Conditional Probability and Independence

P. Sam Johnson

NITK, Surathkal, India

Let us reconsider the difference between choosing an item at random from a lot **with or without replacement**. Consider a lot containing one hundred items, 20 defective and 80 non-defective items. Suppose that we choose two items from this lot, (a) with replacement; (b) without replacement. We define the following two events.

$A = \{\text{the first item is defective}\}$, $B = \{\text{the second item is defective}\}$.

If we are choosing *with* replacement, $P(A) = P(B) = \frac{20}{100} = \frac{1}{5}$. For each time we choose from the lot there are 20 defective items among the total of 100. However, if we are choosing *without* replacement, the results are not quite immediate. It is still true, of course, that $P(A) = \frac{1}{5}$. But what about $P(B)$? It is clear that in order to compute $P(B)$ we should know the composition of the lot *at the time the second item is chosen*. That is, we should know whether A did or did not occur. This example indicates the need to introduce the important concept "Conditional Probability".

Introduction

The importance of conditional probability is twofold. In the first place, we are often interested in calculating probabilities when some partial information concerning the result of an experiment is available; in such a situation, the desired probabilities are conditional.

Second, even when no partial information is available, conditional probabilities can often be used to compute the desired probabilities more easily.

Let A and B be two events associated with an experiment E. We denote by $P(B|A)$ the *conditional probability* of the event B, given that A has occurred.

In the above example, $P(B|A) = \frac{19}{99}$. For, if A has occurred, then on the second drawing there are only 99 items left, 19 of which are defective. Whenever we compute $P(B|A)$ we are essentially computing $P(B)$ with respect to the **reduced sample space A**, rather than with respect to the original sample space S.

When we evaluate $P(B)$ we are asking ourselves how probable it is that we shall be in B, knowing that we must be in S. And when we compute $P(B|A)$ we are asking ourselves how probable it is that we shall be in B, knowing that we must be in A. (That is, the sample space has been *reduced* from S to A.)

Example 1.

Two fair dice are tossed, the outcome being recorded as (x_1, x_2), *where* x_i *is the outcome of the* i*th die,* $i = 1, 2$. *Hence the sample space* S *may be represented by the following array of* 36 *equally likely outcomes.*

$$S = \left\{ \begin{matrix} (1,1) & (1,2) & \cdots & (1,6) \\ (2,1) & (2,2) & \cdots & (2,6) \\ \vdots & & & \vdots \\ (6,1) & (6,2) & \cdots & (6,6) \end{matrix} \right\}.$$

Consider the following two events:

$$A = \{(x_1, x_2) : x_1 + x_2 = 10\}, \quad B = \{(x_1, x_2) : x_1 > x_2\}.$$

Thus $A = \{(5,5), (4,6), (6,4)\}$, *and* $B = \{(2,1), (3,1), (3,2), \ldots, (6,5)\}$.

Example (contd...)

Hence $P(A) = \frac{3}{36}$ and $P(B) = \frac{15}{36}$. And $P(B|A) = \frac{1}{3}$, since the sample space now consists of A (that is, three outcomes), and only one of these three outcomes is consistent with the event B. In a similar way, we may compute $P(A|B) = \frac{1}{15}$.
Finally, let us compute $P(A \cap B)$. The event $A \cap B$ occurs if and only if the sum of the two dice is 10 and the first die shows a larger value than the second die. There is only *one* such outcome, and hence $P(A \cap B) = \frac{1}{36}$. If we take a long careful look at the various numbers we have computed above, we note the following:

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

These relationships did not just happen to arise in the particular example we considered. Rather, they are quite general and give us a means of *formally defining* conditional probability.

Motivation

To motivate this definition, let us return to the concept of relative frequency. Suppose that an experiment E has been repeated n times. Let n_A, n_B, and $n_{A\cap B}$ be the number of times the events A, B, and $A \cap B$, respectively, have occurred among the n repetitions. What is the meaning of $n_{A\cap B}/n_A$? It represents the relative frequency of B among those outcomes in which A occurred. That is, $n_{A\cap B}/n_A$ is the conditional relative frequency of B, given that A occurred.

We may write $n_{A\cap B}/n_A$ as follows:

$$\frac{n_{A\cap B}}{n_A} = \frac{n_{A\cap B}/n}{n_A/n} = \frac{f_{A\cap B}}{f_A},$$

where $f_{A\cap B}$ and f_A are the relative frequencies of the events $A \cap B$ and A, respectively. As we have already indicated, if n, the number of repetitions is large, $f_{A\cap B}$ will be close to $P(A \cap B)$ and f_A will be close to $P(A)$. Hence the above relation suggests that $n_{A\cap B}/n_A$ will be close to $P(B|A)$.

Thus we make the following formal definition.

Definition 2.

$P(B|A) = \frac{P(A \cap B)}{P(A)}$, *provided that* $P(A) > 0$.

Verify that $P(B|A)$, for fixed A, satisfies the various postulates of probability.

(1') $0 \leq P(B|A) \leq 1$,

(2') $P(S|A) = 1$,

(3') $P(B_1 \cup B_2|A) = P(B_1|A) + P(B_2|A) \quad$ if $B_1 \cap B_2 = \emptyset$,

(4') $P(B_1 \cup B_2 \cup \cdots |A) = P(B_1|A) + P(B_2|A) + \cdots \quad$ if $B_i \cap B_j = \emptyset$ for $i \neq j$.

Conditional Probability

(a) If $A = S$, $P(B|S) = P(B \cap S)/P(S) = P(B)$.

(b) With every event $B \subset S$ we can associate two numbers, $P(B)$, the **(unconditional) probability** of B, and $P(B|A)$, the **conditional probability** of B, given that some event A (for which $P(A) > 0$) has occurred. In general, these two probability measures will assign different probabilities to the event B. We shall study an important special case for which $P(B)$ and $P(B|A)$ are the same.

(c) Observe that the conditional probability is defined in terms of the unconditional probability measure P. That is, if we know $P(B)$ for every $B \subset S$, we can compute $P(B|A)$ for every $B \subset S$.

Conditional Probability

Thus we have two ways of computing the conditional probability $P(B|A)$:

(a) Directly, by considering the probability of B with respect to the reduced sample space A.

(b) Using the above definition, where $P(A \cap B)$ and $P(A)$ are computed with respect to the original sample space S.

If $A = S$, we obtain $P(B|S) = P(B \cap S)/P(S) = P(B)$, since $P(S) = 1$ and $B \cap S = B$.

Example

Example 3.

Suppose that we toss 2 *dice, and suppose that each of the* 36 *possible outcomes is equally likely to occur and hence has probability* $\frac{1}{36}$ *. Suppose further that we observe that the first die is a* 3*. Then, given this information, what is the probability that the sum of the* 2 *dice equals* 8*?*
To calculate this probability, we reason as follows: Given that the initial die is a 3*, there can be at most* 6 *possible outcomes of our experiment, namely,* $(3,1),(3,2),(3,3),(3,4),(3,5)$*, and* $(3,6)$*. Since each of these outcomes originally had the same probability of occurring, the outcomes should still have equal probabilities. That is, given that the first die is a* 3*, the (conditional) probability of each of the outcomes* $(3,1),(3,2),(3,3),(3,4),(3,5)$*, and* $(3,6)$ *is* $\frac{1}{6}$*, whereas the (conditional) probability of the other* 30 *points in the sample space is* 0*. Hence, the desired probability will be* $\frac{1}{6}$.

Conditional Probability

If we let E and F denote, respectively, the event that the sum of the dice is 8 and the event that the first die is a 3, then the probability just obtained is called the *conditional probability* that E occurs given that F has occurred and is denoted by

$$P(E|F).$$

Conditional Probability

A general formula for $P(E|F)$ that is valid for all events E and F is derived in the same manner: If the event F occurs, then, in order for E to occur, it is necessary that the actual occurrence be a point both in E and in F; that is, it must be in $E \cap F$. Now, since we know that F has occurred, it follows that F becomes our new, or reduced, sample space; hence, the probability that the event $E \cap F$ occurs will equal the probability of $E \cap F$ relative to the probability of F.

That is,

$$\text{if } P(F) > 0, \text{ then } P(E|F) = \frac{P(E \cap F)}{P(F)}.$$

Example 4.

Suppose that an office has 100 calculating machines. Some of these machines are electric (E) while others are manual (M). And some of the machines are new (N) while others are used (U). The following table gives the number of machines in each category. A person enters the office, picks a machine at random, and discovers that it is new. What is the probability that it is electric?

	E	M	
N	40	30	70
U	20	10	30
	60	40	100

Example (contd...)

Simply considering the reduced sample space N (e.g., the 70 new machines), we have $P(E|N) = \frac{40}{70} = \frac{4}{7}$. Using the definition of conditional probability, we have that

$$P(E|N) = \frac{P(E \cap N)}{P(N)} = \frac{40/100}{70/100} = \frac{4}{7}.$$

Multiplication Theorem

The most important consequence of the above definition of conditional probability is obtained by writing it in the following form:

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

or, equivalently,

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

This is sometimes known as the **multiplication theorem** of probability. We may apply this theorem to compute the probability of the simultaneous occurrence of two events A and B.

Example

Example 5.

Consider a lot consisting of 20 *defective and* 80 *nondefective items. If we choose two items at random,* without replacement, *what is the probability that both items are defective?*
As before, we define the events A and B as follows:

$$A = \{\textit{the first item is defective}\}, \quad B = \{\textit{the second item is defective}\}.$$

Hence we require $P(A \cap B)$*, which we may compute, according to the above formula, as* $P(B|A)P(A)$*. But* $P(B|A) = \frac{19}{99}$*, while* $P(A) = \frac{1}{5}$*. Hence*

$$P(A \cap B) = \frac{19}{495}.$$

The above multiplication theorem may be generalized to more than two events in the following way:

$$P[A_1 \cap A_2 \cap \cdots \cap A_n] = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2) \cdots P(A_n|A_1, \ldots, A_{n-1}).$$

Conditional Probability

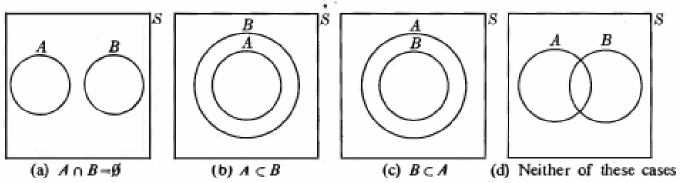
Let us consider for a moment whether we can make a general statement about the **relative magnitude** of $P(A|B)$ and $P(A)$. We shall consider four cases, which are illustrated by the Venn diagrams in the following figures. We have

(a) $P(A|B) = 0 \leq P(A)$, since A cannot occur if B has occurred.

(b) $P(A|B) = P(A \cap B)/P(B) = [P(A)/P(B)] \geq P(A)$, since $0 \leq P(B) \leq 1$.

(c) $P(A|B) = P(A \cap B)/P(B) = P(B)/P(B) = 1 \geq P(A)$.

(d) In this case we cannot make any statement about the relative magnitude of $P(A|B)$ and $P(A)$.



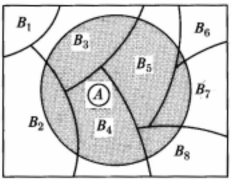
Note that in two of the above cases, $P(A) \leq P(A|B)$; in one case, $P(A) \geq P(A|B)$, and in the fourth case, we cannot make any comparison at all. We used the concept of conditional probability in order to evaluate the probability of the simultaneous occurrence of two events. We can apply this concept in another way to compute the probability of a single event A.

We need the following definition.

Definition 6.

We say that the events $B_1, B_2, \ldots, B_k$ represent a partition of the sample space S if

(a) $B_i \cap B_j = \emptyset$ *for all* $i \neq j$ (b) $\cup_{i=1}^{k} B_i = S$ (c) $P(B_i) > 0$ *for all* i.



