

CHAPTER

4

**Probability Theory,
Part 1***Introduction**Definitions**Theoretical and Historical Methods of Estimation**Samples and Universes**The Concept of a Sample**The Concept of a Universe or Population**The Conventions of Probability**Mutually Exclusive Events—The Addition Rule**Joint Probabilities**The Monte Carlo Simulation Method (Resampling)**The Deductive Formulaic Method**Multiplication Rule**Conditional and Unconditional Probabilities**The Skins Again, Plus Leaving the Game Early*

Introduction

Let's assume we understand the nature of the system or mechanism that produces the uncertain events in which we are interested. That is, the probability of the relevant independent *simple* events is assumed to be known, the way we assume we know the probability of a single "6" with a given die. The task is to determine the probability of various sequences or combinations of the simple events—say, three "6's" in a row with the die. These are the sorts of probability problems dealt with in this chapter.

The resampling method—or just call it simulation or Monte Carlo method, if you prefer—will be illustrated with classic examples. Typically, a single trial of the system is simulated with cards, dice, random numbers, or a computer program. Then trials are repeated again and again to estimate the frequency of occurrence of the event in which we are interested; this is the probability we seek. We can obtain as accurate an estimate of the probability as we wish by increasing the number of trials. The key task in each situation is *designing an experiment that accurately simulates the system in which we are interested*.

This chapter begins the Monte Carlo simulation work that culminates in the resampling method in statistics proper. The chapter deals with problems in probability theory—that is, situations where one wants to estimate the probability of one or more particular events when the basic structure and parameters of the system are known. In later chapters we move on to inferential statistics, where similar simulation work is known as resampling.

Definitions

A few definitions first:

Simple Event: An event such as a single flip of a coin, or one draw of a single card. A simple event cannot be broken down into simpler events of a similar sort.

Simple Probability (also called “primitive probability”): The probability that a simple event will occur; for example, that my favorite football team, the Skins, will win on Sunday.

During a recent season, the “experts” said that the Skins had a 60 percent chance of winning on Opening Day; that estimate is a simple probability. We can *model* that probability by putting into an urn six green balls to stand for wins, and four red balls to stand for losses. (Or we could use 60 and 40 balls, or 600 and 400). For the outcome on any given day, we draw one ball from the urn, and record a simulated win if the ball is green, a loss if the ball is red.

So far the urn has served only as a physical representation of our thoughts. But as we shall see shortly, this representation can help us think clearly about the process of interest to us. It can also give us information that is not yet in our thoughts.

Estimating simple probabilities wisely depends largely upon gathering evidence well. It also helps to skillfully adjust one’s probability estimates to make them internally consistent. Estimating probabilities has much in common with estimating lengths, weights, skills, costs, and other subjects of measurement and judgment.

Composite Event: A composite event is the combination of two or more simple events. Examples include all heads in three throws of a single coin; all heads in one throw of three coins at once; Sunday being a nice day *and* the Skins winning; and the birth of nine females out of the next ten calves born if the chance of a female in a single birth is .48.

Compound Probability: The probability that a composite event will occur.

The difficulty in estimating *simple* probabilities such as the chance of the Skins winning Sunday arises from our lack of understanding of the world around us. The difficulty of estimating *compound* probabilities such as the probability of it being a nice day Sunday *and* the Skins winning is the weakness in our mathematical intuition interacting with our lack of understanding of the world around us. Our task in the study of probability and statistics is to overcome the weakness of our mathematical intuition by using a systematic process of simulation (or the devices of formulaic deductive theory).

Consider now a question about a compound probability: What are the chances of the Skins winning their first *two* games if we think that *each* of those games can be modeled by our urn containing six red and four green balls? If one drawing from the urn represents one game, a second drawing should represent the second game (assuming we replace the first ball drawn in order to keep the chances of winning the same for the two games). If so, two drawings from the urn should represent two games. And we can then estimate the compound probability we seek with a series of two-ball trial experiments.

More specifically, our procedure in this case—the prototype of all procedures in the resampling simulation approach to probability and statistics—is as follows:

1. Put six green (“Win”) and four red (“Lose”) balls in an urn.
2. Draw a ball, record its color, and replace it (so that the probability of winning the second simulated game is the same as the first).
3. Draw another ball and record its color.
4. If both balls drawn were green record “Yes”; otherwise record “No.”
5. Repeat steps 2-4 a thousand times.
6. Count the proportion of “Y”'s to the total number of “Y”'s and “N”'s; the result is the probability we seek.

