CHAPTER 4. CONDITIONAL PROBABILITY.

Let us look again at the experiment of rolling a single symmetrical and homogeneous die. We have seen that a good probability model for this experiment is a probability space which has six points (which we call 1,2,3,4,5,6) and a probability value of $\frac{1}{6}$ for each point. Now take A to be the event that an odd number occurs (the event $\{1,3,5\}$), and take B to be the event that a number greater than 3 occurs (the event $\{4,5,6\}$). We wish to ask the following question. If we repeat the experiment many times and then look only at those trials where the event B occurs, in what proportion of those trials does the event A also occur? The answer to this question is obvious. For B to occur, we must get a 4, 5, or 6, and we expect each equally often. Of these, only 5 is also in the event A, hence the answer to our question must be $\frac{1}{3}$. We can ask the same question for any two events, A and B, in any probability space. To get the answer, we form the quotient (or proportion) between the expected relative frequency of A and B both occurring and the expected relative frequency of B occurring. That is to say, we form the quotient:

$$\frac{P(A \cap B)}{P(B)} .$$

(For the die example, this quotient would be $\frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$.) We give this quotient a special name. We call it the conditional probability of the event A, given that the event B occurs, or, more briefly, the conditional probability of A given B. We use the notation $P(A \mid B)$ for this quantity. We thus have as a definition:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} .$$

(This definition is only used when $P(B) \neq 0$.)

For example, with a die, what is the conditional probability of having an odd number, given that we have a number greater than 1 ? We get the answer as follows:

$$P(\underline{\text{odd}} \mid > 1) = \frac{P(\underline{\text{odd and}} > 1)}{P(> 1)} = \frac{\frac{2}{6}}{\frac{5}{6}} = \frac{2}{5} .$$

We can also, if we wish, ask for the conditional probability of having a number greater than 1 given that we have an odd number. We get:

$$P(> 1 \mid \underline{\text{odd}}) = \frac{P(\underline{\text{odd and}} > 1)}{P(\underline{\text{odd}})} = \frac{\frac{2}{6}}{\frac{1}{2}} = \frac{2}{3} .$$

Each of these conditional probabilities has a definite and clear meaning. The first, $P(\underline{\text{odd}} \mid > 1) = \frac{2}{5}$, says that if we look only at outcomes > 1, then, from among those, we expect an odd number to occur two-fifths of the time. The second, $P(> 1 \mid \underline{\text{odd}}) = \frac{2}{3}$, says that if we look only at odd outcomes, then, from among those, we expect a number greater than 1 to occur two-thirds of the time.

Conditional probabilities are important because, in many probability problems, we begin with information about conditional probabilities. We then use these conditional probabilities to calculate other probability values that we are interested in.

The equation

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

can also be written

$$P(A \cap B) = P(B)P(A \mid B) .$$

This last equation is sometimes called the multiplication law of probability. If we know the values of $P(B)$ and $P(A \mid B)$, we can use this law to find a value for $P(A \cap B)$.

Here is an example. We have two jars, Jar I and Jar II. Jar I has 1 red ball and 2 white balls in it. Jar II has 2 red balls and 1 white ball in it. We carry out the following experiment. We mix up the balls in Jar I, close our eyes, and take a ball from Jar I. We then transfer this ball to Jar II, mix up the four balls now in Jar II, close our eyes, and draw a ball from Jar II. We form a sample space for this experiment by taking, as our four basic outcomes:

transfer red and draw red,

transfer red and draw white,

transfer white and draw red,

transfer white and draw white.

What probability function should we use? We note that certain probabilities and certain conditional probabilities are immediately given. In particular,

$$P(\underline{\text{transfer red}}) = \frac{1}{3}$$

because we are taking from a jar with 2 white and 1 red; and

$$P(\underline{\text{draw red}} \mid \underline{\text{transfer red}}) = \frac{3}{4}$$

because, after transferring red, we will be drawing from a jar with 1 white and 3 red. Similarly we see that

$$P(\underline{\text{transfer white}}) = \frac{2}{3},$$

$$P(\underline{\text{draw white}} \mid \underline{\text{transfer red}}) = \frac{1}{4},$$

$$P(\underline{\text{draw red}} \mid \underline{\text{transfer white}}) = \frac{1}{2},$$

$$P(\underline{\text{draw white}} \mid \underline{\text{transfer white}}) = \frac{1}{2}.$$

