Operations Research

Hiroshi Toyoizumi ¹

March 8, 2013

¹This handout is available at http://www.f.waseda.jp/toyoizumi.
# Contents

1 Introduction of Operations Research  
1.1 A Simple Example 7  
1.2 What is O.R.? 8  
1.3 Decision Making 9  
1.4 History 10  
1.5 Examples of O.R. 10

2 Basics in Probability Theory 17  
2.1 Why Probability? 17  
2.2 Probability Space 17  
2.3 Conditional Probability and Independence 20  
2.4 Random Variables 22  
2.5 Expectation, Variance and Standard Deviation 23  
2.6 How to Make a Random Variable 26  
2.7 News-vendor Problem, “How many should you buy?” 26  
2.8 Covariance and Correlation 27  
2.9 Value at Risk 28  
2.10 References 31  
2.11 Exercises 32

3 Markov chain 33  
3.1 Discrete-time Markov chain 33  
3.2 Stationary State 34  
3.3 Matrix Representation 35  
3.4 Stock Price Dynamics Evaluation Based on Markov Chain 35  
3.5 Google’s PageRank and Markov Chain 36
## Contents

### 4 Birth and Death process and Poisson Process

- 4.1 Definition of Birth and Death Process ........................................... 40
- 4.2 Differential Equations of Birth and Death Process ......................... 41
- 4.3 Infinitesimal Generator ................................................................. 41
- 4.4 System Dynamics ............................................................................ 42
- 4.5 Poisson Process ............................................................................... 42
- 4.6 Why we use Poisson process? ......................................................... 43
- 4.7 Z-transform of Poisson process ...................................................... 44
- 4.8 Independent Increment .................................................................. 44
- 4.9 Interarrival Time ............................................................................ 44
- 4.10 Memory-less property of Poisson process ...................................... 44
- 4.11 PASTA: Poisson Arrival See Time Average .................................... 45
- 4.12 Exercise ......................................................................................... 45

### 5 Introduction of Queueing Systems

- 5.1 Foundation of Performance Evaluation ........................................... 46
- 5.2 Starbucks ....................................................................................... 47
- 5.3 Specification of queueing systems ................................................... 47
- 5.4 Little’s formula ............................................................................... 48
- 5.5 Lindley equation and Loynes variable ........................................... 50
- 5.6 Exercise ......................................................................................... 51

### 6 M/M/1 queue

- 6.1 M/M/1 queue as a birth and death process ..................................... 53
- 6.2 Utilization ....................................................................................... 54
- 6.3 Waiting Time Estimation ............................................................... 54
  - 6.3.1 Waiting Time by Little’s Formula ............................................ 55
  - 6.3.2 Waiting Time Distribution of M/M/1 Queues ......................... 55
- 6.4 Example ......................................................................................... 56
- 6.5 Exercise ......................................................................................... 57

### 7 Reversibility

- 7.1 Output from Queue ....................................................................... 58
- 7.2 Reversibility .................................................................................. 58
  - 7.2.1 Definition of Reversibility ....................................................... 58
  - 7.2.2 Local Balance for Markov Chain ........................................... 58
- 7.3 Output from M/M/1 queue ............................................................. 60
- 7.4 Exercise ......................................................................................... 60
CONTENTS

8 Network of Queues 61
  8.1 Open Network of queues 61
  8.2 Global Balance of Network 61
  8.3 Traffic Equation 62
  8.4 Product Form Solution 62
  8.5 Exercises 64

9 Examples of Queueing System Comparison 65
  9.1 Single Server vs Tandem Servers 65
  9.2 $M/M/2$ queue 66
  9.3 $M/M/1$ VS $M/M/2$ 67
  9.4 Two $M/M/1$ VS One $M/M/2$ 68
  9.5 Exercises 69

10 Repairman Problem in Manufacturing System 70
  10.1 Repairman Problem 70
  10.2 Average Analysis 71
  10.3 Single Repairman in Markov Model 74

11 Production Systems 77
  11.1 Produce-to-Order System 77
  11.2 Produce-to-Stock System 78

12 Utility Function 81
  12.1 Horse Races 81
  12.2 Define the Utility 82
  12.3 Horse Race Revisited 84
  12.4 Risk-averse and Risk-neutral Investors 85
  12.5 Log Utility Function 86

13 Portfolio Analysis 87
  13.1 Portfolio Problem 87
  13.2 Mean and Variance of $W$ 89
  13.3 Portfolio of Two Stocks 90

14 Basics Formulation of Linear Programming 93
  14.1 What is Linear Programming 93
  14.2 Example of Wine Production 93
  14.3 Mathematical Formulation 95
15 Solving LP
  15.1 Can We Always Solve LP? ........................................ 96
  15.2 Slack Variables .................................................. 96
  15.3 Dictionary ....................................................... 97
  15.4 Simplex Method .................................................. 98
Bibliography


Chapter 1

Introduction of Operations Research

Here’s a brief introduction of the concept of Operations Research.

1.1 A Simple Example

Example 1.1 (Behind Monty Hall’s Doors [Mori and Matsui, 2004]). There was a famous quiz show in US, called “Let’s Make a Deal.” The MC of the show was Monty Hall.

Figure 1.1: Monty Hall appeared in “Let’s Make a Deal” adopted from http://www.curtalliaume.com/lmad.html

Problem 1.1. Suppose you’re on a game show, and you’re given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what’s behind the doors, opens another door,
say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

![Figure 1.2: Behind Monty Hall’s Doors from Wiki:Behind Monty Hall’s Doors.](image)

1.2 What is O.R.?

- INFORMS: The Institute for Operations Research and the Management Sciences (INFORMS) [http://www.informs.org](http://www.informs.org)

  Since its inception more than 50 years ago, O.R. has contributed billions of dollars in benefits and savings to corporations, government, and the nonprofit sector. Because the public is largely unacquainted with the sophisticated techniques used by operations researchers, INFORMS established a public relations department to educate those outside the profession.¹


  In a nutshell, operations research (O.R.) is the discipline of applying advanced analytical methods to help make better decisions. By using techniques such as mathematical modeling to analyze complex situations, operations research gives executives the power to make more effective decisions and build more productive systems based on:

  - More complete data

¹ [http://www.informs.org](http://www.informs.org)
Figure 1.3: The Science of Better

- Consideration of all available options
- Careful predictions of outcomes and estimates of risk
- The latest decision tools and techniques

1.3 Decision Making

A uniquely powerful approach to decision making You’ve probably seen dozens of articles and ads about solutions that claim to enhance your decision-making capabilities. O.R. is unique. It’s best of breed, employing highly developed methods practiced by specially trained professionals. It’s powerful, using advanced tools and technologies to provide analytical power that no ordinary software or spreadsheet can deliver out of the box. And it’s tailored to you, because an O.R. professional offers you the ability to define your specific challenge in ways that make the most of your data and uncover your most beneficial options. To achieve these results, O.R. professionals draw upon the latest analytical technologies, including:

- Simulation. Giving you the ability to try out approaches and test ideas for improvement
- Optimization. Narrowing your choices to the very best when there are virtually innumerable feasible options and comparing them is difficult

• Probability and Statistics. Helping you measure risk, mine data to find valuable connections and insights, test conclusions, and make reliable forecasts\(^3\)

Decision making is hard for human beings. What is important for O.R. is that we can use scientific approach towards OUR decision making process.

### 1.4 History

Here’s the brief history how O.R. was born[\textit{Mori and Matsui, 2004}].

1. In 1939, A.P. Rowe started to use the terminology “operational research”\(^4\) for the development of the sophisticated system of radar defense system preventing from Nazi attack to England.

2. Around 1940s, the concept of O.R. was brought to the United States. Americans P. M. Morse and G. E. Kimball published the first book about O.R. (\textit{Morse MP, Kimbal GE. Methods of operations research. Cambridge, MA: MIT Press}).

3. After the world war II, O.R. researchers are starting to use their method to management science.

### 1.5 Examples of O.R.

\textbf{Example 1.2} (More Efficient Planning and Delivery at UPS\(^5\)). The world’s largest package delivery company, UPS relies on the efficient design and operation of its hub-and-spoke air network - seven hubs and nearly 100 additional airports in the U.S. - to move over a million domestic Next Day Air packages every night. Bringing greater efficiency to such an enormous system was a challenge. Known methods for solving large-scale network design problems were inadequate for planning the UPS air network. The primary obstacles were the complexity and immense size of the air operation, which involves over 17,000 origin-destination flows and more than


\(^{4}\)Not operations research

\(^{5}\)http://www.scienceofbetter.org
160 aircraft of nine different types. Solving these problems required operations research expertise in large-scale optimization, integer programming, and routing.

- The O.R. Solution

UPS Air Group worked with an MIT specialist in transportation. The joint research and development effort resulted in an optimization-based planning system for designing the UPS aircraft network. To ensure overnight delivery, the approach simultaneously determined minimal-cost aircraft routes, fleet assignments, and allocation of packages to routes. The project team built integer programming formulations that were equivalent to conventional network design formulations, but yielded greatly improved linear programming-based bounds. This enabled realistic instances of the original planning problem to be solved typically in less than six hours, with many instances solving in less than one hour—substantial time savings.

- The Value

UPS planners now use solutions and insights generated by the system to create improved plans. UPS management credits the system with identifying operational changes that have saved over $87 million to date, and is anticipating additional savings of $189 million over the next ten years. Other benefits include reduced planning time, reduced peak and non-peak costs, reduced aircraft fleet requirements, and improved planning.

Example 1.3 (Seizing Marketplace Initiative with Merrill Lynch Integrated Choice6). The Private Client Group at Merrill Lynch handles brokerage and lending services. In late 1998, Merrill Lynch and other full-service financial service firms were under assault. Electronic trading and the commoditization of trading threatened Merrill Lynch’s value proposition providing advice and guidance through a financial advisor. Management decided to offer investors more choices for doing business with Merrill Lynch by developing an online investment site. The firm had to balance carefully the needs of its brokers and its clients, while planning for considerable financial risk. Doing so would require operations research expertise in data mining and strategic planning.

6http://www.orchampions.org/prove/success_stories/smimlic.htm
Figure 1.4: The Hub-and-Spoke Structure of UPS adopted from http://people.hofstra.edu/geotrans/eng/ch5en/appl5en/hubspokeups.html
1.5. EXAMPLES OF O.R.

- The O.R. Solution
  
  A cross-functional team evaluated alternative product and service structures and pricing, and constructed models to assess individual client behavior. The models showed that revenue at risk to Merrill Lynch ranged from $200 million to $1 billion. The Merrill Lynch Management Science Group worked with executive management and a special pricing task force to:
  
  - Determine the total revenue at risk if the only clients choosing the new pricing options were those who would pay less to Merrill Lynch.
  - Determine a more realistic revenue impact based on a client’s likelihood of adopting one of the new services.
  - Assess the revenue impact for various pricing schedules, minimum fee levels, product combinations, and product features.
  - Assess the impact on each financial consultant and identify those who would be potentially most affected.
  
  All in all, they assessed more than 40 combinations of offer architectures and pricing. Most significantly, they were able to analyze new scenarios with a new set of offerings quickly enough to meet demanding self-imposed deadlines.

- The Value
  
  The resulting Integrated Choice strategy enabled Merrill Lynch to seize the marketplace initiative, changed the financial services landscape, and mitigated the revenue risk. At year-end 2000, client assets reached $83 billion in the new offer, net new assets to the firm totaled $22 billion, and incremental revenue reached $80 million. David Komansky, CEO of Merrill Lynch and Chairman of the Board, called Integrated Choice, "The most important decision we as a firm have made since we introduced the first cash-management account in the 1970s."
Figure 1.5: Merrill Lynch Asset Allocation adopted from [Duffy et al., 2005].
Example 1.4 (Rush Hour in Tokyu Denentoshi Line\textsuperscript{7}). Tokyu Denentoshi Line is notorious for the rush hour because of the congestion of trains bound for Shibuya (see Figure 1.6). A research group in Chuo University led by Professor Taguchi analyzes how to ease the rush hour in Tokyu Denentoshi Line.

They analyze the data of the passenger’s behavioral pattern. Found out that the most passengers are attracted by the express trains, which faster than the ordinary trains. Thus, the passenger overwhelms the capacity of express train, and it requires significant time to get off and get in the express trains. Once the express trains are delayed, eventually, all the other trains will be delayed.

The research group came up with an innovative idea. Delete the express trains so as to equalize the number of passengers on each train.

They develop a model of the train and passenger dynamics and ran intensive simulations, and found out “Delete the express trains” really works.

\textsuperscript{7}http://www.ise.chuo-u.ac.jp/ise-labs/taguchi-lab/member/toriumi/rail.html
Figure 1.7: Delete the express trains in Tokyu Denentoshi Line.
Chapter 2

Basics in Probability Theory

2.1 Why Probability?

Example 2.1. Here’s examples where we use probability:

- Lottery.
- Weathers forecast.
- Gamble.
- Baseball,
- Life insurance.
- Finance.

Problem 2.1. Name a couple of other examples you could use probability theory.

Since our intuition sometimes leads us mistake in those random phenomena, we need to handle them using extreme care in rigorous mathematical framework, called probability theory. (See Exercise 2.1).

2.2 Probability Space

Be patient to learn the basic terminology in probability theory. To determine the probabilistic structure, we need a probability space, which is consisted by a sample space, a probability measure and a family of (good) set of events.
CHAPTER 2. BASICS IN PROBABILITY THEORY

Definition 2.1 (Sample Space). The set of all events is called sample space, and we write it as $\Omega$. Each element $\omega \in \Omega$ is called an event.

Example 2.2 (Lottery). Here’s an example of Lottery.

- The sample space $\Omega$ is \{first prize, second prize,..., lose\}.
- An event $\omega$ can be first prize, second prize,..., lose, and so on.

Sometimes, it is easy to use sets of events in sample space $\Omega$.

Example 2.3 (Sets in Lottery). The following is an example in $\Omega$ of Example 2.2.

$$W = \{\text{win}\} = \{\text{first prize, second prize,..., sixth prize}\} \quad (2.1)$$
$$L = \{\text{lose}\} \quad (2.2)$$

Thus, we can say that “what is the probability of win?”, instead of saying “what is the probability that we have either first prize, second prize,..., or sixth prize?”.

Example 2.4 (Coin tosses). Let us consider tossing coins 3 times. Then, by writing ”up =1” and ”down = 0”, the corresponding sample space $\Omega$ is

$$\Omega = \{\omega = (x_1,x_2,x_3) : x_i = 0 \text{ or } 1\}. \quad (2.3)$$

Problem 2.2. Find the set $S = \{\text{the number of up is } 2\}$ in Example 2.4.

Definition 2.1 (Probability measure). The probability of $A$, $P(A)$, is defined for each set of the sample space $\Omega$, if the followings are satisfied:

1. $0 \leq P(A) \leq 1$ for all $A \subset \Omega$.
2. $P(\Omega) = 1$.
3. For any sequence of mutually exclusive $A_1,A_2...$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \quad (2.4)$$

In addition, $P$ is said to be the probability measure on $\Omega$. The third condition guarantees the practical calculation on probability.
2.2. PROBABILITY SPACE

Mathematically, all function \( f \) which satisfies Definition 2.1 can regarded as probability. In other words, we need to be careful to select which function is suitable for probability.

**Example 2.5** (Probability Measures in Lottery). Suppose we have a lottery such as 10 first prizes, 20 second prizes \( \cdots \) 60 sixth prizes out of total 1000 tickets, then we have a probability measure \( P \) defined by

\[
P(n) = P(\text{win } n\text{-th prize}) = \frac{n}{100} \quad (2.5)
\]
\[
P(0) = P(\text{lose}) = \frac{79}{100}. \quad (2.6)
\]

It is easy to see that \( P \) satisfies Definition 2.1. According to the definition \( P \), we can calculate the probability on a set of events:

\[
P(W) = \text{the probability of win} = P(1) + P(2) + \cdots + P(6) = \frac{21}{100}.
\]

Of course, you can cheat your customer by saying you have 100 first prizes instead of 10 first prizes. Then your customer might have a different \( P \) satisfying Definition 2.1. Thus it is pretty important to select an appropriate probability measure. Selecting the probability measure is a bridge between physical world and mathematical world. Don’t use wrong bridge!