When the experiment E is performed one and only one of the events B_i occurs.

For example, for the tossing of a die $B_1 = \{1, 2\}$, $B_2 = \{3, 4, 5\}$, and $B_3 = \{6\}$ would represent a partition of the sample space, while $C_1 = \{1, 2, 3, 4\}$ and $C_2 = \{4, 5, 6\}$ would not.

Let A be some event with respect to S and let $B_1, B_2, \ldots, B_k$ be a partition of S. The Venn diagram in the above figure illustrates this for $k = 8$. Hence we may write

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k).$$

Theorem on Total Probability

Of course, some of the sets $A \cap B_j$ may be empty, but this does not invalidate the above decomposition of A. The important point is that all the events $A \cap B_1, \ldots, A \cap B_k$ are pairwise mutually exclusive. Hence we may apply the addition property for mutually exclusive events and write

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_k).$$

However each term $P(A \cap B_j)$ may be expressed as $P(A|B_j)P(B_j)$ and hence we obtain what is called the **theorem on total probability**:

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_k)P(B_k).$$

This result represents an extremely useful relationship. For often when $P(A)$ is required, it may be difficult to compute it directly. However, with the additional information that B_j has occurred, we may be able to evaluate $P(A|B_j)$ and then use the above formula.

Example 7.

Consider the lot of 20 defective and 80 nondefective items from which we choose two items without replacement. Again defining A and B as

$$A = \{\textit{the first chosen item is defective}\},$$
$$B = \{\textit{the second chosen item is defective}\},$$

we may now compute $P(B)$ as follows:

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

We find that

$$P(B) = \frac{19}{99} \cdot \frac{1}{5} + \frac{20}{99} \cdot \frac{4}{5} = \frac{1}{5}.$$

Example 8.

A certain item is manufactured by three factories, say 1, 2, and 3. It is known that 1 turns out twice as many items as 2, and that 2 and 3 turn out the same number of items (during a specified production period). It is also known that 2 percent of the items produced by 1 and by 2 are defective, while 4 percent of those manufactured by 3 are defective. All the items produced are put into one stockpile, and then one item is chosen at random. What is the probability that this item is defective?

Solution: Let us introduce the following events:

$$A = \{\text{the item is defective}\}, \qquad B_1 = \{\text{the item came from } 1\},$$
$$B_2 = \{\text{the item came from } 2\}, \qquad B_3 = \{\text{the item came from } 3\}.$$

We require $P(A)$ and, using the above result, we may write

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3).$$

Now $P(B_1) = \frac{1}{2}$, while $P(B_2) = P(B_3) = \frac{1}{4}$. Also $P(A|B_1) = P(A|B_2) = 0.02$, while $P(A|B_3) = 0.04$. Inserting these values into the above expression, we obtain $P(A) = 0.025$.

Theorem on Total Probability

The following analogy to the theorem on total probability has been observed in chemistry: Suppose that we have k beakers in laboratory equipment, a beaker is generally a cylindrical container containing different solutions of the same salt, totaling, say one liter.

Let $P(B_i)$ be the volume of the ith beaker and let $P(A|B_i)$ be the concentration of the solution in the ith beaker. If we combine all the solutions into one beaker and let $P(A)$ denote the concentration of the resulting solution, we obtain,

$$P(A) = P(A|B_1)P(B_1) + \cdots + P(A|B_k)P(B_k).$$

Bayes' Theorem

In the Example (8), suppose that one item is chosen from the stockpile and is found to be defective. What is the probability that it was produced in factory 1?
Using the notation introduced previously, we require $P(B_1|A)$. We can evaluate this probability as a consequence of the following discussion. Let $B_1, \ldots, B_k$ be a partition of the sample space S and let A be an event associated with S. Applying the definition of conditional probability, we may write

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{k} P(A|B_j)P(B_j)} \quad i = 1, 2, \ldots, k.$$

This result is known as Bayes' theorem.

It is also called the formula for the probability of "causes". Since the B_i's are a partition of the sample space, one and only one of the events B_i occurs. (That is, one of the events B_i must occur and only one can occur.) Hence the above formula gives us the probability of a particular B_i (that is, a "cause"), given that the event A has occurred. In order to apply this theorem we must know the values of the $P(B_i)$'s.

Bayes' Theorem

Quite often these values are not known, and this limits the applicability of the result.

Now applying Bayes' theorem, we obtain

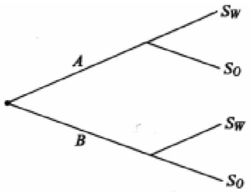
$$P(B_1|A) = \frac{(0.02)(1/2)}{(0.02)(1/2) + (0.02)(1/4) + (0.04)(1/4)} = 0.40.$$

We can again find an analogy, from Chemistry, to Bayes' theorem. In k beakers we have solutions of the same salt, but of different concentrations. Suppose that the total volume of the solutions is one liter. Denoting the volume of the solution in the ith beaker by $P(B_i)$ and denoting the concentration of the salt in the ith beaker by $P(A|B_i)$, we find that Bayes' theorem yields the proportion of the entire amount of salt which is found in the ith beaker.

The following illustration of Bayes' theorem will give us an opportunity to introduce the idea of a tree diagram, a rather useful device for analyzing certain problems.

Example 9.

Suppose that a large number of containers of candy are made up of two types, say A and B. Type A contains 70 percent sweet and 30 percent sour ones while for type B these percentages are reversed. Furthermore, suppose that 60 percent of all candy jars are of type A while the remainder are of type B.
You are now confronted with the following decision problem. A jar of unknown type is given to you. You are allowed to sample one piece of candy, and with this information, you must decide whether to guess that type A or type B has been offered to you. The following "tree diagram" (so called because of the various paths or branches which appear) will help us to analyze the problem. (S_W and S_O stand for choosing a sweet or sour candy, respectively.)



Bayes' Theorem

Let us make a few computations:

$$P(A) = 0.6;\ P(B) = 0.4;\ P(S_W|A) = 0.7;$$
$$P(S_O|A) = 0.3;\ P(S_W|B) = 0.3;\ P(S_O|B) = 0.7.$$

What we really wish to know is $P(A|S_W)$, $P(A|S_O)$, $P(B|S_W)$, and $P(B|S_O)$. That is, suppose we actually pick a sweet piece of candy. What decision would we be most tempted to make? Let us compare $P(A|S_W)$ and $P(B|S_W)$. Using Bayes' formula we have

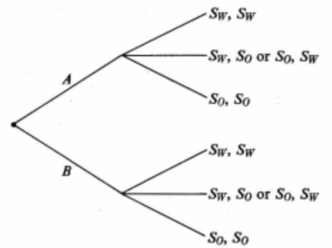
$$P(A|S_W) = \frac{P(S_W|A)P(A)}{P(S_W|A)P(A) + P(S_W|B)P(B)} = \frac{(0.7)(0.6)}{(0.7)(0.6) + (0.3)(0.4)} = \frac{7}{9}.$$

A similar computation yields $P(B|S_W) = 2/9$.
Thus, based on the evidence we have (i.e., the obtaining of a sweet candy) it is $2\frac{1}{2}$ times as likely that we are dealing with a container of type A rather than one of type B. Hence we would, presumably, decide that a container of type A was involved. The point of the above analysis is that we are choosing that alternative which appears most likely based on the limited evidence we have.

Bayes' Theorem

In terms of the tree diagram, what was really required (and done) in the preceding calculation was a "backward" analysis. That is, given what we observed, S_W in this case, how probable is it that type A is involved?

A somewhat more interesting situation arises if we are allowed to choose two pieces of candy before deciding whether type A or type B is involved. In this case, the tree diagram would appear as follows.



Example 10.

A student is taking a one-hour-time-limit makeup examination. Suppose the probability that the student will finish the exam in less than x hours is $x/2$, for all $0 \leq x \leq 1$. Then, given that the student is still working after .75 hour, what is the conditional probability that the full hour is used?

Solution. Let L_x denote the event that the student finishes the exam in less than x hours, $0 \leq x \leq 1$, and let F be the event that the student uses the full hour. Because F is the event that the student is not finished in less than 1 hour, $P(F) = P(L_1^c) = 1 - P(L_1) = 0.5$.

Now, the event that the student is still working at time .75 is the complement of the event $L_{.75}$, so the desired probability is obtained from

$$\begin{aligned} P(F|L_{.75}^c) &= \frac{P(FL_{.75}^c)}{P(L_{.75}^c)} \\ &= \frac{P(F)}{1 - P(L_{.75})} \\ &= \frac{0.5}{0.625} = 0.8. \end{aligned}$$

If each outcome of a finite sample space S is equally likely, then, conditional on the event that the outcome lies in a subset $F \subset S$, all outcomes in F become equally likely.

In such cases, it is often convenient to compute conditional probabilities of the form $P(E|F)$ by using F as the sample space. Indeed, working with this reduced sample space often results in an easier and better understood solution.

Our next few examples illustrate this point.

Example 11.

A coin is flipped twice. Assuming that all four points in the sample space $S = \{(h, h), (h, t), (t, h), (t, t)\}$ are equally likely, what is the conditional probability that both flips land on heads, given that (a) the first flip lands on heads? (b) at least one flip lands on heads?

Solution.

Let $B = \{(h,h)\}$ be the event that both flips land on heads; let $F = \{(h,h),(h,t)\}$ be the event that the first flip lands on heads; and let $A = \{(h,h),(h,t),(t,h)\}$ be the event that at least one flip lands on heads. The probability for (a) can be obtained from

$$P(B|F) = \frac{P(BF)}{P(F)}$$
$$= \frac{P(\{(h,h)\})}{P(\{(h,h),(h,t)\})}$$
$$= \frac{1/4}{2/4} = 1/2.$$

For (b), we have

$$P(B|F) = \frac{P(BA)}{P(A)}$$
$$= \frac{P(\{(h,h)\})}{P(\{(h,h),(h,t),(t,h)\})}$$
$$= \frac{1/4}{3/4} = 1/3.$$

Example (contd...)

Thus, the conditional probability that both flips land on heads given that the first one does is $1/2$, whereas the conditional probability that both flips land on heads given that at least one does is only $1/3$. Many students initially find this latter result surprising. They reason that, given that at least one flip lands on heads, there are two possible results: Either they both land on heads or only one does. Their mistake, however, is in assuming that these two possibilities are equally likely. For, initially, there are 4 equally likely outcomes. Because the information that at least one flip lands on heads is equivalent to the information that the outcome is not (t, t), we are left with the 3 equally likely outcomes $(h, h), (h, t), (t, h)$, only one of which results in both flips landing on heads.

Example 12.

In the card game bridge, the 52 *cards are dealt out equally to* 4 *players - called East, West, North, and South. If North and South have a total of* 8 *spades among them, what is the probability that East has* 3 *of the remaining* 5 *spades?*

Solution. Probably the easiest way to compute the desired probability is to work with the reduced sample space. That is, given that North-South have a total of 8 spades among their 26 cards, there remains a total of 26 cards, exactly 5 of them being spades, to be distributed among the East-West hands. Since each distribution is equally likely, it follows that the conditional probability that East will have exactly 3 spades among his or her 13 cards is

$$\frac{\binom{5}{3}\binom{21}{10}}{\binom{26}{13}} \approx 0.339.$$

Example

Example 13.

A total of n balls are sequentially and randomly chosen, without replacement, from an urn containing r red and b blue balls ($n \leq r + b$). Given that k of the n balls are blue, what is the conditional probability that the first ball chosen is blue?