Much the same procedure could be used to estimate the probability of the Skins winning (say) 3 of their next 4 games. We will return to this illustration again and we will see how it enables us to estimate many other sorts of probabilities.

Experiment or Experimental Trial, or Trial, or Resampling Experiment: A simulation experiment or trial is a randomly-generated composite event which has the same characteristics as the actual composite event in which we are interested (except that in inferential statistics the resampling experiment is generated with the “benchmark” or “null” universe rather than with the “alternative” universe).

Parameter: A numerical property of a universe. For example, the “true” mean (don’t worry about the meaning of “true”), and the range between largest and smallest members, are two of its parameters.

Please see the glossary at the end of the book for a complete list of terms used in the book.

Theoretical and historical methods of estimation

As introduced in Chapter 3, there are two general ways to tackle any probability problem: *theoretical-deductive* and *empirical*, each of which has two sub-types. These concepts have complicated links with the concept of “frequency series” discussed earlier.

Empirical Methods. One empirical method is to look at *actual cases in nature*—for example, examine all (or a sample of) the families in Brazil that have four children and count the proportion that have three girls among them. (This is the most fundamental process in science and in information-getting generally. But in general we do not discuss it in this book and leave it to courses called “research methods.” I regard that as a mistake and a shame, but so be it.) In some cases, of course, we cannot get data in such fashion because it does not exist.

Another empirical method is to manipulate the simple elements in such fashion as to produce hypothetical experience with how the simple elements behave. This is the heart of the resampling method, as well as of physical simulations such as wind tunnels.

Theoretical Methods. The most fundamental theoretical approach is to resort to first principles, working with the elements in their full deductive simplicity, and examining all possibilities. This is what we do when we use a tree diagram to calculate the probability of three girls in families of four children.

The formulaic approach is a theoretical method that aims to avoid the inconvenience of resorting to first principles, and

instead uses calculational shortcuts that have been worked out in the past.

What the Book Teaches. This book teaches you the empirical method using hypothetical cases. Formulas can be misleading for most people in most situations, and should be used as a shortcut only when a person understands exactly which first principles are embodied in the formulas. But most of the time, students and practitioners resort to the formulaic approach without understanding the first principles that lie behind them—indeed, their own teachers often do not understand these first principles—and therefore they have almost no way to verify that the formula is right. Instead they use canned checklists of qualifying conditions.

Samples and universes

The terms “sample” and “universe” (or “population”) [5] were used earlier without definition. But now these terms must be defined.

The concept of a sample

For our purposes, a “sample” is a collection of observations for which you obtain the data to be used in the problem. Almost any set of observations for which you have data constitutes a sample. (You might, or might not, choose to call a complete census a sample.)

The concept of a universe or population

For every sample there must also be a universe “behind” it. But “universe” is harder to define, partly because it is often an *imaginary* concept. A universe is the collection of things or people *that you want to say that your sample was taken from*. A universe can be finite and well defined—“all live holders of the Congressional Medal of Honor,” “all presidents of major universities,” “all billion-dollar corporations in the United States.” Of course, these finite universes may not be easy to pin down; for instance, what is a “major university”? And these universes may contain some elements that are difficult to find; for instance, some Congressional Medal winners may have left the country, and there may not be any public records on some billion-dollar corporations.

Universes that are called “infinite” are harder to understand, and it is often difficult to decide which universe is appropri-

ate for a given purpose. For example, if you are studying a sample of schizophrenics, what is the universe from which the sample comes? Depending on your purposes, the appropriate universe might be all schizophrenics now alive, or it might be all schizophrenics who might *ever* live. The latter concept of the universe of schizophrenics is *imaginary* because some of the universe does not exist. And it is *infinite* because it goes on forever.

Not everyone likes this definition of “universe.” Others prefer to think of a universe, not as the collection of people or things that you *want* to say your sample was taken from, but as the collection that the sample was *actually* taken from. This latter view equates the universe to the “sampling frame” (the actual list or set of elements you sample from) which is always finite and existent. The definition of universe offered here is simply the most practical, in my opinion.