**Problem 2.3.** In Example 2.4, find the appropriate probability measure \( P \) on \( \Omega \). For example, calculate

\[
P(S), \quad (2.7)
\]

where \( S = \{\text{the number of up is } 2\} \).

**Remark 2.1.** There is a more rigorous way to define the probability measure. Indeed, Definition 2.1 is NOT mathematically satisfactory in some cases. If you are familiar with measure theory and advanced integral theory, you may proceed to read [Durrett, 1991].
2.3 Conditional Probability and Independence

Now we introduce one of the most useful and probably most difficult concepts of probability theory.

**Definition 2.2 (Conditional Probability).** Define the probability of $B$ given $A$ by

$$ P(B \mid A) = \frac{P(B \cap A)}{P(A)} = \frac{P(B \cap A)}{P(A)}. \quad (2.8) $$

We can use the conditional probability to calculate complex probability. It is actually the only tool we can rely on. Be sure that the conditional probability $P(B \mid A)$ is different with the regular probability $P(B)$.

**Example 2.6 (Lottery).** Let $W = \{\text{win}\}$ and $F = \{\text{first prize}\}$ in Example 2.5. Then we have the conditional probability that

$$ P(F \mid W) = \frac{P(F \cap W)}{P(W)} = \frac{P(F)}{P(W)} = \frac{10}{1000} = \frac{1}{21}. $$

**Remark 2.2.** Sometimes, we may regard Definition 2.2 as a theorem and call Bayes rule. But here we use this as a definition of conditional probability.

**Problem 2.4 (False positives\(^1\)).** Answer the followings:

1. Suppose there are illegal acts in one in 10000 companies on the average. You as a accountant audit companies. The auditing contains some uncertainty. There is a 1% chance that a normal company is declared to have some problem. Find the probability that the company declared to have a problem is actually illegal.

2. Suppose you are tested by a disease that strikes 1/1000 population. This test has 5% false positives, that mean even if you are not affected by this disease, you have 5% chance to be diagnosed to be suffered by it. A medical operation will cure the disease, but of course there is a mis-operation. Given that your result is positive, what can you say about your situation?

\(^1\)Modified from [Taleb, 2005, p.207].
Problem 2.5. In Example 2.4, find the conditional probability
\[ P(A|S), \] (2.9)
where \( A = \{ \text{the first 4 tosses are all up} \} \) and \( S = \{ \text{the number of up is 5} \} \).

Definition 2.3 (Independence). Two sets of events \( A \) and \( B \) are said to be independent if
\[ P(A \& B) = P(A \cap B) = P(A)P(B) \] (2.10)

Theorem 2.1 (Conditional Probability of Independent Events). Suppose \( A \) and \( B \) are independent, then the conditional probability of \( B \) given \( A \) is equal to the probability of \( B \).

Proof. By Definition 2.2, we have
\[ P(B | A) = \frac{P(B \cap A)}{P(A)} = \frac{P(B)P(A)}{P(A)} = P(B), \]
where we used \( A \) and \( B \) are independent. \( \Box \)

Example 2.7 (Independent two dices). Of course two dices are independent. So
\[ P(\text{The number on the first dice is even while the one on the second is odd}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}. \]

Example 2.8 (More on two dice). Even though the two dices are independent, you can find dependent events. For example,
\[ P(\text{The first dice is bigger than second dice even while the one on the second is even}) = \? \]
How about the following?
\[ P(\text{The sum of two dice is even while the one on the second is odd}) = \? \]
See Exercise 2.4 for the detail.
2.4 Random Variables

The name random variable has a strange and stochastic history\(^2\). Although its fragile history, the invention of random variable certainly contribute a lot to the probability theory.

**Definition 2.4** (Random Variable). The random variable \(X = X(\omega)\) is a real-valued function on \(\Omega\), whose value is assigned to each outcome of the experiment (event).

**Remark 2.3.** Note that probability and random variables is NOT same! Random variables are function of events while the probability is a number. To avoid the confusion, we usually use the capital letter to random variables.

**Example 2.9** (Lottery). A random variable \(X\) can be designed to formulate a lottery.

- \(X = 1\), when we get the first prize.
- \(X = 2\), when we get the second prize.

**Example 2.10** (Bernouilli random variable). Let \(X\) be a random variable with

\[
X = \begin{cases} 
1 \text{ with probability } p, \\
0 \text{ with probability } 1 - p. 
\end{cases}
\]

for some \(p \in [0, 1]\). The random variable \(X\) is said to be a Bernouilli random variable. Coin toss is a typical example of Bernoulli random variable with \(p = 1/2\).

Sometimes we use random variables to indicate the set of events. For example, instead of saying the set that we win first prize, we write as \(\{\omega \in \Omega : X(\omega) = 1\}\), or simply \(\{X = 1\}\).

**Definition 2.5** (Probability distribution). The probability distribution function \(F(x)\) is defined by

\[
F(x) = P\{X \leq x\}.
\]

\(^2\)J. Doob quoted in Statistical Science. (One of the great probabilists who established probability as a branch of mathematics.) While writing my book [Stochastic Processes] I had an argument with Feller. He asserted that everyone said “random variable” and I asserted that everyone said “chance variable.” We obviously had to use the same name in our books, so we decided the issue by a stochastic procedure. That is, we tossed for it and he won.
The probability distribution function fully-determines the probability structure of a random variable $X$. Sometimes, it is convenient to consider the probability density function instead of the probability distribution.

**Definition 2.6** (probability density function). The probability density function $f(t)$ is defined by

$$f(x) = \frac{dF(x)}{dx} = \frac{dP\{X \leq x\}}{dx}. \quad (2.13)$$

Sometimes we use $dF(x) = dP\{X \leq x\} = P(X \in (x, x + dx])$ even when $F(x)$ has no derivative.

**Lemma 2.1.** For a (good) set $A$,

$$P\{X \in A\} = \int_A dP\{X \leq x\} = \int_A f(x)dx. \quad (2.14)$$

**Problem 2.6.** Let $X$ be an uniformly-distributed random variable on $[100, 200]$. Then the distribution function is

$$F(x) = P\{X \leq x\} = \frac{x - 100}{100}, \quad (2.15)$$

for $x \in [100, 200]$.

- Draw the graph of $F(x)$.
- Find the probability function $f(x)$.

### 2.5 Expectation, Variance and Standard Deviation

Let $X$ be a random variable. Then, we have some basic tools to evaluate random variable $X$. First we have the most important measure, the expectation or mean of $X$.

**Definition 2.7** (Expectation).

$$E[X] = \int_{-\infty}^{\infty} xdP\{X \leq x\} = \int_{-\infty}^{\infty} xf(x)dx. \quad (2.16)$$
Remark 2.4. For a discrete random variable, we can rewrite (2.16) as

\[ E[X] = \sum_n x_n P[X = x_n] . \]  

(2.17)

Lemma 2.2. Let \((X_n)_{n=1,\ldots,N}\) be the sequence of possibly correlated random variables. Then we can change the order of summation and the expectation.

\[ E[X_1 + \cdots + X_N] = E[X_1] + \cdots + E[X_N] \]  

(2.18)

Proof. See Exercise 2.6.

\( E[X] \) gives you the expected value of \( X \), but \( X \) is fluctuated around \( E[X] \). So we need to measure the strength of this stochastic fluctuation. The natural choice may be \( X - E[X] \). Unfortunately, the expectation of \( X - E[X] \) is always equal to zero (why?). Thus, we need the variance of \( X \), which is indeed the second moment around \( E[X] \).

Definition 2.8 (Variance).

\[ \text{Var}[X] = E[(X - E[X])^2] . \]  

(2.19)

Lemma 2.3. We have an alternative to calculate \( \text{Var}[X] \).

\[ \text{Var}[X] = E[X^2] - E[X]^2 . \]  

(2.20)

Proof. See Exercise 2.6.

Unfortunately, the variance \( \text{Var}[X] \) has the dimension of \( X^2 \). So, in some cases, it is inappropriate to use the variance. Thus, we need the standard deviation \( \sigma[X] \) which has the order of \( X \).

Definition 2.9 (Standard deviation).

\[ \sigma[X] = \text{Var}[X]^{1/2} . \]  

(2.21)

Example 2.11 (Bernouilli random variable). Let \( X \) be a Bernouilli random variable with \( P[X = 1] = p \) and \( P[X = 0] = 1 - p \). Then we have

\[ E[X] = 1p + 0(1 - p) = p . \]  

(2.22)

\[ \text{Var}[X] = E[X^2] - E[X]^2 = E[X] - E[X]^2 = p(1 - p) , \]  

(2.23)

where we used the fact \( X^2 = X \) for Bernouilli random variables.
In many cases, we need to deal with two or more random variables. When these random variables are independent, we are very lucky and we can get many useful results. Otherwise...

**Definition 2.2.** We say that two random variables $X$ and $Y$ are independent when the sets \( \{ X \leq x \} \) and \( \{ Y \leq y \} \) are independent for all $x$ and $y$. In other words, when $X$ and $Y$ are independent,

\[
P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)
\] (2.24)

**Lemma 2.4.** For any pair of independent random variables $X$ and $Y$, we have

- \( \text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \).

**Proof.** Extending the definition of the expectation, we have a double integral,

\[
E[XY] = \int xydP(X \leq x, Y \leq y).
\]

Since $X$ and $Y$ are independent, we have $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$. Thus,

\[
E[XY] = \int xydP(X \leq x)dP(Y \leq y)
\]

\[
= \int xdP(X \leq x)\int ydP(X \leq y)
\]

\[
= E[X]E[Y].
\]

Using the first part, it is easy to check the second part (see Exercise 2.9.) \(\Box\)

**Example 2.12** (Binomial random variable). Let $X$ be a random variable with

\[
X = \sum_{i=1}^{n} X_i,
\] (2.25)

where $X_i$ are independent Bernoulli random variables with the mean $p$. The random variable $X$ is said to be a Binomial random variable. The mean and variance of $X$ can be obtained easily by using Lemma 2.4 as

\[
E[X] = np,
\] (2.26)

\[
\text{Var}[X] = np(1 - p).
\] (2.27)

**Problem 2.7.** Let $X$ be the number of up’s in 10 tosses.

1. Find $E[X]$ and $\text{Var}[X]$ using 2.12.

2. Find the probability measure $P$, and compute $E[X]$ using Definition 2.7.
2.6 How to Make a Random Variable

Suppose we would like to simulate a random variable \( X \) which has a distribution \( F(x) \). The following theorem will help us.

**Theorem 2.2.** Let \( U \) be a random variable which has a uniform distribution on \([0, 1]\), i.e

\[
P[U \leq u] = u. \tag{2.28}
\]

Then, the random variable \( X = F^{-1}(U) \) has the distribution \( F(x) \).

**Proof.**

\[
P[X \leq x] = P[F^{-1}(U) \leq x] = P[U \leq F(x)] = F(x). \tag{2.29}
\]

\[
\square
\]

2.7 News-vendor Problem, “How many should you buy?”

Suppose you are assigned to sell newspapers. Every morning you buy in \( x \) newspapers at the price \( a \). You can sell the newspaper at the price \( a + b \) to your customers. You should decide the number \( x \) of newspapers to buy in. If the number of those who buy newspaper is less than \( x \), you will be left with piles of unsold newspapers. When there are more buyers than \( x \), you lost the opportunity of selling more newspapers. Thus, there seems to be an optimal \( x \) to maximize your profit.

Let \( X \) be the demand of newspapers, which is not known when you buy in newspapers. Suppose you buy \( x \) newspapers and check if it is profitable when you buy the additional \( \Delta x \) newspapers. If the demand \( X \) is larger than \( x + \Delta x \), the additional newspapers will pay off and you get \( b \Delta x \), but if \( X \) is smaller than \( x + \Delta x \), you will lose \( a \Delta x \). Thus, the expected additional profit is

\[
E[\text{profit from additional } \Delta x \text{ newspapers}] = b \Delta x P\{X \geq x + \Delta x\} - a \Delta x P\{X \leq x + \Delta x\} = b \Delta x - (a + b) \Delta x P\{X \leq x + \Delta x\}.
\]

Whenever this is positive, you should increase the stock, thus the optimum stock \( x \) should satisfy the equilibrium equation;

\[
P\{X \leq x + \Delta x\} = \frac{b}{a + b}, \tag{2.30}
\]
2.8. COVARIANCE AND CORRELATION

for all \( \Delta x > 0 \). Letting \( \Delta x \to 0 \), we have

\[
P\{X \leq x\} = \frac{b}{a+b},
\]

(2.31)

Using the distribution function \( F(x) = P\{X \leq x\} \) and its inverse \( F^{-1} \), we have

\[
x = F^{-1}\left(\frac{b}{a+b}\right),
\]

(2.32)

Using this \( x \), we can maximize the profit of news-vendors.

**Problem 2.8.** Suppose you are a newspaper vender. You buy a newspaper at the price of 70 and sell it at 100. The demand \( X \) has the following uniform distribution,

\[
P\{X \leq x\} = \frac{x-100}{100},
\]

for \( x \in [100, 200] \). Find the optimal stock for you.

### 2.8 Covariance and Correlation

When we have two or more random variables, it is natural to consider the relation of these random variables. But how? The answer is the following:

**Definition 2.10 (Covariance).** Let \( X \) and \( Y \) be two (possibly not independent) random variables. Define the covariance of \( X \) and \( Y \) by

\[
Cov[X, Y] = E[(X - E[X])(Y - E[Y])].
\]

(2.34)

Thus, the covariance measures the multiplication of the fluctuations around their mean. If the fluctuations are tends to be the same direction, we have larger covariance.

**Example 2.13 (The covariance of a pair of independent random variables).** Let \( X_1 \) and \( X_2 \) be the independent random variables. The covariance of \( X_1 \) and \( X_2 \) is

\[
Cov[X_1, X_2] = E[X_1X_2] - E[X_1]E[X_2] = 0,
\]
CHAPTER 2. BASICS IN PROBABILITY THEORY

since $X_1$ and $X_2$ are independent. Thus, more generally, if the two random variables are independent, their covariance is zero. (The converse is not always true. Give some example!)

Now, let $Y = X_1 + X_2$. How about the covariance of $X_1$ and $Y$?

$$\text{Cov}[X_1, Y] = E[X_1Y] - E[X_1]E[Y]$$
$$= E[X_1(X_1 + X_2)] - E[X_1]E[X_1 + X_2]$$
$$= E[X_1^2] - E[X_1]^2$$
$$= \text{Var}[X_1] = np(1 - p) > 0.$$

Thus, the covariance of $X_1$ and $Y$ is positive as can be expected.

It is easy to see that we have

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y],$$

which is sometimes useful for calculation. Unfortunately, the covariance has the order of $XY$, which is not convenience to compare the strength among different pair of random variables. Don’t worry, we have the correlation function, which is normalized by standard deviations.

**Definition 2.11** (Correlation). Let $X$ and $Y$ be two (possibly not independent) random variables. Define the correlation of $X$ and $Y$ by

$$\rho[X, Y] = \frac{\text{Cov}[X, Y]}{\sigma[X]\sigma[Y]}.$$

**Lemma 2.5.** For any pair of random variables, we have

$$-1 \leq \rho[X, Y] \leq 1.$$  \hspace{1cm} \text{(2.37)}

**Proof.** See Exercise 2.11

\hspace{1cm} \Box

### 2.9 Value at Risk

Suppose we have one stock with its current value $x_0$. The value of the stock fluctuates. Let $X_1$ be the value of this stock tomorrow. The rate of return $R$ can be defined by

$$R = \frac{X_1 - x_0}{x_0}.$$  \hspace{1cm} \text{(2.38)}

The rate of return $R$ can be positive or negative. We assume $R$ is normally distributed with its mean $\mu$ and $\sigma$. 
2.9. VALUE AT RISK

Problem 2.9. Why did the rate of return $R$ assume to be a normal random variable, instead of the stock price $X_1$ itself.

