Probability and Statistics
(ENM 503)

Michael A. Carchidi

October 8, 2018

Chapter 4 - Conditional Probability and Independence

The following notes are based on the textbook entitled: *A First Course in Probability* by Sheldon Ross (9th edition) and these notes can be viewed at

https://canvas.upenn.edu/

after you log in using your PennKey user name and Password.

1. Introduction

In this chapter we shall introduce one of the most important concepts in probability, that of *conditional probability*. The importance of this concept 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 and in such a situation, the desired probabilities are called conditional, but it should be noted that extra knowledge does not always mean higher probabilities. Second, and more important, even when no partial information is available, conditional probabilities can often be used to compute desired (unconditional) probabilities more easily.

Example #1: Conditioning Doesn't Always Mean Higher Probabilities

Suppose that we have two six-sided dice and we want to compute the probability that when both are rolled, the total adds to 6. Since the sample space of rolling two six-sided dice consists of the 36 equally likely events

$$\begin{aligned} S \;=\; \{&(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\ &(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\ &(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\ &(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\ &(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\ &(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\} \end{aligned}$$

with

$$E_6 = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$$

being those that add to six, we have

$$P(\text{sum adds to 6}) = \frac{|E_6|}{|S|} = \frac{5}{36}.$$

Now suppose we know that the first die rolled a 4. Then the sample space S is *reduced* to

$$S' = \{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\}$$

and of them, only

$$E_6' = E_6 \cap S' = \{(4,2)\}$$

adds to six and so the probability that the sum adds to six, *given* that (or conditioned on the fact that) the roll of the first die is 4, is

$$P(\text{sum adds to 6 given that the roll of the first die is 4}) = \frac{|E_6'|}{|S'|} = \frac{1}{6}$$

which is larger than $5/36$. Suppose instead that we know that the first die rolled a 6. Then the sample space S is now reduced to

$$S'' = \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$$

and since,

$$E_6'' = E_6 \cap S'' = \emptyset,$$

we see now that none of them add to six and so the probability that the sum adds to six, *given* that (or conditioned on the fact that) the roll of the first die is 6, is

$$P(\text{sum adds to 6 given that the roll of the first die is 6}) = \frac{|E_6''|}{|S''|} = \frac{0}{6} = 0$$

which is now smaller than 5/36. ■

Example #2: Conditioning Doesn't Always Mean Higher Probabilities

Suppose that we have two six-sided dice and we want compute the probability that when both are rolled, the total adds to 11. Since the sample space of rolling two six-sided dice consists of the 36 equally likely events

$$\begin{aligned} S &= \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6), \\ &\quad (2,1),(2,2),(2,3),(2,4),(2,5),(2,6), \\ &\quad (3,1),(3,2),(3,3),(3,4),(3,5),(3,6), \\ &\quad (4,1),(4,2),(4,3),(4,4),(4,5),(4,6), \\ &\quad (5,1),(5,2),(5,3),(5,4),(5,5),(5,6), \\ &\quad (6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\} \end{aligned}$$

with

$$E_{11} = \{(5,6),(6,5)\}$$

being those that add to 11, we have

$$P(\text{sum adds to } 11) = \frac{|E_{11}|}{|S|} = \frac{2}{36} = \frac{1}{18}.$$

Now suppose we know that the first die rolled at 4. Then the sample space S is reduced to

$$S' = \{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\}$$

and of them, none add to 11 so that

$$E'_{11} = E_{11} \cap S' = \emptyset,$$

and so the probability that the sum adds to 11, given that the roll of the first die is 4, is

$$P(\text{sum adds to 11 given that the roll of the first die is 4}) = \frac{|E'_{11}|}{|S'|} = \frac{0}{6}$$

showing, as we would expect that there is no way for the sum to add to 11 if the first die shows a 4. ■

2. Conditional Probabilities

Given two events A and B, we let $P(A|B)$ denote the probability that A occurs *given* that (or conditioned on the face that) B has occurred. To see how this can be computed, suppose that the sample space of an experiment is S. Knowing the B has occurred *reduces* the sample space S to

$$S' = B \cap S = B$$

since $B \subseteq S$. If B has occurred and we now want A to also occur, then this is implying that both A and B (or $A \cap B$) must occur. This would suggest that $P(A|B)$ should equal $P(A \cap B)$, but only with respect to (or relative to) the reduced sample space S', rather than with respect to the complete sample space S. This says that instead of measuring $P(A \cap B)$ relative to $P(S)$, (which equals one since $P(S) = 1$), we should measure $P(A \cap B)$ relative to $P(S') = P(B)$ which is generally less that one. Thus we are motivated to *defined* $P(A|B)$ as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{1}$$

Of course, the occurrence of B guarantees that $P(B) > 0$ so we need not worry about dividing by zero.

Some "Common Sense" Checks

Note that if $B = S$, we have

$$P(A|S) = \frac{P(A \cap S)}{P(S)} = \frac{P(A)}{1} = P(A) \tag{2a}$$

as it should since saying that S has occurred is simply saying that an experiment has been performed without saying any outcome of the experiment. Also we note that if $B = A$, then

$$P(A|A) = \frac{P(A \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1 \tag{2b}$$

which also makes sense since the probability of having A occur, given that it has already occurred, should be one. In a similar way, if $B = \bar{A}$, then

$$P(A|\bar{A}) = \frac{P(A \cap \bar{A})}{P(\bar{A})} = \frac{P(\emptyset)}{P(\bar{A})} = 0 \tag{2c}$$

which also makes sense since the probability of having A occur, given that it A already has not occurred, must be zero. As a final note for now, starting with

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

we use the fact that

$$(\bar{A} \cap B) \cup (A \cap B) = (\bar{A} \cup A) \cap B = S \cap B = B$$

and

$$(\bar{A} \cap B) \cap (A \cap B) = (\bar{A} \cap A) \cap (B\cap) = \emptyset \cap B = \emptyset$$

we see that $\bar{A} \cap B$ and $A \cap B$ make up a 2-set partition of B, and so

$$P(\bar{A} \cap B) + P(A \cap B) = P(B) \qquad \text{or} \qquad P(\bar{A} \cap B) = P(B) - P(A \cap B).$$

Putting this into the above equation for $P(\bar{A}|B)$, we see that

$$P(\bar{A}|B) = \frac{P(B) - P(A \cap B)}{P(B)} = 1 - \frac{P(A \cap B)}{P(B)}$$

which says that

$$P(\bar{A}|B) = 1 - P(A|B), \tag{2d}$$

another result that should be expected.

In fact, we shall show later that the conditional probabilities $P(\bullet|B)$ do satisfy the axioms of probability and so all results that come from these axioms, such as Equation (2b,c,d) must follow, but this is done only after we develope a few properties of $P(\bullet|B)$ and so we shall provide this later in the chapter. But first, let us investigate the consistency of $P(\bullet|B)$ with the long-term relative frequency idea of probability and then we shall look at some examples.

Consistency of $P(A|B)$ With the Long-Term Relative Frequency

Note that our definition of $P(A|B)$ is consistent with the interpretation of probability as being a long-term relative frequency. To see this, suppose that n repetitions of an experiment are to be performed, where n is large. We claim that if we consider only those experiments in which B occurs, then $P(A|B)$ will equal the long-run proportion of them in which A also occurs. To verify this statement, note that since $P(B)$ is the long-term proportion of experiments in which B occurs, it follows that in n repetitions of the experiment (for large n), B should occur approximately $n \times P(B)$ times. Similarly, in approximately $n \times P(A \cap B)$ of these experiments, both A and B will occur. Hence, out of the approximately $n \times P(B)$ experiments in which B occurs, the proportion of them in which A also occurs is approximately equal to

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

and because this probability becomes exact as n becomes larger and larger, we have the appropriate definition of $P(A|B)$.

Example #3: A Surprise

Suppose that Joe is 40% certain that his missing key is in the left-hand pocket of his hanging jacket and he is 40% certain that it is in the right-hand pocket of the same jacket. Since the missing key cannot be in both pockets, we may now say that Joe must be $(40 + 40 - 0)\% = 80\%$ certain that his missing key is in one of the two pockets in his hanging jacket. If the search of the left-hand pocket does not find the key, what is the probability that the key is in the other pocket?

To answer this, let L be the event that the missing key is in the left-hand pocket of the jacket and let R be the event that the missing key is in the right-hand pocket of the jacket. Then we have $P(L) = 0.4$ and $P(R) = 0.4$. Now we want to compute

$$P(R|\bar{L}) = \frac{P(R \cap \bar{L})}{P(\bar{L})} = \frac{P(R \cap \bar{L})}{1 - P(L)}.$$

But $\bar{L}$ means that the key is either in the right-hand pocket R or it is in neither pocket (N), so that $\bar{L} = R \cup N$, and so

$$R \cap \bar{L} = R \cap (R \cup N) = (R \cap R) \cup (R \cap N) = R \cup \emptyset = R.$$

Thus we find that

$$P(R|\bar{L}) = \frac{P(R)}{1 - P(L)} = \frac{0.4}{1 - 0.4} = \frac{2}{3}.$$

Many students would think that $P(R|\bar{L}) = P(R) = 0.4$ by reasoning that knowing that the key is not in the left pocket means if it is in the jacket, then it must be in the right pocket and Joe is only 40% certain that the key is in the right pocket. But remember that Joe is 40% certain that the key is in his right-hand pocket only because he is uncertain that the key is in the left pocket. By now knowing that the key is not in the left pocket, one would think that he should be *more certain* that the missing key is in the right pocket, and he is since $2/3 > 0.4$. ■

Example #4: Another Surprise

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 probability that both flips land on heads, given that: (a) the first flip lands on heads and (b) at least one flip lands on heads?

First 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. It should be clear that

$$P(B) = \frac{|B|}{|S|} = \frac{1}{4} \quad , \quad P(F) = \frac{|F|}{|S|} = \frac{2}{4} \quad \text{and} \quad P(A) = \frac{|A|}{|S|} = \frac{3}{4}.$$

To solve (a), we want

$$P(B|F) = \frac{P(B \cap F)}{P(F)} = \frac{P(B)}{P(F)} = \frac{1/4}{2/4} = \frac{1}{2},$$

which is not a surprise. To solve (b), we want

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(B)}{P(A)} = \frac{1/4}{3/4} = \frac{1}{3} < P(B|F),$$

and many students initially find this 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 lands on heads. Their mistake is assuming that these two events

$$E_1 = \{\text{They both land on heads.}\} \quad \text{and} \quad E_2 = \{\text{Only one lands on heads.}\}$$

are equally likely since they are not. Initially, all four events (H, H), (H, T), (T, H) and (T, T) are equally likely. Because the information that at least one flip lands on heads is equivalent to the outcome that is not (T, T) and we are left with the reduced space consisting of the 3 equally likely outcomes (H, H), (H, T) and (T, H), only one of which results in both landing on heads, so that the probability is $1/3$. ■

Example #5: The Famous Monte-Hall Problem

The Monte-Hall Show was a game show which aired from 1963 to 1976. The end of the show always had a contestant who was given the choice of picking one of three curtains and receiving the "prize" behind the curtain they picked. Behind one curtain was always a very nice prize (such as a brand new car) and behind the other two curtains was always a gag prize (such as a old goat). The player of the game would choose a curtain, say curtain #1, and the host, who knows what's behind the other two curtains, would reveal what was behind, say curtain #3, which has a goat. Of course, the host would never reveal the curtain having the car because then the player of the game would know from the start that they did not win the car. The game show host would then say to the contestant playing the game, "Now that you know that the car is not behind curtain #3, it must be behind the curtain you originally selected (#1) or behind the curtain you did not originally select (#2). Would you like to switch your choice from winning the prize behind curtain #1 to winning the prize behind curtain #2?" The question is, "Is it to the player's advantage to switch their choice to curtain #2 or to stick with their original choice of curtain #1, or does it not matter?

