(\frac{\beta\Delta x}{2D}\right)^2\right] ,$$

where we have used the formulas for p and q above. Finally, going to the limit, and noting again that $t = n\Delta t$ and $2D = (\Delta x)^2/\Delta t$, we conclude that the variance of position is $\sigma^2 = 2Dt$. Thus, in the limit, the distribution of position is approximately normal with

mean $\mu = \beta t$ and variance $\sigma = \sqrt{2Dt}$. This result is the same as the one we obtained earlier through use of the partial differential equation.

A **stochastic process** is a family of random variables $X(t)$, where $X(t)$ represents the state of the process at time t. In our case, the state of the process is the position of the particle along the real line. A process is said to have **stationary increments** if for any t, the distribution of the **increment** $X(s+t) - X(s)$ depends only on t, the length of the time interval. Further, if the increments are independent for any set of disjoint intervals, the process is said to have **independent increments**.

Since the continuous process above, with transition distribution function (1), was derived as the limit of a discrete process that has stationary and independent increments, it is reasonable to expect that the limiting process would also have these two properties. A continuous process with transition distribution function (1), is called an **Einstein-Wiener process**. The parameter β is called the drift coefficient, and the parameter D is called the diffusion coefficient.

2. The Ornstein-Ehrenfest Model For a positive integer a, consider a random walk in which, at each point in time, if the process is at position $j\Delta x$, it moves one step to the right with probability $(a-j)/2a$ and one step to the left with probability $(a+j)/2a$, when $-a < j < a$. If $j = a$ then it moves to the left with probability 1, and if $j = -a$ it moves to the right with probability one.

Denote by $\pi_j^{(n)}$ the probability that the process is at point $j\Delta x$ at time $n\Delta t$. Then, by conditioning on the next state, we can write

$$\pi_j^{(n+1)} = \frac{a-j+1}{2a}\pi_{j-1}^{(n)} + \frac{a+j+1}{2a}\pi_{j+1}^{(n)} ,$$

for each $j = 1, 2, \cdots$, and $n = 0, 1, \cdots$. Consider this process in the limit when the bound a is large, and the Δx and Δt are small. For a fixed position x and time t, subject to $x = j\Delta x$ and $t = n\Delta t$, suppose we have the approximation

$$\pi_j^{(n)} \approx f(x,t)\Delta x ,$$

where f is some function of x and t. From the recurrence formula above, we see that for such an approximation to hold, we need approximately

$$f(x, t+\Delta t) = \frac{a-j+1}{2a} f(x-\Delta x, t) + \frac{a+j+1}{2a} f(x+\Delta x, t) .$$

Assuming f is twice continuously differentiable, the Taylor series expansion yields

$$\begin{aligned} f(x,t) + \Delta t \frac{\partial f}{\partial t} + O(\Delta t)^2 &= \frac{a-j+1}{2a}\left[f(x,t) - \Delta x \frac{\partial f}{\partial x} + \frac{1}{2}(\Delta x)^2 \frac{\partial^2 f}{\partial x^2} \right] \\ &+ \frac{a+j+1}{2a}\left[f(x,t) + \Delta x \frac{\partial f}{\partial x} + \frac{1}{2}(\Delta x)^2 \frac{\partial^2 f}{\partial x^2} \right] + O(\Delta x)^3 , \end{aligned}$$

which, upon simplification, gives

$$\frac{\partial f}{\partial t} = \frac{1}{a\Delta t} f(x,t) + \frac{x}{a\Delta t}\frac{\partial f}{\partial x} + \frac{a+1}{a}\frac{(\Delta x)^2}{2\Delta t}\frac{\partial^2 f}{\partial x^2} + \frac{O(\Delta x)^3}{\Delta t} + O(\Delta t) .$$

In order to obtain a limiting equation, assume that for small Δx and Δt, and large a, we have approximately,

$$a\Delta t = c^{-1} \quad \text{and} \quad \frac{1}{2}\frac{(\Delta x)^2}{\Delta t} = D ,$$

for some constants c and D. Going to the limit then, we arrive at the partial differential equation

$$\frac{\partial f}{\partial t} = c\frac{\partial (xf)}{\partial x} + D\frac{\partial^2 f}{\partial x^2} .$$

This equation is not so easily solved as in the previous case. However, under the boundary conditions

$$xf(x,t) \to 0 , \quad \frac{\partial f(x,t)}{\partial x} \to 0 \ \text{ as } \ x \to \pm\infty ,$$

and using the Fourier transform, the equation can be transformed into a first order, variable coefficient hyperbolic equation. This hyperbolic equation can then be solved using the method of characteristics. This method of solution is outlined in Problem 4 below.

Suppose the particle starts at a point x_0. Thus, the initial density of position is a dirac-delta function centered at the point x_0. Then the solution is found to be

$$f(x,t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma} \right)^2 \right] ,$$

where $\mu = x_0 e^{-ct}$ and

$$\sigma^2 = \frac{D}{c}\left(1 - e^{-2ct} \right) .$$

A continuous process with stationary and independent increments, and having this transitional distribution function, is called an **Ornstein-Ehrenfest process**.

Exercises

1. Derive the Einstein-Wiener process by noting that the position of the particle is $x = j\Delta x$, where $j = X_1 + X_2 + \cdots X_n$ with $t = n\Delta t$, and the X_i are independent and identically distributed random variables which have value $+1$ with probability p, and value -1 with probability $q = 1 - p$. Take $p = q = 1/2$.

2. (a) Use the formulation in the previous exercise to simulate the random walk for $p = q = 1/2$, and a specified diffusion coefficient D. Restrict Δx and Δt so that $D = (\Delta x)^2/2\Delta t$. (b) Use the simulation model to test that in the limit as $\Delta x \to 0$ and $\Delta t \to 0$, subject to $D = (\Delta x)^2/2\Delta t$, the distribution of position, for fixed time t and given D, is normal with mean 0 and variance $\sigma^2 = 2Dt$.

3. Solve the partial differential equation

$$\frac{\partial f}{\partial t} = -\beta \frac{\partial f}{\partial x} + D\frac{\partial^2 f}{\partial x^2} \ .$$

for the Einstein-Wiener process. Use the Fourier transform, and the following steps. (a) The Fourier transform of an absolutely integrable, and piecewise continuous function g on $(-\infty, \infty)$, is defined by

$$F[g](y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(x) e^{-iyx} dx \ .$$

Multiply the differential equation through by e^{-iyx} and integrate with respect to x over the interval $(-\infty, \infty)$. Then use integration by parts twice to obtain the equation

$$\frac{\partial \phi}{\partial t} = \left(-i\beta y - Dy^2\right) \phi(y,t) \ ,$$

where

$$\phi(y,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x,t) e^{-iyx} dx \ .$$

In doing the integrations by parts, assume that

$$f(x,t) \to 0 \ , \quad \frac{\partial f(x,t)}{\partial x} \to 0 \ \text{ as } \ x \to \pm\infty \ .$$

(b) Solve the differential equation for ϕ to get

$$\phi(y,t) = \phi(y,0) \exp\left(-i\beta t y - Dty^2\right) \ .$$

Note that $\phi(y,0)$ is the Fourier transform of $f(x,0)$, the density of the location of the particle at time $t = 0$.

(c) Now make use of the following two properties of the Fourier transform:

$$(1) \quad \text{For} \quad f(x) = e^{-x^2} \ , \quad F[f](y) = \frac{1}{\sqrt{2}} e^{-y^2/4} \ ,$$

and

$$(2) \quad \text{For} \quad g(x) = f(ax+b) \ , \ a \neq 0 \ , \quad F[g](y) = \frac{1}{|a|} e^{iy(b/a)} F[f](y/a) \ .$$

Thus conclude that

$$\text{For} \quad g(x) = e^{-(ax+b)^2} \ , \ a \neq 0 \ , \quad F[g](y) = \frac{1}{\sqrt{2}|a|} e^{iy(b/a)} e^{-y^2/4a^2} \ .$$

(d) Set $b/a = -\beta t$ and $1/4a^2 = Dt$, and use the previous results to conclude that

$$\phi(y,t) = \phi(y,0) \exp\left[\left(-i\beta y - Dy^2\right) t\right] = \phi(y,0) F[h(\cdot,t)](y) \ ,$$

where

$$h(x,t) = \frac{1}{\sigma(t)} \exp\left[-\frac{1}{2}\left(\frac{x-\mu(t)}{\sigma(t)}\right)^2\right] , \quad \mu(t) = \beta t \ , \quad \sigma(t) = \sqrt{2Dt} \ .$$

(e) Finally, the Fourier transform has the property that

$$F[f * g] = \sqrt{2\pi} F[f] F[g] \ , \quad \text{where} \quad f * g(x) = \int_{-\infty}^{\infty} f(u) g(x-u) du \ .$$

Use this property to show that

$$f(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(u,0) h(x-u,t) du \ .$$

In particular, if the particle starts that the origin, then we can view the initial density of position, $f(x,0)$, as a dirac delta function and thus deduce that in this case

$$f(x,t) = \frac{1}{\sqrt{2\pi}} h(x,t) \ ,$$

which is the expression (1) above.

4. Solve the partial differential equation

$$\frac{\partial f}{\partial t} = c \frac{\partial (xf)}{\partial x} + D \frac{\partial^2 f}{\partial x^2} \ .$$

for the Ornstein-Ehrenfest process. Use the Fourier transform (see definition in previous problem), and the following steps. (a) Multiply the differential equation

through by e^{-iyx} and integrate with respect to x over the interval $(-\infty, \infty)$. Then use integration by parts three times to obtain the equation

$$\frac{\partial \phi}{\partial t} = -cy\frac{\partial \phi}{\partial y} - Dy^2\phi \ ,$$

where

$$\phi(y,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x,t)e^{-iyx}dx \ .$$

In doing the integrations by parts, assume that

$$xf(x,t) \to 0 \ , \quad \text{and} \quad \frac{\partial f(x,t)}{\partial x} \to 0 \ , \quad \text{as} \ \ x \to \pm\infty \ .$$

(b) Introduce the integrating factor

$$I(y) = \exp\left(\frac{D}{2c}y^2\right)$$

and show that $u(y,t) = I(y)\phi(y,t)$ satisfies the hyperbolic equation

$$\frac{\partial u}{\partial t} + cy\frac{\partial u}{\partial y} = 0 \ ,$$

subject to the initial condition $u(y,0) = u_0(y) = I(y)\phi(y,0)$ for $-\infty < y < \infty$. Note that $\phi(y,0)$ is the Fourier transform of $f(x,0)$, the density of the location of the particle at time $t = 0$.

(c) Use the method of characteristics to solve the previous hyperbolic equation and conclude that

$$u(y,t) = u_0\left(ye^{-ct}\right) \ , \quad \text{for} \ -\infty < y < \infty \ \text{ and } \ t \geq 0 \ .$$

Then, using the definitions of $u(y,t)$ and $u_0(y)$, deduce that

$$\phi(y,t) = \exp\left(-\frac{1}{2}\sigma^2(t)y^2\right)\phi\left(ye^{-ct}, 0\right) \ , \quad \text{where} \ \ \sigma^2(t) = \frac{D}{c}\left(1 - e^{-2ct}\right) \ .$$

(d) To simplify the analysis, assume that the particle starts at a point x_0. Thus, the initial density of position is a dirac-delta function centered at the point x_0. It follows that

$$\phi(y,0) = \frac{1}{\sqrt{2\pi}}e^{-iyx_0} \ .$$

In this case, use the previous result to conclude that

$$\phi(y,t) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\sigma^2(t)y^2 - iy\mu(t)\right) \ , \quad \text{where} \ \ \mu(t) = x_0e^{-ct} \ .$$

(e) Set $b/a = -\mu(t)$ and $1/4a^2 = (1/2)\sigma^2(t)$, and use the result of part (c), Problem 3, to conclude that

$$\phi(y,t) = F[r(\cdot,t)](y) \ ,$$

where

$$r(x,t) = \frac{1}{\sigma(t)\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu(t)}{\sigma(t)}\right)^2\right] \ .$$

Thus, $f(x,t) = r(x,t)$, which is the solution given above for the Ornstein-Ehrenfest model.

2.5 Monday 2/25/08 notes

Problems to practice solving first order pde using the characteristics method.

Method of Characteristics - Practice Problems

Use the method of characteristics to solve the following initial value problems for $t > 0$ and $-\infty < x < \infty$.

1. $u_t + 4u_x = 0$, and $u(x,0) = e^{x^2}$.
2. $u_t + (xt)u_x = 0$, and $u(x,0) = 2x$.
3. $u_t + (x \sin t)u_x = 0$, and $u(x,0) = 1/(1+x^2)$.
4. $u_t - (tx^2)\, u_x = 0$, and $u(x,0) = 1 + x$.
5. $u_t - u_x = xu$, and $u(x,0) = 2x$.

1

2.6 Monday notes

Craps game and inventory problem. Markov chain computing assignment.

Computer.

Math 504 Assignment 2 3

a week from next monday.
March 3/3/2008

1. Consider the game of "craps": Two die are tossed. The player wins if a sum of 7 or 11 appears, and loses if a sum of 2,3, or 12 appears. Otherwise, the sum on the first toss is designated the player's point score. The player then continues to toss the two die until a sum of 7 appears, in which case the player loses, or a sum equal to the point score appears, in which case the player wins. Model this game as an absorbing Markov chain. Determine the one-step transition matrix P.

2. Write a program to form the one-step transition matrix P for the inventory problem in Example 4.3.2. For the demand, use a function of the form
$$d(x) = c(x+1) \text{ , for } x = 0, 1, 2, \cdots, 5 \text{ ,}$$
and
$$d(x) = 6c(13-x)/8 \text{ , for } x = 6, 7, \cdots, 13 \text{ .}$$
Choose the constant c so that the function d is a probability distribution. Inputs to your program are the values of s and S. Output is the one-step transition matrix P.

Experiment with each of these processes, and investigate what can be said about the long-run behavior. In particular, consider the following questions. (a) Do the powers P^n converge as $n \to \infty$? If so, what can you say about the limit matrix? (b) Is there a limiting state probability distribution in all cases? If so, does this limiting distribution depend on the initial state probability distribution? (c) Assuming your conclusions from parts (a) and (b) are correct, show how the results of part (a) could be used to deduce the conclusions of part (b).

5 pts

In your report, first summarize your numerical results briefly and succinctly. Present these results in such a way that the reader can easily understand the observations that you are drawing from your experiments. Next state your observations based on the numerical results, and indicate any general conclusions that seem to be suggested by the experiments.

2.7 Monday 3/10/2008 notes

Hand out. March 10, 2008 Math 504

Convergent Finite Markov Chains

1 Introduction

Consider a finite state Markov chain with one-step probability transition matrix P and state probability distribution vector $\pi^{(n)}$ at time $n \geq 0$. Then

$$\pi^{(n+1)} = \pi^{(n)} P \ , \quad n = 0, 1, \cdots \ .$$

A fundamental question is whether or not the process approaches a limit in the long-run. In other words, given an arbitrary initial state probability distribution $\pi^{(0)}$, do the state probability distributions $\pi^{(n)}$ converge as $n \to \infty$. Since

$$\pi^{(n)} = \pi^{(0)} P^n \ ,$$

for each $n \geq 0$, and since $\pi^{(0)}$ is arbitrary, this question is equivalent to asking whether the powers of the transition matrix converge as $n \to \infty$. A Markov chain with transition matrix P will be called **convergent** if P^n converges as $n \to \infty$. In this case, we shall also refer to the transition matrix as being convergent.

2 Structure and Properties of Finite Markov Chains

The study of Markov chains hinges on the notion of recurrent and transient states. For a state i, let T_{ii} denote the time until the process first returns to state i, given that it starts in state i. Then state i is said to be **recurrent** if $P(T_{ii} < \infty) = 1$, and otherwise, if $P(T_{ii} < \infty) < 1$, state i is said to be **transient**. Thus, a state is transient if the process, having started in that state and perhaps having returned to that state a number of times, will eventually leave that state forever. It can be shown that if states i and j are transient then

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty \ .$$

This expression is the expected number of visits to state j, given that the process starts in state i. Thus, for transient states i and j, we have

$$p_{ij}^{(n)} \to 0 \,, \quad \text{as} \quad n \to \infty \,. \tag{1}$$

One important consequence of this result is that not all states of a finite chain can be transient.

The process moves among the states randomly, according to the transition probabilities. Two states i and j are said to **communicate** if

$$p_{ij}^{(n)} > 0 \,, \quad \text{and} \quad p_{ji}^{(m)} > 0 \,,$$

for some n and m. In other words, two states communicate if it is possible to travel from one state to the other and back again. If state n is recurrent, and it communicates with state m, then state m must be recurrent also.

A set of states S is **closed** if $p_{ij} = 0$ whenever $i \in S$ and $j \notin S$. Note that if the process enters a closed set, then it will never leave that set. A chain is said to be **irreducible** if there is no proper closed subset. Otherwise, the chain is called **reducible**. For a reducible chain, after possibly re-ordering the states, the transition matrix can be written in the form

$$P = \begin{bmatrix} P_1 & 0 \\ P_2 & P_3 \end{bmatrix} .$$

The matrix P_1 is the transition matrix for the sub-chain consisting of a closed set of states.

It follows from these concepts that for an arbitrary finite Markov chain, the probability transition matrix, again after possibly re-ordering the states, can be written in the so-called **canonical form**:

$$P = \begin{bmatrix} D & 0 \\ R & Q \end{bmatrix} , \tag{2}$$

where the matrix D is block diagonal

$$D = \begin{bmatrix} D_1 & 0 & \cdots & 0 \\ 0 & D_2 & 0 & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & D_k \end{bmatrix} . \tag{3}$$

To see how this form comes about, start with a recurrent state, and find all states that communicate with it. These states form an irreducible, closed set. Let the matrix consisting of the one-step transition probabilities among these states form the first block D_1. Next, find a recurrent state (if any) that is not in the set obtained previously, and determine all states that communicate with it. This set of states forms another irreducible, closed set. Let the matrix consisting of the one-step transition probabilities among these states form the second block D_2. Continue in this fashion until no recurrent states are left. The remaining states are transient, and the matrix consisting of the one-step transition probabilities among these states is the matrix Q. The matrix R in (2) consists of the one-step transition probabilities from transient states to recurrent states. Finally, the matrix of zeros in (2) appears in the upper righthand corner, since it is not possible to go from a recurrent state to a transient state.

The canonical form reveals much about the behavior of the chain. If the process starts in an irreducible class, it will stay there and ultimately approach the limiting behavior of that class, if any. On the other hand, if the process starts in a transient state, it will ultimately move into one of the irreducible classes. In fact, as indicated by (1), $Q^n \to 0$ as $n \to \infty$, since the (i, j)-th entry of this matrix is the probability that the process is in transient state j, after n steps, given that it started in a transient state i. Further, entry (k, s) of the matrix $Q^n R$ is the probability that when the process starts in transient state k, it will enter for the first time, one of the irreducible sub-chains in $n+1$ steps, and will do so at state s. Summing over n, it follows that entry (k, s) of the matrix

$$R + QR + Q^2R + \cdots = (I + Q + Q^2 + \cdots)R\,,$$

is the probability that when the process starts in transient state k, it will ultimately enter the recurrent classes for the first time at state s. Using the Neumann expansion $N = (I - Q)^{-1} = I + Q + Q^2 + \cdots$, we can write this matrix simply as NR.

Example Suppose the one-step transition matrix of a Markov chain is

given by

$$P = \begin{bmatrix} 0.800 & 0.200 & 0 & 0 & 0 & 0 & 0 \\ 0.300 & 0.700 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.00 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.400 & 0.600 & 0 & 0 \\ 0 & 0 & 0 & 0.200 & 0.800 & 0 & 0 \\ 0.100 & 0.200 & 0.050 & 0.150 & 0.200 & 0.100 & 0.200 \\ 0.300 & 0.100 & 0.200 & 0.100 & 0.050 & 0.100 & 0.150 \end{bmatrix}.$$

The states have been numbered so that the matrix is in canonical form. The first recurrence class consists of states 1 and 2, the second recurrence class consists of state 3 only, and is absorbing, while the third and last recurrence class consists of states 4 and 5. States 6 and 7 are transient. The matrices R and Q are

$$R = \begin{bmatrix} 0.100 & 0.200 & 0.050 & 0.150 & 0.200 \\ 0.300 & 0.100 & 0.200 & 0.100 & 0.050 \end{bmatrix}, \quad Q = \begin{bmatrix} 0.150 & 0.200 \\ 0.100 & 0.050 \end{bmatrix}.$$

Thus

$$NR = (I - Q)^{-1}R = \begin{bmatrix} 0.195 & 0.255 & 0.111 & 0.198 & 0.242 \\ 0.376 & 0.148 & 0.248 & 0.141 & 0.087 \end{bmatrix}.$$

Consider the first row of this matrix. The first entry shows that 19.5% of the entities that start in state 6, will enter the recurrent classes for the first time at state 1. Similarly, the second entry shows that 25.5% of the entities that start in state 6, will enter the recurrent classes for the first time at state 2. Thus, 19.5%+25.5% = 45% of the entities that start in state 6 will ultimately enter the first recurrence class. These entities however, once having entered this recurrent class, must then circulate among the states and ultimately become distributed according to limiting behavior of the class. The third entry in the first row shows that 11.1% of the entities that start in state 6 will enter the recurrent classes for the first time at state 3. Since this state is absorbing, the limiting behavior is clear. The entities that enter this state just stay there. A similar analysis can be done for entries 4 and 5 in this row.

In general, we will be able to determine the limiting distributions of the recurrent classes by finding the limit P^n as $n \to \infty$. In this case, the limit

matrix is

$$W = \begin{bmatrix} 0.600 & 0.400 & 0 & 0 & 0 & 0 & 0 \\ 0.600 & 0.400 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.00 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.250 & 0.750 & 0 & 0 \\ 0 & 0 & 0 & 0.250 & 0.750 & 0 & 0 \\ 0.270 & 0.180 & 0.110 & 0.110 & 0.330 & 0 & 0 \\ 0.314 & 0.210 & 0.248 & 0.057 & 0.171 & 0 & 0 \end{bmatrix}.$$

Consider the 6-th row of this matrix. The first two entries show that 27% of the entities that start in state 6 ultimately go to state 1, and that 18% ultimately go to state 2. Thus, $27\% + 18\% = 45\%$ of the entities that start in state 6 ultimately move to the first recurrence class. We saw this result earlier using the matrix NR. From the 2×2 matrix in the upper left corner, we see that in the long-run, 60% of the entities in the first recurrence class will be in state 1, and 40% will be in state 2. In particular, the entities that reach this recurrence class from state 6 are proportioned this way: the fraction 0.270 is 60% of the total fraction 0.45, while the fraction 0.180 is 40% of the total fraction 0.45.

A similar analysis applies to the entities that go from state 6 to the third recurrence class. Indeed, the fourth and fifth entries show that 11% of the entities that start in state 6 ultimately go to state 4, and that 33% ultimately go to state 5. Thus, $11\% + 33\% = 44\%$ of the entities that start in state 6, ultimately move to the third recurrence class. From the 2×2 matrix in the middle of W, we see that in the long-run, 25% of the entities in the third recurrence class will be in state 4, and 75% will be in state 5. In particular, the entities that reach this recurrence class from state 6 are proportioned this way: the fraction 0.11 is 25% of the total fraction 0.44, while the fraction 0.33 is 75% of the total fraction 0.44.

So far, we have accounted for $45\% + 44\% = 89\%$ of the entities that start in transient state 6. But entry $(6, 3)$ of W shows that the remaining 11% go to state 3 and are absorbed. Thus, our accounting of the entities that start in transient state 6 is complete. The same analysis can be used to account for the entities that start in transient state 7.

2.8 Monday 3/17/2008 notes

handout Solution 3/17/08

Chapter 5: Some Solutions

5.7 Consider an irreducible Markov chain with a finite number of states $\{0, 1, 2, \cdots, m\}$. Let E denote the event that the process reaches state m before it reaches state 0. Set $Q_i = P(E \,|\, X_0 = i)$. Then $Q_0 = 0$ and $Q_m = 1$. (*a*) Find a system of $m - 1$ linear equations that is satisfied by $Q_1, Q_2, \cdots, Q_{m-1}$. (*b*) Show that the matrix in this linear system of equations is nonsingular.

Solution (*a*) Conditioning on the next state gives us

$$Q_i = \sum_{j=0}^{m} P(E \,|\, X_1 = j) p_{ij} = \sum_{j=1}^{m-1} Q_j p_{ij} + p_{im} \,,$$

for $i \in I = \{1, 2, \cdots, m-1\}$. Note that the Markov property was used to get the second equation. (*b*) If the matrix in this linear system of equations is singular, then there is a nonzero vector $v \in R^{m-1}$ such that

$$v_i = \sum_{j=1}^{m-1} v_j p_{ij} \,,$$

for $i \in I$. Normalize v so that each component is less than or equal to one, and at least one component is equal to one. Set $J = \{i \in I \,|\, v_i = 1\}$ and $S = \{i \in I \,|\, v_i < 1\}$.

Suppose first that $S \neq \emptyset$. Assume there is a nonzero p_{ir} for some $i \in J$ and some $r \in S$. It then follows that

$$1 = v_i = \sum_{j \in J} v_j p_{ij} + \sum_{j \in S} v_j p_{ij} < \sum_{j=1}^{m-1} p_{ij} \leq 1 \,,$$

which is impossible. Therefore, for any $i \in J$, and any $r \in S$, we must have $p_{ir} = 0$. But then, for each $i \in J$,

$$1 = v_i = \sum_{j \in J} v_j p_{ij} = \sum_{j \in J} p_{ij} \,.$$

This result tells us that once the process enters a state $i \in J$, it must travel to another state in J. Thus, in the original chain, the set of states J, which is not empty, must be a closed set. However, this result is impossible, since the original chain is irreducible. Therefore, $S = \emptyset$.

It now follows that $J = \{1, 2, \cdots, m-1\}$. But then, for each $i \in J$,

$$1 = v_i = \sum_{j \in J} v_j p_{ij} = \sum_{j \in J} p_{ij} \ ,$$

since $v_j = 1$ for each $j \in J$. Thus, as before, the set J is seen to be a closed set, which is a contradiction.

2.9 Wed 3/19/2008 notes

Handout March 19, 2008

Chapter 6 Some Solutions

6.3 For an absorbing Markov chain, let V_{ij} denote the number of visits to transient state j before absorption, given that the process starts in transient state i. Let B denote the matrix whose (i,j)-th entry is $b_{ij} = E(V_{ij})$. (*a*) Show that $B = N = (I - Q)^{-1}$. (*b*) Given that the process starts in state i, give a formula for the expected number of steps until absorption.

Solution (*a*) Condition on the next state to get

$$E(V_{ij}) = \sum_{k \in A} E(V_{ij} \mid X_1 = k) p_{ik} + \sum_{k \in T} E(V_{ij} \mid X_1 = k) p_{ik}$$

where A denotes the set of absorbing states, and T denotes the set of transient states. Let δ_{ij} be the Kronecker delta, which equals one when $i = j$, and equals zero otherwise. Then, in the first sum, $E(V_{ij} \mid X_1 = k) = \delta_{ij}$, since k is an absorbing state. For the second sum, making use of the Markov property, $E(V_{ij} \mid X_1 = k) = \delta_{ij} + E(V_{kj})$, since here k is a transient state. Thus,

$$E(V_{ij}) = \sum_{k \in A} \delta_{ij} p_{ik} + \sum_{k \in T} [\delta_{ij} + E(V_{kj})] p_{ik} = \delta_{ij} + \sum_{k \in T} p_{ik} E(V_{kj}) \; .$$

Therefore, for each pair of states i and j in T,

$$b_{ij} - \sum_{k \in T} p_{ik} b_{kj} = \delta_{ij} \; .$$

In matrix form, these equations are expressed as $B - QB = I$, where I is the identity matrix. Note that this result shows that $I - Q$ is invertible, and that $B = (I - Q)^{-1}$.

another proof that (I-Q) is invertible.

Here is another proof. Note first that

$$E(V_{ij}) = E\left(\sum_{n=0}^{\infty} I_n \mid X_0 = i \right) \; ,$$

where $I_n = 1$ if $X_n = j$ and $I_n = 0$ otherwise. Thus,

$$E(V_{ij}) = E\left(\sum_{n=0}^{\infty} I_n \mid X_0 = i \right) = \sum_{n=0}^{\infty} p_{ij}^{(n)} = \delta_{ij} + \sum_{n=1}^{\infty} q_{ij}^{(n)} \; ,$$

where $q_{ij}^{(n)}$ is the (i,j)-th element of the matrix Q^n. Therefore, $E(V_{ij})$ is the (i,j)-th element of the matrix

$$I + \sum_{n=1}^{\infty} Q^n = (I - Q)^{-1} .$$

(*b*) The sum of the elements in the i-th row of B is the expected value of the random variable

$$Y_i = \sum_{j \in T} V_{ij} ,$$

which is the number of visits to transient states before absorption, given that the process started in transient state i.

6.5 Consider a regular Markov chain, with state space $I = \{1, 2, \cdots, r\}$. Denote by T_{ij} the first entrance time into state j, given that the process starts in state i. Set $m_{ij} = E(T_{ij})$. (*a*) Show that

$$m_{ij} = 1 + \sum_{k \neq j} p_{ik} m_{kj} .$$

(*b*) Let $(w_1, w_2, \cdots, w_r)$ be the stationary probability state vector for the process. Show that $m_{jj} = 1/w_j$, for each state j. (*c*) Give a heuristic argument to justify the result of part (*b*).

Solution (*a*) Conditioning on the next state yields

$$E(T_{ij}) = \sum_{k \in I} E(T_{ij} \mid X_1 = k) p_{ik} = E(T_{ij} \mid X_1 = j) p_{ij} + \sum_{k \neq j} E(T_{ij} \mid X_1 = k) p_{ik} .$$

This equation then becomes

$$E(T_{ij}) = 1 \cdot p_{ij} + \sum_{k \neq j} (1 + E(T_{kj})) p_{ik} = 1 + \sum_{k \neq j} p_{ik} E(T_{kj}) ,$$

which is the result to be shown. (*b*) In the last equation above, for a fixed j, multiply the i-th equation by w_i, and sum over i to get

$$\sum_{i=1}^{r} w_i E(T_{ij}) = \sum_{i=1}^{r} w_i + \sum_{i=1}^{r} \sum_{k \neq j} w_i p_{ik} E(T_{kj}) .$$

We have that $w = wP$, and the sum of the components of w is one. Therefore, interchanging the order of summation, we get

$$\sum_{i=1}^{r} w_i E(T_{ij}) = 1 + \sum_{k \neq j} w_k E(T_{kj}) \ .$$

Cancelling like terms on each side of this equation yields $w_j E(T_{jj}) = 1$, as required. (*c*) Over a long number of time steps T, the average number of times the process is in state j is $w_j T$. But the average duration between these times that the process is in state j is m_{ij}. Hence, in the long-run, $(w_j T) m_{ij} = T$. Thus, dividing by T, yields $m_{jj} = 1/w_j$.

2.10 Wed 4/23/2008 notes

Hand out 4/22/2008 Wed

Chapter 10: Solutions to Selected Problems

10.4 Consider a set of m machines that are in continuous operation, and which fail independently of each other at an exponential rate λ. Assume there are $s : 1 \leq s \leq m$, repair persons that service the machines independently and each at exponential rate μ. Let $X(t)$ denote the number of machines at time t that are not operational; that is, they are in the repair shop. Determine the arrival and departure rates for the birth-and-death model.

Solution Suppose the state of the system is i and that $s \leq i < m$. Then $m - i$ machines are in operation, and s repair persons are busy. Hence, for a small time interval of length h, the probability $p_{i,i+1}(h)$ of an increase of one in the system is

$$\binom{m-i}{1}(\lambda h + o(h))(1-\lambda h + o(h))^{m-i-1}(1-\mu h + o(h))^{s} + o(h) .$$

to take care of b. 2 terms etc..

The first factors covers the event of one breakdown and no service completions, while the last $o(h)$ term covers other less likely possibilities, such as two breakdowns and one service completion, say. For the case $0 \leq i < s$, the same expression holds, except the last factor becomes $(1 - \mu h + o(h))^{i}$, since now i repair persons are busy. Thus, for $0 \leq i < m$, we have

$$p_{i,i+1}(h) = (m-i)\lambda h + o(h) \quad \text{as} \quad h \to 0^{+} .$$

Hence, the birth rates are $\lambda_i = (m-i)\lambda$, for $0 \leq i < m$. For $i = m$, evidently $\lambda_i = 0$ since no arrivals are possible in this case.

For the departure rates, suppose again that the state of the system is i and that $s \leq i < m$. Then $m - i$ machines are in operation, and s repair persons are busy. Hence, for a small time interval of length h, the probability $p_{i,i-1}(h)$ of a decrease of one in the system is

$$\binom{s}{1}(\mu h + o(h))(1-\mu h + o(h))^{s-1}(1-\lambda h + o(h))^{m-i} + o(h) .$$

The first factors covers the event of no breakdowns and one service completion, while the last $o(h)$ term covers other less likely possibilities, such as one breakdown and two service completions, say. For the case $1 \leq i < s$, the probability $p_{i,i-1}(h)$ becomes

$$\binom{i}{1} (\mu h + o(h))(1 - \mu h + o(h))^{i-1}(1 - \lambda h + o(h))^{m-i} + o(h) \ .$$

since now i repair persons are busy. Thus, asymptotically, $p_{i,i-1}(h) = s\mu h + o(h)$, for $s \leq i < m$, while for $1 \leq i < s$, $p_{i,i-1}(h) = i\mu h + o(h)$. Hence, the departure rates are $\mu_i = s\mu$ for $s \leq i < m$, and $\mu_i = i\mu$, for $1 \leq i < s$.

2.11 Mon 4/28/2008 notes

Handout Monday 4/28/2008.

Chapter 10: The Kolmogorov Equations

The purpose of this note is to develop the forward Kolmogorov equations, and also the backward Kolmogorov equations, for the case of a pure-jump, continuous-time Markov chain.

For a stationary Markov chain, define the transition probability

$$p_{ij}(t) = P(X(\tau + t) = j \mid X(\tau) = i) .$$

Denote the state space by I. Then, considering the transition probability $p_{ij}(t+s)$, and partitioning over all intermediate states $k \in I$ at time s, it follows that

$$p_{ij}(t+s) = \sum_{k \in I} p_{ik}(s) p_{kj}(t) .$$

These equations are the Chapman-Kolmogorov equations, which play a fundamental role throughout the analysis of Markov chains. From these equations, we can derive the forward Kolmogorov equations and the backward Kolmogorov equations.

The Forward Kolmogorov Equations We have

$$p_{ij}(t+s) = \sum_{k \in I} p_{ik}(t) p_{kj}(s) = \sum_{k \neq j} p_{ik}(t) p_{kj}(s) + p_{ij}(t) p_{jj}(s) .$$

Thus,

$$p_{ij}(t+s) - p_{ij}(t) = \sum_{k \neq j} p_{ik}(t) p_{kj}(s) + p_{ij}(t)(p_{jj}(s) - 1) .$$

Dividing both sides by s, and proceeding formally, we can take the limit as $s \to 0^+$, to get

$$p'_{ij}(t) = \sum_{k \neq j} p_{ik}(t) q_{kj} - v_j p_{ij}(t) .$$

These equations are called the forward Kolmogorov equations.

The Backward Kolmogorov Equations We have

$$p_{ij}(t+s) = \sum_{k\in I} p_{ik}(s)p_{kj}(t) = \sum_{k\neq i} p_{ik}(s)p_{kj}(t) + p_{ii}(s)p_{ij}(t) \ .$$

Thus,

$$p_{ij}(t+s) - p_{ij}(t) = \sum_{k\in I} p_{ik}(s)p_{kj}(t) = \sum_{k\neq i} p_{ik}(s)p_{kj}(t) + (p_{ii}(s)-1)p_{ij}(t) \ .$$

Dividing both sides by s, and proceeding formally, we can take the limit as $s \to 0^+$, to get

$$p'_{ij}(t) = \sum_{k\neq j} q_{ik}p_{kj}(t) - v_i p_{ij}(t) \ .$$

These equations are called the backward Kolmogorov equations.

Let $P(t)$ be the matrix whose (i,j)-th entry is $p_{ij}(t)$. Denote the state probability vector at time t by $z(t)$. Thus, $z_n(t) = P(X(t) = n)$, for $n \in I$. For any time $t \geq 0$, we then have $z(t) = z(0)P(t)$. Next, define Q to be the matrix whose (i,j)-th entry is q_{ij}, for $i \neq j$, and $q_{ii} = -v_i$. Then the forward Kolmogorov equations can be written

$$P'(t) = P(t)Q \ , \quad \text{for} \quad t > 0 \ ,$$

while the backward Kolmogorov equations can be written

$$P'(t) = QP(t) \ , \quad \text{for} \quad t > 0 \ .$$

In developing mathematical models using continuous time Markov chains, the elements of the matrix Q are typically determined first. See for instance, Example 10.3.1 in the text, and also Problems 10.4, 10.5, and 10.6. Then, in theory at least, the differential equations above can be solved to find the transition matrix $P(t)$ and the state probability vector $z(t)$ defined above.

Example 1 The Poisson process: The Poisson process, with rate constant $\lambda > 0$, is a pure birth process with state space $I = \{0, 1, 2, \cdots\}$, for

which $q_{i,i+1} = \lambda$, $v_i = \lambda$, and $q_{i,j} = 0$ otherwise. Thus, the matrix Q has the form

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 & \cdots \\ 0 & -\lambda & \lambda & \cdots \\ 0 & 0 & -\lambda & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} .$$

The entries in the transition matrix $P(t)$ can be found by induction. However, we will just find the state probability vector $z(t)$, assuming the process starts in state 0. Recalling that $z_n(t) = P(X(t) = n)$, we have $z(0) = (1, 0, 0, \cdots)$. In effect, we are finding the first row of $P(t)$, since at any time $t \geq 0$, $z(t) = z(0)P(t)$. Using the forward Kolmogorov equations gives us

$$z'(t) = z(0)P'(t) = z(0)P(t)Q = z(t)Q .$$

Thus, for $n = 0$, we have $z_0'(t) = -\lambda z_0(t)$, and for $n \geq 1$,

$$z_n'(t) = \lambda z_{n-1}(t) - \lambda z_n(t) .$$

The equation for $n = 0$ yields $z_0(t) = e^{-\lambda t}$, where we have used the initial condition $z_0(0) = 1$. Next, for $n \geq 1$, using the integrating factor $e^{\lambda t}$, we can solve for $z_n(t)$ to get

$$z_n(t) = \lambda e^{-\lambda t} \int_0^t e^{\lambda s} z_{n-1}(s) ds ,$$

where we have invoked the initial condition $z_n(0) = 0$ for $n \geq 1$. Working with these equations successively, starting with z_0, gives us

$$z_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} , \quad \text{for} \quad n \geq 0 . \qquad \blacksquare$$

another derivation for Poisson process

Example 2 A device is either operational (state 1), or being repaired (state 0). If it is in state 1, it can fail in an interval of time $(t, t + h)$ with probability $\mu h + o(h)$. If it is in state 0, it can be repaired and become

operational in an interval of time $(t, t+h)$ with probability $\lambda h + o(h)$. For this model, we have

$$Q = \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix} .$$

The forward Kolmogorov equations $P'(t) = P(t)Q$ are, in the first row:

$$p'_{0,0}(t) = -\lambda p_{0,0}(t) + \mu p_{0,1}(t) \quad \text{and} \quad p'_{0,1}(t) = \lambda p_{0,0}(t) - \mu p_{0,1}(t) ,$$

and in the second row:

$$p'_{1,0}(t) = -\lambda p_{1,0}(t) + \mu p_{1,1}(t) \quad \text{and} \quad p'_{1,1}(t) = \lambda p_{1,0}(t) - \mu p_{1,1}(t) .$$

From the first set of equations, noting that $p_{0,0}(t) + p_{0,1}(t) = 1$, we obtain the single differential equation

$$p'_{0,1}(t) + (\lambda + \mu) p_{0,1}(t) = +\lambda .$$

Using the initial condition $p_{0,1}(0) = 0$, and employing the integrating factor $e^{-(\lambda+\mu)t}$, yields

$$p_{0,1} = \frac{\lambda}{\lambda+\mu} - \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t} ,$$

and thus,

$$p_{0,0} = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t} .$$

In a similar way, it follows that

$$p_{1,0} = \frac{\mu}{\lambda+\mu} - \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t} ,$$

and,

$$p_{1,1} = \frac{\lambda}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t} .$$

Note that by taking the limit as $t \to \infty$, the long-run state probability vector π is found to be

$$\pi = \left(\frac{\mu}{\lambda+\mu} , \frac{\lambda}{\lambda+\mu} \right) .$$

This result could also have been obtained by solving the balance equations $\pi Q = 0$, subject to $\pi_0 + \pi_1 = 1$. ■

Exercises

1. For the machine repair problem in Problem 10.4 of the notes, what is the Q matrix? Assume there is only one repair person ($s = 1$).

2. For the machine repair problem with spares in Problem 10.5 of the notes, what is the Q matrix? Assume there is one repair person ($s = 1$), and one machine ($m = 1$).

3. For the light bulb problem, Problem 10.6 of the notes, what is the Q matrix?

2.12 Wed 5/7/08 notes

Handout Wed. May 7, 2008.

Chapter 9: The Poisson Process - Solutions to Selected Problems

9.1 A randomly occurring event E is called a Poisson event if (i) for some constant λ, the probability that E occurs during a time interval of length h is $\lambda h + o(h)$, (ii) the probability that E occurs two or more times during an interval of length h is $o(h)$, and (iii) the occurrence or lack of occurrence of E during disjoint time intervals are independent events. Let T be the waiting time for a Poisson event E to occur. Show that T is exponentially distribution with parameter λ.

Solution The probability

$$P(T > t) = \lim_{h \to 0} (1 - \lambda h + o(h))^N ,$$

where $Nh = t$. Thus,

$$P(T > t) = \lim_{h \to 0} (1 - \lambda h + o(h))^{t/h} = e^{-\lambda t} .$$

The distribution of the time until the first Poisson event occurs is therefore exponential. Note further that it follows, using the binomial distribution, that during a time interval of length t, the expected number of occurrences of a Poisson event is the limit as $h \to 0^+$ of $N(\lambda h + o(h))$, where again $Nh = t$. This expected value is thus λt.

9.2 Consider two independent Poisson processes, type 1 having rate λ_1, and type 2 having rate λ_2. Show that the combined process, where events are observed without regard to type, is still Poisson and that the rate is $\lambda_1 + \lambda_2$.

Solution Since each of the processes are Poisson, and since each operates independently of the other, it follows that in the combined process, we again have independent and stationary increments. Next, since the sum of two Poisson random variables is again Poisson, with mean equal to the sum of the two means, it follows that property (c) in the first definition of a Poisson process is satisfied.

9.3 Consider a Poisson process such that each time an event occurs, it is of type 1 with probability $p > 0$, or it is type 2 with probability $q = 1-p > 0$. Assume these two types appear independently from event to event. Show that these two processes each are Poisson processes and that they are independent of each other. Help: Let $N_1(t)$ and $N_2(t)$ be the counting processes for each type of event. Find the their joint distribution by conditioning on $N(t)$, where $N(t) = N_1(t) + N_2(t)$.

Solution We can think of the processes N_1 and N_2 as being formed by tossing a coin each time an event in the process N occurs. If the coin is heads (with probability p,) then the event is viewed as one in the N_1 process, and otherwise (with probability q), it is viewed as one in the N_2 process. The coin tossing is done independently on each occasion. It follows that since the original process has stationary increments, so must the processes N_1 and N_2. Indeed, since the number of events for the original process in an interval of time $(s, s+t)$ does not depend on s, and only on t, the same must be true of the derived processes N_1 and N_2. Similarly, since the original process has independent increments, it follows that each of the processes N_1 and N_2 must have independent increments. Indeed, since the number of events in disjoint intervals of the original process are independent, and since the coin tossing takes place independently from each occasion to the next, each of the derived processes must have independent increments. All of these conclusions hold whether or not the original process follows the Poisson process, only the independent and stationary increments of the original process are needed.

To find the distributions of N_1 and N_2, note that for any nonnegative integers n and m, the probability $P(N_1(t) = n, N_2(t) = m)$ is equal to

$$\sum_{\nu=0}^{\infty} P(N_1(t) = n, N_2(t) = m \,|\, N(t) = \nu) P(N(t) = \nu) \ .$$

But each conditional probability in the sum is zero, except for $\nu = n + m$. Therefore, this sum becomes simply

$$P(N_1(t) = n, N_2(t) = m \,|\, N(t) = n + m) P(N(t) = n + m) \ .$$

Next, by the coin tossing interpretation above, it follows that

$$P(N_1(t) = n, N_2(t) = m | N(t) = n + m) = \left(\begin{array}{c} n+m \\ m \end{array} \right) p^n q^m \ .$$

since each of the $n + m$ events corresponds to a Bernoulli trial, and the trials are independent with probability p that a type 1 process occurs. Therefore,

$$P(N_1(t) = n, N_2(t) = m) = \left(\begin{array}{c} n+m \\ m \end{array} \right) p^n q^m \cdot e^{-\lambda t} \frac{(\lambda t)^{n+m}}{(n+m)!} \ .$$

Simplifying this expression, and writing $\lambda = \lambda p + \lambda q$, we find that

$$P(N_1(t) = n, N_2(t) = m) = e^{-\lambda p t} \frac{(\lambda p t)^n}{n!} \, e^{-\lambda q t} \frac{(\lambda q t)^m}{m!} \ .$$

Since the joint distribution of N_1 and N_2 can be written as the product of a function of n alone, and a function of m alone, it follows that $N_1(t)$ and $N_2(t)$ are independent. However, each term in this product is a distribution function itself, and so

$$P(N_1(t) = n) = e^{-\lambda p t} \frac{(\lambda p t)^n}{n!} \quad \text{and} \quad P(N_2(t) = m) = e^{-\lambda q t} \frac{(\lambda q t)^m}{m!} \ .$$

Using these formulas, and the conclusion above that each process is stationary, it follows that property (c) of the first definition in the class notes for a Poisson process holds for N_1 and N_2. Alternatively, these formulas can be used to establish property (c) of the second definition of a Poisson process. Either way, we can conclude that N_1 is Poisson with parameter λp, and N_2 is Poisson with parameter λq.

9.5 A vehicle-controlled traffic light will stay green for τ seconds after a car passes through the intersection, and if no cars arrive during a period of time τ, then it turns red. Let X denote the number of cars that pass through the intersection following dissipation of the initial queue and until the light next turns red. Assume cars arrive according to a Poisson distribution with rate λ. Find the probability distribution of X and find the expected value of X. Help: Note that the event $\{X \geq n\}$ occurs if the inter-arrival times of the next n arriving cars are each no more than τ.

Solution We have

$$P(X = n) = P(T_1 < \tau, \cdots, T_n < \tau, T_{n+1} \geq \tau) ,$$

where the T_i are the inter-arrival times of the next n cars following queue dissipation. Since cars arrive according to a Poisson process, the T_i are independent random variables that are each exponentially distributed with parameter λ. Thus,

$$P(X = n) = \left(1 - e^{-\lambda\tau}\right)^n e^{-\lambda\tau} ,$$

for $n = 0, 1, 2, \cdots$. Not Geometric

To find the expected value of X, note that $Y = X + 1$ has a geometric distribution. Indeed,

$$P(Y = k) = P(X = k - 1) = e^{-\lambda\tau}\left(1 - e^{-\lambda\tau}\right)^{k-1} ,$$

for $k = 1, 2, \cdots$. Therefore, the random variable Y is geometric with parameter $e^{-\lambda\tau}$. Hence,

$$E(X) = E(Y - 1) = E(Y) - 1 = [1/(e^{-\lambda\tau})] - 1 = e^{\lambda\tau} - 1 .$$

not exam

9.6 An investor must decide on just one investment to make during a fixed period of time T. The opportunities are of two types, those of profit H, and those of profit L, where $L < H$. Opportunities appear at random according to a Poisson process of rate λ. Each time an opportunity appears it is worth L with probability p, and worth H with independent probability $q = 1 - p$. Consider the following strategy. For a given time τ, we invest only if an opportunity of profit H occurs. After this time, we invest in the next opportunity that appears, if any. What is the expect profit?

Solution Let Y denote the profit, and let X denote the first time an investment opportunity arrives. Partition the sample space according to the events $E = \{X \leq \tau\}$, $F = \{\tau < X \leq T\}$, and $G = \{T < X\}$. Then

$$E(Y) = E(Y|E)P(E) + E(Y|F)P(F) + E(Y|G)P(G) .$$

Since opportunities arrive according to a Poisson process, say $N(t)$, we see that $P(E) = P(N(\tau) \geq 1) = 1 - P(N(\tau) = 0) = 1 - e^{-\lambda\tau}$. Further, $P(G) = P(N(T) = 0) = e^{-\lambda T}$. Hence

$$P(F) = 1 - \left(1 - e^{-\lambda\tau}\right) - e^{-\lambda T} = e^{-\lambda\tau} - e^{-\lambda T} .$$

Returning to the expression for $E(Y)$ above, note that in the last term, $E(Y|G) = 0$, and so

$$E(Y) = E(Y|E)P(E) + E(Y|F)P(F) .$$

Consider next, in the second term, $E(Y|F)$. Given that the event F has occurred, the next opportunity will yield profit L with probability p and profit H with probability q. Hence $E(Y|F) = pL + qH$. Thus, we have so far

$$E(Y) = E(Y|E)P(E) + (pL + qH)\left(e^{-\lambda\tau} - e^{-\lambda T}\right) .$$

Finally, consider $E(Y \mid E)$. Given that the event E has occurred, namely that at least one opportunity has arrived in the period up to time τ, the probability that n opportunities arrive in this time interval is

$$q_n = \frac{1}{P(E)} P(N(\tau) = n) = \frac{1}{1 - e^{-\lambda\tau}} e^{-\lambda\tau} \frac{(\lambda\tau)^n}{n!} ,$$

and the possible values for n are $1, 2, 3, \cdots$. Therefore, the probability, say Q_L, that all the arriving opportunities yield profit L is

$$Q_L = \sum_1^\infty p^n q_n = \frac{1}{1 - e^{-\lambda\tau}} e^{-\lambda\tau} \sum_1^\infty \frac{(\lambda\tau p)^n}{n!} = \frac{1}{1 - e^{-\lambda\tau}} e^{-\lambda\tau} \left(e^{\lambda\tau p} - 1\right) ,$$

and so,

$$Q_L = \frac{e^{-\lambda\tau q} - e^{-\lambda\tau}}{1 - e^{-\lambda\tau}} .$$

Hence, given that the event E has occurred, the probability that in the period up to time τ, at least one of the arriving opportunities yields profit H is

$$1 - Q_L = 1 - \frac{e^{-\lambda\tau q} - e^{-\lambda\tau}}{1 - e^{-\lambda\tau}} = \frac{1 - e^{-\lambda\tau q}}{1 - e^{-\lambda\tau}} .$$

We now write

$$E(Y|E) = H(1 - Q_L) + 0 \cdot Q_L = H\frac{1 - e^{-\lambda\tau q}}{1 - e^{-\lambda\tau}} .$$

Returning to the last expression for $E(Y)$ we get

$$E(Y) = H\left(1 - e^{-\lambda\tau q}\right) + (pL + qH)\left(e^{-\lambda\tau} - e^{-\lambda T}\right) .$$

Not exam

9.7 Consider the $G/M/1$ model. (a) Is the chain irreducible? Is the chain aperiodic? Help: Write the transition probability matrix with a $+$, say, for each entry that is positive, and a 0 otherwise. (b) Find the transition probabilities when the inter-arrival time distribution is exponential with parameter λ. (c) Under the assumption of part (b), find the stationary distribution?

Solution (a) The transition matrix has nonzero entries on and below the diagonal, and also on the first upper diagonal. Given any two states i and j, it follows that j is accessible from i. Indeed, if $j \leq i+1$, then it is possible to go from state i to state j in one step. However, if $j \geq i+1$, then it is possible to go to state $i+1$ in one step, then from state $i+1$ to state $i+2$, and so on until state j is reached. It follows that any two states communicate with each other, and so the chain is irreducible. Thus all states must be of the same type and each must have the same period. Since $p_{11} > 0$, state 1 has period one, and so all states have period one.

(b) Substitution of the density $f(t) = \lambda e^{-\lambda t}$ into the expressions for the transition probabilities yields

$$p_{i1} = q^i , \quad \text{for} \quad i \geq 1, \quad \text{and} \quad p_{ij} = pq^{i+1-j} \quad \text{for} \quad 1 < j \leq i+1 ,$$

where $q = \mu/(\lambda + \mu)$, and $p = 1 - q$, and where $p_{ij} = 0$ otherwise.

(c) The solution for this case at hand should be the same as the one for the earlier problem concerning the $M/G/1$ queue. Thus, we could consider a solution to the stationary equations of the form $\pi_i = c\rho^i$, where $\rho = \lambda/\mu$, and then verify that such a solution works.

However, it is possible to show that a stationary vector exists in the general for an arbitrary inter-arrival time distribution. Consider a solution of the form $\pi_i = cr^i$, for some constants $c > 0$ and $r : \ 0 < r < 1$, which are to be determined. The stationary equations are

$$\pi_j = \sum_{i=j-1}^{\infty} \pi_i p_{ij} \ , \qquad \text{for } j = 2, 3, \cdots, \quad \text{where} \quad \pi_1 + \pi_2 + \cdots = 1 \ .$$

It suffices to consider only the equations for $j = 2, 3, \cdots$, since the condition $\pi_1 + \pi_2 + \cdots = 1$ implies that the equation for $j = 1$ is a consequence of the other equations. Now, for $j > 1$, plugging the hoped for solution into the j-th equation gives us

$$cr^j = \pi_j = \sum_{i=j-1}^{\infty} \pi_i p_{ij} = \sum_{i=j-1}^{\infty} cr^i \int_0^{\infty} e^{-\mu t} \frac{(\mu t)^{i+1-j}}{(i+1-j)!} f(t)\, dt$$

Making the change of variables $\nu = i - (j-1)$, simplifying, and then using the series expansion for the exponential function yields

$$r = \sum_{\nu=0}^{\infty} \int_0^{\infty} e^{-\mu t} \frac{(\mu r t)^{\nu}}{\nu!} f(t)\, dt = \int_0^{\infty} e^{-\mu(1-r)t} f(t)\, dt \ ,$$

which gives us the equation

$$r = \phi(r) = \int_0^{\infty} e^{-\mu(1-r)t} f(t) dt \ .$$

Note that $\phi(r)$ is a positive and increasing function of r, and that $\phi(1) = 1$. It follows that a solution to this equation for r exists and is unique, provided $\phi'(1) > 1$. But

$$\phi'(r) = \mu \int_0^{\infty} te^{-\mu(1-r)t} f(t) dt \ ,$$

so we need

$$\phi'(1) = \mu \int_0^{\infty} t f(t)\, dt > 1 \ .$$

Note that $\phi'(1)$ can be interpreted as a traffic intensity, defined as the ratio of the arrival rate to the service rate.

It follows now from Theorem 5.5.3 of the notes on classification of states, that when the condition $\phi'(1) > 1$ holds, the Markov chain $G/M/1$ is irreducible, aperiodic and all states are positive recurrent.

2.13 Wed 5/7/08 notes

Key solutions to continuous time Markov chains, chapter from lecture notes, chapter 10.

Handout Wed. May 8, 2008

Chapter 10: Solutions to Selected Problems

No exam

10.2 Find the recurrence formula for the expected values of the first entrance times $T_{i,i-1}$ for the birth-and-death process. Remark: To use such a recurrence formula to find the expected values $E(T_{i,i-1})$, it is necessary to specify boundary conditions. Generally, however, the boundary conditions can only be determined on a case by case basis.

Solution Following the steps in Section 10.4, let T_{ij} denote the first entrance time into state j, given that the process has just entered state i. For a birth-and-death process, we shall find the expected value of $T_{i,i-1}$, $i \geq 1$, by conditioning on the next state. Thus, let S denote the event that the next state is $i-1$. Then

$$E(T_{i,i-1}) = E(T_{i,i-1} \mid S)P(S) + E(T_{i,i-1} \mid \bar{S})P(\bar{S})$$

Note that $\bar{S}$ is the event that the next state is $i+1$. Using the jump probabilities for the birth-and-death process, we have

$$E(T_{i,i-1}) = E(T_{i,i-1} \mid S)\frac{\mu_i}{\lambda_i + \mu_i} + E(T_{i,i-1} \mid \bar{S})\frac{\lambda_i}{\lambda_i + \mu_i} \ ,$$

Now, $E(T_{i,i-1} \mid S) = 1/(\lambda_i + \mu_i)$, because the expected time to reach state $i-1$, given that the next state is $i-1$, is simply the expected time spent in state i, which is $1/(\lambda_i + \mu_i)$. Next,

$$E(T_{i,i-1} \mid \bar{S}) = \frac{1}{\lambda_i + \mu_i} + [E(T_{i+1,i}) + E(T_{i,i-1})] \ ,$$

for if the next state is $i+1$, then the process must spend the expected time in state i, which is $1/(\lambda_i + \mu_i)$, then spend the expected time $E(T_{i+1,i})$ to get back to state i, and then finally spend the expected time $E(T_{i,i-1})$ to get from state i to state $i-1$. Substituting these results, and then solving for $E(T_{i,i+1})$, yields

$$E(T_{i,i-1}) = \frac{1}{\mu_i} + \frac{\lambda_i}{\mu_i} E(T_{i+1,i}) \ .$$

As remarked above, to use this recurrence formula to find the expected values $E(T_{i,i-1})$, it is necessary to specify boundary conditions. Generally, however, the boundary conditions can only be determined on a case by case basis. Such a case appears in the next problem.

Not exam

10.3 (*a*) For the $M/M/1$ queue, find the expected first entrance time $E(T_{1,0})$. This time can be interpreted as the busy time for the queue. Assume $\lambda < \mu$. (*b*) For the $M/M/3$ queue, find the expected values $E(T_{i,i-1})$ for all $i \geq 1$. Help: Use the recurrence formula obtained in problem 10.2 above. Argue that all the values $E(T_{i,i-1})$ are the same for all $i \geq s$, where s is the number of servers.

Solution For the $M/M/s$ queue, when $i \geq s$, we have $\lambda_i = \lambda$ and $\mu_i = s\mu$, and these rates to not depend on the state i. The random variables $T_{i,i-1}$, for $i \geq s$, are concerned only with transitions above state i until the first instant that a transition to state $i-1$ occurs. Thus, the set of sample paths associated with the event $\{T_{i,i-1} = n\}$ is the same for each $i \geq s$, and each sample path for one value of $i \geq s$ is just as likely to occur as for another. Therefore, the stochastic nature of each $T_{i,i-1}$ for $i \geq s$ is the same. Note that this argument could not be made for the $T_{i,i+1}$. In these cases, the random variables are concerned only with transitions below state i until the first instant that a transition to state $i+1$ occurs. The set of sample paths associated with the event $\{T_{i,i+1} = n\}$ now depends on i, as some sample paths go below state s. Moreover, for smaller values of i, transitions that include visits to state 0 are more likely than for larger values of i. Thus, not all sample paths associated with the event $\{T_{i,i+1} = n\}$ are as likely to occur for one i as for another.

Continuing now with the solution, it follows that for $i \geq s$, each $E(T_{i,i-1})$ is a constant, say τ. Substituting τ into the recurrence formula obtained in the solution to Problem 10.2, we get

$$\tau = (1/s\mu) + (\lambda/s\mu)\tau \ ,$$

and solving this equation yields

$$\tau = E(T_{i,i-1}) = 1/(s\mu - \lambda) \ , \quad \text{for} \quad i \geq s.$$

Now, for part (a), when $s = 1$, we have $E(T_{1,0}) = 1/(\mu - \lambda)$. Next, for part (b), when $s = 3$, we have

$$E(T_{3,2}) = 1/(3\mu - \lambda) \ .$$

Using again the recurrence formula obtained in the solution to Problem 10.2, we get

$$E(T_{2,1}) = \frac{1}{2 \cdot \mu} + \frac{\lambda}{2 \cdot \mu} E(T_{3,2}) = \frac{3/2}{3\mu - \lambda} \ ,$$

and finally,

$$E(T_{1,0}) = \frac{1}{1 \cdot \mu} + \frac{\lambda}{1 \cdot \mu} E(T_{2,1}) = \frac{3\mu + (1/2)\lambda}{\mu(3\mu - \lambda)} \ .$$

10.4 Consider a set of m machines that are in continuous operation, and which fail independently of each other at an exponential rate λ. Assume there are $s : 1 \leq s \leq m$, repair persons that service the machines independently and each at exponential rate μ. Let $X(t)$ denote the number of machines at time t that are not operational; that is, they are in the repair shop. Determine the arrival and departure rates for the birth-and-death model.

Solution Suppose the state of the system is i and that $s \leq i < m$. Then $m - i$ machines are in operation, and s repair persons are busy. Hence, for a small time interval of length h, the probability $p_{i,i+1}(h)$ of an increase of one in the system is

$$\binom{m-i}{1} (\lambda h + o(h))(1 - \lambda h + o(h))^{m-i-1}(1 - \mu h + o(h))^s + o(h) \ .$$

The first factors cover the event of one breakdown and no service completions, while the last $o(h)$ term covers other less likely possibilities, such as two

breakdowns and one service completion, say. For the case $0 \leq i < s$, the same expression holds, except the last factor becomes $(1 - \mu h + o(h))^i$, since now i repair persons are busy. Thus, for $0 \leq i < m$, we have

$$p_{i,i+1}(h) = (m - i)\lambda h + o(h) \quad \text{as} \quad h \to 0^+ .$$

Hence, the birth rates are $\lambda_i = (m - i)\lambda$, for $0 \leq i < m$. For $i = m$, evidently $\lambda_i = 0$ since no arrivals are possible in this case.

For the departure rates, suppose again that the state of the system is i and that $s \leq i < m$. Then $m - i$ machines are in operation, and s repair persons are busy. Hence, for a small time interval of length h, the probability $p_{i,i-1}(h)$ of a decrease of one in the system is

$$\binom{s}{1} (\mu h + o(h))(1 - \mu h + o(h))^{s-1}(1 - \lambda h + o(h))^{m-i} + o(h) .$$

The first factors covers the event of no breakdowns and one service completion, while the last $o(h)$ term covers other less likely possibilities, such as one breakdown and two service completions, say. For the case $1 \leq i < s$, the probability $p_{i,i-1}(h)$ becomes

$$\binom{i}{1} (\mu h + o(h))(1 - \mu h + o(h))^{i-1}(1 - \lambda h + o(h))^{m-i} + o(h) .$$

since now i repair persons are busy. Thus, asymptotically, $p_{i,i-1}(h) = s\mu h + o(h)$, for $s \leq i < m$, while for $1 \leq i < s$, $p_{i,i-1}(h) = i\mu h + o(h)$. Hence, the departure rates are $\mu_i = s\mu$ for $s \leq i < m$, and $\mu_i = i\mu$, for $1 \leq i < s$.

10.5 Consider a set of $m+n$ machines which fail independently of each other at an exponential rate λ. It is intended that m machines are to be in operation at any time. The remaining machines serve as spares and are called into operation when an operating machine fails. If more than n machines are in a state of failure, then all the operational machines will be in service. Suppose there are s, $1 \leq s \leq m$, repair persons that service the machines independently and each at exponential rate

μ. Let $X(t)$ denote the number of machines at time t that are not operational; that is, they are in the repair shop. Determine the arrival and departure rates for the birth-and-death model.

Solution Suppose the state of the system is i. Then the number of machines not in the repair shop is $m+n-i$. Thus, if $m+n-i \geq m$, or equivalently $i \leq n$, then the arrival rate to the repair shop is $m\lambda$, while if $m+n-i < m$, or equivalently $i > n$, then the arrival rate to the repair shop is $(m+n-i)\lambda$. Finally, as in the case of the $M/M/s$ queue, the departure rates are $\mu_i = i\mu$ for $i < s$, and $\mu_i = s\mu$ for $i \geq s$.

10.6 Consider a sign that contains N light bulbs, each with a lifetime that follows an exponential distribution with parameter λ. Assume that the bulbs function independently of each other. Suppose it is the policy to allow bulbs to burn out until the moment the r-th bulb expires, and to then replace all burned out bulbs at that time. Define the state of the system $X(t)$ to be the number of burned out bulbs at time t. Argue that this stochastic process (a) has the Markov property, (b) is stationary, and (c) can be represented as a pure jump process. Determine the parameters v_i and q_{ij} of the jump process. (d) Determine the balance equations for this system, and find the long-run probability distribution for the states.

Solution For (a) and (b), suppose it is known that the state of the system is i at time s. Then there are $N-i$ bulbs still functioning. These $N-i$ bulbs operate independently of each other, and each has a lifetime that follows the exponential distribution. However the exponential distribution has the memoryless property. Thus, the situation is the same as if we were faced with $N-i$ new bulbs, whose lifetimes are not influenced by the behavior of the bulbs that have expired previously. The future evolution of the system depends only this information, and therefore depends only on the current state i, and not on how the system arrived at this state, nor on the current time. Hence, the process has the Markov property and is stationary. (c) Set

$$p_{ij}(h) = P(X(t+h) = j \mid X(t) = i) \; .$$

Using the properties of the exponential distribution, it follows that for small h, and $i < r-1$,

$$p_{i,i+1}(h) = \binom{N-i}{1} (\lambda h + o(h))^1 (1 - \lambda h + o(h))^{N-i-1} = (N-i)\lambda h + o(h) .$$

Thus, $q_{i,i+1} = (N-i)\lambda$. For $j \neq i+1$ and $j \neq i$, this transition probability $p_{ij}(h) = o(h)$. Hence,

$$p_{i,i}(h) = 1 - (N-i)\lambda h + o(h) .$$

Therefore, when $i < r-1$, we have $q_{i,i+1} = (N-i)\lambda$, and $v_i = (N-i)\lambda$, while $q_{ij} = 0$ otherwise. Finally, for $i = r-1$, using reasoning similar to above, and because all bulbs are replaced at the instance the r-th bulb expires, we get

$$p_{r-1,0}(h) = (N-r+1)\lambda h + o(h) ,$$

and

$$p_{r-1,r-1}(h) = 1 - (N-r+1)\lambda h + o(h) .$$

Therefore, when $i = r-1$, we have $q_{r-1,0} = (N-r+1)\lambda$, and $v_{r-1} = (N-r+1)\lambda$, while $q_{ij} = 0$ otherwise. (*d*) The balance equations in this case are

$$\pi_{i+1} = \frac{N-i}{N-(i+1)}\pi_i , \quad \text{for} \quad i = 0, 1, 2, \cdots, r-2, \quad \text{and} \quad \pi_{r-1} = \frac{N}{N-r+1}\pi_0 .$$

Requiring that $\pi_0 + \pi_1 + \cdots + \pi_{r-1} = 1$, we get

$$\pi_i = \frac{1}{h(r,N)} \cdot \frac{1}{N-i} , \quad \text{for} \quad i = 0, 1, 2, \cdots, r-1 .$$

where $h(r,N)$ is,

$$h(r,N) = \frac{1}{N} + \frac{1}{N-1} + \cdots + \frac{1}{N-(r-1)} .$$

10.7 For the previous light bulb problem: (*a*) What is the expected time between replacements? (*b*) Suppose for each replacement, it costs $\alpha + \beta r$ dollars to replace the r bulbs, where α and β are constants. Develop a formula that gives a reasonable estimate of the expected cost per unit time.

Solution (*a*) The time between replacements is

$$R = T_1 + T_2 + \cdots + T_r \ ,$$

where T_i is the time between the $(i-1)$-st burnout and the i-th burnout. The random variable T_i is an exponential random variable with rate constant $(N-i)\lambda$. Thus, $E(T_i) = 1/(N-i)\lambda$. Hence,

$$E(R) = \frac{1}{N\lambda} + \frac{1}{(N-1)\lambda} + \cdots + \frac{1}{(N-(r-1))\lambda} \ .$$

(*b*) One could argue as follows. Over a long period of time T, the average number of renewals is approximately $T/E(R)$. The total cost over this period of time is $(\alpha+\beta r)[T/E(R)]$. Dividing by the total time T yields an estimate of the cost per unit time. The final formula is therefore $C(r) = (\alpha+\beta r)/E(R)$.

10.12 For the $M/M/2$ queue, find the long-run probability distribution for the states. Use these probabilities to find $E(X)$, the expected number of customers in the system in the long-run. Help: Let $\rho = \lambda/2\mu$. Then $E(X) = 2\rho/(1-\rho^2)$.

Solution For the $M/M/s$ queue, with arrival rate λ and service rate μ, the steady state probabilities are given by

$$p_n = \frac{(\lambda/\mu)^n}{n!} p_0 \ , \quad \text{for } n = 1, 2, \cdots, s \ , \quad \text{and} \quad p_n = \frac{(\lambda/\mu)^n}{s! s^{n-s}} p_0 \ , \quad \text{for } n \geq s \ ,$$

where:

$$\frac{1}{p_0} = \frac{(\lambda/\mu)^s}{s!} \frac{1}{1-(\lambda/s\mu)} + \sum_{n=0}^{s-1} \frac{(\lambda/\mu)^n}{n!} \ , \quad \text{provided } \rho = \lambda/s\mu < 1 \ .$$

For the case of two servers,

$$L = \sum_{n=0}^{\infty} np_n = \sum_{n=1}^{\infty} np_n = p_1 + \sum_{n=2}^{\infty} np_n \ .$$

where $p_1 = (\lambda/\mu)p_0 = 2\rho p_0$, and for each $n \geq 2$, $p_n = 2\rho^n p_0$. Substituting yields

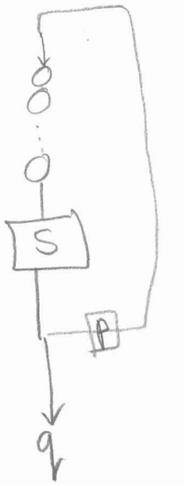
$$\begin{aligned} L &= 2\rho p_0 + 2p_0 \sum_{n=2}^{\infty} n\rho^n = 2\rho p_0 + 2\rho p_0 \sum_{n=2}^{\infty} n\rho^{n-1} \\ &= 2\rho p_0 \left(1 + \sum_{n=2}^{\infty} n\rho^{n-1}\right) = 2\rho p_0 \frac{d}{d\rho}\left(\sum_{n=0}^{\infty} \rho^n\right) = \frac{2\rho p_0}{(1-\rho)^2} \ . \end{aligned}$$

It remains to find p_0. From the formula above,

$$p_0 = \left(\frac{(\lambda/\mu)^2}{2!}\frac{1}{1-\rho} + [1 + (\lambda/\mu)]\right)^{-1} = \left(1 + 2\rho + \frac{2\rho^2}{1-\rho}\right)^{-1} = \frac{1-\rho}{1+\rho}$$

Hence, we get finally $L = 2\rho/(1-\rho^2)$.

10.15 Consider the $m/m/1$ queueing model, with the modification that when a customer finishes being served, there is a probability p that the customer will be returned to the queue to repeat the required service. (*a*) Assume this process follows a birth-and-death model, and find the arrival and departure rates. (*b*) Find the long-run state probability distribution.



Solution The state is the number of customers in the system. Suppose first the state $i = 0$. Then we have $p_{0,1}(h) = \lambda h + o(h)$, and $p_{0,0}(h) = 1 - \lambda h + o(h)$. Suppose now the state $i \geq 1$. Then

$$p_{i,i-1}(h) = (1 - \lambda h + o(h))(\mu h + o(h))q \ .$$

The first factor indicates that no arrival has occurred, and the next two factors indicate that a service completion has taken place, and the customer

was not returned for service. Thus, asymptotically, $p_{i,i-1}(h) = \mu q h + o(h)$. Next, we have

$$p_{i,i+1}(h) = (\lambda h + o(h))(1 - \mu h + o(h)) \ .$$

The first factor indicates that an arrival has occurred, and the next factor indicates that no service completion has taken place. Thus, asymptotically, $p_{i,i+1}(h) = \lambda h + o(h)$. Next, since the process follows a birth-and-death model, we could determine $p_{i,i}(h)$ from $p_{i,i-1}(h)$ and $p_{i,i+1}(h)$. However, it is instructive to find this term directly. Thus, we have

$$p_{i,i}(h) = (1 - \lambda h + o(h))(1 - \mu h + o(h)) + (1 - \lambda h + o(h))(\mu h + o(h))p \ .$$

The first term in the sum indicates that no arrival nor service completion have occurred. The second term in the sum indicates that no arrival has occurred, but a service completion has taken place, and the customer was returned for service. Asymptotically then,

$$p_{i,i}(h) = 1 - \lambda h - \mu h + \mu p h + o(h) = 1 - (\lambda + \mu q)h + o(h) \ .$$

In summary, then, we have $\lambda_i = \lambda$, for all i, and for $i \geq 1$, $\mu_i = \mu q$.

(*b*) Appealing to the solution of the balance equations given in the class notes, we have $p_n = r^n p_0$, for $n = 1, 2, \cdots$, where $r = \lambda/\mu q$. For a steady state to exist, we must require that $r < 1$. With this restriction, we get $p_n = (1 - r)r^n$, for $n = 0, 1, 2, \cdots$.

10.16 For the $m/m/s$ queueing model, assume that when a customer arrives and finds servers idle, the customer selects a server at random from those available. Consider a particular server, say server number 1. Show that the probability, in the long-run, that this server is idle is $1 - \rho$ where $\rho = \lambda/s\mu$. Help: Condition on the number in the system in the long-run.

Not on exam

Solution Let I denote the event that server one is idle. Then, for $n < s$,

$$P(I \mid N = n) = \frac{\binom{s-1}{n}}{\binom{s}{n}} = \frac{s-n}{s} \ ,$$

follow algebra

while $P(I \mid N = n) = 0$ for $n \geq s$. Thus

$$P(I) = \sum_{n=0}^{s-1} P(I \mid N = n)\pi_n = \sum_{n=0}^{s-1} \left(\frac{s-n}{s}\right) \pi_n \ .$$

Using the formulas from Example 10.3.2 for the long-run state probabilities, and setting $r = \lambda/\mu$, gives us

$$P(I) = \sum_{n=0}^{s-1} \left(\frac{s-n}{s}\right) \frac{r^n}{n!}\pi_0 \ ,$$

where

$$\frac{1}{\pi_0} = \sum_{i=0}^{s-1} \frac{r^i}{i!} + \frac{r^s}{s!}\frac{1}{1-\rho} \ .$$

To ease notation, set

$$T_s = \sum_{i=0}^{s-1} \frac{r^i}{i!} \ .$$

Then we can write the previous equation as

$$1 = T_s\pi_0 + \frac{r^s}{s!}\frac{\pi_0}{1-\rho} \ , \quad \text{or equivalently} \quad 1 - \rho = \pi_0(1-\rho)T_s + \pi_0\frac{r^s}{s!} \ .$$

Now,

$$P(I) = \pi_0 \left[\sum_{n=0}^{s-1} \left(\frac{s-n}{s}\right) \frac{r^n}{n!}\right] = \pi_0 \left[T_s - \frac{r}{s}\left(T_s - \frac{r^{s-1}}{(s-1)!}\right)\right] \ .$$

and so

$$P(I) = \pi_0 \left[(1-\rho)T_s + \frac{r^s}{s!}\right] = \pi_0(1-\rho)T_s + \pi_0\frac{r^s}{s!} \ .$$

Finally, recalling the expression for $1-\rho$ in the equation two lines up, we see that $P(I) = 1 - \rho$.