We can now use the multiplication law to get our basic probability values. We get:

$$P(\underline{\text{transfer red and draw red}}) = P(\underline{\text{transfer red}})P(\underline{\text{draw red}} \mid \underline{\text{transfer red}})$$

$$= \frac{1}{3} \cdot \frac{3}{4} = \frac{1}{4},$$

$$P(\underline{\text{transfer red and draw white}}) = \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{12},$$

$$P(\underline{\text{transfer white and draw red}}) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{3},$$

$$P(\underline{\text{transfer white and draw white}}) = \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3}.$$

It is sometimes helpful in problems like this to make a table:

		draw: R	draw: W
transfer	R	$\frac{1}{4}$	$\frac{1}{12}$
	W	$\frac{1}{3}$	$\frac{1}{3}$

This table shows the sample space and the probability function. Each cell in the table is a point in the sample space, and the number appearing in that cell is the corresponding value of the probability function.

Once we have the probability function, we can go on and calculate any other probabilities we wish. For example, we can get:

$$P(\underline{\text{draw red}}) = \frac{1}{4} + \frac{1}{3} = \frac{7}{12} .$$

Or, if we want the conditional probability of <u>transfer red</u> given <u>draw red</u>, we can get:

$$P(\underline{\text{transfer red}} \mid \underline{\text{draw red}}) = \frac{\frac{1}{4}}{\frac{7}{12}} = \frac{3}{7} .$$

(This tells us that if we look at all trials where red is drawn, we expect to find that red has also been transferred in about three-sevenths of those trials.)

If one understands the idea of conditional probability, one is well on one's way to a good understanding of much of probability theory. Here are two more examples to help strengthen the reader's understanding.

<u>Example 1</u>. A special deck of eight cards is formed by taking the four <u>queens</u> and four <u>kings</u> from a regular deck. This special deck is shuffled and two cards are drawn from it. Find the following conditional probabilities:

$$P(\underline{\text{two queens drawn}} \mid \underline{\text{at least one queen drawn}}),$$

and $$P(\underline{\text{two queens drawn}} \mid \underline{\text{at least one black queen drawn}}).$$

Our basic sample space has $\binom{8}{2} = \frac{8 \cdot 7}{2} = 28$ points. (If we prefer, we can think of the two cards as being drawn one after the other rather than at the same time. This would give us a sample space of $7 \cdot 8 = 56$ points. Thus, as we have noted in connection with poker hands in Chapter 3, we have two distinct choices of sample space available in forming our model. Not surprisingly, either choice leads to the same numerical answers to our original problem.) As probability function, we use the equiprobable function, which gives each point the probability value $\frac{1}{28}$.

Now, by counting basic outcomes, we get

$$P(\underline{\text{two queens}}) = \frac{6}{28} = \frac{3}{14}$$

and $$P(\underline{\text{at least one queen}}) = \frac{22}{28} = \frac{11}{14}.$$

Therefore

$$P(\underline{\text{two queens}} \mid \underline{\text{at least one queen}}) = \frac{\frac{3}{14}}{\frac{11}{14}} = \frac{3}{11} = 0.27.$$

(Note that the event two queens and at least one queen is the same as the event two queens.)

Similarly, again counting basic outcomes, we get

$$P(\text{two queens and at least one black queen}) = \frac{5}{28},$$

and $$P(\text{at least one black queen}) = \frac{13}{28}.$$

Hence we get

$$P(\text{two queens} \mid \text{at least one black queen}) = \frac{\frac{5}{28}}{\frac{13}{28}} = \frac{5}{13} = 0.38.$$

(Students are often surprised that the two conditional probabilities calculated in this problem are different, but one can see from the definition of conditional probability that the two conditional probabilities describe quite different situations. For the first, we look at all cases where at least one queen occurs and see in what proportion of these we have two queens. For the second, we look at all cases where at least one black queen occurs, and see in what proportion of these we have two queens. It is evident that P(two queens | at least one red queen) is also equal to 0.37. Consider the following intuitive argument: "Because P(two queens | at least one black queen) and P(two queens | at least one red queen) are both equal to 0.37, and because every queen is either black or red, we must have P(two queens | at least one queen) also equal to 0.37." Our analysis and calculation show that this argument is false.)

Example 2. In a certain country, every day is either sunny or rainy. The probability that a day is sunny is $\frac{2}{3}$, and the conditional probability that a day is sunny given that the previous day was sunny is $\frac{3}{4}$. What is the conditional probability that a day is rainy, given that the previous day was rainy?

