PROBABILITY

AN INTRODUCTION WITH STATISTICAL APPLICATIONS

SECOND EDITION

JOHN J. KINNEY

WILEY

Probability

Probability

An Introduction with Statistical Applications

Second Edition

John J. Kinney

Colorado Springs, CO

WILEY

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

Library of Congress Cataloging-in-Publication Data:

Kinney, John J.
Probability : an introduction with statistical applications / John Kinney, Colorado Springs, CO. – Second edition.
pages cm
Includes bibliographical references and index.
ISBN 978-1-118-94708-1 (cloth)
1. Probabilities–Textbooks. 2. Mathematical statistics–Textbooks. I. Title.
QA273.K493 2015
519.2–dc23
2014020218

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

This book is for
Cherry and Kaylyn

Contents

Preface for the First Edition

HISTORICAL NOTE

The theory of probability is concerned with events that occur when randomness or chance influences the result. When the data from a sample survey or the occurrence of extreme weather patterns are common enough examples of situations where randomness is involved, we have come to presume that many models of the physical world contain elements of randomness as well. Scientists now commonly suppose that their models contain random components as well as deterministic components. Randomness, of course, does not involve any new physical forces; rather than measuring all the forces involved and thus predicting the exact outcome of an experiment, we choose to combine all these forces and call the result random. The study of random events is the subject of this book.

It is impossible to chronicle the first interest in events involving randomness or chance, but we do know of a correspondence between Blaise Pascal and Pierre de Fermat in the middle of the seventeenth century regarding questions arising in gambling games. Appropriate mathematical tools for the analysis of such situations were not available at that time, but interest continued among some mathematicians. For a long time, the subject was connected only to gambling games and its development was considerably restricted by the situations arising from such considerations. Mathematical techniques suitable for problems involving randomness have produced a theory applicable to not only gambling situations but also more practical situations. It has not been until recent years, however, that scientists and engineers have become increasingly aware of the presence of random factors in their experiments and manufacturing processes and have become interested in measuring or controlling these factors.

It is the realization that the statistical analysis of experimental data, based on the theory of probability, is of great importance to experimenters that has brought the theory to the forefront of applicable mathematics. The history of probability and the statistical analysis it makes possible illustrate a prime example of seemingly useless mathematical research that now has an incredibly wide range of practical application. Mathematical models for experimental situations now commonly involve both deterministic and random terms. It is perhaps a simplification to say that science, while interested in deterministic models to explain the physical world, now is interested as well in separating deterministic factors from random factors and measuring their relative importance.

There are two facts that strike me as most remarkable about the theory of probability. One is the apparent contradiction that random events are in reality well behaved and that there are laws of probability. The outcome on one toss of a coin cannot be predicted, but given 10,000 tosses of the same coin, many events can be predicted with a high degree of accuracy. The second fact, which the reader will soon perceive, is the pervasiveness of a probability distribution known as the normal distribution. This distribution, which will be defined and discussed at some length, arises in situations which at first glance have little in common: the normal distribution is an essential tool in statistical modeling and is perhaps the single most important concept in statistical inference.

There are reasons for this, and it is my purpose to explain these in this book.

ABOUT THE TEXT

From the author's perspective, the characteristics of this text which most clearly differentiate it from others currently available include the following:

- Applications to a variety of scientific fields, including engineering, appear in every chapter.
- Integration of computer algebra systems such as Mathematica provides insight into both the structure and results of problems in probability.
- A great variety of problems at varying levels of difficulty provides a desirable flexibility in assignments.
- Topics in statistics appear throughout the text so that professors can include or omit these as the nature of their course warrants.
- Some problems are structured and solved using recursions since computers and computer algebra systems facilitate this.
- Significant and practical topics in quality control and quality production are introduced.

It has been my purpose to write a book that is readable by students who have some background in multivariable calculus. Mathematical ideas are often easily understood until one sees formal definitions that frequently obscure such understanding. Examples allow us to explore ideas without the burden of language. Therefore, I often begin with examples and follow with the ideas motivated first by them; this is quite purposeful on my part, since language often obstructs understanding of otherwise simply perceived notions.

I have attempted to give examples that are interesting and often practical in order to show the widespread applicability of the subject. I have sometimes sacrificed exact mathematical precision for the sake of readability; readers who seek a more advanced explication of the subject will have no trouble in finding suitable sources. I have proceeded in the belief that beginning students want most to know what the subject encompasses and for what it may be useful. More theoretical courses may then be chosen as time and opportunity allow. For those interested, the bibliography contains a number of current references.