2.14 Wed 5/7/08 notes

Hastings metropolis algorithm lecture 11

11. Hastings-Metropolis Algorithm – Lecture 11 – October 8, 2002

11.1 The Hastings-Metropolis Algorithm

1. Begin with an irreducible Markov matrix Q_{ij}, with $i, j = 1, 2, 3, ...n$ (which need not be symmetric).
2. Let $n = 0$ and $X_0 = k$, for some $1 \leq k \leq m$.
3. Generate a new random variable X such that $Prob\{X = j\} = Q_{X_n,j}$.
4. Generate a random number U uniformly distributed on $(0, 1)$. If $U < \frac{[b(X)Q_{X,X_n}]}{[b(X_n)Q_{X_n,X}]}$ then $NS = X$; otherwise $NS = X_n$.
5. Let $n = n + 1$; set $X_n = NS$.
6. Go to step 3.

11.2 Application – Example 10a in Ross

We begin with a large set L of all permutations of $\{1, 2, ..., n\}$ for which $\sum_{j=1}^{n} jx_j > a$ for a given constant a. Another example is the set of all tree subgraphs of a given graph H. We want a limiting probability distriubtion which is uniform.

– First define a Markov chain graph G whose vertices are the elements of L. We will need a notion of neighbors in L, and we join node i to j by an arc if j is accessible from i in one move.

– For $L = S_n$ we put the arcs of G between states or permutations which differ by a transposition.

– Let $N(s) = \{$ the neighbors of a node $s\}$, and let $|N(x)|$ equal the cardinality of $N(s)$. Let $Q_{s,t} = frac1|N(s)|$ if $t \in N(s)$.

– Since we are interested in sampling uniformly from L, we want $\Pi(s) = \Pi(t) = K = \frac{b_s}{\sum b_j}$. Therefore, by setting

$$min(1, \frac{b_t Q_{ts}}{b_s Q_{st}} = min(\frac{|N_s|}{|N_t|}, 1) \tag{11.1}$$

we get an ergodic Markov chain which is reversible:

$$P_{s,t} = \left\{ \begin{array}{c} Q_{s,t} min(1, \frac{b_t S_{ts}}{b_s Q_{st}}) \\ Q_{s,s} + \sum_{r \neq s} Q_{s,r}(1 - min(1, \frac{b_r Q_{rs}}{b_s Q_{sr}})) \end{array} \right\} \tag{11.2}$$

by a theorem in the last lecture.

– The justification for using $min(1, \frac{b_t Q_{ts}}{b_s Q_{st}} = min(1, \frac{|N_s|}{|N_t|}$ as opposed to the earlier formulation $min(1, \frac{b_s}{b_t}$ is based on relaxing the condition that $Q_{st} = Q_{ts}$.

We check that $\sum_{t=1}^{m} P_{st} = 1$; and $P_{ss} > 0$ for some s, which is a generic property that leads to aperiodicity.

– Suppose the current state is $X_n = S$. Choose one of its neighbors randomly (make a random transposition).

If $|N_t| \leq |N_s|$, then accept $X_{n+1} = t$.

If $|N_t| > |N_s|$, then set $X_{n+1} = t$ with probability $q = \frac{|N_s|}{|N_t|}$. Set $X_{n+1} = s$ with probability $1 - q = 1 - \frac{|N_s|}{|N_t|}$.

11.3 Gibbs Sampler (Vector form of Hastings-Metropolis)

Let $\vec{X} = (X_1, x_2, ..., x_n)$ be a random vector with probability mass function $p(\vec{X}) = cg(\vec{x})$. We want to sample from such a distribution of random vectors, where $g(\vec{X})$ is known but c is not.

Consider a Markov chain where the states are $\vec{x} = (x_1, x_2, ..., x_n)$. Let $\vec{x}$ be the curent vector state. Choose $i = 1, 2, ..., n$ randomly and set the random variable $X = x$ with:

$$Prob\{X = x\} = Prob\{X_i = x | X_j = x_j, j \neq i\} \tag{11.3}$$

(which is given a priori by $p(\vec{x})$), since

$$Prob\{X_i = x | X_j = x_j, j \neq i\} = \frac{Prob\{X_i = x, X_j = x_j, j \neq i\}}{Prob\{X_j = x_j, j \neq i\}} \tag{11.4}$$

$$= \frac{Prob\{X_i = x, X_j = x_j, j \neq i\}}{\sum_k Prob\{X_i = k, X_j = x_j, j \neq i\}} \tag{11.5}$$

$$= \frac{p(x_1, x_2, ..., x_{i-1}, x_i, x_{i+1}, ..., x_n)}{\sum_k p(x_1, x_2, ..., x_{i-1}, k, x_{i+1}, ..., x_n)} \tag{11.6}$$

Next, if the random variable $X = x$, then $\vec{y} = (x_1, x_2, ..., x_{i-1}, x, x_{i+1}, ..., x_n)$ is a possible next state vector.

This is equivalent to Hastings-Metropolis with:

$$\begin{aligned} Q(\vec{x}, \vec{y}) &= \frac{1}{n} Prob\{X_i = x | X_j = x_j, j \neq i\} && (11.7) \\ &= \frac{p(\vec{y})}{nProb\{X_j = x_j, j \neq i\}} && (11.8) \\ &= \frac{p(\vec{y})}{n \sum_k p(x_1, x_2, ..., x_{i-1}, k, x_{i+1}, ..., x_n)} && (11.9) \end{aligned}$$

and

$$\begin{aligned} P_{\vec{x},\vec{y}} &= \left\{ \begin{array}{c} Q(\vec{x}, \vec{y}) min(1, \frac{p(\vec{y})Q(\vec{y},\vec{x})}{p(\vec{x})Q(\vec{x},\vec{y})}) \\ Q(\vec{x}, \vec{x}) + \sum_{\vec{z} \neq \vec{x}} Q(\vec{x}, \vec{z})(1 - \frac{p(\vec{z})Q(\vec{z},\vec{x})}{p(\vec{x})Q(\vec{x},\vec{z})}) \end{array} \right\} && (11.10) \\ &= \left\{ \begin{array}{ll} Q(\vec{x}, \vec{y}) & \vec{y} \neq \vec{x} \\ Q(\vec{x}, \vec{x}) & \vec{y} = \vec{x} \end{array} \right\} && (11.11) \end{aligned}$$

11.4 Application (Example 10b)

the problem is to generate n random points on the unit circle, such that no two points are within distance d of each other, where

$$\beta = Prob\{\text{no two points are within distance } d \text{ of each other}\} \qquad (11.12)$$

is assumed to be small.

Do this by applying the Gibbs sampler, starting with n points on the unit sphere $x_1, x_2, ..., x_n$ such that no two are within distance d of each other.

Generate a random number U and let $I = int(nU)+1$. this step picks randomly from $i = 1, 2, ..., n$.

Next, generate a random point on the circle, $X = x$ and if $|x - x_j| > d, j \neq I$, then set $\vec{y} = (x_1, x_2, ..., x_{I-1}, x, x_{I+1}, ..., x_n)$. Otherwise generate another point $X = x'$ and repeat.

11.5 The Metropolis Algorithm for Statistical Mechanics

Construct Aafinal probability distribution

$$\Pi_{\vec{x}} = P_{\vec{x}} = \frac{e^{-\beta E(\vec{x})}}{Z_N} \equiv \frac{b(\vec{x})}{\sum_{\vec{r}} b(\vec{r})} \tag{11.13}$$

where N denotes the number of particles or the number of lattice sites. To be precise let us suppose that the domain is the unit square (with periodic boundary conditions), and the N particles can take any position on a uniform (MxM) grid – let us say $(100x100)$ for this demonstration.

a. Compute the number of states in the problem. For $N = 100$ this number is $(10^4)^M = (10^4)^{100}$ which is huge!

b. So Q is an $M^{2N}xM^{2N}$ matrix $Q(\vec{x},\vec{y}) = \frac{1}{N} Prob\{X_i = x | X_j = x_j, j \neq i\}$.

c. The important point is that in the Metropolis algorithm,

$$P_{\vec{x},\vec{y}} = \left\{ \begin{array}{c} Q(\vec{x},\vec{y}) min(1, \frac{b_{\vec{y}} Q(\vec{y},\vec{x})}{b_{\vec{x}} Q(\vec{x},\vec{y})}) \\ Q(\vec{x},\vec{x}) + \sum_{\vec{z}\neq\vec{x}, b_j < b_i} Q(\vec{x},\vec{z})(1 - \frac{b_{\vec{y}} Q(\vec{y},\vec{x})}{b_{\vec{x}} Q(\vec{x},\vec{y})}) \end{array} \right\} \tag{11.14}$$

d. Finally, we need to analyze what

$$Prob\{X_i = X | X_j = x_j, j \neq i\}; i, j = 1, 2, ...N \tag{11.15}$$

$x_j \in L$, the MxM grid.

This can be treated as a graph G where the nodes are all possible states $\vec{x}$ of which there are M^{2N}. Arcs connect "neighbors" in G of $\vec{x}$ which are defined by states $\vec{y}$ that can be reached from $\vec{x}$ by a Gibbs sampler move, i.e, randomly choose $i = 1, 2, ..., N$ and then change $X_i = x$, leaving $X_j = x_j$ for $j \neq i$ with

$$Prob X_i = x | X_j = x_j = \frac{Prob\{X_i = x, X_j = x_j\}}{Prob\{X_j = x_j, j \neq i\}} \tag{11.16}$$

$$= \frac{e^{-\beta E(\vec{y})}}{\sum_k e^{-\beta E(X_i = k)}} \tag{11.17}$$

e. Putting it together we get

$$Q(\vec{x},\vec{y}) = \frac{e^{-\beta E(\vec{y})}}{N\sum_k e^{-\beta E(X_i=k)}} \tag{11.18}$$

and

$$P_{\vec{x},\vec{\cdot}} = \left\{ \begin{array}{ll} Q(\vec{x},\vec{y}) & \vec{y} \neq \vec{x} \\ Q(\vec{x},\vec{x}) & \vec{y} = \vec{x} \end{array} \right\} \tag{11.19}$$

Chapter 3

study notes,lecture notes

3.1 some code I wrote for testing things...

1. expected_test.m
2. solving_3_dot_6.nb
3. A small note on using the left eigenvector of the one step transition matrix for a regular chain to determine the limiting distribution trying_8_1_problem.nb trying_8_1_problem.pdf

3.2 Definitions

Source of this block unknown and lost from the net:

"Continuous Time Markov Chains

Most of our models will be formulated as continuous time Markov chains. On this page we describe the workings of these processes and how we simulate them.

Definition.

We say that an event happens at rate r(t) if an occurence between times t and t + dt has probability about r(t)dt when dt is small.

Fact. When r(t) is a constant r, the times t[i] between occurrences are independent exponentials with mean 1/r, and we have a Poisson process with rate r.

Markov chains in continuous time are defined by giving the rates q(x,y) at which jumps occur from state x to state y. In many cases (including all our examples) q(x,y) can be written as p(x,y)Q where Q is a constant that represents the total jump rate. In this case we can construct the chain by taking one step according to the transition probability p(x,y) at each point of a Poisson process with rate Q.

If we throw away the information about the exponential holding times in each state, the resulting sequence of states visited is a discrete time Markov chain, which is called the embedded discrete time chain. In our simulations, the total flip rate Q at any one time is a multiple of the number of sites, CQ. Since the number of sites is typically tens of thousands, we lose very little accuracy by simulating TCQ steps and calling the result the state at time T.

To build the discrete time chain we must pick from the various transitions with probabilities proportional to their rates. In our particle systems we can do this by picking a site at random, applying a stochastic updating rule, and then repeating the procedure. Because of this,

continuous time is occasionally referred to as asynchronous updating. This is to distinguish that porcedure from the synchronous updating of a discrete time process which updates all of the sites simultaneously. ”

3.2.1 Regular finite M.C.

definition 1: There exist some n such that $P^{(n)}$ has all positive entries

definition 2: A regular finite chain is one which is irreducible and aperiodic

Notice that this means regular chain has NO transient states.

3.2.2 irreducible M.C.

A M.C. which contains one and only one closed set of states. Note that for finite MC, this means all the states are recurrent. In otherwords, its state space contains no proper subset that is closed.

3.2.3 Stationary distribution

This is the state vector π which contains the probability of each state that the MC could be in the long term. For an irreducible MC, this is independent of the starting $\pi^{(0)}$, however, for a reducible MC, the Stationary distribution will be different for different initial $\pi^{(0)}$

3.2.4 recurrent state

1. $f_{ii} = 1$. In otherwords, the probability of reaching state i eventually, starting from state i is always certain.
2. $\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$, in otherwords, since sum diverges, this means the probability to return back to i starting from i will always exist, not matter how large n is (i.e. sum terms never reach all zeros after some limiting value n)

3.2.5 transient state

1. $f_{ii} < 1$. In otherwords, the probability of reaching state i eventually, starting from state i is not certain. i.e. there will be a chance that starting from i, chain will never again get back to state i.
2. $\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$, in otherwords, since sum converges, this means the probability to return back to i starting from i will NOT always exist (i.e. sum terms reach all zeros after some limiting value n)

3.2.6 Positive recurrent state

A recurrent state where the expected number of steps to return back to the state is finite.

3.2.7 Null recurrent state

A recurrent state where the expected number of steps to return back to the state is infinite.

3.2.8 Period of state

GCD of the integers n such that $p_{ii}^{(n)} > 0$. In otherwords, find all the steps MC will take to return back to the same state, then find the GCD of these values. If the GCD is 1, then the period is 1 and the state is called Aperiodic (does not have a period).

3.2.9 Ergodic state

A state which is Aperiodic and positive recurrent. i.e. a recurrent state (with finite number of steps to return) but it has no period.

3.2.10 First entrance time T_{ij}

The number of steps needed to reach state j (first time) starting from transient state i

3.2.11 $f_{ij}^{(n)}$

This is the probability that it will take n steps to first reach state j starting from transient state i. i..e $f_{ij}^{(n)} = P(T_{ij} = n)$.

3.2.12 f_{ij}

This is the probability of reaching state j (for first time) when starting from transient state i. Hence $f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)} = P(T_{ij} < \infty)$

3.2.13 Closed set

A set of states, where if MC enters one of them, it can't reach a state outside this set. i.e. $P_{ij} = 0$ whenever $i \in S$ and $j \notin S$, then set S is called closed set.

3.2.14 Absorbing M.C.

All none-transient states are absorbing states. Hence the P matrix looks like $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ R & R & q_{11} & q_{12} \\ R & R & q_{21} & q_{22} \end{bmatrix}$ i.e.

$\begin{bmatrix} I & 0 \\ R & Q \end{bmatrix}$

3.2.15 Q Matrix

Properties of a Q matrix are: There is at least one row which sums to less than 1. And there is a way to reach such row(s) from other others. and $Q^n \to 0$ as $n \to \infty$

3.2.16 Balance equations

3.3 HOW TO finite Markov chain

3.3.1 How to find $f_{ij}^{(n)}$

This is the probability it will take n steps to first reach state j from state i. In Below $C(J)$ means the closed set which contains the state j and T means the transient set

$i \in C(J)$, $j \in C(J)$	use formula (1) below
$i \in C(J)$, $j \notin C(J)$	$f_{ij}^{(n)} = 0$
$i \in T$, j Absorbing state	Calculate $A^m = Q^m R$ where $m = n-1$ then the i,j entry of A^m gives $f_{ij}^{(n)}$
$i \in T$, $j \in T$	Normally we are interested in finding expected number of visits to j before absorbing. i.e $E(V_{ij})$. see below. Otherwise use (1)

We can use $p_{ij}^{(n)}$. Notice the subtle difference between $f_{ij}^{(n)}$ and $p_{ij}^{(n)}$.

$f_{ij}^{(n)}$ gives the probability of needing n steps to first reach j from i, while $p_{ij}^{(n)}$ gives the probability of being in state j after n steps leaving i. So with $p_{ij}^{(n)}$ could have reached state j before n steps, but left state j and moved around, then came back, as long as after n steps exactly MC will be in state j. With $f_{ij}^{(n)}$ this is not allowed. The chain must reach state j the very first time in n steps from leaving i. So in a sense, $f_{ij}^{(n)}$ is a more strict probability. Using the recursive formula

$$p_{ij}^{(n)} = f_{ij}^{(n)} + f_{ij}^{(n-1)} p_{jj}^{(1)} + f_{ij}^{(n-2)} p_{jj}^{(2)} + \cdots + f_{ij}^{(1)} p_{jj}^{(n-1)} \tag{1}$$

We can calculate $f_{ij}^{(n)}$. We see that $f_{ij}^{(1)} = p_{ij}^{(1)}$ and so $f_{ij}^{(2)} = p_{ij}^{(2)} - f_{ij}^{(1)} p_{jj}^{(1)}$ and also

$$\begin{aligned} f_{ij}^{(3)} &= p_{ij}^{(3)} - f_{ij}^{(2)} p_{jj}^{(1)} - f_{ij}^{(1)} p_{jj}^{(2)} \\ &= p_{ij}^{(3)} - \left(p_{ij}^{(2)} - f_{ij}^{(1)} p_{jj}^{(1)} \right) p_{jj}^{(1)} - f_{ij}^{(1)} p_{jj}^{(2)} \\ &= p_{ij}^{(3)} - \left(p_{ij}^{(2)} p_{jj}^{(1)} - p_{ij}^{(1)} \left[p_{jj}^{(1)} \right]^2 \right) - p_{ij}^{(1)} p_{jj}^{(2)} \\ &= p_{ij}^{(3)} - p_{ij}^{(2)} p_{jj}^{(1)} + p_{ij}^{(1)} \left[p_{jj}^{(1)} \right]^2 - p_{ij}^{(1)} p_{jj}^{(2)} \end{aligned}$$

and

$$
\begin{aligned}
f_{ij}^{(4)} &= p_{ij}^{(4)} - f_{ij}^{(3)}p_{jj}^{(1)} - f_{ij}^{(2)}p_{jj}^{(2)} - f_{ij}^{(1)}p_{jj}^{(3)} \\
&= p_{ij}^{(4)} - \left(p_{ij}^{(3)} - p_{ij}^{(2)}p_{jj}^{(1)} + p_{ij}^{(1)}\left[p_{jj}^{(1)}\right]^2 - p_{ij}^{(1)}p_{jj}^{(2)} \right) p_{jj}^{(1)} - \left(p_{ij}^{(2)} - f_{ij}^{(1)}p_{jj}^{(1)} \right) p_{jj}^{(2)} - f_{ij}^{(1)}p_{jj}^{(3)} \\
&= p_{ij}^{(4)} - \left(p_{ij}^{(3)}p_{jj}^{(1)} - p_{ij}^{(2)}\left[p_{jj}^{(1)}\right]^2 + p_{ij}^{(1)}\left[p_{jj}^{(1)}\right]^3 - p_{ij}^{(1)}p_{jj}^{(1)}p_{jj}^{(2)} \right) - \left(p_{ij}^{(2)}p_{jj}^{(2)} - p_{ij}^{(1)}p_{jj}^{(1)}p_{jj}^{(2)} \right) - p_{ij}^{(1)}p_{jj}^{(3)} \\
&= p_{ij}^{(4)} - p_{ij}^{(3)}p_{jj}^{(1)} + p_{ij}^{(2)}\left[p_{jj}^{(1)}\right]^2 - p_{ij}^{(1)}\left[p_{jj}^{(1)}\right]^3 + p_{ij}^{(1)}p_{jj}^{(1)}p_{jj}^{(2)} - p_{ij}^{(2)}p_{jj}^{(2)} + p_{ij}^{(1)}p_{jj}^{(1)}p_{jj}^{(2)} - p_{ij}^{(1)}p_{jj}^{(3)} \\
&= p_{ij}^{(4)} - p_{ij}^{(3)}p_{jj}^{(1)} - p_{ij}^{(2)}p_{jj}^{(2)} + p_{ij}^{(2)}\left[p_{jj}^{(1)}\right]^2 - p_{ij}^{(1)}\left[p_{jj}^{(1)}\right]^3 + 2p_{ij}^{(1)}p_{jj}^{(1)}p_{jj}^{(2)} - p_{ij}^{(1)}p_{jj}^{(3)}
\end{aligned}
$$

etc...

Hence knowing just the P matrix, we can always obtain values of the f_{ij} for any powers

However, using the following formula, from lecture notes 6.2 is easier

$$A^{(n)} = Q^n R$$

the i,j entry of $A^{(n)}$ gives the probability of taking $n+1$ steps to first reaching j when starting from transient state i . So use this formula. Just note this formula works only when i is transient.

question: If i is NOT transient, and we asked to find what is the prob. it will take n steps to first reach state j from state i. Then use (1). right?

3.3.2 How to find f_{ij}

This is the probability that chain will eventually reach state j given it starts in state i

$i \in C(J)$ $j \in C(J)$	$f_{ij} = 1$
$i \in C(J)$ $j \notin C(J)$	$f_{ij} = 0$
$i \in T$ j recurrent but not absorbent, hence in a closed set with other states	Use formula in page 5.5 lect z_{ij} =probability that transi
$i \in T$ j is an absorbent	$f_{ij} = \left[(I-Q)^{-1}R\right]_{i,j}$
$i \in T$ $j \in T$	We know eventually $p_{ij} = 0$ f

3.3.3 How to find $E(V_{ij})$ the expected number of visits to j before absorbing?

Here, $i \in T$ and $j \in T$ Then

$$E(V_{ij}) = (I-Q)^{-1}_{i,j}$$

The above gives the average number of visits to state j (also transient) before chain is absorbed for first time.

question: Note that if chain is regular, then all states communicates with each others and then $i \in R, j \in R$ and so $E(V_{ij})$ can be found from the stationary distribution π^{∞} , right?

3.3.4 How to find average number of steps $E(T_{ij})$ between state i and state j?

$regular : i \in R, j \in R$	$E(T_{ij}) = 1 + \sum_{k \neq j} p_{ik} E(T_{kj})$ if $i = j$ then $E(T_{ij}) = \frac{1}{w_{jj}}$ where w is the stationary probability vector
$i \in T, j \in T$	does not make sense to ask this here?

3.3.5 How to find number of visits to a transient state?

Number of visits to transient state is a geometric distribution.

$$\Pr(n) = f_{ii}^{n-1}(1 - f_{ii})$$

The expected number of visits to transient state i is

$$E(X) = \frac{1}{1 - f_{ii}}$$

where f_{ii} is the probability of visiting state i if chain starts in state i

3.4 Some useful formulas

$\lim_{n\to\infty} \left(1 - \frac{z}{n}\right)^n = e^{-z}$

$\lim_{h\to 0} \left(1 - \lambda h + o(h)\right)^{\frac{t}{h}} = e^{-\lambda t}$

3.4.1 Law of total probability

$$\Pr(A) = \sum \Pr(A|B_i) \Pr(B_i)$$

3.4.2 Conditional (Bayes) formula

$$\Pr(A|B) = \frac{\Pr(A, B)}{\Pr(B)}$$

3.4.3 Inverse of a 2 by 2 matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}}{ad - bc}$$

3.4.4 $\omega_j = \frac{1}{\mu_j}$

The above says that for a regular finite MC, where a stationary probability exist (and is unique), then it is inverse of the mean number of steps between visits μ_j to state j in steady state.

3.4.5 $\omega = \omega P$

The above says that for a regular M.C. there exist a stationary probability distribution ω

3.4.6 Poisson random variable

N is number of events that occur over some period of time.

N is a Poisson random variable if

1. $N(0) = 0$
2. Independent increments
3. $P(N = n) = \frac{\lambda^n e^{-\lambda}}{n!}$

Where λ is the average number of events that occur over the same period that we are asking for the probability of this number of events to occur. Hence remember to adjust λ accordingly if we are given λ as rate (i.e. per unit time).

3.4.7 Poisson random Process

$N(t)$ is a Poisson random variable if

1. $N(0) = 0$
2. Independent increments
3. $P(N(t) = n) = \frac{(\lambda t)^n e^{-(\lambda t)}}{n!}$

Where λ is the average number of events that occur in one unit time. So $N(t)$ is random variable which is the number of events that occur during interval of length t

The probability that ONE event occure in the next h interval, when the interval is very small, is $\lambda h + o($

This can be seen by setting $n = 1$ in the definition and using series expansion for $e^{-(\lambda h)}$ and then letting $h \to 0$

Expected value of Poisson random variable: $E(N) = \lambda$. For a process, $E(N(t)) = \lambda t$ where λ is the rate.

3.4.8 Exponential random variable

T is random variable which is the time between events where the number of events occur as Poisson distribution,

pdf: $f(t) = \lambda e^{-\lambda t}$

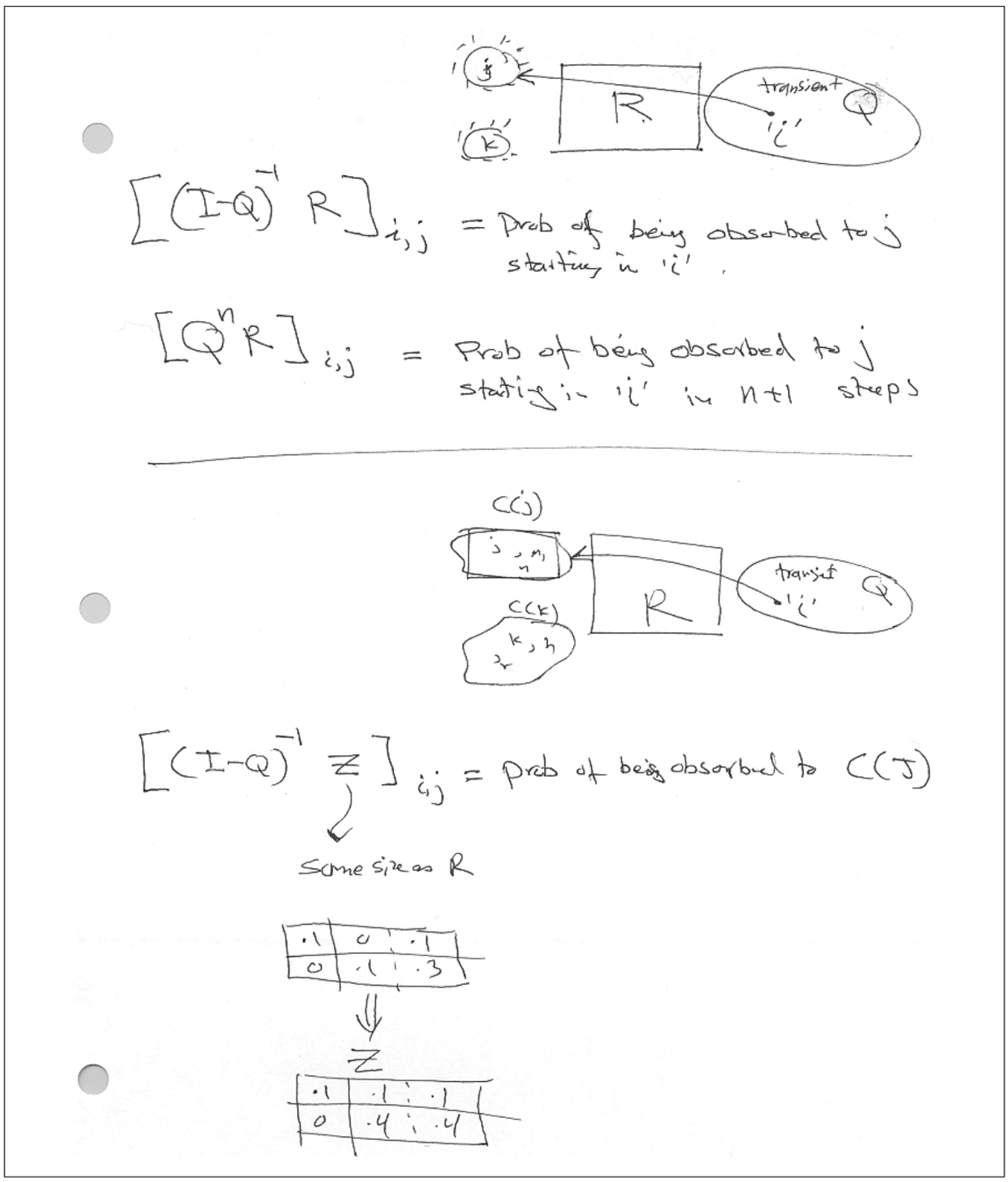
$$P(T > t) = \int_t^\infty \lambda e^{-\lambda s} ds = e^{-\lambda t}$$

$$P(T < t) = \int_0^t \lambda e^{-\lambda s} ds = -\left[e^{-\lambda s}\right]_0^t = -\left[e^{-\lambda t} - 1\right] = 1 - e^{-\lambda t}$$

pdf=derivative of CDF

Probability that the waiting time for n events to occur $\leq t$ is a GAMMA distribution. $g_n(t) = \frac{\lambda}{(n-1)!}(\lambda t)^{n-1} e^{-\lambda t}$

3.5 Diagram to help understanding



3.5.1 Continouse time Markov chain

$$p_{ii}(h) = 1 - v_i h + o(h)$$
$$p_{ij}(h) = q_{ij} h + o(h) \qquad i \neq j$$

v_i is the parameter (rate) for the exponential distributed random variable which represents the time in that state. Hence The probability that system remains in state i for time larger than t is given by

$$\Pr(T_i > t) = e^{-v_i t}$$

◦) Jump probability $Q_{ij} = \frac{q_{ij}}{v_i}$ for $i \neq j$. This is the probability of going from state i to state j (once the process leaves state i)

◦) FOrward Komogolv equation

$P'(t) = P(t)\,Q$, let $z(t) = z(0)\,P(t)$, hence $z'(t) = z(0)\,P'(t)$, hence $z'(t) = z(0)\,P(t)\,Q$ therefore

$$z'(t) = z(t)\,Q$$

∘) Balance equations

$$\pi_j v_j = \sum_{k \neq j} q_{kj} \pi_k$$

This is 'flow out' = 'flow in'.

This equation can also be obtaind more easily I think from $\pi Q = \mathbf{0}$ Where Q is the matrix made up from the $q's$ and the $v's$ on the diagonal. Just write then down, and at the end add $\pi_0 + \pi_1 + \cdots = 1$ to find π_0

Chapter 4

HWs

Local contents

4.1 Computing Assignment 1. Wed 2/7/08

Grade 5/5.

A problem in conditional probability (the first simulation HW, confidence interval, histogram) see first hand out for more details

Handout · 1st lecture 1/22/08

A Problem in Conditional Probability

A number is chosen at random from the interval $[0, 1]$. This value is placed in a box, and twice this value is placed in a second box. One of these boxes is selected at random and opened to reveal the number inside. Given this observed value, what is the probability that this number is the smaller of the two.

1. A Solution Let the random variable X denote the observed number, and let S denote the event that the selected box contains the smaller number. We seek $P(S|X = x)$ for $0 \le x \le 2$. The quantity $P(S|X = x)$ is undefined for other values of x. We will apply Bayes Theorem which gives us

$$P(S|X = x) = \frac{f_X(x|S)P(S)}{f_X(x)},$$

where $f_X(x)$ is the density function of the random variable X, and $f_X(x|S)$ is the conditional density of X given the event S. If $1 < x \le 2$, then evidently we have the larger of the two numbers, and so $P(S|X = x) = 0$ when $1 < x \le 2$. Thus, we need consider only the case $0 \le x \le 1$. Since a box is selected at random, $P(S) = 1/2$. Next, the conditional density of X given the event S is just the uniform density on $[0, 1]$. Thus, $f_X(x|S) = 1$ for $0 \le x \le 1$, and $f_X(x|S) = 0$ otherwise. Finally, to determine the density function of the random variable X, we use

$$f_X(x) = f_X(x|S)P(S) + f_X(x|\bar{S})P(\bar{S}).$$

The conditional density of X given the event $\bar{S}$, is the uniform density on $[0, 2]$. Thus, $f_X(x|\bar{S}) = 1/2$ for $0 \le x \le 2$, and $f_X(x|\bar{S}) = 0$ otherwise. Hence, for $0 \le x \le 1$,

$$f_X(x) = f_X(x|S)P(S) + f_X(x|\bar{S})P(\bar{S}) = 1 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4},$$

while for $1 < x \le 2$,

$$f_X(x) = f_X(x|S)P(S) + f_X(x|\bar{S})P(\bar{S}) = 0 \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4},$$

and otherwise, $f_X(x) = 0$. Returning now to the formula for $P(S|X = x)$ we have for $0 \le x \le 1$,

$$P(S|X = x) = \frac{(1)(1/2)}{3/4} = \frac{2}{3}.$$

This result shows that when we select a box at random, and observe a value between 0 and 1, there is a 2/3 chance that the observed value is the smaller of the two.

Exercise

1. Suppose that when we select a box, and observe the value, we have an opportunity to switch to the other box. The result above suggests that if we observe a value between 0 and 1, then we should switch, and otherwise, hold the value we have. Let the random variable Y denote the reward using such a strategy. Write a simulation program (in MATLAB, say) to estimate the expected value of Y. Use a 95% confidence interval, and determine the sample size so that the relative accuracy of your estimate is about one percent. In your report, explain how you determined your sample size. Also, compare theory and practise; that is, did your confidence interval include the true value.

$$\bar{x} \pm 1.96 \frac{s}{\sqrt{n}}$$ 95% C.I.

2. **Expected Value of Y** Suppose now the strategy is to switch if the observed value is less than or equal to 1, and otherwise to hold. Let Y be the reward using this strategy. Then

$$E(Y) = E(Y \mid X \leq 1)P(X \leq 1) + E(Y \mid X > 1)P(X > 1) .$$

Consider first the events $\{X \leq 1\}$ and $\{X > 1\}$. In order for the event $\{X > 1\}$ to occur, we must select the box with the larger value, which occurs with probability 1/2, and also the original value must be in the interval $(1/2, 1)$, which occurs with probability 1/2. Since these two events are independent, it follows that $P(X > 1) = (1/2)(1/2) = 1/4$, and further, $P(X \leq 1) = 1 - 1/4 = 3/4$.

Next consider the expected value of Y given that the event $\{X > 1\}$ has occurred. Then Y is the observed value X. Given that the event $\{X > 1\}$ has occurred, the random variable X is uniformly distributed over the interval $(1, 2)$. Hence, $E(Y \mid X > 1) = 3/2$.

Consider now the expected value of Y given that the event $\{X \leq 1\}$ has occurred. Here, we will switch to the value in the other box. However, we will either (α) switch to the larger value, which occurs with probability $2/3$, or (β) switch to the smaller value, which occurs with probability $1 - 2/3 = 1/3$. In case (α), Y is the larger value, which is uniformly distributed over the interval $(0, 2)$. Hence its expected value is 1, and so $E(Y|\alpha) = 1$. In case (β), Y is the smaller value, which now, because the event $\{X \leq 1\}$ has taken place, is uniformly distributed over the interval $(0, 1/2)$. Hence, the expected value is $1/4$, and so $E(Y|\beta) = 1/4$. Thus,

$$E(Y \mid X \leq 1) = E(Y|\alpha)P(\alpha) + E(Y|\beta)P(\beta) = 1 \cdot \frac{2}{3} + \frac{1}{4} \cdot \frac{1}{3} = \frac{3}{4}\,.$$

We are ready finally to compute the expected value $E(Y)$. From the formula above, we get

$$E(Y) = E(Y|X \leq 1)P(X \leq 1) + E(Y|X > 1)P(X > 1) = \frac{3}{4} \cdot \frac{3}{4} + \frac{3}{2} \cdot \frac{1}{4} = \boxed{\frac{15}{16}}\,.$$

Exercise

1. Find the density function of the random variable Y.

try to do this $f_Y(y) = f_Y(y|S)P(S) + f_Y(y|\bar{S})P(\bar{S})$

4.1.1 Problem description

A number is chosen at random from the interval $[0, 1]$. This value is placed in a box, and twice this value is placed in a second box. One of these boxes is selected at random and opened to reveal the number inside. Given this observed value, what is the probability that this number is the smaller of the two.

1. Suppose that when we select a box, and observe the value, we have an opportunity to switch to the other box. The result above suggests that if we observe a value between 0 and 1, then we should switch, and otherwise, hold the value we have. Let the random variable Y denote the reward using such a strategy. Write a simulation program (in MATLAB, say) to estimate the expected value of Y. Use a 95% confidence interval, and determine the sample size so that the relative accuracy of your estimate is about one percent. In your report, explain how you determined your sample size. Also, compare theory and practise; that is, did your confidence interval include the true value.

4.1.2 Purpose and design of project

The purpose of this project is to estimate the expected value of a random variable (called Y) which is generated by an experiment that is described in the above problem statement. Each experiment generates one random variable y. The experiment is described well in the above problem statement and no need to repeat it here again.

In addition, we are asked to determine the interval over which we are 95% confident the estimated expected value will lie within. We are asked that the interval should not be wider than 1% of the true mean from either side of the estimated expected value.

The simulation involve a two stage process. In the first stage, an initial simulation was made for $20,000$ experiments in which we obtained an estimate of the population standard deviation s and estimate of the population mean given by the sample mean $\bar{X}$. These 2 values are used to determined the sample size (number of experiments) needed for the second simulation performed to meet the above stated requirement for relative accuracy in expected value of Y. Therefore, once the first simulation is completed, the sample size for the second simulation was found by solving for n (sample size) by setting the expression for the standard error to be 1% of the population mean (in which we are using an estimate of which is $\bar{X}$ as generated by the first simulation). Therefore, we solve for n from

$$1.96\frac{s}{\sqrt{n}} = 0.01\bar{X}$$

Finally, the second simulation was now run using the above computed n, and the confidence interval was found from

$$C.I. = \left\{ \bar{X} - 1.96\frac{s}{\sqrt{n}} \cdots \bar{X} + 1.96\frac{s}{\sqrt{n}} \right\}$$

Where in the above equation the s and $\bar{X}$ are the sample standard deviation and the sample mean resulting from this second simulation (and not the first simulation run used to estimate n).

Next, the histogram Y was plotted to obtain an estimated of the probability density function of Y.

4.1.3 Summary of numerical results

For the initial simulation run, we used $20,000$ experiments and obtained the following estimate of the standard deviation and the population mean

$$\begin{aligned} s &= 0.625760 \\ \bar{X} &= 0.931717 \end{aligned}$$

Now solve for n from

$$1.96\frac{s}{\sqrt{n}} = 0.01\bar{X}$$

we found

$$\boxed{n = 17328}$$

Running the second stage simulation now to estimate the expected value of Y we obtain the following result that the estimate of the expected value of Y is

$$\bar{X} = 0.9311632$$

and the 95% confidence interval was found to be

$$\{0.92190 \cdots 0.94043\}$$

4.1.4 Discussion of numerical results

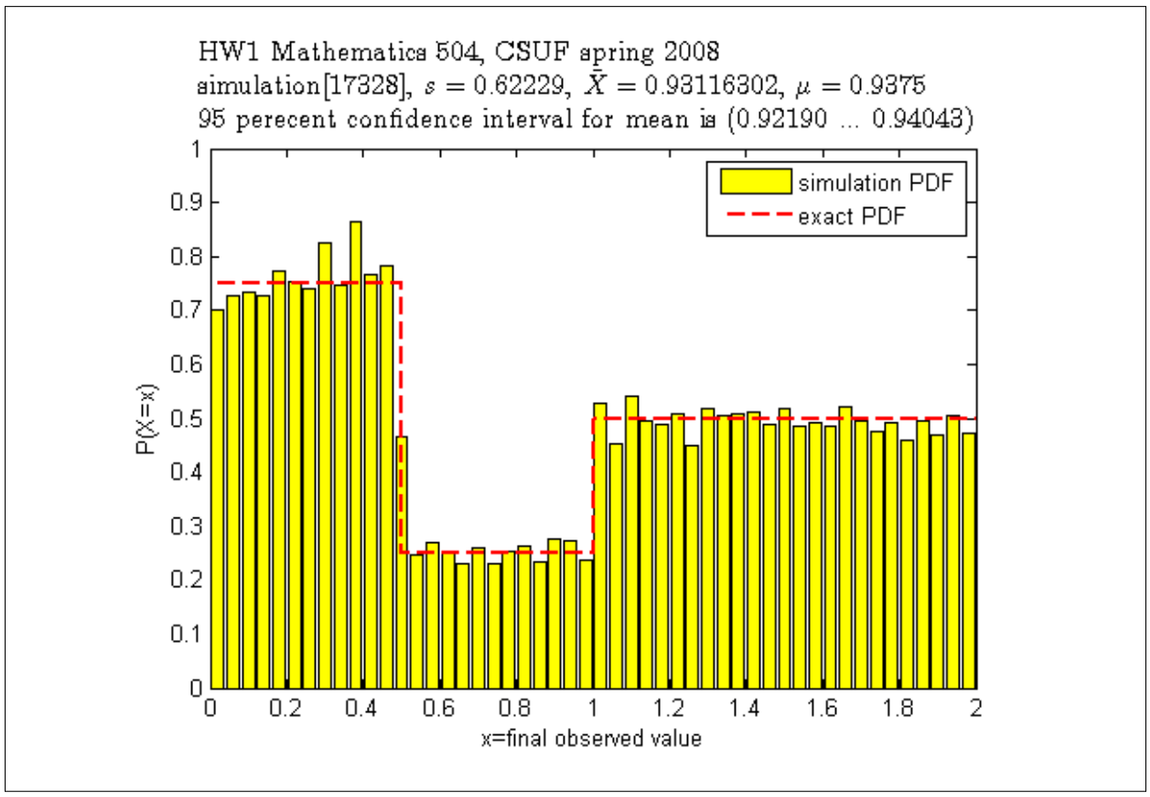
Since we know that the true value of $E(Y) = \frac{15}{16} = 0.937\,5$, we see that the

$$\boxed{\text{95\% confidence interval did contain the true value}}$$

We also notice that the relative error in $\bar{X}$ (the estimate of the expected value) when compared to the true mean $\mu = \frac{15}{16}$ is calculated as $\frac{\mu - \bar{X}}{\mu} = \frac{0.937\,5 - 0.9311632}{0.937\,5} = 0.006759\,3 \simeq 0.7\%$ which is little below the 1% requirement.

We note that the value of the relative error did not come out exactly 1% because we used an estimate of the true mean in order to find the sample size needed for the calculation.

The result of the simulation is the estimate of the PDF of Y which is shown in the plot below. The number of bins used is 50. This was determined by trial and error to obtain the most pleasing looking histogram.



We note that the true PDF is given below (derived in the class) and we see from the above plot that the estimated PDF is very close to the analytical PDF.

$$P(Y=y) = \begin{cases} \frac{3}{4} & 0 \le y \le 0.5 \\ \frac{1}{4} & 0.5 \le y \le 1 \\ \frac{1}{2} & 1 < x \le 2 \end{cases}$$

4.1.5 Code listing

nma_Math504_HW1.m

```
function nma_Math504_HW1(nBins,nSims,seed)

%by Nasser Abbasi, HW1, Math 504
%calculate expected value of y by simulation

%clear all; close all;

%
%  C O N S T A N T S     and   P A R A M E T E R S
%

%nSims = 20000;
%nBins = 50;
%seed  = 01010101;

skip  = round(0.01*nSims);
a = 0;  b = 1;
str = 'HW1 Mathematics 504   CSUF spring 2008';
str =[str '\nsimulation[%d],  std[%4.3f], mean[%6.5f],  true mean[%5.4f]'];
barWidth = 2/nBins;

%
%  I N I T I A L I Z A T I O N
```

```
%
rand('twister',seed);
box = zeros(1,2);
y = zeros(nSims,1);
figure(1);

%
%  L O G I C
%
for i=1:nSims
    box(1)=rand;
    box(2)=2*box(1);
    boxSelected=((a + (b-a).*rand)<=.5)+1;

    y(i)=box(boxSelected);

    if y(i)<=1  %switch box if needed
        if boxSelected==1
            boxSelected=2;
        else
            boxSelected=1;
        end

        y(i)=box(boxSelected);
    end

    [n,x]=hist(y(1:i),nBins);
    currentArea = barWidth*sum(n);
    if mod(i,skip)==0
        bar(x,n/currentArea,'y');  %relative frequency
        title(sprintf(str,i,std(y),mean(y),15/16));
        xlabel('x=final observed value');
        ylabel('P(X=x)');
        xtrue=[0 .5 .5 1 1 2]; ytrue=[.75 .75 .25 .25 .5 .5];
        line(xtrue,ytrue,'Color','r','LineWidth',4);
        ylim([0,1]);
        legend('simulation PDF','exact PDF');
        drawnow;
        pause(0.01)
    end
end

end
```

nma_Math504_HW1_as_script.m

```
function nma_Math504_HW1_part2()

%by Nasser Abbasi, HW1, part 2, Math 504
%calculate expected value of y by simulation

%This scripts simulates the pdf of the observed value from
%the following expeirment:
%
%pick a random number x from uniform[0,1], put this
%number in a box, and put twice the number in a second
%box. Next, pick one of these boxes by random, and look
```

```
%at the number inside. Call this y. If the number is smaller than
%one, then switch the box. Find the pdf of final y observed y, and
%find the estimate of the mean of y.
%Notice that  0<= y <= 2.

clear all; close all;

seed  = 01010101;
rand('twister',seed);

% First do an initial estimate using simulation to estimate
% population mean and standard deviation, and then use these to
% obtain the needed sample size for the error level required

[s,xBar] = initialEstimate();
err = 0.01 * xBar;
sampleSize = ((1.96*s)/err)^2;

%
%  C O N S T A N T S     and   P A R A M E T E R S
%
nBins = 50;
skip  = round(0.01*sampleSize);

barWidth = 2/nBins;

%
%  I N I T I A L I Z A T I O N
%
y   = zeros(sampeSize,1);
figure(1);
set(0,'DefaultTextinterpreter','none');
h=title('');
axpos = get(gca,'pos');
extent = get(h,'extent');
set(gca,'pos',[axpos(1) axpos(2) axpos(3) axpos(4)-.45*extent(4)]);
set(h,'VerticalAlignment','Middle');

%
%  L O G I C
%
for i=1:sampleSize
    y(i) = makeAnObservation();
    % currentArea = barWidth*sum(n);
    if mod(i,skip)==0
       generateOneFrame(y(1:i),nBins,sampleSize);
    end
end

generateFinalResult(y,nBins);


end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
function [s,xbar]=initialEstimate()


sampleSize = 20000;
y = zeros(sampleSize,1);

for i = 1:sampleSize
    y(i) = makeAnObservation()
end

s    = std(y);
xbar = mean(y);

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function y = makeAnObservation()

box = zeros(1,2);
box(1) = rand;
box(2) = 2*box(1);
boxSelected = (rand<.5)+1;       %pick a box by random

y =box(boxSelected);

if y <=1     %switch box if needed
    if boxSelected  == 1
        boxSelected = 2;
    else
        boxSelected = 1;
    end

    y = box(boxSelected);
end

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function generateOneFrame(y,nBins,sampleSize)

currentFrameNumber = length(y);

firstLineTitle  = 'HW1 Mathematics 504, part(2) CSUF spring 2008';
secondLineTitle = 'simulation[$%d$],  $s=%6.5f$, $\\bar{X}=%9.8f$,  $\\mu=%5.4f$';

[n,x]=hist(y,nBins);
bar(x,n/sampleSize,'y');  %relative frequency
ls=sprintf(secondLineTitle,currentFrameNumber,std(y),mean(y),15/16);

h=title(char(firstLineTitle,ls),'fontsize',12,'interpreter','latex');
xlabel('x=final observed value');
ylabel('P(X=x)');
```

```
xtrue=[0 .5 .5 1 1 2]; ytrue=[.75 .75 .25 .25 .5 .5];
line(xtrue,ytrue,'Color','r','LineWidth',2,'LineStyle','--');
ylim([0,1]);
legend('simulation PDF','exact PDF');
drawnow;

end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function generateFinalResult(y,nBins)

sampleSize = length(y);

firstLineTitle  = 'HW1 Mathematics 504, CSUF spring 2008';
secondLineTitle = 'simulation[$%d$],  $s=%6.5f$, $\\bar{X}=%9.8f$,  $\\mu=%5.4f$';

%
% Compare estimated mean with theortical mean
%
estimatedMean = mean(y);
standardError = 1.96*std(y)/sqrt(sampleSize);

lss=sprintf('95 perecent confidence interval for mean is (%6.5f ... %6.5f)',...
    estimatedMean-standardError,estimatedMean+standardError);

ls=sprintf(secondLineTitle,sampleSize,std(y),mean(y),15/16);
title({firstLineTitle,ls,lss},'fontsize',12,'interpreter','latex');

[n,x]=hist(y,nBins);
bar(x,n/sampleSize,'y');  %relative frequency
ls=sprintf(secondLineTitle,sampleSize,std(y),mean(y),15/16);

h=title(char(firstLineTitle,ls),'fontsize',12,'interpreter','latex');
xlabel('x=final observed value');
ylabel('P(X=x)');
xtrue=[0 .5 .5 1 1 2]; ytrue=[.75 .75 .25 .25 .5 .5];
line(xtrue,ytrue,'Color','r','LineWidth',2,'LineStyle','--');
ylim([0,1]);
legend('simulation PDF','exact PDF');
drawnow;

end
```

nma_Math504_HW1_part1_as_script.m

```
% file nma_Math504_HW1_part1_as_script
%by Nasser Abbasi, HW1, Math 504

%This scripts simulates the pdf of the observed value from
%the following expeirment:
%
%pick a random number x from uniform[0,1], put this
%number in a box, and put twice the number in a second
%box. Next, pick one of these boxes by random, and look
```

```
%at the number inside. Call this y. Find the pdf of
%y. Notice that  0<= y <= 2.

clear all; close all;

%
%  C O N S T A N T S     and   P A R A M E T E R S
%

nSims = 20000;
nBins = 50;
seed  = 01010101;   %my seed to reproduce same results.
skip  = round(0.01*nSims);

str = 'HW1 Mathematics 504, part 1. CSUF spring 2008';
str =[str '\nsimulation[%d],  std[%4.3f], mean[%6.5f],  true mean[%5.4f]'];
barWidth = 2/nBins;

%
%  I N I T I A L I Z A T I O N
%
rand('twister',seed);
box = zeros(1,2);
y   = zeros(nSims,1);
figure(1);

%
%  L O G I C
%
for i=1:nSims
    box(1) = rand;
    box(2) = 2*box(1);
    boxSelected = (rand<.5)+1;
    y(i)=box(boxSelected);

    [n,x]=hist(y(1:i),nBins);
    currentArea = barWidth*sum(n);
    if mod(i,skip)==0
        bar(x,n/currentArea,'y');  %relative frequency
        title(sprintf(str,i,std(y),mean(y),3/4));
        xlabel('x=observed value');
        ylabel('P(X=x)');
        xtrue=[0 1 1 2]; ytrue=[.75 .75 .25 .25];
        line(xtrue,ytrue,'Color','r','LineWidth',2,'LineStyle','--');
        ylim([0,1]);
        legend('simulation PDF','exact PDF');
        drawnow;
        %pause(0.01)
    end
end

title(sprintf(str,nSims,std(y),mean(y),3/4));
```

nma_Math504_HW1_part2.m

```
function nma_Math504_HW1_part2()
%function nma_Math504_HW1_part2()
%
%This function simulates the pdf of the observed value from
%the following expeirment:
%
%pick a random number x from uniform[0,1], put this
%number in a box, and put twice the number in a second
%box. Next, pick one of these boxes by random, and look
%at the number inside. Call this y. If the number is smaller than
%one, then switch the box. Find the pdf of final y observed y, and
%find the estimate of the mean of y.
%Notice that  0<= y <= 2.

%by Nasser Abbasi, HW1, part 2, Math 504

clear all; close all;

seed  = 01010101;
rand('twister',seed);

% First do an initial estimate using simulation to estimate
% population mean and standard deviation, and then use these to
% obtain the needed sample size for the error level required

[s,xBar] = initialEstimate();
err = 0.01 * xBar;
sampleSize = round(((1.96*s)/err)^2);
fprintf('s=%f, xBar=%f\n',s,xBar);

%
%  C O N S T A N T S     and   P A R A M E T E R S
%
nBins = 50;
skip  = round(0.01*sampleSize); %for simulation, skip frames

%
%  I N I T I A L I Z A T I O N
%
y   = zeros(sampleSize,1);
figure(1);
set(0,'DefaultTextinterpreter','none');
h=title({'','',''});
axpos = get(gca,'pos');
extent = get(h,'extent');
set(gca,'pos',[axpos(1) axpos(2) axpos(3) axpos(4)-.20*extent(4)]);
set(h,'VerticalAlignment','Middle');

%
%  L O G I C
%
for i=1:sampleSize
    y(i) = makeAnObservation();
    if mod(i,skip)==0
        generateOneFrame(y(1:i),nBins);
    end
```

```
end

generateFinalResult(y,nBins);

end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [s,xbar]=initialEstimate()

sampleSize = 20000;
y = zeros(sampleSize,1);

for i = 1:sampleSize
    y(i) = makeAnObservation();
end

s    = std(y);
xbar = mean(y);

end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function y = makeAnObservation()

box = zeros(1,2);
box(1) = rand;
box(2) = 2*box(1);
boxSelected = (rand<.5)+1;       %pick a box by random

y = box(boxSelected);

if y <=1     %switch box if needed
    if boxSelected  == 1
        boxSelected = 2;
    else
        boxSelected = 1;
    end

    y = box(boxSelected);
end

end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function generateOneFrame(y,nBins)

barWidth   = 2/nBins;
sampleSize = length(y);

firstLineTitle  = 'HW1 Mathematics 504, part(2) CSUF spring 2008';
secondLineTitle = 'simulation[$%d$],  $s=%6.5f$, $\\bar{X}=%9.8f$,  $\\mu=%5.4f$';
```

```
[n,x] = hist(y,nBins);
currentArea = barWidth*sum(n);
bar(x,n/currentArea,'y');  %relative frequency
ls=sprintf(secondLineTitle,sampleSize,std(y),mean(y),15/16);

title(char(firstLineTitle,ls),'fontsize',12,'interpreter','latex');
xlabel('x=final observed value');
ylabel('P(X=x)');
xtrue = [0 .5 .5 1 1 2]; ytrue=[.75 .75 .25 .25 .5 .5];
line(xtrue,ytrue,'Color','r','LineWidth',2,'LineStyle','--');
ylim([0,1]);
legend('simulation PDF','exact PDF');
drawnow;

end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function generateFinalResult(y,nBins)

barWidth = 2/nBins;
sampleSize = length(y);

firstLineTitle  = 'HW1 Mathematics 504, CSUF spring 2008';
secondLineTitle = 'simulation[$%d$],  $s=%6.5f$, $\\bar{X}=%9.8f$,  $\\mu=%5.4f$';

%
% Compare estimated mean with theortical mean
%
estimatedMean = mean(y);
standardError = 1.96*std(y)/sqrt(sampleSize);

lss=sprintf('95 perecent confidence interval for mean is (%6.5f ... %6.5f)',...
    estimatedMean-standardError,estimatedMean+standardError);

ls=sprintf(secondLineTitle,sampleSize,std(y),mean(y),15/16);

[n,x]=hist(y,nBins);
currentArea = barWidth*sum(n);
bar(x,n/currentArea,'y');  %relative frequency
title({firstLineTitle,ls,lss},'fontsize',12,'interpreter','latex');
xlabel('x=final observed value');
ylabel('P(X=x)');
xtrue=[0 .5 .5 1 1 2]; ytrue=[.75 .75 .25 .25 .5 .5];
line(xtrue,ytrue,'Color','r','LineWidth',2,'LineStyle','--');
ylim([0,1]);
legend('simulation PDF','exact PDF');
drawnow;

end
```

rhist.m

```
function [no,xo] = rhist(varargin)
%RHIST   Relative Histogram.
%   N = HIST(Y) bins the elements of Y into 10 equally spaced containers
%   and returns the relative frequency of elements in each container. If Y is a
%   matrix, RHIST works down the columns.
%
%   N = RHIST(Y,M), where M is a scalar, uses M bins.
%
%   N = RHIST(Y,X), where X is a vector, returns the relative freqency of Y
%   among bins with centers specified by X. The first bin includes
%   data between -inf and the first center and the last bin
%   includes data between the last bin and inf. Note: Use HISTC if
%   it is more natural to specify bin edges instead.
%
%   N = RHIST(Y,M,Any_Character) returns relative frequency density of
%   Y among bins.Any_Character is the any character inside single quotation
%   or any numeric value.
%   You can omit second optional argument using single quotation
%   i.e. N = RHIST(Y,'',Any_Character) returns relative frequency density
%   for 10 bins.
%   It is to be noted that sum(N)equals unity for relative frequency
%   while area under curve for relative frequency density equals unity.
%   Note that as size(Y,1) and M increases relative frequency density is
%   close to probability density for continous random variable.
%
%   [N,X] = RHIST(...) also returns the position of the bin centers in X.
%
%   RHIST(...) without output arguments produces a histogram of relative
%   frequency or relative frequency densisty bar plot of the results.
%   The bar edges on the first and last bins may extend to cover the min
%   and max of the data unless a matrix of data is supplied.
%
%   RHIST(AX,...) plots into AX instead of GCA.
%
%   Class support for inputs Y, X:
%      float: double, single
%
%   See also HIST.

%   Copyright 2004-2005 Durga Lal Shrestha.
%   $Revision: 1.0.0 $  $Date: 2005/6/20 14:30:00 $

% Parse possible Axes input

error(nargchk(1,inf,nargin));
[cax,args,nargs] = axescheck(varargin{:});

y = args{1};
if nargs == 1
    x = 10;
elseif nargs == 2
    x = args{2};
else
    if isempty(args{2})
        x = 10;
    else
```

```
        x = args{2};
    end
end
[m,n] = size(y);
[nn,x]=hist(y,x); % frequency
nn = nn./m;     % relative frequency

%  relative frequency density
if nargs == 3
    binwidth = x(2)-x(1);
    nn = nn./binwidth;
end

if nargout == 0
  if ~isempty(cax)
    bar(cax,x,nn,[min(y(:)) max(y(:))],'hist');
  else
    bar(x,nn,[min(y(:)) max(y(:))],'hist');
  end
  xlabel('y')
  if nargs == 3
    ylabel('relative frequency density')
  else
     ylabel('relative frequency')
  end
else
  no = nn;
  xo = x;
end
```

hw1.nb Mathematica

hw1.nb

4.2 Mon 2/5/08

Grade: 2/2.

Derive PDF of Y from an experiment where we switch boxes, uses probability decision tree

4.2.1 Problem description

This problem is a follow up on the problem described in HW1.

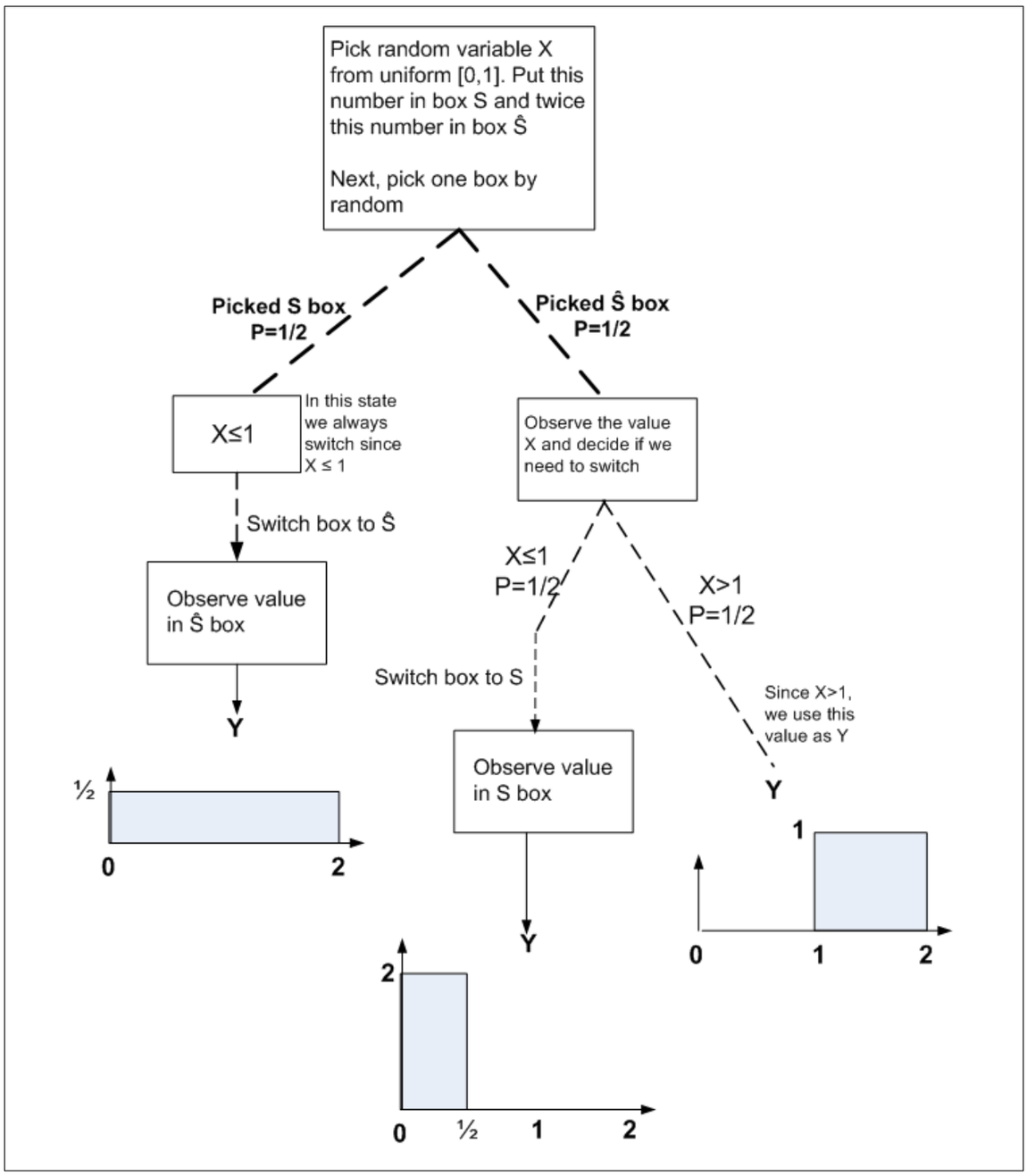
In this problem we are asked to derive analytically the PDF of the random variable Y by conditioning on the box selected.

Y is the random variable which is the observation from the following experiment: Generate random variable X from uniform $[0,1]$. Put this number is box labeled S and put twice this number in a box labeled $\bar{S}$. Next, we pick one of these 2 boxes by random. If the number inside the box selected is found to be greater than 1, then we switch the boxes and pick the number inside the second box. The random variable Y is the final number selected.

4.2.2 Solution

We first note the following known probabilities in this problem. The probability of picking box S or box $\bar{S}$ is $\frac{1}{2}$. Once we pick box S, then we have to switch the box. If we pick the $\bar{S}$ box, then we switch only if the observed X is less than 1.

To help solve this problem, we start by drawing the decision tree describing the possible flow and assign a probability to each branch. At the end of each branch we draw the PDF of Y resulting from traversing that branch only. Next, we combine (add algebraically) all the PDF's together after we scale each PDF by the probabilities found along the edges which lead to the end of the branch.



Using the above diagram as a guide, we now calculate the PDF for Y as follows (starting from the right most branch to the left most branch)

$$P(Y=y)=\frac{1}{2}\times\begin{cases}\frac{1}{2} & 0<y<\frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}<y<1\\ \frac{1}{2} & 1<y<2\end{cases}+\frac{1}{2}\times\frac{1}{2}\times\begin{cases}2 & 0<y<\frac{1}{2}\\ 0 & \frac{1}{2}<y<1\\ 0 & 1<y<2\end{cases}+\frac{1}{2}\times\frac{1}{2}\times\begin{cases}0 & 0<y<\frac{1}{2}\\ 0 & \frac{1}{2}<y<1\\ 1 & 1<y<2\end{cases}$$

$$=\begin{cases}\frac{1}{4} & 0<y<\frac{1}{2}\\ \frac{1}{4} & \frac{1}{2}<y<1\\ \frac{1}{4} & 1<y<2\end{cases}+\begin{cases}\frac{1}{2} & 0<y<\frac{1}{2}\\ 0 & \frac{1}{2}<y<1\\ 0 & 1<y<2\end{cases}+\begin{cases}0 & 0<y<\frac{1}{2}\\ 0 & \frac{1}{2}<y<1\\ \frac{1}{4} & 1<y<2\end{cases}$$

$$=\begin{cases}\frac{3}{4} & 0<y<\frac{1}{2}\\ \frac{1}{4} & \frac{1}{2}<y<1\\ \frac{1}{2} & 1<y<2\end{cases}$$

4.3 Wed 2/20/2008

Grade: 2/2.

The long analytical problem. Problem #4 from handout #3 above. Solving Einstein-Weiner pde using fourier transform

4.3.1 Problem

Solve the partial differential equation

$$\frac{\partial f}{\partial t} = c\frac{\partial (xf)}{\partial x} + D\frac{\partial^2 f}{\partial x^2} .$$

for the Ornstein-Ehrenfest process. Use the Fourier transform (see definition in previous problem), and the following steps. (*a*) Multiply the differential equation through by e^{-iyx} and integrate with respect to x over the interval $(-\infty, \infty)$. Then use integration by parts three times to obtain the equation

$$\frac{\partial \phi}{\partial t} = -cy\frac{\partial \phi}{\partial y} - Dy^2\phi ,$$

where

$$\phi(y,t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(x,t)e^{-iyx}dx .$$

In doing the integrations by parts, assume that

$$xf(x,t) \to 0 , \quad \text{and} \quad \frac{\partial f(x,t)}{\partial x} \to 0 , \quad \text{as} \quad x \to \pm\infty .$$

(*b*) Introduce the integrating factor

$$I(y) = \exp\left(\frac{D}{2c}y^2\right)$$

and show that $u(y,t) = I(y)\phi(y,t)$ satisfies the hyperbolic equation

$$\frac{\partial u}{\partial t} + cy\frac{\partial u}{\partial y} = 0 ,$$

subject to the initial condition $u(y,0) = u_0(y) = I(y)\phi(y,0)$ for $-\infty < y < \infty$. Note that $\phi(y,0)$ is the Fourier transform of $f(x,0)$, the density of the location of the particle at time $t = 0$.

(*c*) Use the method of characteristics to solve the previous hyperbolic equation and conclude that

$$u(y,t) = u_0\left(ye^{-ct}\right) , \quad \text{for} \ -\infty < y < \infty \ \text{and} \ t \geq 0 .$$

Then, using the definitions of $u(y,t)$ and $u_0(y)$, deduce that

$$\phi(y,t) = \exp\left(-\frac{1}{2}\sigma^2(t)y^2\right)\phi\left(ye^{-ct}, 0\right) , \quad \text{where} \ \sigma^2(t) = \frac{D}{c}\left(1 - e^{-2ct}\right) .$$

(*d*) To simplify the analysis, assume that the particle starts at a point x_0. Thus, the initial density of position is a dirac-delta function centered at the point x_0. It follows that

$$\phi(y,0) = \frac{1}{\sqrt{2\pi}}e^{-iyx_0} .$$

In this case, use the previous result to conclude that

$$\phi(y,t) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\sigma^2(t)y^2 - iy\mu(t)\right) , \quad \text{where} \ \mu(t) = x_0e^{-ct} .$$

(*e*) Set $b/a = -\mu(t)$ and $1/4a^2 = (1/2)\sigma^2(t)$, and use the result of part (*c*), Problem 3, to conclude that

$$\phi(y,t) = F[r(\cdot,t)](y) ,$$

where

$$r(x,t) = \frac{1}{\sigma(t)\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{x-\mu(t)}{\sigma(t)}\right)^2\right] .$$

Thus, $f(x,t) = r(x,t)$, which is the solution given above for the Ornstein-Ehrenfest model.

4.3.2 Solution

The PDE equation to solve is

$$\frac{\partial f(x,t)}{\partial t} = c\frac{\partial (xf(x,t))}{\partial x} + D\frac{\partial^2 f(x,t)}{\partial x^2}$$

part A

multiply through by e^{-iyx} and integrate w.r.t. x from $-\infty$ to ∞ we obtain

$$\frac{\partial}{\partial t}\left(\int_{-\infty}^{\infty} f(x,t)\, e^{-iyx} dx\right) = c\int_{-\infty}^{\infty} \frac{\partial (xf(x,t))}{\partial x} e^{-iyx} dx + D\int_{-\infty}^{\infty} \frac{\partial^2 f(x,t)}{\partial x^2} e^{-iyx} dx \tag{A}$$

Now do integration by parts on the first and second terms on the RHS above. We start with the first term

$$\begin{aligned}\int_{-\infty}^{\infty} \overbrace{\frac{\partial}{\partial x}(xf(x,t))}^{dv}\overbrace{e^{-iyx}}^{u} dx &= [vu]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} v du \\ &= \left[xf(x,t)\ \ e^{-iyx}\right]_{-\infty}^{\infty} + iy\int_{-\infty}^{\infty} xf(x,t)\, e^{-iyx} dx\end{aligned}$$

Using the assumption given that $xf(x,t) \to 0$ as $x \to \pm\infty$ then the first term above will vanish leaving

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial x}(xf(x,t))\, e^{-iyx} dx = iy\int_{-\infty}^{\infty} xf(x,t)\, e^{-iyx} dx \tag{1}$$

Now we need to solve the RHS of the above. To do that, we take the derivative of the Fourier transform itself with respect to its variable y and write

$$\begin{aligned}\frac{d}{dy}\mathcal{F}[f](y) &= \frac{d}{dy}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(x,t)\, e^{-iyx} dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \frac{d}{dy}\left(f(x,t)\, e^{-iyx}\right) dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} -ix\ f(x,t)\, e^{-iyx} dx\end{aligned}$$

Therefore we see that

$$\begin{aligned}\int_{-\infty}^{\infty} x\ f(x,t)\, e^{-iyx} dx &= \frac{\sqrt{2\pi}}{-i}\frac{d}{dy}\mathcal{F}[f](y) \\ &= \boxed{i\sqrt{2\pi}\tfrac{d}{dy}\mathcal{F}[f](y)}\end{aligned} \tag{2}$$

We now use this result to complete the solution. Substitute (2) into (1) we obtain

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial x}(xf(x,t))\, e^{-iyx} dx = -y\sqrt{2\pi}\frac{d}{dy}\mathcal{F}[f](y) \tag{3}$$

And using

$$\mathcal{F}[f(x,t)](y) \equiv \phi(y,t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(x,t)\, e^{-iyx} dx$$

Then (3) becomes

$$\boxed{\int_{-\infty}^{\infty} \frac{\partial}{\partial x}(xf(x,t))\, e^{-iyx} dx = -y\sqrt{2\pi}\frac{\partial}{\partial y}\phi(y,t)} \tag{4}$$

Now going back to equation (A) above, we do integration by parts twice on the second term on the RHS of that equation and obtain

$$\begin{aligned} D\int_{-\infty}^{\infty} \frac{\partial^2 f(x,t)}{\partial x^2} e^{-iyx} dx &= D\int_{-\infty}^{\infty} \overbrace{\frac{\partial}{\partial x}\left[\frac{\partial f(x,t)}{\partial x}\right]}^{dv} \overbrace{e^{-iyx}}^{u} dx \\ &= D\left\{ [vu] - \int_{-\infty}^{\infty} v du \right\} \\ &= D\left\{ \left[\frac{\partial f(x,t)}{\partial x} e^{-iyx}\right]_{-\infty}^{\infty} + iy\int_{-\infty}^{\infty} \frac{\partial f(x,t)}{\partial x} e^{-iyx} dx \right\} \end{aligned}$$

The term $\left[\frac{\partial f(x,t)}{\partial x} e^{-iyx}\right]_{-\infty}^{\infty}$ vanishes from the assumption that $\frac{\partial f(x,t)}{\partial x} \to 0$ as $x \to \pm\infty$, hence the above becomes

$$D\int_{-\infty}^{\infty} \frac{\partial^2 f(x,t)}{\partial x^2} e^{-iyx} dx = D\left\{ iy\int_{-\infty}^{\infty} \overbrace{\frac{\partial f(x,t)}{\partial x}}^{dv} \overbrace{e^{-iyx}}^{u} dx \right\}$$

Doing integration by parts again on the above we obtain

$$\begin{aligned} D\int_{-\infty}^{\infty} \frac{\partial^2 f(x,t)}{\partial x^2} e^{-iyx} dx &= iyD\left\{ [vu] - \int_{-\infty}^{\infty} v du \right\} \\ &= iyD\left\{ \left[f(x,t)\, e^{-iyx}\right]_{-\infty}^{\infty} + iy\int_{-\infty}^{\infty} f(x,t)\, e^{-iyx} dx \right\} \end{aligned}$$

The term $\left[f(x,t)\, e^{-iyx}\right]_{-\infty}^{\infty}$ vanishes from the assumption that $f(x,t) \to 0$ as $x \to \pm\infty$, hence the above becomes

$$\begin{aligned} D\int_{-\infty}^{\infty}\frac{\partial^2 f(x,t)}{\partial x^2}e^{-iyx}dx &= -y^2 D\left\{\int_{-\infty}^{\infty} f(x,t)\,e^{-iyx}dx\right\} \\ &= -y^2 D\sqrt{2\pi}\mathcal{F}[f](y) \\ &= -y^2 D\sqrt{2\pi}\phi(y,t) \end{aligned} \tag{5}$$

Substituting (5) and (4) into (A) gives

$$\frac{\partial}{\partial t}\left(\int_{-\infty}^{\infty} f(x,t)\,e^{-iyx}dx\right) = -cy\sqrt{2\pi}\frac{\partial}{\partial y}\phi(y,t) - y^2 D\sqrt{2\pi}\phi(y,t)$$

Hence

$$\sqrt{2\pi}\frac{\partial}{\partial t}\phi(y,t) = -cy\sqrt{2\pi}\frac{\partial}{\partial y}\phi(y,t) - y^2 D\sqrt{2\pi}\phi(y,t)$$

Simplifying

$$\frac{\partial \phi(y,t)}{\partial t} = -cy\frac{\partial \phi(y,t)}{\partial y} - Dy^2\ \phi(y,t)$$

part B

We need to show that

$$u(y,t) = I(y)\,\phi(y,t) \tag{1}$$

satisfies the hyperbolic equation shown below in (2), where $I(y) = \exp\left(\frac{D}{2c}y^2\right)$ and $\phi(y,t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(x,t)\,e^{-iyx}dx$

$$\frac{\partial u}{\partial t} + cy\frac{\partial u}{\partial y} = 0 \tag{2}$$

One way to do this is to plug in the expression for $u(y,t)$ given in (1) into the LHS of (2) and see if that gives zero. Hence the LHS of the above pde becomes

$$\begin{aligned} LHS &= \frac{\partial u}{\partial t} + cy\frac{\partial u}{\partial y} \\ &= \frac{\partial}{\partial t}I(y)\,\phi(y,t) + cy\frac{\partial}{\partial y}I(y)\,\phi(y,t) \\ &= \left[\phi(y,t)\,\frac{\partial I(y)}{\partial t} + I(y)\,\frac{\partial \phi(y,t)}{\partial t}\right] + cy\left[\phi(y,t)\,\frac{\partial I(y)}{\partial y} + I(y)\,\frac{\partial \phi(y,t)}{\partial y}\right] \end{aligned} \tag{3}$$

But $\frac{\partial I(y)}{\partial t} = 0$ and $\frac{\partial I(y)}{\partial y} = \frac{Dy}{c}I(y)$ and $\frac{\partial \phi(y,t)}{\partial t} = -cy\frac{\partial \phi(y,t)}{\partial y} - Dy^2\ \phi(y,t)$ since this is the pde we obtained in part(A), hence putting all these into (3) we obtain

$$\begin{aligned} LHS &= \left[I(y)\left\{-cy\frac{\partial \phi(y,t)}{\partial y} - Dy^2\ \phi(y,t)\right\}\right] + cy\left[\phi(y,t)\,\frac{Dy}{c}I(y) + I(y)\,\frac{\partial \phi(y,t)}{\partial y}\right] \\ &= \overbrace{-cyI(y)\,\frac{\partial \phi(y,t)}{\partial y}} - \underbrace{I(y)\,Dy^2\ \phi(y,t)} + \underbrace{I(y)\,Dy^2\phi(y,t)} + \overbrace{cyI(y)\,\frac{\partial \phi(y,t)}{\partial y}} \end{aligned}$$

Which is zero.