Solution. If we imagine that the balls are numbered, with the blue balls having numbers 1 through b and the red balls $b + 1$ through $b + r$, then the outcome of the experiment of selecting n balls without replacement is a vector of distinct integers $x_1 \ldots, x_n$ where each x_i is between 1 and $r + b$. Moreover, each such vector is equally likely to be the outcome. So, given that the vector contains k blue balls (that is, it contains k values between 1 and b), it follows that each of these outcomes is equally likely. But because the first ball chosen is, therefore, equally likely to be any of the n chosen balls, of which k are blue, it follows that the desired probability is k/n.

Example (contd...)

If we did not choose to work with the reduced sample space, we could have solved the problem by letting B **be the event that the first ball chosen is blue and** B_k **be the event that a total of** k **blue balls are chosen**. Then

$$P(B|B_k) = \frac{P(BB_k)}{P(B_k)} = \frac{P(B_k|B)P(B)}{P(B_k)}.$$

Now, $P(B_k|B)$ is the probability that a random choice of $n-1$ balls from an urn containing r red and $b-1$ blue balls results in a total of $k-1$ blue balls being chosen; consequently,

$$P(B_k|B) = \frac{\binom{b-1}{k-1}\binom{r}{n-k}}{\binom{r+b-1}{n-1}}.$$

Using the preceding formula along with $P(B) = \frac{b}{(r+b)}$ and the hypergeometric probability

$$P(B_k) = \frac{\binom{b}{k}\binom{r}{n-k}}{\binom{r+b}{n}}$$

again yields the result that

$$P(B|B_k) = \frac{k}{n}.$$

Example 14.

Celine is undecided as to whether to take a French course or a chemistry course. She estimates that her probability of receiving an A grade would be $\frac{1}{2}$ in a French course and $\frac{2}{3}$ in a chemistry course. If Celine decides to base her decision on the flip of a fair coin, what is the probability that she gets an A in chemistry?

Solution. Let C be the event that Celine takes chemistry and A denote the event that she receives an A in whatever course she takes, then the desired probability is $P(CA)$, which is calculated as follows:

$$\begin{aligned} P(CA) &= P(C)P(A|C) \\ &= \left(\frac{1}{2}\right)\left(\frac{2}{3}\right) = \frac{1}{3}. \end{aligned}$$

Example 15.

Suppose that an urn contains 8 red balls and 4 white balls. We draw 2 balls from the urn without replacement. (a) If we assume that at each draw each ball in the urn is equally likely to be chosen, what is the probability that both balls drawn are red? (b) Now suppose that the balls have different weights, with each red ball having weight r and each white ball having weight w. Suppose that the probability that a given ball in the urn is the next one selected *is its weight divided by the sum of the weights of all balls currently in the urn. Now what is the probability that both balls are red?*

Solution. Let R_1 and R_2 denote, respectively, the events that the first and second balls drawn are red. Now, given that the first ball selected is red, there are 7 remaining red balls and 4 white balls, so $P(R_2|R_1) = \frac{7}{11}$. As $P(R_1)$ is clearly $\frac{8}{12}$, the desired probability is

$$P(R_1R_2) = P(R_1)P(R_2|R_1) = \left(\frac{2}{3}\right)\left(\frac{7}{11}\right) = \frac{14}{33}.$$

Of course, this probability could have been computed by $P(R_1R_2) = \binom{8}{2} \Big/ \binom{12}{2}$.

For part (b), we again let R_i **be the event that the ith ball chosen is red** and use

$$P(R_1R_2) = P(R_1)P(R_2|R_1).$$

Example (contd...)

Now, number the red balls, and let $B_i, i = 1, \ldots, 8$ be the event that the first ball drawn is red ball number i. Then

$$P(R_1) = P(\cup_{i=1}^{8} B_i) = \sum_{i=1}^{8} P(B_i) = 8\frac{r}{8r + 4w}.$$

Moreover, given that the first ball is red, the urn then contains 7 red and 4 white balls. Thus, by an argument similar to the preceding one,

$$P(R_2|R_1) = \frac{7r}{7r + 4w}.$$

Hence, the probability that both balls are red is

$$P(R_2R_1) = \frac{8r}{8r + 4w}\frac{7r}{7r + 4w}.$$

Example 16.

Suppose that each of N men at a party throws his hat into the center of the room. The hats are first mixed up, and then each man randomly selects a hat. What is the probability that none of the men selects his own hat?

Solution. We first calculate the complementary probability of at least one man's selecting his own hat. Let us denote by $E_i, i = 1, 2, \ldots, N$ the event that the **ith man selects his own hat**.

Now, $P\left(\bigcup_{i=1}^{N} E_i\right)$, the probability that at least one of the men selects his own hat is given by

$$\begin{aligned} P\left(\bigcup_{i=1}^{N} E_i\right) &= \sum_{i=1}^{N} P(E_i) - \sum_{i_1<i_2} P(E_{i_1}E_{i_2}) + \ldots \\ &+ (-1)^{n+1} \sum_{i_1<i_2\cdots<i_n} P(E_{i_1}E_{i_2}\cdots E_{i_n}) \\ &+ \cdots + (-1)^{N+1} P(E_1E_2\cdots E_N). \end{aligned}$$

Example (contd...)

If we regard the outcome of this experiment as a vector of N numbers, where the ith element is the number of the hat drawn by the ith man, then there are $N!$ possible outcomes. [The outcome $(1, 2, 3, ..., N)$ means, for example, that each man selects his own hat.]

Furthermore, $E_{i_1} E_{i_2} \ldots, E_{i_n}$, the event that each of the n men $i_1, i_2, \ldots, i_n$ selects his own hat, can occur in any of $(N-n)(N-n-1)\cdots 3\cdot 2\cdot 1 = (N-n)!$ possible ways; for, of the remaining $N-n$ men, the first can select any of $N-n$ hats, the second can then select any of $N-n-1$ hats, and so on. Hence, assuming that all $N!$ possible outcomes are equally likely, we see that

$$P(E_{i_1} E_{i_2} \cdots E_{i_n}) = \frac{(N-n)!}{N!}.$$

Aslo, as there are $\binom{N}{n}$ terms in $\sum_{i_1 < i_2 \cdots < i_n} P(E_{i_1} E_{i_2} \cdots E_{i_n})$, it follows that

$$\sum_{i_1 < i_2 \cdots < i_n} P(E_{i_1} E_{i_2} \cdots E_{i_n}) = \frac{N!(N-n)!}{(N-n)!n!N!} = \frac{1}{n!}.$$

Example (contd...)

Thus,

$$P\left(\bigcup_{i=1}^{N} E_i\right) = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{N+1}\frac{1}{N!}.$$

Hence, the probability that none of the men selects his own hat is

$$1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^N}{N!}$$

which is approximately equal to $e^{-1} \approx .36788$ for N large. In other words, for N large, the probability that none of the men selects his own hat is approximately 0.37.

We would have thought that the probability would go to 1 **as** $N \to \infty$.

Example

A generalization of the following equation, which provides an expression for the probability of the intersection of an arbitrary number of events, is sometimes referred to as the *multiplication rule*.

Example 17.

Suppose that each of N men at a party throws his hat into the center of the room. The hats are first mixed up, and then each man randomly selects a hat. We have shown that P_N, the probability that there are no matches when N people randomly select from among their own N hats, is given by

$$P_N = \sum_{i=0}^{N} (-1)^i / i!$$

What is the probability that exactly k of the N people have matches?

Solutions. Let us fix our attention on a particular set of k people and determine the probability that these k individuals have matches and no one else does. Letting E denote the event that everyone in this set has a match, and letting G be the event that none of the other $N - k$ people have a match, we have

$$P(EG) = P(E)P(G|E).$$

Example (contd...)

Now, let $F_i, i = 1, \ldots, k$, be the **event that the ith member of the set has a match.** Then

$$\begin{aligned} P(E) &= P(F_1 F_2 \cdots F_k) \\ &= P(F_1)P(F_2|F_1)P(F_3|F_1F_2)\cdots P(F_k|F_1\cdots F_{k-1}) \\ &= \frac{1}{N}\frac{1}{N-1}\frac{1}{N-2}\cdots\frac{1}{N-k+1} = \frac{(N-K)!}{N!}. \end{aligned}$$

Given that everyone in the set of k has a match, the other $N-k$ people will be randomly choosing among their own $N-k$ hats, so the probability that none of them has a match is equal to the probability of **no matches in a problem having $N-k$ people choosing among their own $N-k$ hats.** Therefore, $P(G|E) = P_{N-K} = \sum_{i=0}^{N-K}(-1)^i/i!$ showing that the probability that a specified set of k people have matches and no one else does is

$$P(EG) = \frac{(N-K)!}{N!}P_{N-K}$$

Because there will be exactly k matches if the preceding is true for any of the $\binom{N}{k}$ sets of k individuals, the desired probability is

$$\begin{aligned} P(\text{exactly } k \text{ matches}) &= P_{N-k}/k! \\ &\approx e^{-1}/k! \quad \text{when } N \text{ is large.} \end{aligned}$$

Example 18.

An insurance company believes that people can be divided into two classes: those who are accident prone and those who are not. The company's statistics show that an accident-prone person will have an accident at some time within a fixed 1*-year period with probability* .4*, whereas this probability decreases to* .2 *for a person who is not accident prone. If we assume that* 30 *percent of the population is accident prone, what is the probability that a new policyholder will have an accident within a year of purchasing a policy?*

Solution. We shall obtain the desired probability by first conditioning upon whether or not the policyholder is accident prone. Let A_1 **denote the event that the policyholder will have an accident within a year of purchasing the policy**, and let A denote the event that the policyholder is accident prone. Hence, the desired probability is given by

$$\begin{aligned} P(A_1) &= P(A_1|A)P(A) + P(A_1|A^c)P(A^c) \\ &= (.4)(.3) + (.2)(.7) = .26 \end{aligned}$$

Example 19.

Suppose that a new policyholder has an accident within a year of purchasing a policy. What is the probability that he or she is accident prone?

Solution. The desired probability is

$$P(A|A_1) = \frac{P(AA_1)}{P(A_1)} = \frac{P(A)P(A_1|A)}{P(A_1)} = \frac{(.3)(.4)}{.26} = \frac{6}{13}.$$

Example 20.

In answering a question on a multiple-choice test, a student either knows the answer or guesses. Let p be the probability that the student knows the answer and $1 - p$ be the probability that the student guesses. Assume that a student who guesses at the answer will be correct with probability $1/m$, where m is the number of multiple-choice alternatives. What is the conditional probability that a student knew the answer to a question given that he or she answered it correctly?

Solution. Let C and K denote, respectively, the events that the student answers the question correctly and the event that he or she actually knows the answer. Now,

$$\begin{aligned} P(K|C) &= \frac{P(KC)}{P(C)} \\ &= \frac{P(C|K)P(K)}{P(C|K)P(K) + P(C|K^c)P(K^c)} \\ &= \frac{p}{p + (1/m)(1-p)} = \frac{mp}{1 + (m-1)p}. \end{aligned}$$

For example, if $m = 5, p = \frac{1}{2}$, then the probability that the student knew the answer to a question he or she answered correctly is $\frac{5}{6}$.

Example

Example 21.

