DIFFERENCE AND DIFFERENTIAL EQUATIONS WITH APPLICATIONS IN QUEUEING THEORY

Aliakbar Montazer Haghighi
Department of Mathematics
Prairie View A&M University
Prairie View, TX

Dimitar P. Mishev
Department of Mathematics
Prairie View A&M University
Prairie View, TX

WILEY

DIFFERENCE AND DIFFERENTIAL EQUATIONS WITH APPLICATIONS IN QUEUEING THEORY

DIFFERENCE AND DIFFERENTIAL EQUATIONS WITH APPLICATIONS IN QUEUEING THEORY

Aliakbar Montazer Haghighi
Department of Mathematics
Prairie View A&M University
Prairie View, TX

Dimitar P. Mishev
Department of Mathematics
Prairie View A&M University
Prairie View, TX

WILEY

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:
Haghighi, Aliakbar Montazer.
Difference and differential equations with applications in queueing theory / Aliakbar Montazer Haghighi Department of Mathematics, Prairie View A&M University, Prairie View, Texas, Dimitar P. Mishev Department of Mathematics, Prairie View A&M University, Prairie View, Texas.
pages cm
Includes bibliographical references and index.
ISBN 978-1-118-39324-6 (hardback) 1. Queuing theory. 2. Difference equations. 3. Differential equations. I. Mishev, D. P. (Dimiter P.) II. Title.
QA274.8.H338 2013
519.8'2–dc23
2013001302

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To my better half, Shahin,
who has given me three beautiful grandchildren:
Maya Haghighi, Kayvan Haghighi, and Leila Madison Grant.
—A. M. Haghighi

To my mother, wife, and three children.
—D. P. Mishev

CONTENTS

PREFACE

Topics in difference and differential equations with applications in queueing theory typically span five subject areas: (1) probability and statistics, (2) transforms, (3) differential equations, (4) difference equations, and (5) queueing theory. These are addressed in at least four separate textbooks and taught in four different courses in as many semesters. Due to this arrangement, students needing to take these courses have to wait to take some important and fundamental required courses until much later than should be necessary. Additionally, based on our long experience in teaching at the university level, we find that perhaps not all topics in one subject are necessary for a degree. Hence, perhaps we as faculty and administrators should rethink our traditional way of developing and offering courses. This is another reason for the content of this book, as of the previous one from the authors, to offer several related topics in one textbook. This gives the instructor the freedom to choose topics according to his or her desire to emphasize, yet cover enough of a subject for students to continue to the next course, if necessary.

The methodological content of this textbook is not exactly novel, as "mathematics for engineers" textbooks have reflected this method for long past. However, that type of textbook may cover some topics that an engineering student may already know. Now with this textbook the subject will be reinforced. The need for this practice has generally ignored some striking relations that exist between the seemingly separate areas of a subject, for instance, in statistical concepts such as the estimation of parameters of distributions used in queueing theory that are derived from differential–difference equations. These concepts commonly appear in queueing theory, for instance, in measures on effectiveness in queuing models.

All engineering and mathematics majors at colleges and universities take at least one course in ordinary differential equations, and some go further to take courses in partial differential equations. As mentioned earlier, there are many books on "mathematics for engineers" on the market, and one that contains some applications using Laplace and Fourier transforms. Some also have included topics of probability and statistics, as the one by these authors. However, there is a lack of applications of probability and statistics that use differential equations, although we did it in our book. Hence, we felt that there is an urgent need for a textbook that recognizes the corresponding relationships between the various areas and a matching cohesive course. Particularly, theories of queues and reliability are two of those topics, and this book is designed to achieve just that. Its five chapters, while retaining their individual integrity, flow from selected topics in probability and statistics to differential and difference equations to stochastic processes and queueing theory.

Chapter 1 establishes a strong foundation for what follows in Chapter 2 and beyond. Classical Fourier and Laplace transforms as well as Z-transforms and generating functions are included in Chapter 2. Partial differential equations are often used to construct models of the most basic theories underlying physics and engineering, such as the system of partial differential equations known as Maxwell's equations, from which one can derive the entire theory of electricity and magnetism, including light. In particular, elegant mathematics can be used to describe the vibrating circular membrane. However, our goal here is to develop the most basic ideas from the theory of partial differential equations and to apply them to the simplest models arising from physics and the queueing models. Detailed topics of ordinary and partial differential and difference equations are included in Chapter 3 and Chapter 4 that complete the necessary tools for Chapter 5, which discusses stochastic processes and queueing models. However, we have also included the power series method of solutions of differential equations, which can be applied to, for instance, Bessel's equation.

