QOS IN PACKET NETWORKS

THE KLUWER INTERNATIONAL SERIES IN ENGINEERING AND COMPUTER SCIENCE

QOS IN PACKET NETWORKS

by

Kun I. Park, Ph.D.
The MITRE Corporation USA

Springer

eBook ISBN: 0-387-23390-3
Print ISBN: 0-387-23389-X

Created in the United States of America

Visit Springer's eBookstore at: http://ebooks.kluweronline.com
and the Springer Global Website Online at: http://www.springeronline.com

Dedication

For Meyeon and Kyunja.

Contents

Preface

QoS is an important subject that takes a central place in overall packet network technologies. It is a complex subject and its analysis involves such mathematical disciplines as probability, random variables, stochastic processes, and queuing. These mathematical subjects are abstract and are not easy to grasp for uninitiated persons.

This book is written with two objectives. The first objective is to explain the fundamental mathematical concepts used in QoS analysis in layman's terms and as plainly as possible so that the reader can have a better appreciation of the subject of QoS treated in this book. Second, this book explains in plain language the various parts of QoS in packet networks so that the reader can have a complete view of this complex and dynamic area of communications networking technology.

Kun I. Park
Holmdel, New Jersey

Chapter 1

INTRODUCTION

1. NEED FOR QOS

In recent years, the importance of Quality of Service (QoS) technologies for packet networks has increased rapidly. Today, QoS is undoubtedly one of the central pieces of the overall packet network technologies. How has QoS come to take such an important place in packet networks? This section reviews the recent history of telecommunications network evolution to put this fundamental question underpinning this book in perspective.

Referring to Figure 1.1, in the beginning of telecommunications, there were in general two separate networks, one for voice and one for data. Each network started with a simple goal of transporting a specific type of information. The telephone network, which was introduced with the invention of telephone by Alexander Graham Bell some hundred years ago, was designed to carry voice. The IP network, on the other hand, was designed to carry data.

In the early telephone network, the terminal device was a simple telephone set, which was nothing more than an analog transducer designed to produce an electrical current fluctuating with the speaker's acoustic pressure. For all practical purposes, this was all the function that the terminal device had to perform. The network itself, on the other hand, was more complex than the terminal, and was provided with "intelligence" necessary for providing various types of voice services.

A telephone connection is dedicated to a call during the entire period. Once the call is complete, the circuits are used to set up other calls. The circuits used to set up calls are referred to as trunks as opposed to "loops,"

which are the lines permanently dedicated to individual end users' telephone sets.

In the early telephone network, there were two key measures of service quality. The first was the probability of call blocking, that is, the probability that a call attempt would be blocked because of unavailability of a trunk circuit. Once a call attempt was successful and a connection was established for the call, the next measure of quality was voice quality. Voice quality depended on the transmission quality of the end-to-end connection during a call such as transmission loss, circuit noise, echo, etc.

The original telephone network, therefore, was designed with two main objectives. The first was to make sure that enough trunk circuits were provided to render call blocking probability reasonable, e.g., 1%. The