Consider a medical practitioner pondering the following dilemma: "If I'm at least 80 *percent certain that my patient has this disease, then I always recommend surgery, whereas if I'm not quite as certain, then I recommend additional tests that are expensive and sometimes painful. Now, initially I was only* 60 *percent certain that Jones had the disease, so I ordered the series A test, which always gives a positive result when the patient has the disease and almost never does when he is healthy. The test result was positive, and I was all set to recommend surgery when Jones informed me, for the first time, that he was diabetic. This information complicates matters because, although it doesn't change my original* 60 *percent estimate of his chances of having the disease in question, it does affect the interpretation of the results of the A test. This is so because the A test, while never yielding a positive result when the patient is healthy, does unfortunately yield a positive result* 30 *percent of the time in the case of diabetic patients who are not suffering from the disease. Now what do I do? More tests or immediate surgery?"*

Solution

Solution: In order to decide whether or not to recommend surgery, the doctor should first compute her updated probability that Jones has the disease given that the A test result was positive. Let D denote the event that Jones has the disease and E the event that the A test result is positive. The desired conditional probability is then

$$\begin{aligned} P(D|E) &= \frac{P(DE)}{P(E)} \\ &= \frac{P(D)P(E|D)}{P(E|D)P(D) + P(E|D^c)P(D^c)} \\ &= \frac{(.6)1}{1(.6) + (.3)(.4)} \\ &= .833. \end{aligned}$$

Note that we have computed the probability of a positive test result by conditioning on whether or not Jones has the disease and then using the fact that, because Jones is a diabetic, his conditional probability of a positive result given that he does not have the disease, $P(E|D^c)$, equals .3. Hence, as the doctor should now be over 80 percent certain that Jones has the disease, she should recommend surgery.

Example 22.

At a certain stage of a criminal investigation, the inspector in charge is 60 percent convinced of the guilt of a certain suspect. Suppose, however, that a new piece of evidence which shows that the criminal has a certain characteristic (such as left-handedness, baldness, or brown hair) is uncovered. If 20 percent of the population possesses this characteristic, how certain of the guilt of the suspect should the inspector now be if it turns out that the suspect has the characteristic?

Solution: Letting G denote the event that the suspect is guilty and C the event that he possesses the characteristic of the criminal, we have

$$\begin{aligned} P(G|C) &= \frac{P(GC)}{P(C)} \\ &= \frac{P(C|G)P(G)}{P(C|G)P(G) + P(C|G^c)P(G^c)} \\ &= \frac{1(.6)}{1(.6) + (.2)(.4)} \approx .882 \end{aligned}$$

where we have supposed that the probability of the suspect having the characteristic if he is, in fact, innocent is equal to .2, the proportion of the population possessing the characteristic.

Example

Example 23.

Urn 1 initially has n red molecules and urn 2 has n blue molecules. Molecules are randomly removed from urn 1 in the following manner: After each removal from urn 1, a molecule is taken from urn 2 (if urn 2 has any molecules) and placed in urn 1. The process continues until all the molecules have been removed. (Thus, there are $2n$ removals in all.)

(a) *Find $P(R)$, where R is the event that the final molecule removed from urn 1 is red.*

(b) *Repeat the problem when urn 1 initially has r_1 red molecules and b_1 blue molecules and urn 2 initially has r_2 red molecules and b_2 blue molecules.*

Solution

(a) Focus attention on any particular red molecule, and let F be the event that this molecule is the final one selected. Now, in order for F to occur, the molecule in question must still be in the urn after the first n molecules have been removed (at which time urn 2 is empty). So, letting N_i be the event that this molecule is not the ith molecule to be removed, we have

$$\begin{aligned} P(F) &= P(N_1 \cdots N_n F) \\ &= P(N_1)P(N_2|N_1) \cdots P(N_n|N_1 \cdots N_{n-1})P(F|N_1 \cdots N_n) \\ &= \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{1}{n}\right) \frac{1}{n} \end{aligned}$$

where the preceding formula uses the fact that the conditional probability that the molecule under consideration is the final molecule to be removed, given that it is still in urn 1 when only n molecules remain, is, by symmetry, $1/n$. Therefore, if we number the n red molecules and let R_j be the event that red molecule number j is the final molecule removed, then it follows from the preceding formula that $P(R_j) = \left(1 - \frac{1}{n}\right)^n \frac{1}{n}$. Because the events R_j are mutually exclusive, we obtain

$$P(R) = P\left(\bigcup_{j=1}^{n} R_j\right) = \sum_{j=1}^{n} P(R_j) = (1 - \frac{1}{n})^n \approx e^{-1}.$$

Solution (contd...)

(b) Suppose now that urn i initially has r_i red and b_i blue molecules, for $i = 1, 2$. To find $P(R)$, the probability that the final molecule removed is red, focus attention on any molecule that is initially in urn 1. As in part (a), it follows that the probability that this molecule is the final one removed is

$$p = \left(1 - \frac{1}{r_1 + b_1}\right)^{r_2+b_2} \frac{1}{r_1 + b_1}.$$

That is, $\left(1 - \frac{1}{r_1+b_1}\right)^{r_2+b_2}$ is the probability that the molecule under consideration is still in urn 1 when urn 2 becomes empty, and $\frac{1}{r_1+b_1}$ is the conditional probability, given the preceding event, that the molecule under consideration is the final molecule removed.

Solution (contd...)

Hence, if we let O be the event that the last molecule removed is one of the molecules originally in urn 1, then

$$P(O) = (r_1 + b_1)p = \left(1 - \frac{1}{r_1 + b_1}\right)^{r_2+b_2}.$$

To determine $P(R)$, we condition on whether O occurs, to obtain

$$\begin{aligned} P(R) &= P(R|O)P(O) + P(R|O^c)P(O^c) \\ &= \frac{r_1}{r_1 + b_1}\left(1 - \frac{1}{r_1 + b_1}\right)^{r_2+b_2} + \frac{r_2}{r_2 + b_2}\left(1 - \left(1 - \frac{1}{r_1 + b_1}\right)^{r_2+b_2}\right) \end{aligned}$$

If $r_1 + b_1 = r_2 + b_2 = n$, so that both urns initially have n molecules, then, when n is large,

$$P(L) \approx \frac{r_1}{r_1 + b_1}e^{-1} + \frac{r_2}{r_2 + b_2}(1 - e^{-1}).$$

The change in the probability of a hypothesis when new evidence is introduced can be expressed compactly in terms of the change in the odds of that hypothesis, where the concept of odds is defined as follows.

Definition 24.

The odds of an event A are defined by

$$\frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}.$$

That is, the odds of an event A tell how much more likely it is that the event A occurs than it is that it does not occur.
For instance, if $P(A) = \frac{2}{3}$, *then* $P(A) = 2P(A^c)$, *so the odds are* 2. *If the odds are equal to* α, *then it is common to say that the odds are "*α *to* 1*" in favor of the hypothesis.*

Conditional Probability

Consider now a hypothesis H that is true with probability $P(H)$, and suppose that new evidence E is introduced. Then the conditional probabilities, given the evidence E, that H is true and that H is not true are respectively given by

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \quad P(H^c|E) = \frac{P(E|H^c)P(H^c)}{P(E)}.$$

Therefore, the new odds after the evidence E has been introduced are

$$\frac{P(H|E)}{P(H^c|E)} = \frac{P(H)}{P(H^c)} \frac{P(E|H)}{P(E|H^c)}. \tag{1}$$

That is, the new value of the odds of H is the old value, multiplied by the ratio of the conditional probability of the new evidence given that H is true to the conditional probability given that H is not true. Thus, Equation (1) verifies the result of Example (22), since the odds, and thus the probability of H, increase whenever the new evidence is more likely when H is true than when it is false. Similarly, the odds decrease whenever the new evidence is more likely when H is false than when it is true.

Example 25.

An urn contains two type A coins and one type B coin. When a type A coin is flipped, it comes up heads with probability $1/4$, whereas when a type B coin is flipped, it comes up heads with probability $3/4$. A coin is randomly chosen from the urn and flipped. Given that the flip landed on heads, what is the probability that it was a type A coin?

Solution: Let A be the event that a type A coin was flipped, and let $B = A^c$ be the event that a type B coin was flipped. We want $P(A|\text{heads})$, where heads is the event that the flip landed on heads. From Equation (1), we see that

$$\begin{aligned}\frac{P(A|\text{heads})}{P(A^c|heads)} &= \frac{P(A)}{P(B)}\frac{P(\text{heads}|A)}{P(\text{heads}|B)}\\ &= \frac{2/3}{1/3}\frac{1/4}{3/4}\\ &= 2/3.\end{aligned}$$

Hence, the odds are $2/3 : 1$, or, equivalently, the probability is $2/5$ that a type A coin was flipped.

Example 26.

A plane is missing, and it is presumed that it was equally likely to have gone down in any of 3 *possible regions. Let* $1 - \beta_i$, $i = 1, 2, 3$, *denote the probability that the plane will be found upon a search of the* i*th region when the plane is, in fact, in that region. (The constants* β_i *are called overlook probabilities, because they represent the probability of overlooking the plane; they are generally attributable to the geographical and environmental conditions of the regions.) What is the conditional probability that the plane is in the* i*th region given that a search of region* 1 *is unsuccessful?*

Solution

Let R_i, $i = 1, 2, 3$, be the event that the plane is in region i, and let E be the event that a search of region 1 is unsuccessful. From Bayes's formula, we obtain

$$\begin{aligned} P(R_1|E) &= \frac{P(ER_1)}{P(E)} \\ &= \frac{P(E|R_1)P(R_1)}{\sum_{i=1}^{3} P(E|R_i)P(R_i)} \\ &= \frac{(\beta_1)\frac{1}{3}}{(\beta_1)\frac{1}{3} + (1)\frac{1}{3} + (1)\frac{1}{3}} \\ &= \frac{\beta_1}{\beta_1 + 2} \end{aligned}$$

For $j = 2, 3$,

$$\begin{aligned} P(R_j|E) &= \frac{P(E|R_j)P(R_j)}{P(E)} \\ &= \frac{(1)\frac{1}{3}}{(\beta_1)\frac{1}{3} + \frac{1}{3} + \frac{1}{3}} \\ &= \frac{1}{\beta_1 + 2}. \end{aligned}$$

Solution (contd...)

Note that the updated (that is, the conditional) probability that the plane is in region j, given the information that a search of region 1 did not find it, is greater than the initial probability that it was in region j when $j \neq 1$ and is less than the initial probability when $j = 1$. This statement is certainly intuitive, since not finding the plane in region 1 would seem to decrease its chance of being in that region and increase its chance of being elsewhere. Further, the conditional probability that the plane is in region 1 given an unsuccessful search of that region is an increasing function of the overlook probability β_1. This statement is also intuitive, since the larger β_1 is, the more it is reasonable to attribute the unsuccessful search to "bad luck" as opposed to the plane's not being there. Similarly, $P(R_j|E)$, $j \neq 1$, is a decreasing function of β_1.

Example

The next example has often been used by unscrupulous probability students to win money from their less enlightened friends.

Example 27.

Suppose that we have 3 cards that are identical in form, except that both sides of the first card are colored red, both sides of the second card are colored black, and one side of the third card is colored red and the other side black. The 3 cards are mixed up in a hat, and 1 card is randomly selected and put down on the ground. If the upper side of the chosen card is colored red, what is the probability that the other side is colored black?

Solution: Let RR, BB, and RB denote, respectively, the events that the chosen card is all red, all black, or the red-black card. Also, let R be the event that the upturned side of the chosen card is red. Then the desired probability is obtained by

$$\begin{aligned} P(RB|R) &= \frac{P(RB \cap R)}{P(R)} \\ &= \frac{P(R|RB)P(RB)}{P(R|RR)P(RR) + P(R|RB)P(RB) + P(R|BB)P(BB)} \\ &= \frac{\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)}{1\left(\frac{1}{3}\right) + \left(\frac{1}{2}\right)\left(\frac{1}{3}\right) + 0\left(\frac{1}{3}\right)} = \frac{1}{3} \end{aligned}$$

Solution (contd...)

Hence, the answer is $\frac{1}{3}$. Some students guess $\frac{1}{2}$ as the answer by incorrectly reasoning that, given that a red side appears, there are two equally likely possibilities: that the card is the all-red card or the red-black card. Their mistake, however, is in assuming that these two possibilities are equally likely. For, if we think of each card as consisting of two distinct sides, then we see that there are 6 equally likely outcomes of the experiment-namely, $R_1, R_2, B_1, B_2, R_3, B_3$-where the outcome is R_1 if the first side of the all-red card is turned face up, R_2 if the second side of the all-red card is turned face up, R_3 if the red side of the red-black card is turned face up, and so on. Since the other side of the upturned red side will be black only if the outcome is R_3, we see that the desired probability is the conditional probability of R_3 given that either R_1 or R_2 or R_3 occurred, which obviously equals $\frac{1}{3}$.