In our previous book, we required two semesters of calculus and a semester of ordinary differential equations for a reader to comprehend the contents of the book. In this book, however, knowledge of at least two semesters of calculus that includes some familiarity with terminology such as the gradient, divergence, and curl, and the integral theorems that relate them to each other, are needed. However, we discuss not only the topics in differential equation, but also the difference equations that have vast applications in the theory of signal processing, stochastic analysis, and queueing theory.

Few instructors teach the combined subject areas together due to the difficulties associated with handling such a rigorous course with such hefty materials. Instructors can easily solve this issue by teaching the class as a multi-instructor course.

We should note that throughout the book, we use boldface letters, Greek or Roman (lowercase or capital) for vectors and matrices. We shall write $P(n)$ or P_n to mean P as a function of a discrete parameter n. Thus, we want to make sure that students are well familiar with functions of discrete variables as well as continuous ones. For instance, a vibrating string can be regarded as a continuous object, yet if we look at a fine enough scale, the string is made up of molecules, suggesting a discrete model with a large number of variables. There are many cases in which a discrete model may actually provide a better description of the phenomenon under study than a continuous one.

We also want to make sure that students realize that solution of some problems requires the ability to carry out lengthy calculations with confidence.

Of course, all of these skills are necessary for a thorough understanding of the mathematical terminology that is an essential foundation for the sciences and engineering. We further note that subjects discussed in each chapter could be studied in isolation; however, their cohesiveness comes from a thorough understanding of applications, as discussed in this book.

We hope this book will be an interesting and useful one to both students and faculty in science, technology, engineering, and mathematics.

Aliakbar Montazer Haghighi

Dimitar P. Mishev

Houston, Texas
April 2013

CHAPTER ONE

Probability and Statistics

The organization of this book is such that by the time reader gets to the last chapter, all necessary terminology and methods of solutions of standard mathematical background has been covered. Thus, we start the book with the basics of probability and statistics, although we could have placed the chapter in a later location. This is because some chapters are independent of the others.

In this chapter, the basics of probability and some important properties of the theory of probability, such as discrete and continuous random variables and distributions, as well as conditional probability, are covered.

After the presentation of the basics of probability, we will discuss statistics. Note that there is still a dispute as to whether statistics is a subject on its own or a branch of mathematics. Regardless, statistics deals with gathering, analyzing, and interpreting data. Statistics is an important concept that no science can do without. Statistics is divided in two parts: *descriptive statistics* and *inferential statistics*. Descriptive statistics includes some important basic terms that are widely used in our day-to-day lives. The latter is based on *probability theory*. To discuss this part of the statistics, we include point estimation, interval estimation, and hypothesis testing.

We will discuss one more topic related to both probability and statistics, which is extremely necessary for business and industry, namely *reliability of a system*. This concept is also needed in applications such as queueing networks, which will be discussed in the last chapter.

In this chapter, we cover as much probability and statistics as we will need in this book, except some parts that are added for the sake of completeness of the subject.

1.1. BASIC DEFINITIONS AND CONCEPTS OF PROBABILITY

Nowadays, it has been established in the scientific world that since quantities needed are not quite often predictable in advance, randomness should be

Difference and Differential Equations with Applications in Queueing Theory, First Edition.
Aliakbar Montazer Haghighi and Dimitar P. Mishev.

accounted for in any realistic world phenomenon, and that is why we will consider random experiments in this book.

Determining probability, or chance, is to quantify the variability in the outcome or outcomes of a random experiment whose exact outcome or outcomes cannot be predicted by certainty. Satellite communication systems, such as radar, are built of electronic components such as transistors, integrated circuits, and diodes. However, as any engineer would testify, the components installed usually never function as the designer has anticipated. Thus, not only is the probability of failure to be considered, but the reliability of the system is also quite important, since the failure of the system may have not only economic losses but other damages as well. With probability theory, one may answer the question, "How reliable is the system?"

Definition 1.1.1. ***Basics***

(a) Any result of performing an experiment is called an *outcome* of that experiment. A set of outcomes is called an *event*.

(b) If occurrences of outcomes are not certain or completely predictable, the experiment is called a *chance* or *random experiment*.

(c) In a random experiment, sets of outcomes that cannot be broken down into smaller sets are called *elementary* (or *simple* or *fundamental*) *events*.