The conventions of probability

Let’s review the basic conventions and rules used in the study of probability:

1. Probabilities are expressed as decimals between 0 and 1, like percentages. The weather forecaster might say that the probability of rain tomorrow is .2, or .97.
2. The probabilities of all the possible alternative outcomes in a single “trial” must add to unity. If you are prepared to say that it must either rain or not rain, with no other outcome being possible—that is, if you consider the outcomes to be *mutually exclusive* (a term that will be discussed below), then one of those probabilities implies the other. That is, if you estimate that the probability of rain is .2—written $P(\text{rain}) = .2$ —that implies that you estimate that $P(\text{no rain}) = .8$.

Mutually exclusive events – the addition rule

Definition: If there are just two events a and b and they are “mutually exclusive” or “disjoint,” each implies the absence of the other. Green and red coats are mutually exclusive for you if (but only if) you never wear more than one coat at a time.

To state this idea formally,

$$\text{If } P(a \text{ and } b) = 0$$

then outcomes a and b , and hence outcome a and its own absence (written $P(\bar{a})$), are necessarily mutually exclusive, and hence the two probabilities add to unity:

$$P(A) + P(\bar{A}) = 1.$$

The sales of your store in a given year cannot be both above and below \$1 million. Therefore if $P(\text{sales} > \$1 \text{ mil}) = .2$, $P(\text{sales} \leq \$1 \text{ mil}) = .8$.

This “complements” rule is useful as a consistency check on your estimates of probabilities. If you say that the probability of rain is .2, then you should check that you think that the probability of no rain is .8; if not, reconsider both the estimates. The same for the probabilities of your team winning and losing its next game.

Joint probabilities

Let's return now to the Skins. We said earlier that our best guess of the probability that the Skins will win the first game is .6. Let's complicate the matter a bit and say that the probability of the Skins winning depends upon the weather; on a nice day we estimate a .65 chance of winning, on a nasty (rainy or snowy) day a chance of .55. It is obvious that we then want to know the chance of a nice day, and we estimate a probability of .7. Let's now ask the probability that both will happen—*it will be a nice day and the Skins will win*.

Before getting on with the process of estimation itself, let's tarry a moment to discuss the probability estimates. Where do we get the notion that the probability of a nice day next Sunday is .7? We might have done so by checking the records of the past 50 years, and finding 35 nice days on that date. If we assume that the weather has not changed over that period (an assumption that some might not think reasonable, and the wisdom of which must be the outcome of some non-objective judgment), our probability estimate of a nice day would then be $35/50 = .7$.

Two points to notice here: 1) The source of this estimate is an objective “frequency series.” And 2) the data come to us as the records of 50 days, of which 35 were nice. We would do best to stick with exactly those numbers rather than convert

them into a single number—70 percent. Percentages have a way of being confusing. (When his point score goes up from 2 to 3, my racquetball partner is fond of saying that he has made a “fifty percent increase”; that’s just one of the confusions with percentages.) And converting to a percent loses information: We no longer know how many observations the percent is based upon, whereas $35/50$ keeps that information.

Now, what about the estimate that the Skins have a .65 chance of winning on a nice day—where does that come from? Unlike the weather situation, there is no long series of stable data to provide that information about the probability of winning. Instead, we *construct* an estimate using whatever information or “hunch” we have. The information might include the Skins’ record earlier in this season, injuries that have occurred, what the “experts” in the newspapers say, the gambling odds, and so on. The result certainly is not “objective,” or the result of a stable frequency series. But we treat the .65 probability in quite the same way as we treat the .7 estimate of a nice day. In the case of winning, however, we produce an estimate expressed directly as a percent.

If we are shaky about the estimate of winning—as indeed we ought to be, because so much judgment and guesswork inevitably goes into it—we might proceed as follows: Take hold of an urn and two bags of balls, green and red. Put into the urn some number of green balls—say 10. Now add enough red balls to express your judgment that the *ratio* is the ratio of expected wins to losses on a nice day, adding or subtracting green balls as necessary to get the ratio you want. If you end up with 13 green and 7 red balls, then you are “modeling” a probability of .65, as stated above. If you end up with a different ratio of balls, then you have learned from this experiment with your own mind processes that you think that the probability of a win on a nice day is something other than .65.

Don’t put away the urn. We will be using it again shortly. And keep in mind how we have just been using it, because our use later will be somewhat different though directly related.

One good way to begin the process of producing a compound estimate is by portraying the available data in a “tree diagram” like Figure 4-1. The tree diagram shows the possible events in the order in which they might occur. A tree diagram is extremely valuable whether you will continue with either simulation or the formulaic method.

