SOME THEORY AND PRACTICE OF STATISTICS

by Howard G. Tucker

CHAPTER 1. EVENTS AND PROBABILITIES

1.1 The Notion of Probability. The notion of the probability of an event may be approached by at least three methods. One method, perhaps the first historically, is to repeat many times an experiment or a game in which the event might or might not occur under identical conditions and then compute the relative frequency with which the event occurs. This means: divide the total number of times that the specific event occurs by the total number of times that the experiment is performed or the game is played. This ratio is called the *relative frequency* and is really only an approximation of what would be considered as the probability of the event. For example, if one tosses a penny 25 times, and if it comes up heads 14 times, then we would estimate the probability of of this coin coming up heads as $14/25$ or 0.56. Although this method of arriving at the notion of probability is the most primitive and unsophisticated, it is the most meaningful to the practical individual, and in particular to the working scientist and engineer who must apply the results of probability theory to real-life situations. Accordingly, whatever results one obtains in the theory of probability and statistics, one should be able to interpret them in terms of relative frequency.

A second approach to the notion of probability is from an axiomatic point of view. That is, a minimal list of axioms is set down which assumes certain properties of probabilities. From this minimal set of assumptions the further properties of probability are deduced and applied. The axiomatic approach will be used in this book, realistically tempered by relative frequency interpretations from time to time.

A third approach to the notion of probability is limited in application but is extremely useful. This approach is that of probability in the "equally likely" case. Let us consider some game or experiment which, when played or performed, has among its outcomes a certain event E. For example, in tossing a die once, the event E might be that the outcome is an even number. In general, we might suppose that the experiment or game has a certain number of "equally likely" outcomes. Let us suppose that a certain event E can occur in any one of a specified number of these "equally likely" outcomes. Then the probability of the event is defined to be the number of "equally likely" ways in which this event can occur divided by the total number of possible "equally likely" outcomes. It must be emphasized that the number of equally likely ways in which the event can occur must be from *among* the total number of equally likely outcomes. For example, if as above the experiment or game is the single toss of a die in which the "equally likely" outcomes are the numbers $1, 2, 3, 4, 5, 6$, and if the event E considered is that the outcome is an even number, i.e., is 2, 4 or 6, then the probability of E here is defined to be $3/6$ or 0.5. This approach is limited, as was mentioned above, because in many games or experiments the possible individual outcomes are not "equally likely".

In this chapter an introduction is laid out about events and their probabilities. First we consider probabilities in the all-important equally likely case. Then we develop the notion of event. The axioms of probability are then given, and their immediate properties are deduced. Finally, the notions of dependence and independence are discussed in some detail.

EXERCISES

1. A (possibly loaded) die was tossed 150 times. The number 1 came up 27 times, the number 2 came up 26 times, 3 came up 24 times, 4 came up 20 times, 5 came up 29 times and 6 came up 24 times. (i) Compute the relative frequency of the event that the outcome is 1. (ii) Find the relative frequency of the event that the outcome is an even number. (iii) Find the relative frequency of the event that the outcome is not less than 5.

2. Twenty numbered tags are in a hat. The number 1 is on seven of the tags, the number 2 is on five of the tags and the number 3 is on eight of the tags. The experiment is to stir the tags without looking at them and to select one tag "at random".

(i) What is the total number of equal outcomes of the experiment?

(ii) From among these twenty equally likely outcomes what is the total number of ways in which the outcome is the number 1?

(iii) From among the total number of equally likely outcomes of the experiment, what is the total number of eqally likely outcomes in which one draws the number 3?

(iv) Compute the probability of selecting a tag numbered 1 or 3.

(v) What is the sum of the probabilities obtained in (ii) and (iii)?

1.2 Combinatorial Probability. In this section a brief exposition is given on how to compute probabilities in the equally likely case.

Let us suppose that we have n different objects, and suppose that we wish to arrange k of these in a row (where of course, $k \leq n$). We wish to know in how many ways this can be accomplished. As an example, suppose there are five members of a committee, call them A, B, C, D and E, and we want to know in how many ways we can select a chair of the committee and its secretary. When we select the arrangement (C, A), we mean that individual C is the chair and A is the secretary. In this case, $n = 5$ and $k = 2$. The different arrangements are listed as follows:

$$\begin{array}{llll} (A,B) & (A,C) & (A,D) & (A,E) \\ (B,A) & (B,C) & (B,D) & (B,E) \\ (C,A) & (C,B) & (C,D) & (C,E) \\ (D,A) & (D,B) & (D,C) & (D,E) \\ (E,A) & (E,B) & (E,C) & (E,D) \end{array}$$

One sees that there are twenty such arrangements. The number 20 can also be obtained by the following reasoning: there are five ways in which the chair can be selected (which accounts for the five horizontal rows of pairs), *for each* person selected as chair there are four ways of selecting a secretary (which accounts for the four vertical columns), and consequently there are 20 such pairs. In general, if we want to determine in how many ways we can arrange k out of n objects, we reason as follows. There are n ways of selecting the first object. For each way we select the first object there are $n-1$ ways of selecting the second object. Hence the total number of ways in which the first two objects can be selected is $n(n-1)$. For each way the first two objects are selected there are $n-2$ ways of selecting the third object from those remaining. Thus, the number of ways in which the first three objects can be selected is $n(n-1)(n-2)$. From this one can conclude that the number of ways in which k out of n objects can be laid out in a row is $n(n-1)(n-2)\cdots(n-k+1)$, which can be written as the ratio of factorials: $n!/(n-k)!$. This is also referred to as the number of permutations of n things taken k at a time.

In the above arrangements (or permutations) of n things taken k at a time, we counted each way in which we could arrange the same k objects in a row. (For example, we listed both (A, B) and (B, A).) Suppose, however, that one is interested only in the number of ways k objects can be selected from the n objects and is not interested in the order in which they appear or are drawn. In the case of the committee discussed above, the number of ways in which two members can be selected out of the five to form a subcommittee is as follows:

$$\begin{array}{llll} (A,B) & (A,C) & (A,D) & (A,E) \\ (B,C) & (B,D) & (B,E) & \\ (C,D) & (C,E) & & \\ (D,E) & & & \end{array}$$

We do not list (D, B) as before because the subcommittee denoted by (B, D) is the same as that denoted by (B, D) which is already listed. Now we have only half the number of selections. In general, if we want to find the number of ways in which one can select k objects out of n objects, we reason as follows: as before, there are $n!/(n-k)!$ ways of arranging (or permuting) n objects taken k at a time. However, all $k!$ ways of arranging every subset of k objects are included here. Hence we must divide the $n!/(n-k)!$ ways of arranging k out of n objects by $k!$ to obtain the number of ways in which we can make the k selections. This number of ways in which we can select k objects out of n objects without regard to order is usually referred to as the number of combinations of n objects or things taken k at a time. It is denoted here by the binomial coefficient:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

This binomial coefficient is usually first encountered when learning the binomial theorem:

$$(a+b)^n = \textstyle\sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k},$$

where $0!$ is defined to be 1.

Now we shall apply these two notions to some combinatorial probability problems, i.e., the computation of probabilities in the “equally likely" case. In each problem, the cautious approach is first to determine the number of equally likely outcomes in the game or experiment, Then one computes the number of equally likely ways *from among these* in which the particular event can occur. Then the ratio of the second number to the first number is computed to obtain the probability of the event.

