PRODUCING MODELS USING PROBABILITY AND SIMULATION

This chapter discusses how to use probability as a tool to judge the distribution of data under a given model. In the **multiple-choice section**, this topic appears in eight to 12 out of 40 questions. In the **free-response section**, this topic appears in one or two out of six questions.

PROBABILITY

Words referring to probability or chance are commonly used in conversation. For example, we often come across statements like these:

- It is likely to rain today, so please take your umbrella with you.
- It was an easy test. I'll probably get an A on it.
- The Yankees have a much better chance of winning than the Mets.

Words like "probably," "likely," and "chance" carry similar meanings in conversation. They all convey uncertainty. By using probability, we can also make a numerical statement about uncertainty. For example, bank managers can never know exactly when their depositors will make a withdrawal or exactly how much they'll withdraw. Managers also know that though most loans they've granted will be paid back, some of them will result in defaults—but they can't know exactly which ones. In other words, a variety of outcomes is possible, and therefore bank managers can never know exactly how much money the bank will have at any given moment in the future. However, the bankers can use the rules of probability and their past experience to make a reasonable estimation, and then use that estimation when making business decisions.

WHAT IS "PROBABILITY"?

Probability is a measure of the likelihood of an event. Consider a fair coin toss. What makes this coin toss "fair"? We call it fair if the coin's chance of showing heads when flipped is the same as its chance of showing tails—in other words, if there is a 50 percent chance of its showing heads and a 50 percent chance of its showing tails. Suppose we tossed the coin twice and got two heads. Does that mean this coin toss was not fair? What if we toss the coin three times? What do we expect to happen? Let's toss a coin five, 10, 15, 20, 25, and more times and count the number of heads. Then we can calculate the probability of getting heads in a toss and plot that figure on a graph:

$$P(\text{Heads in a toss}) = \frac{\text{Number of heads}}{\text{Number of tosses}}$$

$$P(\text{Percent heads}) = \frac{\text{Number of heads}}{\text{Number of tosses}} \times 100$$

Table 1 lists the results of one such experiment, and the plot in Figure 1 shows them graphically.

Number of Tosses	Number of Heads	*P*(Heads)	Percent of Heads	Number of Tosses	Number of Heads	*P*(Heads)	Percent of Heads
2	0	0	0	35	17	0.48571	48.571
3	2	0.66667	66.667	40	18	0.45	45
4	3	0.75	75	45	18	0.4	40
5	5	1	100	50	23	0.46	46
6	3	0.5	50	60	32	0.53333	53.333
7	5	0.71429	71.429	70	29	0.41429	41.429
8	5	0.625	62.5	80	34	0.425	42.5
9	7	0.77778	77.778	90	48	0.53333	53.333
10	4	0.4	40	100	49	0.49	49
15	10	0.66667	66.667	150	74	0.49333	49.333
20	9	0.45	45	200	106	0.53	53
25	12	0.48	48	500	264	0.528	52.8
30	17	0.56667	56.667	1,000	508	0.508	50.8

Table 1: Number of heads shown in different numbers of tosses

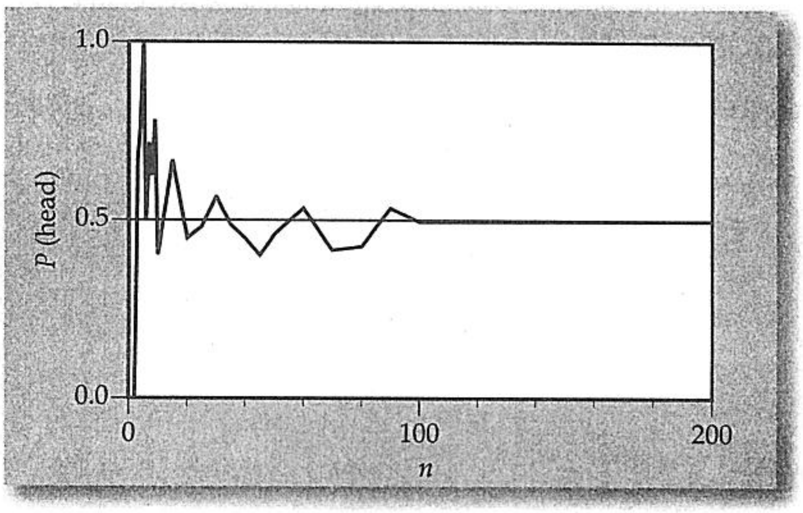

Figure 1: *P*(heads) estimated from different numbers of tosses

Notice that as the number of tosses increases, the percent of times that the coin lands on heads gets closer and closer to 50 percent. In other words, in the long run, the relative frequency of getting heads approaches 0.5, which is what we expected it to be. This relative frequency reflects the concept of probability. In fact, in the long run, the relative frequency of the occurrence of any specific event will always approach the expected value, also known as the probability. Random events are events that cannot be predicted in the short term, but do produce patterns (such as the 50/50 nature of a fair coin toss in the long run).