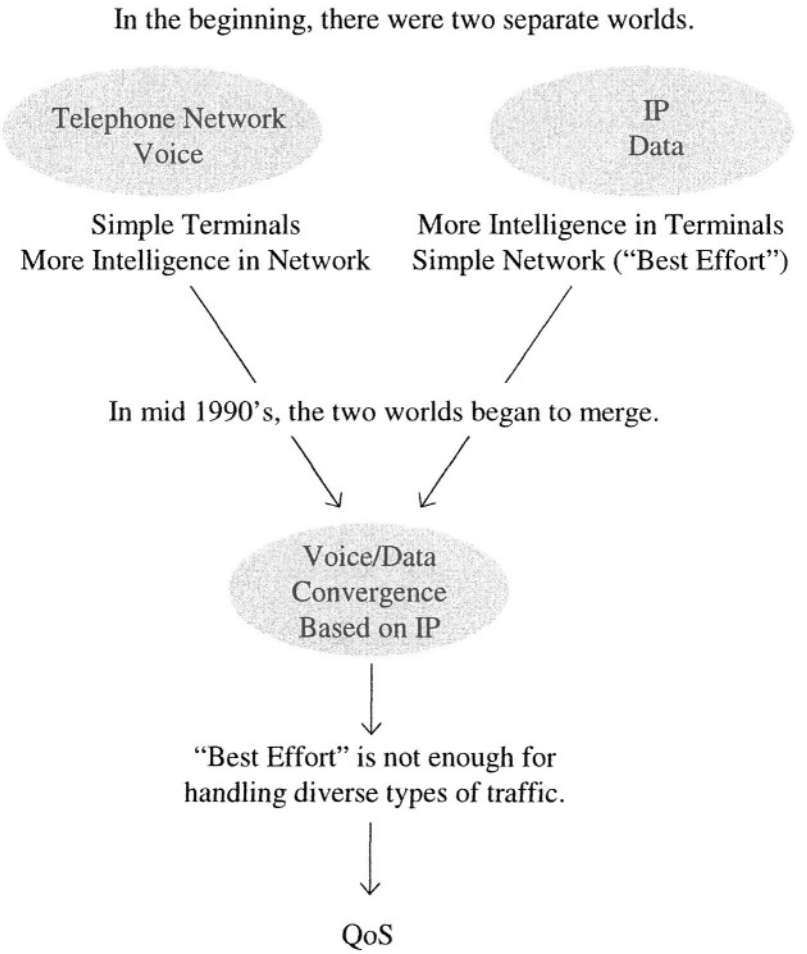

Figure 1-1. Telecommunications network evolution.

second was to design the end to end network with a transmission plan optimized for voice so that the network impairments such as loss, noise, echo, and delay were reasonable. Voice was – and still is – a real time communications service, and there were no queues in the original telephone network to store voice signals for later delivery.

The early IP network was a completely different type of network from the telephone network. First of all, the IP network was designed to carry data. Unlike voice, data was – and still mostly is – a non-real time service. Data could be stored in the network and delivered later. If the data was delivered with error, it could be retransmitted. The data service was sometimes referred to as a "store-and-forward" service.

Since the information carried by the IP network was different from that of the telephone network, the design philosophy used for the IP network was also different from that used for the telephone network.

First, in the original IP network, the network *per se* was designed to be as simple as possible. The main function of the network was to forward packets from one node to the next. All packets were treated the same way and stored in a single buffer and forwarded in a first-in, first-out order.

Second, most of intelligence was placed in the terminal device, which was typically a host computer. For example, if a packet arrived at its destination with error, the receiving terminal would send the sending terminal a negative acknowledgement and the sending terminal would retransmit the packet. The capability of retransmitting lost or errored packets was placed in the terminal, while the network was unaware of the errored packet.

Because the early IP network carried basically one type of information, "store and forward," non-real time data, the network could be designed to operate in the "best effort" mode treating all packets equally, and, as a result, the simple design paradigm described above was possible. The main design objective of the IP network was to make sure that the end user terminal had the appropriate protocols and intelligence to ensure reliable data transmission so that the network could operate as simply as possible.

Although voice and data have distinctly different traffic characteristics and different performance requirements, since the two types of traffic were carried by two separate networks, it was possible to design the networks in the way best suited for the respective payload. In mid 1990's, however, the two separate networks started to merge. A buzz word around this time was "voice and data convergence." The idea was to create a single network to carry both voice and data. Carriers started to plan to consolidate their hodgepodge of separate networks into single "converged" networks for more efficient and economical operation.

At the time, this idea of creating a single converged network for voice and data seemed no more than an engineer's abstract concept. Today, no one can doubt the reality of converged networks for voice and data.

With this convergence, however, a new technical challenge has emerged. In the converged network, the best effort operation of the earlier IP network is no longer good enough to meet diverse performance requirements, often times conflicting, of various types of information carried by the network. QoS is the technology that provides solutions to this technical problem.

2. DEFINITION OF QOS

Figure 1-2 shows an end-to-end network, defines QoS, and the relationships between the various QoS topics treated in this book. The end user represents the terminal devices such as a telephone set, a host computer and other end user communications device. It also represents the human beings who use these terminal devices. The network is a packet network that connects the two end users.