An author has considerable control over the reader by selecting the material, its order of presentation, and the explication. I am hopeful that I have executed these duties with due regard for the reader. While the author may not be described with any sort of precision as the holder of a tightrope, I have been guided by the admonition: "It's not healthy for the tightrope walker to be misunderstood by the person who's holding the rope."[1]

The book makes free use of the now widely available computer algebra systems. I have used Mathematica, Maple, and Derive for various problems and examples in the book, and I hope the reader has access to one of these marvelous mathematical aids. These systems allow us the incredible opportunity to see graphs and surfaces easily, which otherwise would be very difficult and time-consuming to produce. Computer algebra systems make some

[1] *Smilla's Sense of Snow*, by Peter Hoeg (Farrar, Straus and Giroux: New York, 1993).

parts of mathematics visual and thereby add immensely to our understanding. Derivatives, integrals, series expansions, numerical computation, and the solution of recursions are used throughout the book, but the reader will find that only the results are included: in my opinion there is no longer any reason to dwell on calculation of either a numeric or algebraic sort. We can now concentrate on the meaning of the results without being restrained by the often mechanical effort in achieving them; hence our concentration is on the structure of the problem and the insight the solution gives. Graphs are freely drawn and, when appropriate, a geometric view of the problem is given so that the solution and the problem can be visualized. Numerical approximations are given when exact solutions are not feasible. The reader without a computer algebra system can still do the problems; the reader with such a system can reproduce every graph in the book exactly as it appears. I have included a fairly expensive appendix in which computer commands in Mathematica are given for many of the examples in which Mathematica was used; this should also ease the translation to other computer algebra systems. The reader with access to a computer algebra system should refer to Appendix 1 fairly frequently.

Although I hope the book is readable and as completely explanatory as a probability text may be, I know that students often do not read the text, but proceed directly to the problems. There is nothing wrong with this; after all, if the ability to solve practical problems is the goal, then the student who can do this without reading the text is to be admired. Readers are warned, however, that probability problems are rarely repetitive; the solution of one problem does not necessarily give even any sort of hint as to the solution of the next problem. I have included over 840 problems so that a reader who solves the problems can be reasonably assured that the concepts involving them are understood.

The problem sections begin with the easiest problems and gradually work their way up to some reasonably difficult problems while remaining within the scope and level of the book. In discussing a forthcoming examination with my students, I summarize the material and give some suggestions for practice problems, so I have followed each chapter by a Chapter Summary, some suggestions for Review Problems, and finally some Supplementary Problems.

FOR THE INSTRUCTOR

Texts on probability often use generating functions and recursions in the solution of many complex problems; with our use of computer algebra systems, we can determine generating functions, and often their power series expansions, with ease. The structure of generating functions is also used to explain limiting behavior in many situations. Many interesting problems can be best described in terms of recursions; since computer algebra systems allow us to solve such recursions, some discussion of recursive functions is given. Proofs are often given using recursions, a novel feature of the book. Occasionally, the more traditional proofs are given in the exercises.

Although numerous applications of the theory are given in the text and in the problems, the text by no means exhausts the applications of the theory of probability. In addition to solving many practical and varied problems, the theory of probability also provides the basis for the theory of statistical inference and the analysis of data. Statistical analysis is combined with the theory of probability throughout the book. Hypothesis testing, confidence intervals, acceptance sampling, and control charts are considered at various points in the text. The order in which these topics are to be considered is entirely up to the instructor; the book is quite flexible in allowing sections to be skipped, or delayed, resulting in rearrangement of the material. This book will serve as a first introduction to statistics, but the reader who intends to apply statistics should also elect a course in applied statistics. In my opinion, statistics will be the centerpiece of applied mathematics in the twenty-first century.

Preface for the Second Edition

I am pleased to offer a second edition of this text. The reasons for writing the book remain the same and are indicated in the preface for the first edition. While remaining readable and I hope useful for both the student and the instructor, I want to point out some differences between the two editions.