Hence $u(y,t) = I(y)\,\phi(y,t)$ satisfies $\frac{\partial u}{\partial t} + cy\frac{\partial u}{\partial y} = 0$ by direct substitution.

Part C

Now we need to solve the first order hyperbolic PDE equation

$$\frac{\partial u}{\partial t} + cy\frac{\partial u}{\partial y} = 0 \tag{1}$$

In the method of characteristics we convert the PDE to an ODE by looking for parametric path along which solutions for PDE exist. Let $t = t(s)$ and $y = y(s)$ where s is a parameter. So now we can write

$$u \equiv u(t,y) = u(t(s)\,, y(s))$$

Therefore taking full derivative of u w.r.t. to the parameter s we obtain using the chain rule the following

$$\frac{du}{ds} = \frac{\partial u}{\partial t}\frac{dt}{ds} + \frac{\partial u}{\partial y}\frac{dy}{ds} \tag{2}$$

Compare (2) to (1) we see that if we set

$$\frac{dt}{ds} = 1 \tag{2.1}$$

with the initial condition $t(s=0) = 0$ and if we set

$$\frac{dy}{ds} = cy \tag{2.2}$$

with initial condition $y(s=0) = y_0$, then this would make $\frac{du}{ds} = 0$, which means that the solution is constant along each specific parameter s which is what we want. Let the initial condition $u(s=0) \equiv u(y(s=0)\,, t(s=0)) = u_0(y_0, 0)$. Hence solution to $\frac{du}{ds} = 0$ is

$$u = u_0(y_0) \tag{3}$$

Now, from (2.1) we have $t = s$ since $t(0) = 0$ and from (2.2) we have $y = y_0 e^{cs}$where y_0 comes from initial condition of $y(s)$ as above.

Now, since $t = s$ hence we have

$$y = y_0 e^{ct}$$

Hence solve for y_0 we have

$$y_0 = ye^{-ct}$$

Pluging the above into (3) gives

$$u(t,y) = u_0\big(ye^{-ct}\big)$$

Which is the solution for $t \geq 0$ and $-\infty \leq y \leq \infty$

Now to show the final part. From part(B) we showed that

$$u(y,t) = I(y)\,\phi(y,t) \tag{4}$$

But since $u(t,y) = u_0(ye^{-ct})$ then we write (4) as

$$u_0\big(ye^{-ct}\big) = I(y)\,\phi(y,t) \tag{5}$$

But

$$u_0\big(ye^{-ct}\big) = I\big(ye^{-ct}\big)\,\phi\big(ye^{-ct},0\big) \tag{6}$$

which is the initial conditions we are given in the problem statement (where I replaced y by ye^{-ct}) Hence plug (6) into (5) we obtain

$$I\big(ye^{-ct}\big)\,\phi\big(ye^{-ct},0\big) = I(y)\,\phi(y,t)$$

or

$$\phi(y,t) = \frac{I(ye^{-ct})}{I\,(y)}\phi\big(ye^{-ct},0\big) \tag{7}$$

But $\frac{I(ye^{-ct})}{I(y)} = \frac{\exp\left(\frac{D}{2c}\left(ye^{-ct}\right)^2\right)}{\exp\left(\frac{D}{2c}y^2\right)} = \exp\left(\frac{D}{2c}y^2e^{-2ct} - \frac{D}{2c}y^2\right) = \exp\left(-\frac{D}{2c}y^2(1-e^{-2ct})\right)$

Hence (7) becomes

$$\phi(y,t) = \exp\left(-\frac{D}{2c}y^2\left(1-e^{-2ct}\right)\right)\phi\big(ye^{-ct},0\big)$$

Letting $\sigma^2(t) \equiv \frac{D}{c}(1-e^{-2ct})$, then the above can be rewritten as

$$\boxed{\phi(y,t) = \exp\left(-\tfrac{1}{2}\sigma^2(t)\,y^2\right)\phi(ye^{-ct},0)} \tag{8}$$

Part (D)

Since now $\phi(y,0) = \frac{1}{\sqrt{2\pi}}\exp\left(-iyx_0\right)$, then

$$\phi\big(ye^{-ct},0\big) = \frac{1}{\sqrt{2\pi}}\exp\left(-iye^{-ct}x_0\right)$$

where I replaced y by ye^{-ct}

Substitute the above into (8)

$$\begin{aligned}\phi(y,t) &= \exp\left(-\frac{1}{2}\sigma^2(t)\,y^2\right)\frac{1}{\sqrt{2\pi}}\exp\left(-iye^{-ct}x_0\right)\\ &= \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}\sigma^2(t)\,y^2 - iy\mu(t))\end{aligned}$$

Where $\mu(t) = e^{-ct}x_0$

part(E)

We need to show that

$$\frac{1}{\sigma(t)\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{x-\mu(t)}{\sigma(t)}\right)^2\right) \xrightarrow{\mathcal{F}} \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}\sigma^2(t)\,y^2 - iy\mu(t))$$

where $\mathcal{F}$ is the Fourier transform operator.

Since the transform variable is y, we can rewrite the above as needing to show the following

$$\exp\left(-\frac{1}{2}\left(\frac{x-\mu(t)}{\sigma(t)}\right)^2\right) \xrightarrow{\mathcal{F}} \sigma(t)\exp(-\frac{1}{2}\sigma^2(t)\,y^2 - iy\mu(t)) \tag{1}$$

We start by using result from problem (3) part (c) which says the following is true

$$\exp\left(-(ax+b)^2\right) \xrightarrow{\mathcal{F}} \frac{1}{\sqrt{2}\,|a|}\exp(-\frac{y^2}{4a^2} + iy\frac{b}{a}) \tag{2}$$

Hence, we know (2) is true, and we need to show that (1) is true. Using the hint given, we let $a = \frac{1}{\sqrt{2}\sigma(t)}$ and $\frac{b}{a} = -\mu(t)$ in (2) to arrive at (1). hence starting with (2) we write

$$\exp\left(-\left[a\left(x+\frac{b}{a}\right)\right]^2\right) \xrightarrow{\mathcal{F}} \frac{1}{\sqrt{2}\,|a|}\exp(-\frac{y^2}{4a^2} + iy\frac{b}{a})$$

$$\exp\left(-\left[\frac{1}{\sqrt{2}\sigma(t)}(x-\mu(t))\right]^2\right) \xrightarrow{\mathcal{F}} \frac{\left|\sqrt{2}\sigma(t)\right|}{\sqrt{2}}\exp(-\frac{y^2}{4\left(\frac{1}{\sqrt{2}\sigma(t)}\right)^2} - iy\mu(t))$$

Simplify, we obtain

$$\exp\left(-\frac{1}{2}\left(\frac{x-\mu(t)}{\sigma(t)}\right)^2\right) \xrightarrow{\mathcal{F}} \sigma(t)\exp(-\frac{1}{2}\sigma(t)\,y^2 - iy\mu(t))$$

Which is the same as (1). QED

4.4 Computing Assignment #2, Wed 2/27/2008

Grade: 2/2.

The limiting process simulation. Show that random walk final position is normally distributed in the limit under the Einstein-Weiner process (see problem 2 in this handout

Hand out Math 504
Monday 1/28/2008

Continuous Approximations To Random Walks

1. A Simple Random Walk Consider a particle that moves along the real line in such a way that, at each point in time, it makes one step to the right with probability p, and one step to the left with probability $q = 1 - p$. Assume that distance is measured in multiples of an amount Δx, and that time is measured in multiples of Δt. Let X_n denote the position of the particle after n steps. We shall assume the particle starts at the origin.

Let $\pi_j^{(n)} = P(X_n = j\Delta x)$. Thus, $\pi_j^{(n)}$ is the probability that the particle is located at $j\Delta x$ at time $n\Delta t$. Conditioning on the next state, we can write

$$\begin{aligned} P(X_{n+1} = j\Delta x) &= P(X_{n+1} = j\Delta x \mid X_n = (j-1)\Delta x)P(X_n = (j-1)\Delta x) \\ &+ P(X_{n+1} = j\Delta x \mid X_n = (j+1)\Delta x)P(X_n = (j+1)\Delta x) \,, \end{aligned}$$

or equivalently

$$\pi_j^{(n+1)} = p\pi_{j-1}^{(n)} + q\pi_{j+1}^{(n)} \,,$$

for $j = 0, \pm 1, \pm 2, \cdots$, and $n = 0, 1, \cdots$. Consider now a fixed position x and time t, subject to $x = j\Delta x$ and $t = n\Delta t$. Suppose that when Δx and Δt are small, we have the approximation

$$\pi_j^{(n)} \approx f(x,t)\Delta x \,,$$

where f is some function of x and t. Note that for each fixed t, the function f is a density function that describes probabilistically the location of the particle. From the recurrence formula above, we see that for such an approximation to hold, we need approximately

$$f(x, t + \Delta t) \cong pf(x - \Delta x, t) + qf(x + \Delta x, t) \,.$$

Assuming f is twice continuously differentiable, the Taylor series expansion yields

$$\begin{aligned} f(x,t) + \Delta t\frac{\partial f}{\partial t} + O(\Delta t)^2 &= p\left[f(x,t) - \Delta x\frac{\partial f}{\partial x} + \frac{1}{2}(\Delta x)^2\frac{\partial^2 f}{\partial x^2}\right] \\ &+ q\left[f(x,t) + \Delta x\frac{\partial f}{\partial x} + \frac{1}{2}(\Delta x)^2\frac{\partial^2 f}{\partial x^2}\right] + O(\Delta x)^3 \,, \end{aligned}$$

which, upon simplification, gives us

$$\frac{\partial f}{\partial t} = -(p-q)\frac{\Delta x}{\Delta t}\frac{\partial f}{\partial x} + \frac{1}{2}\frac{(\Delta x)^2}{\Delta t}\frac{\partial^2 f}{\partial x^2} + \frac{O(\Delta x)^3}{\Delta t} + O(\Delta t) \,.$$

In order to obtain a limiting equation, assume that for small Δx and Δt, there are constants β and D such that approximately,

$$(p-q)\frac{\Delta x}{\Delta t} = \beta \quad \text{and} \quad \frac{1}{2}\frac{(\Delta x)^2}{\Delta t} = D \,.$$

Recalling that $q = 1 - p$, these approximations tell us that in the limit we need

$$p = \frac{1}{2}\left(1 + \frac{\beta\Delta x}{2D}\right) \quad \text{and} \quad q = \frac{1}{2}\left(1 - \frac{\beta\Delta x}{2D}\right) .$$

Going to the limit then, we arrive at the partial differential equation

$$\frac{\partial f}{\partial t} = -\beta\frac{\partial f}{\partial x} + D\frac{\partial^2 f}{\partial x^2} .$$

Using the Fourier transform, this equation can be solved to get

$$f(x,t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] , \text{ where } \mu = \beta t , \text{ and } \sigma = \sqrt{2Dt} . \qquad (1)$$

The steps of this method are outlined in Problem 3 below.

There is an alternative way to reach this conclusion. After n steps, the particle will have made a certain number of steps to the right, say R_n, and a certain number of steps to the left, say L_n. Then $R_n + L_n = n$ and $X_n = (R_n - L_n)\Delta x$. Hence, $X_n = (2R_n - n)\Delta x$. Note that R_n is a binomial random variable with parameters n and p. Thus, the central limit theorem tells us that the distribution of R_n, and hence of X_n, is approximately normal. Further, the mean of position is

$$E(X_n) = (2E(R_n) - n)\Delta x = (2(np) - n)\Delta x = n\Delta x(2p-1) = n\frac{\beta(\Delta x)^2}{2D} ,$$

where we have used the formula for p above. Hence, the mean of position is

$$\left(n\frac{\beta\Delta x}{2D}\right)\Delta x = (n\Delta t)\frac{\beta(\Delta x)^2}{2D\Delta t} .$$

However, in the limit as n gets larger, we have $t = n\Delta t$ and $(\Delta x)^2/\Delta t = 2D$. Thus, the mean of position is simply $\mu = \beta t$.

Next, continuing this line of reasoning, we argue that the variance of position is

$$Var(X_n) = Var((2R_n - n))\Delta x) = Var(2R_n\Delta x) = 4(\Delta x)^2 Var(R_n) .$$

But the variance of R_n is npq. Hence, the variance of position is

$$4(\Delta x)^2 npq = 4(\Delta x)^2 n\frac{1}{2}\left(1 + \frac{\beta\Delta x}{2D}\right)\frac{1}{2}\left(1 - \frac{\beta\Delta x}{2D}\right) = \frac{(\Delta x)^2}{\Delta t}(n\Delta t)\left[1 - \left(\frac{\beta\Delta x}{2D}\right)^2\right] ,$$

where we have used the formulas for p and q above. Finally, going to the limit, and noting again that $t = n\Delta t$ and $2D = (\Delta x)^2/\Delta t$, we conclude that the variance of position is $\sigma^2 = 2Dt$. Thus, in the limit, the distribution of position is approximately normal with

mean $\mu = \beta t$ and variance $\sigma = \sqrt{2Dt}$. This result is the same as the one we obtained earlier through use of the partial differential equation.

A **stochastic process** is a family of random variables $X(t)$, where $X(t)$ represents the state of the process at time t. In our case, the state of the process is the position of the particle along the real line. A process is said to have **stationary increments** if for any t, the distribution of the **increment** $X(s+t) - X(s)$ depends only on t, the length of the time interval. Further, if the increments are independent for any set of disjoint intervals, the process is said to have **independent increments**.

Since the continuous process above, with transition distribution function (1), was derived as the limit of a discrete process that has stationary and independent increments, it is reasonable to expect that the limiting process would also have these two properties. A continuous process with transition distribution function (1), is called an **Einstein-Wiener process**. The parameter β is called the drift coefficient, and the parameter D is called the diffusion coefficient.

2. The Ornstein-Ehrenfest Model For a positive integer a, consider a random walk in which, at each point in time, if the process is at position $j\Delta x$, it moves one step to the right with probability $(a-j)/2a$ and one step to the left with probability $(a+j)/2a$, when $-a < j < a$. If $j = a$ then it moves to the left with probability 1, and if $j = -a$ it moves to the right with probability one.

Denote by $\pi_j^{(n)}$ the probability that the process is at point $j\Delta x$ at time $n\Delta t$. Then, by conditioning on the next state, we can write

$$\pi_j^{(n+1)} = \frac{a-j+1}{2a}\pi_{j-1}^{(n)} + \frac{a+j+1}{2a}\pi_{j+1}^{(n)} ,$$

for each $j = 1, 2, \cdots$, and $n = 0, 1, \cdots$. Consider this process in the limit when the bound a is large, and the Δx and Δt are small. For a fixed position x and time t, subject to $x = j\Delta x$ and $t = n\Delta t$, suppose we have the approximation

$$\pi_j^{(n)} \approx f(x,t)\Delta x ,$$

where f is some function of x and t. From the recurrence formula above, we see that for such an approximation to hold, we need approximately

$$f(x, t+\Delta t) = \frac{a-j+1}{2a} f(x-\Delta x, t) + \frac{a+j+1}{2a} f(x+\Delta x, t) .$$

Assuming f is twice continuously differentiable, the Taylor series expansion yields

$$\begin{aligned} f(x,t) + \Delta t \frac{\partial f}{\partial t} + O(\Delta t)^2 &= \frac{a-j+1}{2a}\left[f(x,t) - \Delta x \frac{\partial f}{\partial x} + \frac{1}{2}(\Delta x)^2 \frac{\partial^2 f}{\partial x^2} \right] \\ &+ \frac{a+j+1}{2a}\left[f(x,t) + \Delta x \frac{\partial f}{\partial x} + \frac{1}{2}(\Delta x)^2 \frac{\partial^2 f}{\partial x^2} \right] + O(\Delta x)^3 \ , \end{aligned}$$

which, upon simplification, gives

$$\frac{\partial f}{\partial t} = \frac{1}{a\Delta t} f(x,t) + \frac{x}{a\Delta t}\frac{\partial f}{\partial x} + \frac{a+1}{a}\frac{(\Delta x)^2}{2\Delta t}\frac{\partial^2 f}{\partial x^2} + \frac{O(\Delta x)^3}{\Delta t} + O(\Delta t) \ .$$

In order to obtain a limiting equation, assume that for small Δx and Δt, and large a, we have approximately,

$$a\Delta t = c^{-1} \quad \text{and} \quad \frac{1}{2}\frac{(\Delta x)^2}{\Delta t} = D \ ,$$

for some constants c and D. Going to the limit then, we arrive at the partial differential equation

$$\frac{\partial f}{\partial t} = c\frac{\partial (xf)}{\partial x} + D\frac{\partial^2 f}{\partial x^2} \ .$$

This equation is not so easily solved as in the previous case. However, under the boundary conditions

$$xf(x,t) \to 0 \ , \quad \frac{\partial f(x,t)}{\partial x} \to 0 \ \text{ as } \ x \to \pm\infty \ ,$$

and using the Fourier transform, the equation can be transformed into a first order, variable coefficient hyperbolic equation. This hyperbolic equation can then be solved using the method of characteristics. This method of solution is outlined in Problem 4 below.

Suppose the particle starts at a point x_0. Thus, the initial density of position is a dirac-delta function centered at the point x_0. Then the solution is found to be

$$f(x,t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma} \right)^2 \right] \ ,$$

where $\mu = x_0 e^{-ct}$ and

$$\sigma^2 = \frac{D}{c}\left(1 - e^{-2ct} \right) \ .$$

A continuous process with stationary and independent increments, and having this transitional distribution function, is called an **Ornstein-Ehrenfest process**.

Exercises

1. Derive the Einstein-Wiener process by noting that the position of the particle is $x = j\Delta x$, where $j = X_1 + X_2 + \cdots X_n$ with $t = n\Delta t$, and the X_i are independent and identically distributed random variables which have value $+1$ with probability p, and value -1 with probability $q = 1 - p$. Take $p = q = 1/2$.

2. (a) Use the formulation in the previous exercise to simulate the random walk for $p = q = 1/2$, and a specified diffusion coefficient D. Restrict Δx and Δt so that $D = (\Delta x)^2/2\Delta t$. (b) Use the simulation model to test that in the limit as $\Delta x \to 0$ and $\Delta t \to 0$, subject to $D = (\Delta x)^2/2\Delta t$, the distribution of position, for fixed time t and given D, is normal with mean 0 and variance $\sigma^2 = 2Dt$.

3. Solve the partial differential equation

$$\frac{\partial f}{\partial t} = -\beta \frac{\partial f}{\partial x} + D\frac{\partial^2 f}{\partial x^2} \ .$$

for the Einstein-Wiener process. Use the Fourier transform, and the following steps. (a) The Fourier transform of an absolutely integrable, and piecewise continuous function g on $(-\infty, \infty)$, is defined by

$$F[g](y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(x) e^{-iyx} dx \ .$$

Multiply the differential equation through by e^{-iyx} and integrate with respect to x over the interval $(-\infty, \infty)$. Then use integration by parts twice to obtain the equation

$$\frac{\partial \phi}{\partial t} = \left(-i\beta y - Dy^2\right) \phi(y,t) \ ,$$

where

$$\phi(y,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x,t) e^{-iyx} dx \ .$$

In doing the integrations by parts, assume that

$$f(x,t) \to 0 \ , \quad \frac{\partial f(x,t)}{\partial x} \to 0 \ \text{ as } \ x \to \pm\infty \ .$$

(b) Solve the differential equation for ϕ to get

$$\phi(y,t) = \phi(y,0) \exp\left(-i\beta t y - Dty^2\right) \ .$$

Note that $\phi(y,0)$ is the Fourier transform of $f(x,0)$, the density of the location of the particle at time $t = 0$.

(c) Now make use of the following two properties of the Fourier transform:

$$(1) \quad \text{For} \quad f(x) = e^{-x^2} \ , \quad F[f](y) = \frac{1}{\sqrt{2}} e^{-y^2/4} \ ,$$

and

$$(2) \quad \text{For} \quad g(x) = f(ax+b) \ , \ a \neq 0 \ , \quad F[g](y) = \frac{1}{|a|} e^{iy(b/a)} F[f](y/a) \ .$$

Thus conclude that

$$\text{For} \quad g(x) = e^{-(ax+b)^2} \ , \ a \neq 0 \ , \quad F[g](y) = \frac{1}{\sqrt{2}|a|} e^{iy(b/a)} e^{-y^2/4a^2} \ .$$

(d) Set $b/a = -\beta t$ and $1/4a^2 = Dt$, and use the previous results to conclude that

$$\phi(y,t) = \phi(y,0) \exp\left[\left(-i\beta y - Dy^2\right) t\right] = \phi(y,0) F[h(\cdot,t)](y) \ ,$$

where

$$h(x,t) = \frac{1}{\sigma(t)} \exp\left[-\frac{1}{2}\left(\frac{x-\mu(t)}{\sigma(t)}\right)^2\right] , \quad \mu(t) = \beta t \ , \quad \sigma(t) = \sqrt{2Dt} \ .$$

(e) Finally, the Fourier transform has the property that

$$F[f * g] = \sqrt{2\pi} F[f] F[g] \ , \quad \text{where} \quad f * g(x) = \int_{-\infty}^{\infty} f(u) g(x-u) du \ .$$

Use this property to show that

$$f(x,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(u,0) h(x-u,t) du \ .$$

In particular, if the particle starts that the origin, then we can view the initial density of position, $f(x,0)$, as a dirac delta function and thus deduce that in this case

$$f(x,t) = \frac{1}{\sqrt{2\pi}} h(x,t) \ ,$$

which is the expression (1) above.

4. Solve the partial differential equation

$$\frac{\partial f}{\partial t} = c \frac{\partial (xf)}{\partial x} + D \frac{\partial^2 f}{\partial x^2} \ .$$

for the Ornstein-Ehrenfest process. Use the Fourier transform (see definition in previous problem), and the following steps. (a) Multiply the differential equation

through by e^{-iyx} and integrate with respect to x over the interval $(-\infty, \infty)$. Then use integration by parts three times to obtain the equation

$$\frac{\partial \phi}{\partial t} = -cy\frac{\partial \phi}{\partial y} - Dy^2\phi \ ,$$

where

$$\phi(y,t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x,t)e^{-iyx}dx \ .$$

In doing the integrations by parts, assume that

$$xf(x,t) \to 0 \ , \quad \text{and} \quad \frac{\partial f(x,t)}{\partial x} \to 0 \ , \quad \text{as} \ \ x \to \pm\infty \ .$$

(b) Introduce the integrating factor

$$I(y) = \exp\left(\frac{D}{2c}y^2\right)$$

and show that $u(y,t) = I(y)\phi(y,t)$ satisfies the hyperbolic equation

$$\frac{\partial u}{\partial t} + cy\frac{\partial u}{\partial y} = 0 \ ,$$

subject to the initial condition $u(y,0) = u_0(y) = I(y)\phi(y,0)$ for $-\infty < y < \infty$. Note that $\phi(y,0)$ is the Fourier transform of $f(x,0)$, the density of the location of the particle at time $t = 0$.

(c) Use the method of characteristics to solve the previous hyperbolic equation and conclude that

$$u(y,t) = u_0\left(ye^{-ct}\right) \ , \quad \text{for} \ -\infty < y < \infty \ \text{ and } \ t \geq 0 \ .$$

Then, using the definitions of $u(y,t)$ and $u_0(y)$, deduce that

$$\phi(y,t) = \exp\left(-\frac{1}{2}\sigma^2(t)y^2\right)\phi\left(ye^{-ct}, 0\right) \ , \quad \text{where} \ \ \sigma^2(t) = \frac{D}{c}\left(1 - e^{-2ct}\right) \ .$$

(d) To simplify the analysis, assume that the particle starts at a point x_0. Thus, the initial density of position is a dirac-delta function centered at the point x_0. It follows that

$$\phi(y,0) = \frac{1}{\sqrt{2\pi}}e^{-iyx_0} \ .$$

In this case, use the previous result to conclude that

$$\phi(y,t) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\sigma^2(t)y^2 - iy\mu(t)\right) \ , \quad \text{where} \ \ \mu(t) = x_0e^{-ct} \ .$$

(e) Set $b/a = -\mu(t)$ and $1/4a^2 = (1/2)\sigma^2(t)$, and use the result of part (c), Problem 3, to conclude that

$$\phi(y,t) = F[r(\cdot,t)](y) \ ,$$

where

$$r(x,t) = \frac{1}{\sigma(t)\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu(t)}{\sigma(t)}\right)^2\right] \ .$$

Thus, $f(x,t) = r(x,t)$, which is the solution given above for the Ornstein-Ehrenfest model.

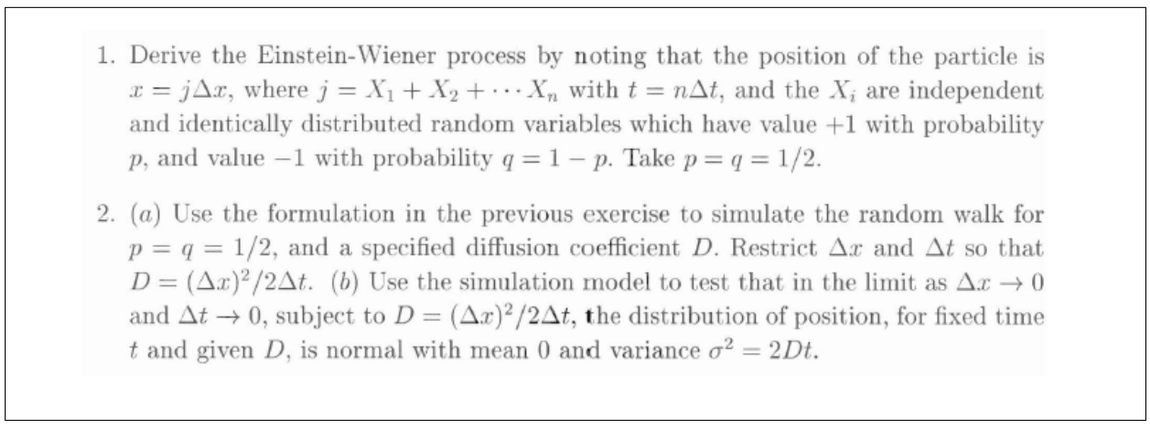
4.4.1 Simulation

A Matlab function was written to simulate a random walk based on the Einstein-Wiener process to verify that the distribution of the final position of the walk is normally distributed with mean $= \beta t$ and variance $= 2Dt$ (more details in the report).

4.4.2 Purpose and design of project

Nature of the project

We are solving problem #2 as described in the following screen shot (taken from the class handout)



1. Derive the Einstein-Wiener process by noting that the position of the particle is $x = j\Delta x$, where $j = X_1 + X_2 + \cdots X_n$ with $t = n\Delta t$, and the X_i are independent and identically distributed random variables which have value $+1$ with probability p, and value -1 with probability $q = 1 - p$. Take $p = q = 1/2$.

2. (a) Use the formulation in the previous exercise to simulate the random walk for $p = q = 1/2$, and a specified diffusion coefficient D. Restrict Δx and Δt so that $D = (\Delta x)^2/2\Delta t$. (b) Use the simulation model to test that in the limit as $\Delta x \to 0$ and $\Delta t \to 0$, subject to $D = (\Delta x)^2/2\Delta t$, the distribution of position, for fixed time t and given D, is normal with mean 0 and variance $\sigma^2 = 2Dt$.

Short background on the problem: In this project we are asked to verify an analytical result derived in a handout given in the class called 'Continuos approximation to random walk'.

A random walk is formulated, by proposing that $\pi_j^{(n)}$which is the probability that the position of a particle at $x = j\Delta x$ and at time $n\Delta t$ can be expressed as $f(x,t)\,\Delta x$, where $f(x,t)$ represents a density per unit length, which gives a measure of the particle being at that position x at time t.

Starting with this and applying a limiting argument lead to a partial differential equation whose solution is the normal distribution function with certain mean and variance. However, the condition for arriving at the PDE was that as we make Δt and Δx small, we needed to keep the ratio $\frac{(\Delta x)^2}{\Delta t}$constant.

In this assignment, we simulate a random walk as Δt and Δx are made smaller and smaller subject to this same condition to verify if the distribution of the final position of the random walk converges to the solution of the PDE which is normal distribution and if the converged distribution will have the same variance of $2Dt$ and same mean of βt as does the solution of the PDE.

The details of the theoretical derivation is shown in the above mentioned handout. A diagram below is made to help illustrate the overall purpose of this assignment. In this assignment, we are working on the flow shown on the right side below.

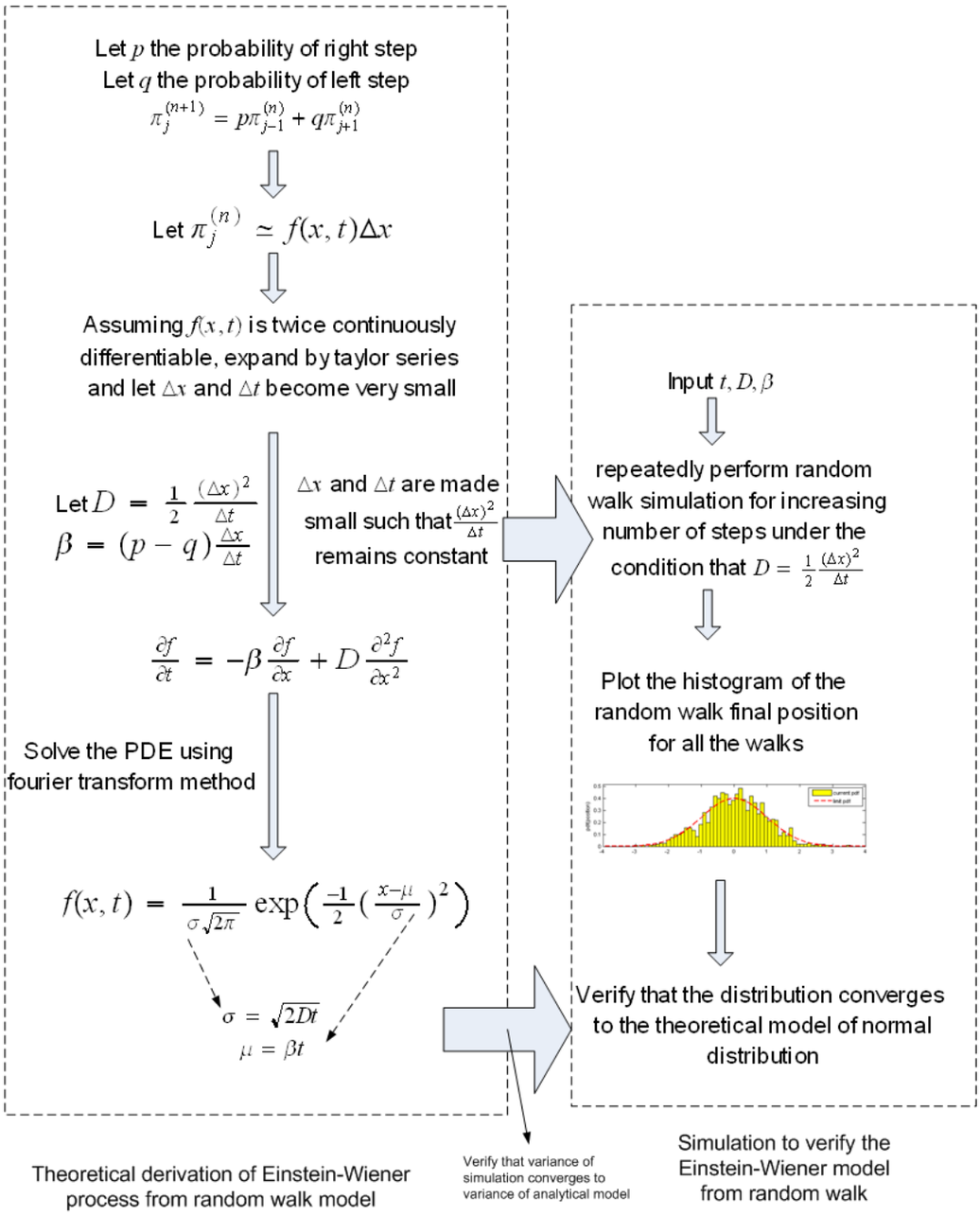


Figure 4.1: Random walk simulation to verify the Einstein-Wiener analytical derivation

Questions we are investigating

These are the questions we are trying to answer in this project

1. Does the distribution of the random walk final position generated by increasing the number of steps for fixed t (total time of the random walk) while keeping the ratio $\frac{(\Delta x)^2}{\Delta t}$constant (equal to $2D$), converges to a normal distribution (which is the solution of the Einstein-Wiener process model)?

2. Does the variance of the above distribution converges, as $\Delta t \to 0$ and $\Delta x \to 0$ under the above mentioned condition of keeping $\frac{(\Delta x)^2}{\Delta t}$, to the analytical variance of $2Dt$ and the theoretical mean of βt?

Few words on the program

The input to the program is t, D, β where t is the total random walk time and D, β represents the terms as shown in the diagram above.

A distribution of the final random walk position is generated by running the random walk simulation a number of times (called the sample size). In each such run, we use a specific number of steps. The number of steps is increased, and we generate another distribution. We keep doing this and plot each distribution as the number of steps is increased.

At the end of the simulation, to verify that the distribution in the limit is normal. A quantile-quantile plot is made to compare the generated histogram with the theoretical standard normal distribution to see if the result is close to a straight line or not. Also a plot is made showing the convergence of the variance of the current distribution as number of steps is increased by keeping track of the relative error in the variance. In addition, the RMS error between the standard normal and the current distribution is calculated and plotted as a function of delta(T) as delta(T) is made smaller and smaller. The program is written in Matlab version 2007a and uses the statistics toolbox.

The following is a description of the algorithm of the program

We simulate a random walk, where each step made is either to the left or to the right with probability q and p respectively.

Let Y_i be either 1 or -1 depending if we make a right or a left step. Hence

$$Y_i = \begin{cases} 1 & probability\ p \\ -1 & probability\ q \end{cases}$$

and now if we let $X_n = Y_1 + Y_2 + \cdots + Y_n$ then the final position of the random walk can be written as

$$X_n = \Delta x \sum_{j=1}^{n} Y_j$$

where Δx is the step size. The step size is found by solving $\Delta x = \sqrt{2D\Delta t}$ where D is the diffusion parameter which is an input, and Δt is the current time step found by dividing the total simulation fixed time t, which is an input, by the current number of steps n.

$$\Delta t = \frac{t}{n}$$

This program handles a general value for β other than zero. To be able to accomplish this, we need to determine the correct starting step size n to avoid the problem with coming up with a value for the probability p being larger than 1. So, this was done in the initialization stage using this formula

$$starting\ n = round\left(\frac{t\beta^2}{2D}\right) + 1$$

And the simulation was started from the above n and not from 1.

A note about the quantile-quantile plot

To answer the first question of this simulation, which is to determine if the final position distribution converges to normal distribution with mean βt and variance $2Dt$, a quantile plot was used. In this plot, the quantile for the standard normal distribution was plotted against the quantile of the distribution of the final position.

The $x-axis$ of the quantile-quantile plot was found as follows

$$\begin{aligned} n &= sample\_size \\ x &= F^{-1}(([1:n]-0.5)/n \end{aligned}$$

Where F^{-1} is the inverse of the CDF for the standard normal distribution (the matlab function norminv() was used for this). While the $y-axis$ is the quantile of the actual data (the sample data of the final distribution of the random walk position). This was found by sorting the data from small to large and then using the resulting sorted vector as the y values. Notice that the distribution was already standardized using

$$z(i) = \frac{y(i)-\mu}{\sigma}$$

Where $\mu = \beta t$ and $\sigma = \sqrt{2Dt}$,

4.4.3 Summary of numerical results

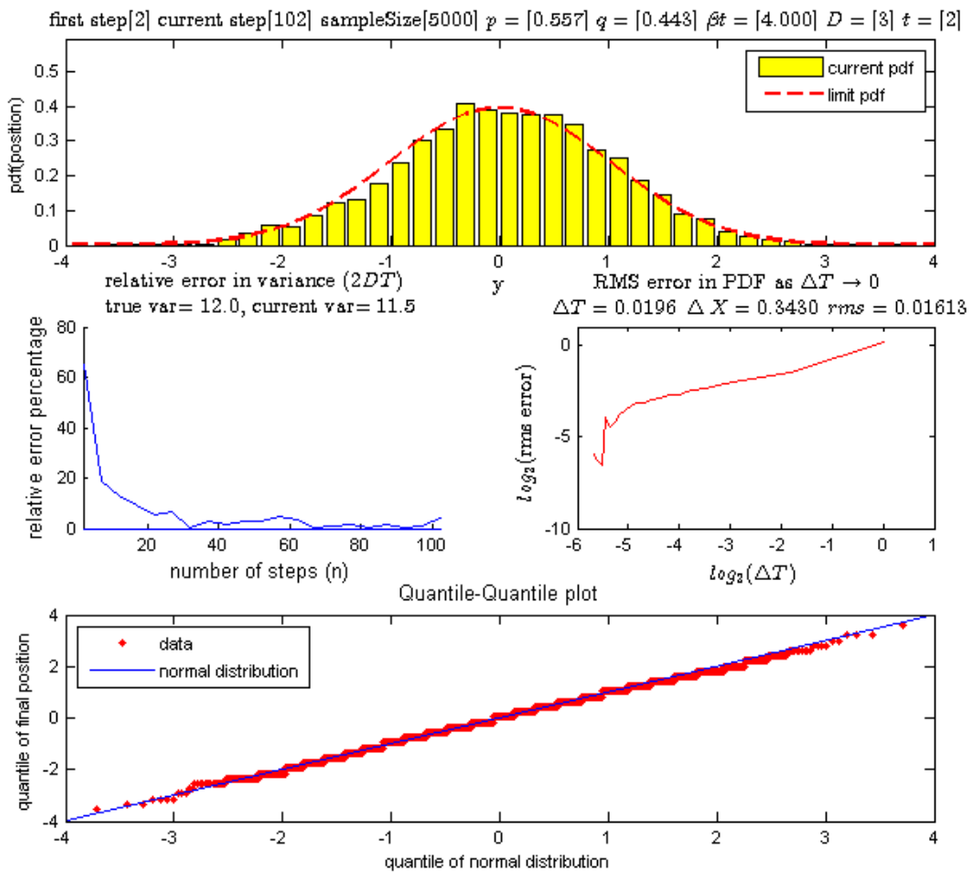
A number of experiments were performed for different input parameters. The table below lists the variance of the distribution of the final position as the number of steps is increased. The run parameters are also shown

Experiment #1 $\beta = 2, t = 2, D = 3, n = 100$

starting step number= 2, $\beta = 2, t = 2, D = 3, final\ p = 0.557, final\ q = 0.443$

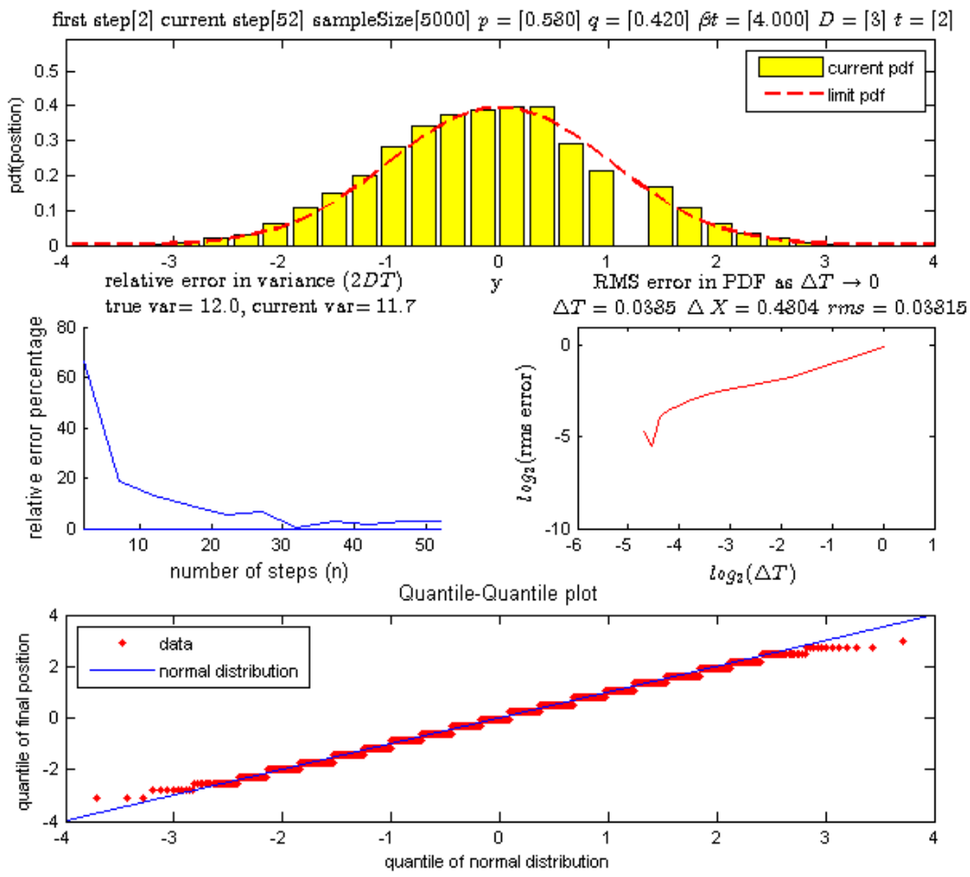
sample size 5000, number of bins 40, seed= 123456

n (number of steps)	Variance	True variance (2Dt)	Δt
2	3.92	12	1
7	9.73	12	0.2857
12	10.43	12	0.1667
17	10.9	12	0.1176
22	11.37	12	0.0909
27	11.19	12	0.0741
32	12.02	12	0.0625
...	...	...	...
67	12.05	12	0.0299
72	11.89	12	0.0278
77	12.16	12	0.0260
82	11.99	12	0.0244
87	11.78	12	0.0230
92	12.03	12	0.0217
97	11.88	12	0.0206
102	11.47	12	0.0196



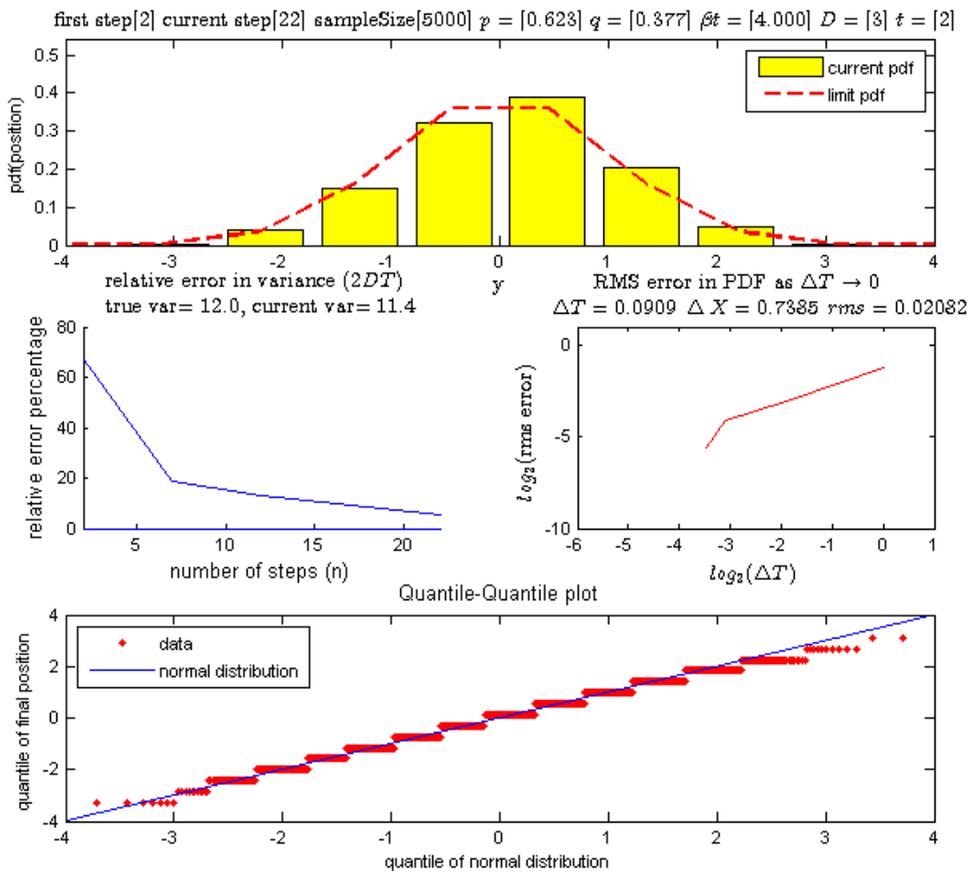
Experiment #2 $\beta = 2, t = 2, D = 3, n = 50$

Since the parameters t, D, β, then running for $n = 50$ will produce the same numerical values already contained in the first experiment when looking at the table above up to $n = 50$ (the program starts by seeding the random number generator, so nothing will change here and we will just produce a subset of the result already produced in first experiment). So I will just show the final plot, showing the convergence of the histogram and the quantile-quantile plot



Experiment #3 $\beta = 2, t = 2, D = 3, n = 20$

Again, as described at the start of experiment 2 above, this is a subset of the first experiment. We will show the final plot only to show how close to the standard normal the final position histogram is.



Experiment #4 $\beta = 2, t = 2, D = 3,\ n = 7000$

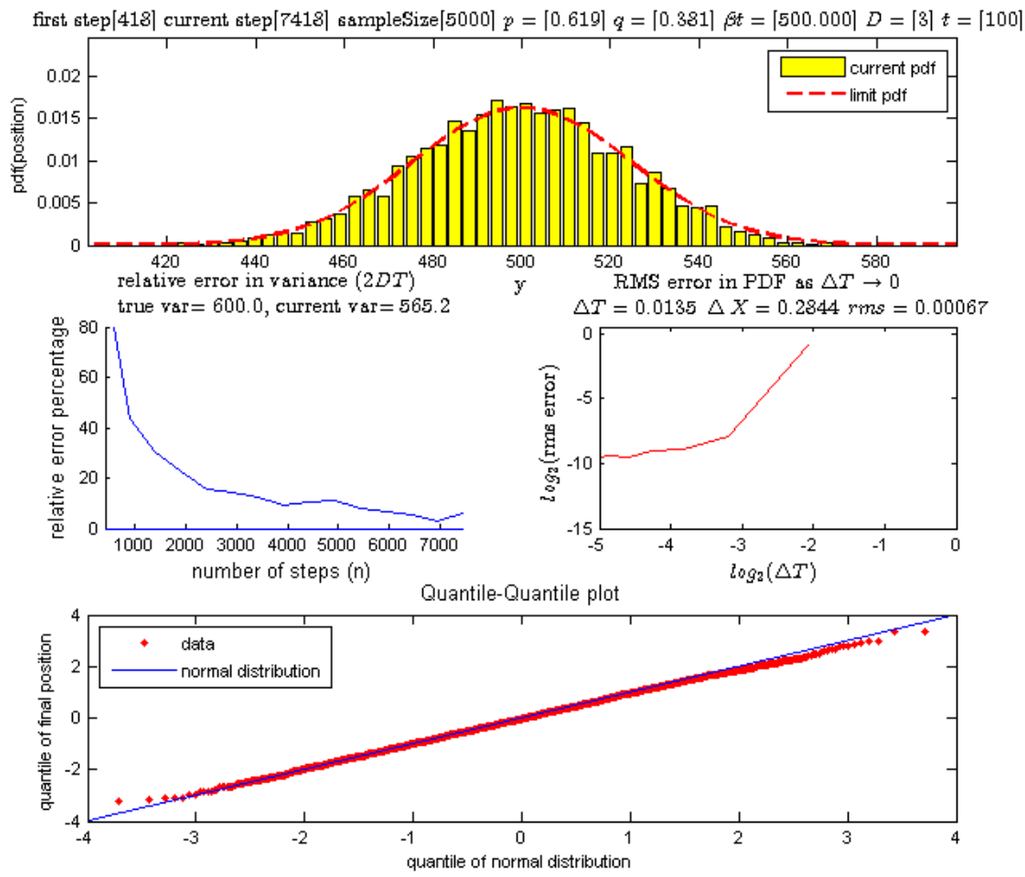
The following 2 experiments are not required to do, but they are extra experiments I already done and included here.

starting step number= 400, $\beta = 5, t = 100, D = 3, final\ p = 0.623, final\ q = .377$

sample size 5000, number of bins 60, seed= 123456

final $\Delta x = 0.2945$ final $\Delta t = 0.0145$

Experiment number	n (number of steps)	Variance	True variance (2Dt)	Δt
1	400	1.89	600	0.2392
2	900	340	600	0.1089
3	1400	420	600	0.0705
4	1900	464	600	0.0521
5	2400	504	600	0.0414
6	2900	514	600	0.0343
7	3400	525	600	0.0293
8	3900	546	600	0.0255
9	4400	536	600	0.0226
10	4900	533	600	0.0203
11	5400	552	600	0.0185
12	5900	558	600	0.0169
13	6400	567	600	0.0156
14	6900	583	600	0.0145



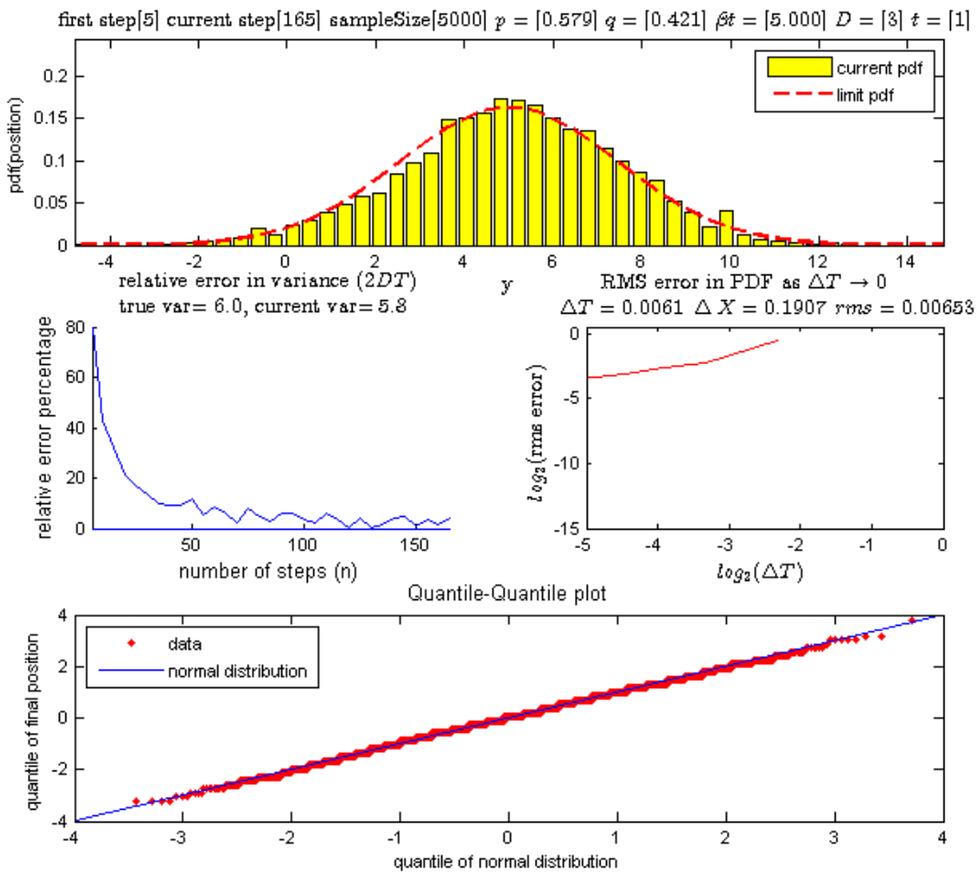
Experiment #5 $n = 160, \beta = 5, t = 1, D = 3$

starting step number= $5, \beta = 5, t = 1, D = 3, final\ p = 0.579, final\ q = 0.421$

sample size 5000, number of bins 50, seed= 123456

final $\Delta x = 0.1907$, final $\Delta t = 0.0061$

Experiment number	n (number of steps)	Variance	True variance ($2Dt$)	Δt
1	5	1.019	6	0.2
2	10	3.4	6	0.1
3	15	4.09	6	0.0667
4	20	4.74	6	0.05
5	25	5	6	0.4
6	30	5.18	6	0.0333
7	35	5.43	6	0.0286
8	40	5.466	6	0.0250
9	45	5.3	6	0.0222
10	50	5.66	6	0.02
11	55	5.4	6	0.0182
12	60	5.85	6	0.0167
...	...	...	...	...
31	150	5.78	6	0.0065
32	155	5.909	6	0.0063
33	160	5.75	6	0.0061



4.4.4 Discussion of numerical results

From the above tables we observe that as Δt becomes smaller, the variance of the sample of the final position becomes closer to the variance predicted by the model which is $2Dt$.

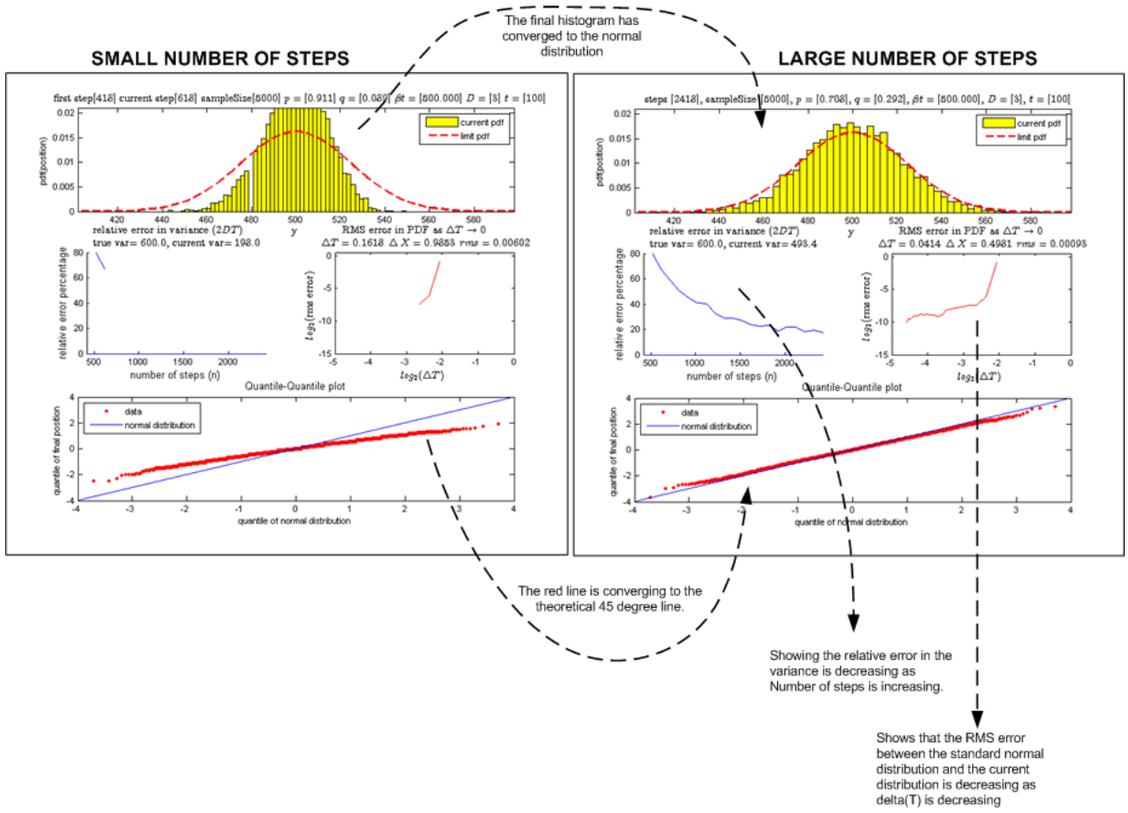
The mean remains the same which is βt.

We observe that if the total walk time is large (experiment #4) , then more steps are needed to bring Δt to be small enough so that the variance becomes close to $2Dt$.. This answers the second question we are set to solve in this project which is Does the variance of the above distribution converges, as $\Delta t \to 0$ and $\Delta x \to 0$ under the above mentioned condition of keeping $\frac{(\Delta x)^2}{\Delta t}$, to the analytical variance of $2Dt$ and the theoretical mean of βt?

Now to answer the first question of convergence of the histogram of the final position to the normal.

Looking at the quantile plots we observe that as more steps are used (hence smaller Δt and smaller Δx) then the quantile-quantile plot was tilting closer and closer to the straight line at 45^0 which would be the case when we plot the quantile of 2 data sets coming from the same distribution. This concludes that the final distribution of the random walk position converges to normal distribution with the above parameters.

The following diagram below shows a run where on the left side there is a plot showing the quantile plot when the number of steps is small. The plot on the right side shows the quantile plot at the end of the run when n was large. We see that the quantile plot line is now almost exactly over the 45^0 line, confirming that the data is coming from normal distribution.



Therefore, we have answered the 2 questions this simulation was designed to answer.

4.4.5 Final observation

In doing the above experiments, it was observed that the relative error in the variance of the final position as n increased does approach the true variance $2Dt$ but the convergence is not smooth. As the relative error (around 5% to 10%), then increasing n more can cause the error to sometimes increase and not decrease as one would expect. Meaning the relative error is not monotonic decreasing as n increases. However, as n becomes very large, the trend is for the relative error is to decrease. I can only contribute this behavior to some sort of statistical error. This needs to be investigated more.

4.4.6 Graded

Limiting process for Einstein-Wiener random walk simulation.

Computer assignment #2, Math 504, CSUF, spring 2008

by Nasser Abbasi

February 26, 2008

5/5 Excellent!

Contents

1 Purpose and design of project

1.1 Nature of the project

We are solving problem #2 as described in the following screen shot (taken from the class handout)

1. Derive the Einstein-Wiener process by noting that the position of the particle is $x = j\Delta x$, where $j = X_1 + X_2 + \cdots X_n$ with $t = n\Delta t$, and the X_i are independent and identically distributed random variables which have value $+1$ with probability p, and value -1 with probability $q = 1 - p$. Take $p = q = 1/2$.

2. (a) Use the formulation in the previous exercise to simulate the random walk for $p = q = 1/2$, and a specified diffusion coefficient D. Restrict Δx and Δt so that $D = (\Delta x)^2/2\Delta t$. (b) Use the simulation model to test that in the limit as $\Delta x \to 0$ and $\Delta t \to 0$, subject to $D = (\Delta x)^2/2\Delta t$, the distribution of position, for fixed time t and given D, is normal with mean 0 and variance $\sigma^2 = 2Dt$.

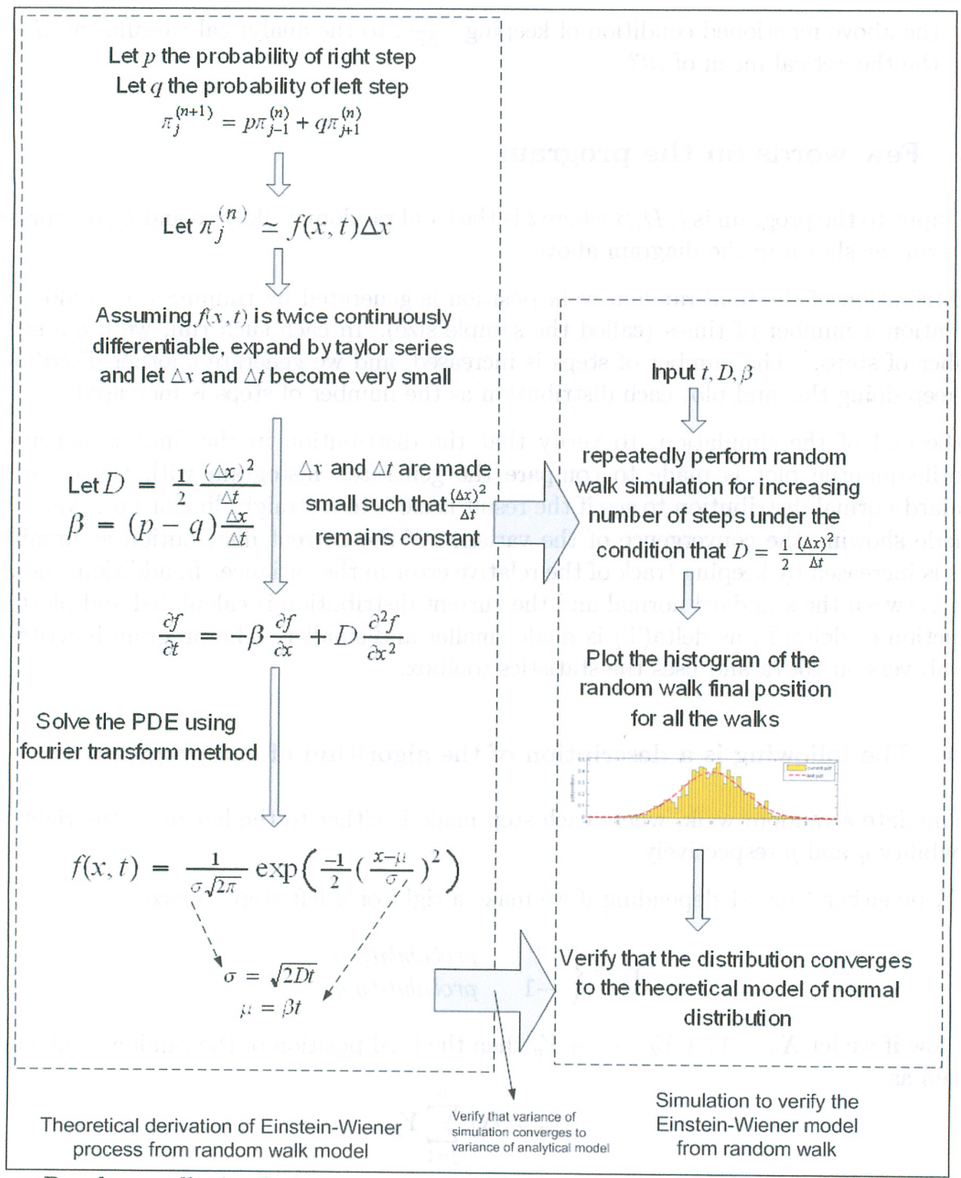
Short background on the problem: In this project we are asked to verify an analytical result derived in a handout given in the class called 'Continuos approximation to random walk'.

A random walk is formulated, by proposing that $\pi_j^{(n)}$ which is the probability that the position of a particle at $x = j\Delta x$ and at time $n\Delta t$ can be expressed as $f(x,t)\,\Delta x$, where $f(x,t)$ represents a density per unit length, which gives a measure of the particle being at that position x at time t.

Starting with this and applying a limiting argument lead to a partial differential equation whose solution is the normal distribution function with certain mean and variance. However, the condition for arriving at the PDE was that as we make Δt and Δx small, we needed to keep the ratio $\frac{(\Delta x)^2}{\Delta t}$ constant.

In this assignment, we simulate a random walk as Δt and Δx are made smaller and smaller subject to this same condition to verify if the distribution of the final position of the random walk converges to the solution of the PDE which is normal distribution and if the converged distribution will have the same variance of $2Dt$ and same mean of βt as does the solution of the PDE.

The details of the theoretical derivation is shown in the above mentioned handout. A diagram below is made to help illustrate the overall purpose of this assignment. In this assignment, we are working on the flow shown on the right side below.



Random walk simulation to verify the Einstein-Wiener analytical derivation

1.2 Questions we are investigating

These are the questions we are trying to answer in this project

1. Does the distribution of the random walk final position generated by increasing the number of steps for fixed t (total time of the random walk) while keeping the ratio $\frac{(\Delta x)^2}{\Delta t}$ constant (equal to $2D$), converges to a normal distribution (which is the solution of the Einstein-Wiener process model)?

2. Does the variance of the above distribution converges, as $\Delta t \to 0$ and $\Delta x \to 0$ under

the above mentioned condition of keeping $\frac{(\Delta x)^2}{\Delta t}$, to the analytical variance of $2Dt$ and the theoretical mean of βt?

1.3 Few words on the program

The input to the program is t, D, β where t is the total random walk time and D, β represents the terms as shown in the diagram above.

A distribution of the final random walk position is generated by running the random walk simulation a number of times (called the sample size). In each such run, we use a specific number of steps. The number of steps is increased, and we generate another distribution. We keep doing this and plot each distribution as the number of steps is increased.

At the end of the simulation, to verify that the distribution in the limit is normal. A quantile-quantile plot is made to compare the generated histogram with the theoretical standard normal distribution to see if the result is close to a straight line or not. Also a plot is made showing the convergence of the variance of the current distribution as number of steps is increased by keeping track of the relative error in the variance. In addition, the RMS error between the standard normal and the current distribution is calculated and plotted as a function of delta(T) as delta(T) is made smaller and smaller. The program is written in Matlab version 2007a and uses the statistics toolbox.

1.3.1 The following is a description of the algorithm of the program

We simulate a random walk, where each step made is either to the left or to the right with probability q and p respectively.

Let Y_i be either 1 or -1 depending if we make a right or a left step. Hence

$$Y_i = \begin{cases} 1 & \textit{probability } p \\ -1 & \textit{probability } q \end{cases}$$

and now if we let $X_n = Y_1 + Y_2 + \cdots + Y_n$ then the final position of the random walk can be written as

$$X_n = \Delta x \sum_{j=1}^{n} Y_j$$

where Δx is the step size. The step size is found by solving $\Delta x = \sqrt{2D\Delta t}$ where D is the diffusion parameter which is an input, and Δt is the current time step found by dividing the total simulation fixed time t, which is an input, by the current number of steps n.

$$\Delta t = \frac{t}{n}$$

This program handles a general value for β other than zero. To be able to accomplish this, we need to determine the correct starting step size n to avoid the problem with coming up with a value for the probability p being larger than 1. So, this was done in the initialization stage using this formula

$$starting\ n = round\left(\frac{t\beta^2}{2D}\right) + 1$$

And the simulation was started from the above n and not from 1.

1.3.2 A note about the quantile-quantile plot

To answer the first question of this simulation, which is to determine if the final position distribution converges to normal distribution with mean βt and variance $2Dt$, a quantile plot was used. In this plot, the quantile for the standard normal distribution was plotted against the quantile of the distribution of the final position.

The $x - axis$ of the quantile-quantile plot was found as follows

$$\begin{aligned} n &= sample\_size \\ x &= F^{-1}(([1:n] - 0.5)/n \end{aligned}$$

Where F^{-1} is the inverse of the CDF for the standard normal distribution (the matlab function norminv() was used for this). While the $y - axis$ is the quantile of the actual data (the sample data of the final distribution of the random walk position). This was found by sorting the data from small to large and then using the resulting sorted vector as the y values. Notice that the distribution was already standardized using

$$z(i) = \frac{y(i) - \mu}{\sigma}$$

Where $\mu = \beta t$ and $\sigma = \sqrt{2Dt}$,

2 Summary of numerical results

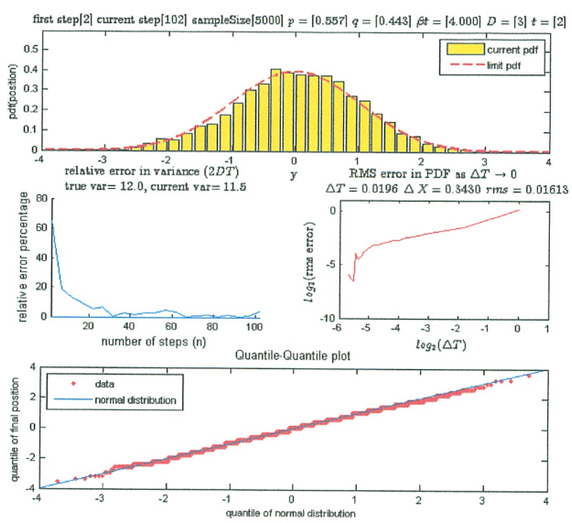
A number of experiments were performed for different input parameters. The table below lists the variance of the distribution of the final position as the number of steps is increased. The run parameters are also shown

2.1 Experiment #1 $\beta = 2, t = 2, D = 3, n = 100$

starting step number= 2, $\beta = 2, t = 2, D = 3, final\ p = 0.557, final\ q = 0.443$

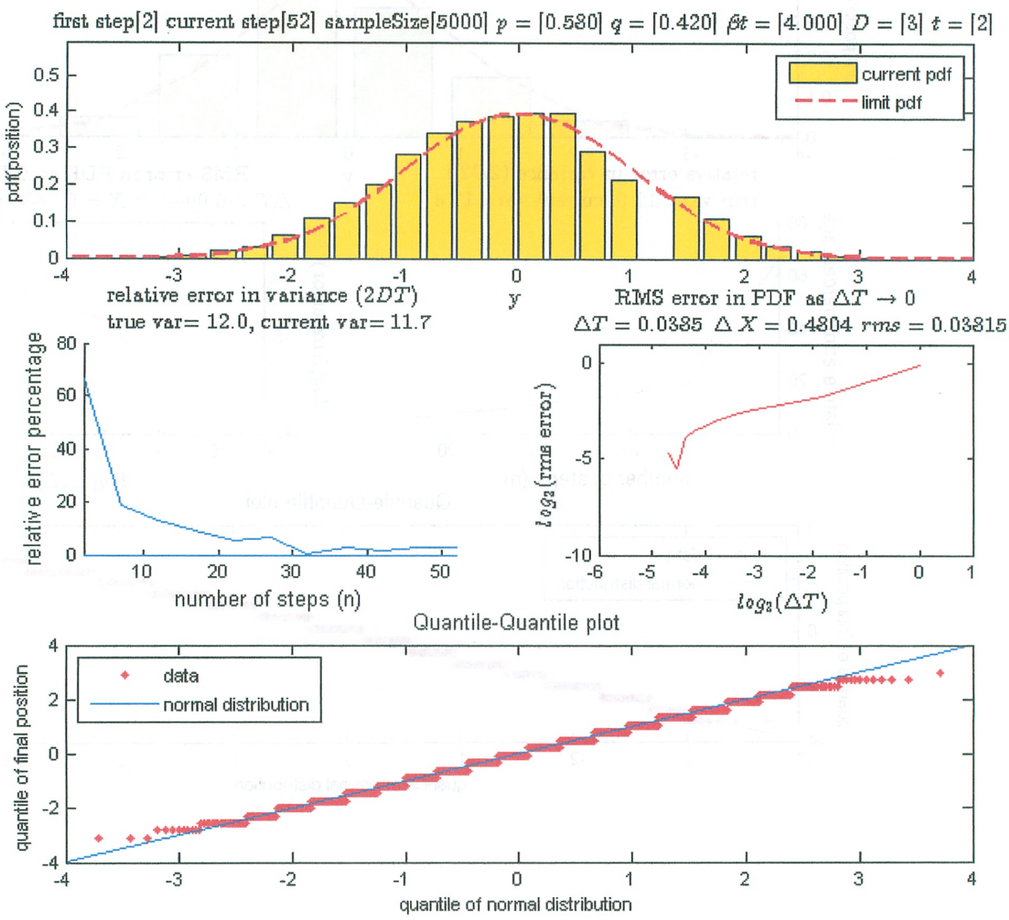
sample size 5000, number of bins 40, seed= 123456

n (number of steps)	Variance	True variance (2Dt)	Δt
2	3.92	12	1
7	9.73	12	0.2857
12	10.43	12	0.1667
17	10.9	12	0.1176
22	11.37	12	0.0909
27	11.19	12	0.0741
32	12.02	12	0.0625
⋮	...	...	...
67	12.05	12	0.0299
72	11.89	12	0.0278
77	12.16	12	0.0260
82	11.99	12	0.0244
87	11.78	12	0.0230
92	12.03	12	0.0217
97	11.88	12	0.0206
102	11.47	12	0.0196



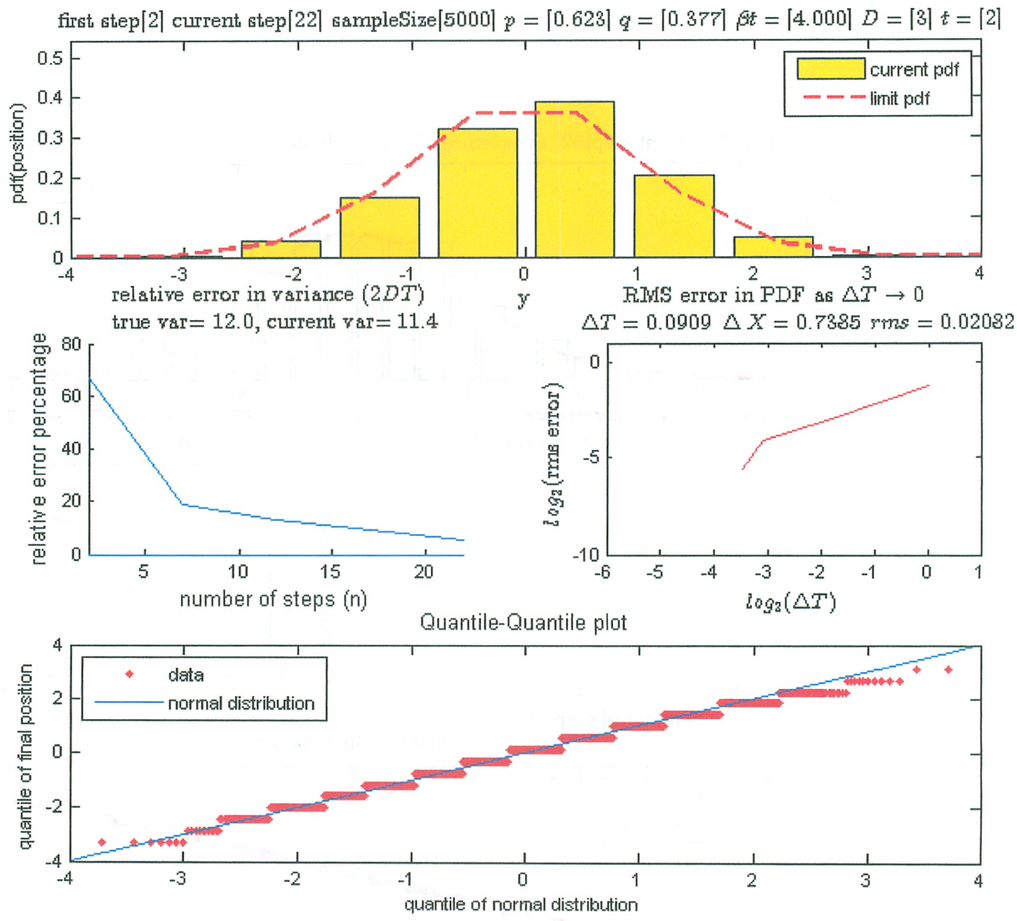
2.2 Experiment #2 $\beta = 2, t = 2, D = 3, n = 50$

Since the parameters t, D, β, then running for $n = 50$ will produce the same numerical values already contained in the first experiment when looking at the table above up to $n = 50$ (the program starts by seeding the random number generator, so nothing will change here and we will just produce a subset of the result already produced in first experiment). So I will just show the final plot, showing the convergence of the histogram and the quantile-quantile plot



2.3 Experiment #3 $\beta = 2, t = 2, D = 3, n = 20$

Again, as described at the start of experiment 2 above, this is a subset of the first experiment. We will show the final plot only to show how close to the standard normal the final position histogram is.