Referring to Figure 1-2, QoS is defined from two points of view: QoS experienced by the end user and the QoS from the point of view of the network. From the end user's perspective, QoS is the end user's perception of the quality that he receives from the network provider for the particular service or application that he subscribes to, e.g., voice, video, and data.

From the network's perspective, the term "QoS" refers to the network's capabilities to provide the QoS perceived by the end user as defined above. Two types of network capabilities are needed to provide QoS in packet networks.

First, to provide QoS, a packet network must be able to differentiate between classes of traffic so that the end users can treat one or more classes of traffic differently than others. Second, once the network differentiates between the traffic classes, it must then be able to treat these classes distinctly by providing resource assurance and service differentiation within the network.

The end user perception of the quality is determined by subjective testing as a function of the network impairments such as delay, jitter, packet loss, and blocking probability. The amount of impairment introduced by a packet network depends on the particular QoS mechanism implemented in the network.

Since a network typically carries a mix of traffic types with different performance requirements, one type of impairment important to a particular service or application may not be as important to other types of service or application and *vice versa.* A QoS mechanism implemented in a network

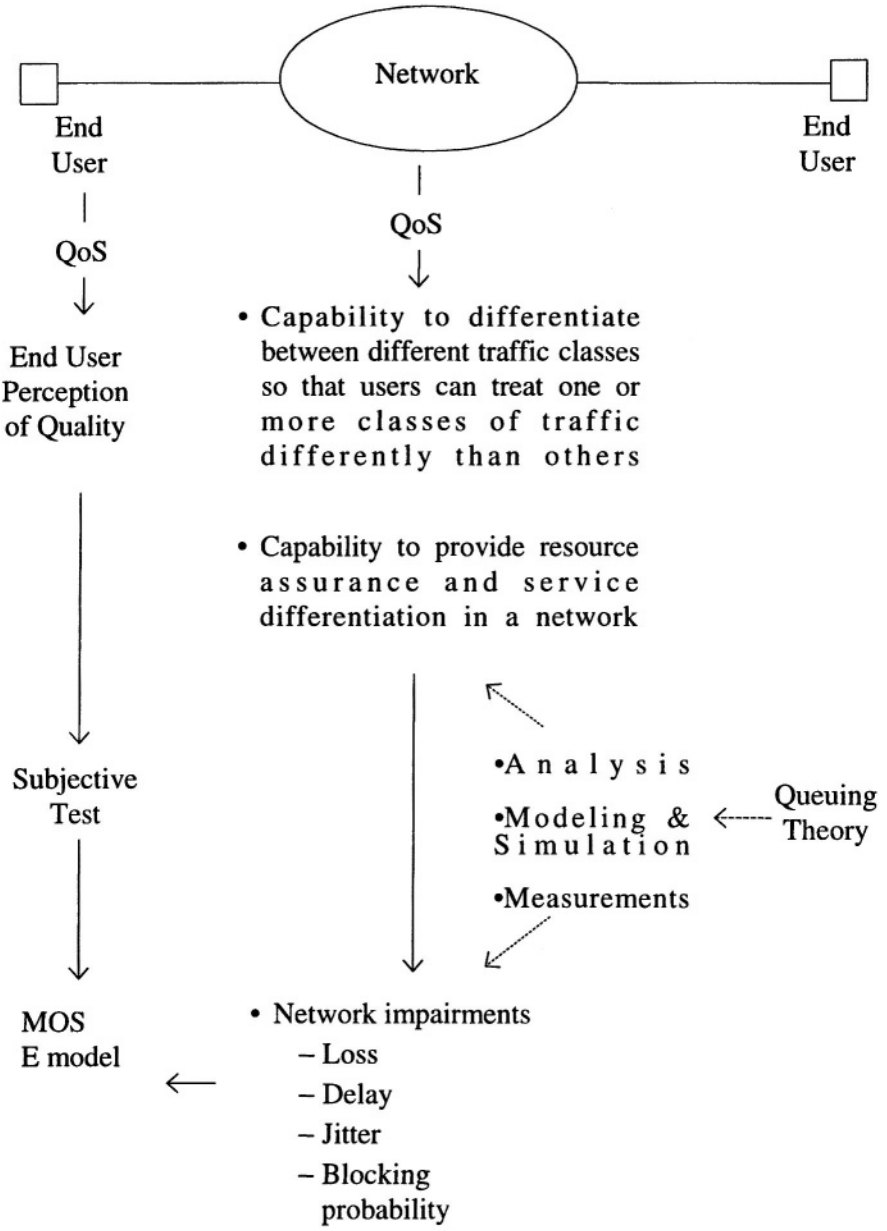

Figure 1-2. Definition of QoS.