Example 28.

A new couple, known to have two children, has just moved into town. Suppose that the mother is encountered walking with one of her children. If this child is a girl, what is the probability that both children are girls?

Solution: Let us start by defining the following events:

G_1: the first (that is, the oldest) child is a girl.

G_2: the second child is a girl.

G: the child seen with the mother is a girl.

Also, let B_1, B_2, and B denote similar events, except that "girl" is replaced by "boy".

Now, the desired probability is $P(G_1G_2|G)$, which can be expressed as follows:

$$P(G_1G_2|G) = \frac{P(G_1G_2G)}{P(G)} = \frac{P(G_1G_2)}{P(G)}.$$

Solution (contd...)

Also,

$$\begin{aligned} P(G) =& P(G|G_1G_2)P(G_1G_2) + P(G|G_1B_2)P(G_1B_2) \\ &+ P(G|B_1G_2)P(B_1G_2) + P(G|B_1B_2)P(B_1B_2) \\ =& P(G_1G_2) + P(G|G_1B_2)P(G_1B_2) + P(G|B_1G_2)P(B_1G_2) \end{aligned}$$

where the final equation used the results $P(G|G_1G_2) = 1$ and $P(G|B_1B_2) = 0$. If we now make the usual assumption that all 4 gender possibilities are equally likely, then we see that

$$\begin{aligned} P(G_1G_2|G) &= \frac{\frac{1}{4}}{\frac{1}{4} + P(G|G_1B_2)/4 + P(G|B_1G_2)/4} \\ &= \frac{1}{1 + P(G|G_1B_2) + P(G|B_1G_2)}. \end{aligned}$$

Solution (contd...)

Thus, the answer depends on whatever assumptions we want to make about the conditional probabilities that the child seen with the mother is a girl given the event G_1B_2 and that the child seen with the mother is a girl given the event G_2B_1. For instance, if we want to assume, on the one hand, that, independently of the genders of the children, the child walking with the mother is the elder child with some probability p, then it would follow that

$$P(G|G_1B_2) = p = 1 - P(G|B_1G_2)$$

implying under this scenario that

$$P(G_1G_2|G) = \frac{1}{2}.$$

If, on the other hand, we were to assume that if the children are of different genders, then the mother would choose to walk with the girl with probability q, independently of the birth order of the children, then we would have

$$P(G|G_1B_2) = P(G|B_1G_2) = q$$

implying that

$$P(G_1G_2|G) = \frac{1}{1+2q}$$

Solution (contd...)

For instance, if we took $q = 1$, meaning that the mother would always choose to walk with a daughter, then the conditional probability that she has two daughters would be $\frac{1}{3}$ because seeing the mother with a daughter is now equivalent to the event that she has at least one daughter.

Hence, as stated, the problem is incapable of solution. Indeed, even when the usual assumption about equally likely gender probabilities is made, we still need to make additional assumptions before a solution can be given.

This is because the sample space of the experiment consists of vectors of the form s_1, s_2, i, where s_1 is the gender of the older child, s_2 is the gender of the younger child, and i identifies the birth order of the child seen with the mother. As a result, to specify the probabilities of the events of the sample space, it is not enough to make assumptions only about the genders of the children; it is also necessary to assume something about the conditional probabilities as to which child is with the mother given the genders of the children.

Independent Events

We have considered events A and B which could not occur simultaneously, that is, $A \cap B = \emptyset$. Such events were called **mutually exclusive**. We have noted previously that if A and B are mutually exclusive, then $P(A|B) = 0$, for the given occurrence of B precludes the occurrence of A. At the other extreme we have the situation, also discussed above, in which $B \supset A$ and hence $P(B|A) = 1$.

In each of the above situations, knowing that B has occurred gave us some very definite information concerning the probability of the occurrence of A. However, there are many situations in which knowing that some event B did occur has no bearing whatsoever on the occurrence or nonoccurrence of A.

Example 29.

Suppose that a fair die is tossed twice. Define the events A and B as follows:

$$A = \{\textit{the first die shows an even number}\},$$
$$B = \{\textit{the second die shows a } 5 \textit{ or a } 6\}.$$

It is intuitively clear that events A and B are totally unrelated*. Knowing that B did occur yields no information about the occurrence of A. In fact, the following computation bears this out. Taking as our sample space the 36 equally likely outcomes, we find that* $P(A) = \frac{18}{36} = \frac{1}{2}$, $P(B) = \frac{12}{36} = \frac{1}{3}$, *while* $P(A \cap B) = \frac{6}{36} = \frac{1}{6}$. *Hence*

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\left(\frac{1}{6}\right)}{\left(\frac{1}{3}\right)} = \frac{1}{2}.$$

Thus we find, as we might have expected, that the unconditional probability is equal to the conditional probability $P(A|B)$. Similarly,

$$P(B|A) = \frac{P(B \cap A)}{P(B)} = \frac{\left(\frac{1}{6}\right)}{\left(\frac{1}{2}\right)} = \frac{1}{3} = P(B).$$

Independent Events

Definition 30.

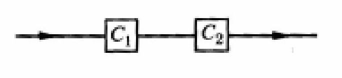
A and B are independent events if and only if

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

This definition is essentially equivalent to the one suggested above, namely, that A and B are independent if $P(B|A) = P(B)$ and $P(A|B) = P(A)$. If either $P(A)$ or $P(B)$ equals zero, this definition is still valid. This latter form is slightly more intuitive, A and B are independent if knowledge of the occurrence of A in no way influences the probability of the occurrence of B.

Example 31.

Suppose that a mechanism is made up of two components hooked up in series as indicated in the following figure. Each component has a probability p of not working. What is the probability that the mechanism does work?



It is clear that the mechanism will work if and only if both components are functioning. Hence Prob (mechanism works) = Prob (C_1 functions and C_2 functions).
The information we are given does not allow us to proceed unless we know (or assume) that the two mechanisms work independently of each other. This may or may not be a realistic assumption, depending on how the two parts are hooked up. If we assume that the two mechanisms work independently, we obtain for the required probability $(1 - p)^2$.

Example (contd...)

It will be important for us to extend the above notion of independence to more than two events.

Let us first consider three events associated with an experiment, say A, B, and C.

If A and B, A and C, B and C are each pairwise independent (in the above sense), then it does not follow, in general, that there exists no dependence between the three events A, B, and C. The following (somewhat artificial) example illustrates this point.

Example 32.

Suppose that we toss two dice. Define the events A, B, and C as follows:
$A = \{$*the first die shows an even number*$\}$,
$B = \{$*the second die shows an odd number*$\}$,
$C = \{$*the two dice show both odd or both even numbers*$\}$.
We have $P(A) = P(B) = P(C) = \frac{1}{2}$. *Furthermore,* $P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}$.
Hence the three events are all pairwise independent. However,
$P(A \cap B \cap C) = 0 \neq P(A)P(B)P(C)$. *This example motivates the following definition.*

Mutually Independent

Definition 33.

We say that the three events A, B, and C are mutually independent if and only if all the following conditions hold:

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

We finally generalize this notion to n events in the following definition.

Definition 34.

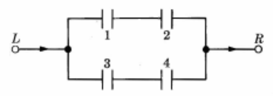
The n events $A_1, A_2, \ldots, A_n$ are mutually independent if and only if we have for $k = 2, 3, \ldots, n$,

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$$

In most applications we need not check all these conditions, for we usually assume independence (based on what we know about the experiment). We then use this assumption to evaluate, say $P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k})$ as $P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$.

Example 35.

The probability of the closing of each relay of the circuit shown in the following figure is given by p. If all relays function independently, what is the probability that a current exists between the terminals L and R?

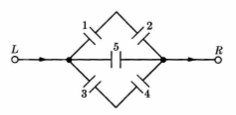


Let A represent the event {relay i is closed}, $i = 1, 2, 3, 4$. Let E represent the event {current flows from L to R.} Hence $E = (A_1 \cap A_2) \cup (A_3 \cap A_4)$. (Note that $A_1 \cap A_2$ and $A_3 \cap A_4$ are not mutually exclusive.) Thus

$$\begin{aligned} P(E) &= P(A_1 \cap A_2) + P(A_3 \cap A_4) - P(A_1 \cap A_2 \cap A_3 \cap A_4) \\ &= p^2 + p^2 - p^4 = 2p^2 - p^4. \end{aligned}$$

Example 36.

Assume again that for the circuit in the following figure, the probability of each relay being closed is p and that all relays function independently. What is the probability that current exists between the terminals L and R?



We have that

$$\begin{aligned}P(E) &= P(A_1 \cap A_2) + P(A_5) + P(A_3 \cap A_4) - P(A_1 \cap A_2 \cap A_5) \\ &\quad - P(A_1 \cap A_2 \cap A_3 \cap A_4) - P(A_5 \cap A_3 \cap A_4) \\ &\quad + P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) \\ &= p^2 + p + p^2 - p^3 - p^4 - p^3 + p^5 = p + 2p^2 - 2p^3 - p^4 + p^5.\end{aligned}$$

Example

Let us discuss a fairly common, but erroneous, approach to a problem.

Example 37.

Suppose that among six bolts, two are shorter than a specified length. If two bolts are chosen at random, what is the probability that the two short bolts are picked? Let A_i *be the event* $\{\text{the ith chosen bolt is short}\}, i = 1, 2$. *Hence we want to evaluate* $P(A_1 \cap A_2)$. *The proper solution is obtained, of course, by writing*

$$P(A_1 \cap A_2) = P(A_2|A_1)P(A_1) = \frac{1}{5} \cdot \frac{2}{6} = \frac{1}{15}.$$

The common but incorrect approach is to write

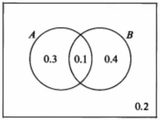
$$P(A_1 \cap A_2) = P(A_2)P(A_1) = \frac{1}{5} \cdot \frac{2}{6} = \frac{1}{15}.$$

Of course, the point is that although the answer is correct, the identification of $\frac{1}{5}$ with $P(A_2)$ is incorrect; $\frac{1}{5}$ represents $P(A_2|A_1)$. To evaluate $P(A_2)$ properly, we write

$$P(A_2) = P(A_2|A_1)P(A_1) + P(A_2|\bar{A}_1 P(\bar{A}_1) = \frac{1}{5} \cdot \frac{2}{6} + \frac{2}{5} \cdot \frac{4}{6} = \frac{1}{3}.$$

Schematic Considerations; Conditional Probability and Independence

The following schematic approach may be useful for understanding the concept of conditional probability. Suppose that A and B are two events associated with a sample space for which the various probabilities are indicated in the Venn diagram given in the following figure.



Hence $P(A \cap B) = 0.1, P(A) = 0.1 + 0.3 = 0.4$, and $P(B) = 0.1 + 0.4 = 0.5$. Next, let us represent the various probabilities by the areas of the rectangles as in the following figure. In each case, the shaded regions indicate the event B: In the left rectangle we are representing $A \cap B$ and in the right one $A' \cap B$.

