CHAPTER 2 Basic Concepts in Probability and Statistics, Part 1

"Uncertainty, in the presence of vivid hopes and fears, is painful, but must be endured if we wish to live without the support of comforting fairy tales."

Bertrand Russell, *A History of Western Philosophy*
(New York: Simon and Schuster, 1945, p. xiv)

Introduction

The central concept for dealing with uncertainty is probability. Hence we must inquire into the "meaning" of the term probability. (The term "meaning" is in quotes because it can be a confusing word.)

You have been using the notion of probability all your life when drawing conclusions about what you expect to happen, and in reaching decisions in your public and personal lives.

You wonder: Will the footballer's kick from the 45 yard line go through the uprights? How much oil can you expect from the next well you drill, and what value should you assign to that prospect? Will you be the first one to discover a completely effective system for converting speech into computer-typed output? Will the next space shuttle end in disaster? Your answers to these questions rest on the probabilities you estimate.

And you act on the basis of probabilities: You place your blanket on the beach where there is a low probability of someone's kicking sand on you. You bet heavily on a poker hand if there is a high probability that you have the best hand. A hospital decides not to buy another ambulance when the administrator judges that there is a low probability that all the other ambulances will ever be in use at once. NASA decides whether or not to send off the space shuttle this morning as scheduled.

This chapter discusses what is meant by such key terms as "probability," "conditional" and "unconditional" probability, "independence," "sample," and "universe." It discusses the nature and the usefulness of the concept of probability as used in this book, and it touches on the source of basic estimates of probability that are the raw material of statistical inferences. The chapter also distinguishes between probability theory and inferential statistics. (Descriptive statistics, the other main branch of statistics, was discussed briefly in the previous chapter.)

The nature and meaning of the concept of probability

The common meaning of the term "probability" is as follows: *Any particular stated probability is an assertion that indicates how likely you believe it is that an event will occur*.

It is confusing and unnecessary to inquire what probability "really" is. (Indeed, the terms "really" and "is," alone or in combination, are major sources of confusion in statistics and in other logical and scientific discussions, and it is often wise to avoid their use.) Various concepts of probability—which correspond to various common definitions of the term—are useful in particular contexts (see Ayer, 1965; Schlaifer, 1961, Chapter 1; for perhaps the best summary of the probability concept, see Barnett, 1973, Chapter 3). This book contains many examples of the use of probability. Work with them will gradually develop a sound understanding of the concept.

There are two major concepts and points of view about probability—frequency and belief. Each is useful in some situations but not in others. Though they may seem incompatible in principle, there almost never is confusion about which is appropriate in a given situation.

1. *Frequency*. The probability of an event can be said to be the proportion of times that the event has taken place in the past, usually based on a long series of trials. Insurance companies use this when they estimate the probability that a thirty-five-year-old postman will die during a period for which he wants to buy an insurance policy. (Notice this shortcoming: Sometimes you must bet upon events that have never or only infrequently taken place before, and so you cannot reasonably reckon the proportion of times they occurred one way or the other in the past.)

2. *Belief*. The probability that an event will take place or that a statement is true can be said to correspond to the odds at which you would bet that the event will take place. (Notice a shortcoming of this concept: You might be willing to accept a five-dollar bet at 2-1 odds that your team will win the game, but you might be unwilling to bet a hundred dollars at the same odds.)

The connection between gambling and immorality or vice troubles some people about gambling examples. On the other hand, the racy aspects can give the subject a special tang. There are several reasons why statistics use so many gambling examples—and especially tossing coins, throwing dice, and playing cards:

1. *Historical*. The theory of probability began with gambling examples of dice analyzed by Cardano, Galileo, and then by Pascal and Fermat.

2. *Generality*. These examples are not related to any particular walk of life, and therefore they can be generalized to applications in any walk of life. Students in any field—business, medicine, science—can feel equally at home with gambling examples.

3. *Sharpness*. These examples are particularly stark, and unencumbered by the baggage of particular walks of life or special uses.

4. *Universality*. Every other text uses these same examples, and therefore the use of them connects up this book with the main body of writing about probability and statistics.

Often we'll begin with a gambling example and then consider an example in one of the professional fields—such as business and other decision-making activities, biostatistics and medicine, social science and natural science—and everyday living. People in one field often can benefit from examples in others; for example, medical students should understand the need for business decision-making in terms of medical practice, as well as the biostatistical examples. And social scientists should understand the decision-making aspects of statistics if they have any interest in the use of their work in public policy.