2.4 Experiment #4 $\beta = 2, t = 2, D = 3$, $n = 7000$

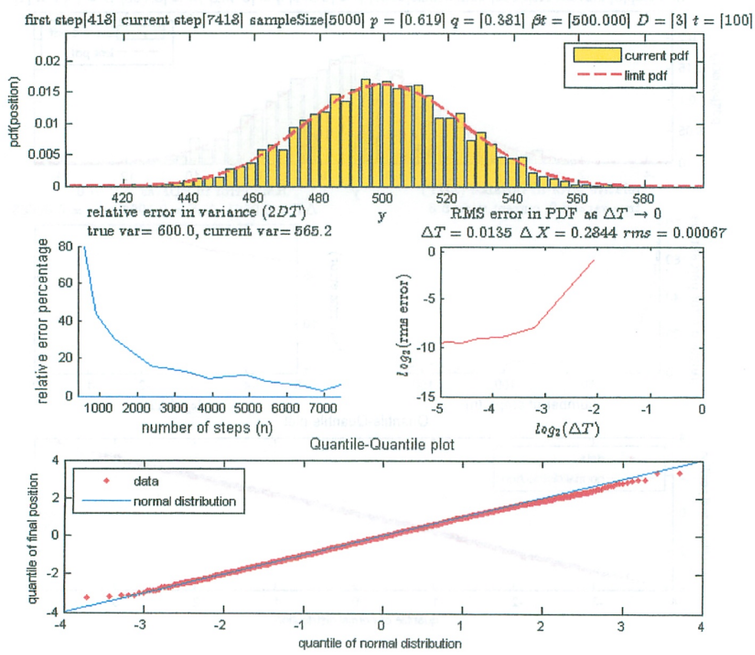
The following 2 experiments are not required to do, but they are extra experiments I already done and included here.

starting step number= 400, $\beta = 5, t = 100, D = 3$, $final\ p = 0.623$, $final\ q = .377$

sample size 5000, number of bins 60, seed= 123456

final $\Delta x = 0.2945$ final $\Delta t = 0.0145$

Experiment number	n (number of steps)	Variance	True variance (2Dt)	Δt
1	400	1.89	600	0.2392
2	900	340	600	0.1089
3	1400	420	600	0.0705
4	1900	464	600	0.0521
5	2400	504	600	0.0414
6	2900	514	600	0.0343
7	3400	525	600	0.0293
8	3900	546	600	0.0255
9	4400	536	600	0.0226
10	4900	533	600	0.0203
11	5400	552	600	0.0185
12	5900	558	600	0.0169
13	6400	567	600	0.0156
14	6900	583	600	0.0145



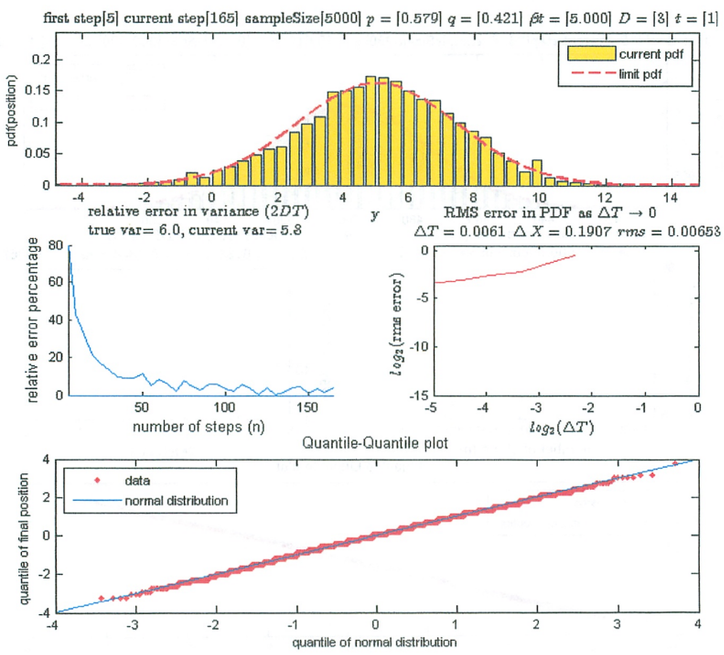
2.5 Experiment #5 $n = 160, \beta = 5, t = 1, D = 3$

starting step number= $5, \beta = 5, t = 1, D = 3, final\ p = 0.579, final\ q = 0.421$

sample size 5000, number of bins 50, seed= 123456

final $\Delta x = 0.1907$, final $\Delta t = 0.0061$

Experiment number	n (number of steps)	Variance	True variance $(2Dt)$	Δt
1	5	1.019	6	0.2
2	10	3.4	6	0.1
3	15	4.09	6	0.0667
4	20	4.74	6	0.05
5	25	5	6	0.4
6	30	5.18	6	0.0333
7	35	5.43	6	0.0286
8	40	5.466	6	0.0250
9	45	5.3	6	0.0222
10	50	5.66	6	0.02
11	55	5.4	6	0.0182
12	60	5.85	6	0.0167
...	...	...	...	...
31	150	5.78	6	0.0065
32	155	5.909	6	0.0063
33	160	5.75	6	0.0061



3 Discussion of numerical results

From the above tables we observe that as Δt becomes smaller, the variance of the sample of the final position becomes closer to the variance predicted by the model which is $2Dt$.

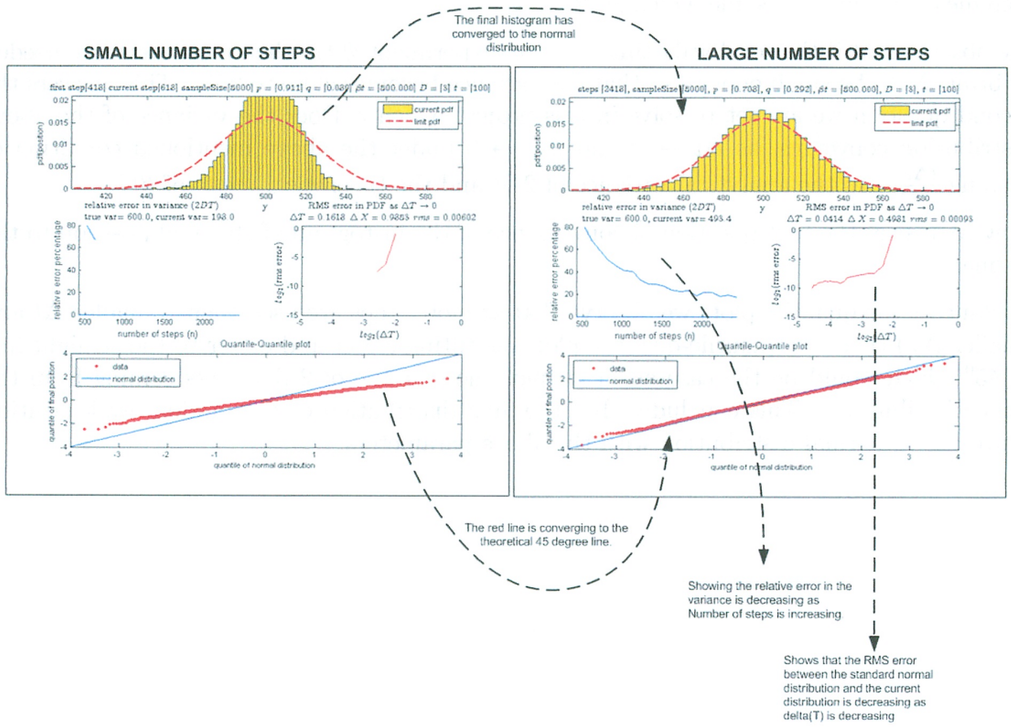
The mean remains the same which is βt.

We observe that if the total walk time is large (experiment #4) , then more steps are needed to bring Δt to be small enough so that the variance becomes close to $2Dt$.. This answers the second question we are set to solve in this project which is Does the variance of the above distribution converges, as $\Delta t \rightarrow 0$ and $\Delta x \rightarrow 0$ under the above mentioned condition of keeping $\frac{(\Delta x)^2}{\Delta t}$, to the analytical variance of $2Dt$ and the theoretical mean of βt?

Now to answer the first question of convergence of the histogram of the final position to the normal.

Looking at the quantile plots we observe that as more steps are used (hence smaller Δt and smaller Δx) then the quantile-quantile plot was tilting closer and closer to the straight line at 45^0 which would be the case when we plot the quantile of 2 data sets coming from the same distribution. This concludes that the final distribution of the random walk position converges to normal distribution with the above parameters.

The following diagram below shows a run where on the left side there is a plot showing the quantile plot when the number of steps is small. The plot on the right side shows the quantile plot at the end of the run when n was large. We see that the quantile plot line is now almost exactly over the 45^0 line, confirming that the data is coming from normal distribution.



Therefore, we have answered the 2 questions this simulation was designed to answer.

3.1 Final observation

In doing the above experiments, it was observed that the relative error in the variance of the final position as n increased does approach the true variance $2Dt$ but the convergence is not smooth. As the relative error (around 5% to 10%), then increasing n more can cause the error to sometimes increase and not decrease as one would expect. Meaning the relative error is not monotonic decreasing as n increases. However, as n becomes very large, the trend is for the relative error is to decrease. I can only contribute this behavior to some sort of statistical error. This needs to be investigated more.

```
function nmaHW4math504()
%function nmaHW4math504()
%Solve problem #2 in second handout, Math 504
%spring 2008 CSUF
%by Nasser Abbasi  feb 8,2008

%Developed on MATLAB Version 7.4.0.287 (R2007a)
%Running on Win XP. Uses statistics toolbox

%
%  M A I N    C O N F I G U R A T I O N     S E C T I O N
%
params.D = 3;           % Diffusion parameter
params.T = 1;           % total running time
params.beta = 5;        % drift parameter

%
%
% These parameters are derived automatically from the above
%
params.trueVar  = 2*params.D*params.T;
params.trueStd  = sqrt(params.trueVar);
params.trueMean = params.beta*params.T;

%
%  internal program   C O N F I G U R A T I O N
%  These are configutation parameters for displying
%  and for setting sample size and number of steps
%  Adjust as needed
%

config.seed  = 12345;
config.nBins = 50;
config.sample_size = 5000;
config.max_number_of_steps = 160;

%
%determine the starting number of steps such that 'p' comes out to be
%less than 1. see report for derivation of this formula
%
config.starting_step = round(params.T*params.beta^2/(2*params.D)) +1;

%add the above number of steps to the starting step to
%obtain max number of steps

config.max_number_of_steps =
config.starting_step+config.max_number_of_steps;

%
% set the number of steps to skip at each simulation else this
% will take too long to run
%
config.n = config.starting_step:5:config.max_number_of_steps;
```

```
%
% Rest are internal data structures to keep track of
% simulation data during runs
%
config.varianceVector = zeros(length(config.n),1);
config.rmsVector      = zeros(length(config.n),1);
config.delT           = zeros(length(config.n),1);
config.standarize     = 0;  %set this to zero if you do not standarize

%
% Determine the theoretical quantiles for the standard normal
% distribution to use in the plots generated
%
config.qp=norminv( ((1:config.sample_size)-.5)/config.sample_size,0,1);

%
% I N I T I A L I Z A T I O N
% Create the figure and seed the random number generator
%

makeFigure();
hold on;
rand('state',config.seed);

%
% Here we go, let start the fun part
%

for i = 1:length(config.n)
    config=simulate_one_walk( params, config, config.n(i), i );
end

end
```

```
%%%%%%%%%%%%%%%%%%%%%%
%
%
%%%%%%%%%%%%%%%%%%%%%%
function config=simulate_one_walk( params, config, number_of_steps, ...
    current_experiement_number )

%
% generate delt(T) and del(X) and p and q from the input
%
config.delT(current_experiement_number) = params.T/number_of_steps;
delX = sqrt(2*params.D*config.delT(current_experiement_number));
p    = 1/2 * (1 + params.beta*delX/(2*params.D));
q    = 1-p;

[normalizedPosition,position] = generate_distribution(p,q,delX,...
    number_of_steps,params,config);

config.varianceVector(current_experiement_number) = var(position);

if config.standarize
    pos=normalizedPosition;
else
    pos=position;
end

%
% Now that we have a distribution generated, lets find the rms error
%
[config,truePDF,estimatedPDF,estimatedFit,xForSimulation,xForNormal]=..
.
    getRMSerrorInCurrentPDF(pos,config,params,p,q,...
    current_experiement_number);

%
%Ok, we have all the data, lets make a plot
%
updatePlots(number_of_steps,normalizedPosition,...
    delX,params,config,truePDF,estimatedPDF,p,q,...
    xForNormal,current_experiement_number);

end
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%  This function generates a sample
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [normalizedPosition,position]=generate_distribution(p,q,...
    delX,number_of_steps,params,config)

%
% generate 2 arrays to hold the final positions in. One
% for standardized position and for not standarized position
% (was not sure which to use at one point, so I keep both)
%
position = zeros(config.sample_size,1);
normalizedPosition = position;

for i = 1:config.sample_size
    y = makeOneRandomWalk(p,q,number_of_steps);
    position(i) = sum(y)*delX;
    normalizedPosition(i) = (position(i)-params.trueMean) /
params.trueStd;
end

end
%%%%%%%%%%%%%%%%
%
%
%%%%%%%%%%%%%%%%
function y=makeOneRandomWalk(p,q,number_of_steps)

y       = rand(number_of_steps,1);
y(y<=q) = -1;
y(y>q)  = 1;

end
```

```
%%%%%%%%%%%%%%%%%%
% Called to obtain the RMS error between the normal distibution
% and the current sample distribution
%
%%%%%%%%%%%%%%%%%%%%
function [config,truePDF,estimatedPDF,estimatedFit,...

xForSimulation,xForNormal]=getRMSerrorInCurrentPDF(position,config,...
    params,p,q,current_experiement_number)

mu=mean(position);
stdd=std(position);

xForSimulation = linspace(mu-4*stdd,...
    mu+4*stdd,config.nBins);
[estimatedPDF,config.xout] = hist(position,xForSimulation);
config.binWidth = config.xout(2)-config.xout(1);
currentArea     = config.binWidth*sum(estimatedPDF);
estimatedPDF    = estimatedPDF/currentArea;

if config.standarize
    stdd = 1;
    mu   = 0;
else
    stdd = params.trueStd;
    mu   = params.trueMean;
end

xForNormal = linspace(mu-4*stdd, mu+4*stdd, config.nBins);
truePDF    = pdf('Normal',xForNormal,mu, stdd);

diffPDF = truePDF-estimatedPDF;

config.rmsError(current_experiement_number) = ...
    norm(diffPDF)/sqrt(length(diffPDF));

estimatedFit = pdf('Normal',xForSimulation,params.trueMean, ...
    stdd*sqrt(4*p*q));

end
%%%%%%%%%%%%%%%%%%
%
%
%%%%%%%%%%%%%%%%%%%%
function makeFigure()

figure;
set(gcf,'Position',[200 100 700 600]);
set(gcf,'Resize','off')
set(0,'DefaultTextinterpreter','none');
axpos = get(gca,'pos');
h = title({'',''});
extent = get(h,'extent');

% position is [left, bottom, width, height];
set(gca,'pos',[axpos(1) axpos(2) axpos(3) axpos(4)-.3*extent(4)]);
end
```

```
%%%%%%%%%%%%%%%
%
%
%%%%%%%%%%%%%%%
function updatePlots(current_number_of_steps,...
    normalizedPosition,delX,params,config,truePDF,...
    estimatedPDF,p,q,xForNormal,current_experiement_number)

subplot(3,2,1:2);
str1='first step[$%d$] current step[$%d$] sampleSize[$%d$] $p=[%4.3f]$
$q=[%4.3f$] $\\beta t=[%4.3f]$ $D = [%d]$ $t = [%d]$';
cla;
bar(config.xout,estimatedPDF,'y');  %relative frequency

line1=sprintf(str1,config.starting_step,current_number_of_steps,...
    config.sample_size,p,q,params.trueMean,params.D,params.T);

title(char(line1),'fontsize',10,'interpreter','latex');
xlabel(sprintf('y'),'fontsize',10,'interpreter','latex');
ylabel('pdf(position)');
set(gca,'FontSize',8);
hold on;
plot(xForNormal,truePDF,'--r','LineWidth',2);

ymax = max(truePDF);
ylim([0,ymax+.5*ymax]);
if config.standarize
    xlim([-4,4]);
else
    xlim([params.trueMean-4*params.trueStd,...
        params.trueMean+4*params.trueStd]);
end
legend('current pdf','limit pdf');
drawnow;

%
% relative error in variance plot
%
subplot(3,2,3);
cla;
line([config.n(1) config.n(end)],[0  0]);
z=repmat(params.trueVar,current_experiement_number,1);
relativeErrorInVar = ...
    ((abs(z-
config.varianceVector(1:current_experiement_number))./z)*100;

line(config.n(1:current_experiement_number),relativeErrorInVar);
xlim([config.n(1) config.n(end)]);
ylim([0 80]);
```

```
line1 = sprintf('relative error in variance ($2DT$)');
line2 = 'true var$=%4.1f$, current var$=%4.1f$';
line2 =
sprintf(line2,params.trueVar,config.varianceVector(current_experiement_
number));
title(char(line1,line2),'fontsize',10,'interpreter','latex');
ylabel('relative error percentage','fontsize',10);
xlabel('number of steps (n)','fontsize',10);
%set(gca,'XTick',1:config.max_number_of_steps);
set(gca,'FontSize',8);
drawnow;

%
% RMS error plot
%
subplot(3,2,4);
cla;

plot(log2(config.delT(1:current_experiement_number)),...
    log2(config.rmsError(1:current_experiement_number)),'r-');
line1=sprintf('\\ \\ \\ \\ \\ \\  RMS error in PDF as
$\\bigtriangleup{T}\\rightarrow{0}$');
line2=sprintf('$\\bigtriangleup{T}=%6.4f \\: \\bigtriangleup{X}=%6.4f
\\:\\: rms=%6.5f$',...
    config.delT(current_experiement_number),delX, ...
    config.rmsError(current_experiement_number));
title(char(line1,line2),'fontsize',10,'interpreter','latex');

ylabel(sprintf('$log_2$(rms
error)'),'fontsize',10,'interpreter','latex');
xlabel(sprintf...

('$log_2(\\bigtriangleup{T})$'),'fontsize',10,'interpreter','latex');

xlim([-10,0]);
ylim([-15,.5]);
set(gca,'FontSize',8);
drawnow;
%
% quantile plot
%
subplot(3,2,5:6);
cla;
sp=sort(normalizedPosition);
plot(config.qp,sp,'r.');

xlabel('quantile of normal distribution');
ylabel('quantile of final position');
title('Quantile-Quantile plot','fontsize',10);
hold on;
line([-4 4],[-4 4]);
legend('data','normal distribution','Location','NorthWest');
set(gca,'FontSize',8);
xlim([-4,4]);
ylim([-4,4]);
drawnow;
end
```

4.4.7 Source code listing

Matlab code is

```
function nmaHW4math504()
%function nmaHW4math504()
%Solve problem #2 in second handout, Math 504
%spring 2008 CSUF
%by Nasser Abbasi  feb 8,2008

%Developed on MATLAB Version 7.4.0.287 (R2007a)
%Running on Win XP. Uses statistics toolbox

%
%  M A I N    C O N F I G U R A T I O N     S E C T I O N
%
params.D = 3;          % Diffusion parameter
params.T = 2;          % total running time
params.beta = 2;       % drift parameter

%
%
% These parameters are derived automatically from the above
%
params.trueVar  = 2*params.D*params.T;
params.trueStd  = sqrt(params.trueVar);
params.trueMean = params.beta*params.T;

%
%  internal program   C O N F I G U R A T I O N
%  These are configutation parameters for displying
%  and for setting sample size and number of steps
%  Adjust as needed
%

config.seed  = 12345;
config.nBins = 10;
config.sample_size = 5000;
config.max_number_of_steps = 20;

%
%determine the starting number of steps such that 'p' comes out to be
%less than 1. see report for derivation of this formula
%
config.starting_step = round(params.T*params.beta^2/(2*params.D)) +1;

%add the above number of steps to the starting step to
%obtain max number of steps

config.max_number_of_steps = config.starting_step+config.max_number_of_steps;

%
% set the number of steps to skip at each simulation else this
% will take too long to run
%
config.n = config.starting_step:5:config.max_number_of_steps;

%
% Rest are internal data structures to keep track of
```

```
% simulation data during runs
%
config.varianceVector = zeros(length(config.n),1);
config.rmsVector      = zeros(length(config.n),1);
config.delT           = zeros(length(config.n),1);
config.standarize     = 1;  %set this to zero if you do not standarize

%
% Determine the theoretical quantiles for the standard normal
% distribution to use in the plots generated
%
config.qp=norminv( ((1:config.sample_size)-.5)/config.sample_size,0,1);

%
% I N I T I A L I Z A T I O N
% Create the figure and seed the random number generator
%

makeFigure();
hold on;
rand('state',config.seed);

%
% Here we go, let start the fun part
%

for i = 1:length(config.n)
    config=simulate_one_walk( params, config, config.n(i), i );
end

end

%%%%%%%%%%%%%%%%%%%%%%
%
%
%%%%%%%%%%%%%%%%%%%%%%
function config=simulate_one_walk( params, config, number_of_steps, ...
    current_experiement_number )

%
% generate delt(T) and del(X) and p and q from the input
%
config.delT(current_experiement_number) = params.T/number_of_steps;
delX = sqrt(2*params.D*config.delT(current_experiement_number));
p    = 1/2 * (1 + params.beta*delX/(2*params.D));
q    = 1-p;

[normalizedPosition,position] = generate_distribution(p,q,delX,...
    number_of_steps,params,config);

config.varianceVector(current_experiement_number) = var(position);

if config.standarize
    pos=normalizedPosition;
else
    pos=position;
end
```

```
%
% Now that we have a distribution generated, lets find the rms error
%
[config,truePDF,estimatedPDF,estimatedFit,xForSimulation,xForNormal]=...
    getRMSerrorInCurrentPDF(pos,config,params,p,q,...
    current_experiement_number);

%
%Ok, we have all the data, lets make a plot
%
updatePlots(number_of_steps,normalizedPosition,...
    delX,params,config,truePDF,estimatedPDF,p,q,...
    xForNormal,current_experiement_number);

end
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
%  This function generates a sample
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [normalizedPosition,position]=generate_distribution(p,q,...
    delX,number_of_steps,params,config)

%
% generate 2 arrays to hold the final positions in. One
% for standardized position and for not standarized position
% (was not sure which to use at one point, so I keep both)
%
position = zeros(config.sample_size,1);
normalizedPosition = position;

for i = 1:config.sample_size
    y = makeOneRandomWalk(p,q,number_of_steps);
    position(i) = sum(y)*delX;
    normalizedPosition(i) = (position(i)-params.trueMean) / params.trueStd;
end

end
%%%%%%%%%%%%%%%%
%
%
%%%%%%%%%%%%%%%%
function y=makeOneRandomWalk(p,q,number_of_steps)

y       = rand(number_of_steps,1);
y(y<=q) = -1;
y(y>q)  = 1;

end
%%%%%%%%%%%%%%%%
%
%
%%%%%%%%%%%%%%%%
function updatePlots(current_number_of_steps,...
    normalizedPosition,delX,params,config,truePDF,...
    estimatedPDF,p,q,xForNormal,current_experiement_number)

subplot(3,2,1:2);
```

```
str1='first step[$%d$] current step[$%d$] sampleSize[$%d$] $p=[%4.3f]$ $q=[%4.3f$] $\\beta t=[%
cla;
bar(config.xout,estimatedPDF,'y');  %relative frequency

line1=sprintf(str1,config.starting_step,current_number_of_steps,...
    config.sample_size,p,q,params.trueMean,params.D,params.T);

title(char(line1),'fontsize',10,'interpreter','latex');
xlabel(sprintf('y'),'fontsize',10,'interpreter','latex');
ylabel('pdf(position)');
set(gca,'FontSize',8);
hold on;
plot(xForNormal,truePDF,'--r','LineWidth',2);

ymax = max(truePDF);
ylim([0,ymax+.5*ymax]);
if config.standarize
    xlim([-4,4]);
else
    xlim([params.trueMean-4*params.trueStd,...
        params.trueMean+4*params.trueStd]);
end
legend('current pdf','limit pdf');
drawnow;

%
% relative error in variance plot
%
subplot(3,2,3);
cla;
line([config.n(1) config.n(end)],[0  0]);
z=repmat(params.trueVar,current_experiement_number,1);
relativeErrorInVar = ...
    ((abs(z-config.varianceVector(1:current_experiement_number)))./z)*100;

line(config.n(1:current_experiement_number),relativeErrorInVar);
xlim([config.n(1) config.n(end)]);
ylim([0 80]);

line1 = sprintf('relative error in variance ($2DT$)');
line2 = 'true var$=%4.1f$, current var$=%4.1f$';
line2 = sprintf(line2,params.trueVar,config.varianceVector(current_experiement_number));
title(char(line1,line2),'fontsize',10,'interpreter','latex');
ylabel('relative error percentage','fontsize',10);
xlabel('number of steps (n)','fontsize',10);
%set(gca,'XTick',1:config.max_number_of_steps);
set(gca,'FontSize',8);
drawnow;

%
% RMS error plot
%
subplot(3,2,4);
cla;

plot(log2(config.delT(1:current_experiement_number)),...
    log2(config.rmsError(1:current_experiement_number)),'r-');
line1=sprintf('\\ \\ \\ \\ \\ \\  RMS error in PDF as $\\bigtriangleup{T}\\rightarrow{0}$');
```

```
line2=sprintf('$\\bigtriangleup{T}=%6.4f \\: \\bigtriangleup{X}=%6.4f \\:\\: rms=%6.5f$',...
    config.delT(current_experiement_number),delX, ...
    config.rmsError(current_experiement_number));
title(char(line1,line2),'fontsize',10,'interpreter','latex');

ylabel(sprintf('$log_2$(rms error)'),'fontsize',10,'interpreter','latex');
xlabel(sprintf...
    ('$log_2(\\bigtriangleup{T})$'),'fontsize',10,'interpreter','latex');

xlim([-6,1]);
ylim([-10,1]);
set(gca,'FontSize',8);
drawnow;

%
% quantile plot
%
subplot(3,2,5:6);
cla;
sp=sort(normalizedPosition);
plot(config.qp,sp,'r.');

xlabel('quantile of normal distribution');
ylabel('quantile of final position');
title('Quantile-Quantile plot','fontsize',10);
hold on;
line([-4 4],[-4 4]);
legend('data','normal distribution','Location','NorthWest');
set(gca,'FontSize',8);
xlim([-4,4]);
ylim([-4,4]);
drawnow;

end
%%%%%%%%%%%%%%%%%%%
% Called to obtain the RMS error between the normal distibution
% and the current sample distribution
%
%%%%%%%%%%%%%%%%%%%%%
function [config,truePDF,estimatedPDF,estimatedFit,...
    xForSimulation,xForNormal]=getRMSerrorInCurrentPDF(position,config,...
    params,p,q,current_experiement_number)

mu=mean(position);
stdd=std(position);

xForSimulation = linspace(mu-4*stdd,...
    mu+4*stdd,config.nBins);
[estimatedPDF,config.xout] = hist(position,xForSimulation);
config.binWidth = abs(  abs(config.xout(2))-abs(config.xout(1)));
currentArea     = config.binWidth*sum(estimatedPDF);
estimatedPDF    = estimatedPDF/currentArea;

if config.standarize
    stdd = 1;
    mu   = 0;
else
    stdd = params.trueStd;
```

```
    mu   = params.trueMean;
end

xForNormal = linspace(mu-4*stdd, mu+4*stdd, config.nBins);
truePDF    = pdf('Normal',xForNormal,mu, stdd);

diffPDF = truePDF-estimatedPDF;

config.rmsError(current_experiement_number) = ...
    norm(diffPDF)/sqrt(length(diffPDF));

estimatedFit = pdf('Normal',xForSimulation,params.trueMean, ...
    stdd*sqrt(4*p*q));

end
%%%%%%%%%%%%%%%%%
%
%
%%%%%%%%%%%%%%%%%%%%%
function makeFigure()

figure;
set(gcf,'Position',[200 100 700 600]);
set(gcf,'Resize','off')
set(0,'DefaultTextinterpreter','none');
axpos = get(gca,'pos');
h = title({'',''});
extent = get(h,'extent');

% position is [left, bottom, width, height];
set(gca,'pos',[axpos(1) axpos(2) axpos(3) axpos(4)-.3*extent(4)]);

end
```

To run, save it to your Matlab working directory and type the command `nmaHW4math504()` from the console.

4.5 Wed 2/27/2008

Grade: 2/2.

Problem 3.9 from handouts (probability distribution related to record time distribution)

4.5.1 Problem

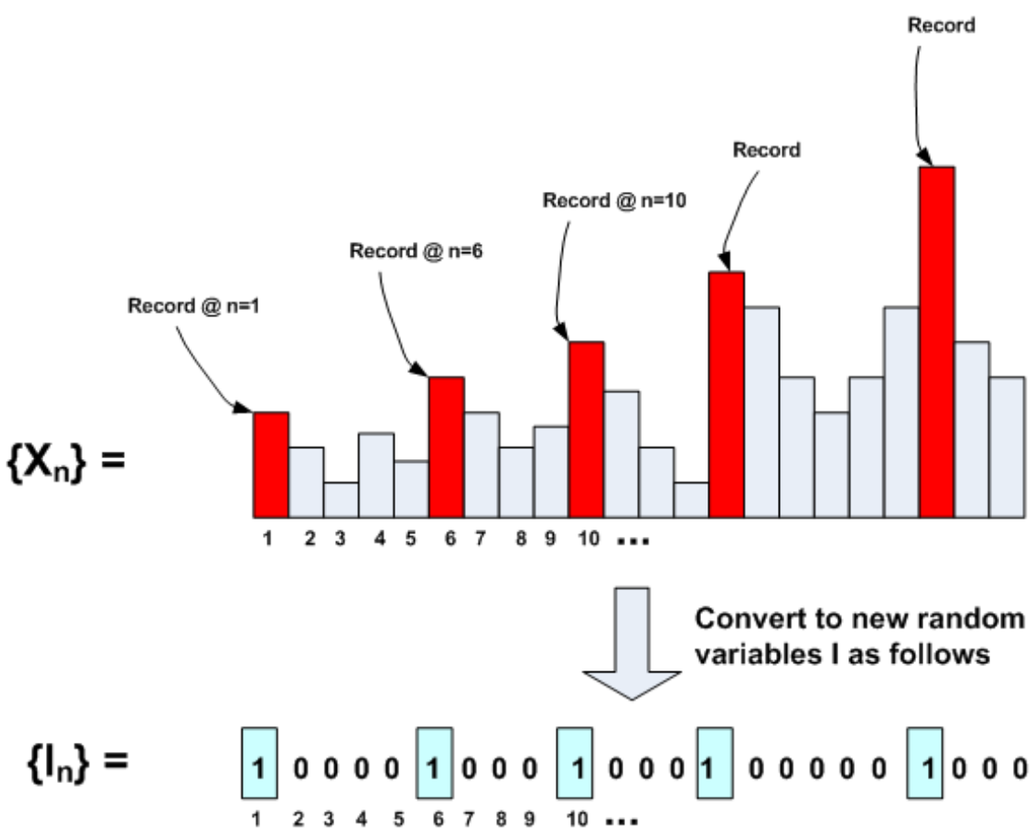
3.9 Let $X_1, X_2, \cdots$ be a sequence of independent, identically distributed, continuous random variables. A record is said to occur at time $n \geq 2$ if $X_n > max\{X_1, X_2, \cdots, X_{n-1}\}$. The value of the first variable X_1 is considered a record at time $n = 1$. Define the random variable I_i to be 1 if a record occurs at time i, and to be zero otherwise. (a) Show that $P(I_i = 1) = 1/i$, for $i \geq 1$. Help: For $i \geq 2$, $P(I_i = 1)$ is equal to $P(X_1 < X_i, X_2 < X_i, \cdots, X_{i-1} < X_i)$. Condition on X_i to solve this equation. (b) Let N_n equal the number of records that occur up to time n. Find the expected value and variance of N_n. Help: Note that $N_n = I_1 + \cdots + I_n$, and assume without proof that the variables I_i are independent. (c) Let T be the first time greater than 1 at which a record occurs. Find the probability distribution of T, and show that $P(T < \infty) = 1$, while $E(T) = \infty$. Help: Note that $P(T > n) = P(X_2 < X_1, X_3 < X_1, \cdots X_n < X_1)$, and condition on X_1.

Figure 4.2: problem 3.9 from lecture notes

4.5.2 Solution

Part (A)

We first covert the sequence of random variables X_n to sequence of random variables I_i as described. A diagram below will also help illustrate this conversion



We need to show that $P(I_i = 1) = \frac{1}{i}$. Using the hint given, we write (for $i \geq 2$)

$$P(I_i = 1) = P(X_1 < X_i, X_2 < X_i, \cdots, X_{i-1} < X_i)$$

Conditioning on X_i, and assuming the pdf of X is given by $f(x)$ we write

$$P(I_i = 1) = \int_{-\infty}^{\infty} P(X_1 < X_i, X_2 < X_i, \cdots, X_{i-1} < X_i | X_i = x)\, f(x)\, dx$$

Since the X_j random variables are independent from each others, we break the above 'and' probabilities to products of probabilities.

$$P(I_i = 1) = \int_{-\infty}^{\infty} P(X_1 < X_i | X_i = x) \quad P(X_2 < X_i | X_i = x) \cdots P(X_{i-1} < X_i | X_i = x) \quad f(x)\, dx$$

But $P(X_j < X_i | X_i = x) = F(x)$, hence the above becomes

$$P(I_i = 1) = \int_{-\infty}^{\infty} [F(x)]^{i-1} \quad f(x)\, dx \tag{1}$$

But $f(x) = F'(x)$, hence the above becomes

$$P(I_i = 1) = \int_{-\infty}^{\infty} [F(x)]^{i-1} \quad F'(x)\, dx$$

Now do integration by parts (let $dv = F'(x)$ and $u = [F(x)]^{i-1}$

$$\begin{aligned} P(I_i = 1) &= [uv]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} v du \\ &= \left[F(x) [F(x)]^{i-1} \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} F(x) (i-1) [F(x)]^{i-2} F'(x)\, dx \\ &= \left[F(x)^i \right]_{-\infty}^{\infty} - (i-1) \int_{-\infty}^{\infty} [F(x)]^{i-1} F'(x)\, dx \end{aligned}$$

But $\left[F(x)^i \right]_{-\infty}^{\infty} = 1 - (0) = 1$ hence the above becomes

$$P(I_i = 1) = 1 - (i-1) \overbrace{\int_{-\infty}^{\infty} [F(x)]^{i-1} F'(x)\, dx}^{P(I_i=1)}$$

But $\int_{-\infty}^{\infty} [F(x)]^{i-1} F'(x)\, dx = P(I_i = 1)$ since it is the integral we started with (see (1)), so move it to the left side, and the above becomes

$$\begin{aligned} P(I_i = 1) &= 1 - (i-1) P(I_i = 1) \\ P(I_i = 1) &+ (i-1) P(I_i = 1) = 1 \\ P(I_i = 1) i &= 1 \end{aligned}$$

Hence

$$P(I_i = 1) = \frac{1}{i}$$

Part(B)

N_n is number of records up to time n. We need to find $E(N_n)$ and $Var(N_n)$

$$\begin{aligned} N_n &= I_1 + I_2 + \cdots + I_n \\ E(N_n) &= E(I_1 + I_2 + \cdots + I_n) \\ &= E(I_1) + E(I_2) + \cdots + E(I_n) \end{aligned}$$

But $E(I_1) = 1 \times P(I_1 = 1) + 0 \times P(I_1 = 0) = P(I_1 = 1)$ and similarly, $E(I_i) = P(I_i = 1)$

Hence

$$\begin{aligned} E(N_n) &= P(I_1 = 1) + P(I_2 = 1) + \cdots + P(I_n = 1) \\ &= 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \\ &= \sum_{i=1}^{n} \frac{1}{i} \end{aligned}$$

So $E(N_n)$ is a harmonic number. In the limit, this sum is $\boxed{\text{infinity}}$. Hence number of records is infinite. i.e. if we wait long enough, we will always obtain a new record.

To find the variance of N_n, we use the hint and assume I_i are independent of each others (i.e. when a record occurs is independent of when previous record occurred), hence the covariance terms drop out (since all zero) and we are left with the sum of variances

$$\begin{aligned} Var(N_n) &= Var(I_1 + I_2 + \cdots + I_n) \\ &= Var(I_1) + Var(I_2) + \cdots + Var(I_n) \end{aligned}$$

But

$$Var(I_i) = E\left(I_i^2\right) - E(I_i)^2$$

But $E(I_i^2) = 1^2 \times P(I_i = 1) + 0^2 \times P(I_1 = 0) = P(I_i = 1) = \frac{1}{i}$

Hence $\boxed{E(I_i^2) = \frac{1}{i}}$, therefore

$$\begin{aligned} Var(I_i) &= \frac{1}{i} - P(I_i = 1)^2 \\ &= \frac{1}{i} - \left(\frac{1}{i}\right)^2 \end{aligned}$$

therefore

$$Var(N_n) = \sum_{i=1}^{n} \frac{1}{i} - \left(\frac{1}{i}\right)^2$$

Since $\sum_{i=1}^{n} \frac{1}{i} = \infty$ as n gets very large, and $\sum_{i=1}^{n} \frac{1}{i^2} = \frac{\pi^2}{6}$ as n gets very large, then

$Var(N_n) = \infty$ as n gets very large.

Part(C)

We need to find $\Pr(T=n)$ where T is the time of the first record (not counting $n=1$ which is always a record ofcourse).

$$\Pr(T=2)=\Pr(I_2=1)=\frac{1}{2}$$

Now

$$\Pr(T=3)=\Pr(\text{no record at T=2,record at T=3})$$

Since having no record at $T=2$ and having a record at $T=3$ are indepdent events the above becomes

$$\begin{aligned}\Pr(T=3)&=\Pr(\text{no record at T=2})\ \ \Pr(\text{record at T=3})\\&=(1-\Pr(I_2=1))\times\Pr(I_3=1)\\&=\left(1-\frac{1}{2}\right)\left(\frac{1}{3}\right)\\&=\frac{1}{2}\times\frac{1}{3}\end{aligned}$$

Similarly,

$$\begin{aligned}\Pr(T=4)&=\Pr(\text{no record at T=2,no record at T=3,record at T=4})\\&=(1-\Pr(I_2=1))\times(1-\Pr(I_3=1))\times\Pr(I_4=1)\\&=\left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right)\left(\frac{1}{4}\right)\\&=\frac{1}{2}\times\frac{2}{3}\times\frac{1}{4}\\&=\frac{1}{3}\times\frac{1}{4}\end{aligned}$$

Similarly,

$$\begin{aligned}\Pr(T=5)&=\Pr(\text{no record at T=2,no record at T=3,no record at T=4,record at T=5})\\&=(1-\Pr(I_2=1))\times(1-\Pr(I_3=1))\times(1-\Pr(I_4=1))\times\Pr(I_5=1)\\&=\left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right)\left(1-\frac{1}{4}\right)\left(\frac{1}{5}\right)\\&=\frac{1}{2}\times\frac{2}{3}\times\frac{3}{4}\times\frac{1}{4}\\&=\frac{1}{4}\times\frac{1}{5}\end{aligned}$$

Hence continuing this way, we see that

$$\Pr(T=n)=\frac{1}{n(n-1)}$$

Hence

$$\begin{aligned}\Pr(T<\infty) &= \lim_{k\to\infty}\sum_{n=2}^{k}\frac{1}{n(n-1)}\\ &= \lim_{k\to\infty}\left(\frac{k-1}{k}\right)\\ &= 1\end{aligned}$$

and

$$\begin{aligned}E(T) &= 2\times\Pr(T=2)+3\times\Pr(T=3)+4\times\Pr(T=4)+\cdots\\ &= 2\left(\frac{1}{2}\right)+3\left(\frac{1}{2}\times\frac{1}{3}\right)+4\left(\frac{1}{3}\times\frac{1}{4}\right)+\cdots\\ &= 1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\cdots\\ &= \sum_{i=1}^{\infty}\frac{1}{n}\\ &= \infty\end{aligned}$$

Hence

$$E(T)=\infty$$

4.6 Computing Assignment #3. Monday 3/3/2008

Craps game and inventory problem. Markov chain Problem description is

Computer.

Math 504 Assignment 2 3

a week from next monday.

March 3/3/2018

1. Consider the game of "craps": Two die are tossed. The player wins if a sum of 7 or 11 appears, and loses if a sum of 2,3, or 12 appears. Otherwise, the sum on the first toss is designated the player's point score. The player then continues to toss the two die until a sum of 7 appears, in which case the player loses, or a sum equal to the point score appears, in which case the player wins. Model this game as an absorbing Markov chain. Determine the one-step transition matrix P.

2. Write a program to form the one-step transition matrix P for the inventory problem in Example 4.3.2. For the demand, use a function of the form
$$d(x) = c(x+1)\ , \text{ for } x = 0, 1, 2, \cdots, 5\ ,$$
and
$$d(x) = 6c(13-x)/8\ , \text{ for } x = 6, 7, \cdots, 13\ .$$
Choose the constant c so that the function d is a probability distribution. Inputs to your program are the values of s and S. Output is the one-step transition matrix P.

Experiment with each of these processes, and investigate what can be said about the long-run behavior. In particular, consider the following questions. (a) Do the powers P^n converge as $n \to \infty$? If so, what can you say about the limit matrix? (b) Is there a limiting state probability distribution in all cases? If so, does this limiting distribution depend on the initial state probability distribution? (c) Assuming your conclusions from parts (a) and (b) are correct, show how the results of part (a) could be used to deduce the conclusions of part (b).

byrs

In your report, first summarize your numerical results briefly and succinctly. Present these results in such a way that the reader can easily understand the observations that you are drawing from your experiments. Next state your observations based on the numerical results, and indicate any general conclusions that seem to be suggested by the experiments.

4.6.1 Problem description

Math 504 Assignment 2 3

1. Consider the game of "craps": Two die are tossed. The player wins if a sum of 7 or 11 appears, and loses if a sum of 2,3, or 12 appears. Otherwise, the sum on the first toss is designated the player's point score. The player then continues to toss the two die until a sum of 7 appears, in which case the player loses, or a sum equal to the point score appears, in which case the player wins. Model this game as an absorbing Markov chain. Determine the one-step transition matrix P.

2. Write a program to form the one-step transition matrix P for the inventory problem in Example 4.3.2. For the demand, use a function of the form

$$d(x) = c(x+1) \text{ , for } x = 0, 1, 2, \cdots, 5 \text{ ,}$$

and

$$d(x) = 6c(13-x)/8 \text{ , for } x = 6, 7, \cdots, 13 \text{ .}$$

Choose the constant c so that the function d is a probability distribution. Inputs to your program are the values of s and S. Output is the one-step transition matrix P.

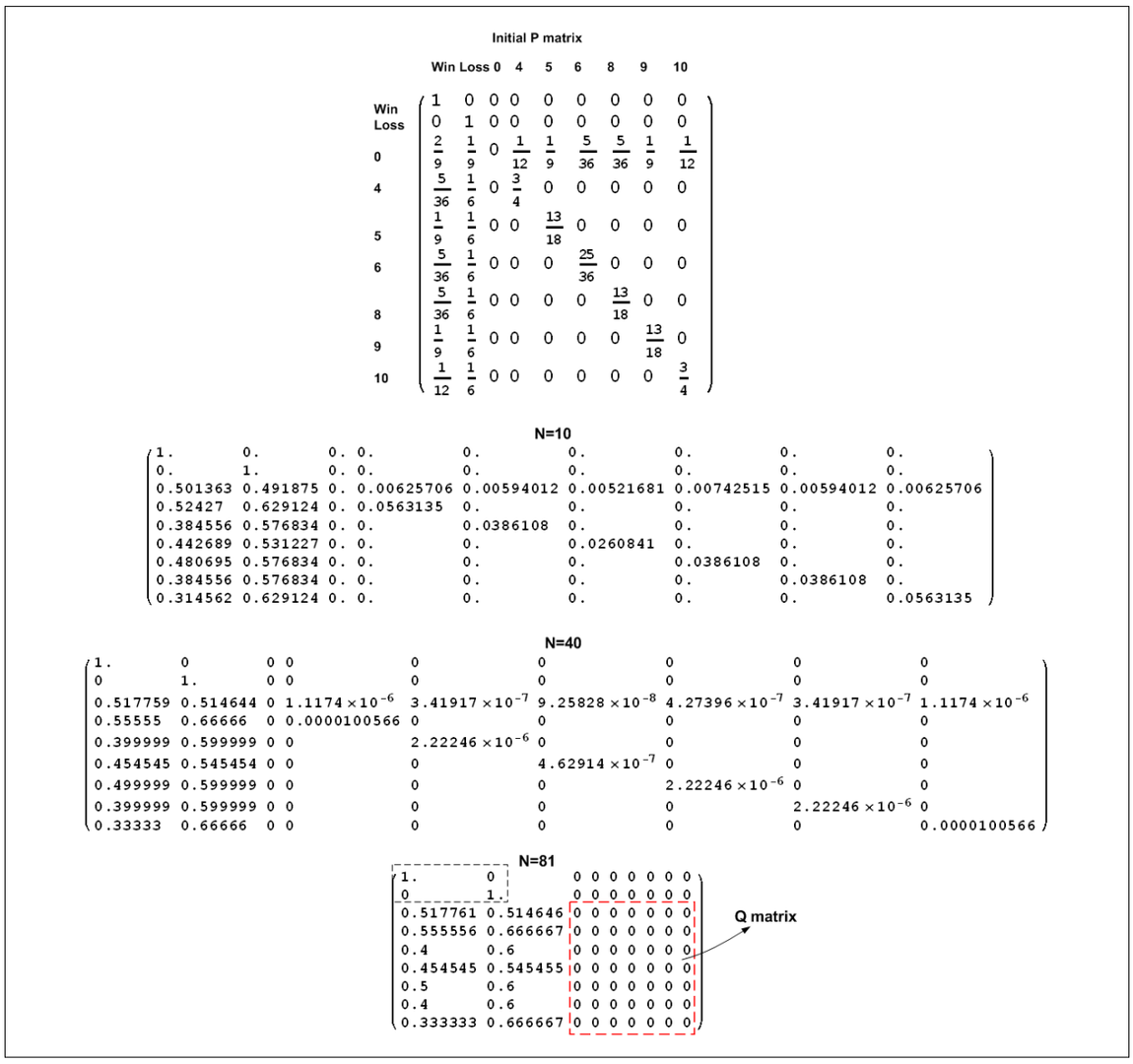
Experiment with each of these processes, and investigate what can be said about the long-run behavior. In particular, consider the following questions. (a) Do the powers P^n converge as $n \to \infty$? If so, what can you say about the limit matrix? (b) Is there a limiting state probability distribution in all cases? If so, does this limiting distribution depend on the initial state probability distribution? (c) Assuming your conclusions from parts (a) and (b) are correct, show how the results of part (a) could be used to deduce the conclusions of part (b).

In your report, first summarize your numerical results briefly and succinctly. Present these results in such a way that the reader can easily understand the observations that you are drawing from your experiments. Next state your observations based on the numerical results, and indicate any general conclusions that seem to be suggested by the experiments.

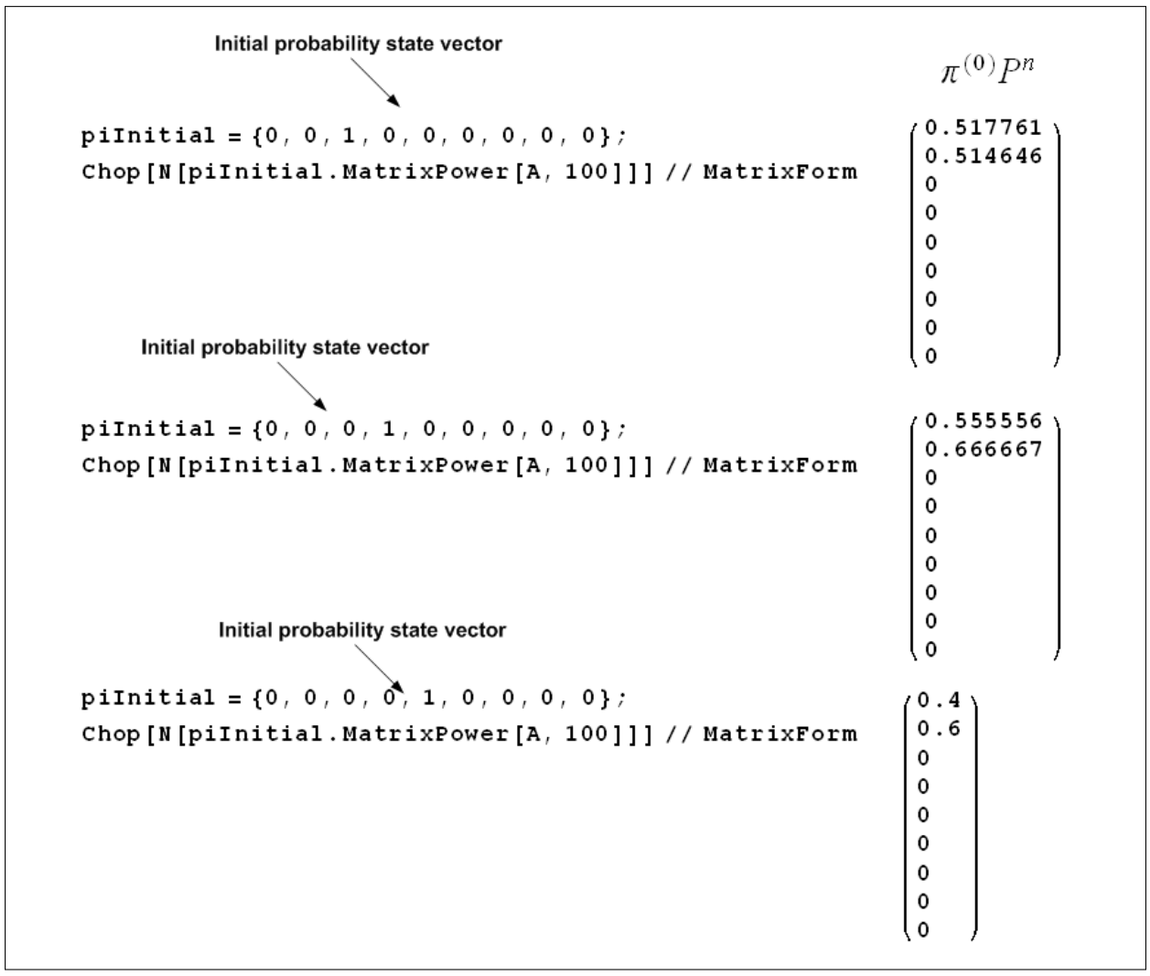
4.6.2 craps game

Summary of numerical results

The state probability transition matrix was entered and then raised to higher powers. This is the numerical result



To answer part (b) below, we need to run the system from different initial state vector (i.e. different $\pi^{(0)}$) and observe if the system probability state vector after a long time (i.e. $\pi^{(\infty)}$) will depend on the initial state vector or not. Here is the result for 3 different initial state vectors. In diagram below we show the $\pi^{(0)}$ and to its right $\pi^{(\infty)}$.



Analysis of numerical results

part (a)

Yes. The powers of P^n converges as $n \to \infty$. This is seen by looking at the above sequence of the P matrix where we see that the matrix P converges to the following limiting matrix at around $n = 81$

We can say the following about the limiting matrix: As $n \to \infty$ the matrix P converges to a fixed value shown above. The entries P_{ij}^n where j is a transient state goes to zero as n gets large.

part (b)

From the above numerical result, we see that depending on the initial system probability state vector $\pi^{(0)}$ we obtain a different system probability state vector $\pi^{(n)}$ as n gets very large. This is because some states are transient (states $\{4,5,6,8,9,10\}$). In the inventory problem below, we see that we obtained a different result for this part since the inventory problem has no transient states.

part (c) Let I be the set of all the possible states the system can be in. Hence from definition, we write

$$\pi_j^{(n)} = \sum_{i \in I} \pi_i^{(0)} P_{ij}^n$$

Where $\pi_j^{(n)}$ means the probability that the system will be in state j after n steps and P_{ij}^n is the n steps transition probability. Now take the limit of the above as $n \to \infty$ we have

$$\begin{aligned} \lim_{n\to\infty} \pi_j^{(n)} &= \lim_{n\to\infty} \sum_{i \in I} \pi_i^{(0)} P_{ij}^n \\ &= \sum_{i \in I} \lim_{n\to\infty} \left(\pi_i^{(0)} P_{ij}^n \right) \end{aligned}$$

Assume there are k states, we can expand

$$\lim_{n\to\infty}\left(\pi_1^{(0)}P_{1j}^n+\pi_2^{(0)}P_{2j}^n+\cdots+\pi_k^{(0)}P_{kj}^n\right)$$

But from part(a) we observed that in the limit, entries of each columns are not equal. Hence $P_{1j}^n\neq P_{2j}^n\neq\cdots\neq P_{kj}^n$ this means the above sum will produce a different value depending on the initial state probability vector $\pi^{(0)}$. (Compare this to the inventory problem below, where each entry in a column is the same, and we could factor it out of the sum and we reached a different conclusion than here).

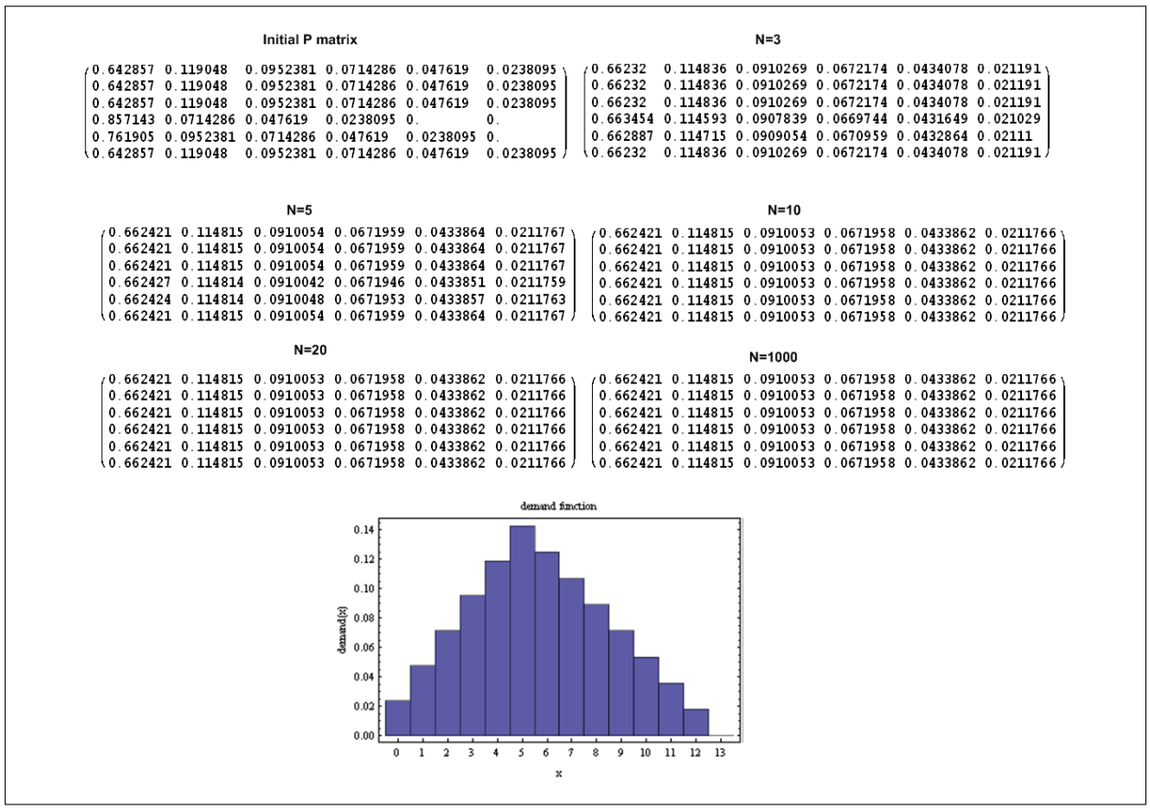
Hence we showed depending on the initial $\pi^{(0)}$ then $\lim_{n\to\infty}\pi_j^{(n)}$ goes to different value as confirmed by the numerical result shown above in part(a). Hence part(a) results could be used to deduce part(b) conclusion.

4.6.3 Inventory problem

Summary of numerical results

An inventory program was written in Mathematica (please see appendix for full source code) which generated the P matrix for an increasing values of n. The specification of the inventory model is described in the question shown above. The value $\boxed{s=3}$ and $\boxed{S=5}$ was used.

The following are few results of the P matrix for an increasing values of n and the histogram of the demand distribution used.



To answer part (b) below, we need to run the system from different initial state vector (i.e. different $\pi^{(0)}$) and observe if the system probability state after a long time (i.e. $\pi^{(\infty)}$) will depend on the initial state vector or not. Since we know that

$$\begin{aligned}\pi^{(1)}&=\pi^{(0)}P\\ \pi^{(2)}&=\pi^{(1)}P=\pi^{(0)}P^{(2)}\\ \pi^{(3)}&=\pi^{(2)}P=\pi^{(0)}P^{(3)}\\ &\cdots\\ \pi^{(n)}&=\pi^{(0)}P^{(n)}\end{aligned}$$

And since $P^{(n)} = P^n$, then all what we have do is pick few $\pi^{(0)}$ vectors, and post multiply them by $P^{(n)}$ for large n and see if we obtain the same $\pi^{(n)}$. Below is the numerical result for this part showing the initial $\pi^{(0)}$ and the final $\pi^{(n)}$. I used $n = 30$ in all cases as this showed it is large enough from the above numerical results. Here are the results. Below we show result of 6 tests. In each one, $\pi^{(0)}$ is shown and to its right $\pi^{(n)}$.

$$\begin{pmatrix} \pi^{(0)} \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} \pi^{(30)} \\ 0.662421 \\ 0.114815 \\ 0.0910053 \\ 0.0671958 \\ 0.0433862 \\ 0.0211766 \end{pmatrix} \quad \begin{pmatrix} \pi^{(0)} \\ 0.5 \\ 0.5 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} \pi^{(30)} \\ 0.662421 \\ 0.114815 \\ 0.0910053 \\ 0.0671958 \\ 0.0433862 \\ 0.0211766 \end{pmatrix} \quad \begin{pmatrix} \pi^{(0)} \\ 0.4 \\ 0.2 \\ 0.4 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} \pi^{(30)} \\ 0.662421 \\ 0.114815 \\ 0.0910053 \\ 0.0671958 \\ 0.0433862 \\ 0.0211766 \end{pmatrix}$$

$$\begin{pmatrix} \pi^{(0)} \\ 0.1 \\ 0.1 \\ 0.4 \\ 0.4 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} \pi^{(30)} \\ 0.662421 \\ 0.114815 \\ 0.0910053 \\ 0.0671958 \\ 0.0433862 \\ 0.0211766 \end{pmatrix} \quad \begin{pmatrix} \pi^{(0)} \\ 0.2 \\ 0.3 \\ 0.1 \\ 0.1 \\ 0.2 \\ 0.1 \end{pmatrix} \begin{pmatrix} \pi^{(30)} \\ 0.662421 \\ 0.114815 \\ 0.0910053 \\ 0.0671958 \\ 0.0433862 \\ 0.0211766 \end{pmatrix} \quad \begin{pmatrix} \pi^{(0)} \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} \pi^{(30)} \\ 0.662421 \\ 0.114815 \\ 0.0910053 \\ 0.0671958 \\ 0.0433862 \\ 0.0211766 \end{pmatrix}$$

Analysis of numerical results

Part (a)

Yes. The powers of P^n converges as $n \to \infty$. This is seen by looking at the above sequence of the P matrix where we see that the matrix P converges to the following limiting matrix at around $n = 20$

$$\begin{pmatrix} 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \end{pmatrix}$$

We can say the following about the limiting matrix: As $n \to \infty$ the matrix P converges to a fixed value shown above. Each column has the same entries in its rows. In addition, all entries are non-zero. This implies that the chain contains no transient states. And since all the values on the converged P matrix are positive, then we have only one closed set in the chain, which contains all the states.

part (b)

Yes. There is a limiting state probability distribution in all cases. This is show by looking at the numerical result above that shows for different initial probability state vector $\pi^{(0)}$ we obtain the same probability state vector $\pi^{(n)}$ when n is large. So the final $\pi^{(\infty)}$ *does not depend* on which state the system starts from.

part (c)

In this part, we need to show given that P^∞ converges to limiting fixed value, then the $\pi_k^{(\infty)}$ is the same for all states k.

Let I be the set of all the possible states the system can be in. Hence from definition, we

write

$$\pi_j^{(n)} = \sum_{i \in I} \pi_i^{(0)} P_{ij}^n$$

Where $\pi_j^{(n)}$ means the probability that the system will be in state j after n steps and P_{ij}^n is the n steps transition probability. Now take the limit of the above as $n \to \infty$ we have

$$\begin{aligned} \lim_{n\to\infty} \pi_j^{(n)} &= \lim_{n\to\infty} \sum_{i \in I} \pi_i^{(0)} P_{ij}^n \\ &= \sum_{i \in I} \lim_{n\to\infty} \left(\pi_i^{(0)} P_{ij}^n \right) \end{aligned}$$

Assume there are k states, we can expand

$$\lim_{n\to\infty} \left(\pi_1^{(0)} P_{1j}^n + \pi_2^{(0)} P_{2j}^n + \cdots + \pi_k^{(0)} P_{kj}^n \right)$$

But from part(a) we observed that $\lim_{n\to\infty} P_{ij}^n$ is a fixed value, which is the limit the transition matrix converged to. In other words, $P_{1j}^n = P_{2j}^n = \cdots = P_{kj}^n$ since all entries in the j column are the same. Call this entry in j^{th} column as k say. So k is a single number which represents the one step transition probability from state i to state j when the system has run for a long time. So we write the above as

$$\lim_{n\to\infty} \pi_j^{(n)} = k \left(\pi_1^{(0)} + \pi_2^{(0)} + \cdots + \pi_k^{(0)} \right)$$

now, $\sum_{i \in I} \pi_i^{(0)}$ is the sum of the probabilities of the system being in all its states at time zero, which must be 1 hence

$$\lim_{n\to\infty} \pi_j^{(n)} = k$$

Hence we showed that regardless of the initial $\pi^{(0)}$ then $\lim_{n\to\infty} \pi_j^{(n)}$ goes to some fixed values. This shows that for any state j the probability that the system will be in that state after a long time converges to a fixed value regardless of the initial state if the system transition matrix converges in the limit. Hence part(a) results could be used to deduce part(b) conclusion.

4.6.4 Graded assignment

COMPUTING ASSIGNMENT 3 MATH 504. SPRING 2008. CSUF

CRAPS AND INVENTORY PROBLEM

by Nasser Abbasi

March 2, 2008

2/2

excellent!

Contents

1 Problem description

Math 504 Assignment 2 3

1. Consider the game of "craps": Two die are tossed. The player wins if a sum of 7 or 11 appears, and loses if a sum of 2,3, or 12 appears. Otherwise, the sum on the first toss is designated the player's point score. The player then continues to toss the two die until a sum of 7 appears, in which case the player loses, or a sum equal to the point score appears, in which case the player wins. Model this game as an absorbing Markov chain. Determine the one-step transition matrix P.

2. Write a program to form the one-step transition matrix P for the inventory problem in Example 4.3.2. For the demand, use a function of the form

$$d(x) = c(x+1)\ , \text{ for } x = 0, 1, 2, \cdots, 5\ ,$$

and

$$d(x) = 6c(13 - x)/8\ , \text{ for } x = 6, 7, \cdots, 13\ .$$

Choose the constant c so that the function d is a probability distribution. Inputs to your program are the values of s and S. Output is the one-step transition matrix P.

Experiment with each of these processes, and investigate what can be said about the long-run behavior. In particular, consider the following questions. (a) Do the powers P^n converge as $n \to \infty$? If so, what can you say about the limit matrix? (b) Is there a limiting state probability distribution in all cases? If so, does this limiting distribution depend on the initial state probability distribution? (c) Assuming your conclusions from parts (a) and (b) are correct, show how the results of part (a) could be used to deduce the conclusions of part (b).

In your report, first summarize your numerical results briefly and succinctly. Present these results in such a way that the reader can easily understand the observations that you are drawing from your experiments. Next state your observations based on the numerical results, and indicate any general conclusions that seem to be suggested by the experiments.

2 craps game

2.1 Summary of numerical results

The state probability transition matrix was entered and then raised to higher powers. This is the numerical result

Initial P matrix

	Win	Loss	0	4	5	6	8	9	10
Win	1	0	0	0	0	0	0	0	0
Loss	0	1	0	0	0	0	0	0	0
0	$\frac{2}{9}$	$\frac{1}{9}$	0	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$
4	$\frac{5}{36}$	$\frac{1}{6}$	0	$\frac{3}{4}$	0	0	0	0	0
5	$\frac{1}{9}$	$\frac{1}{6}$	0	0	$\frac{13}{18}$	0	0	0	0
6	$\frac{5}{36}$	$\frac{1}{6}$	0	0	0	$\frac{25}{36}$	0	0	0
8	$\frac{5}{36}$	$\frac{1}{6}$	0	0	0	0	$\frac{13}{18}$	0	0
9	$\frac{1}{9}$	$\frac{1}{6}$	0	0	0	0	0	$\frac{13}{18}$	0
10	$\frac{1}{12}$	$\frac{1}{6}$	0	0	0	0	0	0	$\frac{3}{4}$

N=10

1.	0.	0.	0.	0.	0.	0.	0.	0.
0.	1.	0.	0.	0.	0.	0.	0.	0.
0.501363	0.491875	0.	0.00625706	0.00594012	0.00521681	0.00742515	0.00594012	0.00625706
0.52427	0.629124	0.	0.0563135	0.	0.	0.	0.	0.
0.384556	0.576834	0.	0.	0.0386108	0.	0.	0.	0.
0.442689	0.531227	0.	0.	0.	0.0260841	0.	0.	0.
0.480695	0.576834	0.	0.	0.	0.	0.0386108	0.	0.
0.384556	0.576834	0.	0.	0.	0.	0.	0.0386108	0.
0.314562	0.629124	0.	0.	0.	0.	0.	0.	0.0563135

N=40

1.	0	0	0	0	0	0	0	0
0	1.	0	0	0	0	0	0	0
0.517759	0.514644	0	1.1174×10^{-6}	3.41917×10^{-7}	9.25828×10^{-8}	4.27396×10^{-7}	3.41917×10^{-7}	1.1174×10^{-6}
0.55555	0.66666	0	0.0000100566	0	0	0	0	0
0.399999	0.599999	0	0	2.22246×10^{-6}	0	0	0	0
0.454545	0.545454	0	0	0	4.62914×10^{-7}	0	0	0
0.499999	0.599999	0	0	0	0	2.22246×10^{-6}	0	0
0.399999	0.599999	0	0	0	0	0	2.22246×10^{-6}	0
0.33333	0.66666	0	0	0	0	0	0	0.0000100566

N=81

1.	0	0	0	0	0	0	0	0
0	1.	0	0	0	0	0	0	0
0.517761	0.514646	0	0	0	0	0	0	0
0.555556	0.666667	0	0	0	0	0	0	0
0.4	0.6	0	0	0	0	0	0	0
0.454545	0.545455	0	0	0	0	0	0	0
0.5	0.6	0	0	0	0	0	0	0
0.4	0.6	0	0	0	0	0	0	0
0.333333	0.666667	0	0	0	0	0	0	0

Q matrix

To answer part (b) below, we need to run the system from different initial state vector (i.e. different $\pi^{(0)}$) and observe if the system probability state vector after a long time (i.e. $\pi^{(\infty)}$) will depend on the initial state vector or not. Here is the result for 3 different initial state vectors. In diagram below we show the $\pi^{(0)}$ and to its right $\pi^{(\infty)}$.

Initial probability state vector

```
piInitial = {0, 0, 1, 0, 0, 0, 0, 0, 0};
Chop[N[piInitial.MatrixPower[A, 100]]] // MatrixForm
```

$\pi^{(0)}P^n$

$$\begin{pmatrix} 0.517761 \\ 0.514646 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Initial probability state vector

```
piInitial = {0, 0, 0, 1, 0, 0, 0, 0, 0};
Chop[N[piInitial.MatrixPower[A, 100]]] // MatrixForm
```

$$\begin{pmatrix} 0.555556 \\ 0.666667 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Initial probability state vector

```
piInitial = {0, 0, 0, 0, 1, 0, 0, 0, 0};
Chop[N[piInitial.MatrixPower[A, 100]]] // MatrixForm
```

$$\begin{pmatrix} 0.4 \\ 0.6 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

2.2 Analysis of numerical results

2.2.1 (a)

Yes. The powers of P^n converges as $n \to \infty$. This is seen by looking at the above sequence of the P matrix where we see that the matrix P converges to the following limiting matrix at around $n = 81$

We can say the following about the limiting matrix: As $n \to \infty$ the matrix P converges to a fixed value shown above. The entries P_{ij}^n where j is a transient state goes to zero as n gets large.

2.2.2 (b)

From the above numerical result, we see that depending on the initial system probability state vector $\pi^{(0)}$ we obtain a different system probability state vector $\pi^{(n)}$ as n gets very large. This is because some states are transient (states $\{4, 5, 6, 8, 9, 10\}$). In the inventory problem below, we see that we obtained a different result for this part since the inventory problem has no transient states.

2.2.3 (c)

Let I be the set of all the possible states the system can be in. Hence from definition, we write

$$\pi_j^{(n)} = \sum_{i \in I} \pi_i^{(0)} P_{ij}^n$$

Where $\pi_j^{(n)}$ means the probability that the system will be in state j after n steps and P_{ij}^n is the n steps transition probability. Now take the limit of the above as $n \to \infty$ we have

$$\begin{aligned} \lim_{n\to\infty} \pi_j^{(n)} &= \lim_{n\to\infty} \sum_{i \in I} \pi_i^{(0)} P_{ij}^n \\ &= \sum_{i \in I} \lim_{n\to\infty} \left(\pi_i^{(0)} P_{ij}^n \right) \end{aligned}$$

Assume there are k states, we can expand

$$\lim_{n\to\infty} \left(\pi_1^{(0)} P_{1j}^n + \pi_2^{(0)} P_{2j}^n + \cdots + \pi_k^{(0)} P_{kj}^n \right)$$

But from part(a) we observed that in the limit, entries of each columns are not equal. Hence $P_{1j}^n \neq P_{2j}^n \neq \cdots \neq P_{kj}^n$ this means the above sum will produce a different value depending on the initial state probability vector $\pi^{(0)}$. (Compare this to the inventory problem below, where each entry in a column is the same, and we could factor it out of the sum and we reached a different conclusion than here).

Hence we showed depending on the initial $\pi^{(0)}$ then $\lim_{n\to\infty} \pi_j^{(n)}$ goes to different value as confirmed by the numerical result shown above in part(a). Hence part(a) results could be used to deduce part(b) conclusion.

3 Inventory problem

3.1 Summary of numerical results

An inventory program was written in Mathematica (please see appendix for full source code) which generated the P matrix for an increasing values of n. The specification of the inventory model is described in the question shown above. The value $\boxed{s=3}$ and $\boxed{S=5}$ was used.

The following are few results of the P matrix for an increasing values of n and the histogram of the demand distribution used.