- The first edition was written when Mathematica was in its fourth release; it is now in its ninth release and while its capabilities have grown, some of the commands, especially those regarding graphs, have changed. Therefore, Appendix 1 is totally new, reflecting the changes in Mathematica.
- Both first and second editions contain about 120 graphs; these have been mostly redrawn.
- The problems are of primary importance to the student. Being able to solve them verifies the student's mastery of the material. The book now contains over 880 problems, 60 or so of which are new.
- Chapter 7, titled "Some Challenging Problems", is new. Five problems, or sets of problems, some of which have been studied by famous mathematicians, are introduced. Open questions are given, some of which will challenge the reader. Problems are almost always capable of extension; the reader may do this while doing a project regarding one of the major problems.

I have profited from comments from both instructors and students who used the first edition. In a sense I owe a debt to every student of mine at Rose–Hulman Institute of Technology. Heartfelt Thank yous go to Sari Freedman and my editor, Susanne Steitz-Filler of John Wiley & Sons. Sangeetha Parthasarathy of LaserWords has been very helpful and patient during the production process. I have been fortunate to rely on the extensive computer skills of my nephew, Scott Carter to whom I owe a big Thank You. But I owe the greatest debt to my wife, Cherry, who has out up with my long hours in the study. I also owe a pat on the head for Ginger who allowed me to refresh while guiding me on long walks through our Old North End neighborhood.

JOHN J. KINNEY

March 4, 2014
Colorado Springs

Chapter 1

Sample Spaces and Probability

1.1 DISCRETE SAMPLE SPACES

Probability theory deals with situations in which there is an element of randomness or chance. Some models of the physical world are *deterministic*, that is, they predict exactly what will happen under certain circumstances. For example, if an object is dropped from a height and given no initial velocity, its distance, s, from the starting point is given by $s = \frac{1}{2} \cdot g \cdot t^2$, where g is the acceleration due to gravity and t is the time. If one tried to apply the formula in a practical situation, one would not find very satisfactory results. The problem is that the formula applies only in a vacuum and ignores the shape of the object and the resistance of the air as well as other factors. Although some of these factors can be determined, we generally combine them and say that the result has a random or chance component. Our model then becomes $s = \frac{1}{2} \cdot g \cdot t^2 + \epsilon$, where ϵ denotes the random component of the model. In contrast with the deterministic model, this model is *stochastic*.

Science often considers stochastic models; in formulating new models, the scientist may try to determine the contributions of both deterministic and random components of the model in predicting accurate results.

The mathematical theory of probability arose in consideration of games of chance, but, as the above-mentioned example shows, it is now widely used in far more practical and applied situations. We encounter other circumstances frequently in everyday life in which we presume that some random factors are at work. Here are some simple examples. What is the chance I will find that all eight traffic lights I pass through on my way to work are green? What are my chances for winning a lottery? I have a ten-volume encyclopedia that I have packed in separate boxes. If the boxes become mixed up and I draw the volumes out at random, what is the chance that my encyclopedia will be in order? My desk lamp has a bulb that is "guaranteed" to last 5000 hours. It has been used for 3000 hours. What is the chance that I must replace it before 2000 more hours are used? Each of these situations involves a random event whose specific outcome is unpredictable in advance.

Probability theory has become important because of the wide variety of practical problems it solves and its role in science. It is also the basis of the statistical analysis of data that is widely used in industry and in experimentation. Consider some examples. A manufacturer of television sets may know that 1% of the television sets manufactured have defects of some kind. What is the chance that a shipment of 200 sets a dealer has received contains 2% defective sets? Solving problems such as these has become important to manufacturers who are anxious to produce high quality products, and indeed such considerations play a central role in what has become known in manufacturing as *statistical process control*.

Probability: An Introduction with Statistical Applications, Second Edition. John J. Kinney.

Sample surveys, in which only a portion of a population or reference set is investigated, have become commonplace. A recent survey, for example, showed that two-thirds of welfare recipients in the United States were not old enough to vote. But surely we do not know that exactly two-thirds of all welfare recipients were not old enough to vote; there is some uncertainty, largely dependent on the size of the sample investigated as well as the manner in which the survey was conducted, connected with this result. How is this uncertainty calculated?

As a final example, consider a scientific investigation into say the relationship between temperature, a catalyst, and pressure in creating a chemical compound. A scientist can only carry out a few experiments in which several combinations of temperatures, amount of catalyst, and level of pressure are investigated. Furthermore, there is an element of randomness (largely due to other, unmeasured, factors) that influence the amount of compound produced. How is the scientist to determine which combination of factors maximizes the amount of chemical compound? We will encounter many of these examples in this book.