We need to evaluate the uncertain risk of this future stock.

Definition 2.12 (Value at Risk). The future risk of a property can be evaluated by Value at Risk (VaR) $z_\alpha$, the decrease of the value of the property in the worst case which has the probability $\alpha$, or

$$P\{X_1 - x_0 \geq -z_\alpha\} = \alpha,$$  \hfill (2.39)

or

$$P\{z_\alpha \geq x_0 - X_1\} = \alpha.$$  \hfill (2.40)

In short, our damage is limited to $z_\alpha$ with the probability $\alpha$.

Figure 2.1: VaR: adopted form http://www.nomura.co.jp/terms/english/v/var.html

By the definition of rate of return (2.38), we have

$$P\left\{ R \geq -\frac{z_\alpha}{x_0} \right\} = \alpha,$$  \hfill (2.41)

or

$$P\left\{ R \leq -\frac{z_\alpha}{x_0} \right\} = 1 - \alpha.$$  \hfill (2.42)
Since \( R \) is assumed to be normal random variable, using the fact that

\[
Z = \frac{R - \mu}{\sigma},
\]

is a standard normal random variable, where \( \mu = E[R] \), and \( \sigma = \sqrt{\text{Var}[R]} \), we have

\[
1 - \alpha = P\left\{ R \leq -\frac{z_\alpha}{x_0} \right\} = P\left\{ Z \leq \frac{-z_\alpha}{x_0} - \frac{\mu}{\sigma} \right\}.
\]

(2.44)

Since the distribution function of standard normal random variable is symmetric, we have

\[
\alpha = P\left\{ Z \leq \frac{z_\alpha}{x_0} + \frac{\mu}{\sigma} \right\}.
\]

(2.45)

Set \( x_\alpha \) as

\[
\alpha = P\left\{ Z \leq x_\alpha \right\},
\]

(2.46)

or

\[
x_\alpha = F^{-1}(\alpha),
\]

(2.47)

which can be found in any standard statistics text book. From (2.45) we have

\[
\frac{z_\alpha}{x_0} + \frac{\mu}{\sigma} = x_\alpha,
\]

(2.48)

or

\[
z_\alpha = x_0(F^{-1}(\alpha) \sigma - \mu).
\]

(2.49)

Now consider the case when we have \( n \) stocks on our portfolio. Each stocks have the rate of return at one day as,

\[
(R_1, R_2, \ldots, R_n).
\]

(2.50)

Thus, the return rate of our portfolio \( R \) is estimated by,

\[
R = c_1R_1 + c_2R_2 + \cdots + c_nR_n,
\]

(2.51)

where \( c_i \) is the number of stocks \( i \) in our portfolio.
Let $q_0$ be the value of the portfolio today, and $Q_1$ be the one for tomorrow. The value at risk (VaR) $Z_\alpha$ of our portfolio is given by

$$P\{Q_1 - q_0 \geq -z_\alpha\} = \alpha. \quad (2.52)$$

We need to evaluate $E[R]$ and $\text{Var}[R]$. It is tempting to assume that $R$ is a normal random variable with

$$\mu = E[R] = \sum_{i=1}^{n} E[R_i], \quad (2.53)$$

$$\sigma^2 = \text{Var}[R] = \sum_{i=1}^{n} \text{Var}[R_i]. \quad (2.54)$$

This is true if $R_1, \ldots, R_n$ are independent. Generally, there may be some correlation among the stocks in portfolio. If we neglect it, it would cause underestimate of the risk.

We assume the vector

$$(R_1, R_2, \ldots, R_n), \quad (2.55)$$

is the multivariate normal random variable, and the estimated rate of return of our portfolio $R$ turns out to be a normal random variable again[Toyoizumi, 2008, p.7].

**Problem 2.10.** Estimate $\text{Var}[R]$ when we have only two different stocks, i.e., $R = R_1 + R_2$, using $\rho[R_1, R_2]$ defined in (2.36).

Using $\mu$ and $\sigma$ of the overall rate of return $R$, we can evaluate the VaR $Z_\alpha$ just like (2.49).

### 2.10 References

There are many good books which useful to learn basic theory of probability. The book [Ross, 1992] is one of the most cost-effective book who wants to learn the basic applied probability featuring Markov chains. It has a quite good style of writing. Those who want more rigorous mathematical frame work can select [Durrett, 1991] for their starting point. If you want directly dive into the topic like stochastic integral, your choice is maybe [Oksendal, 2003].
2.11 Exercises

Exercise 2.1. Find an example that our intuition leads to mistake in random phenomena.

Exercise 2.2. Define a probability space according to the following steps.

1. Take one random phenomena, and describe its sample space, events and probability measure
2. Define a random variable of above phenomena
3. Derive the probability function and the probability density.
4. Give a couple of examples of set of events.

Exercise 2.3. Explain the meaning of (2.4) using Example 2.2

Exercise 2.4. Check $P$ defined in Example 2.5 satisfies Definition 2.1.

Exercise 2.5. Calculate the both side of Example 2.8. Check that these events are dependent and explain why.

Exercise 2.6. Prove Lemma 2.2 and 2.3 using Definition 2.7.

Exercise 2.7. Prove Lemma 2.4.

Exercise 2.8. Let $X$ be the Bernoulli random variable with its parameter $p$. Draw the graph of $E[X]$, $Var[X]$, $\sigma[X]$ against $p$. How can you evaluate $X$?


Exercise 2.10 (Binomial random variable). Let $X$ be a random variable with

$$X = \sum_{i=1}^{n} X_i,$$

(2.56)

where $X_i$ are independent Bernoulli random variables with the mean $p$. The random variable $X$ is said to be a Binomial random variable. Find the mean and variance of $X$.

Exercise 2.11. Prove for any pair of random variables, we have

$$-1 \leq \rho[X,Y] \leq 1.$$ 

(2.57)
Chapter 3

Markov chain

Markov chain is one of the most basic tools to investigate dynamical features of stochastic phenomena. Roughly speaking, Markov chain is used for any stochastic processes for first-order approximation.

**Definition 3.1** (Rough definition of Markov Process). A stochastic process \( X(t) \) is said to be a Markov chain, when the future dynamics depends probabilistically only on the current state (or position), not depend on the past trajectory.

**Example 3.1.** The followings are examples of Markov chains.

- Stock price
- Brownian motion
- Queue length at ATM
- Stock quantity in storage
- Genes in genome
- Population
- Traffic on the internet

### 3.1 Discrete-time Markov chain

**Definition 3.1** (discrete-time Markov chain). \( (X_n) \) is said to be a discrete-time Markov chain if
• state space is at most countable,
• state transition is only at discrete instance,

and the dynamics is probabilistically depend only on its current position, i.e.,
\[
P[X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_1 = x_1] = P[X_{n+1} = x_{n+1} | X_n = x_n]. \tag{3.1}
\]

**Definition 3.2** (time-homogenous) 1-step transition probability).
\[
p_{ij} \equiv P[X_{n+1} = j | X_n = i]. \tag{3.2}
\]
The probability that the next state is \( j \), assuming the current state is in \( i \).

Similarly, we can define the \( m \)-step transition probability:
\[
p_{ij}^m \equiv P[X_{n+m} = j | X_n = i]. \tag{3.3}
\]

We can always calculate \( m \)-step transition probability by
\[
p_{ij}^m = \sum_k p_{ik}^{m-1} p_{kj}. \tag{3.4}
\]

### 3.2 Stationary State

The initial distribution: the probability distribution of the initial state. The initial state can be decided on the consequence of tossing a dice...

**Definition 3.3** (Stationary distribution). The probability distribution \( \pi_j \) is said to be a stationary distribution, when the future state probability distribution is also \( \pi_j \) if the initial distribution is \( \pi_j \).

\[
P\{X_0 = j\} = \pi_j \implies P\{X_n = j\} = \pi_j \tag{3.5}
\]

In Markov chain analysis, to find the stationary distribution is quite important. If we find the stationary distribution, we almost finish the analysis.

**Remark 3.1.** Some Markov chain does not have the stationary distribution. In order to have the stationary distribution, we need Markov chains to be irreducible, positive recurrent. See [Ross, 1992, Chapter 4]. In case of finite state space, Markov chain have the stationary distribution when all state can be visited with a positive probability.
3.3 Matrix Representation

Definition 3.4 (Transition (Probability) Matrix).

\[ P = \begin{pmatrix}
    p_{11} & p_{12} & \cdots & p_{1m} \\
    p_{21} & p_{22} & \cdots & p_{2m} \\
    \vdots & \vdots & \ddots & \vdots \\
    p_{m1} & p_{m2} & \cdots & p_{mm}
\end{pmatrix} \quad (3.6) \]

The matrix of the probability that a state \( i \) to another state \( j \).

Definition 3.5 (State probability vector at time \( n \)).

\[ \pi(n) \equiv (\pi_1(n), \pi_2(n), \ldots) \quad (3.7) \]

\( \pi_i(n) = P[X_n = i] \) is the probability that the state is \( i \) at time \( n \).

Theorem 3.1 (Time Evolution of Markov Chain).

\[ \pi(n) = \pi(0)P^n \quad (3.8) \]

Given the initial distribution, we can always find the probability distribution at any time in the future.

Theorem 3.2 (Stationary Distribution of Markov Chain). When a Markov chain has the stationary distribution \( \pi \), then

\[ \pi = \pi P \quad (3.9) \]

3.4 Stock Price Dynamics Evaluation Based on Markov Chain

Suppose the up and down of a stock price can be modeled by a Markov chain. There are three possibilities: (1) up, (2) down and (3) hold. The price fluctuation tomorrow depends on the today's movement. Assume the following transition probabilities:

\[ P = \begin{pmatrix}
    0 & 3/4 & 1/4 \\
    1/4 & 0 & 3/4 \\
    1/4 & 1/4 & 1/2
\end{pmatrix} \quad (3.10) \]

For example, if today’s movement is “up”, then the probability of “down” again tomorrow is 3/4.
Problem 3.1. 1. Find the steady state distribution.

2. Is it good idea to hold this stock in the long run?

Solutions:

1. 

\[ \pi = \pi P \]  

(3.11)

\[(\pi_1, \pi_2, \pi_3) = (\pi_1, \pi_2, \pi_3) \begin{pmatrix} 0 & 3/4 & 1/4 \\ 1/4 & 0 & 3/4 \\ 1/4 & 1/4 & 1/2 \end{pmatrix} \]  

(3.12)

Using the nature of probability \((\pi_1 + \pi_2 + \pi_3 = 1)\), (3.12) can be solved and

\[(\pi_1, \pi_2, \pi_3) = (1/5, 7/25, 13/25). \]  

(3.13)

2. Thus, you can avoid holding this stock in the long term.

3.5 Google’s PageRank and Markov Chain

Google uses an innovative concept called PageRank\(^1\) to quantify the importance of web pages. PageRank can be understood by Markov chain. Let us take a look at a simple example based on [Langville and Meyer, 2006, Chapter 4].

Suppose we have 6 web pages on the internet\(^2\). Each web page has some links to other pages as shown in Figure 3.1. For example the web page indexed by 1 refers 2 and 3, and is referred back by 3. Using these link information, Google decide the importance of web pages. Here’s how.

Assume you are reading a web page 3. The web page contains 3 links to other web pages. You will jump to one of the other pages by pure chance. That

\(^1\)PageRank is actually Page’s rank, not the rank of pages, as written in http://www.google.co.jp/intl/ja/press/funfacts.html. “The basis of Google’s search technology is called PageRank, and assigns an "importance" value to each page on the web and gives it a rank to determine how useful it is. However, that’s not why it’s called PageRank. It’s actually named after Google co-founder Larry Page.”

\(^2\)Actually, the number of pages dealt by Google has reached 8.1 billion!
means your next page may be 2 with probability 1/3. You may hop the web pages according to the above rule, or transition probability. Now your hop is governed by Markov chain, you can the state transition probability $P$ as

$$
P = (p_{ij}) = 
\begin{pmatrix}
0 & 1/2 & 1/2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\
0 & 0 & 0 & 0 & 1/2 & 1/2 \\
0 & 0 & 0 & 1/2 & 0 & 1/2 \\
0 & 0 & 0 & 1 & 0 & 0 \\
\end{pmatrix}, \quad (3.14)
$$

where

$$
p_{ij} = P\{\text{Next click is web page } j|\text{reading page } i\} \quad (3.15)
$$

$$
p_{ij} = \begin{cases} 
\frac{1}{\text{the number of links outgoing from the page } i} & \text{if } i \text{ has a link to } j, \\
0 & \text{otherwise.} 
\end{cases} \quad (3.16)
$$

Starting from web page 3, you hop around our web universe, and eventually you may reach the steady state. The page rank of web page $i$ is nothing but the steady state probability that you are reading page $i$. The web page where you stay the most likely gets the highest PageRank.
**Problem 3.2.** Are there any flaws in this scheme? What will happen in the long-run?

When you happened to visit a web page with no outgoing link, you may jump to a web page completely unrelated to the current web page. In this case, your next page is purely randomly decided. For example, the web page 2 has no outgoing link. If you are in 2, then the next stop will be randomly selected.

Further, even though the current web page has some outgoing link, you may go to a web page completely unrelated. We should take into account such behavior. With probability $1/10$, you will jump to a random page, regardless of the page link.

Thus the transition probability is modified, and when $i$ has at least one outgoing link,

$$p_{ij} = \frac{9}{10} \frac{1}{\text{the number of links outgoing from the page } i} + \frac{1}{10} \frac{1}{6}.$$  \hfill (3.17)

On the other hand, when $i$ has no outgoing link, we have

$$p_{ij} = \frac{1}{6}. \hfill (3.18)$$

The new transition probability matrix $P$ is

$$P = \frac{1}{60} \begin{pmatrix}
1 & 28 & 28 & 1 & 1 & 1 \\
10 & 10 & 10 & 10 & 10 & 10 \\
19 & 19 & 1 & 1 & 19 & 1 \\
1 & 1 & 1 & 1 & 28 & 28 \\
1 & 1 & 1 & 28 & 1 & 28 \\
1 & 1 & 1 & 55 & 1 & 1 \\
\end{pmatrix}. \hfill (3.19)$$

**Problem 3.3.** Answer the followings:

1. Verify (3.19).

2. Compute $\pi(1)$ using

$$\pi(1) = \pi(0)P, \hfill (3.20)$$

given that initially you are in the web page 3.
By Theorem 3.2, we can find the stationary probability $\pi$ satisfying

$$\pi = \pi P. \quad (3.21)$$

It turns out to be

$$\pi = (0.0372, 0.0539, 0.0415, 0.375, 0.205, 0.286), \quad (3.22)$$

As depicted in Figure 3.2, according to the stationary distribution, we can say that web page 4 has the best PageRank.
Chapter 4

Birth and Death process and Poisson Process

As we saw in Chapter 3, we can analyze complicated system using Markov chains. Essentially, Markov chains can be analyzed by solving a matrix equation. However, instead of solving matrix equations, we may find a fruitful analytical result, by using a simple variant of Markov chains.