The "Meaning" of "Probability"

A probability estimate of .2 indicates that you think there is twice as great a chance of the event happening as if you had estimated a probability of .1. This is the rock-bottom interpretation of the term "probability," and the heart of the concept.

The idea of probability arises when you are not sure about what will happen in an uncertain situation—that is, when you lack information and therefore can only make an estimate. For example, if someone asks you your name, you do not use the concept of probability to answer; you know the answer to a very high degree of surety. To be sure, there is some chance that you do not know your own name, but for all practical purposes you can be quite sure of the answer. If someone asks you who will win tomorrow's ball game, however, there is a considerable chance that you will be wrong no matter what you say. Whenever there is a reasonable chance that your prediction will be wrong, the concept of probability can help you.

The concept of probability helps you to answer the question, "How likely is it that…?" The purpose of the study of probability and statistics is to help you make sound appraisals of statements about the future, and good decisions based upon those appraisals. The concept of probability is especially useful when you have a sample from a larger set of data—a "universe"—and you want to know the probability of various degrees of likeness between the sample and the universe. (The universe of events you are sampling from is also called the "population," a concept to be discussed below.) Perhaps the universe of your study is all high school seniors in 1997. You might then want to know, for example, the probability that the universe's average SAT score will not differ from your sample's average SAT by more than some arbitrary number of SAT points—say, ten points.

I've said that a probability statement is about the future. Well, usually. Occasionally you might state a probability about your future knowledge of past events—that is, "I think I'll find out that..."—or even about the unknown past. (Historians use probabilities to measure their uncertainty about whether events occurred in the past, and the courts do, too, though the courts hesitate to say so explicitly.)

Sometimes one knows a probability, such as in the case of a gambler playing black on an honest roulette wheel, or an insurance company issuing a policy on an event with which it has had a lot of experience, such as a life insurance policy. But often one does not *know* the probability of a future event. Therefore, our concept of probability must include situations where extensive data are not available.

All of the many techniques used to estimate probabilities should be thought of as *proxies* for the actual probability. For example, if Mission Control at Space Central simulates what should and probably will happen in space if a valve is turned aboard a space craft just now being built, the test result on the ground is a proxy for the real probability of what will happen in space.

In some cases, it is difficult to conceive of *any* data that can serve as a proxy. For example, the director of the CIA, Robert Gates, said in 1993 "that in May 1989, the CIA reported that the problems in the Soviet Union were so serious and the situation so volatile that Gorbachev had only a 50-50 chance of surviving the next three to four years unless he retreated from his reform policies" (*The Washington Post*, January 17, 1993, p. A42). Can such a statement be based on solid enough data to be more than a crude guess?

The conceptual probability in any specific situation is *an interpretation of all the evidence that is then available*. For example, a wise biomedical worker's estimate of the chance that a given therapy will have a positive effect on a sick patient should be an interpretation of the results of not just one study in isolation, but of the results of that study plus everything else that is known about the disease and the therapy. A wise policymaker in business, government, or the military will base a probability estimate on a wide variety of information and knowledge. The same is even true of an insurance underwriter who bases a life-insurance or shipping-insurance rate not only on extensive tables of long-time experience but also on recent knowledge of other kinds. The choice of a method of estimating a probability constitutes an *operational definition* of probability.

Digression about Operational Definitions

An operation definition is the all-important intellectual procedure that Einstein employed in his study of special relativity to sidestep the conceptual pitfalls into which discussions of such concepts as probability also often slip. An operational definition is to be distinguished from a property or attribute definition, in which something is defined by saying what it consists of. For example, a crude attribute definition of a college might be “an organization containing faculty and students, teaching a variety of subjects beyond the high-school level.” An operational definition of university might be “an organization found in *The World Almanac's* listing of ‘Colleges and Universities.’” (Simon, 1969, p. 18.)

P. W. Bridgman, the inventor of operational definitions, stated that “the proper definition of a concept is not in terms of its properties but in terms of actual operations.” It was he who explained that definitions in terms of properties had held physics back and constituted the barrier that it took Albert Einstein to crack (Bridgman, 1927, pp. 6-7).

A formal operational definition of “operational definition” may be appropriate. “A definition is an operational definition to the extent that the definer (a) specifies the procedure (including materials used) for identifying or generating the definiendum and (b) finds high reliability for [consistency in application of] his definition” (Dodd, in *Dictionary of Social Science*, p. 476). A. J. Bachrach adds that “the operational definition of a dish ... is its recipe” (Bachrach, 1962, p. 74).