(d) An elementary event is, usually, just a singleton (a set with a single element, such as $\{e\}$). Hence, a combination of elementary events is just an *event*.

(e) When any element (or outcome) of an event happens, we say that the *event occurred*.

(f) The *union* (set of all elements, with no repetition) of all events for a random experiment (or the set of all possible outcomes) is called the *sample space*.

(g) In "set" terminology, an *event* is a *subset* of the sample space. Two events A_1 and A_2 are called *mutually exclusive* if their intersection is the empty set, that is, they are disjoint subsets of the sample space.

(h) Let $A_1, A_2, \ldots, A_n$ be mutually exclusive events such that $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$. The set of $\{A_1, A_2, \ldots, A_n\}$ is then called a *partition* of the sample space Ω.

(i) For an experiment, a collection or a set of all individuals, objects, or measurements of interest is called a (statistical) *population*.

For instance, to determine the average grade of the differential equation course for all mathematics major students in four-year colleges and universities in Texas, the totality of students majoring mathematics in the colleges and universities in the Texas constitutes the population for the study.

Usually, studying the population may not be practically or economically feasible because it may be quite time consuming, too costly, and/or impossible to identify all members of it. In such cases, sampling is being used.

(j) A portion, subset, or a part of the population of interest (finite or infinite number of them) is called a *sample*.

Of course, the sample must be *representative* of the entire population in order to make any prediction about the population.

(k) An element of the sample is called a *sample point*. By *quantification* of the sample we mean changing the sample points to numbers.

(l) The *range* is the difference between the smallest and the largest sample points.

(m) A sample selected such that each element or unit in the population has the same chance to be selected is called a *random sample*.

(n) The *probability of an event* A, denoted by $P(A)$, is a number between 0 and 1 (inclusive) describing likelihood of the event A to occur.

(o) An event with probability 1 is called an *almost sure event*. An event with probability 0 is called a *null* or an *impossible event*.

(p) For a sample space with n (finite) elements, if all elements or outcomes have the same chance to occur, then we assign probability $1/n$ to each member. In this case, the sample space is called *equiprobable*.

For instance, to choose a digit at random from 1 to 5, we mean that every digit of {1, 2, 3, 4, 5} has the same chance to be picked, that is, all elementary events in {1}, {2}, {3}, {4}, and {5} are equiprobable. In that case, we may associate probability 1/5 to each digit singleton.

(q) If a random experiment is repeated, then the chance of occurrence of an outcome, intuitively, will be approximated by the ratio of occurrences of the outcome to the total number of repetitions of the experiment. This ratio is called the *relative frequency*.

Axioms of Probabilities of Events

We now state properties of probability of an event A through *axioms of probability*. The Russian mathematician Kolmogorov originated these axioms in early part of the twentieth century. By an axiom, it is meant a statement that cannot be proved or disproved. Although all probabilists accept the three axioms of probability, there are axioms in mathematics that are still controversial, such as the axiom of choice, and not accepted by some prominent mathematicians.

Let Ω be the sample space, $\mathcal{B}$ the set function containing all possible events drawn from Ω, and P denote the probability of an event. The triplet $(\Omega, \mathcal{B}, P)$ is then called the *probability space*. Later, after we define a random variable, we will discuss this space more rigorously.

Axioms of Probability

Axiom A1. $0 \le P(A) \le 1$ for each event A in $\mathcal{B}$.
Axiom A2. $P(\Omega) = 1$.
Axiom A3. If A_1 and A_2 are *mutually exclusive* events in $\mathcal{B}$, then:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2),$$

where mutually exclusive events are events that have no sample point in common, and the symbol $\cup$ means the union of two sets, that is, the set of all elements in both set without repetition.

Note that the axioms stated earlier are for events. Later, we will define another set of axioms of probability involving random variables.

If the occurrence of an event has influence on the occurrence of other events under consideration, then the probabilities of those events change.

Definition 1.1.2
Suppose $(\Omega, \mathcal{B}, P)$ is a probability space and B is an event (i.e., $B \in \mathcal{B}$) with positive probability, $P(B) > 0$. The *conditional probability of A given B*, denoted by $P(A|B)$, defined on $\mathcal{B}$, is then given by:

$$P(A|B) = \frac{P(AB)}{P(B)}, \text{ for any event } A \text{ in } \mathcal{B}, \text{ and for } P(B) > 0. \tag{1.1.1}$$

If $P(B) = 0$, then $P(A|B)$ is not defined. Under the condition given, we will have a new triple, that is, a new probability space $(\Omega, \mathcal{B}, P(A|B))$. This space is called the *conditional probability space induced on* $(\Omega, \mathcal{B}, P)$*, given B*.

