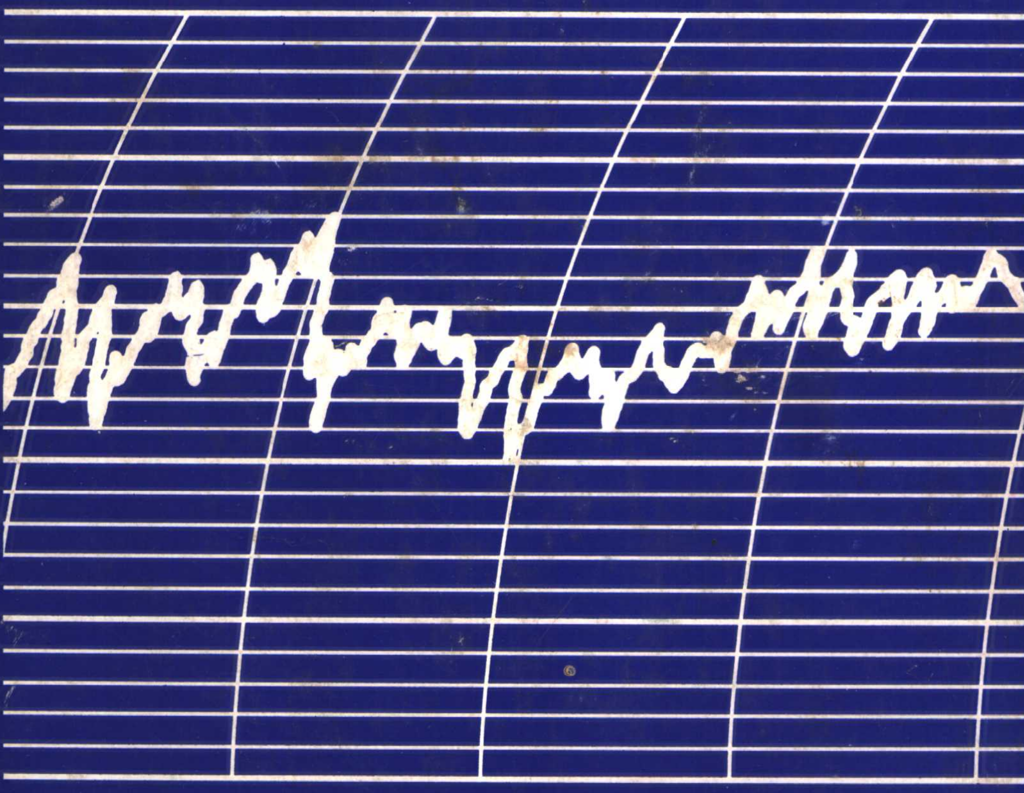
INTRODUCTION TO RANDOM SIGNAL ANALYSIS AND KALMAN FILTERING



Robert Grover Brown

INTRODUCTION TO

RANDOM SIGNAL ANALYSIS AND KALMAN FILTERING

ROBERT GROVER BROWN

Electrical Engineering Department
Iowa State University

JOHN WILEY & SONS

New York • Chichester • Brisbane • Toronto • Singapore

Library of Congress Cataloging in Publication Data:

Brown, Robert Grover.
Introduction to random signal analysis and Kalman filtering.

Includes index.
1. Signal theory (Telecommunication) 2. Random noise theory. 3. Kalman filtering. I. Title.
TK5102.5.B696 1983 621.38'0436 82-19957
ISBN 0-471-08732-7

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Preface

Almost all engineers get involved with instrumentation and signal processing at some time in their career. Along with signal processing comes noise—thus, it is important to understand it and know what to do about it. Noiselike signals are formally described as random processes, and a considerable amount of theory about these processes has been developed over the past 40 years. The primary emphasis here is on linear least-squares filtering. This subject has become especially important since the recursive form of the filter was introduced by R. E. Kalman in 1960. Therefore, much of the book is devoted to a topic that is now known as Kalman filtering. One of my purposes in writing this book was to bring down the level of this subject and make it available to a wider audience. Thus, wherever possible, intuitive arguments are used instead of rigorous proofs, and long derivations are avoided unless they are essential to a full understanding of the limitations of the theory. There is a strong emphasis on examples. They reinforce the theory and demonstrate its applicability to real-life engineering problems.