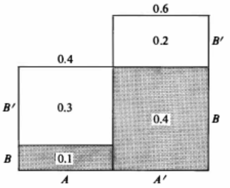


Figure 1

Schematic Considerations; Conditional Probability and Independence

Now suppose we wish to compute $P(B|A)$. Thus we need only to consider A; that is, A' may be ignored in the computation. We note that the proportion of B in A is $1/4$. (We can also check this by applying the equation: $P(B|A) = P(A \cap B)/P(A) = 0.1/0.4 = 1/4$.) Hence $P(B'A) = 3/4$, and our diagram representing this conditional probability would be given by the following figure:

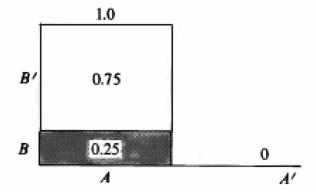
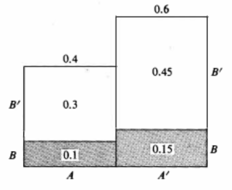


Figure 2

Observe also that if A is given to have occurred, all of the probability (i.e., 1) must be associated with the event A while none of the probability (i.e., 0) is associated with A'. Furthermore note that in the left rectangle, representing A, only the individual entries have changed from Figure 1 to Figure 2 (adding up to 1 instead of 0.4). However, the proportions within the rectangle have remained the same (i.e., 3 : 1).

Schematic Considerations; Conditional Probability and Independence

Let us also illustrate the notion of independence using the schematic approach introduced above. Suppose that the events A and B are as depicted in the following figure. In that case the proportions in the two rectangles, representing A and A', are the same: 3 : 1 in both cases. Thus we have $P(B) = 0.1 + 0.15 = 0.25$, and $P(B \cap A) = 0.1/0.4 = 0.25$.



Finally observe that by simply looking at the first figure, we can also compute the other conditional probabilities: $P(A|B) = 1/5$ (since 1/5 of the total rectangular area representing B is occupied by A), $P(A'|B) = 4/5$.

Example 38.

Two coins are flipped, and all 4 outcomes are assumed to be equally likely. If E is the event that the first coin lands on heads and F the event that the second lands on tails, are E and F independent?

Solution: $P(EF) = P(\{(H, T)\}) = \frac{1}{4}$, whereas $P(E) = P(\{(H, H), (H, T)\}) = \frac{1}{2}$ and $P(F) = P(\{(H, T), (T, T)\}) = \frac{1}{2}$.

Example 39.

Suppose that we toss 2 fair dice. Let E_1 denote the event that the sum of the dice is 6 and F denote the event that the first die equals 4. Then

$$P(E_1F) = P(\{(4,2)\}) = \frac{1}{36}$$

whereas

$$P(E_1)P(F) = \left(\frac{5}{36}\right)\left(\frac{1}{6}\right) = \frac{5}{216}.$$

Are E_1 and F independent?

Solution: If we are interested in the possibility of throwing a 6 (with 2 dice), we shall be quite happy if the first die lands on 4 (or, indeed, on any of the numbers $1, 2, 3, 4$, and 5), for then we shall still have a possibility of getting a total of 6. If, however, the first die landed on 6, we would be unhappy because we would no longer have a chance of getting a total of 6. In other words, our chance of getting a total of 6 depends on the outcome of the first die; thus, E_1 and F cannot be independent.

Example 40.

Suppose that we let E_2 be the event that the sum of the dice equals 7. Is E_2 independent of F?

Solution: The answer is yes, since

$$P(E_2F) = P(\{(4,3)\}) = \frac{1}{36}$$

whereas

$$P(E_2)P(F) = \left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \left(\frac{1}{36}\right).$$

Proposition 41.

If E and F are independent, then so are E and F^c.

Proof: Assume that E and F are independent. Since $E = EF \cup EF^c$ and EF and EF^c are obviously mutually exclusive, we have

$$P(E) = P(EF) + P(EF^c) = P(E)P(F) + P(EF^c)$$

or, equivalently,

$$P(EF^c) = P(E)[1 - P(F)] = P(E)P(F^c)$$

and the result is proved.

Thus, if E is independent of F, then the probability of E's occurrence is unchanged by information as to whether or not F has occurred.

Suppose now that E is independent of F and is also independent of G. Is E then necessarily independent of FG? The answer, somewhat surprisingly, is no, as the following example demonstrates.

Example 42.

Two fair dice are thrown. Let E denote the event that the sum of the dice is 7. Let F denote the event that the first die equals 4 and G denote the event that the second die equals 3. From Example (39), we know that E is independent of F, and the same reasoning as applied there shows that E is also independent of G; but clearly, E is not independent of FG [since $P(E|FG) = 1$].

Definition 43.

Three events E, F, and G are said to be independent if

$$
\begin{aligned}
P(EFG) &= P(E)P(F)P(G) \\
P(EF) &= P(E)P(F) \\
P(EG) &= P(E)P(G) \\
P(FG) &= P(F)P(G).
\end{aligned}
$$

We define an infinite set of events to be independent if every finite subset of those events is independent.

Independent Events

Sometimes, a probability experiment under consideration consists of performing a sequence of subexperiments. For instance, if the experiment consists of **continually tossing a coin**, we may think of each toss as being a subexperiment. In many cases, it is reasonable to assume that the outcomes of any group of the subexperiments have no effect on the probabilities of the outcomes of the other subexperiments. If such is the case, we say that the subexperiments are independent.

More formally, we say that the subexperiments are independent if $E_1, E_2, \ldots, E_n, \ldots$ is necessarily an independent sequence of events whenever E_i is an event whose occurrence is completely determined by the outcome of the ith subexperiment.

If each subexperiment has the same set of possible outcomes, then the subexperiments are often called *trials*.

Example 44.

An infinite sequence of independent trials is to be performed. Each trial results in a success with probability p and a failure with probability $1 - p$. What is the probability that

(a) *at least* 1 *success occurs in the first n trials;*

(b) *exactly k successes occur in the first n trials;*

(c) *all trials result in successes?*

Solution: In order to determine the probability of at least 1 success in the first n trials, it is easiest to compute first the probability of the complementary event: that of no successes in the first n trials. If we let E_i **denote the event of a failure on the ith trial**, then the probability of no successes is, by independence,

$$P(E_1 E_2 \cdots E_n) = P(E_1)P(E_2)\cdots P(E_n) = (1 - p)^n$$

Hence, the answer to part (a) is $1 - (1 - p)^n$.

Solution (contd...)

To compute the answer to part (b), consider any particular sequence of the first n outcomes containing k successes and $n-k$ failures. Each one of these sequences will, by the assumed independence of trials, occur with probability $p^k(1-p)^{n-k}$. Since there are $\binom{n}{k}$ such sequences (there are $n!/k!(n-k)!$ permutations of k successes and $n-k$ failures), the desired probability in part (b) is

$$P\{\text{exactly } k \text{ successes}\} = \binom{n}{k} p^k(1-p)^{n-k}.$$

To answer part (c), we note that, by part (a), the probability of the first n trials all resulting in success is given by

$$P(E_1^c E_2^c \cdots E_n^c) = p^n.$$

Solution (contd...)

Thus, using the continuity property of probabilities, we see that the desired probability is given by

$$\begin{aligned} P\left(\bigcap_{i=1}^{\infty} E_i^c\right) &= P\left(\lim_{n\to\infty} \bigcap_{i=1}^{n} E_i^c\right) \\ &= \lim_{n\to\infty} P\left(\bigcap_{i=1}^{n} E_i^c\right) \\ &= \lim_{n} p^n = \begin{cases} 0 \text{ if } p < 1 \\ 1 \text{ if } p = 1. \end{cases} \end{aligned}$$

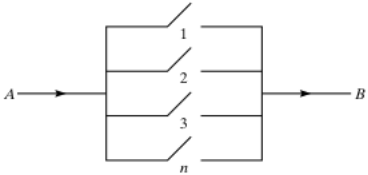
Example 45.

A system composed of n separate components is said to be a parallel system if it functions when at least one of the components functions. For such a system, if component i, which is independent of the other components, functions with probability p_i, $i = 1, \ldots, n$, what is the probability that the system functions?

Solution: Let A_i denote the event that component i functions. Then

$$P\{\text{system functions}\} = 1 - P\{\text{system doesn't function}\} = 1 - P\{\text{all components don't function}\}$$

$$= 1 - P\left(\bigcap_i A_i^c\right) = 1 - \prod_{i=1}^{n}(1 - p_i) \quad \text{by independence}$$



Parallel System: Functions if Current Flows from A to B

Example 46.

Independent trials resulting in a success with probability p and a failure with probability $1-p$ are performed. What is the probability that n successes occur before m failures? If we think of A and B as playing a game such that A gains 1 point when a success occurs and B gains 1 point when a failure occurs, then the desired probability is the probability that A would win if the game were to be continued in a position where A needed n and B needed m more points to win.

Solution: We shall present two solutions. The first is due to Pascal and the second to Fermat. Let us denote by $P_{n,m}$ the probability that n successes occur before m failures. By conditioning on the outcome of the first trial, we obtain

$$P_{n,m} = pP_{n-1,m} + (1-p)P_{n,m-1} \quad n \geq 1, m \geq 1$$

Using the obvious boundary conditions $P_{n,0} = 0$, $P_{0,m} = 1$, we can solve these equations for $P_{n,m}$. Rather than go through the tedious details, let us instead consider Fermat's solution.

Solution (contd...)

Fermat argued that, in order for n successes to occur before m failures, it is necessary and sufficient that there be at least n successes in the first $m+n-1$ trials. (Even if the game were to end before a total of $m+n-1$ trials were completed, we could still imagine that the necessary additional trials were performed.)

This is true, for if there are at least n successes in the first $m+n-1$ trials, there could be at most $m-1$ failures in those $m+n-1$ trials; thus, n successes would occur before m failures. If, however, there were fewer than n successes in the first $m+n-1$ trials, there would have to be at least m failures in that same number of trials; thus, n successes would not occur before m failures.

Hence, the probability of exactly k successes in $m+n-1$ trials is $\binom{m+n-1}{k}p^k(1-p)^{m+n-1-k}$, it follows that the desired probability of n successes before m failures is

$$P_{n,m} = \sum_{k=n}^{m+n-1} \binom{m+n-1}{k} p^k (1-p)^{m+n-1-k}.$$

Example 47.

Suppose that initially there are r players, with player i having n_i units, $n_i > 0$, $i = 1, \ldots, r$. At each stage, two of the players are chosen to play a game, with the winner of the game receiving 1 unit from the loser. Any player whose fortune drops to 0 is eliminated, and this continues until a single player has all $n \equiv \sum_{i=1}^{r} n_i$ units, with that player designated as the victor. Assuming that the results of successive games are independent and that each game is equally likely to be won by either of its two players, find P_i, the probability that player i is the victor?

Solution: To begin, suppose that there are n players, with each player initially having 1 unit. Consider player i. Each stage she plays will be equally likely to result in her either winning or losing 1 unit, with the results from each stage being independent.

In addition, she will continue to play stages until her fortune becomes either 0 or n. Because this is the same for all n players, it follows that each player has the same chance of being the victor, implying that each player has probability $1/n$ of being the victor.

Solution (contd...)

Now, suppose these n players are divided into r teams, with team i containing n_i players, $i = 1, \dots, r$. Then the probability that the victor is a member of team i is n_i/n. But because

(a) team i initially has a total fortune of n_i units, $i = 1, \dots, r$, and

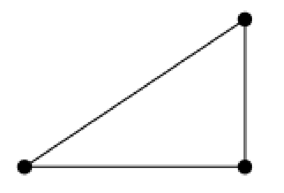
(b) each game played by members of different teams is equally likely to be won by either player and results in the fortune of members of the winning team increasing by 1 and the fortune of the members of the losing team decreasing by 1,

it is easy to see that the probability that the victor is from team i is exactly the probability we desire. Thus, $P_i = n_i/n$. Interestingly, our argument shows that this result does *not* depend on how the players in each stage are chosen.

Example

Example 48.