Initial P matrix

$$\begin{pmatrix} 0.642857 & 0.119048 & 0.0952381 & 0.0714286 & 0.047619 & 0.0238095 \\ 0.642857 & 0.119048 & 0.0952381 & 0.0714286 & 0.047619 & 0.0238095 \\ 0.642857 & 0.119048 & 0.0952381 & 0.0714286 & 0.047619 & 0.0238095 \\ 0.857143 & 0.0714286 & 0.047619 & 0.0238095 & 0. & 0. \\ 0.761905 & 0.0952381 & 0.0714286 & 0.047619 & 0.0238095 & 0. \\ 0.642857 & 0.119048 & 0.0952381 & 0.0714286 & 0.047619 & 0.0238095 \end{pmatrix}$$

N=3

$$\begin{pmatrix} 0.66232 & 0.114836 & 0.0910269 & 0.0672174 & 0.0434078 & 0.021191 \\ 0.66232 & 0.114836 & 0.0910269 & 0.0672174 & 0.0434078 & 0.021191 \\ 0.66232 & 0.114836 & 0.0910269 & 0.0672174 & 0.0434078 & 0.021191 \\ 0.663454 & 0.114593 & 0.0907839 & 0.0669744 & 0.0431649 & 0.021029 \\ 0.662887 & 0.114715 & 0.0909054 & 0.0670959 & 0.0432864 & 0.02111 \\ 0.66232 & 0.114836 & 0.0910269 & 0.0672174 & 0.0434078 & 0.021191 \end{pmatrix}$$

N=5

$$\begin{pmatrix} 0.662421 & 0.114815 & 0.0910054 & 0.0671959 & 0.0433864 & 0.0211767 \\ 0.662421 & 0.114815 & 0.0910054 & 0.0671959 & 0.0433864 & 0.0211767 \\ 0.662421 & 0.114815 & 0.0910054 & 0.0671959 & 0.0433864 & 0.0211767 \\ 0.662427 & 0.114814 & 0.0910042 & 0.0671946 & 0.0433851 & 0.0211759 \\ 0.662424 & 0.114814 & 0.0910048 & 0.0671953 & 0.0433857 & 0.0211763 \\ 0.662421 & 0.114815 & 0.0910054 & 0.0671959 & 0.0433864 & 0.0211767 \end{pmatrix}$$

N=10

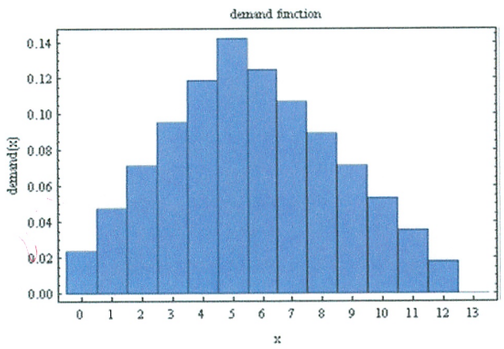
$$\begin{pmatrix} 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \end{pmatrix}$$

N=20

$$\begin{pmatrix} 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \end{pmatrix}$$

N=1000

$$\begin{pmatrix} 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \\ 0.662421 & 0.114815 & 0.0910053 & 0.0671958 & 0.0433862 & 0.0211766 \end{pmatrix}$$



To answer part (b) below, we need to run the system from different initial state vector (i.e. different $\pi^{(0)}$) and observe if the system probability state after a long time (i.e. $\pi^{(\infty)}$) will depend on the initial

3.2.2 (b)

Yes. There is a limiting state probability distribution in all cases. This is show by looking at the numerical result above that shows for different initial probability state vector $\pi^{(0)}$ we obtain the same probability state vector $\pi^{(n)}$ when n is large. So the final $\pi^{(\infty)}$ *does not depend* on which state the system starts from.

3.2.3 (c)

In this part, we need to show given that P^{∞} converges to limiting fixed value, then the $\pi_k^{(\infty)}$ is the same for all states k.

Let I be the set of all the possible states the system can be in. Hence from definition, we write

$$\pi_j^{(n)} = \sum_{i \in I} \pi_i^{(0)} P_{ij}^n$$

Where $\pi_j^{(n)}$ means the probability that the system will be in state j after n steps and P_{ij}^n is the n steps transition probability. Now take the limit of the above as $n \to \infty$ we have

$$\begin{aligned} \lim_{n\to\infty} \pi_j^{(n)} &= \lim_{n\to\infty} \sum_{i \in I} \pi_i^{(0)} P_{ij}^n \\ &= \sum_{i \in I} \lim_{n\to\infty} \left(\pi_i^{(0)} P_{ij}^n \right) \end{aligned}$$

Assume there are k states, we can expand

$$\lim_{n\to\infty} \left(\pi_1^{(0)} P_{1j}^n + \pi_2^{(0)} P_{2j}^n + \cdots + \pi_k^{(0)} P_{kj}^n \right)$$

But from part(a) we observed that $\lim_{n\to\infty} P_{ij}^n$ is a fixed value, which is the limit the transition matrix converged to. In other words, $P_{1j}^n = P_{2j}^n = \cdots = P_{kj}^n$ since all entries in the j column are the same. Call this entry in j^{th} column as k say. So k is a single number which represents the one step transition probability from state i to state j when the system has run for a long time. So we write the above as

$$\lim_{n\to\infty} \pi_j^{(n)} = k \left(\pi_1^{(0)} + \pi_2^{(0)} + \cdots + \pi_k^{(0)} \right)$$

now, $\sum_{i \in I} \pi_i^{(0)}$ is the sum of the probabilities of the system being in all its states at time zero, which must be 1 hence

$$\lim_{n\to\infty} \pi_j^{(n)} = k$$

Hence we showed that regardless of the initial $\pi^{(0)}$ then $\lim_{n\to\infty} \pi_j^{(n)}$ goes to some fixed values. This shows that for any state j the probability that the system will be in that state after a long time converges to a fixed value regardless of the initial state if the system transition matrix converges in the limit. Hence part(a) results could be used to deduce part(b) conclusion.

state vector or not. Since we know that

$$\begin{aligned}\pi^{(1)} &= \pi^{(0)}P\\ \pi^{(2)} &= \pi^{(1)}P = \pi^{(0)}P^{(2)}\\ \pi^{(3)} &= \pi^{(2)}P = \pi^{(0)}P^{(3)}\\ &\cdots\\ \pi^{(n)} &= \pi^{(0)}P^{(n)}\end{aligned}$$

And since $P^{(n)} = P^n$, then all what we have do is pick few $\pi^{(0)}$ vectors, and post multiply them by $P^{(n)}$ for large n and see if we obtain the same $\pi^{(n)}$. Below is the numerical result for this part showing the initial $\pi^{(0)}$ and the final $\pi^{(n)}$. I used $n = 30$ in all cases as this showed it is large enough from the above numerical results. Here are the results. Below we show result of 6 tests. In each one, $\pi^{(0)}$ is shown and to its right $\pi^{(n)}$.

$$\begin{pmatrix}\pi^{(0)}\\ 1\\0\\0\\0\\0\\0\end{pmatrix}\begin{pmatrix}\pi^{(30)}\\ 0.662421\\0.114815\\0.0910053\\0.0671958\\0.0433862\\0.0211766\end{pmatrix}\quad\begin{pmatrix}\pi^{(0)}\\ 0.5\\0.5\\0\\0\\0\\0\end{pmatrix}\begin{pmatrix}\pi^{(30)}\\ 0.662421\\0.114815\\0.0910053\\0.0671958\\0.0433862\\0.0211766\end{pmatrix}\quad\begin{pmatrix}\pi^{(0)}\\ 0.4\\0.2\\0.4\\0\\0\\0\end{pmatrix}\begin{pmatrix}\pi^{(30)}\\ 0.662421\\0.114815\\0.0910053\\0.0671958\\0.0433862\\0.0211766\end{pmatrix}$$

$$\begin{pmatrix}\pi^{(0)}\\ 0.1\\0.1\\0.4\\0.4\\0\\0\end{pmatrix}\begin{pmatrix}\pi^{(30)}\\ 0.662421\\0.114815\\0.0910053\\0.0671958\\0.0433862\\0.0211766\end{pmatrix}\quad\begin{pmatrix}\pi^{(0)}\\ 0.2\\0.3\\0.1\\0.1\\0.2\\0.1\end{pmatrix}\begin{pmatrix}\pi^{(30)}\\ 0.662421\\0.114815\\0.0910053\\0.0671958\\0.0433862\\0.0211766\end{pmatrix}\quad\begin{pmatrix}\pi^{(0)}\\ 0\\1\\0\\0\\0\\0\end{pmatrix}\begin{pmatrix}\pi^{(30)}\\ 0.662421\\0.114815\\0.0910053\\0.0671958\\0.0433862\\0.0211766\end{pmatrix}$$

3.2 Analysis of numerical results

3.2.1 (a)

Yes. The powers of P^n converges as $n \to \infty$. This is seen by looking at the above sequence of the P matrix where we see that the matrix P converges to the following limiting matrix at around $n = 20$

$$\begin{pmatrix}0.662421&0.114815&0.0910053&0.0671958&0.0433862&0.0211766\\0.662421&0.114815&0.0910053&0.0671958&0.0433862&0.0211766\\0.662421&0.114815&0.0910053&0.0671958&0.0433862&0.0211766\\0.662421&0.114815&0.0910053&0.0671958&0.0433862&0.0211766\\0.662421&0.114815&0.0910053&0.0671958&0.0433862&0.0211766\\0.662421&0.114815&0.0910053&0.0671958&0.0433862&0.0211766\end{pmatrix}$$

We can say the following about the limiting matrix: As $n \to \infty$ the matrix P converges to a fixed value shown above. Each column has the same entries in its rows. In addition, all entries are non-zero. This implies that the chain contains no transient states. And since all the values on the converged P matrix are positive, then we have only one closed set in the chain, which contains all the states.

Inventory problem. Computing assignment 3. Math 504, spring 2008. CSUF

by Nasser Abbasi

- **Define constants for the problem (give in class to use)**

```
s = 3;
S = 5;
```

- **define demand function (note the use of conditions on arguments)**

```
demand[x_] := (x + 1) /; x ≤ 5
demand[x_] := 3/4 (13 - x) /; 5 < x ≤ 13;
demand[x_] := 0;
points = Range[0, 13]
```

```
{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13}
```

- **define the and p (i, j) functions (see handout)**

```
p[i_, j_] := Total[Table[c * demand[v], {v, S, Length[points]}]] /; (j == 0 && i < s);
p[i_, j_] := Total[Table[c * demand[v], {v, i, Length[points]}]] /; (j == 0 && i ≥ s);
p[i_, j_] := c * demand[S - j] /; (i < s && (0 < j ≤ S));
p[i_, j_] := c * demand[i - j] /; (i ≥ s && (0 < j ≤ i));
p[i_, j_] := 0;
```

- **define a function which determines the ' c' constant by making at area = 1 under the demand function**

```
demandTable = Table[demand[x], {x, points[[1]], points[[-1]]}];
c = 1 / Total[demandTable];
Print["c=", N[c]];
demandTable = demandTable * c(*normalize it to make it probability*)
```

```
c=0.0238095
```

$$\left\{\frac{1}{42}, \frac{1}{21}, \frac{1}{14}, \frac{2}{21}, \frac{5}{42}, \frac{1}{7}, \frac{1}{8}, \frac{3}{28}, \frac{5}{56}, \frac{1}{14}, \frac{3}{56}, \frac{1}{28}, \frac{1}{56}, 0\right\}$$

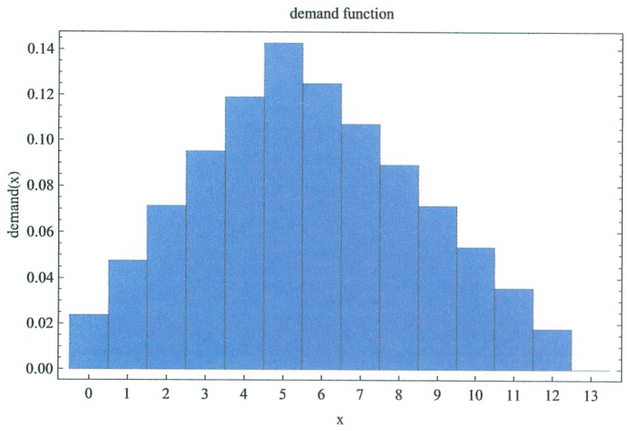
Print the demand function numerical values

```
t = TableForm[N[demandTable], TableHeadings → {Apply[ToString, points, 1]}]
```

0	0.0238095
1	0.047619
2	0.0714286
3	0.0952381
4	0.119048
5	0.142857
6	0.125
7	0.107143
8	0.0892857
9	0.0714286
10	0.0535714
11	0.0357143
12	0.0178571
13	0.

Make a histogram of the demand function

```
Needs["BarCharts`"]
t = Table[{x, demandTable[[x + 1]], 1}, {x, points[[1]], points[[-1]]}];
GeneralizedBarChart[t, AxesLabel → {"x", "demand(x)"},
 Frame → True, FrameTicks → {{Automatic, Automatic}, {points, None}},
 FrameLabel → {"x", "demand(x)", "demand function"}]
```



craps game markov chain P matrix

```
In[1]:= A = {{1, 0, 0, 0, 0, 0, 0, 0, 0}, {0, 1, 0, 0, 0, 0, 0, 0, 0},
    {8/36, 4/36, 0, 3/36, 4/36, 5/36, 5/36, 4/36, 3/36}, {5/36, 6/36, 0, 27/36, 0, 0, 0, 0, 0},
    {4/36, 6/36, 0, 0, 26/36, 0, 0, 0, 0}, {5/36, 6/36, 0, 0, 0, 25/36, 0, 0, 0}, {5/36, 6/36, 0, 0, 0,
     0, 26/36, 0, 0}, {4/36, 6/36, 0, 0, 0, 0, 0, 26/36, 0}, {3/36, 6/36, 0, 0, 0, 0, 0, 0, 27/36}};
MatrixForm[
 A]
```

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{2}{9} & \frac{1}{9} & 0 & \frac{1}{12} & \frac{1}{9} & \frac{5}{36} & \frac{5}{36} & \frac{1}{9} & \frac{1}{12} \\ \frac{5}{36} & \frac{1}{6} & 0 & \frac{3}{4} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{9} & \frac{1}{6} & 0 & 0 & \frac{13}{18} & 0 & 0 & 0 & 0 \\ \frac{5}{36} & \frac{1}{6} & 0 & 0 & 0 & \frac{25}{36} & 0 & 0 & 0 \\ \frac{5}{36} & \frac{1}{6} & 0 & 0 & 0 & 0 & \frac{13}{18} & 0 & 0 \\ \frac{1}{9} & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & \frac{13}{18} & 0 \\ \frac{1}{12} & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{3}{4} \end{pmatrix}$$

- **Raise the matrix for higher power**

```
N[A] // MatrixForm
```

$$\begin{pmatrix} 1. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. \\ 0. & 1. & 0. & 0. & 0. & 0. & 0. & 0. & 0. \\ 0.222222 & 0.111111 & 0. & 0.0833333 & 0.111111 & 0.138889 & 0.138889 & 0.111111 & 0.0833333 \\ 0.138889 & 0.166667 & 0. & 0.75 & 0. & 0. & 0. & 0. & 0. \\ 0.111111 & 0.166667 & 0. & 0. & 0.722222 & 0. & 0. & 0. & 0. \\ 0.138889 & 0.166667 & 0. & 0. & 0. & 0.694444 & 0. & 0. & 0. \\ 0.138889 & 0.166667 & 0. & 0. & 0. & 0. & 0.722222 & 0. & 0. \\ 0.111111 & 0.166667 & 0. & 0. & 0. & 0. & 0. & 0.722222 & 0. \\ 0.0833333 & 0.166667 & 0. & 0. & 0. & 0. & 0. & 0. & 0.75 \end{pmatrix}$$

```
N[MatrixPower[A, 3]] // MatrixForm
```

$$\begin{pmatrix} 1. & 0. & 0. & 0. & 0. & 0. & 0. & 0. & 0. \\ 0. & 1. & 0. & 0. & 0. & 0. & 0. & 0. & 0. \\ 0.363062 & 0.302598 & 0. & 0.046875 & 0.0579561 & 0.0669796 & 0.0724451 & 0.0579561 & 0.046875 \\ 0.321181 & 0.385417 & 0. & 0.421875 & 0. & 0. & 0. & 0. & 0. \\ 0.249314 & 0.373971 & 0. & 0. & 0.376715 & 0. & 0. & 0. & 0. \\ 0.302319 & 0.362783 & 0. & 0. & 0. & 0.334898 & 0. & 0. & 0. \\ 0.311643 & 0.373971 & 0. & 0. & 0. & 0. & 0.376715 & 0. & 0. \\ 0.249314 & 0.373971 & 0. & 0. & 0. & 0. & 0. & 0.376715 & 0. \\ 0.192708 & 0.385417 & 0. & 0. & 0. & 0. & 0. & 0. & 0.421875 \end{pmatrix}$$

```
N[MatrixPower[A, 10]] // MatrixForm
```

1.	0.	0.	0.	0.	0.	0.	0.	0.
0.	1.	0.	0.	0.	0.	0.	0.	0.
0.501363	0.491875	0.	0.00625706	0.00594012	0.00521681	0.00742515	0.00594012	0.00625706
0.52427	0.629124	0.	0.0563135	0.	0.	0.	0.	0.
0.384556	0.576834	0.	0.	0.0386108	0.	0.	0.	0.
0.442689	0.531227	0.	0.	0.	0.0260841	0.	0.	0.
0.480695	0.576834	0.	0.	0.	0.	0.0386108	0.	0.
0.384556	0.576834	0.	0.	0.	0.	0.	0.0386108	0.
0.314562	0.629124	0.	0.	0.	0.	0.	0.	0.0563135

```
Chop[N[MatrixPower[A, 40]]] // MatrixForm
```

1.	0	0	0	0	0	0	0	0
0	1.	0	0	0	0	0	0	0
0.517759	0.514644	0	1.1174×10^{-6}	3.41917×10^{-7}	9.25828×10^{-8}	4.27396×10^{-7}	3.41917×10^{-7}	1.
0.55555	0.66666	0	0.0000100566	0	0	0	0	0
0.399999	0.599999	0	0	2.22246×10^{-6}	0	0	0	0
0.454545	0.545454	0	0	0	4.62914×10^{-7}	0	0	0
0.499999	0.599999	0	0	0	0	2.22246×10^{-6}	0	0
0.399999	0.599999	0	0	0	0	0	2.22246×10^{-6}	0
0.33333	0.66666	0	0	0	0	0	0	0.

```
Chop[N[MatrixPower[A, 81]]] // MatrixForm
```

1.	0	0	0	0	0	0	0	0
0	1.	0	0	0	0	0	0	0
0.517761	0.514646	0	0	0	0	0	0	0
0.555556	0.666667	0	0	0	0	0	0	0
0.4	0.6	0	0	0	0	0	0	0
0.454545	0.545455	0	0	0	0	0	0	0
0.5	0.6	0	0	0	0	0	0	0
0.4	0.6	0	0	0	0	0	0	0
0.333333	0.666667	0	0	0	0	0	0	0

```
piInitial = {0, 0, 1, 0, 0, 0, 0, 0, 0};
Chop[N[piInitial.MatrixPower[A, 100]]] // MatrixForm
```

0.517761
0.514646
0
0
0
0
0
0
0

```
piInitial = {0, 0, 0, 1, 0, 0, 0, 0, 0};
Chop[N[piInitial.MatrixPower[A, 100]]] // MatrixForm
```

$$\begin{pmatrix} 0.555556 \\ 0.666667 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

```
piInitial = {0, 0, 0, 0, 1, 0, 0, 0, 0};
Chop[N[piInitial.MatrixPower[A, 100]]] // MatrixForm
```

$$\begin{pmatrix} 0.4 \\ 0.6 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

```
piInitial = {0, 0, 0, 0, .5, .5, 0, 0, 0};
Chop[N[piInitial.MatrixPower[A, 100]]] // MatrixForm
```

$$\begin{pmatrix} 0.427273 \\ 0.572727 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

4.6.5 Appendix

Source code for craps problem

Mathematica notebook

source code for inventory problem

Mathematica notebook

4.7 Practice problems

Grade: 2/2.

These are 5 problems to practice using method of characteristics to solve first order liner pde. The problems are listed in the handout

Method of Characteristics - Practice Problems

Use the method of characteristics to solve the following initial value problems for $t > 0$ and $-\infty < x < \infty$.

1. $u_t + 4u_x = 0$, and $u(x,0) = e^{x^2}$.
2. $u_t + (xt)u_x = 0$, and $u(x,0) = 2x$.
3. $u_t + (x \sin t)u_x = 0$, and $u(x,0) = 1/(1+x^2)$.
4. $u_t - (tx^2)\, u_x = 0$, and $u(x,0) = 1 + x$.
5. $u_t - u_x = xu$, and $u(x,0) = 2x$.

4.7.1 Problem 2

Solve

$$u_t + (xt)\, u_x = 0 \tag{1}$$

$u(x,0) = 2x$

Solution

Seek solution where $u(s) = u(t(s), x(s)) =$ constant,hence

$$\frac{du}{ds} = \frac{\partial u}{\partial t}\frac{dt}{ds} + \frac{\partial u}{\partial x}\frac{dx}{ds} = 0$$

Compare to (1) we see that $\frac{dt}{ds} = 1$ or $t = s$ and $\frac{dx}{ds} = xt$, but since $t = s$ then $\frac{dx}{ds} = xs$, and this has solution $x = x_0 \exp\left(\frac{s^2}{2}\right)$ but $s = t$, hence

$$x = x_0 \exp\left(\frac{t^2}{2}\right) \tag{2}$$

Now at $t = 0$, the solution is $2x_0$, but this solution is valid any where on this characteristic line and not just when $t = 0$. hence

$$u(x,t) = 2x_0$$

But $x_0 = x \exp\left(\frac{-t^2}{2}\right)$ from (2), hence

$$u(x,t) = 2x \exp\left(\frac{-t^2}{2}\right)$$

4.8 Problem 3

Solve

$u_t + (x \sin t)\, u_x = 0$

$u(x,0) = \frac{1}{1+x^2}$

Solution

Seek solution where $u(s) = u(t(s), x(s)) =$ constant,hence

$$\frac{du}{ds} = \frac{\partial u}{\partial t}\frac{dt}{ds} + \frac{\partial u}{\partial x}\frac{dx}{ds} = 0$$

Compare to (1) we see that $\frac{dt}{ds} = 1$ or $t = s$ and $\frac{dx}{ds} = x \sin t$, but since $t = s$ then $\frac{dx}{ds} = x \sin s$, and this has solution

$$\ln x = \int \sin(s)\, ds$$
$$x = x_0 \exp(-\cos(s))$$

but $s = t$ hence

$$x = x_0 \exp(-\cos(t)) \tag{1}$$

Hence

$$x_0 = x \exp(\cos(t)) \tag{2}$$

At $t = 0$,

$$x = x_0 \exp(-1)$$

Now we are told the solution at $t = 0$ is $\frac{1}{1+x^2}$, or $\frac{1}{1+[x_0 \exp(-1)]^2}$but this solution is valid any where on this characteristic line and not just when $t = 0$. hence

$$u(x,t) = \frac{1}{1 + [x_0 \exp(-1)]^2}$$

Replace the value of x_0 obtained in (2) we obtain

$$\begin{aligned} u(x,t) &= \frac{1}{1 + [x \exp(\cos(t)) \exp(-1)]^2} \\ &= \frac{1}{1 + x^2 \exp(2\cos(t)) \exp(-2)} \end{aligned}$$

Hence

$$u(x,t) = \frac{\exp(2)}{\exp(2) + x^2 \exp(2\cos(t))}$$

4.8.1 Problem 4

Solve

$u_t - (tx^2)\, u_x = 0$

$u(x,0) = 1 + x$

Solution

Seek solution where $u(s) = u(t(s)\,, x(s)) =$ constant,hence

$$\frac{du}{ds} = \frac{\partial u}{\partial t}\frac{dt}{ds} + \frac{\partial u}{\partial x}\frac{dx}{ds} = 0$$

Compare to (1) we see that $\frac{dt}{ds} = 1$ or $t = s$ and $\frac{dx}{ds} = -tx^2$, but since $t = s$ then $\frac{dx}{ds} = -sx^2$ hence we need to solve

$$\begin{aligned}\frac{dx}{x^2} &= -s ds \\ -\frac{1}{x} &= -\frac{s^2}{2} + x_0\end{aligned}$$

but $s = t$ hence

$$-\frac{1}{x} = -\frac{t^2}{2} + x_0 \tag{1}$$

Hence

$$x_0 = -\left(\frac{1}{x} - \frac{t^2}{2}\right) \tag{2}$$

At $t = 0\,,$

$$x_0 = -\frac{1}{x}$$

Now we are told the solution at $t = 0$ is $1 + x$, or $1 - \frac{1}{x_0}$but this solution is valid any where on this characteristic line and not just when $t = 0$. hence

$$u(x,t) = 1 - \frac{1}{x_0}$$

Replace the value of x_0 obtained in (2) we obtain

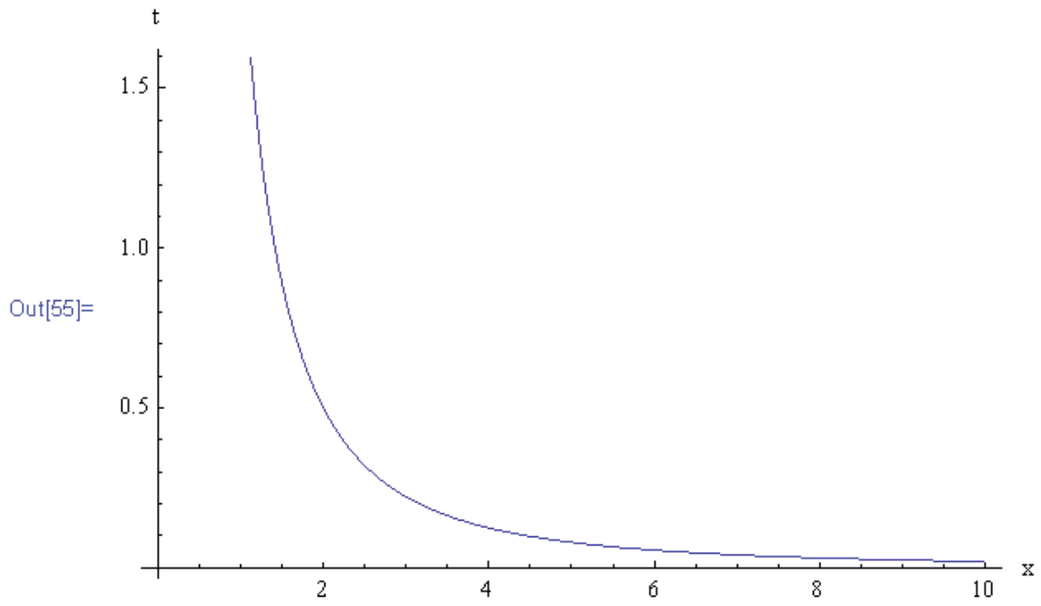
$$\begin{aligned}u(x,t) &= 1 - \frac{1}{-\left(\frac{1}{x} - \frac{t^2}{2}\right)} \\ &= 1 + \frac{2x}{2 - xt^2}\end{aligned}$$

Hence

$$u(x,t) = \frac{2 - xt^2 + 2x}{2 - xt^2}$$

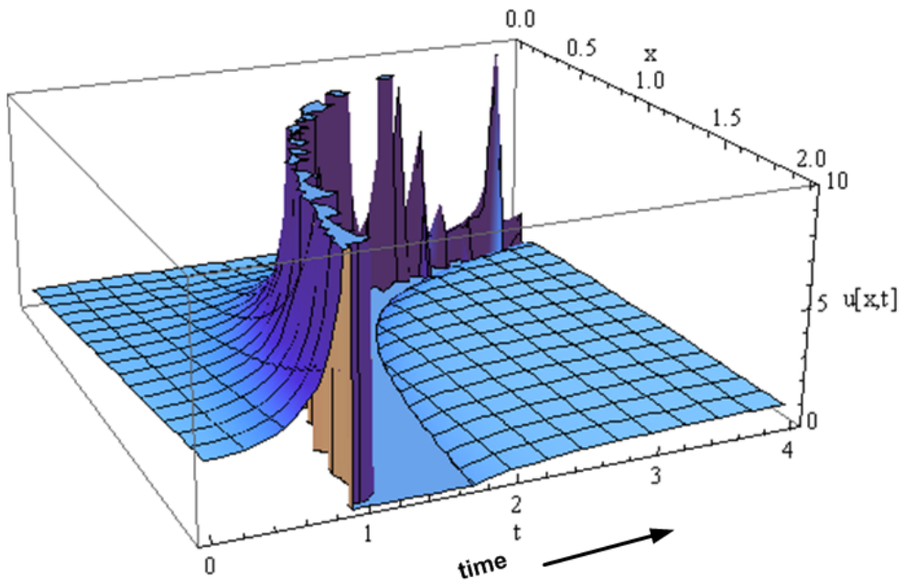
To avoid a solution u which blow up, we need $2 - xt^2 \neq 0$, hence $xt^2 \neq 2$, for example, $x = 2$ and $t = 1$ will not give a valid solution. so all region in $x - t$ plane in which $xt^2 = 2$ is not a valid region to apply this solution at.

The solution breaks down along this line in the $x - t$ plane

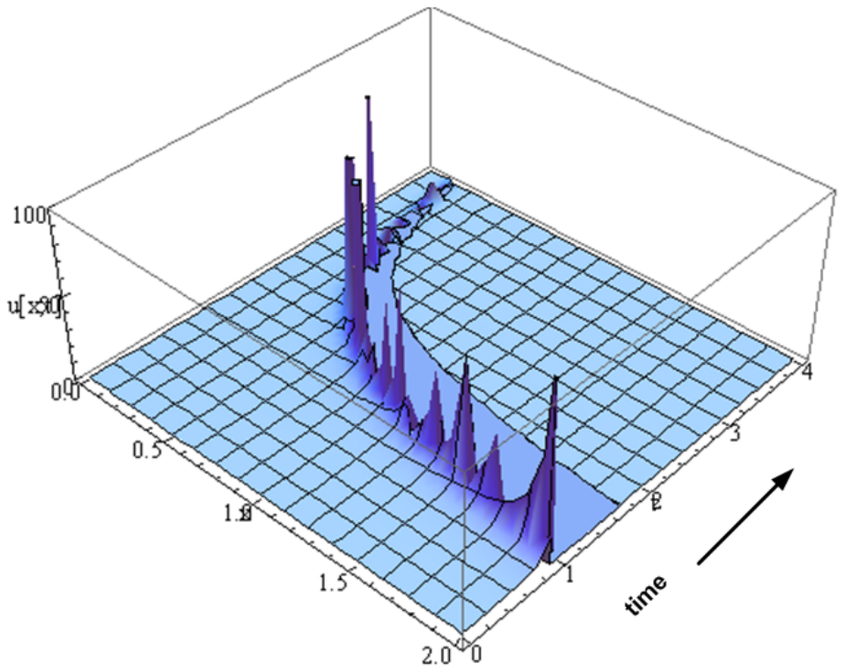


To see it in 3D, here is the $u(x,t)$ solution that includes the above line, and we see that the solution below the line and the above the line are not continuous across it. (I think there is a name to this phenomena that I remember reading about sometime, may be related to shockwaves but do not now know how this would happen in reality)

Out[51]=



Out[54]=



4.8.2 Problem 5

$u_t - u_x = xu$

$u(x,0) = 2x$

Solution

Nonhomogeneous pde first order.

(TO DO)

4.9 Monday 3/10/2008

Grade: 2/2.

Problem 5.7 from lecture notes (Irreducible matrix, analytical problem)

4.9.1 Problem

5.7 Consider an irreducible Markov chain with a finite number of states $\{0, 1, 2, \cdots, m\}$. Let $Q_i = P(\text{visit state } m \text{ before state } 0 \mid \text{start in state } i)$. Then $Q_0 = 0$ and $Q_m = 1$. (a) Find a system of $m-1$ linear equations that is satisfied by $Q_1, Q_2, \cdots, Q_{m-1}$. Help: Condition on the next state. (b) Show that the matrix which arises in part (a) is nonsingular. Help: Assume this matrix, say A, is singular so that there exists a vector $v \neq 0$ such that $Av = 0$. Normalize v so that one component is 1, and the rest are ≤ 1. You will need the irreducible property.

4.9.2 Answer

PART (A) First note that $Q_0 = 0$ and $Q_1 = 1$.

Let us define β as the event $\{visit\ state\ m\ before\ state\ 0\}$, then we write

$$Q_i = \Pr(\beta | X_0 = i)$$

But by conditioning on state of the chain at time 1 instead of time 0, we write[1]

$$\Pr(\beta | X_0 = i) = \sum_{k=0}^{k=m} \Pr(\beta | X_1 = k) \Pr(X_1 = k | X_0 = i)$$

But $\Pr(\beta | X_1 = k) = Q_k$ by definition, and $\Pr(X_1 = k | X_0 = i) = P_{ik}$,Therefore the above becomes

$$\boxed{Q_i = \sum_{k=0}^{k=m} Q_k P_{ik}}$$

Since $Q_k = 0$ and $Q_m = 1$, we can rewrite the above as follows

$$\begin{aligned} Q_i &= 0 + P_{im} + \sum_{k=1}^{k=m-1} Q_k P_{ik} \\ &= P_{im} + \sum_{k=1}^{k=m-1} Q_k P_{ik} \end{aligned}$$

If we examine the sum more closely, we see it is a product of a vector and a matrix. Since if we expand for few terms we see that

$$\begin{aligned} Q_1 &= P_{1,m} + (Q_1 P_{1,1} + Q_2 P_{1,2} + \cdots + Q_{m-1} P_{1,m-1}) \\ Q_2 &= P_{2,m} + (Q_1 P_{2,1} + Q_2 P_{2,2} + \cdots + Q_{m-1} P_{2,m-1}) \\ &\cdots \\ Q_{m-1} &= P_{m-1,m} + (Q_1 P_{m-1,1} + Q_2 P_{m-1,2} + \cdots + Q_{m-1} P_{m-1,m-1}) \end{aligned}$$

[1] Given an event β it is clear we can say

$$\Pr(\beta | X_0 = i) = \sum_{k \in all\ states} \Pr(\beta | X_1 = k) P_{ik}$$

since P_{ik} is the probability of going from state i to state k in one step. This works since we assume the Markov property which says the probability of transition to next state depends only on current state and not on any earlier state (for an order 1 Markov chain).

Which can be written as

$$\begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_{m-1} \end{bmatrix} = \begin{bmatrix} P_{1,m} \\ P_{2,m} \\ \vdots \\ P_{m-1,m} \end{bmatrix} + \begin{bmatrix} P_{1,1} & P_{1,2} & \cdots & P_{1,m-1} \\ P_{2,1} & P_{2,2} & \cdots & P_{2,m-1} \\ \cdots & \vdots & \vdots & \cdots \\ P_{m-1,m} & P_{m-1,2} & \cdots & P_{m-1,m-1} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_{m-1} \end{bmatrix}$$

Let $x = \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_{m-1} \end{bmatrix}$ and let $B = \begin{bmatrix} P_{1,1} & P_{1,2} & \cdots & P_{1,m-1} \\ P_{2,1} & P_{2,2} & \cdots & P_{2,m-1} \\ \cdots & \vdots & \ddots & \cdots \\ P_{m-1,m} & P_{m-1,2} & \cdots & P_{m-1,m-1} \end{bmatrix}$, and let $b = \begin{bmatrix} P_{1,m} \\ P_{2,m} \\ \vdots \\ P_{m-1,m} \end{bmatrix}$, then the above can be written as

$$\begin{aligned} x &= b + Bx \\ x - Bx &= b \\ (I - B)\, x &= b \end{aligned}$$

Where I is the identity matrix of order $m-1$. Now let $A = (I - B)$. hence

$$Ax = b$$

Therefore we can find x (which is the $Q's$) if we can solve the above. i.e. if we can invert the matrix A.(i.e. A is non-singular)

PART(B)

Now we need to show that $(I - B)$ is invertible. Recall that a Matrix A is not invertible if we can find a vector $\mathbf{v} \neq \mathbf{0}$ such that $A\mathbf{v} = 0$.

Let us assume that $(I - B)$ is not invertible. Hence there exist a vector $\mathbf{v} \neq \mathbf{0}$ such that

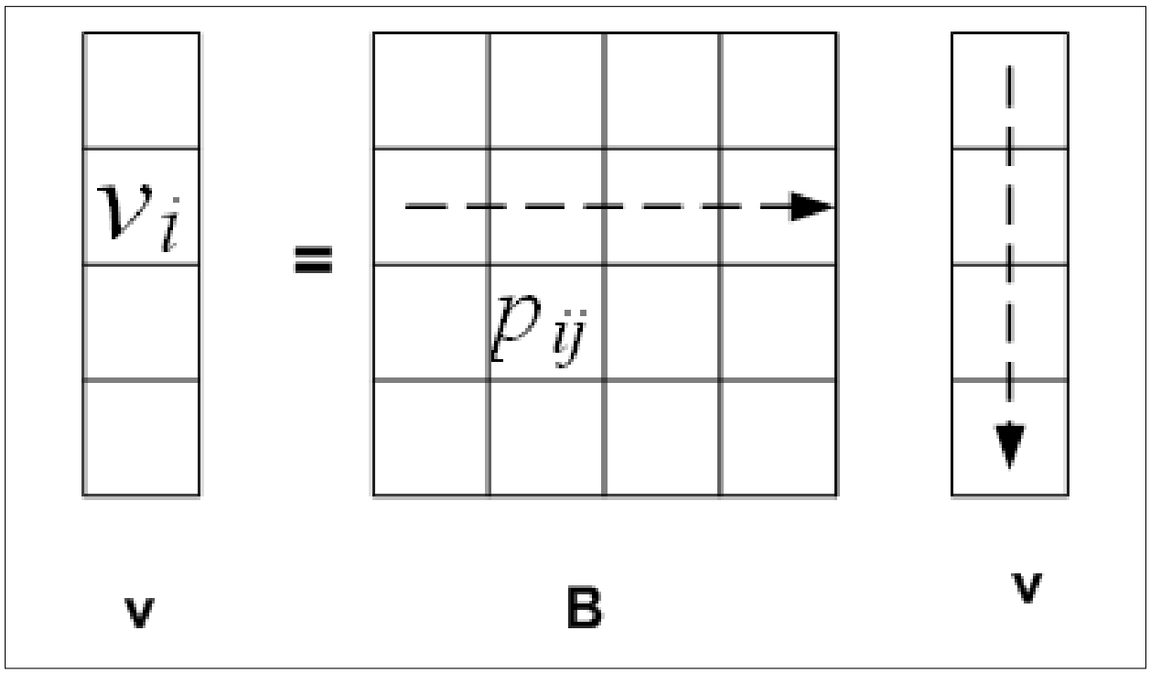
$$(I - B)\, \mathbf{v} = \mathbf{0}$$

In other words

$$\mathbf{v} = B\mathbf{v} \tag{1}$$

Now we show that it is not possible to find such a vector $\mathbf{v}$, showing that $(I - B)$ must therefore has an inverse.

We can always normalized the vector $\mathbf{v}$ in (1) without changing this relation, hence we assume $\mathbf{v}$ is normalized such that its largest component v_i has length 1 (we do this by dividing the vector by the largest component it had). Now (1) can be written in component form as follows



$$v_i = \sum_{j=1}^{m-1} p_{ij} v_j \tag{2}$$

Since $\mathbf{v}$ is normalized, it will have at least one component which is 1 in value, and it can have components which are less than 1 in value (we proof this part below). Let the set of the indices of those components of $\mathbf{v}$ which are 1 be the set J and let the set of the indices of those components which are less than 1 be the set S. In other words

$$J = \{i : v_i = 1\}$$
$$S = \{i : v_i < 1\}$$

First, we show that the set S can not be empty: Proof by contradiction. Assume S is empty. Hence every element in the vector $\mathbf{v}$ is 1. Let us pick one of these elements $v_i = 1$ such that i corresponds to a row number in the matrix B where this row happens to sum to a value less than *one*[2]. Then we write

$$1 = \sum_{j=1}^{m-1} p_{ij} v_i = \sum_{j=1}^{m-1} p_{ij}(1) = \sum_{j=1}^{m-1} p_{ij}$$

But since this row sums to less than one, then the RHS above is less than 1. Hence this is a contradiction, hence $\boxed{\text{the set } S \text{ can not be empty}}$.

Now that we showed the set S is not empty, we can write (2) as a sum over the set J and the set S of indices. (We know the set J is not empty by definition, since the vector v is normalized, so it will have at least one element in the set J). Hence (2) becomes

$$\boxed{v_i = \sum_{j \in K} p_{ij} v_j + \sum_{j \in S} p_{ij} v_j} \tag{3}$$

Let us again pick one of those v_i components which has value 1 (we know there is at least one of these), and try to see if this equality holds for this row i. So (3) becomes

$$1 = \sum_{j \in J} p_{ij} v_j + \sum_{j \in S} p_{ij} v_j$$

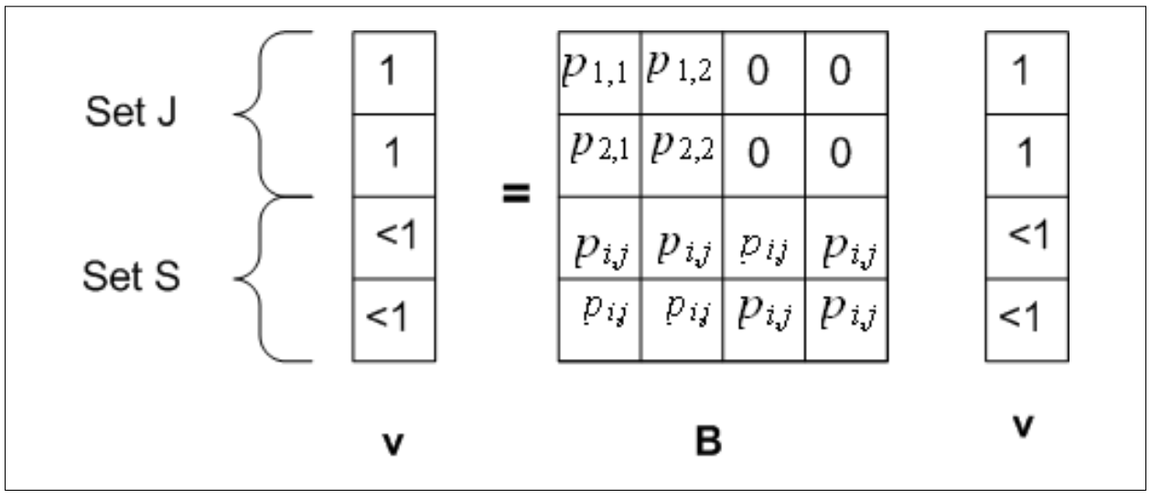
[2]We know that we can find such a row in B using the following argument: Assume that there is no row in B which sums to less than 1. This means B is an irreducable transitional probability matrix. However this is a submatrix of an original probability transition matrix which is irreducable, meaning it has no closed subsets. Hence B can not be irreduacble (closed). Therefore, we can find at least one row in B which sums to less than 1. (Matrix B is similar to a Q matrix, it has at least one row which sums to less than 1).

But all the v_j in the set J have value 1, so the above can be simplified

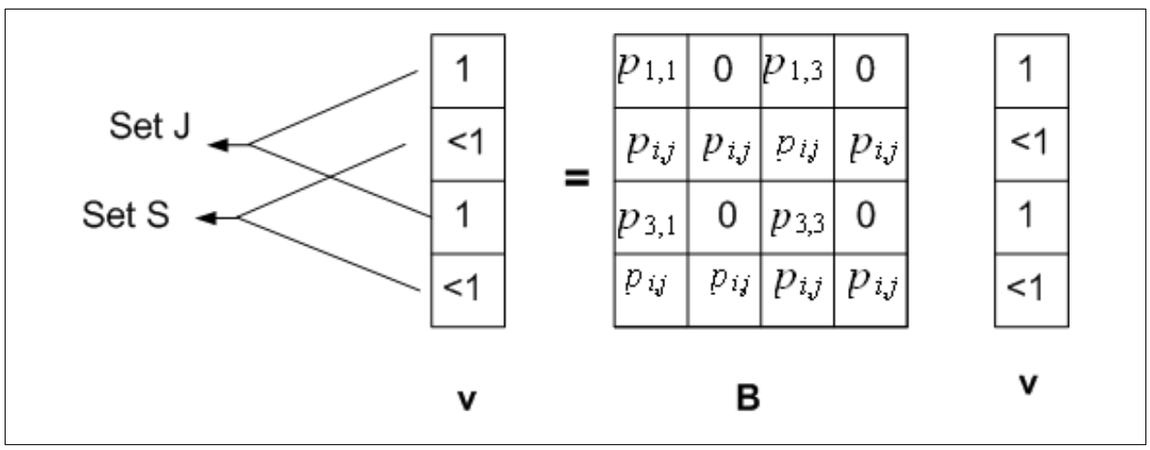
$$1 = \sum_{j \in J} p_{ij} + \overbrace{\sum_{j \in S} p_{ij} v_j}^{Y} \tag{4}$$

Now each v_j in the term labeled Y above is less than 1 (since it is the set S), so this means $\sum_{j \in S} p_{ij} v_j < \sum_{j \in S} p_{ij}$, therefore the sum in (4) could never add to 1 if there are values p_{ij} that are non-zero when j is in the set S. (since the sum $\sum_{j \in \{J \cup S\}} p_{ij}$ is being reduced from its original row sum). So for (4) to be satisfied, we need to have all the $p_{ij} = 0$ when j is in the set S. Hence the sum labeled Y is zero.

What this means is that if $v_i = 1$, then the i^{th}row in the matrix B must have zero entries in the columns which correspond to the indices in the set S. As shown in this diagram as an example



In the above diagram, I showed one example of the conclusion of above argument. Of course the set J the way I drawn it does not have to be 'contiguous', it could be in any pattern, as say the following



Therefore, we see that $p_{js} = 0$ when j correspond to a state whose number is the same as the index value in the set J, and s is a state whose number correspond to a state whose number is the same as the index value in the set S.

What this means is that it is not possible to reach states that correspond to indices in the set J from states which correspond to indices in the set S.

Hence, once the chain is in a state in the set J it is not possible to leave this set.

But this is the same as saying the matrix B contains a closed subset. In other words, B is reducible. However, this is not possible, since the matrix B is taken from a subset of a chain which is irreducible, i.e. it contains no closed subsets, but we found at least one such subset.

Therefore, we conclude that our assumption which lead to this is invalid. Therefore, there exist no vector $\mathbf{v} \neq \mathbf{0}$ such that $(I - B)\,\mathbf{v} = 0$. Hence $(I - B)$ does have an inverse. QED

4.9.3 Key solution

handout Solution 3/17/08

Chapter 5: Some Solutions

5.7 Consider an irreducible Markov chain with a finite number of states $\{0, 1, 2, \cdots, m\}$. Let E denote the event that the process reaches state m before it reaches state 0. Set $Q_i = P(E \mid X_0 = i)$. Then $Q_0 = 0$ and $Q_m = 1$. (a) Find a system of $m - 1$ linear equations that is satisfied by $Q_1, Q_2, \cdots, Q_{m-1}$. (b) Show that the matrix in this linear system of equations is nonsingular.

Solution (a) Conditioning on the next state gives us

$$Q_i = \sum_{j=0}^{m} P(E \mid X_1 = j) p_{ij} = \sum_{j=1}^{m-1} Q_j p_{ij} + p_{im} \,,$$

for $i \in I = \{1, 2, \cdots, m-1\}$. Note that the Markov property was used to get the second equation. (b) If the matrix in this linear system of equations is singular, then there is a nonzero vector $v \in R^{m-1}$ such that

$$v_i = \sum_{j=1}^{m-1} v_j p_{ij} \,,$$

for $i \in I$. Normalize v so that each component is less than or equal to one, and at least one component is equal to one. Set $J = \{i \in I \mid v_i = 1\}$ and $S = \{i \in I \mid v_i < 1\}$.

Suppose first that $S \neq \emptyset$. Assume there is a nonzero p_{ir} for some $i \in J$ and some $r \in S$. It then follows that

$$1 = v_i = \sum_{j \in J} v_j p_{ij} + \sum_{j \in S} v_j p_{ij} < \sum_{j=1}^{m-1} p_{ij} \leq 1 \,,$$

which is impossible. Therefore, for any $i \in J$, and any $r \in S$, we must have $p_{ir} = 0$. But then, for each $i \in J$,

$$1 = v_i = \sum_{j \in J} v_j p_{ij} = \sum_{j \in J} p_{ij} \,.$$

1

This result tells us that once the process enters a state $i \in J$, it must travel to another state in J. Thus, in the original chain, the set of states J, which is not empty, must be a closed set. However, this result is impossible, since the original chain is irreducible. Therefore, $S = \emptyset$.

It now follows that $J = \{1, 2, \cdots, m-1\}$. But then, for each $i \in J$,

$$1 = v_i = \sum_{j \in J} v_j p_{ij} = \sum_{j \in J} p_{ij} \ ,$$

since $v_j = 1$ for each $j \in J$. Thus, as before, the set J is seen to be a closed set, which is a contradiction.

4.10 Monday 3/17/08

Grade: 2/2.

Problems 6.3 and 6.5 from the handout

4.10.1 Problems

6.3 For an absorbing Markov chain, let V_{ij} denote the number of visits made to transient state j before absorption, given that the process starts in transient state i. Let B be the matrix whose (i,j)-th entry is $b_{ij} = E(V_{ij})$. (a) Show that $B = N = (I - Q)^{-1}$. Help: Condition on the next state X_1. Write out the resulting equations for each (i,j), and then appeal to matrix multiplication to get the result. For $i = j$, include this initial condition as one of the visits to state j, among possibly others before absorption. (b) Given that the process starts in transient state i, give a formula for the expected number of steps until absorption.

6.4 Regarding the proof of Theorem 6.3.1, derive equation (6.3.1) and show that $0 < d \leq 1/2$. Further, verify the results listed in the paragraph after the statement of Theorem 6.3.1.

6.5 Consider a regular Markov chain, and denote by T_{ij} the first entrance time into state j, given that the process starts in state i. Set $m_{ij} = E(T_{ij})$. (a) Show that $m_{ij} = 1 + \sum_{k \neq j} p_{ik} m_{kj}$. Help: Use conditional expectation, and condition on the next state. (b) Let $(w_1, w_2, \cdots, w_r)$ be the stationary probability vector for the process. Show that $m_{jj} = 1/w_j$, for each state j. Help: Use the result of (a). Multiply the i-th equation by w_i, and then sum over i. (c) Give a heuristic argument to justify the result of part (b).

4.10.2 Problem 6.3

PART(A)

Let I_n be an indicator variable defined as

$$I_n = \begin{cases} 1 & when\ (X_n = j | X_0 = i) \\ 0 & otherwise \end{cases}$$

Hence

$$E(I_n) = P(X_n = j | X_0 = i)$$

Now we see that

$$E(V_{ij}) = E\left(\sum_{n=0}^{\infty} I_n\right) = \sum_{n=0}^{\infty} E(I_n) = \sum_{n=0}^{\infty} P(X_n = j | X_0 = i)$$

Now, let b_{ij} be entry in matrix B where $b_{ij} = E(V_{ij})$, then the above can be written as

$$b_{ij} = P(X_0 = j | X_0 = i) + P(X_1 = j | X_0 = i) + P(X_2 = j | X_0 = i) + P(X_3 = j | X_0 = i) + \cdots \tag{1}$$

Which is the same as writing

$$b_{ij} = P_{ij}^{(0)} + P_{ij}^{(1)} + P_{ij}^{(2)} + P_{ij}^{(3)} + \cdots$$

When $i = j$, then $P_{ij}^{(0)} = 1$ otherwise it is 0. Hence

$$b_{ij} = \delta_{ij} + P_{ij}^{(1)} + P_{ij}^{(2)} + P_{ij}^{(3)} + P_{ij}^{(4)} + \cdots \tag{2}$$

Let the set of transient states be T, and using chapman-kolmogorov, the above can be written as

$$b_{ij} = \delta_{ij} + P_{ij}^{(1)} + \overbrace{\sum_{k\in T} P_{ik}^{(1)} P_{kj}^{(1)}}^{P_{ij}^{(2)}} + \overbrace{\sum_{k\in T} P_{ik}^{(2)} P_{kj}^{(1)}}^{P_{ij}^{(3)}} + \overbrace{\sum_{k\in T} P_{ik}^{(3)} P_{kj}^{(1)}}^{P_{ij}^{(4)}} + \cdots \tag{2}$$

But $\overbrace{\sum_{k\in T} P_{ik}^{(1)} P_{kj}^{(1)}}^{P_{ij}^{(2)}}$ is multiplying the i^{th} row of the Q matrix by the j^{th} column of the Q matrix. which is the (i,j) entry of the matrix Q^2, and $\overbrace{\sum_{k\in T} P_{ik}^{(2)} P_{kj}^{(1)}}^{P_{ij}^{(3)}}$ is multiplying the i^{th} row of the Q^2 matrix we just obtained, by the j^{th} column of the Q matrix, which is the (i,j) entry of the matrix Q^3. Continue this way, we obtain that $P_{ij}^{(4)}$ is the entry i,j in matrix Q^4 and so on.

Hence we see that b_{ij} is the (i,j) entry of a matrix resulting from $I + Q + Q^2 + Q^3 + \cdots$ QED.

PART(B) From part(A), we obtained that $E(V_{ij})$ is the (i,j) entry in the matrix resulting from the sum $I+Q+Q^2+Q^3+\cdots$. Since this is a Q matrix, then we know its elements will all go to zero an n gets very large, so this is a convergent sum, hence $I+Q+Q^2+Q^3+\cdots.\rightarrow (I-Q)^{-1}$. Therefore

$$\boxed{E(V_{ij}) \text{ is the } (i,j) \text{ entry in the matrix } (I-Q)^{-1}.}$$

Problem 6.5

PART(A) I solve this part in 2 ways, first by conditioning on next state, as required, and then by the counting method explained in the lecture.

by conditioning on next state. Let I be the set of all states. Then

$$E(T_{ij}) = \sum_{k\in I} E(T_{ij}|X_1 = k, X_0 = i)\, P(X_1 = k|X_0 = i) \tag{1}$$

But by Markov property, chain state on next step depends only on current state. Hence $E(T_{ij}|X_1 = k, X_0 = i) = E(T_{ij}|X_1 = k)$ and also since $P(X_1 = k|X_0 = i) = P_{ki}$ then (1) can be written as

$$E(T_{ij}) = \sum_{k\in I} E(T_{kj})\, P_{ki} \tag{2}$$

Now, when $X_1 = j$, then $E(T_{ij}) = 1$ since chain already in state j after one step. Therefore (2) can be rewritten as

$$E(T_{ij}) = 1 + \sum_{\substack{k\in I \\ k\neq j}} E(T_{ij}|X_1 = k)\, P_{ki}$$

But $E(T_{ij}|X_1 = k)$ is the same as writing $E(T_{kj})$, so the above becomes

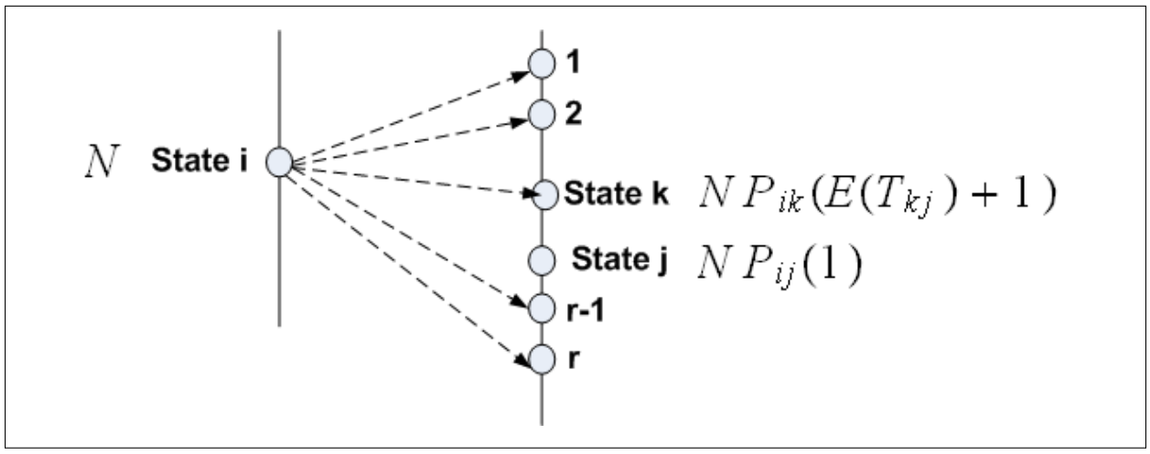
$$E(T_{ij}) = 1 + \sum_{\substack{k\in I \\ k\neq j}} E(T_{kj})\, P_{ki}$$

Using the notation shown in the problem, the above becomes

$$m_{ij} = 1 + \sum_{\substack{k\in I \\ k\neq j}} m_{kj} P_{ki}$$

QED.

Now solve part(a) using first a counting argument, and using the following diagram as a guide



Then we write (letting $E(T_{ij}) = m_{ij}$ and $E(T_{kj}) = m_{kj}$)

$$m_{ij} = \frac{N(P_{ij}) + \sum_{k\neq j} NP_{ik}(m_{kj} + 1)}{N}$$

$$= P_{ij} + \sum_{k\neq j} P_{ik}(m_{kj} + 1)$$

$$= P_{ij} + \sum_{k\neq j} P_{ik} m_{kj} + \sum_{k\neq j} P_{ik}$$

But $P_{ij} + \sum_{k \neq j} P_{ik} = 1$ hence the above becomes

$$m_{ij} = 1 + \sum_{k \neq j} P_{ik} m_{kj} \tag{1}$$

PART(B) We start from the result of part (A) which is

$$m_{ij} = 1 + \sum_{k \neq j} P_{ik} m_{kj}$$

Multiply both sides by w_i and obtain

$$w_i \; m_{ij} = w_i + w_i \sum_{k \neq j} P_{ik} m_{kj}$$

Sum over all possible states i and obtain

$$\sum_{i=1}^{r} w_i \; m_{ij} = \sum_{i=1}^{r} w_i + \sum_{i=1}^{r} \left(w_i \sum_{k \neq j} P_{ik} m_{kj} \right) \tag{2}$$

But $\sum_{i=1}^{r} w_i = 1$ and $\sum_{i=1}^{r} \left(w_i \sum_{k \neq j} P_{ik} m_{kj} \right) = \sum_{i=1}^{r} \left(\sum_{k \neq j} w_i P_{ik} m_{kj} \right) = \sum_{k \neq j} m_{kj} \left(\sum_{i=1}^{r} w_i P_{ik} \right)$, hence (2) becomes

$$\sum_{i=1}^{r} w_i \; m_{ij} = 1 + \sum_{k \neq j} m_{kj} \left(\sum_{i=1}^{r} w_i P_{ik} \right) \tag{3}$$

Now, since $w = \{w_1, w_2, \cdots, w_r\}$ is the stationary state vector, then it satisfies the following relation

$$w = wP$$

Where P is the one step probability transition matrix. The solution to the above is given by

$$w_k = \sum_{i=1}^{r} w_i P_{ik} \tag{4}$$

Where k is any state. Using (4) into RHS of (3), we can rewrite (3) as

$$\sum_{i=1}^{r} w_i \; m_{ij} = 1 + \sum_{\substack{k=1 \\ k \neq j}}^{r} m_{kj} w_k$$

$$\overbrace{\sum_{i=1}^{r} w_i \; m_{ij}}^{A} - \overbrace{\sum_{\substack{k=1 \\ k \neq j}}^{r} m_{kj} w_k}^{B} = 1 \tag{5}$$

Now looking at the LHS, we see that the first sum labeled A counts for all the $w's$ and the second sum labeled B also counts for all the $w's$ except for the j term. Hence if we subtract B from A, only the term $m_{jj} w_j$ will survive. Hence (5) becomes

$$m_{jj} w_j = 1$$

or

$$\boxed{m_{jj} = \frac{1}{w_j}}$$

QED.

PART (C)

If we wait for the chain to arrive at its steady state (i.e. we the chain probability state vector does not change, or $w = wP$), then we observe the chain from that point on, for a long

period of time, say T. The number of times the chain will be in state j during this time T is then given by $w_j T$, since w_j is the probability of the chain being in state j. So, to find the average number of time units (steps) it took for the chain for go from state j back to state j we need to divide T by the number of times the chain was in state j during this time, which we just found as $w_j T$

Hence

$$m_{jj} = \frac{T}{w_j T} = \frac{1}{w_j}$$

Intuitively this makes sense. Since the smaller the probability that the chain will be in state j we would expect the time between the events that the chain is in state j to become larger, So the relation should be an inverse one, as was found. QED

4.10.3 Key solution

Handout March 19, 2008

Chapter 6 Some Solutions

6.3 For an absorbing Markov chain, let V_{ij} denote the number of visits to transient state j before absorption, given that the process starts in transient state i. Let B denote the matrix whose (i, j)-th entry is $b_{ij} = E(V_{ij})$. (*a*) Show that $B = N = (I - Q)^{-1}$. (*b*) Given that the process starts in state i, give a formula for the expected number of steps until absorption.

Solution (*a*) Condition on the next state to get

$$E(V_{ij}) = \sum_{k \in A} E(V_{ij} \mid X_1 = k) p_{ik} + \sum_{k \in T} E(V_{ij} \mid X_1 = k) p_{ik}$$

where A denotes the set of absorbing states, and T denotes the set of transient states. Let δ_{ij} be the Kronecker delta, which equals one when $i = j$, and equals zero otherwise. Then, in the first sum, $E(V_{ij} \mid X_1 = k) = \delta_{ij}$, since k is an absorbing state. For the second sum, making use of the Markov property, $E(V_{ij} \mid X_1 = k) = \delta_{ij} + E(V_{kj})$, since here k is a transient state. Thus,

$$E(V_{ij}) = \sum_{k \in A} \delta_{ij} p_{ik} + \sum_{k \in T} [\delta_{ij} + E(V_{kj})] p_{ik} = \delta_{ij} + \sum_{k \in T} p_{ik} E(V_{kj}) \, .$$

Therefore, for each pair of states i and j in T,

$$b_{ij} - \sum_{k \in T} p_{ik} b_{kj} = \delta_{ij} \, .$$

In matrix form, these equations are expressed as $B - QB = I$, where I is the identity matrix. Note that this result shows that $I - Q$ is invertible, and that $B = (I - Q)^{-1}$.

another proof that (I-Q) is invertible.

Here is another proof. Note first that

$$E(V_{ij}) = E\left(\sum_{n=0}^{\infty} I_n \mid X_0 = i \right) \, ,$$

where $I_n = 1$ if $X_n = j$ and $I_n = 0$ otherwise. Thus,

$$E(V_{ij}) = E\left(\sum_{n=0}^{\infty} I_n \mid X_0 = i \right) = \sum_{n=0}^{\infty} p_{ij}^{(n)} = \delta_{ij} + \sum_{n=1}^{\infty} q_{ij}^{(n)} \, ,$$

1

where $q_{ij}^{(n)}$ is the (i,j)-th element of the matrix Q^n. Therefore, $E(V_{ij})$ is the (i,j)-th element of the matrix

$$I + \sum_{n=1}^{\infty} Q^n = (I - Q)^{-1} .$$

(b) The sum of the elements in the i-th row of B is the expected value of the random variable

$$Y_i = \sum_{j \in T} V_{ij} ,$$

which is the number of visits to transient states before absorption, given that the process started in transient state i.

6.5 Consider a regular Markov chain, with state space $I = \{1, 2, \cdots, r\}$. Denote by T_{ij} the first entrance time into state j, given that the process starts in state i. Set $m_{ij} = E(T_{ij})$. (a) Show that

$$m_{ij} = 1 + \sum_{k \neq j} p_{ik} m_{kj} .$$

(b) Let $(w_1, w_2, \cdots, w_r)$ be the stationary probability state vector for the process. Show that $m_{jj} = 1/w_j$, for each state j. (c) Give a heuristic argument to justify the result of part (b).

Solution (a) Conditioning on the next state yields

$$E(T_{ij}) = \sum_{k \in I} E(T_{ij} \mid X_1 = k) p_{ik} = E(T_{ij} \mid X_1 = j) p_{ij} + \sum_{k \neq j} E(T_{ij} \mid X_1 = k) p_{ik} .$$

This equation then becomes

$$E(T_{ij}) = 1 \cdot p_{ij} + \sum_{k \neq j} (1 + E(T_{kj})) p_{ik} = 1 + \sum_{k \neq j} p_{ik} E(T_{kj}) ,$$

which is the result to be shown. (b) In the last equation above, for a fixed j, multiply the i-th equation by w_i, and sum over i to get

$$\sum_{i=1}^{r} w_i E(T_{ij}) = \sum_{i=1}^{r} w_i + \sum_{i=1}^{r} \sum_{k \neq j} w_i p_{ik} E(T_{kj}) .$$

We have that $w = wP$, and the sum of the components of w is one. Therefore, interchanging the order of summation, we get

$$\sum_{i=1}^{r} w_i E(T_{ij}) = 1 + \sum_{k \neq j} w_k E(T_{kj}) \ .$$

Cancelling like terms on each side of this equation yields $w_j E(T_{jj}) = 1$, as required. (*c*) Over a long number of time steps T, the average number of times the process is in state j is $w_j T$. But the average duration between these times that the process is in state j is m_{ij}. Hence, in the long-run, $(w_j T) m_{ij} = T$. Thus, dividing by T, yields $m_{jj} = 1/w_j$.

4.11 Wed 4/16/2008

Grade: 2/2.

These problem related to Hastings-Meropolis algorithm. And Proofing a Markov chain is irreducible, regular and time inverse. Implemented the simulation using Mathematica

4.11.1 Problem 8.4

8.4 Let $G = (V, E)$ be an undirected, connected graph, with the property that each vertex is connected to at most r vertices. Let f be a positive function defined on V, and let π denote the probability distribution

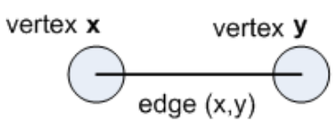
$$\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)} .$$

If $(x, y) \in E$, define the transition probability

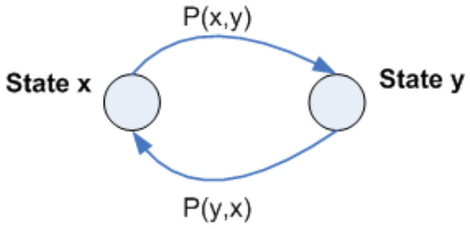
$$p(x, y) = \frac{1}{r} min\left\{1, \frac{f(y)}{f(x)}\right\} ,$$

with $p(x, y) = 0$ otherwise, except that $p(x, x)$ is determined so that the rows sum to one. (i) Show that the Markov chain determined by p is irreducible. (ii) Determine conditions under which the chain is regular. (iii) Show the chain is time reversible with respect to π.

Part(i)



Undirected Graph representation



Markov Chain representation

M.C. is irreducible if there exist no proper closed subset in the state space. Since we are given that the graph G is connected, then this means it is possible to visit each vertex from any other vertex in the graph. But does a connected graph implies no proper closed subset of the corresponding M.C.? The answer is YES. If we view each vertex as state, we just need to show that for each edge in G between 2 vertices x, y, there corresponds a probability of transition from state x to y which is not zero, and also a probability of transition from state y to x which is also not zero. By showing this, we conclude that the M.C. will switch (in some number of steps) to any state from any other state, which implies there is no closed subset, hence P is irreducible.

But from the definition of $p(x, y)$ we see that if there is an edge (x, y) then $p(x, y)$ exist and is not zero, and $p(y, x)$ exist and is not zero (since r is finite). This completes the proof.

Part(ii)

A finite M.C. is regular when, for some integer m, P^m contains only positive elements.

This implies that the one step transition matrix P must have at least one entry along the diagonal P_{ii} that is none-zero (If all elements along the diagonal are zero, then P^m will always contain at least one zero element no matter how large m is). But a diagonal element not being zero is the same as saying that at least one state must be aperiodic (if $P_{ii} > 0$ then the period is one).

Hence the condition for the M.C. to be regular is that at least one state must be aperiodic [3].

To proof that the above chain is regular, we then need to show that at least one state is aperiodic.

This is the proof:

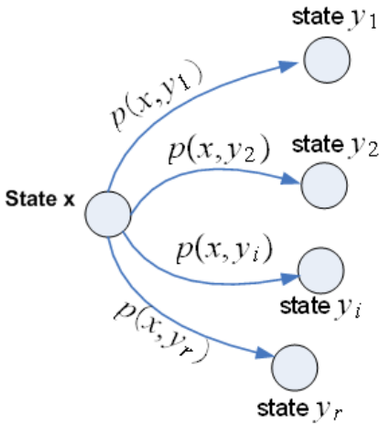
Since at most a vertex can have r edges, then we can find a vertex x with r edges connecting it to vertices $y_1, y_2, \cdots, y_r$ with corresponding one step probability transitions of $p(x, y_1), p(x, y_2), \cdots, p(x, y_r)$. (If we can't find such a vertex, the argument will apply to any other vertex, just replace r with the number of edges on that vertex and the argument will still apply).

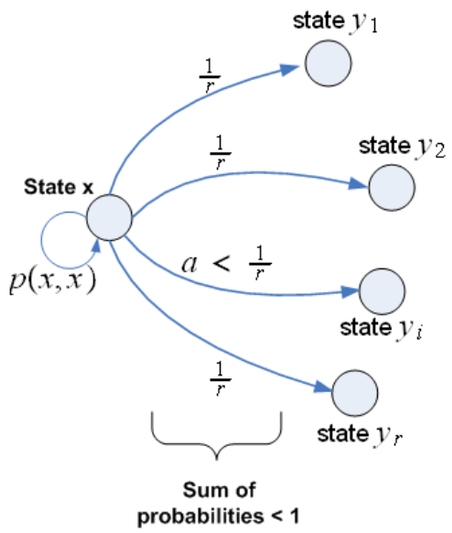
Now let us consider $f(x)$ and compare it to each of the $f(y_i)$ where the y_i is the vertex with direct edge from x. There are 2 cases to consider:

1. $f(x) >$ at least one of the $f(y_i)$, $i = 1 \cdots r$
2. $f(x) <$ all of $f(y_i)$, $i = 1 \cdots r$
3. $f(x) =$ all of $f(y_i)$, $i = 1 \cdots r$

Consider case (1): Since $f(x) > f(y_i)$ for some i, then for this specific y_i, $p(x, y_i) = \frac{1}{r} \min\left\{1, \frac{f(y)}{f(x)}\right\} = \frac{1}{r} k$ where $k < 1$, hence $p(x, y_i) = a$ where $a < \frac{1}{r}$. Lets assume there was only one y_i such that the above is true. I.e. at least one of the vertices connected to x had $f(y_i) < f(x)$ (if more if found, it will not change the argument). Now we add all the probabilities $p(x, y_i)$ and we found that this sum is $\overbrace{\frac{1}{r} + \frac{1}{r} + \cdots + \frac{1}{r}}^{(r-1)\ vertices} + a$ where the a is for that vertex which had $f(y_i) < f(x)$. Now since $a < \frac{1}{r}$ then this sum will be LESS THAN ONE. But the sum of the one step probability transition from each state must be 1, hence to compensate, we must then have $p(x, x)$ added to make up for the difference. Hence we showed that under case (1) we can find p_{ii} which is not zero. This diagram illustrate this case

[3] In addition, since we showed in part (i) that this chain is an irreducible chain, hence each state communicate with each other state, hence all states must be of the same type since all states are in the same communication class (Theorem 5.3.2). Then if one state is aperiodic, then the all states that communicate with it must also be aperiod (to be of the same type). Hence in an irreduible chain, if one state is aperiodic, then all states are aperiodic as well.





Now we consider case (2).

In this case since $f(x) < f(y_i)$ for each i, then $p(x,y_i) = \frac{1}{r}\min\left\{1, \frac{f(y_i)}{f(x)}\right\} = \frac{1}{r}.$, then the sum of the probabilities of transitions from x is $\overbrace{\frac{1}{r}+\frac{1}{r}+\cdots+\frac{1}{r}}^{r \text{ vertices}} = 1$and we do not need to compensate by adding $p(x,x)$ to make up for the deficit. However since now $f(y_i) > f(x)$ then if we view y_i as the x vertex and the x vertex as the y, and consider the probability transitions out of y_i, then we are back to case (1) above. Hence in case (2) as well ,we can find a state in which $p(x,x) > 0$, Hence the chain is aperiodic, and since it is irreducible, then it is regular in this case as well.

Now consider case (3):

In this case $f(x) = f(y_i)$ for $i = 1 \cdots r$. In other words, $f(x)$ is CONSTANT. In this case $p(x,y_i) = \frac{1}{r}\min\left\{1, \frac{f(y_i)}{f(x)}\right\} = \frac{1}{r}$,then the sum of the probabilities of transitions from x is $\overbrace{\frac{1}{r}+\frac{1}{r}+\cdots+\frac{1}{r}}^{r \text{ vertices}} = 1$and we do not need to compensate by adding $p(x,x)$ to make up for the deficit. This will be true for any node. Therefore, it is not possible to find at least one node with the probabilities attached to edges leaving it is less than one. Hence there are no state with $p(x,x) > 0$, hence in this case, the chain is not aperiodic, and hence the chain is NOT regular.

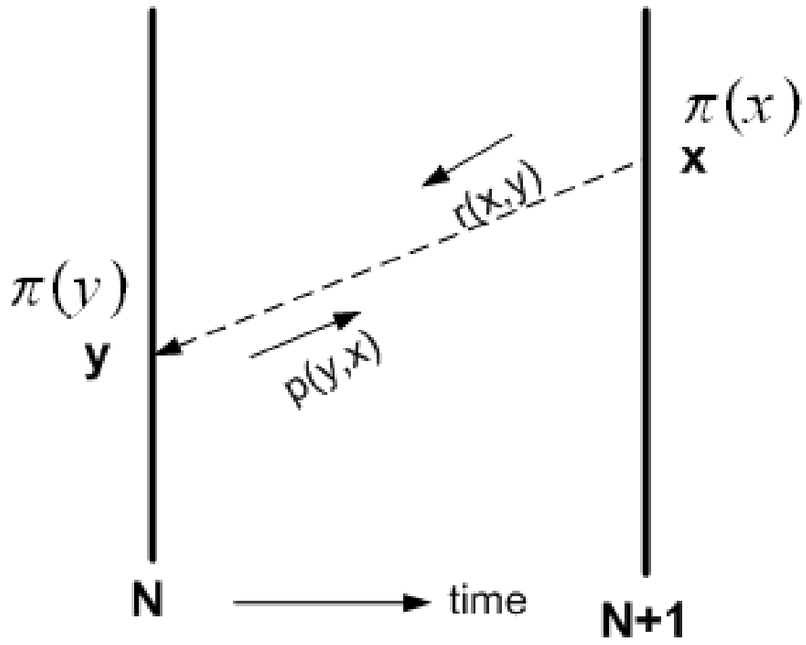
Conclusion: Condition for chain not to be regular is that $f(x)$ be constant.

Part(iii)

Since the chain is irreducible, then there is a reverse Markov chain (proof is on page 8.1 and 8.2 of lecture notes). Hence for an irreducible chain the balance equations hold

$$r(x,y) = \frac{\pi(y)\, p(y,x)}{\pi(x)} \tag{1}$$

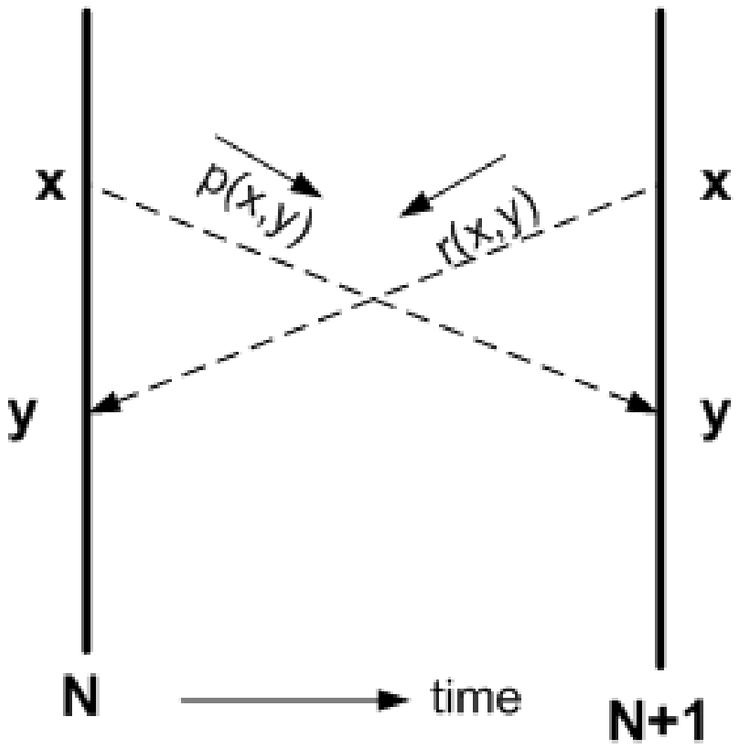
This diagram helps me remember these formulas



$$\pi(x)r(x,y) = \pi(y)p(y,x)$$

BALANCE EQUATION FOR AN IRREDUCIBLE CHAIN

Now if the chain the time reversible as well, then $r(x,y) = p(x,y)$,



$$r(x,y) = p(x,y)$$

Condition for a time reversible irreducible chain

Then the balance equation (1) becomes

$$\pi(x)\,p(x,y) = \pi(y)\,p(y,x) \tag{2}$$

Hence we need to show that the equation above holds to show the chain is time reversible.