*Law of Large Numbers

SAMPLE SPACE

Any process that results in an observation or an outcome is an experiment. An experiment may have more than one possible outcome. A set of all possible outcomes of an experiment is known as a **sample space**. It is generally denoted using the letter S.

- Tossing a coin will result in one of two possible outcomes, heads or tails. Therefore, the sample space of tossing a coin is

$$S = \{\text{Heads, Tails}\}$$

- Throwing a die will result in one of six possible outcomes. The resulting sample space is

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

- Tossing two coins will result in one of four possible outcomes. We can indicate the outcome of each of the two tosses by using a pair of letters, the first letter of which indicates the outcome of tossing the first coin and the second letter the outcome of tossing the second coin. H is for heads and T for tails. Then the resulting sample space is

$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

The outcomes listed in a sample space are never repeated, and no outcome is left out. Two events are said to be equally likely if one does not occur more often than the other. For example, the six possible outcomes for a throw of a die are equally likely.

A **tree diagram** representation is useful in determining the sample space for an experiment, especially if there are relatively few possible outcomes. For example, imagine an experiment in which a die and a quarter are tossed together. What are all the possible outcomes? The six possible outcomes of throwing a die are 1, 2, 3, 4, 5, and 6. The two possible outcomes of tossing a quarter are heads (H) and tails (T). Figure 2 is a tree diagram of the possible outcomes:

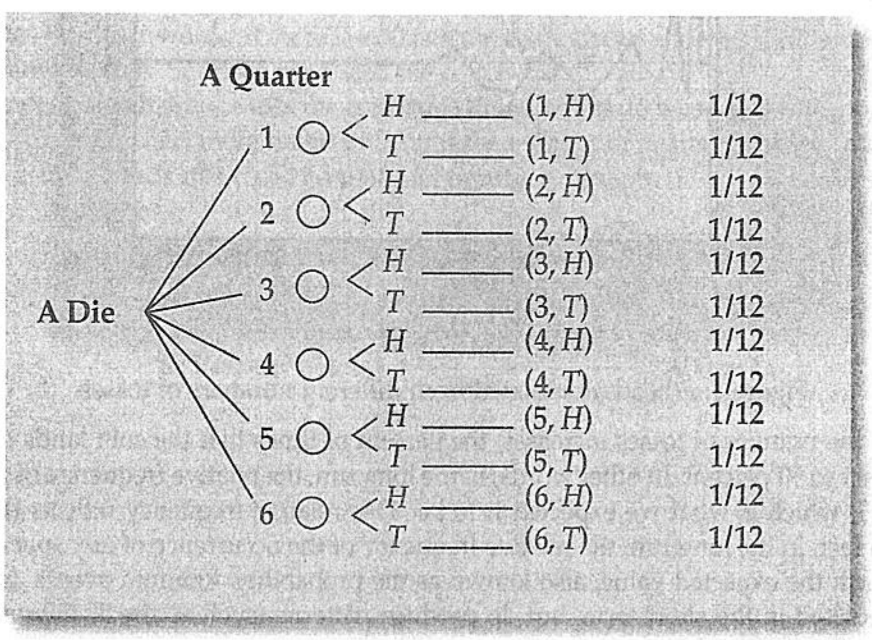

Figure 2: Tree diagram

Looking at the tree diagram, it is easy to see that the sample space is

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

The first letter of each pair represents the outcome of throwing the die and the second represents the outcome of tossing the coin. All 12 outcomes are equally likely. Therefore, the probability of each outcome is $\frac{1}{12}$.

It is common practice to use capital letters to indicate events. For example, one may define

A = getting an even number when a die is thrown = {2, 4, 6}

B = getting two heads when two coins are tossed simultaneously = {(H, H)}

The probability of an event is generally denoted by a capital P followed by the name of the event in parentheses: P(the event). If all the events in a sample space are equally likely, then by using the concept of relative frequency, we can compute the probability of an event as

$$P(\text{An event}) = \frac{\text{Number of outcomes that lead to the event}}{\text{Total number of possible outcomes}}$$

Applying this to the events defined earlier—A (getting an even number in the toss of a die) and B (getting two heads when two coins are tossed)—we get:

- $P(A) = \frac{3}{6} = \frac{1}{2} = 0.5$. The probability of getting an even number when a six-sided die is thrown is 0.5. In other words, there is a 50 percent chance of getting an even number when a six-sided die is thrown.
- $P(B) = \frac{1}{4} = 0.25$. The probability of getting two heads when two coins are tossed simultaneously is 0.25. In other words, there is a 25 percent chance of getting two heads when two coins are tossed simultaneously.