The study of random processes and linear filtering is no more difficult than many other subjects that are taught at the senior level in engineering. Yet, a hierarchy of prerequisite material is required and, for this reason, the subject usually gets pushed into the beginning graduate level. In particular, a working knowledge of linear system theory is needed. This includes both Laplace and Fourier transform methods and, at least, some acquaintance with state space methods. The kind of treatment given in a typical senior-level course in linear control systems is quite adequate for the level of material presented here. Noise must be described in probabilistic terms, so that at least an elementary knowledge of probability is required before proceeding on to random process theory. Since probability is a subject that engineers often miss as undergraduates, Chapter 1 fills this gap. It is a "no frills" treatment that provides the essentials needed for the remaining chapters.

Random signals and linear filtering is interdisciplinary, and this book can be used by all engineers and applied scientists, not just electrical engineers. I have taught the material to mixed groups, and I found that students who are not electrical engineers fare just as well as electrical engineers. The amount of material that can be covered in one semester, of course, depends on the background of the class. If it is first necessary to bring the group "up to speed" on probability, then it is difficult to squeeze all of the material into one semester. I consider Chapters 8 and 9 to be optional material; they can be omitted if necessary. Also, some instructors may prefer to skip Chapter 4 on Wiener filtering and go directly to Kalman filtering. The book is organized so this may be done with little loss in continuity. In this digital age, Kalman filtering is probably the more important of the two subjects; consequently, if something has to give, it should be Wiener filtering. A minimum pedagogical objective would be to cover Chapters 1–3, 5, 6 and Section 9.1, and do it well. I consider this to be the most important material in the book from an applications viewpoint.

As you can see, the book may be used as a text in many ways, depending on the background and interests of the class. I hope the level and style will also appeal to engineers in industry as a self-study reference. Kalman filtering is an especially important topic with many potential applications, and *it is within reach of any B.S.-level engineer* with the usual background in mathematics and linear systems analysis.

I am grateful to my colleagues and students for their encouragement and helpful suggestions during the preparation of the manuscript. I also thank the office staff of the Department of Electrical Engineering for their help in typing classroom notes. I especially thank Jeanne Gehm for preparing the final manuscript.

R. GROVER BROWN

Contents

CHAPTER 1

Probability and Random Variables

1.1 RANDOM SIGNALS

Nearly everyone has some notion of random or noiselike signals. One has only to tune an ordinary AM radio away from a station, turn up the volume, and the result is static, or noise. If one were to look at a strip-chart recording of such a signal, it would appear to wander on aimlessly with no apparent order in its amplitude pattern as shown in Fig. 1.1. Signals of this type cannot be described with explicit mathematical functions such as sine waves, step functions, and the like. Their description must be put in probabilistic terms. Early investigators recognized that random signals could be described loosely in terms of their spectral content, but a rigorous mathematical description of such signals was not formulated until the 1940s, most notably with the work of Wiener and Rice (1,2).

Noise is usually unwanted. The additive noise in the radio signal disturbs our enjoyment of the music or interferes with our understanding of the spoken word; noise in an electronic navigation system induces position errors that can be disasterous in critical situations; noise in a digital data transmission system can cause bit errors with obvious undesirable consequences; and on and on. Any noise that corrupts the desired signal is bad; it is just a question of how bad! Even after the designer has done his best to eliminate all the obvious noise-producing mechanisms, there always seems to be some noise left over that must be suppressed with more subtle means, such as filtering. To do so effectively, one must understand noise in quantitative terms.

Probability plays a key role in the description of noiselike signals. Our treatment of this subject must necessarily be brief and directed toward the specific needs of subsequent chapters. The scope is thus limited in this regard. We make no apology for this, because many fine books have been written on probability in the broader sense. Our main objective here is the study of random signals and optimal filtering, and we wish to move on to this area as quickly as possible. Fitst, though, we must at least review the bare essentials of probability theory.

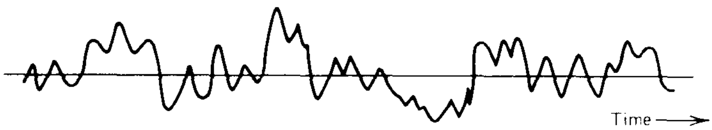


Figure 1.1. Typical noise waveform.