must therefore consider various conflicting performance requirements and optimize the trade-off between the impairments.

3. ORGANIZATION OF THE BOOK

Figure 1-2 also serves as a roadmap for this book. As shown in the figure, designing QoS mechanisms for a packet network involves analysis, modeling, simulation, and measurements of network performance. The fundamental mathematical disciplines employed in QoS studies include probability theory, random variables, stochastic processes, and queuing theory. A basic understanding of these mathematical topics, at least at a conceptual level, will help the reader to gain a better appreciation of the QoS topics treated in this book.

The book appropriately begins with a concise treatment of these concepts. The main focus of Chapter 2 is to explain these concepts in plain terms without necessarily involving rigorous mathematics. Throughout the book, application of the mathematics discussed in this chapter will be discussed when appropriate.

Chapter 3 discusses the performance metrics used for QoS from the points of view of the end user and the network. This chapter examines the basic elements of digital communications systems and packet networks and the various types of network impairments generated by the networks. This chapter also discusses subjective testing and the Erlang B and Erlang C models for calculating blocking probability of connection setup attempts.

Chapter 4 and Chapter 5 deal with IP QoS. Chapter 4 explores the generic functional capabilities required in IP networks to provide QoS. It discusses packet marking, packet classification, traffic policing and shaping, traffic metering and coloring, Active Queue Management (AQM), and packet scheduling. Specific topics in this chapter include the single rate three color marker (srTCM) and the two rate three color marker (trTCM); the Random Early Discarding (RED) and the Weighted RED (WRED); the Explicit Congestion Notification (ECN) method of AQM; and various types of packet scheduling including the Priority Queuing (PQ), the Fair Queuing (FQ), the Weighted Fair Queuing (WFQ), and the Class-Based WFQ.

Chapter 5 examines two specific IP QoS mechanisms referred to as the Integrated Services (IntServ) and the Differentiated Services (DiffServ). It discusses briefly the reservation protocol (RSVP) used for IntServ. For DiffServ, the DiffServ Code Points (DSCP's), the Per Hop Behavior, the Expedited Forwarding (EF), and the Assured Forwarding PHB are discussed.

Chapter 6 explains QoS in the Asynchronous Transfer Mode (ATM) network. It discusses various types of ATM virtual connections such as the Virtual Path Connection (VPC) and the Virtual Channel Connection (VCC), ATM service classes such as the Constant Bit Rate (CBR) and the variable Bit Rate (VBR) services, and Connection Admission Control (CAC) methods.

Finally, Chapter 7 discusses Multi-Protocol Label Switching (MPLS). The discussion includes the architecture, implementation and operation of MPLS as well as how MPLS and DiffServ can be used together.

Chapter 2

BASIC MATHEMATICS FOR QOS

To understand QoS in packet networks, it is important to understand not only the mechanism of providing QoS but also the performance behavior that is produced by the QoS mechanism. This chapter reviews some of the basic mathematics that is needed in the analysis of QoS performance in packet networks. The following topics are reviewed in this chapter:

- probability
- random variables
- stochastic processes
- queuing theory

From the author's experience of teaching, students generally considered the mathematical concepts and disciplines such as probability theory, random variables and stochastic processes to be too abstract and hard to apply to real problems.[1-3] One of the purposes of this chapter is to explain the abstract concepts in layman's terms as much as possible so that they can be applied to real problems such as QoS.

1. PROBABILITY THEORY

1.1 Random experiments, outcomes and events

A random experiment is an experiment that produces random outcomes. For example, throwing a die is a random experiment in which each trial produces a random outcome from six possible outcomes, i.e., faces with one through six spots. The word "experiment" implies that the random situation

under consideration is controlled. However, the word may also be used in a broad sense to mean any random situation that produces random outcomes, let us say, a nature's experiment.