The complete graph having n vertices is defined to be a set of n points (called vertices) in the plane and the $\binom{n}{2}$ lines (called edges) connecting each pair of vertices. The complete graph having 3 vertices is shown in the following figure. Suppose now that each edge in a complete graph having n vertices is to be colored either red or blue. For a fixed integer k, a question of interest is, Is there a way of coloring the edges so that no set of k vertices has all of its $\binom{k}{2}$ connecting edges the same color? It can be shown by a probabilistic argument that if n is not too large, then the answer is yes.



Example (contd...)

The argument runs as follows: Suppose that each edge is, independently, equally likely to be colored either red or blue. That is, each edge is red with probability $\frac{1}{2}$. Number the $\binom{n}{k}$ sets of k vertices and define the events E_i, $i = 1, \ldots, \binom{n}{k}$ as follows:

$$E_i = \{\text{all of the connecting edges of the } i\text{th set of } k \text{ vertices are the same color}\}.$$

Now, since each of the $\binom{k}{2}$ connecting edges of a set of k vertices is equally likely to be either red or blue, it follows that the probability that they are all the same color is

$$P(E_i) = 2\left(\frac{1}{2}\right)^{k(k-1)/2}.$$

Therefore, because

$$P\left(\bigcup_i E_i\right) \leq \sum_i P(E_i) \quad \text{(Boole's inequality)}$$

we find that $P\left(\bigcup_i E_i\right)$, the probability that there is a set of k vertices all of whose connecting edges are similarly colored, satisfies

$$P\left(\bigcup_i E_i\right) \leq \binom{n}{k}\left(\frac{1}{2}\right)^{k(k-1)/2-1}.$$

Example (contd...)

Hence, if

$$\binom{n}{k}\left(\frac{1}{2}\right)^{k(k-1)/2-1} < 1$$

or, equivalently, if

$$\binom{n}{k} < 2^{k(k-1)/2-1}$$

then the probability that at least one of the $\binom{n}{k}$ sets of k vertices has all of its connecting edges the same color is less than 1.

Consequently, under the preceding condition on n and k, it follows that there is a positive probability that no set of k vertices has all of its connecting edges the same color.

But this conclusion implies that there is at least one way of coloring the edges for which no set of k vertices has all of its connecting edges the same color.

Remarks

(a) Whereas the preceding argument established a condition on n and k that guarantees the existence of a coloring scheme satisfying the desired property, it gives no information about how to obtain such a scheme (although one possibility would be simply to choose the colors at random, check to see if the resulting coloring satisfies the property, and repeat the procedure until it does).

(b) The method of introducing probability into a problem whose statement is purely deterministic has been called the *probabilistic method*.

Example 49.

An insurance company believes that people can be divided into two classes: those who are accident prone and those who are not. During any given year, an accident-prone person will have an accident with probability .4, whereas the corresponding figure for a person who is not prone to accidents is .2. What is the conditional probability that a new policyholder will have an accident in his or her second year of policy ownership, given that the policyholder has had an accident in the first year?

Solution: If we let A be the event that the policyholder is accident prone and we let A_i, $i = 1, 2$, be the event that he or she has had an accident in the ith year, then the desired probability $P(A_2|A_1)$ may be obtained by conditioning on whether or not the policyholder is accident prone, as follows:

$$P(A_2|A_1) = P(A_2|AA_1)P(A|A_1) + P(A_2|A^cA_1)P(A^c|A_1).$$

Now,

$$P(A|A_1) = \frac{P(A_1A)}{P(A_1)} = \frac{P(A_1|A)P(A)}{P(A_1)}.$$

Solution (contd...)

However, $P(A)$ is assumed to equal $\frac{3}{10}$, and it was shown that $P(A_1) = .26$. Hence,

$$P(A|A_1) = \frac{(.4)(.3)}{.26} = \frac{6}{13}.$$

Thus,

$$P(A^c|A_1) = 1 - P(A|A_1) = \frac{7}{13}.$$

Since $P(A_2|AA_1) = .4$ and $P(A_2|A^cA_1) = .2$, it follows that

$$P(A_2|A_1) = (.4)\frac{6}{13} + (.2)\frac{7}{13} \approx .29$$

Exercise 50.

Two fair dice are rolled. What is the conditional probability that at least one lands on 6 given that the dice land on different numbers?

Solution.

$$\begin{aligned} P\{6 \mid \text{different}\} &= P\{6, \text{different}\}/P\{\text{different}\} \\ &= \frac{P\{1^{st} = 6, 2^{nd} \neq 6\} + P\{1^{st} \neq 6, 2^{nd} = 6\}}{5/6} \\ &= \frac{2\frac{1}{6}\frac{5}{6}}{5.6} = \frac{1}{3} \end{aligned}$$

could also have been solved by using reduced sample space – for given that outcomes differ it is the same as asking for the probability that 6 is chosen when 2 of the numbers 1, 2, 3, 4, 5, 6 are randomly chosen.

Exercise 51.

If two fair dice are rolled, what is the conditional probability that the first one lands on 6 given that the sum of the dice is i? Compute for all values of i between 2 and 12.

Solution.

$$P\{6 \mid \text{sum of } 7\} = P\{(6,1)\}1/6 = 1/6$$
$$P\{6 \mid \text{sum of } 8\} = P\{(6,2)\}5/36 = 1/5$$
$$P\{6 \mid \text{sum of } 9\} = P\{(6,3)\}4/36 = 1/4$$
$$P\{6 \mid \text{sum of } 10\} = P\{(6,4)\}3/36 = 1/3$$
$$P\{6 \mid \text{sum of } 11\} = P\{(6,5)\}2/36 = 1/2$$
$$P\{6 \mid \text{sum of } 12\} = 1.$$

Exercise 52.

An urn contains 6 white and 9 black balls. If 4 balls are to be randomly selected without replacement, what is the probability that the first 2 selected are white and the last 2 black?

Solution.

$$\frac{6}{15}\frac{5}{14}\frac{9}{13}\frac{8}{12}$$

Exercise 53.

The king comes from a family of 2 children. What is the probability that the other child is his sister?

Solution.

$$P\{1g \text{ and } 1b \mid \text{at least one } b\} = \frac{1/2}{3/4} = 2/3$$

Exercise 54.

A couple has 2 children. What is the probability that both are girls if the older of the two is a girl?

Solution.

$$1/2$$

Exercise 55.

A recent college graduate is planning to take the first three actuarial examinations in the coming summer. She will take the first actuarial exam in June. If she passes that exam, then she will take the second exam in July, and if she also passes that one, then she will take the third exam in September. If she fails an exam, then she is not allowed to take any others. The probability that she passes the first exam is .9. If she passes the first exam, then the conditional probability that she passes the second one is .8, and if she passes both the first and the second exams, then the conditional probability that she passes the third exam is .7.

(a) *What is the probability that she passes all three exams?*

(b) *Given that she did not pass all three exams, what is the conditional probability that she failed the second exam?*

Solution :

(a) $(.9)(.8)(.7) = .504$

(b) Let F_i denote the event that she failed the ith exam.
$P(F_2 \big| (F_1^c F_2^c F_3^c)^c) = \frac{P(F_1^c F_2)}{1-.504} = \frac{(.9)(.5)}{.496} = .3629$

Exercise 56.

Ninety-eight percent of all babies survive delivery. However, 15 percent of all births involve Cesarean (C) sections, and when a C section is performed, the baby survives 96 percent of the time. If a randomly chosen pregnant woman does not have a C section, what is the probability that her baby survives?

Solution. With S being survival and C being C section of a randomly chosen delivery, we have that

$$\begin{aligned} .98 = P(S) &= P(S|C).15 + P(S|C^2).85 \\ &= .96(.15) + P(S|C^2).85 \end{aligned}$$

Hence

$$P(SC^c) \approx .9835.$$

Problems

Exercise 57.

A total of 48 percent of the women and 37 percent of the men that took a certain "quit smoking" class remained nonsmokers for at least one year after completing the class. These people then attended a success party at the end of a year. If 62 percent of the original class was male,

(a) *what percentage of those attending the party were women?*

(b) *what percentage of the original class attended the party?*

Solution : Choose a random member of the class. Let A be the event that this person attends the party and let W be the event that this person is a woman.

(a)

$$P(W|A) = \frac{P(A|W)P(W)}{P(A|W)P(W) + P(A|M)P(M)} \quad \text{where } M = W^c$$
$$= \frac{.48(.38)}{.48(.38) + .37(.62)} \approx .443.$$

Therefore, 44.3 percent of the attendees were women.

(b) $P(A) = .48(.38) + .37(.62) = .4118$. Therefore, 41.18 percent of the class attended.

Exercise 58.

Urn I contains 2 white and 4 red balls, whereas urn II contains 1 white and 1 red ball. A ball is randomly chosen from urn I and put into urn II, and a ball is then randomly selected from urn II. What is

(a) *the probability that the ball selected from urn II is white?*

(b) *the conditional probability that the transferred ball was white given that a white ball is selected from urn II?*

Solution.

$$P\{w \big| w \text{ transferred}\}P\{w \text{ tr.}\} + P\{w \big| R \text{ tr.}\}P\{R \text{ tr.}\} = \frac{2}{3}\frac{1}{3} + \frac{1}{3}\frac{2}{3} = \frac{4}{9}.$$

$$P\{w \text{ transferred} \big| w\} = \frac{P\{w \big| w \text{ tr.}\}P\{w \text{ tr.}\}}{P\{w\}} = \frac{\frac{2}{3}\frac{1}{3}}{\frac{4}{9}} = 1/2$$

Exercise 59.

Suppose that 5 percent of men and .25 percent of women are color blind. A color-blind person is chosen at random. What is the probability of this person being male? Assume that there are an equal number of males and females. What if the population consisted of twice as many males as females?

Solution. Let M be the event that the person is male, and let C be the event that he or she is color blind. Also, let p denote the proportion of the population that is male.

$$P(M|C) = \frac{P(C|M)P(M)}{P(C|M)P(M) + P(C|M^c)P(M^c)} = \frac{(.50)p}{(.50)p + (.0025)(1-p)}$$

Exercise 60.

Suppose that an ordinary deck of 52 cards is shuffled and the cards are then turned over one at a time until the first ace appears. Given that the first ace is the 20th card to appear, what is the conditional probability that the card following it is the

(a) *ace of spades?*

(b) *two of clubs?*

Solution. Let A denote the event that the next card is the ace of spades and let B be the event that it is the two of clubs.

(a)

$$P\{A\} = P\{\text{next card is an ace}\}P\{A \big| \text{next card is an ace}\}$$
$$= \frac{3}{32}\frac{1}{4}\frac{3}{128}$$

(b) Let C be the event that the two of clubs appeared among the first 20 cards.

$$P(B) = P(B \big| C)P(C) + P(B \big| C^c)P(C^c)$$
$$= 0\frac{19}{48} + \frac{1}{32}\frac{29}{48} = \frac{29}{1536}$$

Exercise 61.

On rainy days, Joe is late to work with probability .3; on nonrainy days, he is late with probability .1. With probability .7, it will rain tomorrow.

(a) *Find the probability that Joe is early tomorrow.*

(b) *Given that Joe was early, what is the conditional probability that it rained?*

Exercise 62.

Stores A, B, and C have $50, 75$, and 100 employees, respectively, and $50, 60$, and 70 percent of them respectively are women. Resignations are equally likely among all employees, regardless of sex. One woman employee resigns. What is the probability that she works in store C?

Solution.

$$
\begin{aligned}
P\{C|\text{women}\} &= \frac{P\{\text{women}|C\}P\{C\}}{P\{\text{women}|A\}P\{A\} + P\{\text{women}|B\}P\{B\} + P\{\text{women}|C\}P\{C\}} \\
&= \frac{.7\,\frac{100}{225}}{.5\,\frac{50}{225} + .6\,\frac{75}{225} + .7\,\frac{100}{225}} = \frac{1}{2}
\end{aligned}
$$

Exercise 63.