Example 1. The numbers $1, 2, \cdots, n$ are arranged in random order, i.e., the $n!$ ways in which these numbers can be arranged are assumed to be equally likely. We are to find the probability that the numbers 1 and 2 appear as neighbors in the order indicated.

As was mentioned in the problem, there are $n!$ equally likely outcomes. In order to compute the number of these equally likely ways in which the indicated event can occur, we reason as follows: there are $n-1$ positions permitted for 1, for each poition available for 1 there is only one position available for 2, and for every favorable selection of positions for 1 and 2 there are $(n-2)!$ ways for arranging the remaining $n-2$ integers in the remaining $n-2$ positions. Consequently, there are $(n-1)(n-2)!$ ways in which this event can occur, and its probability is

$$P = \frac{(n-1)(n-2)!}{n!} = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

Before beginning the second example, we should explain what is meant by selecting a random digit. In effect, one takes ten tags and marks 0 on the first tag, 1 on the second tag, $\cdots$, and 9 on the tenth tag. These tags are then placed in a hat or an urn. If we say "select n random digits" or "sample n times with replacement", we mean that one selects a tag at random, notes the number on it and records it, *returns it to the container*, and repeats this action $n-1$ times more.

Example 2. We are to find the probability P that among k random digits neither 0 nor 1 appears.

The total number of possible outcomes is obtained as follows. There are ten possibilities for the first digit. For each possibility for the first digit there are ten ways in which the second digit can be selected. So there are 10^2 ways in which the first two digits can be selected. For each way in which the first two digits are selected, there are ten ways in which the third digit can be selected. Thus there are 10^3 ways of selecting the first three digits. In general, it is not hard to see that there are 10^k ways of selecting these ten digits. Now we consider the event: neither 0 nor 1 appears. In how many ways from among the 10^k equally likely possible outcomes can this event occur? In selecting the k random digits, there are eight ways out of ten in which the first digit selected is neither 0 nor 1. The same goes for the second, third, and on up to the kth random digit. Hence out of the 10^k total number of equally likely outcomes there are 8^k outcomes in which this event can occur. Thus $P = 8^k/10^k$.

Example 3. Now let us determine the probability P that among k random digits the digit 0 appears exactly 3 times (assuming $3 \leq k$).

Again the total number of equally likely outcomes is 10^k. Among the k trials (which we can look at as k different objects) there are $\binom{k}{3}$ ways of selecting the three trials in which the digit 0 appears. For each way of selecting the three trials in which only 0 appears there are 9^{k-3} ways in which the remaining $k-3$ outcomes can occur. Thus $P = \binom{k}{3}9^{k-3}/10^k$.

Example 4. A box contains 90 white balls and 10 red balls. If nine balls are selected at random without replacement, what is the probability that six of them are white?

In this problem there are $\binom{100}{9}$ ways of selecting nine balls out of 100. Since there are $\binom{90}{6}$ ways of selecting six white balls out of 90 white balls, and since for each way in which one selects the six white balls there are $\binom{10}{3}$ ways of selecting three red balls out of the ten red balls, we see that there are $\binom{90}{6}\binom{10}{3}$ ways of selecting six white balls when we select nine of them at random without replacement. Consequently,

$$P = \frac{\binom{90}{6}\binom{10}{3}}{\binom{100}{9}}.$$

Example 5. There are n men standing in a row, among whom are two named A and B. We would like to find the probability P that there are r men between A and B, where $r \leq n-2$. There are two ways of solving this problem. In the first place there are $\binom{n}{2}$ ways in which one can select places for A and B to stand, and among these there are $n-r-1$ ways in which one can select two positions which have r positions betweeen them. So

$$P = \frac{n-r-1}{\binom{n}{2}}.$$

Another way of solving this problem is to say that there are $n!$ ways of arranging these n men in a row, and that among these $n!$ ways there are two ways of selecting one of the men A and B. For each way of selecting one of A or B there are $n-r-1$ ways of placing him among the first $n-r-1$ places, for each way of selecting one of A or B and placing him among the first $n-r-1$ places there is only one place that the other man can be placed in order that there be r men between them, and there are $(n-2)!$ ways of arranging the remaining $n-2$ men in the $n-2$ remaining places. So

$$P = \frac{2(n-r-1)(n-2)!}{n!} = \frac{n-r-1}{\binom{n}{2}}.$$

EXERCISES

1. An urn contains four black balls and six white balls. Two balls are selected at random without replacement. What is the probability that

(i) one ball is black and one ball is white?

(ii) both balls are black?

(iii) both balls are white?

(iv) both balls are the same color?

2. In tossing a pair of fair dice, what is the probability of obtaining a 7 or an 11?

3. Two unbiased coins are tossed simultaneously. What is the probability that

(i) they are both heads?

(ii) they both match?

(iii) one is heads and the other is tails?

4. The numbers $1, 2, \cdots, n$ are placed in a random order in a straight line. Find the probability that

(i) the numbers $1, 2, 3$ appear as neighboors in the order given.

(ii) the numbers $1, 2, 3$ appear as neighbors in any order.

5. Among k random digits, find the probability that

(i) no even digit appears.

(ii) no digit divisible by three appears.

6. Among k random digits ($k \geq 5$) find the probability that

(i) the digit 1 appears exactly five times.

(ii) the digit 0 appears two times and the digit 1 appears three times.

7. A box contains ten white tags and five black tags. Three tags are selected at random without replacement. What is the probability that two are black and one is white?

8. There are n men standing in a circle, among whom are two men named A and B. What is the probability that there are r men between them?

1.3 The Fundamental Probability Set and the Algebra of Events. We now leave the case where all individual outcomes are equally likely and consider the general case. However, before we may adequately discuss probabilities of events we must discuss the events themselves. This study constitutes what is known as the algebra of events.

Connected with any game or experiment is a set or space which is the collection of all possible individual outcomes of that game or experiment.

Such a collection of all possible individual outcomes is called a *fundamental probability set* or *sure event* and will be denoted by the Greek letter Ω. We shall also use the expression *fundamental probability set* for any representation we might construct of the collection of all individual outcomes. For example, in a game of one toss of a coin, a fundamental probability set consists of two individual outcomes which can be conveniently referred to as H (heads) and T (tails). If the game consists of tossing the coin twice, then the fundamental probability set consists of four individual outcomes. One of these outcomes could be denoted (T, H), which means that "tails" occurs on the first toss and "heads" occurs on the second toss. The remaining three individual outcomes may be denoted by (H, H), (H, T) and (T, T), and one can easily determine what these outcomes signify. In general, an arbitrary individual outcome will be denoted by the lower case Greek letter omega, which appears as ω and will be referred to as an "elementary event." Thus Ω denotes the totality of elementary events ω.

An event is simply a collection of elementary events. Different events are different collections of elementary events. (Note: A single elementary event is not necessarily an event.) An event will usually be denoted (with or without a subscript) by an upper case Roman letter near the beginning of the alphabet. Consider the game in which a coin is tossed twice. Then, as was indicated above, Ω consists of the following four elementary events: (H, H), (H, T), (T, H), and (T, T). If A denotes the event: heads occurs on the first toss, then A consists of these two elementary events: (H, H) and (H, T). If B denotes the event: at least one head appears, then B consists of the following three elementary events: (H, H), (H, T) and (T, H). If C denotes the event: no heads appear, then C consists of one elementary event, namely, (T, T). If D denotes the event that heads occurs three times in the two tosses of the coin, this is clearly impossible, and so D is a subset of Ω that contains no elementary events, i.e., it is the empty set. This event is denoted by $\emptyset$

In general, we shall denote the fact that a certain elementary event ω belongs to the collection of elementary events which determine an event A by $\omega \in A$. If an elementary event ω occurs, and if $\omega \in A$, then we shall say that the event A occurs. It must be noted at this point that just because an event A occurs, it does not nmean that no other event occurs. In the previous example, if (H, H) occurs, then A occurs and so does B. The fundamental probability set Ω is called the *sure event*, because whatever elementary event ω does occur, always $\omega \in \Omega$.