Many people who don't know about conditioned probability have incorrectly solved this by arguing in the following way. What is behind curtain #3 has been revealed showing a goat, therefore there are now only two curtains left, namely the original curtain picked (#1) and the other curtain that has not yet been revealed (#2) and the car is certainly behind one of these. Therefore, the probability that the car is behind curtain #1 is $1/2$ and the probability the car is behind curtain #2 is also $1/2$ and so there is no advantage to switching since the player's probability of winning the car is 1/2 whether they stay with the original curtain selected (#1) or switch to the curtain #2.

Many other people who know something about conditional probability (but don't know it correctly) would say the following. Let 1, 2 and 3 be the events that the car is behind curtain #1, #2 and #3, respectively, and we know that when the contestant is initially asked to choose a curtain,

$$P(1) = P(2) = P(3) = \frac{1}{3},$$

since the car is equally likely to be behind one of the three curtains in the first place. Then, using the equations of conditional probability, they would say that

$$P(1|\bar{3}) = \frac{P(1 \cap \bar{3})}{P(\bar{3})} = \frac{P(1)}{1 - P(3)} = \frac{1/3}{1 - 1/3} = \frac{1}{2}$$

and

$$P(2|\bar{3}) = \frac{P(2 \cap \bar{3})}{P(\bar{3})} = \frac{P(2)}{1 - P(3)} = \frac{1/3}{1 - 1/3} = \frac{1}{2}$$

showing that the probabilities that the car is behind curtain #1 or curtain #2, given that it is not behind curtain #3 are both 1/2 and so (once again) there is no advantage to switching.

Now this line of reasoning would be correct *if* (and this is a "big if") the game show host shows you what is behind curtain #3 *before* asking you to choose one of the three curtains. This is what computing $P(1|\bar{3})$ and $P(2|\bar{3})$ gives, because $P(A|B)$ means the probability that event A occurs *given that we already know* that event B has occurred. But when the player of the game is asked to initially choose a curtain, the player of the game does not know that curtain #3 contains a goat behind it and so they had the option of choosing curtain #3 (as well as #1 and #2) in the beginning, and therefore they have a 1/3 probability of winning the car using their initial selection.

If the player of the game had been asked to choose a curtain *after* curtain #3 was revealed to them, then they certainly would not have selected curtain #3 and would have selected either curtain #1 or curtain #2, and they would have a 1/2 probability of winning the car. However, the game show host does not do this. He has the player of the game choose a curtain first and this gives the player a 1/3 probability of choosing the curtain that has the car behind it and a 2/3 probability of choosing a curtain that has a goat behind it. After the player of the game has already chosen a curtain, what the game show host does is now reveal a goat behind one of the other two remaining curtains but the player still had only a 1/3 chance of winning the car. The game show host gives information to player of the game only *after* the player has made their choice, *not before*. Basically, by revealing one of the curtains and asking if the player of the game wants to switch, the game show host is allowing the player to actually pick two of the three curtains, namely the one they have already chosen and one that was revealed to them by the host, and this therefore gives the player of the game a 2/3 probability of winning the car but only if they always choose to switch, *i.e.*, only if they take advantage of the new information revealed to them by the host. Therefore, the player, by always switching their choice, now has a 2/3 chance of winning the car and so it is always to the player's advantage to switch.

Still not convinced? Then the following argument is simple and indisputable. Suppose the player of the game *always chooses not to switch*, *i.e.*, not to use the new information revealed to them by the host, then the only way they will win the car is if they choose the winning curtain right from the start and this has a 1/3 probability of occurring. Therefore if the player's strategy is never to switch curtains, their chances of winning the car is 1/3. Hence, if the player's strategy is always to switch curtains, their chances of winning the car must be $1 - 1/3 = 2/3$ since they can either choose to switch or not choose to switch. We shall run a simple simulation of this later substantiating that 2/3 is indeed the correct answer to the problem. ■

Example #6: Working With the Reduced Sample Space

In the card game of 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?

Probably the easiest way to solve this is by using the reduced sample space consisting of $52 - 13(\text{North}) - 13(\text{South}) = 26$ possible cards for East (E) and West (W), of which 5 are spades and $26 - 5 = 21$ are not spades, and so

$$P = \frac{\binom{5}{3}\binom{21}{10}}{\binom{26}{13}} = \frac{10 \times 352716}{10400600} = \frac{39}{115}$$

or $P \simeq 0.339$. ■

Probability Trees

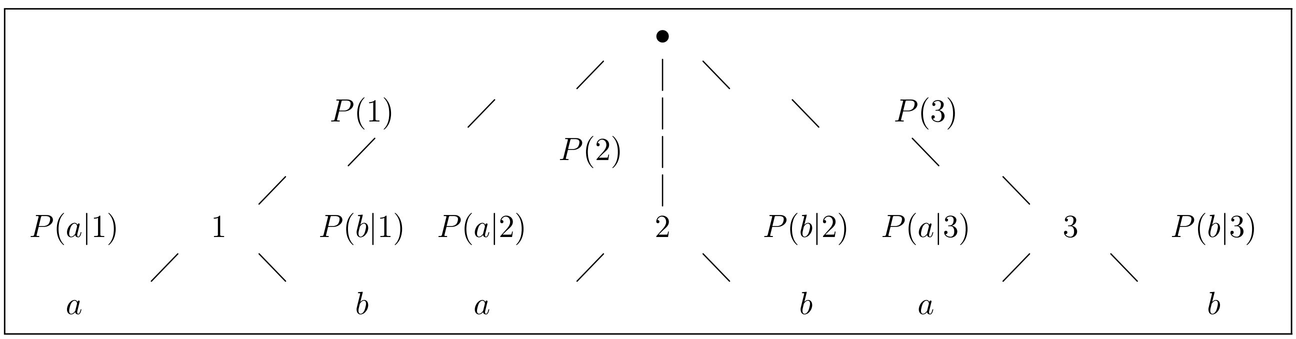
Conditional probabilities are often represented using probability trees. For example, suppose that two experiments E_1 and E_2 are to be performed and suppose that the possible outcomes of these are given by the two sample spaces

$$S_1 = \{1, 2, 3\} \quad \text{and} \quad S_2 = \{a, b\}$$

and suppose we are interested in performing E_1 followed by E_2. Then the probabilities $P(1)$, $P(2)$ and $P(3)$ and the conditional probabilities,

$$P(a|1) \quad , \quad P(a|2) \quad , \quad P(a|3) \quad , \quad P(b|1) \quad , \quad P(b|2) \quad , \quad P(b|3),$$

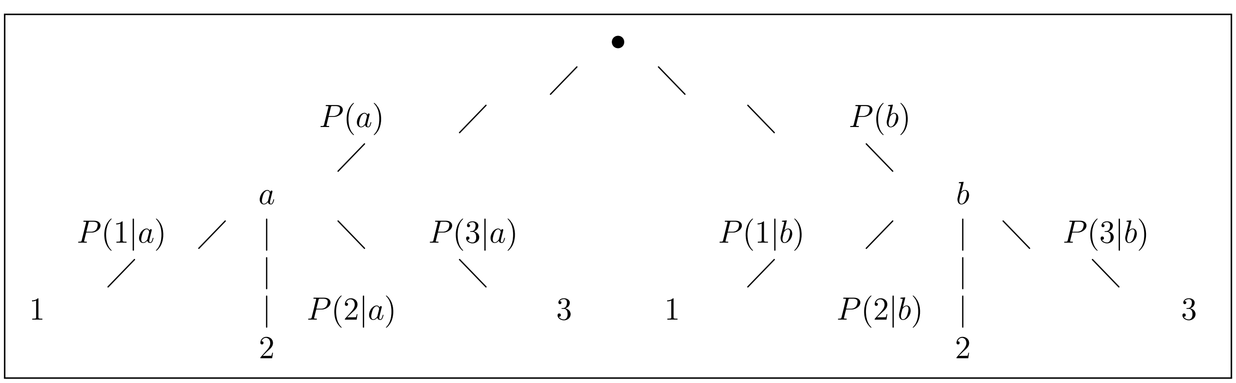
may be represented using a probability tree as illustrated in the following figure.



If we are interested performing E_2 followed by E_1, then we may represent the probabilities $P(a)$, $P(b)$ and the conditional probabilities,

$$P(1|a) \quad , \quad P(2|a) \quad , \quad P(3|a) \quad , \quad P(1|b) \quad , \quad P(2|b) \quad , \quad P(3|b),$$

using a probability tree as illustrated in the following figure.



We shall see that all of $P(1)$, $P(2)$ and $P(3)$, the conditional probabilities,

$$P(a|1) \quad , \quad P(a|2) \quad , \quad P(a|3) \quad , \quad P(b|1) \quad , \quad P(b|2) \quad , \quad P(b|3),$$

the probabilities, $P(a)$, $P(b)$ and the conditional probabilities,

$$P(1|a) \quad , \quad P(2|a) \quad , \quad P(3|a) \quad , \quad P(1|b) \quad , \quad P(2|b) \quad , \quad P(3|b),$$

are all related by the theorem of total probability and Bayes's theorem. To see what these theorems are all about, let use first compute $P(A \cap B)$.

Computing $P(A \cap B)$

We should note that Equation (1) may be rewritten as

$$P(A \cap B) = P(A|B)P(B) \qquad \text{and} \qquad P(B \cap A) = P(B|A)P(A)$$

so that

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) \tag{3}$$

which allows for a calculation of $P(A \cap B)$. Note also that Equation (3) leads to

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}. \tag{4}$$

which allows for the computation of $P(B|A)$ from $P(A|B)$. This last result can be used to invert probability trees as we shall shortly see but first let us look at some more examples.

Example #7: Computing $P(A \cap B)$

Celine is undecided as to whether to take a French course or a Chemistry course. She estimates that her probability of earning an A grade would be 1/2 in the French course and 2/3 in the 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 the Chemistry course?

To answer this, let C be the event that she takes Chemistry and let A be the event that she receives an A in whatever course she takes. Then the desired probability is $P(A \cap C)$, which we may calculate using

$$P(A \cap C) = P(A|C)P(C) = \frac{2}{3} \times \frac{1}{2} = \frac{1}{3}.$$

Note that in using $P(C) = 1/2$, we are using the flip of a fair coin. ■

Example #8: Giving Weights to Balls in an Urn

Suppose that an urn contains 8 red balls and 4 white balls. We draw 2 balls from the urn without replacement. 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?