To apply probability theory, we must (as always) fix an experiment and form a probability space. We fix as our experiment: observe the weather on two successive days. Our sample space then has four points:

sunny then sunny,

sunny then rainy,

rainy then sunny,

rainy then rainy,

We can get the probability function for this sample space as follows. We write the sample space as a table:

second day

first day	S	R
S	x	y
R	z	w

where x,y,z, and w are the unknown probability values. From the statement of the problem we know that

$$P(\underline{\text{first day sunny}}) = \frac{2}{3} = x + y ,$$

$$P(\underline{\text{second day sunny}}) = \frac{2}{3} = x + z ,$$

$$P(\underline{\text{second day sunny}} \mid \underline{\text{first day sunny}}) = \frac{3}{4} = \frac{x}{x+y} .$$ (See

remark below.) Eliminating $x+y$ between the first and third equations, we get $\frac{x}{\frac{2}{3}} = \frac{3}{4}$, and $x = \frac{2}{3} \cdot \frac{3}{4} = \frac{1}{2}$. It follows that $y = \frac{1}{6}$ and $z = \frac{1}{6}$. We then have $w = \frac{1}{6}$, because $x + y + z + w$ must $= 1$.

We now have our probability function and can calculate the answer desired:

$$P(\underline{\text{second day rainy}} \mid \underline{\text{first day rainy}}) = \frac{w}{z+w} = \frac{\frac{1}{6}}{\frac{1}{3}} = \frac{1}{2} .$$

<u>Remark</u>. How does the information that "the probability that a day is sunny is 2/3" permit us to conclude that P(<u>first day sunny</u>) = 2/3 and P(<u>second day sunny</u>) = 2/3 in our experiment? The answer to this question is not obvious and, in effect, makes use of additional information about how the experiment is carried out. If we say that "the probability that a day is sunny is 2/3", we are referring to a different experiment, the experiment of observing a single day. Moreover, because we know that the result on a single day may be related in some way to the result on immediately preceding days, it follows that if we wish to repeat this experiment of observing a single day and to do so under "the same general experimental

conditions", we must do so on rather widely spaced days chosen by some random procedure. (Such procedures are described in Chapter 8.) We would then expect the relative frequency of sunny days to be about 2/3. In the same way, repetitions of the experiment of observing on two successive days would have to be carried out on rather widely spaced and randomly chosen occasions to assure that results of one trial (of the two-successive-day observation) did not influence results of the next. But this would mean that the first days of those repeated two-successive-day trials would themselves be widely spaced and randomly chosen. Hence we would expect P(first day sunny) also to be 2/3. Similarly for P(second day sunny).

Remark. Do we have enough information to calculate the conditional probability that a day is sunny given that the two preceding days are sunny? For the conditional probability P(third day sunny | first two days sunny) we must take, as our experiment, observe three successive days. This gives us a sample space with 8 points. We are not given enough information to find the probability function on this space, or to find the conditional probability asked for. To see this, imagine, for example, a country where there is a constantly repeating six-day cycle of weather as follows

...SSSSRRSSSSRR... .

One can show that this country fits all the information given in the problem and that P(<u>third day sunny</u> | <u>first two days sunny</u>) = $\frac{2}{3}$. Next, imagine a country where there is a constantly repeating twelve-day cycle of weather as follows

...SSSSSSSRRSRRSSSSSSSRRSRR... .

One can again show that this country fits all the information given in the problem, but that P(<u>third day sunny</u>|<u>first two days sunny</u>) = $\frac{5}{6}$.

(<u>**Remark**</u>. In the above situation of observing weather from day to day, we can also think of a single comprehensive experiment in which we observe the weather from day to day for an indefinite period into the future. A basic outcome for this larger experiment can then be taken to be an indefinitely long sequence of specific daily results. Such an experiment is called a <u>time-series</u>. For a mathematical model, we would take a sample space in which each point is a different possible <u>infinite</u> sequence of S's and R's. Such a sample space would be similar, mathematically, to the sample space for Brownian motion mentioned in Chapter 2. To define a probability measure on this sample space requires more advanced mathematical concepts than we use here. As indicated in connection with the three-day experiment, a variety of probability measures are possible satisfying the two-day conditions initially stated for Example 2.

Independent events.

