and then set

$$X_i = \text{number of } j, \quad j = 1, \ldots, n : Y_j = i$$

(That is, the generated value of Y_j represents the result of trial j, and X_i is the number of trials that result in outcome i.)

On the other hand, if n is large relative to r, then $X_1, \ldots, X_r$ can be simulated in sequence. That is, first generate X_1, then X_2, then X_3, and so on. Because each of the n trials independently results in outcome 1 with probability p_1, it follows that X_1 is a binomial random variable with parameters (n, p_1). Therefore, we can use the method of Section 4.3 to generate X_1. Suppose its generated value is x_1. Then, given that x_1 of the n trials resulted in outcome 1, it follows that each of the other $n - x_1$ trials independently results in outcome 2 with probability

$$P\{2|\text{not } 1\} = \frac{p_2}{1 - p_1}$$

Therefore, the conditional distribution of X_2, given that $X_1 = x_1$, is binomial with parameters $(n - x_1, \frac{p_2}{1-p_1})$. Thus, we can again make use of Section 4.3 to generate the value of X_2. If the generated value of X_2 is x_2, then we next need to generate the value of X_3 conditional on the results that $X_1 = x_1, X_2 = x_2$. However, given there are x_1 trials that result in outcome 1 and x_2 trials that result in outcome 2, each of the remaining $n - x_1 - x_2$ trials independently results in outcome 3 with probability $\frac{p_3}{1-p_1-p_2}$. Consequently, the conditional distribution of X_3 given that $X_i = x_i, i = 1, 2$, is binomial with parameters $(n - x_1 - x_2, \frac{p_3}{1-p_1-p_2})$. We then use this fact to generate X_3, and continue on until all the values $X_1, \ldots, X_r$ have been generated. □

Exercises

1. Write a program to generate n values from the probability mass function $p_1 = \frac{1}{3}, p_2 = \frac{2}{3}$.

(a) Let $n = 100$, run the program, and determine the proportion of values that are equal to 1.
(b) Repeat (a) with $n = 1000$.
(c) Repeat (a) with $n = 10{,}000$.

2. Write a computer program that, when given a probability mass function $\{p_j, j = 1, \ldots, n\}$ as an input, gives as an output the value of a random variable having this mass function.

3. Give an efficient algorithm to simulate the value of a random variable X such that

$$P\{X = 1\} = 0.3, \quad P\{X = 2\} = 0.2, \quad P\{X = 3\} = 0.35, \quad P\{X = 4\} = 0.15$$

4. A deck of 100 cards—numbered $1, 2, \ldots, 100$—is shuffled and then turned over one card at a time. Say that a "hit" occurs whenever card i is the ith card to be turned over, $i = 1, \ldots, 100$. Write a simulation program to estimate the expectation and variance of the total number of hits. Run the program. Find the exact answers and compare them with your estimates.

5. Another method of generating a random permutation, different from the one presented in Example 4b, is to successively generate a random permutation of the elements $1, 2, \ldots, n$ starting with $n = 1$, then $n = 2$, and so on. (Of course, the random permutation when $n = 1$ is 1.) Once one has a random permutation of the first $n - 1$ elements—call it $P_1, \ldots, P_{n-1}$—the random permutation of the n elements $1, \ldots, n$ is obtained by putting n in the final position—to obtain the permutation $P_1, \ldots, P_{n-1}, n$—and then interchanging the element in position n (namely, n) with the element in a randomly chosen position which is equally likely to be either position 1, position $2, \ldots$, or position n.

(a) Write an algorithm that accomplishes the above.
(b) Prove by mathematical induction on n that the algorithm works, in that the permutation obtained is equally likely to be any of the $n!$ permutations of $1, 2, \ldots, n$.

6. Using an efficient procedure, along with the text's random number sequence, generate a sequence of 25 independent Bernoulli random variables, each having parameter $p = .8$. How many random numbers were needed?

7. A pair of fair dice are to be continually rolled until all the possible outcomes $2, 3, \ldots, 12$ have occurred at least once. Develop a simulation study to estimate the expected number of dice rolls that are needed.

8. Suppose that each item on a list of n items has a value attached to it, and let $v(i)$ denote the value attached to the ith item on the list. Suppose that n is very large, and also that each item may appear at many different places on the list. Explain how random numbers can be used to estimate the sum of the values of the different items on the list (where the value of each item is to be counted once no matter how many times the item appears on the list).

9. Consider the n events $A_1, \ldots, A_n$ where A_i consists of the following n_i outcomes: $A_i = \{a_{i,1}, a_{i,2}, \ldots, a_{i,n_i}\}$. Suppose that for any given outcome a, $P\{a\}$, the probability that the experiment results in outcome a is known. Explain how one can use the results of Exercise 8 to estimate $P\{\bigcup_{i=1}^{n} A_i\}$, the probability that at least one of the events A_i occurs. Note that the events A_i, $i = 1, \ldots, n$, are not assumed to be mutually exclusive.

10. The negative binomial probability mass function with parameters (r, p), where r is a positive integer and $0 < p < 1$, is given by

$$p_j = \frac{(j-1)!}{(j-r)!(r-1)!} p^r (1-p)^{j-r}, \quad j = r, r+1, \ldots$$

(a) Use the relationship between negative binomial and geometric random variables and the results of Example 4d to obtain an algorithm for simulating from this distribution.
(b) Verify the relation

$$p_{j+1} = \frac{j(1-p)}{j+1-r} p_j$$

(c) Use the relation in part (b) to give a second algorithm for generating negative binomial random variables.
(d) Use the interpretation of the negative binomial distribution as the number of trials it takes to amass a total of r successes when each trial independently results in a success with probability p, to obtain still another approach for generating such a random variable.

11. If Z is a standard normal random variable, show that

$$E[|Z|] = \left(\frac{2}{\pi}\right)^{1/2} \approx 0.798$$

12. Give two methods for generating a random variable X such that

$$P\{X = i\} = \frac{e^{-\lambda}\lambda^i/i!}{\sum_{j=0}^{k} e^{-\lambda}\lambda^j/j!}, \qquad i = 0, \ldots, k$$

13. Let X be a binomial random variable with parameters n and p. Suppose that we want to generate a random variable Y whose probability mass function is the same as the conditional mass function of X given that $X \geq k$, for some $k \leq n$. Let $\alpha = P\{X \geq k\}$ and suppose that the value of α has been computed.

(a) Give the inverse transform method for generating Y.
(b) Give a second method for generating Y.
(c) For what values of α, small or large, would the algorithm in (b) be inefficient?

14. Give a method for simulating X, having the probability mass function $p_j, j = 5, 6, \ldots, 14$, where

$$p_j = \begin{cases} 0.11 & \text{when } j \text{ is odd and } 5 \leq j \leq 13 \\ 0.09 & \text{when } j \text{ is even and } 6 \leq j \leq 14 \end{cases}$$

Use the text's random number sequence to generate X.