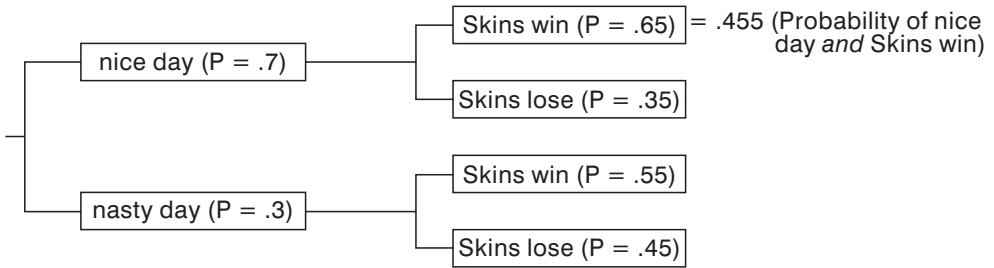


Figure 4-1: Tree Diagram

The monte carlo simulation method (resampling)

The steps we follow to simulate an answer to the compound-probability question are as follows:

1. Put seven blue balls (for “nice day”) and three yellow balls (“not nice”) into an urn labeled A.
2. Draw one ball from urn A. If it is blue, continue; otherwise record “no” and stop.
3. If the ball drawn in step 2 is blue, indicating a nice day and that you should continue the analysis, put 65 green balls (for “win”) and 35 red balls (“lose”) into an urn labeled B.
4. If you have drawn a blue ball from run A, now draw a ball from urn B, and if it is green, record “yes” on a score sheet; otherwise write “no.”
5. Repeat steps 2-4 perhaps 1000 times.
6. Count the number of “yes” trials.
7. Compute the probability you seek as (number of “yeses” / 1000). (This is the same as (number of “yeses” / (number of “yeses” + number of “noes”))

Actually doing the above series of steps by hand is useful to build your intuition about probability and simulation methods. But the procedure can also be simulated with a computer. Using the language RESAMPLING STATS, we produce an estimate as follows:

URN 70#5 30#6 weather

“5” = no rain, “6” = rain

REPEAT 1000

SAMPLE 1 weather a

a draw one “ball” for the weather

IF a = 5

if no rain, check on game outcome

URN 45#7 55#8 winlose

“7” = lose, “8” = win

SAMPLE 1 winlose b

b draw one “ball” for game outcome

IF b = 8

if no rain and a win

SCORE b z

mark a positive result

END

end the inner if-loop

END

end the outer if-loop

END

end the 1000 repeats

COUNT z = 8 k

count the positive results

DIVIDE k 1000 kk

find the proportion positives

PRINT kk

The above procedure gives us the probability that it will be a nice day and the Skins will win—about 38.5 percent.

With the aid of an urn with a different composition—one made by substituting 55 blue and 45 yellow balls in Step 3—a similar procedure yields the chance that it will be a *nasty* day and the Skins will win. With a similar substitution and procedure we could also estimate the probabilities that it will be a nasty day and the Skins will lose, and a nice day and the Skins will lose. The sum of these probabilities should come close to unity, because the sum includes all the possible outcomes. But it will not *exactly* equal unity because of what we call “sampling variation” or “sampling error.”

Please notice that each trial of the procedure begins with the same numbers of balls in the urns as the previous trial. That is, you must replace the balls you draw after each trial in order that the probabilities remain the same from trial to trial. Later we will discuss the general concept of replacement versus non-replacement more fully.

The deductive formulaic method

It also is possible to get an answer with formulaic methods to the question about a nice day and the Skins winning. The following discussion of nice-day-Skins-win handled by formula is a prototype of the formulaic deductive method for handling other problems.

Return now to the tree diagram (Figure 4-1) above. We can read from the tree diagram that 70 percent of the time it will be nice, and of that 70 percent of the time, 65 percent of the games will be wins. That is, $.65 * .7 = .455 =$ the probability of a nice day and a win. That is the answer we seek. The method seems easy, but it also is easy to get confused and obtain the wrong answer.

Multiplication rule

We can generalize what we have just done. The foregoing formula exemplifies what is known as the “multiplication rule”:

$$\mathbf{P(\text{nice day and win}) = P(\text{nice day}) * P(\text{winning} \mid \text{nice day})}$$

where the vertical line in $P(\text{winning} \mid \text{nice day})$ means “conditional upon.” That is, the vertical line indicates a “conditional probability,” a concept we must consider in a minute.