We now introduce some algebraic operations over events. If A is an event, then A^c will denote the event that the event A does not occur. Thus A^c consists of all those elementary events that are not in A. For every elementary event ω in the fundamental probability set and for every event A, one and only one of the following is true: $\omega \in A$ or $\omega \in A^c$. An equivalent way of writing $\omega \in A^c$ is $\omega \notin A$, and we say that ω is not an elementary event of A. Also, A^c is called the negation of A ("not A") or the complement of A.

If A and B are events, then $A \cup B$ will denote the event that at least one of these two events occurs. By this we mean that A occurs but B does not occur or A does not occur and B occurs or both of A and B occur. In the previous example, if E denotes the event that heads occurs in the second trial, then E consists of (H, H) and (T, H), and $A \cup E$ consists of the elementary events (H, H), (H, T) and (T, H). In other words, $A \cup E$ is the event that heads occurs at least once, and we may write $A \cup E = B$.

More generally, if $A_1, A_2, \cdots, A_n$ are any n events, then $A_1 \cup A_2 \cup \cdots \cup A_n$ denotes the event that *at least one* of these n events occurs. We usually write this event with the following notation:

$$\cup_{i=1}^{n} A_i.$$

Also if $A_1, A_2, \cdots, A_n, \cdots$ denotes an infinite sequence of events then

$$A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots$$

or

$$\cup_{i=1}^{\infty} A_i$$

denotes the event that at least one event in this sequence occurs.

Suppose A and B are events which cannot both occur, i.e., if $\omega \in A$, then $\omega \notin B$, and if $\omega \in B$, then $\omega \notin A$. In this case, events A and B are said to be *disjoint* or *incompatible* or *mutually exclusive*.

The notation $A \subset B$ means: if event A occurs, then the event B occurs. Other ways of stating this are: A implies B, B is implied by A, or: if A then B. Technically we would define $A \subset B$ by: if $\omega \in A$, then $\omega \in B$. Thus, in any situation where one is confronted with two events A and B, and one wishes to prove that $A \subset B$, one first takes an arbitrary ω in A and then proves that this ω is also in B. This technique will be illustrated again and again in some of the proofs that follow.

The event that both events A and B occur is denoted by $A \cap B$ or AB. The symbol $\emptyset$ will denote the impossible event, i.e., the event that contains no elementary events. It is clear that if $A \cap B = \emptyset$, then A and B are disjoint events.

Finally, we define the equality of two events. We shall say that events A and B are equal and write $A = B$ if every elementary event in A is an elementary event in B, and if every elementary event in B is also an elementary event in A. It is clear then that $A = B$ if and only if $A \subset B$ and $B \subset A$. One should remember this when trying to prove the equality of two events.

We now prove some properties of the algebra of events.

Proposition 1. *If A is an event, then $A \subset A$.*

Proof: Recall the paragraph before last. Let $\omega \in A$. Then this same $\omega \in A$. Hence every elementary event in the left event is also an elementary event in the event on the right side.

Proposition 2. *If A, B, C are events, if $A \subset B$ and if $B \subset C$, then $A \subset C$.*

Proof: Let $\omega \in A$; we must show that this $\omega \in C$. Since $A \subset B$ and $\omega \in A$, then $\omega \in B$. Now, since $B \subset C$, and since $\omega \in B$, then $\omega \in C$.

Proposition 3. *For every event A, $A \cap A = A$, $A \cup A = A$ and $(A^c)^c = A$.*

Proof: These conclusions are obvious.

Proposition 4. *For every event A, $\emptyset \subset A \subset \Omega$.*

Proof: The trick here involves the fact that any ω that one might find in $\emptyset$ is certainly in A, since $\emptyset$ contains no ω's. The implication $A \subset \Omega$ is obvious.

Proposition 5. *If A and B are events, and if $A \subset B$, then $B^c \subset A^c$. If $A = B$, then $A^c = B^c$.*

Proof: Let $\omega \in B^c$. Then $\omega \notin B$. This implies that $\omega \notin A$, since if to the contrary $\omega \in A$, then the hypothesis $A \subset B$ would imply $\omega \in B$, which contradicts the fact just established that $\omega \notin B$. But $\omega \notin A$ implies $\omega \in A^c$. Thus $B^c \subset A^c$. Next if $A = B$, then $A \subset B$ and $B \subset A$, which by the first conclusion just proved imply $B^c \subset A^c$ and $A^c \subset B^c$, which in turn imply $B^c = A^c$.

Proposition 6. (DeMorgan formulas) *If $A_1, A_2, \cdots, A_n$ are events, then $(\cup_{i=1}^n A_i)^c = \cap_{i=1}^n A_i^c$ and $(\cap_{i=1}^n A_i)^c = \cup_{i=1}^n A_i^c$.*

Proof: In order to prove the first equation, let ω be any elementary event in the left hand side. Then ω is not in the event that at least one of $A_1, A_2, \cdots, A_n$ occurs. This means that ω is not an element of any of the A_i

's, i.e., $\omega \notin A_i$ for all $i = 1, 2, \cdots n$. Hence, $\omega \in A_i^c$ for all i. i.e., $\omega \in \cap_{i=1}^n A_i^c$. Thus we have shown that $(\cup_{i=1}^n A_i)^c \subset \cap_{i=1}^n A_i^c$. Now let ω be any elementary event in $\cap_{i=1}^n A_i^c$. Then $\omega \in A_i^c$ for all i. Hence $\omega \notin A_i$ for $i = 1, 2, \cdots, n$, and when this happens, then certainly $\omega \notin \cup_{i=1}^n A_i$. But this means that $\omega \in (\cup_{i=1}^n A_i)^c$. Thus $\cup_{i=1}^n A_i^c \subset (\cap_{i=1}^n A_i)^c$, and since we have proved the reverse implication, we therefore have the first equality of our proposition.In order to prove the second equation, we use the first one that is already proved and replace each A_i by A_i^c. Thus we obtain $(\cup_{i=1}^n A_i^c)^c = \cap_{i=1}^n (A_i^c)^c$, and by proposition 3 this becomes $(\cup_{i=1}^n A_i^c)^c = \cap_{i=1}^n A_i$. Now take the complement of both sides to obtain the second equation by proposition 5.

Proposition 7. *$\emptyset^c = \Omega$ and $\Omega^c = \emptyset$.*

Proof: Since $\emptyset$ contains no elementary events, its complement must be the set of all elementary events, i.e., Ω. Also, the negative or the complement of the sure event is the event consisting of no elementary events.

Proposition 8.*If A and B are events, and if $A \subset B$, then $A \cap B = A$.*

Proof: Since $A \subset B$, then all the elementary events in both A and B are in A, and every elementary event in A is also in B and therefore in A and B.

Proposition 9: *If A and B are events, then $A \cap B \subset A$.*

Proof: If $\omega \in A \cap B$, then $\omega \in A$ and $\omega \in B$, which implies that $\omega \in A$.