To answer this we let R_1 be the event that the first ball drawn is red and R_2 be the event that the second ball drawn is red. Then

$$P(R_1 \cap R_2) = P(R_2 \cap R_1) = P(R_2|R_1)P(R_1)$$

which says that

$$P(R_1 \cap R_2) = \frac{7}{7+4} \times \frac{8}{8+4} = \frac{14}{33}.$$

As a check, we note that

$$P(R_1 \cap R_2) = \frac{\binom{8}{2}\binom{4}{0}}{\binom{8+4}{2}} = \frac{28}{66} = \frac{14}{33}.$$

Now suppose that the balls have different weights, with each red ball having weight r and each white ball having weight w. Also suppose that the probability that a given ball in the urn is the next one selected is it 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?

Using the approach, we have

$$P(R_1 \cap R_2) = P(R_2|R_1)P(R_1) = \frac{7r}{7r+4w} \times \frac{8r}{8r+4w} = \frac{14r^2}{(7r+4w)(2r+w)}$$

but note that the second approach using the binomial coefficients (when all balls had the same weight) does not work now since the events are no longer equally likely. ■

The Multiplication Rule

We may extend Equation (3) to read

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

or

$$P(A \cap B \cap C) = P(A|B \cap C)P(B|C)P(C) \tag{5a}$$

and since $A \cap B \cap C = C \cap B \cap A$, we may also write Equation (5a) as

$$P(C \cap B \cap A) = P(A)P(B|A)P(C|B \cap A), \tag{5b}$$

and it should be clear how we may extend Equation (5a) to computing

$$\begin{aligned} & P(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n) \\ = \; & P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \\ & \times \cdots \times P(A_{n-1}|A_1 \cap A_2 \cap \cdots \cap A_{n-2})P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \end{aligned} \tag{6a}$$

or

$$\begin{aligned} & P(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n) \\ = \; & P(A_1|A_2 \cap A_3 \cap \cdots \cap A_n)P(A_2|A_3 \cap \cdots \cap A_n) \\ & \times \cdots \times P(A_{n-2}|A_{n-1} \cap A_n)P(A_{n-1}|A_n)P(A_n) \end{aligned} \tag{6b}$$

which is known as the *multiplication rule*.

Example #9: An Example with Playing Cards

An ordinary deck of 52 playing cards is randomly divided into 4 piles of 13 cards each. Compute the probability that each pile has exactly one ace.

To answer this let us define the following four events:

$$\begin{aligned} E_{\text{S}} &= \{\text{the ace of spades is in any one of the piles}\} \\ E_{\text{SH}} &= \{\text{the ace of spades and the ace of hearts are in } \textit{different} \text{ piles}\} \\ E_{\text{SHD}} &= \{\text{the ace of spades, hearts and diamonds are in } \textit{different} \text{ piles}\} \\ E_{\text{SHDC}} &= \{\text{all fours aces are in } \textit{different} \text{ piles}\}. \end{aligned}$$

We desire to compute $P(E_{\text{SHDC}})$ and we note that

$$E_{\text{SHDC}} = E_{\text{SHDC}} \cap E_{\text{SHD}} \cap E_{\text{SH}} \cap E_{\text{S}}$$

since

$$E_{\text{SHDC}} \subset E_{\text{SHD}} \subset E_{\text{SH}} \subset E_{\text{S}}$$

and so

$$\begin{aligned} P(E_{\text{SHDC}} \cap E_{\text{SHD}} \cap E_{\text{SH}} \cap E_{\text{S}}) &= P(E_{\text{SHDC}}|E_{\text{SHD}} \cap E_{\text{SH}} \cap E_{\text{S}}) \\ &\quad \times P(E_{\text{SHD}}|E_{\text{SH}} \cap E_{\text{S}}) P(E_{\text{SH}}|E_{\text{S}}) P(E_{\text{S}}) \end{aligned}$$

or, since $E_{\text{SHDC}} \cap E_{\text{SHD}} \cap E_{\text{SH}} \cap E_{\text{S}} = E_{\text{SHDC}}$, $E_{\text{SHD}} \cap E_{\text{SH}} \cap E_{\text{S}} = E_{\text{SHD}}$, and $E_{\text{SH}} \cap E_{\text{S}} = E_{\text{SH}}$, we find that

$$P(E_{\text{SHDC}}) = P(E_{\text{SHDC}}|E_{\text{SHD}}) P(E_{\text{SHD}}|E_{\text{SH}}) P(E_{\text{SH}}|E_{\text{S}}) P(E_{\text{S}})$$

We first note that $P(E_{\text{S}}) = 1$ since the ace of spades must be in one of the four piles.

Computing $P(E_{\text{SH}}|E_{\text{S}})$ By First Computing $P(\bar{E}_{\text{SH}}|E_{\text{S}})$

To determine $P(E_{\text{SH}}|E_{\text{S}})$, we may first compute $P(\bar{E}_{\text{SH}}|E_{\text{S}})$ by considering the pile that contains the ace of spades (E_{S}). Because the remaining 12 cards (in this pile) are equally likely to be any 12 of the remaining $52 - 1 = 51$ cards (excluding the ace of spades), the probability that the ace of hearts is among them $P(\bar{E}_{\text{SH}}|E_{\text{S}})$, is

$$\frac{\begin{pmatrix} \text{number of ways} \\ \text{of picking the} \\ \text{ace of hearts} \end{pmatrix} \times \begin{pmatrix} \text{number of ways of} \\ \text{picking the other 11} \\ \text{cards out of 50} \end{pmatrix}}{\begin{pmatrix} \text{number of ways} \\ \text{of picking all 12} \\ \text{cards out of 51} \end{pmatrix}} = \frac{\binom{1}{1}\binom{50}{11}}{\binom{51}{12}} = \frac{4}{17}$$

which says that the probability that the ace of hearts is NOT among them is

$$P(E_{\text{SH}}|E_{\text{S}}) = 1 - P(\bar{E}_{\text{SH}}|E_{\text{S}}) = 1 - \frac{4}{17} = \frac{13}{17}.$$

Note that we may compute $P(E_{\mathrm{SH}}|E_{\mathrm{S}})$ directly by saying that the ace of spades is in one pile that has 13 cards. For the ace of hearts to be in a different pile, we may imagine it being placed in a big pile of $52 - 13 = 39$ cards (chosen from $52 - 1 = 51$ cards), with 1 card being the ace of hearts and the other 38 being chosen from $52 - 1 - 1 = 50$ cards and so

$$P(E_{\mathrm{SH}}|E_{\mathrm{S}}) = \frac{\begin{pmatrix}\text{number of ways}\\ \text{of picking the}\\ \text{ace of hearts}\end{pmatrix} \times \begin{pmatrix}\text{number of ways of}\\ \text{picking the other 38}\\ \text{cards out of 50}\end{pmatrix}}{\begin{pmatrix}\text{number of ways}\\ \text{of picking all 39}\\ \text{cards out of 51}\end{pmatrix}}$$

or

$$P(E_{\mathrm{SH}}|E_{\mathrm{S}}) = \frac{\binom{1}{1}\binom{50}{38}}{\binom{51}{39}} = \frac{13}{17},$$

which agrees with the earlier result.

Computing $P(E_{\mathrm{SHD}}|E_{\mathrm{SH}})$ By First Computing $P(\bar{E}_{\mathrm{SHD}}|E_{\mathrm{SH}})$

Given next that the ace of spades and the ace of hearts are now in different piles (E_{SH}), we compute the probability that the ace of diamonds is in *neither* of these two piles by first computing the probability that the ace of diamonds is in *either* of these two piles. Since the remaining $12 + 12 = 24$ cards (excluding the ace of spades and the ace of hearts) are equally likely to be any of the remaining $52 - 2 = 50 = 50$ cards (excluding the ace of spades and the ace of hearts), we see that the probability that the ace of diamonds is among these two piles is

$$\frac{\begin{pmatrix}\text{number of ways}\\ \text{of picking the}\\ \text{ace of diamonds}\end{pmatrix} \times \begin{pmatrix}\text{number of ways of}\\ \text{picking the other 23}\\ \text{cards out of 49}\end{pmatrix}}{\begin{pmatrix}\text{number of ways}\\ \text{of picking all 24}\\ \text{cards out of 50}\end{pmatrix}} = \frac{\binom{1}{1}\binom{49}{23}}{\binom{50}{24}} = \frac{12}{25}$$

which says that the probability that the ace of diamonds is NOT among them is

$$P(E_{\mathrm{SHD}}|E_{\mathrm{SH}}) = 1 - P(\bar{E}_{\mathrm{SHD}}|E_{\mathrm{SH}}) = 1 - \frac{12}{25} = \frac{13}{25}.$$

Note that we may compute $P(E_{\mathrm{SHD}}|E_{\mathrm{SH}})$ directly by saying that the ace of spades is in one pile that has 13 cards and the ace of hearts is in another pile that has 13 cards. For the ace of diamonds to be in a different pile, we may imagine it being placed in a big pile of $52-13-13=26$ cards (chosen from $52-1-1=50$ cards), with 1 card being the ace of diamonds and the other 25 being chosen from $52-1-1-1=49$ cards and so

$$P(E_{\mathrm{SHD}}|E_{\mathrm{SH}}) = \frac{\begin{pmatrix}\text{number of ways}\\ \text{of picking the}\\ \text{ace of diamonds}\end{pmatrix} \times \begin{pmatrix}\text{number of ways of}\\ \text{picking the other 25}\\ \text{cards out of 49}\end{pmatrix}}{\begin{pmatrix}\text{number of ways}\\ \text{of picking all 26}\\ \text{cards out of 50}\end{pmatrix}}$$

or

$$P(E_{\mathrm{SHD}}|E_{\mathrm{SH}}) = \frac{\binom{1}{1}\binom{49}{25}}{\binom{50}{26}} = \frac{13}{25},$$

which agrees with the earlier result.