Sometimes, when we are given two events A and B in a probability space, we find that $P(A)$ and $P(A|B)$ have the same value. If this happens, we say that the event A is independent of the event B. We use the word "independent" because the probability of A is unaffected by information about the occurrence of B on the same trial. For example, in rolling a die, $P(> 4)$ and $P(> 4|\text{odd})$ both have the value $\frac{1}{3}$. Hence we say that the event greater than four is independent of the event odd. For independent events the multiplication law takes the simple form

$$P(A \cap B) = P(B)P(A) = P(A)P(B)$$ (the multiplication law for two independent events.)

If $P(A) = P(A|B)$, then it must also be the case that $P(B) = P(B|A)$, because, by the general multiplication law, $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$. Thus if A is independent of B, it follows that B is also independent of A. We hence say, more briefly, that A and B are independent events. It is also immediate that if A and B are independent, so are A and $\overline{B}$, $\overline{A}$ and B, and $\overline{A}$ and $\overline{B}$, since it follows directly from $P(A) = P(A|B)$ that $P(A) = P(A|\overline{B})$, $P(\overline{A}) = P(\overline{A}|B)$, and $P(\overline{A}) = P(\overline{A}|\overline{B})$, as is easily shown. For example, for any events A and B,

$$P(A) = P(A \cap B) + P(A \cap \overline{B}) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B}).$$

If we assume $P(A) = P(A|B)$, we have

$$P(A) = P(A)P(B) + P(A|\overline{B})P(\overline{B}),$$

$$P(A)(1-P(B)) = P(A|\overline{B})P(\overline{B}),$$

$$P(A)P(\overline{B}) = P(A|\overline{B})P(\overline{B}),$$

$$P(A) = P(A|\overline{B}).$$

(The last step fails only if $P(\overline{B}) = 0$, but in that case $P(A|\overline{B})$ is not defined.) It immediately follows that corresponding forms of the multiplication law also hold if A and B are independent: $P(A \cap \overline{B}) = P(A)P(\overline{B})$; $P(\overline{A} \cap B) = P(\overline{A})P(B)$; and $P(\overline{A} \cap \overline{B}) = P(\overline{A})P(\overline{B})$.

events. We then say that a set of n events is independent if each of the pairs formed by taking two of the given events, or by taking one of the given events and any intersection of two or more of the remaining events, is a pair of independent events. A corresponding simple form of multiplication law for n independent events holds: if $A_1, \ldots, A_n$ are independent, then

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2) \ldots P(A_n);$$

as well as $P(A_1 \cap A_2 \cap \ldots \cap \bar{A}_n) = P(A_1)P(A_2) \ldots P(\bar{A}_n);$
$P(A_1 \cap A_2 \cap \ldots \cap \bar{A}_{n-1} \cap A_n) = P(A_1)P(A_2) \ldots P(\bar{A}_{n-1})P(A_n); \ldots .$

Returning to the case of rolling a die, we now see that if we roll a die 5 times and use the multiplication law for 5 independent events, we get, for example, the following. The probability that 6 does not appear on any roll must be $\frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} = (\frac{5}{6})^5$, and the probability that 6 appears on the first and fourth rolls but not on the other rolls must be $\frac{1}{6} \cdot \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} \cdot \frac{5}{6} = (\frac{1}{6})^2 (\frac{5}{6})^3$.

The above definitions of independence are formal and precise. It is usually enough, however, in applying probability theory, to say that a set of events is independent if, for each event, the probability of that event is unaffected by information about the occurrence or non-occurrence of the other events. As in the case of rolling a die 5 times, it is often intuitively obvious, from the physical situation, that independence must hold. When we can assume independence in this way, we can then use the simple form of the multiplication law as above. Thus the formal definition of independence helps us to define, and make precise, appropriate probability spaces for what were called in Chapter 1 "repeated independent trials of an experiment".

<u>Remark</u>. In Example 2 above, our basic experiment was to observe two successive days of weather. How could we, practically, do a succession of independent trials of this two-day experiment? As each day's weather is influenced by the weather on previous days, we have already noted that we would have to choose the time of the next trial in such a way as to be reasonably sure that the influence of previous trial outcomes was negligibly small. In particular, we would have to allow enough time between trials. If (as in the case of some exactly repeating cycle of weather) this did not suffice, we would have to use some random physical procedure to choose the successive days of observation. (We consider this further in Chapter 8.)

