Applied Probability and Statistics

Mario Lefebvre

Applied Probability and Statistics

With 58 illustrations

Springer

Mario Lefebvre
Département de mathématiques
et de génie industriel
École Polytechnique de Montréal, Québec
C.P. 6070, succ. Centre-ville
Montréal H3C 3A7
Canada
mlefebvre@polymtl.ca

Cover design by Mary Burgess.

Mathematics Subject Classification (2000): 60-01, 62-01

Library of Congress Control Number: 2006923352

ISBN-10 0-387-28454-0 e-ISBN 0-387-28505-9
ISBN-13 978-387-28454-5

Printed on acid-free paper.

Printed in the United States of America. (TXQ/EB)

9 8 7 6 5 4 3 2 1

springer.com

To my parents

Preface

This book is based mainly on the lecture notes that I have been using since 1993 for a course on applied probability for engineers that I teach at the École Polytechnique de Montréal. This course is given to electrical, computer and physics engineering students, and is normally taken during the second or third year of their curriculum. Therefore, we assume that the reader has acquired a basic knowledge of differential and integral calculus.

The main objective of this textbook is to provide a reference that covers the topics that every student in pure or applied sciences, such as physics, computer science, engineering, etc., should learn in probability theory, in addition to the basic notions of stochastic processes and statistics. It is not easy to find a single work on all these topics that is both succinct and also accessible to non-mathematicians.

Because the students, who for the most part have never taken a course on probability theory, must do a lot of exercises in order to master the material presented, I included a very large number of problems in the book, some of which are solved in detail. Most of the exercises proposed after each chapter are problems written especially for examinations over the years. They are not, in general, routine problems, like the ones found in numerous textbooks.

The exercises that can be done after a given section is read are listed in Appendix C. The reader will also find, in Appendix D, the answers to all the multiple choice questions proposed in the manual. (Of course, the student is recommended to first try to solve an exercise before looking at the answer.) Appendix E provides the answers to selected supplementary exercises included in Chapters 6 and 7.

The book contains a few biographical notes on nearly all the mathematicians mentioned in the text. The reader interested in learning more about these great mathematicians can consult the various books or Web sites dedicated to the biographies of scientists.

Most of the figures in this book were realized with the help of a software program that enables one to *draw* curves or diagrams. When the figures involved mathematical functions, such as the exponential function, a mathematical software was used, when possible, to obtain precise curves.

I wish to express my gratitude to my colleagues who taught the course on probability theory for engineers with me during the last ten years. They contributed in providing interesting exercises that are now part of this manual. I am also grateful to Jean-Luc Guilbault, who helped me by typing most of the exercises found in the book. This was made possible by a grant from the Service pédagogique of the École Polytechnique, which I thank as well.

Mario Lefebvre
Montréal, August 2005

Contents

List of Tables

List of Figures

1

Introduction

1.1 The Beginnings of Probability

We often hear that the theory of probability started in the seventeenth century, when a French nobleman, the Chevalier de Méré, proposed the following problem in 1654 to his friend Pascal: Why is one more likely to obtain a "6" in four throws of a die than to obtain a double "6" in 24 throws of two dice? This problem is known as *de Méré's paradox*. We use the word *paradox*, because, based on the fact that there are 6 possible results when we roll a die and 36 possible results when we roll two dice, some people thought that the two *events* above should have the same *probability*. Indeed, notice that the number of throws, divided by the number of possible results, is equal to 2/3 in both cases (4/6 = 24/36 = 2/3). Nowadays, we can easily compute the probability of each event. We find that the probability of obtaining at least one "6" in four rolls of a (fair or non-biased) die is $1 - (5/6)^4 = 671/1296 \simeq 0.5177$, while the probability of getting at least a double "6" in throwing two dice 24 times is $1 - (35/36)^{24} \simeq 0.4914$. We can deduce that the Chevalier de Méré must have spent a lot of time throwing dice to discover such a small difference!