4.1 Definition of Birth and Death Process

Birth and death process is a special continuous-time Markov chain. The very basic of the standard queueing theory. The process allows two kinds of state transitions:

- \( \{ X(t) = j \to j + 1 \} \) : birth
- \( \{ X(t) = j \to j - 1 \} \) : death

Moreover, the process allows no twin, no death at the same instance. Thus, for example,

\[
P\{\text{a birth and a death at } t\} = 0.
\] (4.1)

**Definition 4.1** (Birth and Death process). Define \( X(t) \) be a Birth and Death process with its transition rate;

- \( P[X(t + \Delta t) = j + 1 | X(t) = j] = \lambda_j \Delta t + o(\Delta t) \)
- \( P[X(t + \Delta t) = j - 1 | X(t) = j] = \mu_j \Delta t + o(\Delta t) \).
4.2 Differential Equations of Birth and Death Process

The dynamics of birth and death process is described by a system of differential equations. Using Markov property, for a sufficiently small $\Delta t$, we have

\[ P_j(t + \Delta t) = P_j(t)\{1 - (\lambda_j + \mu_j)\Delta t\} + P_{j-1}(t)\lambda_j\Delta t + P_{j+1}(t)\mu_{j+1}\Delta t + o(\Delta t). \]  

(4.2)

Dividing $\Delta t$ in the both side and letting $\Delta t \to 0$, we have

\[ \frac{d}{dt} P_j(t) = \lambda_j P_{j-1}(t) - (\lambda_j + \mu_j)P_j(t) + P_{j+1}(t)\mu_{j+1}, \]  

(4.3)

for $j \geq 1$. For $j = 0$, we have

\[ \frac{d}{dt} P_0(t) = -\lambda_0 P_0(t) + P_1(t)\mu_1. \]  

(4.4)

**Problem 4.1.** Can you solve this system of differential equations? If not, what kind of data do you need?

4.3 Infinitesimal Generator

Unlike the case of discrete-time, we need the transition rate $q_{ij}$ for continuous-time Markov chains. However, It is also much convenient to use matrix form.

Let $q_{ij}$ be

\[ q_{j+1} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P\{X(t + \Delta t) = j + 1|X(t) = j\} = \lambda_j, \]  

(4.5)

\[ q_{j-1} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P\{X(t + \Delta t) = j - 1|X(t) = j\} = \mu_j \]  

(4.6)

A birth and death process is described by its infinitesimal generator $Q$ as

\[ Q = \begin{pmatrix}
-\lambda_0 & \lambda_0 \\
\mu_1 & -\lambda_1 - \mu_1 \\
\mu_2 & -\lambda_2 - \mu_2 \\
\vdots & \ddots
\end{pmatrix} \]  

(4.7)

Note that in addition to above we need to define the initial condition, in order to know its probabilistic behaviour.
CHAPTER 4. BIRTH AND DEATH PROCESS AND POISSON PROCESS

4.4 System Dynamics

Definition 4.2 (State probability). The state of the system at time $t$ can be defined by the infinite-dimension vector:

$$P(t) = (P_0(t), P_1(t), \cdots),$$

(4.8)

where $P_k(t) = P[X(t) = k]$.

The dynamics of the state is described by the differential equation of matrix:

$$\frac{dP(t)}{dt} = P(t)Q.$$  

(4.9)

Formally, the differential equation can be solved by

$$P(t) = P(0)e^{Qt},$$

(4.10)

where $e^{Qt}$ is matrix exponential defined by

$$e^{Qt} = \sum_{n=0}^{\infty} \frac{(Qt)^n}{n!}.$$  

(4.11)

Remark 4.1. It is hard to solve the system equation, since it is indeed an infinite-dimension equation. (If you are brave to solve it, please let me know!)

4.5 Poisson Process

Definition 4.3 (Poisson Process). Poisson process is a special birth and death process, which has the following parameters:

- $\mu_k = 0$: No death
- $\lambda_k = \lambda$: Constant birth rate

Then, the corresponding system equation is,

$$\frac{dP_k(t)}{dt} = -\lambda P_k(t) + \lambda P_{k-1}(t) \text{ for } k \geq 1: \text{ internal states}$$

(4.12)

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t): \text{ boundary state}$$

(4.13)
4.6. WHY WE USE POISSON PROCESS?

Also, the initial condition,

$$P_k(0) = \begin{cases} 1 & k = 0 \\ 0 & \text{otherwise} \end{cases} \quad (4.14)$$

which means no population initially.

Now we can solve the equation by iteration.

$$P_0(t) = e^{-\lambda t} \quad (4.15)$$

$$P_1(t) = \lambda te^{-\lambda t} \quad (4.16)$$

$$\cdots$$

$$P_k(t) = \frac{(\lambda t)^k}{k!}e^{-\lambda t}, \quad (4.17)$$

which is Poisson distribution!

**Theorem 4.1.** *The population at time t of a constant rate pure birth process has Poisson distribution.*

4.6 Why we use Poisson process?

- IT is Poisson process!
- It is EASY to use Poisson process!

**Theorem 4.2** (The law of Poisson small number). *Poisson process ⇔ Counting process of the number of independent rare events.*

If we have many users who use one common system but not so often, then the input to the system can be Poisson process.

Let us summarize Poisson process as an input to the system:

1. $A(t)$ is the number of arrival during $[0,t)$.

2. The probability that the number of arrival in $[0,t)$ is $k$ is given by

$$P[A(t) = k] = P_k(t) = \frac{(\lambda t)^k}{k!}e^{-\lambda t}.$$ 

3. The mean number of arrival in $[0,t)$ is given by

$$E[A(t)] = \lambda t, \quad (4.18)$$

where $\lambda$ is the arrival rate.
CHAPTER 4. BIRTH AND DEATH PROCESS AND POISSON PROCESS

4.7 Z-transform of Poisson process

Z-transform is a very useful tool to investigate stochastic processes. Here’s some examples for Poisson process.

\[ E[z^{A(t)}] = e^{-\lambda t + \lambda tz} \] (4.19)

\[ E[A(t)] = \frac{d}{dz} E[z^{A(t)}] \bigg|_{z=1} = \lambda \] (4.20)

\[ \text{Var}[A(t)] = \text{Find it!} \] (4.21)

4.8 Independent Increment

Theorem 4.3 (Independent Increment).

\[ P[A(t) = k \mid A(s) = m] = P_{k-m}(t-s) \]

The arrivals after time \( s \) is independent of the past.

4.9 Interarrival Time

Let \( T \) be the interarrival time of Poisson process, then

\[ P[T \leq t] = 1 - P_0(t) = 1 - e^{-\lambda t}; \text{ Exponential distribution.} \] (4.22)

4.10 Memory-less property of Poisson process

Theorem 4.4 (Memory-less property). \( T: \text{exponential distribution} \)

\[ P[T \leq t + s \mid T > t] = P[T \leq s] \] (4.23)

Proof.

\[ P[T \leq t + s \mid T > t] = \frac{P[t < T \leq t + s]}{P[T > t]} = 1 - e^{-\lambda s} \] (4.24)
4.11 **PASTA: Poisson Arrival See Time Average**

Poisson Arrival will see the time average of the system. This is quite important for performance evaluation.

4.12 **Excercise**

1. Find an example of Markov chains which do not have the stationary distribution.

2. In the setting of section 3.4, answer the followings:
   
   (a) Prove (3.13)
   
   (b) When you are sure that your friend was in Aizu-wakamatsu initially on Monday, where do you have to go on the following Wednesday, to join her? Describe why.
   
   (c) When you do not know when and where she starts, which place do you have to go to join her? Describe why.

3. Show $E[A(t)] = \lambda t$, when $A(t)$ is Poisson process, i.e.,

   \[ P[A(t) = k] = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \]

4. When $A(t)$ is Poisson process, calculate $\text{Var}[A(t)]$, using z-transform.

5. Make the graph of Poisson process and exponential distribution, using Mathematica.

6. When $T$ is exponential distribution, answer the following:
   
   (a) What is the mean and variance of $T$?
   
   (b) What is the Laplace transform of $T$? ($E[e^{-sT}]$)
   
   (c) Using the Laplace transform, verify your result of the mean and variance.
Chapter 5

Introduction of Queueing Systems

5.1 Foundation of Performance Evaluation

Queueing system is the key mathematical concept for evaluation of systems. The features for the queueing systems:

1. Public: many customers share the system

2. Limitation of resources: There are not enough resources if all the customers try to use them simultaneously.

3. Random: Uncertainty on the Customer’s behaviors

Many customers use the limited amount of resources at the same time with random environment. Thus, we need to estimate the performance to balance the quality of service and the resource.

Example 5.1. Here are some examples of queueing systems.

- manufacturing system.

- Casher at a Starbucks coffee shop.

- Machines in Lab room.

- Internet.
5.2 Starbucks

Suppose you are a manager of starbucks coffee shop. You need to estimate the customer satisfaction of your shop. How can you observe the customer satisfaction? How can you improve the performance effectively and efficiently?

Problem 5.1. By the way, most Starbucks coffee shops have a pick-up station as well as cashier. Can you give some comment about the performance of the system?

5.3 Specification of queueing systems

System description requirements of queueing systems:

- Arrival process (or input)
- Service time
- The number of server
- Service order

Let $C_n$ be the $n$-th customer to the system. Assume the customer $C_n$ arrives to the system at time $T_n$. Let $X_n = T_{n+1} - T_n$ be the $n$-th interarrival time. Suppose the customer $C_n$ requires to the amount of time $S_n$ to finish its service. We call $S_n$ the service time of $C_n$.

We assume that both $X_n$ and $S_n$ are random variable with distribution functions $P\{X_n \leq x\}$ and $P\{S_n \leq x\}$.

Definition 5.1. Let us define some terminologies:

- $E[X_n] = \frac{1}{\lambda}$: the mean interarrival time.
- $E[S_n] = \frac{1}{\mu}$: the mean service time.
- $\rho = \frac{\lambda}{\mu}$: the mean utilizations. The ratio of the input vs the service. We often assume $\rho < 1$, for the stability.

Let $W_n$ be the waiting time of the $n$-th customer. Define the sojourn time $Y_n$ of $C_n$ by

$$Y_n = W_n + S_n.$$  \hspace{1cm} (5.1)

Problem 5.2. What is $W_n$ and $Y_n$ in Starbucks coffee shop?
5.4 Little’s formula

One of the "must" for performance evaluation. No assumption is needed to prove this!

**Definition 5.1.** Here’s some more definitions:

- \( A(t) \): the number of arrivals in \([0,t)\).
- \( D(t) \): the number of departures in \([0,t)\).
- \( R(t) \): the sum of the time spent by customer arrived before \( t \).
- \( N(t) \): the number of customers in the system at time \( t \).

The relation between the mean sojourn time and the mean queue length.

**Theorem 5.1** (Little’s formula).

\[
E[N(t)] = \lambda E[Y].
\]

**Proof.** Seeing Figure 5.4, it is easy to find

\[
\sum_{n=0}^{A(t)} Y_n = \int_0^t N(t) dt = R(t). \tag{5.2}
\]

Dividing both sides by \( A(t) \) and taking \( t \to \infty \), we have

\[
E[Y(t)] = \lim_{t \to \infty} \frac{1}{A(t)} \sum_{n=0}^{A(t)} Y_n = \lim_{t \to \infty} \frac{t}{A(t)} \int_0^t N(t) dt = \frac{E[N(t)]}{\lambda}, \tag{5.3}
\]

since \( \lambda = \lim_{t \to \infty} A(t)/t \).

**Example 5.2** (Starbucks coffee shop). Estimate the sojourn time of customers, \( Y \), at the service counter.

We don’t have to have the stopwatch to measure the arrival time and the received time of each customer. In stead, we can just count the number of orders not served, and observe the number of customer waiting in front of cashier.
Figure 5.1: Little’s formula
Then, we may find the average number of customer in the system,

\[ E[N(t)] = 3. \] (5.4)

Also, by the count of all order served, we can estimate the arrival rate of customer, say

\[ \lambda = 100. \] (5.5)

Thus, using Little’s formula, we have the mean sojourn time of customers in Starbucks coffee shop,

\[ E[Y] = \frac{E[N]}{\lambda} = 0.03. \] (5.6)

**Example 5.3 (Excercise room).** Estimate the number of students in the room.

- \( E[Y] = 1 \): the average time a student spent in the room (hour).
- \( \lambda = 10 \): the average rate of incoming students (students/hour).
- \( E[N(t)] = \lambda E[Y] = 10 \): the average number of students in the room.

**Example 5.4 (Toll gate).** Estimate the time to pass the gate.

- \( E[N(t)] = 100 \): the average number of cars waiting (cars).
- \( \lambda = 10 \): the average rate of incoming cars (students/hour).
- \( E[Y] = \frac{E[N]}{\lambda} = 10 \): the average time to pass the gate.

### 5.5 Lindley equation and Loynes variable

Here’s the ”Newton” equation for the queueing system.

**Theorem 5.2 (Lindley Equation).** For a one-server queue with First-in-first-out service discipline, the waiting time of customer can be obtained by the following iteration:

\[ W_{n+1} = \max(W_n + S_n - X_n, 0). \] (5.7)
5.6. EXERCISE

The Lindley equation governs the dynamics of the queueing system. Although it is hard to believe, sometimes the following alternative is much easier to handle the waiting time.

**Theorem 5.3** (Loyne’s variable). *Given that* $W_1 = 0$, *the waiting time* $W_n$ *can be also expressed by*

\[
W_n = \max_{j=0,1,...,n-1} \{ \sum_{i=j}^{n-1} (S_i - X_i), 0 \}.
\]  \hspace{1cm} (5.8)

**Proof.** Use induction. It is clear that $W_1 = 0$. Assume the theorem holds for $n-1$. Then, by the Lindley equation,

\[
W_{n+1} = \max(W_n + S_n - X_n, 0)
\]

\[
= \max \left( \sup_{j=1,...,n-1} \{ \sum_{i=j}^{n-1} (S_i - X_i), 0 \} + S_n - X_n, 0 \right)
\]

\[
= \max \left( \sup_{j=1,...,n-1} \{ \sum_{i=j}^{n} (S_i - X_i), S_n - X_n \}, 0 \right)
\]

\[
= \max_{j=0,1,...,n} \{ \sum_{i=j}^{n} (S_i - X_i), 0 \}.
\]

\[\square\]

5.6 Exercise

1. Restaurant Management

Your friend owns a restaurant. He wants to estimate how long each customer spent in his restaurant, and asking you to cooperate him. (Note that the average sojourn time is one of the most important index for operating restaurants.) Your friend said he knows:

- The average number of customers in his restaurant is 10,
- The average rate of incoming customers is 15 per hour.

How do you answer your friend?
2. Modelling PC

Consider how you can model PCs as queueing system for estimating its performance. Describe what is corresponding to the following terminologies in queueing theory.

- customers
- arrival
- service
- the number of customers in the system

3. Web site administration

You are responsible to operate a big WWW site. A bender of PC-server proposes you two plans, which has the same cost. Which plan do you choose and describe the reason of your choice.

- Use 10 moderate-speed servers.
- Use monster machine which is 10 times faster than the moderate one.
Chapter 6

$M/M/1$ queue

The most important queueing process!

- Arrival: Poisson process
- Service: Exponential distribution
- Server: One
- Waiting room (or buffer): Infinite

6.1 $M/M/1$ queue as a birth and death process

Let $N(t)$ be the number of customer in the system (including the one being served). $N(t)$ is the birth and death process with,

- $\lambda_k = \lambda$
- $\mu_k = \mu$

\[
P(\text{Exactly one cusmter arrives in } [t, t + \Delta t]) = \lambda \Delta t \quad (6.1)
\]
\[
P(\text{Given at least one customer in the system, exactly one service completes in } [t, t + \Delta t]) = \mu \Delta t \quad (6.2)
\]

An $M/M/1$ queue is the birth and death process with constant birth rate (arrival rate) and constant death rate (service rate).
Theorem 6.1 (Steady state probability of $M/M/1$ queue). When $\rho < 1$,

$$p_k = P[N(t) = k] = (1 - \rho)\rho^k,$$

(6.3)

where $\rho = \lambda/\mu$.