The language of empirical scientific research is made up of instructions that are descriptions of sets of actions or operations (for instance, “turn right at the first street sign”) that someone can follow accurately. Such instructions are called an “operational definition.” An operational definition contains a specification of all operations necessary to achieve the same result.

The language of science also contains theoretical terms (better called “hypothetical terms”) that are not defined operationally.

Back to Proxies

Example of a proxy: The “probability risk assessments” (PRAs) that are made for the chances of failures of nuclear power plants are based, not on long experience or even on laboratory experiment, but rather on theorizing of various kinds—using pieces of prior experience wherever possible, of course. A PRA can cost a nuclear facility $5 million.

Another example: If a manager looks at the sales of radios in the last two Decembers, and on that basis guesses how likely it is that he will run out of stock if he orders 200 radios, then the last two years' experience is serving as a proxy for future experience. If a sales manager just "intuits" that the odds are 3 to 1 (a probability of .75) that the main competitor will not meet a price cut, then all his past experience summed into his intuition is a proxy for the probability that it will really happen. Whether any proxy is a good or bad one depends on the wisdom of the person choosing the proxy and making the probability estimates.

How does one estimate a probability in practice? This involves practical skills not very different from the practical skills required to estimate with accuracy the length of a golf shot, the number of carpenters you will need to build a house, or the time it will take you to walk to a friend's house; we will consider elsewhere some ways to improve your practical skills in estimating probabilities. For now, let us simply categorize and consider in the next section various ways of estimating an ordinary garden variety of probability, which is called an "unconditional" probability.

The various ways of estimating probabilities

Consider the probability of drawing an even-numbered spade from a deck of poker cards (consider the queen as even and the jack and king as odd). Here are several general methods of estimation, the specifics of which constitute an operational definition of probability in this particular case:

1. Experience.

The first possible source for an estimate of the probability of drawing an even-numbered spade is the purely empirical method of *experience*. If you have watched card games casually from time to time, you might simply guess at the proportion of times you have seen even-numbered spades appear—say, "about 1 in 15" or "about 1 in 9" (which is almost correct) or something like that. (If you watch long enough you might come to estimate something like 6 in 52.)

General information and experience are also the source for estimating the probability that the sales of radios this December will be between 200 and 250, based on sales the last two Decembers; that your team will win the football game tomorrow; that war will break out next year; or that a United States astronaut will reach Mars before a Russian astronaut. You simply put together all your relevant prior experience and knowledge, and then make an educated guess.

Observation of repeated events can help you estimate the probability that a machine will turn out a defective part or that a child can memorize four nonsense syllables correctly in one attempt. You watch repeated trials of similar events and record the results.

Data on the mortality rates for people of various ages in a particular country in a given decade are the basis for estimating the probabilities of death, which are then used by the actuaries of an insurance company to set life insurance rates. This is *systematized experience*—called a *frequency series*.

No frequency series can speak for itself in a perfectly objective manner. Many judgments inevitably enter into compiling every frequency series—deciding which frequency series to use for an estimate, choosing which part of the frequency series to use, and so on. For example, should the insurance company use only its records from last year, which will be too few to provide as much data as is preferable, or should it also use death records from years further back, when conditions were slightly different, together with data from other sources? (Of course, no two deaths—indeed, no events of any kind—are *exactly* the same. But under many circumstances they are *practically* the same, and science is only interested in such "practical" considerations.)

In view of the necessarily judgmental aspects of probability estimates, the reader may prefer to talk about "degrees of belief" instead of probabilities. That's fine, just as long as it is understood that we operate with degrees of belief in exactly the same way as we operate with probabilities; the two terms are working synonyms.

There is no *logical* difference between the sort of probability that the life insurance company estimates on the basis of its "frequency series" of past death rates, and the manager's estimates of the sales of radios in December, based on sales in that month in the past two years.

The concept of a probability based on a frequency series can be rendered meaningless when all the observations are repetitions of a single magnitude—for example, the case of all successes and zero failures of space-shuttle launches prior to the Challenger shuttle tragedy in the 1980s; in those data alone there was no basis to estimate the probability of a shuttle failure. (Probabilists have made some rather peculiar attempts over the centuries to estimate probabilities from the length of a zero-defect time series—such as the fact that the sun has never failed to rise (foggy days aside!)—based on the undeniable fact that the longer such a series is, the smaller the probability of a failure; see e.g., Whitworth, 1897/1965, pp. xix-xli. However, one surely has more information on which to act when one has a long series of observations of the same magnitude rather than a short series).

