Chapter 19

Simulation

Lecture Slides

Case Study: Simulation 1

In a horse race, it is advantageous to have a starting position that is near the inside of the track.

To ensure fairness, starting position is determined by a random draw before the race. All positions are equally likely.

During the summer and autumn of 2007, the members of the Ohio Racing Commission noticed that one trainer appeared to have an unusually good run of luck.

Case Study: Simulation 2

The Ohio Racing Commission believed that the trainer's run of luck was too good to be true. Is it?

A mathematician can show that the chance of receiving one of the three inside positions at least 30 times in 35 races is very small.

The Ohio Racing Commission suspected that cheating had occurred.

Case Study: Simulation 3

But the trainer had entered horses in nearly 1000 races over several years.

Perhaps it was inevitable that at some time over the course of these nearly 1000 races the trainer would have a string of 35 races in which he received one of the three inside positions at least 30 times.

It came to the attention of the Ohio Racing Commission only because it was one of those seemingly surprising coincidences discussed in Chapter 17. Perhaps the accusation of cheating was unfounded. Recall the experiment of Vargas (Ch. 2)!

Case Study: Simulation 4

Calculating the probability that in a sequence of 1000 races with varying numbers of horses, there would occur a string of 35 consecutive races in which one would receive one of the three inside positions at least 30 times is very difficult.

How to go about determining this probability is the subject of this chapter. By the end of the chapter, you will be able to describe how to find this probability.

Where Do Probabilities Come From? 1

Probability can be determined by observing the event in a large number of trials.

Personal probabilities are one's own judgement about the chance of an event occurring.

What about the probability that we get a run of three straight heads somewhere in 10 tosses of a coin?

Recall Varga's experiment!

We can find this probability by calculation from a model that describes tossing coins.

Where Do Probabilities Come From? 2

Probability models allow us to calculate the probabilities of complicated events starting from an assignment of probabilities to simple events such as “heads on one toss.”

This is true whether the model reflects probabilities from data or personal probabilities.

Unfortunately, the math needed to do probability calculations is often tough. We can use a computer to simulate many repetitions.

Where Do Probabilities Come From? 3

Simulation is easier than math and much faster than actually running many repetitions in the real world.

You might compare finding probabilities by simulation to practicing flying in a computer-controlled flight simulator. Both kinds of simulation are in wide use.

Both have similar drawbacks: They are only as good as the model you start with. Flight simulators use a software model of how an airplane reacts. Simulations of probabilities use a probability model.

Where Do Probabilities Come From? 4

Using random digits from a table or from computer software to imitate chance behavior is called **simulation**.

We look at simulation partly because it is how scientists and engineers really do find probabilities in complex situations.

Simulations are used to develop strategies for reducing waiting times in lines to speak to a teller at banks, in lines to check in at airports, and in lines to vote during elections.

Where Do Probabilities Come From? 5

Simulations are used to study the effects of changes in greenhouse gases on the climate.

Simulations are used to study the effects of catastrophic events, such as the failure of a nuclear power plant, the effects on a structure of the explosion of a nuclear device, or the progression of a deadly, infectious disease in a densely populated city.

Where Do Probabilities Come From? 6

We also look at simulation because simulation forces us to think clearly about probability models.

We'll do the hard part (setting up the model) and leave the easy part (telling a computer to do 10,000 repetitions) to those who really need the right probability at the end.

Simulation Basics 1

Simulation is an effective tool for finding probabilities of complex events once we have a trustworthy probability model.

We can use random digits to simulate many repetitions quickly.

The proportion of repetitions on which an event occurs will eventually be close to its probability, so simulation can give good estimates of probabilities.

Example: Doing a Simulation 1

Toss a coin 10 times. What is the probability of a run of at least three consecutive heads or three consecutive tails?

Step 1. Give a probability model. Our model for coin tossing has two parts:

- Each toss has probabilities 0.5 for a head and 0.5 for a tail.
- Tosses are independent of each other. That is, knowing the outcome of one toss does not change the probabilities for the outcomes of any other toss.

