Interplay Between Simulation and Theory

Elliot A Tanis - Holland, Michigan, USA

1. Introduction

Students in a beginning course in probability and statistics often have limited background and experience for knowing how to attack the solution of a problem theoretically. At times, insight into the solution of a problem is provided by writing a computer program that will simulate a solution. At other times, confidence in a solution can be bolstered by the simulation of a solution.

These ideas can be illustrated with a variety of examples. A reason for selecting the four problems given in this paper is that they all have answers that are functions of e and/or π, with one of the answers being e^{π}. The answers and/or solutions to the questions that are posed are perhaps well-known to you. However, try to put yourself in the position of a student who is hearing these problems for the first time.

A computer program written in BASIC to simulate each solution is given and uses a built-in computer "random number generator". In addition, a theoretical solution is outlined for each problem.

2. The four problems

2.1 *Problem 1*

Let the random variable U *have a uniform distribution on the interval* (0,1). *For a sequence of independent observations of* $U : u_1, u_2, u_3, \ldots$, *let X be defined by* $X = \min\{k : u_1 + u_2 + \ldots + u_k > 1\}$. *That is,* X *is the minimum number of random numbers that must be added together so that their sum exceeds 1. Show that* $E(X) = e$.

An outline of a theoretical solution is as follows:

(i) Show that $P(U_1 + U_2 + \ldots + U_k \leq 1) = 1/k!$.

(ii) Show that the distribution function of X, defined at a positive integer x, is $F(x) = P(X \leq x) = 1 - 1/x!$.

(iii) Show that the p.d.f. of X is defined by $f(x) = (x-1)/x!$, $x = 2,3,\dots$.

(iv) Thus

$$E(X) = \sum_{x=2}^{\infty} xf(x) = \sum_{x=2}^{\infty} 1/(x-2)! = e.$$

(v) The following computer program can be used to simulate the value of E(X):

```
100 CLS: RANDOMIZE TIMER: KEY OFF
110 N = 500: REM N is the number of repetitions
120 FOR K = 1 TO N
130     X = 0: SUM = 0
140     SUM = SUM + RND
150     X = X + 1
160     IF SUM < 1 THEN 140
170     SUMOFX = SUMOFX + X
180 NEXT K
190 XBAR = SUMOFX/N
200 PRINT "The sample mean of the x's is"; XBAR
210 END
```

Of course, if you have graphing routines, it is interesting to graph a histogram of the observations of X with the p.d.f. of X superimposed. This problem appeared on The William Lowell Putnam examination in 1958 and is stated in Bush (1961). Its solution is also discussed in Schultz (1979), Schultz and Leonard (1989), and Tanis (1987). It appears as an exercise in Hogg and Tanis (1987).

A question that could be raised is as follows: Suppose that X is the minimum number of random numbers that must be added together so that their sum exceeds 2 (or 3 or 4). What is the value of E(X)?

2.2 *Problem 2*

Generate a sequence of random numbers from (0,1) : $u_1, u_2, u_3, \dots$ *If* $u_1 \le u_2 \le \dots \le u_k$ *and* $u_k > u_{k+1}$, *let the random variable* $X = k + 1$. *Show that* $E(X) = e$.

Given a sequence of random numbers from the interval (0,1) : $u_1, u_2, \dots$, let $V = \max\{k : u_1 \le u_2 \le \dots \le u_k\}$. Then $X = V + 1$. To find the p.d.f. of X:

(i) Show that the distribution function of V, defined at a positive integer k, is

$$G(k) = P(V \le k) = 1 - P(U_1 \le U_2 \le \dots \le U_{k+1}) = 1 - 1/(k+1)!\ .$$

(ii) The p.d.f. of V, for $v = 1,2,3,\dots$, is

$$g(v) = P(V = v) = P(V \le v) - P(V \le v - 1) = v/(v+1)!\ .$$

(iii) Since $X = V + 1$, the p.d.f. of X, for $x = 2,3,4,\dots$, is

$$f(x) = g(x-1) = (x-1)/x!, \quad x = 2,3,4,\ldots$$

(iv) Thus

$$E(X) = \sum_{x=2}^{\infty} x f(x) = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \ldots = e.$$

(v) The following computer program can be used to simulate the value of E(X):

```
100 CLS: RANDOMIZE TIMER: KEY OFF
110 N = 500: REM N is the number of repetitions
120 FOR K = 1 TO N
130     U = RND: X = 1
140     V = RND
150     IF U < V THEN U = V: X = X + 1: GOTO 140
160     X = X + 1
170     SUMOFX = SUMOFX + X
180 NEXT K
190 XBAR = SUMOFX/N
200 PRINT "The sample mean of the x's is"; XBAR
210 END
```

If possible, plot a graph of the histogram of the observations of X with the p.d.f. of X superimposed. This problem is discussed briefly in Schultz and Leonard (1989).