Proposition 10. *If A and B are events, then $A \cup B = B \cup A$ and $A \cap B = B \cap A$.*

Proof: If $\omega \in A \cup B$, then $\omega \in A$ or $\omega \in B$. If $\omega \in A$, then ω is in at least one of the two events B, A, namely A; if $\omega \in B$, then ω is in at least one of the two events B, A, namely B. Thus $\omega \in B \cup A$, and we have shown that $A \cup B \subset B \cup A$. Since this holds for any two events, we may replace A above by B and B by A to obtain $B \cup A \subset A \cup B$. These two inclusions imply $A \cup B = B \cup A$. In order to prove the second equation, replace A and B in the first equation by A^c and B^c, take complements of both sides and apply propositions 5 and 6.

Proposition 11. *If A, B and C are events, then $A \cup (B \cup C) = (A \cup B) \cup C$ and $A \cap (B \cap C) = (A \cap B) \cap C$.*

Proof: If $\omega \in A \cup (B \cup C)$, then at least one of the assertions $\omega \in A$ or $\omega \in B \cup C$ is true. If $\omega \in A$ is true then ω is an element of at least one of A, B, namely A. Hence $\omega \in A \cup B$. This in turn implies that ω is in at least one of $A \cup B$ and C, namely $A \cup B$. Thus, $\omega \in (A \cup B) \cup C$, and we have established the inclusion $A \cup (B \cup C) \subset (A \cup B) \cup C$. In order to establish the reverse inclusion and hence the first equation, we use proposition 10 and the above inclusion to obtain

$$(A\cup B)\cup C = C\cup(A\cup B) = C\cup(B\cup A) \subset (C\cup B)\cup A = (B\cup C)\cup A = A\cup(B\cup C).$$

In order to establish the second equation, replace A, B and C in both sides of the first equation by A^c, B^c and C^c respectively, take the complements of the two sides, and apply propositions 5 and 6 to obtain the conclusion.

Proposition 12. *If A, B and C are events, then $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.*

Proof: If $\omega \in A \cap (B \cup C)$, then $\omega \in A$ and ω is in at least one of B, C. If $\omega \in B$, then $\omega \in A \cap B$; if $\omega \in C$, then $\omega\ A \cap C$. In either case, $\omega \in (A \cap B) \cup (A \cap C)$. If $\omega \in (A \cap B) \cup (A \cap C)$, then ω is in at least one of $A \cap B$ and $A \cap C$. If $\omega \in A \cap B$, then $\omega \in A$ and $\omega \in B$; now $\omega \in B$ implies $\omega \in B \cup C$, and hence $\omega \in A \cap (B \cup C)$. If $\omega \in A \cap C$, then replace B by C in the previous statement to obtain the same conclusion. In order to prove the second equation, replace A, B and C in the first equation by A^c, B^c and C^c respectively, take the complements of both sides and apply propositions 5 and 6.

Proposition 13. *If A and B are events, then $A \cup B = A \cup (A^c \cap B)$, and A and $A^c \cap B$ are disjoint.*

Proof: If $\omega \in A\cup B$, then $\omega \in A$ or $\omega \in B$. If $\omega \in A$, then $\omega \in A\cup(A^c\cap B)$. If $\omega \in B$, then two cases occur: $\omega \in A$ also, or $\omega \notin A$. In the first case, $\omega \in A \cup (A^c \cap B)$. In the second case, $\omega \in A^c$ while yet $\omega \in B$, i.e., $\omega \in A^c \cap B$, and thus $\omega \in A \cup (A^c \cap B)$. Thus $A \cup B \subset A \cup (A^c \cap B)$. Now let $\omega \in A \cup (A^c \cap B)$. Then $\omega \in A$ or $\omega \in A^c \cap B$. If $\omega \in A$, then $\omega \in A \cup B$. If $\omega \in A^c \cap B$, then $\omega \in B$, and hence $\omega \in A \cup B$. Thus, $A \cup (A^c \cap B) \subset A \cup B$, and the equation is established. Also, A and $A^c \cap B$ are disjoint, since if $\omega \in A^c \cap B$, then $\omega \in A^c$ or $\omega \notin A$.

EXERCISES

1. Prove: if B is an event, then $\emptyset \cap B = \emptyset$ and $B \cap \Omega = B$. (See propositions 4 and 8.)

2. If A is an event, use problem 1 and propositions 3 and 7 to prove that $\emptyset \cup A = A$ and $\Omega \cup A = \Omega$.

3. Use only proposition 9 to prove that if A and B are events, then $A \subset A \cup B$.

4. Use problem 3 and two propositions in this section to prove that if A and B are events, then $A \cap B \subset B$.

5. Prove: if A, B, C and D are events, if $A \subset B$ and if $C \subset D$, then $A \cap C \subset B \cap D$.

6 Let $A_1, A_2, \cdots, A_7$ be events. Match these four events:

(i) $A_1^c \cap A_2^c \cap A_3$,

(ii) $A_6 \cap A_7^c$,

(iii) $A_1^c \cap A_2 \cap A_3^c \cap A_4^c \cap A_5 \cap A_6^c \cap A_7^c$, and

(iv)$A_2 \cap A_5$

with the events defined in the following four statements:

(i) A_6 is the last event to occur,

(ii) events A_2 and A_5 occur,

(iii) A_3 is the first event to occur, and

(iv) A_2 and A_5 are the only events that occur.

7. Let A_1, A_2 and A_3 be events. Define B_i to be the event that A_i is the first of these events to occur, $i = 1, 2, 3$. Write each of B_1, B_2, B_3 in terms of A_1, A_2, A_3, prove that B_1, B_2, B_3 are disjoint, and prove that $A_1 \cup A_2 \cup A_3 = B_1 \cup B_2 \cup B_3$.

8. Prove: if A and B are events, then $B = (A \cap B) \cup (A^c \cap B)$. (Hint: first verify that $\Omega = A \cup A^c$, and then apply proposition 12 and problem 1 above.)

1.4. The Axioms of a Probability Space. This section is devoted to the axioms used in this book for a mathematical model for probability. First will come the axioms for the set of all events; this is followed by stating the requirements of the probability measure associated with these events.

It should be emphasized that from here on a *denumerable* set will refer to an infinite set that can be put into one-to-one correspondence with the set of all positive integers. A set will be referred to as being *countable* if it is either finite or denumerable. Now let us consider some fundamental probability set Ω. We shall denote the set of all events to be considered by the upper case caligraphic letter $\mathcal{A}$. Not all subsets of Ω will be events, i.e., will not be in $\mathcal{A}$. However, for that collection of subsets of Ω which we shall refer to as events, we assume the following three properties.

(i) for every event A in $\mathcal{A}$ (or for every $A \in \mathcal{A}$), then also A^c is in $\mathcal{A}$, (or $A^c \in \mathcal{A}$),

(ii) if $A_1, A_2, \cdots, A_n, \cdots$ is any denumerable sequence of events in $\mathcal{A}$, then

$$\cup_{n=1}^{\infty} A_n \in \mathcal{A},$$

and *(iii)* $\emptyset \in \mathcal{A}$.

A collection of events with properties (i), (ii) and (iii) will be called a sigma algebra of events. We shall always assume that every collection of elementary events in a finite or denumerable fundamental probability set is an event. However, in the case where the fundamental probability set is not countable, (e.g., the set of all real numbers in some interval of positive length) then this is not necessarily true.

There are some important consequences of the above three axioms which are obtained below.