Computing $P(E_{\mathrm{SHDC}}|E_{\mathrm{SHD}})$ By First Computing $P(\bar{E}_{\mathrm{SHDC}}|E_{\mathrm{SHD}})$

Finally, given that the ace of spades, hearts and diamonds are in three different piles (E_{SHD}), the probability that the ace of clubs is in *neither* of these piles is computing by first computing the probability that the ace of clubs is in *either* of these piles. Since the remaining $12+12+12=36$ cards are equally likely to be any of the remaining $52-3=49$ cards (excluding the ace of spades, the ace of hearts and the ace of diamonds), and so the probability that the ace of clubs is among them is

$$\frac{\begin{pmatrix}\text{number of ways}\\ \text{of picking the}\\ \text{ace of clubs}\end{pmatrix} \times \begin{pmatrix}\text{number of ways of}\\ \text{picking the other 35}\\ \text{cards out of 48}\end{pmatrix}}{\begin{pmatrix}\text{number of ways}\\ \text{of picking all 36}\\ \text{cards out of 49}\end{pmatrix}} = \frac{\binom{1}{1}\binom{48}{35}}{\binom{49}{36}} = \frac{36}{49}$$

which says that the probability that the ace of clubs is NOT among them is

$$P(E_{\mathrm{SHDC}}|E_{\mathrm{SHD}}) = 1 - P(\bar{E}_{\mathrm{SHDC}}|E_{\mathrm{SHD}}) = 1 - \frac{36}{49} = \frac{13}{49}.$$

Note that we may compute $P(E_{\mathrm{SHDC}}|E_{\mathrm{SHD}})$ directly by saying that the ace of spades is in one pile that has 13 cards, the ace of hearts is in another pile that has 13 cards and the ace of diamonds is in a third pile have 13 cards. For the ace of clubs to be in a different pile, we may imagine it being placed in a pile of $52-13-13-13=13$ cards (chosen from $52-1-1-1=49$ cards), with 1 card being the ace of clubs and the other 12 being chosen from $52-1-1-1-1=48$ cards and so

$$P(E_{\mathrm{SHDC}}|E_{\mathrm{SHD}}) = \frac{\begin{pmatrix}\text{number of ways}\\ \text{of picking the}\\ \text{ace of clubs}\end{pmatrix} \times \begin{pmatrix}\text{number of ways of}\\ \text{picking the other 12}\\ \text{cards out of 48}\end{pmatrix}}{\begin{pmatrix}\text{number of ways}\\ \text{of picking all 13}\\ \text{cards out of 49}\end{pmatrix}}$$

or

$$P(E_{\mathrm{SHDC}}|E_{\mathrm{SHD}}) = \frac{\binom{1}{1}\binom{48}{12}}{\binom{49}{13}} = \frac{13}{49},$$

which agrees with the earlier result.

Computing $P(E_{\mathrm{SHDC}})$ By Putting It All Together

Using the above equation

$$P(E_{\mathrm{SHDC}}) = P(E_{\mathrm{SHDC}}|E_{\mathrm{SHD}})P(E_{\mathrm{SHD}}|E_{\mathrm{SH}})P(E_{\mathrm{SH}}|E_{\mathrm{S}})P(E_{\mathrm{S}})$$

we then have

$$P(E_{\mathrm{SHDC}}) = \frac{13}{49} \times \frac{13}{25} \times \frac{13}{17} \times 1 = \frac{2197}{20825}$$

or $P(E_{\mathrm{SHDC}}) \simeq 0.1055$. ■

3. Partitions and The Theorem of Total Probability

A set of n *non-empty* events $P_{\mathrm{S}} = \{B_1, B_2, B_3, ..., B_n\}$ form a partition of a sample space S if: (a) $B_i \cap B_j = \emptyset$ for all $i \neq j$ and (b) $B_1 \cup B_2 \cup B_3 \cup \cdots \cup B_n = S$. This says that: (a) any two such events are disjoint and (b) the union of all such events is the entire sample space. Thus, whenever an experiment E (of S) is performed, *one and only one* of the events B_i ($i = 1, 2, 3, ..., n$) will occur. It should be clear that if $P_S = \{B_1, B_2, B_3, ..., B_n\}$ represents a partition of a sample space S, and if A is any event, then

$$A = (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3) \cup \cdots \cup (A \cap B_n)$$

and

$$(A \cap B_i) \cap (A \cap B_j) = (A \cap A) \cap (B_i \cap B_j) = A \cap \emptyset = \emptyset$$

for $i \neq j$. Note that the set $\{A \cap B_1, A \cap B_2, A \cap B_3, ..., A \cap B_n\}$ is not necessarily a partition of A since it is possible that $A \cap B_j$ could be empty for some choice (or choices) of $j = 1, 2, 3, ..., n$. However, we may still write that

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3) + \cdots + P(A \cap B_n)$$

or

$$P(A) = \sum_{j=1}^{n} P(A \cap B_j) \tag{7a}$$

since $P(\emptyset) = 0$. Using the fact that $P(A \cap B_j) = P(A|B_j)P(B_j)$, we also may write Equation (7a) as

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

or

$$P(A) = \sum_{j=1}^{n} P(A|B_j)P(B_j). \tag{7b}$$

Equations (7a) and (7b) are known as the *theorems (or laws) of total probability.* As a special case, we note that since $P_S = \{B, \bar{B}\}$, for any non-empty proper event $B \subset S$ forms a partition of S, and so

$$P(A) = P(A \cap B) + P(A \cap \bar{B}) \tag{8a}$$

or

$$P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B}) \tag{8b}$$

which can also be written as

$$P(A) = P(A|B)P(B) + P(A|\bar{B})(1 - P(B)). \tag{8c}$$

Let us now look at some examples.

Example #10 - Accident Prone or Not

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 0.4, whereas this probability decreases to 0.2 for a person who is not accident prone. If we assume that 30% 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?

To answer this, let A_1 denote the event that a policyholder will have an accident within a year of purchasing a policy and let A denote the event that the policyholder is accident prone. We are given that $P(A) = 0.3$ and hence $P(\overline{A}) = 0.7$. We are also given that $P(A_1|A) = 0.4$ and $P(A_1|\overline{A}) = 0.2$. Hence, using Equation (8b), we have

$$P(A_1) = P(A_1|A)P(A) + P(A_1|\overline{A})P(\overline{A}) = (0.4)(0.3) + (0.2)(0.7) = 0.26$$

or 26%.

Now 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?

Here we want to compute $P(A|A_1)$, and using Equation (4), we have

$$P(A|A_1) = \frac{P(A_1|A)P(A)}{P(A_1)} = \frac{(0.4)(0.3)}{0.26} = \frac{6}{13}$$

which is about 46.2%. ■

Example #11: An Interesting Card Game and a Surprise Ending

Consider the following game played with an ordinary deck of 52 playing cards. The cards are shuffled and then turned over one at a time. At any time, the player can guess that the next card to be turned over will be the ace of spades and if it is, then the player wins. Of course, the player wins if the ace of spades has not yet appeared when only one card remains and no guess has yet been made since a correct guess must be made for the last card. A strategy (of which they are 52) is defined as when the person chooses to make their guess. What is the strategy that leads to the highest probability of winning this game? Explain.

The interesting point of this problem is the surprising result that all 52 strategies have the same probability of winning, namely $1/52$. To show this, we will use mathematical induction to prove the stronger result that for an n-card deck of n different cards, one of whose cards is the ace of spades, the probability of winning is $1/n$ no matter what strategy is employed. Since this is clearly true for $n = 1$ card and for $n = 2$ cards, let us assume that it is true for $n - 1$ cards, and now consider a deck of n cards. Let S_1 be the strategy that the person guesses the first card to be the ace of spades and hence $\bar{S}_1$ is the strategy that player does not guess the first card to be the ace of spades, and suppose that $P(S_1) = p$ denotes the probability that the player uses strategy S_1, and hence $P(\bar{S}_1) = 1-p$ is the probability that the player does not use strategy S_1. Given that the player employs strategy S_1, then the player's probability of winning with n cards (W_n) is clearly

$$P(W_n|S_1) = \frac{1}{n}.$$

If, however, the player employs strategy $\bar{S}_1$, then the probability that the player wins is the probability that the first card is not the ace of spades, which is $(n-1)/n$, *times* the conditional probability of winning given that the first card is not the ace of spades, which according to the induction hypothesis is $P(W_{n-1}|\bar{S}_1) = 1/(n-1)$, since there are $n-1$ cards and all strategies are assumed to be the same for $n-1$ cards. Hence,

$$P(W_n|\bar{S}_1) = P\left(\begin{array}{c}\text{first card is}\\ \text{not ace of spades}\end{array}\right) \times P(W_{n-1}|\bar{S}_1)$$

or

$$P(W_n|\bar{S}_1) = \left(\frac{n-1}{n}\right) \times \left(\frac{1}{n-1}\right) = \frac{1}{n}.$$

Then, using Equation (8b), we have

$$P(W_n) = P(W_n|S_1)P(S_1) + P(W_n|\bar{S}_1)P(\bar{S}_1) = \frac{1}{n} \times p + \frac{1}{n} \times (1-p)$$

which reduces to $P(W_n) = 1/n$, independent of the choice of strategy. A bit of a surprise, isn't it? ■

Example #12: Guessing on a Multiple-Choice Question

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 hence $1 - p$ is the probability that the student guesses the answer. Assume that a student who guesses at the answer will get it right with probability $1/m$, where m is the number of multiple-choice alternatives. Let us determine the conditional probability that a student knew the answer to a question that he or she answered correctly.

To answer this, let C be the event that the student answers a question correctly and let K be the event that the student knew the answer to the question. We know that $P(C|K) = 1$, $P(C|\bar{K}) = 1/m$, $P(K) = p$ and $P(\bar{K}) = 1 - p$. We want to compute

$$P(K|C) = \frac{P(C|K)P(K)}{P(C)} = \frac{P(C|K)P(K)}{P(C|K)P(K) + P(C|\bar{K})P(\bar{K})}$$

which leads

$$P(K|C) = \frac{(1)(p)}{(1)(p) + (1/m)(1-p)}$$

which reduces to

$$P(K|C) = \frac{mp}{mp + 1 - p}.$$

For example, if $m = 5$ and $p = 1/2$, then the probability that the student knew the answer to a question he or she answered correctly is

$$P(K|C) = \frac{5(1/2)}{5(1/2) + 1 - 1/2},$$

which reduces to $5/6$. ■

Example #13: Blood Tests and Another Surprise

A laboratory blood test is 95% effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for 1% of the healthy persons tested. That is, if a healthy person is tested, then with probability 0.01, the test result will imply that he or she has the disease. If 0.5%

of the population actually has the disease, what is the probability that a person has the diseased given that the test result is positive?

Let D be the event that the person tested has the disease and let P be the event that the test result is positive. We are given that $P(D) = 0.005$, $P(P|D) = 0.95$, and $P(P|\bar{D}) = 0.01$, and we want to compute $P(D|P)$. Toward this end, we have

$$P(D|P) = \frac{P(P|D)P(D)}{P(P)} = \frac{P(P|D)P(D)}{P(P|D)P(D) + P(P|\bar{D})P(\bar{D})}$$

which says that

$$P(D|P) = \frac{(0.95)(0.005)}{(0.95)(0.005) + (0.01)(1 - 0.005)} = \frac{95}{294} \simeq 0.323$$

or 32.3%, which says that only 32.3% of those persons with test results that are positive, actually have the disease. Many students are often surprised at this result because they expect the percentage to be much higher since the blood test seems to be a good test. We see then that positive test results does not always mean "doom and gloom" and it's always good to get a second opinion or have the test done twice. ■

Conditional probabilities are often useful when one has to reassess one's personal probabilities in the light of additional information. Let us look at the following examples.

Example #14: Detective Work

At a certain stage of a criminal investigation, the inspector in charge is 60% 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% 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?