1.2 INTUITIVE NOTION OF PROBABILITY

Most engineering and science students have had some acquaintance with the intuitive concepts of probability. Typically, with the intuitive approach we first consider all possible outcomes of a chance experiment as being equally likely, and then the probability of a particular event, say event A, is defined as

$$P(A) = \frac{\text{Possible outcomes favoring event } A}{\text{Total possible outcomes}} \tag{1.2.1}$$

where we read $P(A)$ as "probability of event A." This concept is then expanded to include the relative-frequency-of-occurrence or statistical viewpoint of probability. With the relative-frequency concept, we imagine a large number of trials of some chance experiment and then define probability as the relative frequency of occurrence of the event in question. Considerations such as what is meant by "large" and the existence of limits are normally avoided in elementary treatments. This is for good reason. The idea of limit in a probabilistic sense is subtle.

Although the older intuitive notions of probability have limitations, they still play an important role in probability theory. The ratio-of-possible-events concept is a useful problem-solving tool in many instances. The relative-frequency concept is especially helpful in visualizing the statistical significance of the results of probability calculations. That is, it provides the necessary tie between the theory and the physical situation. Two examples that illustrate the usefulness of these intuitive notions of probability should now prove useful.

Example 1.1 In straight poker, each player is dealt 5 cards face down from a deck of 52 playing cards. We pose two questions:

(a) What is the probability of being dealt four of a kind, that is, four aces, four kings, and so forth?

(b) What is the probability of being dealt a straight flush, that is, a continuous sequence of five cards in any suit? ■

Solution to Question (a) This problem is relatively complicated if you think in terms of the sequence of chance events that can take place when the cards are dealt one at a time. Yet the problem is relatively easy when viewed in terms of the ratio of favorable to total number of outcomes. These are easily counted in this case. There are only 48 possible hands containing 4 aces; another 48 containing 4 kings; etc. Thus, there are $13 \cdot 48$ possible four-of-a-kind hands. The total number of possible poker hands of any kind is obtained from the combination formula for "52 things taken 5 at a time" (3). This is given by the binomial coefficient

$$\binom{52}{5} = \frac{52!}{5!(52-5)!} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 2{,}598{,}960 \qquad (1.2.2)$$

Therefore, the probability of being dealt four of a kind is

$$P(\text{Four of a kind}) = \frac{13 \cdot 48}{2{,}598{,}960} = \frac{624}{2{,}598{,}960} \approx .00024 \qquad (1.2.3)$$

Solution to Question (b) Again, the direct itemization of favorable events is the simplest approach. The possible sequences in each of four suits are: AKQJ10, KQJ109, . . . , 5432A. (*Note*. we allow the ace to be counted either high or low.) Thus, there are 10 possible straight flushes in each suit (including the royal flush of the suit) giving a total of 40 possible straight flushes. The probability of a straight flush is, then,

$$P(\text{Straight flush}) = \frac{40}{2{,}598{,}960} \approx .000015 \qquad (1.2.4)$$

We note in passing that in poker a straight flush wins over four of a kind; and, rightly so, since it is the rarer of the two hands.

Example 1.2. Craps is a popular gambling game played in casinos throughout the world (11). The player rolls two dice and plays against the house (i.e., the casino). If the first roll is 7 or 11, the player wins immediately; if it is 2, 3 or 12, the player loses immediately. If the first roll results in 4, 5, 6, 8, 9, or 10, the player continues to roll until either the same number appears, which constitutes a win, or a 7 appears, which results in the player losing. What is the player's probability of winning when throwing the dice?

This example was chosen to illustrate the shortcoming of the direct count-the-outcomes approach. In this case, one cannot enumerate all the possible outcomes. For example, if the player's first roll is a 4, the play continues until another outcome is reached. Presumably, the rolling could continue on ad infinitum without a 4 or 7 appearing, which is what is required to terminate the game. Thus, the direct enumeration approach fails in this situation. On the other hand, the relative-frequency-of-occurrence approach works quite well. Table 1.1 shows the relative frequency of occurrence of the various numbers on the first roll. The numbers in the column labeled "probability" were obtained by enumerating the 36 possible outcomes and allotting $\frac{1}{36}$ for each outcome that