2.3 *Problem 3*

An urn contains n *balls that are numbered from* 1 *to* n. *Take a random sample of size* n *from the urn, one at a time. A match occurs if ball numbered* k *is selected on draw* k. *Let* A *be the event that at least one match is observed. Show that*

$$\lim_{n\to\infty} P(A) = \lim_{n\to\infty}\left[1 - (1 - \tfrac{1}{n})^n\right] = 1 - \frac{1}{e}$$

when sampling with replacement,

$$\lim_{n\to\infty} P(A) = \lim_{n\to\infty}\left[1 - (1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots + \frac{(-1)^n}{n!}\right] = 1 - \frac{1}{e}$$

when sampling without replacement.

(i) Proof when sampling with replacement.
Let A_k be the event that ball k is selected on draw numbered k, k = 1,2,...,n and A'_k its complement. Then $P(A_k) = 1/n$, the n events A_k are independent, and so

$$P(A) = 1 - P(A'_1 \cap A'_2 \cap \ldots \cap A'_n) = 1 - (\frac{n-1}{n})^n \to 1 - \frac{1}{e}.$$

(ii) Proof when sampling without replacement.

Let A_k be the event that ball k is selected on draw k. The proof will be given when n = 4. It should be clear how to extend this proof for a general value of n. We have

$$
\begin{aligned}
P(A) &= P(A_1 \cup A_2 \cup A_3 \cup A_4) \\
&= \sum_{k=1}^{4} P(A_k) - \sum_{j<k} P(A_j \cap A_k) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - P(A_1 \cap A_2 \cap A_3 \cap A_4) \\
&= 4\left(\frac{3!}{4!}\right) - {}^4C_2\left(\frac{2!}{4!}\right) + {}^4C_3\left(\frac{3!}{4!}\right) - \frac{1}{4!} \\
&= 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} = 1 - \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!}\right).
\end{aligned}
$$

Comparing the two limiting probabilities, we note that this example provides an interesting problem for helping students recall two ways to obtain 1/e.

(iii) The following program simulates this problem when sampling with replacement. The PRINT statements are included in lines 180 and 200 so that the student gains a better understanding of this simulation.

```
100 CLS: RANDOMIZE TIMER: KEY OFF
110 M = 500: REM M is the number of repetitions
120 N = 20: REM N is the number of balls in the urn
130 FOR J = 1 TO M
140     MATCH = 0
150     FOR K = 1 TO N
160         BALL = INT(N*RND) + 1
170         IF BALL = K THEN MATCH = 1
180         PRINT USING "###";BALL;
190     NEXT K
200     PRINT USING "####";MATCH
210     NUMMATCH = NUMMATCH + MATCH
220 NEXT J
230 PRINT: PRINT "At least one match was observed on ";NUMMATCH;
"trials".
240 PRINT: PRINT "The proportion of trials on which at least one match
occurred was ";NUMMATCH/M
250 END
```

(iv) When sampling without replacement, add the following lines to the above:

```
105 DIM FLAG(25)
132 FOR L = 1 TO N
134     FLG(L) = 0
```

```
136 NEXT L
165   IF FLAG(BALL) = 1 THEN 160
175   FLAG(BALL) = 1
```

(v) It is interesting to note that, if n is "sufficiently large", sampling with or without replacement yields the same answer, approximately. In both cases the values of P(A) change very slowly with n, but the rates of convergence to 1 - 1/e ≈ 0.632120558 are different. Also, although the program for sampling without replacement is slow when n is large, the convergence of P(A) to 1 - 1/e is very fast so the simulation can be done using small values of n.

Variations of this problem are discussed in Hogg and Tanis (1987), Schultz (1979), and Tanis (1987).

2.4 *Problem 4*

Let $B_n = \{(x_1, x_2, \ldots, x_n) : x_1^2 + x_2^2 + \ldots + x_n^2 \le 1\}$ *denote a ball of radius* 1 *in n-space. Let* V_n *equal the "volume" of* B_n. *Show that* $V_n = \pi^{n/2}/\Gamma(n/2 + 1)$, $n = 0,1,2,\ldots$, *and deduce that*

$$\sum_{k=0}^{\infty} V_{2k} = e^{\pi}.$$

(i) To estimate the value of V_n using simulation, generate m n-tuplets of random numbers, $(u_1, u_2, \ldots, u_n)$, where $0 < u_i < 1$, $i = 1,2,\ldots,n$. That is, each u_i is selected randomly from the interval (0,1). Let

$$Y = \#(\{(u_1, u_2, \ldots, u_n) : u_1^2 + u_2^2 + \ldots + u_n^2 \le 1\}).$$

Then $V_n \approx 2^n\, Y/m$ provided that m is "sufficiently large".

