

Learning Binomial Probability Concepts

With Simulation, Random Numbers

and a Spreadsheet

Dr John A. Rochowicz, Jr

Mathematics Department

Alvernia College

## Abstract

This paper introduces the reader to the concepts of binomial probability and simulation. A spreadsheet is used to illustrate these concepts. Random number generators are great technological tools for demonstrating the concepts of probability. Ideas of approximation, estimation, and mathematical usefulness provide numerous ways of learning mathematics.

The advantages of simulations as well as the disadvantages are discussed. Various examples are illustrated.

Mathematical ideas are enhanced and concepts are motivated through technology use. Comparisons are made between actual results and approximate results.

## Introduction

In everyday applications of mathematics, usually the student or the teacher searches for a model or formula to analyze and solve certain situations. Mathematical fields such as statistics and calculus make use of finding a model that best fits a specific problem. But many times an exact mathematical model cannot be found or is too difficult to use and so simulation is a great alternative. Through the process of simulation approximations to results are achieved.

Simulation presents a way for learning probability concepts. Formulas are not necessary and mathematical intuition is developed about the meaning of concepts and making sense of calculations. The purpose of this paper is to demonstrate a) how a spreadsheet can be used to approximate binomial probabilities; b) how the learner can make connections between formulas and approximations; and c) the value of technology in developing the ability to understand mathematics. A computer or calculator with a random number generator can simulate many probability experiments. The results are so close to the theoretical that they are appropriate for analyzing and predicting future behaviors or actions.

Numerous examples are presented on the spreadsheet EXCEL. Comparisons are made between the simulated values, the actual calculations and the built-in formulas of EXCEL. Looking at mathematics concepts in different ways allows the learner to construct meaningful, individual learning.

### The Binomial Probability Distribution

A probability distribution or mathematical function, for a particular discrete or continuous random variable  $x$  is a table, graph, or formula that displays the values of  $x$

and their respective probabilities. A discrete random variable represents count data such as, finding the number of heads that occurs on the toss of 100 coins. A continuous variable is one that involves measured data. An example of continuous data is the time it takes to complete the toss of 100 coins.

The binomial probability distribution is a discrete distribution that involves  $n$  identical and independent trials in which 2 outcomes such as True/False, Boy/Girl, Success/Failure are possible. In introductory statistics texts the binomial probability distribution is defined as  $P(x) = C(n,k) p^k q^{n-k}$  for values  $k = 0, 1, 2, \dots, n$ .  $C(n,k)$  is the calculation of the combinations of  $n$  things taking  $k$  at a time or  $n! / k!(n-k)!$  (Mendenhall 2004).

#### Example 1

Consider the following example. A researcher believes that 60% of all college students will get a position in their chosen field of study after graduation. Suppose you randomly select 4 alumni from various colleges. Consider the following: a) Find the probability that 3 of the 4 students will get a job in their field of study. Using the binomial distribution formula,  $p(3)$  is calculated. That is,  $p(3) = C(4,3) 0.6^3 0.4^1$  the result is 0.3456. b) Find the probability that at least 3 of the 4 students get a job in their field. Calculating  $p(3) + p(4)$  gives the result 0.4752 This example represents data that is discrete because the number of students that get a job is the variable of interest. This example is also a binomial distribution because there are 4 trials which are identical and independent and the student gets or does not get a job in his chosen field.

#### The Spreadsheet

A spreadsheet is a great technological tool for doing mathematics. Very little interference from learning the software occurs with the learning of the mathematics. Spreadsheets can be used as electronic worksheets for doing mathematics. From charting to calculating to simulating to programming, spreadsheets have a variety of capabilities. The implementation of a spreadsheet for doing mathematics is dynamic. Numerical values can be placed anywhere on the worksheet using correct formulas. Changing one cell on the sheet changes the entire worksheet automatically. The value of using a spreadsheet to the student is the ability of displaying output for different input with little effort on the part of the learner.