Table 1.1 Probabilities in Craps

Number of First Throw	Probability	Result of First Throw	Subsequent Probabilities and Results	Relative Frequency of Winning with Various First Throws
2	$\frac{1}{36}$	Lose		0
3	$\frac{2}{36}$	Lose		0
4	$\frac{3}{36}$	Continue	$P(4 \text{ before } 7) = \frac{1}{3}$ (win) $P(7 \text{ before } 4) = \frac{2}{3}$ (lose)	$\frac{3}{36} \cdot \frac{1}{3}$
5	$\frac{4}{36}$	Continue	$P(5 \text{ before } 7) = \frac{2}{5}$ (win) $P(7 \text{ before } 5) = \frac{3}{5}$ (lose)	$\frac{4}{36} \cdot \frac{2}{5}$
6	$\frac{5}{36}$	Continue	$P(6 \text{ before } 7) = \frac{5}{11}$ (win) $P(7 \text{ before } 6) = \frac{6}{11}$ (lose)	$\frac{5}{36} \cdot \frac{5}{11}$
7	$\frac{6}{36}$	Win		$\frac{6}{36}$
8	$\frac{5}{36}$	Continue	$P(8 \text{ before } 7) = \frac{5}{11}$ (win) $P(7 \text{ before } 8) = \frac{6}{11}$ (lose)	$\frac{5}{36} \cdot \frac{5}{11}$
9	$\frac{4}{36}$	Continue	$P(9 \text{ before } 7) = \frac{2}{5}$ (win) $P(7 \text{ before } 9) = \frac{3}{5}$ (lose)	$\frac{4}{36} \cdot \frac{2}{5}$
10	$\frac{3}{36}$	Continue	$P(10 \text{ before } 7) = \frac{1}{3}$ (win) $P(7 \text{ before } 10) = \frac{2}{3}$ (lose)	$\frac{3}{36} \cdot \frac{1}{3}$
11	$\frac{2}{36}$	Win		$\frac{2}{36}$
12	$\frac{1}{36}$	Lose		0

Total probability of winning $= \frac{244}{495} \approx .4929$

yields a sum corresponding to the number in the first column. For example, a 4 may be obtained with the combinations (1,3), (2,2), or (1,3). For the cases where the game continues after the first throw, the subsequent probabilities were obtained simply by observing the *relative* frequency of occurrence of the numbers involved. For example, a 7 is twice as likely as a 4. Thus, the relative frequency of rolling a 7 before a 4 should be twice that of "4 before 7," and the respective probabilities are $\frac{2}{3}$ and $\frac{1}{3}$. The total probability of winning with a 4 on the first throw was reasoned as follows. A 4 only appears on the first roll $\frac{3}{36}$ of the time; and, of this fraction, only $\frac{1}{3}$ of the time will this result in an ultimate win. Thus, the relative frequency of winning via this route is the product of $\frac{3}{36} \cdot \frac{1}{3}$. Admittedly, this line of reasoning is quite intuitive, but that is the very nature of the relative-frequency-of-occurrence approach to probability.

For the benefit of those who like to gamble, it should be noted that craps is a very close game. The edge in favor of the house is only about $1\frac{1}{2}$ percent. (Also see Problem 1.7.) ■

1.3 AXIOMATIC PROBABILITY

It should be apparent that the intuitive concepts of probability have their limitations. The ratio-of-outcomes approach requires the equal-likelihood assumption for all outcomes. This may fit many situations, but often we wish to consider "unfair" chance situations as well as "fair" ones. Also, as demonstrated in Example 1.2, there are many problems for which all possible outcomes simply cannot be enumerated. The relative-frequency approach is intuitive by its very nature. Intuition should never be ignored; but, on the other hand, it can lead one astray in complex situations. For these reasons, the axiomatic formulation of probability theory is now almost universally favored among both applied and theoretical scholars in this area. As we would expect, axiomatic probability is compatible with the older, more heuristic probability theory.

Axiomatic probability begins with the concept of a *sample space*. We first imagine a conceptual chance experiment. The sample space is the set of all possible *outcomes* of this experiment. The individual outcomes are called *elements* or *points* in the sample space. We denote the sample space as S and its set of elements as $\{s_1, s_2, s_3, \ldots\}$. The number of points in the sample space may be finite, countably infinite, or simply infinite, depending on the experiment under consideration. A few examples of sample spaces should be helpful at this point.