A trial is a single instantiation of a random experiment. If a die is thrown ten times, there would be ten trials. The key concept to note here is that each trial produces exactly one outcome.

Another term frequently used in probability is a random "event." A random event is a higher level outcome that may depend on multiple experiments and multiple outcomes of the experiments. For example, consider a game consisting of two random experiments, "throwing a die" and "throwing a coin." A player is to throw the die twice and the coin once. A player who gets the face with one spot in both die-throwings and a "head" in the coin-throwing wins the grand prize. In this game, the random "event" of interest is "winning the grand prize." This event would "occur," if the trials produce the following outcomes: one spot in both of the die-throwings and a "head" in the coin-throwing. In this example, the event depends on multiple experiments and multiple outcomes.

In set theory, a set is defined by the elements contained in the set, e.g., a set of all integers, a set of all even integers, and a set of positive numbers. Using set theory, an event is defined as a set containing the outcomes that make the event happen. For example, in the die-throwing experiment, an event called "face with an even number of spots" may be defined by a set denoted by say E as follows: E= *{"two," "four," "six"},* where *"two" "four"* and *"six"* denote the number of spots on the face of the die.

A random event defined by a set containing a single outcome is referred to as an "elementary event." For example, in the die throwing example, there are six possible random outcomes: *"one," "two," "three," "four," "five,"* and *"six"*. If each of these possible outcomes is defined to be an event, the six possible outcomes produce six elementary events: *{"one"}, {"two"}, {"three"}, {"four"}, {"five"},* and *{"six"}.*

The distinction between the outcome, e.g., *"one,"* and the event, e.g., *{"one"},* is significant and fundamental in the construct of probability theory because, as we shall see in Section 1.3, probability is defined for an event given the probabilities of the underlying random outcomes. *"One"* is an element of a set, whereas *{"one"}* is a set containing one element, *"one."* The probabilities of elementary events would then be equal to the probabilities of the random outcomes.

1.2 Definition of probability

What is probability? Mathematicians attempted to define this seemingly simple term without much success in reaching a consensus for a long time

until Kolmogorov presented his celebrated theory referred to as the "axiomatic approach." The power of the axiomatic approach is in its simplicity.

First, consider the debate that went on before Kolmogorov. A probability was defined as a frequency of occurrence. Consider 1,000 trials in the coin throwing experiment. If the head shows up 400 times, it is concluded that the "probability" of a head is 0.4. The dilemma of this definition of probability is that unless the coin is thrown many times and the outcomes are observed, there is no way of telling the probability.

Some would say that the probability of head should be 0.5 but then others would argue that, unless the coin is minted "perfectly" with identical sides, no one can say that its probability is 0.5 even though it may be "close," etc., etc. Mathematicians had difficulty overcoming the arguments such as this and, as a result, probability theory could not be developed into a useful discipline that could be applied to practical problems.

Most reasonable persons could agree, deep in their hearts, that it should be good enough to take the probability of, for example, a particular face in die throwing is 1/6 and move on to solve other probability problems associated with die throwing. If the 1/6 probability for a face is accepted, then one can find, for example, the probability of a face with an even number of spots, which would be 0.5, etc. With the frequency definition of probability, this simple solution would not be possible. Such an approach is possible because human beings are given this innate capability of *a priori* reasoning.

Kolmogorov presented this simple idea based on *a priori* reasoning that freed everyone interested in probability from the endless arguments. His approach is referred to as the "axiomatic probability theory" and is based on set theory and measure theory. His idea was that there was no need to determine whether a coin was minted perfectly to discuss its probability. He simply turned the table around and asserted that one could "assign" probabilities to the outcomes based on the *a priori* knowledge of the outcomes and let the probabilities initially assigned be the starting point for developing more complex probability theory just like accepting 1/6 as the probability of a face in die throwing.