The multiplication rule is a formula that produces the probability of the *combination (juncture) of two or more events*. More discussion of it will follow below.

Conditional and unconditional probabilities

Two kinds of probability statements—*conditional* and *unconditional*—must now be distinguished.

It is the appropriate concept when many factors, all small relative to each other rather than one force having an overwhelming influence, affect the outcome.

A *conditional* probability is formally written $P(\text{Skins win} \mid \text{rain}) = .65$, and it is read “The probability that the Skins will win if (given that) it rains is .65.” It is the appropriate concept when there is one (or more) major event of interest in decision contexts.

Let's use another football example to explain conditional and unconditional probabilities. In the year this is being written, the University of Maryland has an unpromising football team. Someone may nevertheless ask what chance the team has of winning the post season game at the bowl to which only the best team in the University of Maryland's league is sent. One may say that if by some miracle the University of Maryland does get to the bowl, its chance would be a bit less than 50-50—say, .40. That is, the probability of its winning, *conditional* on getting to the bowl is .40. But the chance of its getting to the bowl at all is very low, perhaps .01. If so, the unconditional probability of winning at the bowl is the probability of its getting there multiplied by the probability of winning if it gets there; that is, $.01 \times .40 = .004$. (It would be even better to say that .004 is the probability of winning conditional only on having a team, there being a league, and so on, all of which seem almost sure things.) Every probability is conditional on many things—that war does not break out, that the sun continues to rise, and so on. But if all those unspecified conditions are very sure, and can be taken for granted, we talk of the probability as unconditional.

A conditional probability is a statement that the probability of an event is such-and-such *if* something else is so-and-so; it is the "if" that makes a probability statement conditional. True, in *some* sense all probability statements are conditional; for example, the probability of an even-numbered spade is $6/52$ *if* the deck is a poker deck and not necessarily if it is a pinochle deck or Tarot deck. But we ignore such conditions for most purposes.

Most of the use of the concept of probability in the social sciences is conditional probability. All hypothesis-testing statistics (discussed starting in Chapter 14) are conditional probabilities.

Here is the typical conditional-probability question used in social-science statistics: What is the probability of obtaining this sample S (by chance) *if* the sample were taken from universe A ? For example, what is the probability of getting a sample of five children with I.Q.s over 100 *by chance* in a sample randomly chosen from the universe of children whose average I.Q. is 100?

One way to obtain such conditional-probability statements is by examination of the results generated by universes like the conditional universe. For example, assume that we are considering a universe of children where the average I.Q. is 100.

Write down “over 100” and “under 100” respectively on many slips of paper, put them into a hat, draw five slips several times, and see how often the first five slips drawn are all over 100. This is the resampling (Monte Carlo simulation) method of estimating probabilities.

Another way to obtain such conditional-probability statements is formulaic calculation. For example, if half the slips in the hat have numbers under 100 and half over 100, the probability of getting five in a row above 100 is .03125—that is, $.5^5$, or $.5 \times .5 \times .5 \times .5 \times .5$, using the multiplication rule introduced above. But if you are not absolutely sure you know the proper mathematical formula, you are more likely to come up with a sound answer with the simulation method.

Let’s illustrate the concept of conditional probability with four cards—two aces and two 3’s (or two black and two red). What is the probability of an ace? Obviously, .5. If you first draw an ace, what is the probability of an ace now? That is, what is the probability of an ace *conditional on* having drawn one already? Obviously not .5.

This change in the conditional probabilities is the basis of mathematician Edward Thorp’s famous system of card-counting to beat the casinos at blackjack (Twenty One).

Casinos can defeat card counting by using many decks at once so that conditional probabilities change more slowly, and are not very different than unconditional probabilities. Looking ahead, we will see that sampling with replacement, and sampling without replacement from a huge universe, are much the same in practice, so we can substitute one for the other at our convenience.

Let’s further illustrate the concept of conditional probability with a puzzle (from Gardner, 1983, p. 42). “Shuffle a packet of four cards—two red, two black—and deal them face down in a row. Two cards are picked at random, say by placing a penny on each. What is the probability that those two cards are the same color?”