<u>Remark</u>. The discussion above shows why, in the birth-month examples in Chapter 2, we used the equiprobable probability function. If we stop three people on the street and ask their birth-months, we expect the answer of each to be independent of the others. Hence we can use the simple multiplication rule to calculate the probability of each point in the sample space. We get, for each point, $\frac{1}{12} \cdot \frac{1}{12} \cdot \frac{1}{12} = \frac{1}{1728}$.

Bayes' Theorem. If we look back at the example of the jars above, we see that we obtained the unknown conditional probability P(transfer red|draw red) by a method that can be described as follows: (1) we make a table to represent points of the sample space; (2) to get values for the table, we apply the multiplication rule to certain known probabilities and conditional probabilities; (3) we then use these values to get the unknown conditional probability that we want. This general, and usually very simple, method for getting an unknown conditional probability from known probabilities and conditional probabilities is sometimes called Bayes' Theorem. By custom, it is most often called Bayes' Theorem when it is used to obtain an unknown conditional probability $P(A|B)$ in which A refers to a part of the experiment that, in any given physical trial, occurs at an earlier time than the part referred to by B. (For example, transfer red occurs at an earlier time than draw red in any trial of the jar experiment.) In Example 2 above, if we calculate P(rain on first day | rain on second day) = $\frac{w}{y+w} = \frac{1/6}{1/3} = 1/2$, we would say that we have made the calculation by Bayes' Theorem.

Stochastic processes. In Example 2 above, let S_1 be the event sunny on the first day and let S_2 be the event sunny on the second day. Note that the events S_1 and S_2 are not independent (since $P(S_2) = 2/3$ and $P(S_2|S_1) = 3/4$.) When an outcome of an experiment can be viewed as a succession (finite or infinite) of events in time which occur with certain probabilities and conditional probabilities (and where independence may or may not hold), the probability space for the experiment is sometimes called a <u>stochastic process</u>. The probability space in Example 2 (observing the weather on two successive days) is an example of a stochastic process. A probability space for the larger time-series experiment of observing the succession of sunny and rainy days for an indefinitely long period would also be a stochastic process, as would a probability space for Brownian motion paths as described in Chapter 2. The word "stochastic" is from a Greek word meaning "guess", and the word "process" indicates development through time.

<u>Conceptual and philosophical note</u>. In choosing a particular probability function for a given experiment, we have explained and justified our choice by saying that we seek for each sample point a probability value that is the limiting value of the relative frequency for that outcome when the experiment is repeated many times. In fact, the conceptual situation is more subtle and complicated than this. For example, in the case of poker hands in Chapter 3, we assigned a probability value of

$$\frac{1}{2{,}598{,}960} = 0.0000004$$

to each possible (unordered) hand. Did this mean that we had conducted a huge number of trials and had in fact observed that each different hand occurred about 0.0000004 of the time? Clearly, it could not mean this. Instead, the conceptual situation is as follows. We begin with a major assumption: that there is a probability space for the experiment. Once we assume this, we then go on to assume certain further mathematical facts about this space on the basis of empirical observations. Finally, from these mathematical facts, we deduce that the probability space has a certain unique probability function. For example, in the case of poker hands, we begin with the assumption that there _is_ a probability function on the sample space of all possible poker hands. It is convenient here first to take our underlying sample space to be the space of all _ordered_ poker hands (there are $52 \cdot 51 \cdot 50 \cdot 49 \cdot 48 = 311{,}875{,}200$ points in this space), and to think of any given hand as being dealt in order from a shuffled deck. We now make the following _empirical_ comments. (1) Physical symmetry and observed relative frequencies suggest that the probability of getting a particular card as the _first_ card dealt should be 1/52. (2) Given that a certain card has already appeared as the first card, physical symmetry and observed relative frequencies suggest that the conditional probability of then getting a particular card as the _second_ card dealt should be 1/51. (3) Similarly, the conditional probability of each possible _third_ card, given the first two cards, should be 1/50, and the corresponding conditional probabilities for the _fourth_ and _fifth_ cards should be 1/49 and 1/48. But now, with these conditional probabilities, the general multiplication law enables us immediately to deduce that the probability of getting five specified cards in order must be $\frac{1}{52} \cdot \frac{1}{51} \cdot \frac{1}{50} \cdot \frac{1}{49} \cdot \frac{1}{48} = \frac{1}{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}$. Finally, going to unordered hands, we see that the probability of a specified unordered hand must be $\frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48} = 0.0000004$, since that unordered hand can appear as an ordered hand in 5! ways. This is the same probability function that we chose before, but now we have deduced it from empirical knowledge and observation about the dealing of poker hands (rather than base it on an unwarranted claim that we have observed each hand to occur 0.0000004 of the time.