The key concept is in the word "assign." In this approach, probability "begins" with the assignment of it based on one's own judgment about the likelihood of the outcome. In the axiomatic approach, one can start with "assigning" 1/6 each as the probability of a face in the die-throwing experiment. Once this initial assignment of probability is "accepted" (as an axiom, so to speak), it is now possible to solve all kinds of complex and interesting probability problems associated with die-throwing.

For example, what is the probability of getting an even number of spots? Since the 1/6 probability is "accepted," one can proceed to find its answer, which is 0.5. What is the probability of getting a face with more than four spots? Since either five or six spots would make this event happen, the answer would be 2/6.

1.3 Axiomatic approach to probability

A mathematical system, e.g., linear algebra, set theory, and group theory, is simply an artifact that is useful because it provides a structure for drawing meaningful inferences. The axiomatic probability theory is such a mathematical system.

Consider a random experiment with n possible outcomes, $\xi_1, \xi_2, \ldots, \xi_n$. The probability space S is defined as the set of all possible random outcomes of a random experiment as follows:

$$S = \{ \xi_1, \xi_2, \ldots\ldots, \xi_n \} \tag{2-1}$$

A "measure" is "assigned" to each outcome, ξ_i. This measure is referred to as "probability." Denote this measure by p_i. The measure chosen is a real number between 0 and 1 as follows:

$$0 \le p_i \le 1 \tag{2-2}$$

$$p_i = P(\xi_i) = \textit{probability of random outcome } \xi_i \tag{2-3}$$

The word "probability" was difficult to define because of the attempts to define its meaning semantically and in some instances philosophically. In the axiomatic probability theory, its definition is simply a "measure" that is assigned to an outcome. In fact, this measure does not have to be a number between 0 and 1. It can be a number between 0 and 100 or any number for that matter without changing the axiomatic theory. It is conventional though to use a number between 0 and 1 as a probability measure.

An axiom is a statement accepted as a truth or a rule as a basis of inference. Given the probability space S of (2-1) and the probability measures of the random outcomes of the experiment of (2-3), the axiomatic probability theory is based on the following three simple axioms:

$$\textit{Axiom I} \quad P(A) \ge 0 \tag{2-4}$$

$$\textit{Axiom II} \quad P(S) = 1 \tag{2-5}$$

$$\textit{Axiom III} \quad \textit{If } A \cap B = \{\phi\}, P(A \cup B) = P(A) + P(B) \tag{2-6}$$

In the above equations, S is a set referred to as the probability space defined earlier. A and B are subsets of S and define the random events of interest. Since A and B define the events, they are sometimes simply referred to as "events." S is also a set and, as such, also an event. Since S includes all possible outcomes, any outcome will make S happen and so S is referred to as a certain event. Similarly, $\{\phi\}$ is a set that contains no element. No outcome will make $\{\phi\}$ happen, and $\{\phi\}$ is referred to as an impossible event. Two set operations are used in these axioms. $A \cap B$ is an intersection of A and B, a set of elements belonging to both A and B. $A \cup B$ is a union of A and B, a set of elements belonging to either A or B.

Axiom I states that any event defined in the probability space is assigned a non-negative measure or probability. This is simply an agreement to start the theory. It is entirely possible in the axiomatic theory to use negative numbers for probability as long as that is agreed to at the beginning of the framework because probability is simply nothing more than a numerical measure in the axiomatic theory. However, it would be cumbersome to think in negative numbers when one considers probability.

Axiom I defines the starting point of development of a probabilistic framework of a random experiment under consideration. First, define the elementary events $\{\xi_i\}$ and assign probabilities to them, $P(\{\xi_i\})$. Note the distinction between $P(\{\xi_i\})$ and $P(\xi_i)$. The former is the probability of the elementary event $\{\xi_i\}$ and the latter, that of a random outcome ξ_i. It is important to note that the starting point of the axiomatic framework, i.e., Axiom I, is $P(\{\xi_i\})$ and not $P(\xi_i)$.

Axiom II states that the probability of the space S is one. The space S is a set that contains all possible outcomes under consideration and it would be reasonable to accept as a basic truth that the probability of all possible outcomes is one.

