THE INSTRUCTIONAL USE OF SIMULATION OF RANDOM EXPERIMENTS

DR. JAMES M. MEEHAN
DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
ILLINOIS BENEDICTINE COLLEGE
LISLE, ILLINOIS 60532

INTRODUCTION

In order to illustrate concretely the concepts encountered in a course in probability and statistics it is necessary to perform probabilistic experiments repeatedly. Simple examples are flipping coins and rolling dice. It is common that instructors have their students actually perform such experiments many times to illustrate a particular theoretical result. An obvious drawback of such procedures is that the collection of a sizable amount of data is tedious and time consuming. If fast, accurate methods for obtaining sample data were available, time could be saved, boredom eliminated and the learning process enhanced.

Here, of course, is where the computer is useful. All of the common probability distributions can be simulated by the computer via the random number generator. Furthermore, even today's inexpensive microcomputers are equipped with random number generators which are adequate for this purpose. The generator produces a sequence of independent observations which are uniformly distributed on the unit interval. (Henceforth, the notation U(0,1) will be used for this distribution.) These in turn can be transformed, using results from probability theory, into independent observations from other distributions.

Given the capability of generating sample data from a desired distribution, the instructor can begin to design laboratory experiments which use the data to teach concepts. Elliot Tanis of Hope College in Holland, Michigan has written a laboratory manual [1] consisting of such experiments. The manual is used in a one credit laboratory which students take concurrently with a three credit lecture course. The purpose of this paper is to present examples of methods for simulating distributions, and to suggest techniques useful for designing experiments, like those of Tanis, which can be performed by students on their own computers or in the laboratory. It is intended that the experiments provide not only concrete illustrations but also the motivation to study and understand the underlying theory.

SIMULATING DISTRIBUTIONS

For the most part students are able to develop simple algorithms of their own for simulating the common discrete distributions, since most of these depend on repeated Bernoulli trials. The binomial, geometric, negative binomial and hypergeometric are examples of such distributions. Of the common distributions, the Poisson is an exception; it cannot be simulated (exactly) using Bernoulli trials. A technique for simulating the Poisson distribution is presented in the next section.

The theory behind simulating the continuous distributions is richer and more rewarding than that of the discrete distributions. One of the most useful theorems in simulating continuous distributions is the following:

Theorem 1. If the continuous distribution function $F(x)$ is strictly increasing over $0<F(x)<1$, and if random variable U has the U(0,1) distribution, then the random variable $X=F^{-1}(U)$ has distribution function $F(x)$.

The theorem is presented in several text books and tends to be difficult for students to understand. If the student has the opportunity to make use of the result, however, he/she may be better motivated toward an appreciation of the result. Two examples of how this theorem may be used to simulate distributions follow. In each case random variable X is expressed as a function of the random (number) variable V and X has the desired distribution.

Example 1. The Uniform Distribution, U(a,b).

In this case $F(x) = (x-a)/(b-a)$ for $a \leq x \leq b$. Solving $U = F(X)$ yields

$$X = a + (b-a)U.$$

Example 2. The Exponential Distribution, E(a).

Here $F(x) = 1-e^{-x/a}$ for $x \geq 0$, $a>0$.

Again solving $U = F(X)$ for X results in

$$X = - a \ln(1-U).$$

The simpler statement $X = - a \ln U$ can be substituted since $1-U$ also has the U(0,1) distribution.

In order to use this method the equation $U=F(X)$ must be solved analytically for X in terms of U. Closed form solutions do not exist for some of the common distributions. The Gamma and Normal distributions are examples. A Gamma distribution with integral shape parameter, however, can be simulated as a sum of exponentials. The Normal distribution can be simulated using the Box-Muller transformation, which is discussed in the next section. Of course, from the Normal distribution can be obtained the t, F and Chi-Square distributions.

EXAMPLES OF EXPERIMENTS AND TECHNIQUES

One of the most illuminating techniques for studying the distribution of random data is the histogram. Consider an experiment in which a histogram of data from a Poisson distribution is constructed. The Poisson distribution is simulated by making use of the fact that, in the Poisson process, interarrival times have an exponential distribution. Simulation of the exponential distribution was discussed in the previous section. Exponential values are observed one at a time until their sum exceeds one. Subtracting one from the required number of observations yields an observation from a Poisson distribution with mean equal to the reciprocal of that of the exponential.

Appearing in Figure 1 is a histogram of 1000 Poisson values whose distribution has theoretical mean equal to three. A graph representing the expected number of observations has been superimposed in order to provide a comparison of observed versus expected. The sample mean and variance were computed and displayed for the purpose of comparison with the theoretical values. The same procedures can be applied to any distribution which can be simulated.