In some situations, we could measure all the forces involved and predict the outcome precisely but very often choose not to do so. In the traffic light example, we could, by knowledge of the timing of the lights, my speed, and the traffic pattern, predict precisely the color of each light as I approach it. While this is possible, it is probably not worth the effort, so we combine all the forces involved and call the result "chance." So "chance" as we use it does not imply any new or unknown physical forces; it is simply an umbrella under which we put forces we choose not to measure.

How do we then measure the probability of events such as those described earlier? How do we determine how likely such events are? Such probability problems may be puzzling to us since we lack a framework in which to solve them. We lack a strategy for dealing with the randomness involved in these situations. A sensible way to begin is to consider all the possibilities that could occur. Such a list, or set, is called a *sample space*.

We begin here with some situations that are admittedly much simpler than some of those described earlier; more complex problems will also be encountered in this book.

We will consider situations that we call *experiments*. These are situations that can be repeated under identical circumstances. Those of interest to us will involve some randomness so that the outcomes cannot be precisely predicted in advance. As examples, consider the following:

- Two people are chosen at random from a group of five people.
- Choose one of two brands of breakfast cereal at random.
- Throw two fair dice.
- Take an actuarial examination until it is passed for the first time.
- Any laboratory experiment.

Clearly, the first four of these experiments involve random factors. Laboratory experiments involve random factors as well and we would probably choose not to measure all the factors so as to be able to predict the exact outcome in advance.

Once the conditions for the experiment are set, and we are assured that these conditions can be repeated exactly, we can form the *sample space*, which we define as follows:

Definition A *sample space* is a set of all the possible outcomes from an experiment.

Example 1.1.1

The sample spaces for the first four experiments mentioned above are as follows:

(a) (Choose two people at random from a group of five people.) Denoting the five people as A, B, C, D, and E, we find, if we disregard the order in which the persons are chosen, that there are ten possible samples of two people:

$$S = \{AB, AC, AD, AE, BC, BD, BE, CD, CE, DE\}.$$

This set, S, then comprises the sample space for the experiment.

If we consider the choice of people as random, we might expect that each of these ten samples occurs about 10% of the time. Further, we see that any particular person, say B, occurs in exactly four of the samples, so we say the *probability* that any particular person is in the sample is $\frac{4}{10} = \frac{2}{5}$. The reader may be interested to show that if three people were selected from a group of five people, then the probability a particular person is in the sample is $\frac{3}{5}$. Here, there is a pattern that we can establish with some results to be developed later in this chapter.

(b) (Choose one of two brands of breakfast cereal at random.) Denote the brands as K and P. We take the sample space as

$$S = \{K, P\},$$

where the set S contains each of the *elementary* outcomes, K and P.

(c) (Toss two fair dice.) In contrast with the first two examples, we might consider several different sample spaces. Suppose first that we distinguish the two dice by color, say one is red and the other is green. Then we could write the result of a toss as an ordered pair indicating the outcome on each die, giving say the result on the red die first and the result on the green die second. Let a sample space be

$$S_1 = \{(1,1), (1,2), ..., (1,6), (2,1), (2,2), ..., (2,6), ..., (6,6)\}.$$

It is useful to see this sample space as a geometric space as in Figure 1.1.

Note that the 36 dots represent the only possible outcomes from the experiment. The sample space is not continuous in any sense in this case and may differ from our notions of a geometric space.

We could also describe all the possible outcomes from the experiment by the set

$$S_2 = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

since one of these *sums* must occur when the two dice are thrown.

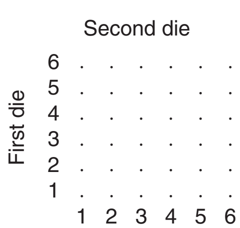


Figure 1.1 Sample space for tossing two dice.

Which sample space should be chosen? Note that each point in S_2 represents at least one point in S_1. So, while we might consider each of the 36 points in S_1 to occur with equal frequency if we threw the dice a large number of times, we would not consider that to be true if we chose sample space S_2. A sum of 7, for example, occurs on 6 of the points in S_1 while a sum of 2 occurs at only one point in S_1. The choice of sample space is largely dependent on what sort of outcomes are of interest when the experiment is performed. It is not uncommon for an experiment to admit more than one sample space. We generally select the sample space most convenient for the analysis of the probabilities involved in the problem.