Theorem 1. *If $A_1, A_2, \cdots, A_n$ is any finite sequence of events in $\mathcal{A}$, then*

$$\cup_{i=1}^{n} A_i \in \mathcal{A}.$$

Proof: Since $\emptyset \in \mathcal{A}$, let us define $\emptyset = A_{n+1} = A_{n+2} = \cdots$. Then by hypothesis, by (ii) and by (iii), $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$. But $\cup_{i=1}^{\infty} A_i = \cup_{i=1}^{n} A_i$, which proves the theorem.

Theorem 2. *The sure event Ω is always in $\mathcal{A}$, i.e., $\Omega \in \mathcal{A}$.*

Proof: By (iii) $\emptyset \in \mathcal{A}$, and thus, by (i), $\emptyset^c \in \mathcal{A}$. But $\emptyset^c = \Omega$.

Theorem 3. *If $A_1, A_2, \cdots, A_n, \cdots$ is any denumerable sequence of events in $\mathcal{A}$, then $\cap_{n=1}^{\infty} A_n \in \mathcal{A}$.*

Proof Since $A_n \in \mathcal{A}$ for $n = 1, 2, \cdots$, then by (i) it follows that $A_n^c \in \mathcal{A}$ for $n = 1, 2, \cdots$. Then by axiom (ii), $\cup_{n=1}^{\infty} A_n^c \in \mathcal{A}$. Axiom (i) then implies that $(\cup_{n=1}^{\infty} A_n^c)^c \in \mathcal{A}$. By one of the DeMorgan laws (proposition 6 in section 1.3),

$$(\cup_{n=1}^{\infty} A_n^c)^c = \cap_{n=1}^{\infty} (A_n^c)^c = \cap_{n=1}^{\infty} A_n,$$

which proves the theorem.

Corollary to Theorem 3. *If $A_1, \cdots A_n$ is any finite sequence of events in $\mathcal{A}$, then $\cap_{k=1}^{n} A_k \in \mathcal{A}$.*

Proof: Let $\Omega = A_{n+1} = A_{n+2} = \cdots$. Then apply theorem 3.

Definition. *A probability P is a function that assigns to every event A in $\mathcal{A}$ a finite number $P(A)$, called the probability of the event A, for which the following axioms are assumed:*

(i) $P(A) \geq 0$ for every $A \in \mathcal{A}$,

(ii) $P(\Omega) = 1$, and

(iii) for every denumerable sequence of disjoint events $A_1, A_2, \cdots A_n, \cdots$,

$$P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n).$$

It should be stressed that the above system of axioms for a probability is assumed to hold for any game or experiment, whether the elementary events are equally likely or not. From the above axioms there are some immediate and useful consequences, which we now prove.

Theorem 4. $P(\emptyset) = 0$.

Proof: Let us define $A_n = \emptyset$ for $n = 1, 2, \cdots$. Clearly the denumerable sequence of events $\{A_n\}$ is a sequence of disjoint events whose union is $\emptyset$. Thus, by (iii) above,

$$P(\emptyset) = P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} P(\emptyset).$$

In order for this equation to hold, the only finite value that $P(\emptyset)$ can have is zero.

Theorem 5. *If $A_1, A_2, \cdots, A_n$ are any n disjoint events, then* $P(\cup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k)$.

Proof: Let us define $A_j = \emptyset$ for all $j > n$. Then the denumerable sequence $A_1, A_2, \cdots, A_n, \cdots$ is a denumerable sequence of disjoint events. By (iii) and theorem 4,

$$P(\cup_{i=1}^{n} A_i) = P(\cup_{i=1}^{\infty} A_i) = \sum_{n=1}^{\infty} P(A_n) = \sum_{i=1}^{n} P(A_i),$$

which proves the theorem.

Theorem 6. *For every two events A and B,*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Proof: Since $A \cup B = A \cup (A^c \cap B)$ (by proposition 13 in section 1.3), then by theorem 5 we have $P(A \cup B) = P(A) + P(A^c \cap B)$. But by problem 8 of section 1.1, $B = (A \cap B) \cup (A^c \cap B)$, so $P(B) = P(A \cap B) + P(A^c \cap B)$, or $P(A^c \cap B) = P(B) - P(A \cap B)$. Thus $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, which was to be proved.

First Corollary to Theorem 6. (Boole's Inequality) *If A and B are events, then $P(A \cup B) \leq P(A) + P(B)$.*

Proof: This easily follows from theorem 6 and the fact that $P(A \cap B) \geq 0$.

Second Corollary to Theorem 6. *If A, B and C are events, then*

$$P(A \cup B \cup C) \leq P(A) + P(B) + P(C).$$

Proof: We need only apply theorem 6 twice to obtain

$$P(A \cup B \cup C) \leq P(A) + P(B \cup C) \leq P(A) + P(B) + P(C).$$

Theorem 7. *If A and B are events, and if $A \subset B$, then $P(A) \leq P(B)$.*

Proof: Since $B = (B \cap A) \cup (B \cap A^c)$ and since $B \cap A = A$, we have $B = A \cup (B \cap A^c)$. Since A and $B \cap A^c$ are disjoint, we have $P(B) = P(A) + P(B \cap A^c)$. The conclusion follows by noting that $P(B \cap A^c) \geq 0$.

Corollary to Theorem 7. *For every event A, $P(A) \leq 1$.*

Proof: Since every event E satisfies $E \subset \Omega$, the conclusion follows from theorem 7 and the fact that $P(\Omega) = 1$.

Theorem 8. *For every event A, $P(A^c) = 1 - P(A)$ and $P(A) = 1 - P(A^c)$.*

Proof: Since $A \cup A^c = \Omega$, we have $1 = P(\Omega) = P(A) + P(A^c)$, from which the two conclusions follow.

The preceding eight theorems and their corollaries are all very important and will be used repeatedly in the subsequent development.

EXERCISES

1. Prove by mathematical induction: if $A_1, A_2, \cdots, A_n$ are any n events, then $P(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$.

2. Prove: if A, B and C are events, then

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

3. A fair coin is tossed until the first head appears.

(i) What are all the elementary events? (Hint: one of the elementary events is (T, T, T, H).)

(ii) In part (i), are all the elementary events equally likely?

(iii) Let H_n denote the (elementary) event: the first head occurs on the nth trial. Show that $P(H_n) = 1/2^n$.

(iv) Let H denote the event: the tossing eventually stops. Show that $P(H) = \sum_{n=1}^\infty P(H_n) = 1$.

4. A fair coin is tossed until a head appears for the second time.

(i) What are the elementary events, and are they equally likely?

(ii) What elementary events are in the event A defined by: the first head occurs on the second trial?

(iii) What elementary events are in the event B defined by: the tossing stops on or before the fourth trial?

(iv) Compute $P(A)$, $P(B)$ and $P(A \cap B)$.

(v) If C is the event that at least one head occurs on or before the third trial, compute $P(C)$ and $P(B \cap C)$.

(vi) Let D denote the event that the coin comes up heads on or before the second trial, and let E denote the event that the second head occurs *after* the fourth trial. Compute $P(D)$, $P(E)$, $P(D \cap E)$ and $P(D \cup E)$.

5. Two fair dice are tossed. Compute the probabilities for the following events.

(i) the outcome is a 2 and a 3.

(ii) the outcome consists of two fives.

(iii) at least one of the dice comes up a five.

(iv) no fives appear.

6. An urn contains five white balls and three black balls. Three balls are selected at random without replacement. Compute the probabilities of the following two events:

(i) the sample contains exactly two white balls.