According to some historians, this problem was not proposed by de Méré. William Feller, who wrote two books on probability theory which are considered as true classics, mentions that the problem was in fact first treated by Cardano[1] in the preceding century. Gerolamo Cardano, or Jerome Cardan as he is called in English, was a colorful character who, in addition to being a mathematician, was also a doctor and an astrologer. He was an inveterate gambler, which caused him many problems. He also analyzed dice games and a card game similar to poker. Furthermore, he made astrological predictions. It is said that he had predicted the day of his death. Since he was in good health when the day in question arrived, he would

[1] Gerolamo Cardano, 1501–1576, was born and died in Italy. He became interested in the domain of probability to gain an advantage over his opponents in card and dice games. He also worked on algebraic equations. He gave the resolution method for third- and fourth-degree equations. The formula for the solution of third-degree equations had been previously obtained by the Italian mathematician Tartaglia.

certainly have committed suicide so as not to lose face! Note that another version of the story says that he managed to die of hunger on that day.

Remark. In order to pay homage to the great mathematicians who left their mark on the history of probability, we have included some biographical notes on almost every one whose name appears in the book, for example, Poisson, who gave his name to a random variable and an important stochastic process.

One thing is certain, Pascal[2] exchanged correspondence with Fermat[3] concerning the above-mentioned problem (see reference [5], p. 128, where an excerpt of a letter is reproduced). They also exchanged letters about other games of chance, including one known as the *problem of points*, which contributed greatly to the development of the domain of probability.

The first complete treatise on the calculus of probabilities was written by Huygens[4] in the seventeenth century. It is however James (or Jacob) Bernoulli (see p. 70), in posthumous works published in 1713, who really founded the calculus of probabilities. Afterward, Laplace (see p. 85), with his book *Théorie analytique des probabilités* written between the years 1812 and 1820, developed the theory of probability in a more rigorous way. Finally, in the twentieth century, Kolmogorov (see p. 227) gave the domain of probability its modern formulation.

1.2 Examples of Applications

If the conditions under which an experiment is carried out determine the result of this experiment, then it is said to be *deterministic*. For example, suppose that we observe an object moving in the sky along a decreasing exponential trajectory. Suppose also that we can control this object from a distance. Let x be the height of the object with respect to the ground at time 0. Then, if there are no perturbations, we can use the following model to determine the value of $x(t)$, the height of the object at time t:

[2] Blaise Pascal, 1623–1662, was born and died in France. His father, Étienne, was his professor. He invented a *calculating machine* to help his father with his work as a tax collector. Along with Fermat, he is one of the founders of the theory of probability. He was also interested in geometry and physics. From 1654 on, he became deeply religious and published books on philosophy and theology.

[3] Pierre de Fermat, 1601–1665, was born and died in France. He is especially known for his work on number theory, in particular his famous "last theorem." He has been a precursor in the domains of probability, differential calculus and analytic geometry. He was also a lawyer and, in addition to doing his research in mathematics, he was a councillor for the parliament at Toulouse.

[4] Christiaan Huygens, 1629–1695, was born and died in the Netherlands. He studied law and mathematics at the university of Leiden. Descartes showed interest in his mathematical progress. He worked, in particular, in astronomy, mechanics and optics. In 1655, using instruments he had made himself, he discovered Titan, the largest satellite of Saturn. The theory of pendulum motion is also due to him.

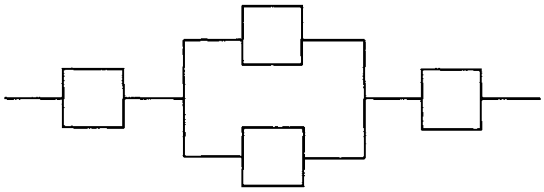

Figure 1.1. Example of a system.

$$\begin{aligned} dx(t) &= -ax(t)\,dt + bu(t)\,dt, \\ x(0) &= x, \end{aligned} \tag{1.1}$$

where a (> 0) and b are constants and the variable $u(t)$ is called the control.

Whittle (see [24]) and the author used the equations above in research papers as a rudimentary model for the displacement of an airplane that is preparing to land.

When we cannot predict the result of an experiment repeated under the same conditions, we say that it is a *random* experiment. We can then make a list of all possible outcomes and try to compute the *probability* of each of these possible outcomes. In this textbook, we are only interested in random experiments.