Example: Doing a Simulation 2

Step 2. Assign digits to represent outcomes. Digits in Table A of random digits will stand for the outcomes in a way that matches the probabilities from Step 1.

We know that each digit in Table A has probability 0.1 of being any one of 0, 1, 2, 3, 4, 5, 6, 7, 8, or 9, and that successive digits in the table are independent.

Here is one assignment of digits for coin tossing:

- One digit simulates one toss of the coin.
- Odd digits represent heads; even digits represent tails.

Example: Doing a Simulation 3

Step 3. Simulate many repetitions. Ten digits simulate 10 tosses, so looking at 10 consecutive digits in Table A simulates one repetition. Keep track of whether or not the event we want (a run of three heads or three tails) occurs on each repetition.

Here are the first three repetitions, starting at line 101 in Table A.

Repetition 1	Repetition 2	Repetition 3
1 9 2 2 3 9 5 0 3 4	0 5 7 5 6 2 8 7 1 3	9 6 4 0 9 1 2 5 3 1
H H T T H H H T H T	T H H H T T T H H H	H T T T H H T H H H
YES	YES	YES

Example: Doing a Simulation 4

Continuing in Table A, we did 25 repetitions; 23 of them did have a run of 3 or more heads or tails. So we estimate the probability of a run by the proportion

$proportion = \frac{23}{25} = 0.92$.

Twenty-five repetitions are not enough to be confident that our estimate is accurate. We can have a computer to do many thousands of repetitions.

A long simulation (or hard mathematics) finds that the true probability is about 0.826, so even our small simulation didn't do too badly.

Simulation Basics 2

Two random phenomena are **independent** if knowing the outcome of one does not change the probabilities for outcome of the other.

It is plausible that repeated tosses of a coin are independent (the coin has no memory), and observation shows that they are.

Step 2 (assigning digits) rests on the properties of the random digit table. Here are some examples of this step.

Example: Assigning Digits for Simulation 1

In the United States, the Age Discrimination Employment Act (ADEA) forbids age discrimination against people who are age 40 and over.

Terminating an “unusually” large proportion of employees age 40 and over can trigger legal action.

Simulation can help determine what might be an “unusual” pattern of terminations.

How might we set up such a simulation?

Example: Assigning Digits for Simulation 2

(a) Choose one employee at random from a group of which 40% are age 40 and over. One digit simulates one employee:

40% 0, 1, 2, 3 = age 40 and over

60% 4, 5, 6, 7, 8, 9 = under age 40

(b) Choose one employee at random from a group of which 43% are age 40 and over. Now two digits simulate one person:

43% 00, 01, 02, . . . , 42 = age 40 and over

57% 43, 74, 75, . . . , 99 = under age 40

Example: Assigning Digits for Simulation 3

(c) Choose one employee at random from a group of which 30% are age 40 and over and have no plans to retire, 10% are age 40 and over and plan to retire in the next few months, and 60% are under age 40. There are now three possible outcomes, but the principle is the same. One digit simulates one person:

0, 1, 2 = age 40 and over and have no plans to retire

3 = age 40 and over and plan to retire in the next few months

4, 5, 6, 7, 8, 9 = under age 40

Thinking about Independence 1

Independence can be verified only by observing many repetitions of random phenomena.

But it is probably more accurate to say that a lack of independence can be verified only by observing many repetitions of random phenomena.

How does one recognize that two random phenomena are not independent?

One approach might be to apply the definition of "independence."

Thinking about Independence 2

For a sequence of tosses of a fair coin, one could compute the proportion of times in the sequence that a toss is followed by the same outcome: in other words, the frequency with which a head is followed by a head or a tail is followed by a tail.

This proportion should be close to 0.5 if tosses are independent (knowing the outcome of one toss does not change the probabilities for outcomes of the next) and if many tosses have been observed.