We continue now with further examples of experiments involving randomness.

(d) (Take an actuarial examination until it is passed for the first time.) Letting P and F denote passing and failing the examination, respectively, we note that the sample space here is infinite:

$$S = \{P, FP, FFP, FFFP, \dots\}.$$

However, S here is a *countably infinite* sample space since its elements can be counted in the sense that they can be placed in a one-to-one correspondence with the set of natural numbers $\{1, 2, 3, 4, \dots\}$ as follows:

$$\begin{aligned} P &\leftrightarrow 1 \\ FP &\leftrightarrow 2 \\ FFP &\leftrightarrow 3 \\ &. \\ &. \\ &. \end{aligned}$$

The rule for the one-to-one correspondence is as follows: given an entry in the left column, the corresponding entry in the right column is the number of the attempt on which the examination is passed; given an entry in the right column, say n, consider $n - 1F$'s followed by P to construct the corresponding entry in the left column. Hence, the correspondence with the set of natural numbers is one-to-one. Such sets are called *countable* or *denumerable*. We will consider countably infinite sets in much the same way that we will consider finite sets. In the next chapter, we will encounter infinite sets that are not countable.

(e) Sample spaces for laboratory experiments are usually difficult to enumerate and may involve a combination of finite and infinite factors.

Example 1.1.2

As a more difficult example, consider observing single births in a hospital until two girls are born in a row.

The sample space now is a bit more challenging to write down than the sample spaces for the situations considered in Example 1.1.1.

For convenience, we write the points, showing the births in order and grouped by the total number of births.

Number of Births	Sample Points	Number of Sample Points
2	*GG*	1
3	*BGG*	1
4	*BBGG*	2
	GBGG	
5	*BBBGG* *BGBGG* *GBBGG*	4
6	*BBBBGG* *BBGBGG* *BGBBGG* *GBBBGG* *GBGBGG*	6

and so on. We note that the number of sample points as we have grouped them follows the sequence 1, 1, 2, 4, 6, … , which we recognize as the beginning of the Fibonacci sequence. The Fibonacci sequence is found by starting with the sequence 1, 1. Subsequent entries are found by adding the two immediately preceding entries. However, we only have evidence that the Fibonacci sequence applies to a few of the groups of points in the sample space. We will have to establish the general pattern in this example before concluding that the Fibonacci sequence does indeed give the number of sample points in the sample space. The reader may wish to do that before reading the following paragraphs!

Here is the reason the Fibonacci sequence occurs: consider a sequence of *B*'s and *G*'s in which *GG* occurs for the first time at the *n*th birth. Let a_n denote the number of ways in which this can occur. If *GG* occurs for the first time on the *n*th birth, there are two possibilities for the beginning of the sequence. These possibilities are mutually exclusive, that is, they cannot occur together.

One possibility is that the sequence begins with a *B* and is followed for the first time by the occurrence of *GG* in $n-1$ births. Since we are requiring the sequence *GG* to occur for the first time at the $n-1$st birth, this can occur in a_{n-1} ways.

The other possibility for the beginning of the sequence is that the sequence begins with *G*, which must then be followed by *B* (else the pattern *GG* will occur in two births) and then the pattern *GG* occurs in $n-2$ births. This can occur in a_{n-2} ways. Since the sequence begins either with *B* or *G*, it follows that

$$a_n = a_{n-1} + a_{n-2},\ n \geq 4,$$
$$\text{where } a_2 = a_3 = 1, \tag{1.1}$$

which describes the Fibonacci sequence.

The sequences for which *GG* occurs for the first time in 7 births can then be found by writing *B* followed by the sequences for 6 births and by writing *GB* followed by *GG* in 5 births:

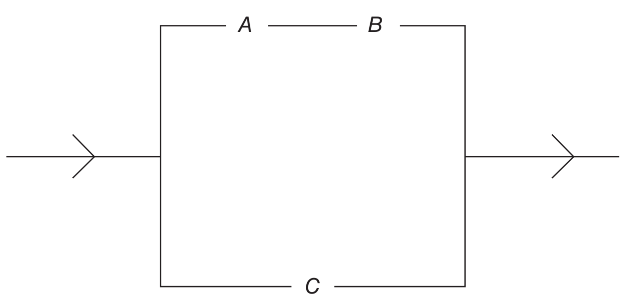
$$B|BBBBGG$$
$$B|BBGBGG$$
$$B|BGBBGG$$
$$B|GBBBGG$$
$$B|GBGBGG$$

$$GB|BBBGG$$
$$GB|BGBGG$$
$$GB|GBBGG$$

Formulas such as ((1.1)) often describe a problem in a very succinct manner; they are called *recursions* because they describe one value of a function, here a_n, in terms of other values of the same function; in addition, they are easily programmed. Computer algebra systems are especially helpful in giving large number of terms determined by recursions. One can find, for example, that there are 46,368 ways for the sequence *GG* to occur for the first time on the 25th birth. It is difficult to imagine determining this number without the use of a computer.

EXERCISES 1.1

1. Show the sample space when 3 people are selected from a group of 5 people. Verify the fact that any particular person in the selected group is 3/5.
2. In Example 1.1.2, show all the sample points where the births of two girls in a row occur in 8 or 9 births.
3. An experiment consists of drawing two numbered balls from a box of balls numbered from 1 to 9. Describe the sample space if
   (a) the first ball is not replaced before the second is drawn.
   (b) the first ball is replaced before the second is drawn.
4. In the diagram below, A, B, and C are switches that may be closed (current flows through the switch) or open (current cannot flow through the switch). Show the sample space indicating all the possible positions of the switches in the circuit.



5. Items being produced on an assembly line can be good (G) or not meeting specifications (N). Show the sample space for the next five items produced by the assembly line.

6. A student decides to take an actuarial examination until it is passed, but will attempt the test at most five times. Show the sample space.

7. In the World Series, games are played until one of the teams has won four games. Show all the points in the sample space in which the American League (A) wins the series over the National League (N) in at most six games.

8. We are interested in the sequence of male and female births in five-child families. Show the sample space.

9. Twelve chips numbered 1 through 12 are mixed in a bowl. Two chips are drawn successively and without replacement. Show the sample space for the experiment.

10. An assembly line is observed until items of both types—good (G) items and items not meeting specification (N)—are observed. Show the sample space.

11. Two numbers are chosen without replacement from the set $\{2, 3, 4, 5, 6, 7\}$, with the additional restriction that the second number chosen must be smaller than the first. Describe an appropriate sample space for the experiment.

12. Computer chips coming off an assembly line are marked defective (D) or nondefective (N). The chips are tested and their condition listed. This is continued until two consecutive defectives are produced or until four chips have been tested, whichever occurs first. Show a sample space for the experiment.

13. A coin is tossed five times and a running count of the heads and tails is kept (so the number of heads and the number of tails tossed so far is recorded at each toss). Show all the sample points where the heads count *always* exceeds the tails count.

14. A sample space consists of all the linear arrangements of the integers 1, 2, 3, 4, and 5. (These linear arrangements are called *permutations*).

(a) Use your computer algebra system to list all the sample points.

(b) If the sample points are equally likely, what is the probability that the number 3 is in the third position?

(c) What is the probability that none of the integers occupies its natural position?

1.2 EVENTS; AXIOMS OF PROBABILITY

After establishing a sample space, we are often interested in particular points, or sets of points, in that sample space. Consider the following examples:

(a) An item is selected at random from a production line. We are interested in the selection of a good item.

(b) Two dice are tossed. We are interested in the occurrence of a sum of 5.

(c) Births are observed until a girl is born. We are interested in this occurring in an even number of births.

Let us begin by defining an *event*.

Definition An *event* is a subset of a sample space.

Events then contain one or more elementary outcomes in the sample space.

In the earlier examples, “a good item is selected,” “the sum is 5,” and “an even number of births was observed” can be described by subsets of the appropriate sample space and are, therefore, *events*.

We say that an event *occurs* if any of the elementary outcomes contained in the event occurs.

We will be interested in *the relative frequency* with which these events occur. In example (a), we would most likely say, if 99% of the items produced in the production line are good, then a good item will be selected about 99% of the time the experiment is performed, but we would expect some variation from this figure. In example (b), such a calculation is more complex since the event “the sum of the spots showing on the dice is 5” comprises several more elementary events. If the sample space distinguishing a red and a green die is

$$S = \{(1,1),(1,2),...,(1,6),(2,1),...,(6,6)\},$$

then the points where the sum is 5 are

$$(1,4),(2,3),(3,2),(4,1).$$

If the dice are fair, then each of the 36 points in S occurs about 1/36 of the time, so we conclude that the sum of the spots showing 5 occurs about $4 \cdot \frac{1}{36} = \frac{1}{9}$ of the time.