Probability is used in practically every pure or applied science. Examples in engineering, particularly in electrical engineering, where we must resort to the calculus of probabilities are the following.

a) Many systems may be represented by a number of components placed in series or in parallel. For example, consider the system described by the diagram in Fig. 1.1. In **reliability**, we must be able to calculate the probability that such a system will function during a certain period of time, or at a given instant. In the latter case, we must then know, for each component, the probability that it functions at this instant and take into account the fact that the components perhaps do not operate independently from one another. In the case of reliability during a given period of time, we must know the distribution of the lifetime of each component.

b) In **communication**, we must often take into account the "noise" present in a system. For example, suppose that a system transmits either a 0 or a 1, and that there is a risk p that the number transmitted will not be received correctly (see Fig. 1.2). We may be interested in computing the probability that a 0 has been transmitted, given that a 0 has been received, or that a transmission error has occurred, etc.

c) In **automatic control**, if we take the random perturbations into account, then the model (1.1) above becomes

$$dx(t) = -ax(t)\,dt + bu(t)\,dt + dW(t), \tag{1.2}$$

where $W(t)$ is called a Brownian motion (or a Wiener process). We also say that $dW(t)/dt$ is a Gaussian white noise. Equation (1.2) is an example of a *stochastic differential equation*.

d) In **computer science**, probability is used, in particular, to help us make decisions in expert systems. We also make use of probability in simulation and in artificial intelligence, as well as in the field of queueing theory.

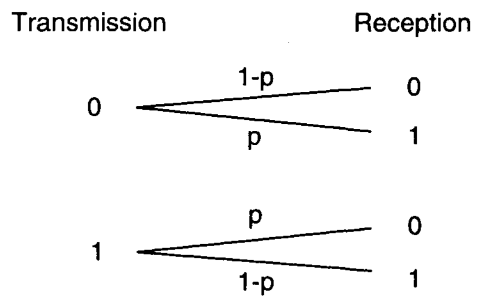

Figure 1.2. Example of a communication system.

e) In **physics**, the domain known as *statistical physics* requires some knowledge of the theory of probability, and so does that of *quantum mechanics*. In this last case, we must be familiar with diffusion processes, such as the Wiener process.

1.3 Relative Frequencies

To obtain the probability of one of the possible outcomes when we perform a random experiment, we repeat this experiment a large number of times. Let

$$f_k(n) := \frac{N_k(n)}{n}, \tag{1.3}$$

where $N_k(n)$ is the number of times that the possible outcome k has occurred during n repetitions of the experiment. The quantity $f_k(n)$ is called the **relative frequency** of outcome k. If there are K possible outcomes, which we denote by $1, 2, \ldots, K$, then we may write that

$$0 \leq f_k(n) \leq 1 \quad \text{for } k = 1, 2, \ldots, K \tag{1.4}$$

and that

$$\sum_{k=1}^{K} f_k(n) = 1. \tag{1.5}$$

Indeed we have, of course: $N_k(n) \in \{0, 1, \ldots, n\}$, so that $f_k(n) \in [0, 1]$, and

$$\sum_{k=1}^{K} f_k(n) = \frac{N_1(n) + \cdots + N_K(n)}{n} = \frac{n}{n} = 1.$$

Moreover, if A is a set that contains two possible outcomes, j and k, then

$$f_A(n) = f_j(n) + f_k(n), \tag{1.6}$$

because j and k cannot occur on the *same* repetition of the experiment.

For instance, if we consider the random experiment that consists in observing the outcome of the roll of a die, then there are $K = 6$ possible results: $1, 2, \ldots, 6$. Let A be the set $\{1, 6\}$. Because we cannot obtain both "1" and "6" on the same roll, we may write that $f_A(n) = f_1(n) + f_6(n)$.

Finally, the **probability** of the outcome k is obtained by taking the limit of $f_k(n)$ as the number n of repetitions tends to infinity:

$$P[\{k\}] := \lim_{n \to \infty} f_k(n). \tag{1.7}$$

2

Elementary Probabilities

2.1 Basic Concepts

Definition 2.1.1. *An experiment that can be repeated under the same conditions and whose outcome cannot be predicted with certainty is called a* **random** *experiment.*

Example 2.1.1 A box contains 10 brand A transistors and 10 brand B transistors. We consider the following four random experiments:
E_1: three transistors are taken, at random and with replacement, and the number of brand A transistors (among the three selected) is counted.