2. Simulated experience.

A second possible source of probability estimates is empirical scientific investigation with repeated trials of the phenomenon. This is an empirical method even when the empirical trials are simulations. In the case of the even-numbered spades, the empirical scientific procedure is to shuffle the cards, deal one card, record whether or not the card is an even-number spade, replace the card, and repeat the steps a good many times. The proportions of times you observe an even-numbered spade come up is a probability estimate based on a frequency series.

You might reasonably ask why we do not just *count* the number of even-numbered spades in the deck of fifty-two cards. No reason at all. But that procedure would not work if you wanted to estimate the probability of a baseball batter getting a hit or a cigarette lighter producing flame.

Some varieties of poker are so complex that experiment is the only feasible way to estimate the probabilities a player needs to know.

The resampling approach to statistics produces estimates of most probabilities with this sort of experimental "Monte Carlo" method. More about this later.

3. Sample space analysis and first principles.

A third source of probability estimates is *counting the possibilities*—the quintessential theoretical method. For example, by examination of an ordinary die one can determine that there are six different numbers that can come up. One can then determine that the probability of getting (say) either a "1" <u>or</u> a "2," on a single throw, is 2/6 = 1/3, because two among the six possibilities are "1" or "2." One can similarly determine that there are two possibilities of getting a "1" *plus* a "6" out of thirty-six possibilities when rolling two dice, yielding a probability estimate of 2/36 = 1/18.

Estimating probabilities by counting the possibilities has two requirements: 1) that the possibilities all be known (and therefore limited), and few enough to be studied easily; and 2) that the probability of each particular possibility be known, for example, that the probabilities of all sides of the dice coming up are equal, that is, equal to 1/6.

4. Mathematical shortcuts to sample-space analysis.

A fourth source of probability estimates is *mathematical calculations*. If one knows by other means that the probability of a spade is 1/4 and the probability of an even-numbered card is 6/13, one can then calculate that the probability of turning up an even-numbered spade is 6/52 (that is, 1/4 x 6/13). If one knows that the probability of a spade is 1/4 and the probability of a heart is 1/4, one can then calculate that the probability of getting a heart *or* a spade is 1/2 (that is 1/4 + 1/4). The point here is not the particular calculation procedures, but rather that one can often calculate the desired probability on the basis of already-known probabilities.

It is possible to estimate probabilities with mathematical calculation only if one knows *by other means* the probabilities of some related events. For example, there is no possible way of mathematically calculating that a child will memorize four nonsense syllables correctly in one attempt; empirical knowledge is necessary.

5. Kitchen-sink methods.

In addition to the above four categories of estimation procedures, the statistical imagination may produce estimates in still other ways such as a) the salesman's seat-of-the-pants estimate of what the competition's price will be next quarter, based on who-knows-what gossip, long-time acquaintance with the competitors, and so on, and b) the probability risk assessments (PRAs) that are made for the chances of failures of nuclear power plants based, not on long experience or even on laboratory experiment, but rather on theorizing of various kinds—using pieces of prior experience wherever possible, of course. Any of these methods may be a combination of theoretical and empirical methods.

Consider the estimation of the probability of failure for the tragic flight of the Challenger shuttle, as described by the famous physicist Nobelist Richard Feynman. This is a very real case that includes just about every sort of complication that enters into estimating probabilities.

> …Mr. Ullian told us that 5 out of 127 rockets that he had looked at had failed—a rate of about 4 percent. He

took that 4 percent and divided it by 4, because he assumed a manned flight would be safer than an unmanned one. He came out with about a 1 percent chance of failure, and that was enough to warrant the destruct charges.

But NASA [the space agency in charge] told Mr. Ullian that the probability of failure was more like 1 of 10^5.

I tried to make sense out of that number. "Did you say 1 in 10^5?"

"That's right; 1 in 100,000."

"That means you could fly the shuttle *every day* for an average of *300 years* between accidents—every day, one flight, for 300 years—which is obviously crazy!"

"Yes, I know," said Mr. Ullian. "I moved my number up to 1 in 1000 to answer all of NASA's claims—that they were much more careful with manned flights, that the typical rocket isn't a valid comparison, etcetera...."