Suppose we have 10 coins such that if the ith coin is flipped, heads will appear with probability $i/10, i = 1, 2, \ldots, 10$. *When one of the coins is randomly selected and flipped, it shows heads. What is the conditional probability that it was the fifth coin?*

Solution.

$$P\{5th \big| \text{heads}\} = \frac{P\{\text{heads} \big| 5^{th}\} P\{5^{th}\}}{\sum_i P\{h \big| i^{th}\} P\{i^{th}\}}$$

$$= \frac{\frac{5}{10}\frac{1}{10}}{\sum_{i=1}^{10} \frac{i}{10}\frac{i}{10}} = \frac{1}{11}.$$

Exercise 64.

Each of 2 cabinets identical in appearance has 2 drawers. Cabinet A contains a silver coin in each drawer, and cabinet B contains a silver coin in one of its drawers and a gold coin in the other. A cabinet is randomly selected, one of its drawers is opened, and a silver coin is found. What is the probability that there is a silver coin in the other drawer?

Solution. (a) $P\{\text{silver in other}|\text{silver found}\}$

$$= \frac{P\{S \text{ in other, S found}\}}{P\{S \text{ found}\}}.$$

To compute these probabilities, condition on the cabinet selected.

$$= \frac{1/2}{P\{S \text{ found}|A\}1/2 + P\{S \text{ found}|B\}1/2}$$
$$= \frac{1}{1+1/2} = \frac{2}{3}.$$

Exercise 65.

A parallel system functions whenever at least one of its components works. Consider a parallel system of n components, and suppose that each component works independently with probability $\frac{1}{2}$. Find the conditional probability that component 1 works given that the system is functioning.

Solution. Let W and F be the events that component 1 works and that the system functions.

$$P(W|F) = \frac{P(WF)}{P(F)} = \frac{P(W)}{1 - P(F^c)} = \frac{1/2}{1 - (1/2)^{n-1}}$$

Exercise 66.

In a class, there are 4 freshman boys, 6 freshman girls, and 6 sophomore boys. How many sophomore girls must be present if sex and class are to be independent when a student is selected at random?

Solution.

$$P\{\text{Boy}, F\} = \frac{4}{16+x} \quad P\{\text{Boy}\} = \frac{10}{16+x} \quad P\{F\} = \frac{10}{16+x}$$

So independence $\Rightarrow 4 = \frac{10.10}{16+x} \Rightarrow 4x = 36$ or $x = 9$.

A direct check now shows that 9 sophomore girls (which the above shows is necessary) is also sufficient for independence of sex and class.

Exercise 67.

Independent flips of a coin that lands on heads with probability p are made. What is the probability that the first four outcomes are

(a) H, H, H, H?

(b) T, H, H, H?

(c) *What is the probability that the pattern T, H, H, H occurs before the pattern H, H, H, H?*

Hint for part (c): How can the pattern H, H, H, H occur first?

Solution.

(a) $1/6$

(b) $1/16$

(c) The only way in which the pattern H, H, H, H can occur first is for the first 4 flips to all be heads, for once a tail appears it follows that a tail will precede the first run of 4 heads (and so T, H, H, H will appear first). Hence, the probability that T, H, H, H occurs first is $15/16$.

Exercise 68.

A true-false question is to be posed to a husbandand-wife team on a quiz show. Both the husband and the wife will independently give the correct answer with probability p. Which of the following is a better strategy for the couple?

(a) *Choose one of them and let that person answer the question.*

(b) *Have them both consider the question, and then either give the common answer if they agree or, if they disagree, flip a coin to determine which answer to give.*

Solution.

If use (a) will win with probability p. If use strategy (b) then

$$\begin{aligned} P\{win\} &= P\{win \big| \text{both correct}\}p^2 + P\{win \big| \text{exactly 1 correct}\}2p(1-p) \\ &+ P\{win \big| \text{neither correct}\}(1-p)^2 \\ &= p^2 + p(1-p) + 0 = p \end{aligned}$$

Thus, both strategies give the same probability of winning.

Exercise 69.

In a certain village, it is traditional for the eldest son (or the older son in a two-son family) and his wife to be responsible for taking care of his parents as they age. In recent years, however, the women of this village, not wanting that responsibility, have not looked favorably upon marrying an eldest son.

(a) *If every family in the village has two children, what proportion of all sons are older sons?*

(b) *If every family in the village has three children, what proportion of all sons are eldest sons?*

Assume that each child is, independently, equally likely to be either a boy or a girl.

Solution. (a) The probability that a family has 2 sons is $1/4$; the probability that a family has exactly 1 son is $1/2$. Therefore, on average, every four families will have one family with 2 sons and two families with 1 son. Therefore, three out of every four sons will be eldest sons.

Another argument is to choose a child at random. Letting E be the event that the child is an eldest son, letting S be the event that it is a son, and letting A be the event that the child's family has at least one son,

$$P(E|S) = \frac{P(ES)}{P(S)} = 2P(E)$$

$$= 2\left[P(E|A)\frac{3}{4} + P(E|A^c)\frac{1}{4}\right] = 2\left[\frac{1}{2}\frac{3}{4} + 0\frac{1}{4}\right] = 3/4$$

Solution (contd...)

(*b*) Using the preceding notation

$$
\begin{aligned}
P(E|S) &= \frac{P(ES)}{P(S)} \\
&= 2P(E) \\
&= 2\left[P(E|A)\frac{7}{8} + P(E|A^c)\frac{1}{8}\right] \\
&= 2\left[\frac{1}{3}\frac{7}{8}\right] = 7/12
\end{aligned}
$$

Exercise 70.

In successive rolls of a pair of fair dice, what is the probability of getting 2 sevens before 6 even numbers?

Solution. Each roll that is either a 7 or an even number will be a 7 with probability

$$p = \frac{P(7)}{P(7) + P(\text{even})} = \frac{1/6}{1/6 + 1/2} = 1/4.$$

Hence, the desired probability is

$$\sum_{i=2}^{7} \binom{7}{i} (1/4)^i (3/4)^{7-i} = 1 - (3/4)^7 - 7(3/4)^6(1/4).$$

Exercise 71.

An urn contains 12 balls, of which 4 are white. Three players A, B, and C successively draw from the urn, A first, then B, then C, then A, and so on. The winner is the first one to draw a white ball. Find the probability of winning for each player if

(a) *each ball is replaced after it is drawn;*

(b) *the balls that are withdrawn are not replaced.*

Solution.

$$P(A \text{ wins}) = \frac{4}{12} + \frac{8}{12}\frac{7}{11}\frac{6}{10}\frac{4}{9} + \frac{8}{12}\frac{7}{11}\frac{6}{10}\frac{5}{9}\frac{4}{8}\frac{3}{7}\frac{4}{6} + \frac{8}{12}\frac{7}{11}\frac{6}{10}\frac{5}{9}\frac{4}{8}\frac{3}{7}$$

$$P(B \text{ wins}) = \frac{8}{12}\frac{4}{11} + \frac{8}{12}\frac{7}{11}\frac{6}{10}\frac{5}{9}\frac{4}{8} + \frac{8}{12}\frac{7}{11}\frac{6}{10}\frac{5}{9}\frac{4}{8}\frac{3}{7}\frac{2}{6}\frac{4}{5}$$

$$P(C \text{ wins}) = \frac{8}{12}\frac{7}{11}\frac{4}{10} + \frac{8}{12}\frac{7}{11}\frac{6}{10}\frac{5}{9}\frac{4}{8}\frac{4}{7} + \frac{8}{12}\frac{7}{11}\frac{6}{10}\frac{5}{9}\frac{4}{8}\frac{3}{7}\frac{2}{6}\frac{1}{5}$$

Exercise 72.

Suppose that n independent trials, each of which results in any of the outcomes $0, 1$, or 2, with respective probabilities p_0, p_1, and $p_2, \sum_{i=0}^{2} p_i = 1$, are performed. Find the probability that outcomes 1 and 2 both occur at least once.

Solution. Let N_i denote the event that none of the trials result in outcome $i, i = 1, 2$. Then

$$\begin{aligned} P(N_1 \cup N_2) &= P(N_1) + P(N_2) - P(N_1 N_2) \\ &= (1 - p_1)^n + (1 - p_2)^n - (1 - p_1 - p_2)^n. \end{aligned}$$

Hence, the probability that both outcomes occur at least once is $1 - (1 - p_1)^n - (1 - p_2)^n + (p_0)^n$.

Exercise 73.

Show that if $P(A) > 0$, *then*

$$P(AB|A) \geq P(AB|A \cup B)$$

Solution.

$$P(AB|A) = \frac{P(AB)}{P(A)} \geq \frac{P(AB)}{P(A \cup B)} = P(AB|A \cup B)$$

Exercise 74.

Let $A \subset B$. Express the following probabilities as simply as possible:

$$P(A|B), \quad P(A|B^c), \quad P(B|A), \quad P(B|A^c)$$

Solution. If $A \subset B$, then

$$P(A|B) = \frac{P(A)}{P(B)}, P(A|B^c) = 0, \quad P(B|A) = 1, \quad P(B|A^c) = \frac{P(BA^c)}{P(A^c)}.$$

Exercise 75.

An event F is said to carry negative information about an event E, and we write $F \searrow E$, if

$$P(E|F) \leq P(E).$$

Prove or give counterexamples to the following assertions:

(a) *If $F \searrow E$, then $E \searrow F$.*

(b) *If $F \searrow E$ and $E \searrow G$, then $F \searrow G$.*

(c) *If $F \searrow E$ and $G \searrow E$, then $FG \searrow E$.*

Repeat parts $(a), (b)$, and (c) when $\searrow$ is replaced by $\nearrow$, where we say that F carries positive information about E, written $F \nearrow E$, when $P(E|F) \geq P(E)$.

Solution. None are true.

Exercise 76.

Independent trials that result in a success with probability p are successively performed until a total of r successes is obtained. Show that the probability that exactly n trials are required is

$$\binom{n-1}{r-1} p^r (1-p)^{n-r}.$$

Solution.

$$\begin{aligned} P\{r \text{ successes before } m \text{ failures}\} &= P\{r^{th} \text{ success occurs before trail } m+r\} \\ &= \sum_{n=r}^{m+r-1} \binom{n-1}{r-1} p^r (1-p)^{n-r}. \end{aligned}$$

Exercise 77.

Independent trials that result in a success with probability p and a failure with probability $1-p$ are called Bernoulli trials. Let P_n denote the probability that n Bernoulli trials result in an even number of successes (0 being considered an even number). Show that

$$P_n = p(1 - P_{n-1}) + (1-p)P_{n-1} n \quad \geq 1$$

and use this formula to prove (by induction) that

$$P_n = \frac{1 + (1-2p)^n}{2}.$$

Solution. If the first trial is a success, then the remaining $n-1$ must result in an odd number of successes, whereas if it is a failure, then the remaining $n-1$ must result in an even number of successes.

Exercise 78.

Prove directly that

$$P(E|F) = P(E|FG)P(G|F) + P(E|FG^c)P(G^c|F).$$

Solution.

$$P(E\big|F) = P(EF)/P(F)$$

$$P(E\big|FG)P(G\big|F) = \frac{P(EFG)}{P(FG)}\frac{P(FG)}{P(F)} = \frac{P(EFG)}{P(F)}$$

$$P(E\big|FG^c)P(G^c\big|F) = \frac{P(EFG^c)}{P(F)}.$$

The result now follows since

$$P(EF) = P(EFG) + P(EFG^c).$$

References

1. Meyer, *Introductory Probability and Statistical Applications*, 2nd Edition, Oxford & IBH Publishing Company, India.
2. Sheldon Ross, *First Course in Probability*, Sixth Edition, Pearson Publisher, India.
3. Dimitri P. Bertsekas and John N. Tsitsiklis, *Introduction to Probability*, Athena Scientific, Belmont, Massachusetts.