EXERCISES FOR CHAPTER 4.

4-1. A jar contains w white balls and r red balls. Two balls are drawn at random, one after the other, without replacement. Find the following:

(a) P(second ball red | first ball white);

(b) P(second ball red);

(c) P(first ball red | second ball red).

4-2. Suppose $P(A) = 0.3$, $P(B) = 0.2$, and $P(A \cup B) = 0.4$. Find the following

(a) $P(A \cap B)$;

(b) $P(A|B)$;

(c) $P(B|A)$.

4-3. Successive cards are dealt, face up, from a bridge deck. What is the probability that the first spade occurs at the fifth card?

4-4. A bag contains 3 white marbles and 4 red ones. In succession, three persons each draw a marble without replacing it in the bag. The first person who draws a white marble wins. They continue drawing, in turn, until someone wins. What are the respective probabilities of winning for the person drawing first, for the person drawing second, and for the person drawing third?

4-5. A commuter uses either a tunnel or a bridge to get home. He chooses the tunnel with probability 1/3 and the bridge with probability 2/3. If he goes by tunnel, he gets home by 6 PM 75 percent of the time. If he goes by bridge (which is easier driving for him) he gets home by 6PM only 70 percent of the time. On a certain night, he gets home after 6 PM. What is the probability that he used the bridge?

4-6. Assume that boys and girls have an equal probability of being born. If a family having 4 children is known to have at least 1 boy, what is the probability that it has exactly 2 boys? What additional assumption are you making?

What is the probability that a family with four children has exactly two boys?

4-7. Three drivers, Arthur, Ben, and Charlie, share the driving of a taxi. Arthur drives it 40 percent of the time, and Ben and Charlie each 30 percent. The probability that Arthur has an accident on a day that he is driving is 0.02. For Ben and Charlie, the probabilities are 0.03 and 0.02. The owner of the taxi learns that an accident has occurred. What is the probability that the accident occurred on a day that Arthur was driving? On a day that Ben was driving? On a day that Charlie was driving?

4-8. In a freshman class, 10 percent of the students fail mathematics, 12 percent fail physics, and 4 percent fail both mathematics and physics. A student is selected at random from the class. Are the events <u>student failed mathematics</u> and <u>student failed physics</u> independent?

4-9. Two dice are rolled. Let us assume, for the purpose of our analysis, that one die is colored red and that the other is blue. In a single roll of the two dice, let u = number on red die, v = number on blue die, and $w = u + v$. Assume that each of the 36 possible outcomes (u,v) is equally probable.

(a) Make a table showing the values of $P(w = 1)$, $P(w = 2)$, ... , $P(w = 12)$.

(b) Show that when the two dice are repeatedly rolled, P(six appears before seven and for the first time on roll x) = $(\frac{25}{36})^{x-1} \frac{5}{36}$, and P(seven appears before six and for the first time on roll y) = $(\frac{25}{36})^{y-1} \frac{6}{36}$. (This result was used in Exercise 2-9.)

(c) Find P(six appears before seven) and P(seven appears before six). (This was Exercise 2-9.)

(d) Make a table showing P(seven appears before w) for w = 4,5,6,8,9,10.

4-10. The game of craps is described in Exercise 1-5. Exercise 4-9 above now puts us in a position to calculate the probability of pass in a game of craps (and hence to verify the value of λ given in Exercise 1-5). We proceed as follows. Calculate the following probabilities and conditional probabilities:

(a) P(pass occurs on first roll);

(b) P(don't pass occurs on first roll);

(c) P(first roll gives point w) for w = 4,5,6,8,9,10;

(d) P(pass occurs after first roll | point is w) for w = 4,5,6,8,9,10 (see Exercise 4-9a);

(e) P(pass occurs but not on first roll) (see 4-9c);

(f) P(pass occurs).

4-11. The game of roulette is described in Exercises 1-4 and 2-1. In North American roulette, which of the following pairs of events are independent?

(a) red; even.

(b) odd; $\geq$ 19.

(c) {0,20}; $\geq$ 18.