Example 1.3 *The experiment:* Make a single draw from a deck of 52 playing cards. Since there are 52 possible outcomes, the sample space contains 52 discrete points. If we wished, we could enumerate them as Ace of Clubs, King of Clubs, Queen of Clubs, and so forth. Note that the points of the sample space in this case are "things," not numbers. ■

Example 1.4 *The experiment:* Two fair dice are thrown and the number of dots on the top of each is observed. There are 36 discrete outcomes that can be enumerated as (1,1), (1,2), (1,3), . . . , (6,5), (6,6). The first number in parentheses identifies the number of dots on die 1 and the second is the number on die 2. Thus, 36 distinct 2-tuples describe the possible outcomes, and our sample space contains 36 points or elements. Note that the points in this sample space retain the identity of each individual die and the number of dots shown on its top face. ■

Example 1.5 *The experiment:* Two fair dice are thrown and the sum of the number of dots is observed. In this experiment, we do not wish to retain the identity of the numbers on each die; only the sum is of interest. Therefore, it would be perfectly proper to say the possible outcomes of the experiment are {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}. Thus, the sample space would contain 11 discrete elements. From this and the preceding example, it can be seen that we have some discretion in how we define the sample space corresponding to a certain experiment. It depends to some extent on what we wish to observe. If certain details of the experiment are not of interest, they often may be suppressed with some resultant simplification. However, once we agree on what items are to be grouped together and called *outcomes*, the sample space must include all the defined outcomes; and, similarly, the result of an experiment must always yield one of the defined outcomes. ■

Example 1.6 *The experiment:* A dart is thrown at a target and the location of the hit is observed. In this experiment we imagine the random mechanisms affecting the throw are such that we get a continuous spread of data centered around the bull's-eye when the experiment is repeated over and over. In this case, even if we bound the hit locations within a certain region determined by reasonableness, we still cannot enumerate all possible hit locations. Thus, we have an infinite number of points in our sample space in this example. Even though we cannot enumerate the points one by one, they are, of course, identifiable in terms of either rectangular or polar coordinates. ■

It should be noted that elements of a sample space must always be *mutually exclusive* or *disjoint*. On a given trial, the occurrence of one excludes the occurrence of another. There is no overlap of points in a sample space.

In axiomatic probability, the term event has special meaning and should not be used interchangably with outcome. An *event* is a special subset of the sample space S. We usually wish to consider various events defined on a sample space, and they will be denoted with uppercase letters such as A, B, C, . . . , or perhaps A_1, A_2, . . . , etc. Also, we will have occasion to consider the set of operations of union, intersection, and complement of our defined events. Thus, we must be careful in our definition of events to make the set sufficiently complete such that these set operations also yield properly defined events. In discrete problems, this can always be done by defining the set of events under consideration to be all possible subsets of the sample space S. We will tacitly assume that the null set is a subset of every set, and that every set is a subset of itself.

One other comment about events is in order before proceeding to the basic axioms of probability. The event A is said to occur if *any* point in A occurs.

The three axioms of probability may now be stated. Let S be the sample space and A be any event defined on the sample space. The first two axioms are

$$\text{Axiom 1:} \qquad P(A) \geq 0 \tag{1.3.1}$$

$$\text{Axiom 2:} \qquad P(S) = 1 \tag{1.3.2}$$

Now, let $A_1, A_2, A_3, \ldots$ be mutually exclusive (disjoint) events defined on S. The sequence may be finite or countably infinite. The third axiom is then

$$\text{Axiom 3:} \qquad \begin{aligned} &P(A_1 \cup A_2 \cup A_3 \cup \ldots) \\ &= P(A_1) + P(A_2) + P(A_3) + \cdots \end{aligned} \tag{1.3.3}$$

Axiom 1 simply says that the probability of an event cannot be negative. This certainly conforms to the relative-frequency-of-occurrence concept of probability. Axiom 2 says that the event S, which includes all possible outcomes, must have a probability of unity. It is sometimes called the certain event. The first two axioms are obviously necessary if axiomatic probability is to be compatible with the older relative-frequency probability theory. The third axiom is not quite so obvious, perhaps, and it simply must be assumed. In words, it says that when we have nonoverlapping (disjoint) events, the probability of the union of these events is the sum of the probabilities of the individual events. If this were not so, one could easily think of counterexamples that would not be compatible with the relative-frequency concept. This would be most undesirable.

We now recapitulate. There are three essential ingredients in the formal approach to probability. First, a sample space must be defined that includes all possible outcomes of our conceptual experiment. We have some discretion in what we call outcomes, but caution is in order here. The outcomes must be disjoint and all-inclusive such that $P(S) = 1$. Second, we must carefully define a set of events on the sample space, and the set must be closed such that the operations of union, intersection, and complement also yield events in the set. Finally, we must assign probabilities to all events in accordance with the basic axioms of probability. In physical problems, this assignment is chosen to be compatible with what we feel to be reasonable in terms of relative frequency of occurrence of the events. If the sample space S contains a finite number of elements, the probability assignment is usually made directly on the elements of S. They are, of course, elementary events themselves. This, along with Axiom 3, then indirectly assigns a probability to all other events defined on the sample space. However, if the sample space consists of an infinite "smear" of points, the probability assignment must be made on events and not on points in the sample space. This will be illustrated later in Example 1.8.