(ii) It is well known that $V_1 = 2$, $V_2 = \pi$, and $V_3 = 4\pi/3$. But the values of V_4, V_5,... are not so well known. In the simulation, we see that 2^n increases as n increases but, for a fixed m, Y will tend to be smaller for larger values of n. Hence it is *not* obvious what happens to the values of V_n as n increases. It can be shown that the maximum volume is $V_5 = 8\pi^2/15$.
An interesting problem for a statistics class is to find the sample size required to show, using simulation and with 95% confidence, that the maximum does occur in 5-space. Because $V_6 = \pi^3/6$, the difference, $V_5 - V_6 = 0.096076$, is not very large.

(iii) An outline for a theoretical solution is now given. This proof uses material from a first year course in calculus. It is based on a solution that was written by my colleague, Dr Herbert Dershem, and myself in 1974 for the use of our students. Other solutions are given by Salgia (1983) and Smith (1989).
For notation, for a positive integer n and a positive real number r, let

$$B_n(r) = \{(x_1, x_2, \ldots, x_n) : x_1^2 + x_2^2 + \ldots + x_n^2 \le r\}.$$

Let $V_n(r)$ equal the "volume" of $B_n(r)$.

Lemma 1: $V_n(r) = r^n V_n(1)$.

Lemma 2: $V_n(1) = a_n V_{n-1}(1)$ where

$$a_n = \pi \int_0^1 (\cos \frac{\pi t}{2})^n \, dt.$$

Lemma 3: $a_n = (n-1)/n \; a_{n-2}$.

The next two lemmas use properties of the gamma function.

Lemma 4: When n = 2k, an even integer,

$$a_n = a_{2k} = \Gamma(\frac{2k-1}{2} + 1)\Gamma(\frac{1}{2}) / \Gamma(\frac{2k}{2} + 1).$$

Lemma 5: When n = 2k + 1, an odd integer,

$$a_n = a_{2k+1} = \Gamma(\frac{2k}{2} + 1)\Gamma(\frac{1}{2}) / \Gamma(\frac{2k+1}{2} + 1).$$

Theorem: The "volume" of a "ball" of radius r in n-dimensional space is given by $V_n(r) = r^n \pi^{n/2} / \Gamma(n/2 + 1)$, n = 0,1,2,3,... .

(iv) From the theorem, we see that a formula for the volume of a ball of radius 1 in even dimensional space, n = 2k, is $V_{2k}(1) = \pi^k/k!$. Thus the sum of the volumes of all balls of radius 1 in even dimensional space is given by

$$\sum_{k=0}^{\infty} V_{2k}(1) = \sum_{k=0}^{\infty} \frac{\pi^k}{k!} = e^{\pi}.$$

(v) The following computer program can be used to simulate values of $V_n(1)$:

```
100 CLS: RANDOMIZE TIMER: KEY OFF
110 INPUT "Input the value of n, the dimension of the space:  n = ";N
120 INPUT "Input the number of n-tuples that you would like to
choose:  m = ";M
130 NUM = 0
140 FOR K = 1 TO M
150     SUM = 0
160     FOR J = 1 TO N
170         SUM = SUM + RND^2
```

```
180     NEXT J
190     IF SUM <= 1 THEN NUM = NUM + 1
200 NEXT K
210 PRINT "The number of successes is ";NUM
220 PRINT "The point estimate of the volume is ";2^N*NUM/M
230 END
```

3. Conclusions

Hopefully the examples in this paper give you some ideas about using the computer as a laboratory tool for helping students understand and appreciate theoretical results. Perhaps they suggest additional examples or modifications. The author is always looking for new and better examples and would appreciate receiving these from you.

The programs that are listed in this paper, as well as some others, are available on a computer disk for an IBM (compatible) computer. Some of the programs on the disk include graphical output. If you would like to receive a copy of this disk, please send $10 to the author to cover the cost. The user can easily modify the programs.

A copy of this paper with some of the proofs is also available from the author.

References

Bush, L E (1961) The William Lowell Putnam Mathematics Competition. *American Mathematical Monthly* **68**, 18-33.

Hogg, Robert V and Tanis, Elliot A (1987) *Probability and Statistical Inference* (3rd ed). Macmillan Publishing Co Inc, New York.

Salgia, Ravi (1983) Volume of an n-dimensional unit sphere. *Pi Mu Epsilon Journal* **7(8)**, 496-501.

Schultz, H S (1979) An expected value problem. *Two-Year College Mathematics Journal* **10**, 277-278.

Schultz, Harris S and Leonard, Bill (1989) Unexpected occurrences of the number e. *Mathematics Magazine* **62(4)**, 269-271.

Smith, David J and Vamanamurthy, Mavina K (1989) How small is a unit ball? *Mathematics Magazine* **62(2)**, 101-107.

Tanis, Elliot A (1987) The answer is 1 - 1/e. What is the question? *Pi Mu Epsilon Journal* **8(6)**, 387-389.

Tanis, Elliot A (1987) Computer simulations to motivate and/or confirm theoretical concepts. *American Statistical Association 1987 Proceedings of the Section on Statistical Education*, 27-32.