But then a new problem came up: the Jupiter probe, *Galileo*, was going to use a power supply that runs on heat generated by radioactivity. If the shuttle carrying *Galileo* failed, radioactivity could be spread over a large area. So the argument continued: NASA kept saying 1 in 100,000 and Mr. Ullian kept saying 1 in 1000, at best.

Mr. Ullian also told us about the problems he had in trying to talk to the man in charge, Mr. Kingsbury: he could get appointments with underlings, but he never could get through to Kingsbury and find out how NASA got its figure of 1 in 100,000 (Feynman, 1989, pp. 179–180). Feynman tried to ascertain more about the origins of the figure of 1 in 100,000 that entered into NASA's calculations. He performed an experiment with the engineers:

…"Here's a piece of paper each. Please write on your paper the answer to this question: what do you think is the probability that a flight would be uncompleted due to a failure in this engine?"

They write down their answers and hand in their papers. One guy wrote "99-44/100% pure" (copying the Ivory soap slogan), meaning about 1 in 200. Another guy wrote something very technical and highly quantitative in the standard statistical way, carefully defining everything, that I had to translate—which also meant about 1 in 200. The third guy wrote, simply, "1 in 300."

Mr. Lovingood's paper, however, said,

Cannot quantify. Reliability is judged from:

- past experience
- quality control in manufacturing
- engineering judgment

"Well," I said, "I've got four answers, and one of them weaseled." I turned to Mr. Lovingood: "I think you weaseled."

"I don't think I weaseled."

"You didn't tell me *what* your confidence was, sir; you told me *how* you determined it. What I want to know is: after you determined it, what *was* it?"

He says, "100 percent"—the engineers' jaws drop, my jaw drops; I look at him, everybody looks at him—"uh, uh, minus epsilon!"

So I say, "Well, yes; that's fine. Now, the only problem is, WHAT IS EPSILON?"

He says, "10^{-5}." It was the same number that Mr. Ullian had told us about: 1 in 100,000.

I showed Mr. Lovingood the other answers and said, "You'll be interested to know that there *is* a difference between engineers and management here—a factor of more than 300."

He says, "Sir, I'll be glad to send you the document that contains this estimate, so you can understand it."

*Later, Mr. Lovingood sent me that report. It said things like "The probability of mission success is necessarily very close to 1.0"—does that mean it is close to 1.0, or it *ought to be* close to 1.0?—and "Historically, this high degree of mission success has given rise to a difference in philosophy between unmanned and manned space flight programs; i.e., numerical probability versus engineering judgment." As far as I can tell, "engineering judgment" means they're just going to make up numbers! The probability of an engine-blade failure was given as a universal constant, as if all the blades were exactly the same, under the same conditions. The whole

> paper was quantifying everything. Just about every nut and bolt was in there: "The chance that a HPHTP pipe will burst is 10^{-7}." You can't estimate things like that; a probability of 1 in 10,000,000 is almost impossible to estimate. It was clear that the numbers for each part of the engine were chosen so that when you add everything together you get 1 in 100,000. (Feynman, 1989, pp. 182-183).

We see in the Challenger shuttle case very mixed kinds of inputs to actual estimates of probabilities. They include frequency series of past flights of other rockets, judgments about the relevance of experience with that different sort of rocket, adjustments for special temperature conditions (cold), and much much more. There also were complex computational processes in arriving at the probabilities that were made the basis for the launch decision. And most impressive of all, of course, are the extraordinary differences in estimates made by various persons (or perhaps we should talk of various statuses and roles) which make a mockery of the notion of objective estimation in this case.

Working with different sorts of estimation methods in different sorts of situations is not new; practical statisticians do so all the time. The novelty here lies in making no apologies for doing so, and for raising the practice to the philosophical level of a theoretically-justified procedure—the theory being that of the operational definition.

The concept of probability varies from one field of endeavor to another; it is different in the law, in science, and in business. The concept is most straightforward in decision-making situations such as business and gambling; there it is crystal-clear that one's interest is entirely in making accurate predictions so as to advance the interests of oneself and one's group. The concept is most difficult in social science, where there is considerable doubt about the aims and values of an investigation. In sum, one should not think of what a probability "is" but rather how best to estimate it. In practice, neither in actual decision-making situations nor in scientific work—nor in classes—do people experience difficulties estimating probabilities because of philosophical confusions. Only philosophers and mathematicians worry—and even they really do not *need* to worry—about the "meaning" of probability.

This topic is continued in the following chapter.