Letting G denote the event that the suspect is guilty and C be the event that the suspect possesses the new characteristic of the criminal, we have

$$P(G|C) = \frac{P(C|G)P(G)}{P(C)} = \frac{P(C|G)P(G)}{P(C|G)P(G) + P(C|\bar{G})P(\bar{G})}$$

with $P(G) = 0.6$, $P(C|G) = 1$, $P(\bar{G}) = 0.4$ and $P(C|\bar{G}) = 0.2$, and so

$$P(G|C) = \frac{(1)(0.6)}{(1)(0.6) + (0.2)(0.4)} = \frac{15}{17} \simeq 0.882$$

or 88.2%, where we have supposed that the probability that the suspect having the characteristic if he is, in fact innocent, is equal to 0.2, the proportion of the suspects possessing the characteristic. Note that

$$P(C) = P(C|G)P(G) + P(C|\bar{G})P(\bar{G}) = (1)(0.6) + (0.2)(0.4) = 0.68$$

which seems to say that 68% of the suspects possesses the characteristic C. Of course,

$$P(G|\bar{C}) = \frac{P(\bar{C}|G)P(G)}{P(\bar{C})} = \frac{(1 - P(C|G))P(G)}{1 - P(C)} = \frac{(1-1)(0.6)}{1 - 0.68} = 0$$

which would say that the suspect should go free if they do not have the characteristic C. ■

Example #15: Identical Versus Fraternal Twins

Twins can be either identical or fraternal. Identical (also called monozygotic) twins form when a single fertilized egg splits into two genetically identical parts. Consequently, identical twins always have the same set of genes. Fraternal (also called dizygotic) twins develop when two eggs are fertilized and implanted in the uterus. The genetic connection of fraternal twins is no more or less the same as siblings born at separate times. A Los Angeles County scientist wishing to know the current fraction of twin pairs born in the county that are identical twins has assigned a county statistician to record all twin births, indicating whether or not the resulting twins were identical. The hospitals, however, told her (the statistician) that to determine whether newborn twins were identical was not a simple task, as it involved the permission of the twins' parents to perform complicated and expensive DNA studies that the hospitals could not afford. After some deliberation, the statistician just asked the hospitals for data listing twin births along with an indication as to whether the twins were of the same sex. When such data indicated that approximately 64% of the twin births were same-sexed, the statistician declared that approximately 28% of all twins were identical. Explain how the statistician came to this conclusion.

The statistician reasoned that identical twins are always of the same sex, whereas fraternal twins, having the same relationship to each other as any pair of siblings, will have probability 1/2 of being the same sex. Letting I be the event that a pair of twins is identical and SS be the event that a pair of twins have the same sex, we may say that

$$P(SS) = P(SS|I)P(I) + P(SS|\bar{I})P(\bar{I}) = P(SS|I)P(I) + P(SS|\bar{I})(1 - P(I))$$

and since $P(SS|I) = 1$ and $P(SS|\bar{I}) = 1/2$, we have

$$P(SS) = (1)P(I) + (1/2)(1 - P(I)) = \frac{1 + P(I)}{2},$$

or

$$P(I) = 2P(SS) - 1.$$

Since the hospital says that $P(SS) = 0.64$, we therefore have

$$P(I) = 2(0.64) - 1 = 0.28$$

or $P(I) = 28\%$. ■

The Odds of an Event A

The odds of an event A is defined by the probability of the event occurring compared to the probability of the event not occurring, *i.e.*,

$$O(A) = \frac{P(A)}{P(\bar{A})} = \frac{P(A)}{1 - P(A)} \tag{9a}$$

which we may also interpret as how much more likely it is that the event A occurs than it is for the event A not to occur. It is common to say that the odds of A are "$O(A)$ to one" in favor (or not in favor) of A occurring when $O(A) \geq 1$ (when $O(A) < 1$) so that if $P(A) = 0.8$ and $P(\bar{A}) = 0.2$, then the odds are $0.8/0.2 = 4$ to one of A occurring to A not occurring. Note also that we may solve Equation (9a) for $P(A)$ in terms of $O(A)$ and get

$$P(A) = \frac{O(A)}{1 + O(A)}, \tag{9b}$$

so that if the occurrence of an event is "3 to 1" odds to the event occurring, then $P(A) = 3/(1+3) = 3/4$.

Introduction of New Evidence

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

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \qquad \text{and} \qquad P(\bar{H}|E) = \frac{P(E|\bar{H})P(\bar{H})}{P(E)}.$$

Therefore the new odds after the evidence E has been introduced is

$$O(H|E) = \frac{P(H|E)}{P(\bar{H}|E)} = \frac{P(E|H)P(H)}{P(E)} \times \frac{P(E)}{P(E|\bar{H})P(\bar{H})} = \frac{P(E|H)P(H)}{P(E|\bar{H})P(\bar{H})}$$

or

$$O(H|E) = \frac{P(E|H)}{P(E|\bar{H})} \times O(H). \tag{10}$$

This says that 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 of the new evidence given that H is not true. Therefore, the odds of H increases whenever the new evidence is more likely when H is true that when H is false. Similarly, the odds of H decrease whenever the new evidence is more likely when H is false than when H is true.

Example #16: Which Coin Is It?

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 lands on heads, what is the probability that it was a type-A coin?

Let A (B) be the event that a type-A (type-B) coin was flipped so that $B = \bar{A}$, and let H be the event that the flipped coin shows heads. We want to compute $P(A|H)$ and using Equation (10), we have

$$O(A|H) = \frac{P(H|A)}{P(H|\bar{A})} \times O(A) = \frac{1/4}{3/4} \times \frac{2/3}{1-2/3} = \frac{2}{3}$$

which then says via Equation (9b), that

$$P(A|H) = \frac{O(A|H)}{1 + O(A|H)} = \frac{2/3}{1 + 2/3} = \frac{2}{5}$$

or $P(A|H) = 40\%$. Of course we may also do this using

$$P(A|H) = \frac{P(H|A)P(A)}{P(H)} = \frac{P(H|A)P(A)}{P(H|A)P(A) + P(H|\bar{A})P(\bar{A})},$$

which leads to

$$P(A|H) = \frac{(1/4)(2/3)}{(1/4)(2/3) + (3/4)(1/3)} = \frac{2}{5}$$

or $P(A|H) = 40\%$.■

4. Bayes' Theorem

Let B_1, B_2, B_3, ..., B_n be a partition of a sample space S. If we combine Equations (7b)

$$P(A) = \sum_{j=1}^{n} P(A|B_j)P(B_j),$$

and Equation (4),

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)},$$

we have

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum\limits_{j=1}^{n} P(A|B_j)P(B_j)} \tag{11}$$

for $i = 1, 2, 3, ..., n$, which is known as Bayes's Theorem. This is also called the probability of "causes". Since the B_i's are a partition of the sample space S, one and only one of the events B_i occurs. Hence Equation (11) gives the probability of a particular event B_i (that is, a cause), given that the event A has occurred.

Example #17: Finding a Missing Plane

A plane is missing, and it is presumed that it was equally likely to have gone down in any of 3 possible regions. Let $1 - \beta_i$ (for $i = 1, 2, 3$) denote the probability that the plane will be found upon a search of the ith 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 and they are generally attributable to the geographical and environmental conditions of the regions. What is the conditional probability that the plane is in the ith region (for $i = 2, 3$) given that a search of region 1 is unsuccessful?

Let R_i (for $i = 1, 2, 3$) be the event that the plane is in region i, and let E be the event that a search of region 1 is unsuccessful. We are given that

$$P(E|R_i) = 1 - P(\bar{E}|R_i) = 1 - (1 - \beta_i) = \beta_i$$

for $i = 1, 2, 3$, and $P(R_i) = 1/3$ for $i = 1, 2, 3$, and want to compute $P(R_i|E)$ for $i = 1, 2, 3$. Using Bayes's theorem, we have

$$P(R_1|E) = \frac{P(E|R_1)P(R_1)}{P(E|R_1)P(R_1) + P(E|R_2)P(R_2) + P(E|R_3)P(R_3)}$$

or

$$P(R_1|E) = \frac{\beta_1(1/3)}{\beta_1(1/3) + (1)(1/3) + (1)(1/3)} = \frac{\beta_1}{\beta_1 + 2}.$$

We also have

$$P(R_2|E) = \frac{P(E|R_2)P(R_2)}{P(E|R_1)P(R_1) + P(E|R_2)P(R_2) + P(E|R_3)P(R_3)}$$

or

$$P(R_2|E) = \frac{(1)(1/3)}{\beta_1(1/3) + (1)(1/3) + (1)(1/3)} = \frac{1}{\beta_1 + 2}$$

and

$$P(R_3|E) = \frac{P(E|R_3)P(R_3)}{P(E|R_1)P(R_1) + P(E|R_2)P(R_2) + P(E|R_3)P(R_3)}$$

or

$$P(R_3|E) = \frac{(1)(1/3)}{\beta_1(1/3) + (1)(1/3) + (1)(1/3)} = \frac{1}{\beta_1 + 2}.$$

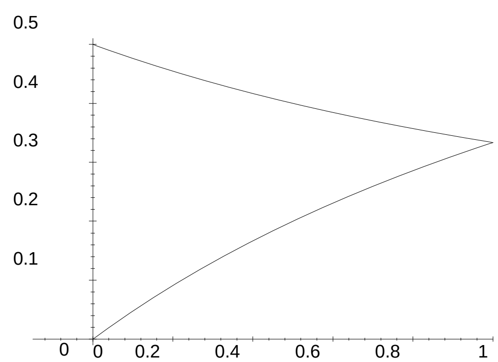
We note that $P(R_1|E) + P(R_2|E) + P(R_3|E) = 1$, as it must. We also note that the updated (*i.e.*, the conditional) probability that the plane is in region i, given the information that a search of region 1 did not find it, is greater than the initial probability when $i \neq 1$, since

$$\frac{1}{\beta_1 + 2} > \frac{1}{3}$$

and the updated (*i.e.*, the conditional) probability that the plane is in region i, given the information that a search of region 1 did not find it, is less than the initial probability when $i = 1$, since

$$\frac{\beta_1}{\beta_1 + 2} < \frac{1}{3}.$$

These statements are certainly intuitive, since not finding the plane in region 1 must decrease its chances of being in region 1 and it must increase its chances of being in regions 2 and 3. Note also that plots of $P(R_i|E)$ versus β_1 are shown in the figure below.



Plots of $P(R_1|E)$ (solid) and
$P(R_i|E)$ for $i = 2, 3$ (dashed)

It says that $P(R_1|E)$ is an increasing function of β_1 while $P(R_2|E)$ and $P(R_3|E)$ are decreasing functions of β_1. These are 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. ■

Example #19: Tricking Your Friends

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

This is like that Monte Hall problem and most people would say that the answer is 1/2 since it can't be the card that has two black sides (BB) and hence must be either the two remaining cards, namely the card that has a red and black side (RB) or the card that has two red sides (RR). To solve this correctly, let RR, BB and RB denote, respectively, the events that the chosen card is all red, all black, or half red and half black. Also let R be the event that the upturned side of the chosen card is red. Then, the desired probability is

$$P(RB|R) = \frac{P(R|RB)P(RB)}{P(R|RR)P(RR) + P(R|RB)P(RB) + P(R|BB)P(BB)}$$

which gives

$$P(RB|R) = \frac{(1/2)(1/3)}{(1)(1/3) + (1/2)(1/3) + (0)(1/3)} = \frac{1}{3}$$

and so $P(RR|R) = 2/3$. Hence the answer for $P(RB|R)$ is 1/3 which is smaller than 1/2, resulting in the possibility of winning money from less enlightened people.