(ii) the sample contains at least two white balls.

7. Prove: if A and B are events, if $A \subset B$ and if $P(B) = 0$, then $P(A) = 0$.

8. Prove: if C and D are disjoint events, if E is an event, and if $E \subset C$, then E and D are disjoint.

1.5 Conditional Probability. The notion of conditional (or relative) probability arises in the following manner. One has at his disposal some information in the form of the occurrence of a certain event, H, which we shall refer to as the “cause". He wants to know, on the basis of this information, what the probability of another event, say E, is. What does such a probability mean to the practical person when he considers the probability of this subsequent event E given the occurence of a causal event (or “hypothesis") H? An interpretation is as follows. Suppose the game or experiment is repeated under identical conditions a large number of times, say, N times. Among these N repetitions of the experiment or game he would note the number N_H of times that the event H occurs. From among these N_H outcomes he would count the number $N_{E\cap H}$ of times in which the event E occurs. Then he would observe the ratio $N_{E\cap H}/N_H$. This ratio is called the *relative frequency* that E occurs among those times that H occurs. In a practical sense this ratio approximates what we would like to call the conditional probability that E occurs, given the information that H occurs. At this point one should notice that $N_{E\cap H}$ is the number of times among the N trials that both E and H occur. But

$$N_{E\cap H}/N_H = \frac{N_{E\cap H}/N}{N_H/N}.$$

As was discussed in the first section, $N_{E\cap H}/N$ approximates the probability that the event $E \cap H$ (in the relative frequency sense) and N_H/N approximates the probability of H. This consideration leads us to the following definition.

Definition. *If A and B are events, and if $P(B) > 0$, then we define the conditional probaility of A given B, $P(A|B)$, by*

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

With this formal definition we are able to obtain some useful properties of conditional probability.

Theorem 1. *If B is an event, and if $P(B) > 0$, then $P(\cdot|B)$ is a probability, i.e.,*

(i) $P(A|B) \geq 0$ for every $A \in \mathcal{A}$,

(ii) $P(\Omega|B) = 1$, and

(iii) for every denumerable sequence of disjoint events $\{A_n\}$,

$$P(\cup_{n=1}^{\infty} A_n|B) = \sum_{n=1}^{\infty} P(A_n|B).$$

Proof: (i) Since $A\cap B \in \mathcal{A}$, it follows that $P(A\cap B) \geq 0$. Dividing both sides of this inequality by $P(B)$ and applying the definition of conditional probability yields the conclusion.

(ii) $P(\Omega|B) = \frac{P(\Omega\cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$

(iii) One easily obtains

$$\begin{aligned} P(\cup_{n=1}^{\infty} A_n|B) &= P(\cup_{n=1}^{\infty} A_n \cap B)/P(B) \\ &= \{\sum_{n=1}^{\infty} P(A_n \cap B)\}/P(B) \\ &= \sum_{n=1}^{\infty} P(A_n|B), \end{aligned}$$

and thus all three assertions of the theorem are proved.

The above theorem states that every conditional probability is a probability. Thus every theorem that was proved for an ordinary probability will also be true for a conditional probability. There are three important results for conditional probability which we now prove: they are the *multiplication rule*, the *theorem of total probabilities* and *Bayes' rule*.

Theorem 2. (Multiplication Rule) *If* $A_0, A_1, \cdots, A_n$ *are* $n+1$ *events, and if* $P(A_0 \cap A_1 \cap \cdots \cap A_{n-1}) > 0$, *then*

$$P(\cap_{i=0}^n A_i) = P(A_0)P(A_1|A_0)P(A_2|A_0 \cap A_1) \cdots P(A_n|A_0 \cap A_1 \cap \cdots \cap A_{n-1}).$$

Proof: Since $\cap_{i=0}^{n-1} A_i \subset \cap_{i=0}^{n-2} A_i \subset \cdots \subset A_0 \cap A_1 \subset A_0$ and $0 < P(\cap_{i=0}^{n-1} A_i)$, we have

$$0 < P(\cap_{i=0}^{n-1} A_i) \leq P(\cap_{i=0}^{n-2} A_i) \leq \cdots \leq P(A_0 \cap A_1) \leq P(A_0),$$

and consequently all the conditional probabilities involved in the statement of the theorem can be defined. The proof of the theorem will be accomplished by use of the axiom of induction (i.e., by so-called "mathematical induction"). The axiom of induction states that if S is a set of integers, if $1 \in S$, and if for every $n \in S$, then also $n+1 \in S$, then S contains all positive integers. In order to prove the theorem, let S denote the set of values of n for which the theorem is true. Clearly $1 \in S$, since $P(A_0 \cap A_1) = P(A_0)P(A_1|A_0)$ by the very definition of conditional probability. Let n be any integer in S; this set is not empty, since we have already proved that it contains 1. Then by the definition of conditional probability,

$$P(\cap_{i=0}^{n+1} A_i) = P(\cap_{i=0}^n A_i) P(A_{n+1}| \cap_{i=0}^n A_i).$$

Since $n \in S$, then the first term in the product on the right side is

$$P(\cap_{i=0}^n A_i) = P(A_0)P(A_1|A_0)P(A_2|A_0 \cap A_1) \cdots P(A_n|A_0 \cap A_1 \cap \cdots \cap A_{n-1}).$$

Consequently $n+1 \in S$, and by the axiom of induction, S contains all the positive integers. This proves the theorem.

The multiplication rule is used for problems where one has a finite sequence of events. In such a problem, one knows the probability of an event given that all the previous events have occurred, and one wants to compute the probability that they all occur. Using the multiplication rule, one would simply multiply all the conditional probabilities of the events given the occurrence of their predecessors.

Example: In Polya's urn scheme, an urn initially contains r red balls and b black balls. At each trial a ball is selected at random, its color is noted, and it is *replaced along with* c additional balls of the same color. We shall compute the probability of obtaining a red ball in each of the first three trials. Let R_i denote the event that the outcome of the ith trial is a red ball. Then the problem is to compute $P(R_1 \cap R_2 \cap R_3)$. By the muliplication rule,

$$\begin{aligned} P(R_1 \cap R_2 \cap R_3) &= P(R_1)P(R_2|R_1)P(R_3|R_1 \cap R_2) \\ &= \frac{r}{r+b}\frac{r+c}{r+b+c}\frac{r+2c}{r+b+2c}. \end{aligned}$$

Theorem 3. (Theorem of Total Probabilities) *If* $\{H_n\}$ *is a finite or denumerable sequence of disjoint events, if* $P(H_n) > 0$ *for every* n*, and if* A *is an event that satisfies* $A \subset \cup_{n\geq 1} H_n$*, then*

$$P(A) = \sum_{n\geq 1} P(A|H_n)P(H_n).$$

Proof: Since $A \subset \cup_{n\geq 1} H_n$, and since events in the sequence of intersections $\{A \cap H_n\}$ are disjoint events, it follows that

$$\begin{aligned} P(A) &= P(A \cap (\cup_{n\geq 1} H_n)) \\ &= P(\cup_{n\geq 1}(A \cap H_n)) \\ &= \sum_{n\geq 1} P(A \cap H_n) \\ &= \sum_{n\geq 1} P(A|H_n)P(H_n), \end{aligned}$$

which completes the proof.

The theorem of total probabilities is used in those cases where an event A occurs only because of some event H_1 or H_2 or $\cdots$ which causes it. In such cases, the causes are disjoint, the conditional probability of the event, given each cause, is known, and the probability of each cause is known.