Proof. The balance equation:

$$\lambda p_{k-1} = \mu p_k$$

(6.4)

Solving this,

$$p_k = \left(\frac{\lambda}{\mu}\right)^k p_0.$$  

(6.5)

Then, by the fact $\sum p_k = 1$, we have

$$p_k = (1 - \rho)\rho^k.$$  

(6.6)

It is easy to see,

$$E[N(t)] = \frac{\rho}{(1 - \rho)}.$$  

(6.7)

6.2 Utilization

The quantity $\rho$ is said to be a utilization of the system.

Since $p_0$ is the probability of the system is empty,

$$\rho = 1 - p_0$$  

(6.8)

is the probability that the em is under service.

When $\rho \geq 1$, the system is unstable. Indeed, the number of customers in the system will go to $\infty$.

6.3 Waiting Time Estimation

For the customer of the system the waiting time is more important than the number of customers in the system. We have two different ways to obtain the information of waiting time.
6.3. WAITING TIME ESTIMATION

6.3.1 Waiting Time by Little’s Formula

Let $S_n$ be the service time of $n$-th customer and $W_n$ be his waiting time. Define the sojourn time $Y_n$ by

$$Y_n = W_n + S_n.$$  \hspace{1cm} (6.9)

**Theorem 6.2.**

$$E[W] = \frac{\rho/\mu}{1 - \rho}.$$  \hspace{1cm} (6.10)

**Proof.** By Little’s formula, we have

$$E[N] = \lambda E[Y].$$  \hspace{1cm} (6.11)

Thus,

$$E[W] = E[Y] - E[S] = \frac{E[N]}{\lambda} - \frac{1}{\mu}$$

$$= \frac{\rho/\mu}{1 - \rho}.$$  \hspace{1cm} \qed

6.3.2 Waiting Time Distribution of $M/M/1$ Queues

Little’s formula gives us the estimation of mean waiting time. But what is the distribution?

**Lemma 6.1** (Erlang distribution). Let $\{S_i\}_{i=1,...,m}$ be a sequence of independent and identical random variables, which has the exponential distribution with its mean $1/\mu$. Let $X = \sum_{i=1}^{m} S_i$, then we have

$$dP\{X \leq t\} = \frac{\mu(\mu t)^{m-1}}{(m-1)!}e^{-\mu t}.$$  \hspace{1cm} (6.12)

**Proof.** Consider a Poisson process with the rate $\mu$. Let $A(t)$ be the number of arrivals by time $t$. Since $\{t < X \leq t+h\} = \{A(t) = m-1, A(t+h) = m\}$, we have

$$P\{t < X \leq t+h\} = P\{A(t) = m-1, A(t+h) = m\}$$

$$= P\{A(t+h) = m \mid A(t) = m-1\} P\{A(t) = m-1\}$$

$$= (\mu h + o(h)) \frac{(\mu t)^{m-1}}{(m-1)!} e^{-\mu t}.$$

Dividing both sides by $h$, and taking $h \to 0$, we have (6.12).  \hspace{1cm} \qed
Theorem 6.3. Let $W$ be the waiting time of the $M/M/1$ queue. Then we have

$$P[W \leq w] = 1 - \rho e^{-\mu(1-\rho)w}. \quad (6.13)$$

Further, let $V$ be the sojourn time (waiting time plus service time) of the $M/M/1$ queue, then $V$ is exponential random variable, i.e.,

$$P[V \leq v] = 1 - e^{-\mu(1-\rho)v}. \quad (6.14)$$

Proof. By condition on the number of customers in the system $N$ at the arrival, we have

$$P[W \leq w] = \sum_{n=0}^{\infty} P[W \leq w \mid N = n] P[N = n]$$

Let $S_i$ be the service time of each customer in the system at the arrival. By the lack of memory, the remaining service time of the customer in service is again exponentially distributed. $P[W \leq w \mid N = n] = P[S_1 + \ldots + S_n \leq w]$ for $n > 0$. Using Lemma 6.1, we have

$$P[W \leq w] = \sum_{n=1}^{\infty} \int_0^w \frac{\mu (\mu t)^{n-1}}{(n-1)!} e^{-\mu t} dt P[N = n] + P[N = 0]$$

$$= \int_0^w \sum_{n=1}^{\infty} \frac{\mu (\mu t)^{n-1}}{(n-1)!} e^{-\mu t} (1-\rho) \rho^n dt + 1 - \rho$$

$$= 1 - \rho e^{-\mu(1-\rho)w}. \quad (6.15)$$

\[\square\]


### 6.4 Example

Consider a WWW server. Users access the server according to Poisson process with its mean 10 access/sec. Each access is processed by the server according to the time, which has the exponential distribution with its mean 0.01 sec.

- $\lambda = 10$.
- $1/\mu = 0.01.$
6.5. EXCERCISE

- $\rho = \lambda / \mu = 10 \times 0.01 = 0.1$.

Thus the system has the utilization 0.1 and the number of customer waiting is distributed as,

$$P[N(t) = n] = (1 - \rho)\rho^n = 0.9 \times 0.1^n \quad (6.16)$$

$$E[N(t)] = \frac{\rho}{(1 - \rho)} = 1/9. \quad (6.17)$$

6.5 Excercise

1. For an $M/M/1$ queue,
   
   (a) Calculate the mean and variance of $N(t)$.
   (b) Show the graph of $E[N(t)]$ as a function of $\rho$. What can you say when you see the graph?
   (c) Show the graph of $P[W \leq w]$ with different set of $(\lambda, \mu)$. What can you say when you see the graph?

2. In Example 6.4, consider the case when the access rate is 50/sec
   
   (a) Calculate the mean waiting time.
   (b) Draw the graph of the distribution of $N(t)$.
   (c) Draw the graph of the distribution of $W$.
   (d) If you consider 10 seconds is the maximum time to wait for the WWW access, how much access can you accept?

3. Show the equation (6.15).
Chapter 7

Reversibility

7.1 Output from Queue

What is output from a queue?

To consider a network of queues, this is very important question, since the output from a queue is indeed the input for the next queue.

7.2 Reversibility

Reversibility of stochastic process is useful to consider the output for the system. If the process is reversible, we may reverse the system to get the output.

7.2.1 Definition of Reversibility

Definition 7.1. $X(t)$ is said to be reversible when $(X(t_1), X(t_2), ..., X(t_m))$ has the same distribution as $(X(\tau - t_1), ..., X(\tau - t_m))$ for all $\tau$ and for all $t_1, ..., t_m$.

The reverse of the process is stochastically same as the original process.

7.2.2 Local Balance for Markov Chain

Transition rate from $i$ to $j$:

$$q_{ij} = \lim_{\tau \to 0} \frac{P[X(t+\tau) = j \mid X(t) = i]}{\tau} \quad (i \neq j) \quad (7.1)$$

$$q_{ii} = 0 \quad (7.2)$$
7.2. **REVERSIBILITY**

We may have two balance equations for Markov chain.

1. Global balance equation:

   \[ \sum_j p_i q_{ij} = \sum_j p_j q_{ji}. \]  
   \(7.3\)

2. Local balance equation:

   \[ p_i q_{ij} = p_j q_{ji}. \]  
   \(7.4\)

**Theorem 7.1** (Local Balance and Reversibility). A stationary Markov chain is reversible, if and only if there exists a sequence \(\{p_i\}\) with

\[\sum_i p_i = 1 \quad \text{and} \quad p_i q_{ij} = p_j q_{ji}. \]  
\(7.5\) \(7.6\)

**Proof.**  
1. Assume the process is reversible, then we have

   \[ P[X(t) = i, X(t + \tau) = j] = P[X(t) = j, X(t + \tau) = i]. \]

   Using Bayes theorem and taking \(\tau \to 0\), we have

   \[ p_i q_{ij} = p_j q_{ji}. \]

2. Assume there exists \(\{p_i\}\) satisfying (7.5) and (7.6). Since the process is Markov, we have

   \[ P[X(u) \equiv i \text{ for } u \in [0,t), X(u) \equiv j \text{ for } u \in [t,t+s)] = p_i e^{-q_i t} q_{ij} e^{-q_j s}, \]

   where \(q_i = \sum_j q_{ij}\).

   Using the local balance, we have

   \[ p_i e^{-q_i t} q_{ij} e^{-q_j s} = p_j e^{-q_j s} q_{ji} e^{-q_i t}. \]

   Thus

   \[ P[X(u) \equiv i \text{ for } u \in [0,t), X(u) \equiv j \text{ for } u \in [t,t+s)] = P[X(u) \equiv i \text{ for } u \in [-t,0), X(u) \equiv j \text{ for } u \in [-t-s,-t)] \].
7.3 Output from $M/M/1$ queue

**Theorem 7.2** (Burke’s Theorem). If $\rho = \lambda / \mu < 1$, the output from an equilibrium $M/M/1$ queue is a Poisson process.

**Proof.** Let $N(t)$ be the number of the customers in the system at time $t$. The point process of upward changes of $N(t)$ corresponds to the arrival process of the queue, which is Poisson. Recall the transition rate of $N(t)$ of the $M/M/1$ queue satisfies

$$\lambda P_{n-1} = \mu P_n,
$$

which is the local balance equation. Thus, by Theorem 7.1 $N(t)$ is reversible and $N(-t)$ is stochastically identical to $N(t)$. Thus the upward changes of $N(t)$ and $N(-t)$ are stochastically identical. In fact, the upward change of $N(-t)$ corresponds to the departures of the $M/M/1$ queue. Thus, the departure process is Poisson.

7.4 Exercise

1. If you reverse an $M/M/1$ queue, what is corresponding to the waiting time of $n$-th customer in the original $M/M/1$? (You may make the additional assumption regarding to the server, if necessary.)

2. Consider two queues, whose service time is exponentially distributed with their mean $1/\mu$. The input to one queue is Poisson process with rate $\lambda$, and its output is the input to other queue. What is the average total sojourn time of two queues?

3. Can you say that $M/M/2$ is reversible? If so, explain why.

4. Can you say that $M/M/m/m$ is reversible? If so, explain why. (You should clearly define the output from the system.)
Chapter 8

Network of Queues

8.1 Open Network of queues

We will consider an open Markovian-type network of queues (Jackson network).

Specifications:

- \( M \): the number of nodes (queues) in the network.
- Source and sink: One for entire network.
- \( i \): index of node.
- \( r_{ij} \): routing probability from node \( i \) to \( j \) (Bernoulli trial).
- Open network: allow the arrival and departure from outside the network.
  - \( r_{is} \): routing probability from node \( i \) to sink (leaving the network).
  - \( r_{si} \): probability of arrival to \( i \) from outside (arrival to the network).
- The arrival from the outside: Poisson process (\( \lambda \))
- The service time of the each node: exponential (\( \mu \))

8.2 Global Balance of Network

- \( (N \cup \{0\})^M \): State space (\( M \)-dimensional).
- \( \bar{n} = (n_1, ..., n_M) \): the vector of the number customers in each node.
• $1_i = (0, \ldots, 0, 1, 0, \ldots, 0)$

• $p(n)$: the probability of the number customers in the network is $n$.

The global balance equation of Network:

**Theorem 8.1** (Global Balance Equation of Jackson Network). *In the steady state of a Jackson network, we have*

$$[\lambda + \sum_{i=1}^{M} \mu_i] p(n) = \sum_{i=1}^{M} \lambda r_{si} p(n-1_i) + \sum_{i=1}^{M} \mu_i r_{is} p(n + 1_i) + \sum_{i=1}^{M} \sum_{j=1}^{M} \mu_j r_{ji} p(n + 1_j - 1_i).$$

(8.1)

**Proof.** The left-hand side correspond to leaving the state $n$ with an arrival or departure. Each terms on the right correspond to entering the state $n$ with an arrival, departure and routing. Thus, the both side should be balanced in the steady state. \qed

### 8.3 Traffic Equation

Let $\Lambda_i$ be the average output (throughput) of a queue $i$.

**Theorem 8.2** (Traffic Equations).

$$\Lambda_i = \lambda r_{si} + \sum_{j=1}^{M} r_{ji} \Lambda_j$$

(8.2)

**Proof.** Consider the average input and output from a queue $i$. The output from $i$ should be equal to the internal and external input to $i$. \qed

**Remark 8.1.** The average output ($\Lambda_i$) can be obtained by solving the system of equations (8.2).

### 8.4 Product Form Solution

**Theorem 8.3** (The steady state probabilities of Jackson network). *The solution of the global balance equation (8.1):*
8.4. PRODUCT FORM SOLUTION

\[ p(n) = \prod_{i=1}^{M} p_i(n_i), \quad (8.3) \]

and

\[ p_i(n_i) = (1 - \rho_i) \rho_i^{n_i}, \quad (8.4) \]

where \( \rho_i = \Lambda_i / \mu_i \).

Proof. First we will show \( p_i \), which satisfy the "local balance"-like equation

\[ \Lambda_i p(n - 1_i) = \mu_i p(n), \quad (8.5) \]

also satisfyies (8.1). Rearranging (8.2), we have

\[ \lambda r_{si} = \Lambda_i - \sum_{j=1}^{M} r_{ji} \Lambda_j. \quad (8.6) \]

Use this to eliminate \( \lambda r_{si} \) from (8.1), then we have

\[
[\lambda + \sum_{i=1}^{M} \mu_i] p(n) = \sum_{i=1}^{M} \Lambda_i p(n - 1_i) - \sum_{i=1}^{M} \sum_{j=1}^{M} r_{ji} \Lambda_i p(n - 1_i) \\
+ \sum_{i=1}^{M} \mu_i r_{is} p(n + 1_i) + \sum_{i=1}^{M} \sum_{j=1}^{M} \mu_j r_{ji} p(n + 1_j - 1_i).
\]

Using (8.5), we have

\[ \lambda = \sum_{i=1}^{M} r_{is} \Lambda_i, \quad (8.7) \]

which is true because the net flow into the network is equal to the net flow out of the network. Thus, a solution satisfying the local balance equation will be automatically the solution of the global balance equation. By solving the local balance equation, we have

\[
p(n) = \frac{\Lambda_i}{\mu_i} p(n - 1_i) \\
= (\Lambda_i / \mu_i)^{n_i} p(n_1, ..., 0, ..., n_M).
\]
CHAPTER 8. NETWORK OF QUEUES

Using this argument for different \( i \), we have

\[
p(n) = \prod_{i=1}^{M} \left( \frac{\Lambda_i}{\mu_i} \right)^n p(\emptyset).
\]

Taking summation over \( n \) and find \( p(\emptyset) \) to be the product of

\[
p_i(0) = 1 - \frac{\Lambda_i}{\mu_i}.
\]

(8.8)

Remark 8.2. The steady state probability of Jackson network is the product of the steady state probability of \( M/M/1 \) queues.

8.5 Exercises

1. Consider two queues in the network.

   - \( \lambda \): external arrival rate, \( \mu_i \): the service rate of the node \( i \).
   - \( r_{s1} = 1, r_{s2} = 0 \)
   - \( r_{11} = 0, r_{12} = 1, r_{1s} = 0 \)
   - \( r_{21} = 3/8, r_{22} = 2/8, r_{2s} = 3/8 \)

   (a) Estimate the throughputs \( \Lambda_i \) for each queue.
   (b) Estimate the steady state probability of the network.
   (c) Draw the 3-D graph of the steady state probability for \( \lambda = 1 \) and \( \mu_1 = 3 \) and \( \mu_2 = 4 \).
   (d) Estimate the mean sojourn time of the network of (1c) by using Little’s Law.
Chapter 9

Examples of Queueing System Comparison

Problem 9.1. Which is better,

1. one fast server or two moderate servers?

2. tandem servers or parallel servers?

3. independent servers or joint servers?

4. Starbucks or Doutor Coffee?

9.1 Single Server vs Tandem Servers

Example 9.1 (Starbucks vs Doutor Coffee). Suppose we have a stream of customer arrival as Poisson process with the rate $\lambda$ to a coffee shop.

