CHAPTER 10 The Procedures of Monte Carlo Simulation (and Resampling)

Until now, the steps to follow in solving particular problems have been chosen to fit the specific facts of that problem. And so they always must. Now let's generalize what we have done in the previous chapters on probability into a general procedure for such problems, which will in turn become the basis for a detailed procedure for resampling simulation in statistics. The generalized procedure describes what we are doing when we estimate a probability using Monte Carlo simulation problem-solving operations.

A definition and general procedure for Monte Carlo simulation

This is what we shall mean by the term *Monte Carlo simulation* when discussing problems in probability: *Using the given data-generating mechanism (such as a coin or die) that is a model of the process you wish to understand, produce new samples of simulated data, and examine the results of those samples*. That's it in a nutshell. In some cases, it may also be appropriate to amplify this procedure with additional assumptions.

This definition fits both problems in pure probability as well as problems in statistics, but in the latter case the process is called *resampling*. The reason that the same definition fits is that *at the core of every problem in inferential statistics lies a problem in probability*; that is, the procedure for handling every statistics problem is the procedure for handling a problem in probability. (There is related discussion of definitions in Chapters 4 and 14.)

The following series of steps should apply to all problems in probability. I'll first state the procedure straight through without examples, and then show how it applies to individual examples.

Step A. Construct a simulated "universe" of cards or dice or some other randomizing mechanism whose composition is similar to the universe whose behavior we wish to describe and investigate. The term "universe" refers to the system that is relevant for a single simple event.

Step B. Specify the procedure that produces a pseudo-sample which simulates the real-life sample in which we are interested. That is, specify the procedural rules by which the sample is drawn from the simulated universe. These rules must correspond to the behavior of the real universe in which you are interested. To put it another way, the simulation procedure must produce simple experimental events with the same probabilities that the simple events have in the real world.

Step C. If several simple events must be combined into a composite event, and if the composite event was not described in the procedure in step B, describe it now.

Step D. Calculate the probability of interest from the tabulation of outcomes of the resampling trials.

Now let us apply the general procedure to some examples to make it more concrete.

Here are four problems to be used as illustrations:

1. If on average 3 percent of the gizmos sent out are defective, what is the chance that there will be more than 10 defectives in a shipment of 200?

2. What are the chances of getting three or more girls in the first four children, if the probability of a female birth is 106/206?

3. What are the chances of Joe Hothand scoring 20 or fewer baskets in 57 shots if his long-run average is 47 percent?

4. What is the probability of two or more people in a group of 25 persons having the same birthday—i. e., the same month and same day of the month?

Step A. Construct a simulated "universe" of cards or dice or some other randomizing mechanism whose composition is similar to the universe whose behavior we wish to describe and investigate. The term "universe" refers to the system that is relevant for a single simple event. For example:

> 1. A random drawing with replacement from the set of numbers 1...100 with 1...3 designated as defective

simulates the system that produces 3 defective gizmos among 100.

2. A coin with two sides, or a random drawing from two sets of random numbers “1-105” and “106-205,” simulates the system that produces a single male or female birth, when we are estimating the probability of three girls in the first four children. Notice that in this universe the probability of a girl remains the same from trial event to trial event—that is, the trials are independent—demonstrating a universe from which we sample with replacement.

3. A random drawing with replacement from an urn containing a hundred balls, 47 red and 53 black, simulates the system that produces 47 percent baskets for Joe Hothand.

4. A random drawing with replacement from the numbers 1...365 simulates the system that produces a birthday.

This step A includes two operations:

1. Decide which symbols will stand for the elements of the universe you will simulate.

2. Determine whether the sampling will be with or without replacement. (This can be ambiguous in a complex modeling situation.)

Hard thinking is required in order to determine the appropriate “real” universe whose properties interest you.

Step B. Specify the procedure that produces a pseudo-sample that simulates the real-life sample in which we are interested. That is, specify the procedural rules by which the sample is drawn from the simulated universe. These rules must correspond to the behavior of the real universe in which you are interested. To put it another way, the simulation procedure must produce simple experimental events with the same probabilities that the simple events have in the real world.

For example:

1. For a single gizmo, you can draw a single number from an infinite universe. Or one can use a finite set with replacement and shuffling,

2. In the case of three or more daughters among four children, you can draw a card and then replace it if

you are using a deck of red and black cards and you are assuming a female birth is 50-50. Or if you are using a random-numbers table, the random numbers automatically simulate replacement. Just as the chances of having a boy or a girl do not change depending on the sex of the preceding child, so we want to ensure through sampling with replacement that the chances do not change each time we choose from the deck of cards.

3. In the case of Joe Hothand's shooting, the procedure is to consider the numbers "1-47" as "baskets," and "48-100" as "misses," with the same other considerations as the gizmos.

4. In the case of the birthday problem, the drawing obviously must be with replacement.

Recording the outcome of the sampling must be indicated as part of this step, e.g., "record 'yes' if girl or basket, 'no' if a boy or a miss."

Step C. If several simple events must be combined into a composite event, and if the composite event was not described in the procedure in step B, describe it now. For example:

1. For the gizmos, draw a sample of 200.

2. For the three or more girls among four children, the procedure for each simple event of a single birth was described in step B. Now we must specify repeating the simple event four times, and counting whether the outcome is or is not three girls.

3. In the case of Joe Hothand's shots, we must draw 57 numbers to make up a sample of shots, and examine whether there are 20 or more misses.

Recording the results as "ten or more defectives," "three or more girls" or "two or less girls," and "20 or more misses" or "19 or fewer," is part of this step. This record indicates the results of all the trials and is the basis for a tabulation of the final result.

Step D. Calculate the probability of interest from the tabulation of outcomes of the resampling trials. For example: the proportions of "yes" and "no," and "20 or more" and "19 or fewer" estimate the probability we seek in step C.

The above procedure is similar to the procedure followed with the analytic formulaic method except that the latter method constructs notation and manipulates it.

Summary

This chapter describes more generally the specific steps used in prior chapters to solve problems in probability.