Once we have specified the sample space, the set of events, and the probabilities associated with the events, we have what is known as a *probability space*. This provides the theoretical structure for the formal solution of a wide variety of probability problems.

Example 1.7 Consider a single throw of two dice, and let us say we are only interested in the sum of the dots that appear on the top faces. This chance situation fits many games that are played with dice. In this case, we will define our sample space to be

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

and it is seen to contain 11 discrete points. Next, we define the set of possible events to be all subsets of S, including the null set and S itself. Note that the elements of S are elementary events, and they are disjoint, as they should be. Also, $P(S) = 1$. Finally, we need to assign probabilities to the events. This could be done arbitrarily (within the constraints imposed by the axioms of probability), but in this case we want the results of our formal analysis to coincide with the relative-frequency approach. Therefore, we will assign probabilities to the elements in accordance with Table 1.2, which, in turn, indirectly specifies probabilities for all other events defined on S. We now have a properly defined probability space, and we can pose a variety of questions relative to the single throw of two dice.

Suppose we ask: What is the probability of throwing either a 7 or an 11? From Axiom 3, and noting that "7 or 11" is the equivalent of saying "$7 \cup 11$," we have

$$P(7 \text{ or } 11) = \frac{6}{36} + \frac{2}{36} = \frac{2}{9} \tag{1.3.4}$$

Table 1.2 Probabilities for Two-Dice Example

Sum of Two Dice	Assigned Probability
2	$\frac{1}{36}$
3	$\frac{2}{36}$
4	$\frac{3}{36}$
5	$\frac{4}{36}$
6	$\frac{5}{36}$
7	$\frac{6}{36}$
8	$\frac{5}{36}$
9	$\frac{4}{36}$
10	$\frac{3}{36}$
11	$\frac{2}{36}$
12	$\frac{1}{36}$

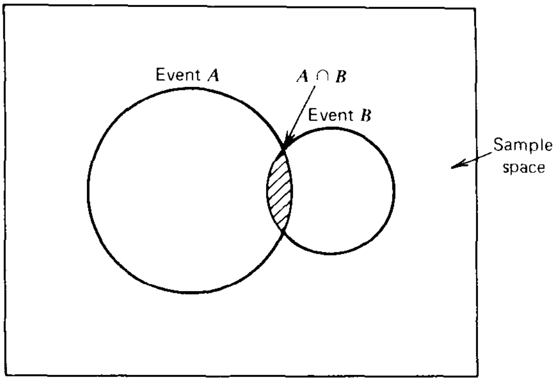


Figure 1.2. Venn diagram for two events A and B.

Next, suppose we ask: What is the probability of not throwing 2, 3, or 12? This calls for the complement of event "2 or 3 or 12" which is the set {4,5,6,7,8,9,10,11}. Recall that we say the event occurs if *any* element in the set occurs. Therefore, again using Axiom 3, we have

$$P(\text{Not throwing 2, 3, or 12}) = \frac{3+4+5+6+5+4+3+2}{36} = \frac{8}{9} \tag{1.3.5}$$

Suppose we now pose the further question: What is the probability that two "4s" are thrown? In our definition of the sample space, we suppressed the identity of the individual dice, so this simply is not a proper question for the probability space, as defined. This example will be continued, but first we digress for a moment to consider intersection of events. ■

In addition to the set operations of union and complementation, the operation of intersection is also useful in probability theory. The intersection of two events A and B is the event containing points that are common to both A and B. This is illustrated in Fig. 1.2 with what is sometimes called a Venn diagram. The points lying within the heavy contour comprise the union of A and B, denoted as $A \cup B$ or "A or B." The points within the shaded region are the event "A intersection B," which is denoted $A \cap B$, or sometimes just "A and B."* The following relationship should be apparent just from the geometry of

* In many references, the notation for "A union B" is "$A + B$," and "A intersection B" is shortened to just "AB." We will be proceeding to the study of random variables shortly, and then the chance occurrences will be related to real numbers, not "things." Thus, in order to avoid confusion, we will stay with the more cumbersome notation of $\cup$ and $\cap$ for set operations, and reserve $X + Y$ and XY to mean the usual arithmetic operations on real variables.