1. Play the game with the cards 100 times, and estimate the probability sought.

OR

1. Put slips with the numbers “1,” “1,” “2,” and “2” in a hat, or in an array on a computer named N.

2. Shuffle the slips or the array
3. Take the first number in the hat or array and store it someplace—perhaps in a location called A.
4. Take the second number and store it in B.
5. Subtract the numbers in A and B. If the result is 0, record “Y,” otherwise “N.”
6. Repeat (1-5) 1000 times, and count the proportion of “Y’s.” That proportion equals the probability we seek to estimate.

Now let’s play the game differently, first picking one card and *putting it back and shuffling* before picking a second card. What are the results now? You can try it with the cards, or with a computer program similar to the above.

Why do you get different results in the two cases? Let’s ask the question differently: What is the probability of first picking a black card? Clearly, it is 50-50, or .5. Now, if you first pick a black card, what is the probability in the first game above of getting a second black card? There are two red and one black cards left, so now $p = 1/3$.

But in the second game, what is the probability of picking a second black card if the first one you pick is black? It is still .5 because we are *sampling with replacement*.

The probability of picking a second black card *conditional on picking a first black card* in the first game is $1/3$, and it is different from the unconditional probability of picking a black card first. But in the second game the probability of the second black card conditional on first picking a black card is the same as the probability of the first black card.

So the reason you lose money if you play the first game at even odds against a carnival game operator is because the conditional probability is different than the original probability.

And an illustrative joke: The best way to avoid there being a live bomb aboard your plane flight is to take an inoperative bomb aboard with you; the probability of one bomb is very low, and by the multiplication rule, *the probability of two bombs is very very low*. Two hundred years ago the same joke was told about the midshipman who, during a battle, stuck his head through a hole in the ship’s side that had just been made by an enemy cannon ball because he had heard that the probability of two cannonballs striking in the same place was one in a million.

What's wrong with the logic in the joke? The probability of there being a bomb aboard already, *conditional on* your bringing a bomb aboard, is the same as the conditional probability if you do *not* bring a bomb aboard. Hence you change nothing by bringing a bomb aboard, and do not reduce the probability of an explosion.

The skins again, plus leaving the game early

Let's carry exactly the same process one tiny step further. Assume that if the Skins win, there is a .3 chance you will leave the game early. Now let us ask the probability of a nice day, the Skins winning, and you leaving early. You should be able to see that this probability can be estimated with three urns instead of two. Or it can be computed with the multiplication rule as $.65 * .7 * .3 = .1365$ (about .14)—the probability of a nice day and a win and you leave early.

The book shows you the formal method—the multiplication rule, in this case—for several reasons: 1) Simulation is weak with very low probabilities, e.g. $P(50 \text{ heads in } 50 \text{ throws})$. But—a big but—*statistics* and probability is seldom concerned with very small probabilities. Even for games like poker, the orders of magnitude of 5 aces in a wild game with joker, or of a royal flush, matter little. 2) The multiplication rule is wonderfully handy and convenient for quick calculations in a variety of circumstances. A back-of-the-envelope calculation can be quicker than a simulation. And it can also be useful in situations where the probability you will calculate will be very small, in which case simulation can require considerable computer time to be accurate. (We will shortly see this point illustrated in the case of estimating the rate of transmission of AIDS by surgeons.) 3) It is useful to know the theory so that you are able to talk to others, or if you go on to other courses in the mathematics of probability and statistics.

The multiplication rule also has the drawback of sometimes being confusing, however. If you are in the slightest doubt about whether the circumstances are correct for applying it, you will be safer to perform a simulation as we did earlier with the Skins, though in practice you are likely to simulate with the aid of a computer program, as we shall see shortly. So use the multiplication rule only when there is no possibility of confusion. Usually that means using it only when the events under consideration are independent.

Notice that the same multiplication rule gives us the probability of *any particular sequence* of hits and misses—say, a miss, then a hit, then a hit if the probability of a single miss is $2/3$. Among the $2/3$ of the trials with misses on the first shot, $1/3$ will next have a hit, so $2/3 \times 1/3$ equals the probability of a miss then a hit. Of those $2/9$ of the trials, $1/3$ will then have a hit, or $2/3 \times 1/3 \times 1/3 = 2/27$ equals the probability of the sequence miss-hit-hit.

The multiplication rule is very useful in everyday life. It fits closely to a great many situations such as “What is the chance that it will rain (.3) and that (if it does rain) the plane will not fly (.8)?” Hence the probability of your not leaving the airport today is $.3 \times .8 = .24$.