Let the LHS of (2) be $\pi(x)\,p(x,y)$ and let RHS of (2) be $\pi(y)\,p(y,x)$. Then we will show that LHS=RHS for the following 3 cases:

1. $f(x) = f(y)$
2. $f(x) < f(y)$
3. $f(x) > f(y)$

Case(1): Since $f(x) = f(y)$ let these be some value, say z

$$\begin{aligned} LHS &= \pi(x)\,p(x,y) \\ &= \frac{f(x)}{\sum_{v\in V} f(v)}\left(\frac{1}{r}\min\left\{1, \frac{f(y)}{f(x)}\right\}\right) \\ &= \frac{z}{\sum_{v\in V} f(v)}\frac{1}{r} \end{aligned} \tag{3}$$

and

$$\begin{aligned} RHS &= \pi(y)\, p(y,x) \\ &= \frac{f(y)}{\sum_{v\in V} f(v)} \left(\frac{1}{r} \min\left\{1, \frac{f(x)}{f(y)}\right\} \right) \\ &= \frac{z}{\sum_{v\in V} f(v)} \frac{1}{r} \end{aligned} \tag{4}$$

We see that (3) is the same as (4), hence $\boxed{\text{LHS=RHS for case (1)}}$.

case(2): $f(x) < f(y)$

$$\begin{aligned} LHS &= \pi(x)\, p(x,y) \\ &= \frac{f(x)}{\sum_{v\in V} f(v)} \left(\frac{1}{r} \min\left\{1, \frac{f(y)}{f(x)}\right\} \right) \\ &= \frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{r} \end{aligned} \tag{5}$$

and

$$\begin{aligned} RHS &= \pi(y)\, p(y,x) \\ &= \frac{f(y)}{\sum_{v\in V} f(v)} \left(\frac{1}{r} \min\left\{1, \frac{f(x)}{f(y)}\right\} \right) \\ &= \frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{r} \frac{f(x)}{f(y)} \\ &= \frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{r} \end{aligned} \tag{6}$$

Hence we see that (5) is the same as (6). Hence $\boxed{\text{RHS=LHS for case(2)}}$.

case (3):$f(x) > f(y)$

$$\begin{aligned} LHS &= \pi(x)\, p(x,y) \\ &= \frac{f(x)}{\sum_{v\in V} f(v)} \left(\frac{1}{r} \min\left\{1, \frac{f(y)}{f(x)}\right\} \right) \\ &= \frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{r} \frac{f(y)}{f(x)} \\ &= \frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{r} \end{aligned} \tag{7}$$

and

$$\begin{aligned} RHS &= \pi(y)\, p(y,x) \\ &= \frac{f(y)}{\sum_{v\in V} f(v)} \left(\frac{1}{r} \min\left\{1, \frac{f(x)}{f(y)}\right\} \right) \\ &= \frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{r} \end{aligned} \tag{8}$$

We see that (7) is the same as (8), hence LHS=RHS for case (3) as well.

Hence we showed the balance equation for the time reversible condition is satisfied. QED.

4.11.2 Problem 8.5

8.5 Suppose $G = (V, E)$ is an undirected, connected graph. For each vertex $v \in V$, let $edge(v)$ denote the number of edges that are connected to v. Let f be a positive function defined on V, and let π denote the probability distribution

$$\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)} .$$

(a) Implement the Hastings-Metropolis method to find a regular Markov chain whose limiting distribution is π. Start with the initial irreducible chain defined by

$$q(x, y) = \frac{1}{edge(x)} , \quad \text{whenever } (x, y) \in E .$$

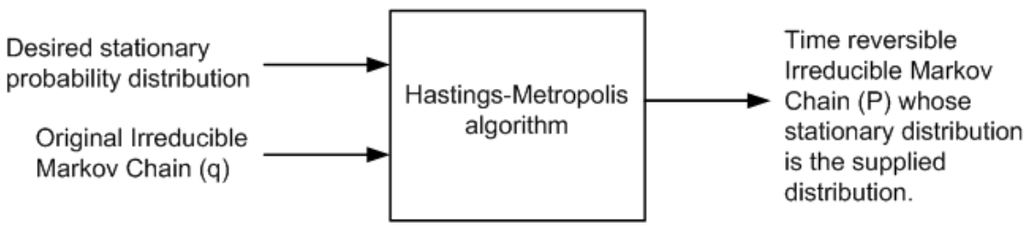
Note that the Markov chain with this one-step transition matrix is traversed by selecting at random one of the edges connected to x, and then making the transition to the corresponding node. $(a1)$ Show the Markov chain is irreducible. $(a2)$ Determine conditions under which the chain is regular. $(a3)$ Show the chain is time reversible with respect to π. (b) Write a MATLAB program that determines the one-step probability matrix resulting from this method. The input to this program is the function f, and the graph, represented by an adjacency matrix. An adjacency matrix is an $n \times n$ matrix, where n is the number of nodes in the graph, and where entry (i, j) is one if there is an edge connecting nodes i and j, and is zero otherwise. Use this adjacency matrix to compute the function $edge(v)$ at each node. Apply your program to the graph $G = (V, E)$, where $V = \{1, 2, 3, 4\}$, and $E = \{(1, 2), (1, 3), (2, 3), (2, 4), (3, 4)\}$,

and where $f(1) = 2$, $f(2) = 8$, $f(3) = 6$, and $f(4) = 4$. Verify (using MATLAB) that the resulting chain is regular and has the required limiting state probability distribution.

Part(a)

The following is the Hastings-Metrpolois algorithm implementation.

This algorithm generates a time-reversible M.C. (referred to as p in the lecture notes) given an irreducible M.C. (called q or the original chain) and given a stationary distribution π for that chain.



Input: $f(x)$ defined over the states x, and $edge(x)$ which represents the number of edges connected to x

1. For each state x calculate $\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)}$ and for each state x calculate $edge(x)$
2. compute $q(x, y) = \frac{1}{edge(x)}$ whenever $edge(x) \neq 0$ else set $q(x, y) = 0$
3. Select a state x by random to start from.
4. Let $n = 1$ and let $X_1 = x$
5. Let S be the set of all states that can be reached in one step from x. These will be the states y in which $q(x, y) \neq 0$

6. Select a state y from S by random (using a uniform $U[0,1]$ random number generator)
7. Calculate $\beta(x,y)=\min\left\{1,\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right\}$
8. Generate a random number u from $U[0,1]$
9. Let $n=n+1$
10. Compare u to $\beta(x,y)$.
11. IF $u<\beta(x,y)$ THEN $X_n=y$ (select the new state) ELSE $X_n=X_{n-1}$ (stay in same state) ENDIF
12. Let $x=X_n$
13. If $n>$ some Max number of iterations or if we reached some convergence limit Then go to 15
14. GOTO 5
15. Algorithm is complete. Now generate the time reversible MC as follows
    (a) Scan the state path generate X_n and count how many times state x switches to state y in one step
    (b) Do the above for all the states x
    (c) Divide the above number by the total number of steps made to generate $p(x,y)$

Since the problem now asks to *implement* Hastings-Metropolis, then I used the data given at the end of the problem and implemented the above simulation using that data[4]. Please see appendix for code and final P matrix generated.

Part (a1)

This is similar the problem 8.4 part(I). To show that the p (final M.C.) is irreducible, we need to show that there exist no closed proper subsets. Since the graph G is connected, then we just need to show whenever there is an edge between vertex x and y then there corresponds in the chain representation of the final p matrix a non-zero $p(x,y)$ and also a non-zero $p(y,x)$. This will insure that the each state can transition to each other state, just as each vertex can be visited from each other vertex (since it is a connected graph).

Let us consider any 2 vertices say x,y with a direct edge between them (this is the only case we need to consider due to the argument above). We need to show the resulting $p(x,y)$ and $p(y,x)$ are non-zero

Consider $p(x,y)$ first. Since

$$\begin{aligned}p(x,y)&=q(x,y)\,\beta(x,y)\\&=\frac{1}{edge\,(x)}\min\left\{1,\frac{\pi(y)\,q(y,x)}{\pi\,(x)\,q\,(x,y)}\right\}\\&=\frac{1}{edge\,(x)}\min\left\{1,\frac{\frac{f(y)}{\sum_{v\in V}f(v)}\frac{1}{edge(y)}}{\frac{f(x)}{\sum_{v\in V}f(v)}\frac{1}{edge(x)}}\right\}\end{aligned}$$

Hence

$$\boxed{p(x,y)=\tfrac{1}{edge(x)}\min\left\{1,\tfrac{f(y)edge(x)}{f(x)edge(y)}\right\}}\tag{1}$$

[4]I allready had the code for the simulation written, just needed to feed the new data for this problem.

Then it is clear that whenever there is an edge between x, y then $p(x,y) \neq 0$ since both $f(x)$ and $f(y)$ are positive (not zero) and also edge(x) and edge(y) are non-zero as well. Hence we see that $p(x,y) \neq 0$. Similar argument shows that $p(y,x) \neq 0$.

This shows that M.C. represented by P is irreducible.

Part (a2)

The condition for regular chain P is that there exist at least one state x such that $p(x,x) > 0$.From (1) above we can decide under what conditions this will occur.

Consider a vertex x with $edge(x)$ edges from it connected to vertices $y_1, y_2, \cdots, y_r$. Then from (1) we see that

$$\begin{aligned} p(x,y_i) &= \frac{1}{edge\,(x)} \min\left\{1, \frac{f(y_i)\,edge(x)}{f\,(x)\,edge\,(y_i)}\right\} \\ &= \frac{1}{edge\,(x)} \min\left\{1, \frac{\frac{f(y_i)}{edge(y_i)}}{\frac{f(x)}{edge(x)}}\right\} \end{aligned}$$

The condition for having $p(x,x) > 0$ is that $\min\left\{1, \frac{\frac{f(y_i)}{edge(y_i)}}{\frac{f(x)}{edge(x)}}\right\} < 1$, since this will cause $p(x,y_i)$ to be some quantity less than $\frac{1}{r}$ and so when summing over all r there will be a deficit in the sum and we have to compensate for it to make it 1 by adding $p(x,x)$. But for $\min\left\{1, \frac{\frac{f(y_i)}{edge(y_i)}}{\frac{f(x)}{edge(x)}}\right\}$ to be less than ONE means that $\frac{f(y_i)}{edge(y_i)} < \frac{f(x)}{edge(x)}$

Hence the condition for finding an Aperiodic state is finding a vertex x such that the above holds for one of the vertices y_i this vertex is directly connected to. For example, if y_i had the same number of edges from it as does x, then the condition will be that $f(y_i) < f(x)$. And if y_i has less or more edges from it than x has, then we need the ratio $\frac{f(y_i)}{edge(y_i)}$ to be less than $\frac{f(x)}{edge(x)}$.

The above is the same as saying $\frac{f(x)}{edge(x)}$ must be constant for the p not to be regular.

Part(A3)

Since the chain is irreducible, then there is a reverse Markov chain (proof is on page 8.1 and 8.2 of lecture notes). Hence for an irreducible chain the balance equations hold

$$r(x,y) = \frac{\pi(y)\,p(y,x)}{\pi\,(x)} \tag{2}$$

Now if the chain the time reversible as well, then $r(x,y) = p(x,y)$, Then the balance equation (1) becomes

$$\begin{aligned} \pi(x)\,p(x,y) &= \pi(y)\,p(y,x) \\ \frac{f(x)}{\sum_{v\in V} f\,(v)} q(x,y)\,\beta(x,y) &= \frac{f(y)}{\sum_{v\in V} f\,(v)} q(y,x)\,\beta(y,x) \\ \frac{f(x)}{\sum_{v\in V} f\,(v)} \frac{1}{edge\,(x)} \left(\min\left\{1, \frac{f(y)\,edge(x)}{f\,(x)\,edge\,(y)}\right\}\right) &= \frac{f(y)}{\sum_{v\in V} f\,(v)} \frac{1}{edge\,(y)} \left(\min\left\{1, \frac{f(x)\,edge(y)}{f\,(y)\,edge\,(x)}\right\}\right) \\ \frac{f(x)}{\sum_{v\in V} f\,(v)} \frac{1}{edge\,(x)} \left(\min\left\{1, \frac{\frac{f(y)}{edge(y)}}{\frac{f(x)}{edge(x)}}\right\}\right) &= \frac{f(y)}{\sum_{v\in V} f\,(v)} \frac{1}{edge\,(y)} \left(\min\left\{1, \frac{\frac{f(x)}{edge(x)}}{\frac{f(y)}{edge(y)}}\right\}\right) \end{aligned} \tag{3}$$

Hence we need to show that the equation (3) above holds to show the chain is time reversible.

There are 3 cases to consider:

1. $\frac{f(y)}{edge(y)} = \frac{f(x)}{edge(x)}$
2. $\frac{f(y)}{edge(y)} < \frac{f(x)}{edge(x)}$
3. $\frac{f(y)}{edge(y)} > \frac{f(x)}{edge(x)}$

For case (1), LHS of equation (3) simplifies to $\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}$ and the RHS of (3) simplifies to $\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}$, but since $\frac{f(y)}{edge(y)} = \frac{f(x)}{edge(x)}$, then LHS=RHS.

Hence balance equation (3) is satisfied for case (1).

For case(2), LHS of (3) simplifies $\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)} \left(\frac{\frac{f(y)}{edge(y)}}{\frac{f(x)}{edge(x)}} \right) = \frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}$ and RHS of (3) simplifies to $\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}$,then LHS=RHS.

Hence balance equation (3) is satisfied for case (2).

For case (3), LHS of (3) simplifies $\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}$ and RHS of (3) simplifies to $\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)} \left(\frac{\frac{f(x)}{edge(x)}}{\frac{f(y)}{edge(y)}} \right) = \frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}$,then LHS=RHS.

Hence balance equation (3) is satisfied for case (3).

Hence in all 3 cases we showed the balance equation is satisfied.

Hence M.C. is time reversible.

Part(b)

A small program written to construct the P matrix directly following instructions on page 8.4 of lecture notes. The following is the resulting P matrix

$$\begin{pmatrix} 0. & 0.5 & 0.5 & 0. \\ 0.0625 & 0.4375 & 0.25 & 0.25 \\ 0.0833333 & 0.333333 & 0.25 & 0.333333 \\ 0. & 0.5 & 0.5 & 0. \end{pmatrix}$$

Now to check that the final chain P is regular, it was raised to some high power to check that all entries in the $P^m > 0$. This is the result

```
In[17]:= MatrixPower[p, 50] // N // MatrixForm

Out[17]//MatrixForm=
```

$$\begin{pmatrix} 0.0526316 & 0.421053 & 0.315789 & 0.210526 \\ 0.0526316 & 0.421053 & 0.315789 & 0.210526 \\ 0.0526316 & 0.421053 & 0.315789 & 0.210526 \\ 0.0526316 & 0.421053 & 0.315789 & 0.210526 \end{pmatrix}$$

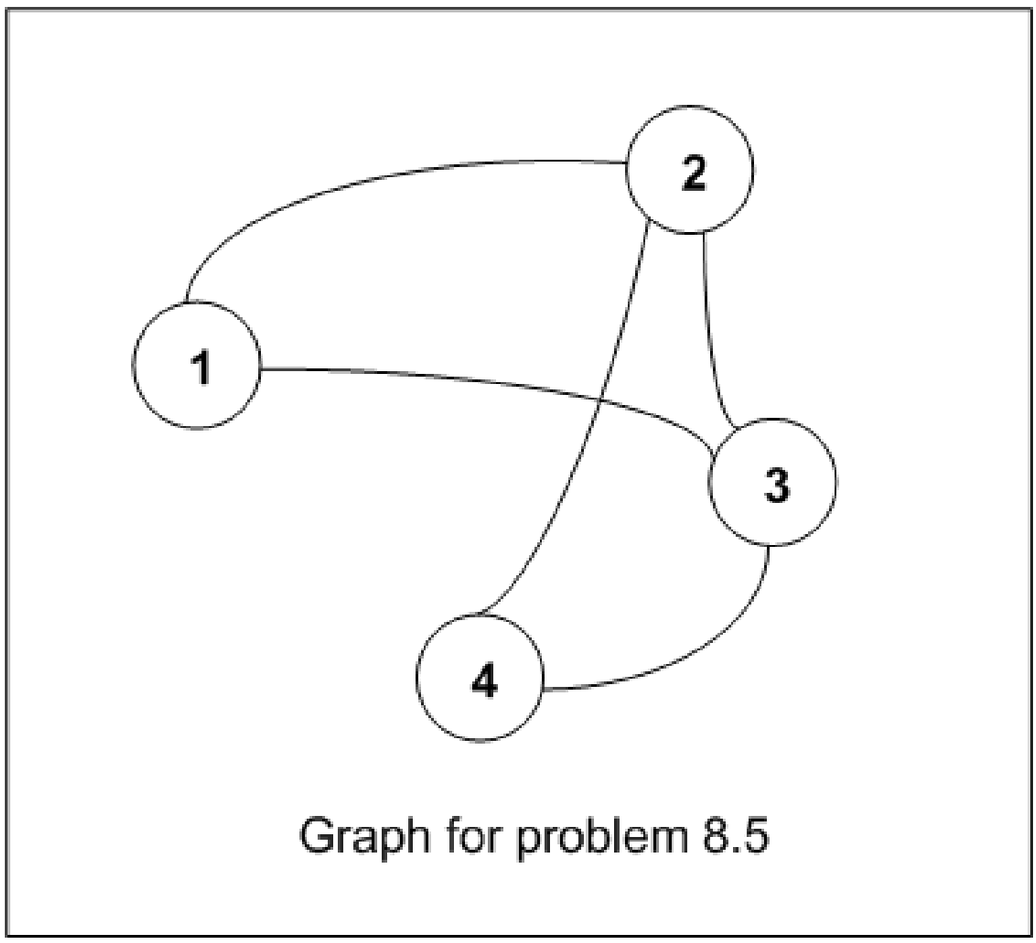
The above verifies that the final matrix p is regular.

Using the Hastings-Metropolis simulation algorithm, the convergence to the above matrix was slow. Had to make 2 million observation to be within 3 decimal points from the above. Here is the P matrix generated from Hastings algorithm for $N = 2,000,000$

$$\begin{pmatrix} 0. & 0.500114 & 0.499886 & 0. \\ 0.0625897 & 0.437179 & 0.249784 & 0.250448 \\ 0.0831875 & 0.333962 & 0.248524 & 0.334326 \\ 0. & 0.499297 & 0.500703 & 0. \end{pmatrix}$$

4.11.3 Appendix (Implementation of part(a) and part(b))

The graph for part(a) and part(b) is the following



4.11.4 code

Mathematica notebook

Mathematica notebook

4.11.5 Key solution

Key Solution For problem 8.5, 8.4

Chapter 8: Markov Chain Monte Carlo Methods - Solutions to Selected problems

8.4 Let $G = (V, E)$ be an undirected, connected graph with the property that each vertex is connected to at most r vertices. Let f be a positive function defined on V and let π denote the probability distribution

$$\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)} .$$

If $(x, y) \in E$, define the transition probability

$$p(x, y) = \frac{1}{r} min \left\{ 1, \frac{f(y)}{f(x)} \right\} .$$

with $p(x, y) = 0$ otherwise, except that $p(x, x)$ is determined so that the rows sum to one. (i) Show that the Markov chain determined by p is irreducible. (ii) Determine conditions under which the chain is regular. (iii) Show the chain is time reversible with respect to π.

Solution (i) To show that the chain is irreducible, note first that G is connected. In other words, in G there is a path from any one node to any other; that is, given any two nodes, say a and b in V, there is a sequence of nodes, say $x_1, x_2, \cdots, x_n$, in V such that $(a, x_1) \in E$, $(x_i, x_{i+1}) \in E$, for each $i = 1, 2, \cdots, n$, and $(x_n, b) \in E$. While the graph G is undirected, the graph of the Markov chain is directed. However, corresponding to each arc in G there are two arcs in the graph of the Markov chain, one in each direction, and each with nonzero probability. Indeed, if $(x, y) \in E$, then there is an arc in the graph of the Markov chain that points from x to y with associated probability $p(x, y) > 0$, determined by the formula above, and there is another arc that points from y to x with associated probability $p(y, x) > 0$, again determined by the formula above. It follows that in the graph of the Markov chain, between any two nodes (now states of the chain), there is a path between these states that can be traversed following the arcs in the required directions. In other words, any two states of the Markov chain communicate. Hence, the chain is irreducible.

(ii) Although the Markov chain is irreducible, it may be periodic, and hence not regular. As a simple example, consider the graph $G = (V, E)$ with vertex set $V = \{1, 2\}$ and edge set $E = \{(1, 2)\}$. Then $r = 1$. Suppose that f is the constant function. Then the associated Markov chain has one-step probability transition matrix

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} .$$

This chain is periodic with period 2. Suppose however that f is not constant. For example, let $f(1) = 1$ and $f(2) = 3$. Then the associated Markov chain has one-step probability transition matrix

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

This chain is aperiodic. However, more generally, whenever f is not the constant function, the associated Markov chain will be aperiodic, and hence regular. To see this result, note that there must be some vertex x such that $(x, y) \in E$, and $f(y) < f(x)$. For this state x, the sum of the off-diagonal elements will be less than one, since there are at most r nonzero off diagonal entries. Hence, for this row, $p(x, x) \neq 0$. Thus, state x is aperiodic, and since the chain is irreducible, all states are aperiodic, and so the chain is aperiodic.

As another condition which implies regularity, suppose that at least one node of the graph G is connected directly to fewer than r nodes. Then, whether f is the constant function or not, that node will become a state in the chain that is aperiodic. Indeed, in the one-step transition matrix, the row corresponding to this state will be such that the sum of the off-diagonal elements will be less than one, and hence the diagonal element will be nonzero. Thus, since the chain is irreducible, and one state is aperiodic, all states are aperiodic.

(*iii*) To show that the balance equations hold, we need to show that $\pi(x)p(x, y) = \pi(y)p(y, x)$ for each pair of states x and y. First, if $p(x, y) = 0$, then $p(y, x) = 0$ also, since $p(x, y) = 0$ only when there is no edge of the graph G that connects x and y. Next, when $(x, y) \in E$,

$$\pi(x)p(x, y) = \frac{f(x)}{rC} min\left\{1, \frac{f(y)}{f(x)}\right\} = \frac{1}{rC} min\left\{f(x), f(y)\right\} ,$$

where C is the sum appearing in the denominator of π. Similarly, we have

$$\pi(y)p(y, x) = \frac{f(y)}{rC} min\left\{1, \frac{f(x)}{f(y)}\right\} = \frac{1}{rC} min\left\{f(y), f(x)\right\} ,$$

These two expressions are the same, which is the desired conclusion.

8.5 Suppose $G = (V, E)$ is an undirected connected graph. For each vertex $v \in V$, let $edge(v)$ denote the number of edges that are connected to v. Let f be a positive function defined on V, and letπ denote the probability distribution

$$\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)} .$$

(*a*) Implement the Hastings-Metropolis method to find a regular Markov chain whose limiting distribution is π. Start with the initial irreducible chain defined by

$$q(x, y) = \frac{1}{edge(x)} , \quad \text{whenever} \quad (x, y) \in E .$$

Note that the Markov chain with this one-step transition matrix is traversed by selecting at random one of the edges connected to x, and then making the transition to the corresponding node. $(a1)$ Show that the Markov chain determined by this method is irreducible. $(a2)$ Determine conditions under which the chain is regular. $(a3)$ Show the chain is time reversible with respect to π. (b) Write a MATLAB program that determines the one-step probability matrix resulting from this method. The input to this program is the function f and the graph, represented by an adjacency matrix. An adjacency matrix is an $n \times n$ matrix, where n is the number of nodes in the graph, and where entry (i, j) is one if there is an edge connecting nodes i and j, and is zero otherwise. Use this adjacency matrix to compute the function $edge(v)$ at each node. Apply your program to the graph $G = (V, E)$ where $V = \{1, 2, 3, 4\}$, and $E = \{(1, 2), (1, 3), (2, 3), (2, 4), (3, 4)\}$, and where $f(1) = 2$, $f(2) = 8$, $f(3) = 6$, and $f(4) = 4$. Verify (using MATLAB) that the resulting chain is regular and has the required limiting state probability distribution.

Solution (a) The one-step transition probabilities are

$$p(x, y) = \frac{1}{edge(x)}\beta(x, y) \quad \text{for } (x, y) \in E\,, \quad \text{with} \quad p(x, x) = 1 - \sum_{y \neq x} p(x, y)\,,$$

where $\beta(x, y)$ is given by

$$\beta(x, y) = min\left\{1, \frac{f(y)edge(x)}{f(x)edge(y)}\right\}\,.$$

Otherwise $p(x, y) = 0$.

$(a1)$ These formulas show that for each arc $(x, y) \in E$, we have $p(x, y) > 0$ and $p(y, x) > 0$. Thus, between any two nodes that are connected by an arc in G, the resulting Markov chain has two corresponding states, x and y, and there are two arcs connecting these states which point in opposite directions. Hence, since the original graph is connected, it is therefore possible, in the Markov chain, to travel from any one state to any other. Thus, the chain is irreducible.

$(a2)$ For the setting of this problem, the Markov chain produced by the Hasting-Metropolis algorithm may be periodic, and hence not regular. For example, consider the graph $G = (V, E)$ with vertex set $V = \{1, 2\}$ and edge set $E = \{(1, 2)\}$. Suppose that f is the constant function. Then the resulting Markov chain has one-step probability transition matrix

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\,,$$

and the chain is periodic with period 2.

However, when $f(x)/edge(x)$ is not the constant function on V, the chain is aperiodic. To see this result, note first that since the graph is connected, there must be two vertices x and y such that $(x, y) \in E$, and $f(y)/edge(y) < f(x)/edge(x)$. For these states x and y we will have $\beta(x, y) < 1$. Therefore, in the one step transition matrix for the Markov chain, the sum of the off-diagonal elements in the row for state x is less than one. Hence, state x is aperiodic. Since the chain is irreducible, the chain is therefore also aperiodic. Thus, in this case when $f(x)/edge(x)$ is not the constant function on V, the chain is irreducible and aperiodic, and hence regular.

$(a3)$ To show that the balance equations hold, the same argument used for the previous problem carries over. we need to show that $\pi(x)p(x, y) = \pi(y)p(y, x)$ for each pair of states x and y. First, if $p(x, y) = 0$, then $p(y, x) = 0$ also, since $p(x, y) = 0$ only when there is no edge of the graph G that connects x and y. Next, when $(x, y) \in E$,

$$\pi(x)p(x, y) = \frac{f(x)}{edge(x)C} min\left\{1, \frac{f(y)edge(x)}{f(x)edge(y)}\right\} = \frac{1}{C} min\left\{\frac{f(x)}{edge(x)}, \frac{f(y)}{edge(y)}\right\} ,$$

where C is the sum appearing in the denominator of π. Similarly, we have

$$\pi(y)p(y, x) = \frac{f(y)}{edge(y)C} min\left\{1, \frac{f(x)edge(y)}{f(y)edge(x)}\right\} = \frac{1}{C} min\left\{\frac{f(y)}{edge(y)}, \frac{f(x)}{edge(x)}\right\} ,$$

These two expressions are the same, which is the desired conclusion.

4.11.6 my graded solution

HW problems 8.4 and 8.5, Mathematics 504

CSUF, spring 2008

by Nasser Abbasi

April 16, 2008

Contents

1 Problem 8.4

8.4 Let $G = (V, E)$ be an undirected, connected graph, with the property that each vertex is connected to at most r vertices. Let f be a positive function defined on V, and let π denote the probability distribution

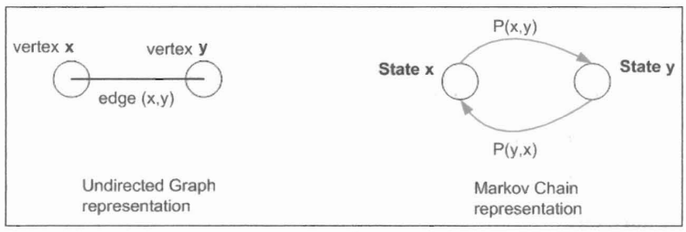
$$\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)} .$$

If $(x, y) \in E$, define the transition probability

$$p(x, y) = \frac{1}{r} min\left\{1, \frac{f(y)}{f(x)}\right\} ,$$

with $p(x, y) = 0$ otherwise, except that $p(x, x)$ is determined so that the rows sum to one. (i) Show that the Markov chain determined by p is irreducible. (ii) Determine conditions under which the chain is regular. (iii) Show the chain is time reversible with respect to π.

1.1 Part(i)



M.C. is irreducible if there exist no proper closed subset in the state space. Since we are given that the graph G is connected, then this means it is possible to visit each vertex from any other vertex in the graph. But does a connected graph implies no proper closed subset of the corresponding M.C.? The answer is YES. If we view each vertex as state, we just need to show that for each edge in G between 2 vertices x, y, there corresponds a probability of transition from state x to y which is not zero, and also a probability of transition from state y to x which is also not zero. By showing this, we conclude that the M.C. will switch (in some number of steps) to any state from any other state, which implies there is no closed subset, hence P is irreducible.

But from the definition of $p(x, y)$ we see that if there is an edge (x, y) then $p(x, y)$ exist and is not zero, and $p(y, x)$ exist and is not zero (since r is finite). This completes the proof.

1.2 Part(ii)

A finite M.C. is regular when, for some integer m, P^m contains only positive elements.

This implies that the one step transition matrix P must have at least one entry along the diagonal P_{ii} that is none-zero (If all elements along the diagonal are zero, then P^m will always contain at least one zero element no matter how large m is). But a diagonal element not being zero is the same as saying that at least one state must be aperiodic (if $P_{ii} > 0$ then the period is one).

Hence the condition for the M.C. to be regular is that at least one state must be aperiodic [1].

To proof that the above chain is regular, we then need to show that at least one state is aperiodic.

This is the proof:

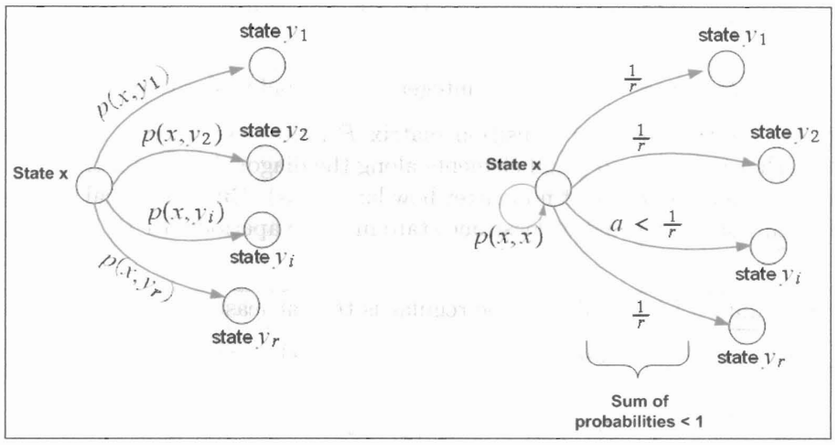
Since at most a vertex can have r edges, then we can find a vertex x with r edges connecting it to vertices $y_1, y_2, \cdots, y_r$ with corresponding one step probability transitions of $p(x, y_1), p(x, y_2), \cdots, p(x, y_r)$. (If we can't find such a vertex, the argument will apply to any other vertex, just replace r with the number of edges on that vertex and the argument will still apply).

Now let us consider $f(x)$ and compare it to each of the $f(y_i)$ where the y_i is the vertex with direct edge from x. There are 2 cases to consider:

1. $f(x) >$ at least one of the $f(y_i)$, $i = 1 \cdots r$
2. $f(x) <$ all of $f(y_i)$, $i = 1 \cdots r$
3. $f(x) =$ all of $f(y_i)$, $i = 1 \cdots r$

Consider case (1): Since $f(x) > f(y_i)$ for some i, then for this specific y_i, $p(x, y_i) = \frac{1}{r} \min\left\{1, \frac{f(y)}{f(x)}\right\} = \frac{1}{r}k$ where $k < 1$, hence $p(x, y_i) = a$ where $a < \frac{1}{r}$. Lets assume there was only one y_i such that the above is true. I.e. at least one of the vertices connected to x had $f(y_i) < f(x)$ (if more if found, it will not change the argument). Now we add all the probabilities $p(x, y_i)$ and we found that this sum is $\overbrace{\frac{1}{r} + \frac{1}{r} + \cdots + \frac{1}{r}}^{(r-1)\ vertices} + a$ where the a is for that vertex which had $f(y_i) < f(x)$. Now since $a < \frac{1}{r}$ then this sum will be LESS THAN ONE. But the sum of the one step probability transition from each state must be 1, hence to compensate, we must then have $p(x, x)$ added to make up for the difference. Hence we showed that under case (1) we can find p_{ii} which is not zero. This diagram illustrate this case

[1] In addition, since we showed in part (i) that this chain is an irreducible chain, hence each state communicate with each other state, hence all states must be of the same type since all states are in the same communication class (Theorem 5.3.2). Then if one state is aperiodic, then the all states that communicate with it must also be aperiod (to be of the same type). Hence in an irreduible chain, if one state is aperiodic, then all states are aperiodic as well.



Now we consider case (2). In this case since $f(x) < f(y_i)$ for each i, then $p(x,y_i) = \frac{1}{r}\min\left\{1, \frac{f(y_i)}{f(x)}\right\} = \frac{1}{r}$., then the sum of the probabilities of transitions from x is $\overbrace{\frac{1}{r}+\frac{1}{r}+\cdots+\frac{1}{r}}^{r \text{ vertices}} =$ 1and we do not need to compensate by adding $p(x,x)$ to make up for the deficit. However since now $f(y_i) > f(x)$ then if we view y_i as the x vertex and the x vertex as the y, and consider the probability transitions out of y_i, then we are back to case (1) above. Hence in case (2) as well ,we can find a state in which $p(x,x) > 0$, Hence the chain is aperiodic, and since it is irreducible, then it is regular in this case as well.

Now consider case (3): In this case $f(x) = f(y_i)$ for $i = 1 \cdots r$. In other words, $f(x)$ is CONSTANT. In this case $p(x,y_i) = \frac{1}{r}\min\left\{1, \frac{f(y_i)}{f(x)}\right\} = \frac{1}{r}$,then the sum of the probabilities of transitions from x is $\overbrace{\frac{1}{r}+\frac{1}{r}+\cdots+\frac{1}{r}}^{r \text{ vertices}} =$ 1and we do not need to compensate by adding $p(x,x)$ to make up for the deficit. This will be true for any node. Therefore, it is not possible to find at least one node with the probabilities attached to edges leaving it is less than one. Hence there are no state with $p(x,x) > 0$, hence in this case, the chain is not aperiodic, and hence the chain is NOT regular.

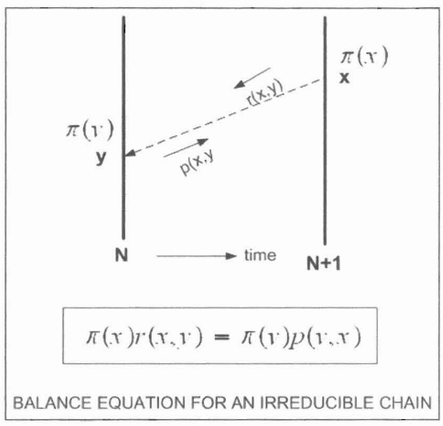
Conclusion: Condition for chain not to be regular is that $f(x)$ be constant.

1.3 Part(iii)

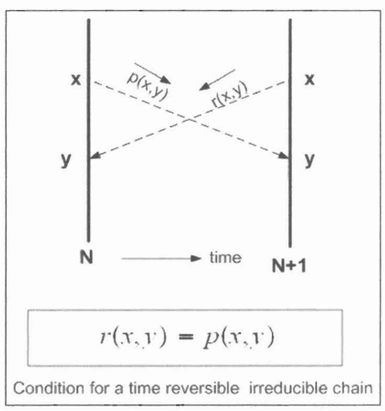
Since the chain is irreducible, then there is a reverse Markov chain (proof is on page 8.1 and 8.2 of lecture notes). Hence for an irreducible chain the balance equations hold

$$r(x,y) = \frac{\pi(y)\,p(y,x)}{\pi(x)} \tag{1}$$

This diagram helps me remember these formulas



Now if the chain the time reversible as well, then $r(x,y) = p(x,y)$,



Then the balance equation (1) becomes

$$\pi(x)\, p(x,y) = \pi(y)\, p(y,x) \tag{2}$$

Hence we need to show that the equation above holds to show the chain is time reversible.

Let the LHS of (2) be $\pi(x)\, p(x,y)$ and let RHS of (2) be $\pi(y)\, p(y,x)$. Then we will show that LHS=RHS for the following 3 cases:

1. $f(x) = f(y)$
2. $f(x) < f(y)$
3. $f(x) > f(y)$

Case(1): Since $f(x) = f(y)$ let these be some value, say z

$$\begin{aligned} LHS &= \pi(x)\, p(x,y) \\ &= \frac{f(x)}{\sum_{v \in V} f(v)} \left(\frac{1}{r} \min\left\{1, \frac{f(y)}{f(x)}\right\}\right) \\ &= \frac{z}{\sum_{v \in V} f(v)} \frac{1}{r} \end{aligned} \tag{3}$$

and

$$\begin{aligned} RHS &= \pi(y)\, p(y,x) \\ &= \frac{f(y)}{\sum_{v \in V} f(v)} \left(\frac{1}{r} \min\left\{1, \frac{f(x)}{f(y)}\right\}\right) \\ &= \frac{z}{\sum_{v \in V} f(v)} \frac{1}{r} \end{aligned} \tag{4}$$

We see that (3) is the same as (4), hence LHS=RHS for case (1).

case(2): $f(x) < f(y)$

$$\begin{aligned} LHS &= \pi(x)\, p(x,y) \\ &= \frac{f(x)}{\sum_{v \in V} f(v)} \left(\frac{1}{r} \min\left\{1, \frac{f(y)}{f(x)}\right\}\right) \\ &= \frac{f(x)}{\sum_{v \in V} f(v)} \frac{1}{r} \end{aligned} \tag{5}$$

and

$$\begin{aligned} RHS &= \pi(y)\, p(y,x) \\ &= \frac{f(y)}{\sum_{v \in V} f(v)} \left(\frac{1}{r} \min\left\{1, \frac{f(x)}{f(y)}\right\}\right) \\ &= \frac{f(y)}{\sum_{v \in V} f(v)} \frac{1}{r} \frac{f(x)}{f(y)} \\ &= \frac{f(x)}{\sum_{v \in V} f(v)} \frac{1}{r} \end{aligned} \tag{6}$$

Hence we see that (5) is the same as (6). Hence RHS=LHS for case(2).

case (3):$f(x) > f(y)$

$$\begin{aligned} LHS &= \pi(x)\, p(x,y) \\ &= \frac{f(x)}{\sum_{v\in V} f(v)} \left(\frac{1}{r} \min\left\{1, \frac{f(y)}{f(x)}\right\} \right) \\ &= \frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{r} \frac{f(y)}{f(x)} \\ &= \frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{r} \end{aligned} \tag{7}$$

and

$$\begin{aligned} RHS &= \pi(y)\, p(y,x) \\ &= \frac{f(y)}{\sum_{v\in V} f(v)} \left(\frac{1}{r} \min\left\{1, \frac{f(x)}{f(y)}\right\} \right) \\ &= \frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{r} \end{aligned} \tag{8}$$

We see that (7) is the same as (8), hence LHS=RHS for case (3) as well.

Hence we showed the balance equation for the time reversible condition is satisfied. QED.

2 Problem 8.5

8.5 Suppose $G = (V, E)$ is an undirected, connected graph. For each vertex $v \in V$, let $edge(v)$ denote the number of edges that are connected to v. Let f be a positive function defined on V, and let π denote the probability distribution

$$\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)} .$$

(a) Implement the Hastings-Metropolis method to find a regular Markov chain whose limiting distribution is π. Start with the initial irreducible chain defined by

$$q(x, y) = \frac{1}{edge(x)} , \quad \text{whenever } (x, y) \in E .$$

Note that the Markov chain with this one-step transition matrix is traversed by selecting at random one of the edges connected to x, and then making the transition to the corresponding node. $(a1)$ Show the Markov chain is irreducible. $(a2)$ Determine conditions under which the chain is regular. $(a3)$ Show the chain is time reversible with respect to π. (b) Write a MATLAB program that determines the one-step probability matrix resulting from this method. The input to this program is the function f, and the graph, represented by an adjacency matrix. An adjacency matrix is an $n \times n$ matrix, where n is the number of nodes in the graph, and where entry (i, j) is one if there is an edge connecting nodes i and j, and is zero otherwise. Use this adjacency matrix to compute the function $edge(v)$ at each node. Apply your program to the graph $G = (V, E)$, where $V = \{1, 2, 3, 4\}$, and $E = \{(1, 2), (1, 3), (2, 3), (2, 4), (3, 4)\}$,

and where $f(1) = 2$, $f(2) = 8$, $f(3) = 6$, and $f(4) = 4$. Verify (using MATLAB) that the resulting chain is regular and has the required limiting state probability distribution.

2.1 Part(a)

The following is the Hastings-Metrpolois algorithm implementation.

This algorithm generates a time-reversible M.C. (referred to as p in the lecture notes) given an irreducible M.C. (called q or the original chain) and given a stationary distribution π for that chain.

Input: $f(x)$ defined over the states x, and $edge(x)$ which represents the number of edges connected to x

1. For each state x calculate $\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)}$ and for each state x calculate $edge(x)$
2. compute $q(x,y) = \frac{1}{edge(x)}$ whenever $edge(x) \neq 0$ else set $q(x,y) = 0$
3. Select a state x by random to start from.
4. Let $n = 1$ and let $X_1 = x$
5. Let S be the set of all states that can be reached in one step from x. These will be the states y in which $q(x,y) \neq 0$
6. Select a state y from S by random (using a uniform $U[0,1]$ random number generator)
7. Calculate $\beta(x,y) = \min\left\{1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right\}$
8. Generate a random number u from $U[0,1]$
9. Let $n = n + 1$
10. Compare u to $\beta(x,y)$.
11. IF $u < \beta(x,y)$ THEN $X_n = y$ (select the new state) ELSE $X_n = X_{n-1}$ (stay in same state) ENDIF
12. Let $x = X_n$
13. If $n >$ some Max number of iterations or if we reached some convergence limit Then go to 15
14. GOTO 5
15. Algorithm is complete. Now generate the time reversible MC as follows
    (a) Scan the state path generate X_n and count how many times state x switches to state y in one step
    (b) Do the above for all the states x
    (c) Divide the above number by the total number of steps made to generate $p(x,y)$

Since the problem now asks to *implement* Hastings-Metropolis, then I used the data given at the end of the problem and implemented the above simulation using that data[2]. Please see appendix for code and final P matrix generated.

[2]I allready had the code for the simulation written, just needed to feed the new data for this problem.

2.1.1 Part (a1)

This is similar the problem 8.4 part(I). To show that the p (final M.C.) is irreducible, we need to show that there exist no closed proper subsets. Since the graph G is connected, then we just need to show whenever there is an edge between vertex x and y then there corresponds in the chain representation of the final p matrix a non-zero $p(x,y)$ and also a non-zero $p(y,x)$. This will insure that the each state can transition to each other state, just as each vertex can be visited from each other vertex (since it is a connected graph).

Let us consider any 2 vertices say x, y with a direct edge between them (this is the only case we need to consider due to the argument above). We need to show the resulting $p(x,y)$ and $p(y,x)$ are non-zero

Consider $p(x,y)$ first. Since

$$\begin{aligned} p(x,y) &= q(x,y)\,\beta(x,y) \\ &= \frac{1}{edge(x)} \min\left\{1, \frac{\pi(y)\,q(y,x)}{\pi(x)\,q(x,y)}\right\} \\ &= \frac{1}{edge(x)} \min\left\{1, \frac{\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}}{\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}}\right\} \end{aligned}$$

This is what is asked for in part (a1)..

Hence

$$\boxed{p(x,y) = \frac{1}{edge(x)} \min\left\{1, \frac{f(y)edge(x)}{f(x)edge(y)}\right\}} \tag{1}$$

Then it is clear that whenever there is an edge between x, y then $p(x,y) \neq 0$ since both $f(x)$ and $f(y)$ are positive (not zero) and also edge(x) and edge(y) are non-zero as well. Hence we see that $p(x,y) \neq 0$. Similar argument shows that $p(y,x) \neq 0$.

This shows that M.C. represented by P is irreducible.

2.1.2 Part (a2)

The condition for regular chain P is that there exist at least one state x such that $p(x,x) > 0$.From (1) above we can decide under what conditions this will occur.

Consider a vertex x with $edge(x)$ edges from it connected to vertices $y_1, y_2, \cdots, y_r$. Then from (1) we see that

$$\begin{aligned} p(x,y_i) &= \frac{1}{edge(x)} \min\left\{1, \frac{f(y_i)\,edge(x)}{f(x)\,edge(y_i)}\right\} \\ &= \frac{1}{edge(x)} \min\left\{1, \frac{\frac{f(y_i)}{edge(y_i)}}{\frac{f(x)}{edge(x)}}\right\} \end{aligned}$$

The condition for having $p(x,x) > 0$ is that $\min\left\{1, \frac{\frac{f(y_i)}{edge(y_i)}}{\frac{f(x)}{edge(x)}}\right\} < 1$, since this will cause $p(x,y_i)$ to be some quantity less than $\frac{1}{r}$ and so when summing over all r there will be a deficit in the sum and we have to compensate for it to make it 1 by adding $p(x,x)$. But for $\min\left\{1, \frac{\frac{f(y_i)}{edge(y_i)}}{\frac{f(x)}{edge(x)}}\right\}$ to be less than ONE means that $\boxed{\frac{f(y_i)}{edge(y_i)} < \frac{f(x)}{edge(x)}}$

Hence the condition for finding an Aperiodic state is finding a vertex x such that the above holds for one of the vertices y_i this vertex is directly connected to. For example, if y_i had the same number of edges from it as does x, then the condition will be that $f(y_i) < f(x)$. And if y_i has less or more edges from it than x has, then we need the ratio $\frac{f(y_i)}{edge(y_i)}$ to be less than $\frac{f(x)}{edge(x)}$.

The above is the same as saying $\frac{f(x)}{edge(x)}$ must be constant for the p not to be regular.

2.1.3 Part(A3)

Since the chain is irreducible, then there is a reverse Markov chain (proof is on page 8.1 and 8.2 of lecture notes). Hence for an irreducible chain the balance equations hold

$$r(x,y) = \frac{\pi(y)\,p(y,x)}{\pi(x)} \tag{2}$$

Now if the chain the time reversible as well, then $r(x,y) = p(x,y)$, Then the balance equation (1) becomes

$$\begin{aligned}
\pi(x)\,p(x,y) &= \pi(y)\,p(y,x)\\
\frac{f(x)}{\sum_{v\in V} f(v)} q(x,y)\,\beta(x,y) &= \frac{f(y)}{\sum_{v\in V} f(v)} q(y,x)\,\beta(y,x)\\
\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)} \left(\min\left\{1, \frac{f(y)\,edge(x)}{f(x)\,edge(y)}\right\}\right) &= \frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)} \left(\min\left\{1, \frac{f(x)\,edge(y)}{f(y)\,edge(x)}\right\}\right)\\
\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)} \left(\min\left\{1, \frac{\frac{f(y)}{edge(y)}}{\frac{f(x)}{edge(x)}}\right\}\right) &= \frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)} \left(\min\left\{1, \frac{\frac{f(x)}{edge(x)}}{\frac{f(y)}{edge(y)}}\right\}\right)
\end{aligned} \tag{3}$$

Hence we need to show that the equation (3) above holds to show the chain is time reversible.

There are 3 cases to consider:

1. $\frac{f(y)}{edge(y)} = \frac{f(x)}{edge(x)}$
2. $\frac{f(y)}{edge(y)} < \frac{f(x)}{edge(x)}$

3. $\frac{f(y)}{edge(y)} > \frac{f(x)}{edge(x)}$

For case (1), LHS of equation (3) simplifies to $\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}$ and the RHS of (3) simplifies to $\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}$, but since $\frac{f(y)}{edge(y)} = \frac{f(x)}{edge(x)}$, then LHS=RHS.

Hence balance equation (3) is satisfied for case (1).

For case(2), LHS of (3) simplifies $\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)} \left(\frac{\frac{f(y)}{edge(y)}}{\frac{f(x)}{edge(x)}} \right) = \frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}$ and RHS of (3) simplifies to $\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}$,then LHS=RHS.

Hence balance equation (3) is satisfied for case (2).

For case (3), LHS of (3) simplifies $\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}$ and RHS of (3) simplifies to $\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)} \left(\frac{\frac{f(x)}{edge(x)}}{\frac{f(y)}{edge(y)}} \right) = \frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}$,then LHS=RHS.

Hence balance equation (3) is satisfied for case (3).

Hence in all 3 cases we showed the balance equation is satisfied.

Hence M.C. is time reversible.

2.2 Part(b)

A small program written to construct the P matrix directly following instructions on page 8.4 of lecture notes. The following is the resulting P matrix

$$\begin{pmatrix} 0. & 0.5 & 0.5 & 0. \\ 0.0625 & 0.4375 & 0.25 & 0.25 \\ 0.0833333 & 0.333333 & 0.25 & 0.333333 \\ 0. & 0.5 & 0.5 & 0. \end{pmatrix}$$

Now to check that the final chain P is regular, it was raised to some high power to check that all entries in the $P^m > 0$. This is the result

```
In[17]:= MatrixPower[p, 50] // N // MatrixForm

Out[17]//MatrixForm=
0.0526316 0.421053 0.315789 0.210526
0.0526316 0.421053 0.315789 0.210526
0.0526316 0.421053 0.315789 0.210526
0.0526316 0.421053 0.315789 0.210526
```

The above verifies that the final matrix p is regular.

Using the Hastings-Metropolis simulation algorithm, the convergence to the above matrix was slow. Had to make 2 million observation to be within 3 decimal points from the above. Here is the P matrix generated from Hastings algorithm for $N = 2,000,000$

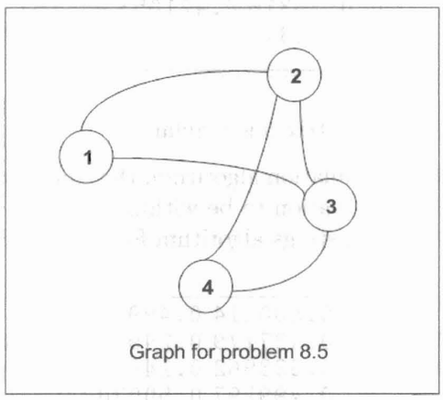
```
0.          0.500114 0.499886 0.
0.0625897   0.437179 0.249784 0.250448
0.0831875   0.333962 0.248524 0.334326
0.          0.499297 0.500703 0.
```

each row should be π:

$$\pi = \frac{1}{20}(2, 8, 6, 4)$$

3 Appendix (Implementation of part(a) and part(b))

The graph for part(a) and part(b) is the following



Graph for problem 8.5

Hastings - Metropolis Algorithm implementation For Problem 8.5 part(a)

This below is an implementation of the Hastings - Metropolis algorithm. A simple GUI interface allows the user to specify the number of steps to run the algorithm for. At each step, the current P matrix and the current calculated stationary distribution for this P matrix are shown to help observe the convergence.

The input to this run below is that of problem 8.5 part(b)

Few seed the random number generator and display the q and the π distribution used

```
SeedRandom[121 212];
```

Define the data given in the problem

```
v = {1, 2, 3, 4};
edges = {{1, 2}, {1, 3}, {2, 3}, {2, 4}, {3, 4}};
f = {1, 8, 6, 4};
```

Define the functions π(x) and q(x,y) to use in the implementation

```
pi[x_, f_] := f[[x]] / Sum_{i=1}^{Length[f]} f[[i]]

q[x_, y_] := Module[{r},
  r = Count[edges, {x, any_}] + Count[edges, {any_, x}];
  If[(Count[edges, {x, y}] > 0 || Count[edges, {y, x}] > 0), 1 / r, 0]
 ]
```

Find the stationary distribution

```
w = Table[pi[x, f], {x, 1, v[[-1]]}]
```

$$\left\{\frac{1}{19}, \frac{8}{19}, \frac{6}{19}, \frac{4}{19}\right\}$$

Generate the original q matrix

```
(originalMatrix = Table[q[x, y], {x, 1, v[[-1]]}, {y, 1, v[[-1]]}]) // MatrixForm
```

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

Bring up the user interface

```
m = Manipulate[First@{x = hastings[maxN]; Grid[{{"stationary distribution w=", N[w]},
      { "current stationary distribution=", MatrixPower[N[x], 100][[1, All]]},
      {, }, {"Current P Matrix=", N[MatrixForm[x]]}}, Alignment → Left]},
  {{maxN, 1, "number of steps"}, 1, 2 000 000, 100, ContinuousAction → False,
   Appearance → "Labeled"},
  AutorunSequencing → {{1, 300}}
 ]
```

number of steps 273401

stationary distribution w= {0.0526316, 0.421053, 0.315789, 0.210526}

current stationary distribution= {0.053453, 0.420567, 0.316001, 0.209979}

Current P Matrix=

$$\begin{pmatrix} 0. & 0.502771 & 0.497229 & 0. \\ 0.0637317 & 0.435999 & 0.251526 & 0.248743 \\ 0.0843336 & 0.331767 & 0.250466 & 0.333434 \\ 0. & 0.502369 & 0.497631 & 0. \end{pmatrix}$$

Define a function for cummulative sum

```
cumSum[list_] := Module[{i, sum, s, k},
  sum = 0;
  k = Length[list];
  s = Table[0, {k}];
  For[i = 1, i ≤ k, i++,
   {
    sum = sum + list[[i]];
    s[[i]] = sum;
   }
  ];
  s
 ]
```

Function to calculate β (x, y)

```
beta[x_, y_, pi_, q_] := Module[{},
  Min[1, (pi[[y]] q[[y, x]]) / (pi[[x]] q[[x, y]])]
 ]
```

Function called at the end of the run to generate P from the path of states travelled

```
generatePMatrixFromStatePath[nStates_, x_] := Module[{i, j, p, allPairs, n, m},
  n = Length[x];
  (*Print["X=",x];*)
  p = Table[0, {nStates}, {nStates}];
  allPairs = Partition[x, 2, 1];
  For[i = 1, i ≤ nStates, i++,
   {
    m = Count[allPairs, {i, y_}];
    For[j = 1, j ≤ nStates, j++,
     If[m ≠ 0, p[[i, j]] = Count[allPairs, {i, j}] / m, p[[i, j]] = 0]
    ]
   }
  ];

  p
 ]
```

Function to sample from q using uniform distribution

```
sampleFromQConditional[q_, x_] := Module[{s, found, j, k, sample, y},
  s = Flatten[Position[q[[x, All]], Except[0], 1, Heads → False] ];
  sample = q[[x, s]];
  sample = cumSum[sample];
  y = RandomReal[];
  found = False;
  For[j = 1, j ≤ Length[sample], j++,
   If[ Not[found], If[y ≤ sample[[j]], {k = j; found = True}]]
  ];

  y = s[[k]]
 ]
```

The Hastings algorithm main loop

```
hastings[maxN_] := Module[{i, j, nStates, n, s, y, α, u, x, sample, pts, sum, k, found},
  nStates = Length[originalMatrix];
  n = 1;
  x = Table[0, {maxN}];
  x[[n]] = 1;  (*pick any state to start from*)
  i = 1;
  While[i < maxN,
   {
    y = sampleFromQConditional[originalMatrix, x[[n]]];
    α = beta[x[[n]], y, w, originalMatrix];
    u = RandomReal[];
    n++;
    If[u ≤ α, x[[n]] = y, x[[n]] = x[[n - 1]]]; (*acceptance step*)
    i++;
   }
  ];

  generatePMatrixFromStatePath[nStates, x]
 ]
```

Problem 8.5 part(b)

by Naser Abbasi. Mathematics 504, Spring 2008. CSUF

This below construct the P matrix using direct computation of the Hasting-Meropolis method

Define the data given in the problem

```
v = {1, 2, 3, 4};
edges = {{1, 2}, {1, 3}, {2, 3}, {2, 4}, {3, 4}};
f = {1, 8, 6, 4};
```

Define the functions π(x) to use in the implementation

```
pi[x_, f_] := f[[x]] / Sum[f[[i]], {i, 1, Length[f]}]
```

Define the functions q(x,y) to use in the implementation. This does something similar to the adjcancy matrix normally used. I used the Count[] function in *Mathematica* which automatically counts the edges from the edges list above, so there is really no need to construct an adjancy matrix as such.

```
q[x_, y_] := Module[{r},
  r = Count[edges, {x, any_}] + Count[edges, {any_, x}];
  If[(Count[edges, {x, y}] > 0 || Count[edges, {y, x}] > 0), 1 / r, 0]
 ]
```

Find the stationary distribution

```
w = Table[pi[x, f], {x, 1, v[[-1]]}]
```

$$\left\{\frac{1}{19}, \frac{8}{19}, \frac{6}{19}, \frac{4}{19}\right\}$$

Generate the original q matrix

```
(originalMatrix = Table[q[x, y], {x, 1, v[[-1]]}, {y, 1, v[[-1]]}]) // MatrixForm
```

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

define Beta function

```
beta[x_, y_, pi_, q_] := Module[{},
  Min[1, pi[[y]] q[[y, x]] / pi[[x]] q[[x, y]]]
 ]
```

define a function to calculate a non - diagonal entry in the P matrix

```
makeEntryInPMatrix[x_, y_, pi_, q_] := Module[{},
  If[q[[x, y]] == 0, 0, q[[x, y]] beta[x, y, pi, q]]
 ]
```

Construct the P matrix for the off-diagonal elements only

```
nStates = Length[originalMatrix];
p = Table[0, {nStates}, {nStates}];

For[i = 1, i ≤ nStates, i++,
  For[j = 1, j ≤ nStates , j++,
   If[i ≠ j, p[[i, j]] = makeEntryInPMatrix[i, j, w, originalMatrix]]
  ]
 ];
```

Now calculate the diagonal elements of the P matrix

```
For[i = 1, i ≤ nStates, i++,
  For[j = 1, j ≤ nStates , j++,
   If[i == j, p[[i, j]] = 1 - Total[p[[i, All]]]]
  ]
 ];
```

Print the P matrix

```
p // N // MatrixForm
```

$$\begin{pmatrix} 0. & 0.5 & 0.5 & 0. \\ 0.0625 & 0.4375 & 0.25 & 0.25 \\ 0.0833333 & 0.333333 & 0.25 & 0.333333 \\ 0. & 0.5 & 0.5 & 0. \end{pmatrix}$$

Raise the final p matrix to some large power to verify it is regular

```
MatrixPower[p, 50] // N // MatrixForm
```

$$\begin{pmatrix} 0.0526316 & 0.421053 & 0.315789 & 0.210526 \\ 0.0526316 & 0.421053 & 0.315789 & 0.210526 \\ 0.0526316 & 0.421053 & 0.315789 & 0.210526 \\ 0.0526316 & 0.421053 & 0.315789 & 0.210526 \end{pmatrix}$$

HW problems 8.4 and 8.5, Mathematics 504

CSUF, spring 2008

by Nasser Abbasi

May 8, 2008

Contents

1 Problem 8.4

8.4 Let $G = (V, E)$ be an undirected, connected graph, with the property that each vertex is connected to at most r vertices. Let f be a positive function defined on V, and let π denote the probability distribution

$$\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)} .$$

If $(x, y) \in E$, define the transition probability

$$p(x, y) = \frac{1}{r} min\left\{1, \frac{f(y)}{f(x)}\right\} ,$$

with $p(x, y) = 0$ otherwise, except that $p(x, x)$ is determined so that the rows sum to one. (i) Show that the Markov chain determined by p is irreducible. (ii) Determine conditions under which the chain is regular. (iii) Show the chain is time reversible with respect to π.

1.1 Part(i)

M.C. is irreducible if there exist no proper closed subset in the state space. Since we are given that the graph G is connected, then this means it is possible to visit each vertex from any other vertex in the graph. But does a connected graph implies no proper closed subset of the corresponding M.C.? The answer is YES. If we view each vertex as state, we just need to show that for each edge in G between 2 vertices x, y, there corresponds a probability of transition from state x to y which is not zero, and also a probability of transition from state y to x which is also not zero. By showing this, we conclude that the M.C. will switch (in some number of steps) to any state from any other state, which implies there is no closed subset, hence P is irreducible.

But from the definition of $p(x, y)$ we see that if there is an edge (x, y) then $p(x, y)$ exist and is not zero, and $p(y, x)$ exist and is not zero (since r is finite). This completes the proof.

1.2 Part(ii)

A finite M.C. is regular when, for some integer m, P^m contains only positive elements.

This implies that the one step transition matrix P must have at least one entry along the diagonal P_{ii} that is none-zero (If all elements along the diagonal are zero, then P^m will always contain at least one zero element no matter how large m is). But a diagonal element not being zero is the same as saying that at least one state must be aperiodic (if $P_{ii} > 0$ then the period is one).

Hence the condition for the M.C. to be regular is that at least one state must be aperiodic[1].

To proof that the above chain is regular, we then need to show that at least one state is aperiodic.

This is the proof:

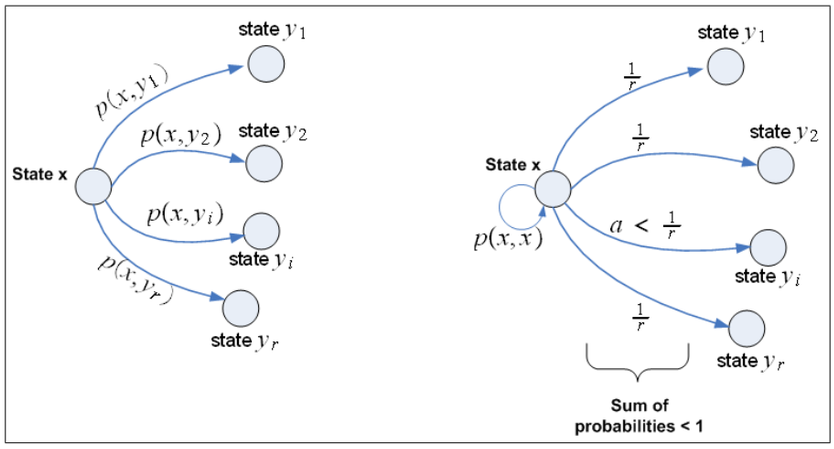
Since at most a vertex can have r edges, then we can find a vertex x with r edges connecting it to vertices $y_1, y_2, \cdots, y_r$ with corresponding one step probability transitions of $p(x, y_1), p(x, y_2), \cdots, p(x, y_r)$. (If we can't find such a vertex, the argument will apply to any other vertex, just replace r with the number of edges on that vertex and the argument will still apply).

Now let us consider $f(x)$ and compare it to each of the $f(y_i)$ where the y_i is the vertex with direct edge from x. There are 2 cases to consider:

1. $f(x) >$ at least one of the $f(y_i)$, $i = 1 \cdots r$
2. $f(x) <$ all of $f(y_i)$, $i = 1 \cdots r$
3. $f(x) =$ all of $f(y_i)$, $i = 1 \cdots r$

Consider case (1): Since $f(x) > f(y_i)$ for some i, then for this specific y_i, $p(x, y_i) = \frac{1}{r} \min\left\{1, \frac{f(y)}{f(x)}\right\} = \frac{1}{r}k$ where $k < 1$, hence $p(x, y_i) = a$ where $a < \frac{1}{r}$. Lets assume there was only one y_i such that the above is true. I.e. at least one of the vertices connected to x had $f(y_i) < f(x)$ (if more if found, it will not change the argument). Now we add all the probabilities $p(x, y_i)$ and we found that this sum is $\overbrace{\frac{1}{r} + \frac{1}{r} + \cdots + \frac{1}{r}}^{(r-1)\ vertices} + a$ where the a is for that vertex which had $f(y_i) < f(x)$. Now since $a < \frac{1}{r}$ then this sum will be LESS THAN ONE. But the sum of the one step probability transition from each state must be 1, hence to compensate, we must then have $p(x, x)$ added to make up for the difference. Hence we showed that under case (1) we can find p_{ii} which is not zero. This diagram illustrate this case

[1] In addition, since we showed in part (i) that this chain is an irreducible chain, hence each state communicate with each other state, hence all states must be of the same type since all states are in the same communication class (Theorem 5.3.2). Then if one state is aperiodic, then the all states that communicate with it must also be aperiod (to be of the same type). Hence in an irreduible chain, if one state is aperiodic, then all states are aperiodic as well.



Now we consider case (2). In this case since $f(x) < f(y_i)$ for each i, then $p(x,y_i) = \frac{1}{r}\min\left\{1, \frac{f(y_i)}{f(x)}\right\} = \frac{1}{r}$., then the sum of the probabilities of transitions from x is $\overbrace{\frac{1}{r}+\frac{1}{r}+\cdots+\frac{1}{r}}^{r \text{ vertices}} =$ 1and we do not need to compensate by adding $p(x,x)$ to make up for the deficit. However since now $f(y_i) > f(x)$ then if we view y_i as the x vertex and the x vertex as the y, and consider the probability transitions out of y_i, then we are back to case (1) above. Hence in case (2) as well ,we can find a state in which $p(x,x) > 0$, Hence the chain is aperiodic, and since it is irreducible, then it is regular in this case as well.

Now consider case (3): In this case $f(x) = f(y_i)$ for $i = 1 \cdots r$. In other words, $f(x)$ is CONSTANT. In this case $p(x,y_i) = \frac{1}{r}\min\left\{1, \frac{f(y_i)}{f(x)}\right\} = \frac{1}{r}$,then the sum of the probabilities of transitions from x is $\overbrace{\frac{1}{r}+\frac{1}{r}+\cdots+\frac{1}{r}}^{r \text{ vertices}} =$ 1and we do not need to compensate by adding $p(x,x)$ to make up for the deficit. This will be true for any node. Therefore, it is not possible to find at least one node with the probabilities attached to edges leaving it is less than one. Hence there are no state with $p(x,x) > 0$, hence in this case, the chain is not aperiodic, and hence the chain is NOT regular.

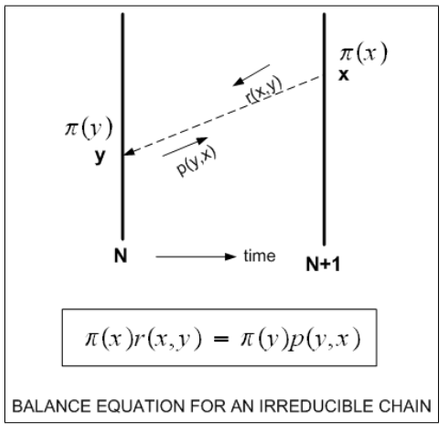
Conclusion: Condition for chain not to be regular is that $f(x)$ be constant.

1.3 Part(iii)

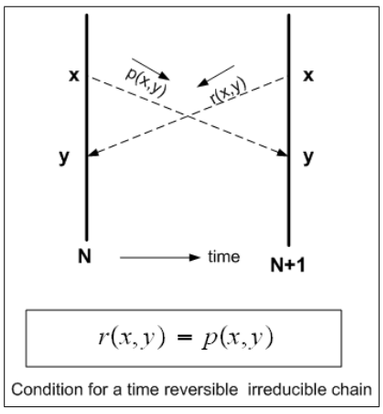
Since the chain is irreducible, then there is a reverse Markov chain (proof is on page 8.1 and 8.2 of lecture notes). Hence for an irreducible chain the balance equations hold

$$r(x,y) = \frac{\pi(y)\, p(y,x)}{\pi(x)} \tag{1}$$

This diagram helps me remember these formulas



Now if the chain the time reversible as well, then $r(x,y) = p(x,y)$,



Then the balance equation (1) becomes

$$\pi(x)\,p(x,y) = \pi(y)\,p(y,x) \tag{2}$$

Hence we need to show that the equation above holds to show the chain is time reversible.

Let the LHS of (2) be $\pi(x)\,p(x,y)$ and let RHS of (2) be $\pi(y)\,p(y,x)$. Then we will show that LHS=RHS for the following 3 cases:

1. $f(x) = f(y)$
2. $f(x) < f(y)$
3. $f(x) > f(y)$

Case(1): Since $f(x) = f(y)$ let these be some value, say z

$$\begin{aligned} LHS &= \pi(x) p(x,y) \\ &= \frac{f(x)}{\sum_{v \in V} f(v)} \left(\frac{1}{r} \min \left\{ 1, \frac{f(y)}{f(x)} \right\} \right) \\ &= \frac{z}{\sum_{v \in V} f(v)} \frac{1}{r} \end{aligned} \tag{3}$$

and

$$\begin{aligned} RHS &= \pi(y) p(y,x) \\ &= \frac{f(y)}{\sum_{v \in V} f(v)} \left(\frac{1}{r} \min \left\{ 1, \frac{f(x)}{f(y)} \right\} \right) \\ &= \frac{z}{\sum_{v \in V} f(v)} \frac{1}{r} \end{aligned} \tag{4}$$

We see that (3) is the same as (4), hence $\boxed{\text{LHS=RHS for case (1)}}$.

case(2): $f(x) < f(y)$

$$\begin{aligned} LHS &= \pi(x) p(x,y) \\ &= \frac{f(x)}{\sum_{v \in V} f(v)} \left(\frac{1}{r} \min \left\{ 1, \frac{f(y)}{f(x)} \right\} \right) \\ &= \frac{f(x)}{\sum_{v \in V} f(v)} \frac{1}{r} \end{aligned} \tag{5}$$

and

$$\begin{aligned} RHS &= \pi(y) p(y,x) \\ &= \frac{f(y)}{\sum_{v \in V} f(v)} \left(\frac{1}{r} \min \left\{ 1, \frac{f(x)}{f(y)} \right\} \right) \\ &= \frac{f(y)}{\sum_{v \in V} f(v)} \frac{1}{r} \frac{f(x)}{f(y)} \\ &= \frac{f(x)}{\sum_{v \in V} f(v)} \frac{1}{r} \end{aligned} \tag{6}$$

Hence we see that (5) is the same as (6). Hence $\boxed{\text{RHS=LHS for case(2)}}$.

case (3):$f(x) > f(y)$

$$\begin{aligned} LHS &= \pi(x) p(x,y) \\ &= \frac{f(x)}{\sum_{v \in V} f(v)} \left(\frac{1}{r} \min \left\{ 1, \frac{f(y)}{f(x)} \right\} \right) \\ &= \frac{f(x)}{\sum_{v \in V} f(v)} \frac{1}{r} \frac{f(y)}{f(x)} \\ &= \frac{f(y)}{\sum_{v \in V} f(v)} \frac{1}{r} \end{aligned} \tag{7}$$

and

$$\begin{aligned} RHS &= \pi(y) p(y,x) \\ &= \frac{f(y)}{\sum_{v \in V} f(v)} \left(\frac{1}{r} \min \left\{ 1, \frac{f(x)}{f(y)} \right\} \right) \\ &= \frac{f(y)}{\sum_{v \in V} f(v)} \frac{1}{r} \end{aligned} \tag{8}$$

We see that (7) is the same as (8), hence $\boxed{\text{LHS=RHS for case (3) as well}}$.

$\boxed{\text{Hence we showed the balance equation for the time reversible condition is satisfied}}$. QED.

2 Problem 8.5

8.5 Suppose $G = (V, E)$ is an undirected, connected graph. For each vertex $v \in V$, let $edge(v)$ denote the number of edges that are connected to v. Let f be a positive function defined on V, and let π denote the probability distribution

$$\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)} .$$

(a) Implement the Hastings-Metropolis method to find a regular Markov chain whose limiting distribution is π. Start with the initial irreducible chain defined by

$$q(x, y) = \frac{1}{edge(x)} , \quad \text{whenever } (x, y) \in E .$$

Note that the Markov chain with this one-step transition matrix is traversed by selecting at random one of the edges connected to x, and then making the transition to the corresponding node. $(a1)$ Show the Markov chain is irreducible. $(a2)$ Determine conditions under which the chain is regular. $(a3)$ Show the chain is time reversible with respect to π. (b) Write a MATLAB program that determines the one-step probability matrix resulting from this method. The input to this program is the function f, and the graph, represented by an adjacency matrix. An adjacency matrix is an $n \times n$ matrix, where n is the number of nodes in the graph, and where entry (i, j) is one if there is an edge connecting nodes i and j, and is zero otherwise. Use this adjacency matrix to compute the function $edge(v)$ at each node. Apply your program to the graph $G = (V, E)$, where $V = \{1, 2, 3, 4\}$, and $E = \{(1, 2), (1, 3), (2, 3), (2, 4), (3, 4)\}$,

and where $f(1) = 2$, $f(2) = 8$, $f(3) = 6$, and $f(4) = 4$. Verify (using MATLAB) that the resulting chain is regular and has the required limiting state probability distribution.

2.1 Part(a)

The following is the Hastings-Metrpolois algorithm implementation.

This algorithm generates a time-reversible M.C. (referred to as p in the lecture notes) given an irreducible M.C. (called q or the original chain) and given a stationary distribution π for that chain.

Input: $f(x)$ defined over the states x, and $edge(x)$ which represents the number of edges connected to x

1. For each state x calculate $\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)}$ and for each state x calculate $edge(x)$
2. compute $q(x,y) = \frac{1}{edge(x)}$ whenever $edge(x) \neq 0$ else set $q(x,y) = 0$
3. Select a state x by random to start from.
4. Let $n = 1$ and let $X_1 = x$
5. Let S be the set of all states that can be reached in one step from x. These will be the states y in which $q(x,y) \neq 0$
6. Select a state y from S by random (using a uniform $U[0,1]$ random number generator)
7. Calculate $\beta(x,y) = \min\left\{1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right\}$
8. Generate a random number u from $U[0,1]$
9. Let $n = n + 1$
10. Compare u to $\beta(x,y)$.
11. IF $u < \beta(x,y)$ THEN $X_n = y$ (select the new state) ELSE $X_n = X_{n-1}$ (stay in same state) ENDIF
12. Let $x = X_n$
13. If $n >$ some Max number of iterations or if we reached some convergence limit Then go to 15
14. GOTO 5
15. Algorithm is complete. Now generate the time reversible MC as follows
    (a) Scan the state path generate X_n and count how many times state x switches to state y in one step
    (b) Do the above for all the states x
    (c) Divide the above number by the total number of steps made to generate $p(x,y)$

Since the problem now asks to *implement* Hastings-Metropolis, then I used the data given at the end of the problem and implemented the above simulation using that data[2]. Please see appendix for code and final P matrix generated.