At Doutor Coffee, they take ordering, payment and coffee serving at the same time. Assume it will take an exponentially-distributed service time with its mean $1/\mu$. On the other hand, Starbucks coffee separate the ordering and coffee service. Assume both will take independent exponentially-distributed service times with its mean $1/2\mu$. Thus, the mean overall service time in both coffee shop is same and equal to $1/\mu$.

Problem 9.2. Which coffee shop is better?
At Doutor Coffee, we have only one $M/M/1$ queue, and as we saw in Chapter 6 the expected number of customer in line including the one serving is obtained by

$$E[N_{Doutor}] = \frac{\rho}{1 - \rho}, \quad (9.1)$$

where $\rho = \lambda/\mu$ as usual. At Starbucks, we have two queues $N_1$ at the ordering area, and $N_2$ at the serving area. This can be modeled by a tandem queue. By Theorem 7.2, we know that the output of $M/M/1$ queue ordering area is again Poisson process with the rate $\lambda$. Thus, the second queue at the serving area is again modeled by $M/M/1$ queue. Since the service rate of both queue is $2\mu$, we have the expected number of customers at Starbucks as

$$E[N_{Starbucks}] = E[N_1] + E[N_2] = \frac{\lambda/2\mu}{1 - \lambda/2\mu} + \frac{\lambda/2\mu}{1 - \lambda/2\mu} = \frac{2\rho}{2 - \rho}, \quad (9.2)$$

Consequently,

$$E[N_{Doutor}] = \frac{\rho}{1 - \rho} > \frac{2\rho}{2 - \rho} = E[N_{Starbucks}], \quad (9.3)$$

and by Little’s formula, the expected sojourn time is obtained by

$$E[Y_{Doutor}] > E[Y_{Starbucks}]. \quad (9.4)$$

**Problem 9.3.** Give an intuitive reason of this consequence.

### 9.2 $M/M/2$ queue

Specifications:

- Arrival: Poisson process
- Service: Exponential distribution
- Server: Two
9.3. $M/M/1$ VS $M/M/2$

- Waiting room (or buffer): Infinite

Birth and death coefficients:

$$\lambda_k = \lambda$$  \hspace{1cm} (9.5)

$$\mu_k = \begin{cases} 
\mu & k = 1 \\
2\mu & k \geq 2 
\end{cases}$$  \hspace{1cm} (9.6)

The balance equations for stationary state probability:

$$-\lambda P_0 + \mu P_1 = 0$$  \hspace{1cm} (9.7)

$$-(\lambda + \mu)P_1 + \lambda P_0 + 2\mu P_2 = 0$$  \hspace{1cm} (9.8)

$$-(\lambda + 2\mu)P_k + \lambda P_{k-1} + 2\mu P_{k+1} = 0 \ (k \geq 2),$$  \hspace{1cm} (9.9)

or

$$\mu P_1 = \lambda P_0$$  \hspace{1cm} (9.10)

$$2\mu P_k = \lambda P_{k-1} \ (k \geq 1).$$  \hspace{1cm} (9.11)

Solving the equations inductively:

$$p_k = 2p_0 \left( \frac{\lambda}{2\mu} \right)^k \text{ for } k \geq 1$$  \hspace{1cm} (9.12)

Using $\Sigma p_k = 1$, we have

$$p_0 = \frac{1 - \frac{\lambda}{2\mu}}{1 + \frac{\lambda}{2\mu}}$$  \hspace{1cm} (9.13)

The average number of customers in the system:

$$E[N(t)] = \frac{\lambda/\mu}{1 - (\lambda/2\mu)^2}$$  \hspace{1cm} (9.14)

9.3  $M/M/1$ VS $M/M/2$

Consider two servers ($A_1, A_2$) and one server ($B$) which is two-times faster than the two servers. So the average power of $M/M/2$ ($A_1, A_2$) and $M/M/1$ ($B$) are equivalent.
• \( \lambda \): The mean arrival rate for both queues.
• \( 2/\mu \): The mean service time of the servers \( A_i \).
• \( 1/\mu \): The mean service time of the servers \( B \).

Set \( \rho = \lambda / \mu \).

The mean queue lengths:

• \( E[N_{M/M/2}] = \frac{2\rho}{1-\rho^2} \)
• \( E[N_{M/M/1}] = \frac{\rho}{1-\rho} \)

It is easy to see

\[
E[N_{M/M/2}] = \frac{2\rho}{1-\rho^2} > \frac{\rho}{1-\rho} = E[N_{M/M/1}]. \quad (9.15)
\]

Thus, by using Little’s formula, the sojourn time of the system (delay) can be evaluated by:

\[
E[Y_{M/M/2}] > E[Y_{M/M/1}]. \quad (9.16)
\]

**Problem 9.4.** Give an intuitive reason of this consequence.

### 9.4 Two \( M/M/1 \) VS One \( M/M/2 \)

Consider two systems which have the same arrival rate:

1. Two independent \( M/M/1 \) queue.
2. One \( M/M/2 \) queue, i.e. sharing two servers forming one queue.

Four servers have the same speed.

• \( \lambda \): The mean arrival rate for both systems.
• \( 1/\mu \): The mean service time of the servers.

Set \( \rho = \lambda / (2\mu) \).

\[
E[N_{M/M/2}] = \frac{\rho}{1-\rho^2/4} < \frac{\rho}{1-\rho/2} = E[N_{M/M/1}] + E[N_{M/M/1}]. \quad (9.17)
\]

Thus, by using Little’s formula, the sojourn time of the system (delay) can be evaluated by:

\[
E[Y_{M/M/2}] < E[Y_{2(M/M/1)}]. \quad (9.18)
\]
9.5 Exercises

1. In the setting of Section 9.3:
   
   (a) Draw the graph of the steady state probabilities of $M/M/1$ and $M/M/2$ vs $\rho$.
   
   (b) Draw the graph of the mean sojourn time of $M/M/1$ and $M/M/2$ vs $\rho$.
   
   (c) What can you say about when $\rho \to 1$?

2. Web site administration (revisited) You are responsible to operate a big WWW site. A bender of PC-server proposes you two plans, which has the same cost. Which plan do you choose and describe the reason of your choice.
   
   - Use 2 moderate-speed servers.
   - Use monster machine which is 2 times faster than the moderate one.

3. Web site administration (extended) You are responsible to operate a big WWW site. Your are proposed two plans, which has the same cost. Which plan do you choose and describe the reason of your choice.
   
   - Use 2 moderate-speed servers and hook them to different ISPs.
   - Use 2 moderate-speed servers and hook them to single ISP, using a load-balancer. (Assume the Load-balancer ideally distribute the load to each server.)
Chapter 10

Repairman Problem in Manufacturing System

This part is based on the book: John A. Buzacott and J. George Shanthikumar, Stochastic Models of Manufacturing Systems [Buzacott and Shanthikumar, 1992].

10.1 Repairman Problem

Suppose we have machines in our factory to make some products. All machines are equivalent and its production rate is $h$ unit per unit time. Unfortunately, our machines are sometimes fail and needs to be repaired. We have repairmen to fix them.

Each repairmen are responsible to some fixed number of machines, say $m$. Assume we have large $m$, then we have considerable probability that more than two machines are failed at the same time. The repairman is busy fixing one machine and the rest of the failed machines are waiting to be repaired. So, the production rate per machine (we call it “efficiceny”) will be down. On the other hand, for large $m$, we may have a good production efficiency, but we may find our repairman is idle quite often.

Thus, we have the basic tradeoff of the performance of repairman and the productivity of each machines.

Problem 10.1. How can you set the approrpriate number of machines $m$ assigned to one repairman satisfying the above tradeoff?
10.2 Average Analysis

In this section, we will see what we can say about our repairman problem without using stochastic argument.

Let \( P_n \) be the proportional time of \( n \) machines are down. By observing our factory, we can easily estimate \( P_n \). Once we obtain the information \( P_n \), we can estimate various performance measures of our factory. For example, the total production rate in our factory can be calculated by

\[
H = \sum_{n=0}^{m} (m-n)P_nh.
\]  

(10.1)

**Definition 10.1** (System efficiency). Let \( G = mh \) be the total production rate of our factory assuming that all machines are available. Then, the system efficiency of our factory \( \eta \) is defined by

\[
\eta = \frac{H}{G} = 1 - \sum_{n=0}^{m} \frac{n P_n}{m}.
\]  

(10.2)

**Problem 10.2.** Derive the right-hand side of (10.2).

**Remark 10.1.** Using the system efficiency \( \eta \), we can estimate the impact of the failure of our machines. Of course, we can expect the system efficiency \( \eta \) will be improved if we reduce the number of machine assigned to a repairman.

Here’s another performance measure for our factory.

**Definition 10.2.** Define the operator utility \( \rho \) by

\[
\rho = \sum_{n=1}^{m} P_n = 1 - P_0,
\]  

(10.3)

which estimate the time ratio during our repairman is working.

Now, suppose the expected repair time of a machine is \( 1/\mu \). Let \( D(t) \) be the number of repaired machines up to time \( t \). Then, in the long run (for large \( t \)), we have

\[
D(t) = \mu \rho t,
\]  

(10.4)
since operator is working for $\rho t$. Further, let $1/\hat{\lambda}$ be the mean time to a failure of a single machine (we say $\hat{\lambda}$ is the failure rate). Then, $A(t)$, the number of machine failed up to time $t$ can be estimated by

$$A(t) = \lambda \sum_{n=0}^{m} (m-n) P_n t$$

$$= \lambda \eta mt. \quad (10.5)$$

**Problem 10.3.** Derive (10.5).

In the long run, we can expect the number of failed machine is equal to the number of repaired machines.

**Problem 10.4.** Why?

Thus, for larger $t$, we have

$$A(t) = D(t),$$

$$\lambda \eta mt = \mu \rho t.$$  

Rearranging the above, we have the relation between the system efficiency $\eta$ and the operator utility $\rho$.

$$\eta = \rho \frac{\mu}{m \lambda}. \quad (10.6)$$

Further, using this relation, the production rate $H$ can be represented by

$$H = \eta G = \eta mh = \rho \frac{\mu h}{\lambda}. \quad (10.7)$$

**Remark 10.2.** The equation (10.6) clearly implies that the reduction of number of machines assigned to one repairman will result the better system efficiency. However, the situation here is not so easy. If we decrease $m$, it will affect the the operator utility $\rho$ (typically we may see smaller $\rho$). We will see this effect in the following section.

As a factory manager, you may not be satisfied with the good operator utility. You may want to know the number of machine waiting for repair, $L$. The
10.2. AVERAGE ANALYSIS

expectation of $L$ can be given by

$$E[L] = \sum_{n=1}^{m} (n-1)P_n$$

$$= \sum_{n=1}^{m}nP_n - \sum_{n=1}^{m}P_n$$

$$= m(1-\eta) - \rho$$

$$= m - \rho \frac{\mu}{\lambda} - \rho,$$

where we used the relation (10.6). Thus,

$$E[L] = m(1-\eta) - \rho = m - \rho (1 + \frac{\mu}{\lambda}).$$

(10.8)

Also, let $T$ be the down time of a machine. Since the total down time in $[0,t)$

is given by

$$\sum_{n=1}^{m}nP_nt = m(1-\eta)t,$$

(10.9)

the average down time of a machine is

$$E[T] = \frac{m(1-\eta)t}{A(t)}$$

(10.10)

$$= \frac{1-\eta}{\lambda \eta}.$$ 

(10.11)

Let $N(t)$ be the number of machine not operating at time $t$, then

$$E[N(t)] = \sum_{n=1}^{m}nP_n = m(1-\eta).$$

(10.12)

Setting $\lambda^*$ as the average number of machine failure per unit time, we have

$$\lambda^* = \lim_{t \to \infty} \frac{A(t)}{t} = \lambda \eta m.$$ 

(10.13)

Combining (10.10), (10.12) and (10.13), we have a relation between these quantities:

$$E[N(t)] = \lambda^* E[T],$$

(10.14)

which turns out to be familiar Little’s formula (see Theorem 5.1).

Problem 10.5. Use Theorem 5.1 to prove (10.14).
10.3 Single Repairman in Markov Model

Assume we have $m$ machines. Machines can be failed and one repairman is responsible to fix it. Take one specific machine. Let $T$ be the time until this machine is failed and assume $T$ to be an exponential random variable with its mean $1/\lambda$, i.e.

$$P\{T \leq x\} = 1 - e^{-\lambda x}. \quad (10.15)$$

Also, our repairman will take the time $S$ to fix this machine. We assume $S$ is also an exponential random variable with its mean $1/\mu$, i.e.

$$P\{S \leq x\} = 1 - e^{-\mu x}. \quad (10.16)$$

As we discussed before, those exponential random variables have memoryless property. Thus, we make the Markov chain by modeling our $m$ machines. Let $N(t)$ be the number of machines down at time $t$. Then $N(t)$ is a Markov chain with its infinitesimal operator $Q$ as

$$Q = \begin{pmatrix}
-m\lambda & m\lambda & -\mu - (m-1)\lambda & (m-1)\lambda \\
\mu & -\mu & \ddots & \ddots \\
\mu & \ddots & \ddots & \lambda \\
\mu & \cdots & \cdots & -\mu
\end{pmatrix}. \quad (10.17)$$

Note that we have $m$ machines and one repairman. Thus, as the number of machines down increases, the rate of having one more machine failed decreases, while the repair rate remained same.

Let $P_n$ be the stationary probability of $N(t)$, i.e.

$$P_n = \lim_{t \to \infty} P\{N(t) = n\}. \quad (10.18)$$

Then, the balance equations can be found as

$$m\lambda p_0 = \mu p_1,$$

$$(m-1)\lambda p_1 = \mu p_2,$$

$$\ldots$$

$$(m-n+1)\lambda p_{n-1} = \mu p_n,$$

$$\ldots$$

$$\lambda p_{m-1} = \mu p_m. \quad (10.19)$$
Thus, we have
\[ P_n = \left( \frac{\lambda}{\mu} \right) (m-n+1) P_{n-1} \]
\[ = \left( \frac{\lambda}{\mu} \right)^2 (m-n+1)(m-n+2) P_{n-1} \]
\[ = \left( \frac{\lambda}{\mu} \right)^n (m-n+1) \cdots m P_0 \]
\[ = \left( \frac{\lambda}{\mu} \right)^n \frac{m!}{(m-n)!} P_0, \]
for \( n = 0, \ldots, m \). Here, \( P_0 \) is still unknown. Using the usual normalization condition \( \sum p_n = 1 \), we have
\[ 1 = P_0 \sum_{n=0}^{m} \left( \frac{\lambda}{\mu} \right)^n \frac{m!}{(m-n)!}. \]
Thus, we have the unknown \( P_0 \) as
\[ P_0 = \frac{1}{\sum_{n=0}^{m} \left( \frac{\lambda}{\mu} \right)^n \frac{m!}{(m-n)!}}, \quad (10.20) \]
and for \( n = 0, \ldots, m \),
\[ P_n = \left( \frac{\lambda}{\mu} \right)^n \frac{m!}{(m-n)!} P_0. \quad (10.21) \]

Now, what we want to know is the trade-off between the production rate \( H \) and the operator utility \( \rho \). From (10.3), the operator utility \( \rho \) is calculated by
\[ \rho = 1 - P_0. \quad (10.22) \]
Also, by (10.6), we have
\[ \eta = \rho \frac{\mu}{m \lambda} = \left( 1 - P_0 \right) \frac{\mu}{m \lambda}. \quad (10.23) \]

See Figure 10.1 and 10.2. Here we set \( \lambda / \mu = 100 \). We can see what will happen when we increase \( m \), the number of machines assigned to our repairman. Of course, the utility of our repairman increases as \( m \) increases. However, at some point, the utility hits 1, which means he is too busy working on repairing machine, and at this point the number of machines waiting to be repaired will explode. So the productivity of our machine decreases dramatically.

**Problem 10.6.** How can you set the appropriate number of machines assigned to a single repairman?
Figure 10.1: Operator utility $\rho$.