In effect, Axiom II simply states that the probability of certainty is one. One may then ask what about the probability of impossibility, i.e., a null event. Don't we need an axiom, say Axiom IIa that states $P(\{\phi\}) = 0$? It can be shown that the three axioms cover this axiom and adding it would be superfluous because it can be derived from Axioms II and III as follows.

From set theory, the union of the space S and the null set $\{\phi\}$ is the space S and the intersection of the space S and the null set $\{\phi\}$ is the null set $\{\phi\}$:

$$S \cup \{\phi\} = S \tag{2-7}$$

$$S \cap \{\phi\} = \{\phi\} \tag{2-8}$$

From Equation (2-7), it follows that:

$$P(S) = P(S \cup \{\phi\}) \tag{2-9}$$

Equation (2-8) satisfies the condition for Axiom III. Hence, from Axiom III and Equation (2-9), it follows that:

$$P(S) = P(S \cup \{\phi\}) = P(S) + P(\{\phi\}) \tag{2-10}$$

From Axiom II and Equation (2-10), it follows that:

$$P(S) = P(S \cup \{\phi\}) = P(S) + P(\{\phi\}) = 1 \tag{2-11}$$

Finally, from Equation (2-11), it follows that:

$$P(\{\phi\}) = 1 - P(S) = 0 \tag{2-12}$$

Note that Axiom I states $P(A) \geq 0$ but it does not include $P(A) \leq 1$. Once again, the reason is because it can be derived from other axioms and including $P(A) \leq 1$ would be superfluous.

Example 1

A box contains a total of 10 balls of different colors as follows: two white balls, three red balls and five black balls. A player is to withdraw a ball, and, if the ball withdrawn is either red or black, the player wins a piece of candy. What is the probability of winning a piece of candy by playing this game?

Solution

There are eight red or black balls out of a total of 10 balls, and so the probability of winning the grand prize is 0.8. This is a simple problem and one can get the answer quickly in the head without going through the rigor of axiomatic formulation.

However, we shall formulate and solve this problem using the axiomatic approach to illustrate how a probability problem can be formulated and solved systematically. For more complex problems, the disciplined way of dealing with the problem using the axiomatic approach is helpful.

First define the random experiment. There are two alternative ways of defining the space and random outcomes for this problem. Either method should yield the same answer.

Formulation 1. A more direct way of formulation is to define the outcomes of ball drawing like the outcomes of die throwing. Imagine that the individual balls can be distinguished (e.g., by numbering them) as the faces of a die are distinguished. Then there are ten possible outcomes with an equal probability as follows:

$$S = \{ \xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \xi_6, \xi_7, \xi_8, \xi_9, \xi_{10} \} \tag{2-13}$$

$$p_i = P(\xi_i) = 1/10; \qquad i = 1, \ldots, 10 \tag{2-14}$$

where ξ_1 and ξ_2 are drawing a white ball, ξ_3, ξ_4 and ξ_5, a red ball and ξ_6 through ξ_{10}, a black ball.

The next step is to define the event. The event of interest is "winning a candy" and is defined as a set denoted by *W*. In set theory, a set is defined by its members or a member is "qualified" to be included in the event set, if it makes that event happen. *W* in turn depends on the following two events:

$$R = \text{"ball withdrawn is red"} = \{ \xi_3, \xi_4, \xi_5 \} \tag{2-15}$$

$$B = \text{"ball withdrawn is black"} = \{ \xi_6, \xi_7, \xi_8, \xi_9, \xi_{10} \} \tag{2-16}$$

Since $\{\xi_i\}$'s are mutually exclusive, i.e., $\{\xi_i\} \cap \{\xi_j\} = \{\phi\}$ *for i, j = 3 - 8,* it follows that:

$$R = \{ \xi_3, \xi_4, \xi_5 \} = \{ \xi_3 \} \cup \{ \xi_4 \} \cup \{ \xi_5 \}$$

$$= [\{ \xi_3 \} \cup \{ \xi_4 \}] \cup \{ \xi_5 \} \tag{2-17}$$

Applying Axiom III twice, it follows that:

$$P(R) = P(\{ \xi_3, \xi_4, \xi_5 \}) = P([\{ \xi_3 \} \cup \{ \xi_4 \}] + P(\{ \xi_5 \})$$