Example. A box contains n_1 tags numbered 1 and n_2 tags numbered 2. A tag is selected at random. If it is a number 1 tag, one goes to urn number 1 which contains r_1 red balls and b_1 black balls and selects a ball at random. If it is a number 2 tag, one goes to urn number 2 which contains r_2 red balls and b_2 black balls and selects a ball at random. The problem is to find the probability that the ball selected is a red ball. Accordingly, let R denote the event that a red ball is selected, let H_1 denote the event that a number 1 tag is selected, and let H_2 denote the event that a number 2 tag is selected. By the theorem of total probabilities,

$$\begin{aligned} P(R) &= P(R|H_1)P(H_1) + P(R|H_2)P(H_2) \\ &= \frac{r_1}{r_1+b_1}\frac{n_1}{n_1+n_2} + \frac{r_2}{r_2+b_2}\frac{n_2}{n_1+n_2}. \end{aligned}$$

It should be noted that this problem would be more difficult to solve if one used the “equally likely outcomes" method on it.

Theorem 4. (Bayes' Rule) *If $\{H_n\}$ is a finite or denumerable sequence of disjoint events, if $P(H_n) > 0$ for every n, if A is an event that satisfies $A \subset \cup_{n\geq 1} H_n$ and if $P(A) > 0$, then*

$$P(H_j|A) = \frac{P(A|H_j)P(H_j)}{\sum_{n\geq 1} P(A|H_n)P(H_n)}$$

for every integer j.

Proof: Using the definition of conditional probability and the theorem of total probabilities, we obtain

$$\begin{aligned} P(H_j|A) &= \frac{P(A\cap H_j)}{P(A)} \\ &= \frac{P(A|H_j)P(H_j)}{\sum_{n\geq 1} P(A|H_n)P(H_n)}, \end{aligned}$$

which proves the theorem.

Example. A box contains $n_1 + n_2 + n_3$ tags, where n_1 tags are numbered 1, n_2 tags are numbered 2 and n_3 tags are numbered 3. There are three urns, and they are numbered 1, 2 and 3. Urn number i contains r_i red balls and b_i black balls. A tag is selected at random from the box, and then a ball is selected from the urn with the same number as the tag selected. The problem is: if the ball selected is red, what is the probability that it came from urn number 2? Let R denote the event that the ball selected is red, and let H_i denote the event that the tag selected from the box is tag number i, $i = 1, 2, 3$. In this problem we are interested in computing $P(H_2|R)$. Using Bayes' Rule we obtain

$$\begin{aligned} P(H_2|R) &= \frac{P(R|H_2)P(H_2)}{\sum_{i=1}^{3} P(R|H_i)P(H_i)} \\ &= \frac{\frac{r_2}{r_2+b_2}\frac{n_2}{n_1+n_2+n_3}}{\sum_{m=1}^{3}\frac{r_m}{r_m+b_m}\frac{n_m}{n_1+n_2+n_3}}. \end{aligned}$$

EXERCISES

1. In urn number 1 there are three white balls and four black balls, and in urn number 2 there are two white balls and five black balls. A ball is drawn at random from urn number 1 and placed in urn number 2. Then a ball is drawn at random from urn number 2.

(i) Find the probability that the ball drawn from urn number 1 is black.

(ii) Compute the conditional probability that a white ball is drawn from urn 2, given that a black ball was drawn from urn number 1.

(iii) Find the probability that a black ball is drawn from urn number 1 and a white ball is drawn from urn number 2.

(iv) Compute the conditional probability that a white ball is drawn from urn number 2, given that a white ball was drawn from urn number 1.

(v) Find the probability that a white ball is drawn from urn number 2.

(vi) Find the probability of drawing a black ball from urn number 2.

(vii) Compute the conditional probability that a black ball was selected from urn number 1, given that a white ball is drawn from urn number 2.

(viii) Give relative frequency interpretations to the answers you gave in problems (iii), (iv) and (vii).

2. In Polya's urn scheme developed in this section, find the probability that

(i) a red ball is selected in both the first and second trials, and

(ii) a black ball is selected in the second trial.

3. In Polya's urn scheme developed in this section, find

(i) the conditional probability that a red ball is selected in the second trial, given that a black ball was selected in the first trial, and

(ii) the conditional probability that a black ball was selected in the first trial, given that a red ball is selected in the second trial.

4. In tossing a fair coin, find the conditional probability that a head occurs for the first time on the fifth trial, given the event that at least one head occurs in the first eight trials.

5. An urn contains four black balls and three white balls. Two balls are selected at random without replacement. Compute the probability that both are black, given that at least one is black. Give a relative frequency interpretation of what this means.

6. A boy has a penny, a nickel, a dime and a quarter in a piggy bank. He shakes them out one by one until he shakes out the quarter. Find the

probability that he shakes out all four coins.

1.6 Stochastic Independence. The notion of stochastic independence undoubtedly arose in those situations where the individual outcomes are equally likely. The following general example illustrates and introduces the subsequent mathematical development.

Let G_1 denote some game in which there are N_1 equally likely outcomes. If A is an event that might or might not occur when game G_1 is played, and if N_A is the number of equally likely ways in which A can occur when the game is played, then the probaility of the event A is N_A/N_1. Now let G_2 and G_3 denote two additional games in which there are N_2 and N_3 equally likely outcomes, respectively. Suppose that each game can be played in a manner which does not depend on the outcomes of the others. In such a situation, the occurrence or nonoccurrence of the event A does not depend on and does not influence the occurrence of an event B in game G_2, and the occurrence or nonoccurrence of either or both of events A and B does not influence and is not influenced by the occurrence of some event C in game G_3. In other words, we might say that the events A, B and C are *independent* of each other.

Now let us consider a supergame, G, which consists of first playing game G_1, then G_2, and finally G_3. Let N_B denote the number of equally likely ways in which event B can occur when playing G_2, and let N_C denote the number of equally likely ways in which C occurs when G_3 is played. Thus there are $N_1N_2N_3$ equally likely outcomes when G is played. Also observe that when game G is played, there are $N_AN_2N_3$ equally likely ways in which event A occurs in part 1 of the game; the number of equally likely ways in which B occurs and in which C occurs is $N_1N_BN_3$ and $N_1N_2N_C$ respectively. Thus

$$P(A) = \frac{N_AN_2N_3}{N_1N_2N_3} = \frac{N_A}{N_1},$$

$$P(B) = \frac{N_1N_BN_3}{N_1N_2N_3} = \frac{N_B}{N_2},$$

and

$$P(C) = \frac{N_1N_2N_C}{N_1N_2N_3} = \frac{N_C}{N_3}.$$

Since A, B and C can all occur in $N_AN_BN_C$ equally likely ways, it follows that

$$P(A \cap B \cap C) = \frac{N_AN_BN_C}{N_1N_2N_3} = \frac{N_A}{N_1}\frac{N_B}{N_2}\frac{N_C}{N_3}.$$

Thus

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

Similarly one obtains $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, and $P(B \cap C) = P(B)P(C)$. In other words, if three events occur independently of each other in this "equally likely" situation, we shall wish to write that the probability of the joint occurrence of any number of them is equal to the product of their individual probabilities.

Now we begin a formal treatment.

Definition. Let $\mathcal{C}$ denote a collection of events. The events in $\mathcal{C}$ are said to be *independent* if the probability of the joint occurrence of any finite number of them is equal to the product of their probabilities.