Definition 1.1.3
For any two events A and B with conditional probability $P(B | A)$ or $P(A | B)$, we have the *multiplicative law*, which states:

$$P(AB) = P(B|A)P(A) = P(A|B)P(B). \tag{1.1.2}$$

We leave it as an exercise to show that for n events $A_1, A_2, \ldots, A_n$, we have:

$$P(A_1 A_2 \ldots A_n) = P(A_1) \times P(A_2|A_1) \times P(A_3|A_1 A_2) \times \cdots \times P(A_n|A_1 A_2 \ldots A_{n-1}). \tag{1.1.3}$$

Definition 1.1.4
We say that events A and B are *independent* if and only if:

$$P(AB) = P(A)P(B). \tag{1.1.4}$$

It will be left as an exercise to show that if events A and B are independent and $P(B) > 0$, then:

$$P(A \mid B) = P(A). \tag{1.1.5}$$

It can be shown that if $P(B) > 0$ and (1.1.5) is true, then A and B are independent. For proof, see Haghighi et al. (2011a, p. 139).

The concept of independence can be extended to a finite number of events.

Definition 1.1.5
Events $A_1, A_2, \ldots, A_n$ are *independent* if and only if the probability of the intersection of any subset of them is equal to the product of corresponding probabilities, that is, for every subset $\{i_1, \ldots, i_k\}$ of $\{1, \ldots, n\}$ we have:

$$P\{(A_{i_1} A_{i_2} \ldots A_{i_n})\} = P(A_{i_1}) \times P(A_{i_2}) \times \cdots \times P(A_{i_k}). \tag{1.1.6}$$

As one of the very important applications of conditional probability, we state the following theorem, whose proof may be found in Haghighi et al. (2011a):

Theorem 1.1.1. ***The Law of Total Probability***
Let $A_1, A_2, \ldots, A_n$ be a partition of the sample space Ω. For any given event B, we then have:

$$P(B) = \sum_{i=1}^{n} P(A_i) P(B \mid A_i). \tag{1.1.7}$$

Theorem 1.1.1 leads us to another important application of conditional probability. Proof of this theorem may also found in Haghighi et al. (2011a).

Theorem 1.1.2. ***Bayes' Formula***
Let $A_1, A_2, \ldots, A_n$ be a partition of the sample space Ω. If an event B occurs, the probability of any event A_j given an event B is:

$$P(A_j \mid B) = \frac{P(A_j) P(B \mid A_j)}{\sum_{i=1}^{n} P(A_i) P(B \mid A_i)}, \quad j = 1, 2, \ldots, n. \tag{1.1.8}$$

Example 1.1.1
Suppose in a factory three machines A, B, and C produce the same type of products. The percent shares of these machines are 20, 50, and 30, respectively. It is observed that machines A, B, and C produce 1%, 4%, and 2% defective items, respectively. For the purpose of quality control, a produced item is chosen at random from the total items produced in a day. Two questions to answer:

1. What is the probability of the item being defective?
2. Given that the item chosen was defective, what is the probability that it was produced by machine B?

Answers

To answer the first question, we denote the event of defectiveness of the item chosen by D. By the law of total probability, we will then have:

$$\begin{aligned} P(D) &= P(\mathrm{A})P(D|\mathrm{A}) + P(\mathrm{B})P(D|B) + P(\mathrm{C})P(D|C) \\ &= 0.20 \times 0.01 + 0.50 \times 0.04 + 0.30 \times 0.20 \\ &= 0.002 + 0.020 + 0.006 = 0.028. \end{aligned}$$

Hence, the probability of the produced item chosen at random being defective is 2.8%.

To answer the second question, let the conditional probability in question be denoted by $P(\mathrm{B} \mid D)$. By Bayes' formula and answer to the first question, we then have:

$$P(\mathrm{B}|D) = \frac{P(\mathrm{B})P(D|\mathrm{B})}{P(D)} = \frac{0.50 \times 0.04}{0.028} = 0.714.$$

Thus, the probability that the defective item chosen be produced by machine C is 71.4%.

Example 1.1.2

Suppose there are three urns that contain black and white balls as follows:

$$\begin{cases} \text{Urn } 1: & 2\ \textit{blacks} \\ \text{Urn } 2: & 2\ \textit{whites} \\ \text{Urn } 1: & 1\ \textit{black} \text{ and } 1\ \textit{white}. \end{cases} \tag{1.1.9}$$

A ball is drawn randomly and it is "white." Discuss possible probabilities.