As another example consider the task of illustrating the Central Limit Theorem, one of the subjects most important results. As a limit theorem, it is difficult for undergraduate students to understand. Loosely speaking, the theorem guarantees that the sample mean from any reasonably well behaved distribution, when standardized, is approximately standard normal in distribution. The quality of the approximation (or the rapidity of convergence) is usually not discussed in undergraduate texts.

Let random variable W be the sum of n independent U(0,1) random variables. If n=2 the density is triangular in shape. For n=k the density consists of k polynomial arcs, each of degree k-1, pieced together. Even for a value of n as small as n=4, the density of the sum is quite "normal" in appearance. Histograms representing 250,000 values obtained for the n=2 and n=4 cases are given in Figures 2 and 3. The data have been standardized and the standard normal density superimposed. The triangular nature of the density is clearly apparent in the histogram in Figure 2. The histogram in Figure 3 indicates the goodness of the approximation even for a very small sample size.

The importance of the normal probability distribution is evidenced by the Central Limit Theorem. Clearly it is desirable to be able to simulate this distribution. Data from the normal distribution can be used to design a variety of experiments for students of probability and statistics. Problems involving estimation, sampling, confidence intervals, regression and analysis of variance are some examples of topics whose concepts can be illustrated using normal sample data.

As mentioned in the previous section, normal sample data cannot be obtained by the method of inverting the distribution function. There is a method which transforms a pair of independent U(0,1) random variables into a pair of standard normal random variables. The method is due to Box and Muller [2], and the transformation is given by

$$X = \sqrt{-2\ln U}\ \cos 2\pi V$$

$$Y = \sqrt{-2\ln U}\ \sin 2\pi V$$

Here U and V are the U(0,1) variables and X and Y are the independent standard normals.

The problem of showing that the transformation actually works appears in several texts. It is a rather straightforward exercise using the Jacobian to obtain the bivariate density. Some texts may mention that the transformation is useful for simulation, but students often want to know how the authors came up with the equations.

Once it is pointed out that it is merely the polar coordinate transformation students seem satisfied. The random variable $\Theta = 2\pi V$ is uniform on the interval $(0, 2\pi)$. Since $X^2 + Y^2 = R^2$, the random variable R is the square root of a Chi-square random variable with 2 degrees of freedom. The latter distribution is exponential so that both Θ and R are easily simulated.

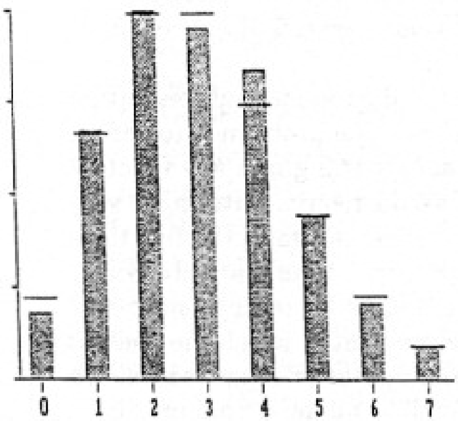

Figure 1. Poisson Histogram

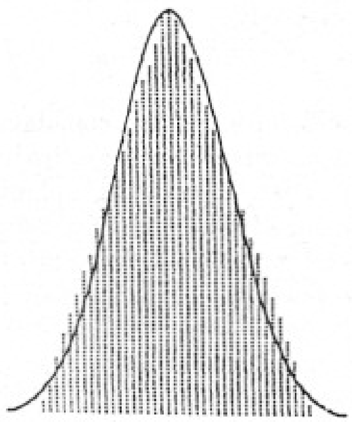

Figure 2. U_1+U_2

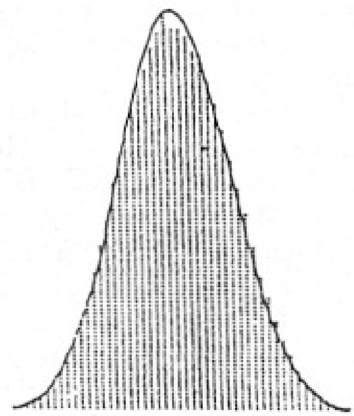

Figure 3. $U_1+U_2+U_3+U_4$

CONCLUDING REMARKS

The previous example, although not actually a simulation experiment itself, reveals the primary purpose of the experiments. The need to find a transformation which yields data from a particular distribution is realized. A transformation which does the job is presented in class along with its proof. It is likely that the student will have a greater appreciation for a mathematical result and the underlying theory if he/she has direct experience involving its application. In general, carefully designed experiments can lead to desirable educational outcomes. Thoughtful instructors may find the time investment, both on their part and on the part of their students, yields valuable dividends in motivation and learning.

REFERENCES

[1] Tanis, E.A. (1978), Laboratory Manual for Probability and Statistical Inference, CONDUIT, Iowa City.

[2] Box, G.E.P., and N.E. Muller (1958), A Note or Generation of Normal Deviates, AMS 28, 610-611.