### Random Numbers

With technology such as spreadsheets, simulations are developed with a random number generator. The commands for generating random numbers on a spreadsheet are easy to learn. To generate a random number between 0 and 1 use “@RAND” in LOTUS and “=RAND()” in EXCEL. To generate a random number on an interval (a, b) use “b\*@RAND\*+a” in LOTUS and use “=b\*RAND()+a” in EXCEL. If an integer on an interval (a, b) is desired use “@INT(b\*@RAND+a)” in Lotus and “=INT(b\*RAND()+a)” in EXCEL. For example to generate in integer between 1 and 10 in EXCEL use “=INT(10\*RAND()+1)”. Many technologies including spreadsheets, programming languages and programmable/graphing calculators have the capabilities to generate random numbers.

### Simulating Binomial Probabilities

The following steps provide a method of calculating binomial probabilities in EXCEL: a) Generate a column of integers to count the number of times the simulation is

to be done. Example 1 was performed 500 times. Suppose you start in cell A6, enter a 1 there. In cell A7 enter “=A6+1” and get 2. Copy and paste the contents of cell A7 down for 500 values. You can start anywhere on the spreadsheet.

b) In column B and column C generate the values for the first trial,  $n = 1$ . Enter a random number on the desired interval. In example 1 generate an integer between 1 and 10.

Suppose you use cell B6 and wish to generate a random integer between 1 and 10. The command “=INT(10\*RAND()+1)” generates an integer between 1 and 10. Copy and paste contents of cell B6 for 500 values. In cell C6 count successes. A success is any number generated less than or equal to the given probability  $p$ . In example 1,  $p = .6$ . In cell C6 enter the formula “=IF(B6<=6,1, 0)”. This formula outputs a 1 for a success (any number less than or equal to 6) and a 0 if not. Copy and Paste what is in cell C6 down the spreadsheet for 500 values.

c) If the number of trials,  $n$  is more than 1, repeat what is calculated to get column B and C in the next columns. If  $n = 4$  as in example 1, what is performed in Column B and C must be calculated 3 more times through columns H and I.

d) Next add every success value from every column across. I placed this sum in column J. If the sum is the desired value count it, if not do not. In example 1, the probability of 3 in a group of 4 is desired. So if the sums in column J are equal to 3 place 1 if not place a 0 in column K. In EXCEL enter the command in column K “=IF(J6 =3,1,0)”. Copy and paste what is in cell K6 down 500 times.

e) Add all these counted values and divide by the number of times the simulation is done. The result is the estimated probability. In Example 1, the number of success found for 500 experiments was 168. The estimated probability calculated was  $168/500$  or 0.336. In

EXCEL for “=BINOMIALDIST( 3,4,0.4, FALSE)”, 3 is the value of interest, n = 4, p is 0.4 and False means calculation is not cumulative. In applying this EXCEL formula the result is 0.3456. Similarly the probability of at least 3 in a group of 4 is calculated and simulated. The result by using the BINOMIALDIST command is 0.4752 and by using simulation is  $227/500 = 0.454$ . The results from each simulation will vary, but are close to the actual value found using the BINOMIALDIST command in EXCEL.

The process illustrated in Table 1 demonstrates how binomial probabilities can be simulated for Example 1. Pressing F9 will automatically conduct each simulation.

Table 1

A	B	C	D	E	F.....	J
n =	4	p =	6	out of	10	
		p =	.6			
	n = 1					
	Random	Success		Sums	p(3)	p(x.>=3)
	Number			of Successes		
1	6	1		2	0	
2	9	0		1	0	
500				Total		
				Successes	168	227
				estimated		
				168/500 =	0.336	227/500 = 0.454
				actual	0.3456	0.4752

Note: The EXCEL command “= INT(10\*RAND()+1)” generates numbers 1 to 10 when p = .1 to .9. Use the command “=INT(100\*RAND()+1)” to generate a number between 1 and 100 when p = .01 to .99. In this way the learner can simulate any probability in tenths, hundredths, thousandths, and so on.