Discussion

The sample space Ω is the set of all pairs $(\cdot,\cdot)$, where the first dot represents the urn number (1, 2, or 3) and the second represents the color (black or white). Let U_1, U_2 and U_3 denote events that drawing was chosen from, respectively. Assuming that urns are identical and balls have equal chances to be chosen, we will then have:

$$P(U_1) = P(U_2) = P(U_3) = \frac{1}{3}. \tag{1.1.10}$$

Also, $U_1 = (1,\cdot)$, $U_2 = (2,\cdot)$, $U_3 = (3,\cdot)$.

Let W denote the event that a white ball was drawn, that is, $W = \{(\cdot, w)\}$. From (1.1.9), we have the following conditional probabilities:

$$P(W|U_1) = 0, \quad P(W|U_2) = 1, \quad P(U_3) = \frac{1}{2}. \tag{1.1.11}$$

From Bayes' rule, (1.1.9), (1.1.10), and (1.1.11), we have:

$$P(U_1|W) = \frac{P(W|U_1)P(U_1)}{P(W|U_1)P(U_1) + P(W|U_2)P(U_2) + P(W|U_3)P(U_3)}, \tag{1.1.12}$$

$$= 0. \tag{1.1.13}$$

Note that denominator of (1.1.12) is:

$$0 + (1)\left(\frac{1}{3}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{3}\right) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}. \tag{1.1.14}$$

Using (1.1.14), we have:

$$\begin{aligned} P(U_2|W) &= \frac{P(W|U_2)P(U_2)}{P(W|U_1)P(U_1) + P(W|U_2)P(U_2) + P(W|U_3)P(U_3)} \\ &= \frac{(1)\left(\frac{1}{3}\right)}{\frac{1}{2}} = \frac{2}{3}. \end{aligned} \tag{1.1.15}$$

Again, using (1.1.14), we have:

$$\begin{aligned} P(U_3|W) &= \frac{P(W|U_3)P(U_3)}{P(W|U_1)P(U_1) + P(W|U_2)P(U_2) + P(W|U_3)P(U_3)} \\ &= \frac{\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)}{\frac{1}{2}} = \frac{1}{3}. \end{aligned} \tag{1.1.16}$$

Now, observing from (1.1.13), (1.1.15), and (1.1.16), there is a better chance that the ball was drawn from the second urn. Hence, if we assume that the ball was drawn from the second urn, there is one white ball that remains in it. That is, we will have the three urns with 0, 1, and 1 white ball, respectively, in urns 1, 2, and 3.

1.2. DISCRETE RANDOM VARIABLES AND PROBABILITY DISTRIBUTION FUNCTIONS

As we have seen so far, elements of a sample space are not necessarily numbers. However, for convenience, we would rather have them so. This is done through what is called a *random variable*. In other words, a *random variable* quantifies the sample space. That is, a *random variable* assigns numerical (or set) labels to the sample points. Formally, we define a random variable as follows:

Definition 1.2.1
A *random variable* is a function (or a mapping) on the sample space.

We note that a random variable is really neither a variable (as known independent variable) nor random, but as mentioned, it is just a function. Also note that sometimes the range of a random variable may not be numbers. This is simply because we defined a random variable as a mapping. Thus, it maps elements of a set into some elements of another set. Elements of either set do not have to necessarily be numbers.

There are two main types of random variables, namely, *discrete* and *continuous*. We will discuss each in detail.

Definition 1.2.2
A *discrete random variable* is a function, say X, from a countable sample space, Ω (that could very well be a numerical set), into the set of real numbers.

Example 1.2.1
Suppose we are to select two digits from 1 to 6 such that the sum of the two numbers selected equals 7. Assume that repetition is not allowed. The sample space under consideration will then be $S = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$, which is discrete. This set can also be described as $S = \{(i, j): i + j = 7, i, j = 1, 2, \dots, 6\}$.

Now, the random variable X can be defined by $X((i, j)) = k, k = 1, 2, \dots, 6$. That is, the range of X is the set $\{1, 2, 3, 4, 5, 6\}$ such that, for instance, $X((1, 6)) = 1$, $X((2, 5)) = 2$, $X((3, 4)) = 3$, $X((4, 3)) = 4$, $X((5, 2)) = 5$, and $X((6, 1)) = 6$. In other words, the discrete random variable X has quantified the set of ordered pairs S to a set of positive integers from 1 to 6.