Once again, many people guess 1/2 as the answer by incorrectly reasoning that given that a red side appears, there are two equally likely possibilities in that the card is either the all red card (RR) or the red/black card (RB). 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 just performed - namely R_1, R_2, B_1, B_2, R_3 and 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 for B_1, B_2 and B_3. 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 probability of R_3 given the "reduced" sample space

$$S' = \{R_1, R_2, R_3\},$$

which obviously equals 1/3. Very sneaky! ■

Example #20: Another Surprise

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?

Once again, most people would say that the answer is 1/2 since the sex of the child seen should not have any influence on the sex of the child unseen but let's see what we get. Let us start by defining the following events:

$$\begin{aligned} G_1 &= \text{the first (that is the oldest) child is a girl} \\ G_2 &= \text{the second (that is the youngest) child is a girl} \\ G_s &= \text{the child seen with the mother is a girl.} \end{aligned}$$

Also let

$$\begin{aligned} B_1 &= \text{the first (that is the older) child is a boy} \\ B_2 &= \text{the second (that is the younger) child is a boy} \\ B_s &= \text{the child seen with the mother is a boy.} \end{aligned}$$

Now the desired probability is $P(G_1 \cap G_2 | G_s)$, which can be expressed as

$$P(G_1 \cap G_2 | G_s) = \frac{P(G_1 \cap G_2 \cap G_s)}{P(G_s)} = \frac{P(G_1 \cap G_2)}{P(G_s)}.$$

Also,

$$\begin{aligned} P(G_s) &= P(G_s | G_1 \cap G_2) P(G_1 \cap G_2) + P(G_s | G_1 \cap B_2) P(G_1 \cap B_2) \\ &\quad + P(G_s | B_1 \cap G_2) P(B_1 \cap G_2) + P(G_s | B_1 \cap B_2) P(B_1 \cap B_2) \\ &= (1) P(G_1 \cap G_2) + P(G_s | G_1 \cap B_2) P(G_1 \cap B_2) \\ &\quad + P(G_s | B_1 \cap G_2) P(B_1 \cap G_2) + (0) P(B_1 \cap B_2) \end{aligned}$$

or

$$\begin{aligned} P(G_s) &= P(G_1 \cap G_2) + P(G_s | G_1 \cap B_2) P(G_1 \cap B_2) \\ &\quad + P(G_s | B_1 \cap G_2) P(B_1 \cap G_2). \end{aligned}$$

If we now make the usual assumption that all 4 gender possibilities are equally likely, then we find that

$$P(G_1 \cap G_2) = P(G_1 | G_2) P(G_2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

and

$$P(G_1 \cap B_2) = P(G_1|B_2)P(B_2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

and

$$P(B_1 \cap G_2) = P(B_1|G_2)P(G_2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

and so

$$P(G_s) = \frac{1}{4} + \frac{1}{4}P(G_s|G_1 \cap B_2) + \frac{1}{4}P(G_s|B_1 \cap G_2),$$

and then putting these into

$$P(G_1 \cap G_2|G_s) = \frac{P(G_1 \cap G_2)}{P(G_s)},$$

we have

$$P(G_1 \cap G_2|G_s) = \frac{1/4}{\frac{1}{4} + \frac{1}{4}P(G_s|G_1 \cap B_2) + \frac{1}{4}P(G_s|B_1 \cap G_2)}$$

which reduces to

$$P(G_1 \cap G_2|G_s) = \frac{1}{1 + P(G_s|G_1 \cap B_2) + P(G_s|B_1 \cap G_2)}$$

showing that 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 that the older child is a girl and the younger is a boy or that the child seen with the mother is a girl given that the older child is a boy and the younger is a girl. For example, suppose we assume with probability p that the probability that the child walking with the mother is the elder child, then

$$P(G_s|G_1 \cap B_2) = p \qquad \text{and} \qquad P(G_s|B_1 \cap G_2) = 1 - p$$

and then

$$P(G_1 \cap G_2|G_s) = \frac{1}{1 + p + 1 - p} = \frac{1}{2}.$$

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

$$P(G_s|G_1 \cap B_2) = q \qquad \text{and} \qquad P(G_s|B_1 \cap G_2) = q$$

and then

$$P(G_1 \cap G_2 | G_s) = \frac{1}{1+q+q} = \frac{1}{1+2q}.$$

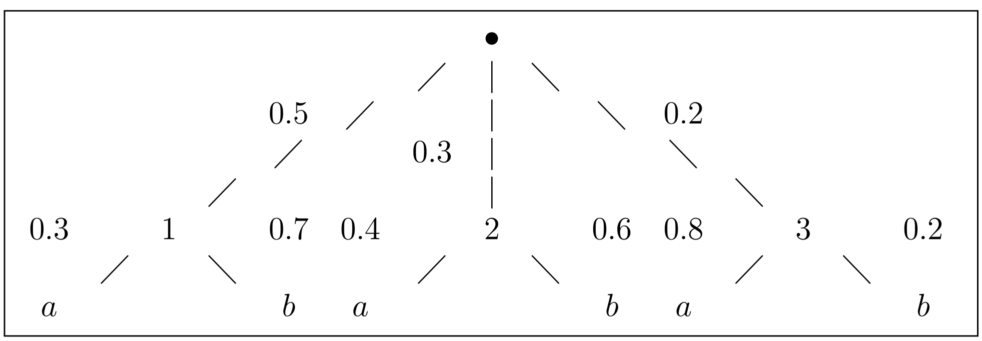
For instance, if we took $q = 1$ so that the mother always chooses to walk with her daughter, then the conditional probability that she has two daughters would be $1/3$, which is in accord with the fact that seeing the mother walk with a daughter says that she has at least one daughter. So why isn't $P(G_s) = 1$ since the problem says that the child seen with the mother is a girl? ■

Example #21: Inverting a Probability Tree

Suppose that two experiments E_1 and E_2 are to be performed and suppose that the possible outcomes of these are given by the two sample spaces

$$S_1 = \{1, 2, 3\} \quad \text{and} \quad S_2 = \{a, b\}.$$

Suppose we are interested in performing E_1 followed by E_2 and we are given the following probability tree.



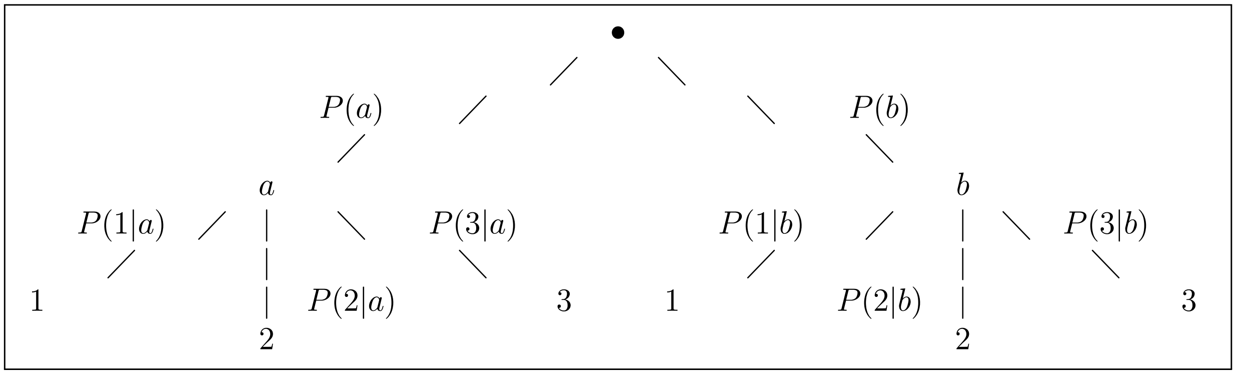
so that $P(1) = 0.5$, $P(2) = 0.3$ and $P(3) = 0.2$, and

$$P(a|1) = 0.3 \quad , \quad P(b|1) = 0.7 \quad , \quad P(a|2) = 0.4 \quad , \quad P(b|2) = 0.6$$

and

$$P(a|3) = 0.8 \quad \text{and} \quad P(b|3) = 0.2$$

We would like to invert this probability tree to form the probability tree



so that we are interested in performing E_2 followed by E_1, which means we need to compute $P(a)$, $P(b)$ and the conditional probabilities

$$P(1|a) \quad , \quad P(2|a) \quad , \quad P(3|a) \quad , \quad P(1|b) \quad , \quad P(2|b) \quad , \quad P(3|b).$$

Toward this end, we first compute

$$\begin{aligned} P(a) &= P(a|1)P(1) + P(a|2)P(2) + P(a|3)P(3) \\ &= (0.3)(0.5) + (0.4)(0.3) + (0.8)(0.2) = 0.43 \end{aligned}$$

and

$$\begin{aligned} P(b) &= P(b|1)P(1) + P(b|2)P(2) + P(b|3)P(3) \\ &= (0.7)(0.5) + (0.6)(0.3) + (0.2)(0.2) = 0.57. \end{aligned}$$

Then

$$P(1|a) = \frac{P(a|1)P(1)}{P(a)} = \frac{(0.3)(0.5)}{0.43} = \frac{15}{43}$$

and

$$P(1|b) = \frac{P(b|1)P(1)}{P(b)} = \frac{(0.7)(0.5)}{0.57} = \frac{35}{57}$$

and

$$P(2|a) = \frac{P(a|2)P(2)}{P(a)} = \frac{(0.4)(0.3)}{0.43} = \frac{12}{43}$$

and

$$P(2|b) = \frac{P(b|2)P(2)}{P(b)} = \frac{(0.6)(0.3)}{0.57} = \frac{18}{57}$$

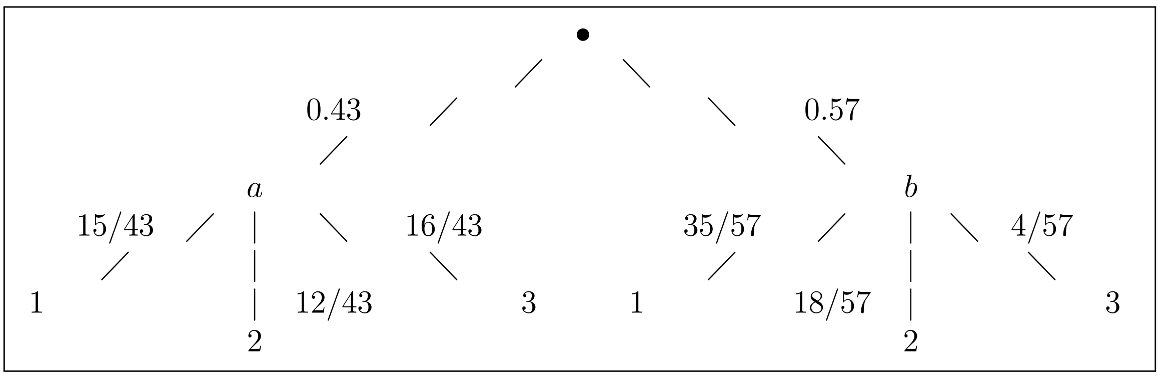
and

$$P(3|a) = \frac{P(a|3)P(3)}{P(a)} = \frac{(0.8)(0.2)}{0.43} = \frac{16}{43}$$

and

$$P(3|b) = \frac{P(b|3)P(3)}{P(b)} = \frac{(0.2)(0.2)}{0.57} = \frac{4}{57}.$$

This leads to the following inverted probability tree.