Another way of stating that the events in $\mathcal{C}$ are independent is: for every integer $n \geq 2$ (where n is equal to or less than the number of events in $\mathcal{C}$ in case $\mathcal{C}$ is finite) and every set of distinct events $A_1, A_2, \cdots, A_n$ in $\mathcal{C}$, the equation

$$P(A_1 \cap A_2 \cap \cdots A_n) = P(A_1)P(A_2) \cdots P(A_n)$$

is satisfied. If $\mathcal{C}$ consists of only two events, A and B, then these events are said to be independent if and only if $P(A \cap B) = P(A)P(B)$. If $\mathcal{C}$ consists of four distinct events, A, B, C and D, then these events satisfy the definition of independence if and only if the probability of the joint occurrence of every pair of them, that of every triple of them and that of all four equals the corresponding product of probabilities. If $\mathcal{C}$ contains only a finite number of events, then one can easily list the finite number of equations that must be satisfied in order that these events are independent.

We now obtain some consequences of the definition of independence.

Lemma 1. *If $B, A_1, \cdots, A_n$ are independent events, then $B^c, A_1, \cdots, A_n$ are also independent events.*

Proof: Let $1 \leq i_1 < \cdots < i_k \leq \dot{n}$ be integers, with $1 \leq k \leq n$. It is clearly sufficient to prove that

$$P(B^c \cap A_{i_1} \cap \cdots \cap A_{i_k}) = P(B^c)P(A_{i_1}) \cdots P(A_{i_k}).$$

We observe that

$$\begin{aligned}
\textstyle\prod_{j=1}^k P(A_{i_j}) &= P\left(\textstyle\bigcap_{j=1}^k A_{i_j}\right) \\
&= P\left(B \cap \textstyle\bigcap_{j=1}^k A_{i_j}\right) + P\left(B^c \cap \textstyle\bigcap_{j=1}^k A_{i_j}\right) \\
&= P(B) \textstyle\prod_{j=1}^k P(A_{i_j}) + P\left(B^c \cap \textstyle\bigcap_{j=1}^k A_{i_j}\right).
\end{aligned}$$

Thus

$$\begin{aligned} P\left(B^c \cap \bigcap_{j=1}^k A_{i_j}\right) &= (1 - P(B)) \prod_{j=1}^k P(A_{i_j}) \\ &= P(B^c) \prod_{j=1}^k P(A_{i_j}), \end{aligned}$$

which proves the lemma.

Theorem 1. *Let $\mathcal{C}$ be a nonempty class of independent events. If any or all of the events in $\mathcal{C}$ are replaced by their negations, then the events in the resulting class are also independent.*

Proof: Let $\mathcal{D}$ be the same as $\mathcal{C}$ except that some of the events A in $\mathcal{C}$ have been replaced by their complements A^c. Then take any finite subset $\mathcal{U}$ of $\mathcal{D}$ which in turn corresponds to a subset $\mathcal{V}$ of $\mathcal{C}$ except that some of the elements of $\mathcal{U}$ are complements of the corresponding elements of $\mathcal{V}$. Then, one at a time, replace each of these corresponding elements of $\mathcal{V}$ by its complement, noting that by lemma 1 the resulting events are independent; in a finite number of steps, the resulting set of events becomes $\mathcal{U}$. Thus the events of $\mathcal{U}$ are independent.

The connection between independence and conditional probability is given by the following theorem.

Theorem 2. *Let A and B be events, and suppose $P(B) > 0$. Then A and B are independent if and only if $P(A|B) = P(A)$.*

Proof: We first show that the condition is necessary. Assuming that A and B are independent, we obtain

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

In order to show that the condition is sufficient, we assume that $P(A|B) = P(A)$ and obtain

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

At this point one might inquire whether we required too much in our definition of independence. In other words, the question that might arise is whether we could not have satisfied the requirements specified by this definition by simply requiring that the probability of the joint occurrence of every pair of events equal the product of their probabilities. However, this is not enough, as can be shown with the following example due to S. N. Bernstein. In his example the sure event contains four equally likely elementary events. However, three events can be defined such that the probability of the joint occurrence each of the three pairs of them is equal to the product of the probabilities of the individual events, but the probability of the joint occurrence of all three events is not. Here is the example due to S. N. Bernstein. Suppose Ω consists only of four equally likely elementary events: $\omega_1, \omega_2, \omega_3$ and ω_4. Let $A = \{\omega_1, \omega_2\}$, let $B = \{\omega_1, \omega_3\}$ and let $C = \{\omega_1, \omega_4\}$. Then $P(A) = 1/2$, $P(B) = 1/2$ and $P(C) = 1/2$. Since $A \cap B = A \cap C = B \cap C = \{\omega_1\}$, we have $P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4} = P(A)P(B) = P(A)P(C) = P(B)P(C)$. However $P(A \cap B \cap C) = P(\{\omega_1\}) \neq P(A)P(B)P(C) = \frac{1}{8}$. Thus in our definition of independence of a collection of events we must state that the probability of the joint occurrence of every finite collection of events equals the product of their individual probabilities.

EXERCISES

1. Prove: if $\mathcal{C}$ is a collection of events, these events are independent if and only if the events of every finite subset are independent.

2. Prove: if A and B are events, and if $P(A) = 0$ or 1, then A and B are independent.

3. If $\mathcal{C}$ consists of events A, B, C and D, how many equations must they be required to satisfy in order that they be independent?

4. If $\mathcal{C}$ consists of events $\{A_1, A_2, \cdots, A_n\}$, how many equations must they be required to satisfy in order for them to be independent?

5. If A, B and C are independent events, prove that A and $B \cap C$ are independent and A and $B \cup C$ are independent.

6. A fair die is tossed twice. Let A denote the event that the outcome on the first toss is even, and let B denote the event that the outcome on the second toss is greater than 4. Show that A and B are independent events.

7. Show that if two events have positive probabilities and are independent they are never disjoint. (N.B. A common error among students is to confuse independence with disjointness.)

8. In this Polya urn scheme, the urn initially contains five black balls and two red balls. At each trial, a ball is selected at random from the urn, its color is noted, and it is replaced in the urn *along with* two balls of the same color. Let R_j denote the event that a red ball is selected in the jth trial, and let B_j denote the event that a black ball is selected in the jth trial.

(i) Compute $P(B_1)$, $P(B_2)$ and $P(B_2|B_1)$.

(ii) Determine if B_1 and B_2 are independent.

(iii) Determine if B_1 and B_3 are independent.

9. An urn contains four white balls and three black balls. A ball is selected at random, and it is replaced. (This is sampling with replacement.) Let W_j denote the event that a white ball is selected at the jth trial, and let B_j denote the event that a black ball is selected at the jth trial.Suppose that the sampling occurs three times.

(i) Compute $P(B_1)$, $P(B_2)$, $P(B_3)$ and $P(B_2 \cap B_3)$.

(ii) Determine if the events B_1, B_2 and B_3 are independent.

(iii) Determine if B_1, W_2 and W_3 are independent.

10. An urn contains four white balls and three black balls. At each trial, a ball is selected at random, its color is noted, but it is not replaced. (This is sampling without replacement.) Let W_j and B_j be defined as in exercise 9. Suppose that the sampling occurs three times.

(i) Compute $P(B_1)$, $P(B_2)$, $P(B_3)$ and $P(B_1 \cap B_2)$.

(ii) Determine if $\{B_1, B_2, B_3\}$ are independent.

(iii) Determine if $\{B_1, W_2, W_3\}$ are independent.