Example 1.2.2
Toss a fair coin three times. Denoting heads by H and tails by T, the sample space will then contain eight triplets as $\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$. Each tossing will result in either heads or tails. Thus, we might define the random variable X to take values 1 and 0 for heads and tails, respectively, at the jth tossing. In other words,

$$X_j = \begin{cases} 1, & \text{if } j\text{th outcome is heads,} \\ 0, & \text{if } j\text{th outcome is tails.} \end{cases}$$

Hence, $P\{X_j = 0\} = 1/2$ and $P\{X_j = 1\} = 1/2$. Now from the sample space we see that the probability of the element HTH is:

$$P\{X_1 = 1, X_2 = 0, X_3 = 1\} = \frac{1}{8}. \quad (1.2.1)$$

In contrast, product of individual probabilities is:

$$P\{X_1 = 1\} \times P\{X_2 = 0\} \times P\{X_3 = 1\} = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}. \tag{1.2.2}$$

From (1.2.1) and (1.2.2), we see that X_1, X_2, and X_3 are mutually independent.

Now suppose we define X and Y as the total number of heads and tails, respectively, after the third toss. The probability, then, of three heads and three tails is obviously zero, since these two events cannot occur at the same time, that is, $P\{X = 3, Y = 3\} = 0$. However, from the sample space probabilities of individual events are $P\{X = 3\} = 1/8$ and $P\{Y = 3\} = 1/8$. Thus, the product is:

$$P\{X = 3\} \times P\{Y = 3\} = \frac{1}{8} \times \frac{1}{8} = \frac{1}{64} \neq 0.$$

Hence, X and Y, in this case, are not independent.

One of the useful concepts using random variable is the *indicator function* (or *indicator random variable* that we will define in the next section.

Definition 1.2.3
Let A be an event from the sample space Ω. The random variable $I_A(\omega)$ for $\omega \in A$ defined as:

$$I_A(\omega) = \begin{cases} 1, & \text{if} \quad \omega \in A, \\ 0, & \text{if} \quad \omega \in A^c, \end{cases} \tag{1.2.3}$$

is called the indicator function (or indicator random variable).

Note that for every $\omega \in \Omega$, $I_\Omega(\omega) = 1$ and $I_\phi(\omega) = 0$.

We leave it as an exercise for the reader to show the following properties of random variables:

(a) if X and Y are two discrete random variables, then $X \pm Y$ and XY are also random variables, and
(b) if $\{Y = 0\}$ is empty, X/Y is also a random variable.

The way probabilities of a random variable are distributed across the possible values of that random variable is generally referred to as the *probability distribution* of that random variable. The following is the formal definition.

Definition 1.2.4
Let X be a discrete random variable defined on a sample space Ω and x is a typical element of the range of X. Let p_x denote the probability that the random variable X takes the value x, that is,

$$p_x = P([X = x]) \quad \text{or} \quad p_x = P(X = x), \tag{1.2.4}$$

where p_X is called the *probability mass function* (pmf) of X and also referred to as the (*discrete*) *probability density function* (pdf) of X.

Note that $\sum_x p_x = 1$, where x varies over all possible values for X.

Example 1.2.3

Suppose a machine is in either "good working condition" or "not good working condition." Let us denote "good working condition" by 1 and " not good working condition" by 0. The sample space of states of this machine will then be $\Omega = \{0, 1\}$. Using a random variable X, we define $P([X = 1])$ as the probability that the machine is in "good working condition" and $P([X = 0])$ as the probability that the machine is not in "good working condition." Now if $P([X = 0]) = 4/5$ and $P([X = 0]) = 1/5$, then we have a distribution for X.

Definition 1.2.5

Suppose X is a discrete random variable, and x is a real number from the interval $(-\infty, x]$. Let us define $F_X(x)$ as:

$$F_X(x) = P([X \le x]) = \sum_{n=-\infty}^{x} p_n, \tag{1.2.5}$$

where p_n is defined as $P([X = n])$ or $P(X = n)$. $F_X(x)$ is then called the *cumulative distribution function* (*cdf*) *for X*.

Note that from the set of axioms of probability mentioned earlier, for all x, we have:

$$p_x \ge 0, \quad \text{and} \quad \sum_x p_x = 1. \tag{1.2.6}$$

We now discuss selected important discrete probability distribution functions. Before that, we note that a random experiment is sometimes called a *trial*.