4-12. Consider the experiment of rolling a single die twice. Let u = number appearing on first roll and v = number appearing on second roll. Let w = u + v = total on the two rolls. Which of the following pairs of events are independent and which are dependent?

(a) u = 3; v = 5.

(b) w = 9; w $\geq$ 7.

(c) w = 9; u = 5.

(d) w = 7; v = 1.

4-13. For the experiment in Exercise 4-12, calculate the following conditional probabilities:

(a) $P(u = 4 \mid v > 4)$;

(b) $P(u = 3 \mid w = 4)$;

(c) $P(w = 4 \mid u = 3)$.

4-14. For the experiment of Exercise 4-12, consider the three events $w = 7$, $v = 3$, and $u = 4$.

(a) Show that each of the three pairs formed from these three events is independent.

(b) Show that the three events together do not form an independent set.

4-15. Show that the following analogues to the laws of probability hold for conditional probability:

(a) $P(A|C) + P(\overline{A}|C) = 1$;

(b) $P(A \cup B|C) = P(A|C) + P(B|C) - P(A \cap B|C)$;

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

4-16. Assume that $S = A_1 \cup A_2 \cup \ldots \cup A_n$, and that $A_i \cap A_j = \emptyset$ for all $i < j \leq n$. Show that:

(a) $B = \bigcup_{i \leq n} (A_i \cap B)$;

(b) $P(B) = \sum_{i \leq n} P(A_i)P(B|A_i)$.

(c) Replace the assumption that $S = A_1 \cup \ldots \cup A_n$ with the assumption that $C = A_1 \cup \ldots \cup A_n$. Show that

$$P(B|C) = \frac{\sum_{i \leq n} P(A_i)P(B|A_i)}{\sum_{i \leq n} P(A_i)} .$$

4-17. Assume that A, B, and C form an independent set of three events.

(a) Use Exercise 4-15 to conclude that

$$\begin{aligned} P(A \cap B \cap \overline{C}) &= P(A)P(B)(1 - P(C|A \cap B)) \\ &= P(A)P(B)P(\overline{C}|A \cap B) \\ &= P(A)P(B)P(\overline{C}). \end{aligned}$$

(b) Conclude that independence holds for the set of events $\{A,B,\overline{C}\}$.

(c) Use (b) to conclude that the following sets of three events are independent: $\{A,\overline{B},\overline{C}\}$ and $\{\overline{A},\overline{B},\overline{C}\}$.

4-18. (a) Assume that $P(A \cap B) = P(A)P(B)$. Show that A and B are independent.

(b) Show that the set $\{A,B,C\}$ may not be independent, even though $P(A \cap B \cap C) = P(A)P(B)P(C)$. (*Hint*. Take the equiprobable measure on $S = \{1,2,3,4,5,6,7,8\}$. Let $A = \{1,2,3,4\}$, $B = \{1,5,6,7\}$, and $C = \{1,2,6,7\}$.)

4-19. Three small boxes are given. Each box has two drawers. In the first box, both drawers contain gold coins. In the second box, one drawer contains a gold coin and one contains a silver coin. In the third box, both drawers contain silver coins. One of the three boxes is chosen at random, and then one of its two drawers is opened at random. It contains a silver coin. What is the conditional probability that the other drawer in the same box contains a gold coin? (This is a well-known example. The following argument is plausible but incorrect: "the other drawer in the same box must contain either a gold or a silver coin; because the probability of initially choosing the silver-gold box is equal to the probability of choosing the silver-silver box, the desired conditional probability must be 1/2." Reference to the intended intuitive meaning of conditional probability, as described on page 102, will show the reader why this argument and conclusion are incorrect.)

4-20. A student is studying for an examination on two chapters of her textbook. Each chapter consists of five sections. She has time to study eight of the ten sections she is responsible for. She knows that the examination will consist of three questions on three different sections, with two sections from one chapter and one section from the other. She decides to assume that the examiner will toss a coin to decide which chapter to take the two sections from and that the sections from each chapter will then be chosen at random.

(a) If the student wishes to maximize the probability of getting all three questions right, how should she divide her studying between the two chapters? What is this probability?

(b) If she wishes to maximize the probability of getting at least two questions right, how should she divide her studying between the two chapters? What is this probability?

4-21. Detective Hawkshaw decides to use a probability space to help analyze a murder case. In this probability space, he assigns certain probabilities as follows. The probability that the butler is guilty is 0.05. If the butler is innocent, the probability that he is seen near the crime is 0.02. If the butler is guilty, the probability that he is seen near the crime is 0.5.