Figure 10.2: Efficiency of machine $\eta$. 
Chapter 11

Production Systems

In this chapter, we will discuss how we can model and analyze two basic production systems, namely produce-to-order and produce-to-stock. This part is based on the book: John A. Buzacott and J. George Shanthikumar, Stochastic Models of Manufacturing Systems [Buzacott and Shanthikumar, 1992].

11.1 Produce-to-Order System

Assume customers arrive according to Poisson Process with the rate $\lambda$ to a single server system. After the arrival of a customers, the system starts to process the job. The service time of the job (or production time) is assumed to be an exponential random variable with its mean $1/\mu$. Upon arrival of a customer, if there are any other customers waiting, the customer join the queue. We assume the order of the service is fist-come-first-serve.

Then, this system is indeed an $M/M/1$ queue that we discussed in Chapter 6. Let $N(t)$ be the number of customers in the system (both waiting in the queue and in service). What we want to know is the stationary distribution of $N(t)$, i.e.,

$$P_n = \lim_{t \to \infty} N(t).$$

(11.1)

As in Theorem 6.1, we have

$$P_n = (1 - \rho)\rho^n,$$

(11.2)

where $\rho = \lambda/\mu$ representing the system utility. Also, as in (6.7), we can derive
the expectation of $N(t)$,

$$E[N(t)] = \frac{\rho}{(1 - \rho)}.$$  \hspace{1cm} (11.3)

Let $W$ be the waiting time in the queue and $V = W + S$ be the sojourn time. The $W$ corresponds to the time until the customer receives the product. We can find the sojourn time distribution as Theorem 6.3,

$$P[V \leq x] = 1 - e^{-\mu(1-\rho)x},$$  \hspace{1cm} (11.4)

and its mean as

$$E[V] = \frac{1}{\mu(1-\rho)}.$$  \hspace{1cm} (11.5)

11.2 Produce-to-Stock System

If you order a product in the system like produce-to-order, you will find a large waiting time. One of the simplest way to reduce the waiting time and enhance the quality of service is to have stock. Of course, stocking products will incur storage cost. So, balancing the quality and the cost is our problem.

Here we assume a simple scheme. We allow at most $z$ products in our storage. We attach a tag to each product. Thus, the number of tags is $z$. Whenever there are some products in our storage upon an arrival of customers, the customer will take one product, and we will take off its tag. The tag now become the order of a new product.

Depending on the production time, occasionally, our products may go out of stock on arrival of a customer. In this case, the customer has to wait. Our problem is to set appropriate level of the number of tags, $z$.

Assume, as assumed in Section 11.1, the arrival of customer is Poisson process with the rate $\lambda$. The service time of the job (or production time) is assumed to be an exponential random variable with its mean $1/\mu$.

Let $I(t)$ be the number of products stocked in our storage at time $t$, and let $C(t)$ be the number of orders in our system (including the currently-processed item). Since we have $z$ tags,

$$I(t) + C(t) = z.$$  \hspace{1cm} (11.6)
In addition, let $B(t)$ be the number of backlogged customers waiting for their products. Define $N(t)$ by the sum of the backlogged customers, the waiting orders and the processing order, i.e.

$$N(t) = C(t) + B(t).$$ \hfill (11.7)

Then, $N(t)$ can be regarded as the number of customers in an $M/M/1$ queue, since at the arrival of customers, it will increase one, and at the service completion, it will decrease one. Thus, we have the stationary distribution of $N(t)$ as

$$P_n = \lim_{t \to \infty} P\{N(t) = n\} = (1 - \rho) \rho^n,$$ \hfill (11.8)

for $\rho = \lambda / \mu < 1$, as found in Theorem 6.1. Since $N(t)$ is the sum of waiting orders, we have $z - N(t)$ products in our storage if and only if $N(t) < z$. Thus, the number of products in our storage can be written by

$$I(t) = (z - N(t))^+. \hfill (11.9)$$

Using (11.8), we can obtain the probability distribution of the number of products in storage as

$$P\{I(t) = n\} = P\{(z - N(t))^+ = n\}$$

$$= \begin{cases} 
P\{N(t) = z - n\} = (1 - \rho) \rho^{z-n} & \text{for } n = 1, 2, \ldots, z, \\
P\{N(t) > z\} = \rho^z & \text{for } n = 0.
\end{cases} \hfill (11.10)$$

When $I(t) > 0$, the customer arrived can receive a product. Thus,

$$P\{\text{No wait}\} = P\{I(t) > 0\}$$

$$= 1 - P\{I(t) = 0\}$$

$$= 1 - \rho^z. \hfill (11.11)$$

Thus, as the number of tag increase, we can expect that our customers do not have to wait.

Also, we can evaluate the expected number of products in our storage can estimated by

$$E[I(t)] = \sum_{n=1}^{z} nP\{I(t) = n\}$$

$$= z - \rho \frac{(1 - \rho^z)}{1 - \rho}. \hfill (11.12)$$

As we can easily see that the second term is rapidly converges to $\rho / (1 - \rho)$ for large $z$, the storage cost is almost linear to the number of tag $z$. 

Problem 11.1. Show (11.10) and (11.12).

Figure 11.1: Probability of no wait: $P\{\text{No wait}\}$.

Figure 11.2: Expected number of Stocks: $E[I(t)]$.

Figure 11.1 and 11.2 shows the trade-off between quality of service and storage cost when $\rho = 0.2$.

Problem 11.2. How can you set the number of tag $z$?
Chapter 12

Utility Function

This part is based on the book of Sheldon Ross [2002].

12.1 Horse Races

There are 10 horses on the race, and you can bet the winning horse by buying a one-dollar ticket. There seems to be no advantage among horses. The odds are the same among the horses, and you can get $9 if you successfully bet on the winning horse.

However, by a deep (and mysterious) analysis, you find the horse No.7 has some edge over other horses. In fact, you find out the probability that No.7 wins the race is 1/2.

Remark 12.1. There were some real cases that you could find the edge in the horse races?.

Problem 12.1. Suppose you have $100 on hand. How are you going to bet on the race?

There seems to be two typical and (seemingly) logical answers.

A: Bet whole $100 on No.7 so as to maximize your profit if you win the race.

B: Betting the whole money on No.7 is too risky. You should also place bets on other horses.
Depending on your strategies, you will have different outcome. The outcome also depends on the outcome of race. Thus, you should take uncertainty into account when you evaluate both strategies.

Further, your choice of strategy depends on your situation and your preference to the risk.

**Problem 12.2.** Do you prefer the risk? In what situation, you should take risk?

### 12.2 Define the Utility

Suppose we have two investment options: A and B. There are \( n \) different consequences of the investments,

\[
C_1, C_2, \ldots, C_n. \tag{12.1}
\]

This corresponds how much we win in the horse race.

The money we can get in the horse race depends on how we invest money on the tickets. Thus, the probability of the consequence is different as our choice of investment, and

\[
p_i = P\{C_i|A\}, \tag{12.2}
\]
\[
q_i = P\{C_i|B\}. \tag{12.3}
\]

Let’s put numerical values for the consequences \( C_1, C_2, \ldots, C_n \) based on our preference. We put \( C_1, C_2, \ldots, C_n \) as their value. Note that our preference may not be linear to the money we can get.

**Problem 12.3.** In what situation can you say that $100 is not worth twice as much as $50?

Assume \( C_1 \) is the worst consequence and set \( C_1 = 0 \). Also the most preferable consequence is \( C_n \) and set \( C_n = 1 \). The rest will be determined as follows. Consider a thought-experiment. Imagine you can chose one of the two following options:

1. Get \( C_i \).

2. Get \( C_n \) with probability \( u \) or \( C_1 \) with probability \( 1 - u \).
12.2. DEFINE THE UTILITY

Assume you chose 2 for example. As \( u \) decreases, the choice will be switched from 2 to 1. There is the point of \( u \) where the choices are essentially same value. Using such \( u_i \), we can set \( C_i = u_i \).

**Definition 12.1.** \( C_i = u_i \) is said to be the utility of the consequence of \( C_i \).

Consider we chose the investment \( A \). Then, with probability \( p_i \), we have \( C_i \), which is equivalent to have

\[
\begin{aligned}
&\{ C_n \text{ with probability } u_i, \\
&\{ C_1 \text{ with probability } 1 - u_i. 
\end{aligned}
\]

Summing up the probability of these equivalent events for all \( i \), the investment \( A \) can be regard as

\[
\begin{aligned}
&\{ C_n \text{ with probability } \sum p_i u_i, \\
&\{ C_1 \text{ with probability } 1 - \sum p_i u_i. 
\end{aligned}
\]

**Problem 12.4.** Check (12.5).

Similarly, we can consider the investment \( B \) and it is equivalent to

\[
\begin{aligned}
&\{ C_n \text{ with probability } \sum q_i u_i, \\
&\{ C_1 \text{ with probability } 1 - \sum q_i u_i. 
\end{aligned}
\]

Thus, if we have

\[
\sum_{i=1}^{n} q_i u_i < \sum_{i=1}^{n} p_i u_i,
\]

we can say that the investment \( A \) is better than \( B \).

**Theorem 12.1.** The value of investment can be evaluated by the expectation of the utility.

Let \( u(x) \) be the investor’s utility of receiving \( x \). We have two investments \( X \) and \( Y \), and

\[
E[u(Y)] < E[u(X)].
\]

Then we will chose \( X \).
12.3 Horse Race Revisited

Suppose you have $100. You are desperate to enter a graduate school. Fortunately, you are allowed to enter the school, but the graduate school requires you to pay $900 for the initial tuition by tomorrow. Your qualification will expire after tomorrow. Your last resort is the horse race!

**A:** Bet whole $100 on No.7 so as to maximize your profit if you win the race.

**B:** Betting the whole money on No.7 is too risky. Place some of your money on other horses. Let’s say $a$ on the other horses, and $100 - 9a$ on No.7.

Let $X$ be the consequence of the strategy A and $Y$ be the consequence of the strategy B, that is

\[
X = \begin{cases} 
900 & \text{No.7 wins the race}, \\
0 & \text{otherwise}. 
\end{cases} \quad (12.9)
\]

\[
Y = \begin{cases} 
9(100 - 9a) & \text{No.7 wins the race}, \\
9a & \text{otherwise}. 
\end{cases} \quad (12.10)
\]

Because of your situation, you need $900. Money no less than $900 is nothing to you. Thus,

\[
u(x) = \begin{cases} 
1 & x = 900, \\
0 & \text{otherwise}. 
\end{cases} \quad (12.11)
\]

In this case, we have

\[
0 = E[u(Y)] < E[u(X)] = \frac{1}{2}, \quad (12.12)
\]

for any positive amount of $a$. 
12.4 Risk-averse and Risk-neutral Investors

The utility function $u(x)$ should be non-decreasing.

**Problem 12.5.** Why?

Further, $u(x)$ may be concave.

**Definition 12.2** (Risk-averse investor). When an investor has concave function, he is said to be risk-averse.

**Lemma 12.1** (Jensen’s inequality).

$$E[u(X)] \leq u(E[X]),$$ (12.13)

for a concave function $u(x)$.

Consider the following two investments strategy,

1. The return of the investment is stochastic and $X$.
2. The return of the investment has no risk and always the mean $E[X]$.

Suppose you have concave utility function $u(x)$. Because of Jensen’s inequality,

$$E[u(X)] \leq u(E[X]),$$ (12.14)

which means the investment 2 is always better than 1. Thus, you are always prefer not having risk, when you have the same expected return.

**Definition 12.3** (Risk-neutral investor). Those investors who has a linear utility function are said to be risk-neutral or risk indifferent.

Suppose you have a linear utility function of the form like

$$u(x) = ax + b,$$ (12.15)

for some $a > 0$ and $b$.

**Problem 12.6.** Why $a > 0$?

In this case,

$$E[u(X)] = a + bE[X],$$ (12.16)

and your preference is only depend on $E[X]$ not on the stochastic behavior of $X$. Thus, you are not indifferent with the risk to have low return.
12.5 Log Utility Function

Here’s one logical explanation of the usage of log utility function.

Suppose you are investing some asset long time. Let $W_n$ be your wealth at time $n$. Assume we have the following recursive relation:

$$W_n = X_n W_{n-1}, \quad (12.17)$$

where $X_n$ is the multiplication factor and supposed to be i.i.d. Then,

$$W_n = X_n X_{n-1} \cdots X_1 W_0. \quad (12.18)$$

Let $R_n$ be the “average” rate of return from $n$ investments, then

$$W_n = (1 + R_n)^n W_0. \quad (12.19)$$

Rearranging the terms, we have

$$(1 + R_n)^n = \frac{W_n}{W_0} = X_n X_{n-1} \cdots X_1. \quad (12.20)$$

Taking log on both sides, we have

$$\log(1 + R_n) = \frac{1}{n} \sum_{i=1}^{n} \log X_i \to E[\log X], \quad (12.21)$$

as $n \to \infty$, by the Strong Law of Large Numbers. Thus, for large $n$, we have an approximation,

$$\log(1 + R_n) \approx E[\log X]. \quad (12.22)$$

Since $f(x) = \log(1 + x)$ is increasing in $x$, if we would like to maximize the “average” rate of return from $n$ investments, $R_n$, it is enough to consider the maximize $E[\log X]$.

Thus, intuitively speaking, someone who has the log utility function is to value the long run average rate of return.
Chapter 13

Portfolio Analysis

This part is based on the book of Sheldon Ross [2002].

13.1 Portfolio Problem

Suppose we can invest the amount $w$ over $n$ different stocks. If we invest $a$ on the stock $i$, at the end of the next period, we have

$$aX_i,$$  \hspace{1cm} (13.1)

where $X_i$ is some non-negative random variable (growth factor). Define the rate of return of the stock $i$ by

$$R_i = \frac{aX_i - a}{a} = X_i - 1.$$  \hspace{1cm} (13.2)

Assume we invest $w_i$ on the stock $i$, where

$$w = \sum_{i=1}^{n} w_i.$$  \hspace{1cm} (13.3)

Then, at the end of the next period, we have the wealth as

$$W = \sum_{i=1}^{n} w_iX_i.$$  \hspace{1cm} (13.4)

Now we can formulate Portfolio problem:
We want to choose our portfolio $w_1, w_2, \ldots, w_n$ such that
\begin{align*}
w_i &\geq 0, \quad (13.5) \\
w &= \sum_{i=1}^{n} w_i, \quad (13.6)
\end{align*}
and maximize our expected utility
\[E[U(W)] = E[U(\sum_{i=1}^{n} w_i X_i)], \quad (13.7)\]
where $U$ is our utility function.

Without the nice assumption of $W = \sum_{i=1}^{n} w_i X_i$, it is hard to evaluate our expected utility. Thus, we will assume that $W$ is a normal distribution, i.e.,
\[W \sim N(\mu, \sigma^2). \quad (13.8)\]
Here’s a couple of validation of our assumption.

1. $X_i$ may be independent each other. If so, by the Central Limit Theorem, $W$ is approximated by normal distribution.

2. If $X_i$ is a multivariate normal random variable, $W$, the sum of the multivariate normal random variables, should be normal random variable.

Let us further assume that our utility function is an exponential function,
\[U(x) = 1 - e^{-bx}, \quad (13.9)\]
where $b > 0$.

**Lemma 13.1.** For a normal random variable $Z$, we have
\[E[e^Z] = e^{E[Z] + \text{Var}[Z]/2}. \quad (13.10)\]

**Proof.** See Toyoizumi Toyoizumi [2008].

Thus, the expectation of our utility function is
\begin{align*}
E[U(W)] &= 1 - E[e^{-bW}] \\
&= 1 - e^{E[-bW] + \text{Var}[-bW]/2} \\
&= 1 - e^{-b(\mu - b\sigma^2/2)}.
\end{align*}
13.2. MEAN AND VARIANCE OF $W$

Since $f(x) = 1 - e^{-bx}$ is increasing in $x$, the maximum of our utility function $E[U(W)]$ will be attained when we choose the portfolio such that

$$
\mu - b\sigma^2 / 2
$$

is maximized.