50%

$P(TT) + P(HH) = P(T)P(T) + P(H)P(H) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

Thinking about Independence 3

Another method for assessing independence is based on the concept of correlation, which we discussed in Chapter 14.

If two random phenomena have numerical outcomes, and we observe both phenomena in a sequence of *n* trials, we can compute the correlation for the resulting data.

If the random phenomena are independent, there will be no straight-line relationship between them, and the correlation should be close to 0.

Thinking about Independence 4

It is not necessarily true that two random phenomena are independent if their correlation is 0.

Independence implies no relationship at all, but correlation measures the strength of only a straight-line relationship.

Because independence implies no relationship, we would expect to see no overall pattern in a scatterplot of the data if the variables are independent.

Looking at scatterplots is another method for determining if independence is lacking.

More Elaborate Simulations 1

The building and simulation of random models constitute a powerful tool of contemporary science, yet a tool that can be understood without advanced mathematics.

Having in mind these two goals of understanding simulation for itself and understanding simulation to understand probability, let us look at two more elaborate examples.

The first still has independent trials, but there is no longer a fixed number of trials as there was when we tossed a coin 10 times.

Example: We Want a Girl 1

A couple plan to have children until they have a girl or until they have three children, whichever comes first. What is the probability that they will have a girl among their children?

Step 1. The probability model is like that for coin tossing:

- Each child has probability 0.49 of being a girl and 0.51 of being a boy. (Yes, more boys than girls are born. Boys have higher infant mortality, so the sexes even out soon.)
- The sexes of successive children are independent.

Example: We Want a Girl 2

Step 2. Two digits simulate the sex of one child. We assign 49 of the 100 pairs to “girl” and the remaining 51 to “boy”:

00, 01, 02, . . . , 48 = girl
49, 50, 51, . . . , 99 = boy

Example: We Want a Girl 3

Step 3. To simulate one repetition of this childbearing strategy, read pairs of digits from Table A until the couple have either a girl or three children.

The number of pairs needed to simulate one repetition depends on how quickly the couple get a girl.

Example: We Want a Girl 4

Here are 10 repetitions, simulated using line 130 of Table A.

To interpret the pairs of digits, we have written G for girl and B for boy under them, have added space to separate repetitions, and under each repetition have written "+" if a girl was born and "−" if not.

00, ..., 48 ← G
49, ..., 99 ← B

6905	16	48	17	8717	40	9517	845340	648987	20
B G	G	G	G	B G	G	B G	B B G	B B B	G
+	+	+	+	+	+	+	+	−	+
1	2	3	4	5	6	7	8	9	10

Example: We Want a Girl 5

In these 10 repetitions, a girl was born 9 times.

Our estimate of the probability that this strategy will produce a girl is therefore

$$estimated\ probability = \frac{9}{10} = 0.9$$

Some mathematics shows that, if our probability model is correct, the true probability of having a girl is 0.867.

Unless the couple are unlucky, they will succeed in having a girl.

More Elaborate Simulations 2

Our final example has stages that are *not independent*.

That is, the probabilities at one stage depend on the outcome of the preceding stage.

Example: Kidney Transplant 1

Morris's kidneys have failed, and he is awaiting a kidney transplant. There is a 90% chance he will survive the transplant operation and 10% he will die.

STAGE 1

The transplant succeeds in 60% of those who survive, and the other 40% must return to kidney dialysis.