Notice the use of Bayes's theorem here, and as a check we note that

$$P(1) = P(1|a)P(a) + P(1|b)P(b) = \frac{15}{43} \times 0.43 + \frac{35}{57} \times 0.57 = 0.5$$

and

$$P(2) = P(2|a)P(a) + P(2|b)P(b) = \frac{12}{43} \times 0.43 + \frac{18}{57} \times 0.57 = 0.3$$

and

$$P(3) = P(3|a)P(a) + P(3|b)P(b) = \frac{16}{43} \times 0.43 + \frac{4}{57} \times 0.57 = 0.2$$

which are the results given in the initial tree. ■

Example #22

A bin contains 3 types of flashlights. The probability that a type 1 flashlight will give more than 100 hours of light is 0.7, with the corresponding probabilities for type 2 and type 3 flashlights being 0.4 and 0.3, respectively. Suppose that 20% of the flashlights in the bin are type 1, 30% are type 2 and 50% are type 3: (a) what is the probability that a randomly chosen flashlight will give more than

100 hours of use, and (b) given that a flashlight lasted more than 100 hours, what is the probability that it was a type j flashlight for $j = 1, 2, 3$?

To solve part (a), let A denote the event that the flashlight chosen will give more than 100 hours of use and let F_j be the event that a type j flashlight is chosen for $j = 1, 2, 3$. To compute $P(A)$, we have

$$\begin{aligned} P(A) &= P(A|F_1)P(F_1) + P(A|F_2)P(F_2) + P(A|F_3)P(F_3) \\ &= (0.7)(0.2) + (0.4)(0.3) + (0.3)(0.5) = 0.41 \end{aligned}$$

so that there is a 41% chance that the flashlight will last for more than 100 hours. To solve part (b), we use Bayes's formula and write

$$P(F_1|A) = \frac{P(A|F_1)P(F_1)}{P(A)} = \frac{(0.7)(0.2)}{0.41} = \frac{14}{41}$$

and

$$P(F_2|A) = \frac{P(A|F_2)P(F_2)}{P(A)} = \frac{(0.4)(0.3)}{0.41} = \frac{12}{41}$$

and

$$P(F_3|A) = \frac{P(A|F_3)P(F_3)}{P(A)} = \frac{(0.3)(0.5)}{0.41} = \frac{15}{41}$$

and note that $P(F_1|A) + P(F_2|A) + P(F_3|A) = 1$, as it must. ■

5. Independent Events

We note that in general $P(A|B) \neq P(A)$ since information about B will usually change the chances of A occurring, making these chances going up or down. In the special case where $P(A|B) = P(A)$, we say that A is *independent* of B. In fact we note that if $P(A|B) = P(A)$, then

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

showing that B must also be independent of A. Thus we say that two events A and B are *independent* of each other if the outcome of one has no effect on the other. That is, when

$$P(A|B) = P(A) \qquad \text{and} \qquad P(B|A) = P(B). \tag{12}$$

Since

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

we see that A and B are two independent events *if and only if*

$$P(A \cap B) = P(A)P(B). \tag{13}$$

We say that three events A, B and C are *mutually independent* if and only if

$$P(A \cap B) = P(A)P(B) \quad , \quad P(A \cap C) = P(A)P(C) \quad , \quad P(B \cap C) = P(B)P(C)$$

and

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

In this case, we note that

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

and

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

and

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

and

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

and

$$\begin{aligned}
P(A|B \cup C) &= \frac{P(A \cap (B \cup C))}{P(B \cup C)} = \frac{P((A \cap B) \cup (A \cap C))}{P(B \cup C)} \\
&= \frac{P(A \cap B) + P(A \cap C) - P((A \cap B) \cap (A \cap C))}{P(B) + P(C) - P(B \cap C)} \\
&= \frac{P(A)P(B) + P(A)P(C) - P((A \cap B \cap C))}{P(B) + P(C) - P(B)P(C)} \\
&= \frac{P(A)P(B) + P(A)P(C) - P(A)P(B)P(C)}{P(B) + P(C) - P(B)P(C)} \\
&= P(A)\left(\frac{P(B) + P(C) - P(B)P(C)}{P(B) + P(C) - P(B)P(C)}\right) = P(A)
\end{aligned}$$

and so on for all the other combinations. Note also that in order for A, B and C to be mutually independent, it is NOT enough to simply require that

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

Example #23a: A Parallel Combination

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 (for $i = 1, 2, 3, ..., n$), determine the probability that the system functions. To do this, it is easier to determine the probability that the system does not function which is

$$\bar{P} = \bar{P}_1 \bar{P}_2 \cdots \bar{P}_n = (1 - p_1)(1 - p_2) \cdots (1 - p_n) = \prod_{k=1}^{n} (1 - p_k)$$

which then says that

$$P = 1 - \bar{P} = 1 - \prod_{k=1}^{n} (1 - p_k).$$

Example #23b: A Series Combination

A system composed of n separate components is said to be a series system if it functions only when all of the components functions. For such a system, if component i, which is independent of the other components functions with probability p_i (for $i = 1, 2, 3, ..., n$), determine the probability that the system functions. To do this, we simply write

$$P = P_1 P_2 \cdots P_n = (p_1)(p_2) \cdots (p_n) = \prod_{k=1}^{n} p_k.$$

Example #24

Independent trials consisting of rolling a pair of fair dice are performed. What is the probability that an outcome of 5 appears before an outcome of 7 when the outcome of a roll is the sum of the dice? To answer this we let E_n denote the event that no 5 or 7 appears on the first $n-1$ trials and then a 5 appears on the nth trial, and then the desired probability is

$$P = P(E_1 \cup E_2 \cup E_3 \cup \cdots \cup E_n \cup \cdots) = \sum_{k=1}^{\infty} P(E_k)$$

and since $P(\text{roll of } 5) = 4/36$ on any trial and $P(\text{roll of } 7) = 6/36$ on any trial, we have

$$P(E_n) = \left(1 - \frac{4}{36} - \frac{6}{36}\right)^{n-1} \frac{4}{36} = \left(\frac{13}{18}\right)^{n-1} \frac{1}{9}.$$

Then

$$P = \sum_{k=1}^{\infty} \left(\frac{13}{18}\right)^{k-1} \frac{1}{9} = \frac{1}{9}\left(\frac{1}{1 - 13/18}\right) = \frac{2}{5}.$$

Example #25: A General Result

Independent trials consisting of disjoint events E and F with $P(E)$ being the probability that E occurs and $P(F)$ being the probability that F occurs and

$$P(N) = P(\bar{E} \cap \bar{F}) = 1 - P(\overline{\bar{E} \cap \bar{F}}) = 1 - P(E \cup F) = 1 - P(E) - P(F)$$

being the probability that neither E nor F occurs. What is the probability that E occurs before F? To answer this we let E_n denote the event that neither E nor F occurs on the first $n-1$ trials and then E occurs on the nth trial. Then

$$P_n = P(E_n) = P(N \cap N \cap N \cap \cdots \cap N \cap E) = (P(N))^{n-1} P(E)$$

is the probability that E occurs before F on the nth trial the desired probability that E occurs before F on some trial is

$$P = \sum_{k=1}^{\infty} P_k = \sum_{k=1}^{\infty} (P(N))^{k-1} P(E) = P(E)\left(\frac{1}{1 - P(N)}\right),$$

or

$$P = \frac{P(E)}{P(E) + P(F)}.$$

If should be noted that the solution to Example #25 follows from this by writing

$$P = \frac{P(\text{roll of } 5)}{P(\text{roll of } 5) + P(\text{roll of } 7)} = \frac{4/36}{4/36 + 6/36} = \frac{2}{5}.$$

Example #26: The Problem of the Points

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.

We consider two solutions. The first is due to Pascal who assumed that $P_{n,m}$ is the probability that n successes occur before m failures. By conditioning on the outcome of the first trial, we obtain

$$\begin{aligned} & P(n \text{ successes before } m \text{ failures}) \\ = \; & P(n-1 \text{ successes before } m \text{ failures}) P(\text{success}) \\ & + P(n \text{ successes before } m-1 \text{ failures}) P(\text{failure}) \end{aligned}$$

or

$$P_{n,m} = pP_{n-1,m} + (1-p)P_{n.m-1}$$

for $n \geq 1$ and $m \geq 1$. Using the obvious boundary conditions $P_{n,0} = 0$ and $P_{0,m} = 1$, we can solve for $P_{n,m}$.

The second argument is due to Fermat who 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 $n+m-1$ 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 and so n successes would not occur before m failures. Now the probability that there are exactly k successes in $m+n-1$ trials is

$$p^k(1-p)^{m+n-1-k} \times \left(\begin{array}{c} \text{the number of ways in choosing} \\ k \text{ successes in } m+n-1 \text{ trials} \\ \text{where order does not matter} \end{array} \right)$$

which gives

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

If follows that the desired probability of n successes to occur before m failures is

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

The student should check that this satisfies Pascal's equation and boundary conditions above. ■

Example #27: The Gambler's Ruin Problem

Two gambler's, A and B, bet on the outcomes of successive flips of a coin. On each flip, if the coin comes up heads, A collects 1 unit from B, whereas if it comes up tails A pays 1 unit to B. They continue to do this until one of them runs out of money. If it is assumed that the successive flips of the coin are independent and each flip results in a heads with probability p, what is the probability that A ends up with all the money if he starts with i units and B starts with $N-i$ units? To answer this, let A_i denote the event that A ends up with all the money given that he starts with i and B starts with $N-i$, and to make clear the dependence of the initial fortune of A, let $P_i = P(A_i)$. We shall obtain an expression for $P(A_i)$ by conditioning on the outcome of the first flip as follows. Let H denote the event that the first flip lands on heads, then

$$P_i = P(A_i) = P(A_i|H)P(H) + P(A_i|\bar{H})P(\bar{H}) = pP(A_i|H) + (1-p)P(A_i|\bar{H}).$$

Now, given that the first flip lands on heads, the situation after the first bet is that A has $i+1$ units and B has $N-(i+1)$ units. Since the successive flips are assumed to be independent with a common probability p of heads, it follows that from that point on, A's probability of winning all the money is exactly the same as if the game were just starting with A having an initial fortune of $i+1$ and B having an initial fortune of $N-(i+1)$. Therefore

$$P(A_i|H) = P(A_{i+1}) = P_{i+1} \quad \text{and similarly} \quad P(A_i|\bar{H}) = P(A_{i-1}) = P_{i-1}.$$

Hence, letting $q = 1-p$, we obtain

$$P_i = pP_{i+1} + qP_{i-1}$$

for $i = 1, 2, 3, ..., N-1$. By making use of the obvious boundary conditions $P_0 = 0$ and $P_N = 1$, we may now solve for P_i.