Table 2: Sample Output for Example 1

Row	Label	Value	Value	Value	Value	Value
507						
508	Number of successes	180	446	234	483	500
509	Simulated	0.36	0.892	0.468	0.966	1
510						
511						
512	Actual	0.3456	0.8704	0.4752	0.9744	1
513						
514						
515						
516						
517						
518						
519						
520						
521						
522						
523						
524						
525						

Note: The following probabilities are shown:  $P(x>0)$ ,  $P(x=3)$ ,  $P(x\leq 3)$ ,  $P(x=2)$ , and  $P(x\geq 3)$ .

### Example 2

A researcher knows that a certain pain reliever is ineffective for 35% of those patients who take it. In a random sample of 6 patients find the probability that the pain reliever is ineffective in a) exactly one patient; b) all 6 patients; c) at most 3 patients. If the student needs to simulate an experiment where  $n = 6$  do the same as above but



generate values for columns B, C, D, E, F, G, H, I, J, K, L, and M. Then count the successes and add them and divide by the number of experiments or simulations. The actual results by using the binomial formula are: a) 0.2437; b) 0.0018; and c) 0.8826. By the simulation techniques demonstrated in Example 1 the results are: a) 0.26; b) 0.002; and c) 0.8700.

### Teaching/Learning Using Simulations

Applying simulations to learning probability enables the student to understand the ideas behind the meanings to uncertainty. The student can also analyze practical and realistic cases without special formulas. Realizing that the use of random number generators can produce accurate results is important. Comparing approximate values with calculated values also allows the learner to see concepts in different ways. Through the use of simulation less stress can be placed on calculations. A mix of both simulation and calculation would be ideal for the learner. In this way, understanding and making connections between mathematics and technology becomes possible.

Also, in teaching the concepts of probability, the ideas of experiment, sample space, and event become better understood. Through this understanding, the learner can do actual experiments and obtain accurate results. For example 1, the experiment is the student getting a job, the sample space is the set of all possible alumni, and the event is the success in getting a job in chosen field.

Consider tossing 1 or many coins. The probability of getting a heads is the number of items in the set of outcomes divided by the total number of items, that is  $1/2$  or 50% is the probability of getting a heads on the toss of one coin. But what does this mean to the learner? There is no clear illustration other than to actually do an activity where

students get a sense of the meaning. These ideas become even more challenging when one tosses 10 coins. What is the probability of obtaining all heads on the toss of 10 coins? How can we approach this in order for the learner to understand probability? One way is to have the student toss a coin or many coins many times about a 1000 or 100,000 times and keep a track of the number of heads that appear. Clearly this is a difficult task to perform. With the use of technology, computers and calculators can be programmed to simulate the toss of 1 coin or many coins and do a particular experiment many times. Technology can not only do the simulation but can also be used to track the number of times the experiment is done.

### Conclusions

What does the student learn from simulation? Things are not always the way they appear. With confidence, probability results that cannot be modeled with a formula become trusted as reliable. In mathematics a formula is not always necessary. In realistic situations, simulations are usually preferred and actually more beneficial. With technology use, estimation and approximation of results become meaningful. Computing a number of examples by formulas and calculator and comparing these examples to simulations performed with a computer show the value and closeness of the simulated values. Hand-held calculations do not become obsolete because the learner must check the correctness of results by doing hand-held calculations.

Using spreadsheets in doing mathematics is invaluable. Learning concepts occurs in different and various ways. Learning is interactive. Also simulations on a spreadsheet enhance learning because the student witnesses a dynamic approach to finding

probabilities. Changing one cell on a spreadsheet changes the entire spreadsheet automatically.

Other probability distributions, discrete as well as continuous can be simulated. With the copy/paste capabilities of windows software, ideas can be generalized from a small number of cases to a large number of cases.

## References

Mendenhall, William, Beaver, Robert J. & Beaver, Barbara M.(2003). *Introduction of Probability and Statistics* (11 edition). Pacific Grove, CA: Thomson Learning.