Definition 1.2.6

A *Bernoulli trial* is a trial with exactly two possible outcomes. The two possible outcomes of a Bernoulli trial are often referred to as *success* and *failure* denoted by s and f, respectively. If a Bernoulli trial is repeated independently n times with the same probabilities of success and failure on each trial, then the process is called *Bernoulli trials*.

Notes:

(1) From Definition 1.2.6, if the probability of s is p, $0 \le p \le 1$, then, by the second axiom of probability, the probability of f will be $q = 1 - p$.

(2) By its definition, in a Bernoulli trial, the sample space for each trial has two sample points.

Definition 1.2.7
Now, let X be a random variable taking values 1 and 0, corresponding to success and failure, respectively, of the possible outcome of a Bernoulli trial, with p ($p > 0$) as the probability of success and q as probability of failure. We will then have:

$$P(X = k) = p^k q^{1-k}, \quad k = 0, 1. \tag{1.2.7}$$

Formula (1.2.7) is the probability distribution function (pmf) of the Bernoulli random variable X.

Note that (1.2.7) is because first of all, $p^k q^{1-k} > 0$, and second, $\sum_{k=0}^{1} p^k q^{1-k} = p + q = 1$.

Example 1.2.4
Suppose we test 6 different objects for strength, in which the probability of breakdown is 0.2. What is the probability that the third object test be successful is, that is, does not breakdown?

Answer
In this case, we have a sequence of six Bernoulli trials. Let us assume 1 for a success and 0 for a failure. We would then have a 6-tuple (001000) to symbolize our objective. Hence, the probability would be (0.2)(0.2)(0.8)(0.2)(0.2)(0.2) = 0.000256.

Now suppose we repeat a Bernoulli trial independently finitely many times. We would then be interested in the probability of given number of times that one of the two possible outcomes occurs regardless of the order of their occurrences. Therefore, we will have the following definition:

Definition 1.2.8
Suppose X_n is the random variable representing the number of successes in n independent Bernoulli trials. Denote the *pmf* of X_n by $\mathrm{B_k} = b(k; n, p)$. $\mathrm{B_k} = b(k; n, p)$ is called the *binomial distribution function* with *parameters n* and *p* of the random variable X, where the parameters n, p and the number k refer to the number of independent trials, probability of *success* in each trial, and the number of successes in n trials, respectively. In this case, X is called the *binomial random variable*. The notation $X \sim b(k; n, p)$ is used to indicate that X is a binomial random variable with parameters n and p.

We leave it as an exercise to prove that:

$$\mathrm{B_k} = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, \ldots, n, \tag{1.2.8}$$

where $q = 1 - p$.

Example 1.2.5
Suppose two identical machines run together, each to choose a digit from 1 to 9 randomly five times. We want to know what the probability that a sum of 6 or 9 appears k times ($k = 0, 1, 2, 3, 4, 5$) is.

Answer
To answer the question, note that we have five independent trials. The sample space in this case for one trial has 81 sample points and can be written in a matrix form as follows:

$$\begin{pmatrix} (1,1) & (1,2) & \cdots & (1,8) & (1,9) \\ (2,1) & (2,2) & \cdots & (2,8) & (2,9) \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ (8,1) & (8,2) & \ddots & (8,8) & (8,9) \\ (9,1) & (9,2) & \cdots & (9,8) & (9,9) \end{pmatrix}.$$

There are 13 sample points, where the sum of the components is 6 or 9. They are:

$$(1,5),(2,4),(3,3),(4,2),(5,1),(1,8),(2,7),(3,6),(4,5),(5,4),(6,3),(7,2),(8,1).$$

Hence, the probability of getting a sum as 6 or 9 on one selection of both machines together (i.e., probability of a success) is $p = 13/81$. Now let X be the random variable representing the total times a sum as 6 or 9 is obtained in 5 trials. Thus, from (1.2.8), we have:

$$P([X = k]) = \binom{5}{k}\left(\frac{13}{81}\right)^k \left(\frac{68}{81}\right)^{5-k}, \quad k = 0, 1, 2, 3, 4, 5.$$

For instance, the probability that the sum as 6 or 9 does not appear at all will be $(68/81)^5 = 0.42$, that is, there is a $(100 - 42) = 58\%$ chance that we do get at least a sum as 6 or 9 during the five trials.