Using the methods of Chapter #1, we set $P_i = r^i$ and get

$$r^i = pr^{i+1} + qr^{i-1} \qquad \text{or} \qquad pr^2 - r + q = 0$$

which says that

$$r_\pm = \frac{1 \pm \sqrt{1-4pq}}{2p} = \frac{1 \pm \sqrt{1-4p(1-p)}}{2p} = \frac{1 \pm \sqrt{(2p-1)^2}}{2p}$$

or

$$r_\pm = \frac{1 \pm (2p-1)}{2p}$$

so that $r_+ = 1$ and $r_- = (1-p)/p = q/p$. When $p \neq 1/2$ (so that $q/p \neq 1$), we have

$$P_i = A_+ r_+^i + A_- r_-^i = A_+ + A_- \left(\frac{q}{p}\right)^i.$$

Using the boundary conditions $P_0 = 0$, we find that $A_+ + A_- = 0$ so that $A_- = -A_+$. Then

$$P_i = A_+ - A_+\left(\frac{q}{p}\right)^i = A_+\left(1 - \left(\frac{q}{p}\right)^i\right).$$

Then using $P_N = 1$, we have

$$A_+\left(1 - \left(\frac{q}{p}\right)^N\right) = 1 \qquad \text{which says that} \qquad A_+ = \frac{1}{1-(q/p)^N}.$$

Thus we find that

$$P_i = \frac{1-(q/p)^i}{1-(q/p)^N}$$

for $i = 0, 1, 2, ..., N$, provided that $p \neq 1/2$. If $p = 1/2$, then $q = p$ and $r_+ = r_- = 1$, and using the results from Chapter #1, we have

$$P_i = A + Bi$$

and setting $P_0 = A = 0$, we have $P_i = Bi$. Then $P_N = BN = 1$ so that $B = 1/N$ and then $P_i = i/N$. Thus we find that

$$P_i = \frac{1-(q/p)^i}{1-(q/p)^N} \qquad \text{for} \qquad p \neq \frac{1}{2}$$

and
$$P_i = \frac{i}{N} \qquad \text{for} \qquad p = \frac{1}{2}$$
and now the problem is complete. Of course, the probability that B wins is $1 - P_i$.
■

6. $P(\bullet|F)$ Is a Valid Probability Function

Conditional probabilities satisfy all of the properties of ordinary probabilities, as we now show. First we note that
$$P(E|F) = \frac{P(E \cap F)}{P(F)} \geq 0$$
since $P(E \cap F) \geq 0$ and $P(F) > 0$. We also note that since $(E \cap F) \subseteq F$, we have
$$P(E \cap F) \leq P(F)$$
from a result in Chapter #2, and hence we find that $P(E|F) \leq 1$ and so we find that
$$0 \leq P(E|F) \leq 1$$
and so $P(E|F)$ satisfies Axiom #1. Next we note that
$$P(S|F) = \frac{P(S \cap F)}{P(F)} = \frac{P(F)}{P(F)} = 1$$
showing that Axiom #2 is now true. Finally if E_i for $i = 1, 2, 3, \ldots$, are *mutually exclusive* events, then
$$P\left(\bigcup_{k=1}^{\infty} E_k|F\right) = \frac{P\left(\left(\bigcup_{k=1}^{\infty} E_k\right) \cap F\right)}{P(F)} = \frac{P\left(\bigcup_{k=1}^{\infty} (E_k \cap F)\right)}{P(F)} = \frac{\sum_{k=1}^{\infty} P(E_k \cap F)}{P(F)}$$
But
$$(E_i \cap F) \cap (E_j \cap F) = (E_i \cap E_j) \cap (F \cap F) = \emptyset \cap F = \emptyset$$
and so
$$P\left(\bigcup_{k=1}^{\infty} (E_k \cap F)\right) = \sum_{k=1}^{\infty} P(E_k \cap F)$$

and so

$$P\left(\bigcup_{k=1}^{\infty} E_k|F\right) = \frac{\sum\limits_{k=1}^{\infty} P(E_k \cap F)}{P(F)}$$

or

$$P\left(\bigcup_{k=1}^{\infty} E_k|F\right) = \sum_{k=1}^{\infty} \frac{P(E_k \cap F)}{P(F)} = \sum_{k=1}^{\infty} P(E_k|F),$$

which shows that Axiom #3 is satisfied. Since Axioms #1, #2 and #3 are satisfied we see that $P(\bullet|F)$ is a valid probability function.

Example #28: Another Problem on Runs

In dependent trials, each resulting in a success with probability p or a failure with probability $q = 1 - p$, are performed. We are interested in computing the probability that a run of n consecutive successes occurs before a run of m consecutive failures.

To solve this let E be the event that a run of n consecutive successes occurs before a run of m consecutive failures. To obtain $P(E)$, we start by conditioning on the outcome of the first trial. That is, letting H denote the event that the first trial results in a success, we obtain

$$P(E) = P(E|H)P(H) + P(E|\bar{H})P(\bar{H}) = pP(E|H) + qP(E|\bar{H}).$$

Now, given that the first trial was successful, one way we can get run of n successes before a run of m failures would be to have the next $n-1$ trials all result in successes. So, let us condition on whether or not that occurs. That is, letting F be the event that trials 2 through n are all successes, we obtain

$$P(E|H) = P(E|H \cap F)P(F|H) + P(E|H \cap \bar{F})P(\bar{F}|H).$$

Now clearly $P(E|F \cap H) = 1$ and on the other hand, if the event $\bar{F} \cap H$ occurs, then the first trial would result in a success but there would be at least one failure some time during the next $n-1$ trials. However when this failure occurs, it would wipe out all of the previous successes, and the situation would be exactly as if we started out with a failure. Hence

$$P(E|H \cap \bar{F}) = P(E|\bar{H}).$$

Because the independence of the trials implies that F and H are independent, and because

$$P(F|H) = P(F) = p^{n-1} \qquad \text{and} \qquad P(\bar{F}|H) = P(\bar{F}) = 1 - p^{n-1}$$

it follows that

$$P(E|H) = (1)p^{n-1} + (1 - p^{n-1})P(E|\bar{H}) = p^{n-1} + (1 - p^{n-1})P(E|\bar{H}).$$

We now obtain an expression for $P(E|\bar{H})$ in a similar manner. That is, we let G denote the event that trials 2 through m are all failures. Then

$$P(E|\bar{H}) = P(E|\bar{H} \cap G)P(G|\bar{H}) + P(E|\bar{H} \cap \bar{G})P(\bar{G}|\bar{H}).$$

But $\bar{H} \cap G$ is the event that the first m trials all result in failures and so $P(E|\bar{H} \cap G) = 0$ and if $\bar{H} \cap \bar{G}$ occurs, then the first trial is a failure and there is at least one success in the next $m - 1$ trials. Hence, since this success wipes out all previous failures, we must have

$$P(E|\bar{H} \cap \bar{G}) = P(E|H).$$

Then, because $P(\bar{G}|\bar{H}) = P(\bar{G}) = 1 - q^{m-1}$, we have

$$P(E|\bar{H}) = (0)P(G|\bar{H}) + (1 - q^{m-1})P(E|H) = (1 - q^{m-1})P(E|H).$$

Putting this into the earlier equation for $P(E|H)$, we have

$$P(E|H) = p^{n-1} + (1 - p^{n-1})(1 - q^{m-1})P(E|H)$$

which says that

$$P(E|H) = \frac{p^{n-1}}{1 - (1 - p^{n-1})(1 - q^{m-1})} = \frac{p^{n-1}}{q^{m-1} + p^{n-1} - p^{n-1}q^{m-1}}$$

and then

$$P(E|\bar{H}) = \frac{p^{n-1}(1 - q^{m-1})}{q^{m-1} + p^{n-1} - p^{n-1}q^{m-1}}.$$

Thus we find that

$$P(E) = pP(E|H) + qP(E|\bar{H})$$

leads to

$$P(E) = p\frac{p^{n-1}}{q^{m-1} + p^{n-1} - p^{n-1}q^{m-1}} + q\frac{p^{n-1}(1 - q^{m-1})}{q^{m-1} + p^{n-1} - p^{n-1}q^{m-1}}$$

or

$$P(E) = \frac{p^n + qp^{n-1} - p^{n-1}q^m}{q^{m-1} + p^{n-1} - p^{n-1}q^{m-1}} = \frac{p^n + (1-p)p^{n-1} - p^{n-1}q^m}{q^{m-1} + p^{n-1} - p^{n-1}q^{m-1}}$$

or

$$P(E) = \frac{p^{n-1}(1-q^m)}{p^{n-1} + q^{m-1} - p^{n-1}q^{m-1}}$$

It is interesting to note that by the symmetry of the problem, the probability of obtaining a run of m failures before a run of n successes would be given by the same result with p and q interchanged and n and m interchanged, so that

$$P\left(\begin{array}{c}\text{run of } m \text{ failures}\\ \text{before a run of } n \text{ successes}\end{array}\right) = \frac{q^{m-1}(1-p^n)}{p^{n-1} + q^{m-1} - p^{n-1}q^{m-1}}.$$

But since

$$\begin{aligned} p^{n-1}(1-q^m) + q^{m-1}(1-p^n) &= p^{n-1} - p^{n-1}q^m + q^{m-1} - q^{m-1}p^n \\ &= p^{n-1} + q^{m-1} - (p^{n-1}q^m + q^{m-1}p^n) \\ &= p^{n-1} + q^{m-1} - p^{n-1}q^{m-1}(q+p) \\ &= p^{n-1} + q^{m-1} - p^{n-1}q^{m-1} \end{aligned}$$

we see that

$$P\left(\begin{array}{c}\text{run of } m \text{ failures}\\ \text{before a run of } n \text{ successes}\end{array}\right) + P\left(\begin{array}{c}\text{run of } n \text{ failures}\\ \text{before a run of } m \text{ successes}\end{array}\right) = 1$$

it follows that, with probability 1, either a run of n successes or a run of m failures will eventually occur. As a specific example in tossing a fair coin so that $p = q = 1/2$, we have

$$P\left(\begin{array}{c}\text{run of } n \text{ failures}\\ \text{before a run of } m \text{ successes}\end{array}\right) = \frac{(1/2)^{n-1}(1-(1/2)^m)}{(1/2)^{n-1} + (1/2)^{m-1} - (1/2)^{n-1}(1/2)^{m-1}}$$

and

$$P\left(\begin{array}{c}\text{run of } m \text{ failures}\\ \text{before a run of } n \text{ successes}\end{array}\right) = \frac{(1/2)^{m-1}(1-(1/2)^n)}{(1/2)^{n-1} + (1/2)^{m-1} - (1/2)^{n-1}(1/2)^{m-1}}$$

and so the probability that a run of 2 heads will precede a run of 3 tails is

$$P\left(\begin{array}{c}\text{run of 2 heads}\\ \text{before a run of 3 tails}\end{array}\right) = \frac{(1/2)^{2-1}(1-(1/2)^3)}{(1/2)^{2-1} + (1/2)^{3-1} - (1/2)^{2-1}(1/2)^{3-1}} = \frac{7}{10}$$

and so 2 heads before 4 tails, we have

$$P\left(\begin{array}{c}\text{run of 2 heads}\\ \text{before a run of 4 tails}\end{array}\right) = \frac{(1/2)^{2-1}(1-(1/2)^4)}{(1/2)^{2-1}+(1/2)^{4-1}-(1/2)^{2-1}(1/2)^{4-1}} = \frac{5}{6}$$

which is 19% larger. ■