STAGE 2

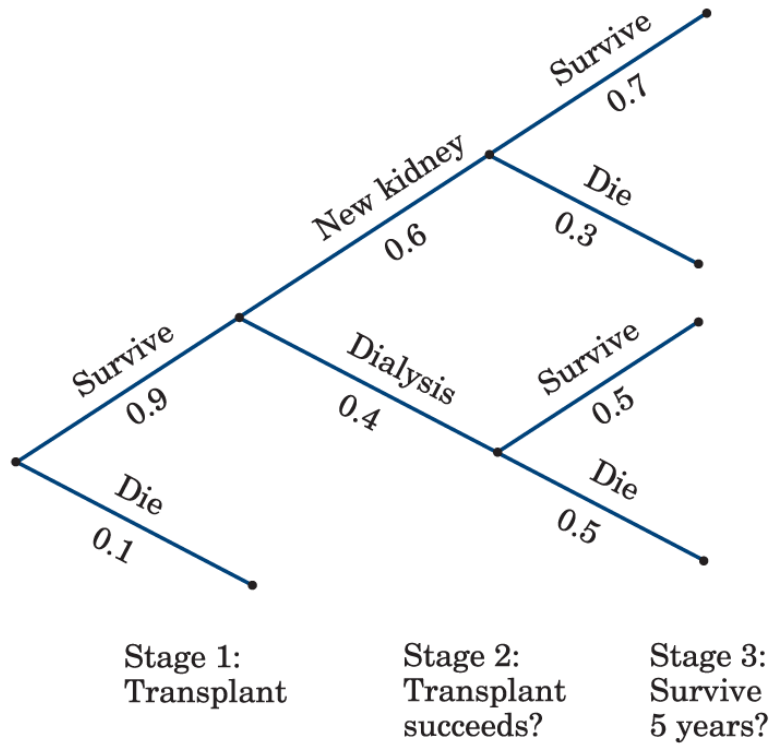
The proportions who survive for at least 5 years are 70% for those with a new kidney and 50% for those who return to dialysis.

STAGE 3

Morris wants to know the probability that he will survive for at least 5 years.

Example: Kidney Transplant 2

Step 1. This tree diagram helps organize the probabilities.



Moore/Notz, *Statistics: Concepts and Controversies*, 10e,

Example: Kidney Transplant 3

Step 2. Assign digits to outcomes:

Stage 1: 0 = die

1, 2, 3, 4, 5, 6, 7, 8, 9 = survive

Stage 2: 0, 1, 2, 3, 4, 5 = transplant succeeds

6, 7, 8, 9 = return to dialysis

Example: Kidney Transplant 4

Step 2. Assign digits to outcomes:

Stage 3 with kidney:

0, 1, 2, 3, 4, 5, 6 = survive 5 years
7, 8, 9 = die

Stage 3 with dialysis:

0, 1, 2, 3, 4 = survive 5 years
5, 6, 7, 8, 9 = die

The assignment of digits at Stage 3 depends on the outcome of Stage 2. That's lack of independence.

Example: Kidney Transplant 5

Step 3. Here are simulations of several repetitions, each arranged vertically. We used random digits from line 110 of Table A. 110: $\underbrace{384}\ \underbrace{484}\ \underbrace{878}\ \underbrace{918}$ .....

	Repetition 1		Repetition 2		Repetition 3		Repetition 4	
Stage 1	3	Survive	4	Survive	8	Survive	9	Survive
Stage 2	8	Dialysis	8	Dialysis	7	Dialysis	1	Kidney
Stage 3	4	Survive	4	Survive	8	Die	8	Die

Morris survives 5 years in two of our four repetitions. Use a computer to do many repetitions!

From a long simulation or from mathematics, we find that Morris has probability 0.558 of living for at least 5 years.

Statistics in Summary 1

- We can use random digits to **simulate** random outcomes if we know the probabilities of the outcomes. Use the fact that each random digit has probability 0.1 of taking any one of the 10 possible digits and that all digits in the random number table are **independent** of each other.

Statistics in Summary 2

- To simulate more complicated random phenomena, string together simulations of each stage. A common situation is several independent trials with the same possible outcomes and probabilities for each trial. Other simulations may require varying numbers of trials or different probabilities at each stage or may have stages that are not independent, so probabilities at some stage depend on the outcome of earlier stages.
- Key to successful simulation is thinking carefully about the probability model. A **tree diagram** can be helpful by giving the probability model in graphical form.