In example (c), observing births until a girl is born, the event “an even number of births is observed” is much more complex than examples (a) and (b) since there is an infinity of possibilities. How are we to judge the frequency of occurrence of each one? We cannot answer this question at this time, but we will consider it later.

Now we consider a structure so that we can deal with such questions, as well as many others far more complex than those considered so far. We start with some assumptions about any sample space.

Axioms of Probability

We consider the *long-range relative frequency* or *probability* of an event in a sample space. If we perform an experiment 120 times and an event, A, occurs 30 times, then we say that the *relative frequency* of A is $30/120 = 1/4$. In general, if in n trials an event A occurs $n(A)$times, then we say that the *relative frequency* of A is $\frac{n(A)}{n}$. Of course, if we perform the experiment another n times, we do not expect A to occur exactly the same number of times as before, giving another relative frequency for the event A. We do expect these variable ratios representing relative frequencies to settle down in some manner as n grows large. If A is an event, we denote this limiting relative frequency by the *probability* of A and denote this by $P(A)$.

Definition If A is an event, then the *probability* of A is

$$P(A) = \lim_{n\to\infty} \frac{n(A)}{n}.$$

We assume at this point that the limit exists. We will discuss this in detail in Chapter 4.

In considering events, it is most convenient to use the language and notation of sets where the following notations are common:

The *union* of sets A and B is denoted by $A \cup B$ where

$$A \cup B = \{x \mid x\epsilon A \text{ or } x\epsilon B\},$$

where the word "or" is used in the inclusive sense, that is, an element in both sets A and B is included in the union of the sets.

The *intersection* of sets A and B is denoted by $A \cap B$ where

$$A \cap B = \{x \mid x\epsilon A \text{ and } x\epsilon B\}.$$

We will consider the following as axiomatic or self-evident:

(1) $P(A) \geq 0$, where A is an event,

(2) $P(S) = 1$, where S is the sample space, and

(3) If $A_1, A_2, \ldots$ are *disjoint* or *mutually exclusive*, that is, they have no sample points in common, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Axioms of probability, of course, should reflect our common intuition about the occurrence of events. Since an event cannot occur with a negative relative frequency, (1) is evident. Since *something* must occur when the experiment is done and since S denotes the entire sample space, S must occur with relative frequency 1, hence assumption (2). Now suppose A and B are events with no sample points in common. We can illustrate events in a graphic manner by drawing a rectangle that represents all the points in S; events are subsets of this sample space. A diagram showing the event A, that is, the set of all elements of S that are in the event A, is shown in Figure 1.2. Illustrations of sets and their relationships with each other are called *Venn diagrams*.

The event A or B consists of all points in A or in B and so its relative frequency is the sum of the relative frequencies of A or B. This is assumption (3). Figure 1.3 shows a Venn diagram illustrating the disjoint events A and B.

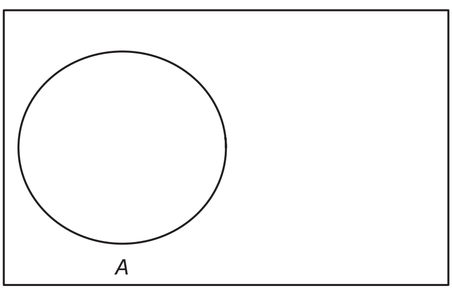


Figure 1.2 Venn diagram showing the event A.

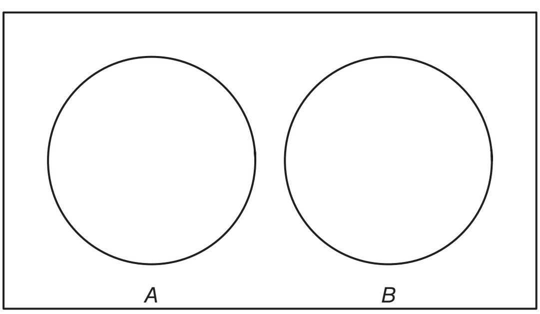


Figure 1.3 Venn diagram showing disjoint events A and B.