What does this mean? If we have larger $\mu = E[W]$, then we have larger expected return. So, it is quite natural to prefer the investment that have a larger $\mu$. How about $\sigma^2 = Var[W]$? From (13.11), smaller $\sigma^2$ is better than large one. This means that we prefer to minimize the variability of $W$, and avoid the risk.

**Problem 13.1.** How can you reduce $\sigma^2 = Var[W]$ in your portfolio?

### 13.2 Mean and Variance of $W$

From the arguments in Section 13.1, we learned that the expectation $\mu$ and the variance $\sigma^2$ of our investment are important for our decision.

Let $R_i = X_i - 1$ be the rate of return of stock $i$ and

$$
r_i = E[R_i],
$$

$$
\sigma_i^2 = Var[R_i].
$$

Since

$$
W = \sum_{i=1}^n w_i X_i = W + \sum_{i=1}^n w_i R_i,
$$

we have the expectation of our investment $W$ as

$$
E[W] = E[\sum_{i=1}^n w_i R_i]
$$

$$
= w + \sum_{i=1}^n w_i r_i.
$$
Also,

\[
Var[W] = \text{Var} \left[ \sum_{i=1}^{n} w_i R_i \right] \\
= \sum_{i=1}^{n} \text{Var}[w_i R_i] + \sum_{i=1}^{n} \sum_{j \neq i} \text{Cov}(w_i R_i, w_j R_j) \\
= \sum_{i=1}^{n} w_i^2 v_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} w_i w_j c(i, j),
\]

where \( c(i, j) = \text{Cov}(R_i, R_j) \). 

### 13.3 Portfolio of Two Stocks

Suppose we have \(a\) and invest it on two stocks. Let \(R_i\) be the rate of return of stock \(i\). Assume we know the information:

\[
\begin{align*}
  r_i &= E[R_i], \\
v_i^2 &= \text{Var}[R_i].
\end{align*}
\]

Further, we know the correlation of the rate of returns:

\[
\rho = \frac{\text{Cov}(R_1, R_2)}{v_1 v_2}.
\]

We have the following exponential utility function:

\[
U(x) = 1 - e^{-bx}.
\]

**Problem 13.2.** How can we allocate our money to maximize our expected utility. In other word, find out the best portfolio \((w_1, w_2) = (y, a-y)\) to maximize

\[
E[U(W)] = E \left[ U \left( \sum_{i=1}^{n} w_i X_i \right) \right].
\]

Set \(w_1 = y\) and \(w_2 = a-y\). Then, by using (13.15), we have the expectation of our portfolio as

\[
E[W] = w + \sum_{i=1}^{n} w_i r_i \\
= w + r_1 y + r_2 (a-y) \\
= w + r_2 a + (r_1 - r_2)y.
\]
By (13.16), the variance of our portfolio can be calculated by

\[
\text{Var}[W] = \sum_{i=1}^{n} w_i^2 v_i^2 + \sum_{i \neq j}^{n} w_i w_j c(i, j)
\]

\[
= y^2 v_1^2 + (a - y)^2 v_2^2 + 2y(a - y)\rho v_1 v_2 \\
= y^2 v_1^2 + (a^2 - 2ay + y^2)v_2^2 + 2\rho v_1 v_2 (ay - y^2) \\
= y^2(v_1^2 + v_2^2 - 2\rho v_1 v_2) + 2ay(\rho v_1 v_2 - v_2^2) + a^2 v_2^2. \tag{13.22}
\]

As discussed in Section 13.1, it is reasonable to allocate our investment so as to maximize (13.11). Set a function \( f(y) \) as

\[
f(y) = E[W] - \frac{b\text{Var}[W]}{2} \\
= w + r_2 a + (r_1 - r_2)y - \frac{b}{2} \left\{ y^2(v_1^2 + v_2^2 - 2\rho v_1 v_2) + 2ay(\rho v_1 v_2 - v_2^2) + a^2 v_2^2 \right\} \\
= -\frac{b}{2}(v_1^2 + v_2^2 - 2\rho v_1 v_2)y^2 + \left\{ (r_1 - r_2) - ab(\rho v_1 v_2 - v_2^2) \right\} y + w + r_2 a - \frac{b}{2} a^2 v_2^2. \tag{13.23}
\]

We need to find \( y \) to maximize \( f(y) \). By differentiating \( f(y) \), it is easy to find the optimal portfolio as

\[
y = \frac{(r_1 - r_2) + ab(v_2^2 - \rho v_1 v_2)}{b(v_1^2 + v_2^2 - 2\rho v_1 v_2)} \tag{13.24}
\]

will maximize \( f(y) \) and maximize our utility function.

In particular, if the stocks have the same expected rate of return, i.e. \( r_1 = r_2 \), then the optimal portfolio is

\[
y = \frac{v_2^2 - \rho v_1 v_2}{v_1^2 + v_2^2 - 2\rho v_1 v_2} a. \tag{13.25}
\]

Thus, in this case, we can find the optimal ratio of the stocks in our portfolio, which is indifferent with the shape of our utility function (especially \( b \)).

Remark 13.1. It turns out that (13.25) minimize \( \text{Var}[W] \), and minimize the expected utility function as long as its utility function is concave.

Further, if the stocks are independent, we have $\rho = 0$ and we have the optimal ratio of stock 1 in our portfolio as
\[
\frac{v_2^2}{v_1^2 + v_2^2} = \frac{1/v_1^2}{1/v_1^2 + 1/v_2^2}.
\] (13.26)
This idea is easy to extend $n$-stock portfolio, and the optimal ratio of stock $i$ is
\[
\frac{1/v_i^2}{\sum_{i=1}^{n} 1/v_i^2}.
\] (13.27)

**Example 13.1** (Portfolio of two stocks). Here is an example how we chose portfolio given the information of rate of return.

Suppose we have $100 and will make a portfolio of two stocks. We found out the statistic of our stocks as
\[
r_1 = 0.15, v_1 = 0.20,
\] (13.28)
\[
r_2 = 0.18, v_2 = 0.25,
\] (13.29)
\[
\rho = -0.4.
\] (13.30)
The second stock is relatively risky (high volatility) and we can expect higher return. Also we can see that two stocks have negative correlation. Thus, any combination of positive allocation will reduce the variance of our investment.

Assume we are a risk-averse investor, and our utility function is exponential and
\[
U(x) = 1 - e^{-bx},
\] (13.31)
where $b = 0.005$.

By using (13.24), if we invest our money
\[
y = 15.789
\] (13.32)
to the stock 1 and the rest to the stock 2, our expected utility is maximized and
\[
E[U[W]] = 0.4416.
\] (13.33)

**Problem 13.4.** Estimate the expected utility when we invest $100 on one stock. Compare the result with our optimal portfolio.
Chapter 14

Basics Formulation of Linear Programing

This part is based on the textbook [Mori and Matsui, 2004].

14.1 What is Linear Programing

Intuitively, Linear Programing is a method of finding the optimal solution given by some constraint.

Formally,

**Definition 14.1** (Linear Programing). Linear Programing (LP) is the problem to find the maximum (or minimum) point of a given linear function (objective function) subject to a specific set of possible points (admissible set).

14.2 Example of Wine Production

**Example 14.1** (Wine production). Suppose you are going to make a plan of the next-year wine production mix. You need to maximize the profit. However, there is a certain restriction. You can expect the grape production next year as in Table 14.1.

From the grapes, you may produce a couple of different wine with different taste and (of course) different price (see Table 14.2). Your objective is to maximize your profit producing appropriate variety of wine.
Problem 14.1. What is your strategy to maximize your profit?

The most profitable wine is white, so natural choice is to produce as much white wine as possible. You have 6 ton of Semmillon, so you can produce the white wine as much as,

\[ 6 / 3 = 2 \text{(ton)}. \]  

(14.1)

The resulted profit from white wine is

\[ 4 \times 2 = 8 \text{(million yen)}. \]  

(14.2)

The next choice is red. To produce red wine, you mix Cabernet and Merlot with the ratio 2 : 1. You have 4-ton Cabernet and 8-ton Merlot. You may run out of Cabernet by producing red wine as much as,

\[ 4 / 2 = 2 \text{(ton)}, \]  

(14.3)

with the profit of red wine:

\[ 3 = 6 \text{(million yen)}. \]  

(14.4)

Unfortunately, you are running out of Cabernet, and you can not produce Rose...

Thus, overall profit of your wine production is

\[ \text{white wine} + \text{red wine} + \text{rose wine} = 8 + 6 + 0 = 14 \text{(million yen)}. \]  

(14.5)

Problem 14.2. Can you think of any other production mix better than the above?
14.3 Mathematical Formulation

Problem 14.3. Describe the objective function and admissible set in Example 14.1

Let \( x_1, x_2, x_3 \) be the production volume of red, white and rose wine, respectively. Then, the total profit \( R = R(x_1, x_2, x_3) \) is written by

\[
R(x_1, x_2, x_3) = 3x_1 + 4x_2 + 2x_3.
\] (14.6)

You objective is to maximize the function \( R(x_1, x_2, x_3) \). However, you have the limited resource of wines. So your product mix \((x_1, x_2, x_3)\) may not have arbitrary value. For example, we have 4-ton Cabernet, which will be use for producing red wine, thus

\[
2x_1 \leq 4.
\] (14.7)

On the other hand, we have 8-ton Merlot, which can be used for both red and rose. So we have the constraint:

\[
x_1 + 2x_3 \leq 8.
\] (14.8)

Similarly, we have the constraint for Semmilon:

\[
3x_2 + x_3 \leq 6.
\] (14.9)

Of course, production volume should be positive. So,

\[
x_i \geq 0,
\] (14.10)

for all \( i \).

Now, we have the formal Linear Programing (LP) problem of wine production:

\[
\max R(x_1, x_2, x_3) = 3x_1 + 4x_2 + 2x_3
\] (14.11)

\[ s.t. \]

\[
2x_1 \leq 4,
\] (14.12)

\[
x_1 + 2x_3 \leq 8,
\] (14.13)

\[
3x_2 + x_3 \leq 6,
\] (14.14)

\[
x_1, x_2, x_3 \geq 0.
\] (14.15)
Chapter 15

Solving LP

15.1 Can We Always Solve LP?

Problem 15.1. Can we always solve LP? Describe two typical cases when we can not solve LP.

15.2 Slack Variables

Let us consider solving a (slightly different) LP problem:

\[
\begin{align*}
\text{max } & R(x_1, x_2, x_3) = 3x_1 + 2x_2 + 4x_3 \\
\text{s.t. } & x_1 + x_2 + 2x_3 \leq 4, \\
& 2x_1 + 2x_3 \leq 5, \\
& x_1, x_2, x_3 \geq 0.
\end{align*}
\]  

(15.1)

(15.2)

(15.3)

(15.4)

We introduce two additional variables \( x_4, x_5 \) to compensate the inequalities (15.2) and (15.3).

Definition 15.1 (slack variables). These additional variables are called slack variables.

Then the LP problem is rewritten by an equivalent LP problem with equal
15.3. DICTIONARY

constraints.

\[
\max R(x_1, x_2, x_3, x_4, x_5) = 3x_1 + 2x_2 + 4x_3 \\
s.t. \\
x_1 + x_2 + 2x_3 + x_4 = 4, \quad (15.6) \\
2x_1 + 2x_3 + x_5 = 5, \quad (15.7) \\
x_1, x_2, x_3 \geq 0. \quad (15.8)
\]

15.3 Dictionary

By introducing the variable \( z \) representing the objective value, we have the relationship between the variables:

\[
z = 3x_1 + 2x_2 + 4x_3, \\
x_4 = 4 - x_1 - x_2 - 2x_3, \\
x_5 = 5 - 2x_1 - 2x_3 - x_5.
\]

Definition 15.2 (Dictionary). A set of equations like above is called a “Dictionary” of LP problem.

You can find the value of left hand side (including our objective \( z \)) from the value of right hand side of a dictionary. For example, let \((x_1, x_2, x_3) = (0, 3, 1)\), then

\[
(z, x_4, x_5) = (10, -1, 3) \quad (15.9)
\]

Definition 15.3 (basic and non-basic variables). The left hand side is called basic variables and the right hand side is called non-basic variables.

Problem 15.2. Given non-basic variables as \((x_1, x_2, x_3) = (0, 0, 0)\), find the value of basic variables.

Definition 15.4 (basic solution). Given all–zero non-basic variables, the corresponding variables are called the basic solution.

Problem 15.3. In what situation does the basic solution become the solution of LP?
15.4 Simplex Method

It is easy to see that our dictionary
\[ z = 3x_1 + 2x_2 + 4x_3, \]  
\[ x_4 = 4 - x_1 - x_2 - 2x_3, \]  
\[ x_5 = 5 - 2x_1 - 2x_3 - x_5. \]  

is not the case where the basic solution:

\[ (z;x_1,x_2,x_3,x_4,x_5) = (0;0,0,4,5) \]  

(15.13)

gives us the solution of LP. It is obvious that \( x_3 \) contributes \( z \) greater than any other non-basic variables. So, temporarily, set \( x_1, x_2 = 0 \) and increase \( x_3 \) as

\[ (x_1,x_2,x_3) = (0,0,\varepsilon). \]  

(15.14)

Problem 15.4. Find the best value of \( \varepsilon \).

Now, we can find that

\[ (z;x_1,x_2,x_3,x_4,x_5) = (8;0,0,2,0,1) \]  

(15.15)

is the best among \( (x_1,x_2,x_3) = (0,0,\varepsilon) \). Note, in (15.15), instead of having \( (x_1,x_2,x_3) = (0,0,0) \), we have \( (x_1,x_2,x_4) = (0,0,0) \). Thus, it seems (15.15) is the basic solution for some other dictionary. We modify the original dictionary to a new equivalent dictionary whose basic solution is (15.15), as

\[ z = 9 - x_5 + x_2 - x_4, \]  
\[ x_3 = (3/2) + (1/2)x_5 - x_2 - x_4, \]  
\[ x_1 = 1 - x_5 + x_2 + x_4. \]  

(15.16)  
\( (15.17) \)  
\( (15.18) \)

Problem 15.5. Verify the modification, using the relationship \( x_4 = 4 - x_1 - x_2 - 2x_3 \).

We can modify the dictionary iteratively:

1. Derive the dictionary whose basic solution is \( (z;x_1,x_2,x_3,x_4,x_5) = (8;1,0,3/2,0,0) \)

as

\[ z = 9 - x_5 + x_2 - x_4, \]  
\[ x_3 = (3/2) + (1/2)x_5 - x_2 - x_4, \]  
\[ x_1 = 1 - x_5 + x_2 + x_4. \]  

(15.19)  
\( (15.20) \)  
\( (15.21) \)
2. Derive the dictionary whose basic solution is \((z; x_1, x_2, x_3, x_4, x_5) = (21/2; 5/2, 3/2, 0, 0, 0)\) as

\[
\begin{align*}
z &= 21/2 - (1/2)x_5 - x_3 - 2x_4, \\
x_2 &= (3/2) + (1/2)x_5 - x_3 - x_4, \\
x_1 &= (5/2) - (1/2)x_5 - x_3.
\end{align*}
\]

Indeed, the last basic solution gives us the solution of a LP problem:

\[
\begin{align*}
\text{max } 21/2 - (1/2)x_5 - x_3 - 2x_4, \\
s.t.
&(3/2) + (1/2)x_5 - x_3 - x_4 \geq 0, \\
&(5/2) - (1/2)x_5 - x_3 \geq 0, \\
x_3, x_4, x_5 \geq 0,
\end{align*}
\]

which is equivalent to our original LP (15.1).

**Problem 15.6.** Find the best wine production mix in Example 14.1.