Remark. Sampling *with* (respectively *without*) replacement means that the object that has been selected *is* (resp. *is not*) replaced in the box before taking the next one. Therefore, in the case of sampling with (resp. without) replacement, the same object can (resp. cannot) be selected more than once.

E_2: three transistors are taken, at random and with replacement, and the brand of each transistor is noted.

E_3: transistors are taken one at a time, at random and with replacement, until a brand A transistor has been obtained; the number of brand B transistors taken before obtaining a brand A transistor is counted.

E_4: a transistor is taken at random and its lifetime is measured (in hours).

Definition 2.1.2. *The set S of all possible outcomes of a random experiment is called the* **sample space** *of this experiment. Each possible outcome is also called an* **elementary event**.

When a repetition of a random experiment is performed, one and only one of the elementary events occurs. That is, the elementary events are **incompatible** (or mutually exclusive) and **exhaustive**.

Example 2.1.1 (continued) Corresponding to the random experiments above, we have the following sample spaces:

$S_1 = \{0, 1, 2, 3\}$.
$S_2 = \{AAA, AAB, ABA, BAA, ABB, BAB, BBA, BBB\}$.
$S_3 = \{0, 1, \dots\}$.
$S_4 = [0, \infty)$.

The number of elementary events in a sample space may be finite (S_1 and S_2), denumerably infinite (S_3), or non-denumerably infinite (S_4).

Remark. A set is called denumerably (or countably) infinite if we can establish a one-to-one relationship between its elements and the positive integers.

A sample space that is finite or denumerably infinite is said to be **discrete**, whereas if S is non-denumerably infinite, we say that it is **continuous**.

Remark. In this chapter, we will not consider the case when the sample space S would be the union of a finite or denumerably infinite set of points and a non-denumerably infinite set of points. For instance, let E be the following random experiment: first a coin is tossed; if we get "tails," then the point 0 is chosen, otherwise a point is taken at random in the interval $[1, 2]$. In this case, we would have $S = \{0\} \cup [1, 2]$. An example like this will be called a "mixed" type in Chapter 3.

Definition 2.1.3. *An* **event** *is a subset of the sample space S. Thus, the empty set $\emptyset$ and the sample space S itself are events.*

Remarks. i) The empty set $\emptyset$ is called the **impossible** (or null) event and S is the **certain** event.
ii) There are 2^n events that can be defined with n elementary events.
iii) We generally use capital letters like A, B, etc., to denote events.

Example 2.1.1 (continued) Events defined with respect to the sample spaces associated with the random experiments above are the following:

A_1: exactly one brand A transistor is obtained; that is, $A_1 = \{1\}$.

A_2: one brand A transistor and two brand B transistors are obtained; that is, $A_2 = \{ABB, BAB, BBA\}$.

A_3: five or six brand B transistors are picked before a (first) brand A transistor is obtained; that is, $A_3 = \{5, 6\}$.

A_4: the selected transistor lasts more than 200 hours; that is, $A_4 = (200, \infty)$.

Operations with Sets (see Fig. 2.1, p. 9)

Union: $A \cup B$ denotes the set of outcomes that belong to A **or** to B (or to both). Similarly, in general, for n events we write:

$$A_1 \cup A_2 \cup \dots \cup A_n = \bigcup_{i=1}^{n} A_i.$$

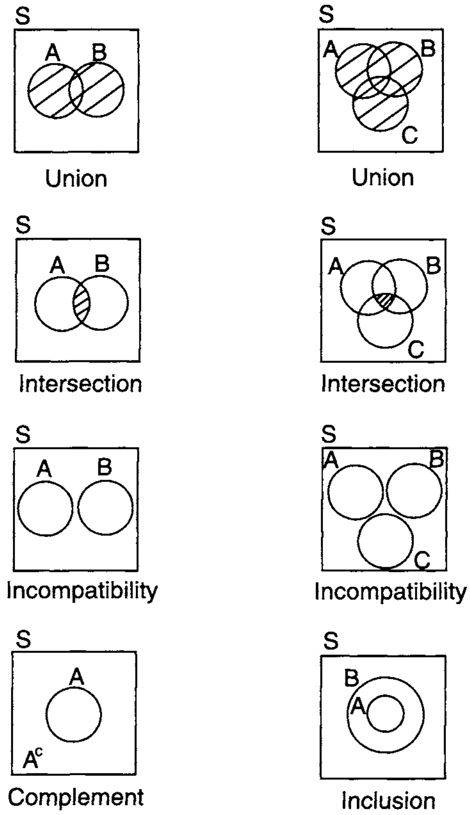

Figure 2.1. Set theory: Venn diagrams.

Intersection: $A \cap B$ denotes the set of outcomes that belong to both A **and** B. In general, we have:

$$A_1 \cap A_2 \cap \ldots \cap A_n = \bigcap_{i=1}^{n} A_i.$$

If two events are incompatible, then we write: $A \cap B = \emptyset$.

Complement: The complement of an event A is the set of outcomes that do not belong to A; it is denoted by A^c.

Inclusion: If all the outcomes that belong to event A also belong to event B, then we say that A is included in B and we write: $A \subset B$.

Equality: Two events are said to be equal if they contain the same outcomes; we then write: $A = B$.

We can easily show the following relationships:

1) $A \cup B = B \cup A$ and $A \cap B = B \cap A$ (Commutativity).

2) $A \cup (B \cup C) = (A \cup B) \cup C$ and $A \cap (B \cap C) = (A \cap B) \cap C$ (Associativity).

3) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ and $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (Distributivity).

4) $(A \cap B)^c = A^c \cup B^c$ and $(A \cup B)^c = A^c \cap B^c$ (De Morgan's laws).

2.2 Probability

Definition 2.2.1. *Let E be a random experiment and S a sample space associated with E. To each event A in S we assign a real number noted $P[A]$, called the* **probability** *of A, so that the following properties are satisfied:*

Axiom I: $P[A] \geq 0 \; \forall \; A \subset S$;

Axiom II: $P[S] = 1$;

Axiom III: *If $A_1, A_2, \ldots$ is a sequence of incompatible events, then*

$$P\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} P[A_k]. \tag{2.1}$$

Remarks. i) We deduce from Axiom III that if $A \cap B = \emptyset$, then

$$P[A \cup B] = P[A] + P[B]. \tag{2.2}$$

Indeed, we only have to take $A_1 = A$, $A_2 = B$ and $A_k = \emptyset$ for $k \geq 3$, because $P[\emptyset] = 0$ (see Proposition 2.2.1).

ii) The function P is called a *probability measure*; it is a function from S into the interval $[0, 1]$.

With the help of the three axioms above, we easily show the following proposition.

Proposition 2.2.1. *We have:*

1) $P[A^c] = 1 - P[A]$.

2) $P[A] \leq 1$.

3) $P[\emptyset] = 0$.

4)

$$P\left[\bigcup_{k=1}^{n} A_k\right] = \sum_{k=1}^{n} P[A_k] - \sum_{j<k} P[A_j \cap A_k] + \cdots + (-1)^{n+1} P\left[\bigcap_{k=1}^{n} A_k\right].$$

5) *If $A \subset B$, then $P[A] \leq P[B]$.*

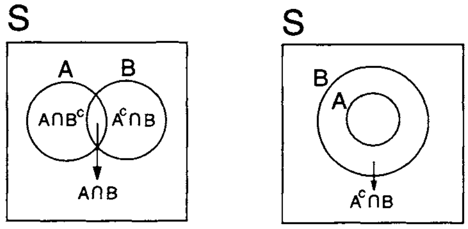

Figure 2.2. Figures for the proof of parts 4) and 5) of Proposition 2.2.1.

Proof.
1) Since $A^c \cup A = S$ and $A^c \cap A = \emptyset$, we deduce from Axioms II and III that

$$P[A^c \cup A] \overset{\text{II}}{=} 1 \quad \text{and} \quad P[A^c \cup A] \overset{\text{III}}{=} P[A^c] + P[A] \;\Rightarrow\; P[A^c] = 1 - P[A].$$

2) $P[A] = 1 - P[A^c] \leq 1$ because $P[A^c] \geq 0$ (by Axiom I).