[2]I allready had the code for the simulation written, just needed to feed the new data for this problem.

2.1.1 Part (a1)

This is similar the problem 8.4 part(I). To show that the p (final M.C.) is irreducible, we need to show that there exist no closed proper subsets. Since the graph G is connected, then we just need to show whenever there is an edge between vertex x and y then there corresponds in the chain representation of the final p matrix a non-zero $p(x,y)$ and also a non-zero $p(y,x)$. This will insure that the each state can transition to each other state, just as each vertex can be visited from each other vertex (since it is a connected graph).

Let us consider any 2 vertices say x, y with a direct edge between them (this is the only case we need to consider due to the argument above). We need to show the resulting $p(x,y)$ and $p(y,x)$ are non-zero

Consider $p(x,y)$ first. Since

$$\begin{aligned} p(x,y) &= q(x,y)\,\beta(x,y) \\ &= \frac{1}{edge(x)} \min\left\{1, \frac{\pi(y)\,q(y,x)}{\pi(x)\,q(x,y)}\right\} \\ &= \frac{1}{edge(x)} \min\left\{1, \frac{\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}}{\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}}\right\} \end{aligned}$$

Hence

$$\boxed{p(x,y) = \frac{1}{edge(x)} \min\left\{1, \frac{f(y)edge(x)}{f(x)edge(y)}\right\}} \tag{1}$$

Then it is clear that whenever there is an edge between x, y then $p(x,y) \neq 0$ since both $f(x)$ and $f(y)$ are positive (not zero) and also edge(x) and edge(y) are non-zero as well. Hence we see that $p(x,y) \neq 0$. Similar argument shows that $p(y,x) \neq 0$.

This shows that M.C. represented by P is irreducible.

2.1.2 Part (a2)

The condition for regular chain P is that there exist at least one state x such that $p(x,x) > 0$.From (1) above we can decide under what conditions this will occur.

Consider a vertex x with $edge(x)$ edges from it connected to vertices $y_1, y_2, \cdots, y_r$. Then from (1) we see that

$$\begin{aligned} p(x,y_i) &= \frac{1}{edge(x)} \min\left\{1, \frac{f(y_i)\,edge(x)}{f(x)\,edge(y_i)}\right\} \\ &= \frac{1}{edge(x)} \min\left\{1, \frac{\frac{f(y_i)}{edge(y_i)}}{\frac{f(x)}{edge(x)}}\right\} \end{aligned}$$

The condition for having $p(x,x)>0$ is that $\min\left\{1,\frac{\frac{f(y_i)}{edge(y_i)}}{\frac{f(x)}{edge(x)}}\right\}<1$, since this will cause $p(x,y_i)$ to be some quantity less than $\frac{1}{r}$ and so when summing over all r there will be a deficit in the sum and we have to compensate for it to make it 1 by adding $p(x,x)$. But for $\min\left\{1,\frac{\frac{f(y_i)}{edge(y_i)}}{\frac{f(x)}{edge(x)}}\right\}$ to be less than ONE means that $\boxed{\frac{f(y_i)}{edge(y_i)}<\frac{f(x)}{edge(x)}}$

Hence the condition for finding an Aperiodic state is finding a vertex x such that the above holds for one of the vertices y_i this vertex is directly connected to. For example, if y_i had the same number of edges from it as does x, then the condition will be that $f(y_i)<f(x)$. And if y_i has less or more edges from it than x has, then we need the ratio $\frac{f(y_i)}{edge(y_i)}$ to be less than $\frac{f(x)}{edge(x)}$.

The above is the same as saying $\frac{f(x)}{edge(x)}$must be constant for the p not to be regular.

2.1.3 Part(A3)

Since the chain is irreducible, then there is a reverse Markov chain (proof is on page 8.1 and 8.2 of lecture notes). Hence for an irreducible chain the balance equations hold

$$r(x,y)=\frac{\pi(y)\,p(y,x)}{\pi(x)} \tag{2}$$

Now if the chain the time reversible as well, then $r(x,y)=p(x,y)$, Then the balance equation (1) becomes

$$\begin{aligned}
\pi(x)\,p(x,y)&=\pi(y)\,p(y,x)\\
\frac{f(x)}{\sum_{v\in V}f(v)}q(x,y)\,\beta(x,y)&=\frac{f(y)}{\sum_{v\in V}f(v)}q(y,x)\,\beta(y,x)\\
\frac{f(x)}{\sum_{v\in V}f(v)}\frac{1}{edge(x)}\left(\min\left\{1,\frac{f(y)\,edge(x)}{f(x)\,edge(y)}\right\}\right)&=\frac{f(y)}{\sum_{v\in V}f(v)}\frac{1}{edge(y)}\left(\min\left\{1,\frac{f(x)\,edge(y)}{f(y)\,edge(x)}\right\}\right)\\
\frac{f(x)}{\sum_{v\in V}f(v)}\frac{1}{edge(x)}\left(\min\left\{1,\frac{\frac{f(y)}{edge(y)}}{\frac{f(x)}{edge(x)}}\right\}\right)&=\frac{f(y)}{\sum_{v\in V}f(v)}\frac{1}{edge(y)}\left(\min\left\{1,\frac{\frac{f(x)}{edge(x)}}{\frac{f(y)}{edge(y)}}\right\}\right)
\end{aligned} \tag{3}$$

Hence we need to show that the equation (3) above holds to show the chain is time reversible.

There are 3 cases to consider:

1. $\frac{f(y)}{edge(y)}=\frac{f(x)}{edge(x)}$
2. $\frac{f(y)}{edge(y)}<\frac{f(x)}{edge(x)}$

3. $\frac{f(y)}{edge(y)} > \frac{f(x)}{edge(x)}$

For case (1), LHS of equation (3) simplifies to $\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}$ and the RHS of (3) simplifies to $\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}$, but since $\frac{f(y)}{edge(y)} = \frac{f(x)}{edge(x)}$, then LHS=RHS.

Hence balance equation (3) is satisfied for case (1).

For case(2), LHS of (3) simplifies $\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)} \left(\frac{\frac{f(y)}{edge(y)}}{\frac{f(x)}{edge(x)}} \right) = \frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}$ and RHS of (3) simplifies to $\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}$,then LHS=RHS.

Hence balance equation (3) is satisfied for case (2).

For case (3), LHS of (3) simplifies $\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}$ and RHS of (3) simplifies to $\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)} \left(\frac{\frac{f(x)}{edge(x)}}{\frac{f(y)}{edge(y)}} \right) = \frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}$,then LHS=RHS.

Hence balance equation (3) is satisfied for case (3).

Hence in all 3 cases we showed the balance equation is satisfied.

Hence M.C. is time reversible.

2.2 Part(b)

A small program written to construct the P matrix directly following instructions on page 8.4 of lecture notes. The following is the resulting P matrix

$$\begin{pmatrix} 0. & 0.5 & 0.5 & 0. \\ 0.0625 & 0.4375 & 0.25 & 0.25 \\ 0.0833333 & 0.333333 & 0.25 & 0.333333 \\ 0. & 0.5 & 0.5 & 0. \end{pmatrix}$$

Now to check that the final chain P is regular, it was raised to some high power to check that all entries in the $P^m > 0$. This is the result

```
In[17]:= MatrixPower[p, 50] // N // MatrixForm

Out[17]//MatrixForm=
⎛0.0526316 0.421053 0.315789 0.210526⎞
⎜0.0526316 0.421053 0.315789 0.210526⎟
⎜0.0526316 0.421053 0.315789 0.210526⎟
⎝0.0526316 0.421053 0.315789 0.210526⎠
```

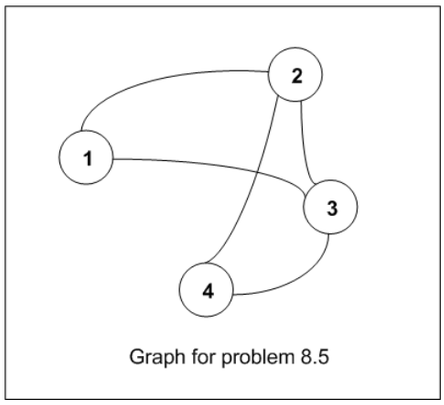
The above verifies that the final matrix p is regular.

Using the Hastings-Metropolis simulation algorithm, the convergence to the above matrix was slow. Had to make 2 million observation to be within 3 decimal points from the above. Here is the P matrix generated from Hastings algorithm for $N = 2,000,000$

```
⎛0.        0.500114 0.499886 0.       ⎞
⎜0.0625897 0.437179 0.249784 0.250448⎟
⎜0.0831875 0.333962 0.248524 0.334326⎟
⎝0.        0.499297 0.500703 0.       ⎠
```

3 Appendix (Implementation of part(a) and part(b))

The graph for part(a) and part(b) is the following



Graph for problem 8.5

Hastings - Metropolis Algorithm implementation For Problem 8.5 part(a)

This below is an implementation of the Hastings - Metropolis algorithm. A simple GUI interface allows the user to specify the number of steps to run the algorithm for. At each step, the current P matrix and the current calculated stationary distribution for this P matrix are shown to help observe the convergence.

The input to this run below is that of problem 8.5 part(b)

Few seed the random number generator and display the q and the π distribution used

```
SeedRandom[121 212];
```

Define the data given in the problem

```
v = {1, 2, 3, 4};
edges = {{1, 2}, {1, 3}, {2, 3}, {2, 4}, {3, 4}};
f = {1, 8, 6, 4};
```

Define the functions π(x) and q(x,y) to use in the implementation

```
pi[x_, f_] := f[[x]] / Sum[f[[i]], {i, 1, Length[f]}]

q[x_, y_] := Module[{r},
  r = Count[edges, {x, any_}] + Count[edges, {any_, x}];
  If[(Count[edges, {x, y}] > 0 || Count[edges, {y, x}] > 0), 1 / r, 0]
 ]
```

Find the stationary distribution

```
w = Table[pi[x, f], {x, 1, v[[-1]]}]
```

$$\left\{\frac{1}{19}, \frac{8}{19}, \frac{6}{19}, \frac{4}{19}\right\}$$

Generate the original q matrix

```
(originalMatrix = Table[q[x, y], {x, 1, v[[-1]]}, {y, 1, v[[-1]]}]) // MatrixForm
```

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

Bring up the user interface

```
m = Manipulate[First@{x = hastings[maxN]; Grid[{{"stationary distribution w=", N[w]},
      {"current stationary distribution=", MatrixPower[N[x], 100][[1, All]]},
      {, }, {"Current P Matrix=", N[MatrixForm[x]]}}, Alignment → Left]},
  {{maxN, 1, "number of steps"}, 1, 2 000 000, 100, ContinuousAction → False,
   Appearance → "Labeled"},
  AutorunSequencing → {{1, 300}}
 ]
```

number of steps 273 401

stationary distribution w= {0.0526316, 0.421053, 0.315789, 0.210526}

current stationary distribution= {0.053453, 0.420567, 0.316001, 0.209979}

Current P Matrix=

$$\begin{pmatrix} 0. & 0.502771 & 0.497229 & 0. \\ 0.0637317 & 0.435999 & 0.251526 & 0.248743 \\ 0.0843336 & 0.331767 & 0.250466 & 0.333434 \\ 0. & 0.502369 & 0.497631 & 0. \end{pmatrix}$$

Define a function for cummulative sum

```
cumSum[list_] := Module[{i, sum, s, k},
  sum = 0;
  k = Length[list];
  s = Table[0, {k}];
  For[i = 1, i ≤ k, i++,
   {
    sum = sum + list[[i]];
    s[[i]] = sum;
   }
  ];
  s
 ]
```

Function to calculate β (x, y)

```
beta[x_, y_, pi_, q_] := Module[{},
  Min[1, pi[[y]] q[[y, x]] / pi[[x]] q[[x, y]]]
 ]
```

Function called at the end of the run to generate P from the path of states travelled

```
generatePMatrixFromStatePath[nStates_, x_] := Module[{i, j, p, allPairs, n, m},
  n = Length[x];
  (*Print["X=",x];*)
  p = Table[0, {nStates}, {nStates}];
  allPairs = Partition[x, 2, 1];
  For[i = 1, i ≤ nStates, i++,
   {
    m = Count[allPairs, {i, y_}];
    For[j = 1, j ≤ nStates, j++,
     If[m ≠ 0, p[[i, j]] = Count[allPairs, {i, j}] / m, p[[i, j]] = 0]
    ]
   }
  ];

  p
 ]
```

Function to sample from q using uniform distribution

```
sampleFromQConditional[q_, x_] := Module[{s, found, j, k, sample, y},
  s = Flatten[Position[q[[x, All]], Except[0], 1, Heads → False] ];
  sample = q[[x, s]];
  sample = cumSum[sample];
  y = RandomReal[];
  found = False;
  For[j = 1, j ≤ Length[sample], j++,
   If[ Not[found], If[y ≤ sample[[j]], {k = j; found = True}]]
  ];

  y = s[[k]]
 ]
```

The Hastings algorithm main loop

```
hastings[maxN_] := Module[{i, j, nStates, n, s, y, α, u, x, sample, pts, sum, k, found},
  nStates = Length[originalMatrix];
  n = 1;
  x = Table[0, {maxN}];
  x[[n]] = 1;  (*pick any state to start from*)
  i = 1;
  While[i < maxN,
   {
    y = sampleFromQConditional[originalMatrix, x[[n]]];
    α = beta[x[[n]], y, w, originalMatrix];
    u = RandomReal[];
    n++;
    If[u ≤ α, x[[n]] = y, x[[n]] = x[[n - 1]]]; (*acceptance step*)
    i++;
   }
  ];

  generatePMatrixFromStatePath[nStates, x]
 ]
```

Problem 8.5 part(b)

by Naser Abbasi. Mathematics 504, Spring 2008. CSUF

This below construct the P matrix using direct computation of the Hasting-Meropolis method

Define the data given in the problem

```
v = {1, 2, 3, 4};
edges = {{1, 2}, {1, 3}, {2, 3}, {2, 4}, {3, 4}};
f = {1, 8, 6, 4};
```

Define the functions π(x) to use in the implementation

```
pi[x_, f_] := f[[x]] / Sum[f[[i]], {i, 1, Length[f]}]
```

Define the functions q(x,y) to use in the implementation. This does something similar to the adjcancy matrix normally used. I used the Count[] function in *Mathematica* which automatically counts the edges from the edges list above, so there is really no need to construct an adjancy matrix as such.

```
q[x_, y_] := Module[{r},
  r = Count[edges, {x, any_}] + Count[edges, {any_, x}];
  If[(Count[edges, {x, y}] > 0 || Count[edges, {y, x}] > 0), 1 / r, 0]
 ]
```

Find the stationary distribution

```
w = Table[pi[x, f], {x, 1, v[[-1]]}]
```

$$\left\{\frac{1}{19}, \frac{8}{19}, \frac{6}{19}, \frac{4}{19}\right\}$$

Generate the original q matrix

```
(originalMatrix = Table[q[x, y], {x, 1, v[[-1]]}, {y, 1, v[[-1]]}]) // MatrixForm
```

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

define Beta function

```
beta[x_, y_, pi_, q_] := Module[{},
  Min[1, pi[[y]] q[[y, x]] / (pi[[x]] q[[x, y]])]
 ]
```

define a function to calculate a non - diagonal entry in the P matrix

```
makeEntryInPMatrix[x_, y_, pi_, q_] := Module[{},
  If[q[[x, y]] == 0, 0, q[[x, y]] beta[x, y, pi, q]]
 ]
```

Construct the P matrix for the off-diagonal elements only

```
nStates = Length[originalMatrix];
p = Table[0, {nStates}, {nStates}];

For[i = 1, i ≤ nStates, i++,
  For[j = 1, j ≤ nStates , j++,
   If[i ≠ j, p[[i, j]] = makeEntryInPMatrix[i, j, w, originalMatrix]]
  ]
 ];
```

Now calculate the diagonal elements of the P matrix

```
For[i = 1, i ≤ nStates, i++,
  For[j = 1, j ≤ nStates , j++,
   If[i == j, p[[i, j]] = 1 - Total[p[[i, All]]]]
  ]
 ];
```

Print the P matrix

```
p // N // MatrixForm
```

$$\begin{pmatrix} 0. & 0.5 & 0.5 & 0. \\ 0.0625 & 0.4375 & 0.25 & 0.25 \\ 0.0833333 & 0.333333 & 0.25 & 0.333333 \\ 0. & 0.5 & 0.5 & 0. \end{pmatrix}$$

Raise the final p matrix to some large power to verify it is regular

```
MatrixPower[p, 50] // N // MatrixForm
```

$$\begin{pmatrix} 0.0526316 & 0.421053 & 0.315789 & 0.210526 \\ 0.0526316 & 0.421053 & 0.315789 & 0.210526 \\ 0.0526316 & 0.421053 & 0.315789 & 0.210526 \\ 0.0526316 & 0.421053 & 0.315789 & 0.210526 \end{pmatrix}$$

4.12 Wed 4/16/2008

Grade: 4/4.

4.12.1 Problem 8.4

8.4 Let $G = (V, E)$ be an undirected, connected graph, with the property that each vertex is connected to at most r vertices. Let f be a positive function defined on V, and let π denote the probability distribution

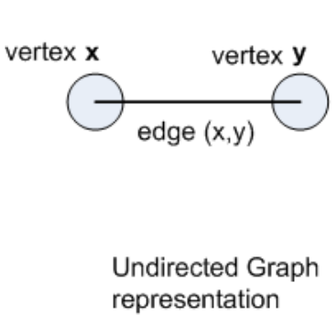
$$\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)} .$$

If $(x, y) \in E$, define the transition probability

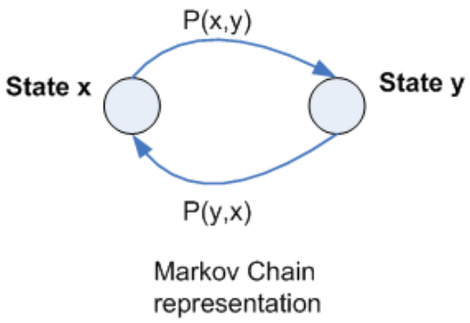
$$p(x, y) = \frac{1}{r} min\left\{1, \frac{f(y)}{f(x)}\right\} ,$$

with $p(x, y) = 0$ otherwise, except that $p(x, x)$ is determined so that the rows sum to one. (i) Show that the Markov chain determined by p is irreducible. (ii) Determine conditions under which the chain is regular. (iii) Show the chain is time reversible with respect to π.

Part(i)



Undirected Graph representation



Markov Chain representation

M.C. is irreducible if there exist no proper closed subset in the state space. Since we are given that the graph G is connected, then this means it is possible to visit each vertex from any other vertex in the graph. But does a connected graph implies no proper closed subset of the corresponding M.C.? The answer is YES. If we view each vertex as state, we just need to show that for each edge in G between 2 vertices x, y, there corresponds a probability of transition from state x to y which is not zero, and also a probability of transition from state y to x which is also not zero. By showing this, we conclude that the M.C. will switch (in some number of steps) to any state from any other state, which implies there is no closed subset, hence P is irreducible.

But from the definition of $p(x, y)$ we see that if there is an edge (x, y) then $p(x, y)$ exist and is not zero, and $p(y, x)$ exist and is not zero (since r is finite). This completes the proof.

Part(ii)

A finite M.C. is regular when, for some integer m, P^m contains only positive elements.

This implies that the one step transition matrix P must have at least one entry along the diagonal P_{ii} that is none-zero (If all elements along the diagonal are zero, then P^m will always contain at least one zero element no matter how large m is). But a diagonal element not being zero is the same as saying that at least one state must be aperiodic (if $P_{ii} > 0$ then the period is one).

Hence the condition for the M.C. to be regular is that at least one state must be aperiodic [5].

To proof that the above chain is regular, we then need to show that at least one state is aperiodic.

This is the proof:

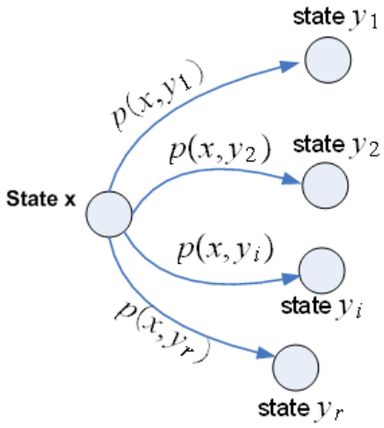
Since at most a vertex can have r edges, then we can find a vertex x with r edges connecting it to vertices $y_1, y_2, \cdots, y_r$ with corresponding one step probability transitions of $p(x, y_1), p(x, y_2), \cdots, p(x, y_r)$. (If we can't find such a vertex, the argument will apply to any other vertex, just replace r with the number of edges on that vertex and the argument will still apply).

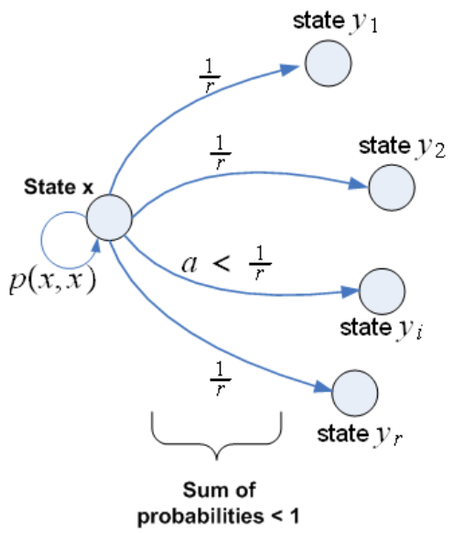
Now let us consider $f(x)$ and compare it to each of the $f(y_i)$ where the y_i is the vertex with direct edge from x. There are 2 cases to consider:

1. $f(x) >$ at least one of the $f(y_i)$, $i = 1 \cdots r$
2. $f(x) <$ all of $f(y_i)$, $i = 1 \cdots r$
3. $f(x) =$ all of $f(y_i)$, $i = 1 \cdots r$

Consider case (1): Since $f(x) > f(y_i)$ for some i, then for this specific y_i, $p(x, y_i) = \frac{1}{r} \min\left\{1, \frac{f(y)}{f(x)}\right\} = \frac{1}{r} k$ where $k < 1$, hence $p(x, y_i) = a$ where $a < \frac{1}{r}$. Lets assume there was only one y_i such that the above is true. I.e. at least one of the vertices connected to x had $f(y_i) < f(x)$ (if more if found, it will not change the argument). Now we add all the probabilities $p(x, y_i)$ and we found that this sum is $\overbrace{\frac{1}{r} + \frac{1}{r} + \cdots + \frac{1}{r}}^{(r-1)\ vertices} + a$ where the a is for that vertex which had $f(y_i) < f(x)$. Now since $a < \frac{1}{r}$ then this sum will be LESS THAN ONE. But the sum of the one step probability transition from each state must be 1, hence to compensate, we must then have $p(x, x)$ added to make up for the difference. Hence we showed that under case (1) we can find p_{ii} which is not zero. This diagram illustrate this case

[5] In addition, since we showed in part (i) that this chain is an irreducible chain, hence each state communicate with each other state, hence all states must be of the same type since all states are in the same communication class (Theorem 5.3.2). Then if one state is aperiodic, then the all states that communicate with it must also be aperiod (to be of the same type). Hence in an irreduible chain, if one state is aperiodic, then all states are aperiodic as well.





Now we consider case (2).

In this case since $f(x) < f(y_i)$ for each i, then $p(x,y_i) = \frac{1}{r}\min\left\{1, \frac{f(y_i)}{f(x)}\right\} = \frac{1}{r}.$, then the sum of the probabilities of transitions from x is $\overbrace{\frac{1}{r}+\frac{1}{r}+\cdots+\frac{1}{r}}^{r \text{ vertices}} = 1$and we do not need to compensate by adding $p(x,x)$ to make up for the deficit. However since now $f(y_i) > f(x)$ then if we view y_i as the x vertex and the x vertex as the y, and consider the probability transitions out of y_i, then we are back to case (1) above. Hence in case (2) as well ,we can find a state in which $p(x,x) > 0$, Hence the chain is aperiodic, and since it is irreducible, then it is regular in this case as well.

Now consider case (3):

In this case $f(x) = f(y_i)$ for $i = 1 \cdots r$. In other words, $f(x)$ is CONSTANT. In this case $p(x,y_i) = \frac{1}{r}\min\left\{1, \frac{f(y_i)}{f(x)}\right\} = \frac{1}{r}$,then the sum of the probabilities of transitions from x is $\overbrace{\frac{1}{r}+\frac{1}{r}+\cdots+\frac{1}{r}}^{r \text{ vertices}} = 1$and we do not need to compensate by adding $p(x,x)$ to make up for the deficit. This will be true for any node. Therefore, it is not possible to find at least one node with the probabilities attached to edges leaving it is less than one. Hence there are no state with $p(x,x) > 0$, hence in this case, the chain is not aperiodic, and hence the chain is NOT regular.

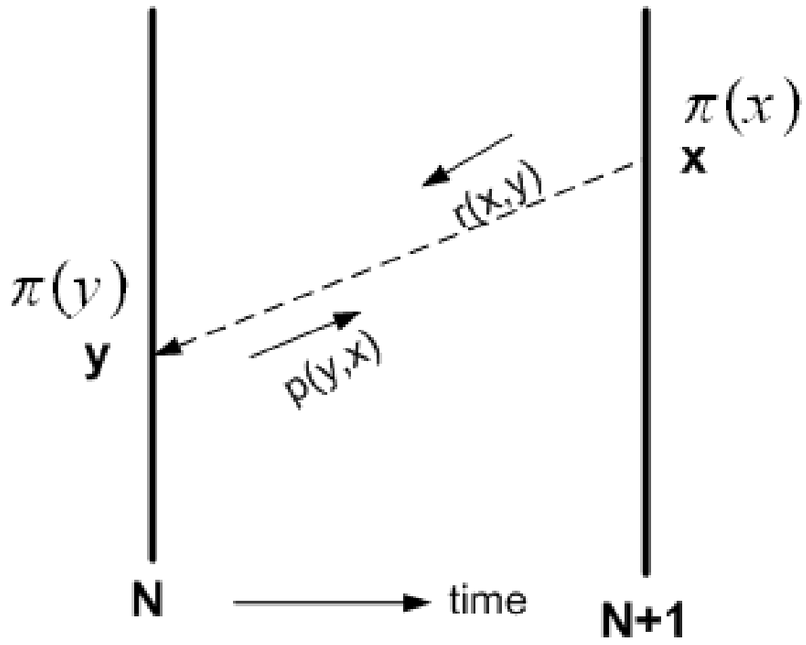
Conclusion: Condition for chain not to be regular is that $f(x)$ be constant.

Part(iii)

Since the chain is irreducible, then there is a reverse Markov chain (proof is on page 8.1 and 8.2 of lecture notes). Hence for an irreducible chain the balance equations hold

$$r(x,y) = \frac{\pi(y)\, p(y,x)}{\pi(x)} \tag{1}$$

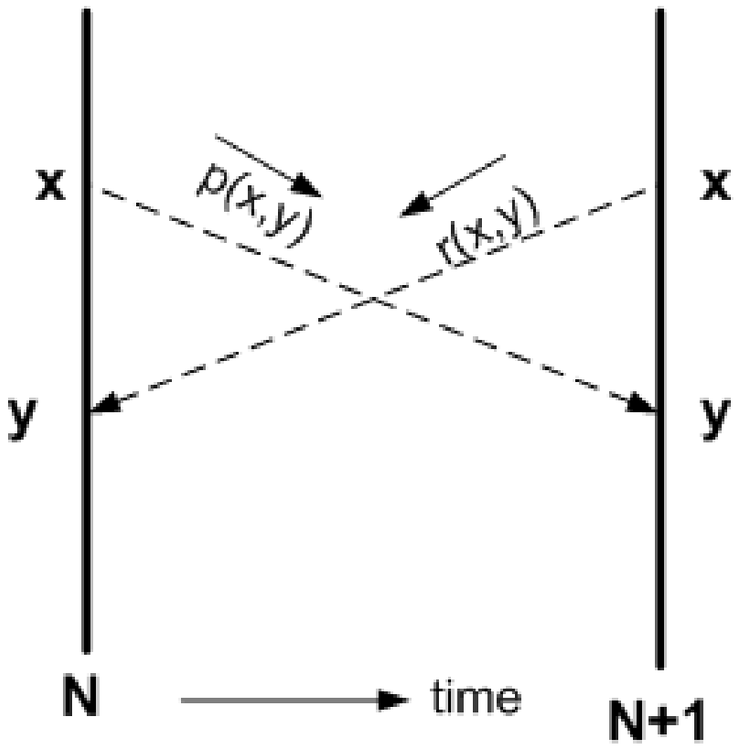
This diagram helps me remember these formulas



$$\pi(x)r(x,y) = \pi(y)p(y,x)$$

BALANCE EQUATION FOR AN IRREDUCIBLE CHAIN

Now if the chain the time reversible as well, then $r(x,y) = p(x,y)$,



$$r(x,y) = p(x,y)$$

Condition for a time reversible irreducible chain

Then the balance equation (1) becomes

$$\pi(x)\,p(x,y) = \pi(y)\,p(y,x) \tag{2}$$

Hence we need to show that the equation above holds to show the chain is time reversible.

Let the LHS of (2) be $\pi(x)\,p(x,y)$ and let RHS of (2) be $\pi(y)\,p(y,x)$. Then we will show that LHS=RHS for the following 3 cases:

1. $f(x) = f(y)$
2. $f(x) < f(y)$
3. $f(x) > f(y)$

Case(1): Since $f(x) = f(y)$ let these be some value, say z

$$\begin{aligned} LHS &= \pi(x)\,p(x,y) \\ &= \frac{f(x)}{\sum_{v\in V} f(v)}\left(\frac{1}{r}\min\left\{1, \frac{f(y)}{f(x)}\right\}\right) \\ &= \frac{z}{\sum_{v\in V} f(v)}\frac{1}{r} \end{aligned} \tag{3}$$

and

$$\begin{aligned} RHS &= \pi(y)\, p(y,x) \\ &= \frac{f(y)}{\sum_{v\in V} f(v)} \left(\frac{1}{r} \min\left\{1, \frac{f(x)}{f(y)}\right\} \right) \\ &= \frac{z}{\sum_{v\in V} f(v)} \frac{1}{r} \end{aligned} \tag{4}$$

We see that (3) is the same as (4), hence $\boxed{\text{LHS=RHS for case (1)}}$.

case(2): $f(x) < f(y)$

$$\begin{aligned} LHS &= \pi(x)\, p(x,y) \\ &= \frac{f(x)}{\sum_{v\in V} f(v)} \left(\frac{1}{r} \min\left\{1, \frac{f(y)}{f(x)}\right\} \right) \\ &= \frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{r} \end{aligned} \tag{5}$$

and

$$\begin{aligned} RHS &= \pi(y)\, p(y,x) \\ &= \frac{f(y)}{\sum_{v\in V} f(v)} \left(\frac{1}{r} \min\left\{1, \frac{f(x)}{f(y)}\right\} \right) \\ &= \frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{r} \frac{f(x)}{f(y)} \\ &= \frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{r} \end{aligned} \tag{6}$$

Hence we see that (5) is the same as (6). Hence $\boxed{\text{RHS=LHS for case(2)}}$.

case (3):$f(x) > f(y)$

$$\begin{aligned} LHS &= \pi(x)\, p(x,y) \\ &= \frac{f(x)}{\sum_{v\in V} f(v)} \left(\frac{1}{r} \min\left\{1, \frac{f(y)}{f(x)}\right\} \right) \\ &= \frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{r} \frac{f(y)}{f(x)} \\ &= \frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{r} \end{aligned} \tag{7}$$

and

$$\begin{aligned} RHS &= \pi(y)\, p(y,x) \\ &= \frac{f(y)}{\sum_{v\in V} f(v)} \left(\frac{1}{r} \min\left\{1, \frac{f(x)}{f(y)}\right\} \right) \\ &= \frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{r} \end{aligned} \tag{8}$$

We see that (7) is the same as (8), hence LHS=RHS for case (3) as well.

Hence we showed the balance equation for the time reversible condition is satisfied. QED.

4.12.2 Problem 8.5

8.5 Suppose $G = (V, E)$ is an undirected, connected graph. For each vertex $v \in V$, let $edge(v)$ denote the number of edges that are connected to v. Let f be a positive function defined on V, and let π denote the probability distribution

$$\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)} .$$

(a) Implement the Hastings-Metropolis method to find a regular Markov chain whose limiting distribution is π. Start with the initial irreducible chain defined by

$$q(x, y) = \frac{1}{edge(x)} , \quad \text{whenever } (x, y) \in E .$$

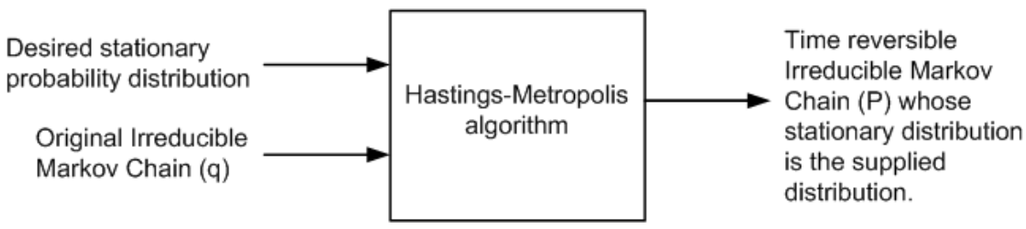
Note that the Markov chain with this one-step transition matrix is traversed by selecting at random one of the edges connected to x, and then making the transition to the corresponding node. $(a1)$ Show the Markov chain is irreducible. $(a2)$ Determine conditions under which the chain is regular. $(a3)$ Show the chain is time reversible with respect to π. (b) Write a MATLAB program that determines the one-step probability matrix resulting from this method. The input to this program is the function f, and the graph, represented by an adjacency matrix. An adjacency matrix is an $n \times n$ matrix, where n is the number of nodes in the graph, and where entry (i, j) is one if there is an edge connecting nodes i and j, and is zero otherwise. Use this adjacency matrix to compute the function $edge(v)$ at each node. Apply your program to the graph $G = (V, E)$, where $V = \{1, 2, 3, 4\}$, and $E = \{(1, 2), (1, 3), (2, 3), (2, 4), (3, 4)\}$,

and where $f(1) = 2$, $f(2) = 8$, $f(3) = 6$, and $f(4) = 4$. Verify (using MATLAB) that the resulting chain is regular and has the required limiting state probability distribution.

Part(a)

The following is the Hastings-Metrpolois algorithm implementation.

This algorithm generates a time-reversible M.C. (referred to as p in the lecture notes) given an irreducible M.C. (called q or the original chain) and given a stationary distribution π for that chain.



Input: $f(x)$ defined over the states x, and $edge(x)$ which represents the number of edges connected to x

1. For each state x calculate $\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)}$ and for each state x calculate $edge(x)$
2. compute $q(x, y) = \frac{1}{edge(x)}$ whenever $edge(x) \neq 0$ else set $q(x, y) = 0$
3. Select a state x by random to start from.
4. Let $n = 1$ and let $X_1 = x$
5. Let S be the set of all states that can be reached in one step from x. These will be the states y in which $q(x, y) \neq 0$

6. Select a state y from S by random (using a uniform $U[0,1]$ random number generator)
7. Calculate $\beta(x,y)=\min\left\{1,\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right\}$
8. Generate a random number u from $U[0,1]$
9. Let $n=n+1$
10. Compare u to $\beta(x,y)$.
11. IF $u<\beta(x,y)$ THEN $X_n=y$ (select the new state) ELSE $X_n=X_{n-1}$ (stay in same state) ENDIF
12. Let $x=X_n$
13. If $n>$ some Max number of iterations or if we reached some convergence limit Then go to 15
14. GOTO 5
15. Algorithm is complete. Now generate the time reversible MC as follows
    (a) Scan the state path generate X_n and count how many times state x switches to state y in one step
    (b) Do the above for all the states x
    (c) Divide the above number by the total number of steps made to generate $p(x,y)$

Since the problem now asks to *implement* Hastings-Metropolis, then I used the data given at the end of the problem and implemented the above simulation using that data[6]. Please see appendix for code and final P matrix generated.

Part (a1)

This is similar the problem 8.4 part(I). To show that the p (final M.C.) is irreducible, we need to show that there exist no closed proper subsets. Since the graph G is connected, then we just need to show whenever there is an edge between vertex x and y then there corresponds in the chain representation of the final p matrix a non-zero $p(x,y)$ and also a non-zero $p(y,x)$. This will insure that the each state can transition to each other state, just as each vertex can be visited from each other vertex (since it is a connected graph).

Let us consider any 2 vertices say x,y with a direct edge between them (this is the only case we need to consider due to the argument above). We need to show the resulting $p(x,y)$ and $p(y,x)$ are non-zero

Consider $p(x,y)$ first. Since

$$\begin{aligned}p(x,y)&=q(x,y)\,\beta(x,y)\\&=\frac{1}{edge\,(x)}\min\left\{1,\frac{\pi(y)\,q(y,x)}{\pi\,(x)\,q\,(x,y)}\right\}\\&=\frac{1}{edge\,(x)}\min\left\{1,\frac{\frac{f(y)}{\sum_{v\in V}f(v)}\frac{1}{edge(y)}}{\frac{f(x)}{\sum_{v\in V}f(v)}\frac{1}{edge(x)}}\right\}\end{aligned}$$

Hence

$$\boxed{p(x,y)=\tfrac{1}{edge(x)}\min\left\{1,\tfrac{f(y)edge(x)}{f(x)edge(y)}\right\}}\tag{1}$$

[6]I allready had the code for the simulation written, just needed to feed the new data for this problem.

Then it is clear that whenever there is an edge between x, y then $p(x,y) \neq 0$ since both $f(x)$ and $f(y)$ are positive (not zero) and also edge(x) and edge(y) are non-zero as well. Hence we see that $p(x,y) \neq 0$. Similar argument shows that $p(y,x) \neq 0$.

This shows that M.C. represented by P is irreducible.

Part (a2)

The condition for regular chain P is that there exist at least one state x such that $p(x,x) > 0$.From (1) above we can decide under what conditions this will occur.

Consider a vertex x with $edge(x)$ edges from it connected to vertices $y_1, y_2, \cdots, y_r$. Then from (1) we see that

$$\begin{aligned} p(x,y_i) &= \frac{1}{edge\,(x)} \min\left\{1, \frac{f(y_i)\,edge(x)}{f\,(x)\,edge\,(y_i)}\right\} \\ &= \frac{1}{edge\,(x)} \min\left\{1, \frac{\frac{f(y_i)}{edge(y_i)}}{\frac{f(x)}{edge(x)}}\right\} \end{aligned}$$

The condition for having $p(x,x) > 0$ is that $\min\left\{1, \frac{\frac{f(y_i)}{edge(y_i)}}{\frac{f(x)}{edge(x)}}\right\} < 1$, since this will cause $p(x,y_i)$ to be some quantity less than $\frac{1}{r}$ and so when summing over all r there will be a deficit in the sum and we have to compensate for it to make it 1 by adding $p(x,x)$. But for $\min\left\{1, \frac{\frac{f(y_i)}{edge(y_i)}}{\frac{f(x)}{edge(x)}}\right\}$ to be less than ONE means that $\frac{f(y_i)}{edge(y_i)} < \frac{f(x)}{edge(x)}$

Hence the condition for finding an Aperiodic state is finding a vertex x such that the above holds for one of the vertices y_i this vertex is directly connected to. For example, if y_i had the same number of edges from it as does x, then the condition will be that $f(y_i) < f(x)$. And if y_i has less or more edges from it than x has, then we need the ratio $\frac{f(y_i)}{edge(y_i)}$ to be less than $\frac{f(x)}{edge(x)}$.

The above is the same as saying $\frac{f(x)}{edge(x)}$ must be constant for the p not to be regular.

Part(A3)

Since the chain is irreducible, then there is a reverse Markov chain (proof is on page 8.1 and 8.2 of lecture notes). Hence for an irreducible chain the balance equations hold

$$r(x,y) = \frac{\pi(y)\,p(y,x)}{\pi\,(x)} \tag{2}$$

Now if the chain the time reversible as well, then $r(x,y) = p(x,y)$, Then the balance equation (1) becomes

$$\begin{aligned} \pi(x)\,p(x,y) &= \pi(y)\,p(y,x) \\ \frac{f(x)}{\sum_{v\in V} f\,(v)} q(x,y)\,\beta(x,y) &= \frac{f(y)}{\sum_{v\in V} f\,(v)} q(y,x)\,\beta(y,x) \\ \frac{f(x)}{\sum_{v\in V} f\,(v)} \frac{1}{edge\,(x)} \left(\min\left\{1, \frac{f(y)\,edge(x)}{f\,(x)\,edge\,(y)}\right\}\right) &= \frac{f(y)}{\sum_{v\in V} f\,(v)} \frac{1}{edge\,(y)} \left(\min\left\{1, \frac{f(x)\,edge(y)}{f\,(y)\,edge\,(x)}\right\}\right) \\ \frac{f(x)}{\sum_{v\in V} f\,(v)} \frac{1}{edge\,(x)} \left(\min\left\{1, \frac{\frac{f(y)}{edge(y)}}{\frac{f(x)}{edge(x)}}\right\}\right) &= \frac{f(y)}{\sum_{v\in V} f\,(v)} \frac{1}{edge\,(y)} \left(\min\left\{1, \frac{\frac{f(x)}{edge(x)}}{\frac{f(y)}{edge(y)}}\right\}\right) \end{aligned} \tag{3}$$

Hence we need to show that the equation (3) above holds to show the chain is time reversible.

There are 3 cases to consider:

1. $\frac{f(y)}{edge(y)} = \frac{f(x)}{edge(x)}$
2. $\frac{f(y)}{edge(y)} < \frac{f(x)}{edge(x)}$
3. $\frac{f(y)}{edge(y)} > \frac{f(x)}{edge(x)}$

For case (1), LHS of equation (3) simplifies to $\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}$ and the RHS of (3) simplifies to $\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}$, but since $\frac{f(y)}{edge(y)} = \frac{f(x)}{edge(x)}$, then LHS=RHS.

Hence balance equation (3) is satisfied for case (1).

For case(2), LHS of (3) simplifies $\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)} \left(\frac{\frac{f(y)}{edge(y)}}{\frac{f(x)}{edge(x)}} \right) = \frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}$ and RHS of (3) simplifies to $\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}$,then LHS=RHS.

Hence balance equation (3) is satisfied for case (2).

For case (3), LHS of (3) simplifies $\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}$ and RHS of (3) simplifies to $\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)} \left(\frac{\frac{f(x)}{edge(x)}}{\frac{f(y)}{edge(y)}} \right) = \frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}$,then LHS=RHS.

Hence balance equation (3) is satisfied for case (3).

Hence in all 3 cases we showed the balance equation is satisfied.

Hence M.C. is time reversible.

Part(b)

A small program written to construct the P matrix directly following instructions on page 8.4 of lecture notes. The following is the resulting P matrix

$$\begin{pmatrix} 0. & 0.5 & 0.5 & 0. \\ 0.0625 & 0.4375 & 0.25 & 0.25 \\ 0.0833333 & 0.333333 & 0.25 & 0.333333 \\ 0. & 0.5 & 0.5 & 0. \end{pmatrix}$$

Now to check that the final chain P is regular, it was raised to some high power to check that all entries in the $P^m > 0$. This is the result

```
In[17]:= MatrixPower[p, 50] // N // MatrixForm

Out[17]//MatrixForm=
0.0526316 0.421053 0.315789 0.210526
0.0526316 0.421053 0.315789 0.210526
0.0526316 0.421053 0.315789 0.210526
0.0526316 0.421053 0.315789 0.210526
```

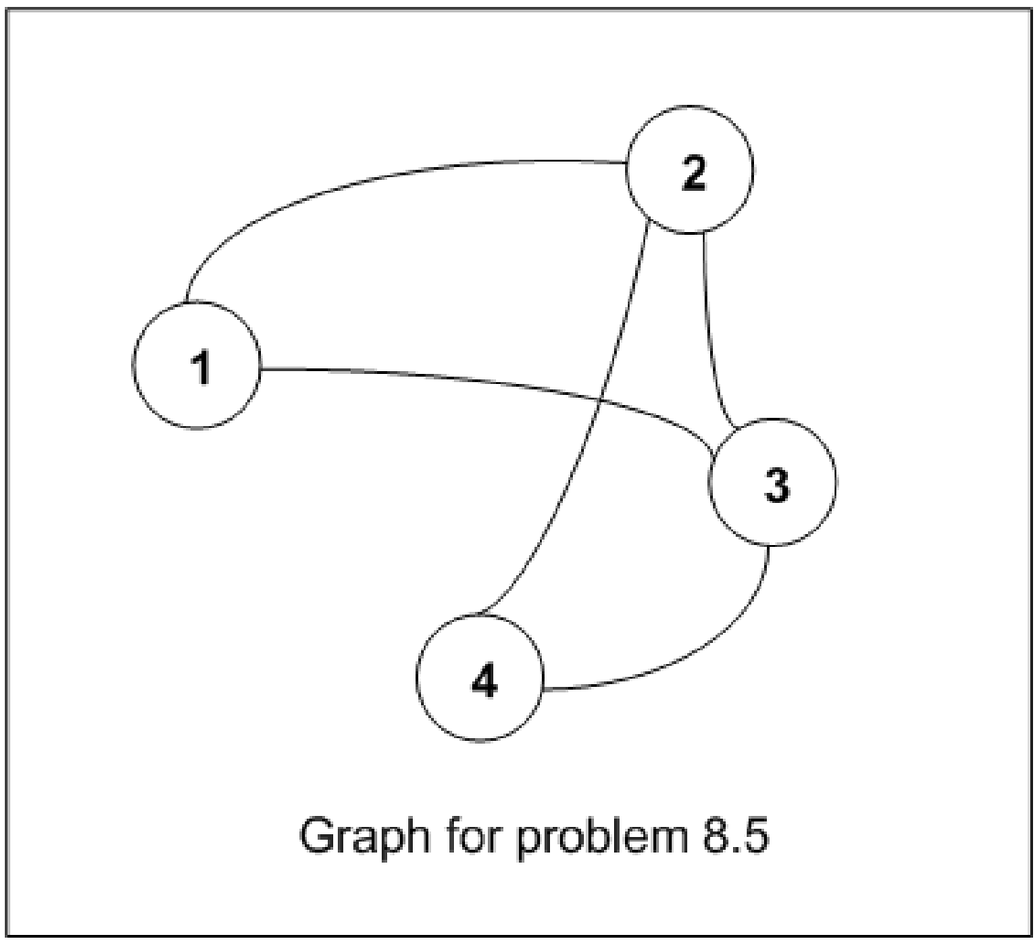
The above verifies that the final matrix p is regular.

Using the Hastings-Metropolis simulation algorithm, the convergence to the above matrix was slow. Had to make 2 million observation to be within 3 decimal points from the above. Here is the P matrix generated from Hastings algorithm for $N = 2,000,000$

$$\begin{pmatrix} 0. & 0.500114 & 0.499886 & 0. \\ 0.0625897 & 0.437179 & 0.249784 & 0.250448 \\ 0.0831875 & 0.333962 & 0.248524 & 0.334326 \\ 0. & 0.499297 & 0.500703 & 0. \end{pmatrix}$$

4.12.3 Appendix (Implementation of part(a) and part(b))

The graph for part(a) and part(b) is the following



4.12.4 code

Mathematica notebook

Mathematica notebook

4.12.5 Key solution

Key Solution For problem 8.5, 8.4

Chapter 8: Markov Chain Monte Carlo Methods - Solutions to Selected problems

8.4 Let $G = (V, E)$ be an undirected, connected graph with the property that each vertex is connected to at most r vertices. Let f be a positive function defined on V and let π denote the probability distribution

$$\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)} .$$

If $(x, y) \in E$, define the transition probability

$$p(x, y) = \frac{1}{r} min \left\{ 1, \frac{f(y)}{f(x)} \right\} .$$

with $p(x, y) = 0$ otherwise, except that $p(x, x)$ is determined so that the rows sum to one. (i) Show that the Markov chain determined by p is irreducible. (ii) Determine conditions under which the chain is regular. (iii) Show the chain is time reversible with respect to π.

Solution (i) To show that the chain is irreducible, note first that G is connected. In other words, in G there is a path from any one node to any other; that is, given any two nodes, say a and b in V, there is a sequence of nodes, say $x_1, x_2, \cdots, x_n$, in V such that $(a, x_1) \in E$, $(x_i, x_{i+1}) \in E$, for each $i = 1, 2, \cdots, n$, and $(x_n, b) \in E$. While the graph G is undirected, the graph of the Markov chain is directed. However, corresponding to each arc in G there are two arcs in the graph of the Markov chain, one in each direction, and each with nonzero probability. Indeed, if $(x, y) \in E$, then there is an arc in the graph of the Markov chain that points from x to y with associated probability $p(x, y) > 0$, determined by the formula above, and there is another arc that points from y to x with associated probability $p(y, x) > 0$, again determined by the formula above. It follows that in the graph of the Markov chain, between any two nodes (now states of the chain), there is a path between these states that can be traversed following the arcs in the required directions. In other words, any two states of the Markov chain communicate. Hence, the chain is irreducible.

(ii) Although the Markov chain is irreducible, it may be periodic, and hence not regular. As a simple example, consider the graph $G = (V, E)$ with vertex set $V = \{1, 2\}$ and edge set $E = \{(1, 2)\}$. Then $r = 1$. Suppose that f is the constant function. Then the associated Markov chain has one-step probability transition matrix

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} .$$

This chain is periodic with period 2. Suppose however that f is not constant. For example, let $f(1) = 1$ and $f(2) = 3$. Then the associated Markov chain has one-step probability transition matrix

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

This chain is aperiodic. However, more generally, whenever f is not the constant function, the associated Markov chain will be aperiodic, and hence regular. To see this result, note that there must be some vertex x such that $(x, y) \in E$, and $f(y) < f(x)$. For this state x, the sum of the off-diagonal elements will be less than one, since there are at most r nonzero off diagonal entries. Hence, for this row, $p(x, x) \neq 0$. Thus, state x is aperiodic, and since the chain is irreducible, all states are aperiodic, and so the chain is aperiodic.

As another condition which implies regularity, suppose that at least one node of the graph G is connected directly to fewer than r nodes. Then, whether f is the constant function or not, that node will become a state in the chain that is aperiodic. Indeed, in the one-step transition matrix, the row corresponding to this state will be such that the sum of the off-diagonal elements will be less than one, and hence the diagonal element will be nonzero. Thus, since the chain is irreducible, and one state is aperiodic, all states are aperiodic.

(*iii*) To show that the balance equations hold, we need to show that $\pi(x)p(x, y) = \pi(y)p(y, x)$ for each pair of states x and y. First, if $p(x, y) = 0$, then $p(y, x) = 0$ also, since $p(x, y) = 0$ only when there is no edge of the graph G that connects x and y. Next, when $(x, y) \in E$,

$$\pi(x)p(x, y) = \frac{f(x)}{rC} min\left\{1, \frac{f(y)}{f(x)}\right\} = \frac{1}{rC} min\left\{f(x), f(y)\right\} ,$$

where C is the sum appearing in the denominator of π. Similarly, we have

$$\pi(y)p(y, x) = \frac{f(y)}{rC} min\left\{1, \frac{f(x)}{f(y)}\right\} = \frac{1}{rC} min\left\{f(y), f(x)\right\} ,$$

These two expressions are the same, which is the desired conclusion.

8.5 Suppose $G = (V, E)$ is an undirected connected graph. For each vertex $v \in V$, let $edge(v)$ denote the number of edges that are connected to v. Let f be a positive function defined on V, and letπ denote the probability distribution

$$\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)} .$$

(*a*) Implement the Hastings-Metropolis method to find a regular Markov chain whose limiting distribution is π. Start with the initial irreducible chain defined by

$$q(x, y) = \frac{1}{edge(x)} , \quad \text{whenever} \quad (x, y) \in E .$$

Note that the Markov chain with this one-step transition matrix is traversed by selecting at random one of the edges connected to x, and then making the transition to the corresponding node. $(a1)$ Show that the Markov chain determined by this method is irreducible. $(a2)$ Determine conditions under which the chain is regular. $(a3)$ Show the chain is time reversible with respect to π. (b) Write a MATLAB program that determines the one-step probability matrix resulting from this method. The input to this program is the function f and the graph, represented by an adjacency matrix. An adjacency matrix is an $n \times n$ matrix, where n is the number of nodes in the graph, and where entry (i, j) is one if there is an edge connecting nodes i and j, and is zero otherwise. Use this adjacency matrix to compute the function $edge(v)$ at each node. Apply your program to the graph $G = (V, E)$ where $V = \{1, 2, 3, 4\}$, and $E = \{(1, 2), (1, 3), (2, 3), (2, 4), (3, 4)\}$, and where $f(1) = 2$, $f(2) = 8$, $f(3) = 6$, and $f(4) = 4$. Verify (using MATLAB) that the resulting chain is regular and has the required limiting state probability distribution.

Solution (a) The one-step transition probabilities are

$$p(x, y) = \frac{1}{edge(x)}\beta(x, y) \quad \text{for } (x, y) \in E\,, \quad \text{with} \quad p(x, x) = 1 - \sum_{y \neq x} p(x, y)\,,$$

where $\beta(x, y)$ is given by

$$\beta(x, y) = min\left\{1, \frac{f(y)edge(x)}{f(x)edge(y)}\right\}\,.$$

Otherwise $p(x, y) = 0$.

$(a1)$ These formulas show that for each arc $(x, y) \in E$, we have $p(x, y) > 0$ and $p(y, x) > 0$. Thus, between any two nodes that are connected by an arc in G, the resulting Markov chain has two corresponding states, x and y, and there are two arcs connecting these states which point in opposite directions. Hence, since the original graph is connected, it is therefore possible, in the Markov chain, to travel from any one state to any other. Thus, the chain is irreducible.

$(a2)$ For the setting of this problem, the Markov chain produced by the Hasting-Metropolis algorithm may be periodic, and hence not regular. For example, consider the graph $G = (V, E)$ with vertex set $V = \{1, 2\}$ and edge set $E = \{(1, 2)\}$. Suppose that f is the constant function. Then the resulting Markov chain has one-step probability transition matrix

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\,,$$

and the chain is periodic with period 2.

However, when $f(x)/edge(x)$ is not the constant function on V, the chain is aperiodic. To see this result, note first that since the graph is connected, there must be two vertices x and y such that $(x, y) \in E$, and $f(y)/edge(y) < f(x)/edge(x)$. For these states x and y we will have $\beta(x, y) < 1$. Therefore, in the one step transition matrix for the Markov chain, the sum of the off-diagonal elements in the row for state x is less than one. Hence, state x is aperiodic. Since the chain is irreducible, the chain is therefore also aperiodic. Thus, in this case when $f(x)/edge(x)$ is not the constant function on V, the chain is irreducible and aperiodic, and hence regular.

$(a3)$ To show that the balance equations hold, the same argument used for the previous problem carries over. we need to show that $\pi(x)p(x, y) = \pi(y)p(y, x)$ for each pair of states x and y. First, if $p(x, y) = 0$, then $p(y, x) = 0$ also, since $p(x, y) = 0$ only when there is no edge of the graph G that connects x and y. Next, when $(x, y) \in E$,

$$\pi(x)p(x,y) = \frac{f(x)}{edge(x)C} min\left\{1, \frac{f(y)edge(x)}{f(x)edge(y)}\right\} = \frac{1}{C} min\left\{\frac{f(x)}{edge(x)}, \frac{f(y)}{edge(y)}\right\} ,$$

where C is the sum appearing in the denominator of π. Similarly, we have

$$\pi(y)p(y,x) = \frac{f(y)}{edge(y)C} min\left\{1, \frac{f(x)edge(y)}{f(y)edge(x)}\right\} = \frac{1}{C} min\left\{\frac{f(y)}{edge(y)}, \frac{f(x)}{edge(x)}\right\} ,$$

These two expressions are the same, which is the desired conclusion.

4.12.6 my graded solution

HW problems 8.4 and 8.5, Mathematics 504

CSUF, spring 2008

by Nasser Abbasi

April 16, 2008

Contents

1 Problem 8.4

8.4 Let $G = (V, E)$ be an undirected, connected graph, with the property that each vertex is connected to at most r vertices. Let f be a positive function defined on V, and let π denote the probability distribution

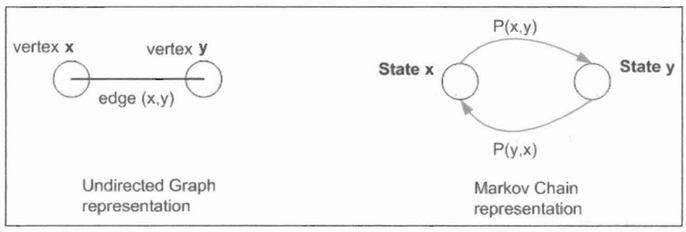
$$\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)} .$$

If $(x, y) \in E$, define the transition probability

$$p(x, y) = \frac{1}{r} min\left\{1, \frac{f(y)}{f(x)}\right\} ,$$

with $p(x, y) = 0$ otherwise, except that $p(x, x)$ is determined so that the rows sum to one. (i) Show that the Markov chain determined by p is irreducible. (ii) Determine conditions under which the chain is regular. (iii) Show the chain is time reversible with respect to π.

1.1 Part(i)



M.C. is irreducible if there exist no proper closed subset in the state space. Since we are given that the graph G is connected, then this means it is possible to visit each vertex from any other vertex in the graph. But does a connected graph implies no proper closed subset of the corresponding M.C.? The answer is YES. If we view each vertex as state, we just need to show that for each edge in G between 2 vertices x, y, there corresponds a probability of transition from state x to y which is not zero, and also a probability of transition from state y to x which is also not zero. By showing this, we conclude that the M.C. will switch (in some number of steps) to any state from any other state, which implies there is no closed subset, hence P is irreducible.

But from the definition of $p(x, y)$ we see that if there is an edge (x, y) then $p(x, y)$ exist and is not zero, and $p(y, x)$ exist and is not zero (since r is finite). This completes the proof.

1.2 Part(ii)

A finite M.C. is regular when, for some integer m, P^m contains only positive elements.

This implies that the one step transition matrix P must have at least one entry along the diagonal P_{ii} that is none-zero (If all elements along the diagonal are zero, then P^m will always contain at least one zero element no matter how large m is). But a diagonal element not being zero is the same as saying that at least one state must be aperiodic (if $P_{ii} > 0$ then the period is one).

Hence the condition for the M.C. to be regular is that at least one state must be aperiodic[1].

To proof that the above chain is regular, we then need to show that at least one state is aperiodic.

This is the proof:

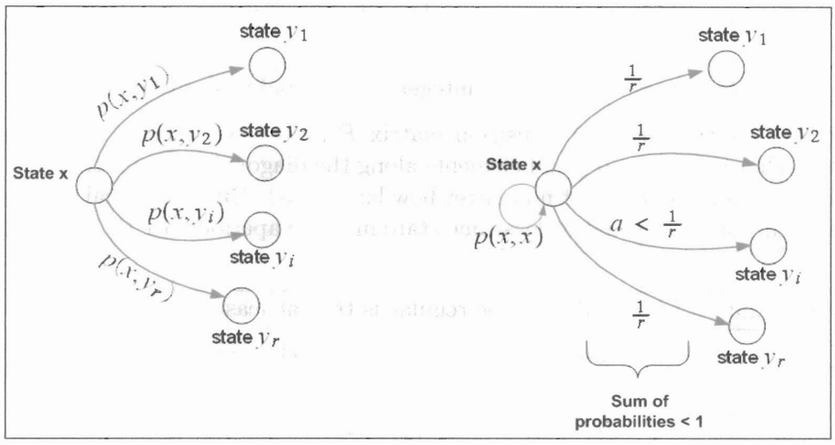
Since at most a vertex can have r edges, then we can find a vertex x with r edges connecting it to vertices $y_1, y_2, \cdots, y_r$ with corresponding one step probability transitions of $p(x, y_1), p(x, y_2), \cdots, p(x, y_r)$. (If we can't find such a vertex, the argument will apply to any other vertex, just replace r with the number of edges on that vertex and the argument will still apply).

Now let us consider $f(x)$ and compare it to each of the $f(y_i)$ where the y_i is the vertex with direct edge from x. There are 2 cases to consider:

1. $f(x) >$ at least one of the $f(y_i)$, $i = 1 \cdots r$
2. $f(x) <$ all of $f(y_i)$, $i = 1 \cdots r$
3. $f(x) =$ all of $f(y_i)$, $i = 1 \cdots r$

Consider case (1): Since $f(x) > f(y_i)$ for some i, then for this specific y_i, $p(x, y_i) = \frac{1}{r} \min\left\{1, \frac{f(y)}{f(x)}\right\} = \frac{1}{r}k$ where $k < 1$, hence $p(x, y_i) = a$ where $a < \frac{1}{r}$. Lets assume there was only one y_i such that the above is true. I.e. at least one of the vertices connected to x had $f(y_i) < f(x)$ (if more if found, it will not change the argument). Now we add all the probabilities $p(x, y_i)$ and we found that this sum is $\overbrace{\frac{1}{r} + \frac{1}{r} + \cdots + \frac{1}{r}}^{(r-1)\ vertices} + a$ where the a is for that vertex which had $f(y_i) < f(x)$. Now since $a < \frac{1}{r}$ then this sum will be LESS THAN ONE. But the sum of the one step probability transition from each state must be 1, hence to compensate, we must then have $p(x, x)$ added to make up for the difference. Hence we showed that under case (1) we can find p_{ii} which is not zero. This diagram illustrate this case

[1] In addition, since we showed in part (i) that this chain is an irreducible chain, hence each state communicate with each other state, hence all states must be of the same type since all states are in the same communication class (Theorem 5.3.2). Then if one state is aperiodic, then the all states that communicate with it must also be aperiod (to be of the same type). Hence in an irreduible chain, if one state is aperiodic, then all states are aperiodic as well.



Now we consider case (2). In this case since $f(x) < f(y_i)$ for each i, then $p(x,y_i) = \frac{1}{r}\min\left\{1, \frac{f(y_i)}{f(x)}\right\} = \frac{1}{r}$., then the sum of the probabilities of transitions from x is $\overbrace{\frac{1}{r}+\frac{1}{r}+\cdots+\frac{1}{r}}^{r \text{ vertices}} =$ 1and we do not need to compensate by adding $p(x,x)$ to make up for the deficit. However since now $f(y_i) > f(x)$ then if we view y_i as the x vertex and the x vertex as the y, and consider the probability transitions out of y_i, then we are back to case (1) above. Hence in case (2) as well ,we can find a state in which $p(x,x) > 0$, Hence the chain is aperiodic, and since it is irreducible, then it is regular in this case as well.

Now consider case (3): In this case $f(x) = f(y_i)$ for $i = 1 \cdots r$. In other words, $f(x)$ is CONSTANT. In this case $p(x,y_i) = \frac{1}{r}\min\left\{1, \frac{f(y_i)}{f(x)}\right\} = \frac{1}{r}$,then the sum of the probabilities of transitions from x is $\overbrace{\frac{1}{r}+\frac{1}{r}+\cdots+\frac{1}{r}}^{r \text{ vertices}} =$ 1and we do not need to compensate by adding $p(x,x)$ to make up for the deficit. This will be true for any node. Therefore, it is not possible to find at least one node with the probabilities attached to edges leaving it is less than one. Hence there are no state with $p(x,x) > 0$, hence in this case, the chain is not aperiodic, and hence the chain is NOT regular.

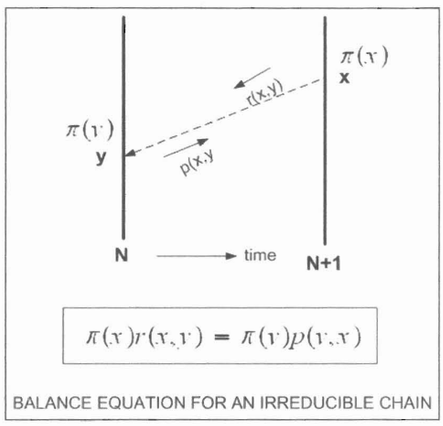
Conclusion: Condition for chain not to be regular is that $f(x)$ be constant.

1.3 Part(iii)

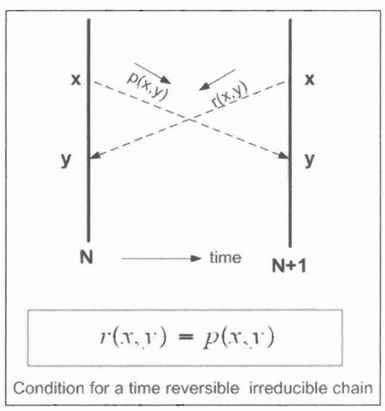
Since the chain is irreducible, then there is a reverse Markov chain (proof is on page 8.1 and 8.2 of lecture notes). Hence for an irreducible chain the balance equations hold

$$r(x,y) = \frac{\pi(y)\,p(y,x)}{\pi(x)} \tag{1}$$

This diagram helps me remember these formulas



Now if the chain the time reversible as well, then $r(x,y) = p(x,y)$,



Then the balance equation (1) becomes

$$\pi(x)\, p(x,y) = \pi(y)\, p(y,x) \tag{2}$$

Hence we need to show that the equation above holds to show the chain is time reversible.

Let the LHS of (2) be $\pi(x)\, p(x,y)$ and let RHS of (2) be $\pi(y)\, p(y,x)$. Then we will show that LHS=RHS for the following 3 cases:

1. $f(x) = f(y)$
2. $f(x) < f(y)$
3. $f(x) > f(y)$

Case(1): Since $f(x) = f(y)$ let these be some value, say z

$$\begin{aligned} LHS &= \pi(x) p(x,y) \\ &= \frac{f(x)}{\sum_{v \in V} f(v)} \left(\frac{1}{r} \min \left\{ 1, \frac{f(y)}{f(x)} \right\} \right) \\ &= \frac{z}{\sum_{v \in V} f(v)} \frac{1}{r} \end{aligned} \tag{3}$$

and

$$\begin{aligned} RHS &= \pi(y) p(y,x) \\ &= \frac{f(y)}{\sum_{v \in V} f(v)} \left(\frac{1}{r} \min \left\{ 1, \frac{f(x)}{f(y)} \right\} \right) \\ &= \frac{z}{\sum_{v \in V} f(v)} \frac{1}{r} \end{aligned} \tag{4}$$

We see that (3) is the same as (4), hence $\boxed{\text{LHS=RHS for case (1)}}$.

case(2): $f(x) < f(y)$

$$\begin{aligned} LHS &= \pi(x) p(x,y) \\ &= \frac{f(x)}{\sum_{v \in V} f(v)} \left(\frac{1}{r} \min \left\{ 1, \frac{f(y)}{f(x)} \right\} \right) \\ &= \frac{f(x)}{\sum_{v \in V} f(v)} \frac{1}{r} \end{aligned} \tag{5}$$

and

$$\begin{aligned} RHS &= \pi(y) p(y,x) \\ &= \frac{f(y)}{\sum_{v \in V} f(v)} \left(\frac{1}{r} \min \left\{ 1, \frac{f(x)}{f(y)} \right\} \right) \\ &= \frac{f(y)}{\sum_{v \in V} f(v)} \frac{1}{r} \frac{f(x)}{f(y)} \\ &= \frac{f(x)}{\sum_{v \in V} f(v)} \frac{1}{r} \end{aligned} \tag{6}$$

Hence we see that (5) is the same as (6). Hence RHS=LHS for case(2).

case (3):$f(x) > f(y)$

$$\begin{aligned} LHS &= \pi(x) p(x,y) \\ &= \frac{f(x)}{\sum_{v \in V} f(v)} \left(\frac{1}{r} \min \left\{ 1, \frac{f(y)}{f(x)} \right\} \right) \\ &= \frac{f(x)}{\sum_{v \in V} f(v)} \frac{1}{r} \frac{f(y)}{f(x)} \\ &= \frac{f(y)}{\sum_{v \in V} f(v)} \frac{1}{r} \end{aligned} \tag{7}$$

and

$$\begin{aligned} RHS &= \pi(y) p(y,x) \\ &= \frac{f(y)}{\sum_{v \in V} f(v)} \left(\frac{1}{r} \min \left\{ 1, \frac{f(x)}{f(y)} \right\} \right) \\ &= \frac{f(y)}{\sum_{v \in V} f(v)} \frac{1}{r} \end{aligned} \tag{8}$$

We see that (7) is the same as (8), hence LHS=RHS for case (3) as well.

Hence we showed the balance equation for the time reversible condition is satisfied. QED.

2 Problem 8.5

8.5 Suppose $G = (V, E)$ is an undirected, connected graph. For each vertex $v \in V$, let $edge(v)$ denote the number of edges that are connected to v. Let f be a positive function defined on V, and let π denote the probability distribution

$$\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)} .$$

(a) Implement the Hastings-Metropolis method to find a regular Markov chain whose limiting distribution is π. Start with the initial irreducible chain defined by

$$q(x, y) = \frac{1}{edge(x)} , \quad \text{whenever } (x, y) \in E .$$

Note that the Markov chain with this one-step transition matrix is traversed by selecting at random one of the edges connected to x, and then making the transition to the corresponding node. $(a1)$ Show the Markov chain is irreducible. $(a2)$ Determine conditions under which the chain is regular. $(a3)$ Show the chain is time reversible with respect to π. (b) Write a MATLAB program that determines the one-step probability matrix resulting from this method. The input to this program is the function f, and the graph, represented by an adjacency matrix. An adjacency matrix is an $n \times n$ matrix, where n is the number of nodes in the graph, and where entry (i, j) is one if there is an edge connecting nodes i and j, and is zero otherwise. Use this adjacency matrix to compute the function $edge(v)$ at each node. Apply your program to the graph $G = (V, E)$, where $V = \{1, 2, 3, 4\}$, and $E = \{(1, 2), (1, 3), (2, 3), (2, 4), (3, 4)\}$,

and where $f(1) = 2$, $f(2) = 8$, $f(3) = 6$, and $f(4) = 4$. Verify (using MATLAB) that the resulting chain is regular and has the required limiting state probability distribution.

2.1 Part(a)

The following is the Hastings-Metrpolois algorithm implementation.

This algorithm generates a time-reversible M.C. (referred to as p in the lecture notes) given an irreducible M.C. (called q or the original chain) and given a stationary distribution π for that chain.

Input: $f(x)$ defined over the states x, and $edge(x)$ which represents the number of edges connected to x

1. For each state x calculate $\pi(x) = \frac{f(x)}{\sum_{v \in V} f(v)}$ and for each state x calculate $edge(x)$
2. compute $q(x,y) = \frac{1}{edge(x)}$ whenever $edge(x) \neq 0$ else set $q(x,y) = 0$
3. Select a state x by random to start from.
4. Let $n = 1$ and let $X_1 = x$
5. Let S be the set of all states that can be reached in one step from x. These will be the states y in which $q(x,y) \neq 0$
6. Select a state y from S by random (using a uniform $U[0,1]$ random number generator)
7. Calculate $\beta(x,y) = \min\left\{1, \frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}\right\}$
8. Generate a random number u from $U[0,1]$
9. Let $n = n + 1$
10. Compare u to $\beta(x,y)$.
11. IF $u < \beta(x,y)$ THEN $X_n = y$ (select the new state) ELSE $X_n = X_{n-1}$ (stay in same state) ENDIF
12. Let $x = X_n$
13. If $n >$ some Max number of iterations or if we reached some convergence limit Then go to 15
14. GOTO 5
15. Algorithm is complete. Now generate the time reversible MC as follows
    (a) Scan the state path generate X_n and count how many times state x switches to state y in one step
    (b) Do the above for all the states x
    (c) Divide the above number by the total number of steps made to generate $p(x,y)$

Since the problem now asks to *implement* Hastings-Metropolis, then I used the data given at the end of the problem and implemented the above simulation using that data[2]. Please see appendix for code and final P matrix generated.

[2]I allready had the code for the simulation written, just needed to feed the new data for this problem.

2.1.1 Part (a1)

This is similar the problem 8.4 part(I). To show that the p (final M.C.) is irreducible, we need to show that there exist no closed proper subsets. Since the graph G is connected, then we just need to show whenever there is an edge between vertex x and y then there corresponds in the chain representation of the final p matrix a non-zero $p(x,y)$ and also a non-zero $p(y,x)$. This will insure that the each state can transition to each other state, just as each vertex can be visited from each other vertex (since it is a connected graph).

Let us consider any 2 vertices say x, y with a direct edge between them (this is the only case we need to consider due to the argument above). We need to show the resulting $p(x,y)$ and $p(y,x)$ are non-zero

Consider $p(x,y)$ first. Since

$$\begin{aligned} p(x,y) &= q(x,y)\,\beta(x,y) \\ &= \frac{1}{edge(x)} \min\left\{1, \frac{\pi(y)\,q(y,x)}{\pi(x)\,q(x,y)}\right\} \\ &= \frac{1}{edge(x)} \min\left\{1, \frac{\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}}{\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}}\right\} \end{aligned}$$

This is what is asked for in part (a) --

Hence

$$\boxed{p(x,y) = \frac{1}{edge(x)} \min\left\{1, \frac{f(y)edge(x)}{f(x)edge(y)}\right\}} \tag{1}$$

Then it is clear that whenever there is an edge between x, y then $p(x,y) \neq 0$ since both $f(x)$ and $f(y)$ are positive (not zero) and also edge(x) and edge(y) are non-zero as well. Hence we see that $p(x,y) \neq 0$. Similar argument shows that $p(y,x) \neq 0$.

This shows that M.C. represented by P is irreducible.

2.1.2 Part (a2)

The condition for regular chain P is that there exist at least one state x such that $p(x,x) > 0$.From (1) above we can decide under what conditions this will occur.