Based on a sequence of independent Bernoulli trials, we now define two other important discrete random variables. Consider a sequence of independent Bernoulli trials with probability of success in each trial as p, $0 \le p \le 1$. Suppose we are interested in the total number of trials required to have the rth success, r being a fixed positive integer. The answer is in the following definition:

Definition 1.2.9
Let X be a random variable with *pmf* as:

$$f(k; r, p) = \binom{r+k-1}{k} p^r q^k, \quad k = 0, 1, \ldots. \tag{1.2.9}$$

Formula (1.2.9) is then called a *negative binomial* (or *Pascal*) *probability distribution function* (or *binomial waiting time*). In particular, if $r = 1$ in (1.2.9), then we will have:

$$f(k; 1, p) = P(x = k + 1) = pq^k, \quad k = 0, 1, \ldots . \tag{1.2.10}$$

The pmf given by (1.2.10) is called a geometric probability distribution function.

Example 1.2.6

As an example, suppose a satellite company finds that 40% of call for services received need advanced technology service. Suppose also that on a particular crazy day, all tickets written are put in a pool and requests are drawn randomly for service. Finally, suppose that on that particular day there are four advance service personnel available. We want to find the probability that the fourth request for advanced technology service is found on the sixth ticket drawn from the pool.

Answer

In this problem, we have independent trials with $p = 0.4$ as probability of success, that is, in need of advanced technology service, on any trial. Let X represent the number of the tickets on which the fourth request in question is found. Thus,

$$P(X = 4) = \binom{6}{4}(0.4)^4 (0.6)^2 = 0.09216.$$

Example 1.2.7

We now want to derive (1.2.9) differently. Suppose treatment of a cancer patient may result in "response" or "no response." Let the probability of a response be p and for a no response be $1 - p$. Hence, the simple space in this case has two outcomes, simply, "response" and "no response." We now repeatedly treat other patients with the same medicine and observe the reactions. Suppose we are looking for the probability of the number of trials required to have exactly k "responses."

Answer

Denoting the sample space by S, $S = \{\text{response, no response}\}$. Let us define the random variable X on S to denote the number of trials needed to have exactly k responses. Let A be the event, in S, of observing $k - 1$ responses in the first $x - 1$ treatments. Let B be the event of observing a response at the xth treatment. Let also C be the event of treating x patients to obtain exactly k responses. Hence, $C = A \cap B$. The probability of C is:

$$P(C) = P(A \cap B) = P(A) \cdot P(B|A).$$

In contrast, $P(B \mid A) = p$ and:

$$P(A) = \binom{x-1}{k-1} p^{k-1}(1-p)^{x-k}.$$

Moreover, $P(X = x) = P(C)$. Hence:

$$P(X = x) = \binom{x-1}{k-1} p^{k}(1-p)^{x-k}, \quad x = k, k+1, \ldots . \tag{1.2.11}$$

We leave it as an exercise to show that (1.2.11) is equivalent to (1.2.9).

Definition 1.2.10

Let n represent a sample (sampling without replacement) from a finite population of size N that consists of two types of items n_1 of "defective," say, and n_2 of "nondefective," say, $n_1 + n_2 = n$. Suppose we are interested in the probability of selecting x "defective" items from the sample. n_1 must be at least as large as x. Hence, x must be less than or equal to the smallest of n and n_1. Thus,

$$p_x \equiv P(X = x) = \frac{\binom{n_1}{x} \times \binom{N-n_1}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, 2, \ldots, \min(n, n_1), \tag{1.2.12}$$

defines the general form of *hypergeometric pmf* of the random variable X.

Notes:

i. If sampling would have been with replacement, distribution would have been binomial.
ii. p_x is the probability of waiting time for the occurrence of exactly x "defective" outcomes. We could think of this scenario as an urn containing N white and green balls. From the urn, we select a random sample (a sample selected such that each element has the same chance to be selected) of size n, one ball at a time without replacement. The sample consists of n_1 white and n_2 green balls, $n_1 + n_2 = n$. What is the probability of having x white balls drawn in a row? This model is called an *urn model*.
iii. If we let x_i equal to 1 if a defective item is selected and 0 if a nondefective item is selected, and let x be the total number of defectives selected, then $x = \sum_{i=1}^{n} x_i$. Now, if we consider selection of a defective item as a success, for instance, then we could also interoperate (1.2.12) as:

$$p_x = \frac{(\text{number of ways for } x \text{ successes}) \times (\text{number of ways for } n-x \text{ failures})}{\text{total number of ways to select}}. \tag{1.2.13}$$