3) We have: $S^c = \emptyset$ and $P[S^c] = 1 - P[S] \overset{\text{II}}{=} 1 - 1 = 0$.

4) For $n = 2$, let us write $A_1 = A$ and $A_2 = B$; we then have (see Fig. 2.2):

$$P[A \cup B] = P[A \cap B^c] + P[A \cap B] + P[A^c \cap B],$$

because the three events are incompatible. Moreover,

$$P[A \cap B^c] = P[A] - P[A \cap B] \quad \text{and} \quad P[A^c \cap B] = P[B] - P[A \cap B].$$

So, we have:

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]. \tag{2.3}$$

Next, let $D := B \cup C$; then, we may write that

$$\begin{aligned} P[A \cup B \cup C] &= P[A \cup D] = P[A] + P[D] - P[A \cap D] \\ &= P[A] + P[B] + P[C] - P[B \cap C] - P[A \cap (B \cup C)] \\ &= P[A] + P[B] + P[C] - P[B \cap C] - P[(A \cap B) \cup (A \cap C)]. \end{aligned}$$

Hence, we obtain the following formula:

$$\begin{aligned} P[A \cup B \cup C] = P[A] + P[B] + P[C] - P[A \cap B] \\ - P[A \cap C] - P[B \cap C] + P[A \cap B \cap C]. \end{aligned} \tag{2.4}$$

To prove the formula in the general case, we proceed by induction (that is, we assume that the formula is true for the case of n events and we try to show that it is then also valid for $n + 1$ events).

5) If $A \subset B$, we may write (see Fig. 2.2) that

$$P[A] = P[B] - P[A^c \cap B] \Rightarrow P[A] \leq P[B]. \qquad \square$$

Discrete Sample Spaces

If S is a discrete sample space, we may write that

$$S = \{e_1, e_2, \ldots\},$$

where e_k is a possible outcome (or an elementary event). Let $A \subset S$; then the probability of event A can be obtained by making use of the following formula:

$$P[A] = \sum_{k:e_k \in A} P[\{e_k\}]. \tag{2.5}$$

If the number n of elementary events e_k is *finite* and if the e_k's are *equiprobable* (or equally likely), so that $P[\{e_k\}] = 1/n \ \forall k$, then we may write that

$$P[A] = n(A)/n, \tag{2.6}$$

where $n(A)$ is the number of elementary events in A.

Example 2.2.1 i) In the case of the sample space S_1 in Example 2.1.1, the four elementary events are *not* equiprobable. If we denote the probability $P[\{0\}]$ by p, then we may write that

$$P[\{1\}] = P[\{2\}] = 3p \quad \text{and} \quad P[\{3\}] = p.$$

Moreover,

$$\sum_{k=0}^{3} P[\{k\}] = 1 \Rightarrow p = 1/8.$$

Hence, we find that

$$P[A_1] = P[\{1\}] = 3/8 \quad (\neq 1/4).$$

On the other hand, the e_k's are *equally likely* in the case of S_2 and we may write, directly, that

$$P[A_2] = P[\{ABB, BAB, BBA\}] = 3/8.$$

Note that the probabilities $P[A_1]$ and $P[A_2]$ must be equal because the events A_1 and A_2 correspond to the same outcome of the three draws in the random experiments E_1 and E_2. Thus, to calculate the probability of getting exactly one brand A transistor in three draws at random and with replacement from a box containing 10 brand A and 10 brand B transistors, it is simpler to consider the sample space S_2 for which the outcomes are equally likely.

ii) Since the sample space S_3 is *denumerably infinite*, the e_k's cannot be equiprobable. We can show that

$$P[A_3] = P[\{5, 6\}] = P[\{5\}] + P[\{6\}] = (10/20)^5(10/20) + (1/2)^6(1/2).$$

The second equality above is obtained by incompatibility of the events $\{5\}$ and $\{6\}$, while the third equality results from the notion of independence (which will be discussed in Section 2.5).

sample content of Applied Probability and Statistics