Consider a vertex x with $edge(x)$ edges from it connected to vertices $y_1, y_2, \cdots, y_r$. Then from (1) we see that

$$\begin{aligned} p(x,y_i) &= \frac{1}{edge(x)} \min\left\{1, \frac{f(y_i)\,edge(x)}{f(x)\,edge(y_i)}\right\} \\ &= \frac{1}{edge(x)} \min\left\{1, \frac{\frac{f(y_i)}{edge(y_i)}}{\frac{f(x)}{edge(x)}}\right\} \end{aligned}$$

The condition for having $p(x,x) > 0$ is that $\min\left\{1, \frac{\frac{f(y_i)}{edge(y_i)}}{\frac{f(x)}{edge(x)}}\right\} < 1$, since this will cause $p(x,y_i)$ to be some quantity less than $\frac{1}{r}$ and so when summing over all r there will be a deficit in the sum and we have to compensate for it to make it 1 by adding $p(x,x)$. But for $\min\left\{1, \frac{\frac{f(y_i)}{edge(y_i)}}{\frac{f(x)}{edge(x)}}\right\}$ to be less than ONE means that $\boxed{\frac{f(y_i)}{edge(y_i)} < \frac{f(x)}{edge(x)}}$

Hence the condition for finding an Aperiodic state is finding a vertex x such that the above holds for one of the vertices y_i this vertex is directly connected to. For example, if y_i had the same number of edges from it as does x, then the condition will be that $f(y_i) < f(x)$. And if y_i has less or more edges from it than x has, then we need the ratio $\frac{f(y_i)}{edge(y_i)}$ to be less than $\frac{f(x)}{edge(x)}$.

The above is the same as saying $\frac{f(x)}{edge(x)}$ must be constant for the p not to be regular.

2.1.3 Part(A3)

Since the chain is irreducible, then there is a reverse Markov chain (proof is on page 8.1 and 8.2 of lecture notes). Hence for an irreducible chain the balance equations hold

$$r(x,y) = \frac{\pi(y)\, p(y,x)}{\pi(x)} \tag{2}$$

Now if the chain the time reversible as well, then $r(x,y) = p(x,y)$, Then the balance equation (1) becomes

$$\begin{aligned}
\pi(x)\, p(x,y) &= \pi(y)\, p(y,x) \\
\frac{f(x)}{\sum_{v\in V} f(v)} q(x,y)\, \beta(x,y) &= \frac{f(y)}{\sum_{v\in V} f(v)} q(y,x)\, \beta(y,x) \\
\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)} \left(\min\left\{1, \frac{f(y)\, edge(x)}{f(x)\, edge(y)}\right\}\right) &= \frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)} \left(\min\left\{1, \frac{f(x)\, edge(y)}{f(y)\, edge(x)}\right\}\right) \\
\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)} \left(\min\left\{1, \frac{\frac{f(y)}{edge(y)}}{\frac{f(x)}{edge(x)}}\right\}\right) &= \frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)} \left(\min\left\{1, \frac{\frac{f(x)}{edge(x)}}{\frac{f(y)}{edge(y)}}\right\}\right)
\end{aligned} \tag{3}$$

Hence we need to show that the equation (3) above holds to show the chain is time reversible.

There are 3 cases to consider:

1. $\frac{f(y)}{edge(y)} = \frac{f(x)}{edge(x)}$
2. $\frac{f(y)}{edge(y)} < \frac{f(x)}{edge(x)}$

3. $\frac{f(y)}{edge(y)} > \frac{f(x)}{edge(x)}$

For case (1), LHS of equation (3) simplifies to $\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}$ and the RHS of (3) simplifies to $\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}$, but since $\frac{f(y)}{edge(y)} = \frac{f(x)}{edge(x)}$, then LHS=RHS.

Hence balance equation (3) is satisfied for case (1).

For case(2), LHS of (3) simplifies $\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)} \left(\frac{\frac{f(y)}{edge(y)}}{\frac{f(x)}{edge(x)}} \right) = \frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}$ and RHS of (3) simplifies to $\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)}$,then LHS=RHS.

Hence balance equation (3) is satisfied for case (2).

For case (3), LHS of (3) simplifies $\frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}$ and RHS of (3) simplifies to $\frac{f(y)}{\sum_{v\in V} f(v)} \frac{1}{edge(y)} \left(\frac{\frac{f(x)}{edge(x)}}{\frac{f(y)}{edge(y)}} \right) = \frac{f(x)}{\sum_{v\in V} f(v)} \frac{1}{edge(x)}$,then LHS=RHS.

Hence balance equation (3) is satisfied for case (3).

Hence in all 3 cases we showed the balance equation is satisfied.

Hence M.C. is time reversible.

2.2 Part(b)

A small program written to construct the P matrix directly following instructions on page 8.4 of lecture notes. The following is the resulting P matrix

$$\begin{pmatrix} 0. & 0.5 & 0.5 & 0. \\ 0.0625 & 0.4375 & 0.25 & 0.25 \\ 0.0833333 & 0.333333 & 0.25 & 0.333333 \\ 0. & 0.5 & 0.5 & 0. \end{pmatrix}$$

Now to check that the final chain P is regular, it was raised to some high power to check that all entries in the $P^m > 0$. This is the result

```
In[17]:= MatrixPower[p, 50] // N // MatrixForm

Out[17]//MatrixForm=
0.0526316 0.421053 0.315789 0.210526
0.0526316 0.421053 0.315789 0.210526
0.0526316 0.421053 0.315789 0.210526
0.0526316 0.421053 0.315789 0.210526
```

The above verifies that the final matrix p is regular.

Using the Hastings-Metropolis simulation algorithm, the convergence to the above matrix was slow. Had to make 2 million observation to be within 3 decimal points from the above. Here is the P matrix generated from Hastings algorithm for $N = 2,000,000$

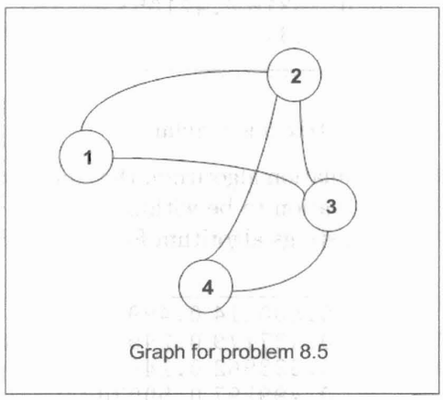
```
0.          0.500114 0.499886 0.
0.0625897   0.437179 0.249784 0.250448
0.0831875   0.333962 0.248524 0.334326
0.          0.499297 0.500703 0.
```

each row should be π:

$$\pi = \frac{1}{20}\cdot(2, 8, 6, 4)$$

3 Appendix (Implementation of part(a) and part(b))

The graph for part(a) and part(b) is the following



Graph for problem 8.5

Hastings - Metropolis Algorithm implementation For Problem 8.5 part(a)

This below is an implementation of the Hastings - Metropolis algorithm. A simple GUI interface allows the user to specify the number of steps to run the algorithm for. At each step, the current P matrix and the current calculated stationary distribution for this P matrix are shown to help observe the convergence.

The input to this run below is that of problem 8.5 part(b)

Few seed the random number generator and display the q and the π distribution used

```
SeedRandom[121 212];
```

Define the data given in the problem

```
v = {1, 2, 3, 4};
edges = {{1, 2}, {1, 3}, {2, 3}, {2, 4}, {3, 4}};
f = {1, 8, 6, 4};
```

Define the functions π(x) and q(x,y) to use in the implementation

```
pi[x_, f_] := f[[x]] / Sum[f[[i]], {i, 1, Length[f]}]

q[x_, y_] := Module[{r},
  r = Count[edges, {x, any_}] + Count[edges, {any_, x}];
  If[(Count[edges, {x, y}] > 0 || Count[edges, {y, x}] > 0), 1 / r, 0]
 ]
```

Find the stationary distribution

```
w = Table[pi[x, f], {x, 1, v[[-1]]}]
```

$$\left\{\frac{1}{19}, \frac{8}{19}, \frac{6}{19}, \frac{4}{19}\right\}$$

Generate the original q matrix

```
(originalMatrix = Table[q[x, y], {x, 1, v[[-1]]}, {y, 1, v[[-1]]}]) // MatrixForm
```

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

Bring up the user interface

```
m = Manipulate[First@{x = hastings[maxN]; Grid[{{"stationary distribution w=", N[w]},
      { "current stationary distribution=", MatrixPower[N[x], 100][[1, All]]},
      {, }, {"Current P Matrix=", N[MatrixForm[x]]}}, Alignment → Left]},
  {{maxN, 1, "number of steps"}, 1, 2 000 000, 100, ContinuousAction → False,
   Appearance → "Labeled"},
  AutorunSequencing → {{1, 300}}
 ]
```

number of steps 273401

stationary distribution w= {0.0526316, 0.421053, 0.315789, 0.210526}

current stationary distribution= {0.053453, 0.420567, 0.316001, 0.209979}

Current P Matrix=

$$\begin{pmatrix} 0. & 0.502771 & 0.497229 & 0. \\ 0.0637317 & 0.435999 & 0.251526 & 0.248743 \\ 0.0843336 & 0.331767 & 0.250466 & 0.333434 \\ 0. & 0.502369 & 0.497631 & 0. \end{pmatrix}$$

Define a function for cummulative sum

```
cumSum[list_] := Module[{i, sum, s, k},
  sum = 0;
  k = Length[list];
  s = Table[0, {k}];
  For[i = 1, i ≤ k, i++,
   {
    sum = sum + list[[i]];
    s[[i]] = sum;
   }
  ];
  s
 ]
```

Function to calculate β (x, y)

```
beta[x_, y_, pi_, q_] := Module[{},
  Min[1, (pi[[y]] q[[y, x]]) / (pi[[x]] q[[x, y]])]
 ]
```

Function called at the end of the run to generate P from the path of states travelled

```
generatePMatrixFromStatePath[nStates_, x_] := Module[{i, j, p, allPairs, n, m},
  n = Length[x];
  (*Print["X=",x];*)
  p = Table[0, {nStates}, {nStates}];
  allPairs = Partition[x, 2, 1];
  For[i = 1, i ≤ nStates, i++,
   {
    m = Count[allPairs, {i, y_}];
    For[j = 1, j ≤ nStates, j++,
     If[m ≠ 0, p[[i, j]] = Count[allPairs, {i, j}] / m, p[[i, j]] = 0]
    ]
   }
  ];

  p
 ]
```

Function to sample from q using uniform distribution

```
sampleFromQConditional[q_, x_] := Module[{s, found, j, k, sample, y},
  s = Flatten[Position[q[[x, All]], Except[0], 1, Heads → False] ];
  sample = q[[x, s]];
  sample = cumSum[sample];
  y = RandomReal[];
  found = False;
  For[j = 1, j ≤ Length[sample], j++,
   If[ Not[found], If[y ≤ sample[[j]], {k = j; found = True}]]
  ];

  y = s[[k]]
 ]
```

The Hastings algorithm main loop

```
hastings[maxN_] := Module[{i, j, nStates, n, s, y, α, u, x, sample, pts, sum, k, found},
  nStates = Length[originalMatrix];
  n = 1;
  x = Table[0, {maxN}];
  x[[n]] = 1;  (*pick any state to start from*)
  i = 1;
  While[i < maxN,
   {
    y = sampleFromQConditional[originalMatrix, x[[n]]];
    α = beta[x[[n]], y, w, originalMatrix];
    u = RandomReal[];
    n++;
    If[u ≤ α, x[[n]] = y, x[[n]] = x[[n - 1]]]; (*acceptance step*)
    i++;
   }
  ];

  generatePMatrixFromStatePath[nStates, x]
 ]
```

Problem 8.5 part(b)

by Naser Abbasi. Mathematics 504, Spring 2008. CSUF

This below construct the P matrix using direct computation of the Hasting-Meropolis method

Define the data given in the problem

```
v = {1, 2, 3, 4};
edges = {{1, 2}, {1, 3}, {2, 3}, {2, 4}, {3, 4}};
f = {1, 8, 6, 4};
```

Define the functions π(x) to use in the implementation

```
pi[x_, f_] := f[[x]] / Sum[f[[i]], {i, 1, Length[f]}]
```

Define the functions q(x,y) to use in the implementation. This does something similar to the adjcancy matrix normally used. I used the Count[] function in *Mathematica* which automatically counts the edges from the edges list above, so there is really no need to construct an adjancy matrix as such.

```
q[x_, y_] := Module[{r},
  r = Count[edges, {x, any_}] + Count[edges, {any_, x}];
  If[(Count[edges, {x, y}] > 0 || Count[edges, {y, x}] > 0), 1 / r, 0]
 ]
```

Find the stationary distribution

```
w = Table[pi[x, f], {x, 1, v[[-1]]}]
```

$$\left\{\frac{1}{19}, \frac{8}{19}, \frac{6}{19}, \frac{4}{19}\right\}$$

Generate the original q matrix

```
(originalMatrix = Table[q[x, y], {x, 1, v[[-1]]}, {y, 1, v[[-1]]}]) // MatrixForm
```

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

define Beta function

```
beta[x_, y_, pi_, q_] := Module[{},
  Min[1, pi[[y]] q[[y, x]] / pi[[x]] q[[x, y]]]
 ]
```

define a function to calculate a non - diagonal entry in the P matrix

```
makeEntryInPMatrix[x_, y_, pi_, q_] := Module[{},
  If[q[[x, y]] == 0, 0, q[[x, y]] beta[x, y, pi, q]]
 ]
```

Construct the P matrix for the off-diagonal elements only

```
nStates = Length[originalMatrix];
p = Table[0, {nStates}, {nStates}];

For[i = 1, i ≤ nStates, i++,
  For[j = 1, j ≤ nStates , j++,
   If[i ≠ j, p[[i, j]] = makeEntryInPMatrix[i, j, w, originalMatrix]]
  ]
 ];
```

Now calculate the diagonal elements of the P matrix

```
For[i = 1, i ≤ nStates, i++,
  For[j = 1, j ≤ nStates , j++,
   If[i == j, p[[i, j]] = 1 - Total[p[[i, All]]]]
  ]
 ];
```

Print the P matrix

```
p // N // MatrixForm
```

$$\begin{pmatrix} 0. & 0.5 & 0.5 & 0. \\ 0.0625 & 0.4375 & 0.25 & 0.25 \\ 0.0833333 & 0.333333 & 0.25 & 0.333333 \\ 0. & 0.5 & 0.5 & 0. \end{pmatrix}$$

Raise the final p matrix to some large power to verify it is regular

```
MatrixPower[p, 50] // N // MatrixForm
```

$$\begin{pmatrix} 0.0526316 & 0.421053 & 0.315789 & 0.210526 \\ 0.0526316 & 0.421053 & 0.315789 & 0.210526 \\ 0.0526316 & 0.421053 & 0.315789 & 0.210526 \\ 0.0526316 & 0.421053 & 0.315789 & 0.210526 \end{pmatrix}$$

4.13 Wed 5/7/2008

Grade: 4/4.

Problems 10.5 and 10.6 These deal with continues time markov chains. To determine rate of arrival and departure for birth/death process

4.13.1 Problem 10.5

> 10.5 Consider a set of $m+n$ machines which fail independently of each other at an exponential rate λ. It is intended that m machines are to be in operation at any time. The remaining machines serve as spares and are called into operation when an operating machine fails. If more than n machine are in a state of failure, then all the operational machines will be in service. Suppose there are s, $1 \leq s \leq m$, repair persons that service the machines independently and each at exponential rate μ. Let $X(t)$ denote the number of machines at time t that are *not* operational; that is, they are in the repair shop. Determine the arrival and departure rates for the birth-and-death model.

SOLUTION

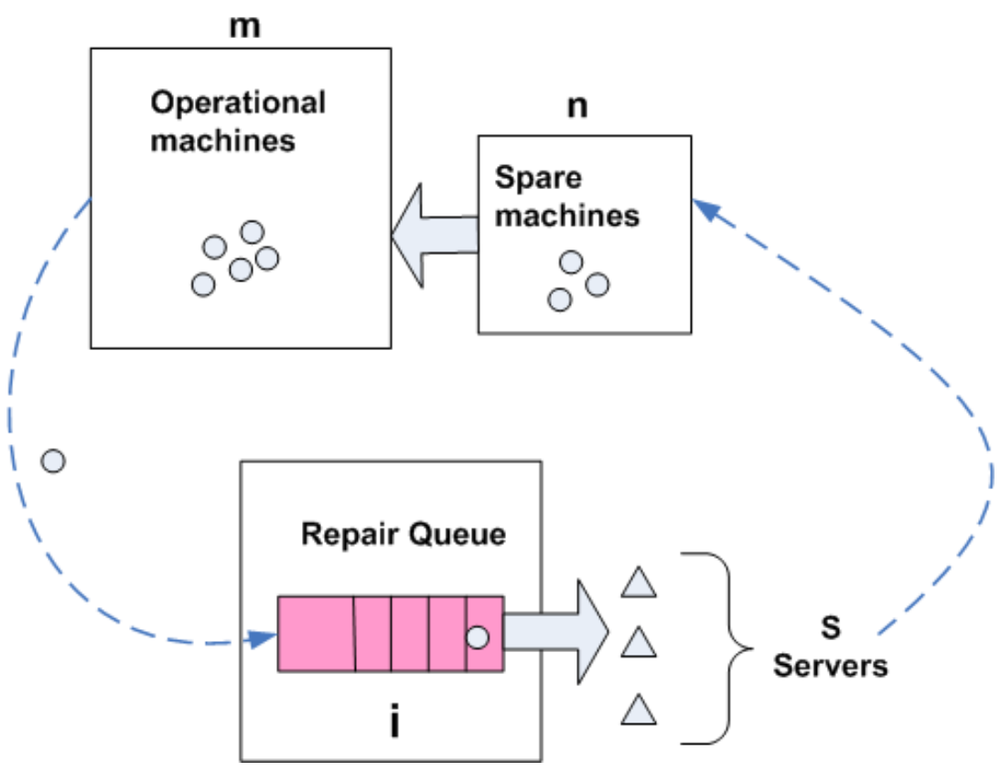


Figure 4.3: Illustrating model diagram for problem 10.5

In the above, i is the number of broken machines in the queue. m is maximum capacity of the operating room. The goal is to keep this room filled to its capacity. In other words, to keep m machines in operations. n is the capacity of the spare room.

Calculating arrival rates:

Need to determine $p_{i,i+1}(h)$. This can happen when one machine fails, but no server completes its service meanwhile. Hence we do not need to consider the servers part in this analysis. There are 2 cases to consider:

1. $i \leq n$ (there are m machines in operations)

$$p_{i,i+1}(h) = \binom{m}{1} (\lambda h + o(h)) (1 - \lambda h + o(h))^{m-1} + o(h)$$

In the above, the last term $o(h)$ accounts for other possible conditions under which i can increase by one but which is considered to be less likely, such as 2 machines break down and one server completes its service. In the above, $(1 - \lambda h + o(h))^{m-1}$ simplifies to zero when h is very small, hence the above equation becomes

$$p_{i,i+1}(h) = m\lambda h + o(h)$$

Comparing the above with the Hence we see $p_{ij}(h) = q_{ij}h + o(h)$ we see that $q_{i,i+1} = m\lambda$ or in other words,

$$\boxed{\lambda_i = m\lambda \qquad i \leq n}$$

2. $n < i \leq n + m$ (there are less than m machines in operations)

$$\begin{aligned} p_{i,i+1}(h) &= \binom{m-(i-n)}{1} (\lambda h + o(h)) (1 - \lambda h + o(h))^{m-(i-n)-1} + o(h) \\ &= (m + n - i)\lambda h + o(h) \end{aligned}$$

Hence we see that $q_{i,i+1} = (m + n - i)$ or in other words,

$$\boxed{\lambda_i = (m + n - i) \qquad n < i \leq n + m}$$

Calculating departure rates:

Need to determine $p_{i,i-1}$, this can happen when a server completes its job but no machine fails meanwhile, Hence we only need to consider the servers. There are 2 cases to consider:

1. $1 \leq i < s$ (Queue is empty and not all servers at working on fixing machines at hand)

$$\begin{aligned} p_{i,i-1}(h) &= \binom{i}{1} (\mu h + o(h)) (1 - \mu h + o(h))^{i-1} + o(h) \\ &= i\mu h + o(h) \end{aligned}$$

Hence $q_{i,i-1} = i\mu$, or since this is a birth/death process, we write

$$\mu_i = i\mu \qquad 1 \leq i < s$$

2. $s \leq i$ (All servers at busy)

$$\begin{aligned} p_{i,i-1}(h) &= \binom{s}{1} (\mu h + o(h)) (1 - \mu h + o(h))^{s-1} (1 - \lambda h + o(h))^{m-(i-n)} + o(h) \\ &= s\mu h + o(h) \end{aligned}$$

Hence $q_{i,i-1} = s\mu$, Hence

$$\mu_i = s\mu \qquad s \leq i$$

Therefore, we summarize all the above as follows

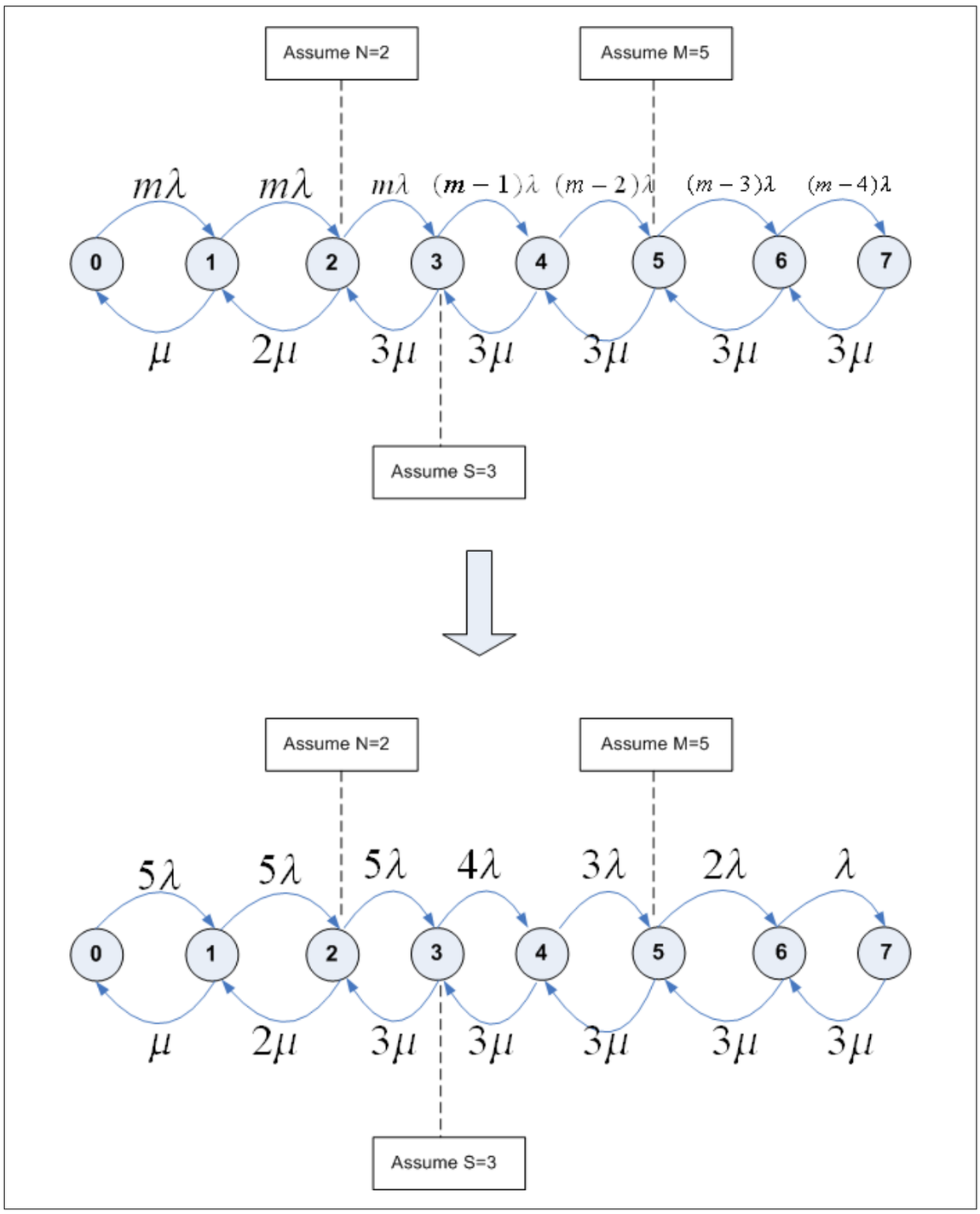
Arrival rate $\lambda_i = m\lambda$ for $i \leq n$ and $\lambda_i = (m + n - i)\ \lambda$ for $\quad n < i \leq n + m$.

Departure rate $\mu_i = i\mu$ for $\ 1 \leq i < s$ and $\mu_i = s\mu$ for $s \leq i$

Notice that arrival rate does not depend on the number of servers s.

The following state transition diagram illustrates the above result, with arrows leaving/entering states show the rate of arrival and departure on them per the above result. To make the diagram easier to make, I assume the following values: $s = 3, m = 5, n = 2$

Notice that $\mu_0 = 0$ and $\lambda_{n+m} = 0$ as expected.



Now compute the steady state distribution π (This is not asked for in this problem, but need to do this to solve problem 12.3 later on and implement it)

Starting with the balance equation, where to balance the rate out of a state, with the rate into a state. We have

$$\pi_j v_j = \sum_{k \neq j} q_{kj} \pi_k$$

Hence for state $i = 0$ we have

$$\pi_0 v_0 = q_{1,0}\pi_1$$

But $v_0 = \lambda_0$ and $q_{1,0} = \mu_1$ hence

$$\pi_0 \lambda_0 = \mu_1 \pi_1 \tag{1}$$

For state $i = 1$ we have

$$\pi_1 v_1 = q_{0,1}\pi_0 + q_{2,1}\pi_2$$

but $v_1 = \lambda_1 + \mu_1$, $q_{0,1} = \lambda_0, q_{2,1} = \mu_2$,hence the above becomes

$$\begin{aligned}\pi_1(\lambda_1 + \mu_1) &= \lambda_0\pi_0 + \mu_2\pi_2 \\ \pi_1\lambda_1 + \pi_1\mu_1 &= \lambda_0\pi_0 + \mu_2\pi_2\end{aligned}$$

But from (1) we have $\mu_1\pi_1 = \lambda_0\pi_0$, hence the above becomes

$$\begin{aligned}\pi_1\lambda_1 + \lambda_0\pi_0 &= \lambda_0\pi_0 + \mu_2\pi_2 \\ \pi_1\lambda_1 &= \mu_2\pi_2\end{aligned} \tag{2}$$

Continue this way, we obtain that

$$\pi_i\lambda_i = \mu_{i+1}\pi_{i+1} \qquad\qquad i = 0, 1, 2, \cdots, n+m$$

From the above, if we solve in terms of π_0 we obtain that

$$\pi_i = \frac{\lambda_0\lambda_1\cdots\lambda_{i-1}}{\mu_1\mu_2\cdots\mu_i}\pi_0 \qquad\qquad i = 0, 1, 2, \cdots, n+m \tag{3}$$

and with the equation $\pi_0 + \pi_1 + \cdots + \pi_{n+m} = 1$ we can now solve for all π_i as follows

$$\begin{aligned}\pi_0 &= 1 - (\pi_1 + \cdots + \pi_{n+m}) \\ &= 1 - \left(\frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \cdots + \frac{\lambda_0\lambda_1\cdots\lambda_{n+m-1}}{\mu_1\mu_2\cdots\mu_{n+m}}\right)\pi_0\end{aligned}$$

Hence

$$\pi_0\left(1 + \left(\frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \cdots + \frac{\lambda_0\lambda_1\cdots\lambda_{n+m-1}}{\mu_1\mu_2\cdots\mu_{n+m}}\right)\right) = 1$$

$$\pi_0 = \frac{1}{1 + \left(\frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \cdots + \frac{\lambda_0\lambda_1\cdots\lambda_{n+m-1}}{\mu_1\mu_2\cdots\mu_{n+m}}\right)}$$

Now that π_0 is found, we can find the remaining π_i using (3)

4.13.2 Problem 10.6

> 10.6 Consider a sign that contains N light bulbs, each with a lifetime that follows an exponential distribution with parameter λ. Assume that the bulbs function independently of each other. Suppose it is the policy to allow bulbs to burn out until the moment the r-th bulb expires, and to then replace all burned out bulbs at that time. Define the state of the system $X(t)$ to be the number of burned out bulbs at time t. Argue that this stochastic process (a) has the Markov property, (b) is stationary, and (c) can be represented as a pure jump process. Determine the parameters v_i and q_{ij} of the jump process. (d) Determine the balance equations for this system, and find the long-run probability distribution for the states.

Review of the problem setup: Imagine there is a queue of length r. Burned out bulbs enter the queue (with inter-arrival time which is a random variable distributed as an exponential with rate λ). Bulbs continue to enter the queue until the queue is full, then at that moment we imagine a single server processing the bulbs in the queue all at once and immediately all r bulbs become operational again and the queue is now empty. This process repeats again and again.

PART A

A stochastic process $X(t)$ is defined to have the Markov property if its transition to the next state depends only on the current state and not on any earlier states. In other words it satisfies the following

$$\Pr\{X(s+t)=j|X(s)=i,X(u)\ \text{for any}\ u<i\}=\Pr\{X(s+t)=j|X(s)=i\}$$

In this problem $X(t)$ is the number of burned out bulbs in the queue at any time t. When $X(t)<r$ then $X(t)$ can be viewed as a counting process (or pure birth process) or a Poisson process (until the queue become full).

Therefore, The time between each successive events (where an event causes the count to increase by one) is a random variable with exponential distribution (we are also given this fact in the problem). But the exponential distribution is memoryless[7] by definition. Therefore it does not depend on clock time but only on the length of the time interval. Hence the process satisfies the Markov property.

[7] A memorless random variable X is one in which $\Pr(X>t+h|X>t)=\Pr(X>h)$

PART B

A stochastic process $X(s)$ is defined to be stationary[8] if its state transition $p_{ij}(t)$ do not depend on when the transitions happen but only on the time interval t. In other words, random process $X(s)$ is stationary if

$$\Pr\{X(s+t)=j|X(s)=i\}=\Pr\{X(c+t)=j|X(c)=i\}$$

For any $c,s\geq 0$. So, letting $c=0$, the system is stationary if

$$\Pr\{X(s+t)=j|X(s)=i\}=\Pr\{X(t)=j|X(0)=i\}$$

This process is clearly stationary, since it is a counting process (when $X(t)\leq r-1$). A counting Poisson process is stationary since it does not depend on clock time as was argued in part (A). To show this more clearly, since this is a counting process, then by definition of the Poisson process

$$\Pr\left(X(s+t)-X(s)=n\right)=e^{-\lambda t}\frac{(\lambda t)^n}{n!}\qquad\qquad n=0,1,\cdots,r-1$$

We see that the probability of $X=n$ does not depend on s and depends only on the time interval t. If this was a non-stationary process, then s would appear in the RHS above. I.e. the probability of the random variable would depend on clock time, but we see from the above definition that it does not.

PART(C)

A stochastic process is a pure jump process if the transition probabilities can be written as

$$p_{ii}(h)=1-v_ih+o(h)\text{ and }p_{ij}(h)=q_{ij}h+o(h)\text{ as }h\to 0^+$$

In this problem $p_{ii}(h)$ is the probability than no bulb burns out during an interval h. This is given by the probability than no bulb burns out from the current number of functional bulbs which is $N-i$. Due to independence, we obtain

$$p_{ii}(h)=(1-\lambda h+o(h))^{(N-i)}$$

Applying Binomial expansion $(a+b)^n=\sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k$, to the above, and taking $a=1,b=-\lambda h+o(h),n=N-i$ we obtain

$$\begin{aligned}p_{ii}(h)&=1+(N-i)\left(-\lambda h+o(h)\right)+\text{higher order terms in }o(h)\text{ which can be ignored when }h\to 0^+\\&=1-N\lambda h+i\lambda h+o(h)\\&=1-h\lambda(N-i)+o(h)\end{aligned}$$

Hence we can write $p_{ii}(h)=1-v_ih+o(h)$ where $\boxed{v_i=\lambda(N-i)}$

Now $p_{ij}(h)$ is the probability that there will be j failed bulbs after h units of time given that there is already i failed bulbs. For this to occur, then we need to have $j-i$ bulbs fail in h units of time. We can solve for the general case when $j-i>1,$ but since we will let $h\to 0^+$ it is most likely that there will be only one event occur (one bulb fail) during this time, and we can collect all other less likely probabilities in the $o(h)$ term. Hence we will only consider $p_{i,i+1}$ in the following.

$$\begin{aligned}p_{i,i+1}(h)&=\binom{N-i}{1}\left(\lambda h+o(h)\right)\left(1-\lambda h+o(h)\right)^{N-i-1}+o(h)\\&=(N-i)\,\lambda h+o(h)\end{aligned}$$

[8]In Linear system theory, the term time-invariant is used.

Therefore from $p_{ij}(h) = q_{ij}h + o(h)$ we see that $q_{i,i+1} = (N-i)\,\lambda$ i.e. $\boxed{\lambda_i = (N-i)\,\lambda \qquad i = 0, 1, \cdots r-1}$

Hence the Q matrix (The rate matrix) is

$$Q = \begin{bmatrix} & 0 & 1 & 2 & \cdots & \text{r-1} \\ 0 & -\lambda N & \lambda N & 0 & 0 & 0 \\ 1 & 0 & -\lambda(N-1) & \lambda(N-1) & 0 & 0 \\ 2 & 0 & 0 & -\lambda(N-2) & \lambda(N-1) & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \text{r-1} & \lambda(N-r+1) & & & & -\lambda(N-r+1) \end{bmatrix}$$

PART (D)

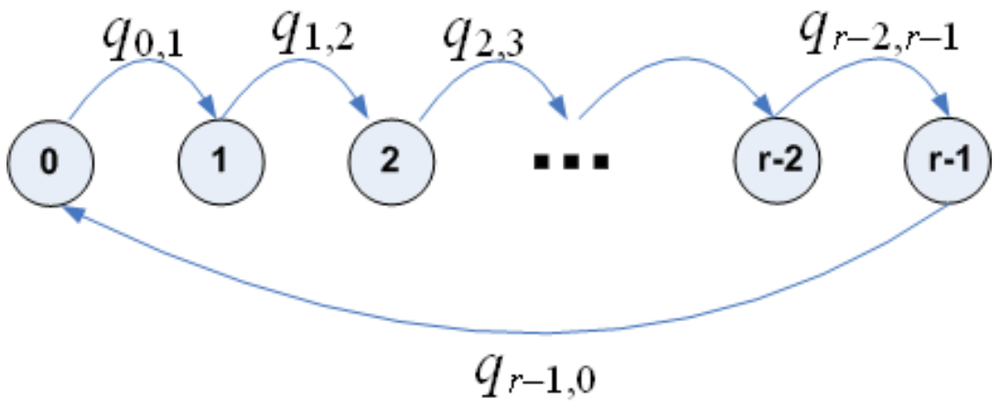
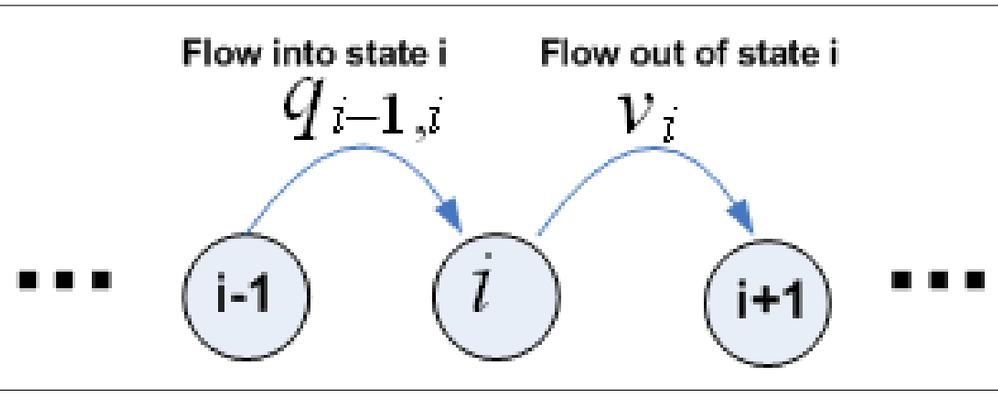


Figure 4.4: rate flow diagram for problem 10.6

The balance equation can be obtained from balancing the flow out rate of a state i (which is given by v_i) by all the flow in rate into the state which is given by $\sum_{v \neq i} q_{v,i}$ as illustrated below for the above problem



Hence we write

$$v_i \pi_i = q_{i-1,i} \pi_{i-1} \qquad i = 1, 2, \cdots r-1$$

and for state $i = 0$ we have

$$v_0 \pi_0 = q_{r-1,0} \pi_{r-1}$$

Therefore we obtain

$$\begin{aligned}\lambda(N-i)\,\pi_i &= (N-i+1)\,\lambda\pi_{i-1} && i=1,2,\cdots r-1\\ \lambda N\pi_0 &= (N-r+1)\,\lambda\,\pi_{r-1} && i=0\end{aligned}$$

Hence we have for $i=1,2,\cdots r-1$

$$\begin{aligned}\lambda(N-1)\,\pi_1 &= N\lambda\pi_0\\ \lambda(N-2)\,\pi_2 &= \lambda(N-1)\,\pi_1\\ \lambda(N-3)\,\pi_3 &= \lambda(N-2)\,\pi_2\\ &\vdots\\ \lambda(N-r+1)\,\pi_{r-1} &= \lambda(N-r)\,\pi_{r-2}\end{aligned}$$

Therefore we have

$$\begin{aligned}\pi_1 &= \frac{N}{N-1}\pi_0\\ \pi_2 &= \frac{N-1}{N-2}\pi_1\\ \pi_3 &= \frac{N-2}{N-3}\pi_2\\ &\vdots\\ \pi_{r-1} &= \frac{N-r}{N-r+1}\pi_{r-2}\end{aligned}$$

back substitute, we obtain

$$\begin{aligned}\pi_1 &= \frac{N}{N-1}\pi_0\\ \pi_2 &= \frac{N-1}{N-2}\frac{N}{N-1}\pi_0\\ \pi_3 &= \frac{N-2}{N-3}\frac{N-1}{N-2}\frac{N}{N-1}\pi_0\\ &\vdots\\ \pi_{r-1} &= \frac{N-r}{N-r+1}\frac{N-r-1}{N-r}\cdots\frac{N}{N-1}\pi_0\end{aligned}$$

Hence

$$\begin{aligned}\pi_1 &= \frac{N}{N-1}\pi_0\\ \pi_2 &= \frac{N}{N-2}\pi_0\\ \pi_3 &= \frac{N}{N-3}\pi_0\\ &\vdots\\ \pi_{r-1} &= \frac{N}{N-r+1}\pi_0\end{aligned}$$

We notice that the last equation above, is the same as for the case $i=0$. Hence we have one of the r equations duplicated. Hence we need one more equation to solve for the unknowns π_i and for that we use $\sum_{i=0}^{r}\pi_i = 1$

Therefore, the general expression for π_i is

$$\pi_i = \frac{N}{N-i}\pi_0 \qquad\qquad i=1,2,\cdots,r-1 \tag{1}$$

Now since $\pi_0+\pi_1+\cdots+\pi_{r-1}=1$, then we write

$$\begin{aligned}
\pi_0 + \sum_{i=1}^{r}\frac{N}{N-i}\pi_0 &= 1\\
\pi_0\left(1+N\sum_{i=1}^{r}\frac{1}{N-i}\right) &= 1\\
\pi_0 &= \frac{1}{1+N\sum_{i=1}^{r}\frac{1}{N-i}}
\end{aligned} \tag{2}$$

So now that we know π_0 from (2), we substitute it into (1) and solve for the remaining π_i

$$\begin{aligned}
\pi_i &= \frac{N}{N-i}\frac{1}{\left(1+N\sum_{i=1}^{r}\frac{1}{N-i}\right)}\\
&= \frac{N}{(N-i)+(N-i)\,N\sum_{i=1}^{r}\frac{1}{N-i}}
\end{aligned} \tag{3}$$

But

$$\sum_{i=1}^{r}\frac{1}{N-i} = \frac{1}{N-1}+\frac{1}{N-2}+\cdots+\frac{1}{N-r} = \sum_{k=1}^{N-1}\frac{1}{k}-\sum_{k=1}^{N-r-1}\frac{1}{k}$$

Which is the difference between 2 partial sums of harmonic numbers. Let $H_n = \sum_{k=1}^{n}\frac{1}{k}$, then $\sum_{i=1}^{r}\frac{1}{N-i} = H_{N-1}-H_{N-r-1}$ hence (3) becomes

$$\begin{aligned}
\pi_i &= \frac{N}{(N-i)\left(1+N\left(H_{N-1}-H_{N-r-1}\right)\right)} \qquad\qquad i=1,2,\cdots,r-1\\
&= \frac{1}{(N-i)\left(\frac{1}{N}+H_{N-1}-H_{N-r-1}\right)} \qquad\qquad i=1,2,\cdots,r-1
\end{aligned}$$

Hence

$$\pi_i = \frac{1}{(N-i)\left(H_N-H_{N-r-1}\right)} \qquad\qquad i=1,2,\cdots,r-1$$

This is a small program which show the long term π for $N=100, r=10$ using the above equation

Verify solution to problem 10.6

Nasser Abbasi, Math 504, CSUF spring 2008.

define the data. Assume n=100 bulbs and r=10

In[1]:=
```
n = 100;
r = 10;
```

calculate π_0 which is special case

In[3]:=
$$\text{piZero} = \frac{1}{1 + n \sum_{i=1}^{r} \frac{1}{n-i}} \text{ // N}$$

Out[3]=
```
0.0862679
```

Now calculate the rest π_1, π_2, ... π_{r-1}.

In[4]:=
```
z = HarmonicNumber[n - 1] - HarmonicNumber[n - r - 1];
restPi = Table[{i, n/((n - i) (1 + n * z))}, {i, 1, r}];
```

Print result in table form

In[6]:=
```
longTerm = Prepend[restPi, {0, piZero}] // N;
TableForm[longTerm, TableHeadings → {None, {"i", "π_i"}}]
```

Out[7]//TableForm=

i	π_i
0.	0.0862679
1.	0.0871393
2.	0.0880284
3.	0.0889359
4.	0.0898624
5.	0.0908083
6.	0.0917743
7.	0.0927611
8.	0.0937694
9.	0.0947998
10.	0.0958532

Verify the sum is ONE

In[80]:=
```
Total[longTerm[[All, 2]]]
```

Out[80]=
```
1.
```

4.14 Wed 5/7/2008

Grade: 4/4.

Computer problem, problem 12.3 in lecture notes. Simulation of problem 10.5 in above HW. Repair shop problem

Problem 12.3 lecture notes, Mathematics 504
Spring 2008, CSUF
By Nasser Abbasi

Problem

Problem
with mean 2 hours. Write an Extend model for this system.

12.3 Consider again the machine-repair problem of Section 12.3.2. Recall that there is a set of $m+n$ machines which fail independently of each other, and assume now that they fail at an exponential rate λ. Recall that m machines are to be in operation at any time, while the remaining machines serve as spares and are called into operation when an operating machine fails. If more than n machine are in a state of failure, then all the operational machines will be in service. Suppose there are s, $1 \leq s \leq m$, repair persons who service the machines independently and each at exponential rate μ. Let $X(t)$ denote the number of machines at time t that are *not* operational; that is, they are in the repair shop. Derive the arrival and departure rates for the birth-and-death model of the machine-repair problem with spares. Use these rates to develop a formula for the average number of machines that are not operational. Write a program, in MATLAB say, which computes the long-run state probabilities and implements your formula for the average number of machine that are not operational. Use the model developed in the assignment above to compare your results with theory. In particular, consider the following cases: (a) two servers, six machines, and two spares, and (b) three servers, six machines and two spares. For each case, take the mean service times to be 8 days, and the mean time between breakdowns to be 4 days for each run. Study also how the computed averages change with increasing duration of the simulation.

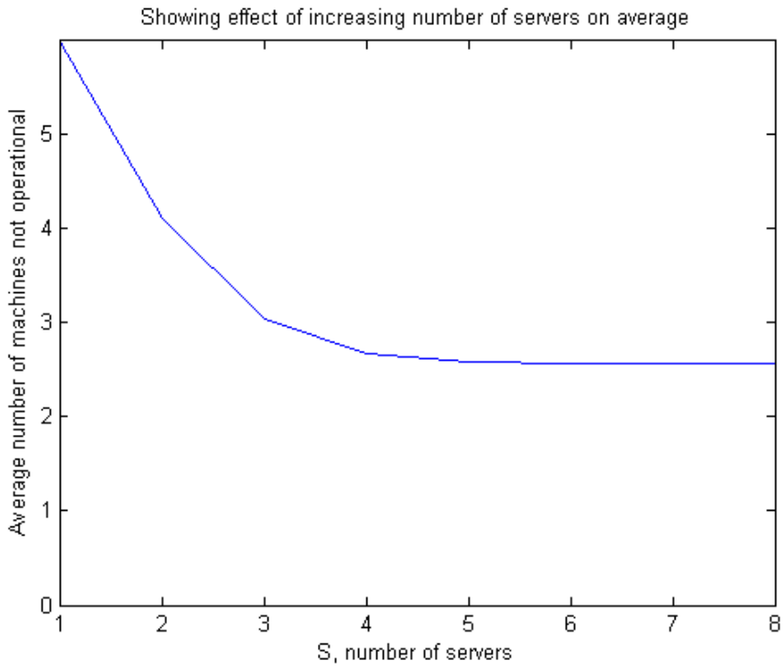
Solution

A Matlab function written to implement the above 2 cases. A driver script was written to call the function and display result. This is the result of the driver run, followed by the Matlab source code listing for the driver script and for the function which calculates the average number of machines not operational.

Result

```
CASE(a)
mu=[0.500000], lambda=[0.250000], m=6, n=2, s=2
Average number of machines not operational [4.117902]
long Term stationary distribution vector

longTermPi =

    0.0236
    0.0707
    0.1061
    0.1592
    0.1990
    0.1990
    0.1492
    0.0746
    0.0187

CASE(b)
mu=[0.500000], lambda=[0.250000], m=6, n=2, s=3
Average number of machines not operational [3.048312]
long Term stationary distribution vector

longTermPi =

    0.0477
    0.1430
    0.2144
    0.2144
    0.1787
    0.1191
    0.0596
    0.0199
    0.0033
```

We now study how the average number of machines not operational changes with increasing number of servers for case(a) parameter. This is done by changing the number of servers from s=1 all the way up to s=n+m while keeping all the other parameters the same.



Source code

Driver script

This is listing of the server script which generated the above results

```
close all;
clear all;

mu=0.5;
lambda=0.25;

%CASE(a)
s=2;
m=6;
n=2;

[av,longTermPi]=nma_HW_12_3(lambda,mu,m,n,s);
fprintf('CASE(a)\n');
fprintf('mu=[%f], lambda=[%f], m=%d, n=%d, s=%d\n',mu,lambda,m,n,s);

fprintf('Average number of machines not operational [%f]\n',av);
fprintf('long Term stationary distribution vector\n');
longTermPi
```

```
%CASE(b)
s=3;
m=6;
n=2;

[av,longTermPi]=nma_HW_12_3(lambda,mu,m,n,s);
fprintf('CASE(b)\n');
fprintf('mu=[%f], lambda=[%f], m=%d, n=%d, s=%d\n',mu,lambda,m,n,s);

fprintf('Average number of machines not operational [%f]\n',av);
fprintf('long Term stationary distribution vector\n');
longTermPi


%
%
%
m=6;
n=2;

s=1:n+m;
averages=zeros(length(s),1);

for i=1:length(s)
    [averages(i),longTermPi]=nma_HW_12_3(lambda,mu,m,n,s(i));
end

plot(s,averages);
title('Showing effect of increasing number of servers on average');
xlabel('S, number of servers');
ylabel('Average number of machines not operational');
ylim([0,max(averages)]);
```

Matlab function

```
function [av,steadyStatePI] = nma_HW_12_3(lambda,mu,m,n,s)
%
%function nma_HW_12_3(lambda,mu,m,n,s)
%solves problem 12.3 in lecture notes by Professor Gearhart,
%CSUF Mathematics 504, spring 2008
%
%by Nasser Abbasi
%
%INPUT:
%  lambda, mean time between breakdown of machines
%  mu,     mean service time
%  m,      maximum number of operating machines
%  n,      maximum number of spare machines
%  s,      number of servers.  1<=s<=m
%
%OUTPUT:
%  av,     The expected number of days a machine stays in the
%          queue (or in the repair shop)

av    = -1;
DEBUG = 0;

[msg,status] = validInput(lambda,mu,m,n,s);
if ~status
    error(msg);
end

%
% Allocate data storage
%
nStates       = n+m+1;
steadyStatePI = zeros(nStates,1);
factors       = zeros(nStates-1,1);
mus           = zeros(nStates-1,1);
lambdas       = zeros(nStates,1);

%
%  Make the Lambda vector
%
for i=0:n+m
    if i<=n
        lambdas(i+1)=m*lambda;
    else
        lambdas(i+1)=(m+n-i)*lambda;
    end
end

%
%  Make the mu vector
%
for i=1:n+m
    if i<s
        mus(i)=i*mu;
    else
        mus(i)=s*mu;
    end
end
```

```
%
%  build the factors lambda/mu terms
%
factors(1)=lambdas(1)/mus(1);
for i=2:n+m
    factors(i)=factors(i-1)*lambdas(i)/mus(i);
end

%
% Find mu_0 and initialize PI vector
%
muZero=1/(1+sum(factors));
steadyStatePI(1)=muZero;

%
% calculate the rest of the steady state PI vector
%
for i=2:nStates
    steadyStatePI(i)=factors(i-1)*muZero;
end

%
% verify sum is ONE
%
if DEBUG
    fprintf('sum of PI vector is %f\n',sum(steadyStatePI));
end

%Find expected value
av=0;
for i=0:n+m
    av=av+i*steadyStatePI(i+1);
end

%%%%%%%%%%%%%
% Function to validate input
%
%%%%%%%%%%%%%
function [msg,status] = validInput(lambda,mu,m,n,s)
VALID     = 1;
NOT_VALID = 0;

status = NOT_VALID;
msg    = '';

if ~isnumeric(s) || ~isnumeric(m) || ~isnumeric(n) ...
        ||  ~isnumeric(mu) ||  ~isnumeric(lambda)
    msg='not a numeric value in input. correct';
    return;
end

if s<=0 || m<=0 || n<0 || lambda<=0 || mu <=0
    msg='negative value in input. correct';
    return;
end

if s>n+m
    msg='number of servers must be less than n+m';
    return;
end
status=VALID;
```

key Matlab code given

```
close all;
clear all;

mu=0.5;
lambda=0.25;

%CASE(a)
s=2;
```

```
m=6;
n=2;

[av,longTermPi]=nma_HW_12_3(lambda,mu,m,n,s);
fprintf('CASE(a)\n');
fprintf('mu=[%f], lambda=[%f], m=%d, n=%d, s=%d\n',mu,lambda,m,n,s);

fprintf('Average number of machines not operational [%f]\n',av);
fprintf('long Term stationary distribution vector\n');
longTermPi


%CASE(b)
s=3;
m=6;
n=2;

[av,longTermPi]=nma_HW_12_3(lambda,mu,m,n,s);
fprintf('CASE(b)\n');
fprintf('mu=[%f], lambda=[%f], m=%d, n=%d, s=%d\n',mu,lambda,m,n,s);

fprintf('Average number of machines not operational [%f]\n',av);
fprintf('long Term stationary distribution vector\n');
longTermPi



%
%
%
m=6;
n=2;

s=1:n+m;
averages=zeros(length(s),1);

for i=1:length(s)
    [averages(i),longTermPi]=nma_HW_12_3(lambda,mu,m,n,s(i));
end

plot(s,averages);
title('Showing effect of increasing number of servers on average');
xlabel('S, number of servers');
ylabel('Average number of machines not operational');
ylim([0,max(averages)]);
```

Matlab function

```
function [av,steadyStatePI] = nma_HW_12_3(lambda,mu,m,n,s)
%
%function nma_HW_12_3(lambda,mu,m,n,s)
%solves problem 12.3 in lecture notes by Professor Gearhart,
%CSUF Mathematics 504, spring 2008
%
%by Nasser Abbasi
%
%INPUT:
%  lambda, mean time between breakdown of machines
```

```
%  mu,     mean service time
%  m,      maximum number of operating machines
%  n,      maximum number of spare machines
%  s,      number of servers.  1<=s<=m
%
%OUTPUT:
%  av,     The expected number of days a machine stays in the
%          queue (or in the repair shop)

av    = -1;
DEBUG = 0;

[msg,status] = validInput(lambda,mu,m,n,s);
if ~status
    error(msg);
end

%
% Allocate data storage
%
nStates       = n+m+1;
steadyStatePI = zeros(nStates,1);
factors       = zeros(nStates-1,1);
mus           = zeros(nStates-1,1);
lambdas       = zeros(nStates,1);

%
%  Make the Lambda vector
%
for i=0:n+m
    if i<=n
        lambdas(i+1)=m*lambda;
    else
        lambdas(i+1)=(m+n-i)*lambda;
    end
end

%
%  Make the mu vector
%
for i=1:n+m
    if i<s
        mus(i)=i*mu;
    else
        mus(i)=s*mu;
    end
end

%
%  build the factors lambda/mu terms
%
factors(1)=lambdas(1)/mus(1);
for i=2:n+m
    factors(i)=factors(i-1)*lambdas(i)/mus(i);
end

%
% Find mu_0 and initialize PI vector
```

```
%
muZero=1/(1+sum(factors));
steadyStatePI(1)=muZero;

%
% calculate the rest of the steady state PI vector
%
for i=2:nStates
    steadyStatePI(i)=factors(i-1)*muZero;
end

%
% verify sum is ONE
%
if DEBUG
    fprintf('sum of PI vector is %f\n',sum(steadyStatePI));
end

%Find expected value
av=0;
for i=0:n+m
    av=av+i*steadyStatePI(i+1);
end

%%%%%%%%%%%%%%
% Function to validate input
%
%%%%%%%%%%%%%%
function [msg,status] = validInput(lambda,mu,m,n,s)
VALID     = 1;
NOT_VALID = 0;

status = NOT_VALID;
msg    = '';

if ~isnumeric(s) || ~isnumeric(m) || ~isnumeric(n) ...
        ||  ~isnumeric(mu) ||  ~isnumeric(lambda)
    msg='not a numeric value in input. correct';
    return;
end

if s<=0 || m<=0 || n<0 || lambda<=0 || mu <=0
    msg='negative value in input. correct';
    return;
end

if s>n+m
    msg='number of servers must be less than n+m';
    return;
end

status=VALID;
```

4.15 Wed 5/7/2008

Problems 9.3 and 9.5 (On Poisson process)

4.15.1 Problem 9.3

> 9.3 Consider a Poisson process such that each time an event occurs, it is of type 1 with probability $p > 0$, or it is type 2 with probability $q = 1 - p > 0$. Assume these two types appear independently from event to event. Show that these two processes each are Poisson processes and that they are independent of each other. Help: Let $N_1(t)$ and $N_2(t)$ be the counting processes for each type of event. Find the joint distribution by conditioning on $N(t)$, where $N(t) = N_1(t) + N_2(t)$.

Start by showing that the processes $N_1(t)$ and $N_2(t)$ are each a Poisson process. Next show that they are independent by showing that the product of these 2 distributions is equal to the joined distribution.

Given: $N(t) = N_1(t) + N_2(t)$, Where are told that $N(t)$ is a Poisson process. Need to find $\Pr(N_1(t) = n)$ and $\Pr(N_2(t) = m)$.

By law of total probabilities

$$\begin{aligned}\Pr(N_1(t) = n) = &\Pr(N_1(t) = n, N_2(t) = 0)\\ &or \Pr(N_1(t) = n, N_2(t) = 1)\\ &or \Pr(N_1(t) = n, N_2(t) = 2)\\ &\vdots\\ &or \Pr(N_1(t) = n, N_2(t) = \infty)\end{aligned}$$

Hence

$$\Pr(N_1(t) = n) = \sum_{m=0}^{\infty} \Pr(N_1(t) = n, N_2(t) = m) \tag{A1}$$

Similarly,

$$\Pr(N_2(t) = m) = \sum_{n=0}^{\infty} \Pr(N_1(t) = n, N_2(t) = m) \tag{A2}$$

Now find expression for the joined distribution $\Pr(N_1(t) = n, N_2(t) = m)$ to complete the above evaluation. Condition on $N(t)$ hence we obtain

$$\begin{aligned}\Pr(N_1(t) = n, N_2(t) = m) = &\Pr(N_1(t) = n, N_2(t) = m \mid N(t) = 0) \Pr(N(t) = 0)\\ &or \Pr(N_1(t) = n, N_2(t) = m \mid N(t) = 0) \Pr(N(t) = 1)\\ &or \Pr(N_1(t) = n, N_2(t) = m \mid N(t) = 0) \Pr(N(t) = 2)\\ &\vdots\\ &or \Pr(N_1(t) = n, N_2(t) = m \mid N(t) = 0) \Pr(N(t) = \infty)\end{aligned}$$

or

$$\Pr\left(N_1(t)=n,N_2(t)=m\right)=\sum_{k=0}^{\infty}\Pr\left(N_1(t)=n,N_2(t)=m\mid N(t)=k\right)\Pr\left(N(t)=k\right)$$

But since $N(t)=N_1(t)+N_2(t)$, then the above reduces to one case which is

$$\Pr\left(N_1(t)=n,N_2(t)=m\right)=\Pr\left(N_1(t)=n,N_2(t)=m\mid N(t)=n+m\right)\Pr\left(N(t)=n+m\right) \tag{1}$$

And all the other probabilities must be zero.

Now in (1), we are given that $\Pr\left(N(t)=n+m\right)$ is a Poisson process with some rate λ Hence the rate adjusted for duration t must be λt, hence from definition of Poisson process with rate λt we write

$$\Pr\left(N(t)=n+m\right)=\frac{(\lambda t)^{n+m}e^{-(\lambda t)}}{(n+m)!} \tag{2}$$

Now we need to evaluate the term $\Pr\left(N_1(t)=n,N_2(t)=m\mid N(t)=n+m\right)$ in (1). This terms asks for the probability of getting the sum $(n+m)$. If we think of n as number of successes and m as number of failures, then this is asking for probability of getting n success out of $n+m$ trials. But this is given by Binomial distribution

$$\Pr\left(X=n\right)=\binom{n+m}{n}p^nq^m$$

Where p is the probability of event type I, and q is the probability of not getting this event, which is the probability of event II which is given by $q=(1-p)$ hence the above becomes

$$\Pr\left(N_1(t)=n,N_2(t)=m\mid N(t)=n+m\right)=\binom{n+m}{n}p^nq^m \tag{3}$$

Substitute (2) and (3) into (1) we obtain

$$\begin{aligned}\Pr\left(N_1(t)=n,N_2(t)=m\right)&=\binom{n+m}{n}p^nq^m\frac{(\lambda t)^{n+m}e^{-(\lambda t)}}{(n+m)!}\\&=\frac{(n+m)!}{m!n!}p^nq^m\frac{(\lambda t)^{n+m}e^{-(\lambda t)}}{(n+m)!}\\&=\frac{(p\lambda t)^n}{n!}\frac{(q\lambda t)^m}{m!}e^{-(\lambda t)}\end{aligned} \tag{4}$$

But $p+q=1$, hence $e^{-(\lambda t)}=e^{-(\lambda t(p+q))}=e^{-((\lambda tp)+(\lambda tq))}=e^{-(\lambda tp)}e^{-(\lambda tq)}$ hence (4) becomes

$$\Pr\left(N_1(t)=n,N_2(t)=m\right)=\left(\frac{(p\lambda t)^n}{n!}e^{-(\lambda tp)}\right)\left(\frac{(q\lambda t)^m}{m!}e^{-(\lambda tq)}\right) \tag{5}$$

The above is the joined probability of $N_1(t)$ and $N_2(t)$. We know can determine the probability distribution of $N_1(t)$ and $N_2(t)$ from substituting (5) into (A1) and (A2)

$$\begin{aligned}\Pr\left(N_2(t)=m\right)&=\sum_{n=0}^{\infty}\Pr\left(N_1(t)=n,N_2(t)=m\right)\\&=\sum_{n=0}^{\infty}\left(\frac{(p\lambda t)^n}{n!}e^{-(\lambda tp)}\right)\left(\frac{(q\lambda t)^m}{m!}e^{-(\lambda tq)}\right)\end{aligned}$$

We remove terms outside sum which do not depend on n and obtain

$$\Pr\left(N_2(t)=m\right)=\left(\frac{(q\lambda t)^m}{m!}e^{-(\lambda tq)}\right)e^{-(\lambda tp)}\sum_{n=0}^{\infty}\frac{(p\lambda t)^n}{n!}$$

But $\sum_{n=0}^{\infty}\frac{(p\lambda t)^n}{n!}=e^{p\lambda t}$ by definition, hence the above becomes

$$\Pr\left(N_2(t)=m\right)=\frac{(q\lambda t)^m}{m!}e^{-(\lambda tq)}$$

Therefore, we see that $\Pr\left(N_2(t)=m\right)$ satisfies the Poisson formula. To show it is a Poisson distribution, we must also show that it satisfies the following

1. $N_2(0)=0$. We see that at $t=0$, the above becomes $\Pr\left(N_2(0)=0\right)=\frac{(q\lambda\times 0)^0}{0!}e^{-(\lambda q\times 0)}=0^0\times 1$,But[9] $0^0=1$, hence $\Pr\left(N_2(0)=0\right)=1$, Therefore $N_2(0)=0$.

2. Increments are independents of each others. Since the original process $N(t)$ is already given to be Poisson process, then the increments of $N(t)$ are independent of each others. But $N_2(t)$ increments are a subset of those increments. Therefore, $N_2(t)$ increments must by necessity be independent of each others.

Similar arguments show that

$$\Pr\left(N_1(t)=n\right)=\frac{(q\lambda t)^n}{n!}e^{-(\lambda tq)}$$

and that it satisfies the Poisson definition.We now need to show independence. We see that

$$\Pr\left(N_2(t)=m\right)\Pr\left(N_1(t)=n\right)=\frac{(q\lambda t)^m}{m!}e^{-(\lambda tq)}\frac{(q\lambda t)^n}{n!}e^{-(\lambda tq)}$$

But from (5) above, we see this is the same as $\Pr\left(N_1(t)=n,N_2(t)=m\right)$, therefore

$$\Pr\left(N_2(t)=m\right)\Pr\left(N_1(t)=n\right)=\Pr\left(N_1=n,N_2=m\right)$$

Hence $N_1(t)$ and $N_2(t)$ are 2 independent Poisson processes.

4.15.2 Problem 9.5

> 9.5 A vehicle-controlled traffic light will stay green for τ seconds after a car passes through the intersection, and if no cars arrive during a period of time τ, then it turns red. Let X denote the number of cars that pass through the intersection following dissapation of the initial queue and until the light next turns red. Assume cars arrive according to a Poisson process with rate λ. Find the probability distribution of X and find the expected value of X. Help: Note that the event $\{X\geq n\}$ occurs if the interarrival times of the next n arriving cars are each no more than τ.

Let the interr arival time between each car be T_i where i is the interval as indicated by this diagram

[9] 0^0 depends on the context. I checked a reference that in this context, it is ok to define $0^0=1$ otherwise, 0^0 is taken as undefined.

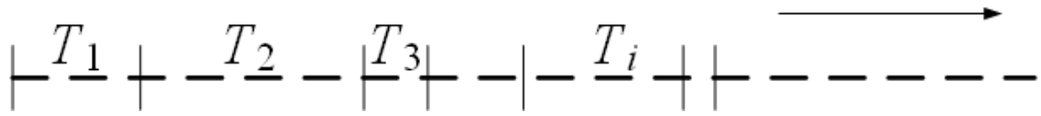


Figure 4.5: Inter-arrival times are random variable which is exponential distributed

For the number of cars that pass through the intersection to be n it must imply that the interval between the first n cars was less than τ and that the $(n+1)^{th}$ car arrived after than n^{th} car after more than τ units of time. Therefore

$$\Pr(X=n) = \Pr(T_1 < \tau, T_2 < \tau, T_3 < \tau, \cdots, T_n < \tau, T_{n+1} \geq \tau) \tag{1}$$

But since X is a Poisson random number with parameter λ , then the time between increment T is an exponential random number with parameter λ (and they are independent from each others). Hence

$$\Pr(T_i \geq \tau) = e^{-\lambda\tau}$$

and

$$\Pr(T_i < \tau) = 1 - e^{-\lambda\tau}$$

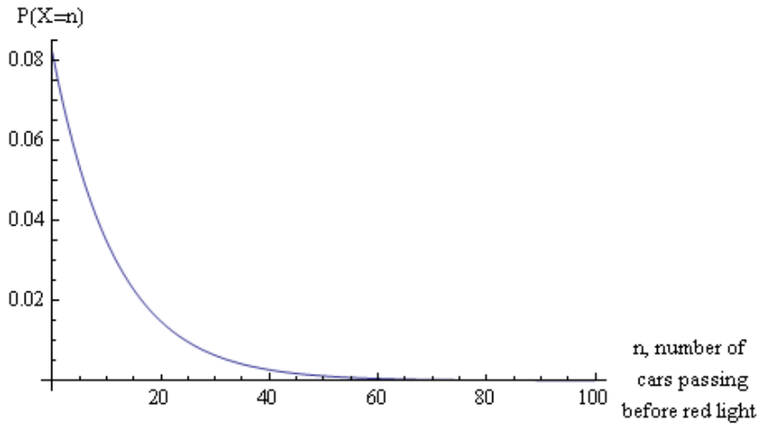
Hence (1) becomes

$$\Pr(X=n) = \left(1-e^{-\lambda\tau}\right)^n e^{-\lambda\tau}$$

This is a small program which plots the probability above as function of n for some fixed λ, τ. It shows as expected the probability of n becomes smaller the larger n gets.

```
In[96]:= pr[n_] := (1 - Exp[-λ τ])^n Exp[-λ τ]
         data = Table[{n, pr[n]}, {n, 0, 100}];
         ListPlot[data, Joined → True, AxesOrigin → {0, 0}, PlotRange → All,
          AxesLabel → {"n, number of\n cars passing\nbefore red light",
            "P(X=n)"}]
```

Out[98]=



Now

$$\begin{aligned} E(X) &= \sum_{n=0}^{\infty} n \Pr(X=n) \\ &= \sum_{n=0}^{\infty} n\left(1-e^{-\lambda\tau}\right)^n e^{-\lambda\tau} \\ &= e^{-\lambda\tau} \sum_{n=1}^{\infty} n\left(1-e^{-\lambda\tau}\right)^n \end{aligned}$$

Let $1 - e^{-\lambda\tau} = z$ then the above becomes

$$E(X) = e^{-\lambda\tau}\sum_{n=1}^{\infty} nz^n$$

The above sum converges since by ratio test the $(k)^{th}$ term over the $(k+1)^{th}$ term is less than one. (I can find a closed form expression for this sum?)

4.16 Challenge Problems

These are extra problems relating to first midterm the instructor gave the class to try to work out. Here are the questions

4.16.1 Problems

Math 504 Exam 1 Extra Problems

1. A stochastic process has states $0, 1, 2, \cdots, r$. If the process is in state $i \geq 1$, then during the next time step, it goes to a state that is selected at random from among the states $0, 1, 2, \cdots, i-1$. If the process is in state 0, it stays there. This process can be modelled as a Markov chain. (a) Specify the one-step transition probability matrix. (b) Let Y_i be the number of steps that the process takes to reach state 0, given that it starts in state i. Derive a linear system of equations that determines the expected values $E(Y_i)$, for $i = 1, 2, \cdots, r$. Write the linear system in the form $Ax = b$, where the components of the vector x are the unknown expected values. Specify the matrices A and b.

Problem 1 Solve explicitly for the expected values $E(Y_i)$, $i = 1, 2, \cdots, r$. Of particular interest is $E(Y_r)$, when r is large. What is an asymptotic estimate of this quantity?

3. Suppose N red balls and N white balls are placed in two urns so that each urn contains N balls. Label one urn A and the other B. The following action is performed repeatedly and independently: A ball is selected at random from urn A and independently a ball is selected at random from urn B. Each ball is then transferred to the other urn. The state of the system at the beginning of each trial is the number of white balls in urn B. This process can be modelled as a regular Markov chain. Determine the one-step transition probability matrix.

Problem 3 Find an explicit formula for the long-run state probability distribution. In other words, find the probability vector w that satisfies the equation $w = wP$, where P is the one-step transition probability matrix.

Figure 4.6: Problem 1

4. A machine component has a random lifetime L with given discrete distribution $P(L = k) = \beta_k$, for $k = 1, 2, \cdots$. Assume $\beta_k > 0$ for all k. A component is put into operation, and replaced when it burns out. However, if the component in operation has age N, it is instantly replaced by a new component. This process can be modelled as a regular Markov chain, where the state X_n at time $n = 0, 1, 2, \cdots$, is defined by $X_n = i$ if the component in operation has age i. The possible states are $i = 0, 1, 2, \cdots, N-1$. The state $i = 0$ means that a new component has just been placed into operation. (a) Determine the one-step transition probabilities. Help: For $i = 0, 1, 2, \cdots, N-2$,

$$a_i = p_{i,i+1} = P(L > i+1 \,|\, L > i) \ , \quad \text{and} \quad p_{N-1,0} = 1 \ .$$

First, find the probabilities $\{a_i\}$, $i = 0, 1, 2, \cdots, N-2$. Then, in terms of the $\{a_i\}$, show the one-step transition probability matrix. (b) Suppose we wish to study the time between replacements. Interpret this time as a first entrance time. (c) Derive a linear system of equations that determines the expected first entrance times from states $i = 1, 2, \cdots, N-1$ to state zero. Express these equations in terms of the $\{a_i\}$. (d) Suppose the expected first entrance times in part (c) have been determined. Find a single formula for the expected first entrance time from state 0 to state 0. Express this equation in terms of the $\{a_i\}$, and the expected first entrance times from part (c).

Problem 4 (i) Find an explicit formula for the long-run state probability distribution. In other words, find the probability vector w that satisfies the equation $w = wP$, where P is the one-step transition probability matrix. (ii) Give an interpretation for the component w_0, and for w_{N-1}, and interpret $w_0 - w_{N-1}$. (iii) Determine the expected first entrance time from state 0 to state 0. Help: Use w_0.

Figure 4.7: Problem 2

4.16.2 Problem 1

1. A stochastic process has states $0, 1, 2, \cdots, r$. If the process is in state $i \geq 1$, then during the next time step, it goes to a state that is selected at random from among the states $0, 1, 2, \cdots, i-1$. If the process is in state 0, it stays there. This process can be modelled as a Markov chain. (a) Specify the one-step transition probability matrix. (b) Let Y_i be the number of steps that the process takes to reach state 0, given that it starts in state i. Derive a linear system of equations that determines the expected values $E(Y_i)$, for $i = 1, 2, \cdots, r$. Write the linear system in the form $Ax = b$, where the components of the vector x are the unknown expected values. Specify the matrices A and b.

Problem 1 Solve explicitly for the expected values $E(Y_i)$, $i = 1, 2, \cdots, r$. Of particular interest is $E(Y_r)$, when r is large. What is an asymptotic estimate of this quantity?

Solution:

We start with the solution we already[10] obtained for $E(Y_i)$ which is

$$E(Y_i) = 1 + \sum_{k=1}^{i-1} E(Y_k)\, P_{ik}$$

Let $E(Y_i) = x_i$ hence the above can be written as

$$x_i = 1 + \sum_{k=1}^{i-1} x_k P_{ik}$$

[10] See my midterm solution for this problem

But $P_{ik} = \frac{1}{i}$ then the above becomes

$$x_i = 1 + \frac{1}{i}\sum_{k=1}^{i-1} x_k$$

Multiply by i the above becomes

$$i\ x_i = i + \sum_{k=1}^{i-1} x_k$$

Therefore, we obtain the following equations for $i = 1 \cdots r$

For $i = 1$

$$x_1 = 1 \tag{1}$$

For $i = 2$

$$2x_2 = 2 + x_1 \tag{2}$$

For $i = 3$

$$3x_3 = 3 + x_1 + x_2 \tag{3}$$

For $i = 4$

$$4x_4 = 4 + x_1 + x_2 + x_3 \tag{4}$$

etc...

Now evaluate (2)-(1) and (3)-(2) and (4)-(3), etc... we obtain the following equations

(2)-(1) gives

$$\begin{aligned} 2x_2 - x_1 &= 2 + x_1 - 1 \\ x_2 &= \frac{1 + 2x_1}{2} \end{aligned} \tag{5}$$

(3)-(2) gives

$$\begin{aligned} 3x_3 - 2x_2 &= 3 + x_1 + x_2 - 2 - x_1 \\ 3x_3 - 2x_2 &= 1 + x_2 \\ x_3 &= \frac{1 + 3x_2}{3} \end{aligned}$$

etc... Hence we see that for the r^{th} term we obtain

$$x_r = \frac{1 + r\ x_{r-1}}{r}$$

Hence

$$x_r = \frac{1}{r} + x_{r-1} \tag{6}$$

Now replace r by $r - 1$ in the above we obtain

$$x_{r-1} = \frac{1}{r-1} + x_{r-2}$$

replace the above in (6) we obtain

$$x_r = \frac{1}{r} + \left(\frac{1}{r-1} + x_{r-2} \right) \tag{7}$$

And again, in the above, $x_{r-2} = \frac{1}{r-2} + x_{r-3}$ hence (7) becomes

$$x_r = \frac{1}{r} + \left(\frac{1}{r-1} + \left(\frac{1}{r-2} + x_{r-3} \right) \right) \tag{8}$$

and so on, until we get to $x_1 = 1$, hence we obtain

$$\boxed{x_r = \tfrac{1}{r} + \tfrac{1}{r-1} + \tfrac{1}{r-2} + \cdots + 1}$$

Hence

$$x_r = \sum_{k=1}^{r} \frac{1}{k}$$

Which is the harmonic series. Now, it is known that[11]

Please see `http://en.wikipedia.org/wiki/Harmonic_number`

$$\lim_{r \to \infty} x_r = \log(r) - \gamma$$

Where γ is Euler Gamma constant given by

```
In[47]:= N[EulerGamma, 16]

Out[47]= 0.5772156649015329
```

[11] Do I have to proof this?

4.17 Links

1. Mathworks SimEvents `http://www.mathworks.com/products/simevents/description2.html`

2. Free demo of extend `http://www.extendsim.com/prods_demo.html`

3. Started to make comparison between some simulation packages. This is not complete

This note compares features between SimEvents™ discrete event simulator from The MathWorks and ExtendSim from Imagine Than inc.

(I need to add Rockwell Arena also) http://www.arenasimulation.com/ check arean basic edition. Arena 10 for students can be downloaded for free from

http://pl.wordpress.com/tag/simulation-software/

(when installing, use STUDENT as serial number)

By Nasser Abbasi, last updated 12/22/07

		SimEvents	ExtendSim
Queue type	Sorted queue FIFO	X	X
	Sorted queue LIFO	X	X
	Sorted queue attribute value		X
	Sorted queue priority	X	X
	Resource pool queue		X
Server Type	Infinite server	X	X
	N-Server	X	?
	Single server	X	X
	Select server delay time statistics from known distributions?	No	Yes

I think extendSim has more capabilities, but I need more time to study this. I do not have an evaluation version of SimEvents, but I have access to a demo version of ExtendSim.

Will update this as time permits.

Links
http://www.wintersim.org/ winter simulation conference
http://www.wintersim.org/pastprog.htm papers from winder simulation conferences online