(a) Hawkshaw then discovers that the butler was seen near the crime. He calculates a new probability that the butler is guilty. What is its value? (_Hint_. Let B = _butler guilty_ and V = _butler was seen_. Find $P(B|V)$ using values for $P(B)$, $P(V|B)$, and $P(V|\overline{B})$.)

(b) After discovering that the butler was seen near the crime, Hawkshaw then searches the butler's room and finds a gun. He assumes that the probability of finding a gun is 0.04 if the butler is innocent and 0.25 if the butler is guilty. He now calculates a still newer probability that the butler is guilty. What is its value? (_Hint_. Let G = _gun in butler's room_. Proceed as in (a), but using, in place of 0.05, the new probability value calculated in (a) for the initial probability that the butler is guilty. More formally and rigorously, this amounts to the following. (i) Verify that $P(B|G \cap V) = \frac{P(B \cap G|V)}{P(G|V)}$ and that $P(B \cap G|V) = P(B|V)P(G|B \cap V)$. (ii) Assume that $P(G|B \cap V) = P(G|B)$. The result follows directly from these formulas.)

Note. In any probability space, an event C with $P(C) \neq 0$ can be taken as a probability space in its own right if we define a probability measure P_C on this new space by setting $P_C(A) = P(A|C)$ for every subset A of C. (See also Exercise 4-15.) This new probability space is called the conditional probability space determined by C. Fact (i) in the above hint is then just the calculation of a conditional probability in the conditional probability space V, and (ii) is the assertion that the events $G \cap B$ and $V \cap B$ are independent in the conditional space B.

Note. Early in the history of probability theory, it was hoped that probability theory would supply a mathematical calculus that could be used to quantify and weigh evidence in courts of law in exactly the manner indicated in the above example. Although such calculations have not been given recognized legal status, some probabilists and statisticians today believe that calculations of this kind can provide a useful normative reconstruction of the less formal processes of analysis and thought that go on in the minds of jurors and judges. It should be noted at this point in our study that Hawkshaw has no apparent information in the form of observed relative frequencies upon which to base his choice of an underlying probability measure and that the meaning of the probabilities which he calculates is therefore not entirely clear. Indeed, he does not appear to be concerned at all here with experiments that can be "repeated under the same general experimental conditions." The use of a probability space and of probability calculations in such circumstances will be considered further in Chapter 20. The lack of agreed-on criteria for assigning these underlying probabilities is perhaps one reason that probability calculus has not been recognized as a formal instrument for the analysis of evidence.

Note. The calculations in the above exercise are Bayes' Theorem calculations. In (a), the calculation is done in an underlying given probability space. In (b) it is done in a certain conditional probability space. In such a calculation, the initial probability for an event is called an a priori or prior probability, and the conditional probability being calculated using new evidence is called an a posteriori or posterior probability. For the calculation in (a), the prior probability is 0.05 and the posterior probability is the answer to (a). For the calculation in (b), the prior probability is the answer to (a), and the posterior probability is the answer to (b).

Additional Homework Problems for Chapter 4.

4-22. In the Massachusetts Megabucks Lottery, each entry selects six distinct integers from the set $L = \{1,2,\ldots,35,36\}$. A sample of 6 integers is then drawn at random, without replacement, from L.

(a) Define p_x to be the probability that a particular entry has exactly x integers in common with the random sample. Make a table of values for p_x for the possible values $x = 0,1,\ldots,6$.

(b) Each entry costs one dollar. Assuming that there is no payoff for $x < 5$, what should the payoff be for $x = 6$ to make this a fair bet? (Note. In the present lottery, there is, in fact, a payoff of \$40 for $x = 4$ and a payoff of \$400 for $x = 5$. We shall later consider this example in more detail.)

4-23. To bet on the Suffolk Downs Twin Trifecta, a bettor selects (i) an ordered sample (without replacement) of three horses from the twelve horses about to run in the 8th race and (ii) an ordered sample (without replacement) of three horses from the twelve horses running in the 10th race. The bettor wins the bet if and only if, after the two races are completed, the two ordered samples in the bet coincide, in order, with the horses who finish 1st, 2nd, and 3rd in the respective races.

(a) Assume that the bettor makes his or her selections by choosing the two samples at random. What is the probability that the bettor will win?

(b) The bet costs three dollars. What should the payoff be to make this a fair bet?