■ RICK HESSE, Feature Editor, Graziadia Graduate School of Business, Pepperdine University

Triangle Distribution: Mathematica Link for Excel

Rick Hesse, Feature Editor

I have always tried to present templates in Excel that can be used by anyone, without the benefit of macros, add-ins, or extra programs. One of the interesting things that can be done with Excel is to make use of the uniform random number function [**RAND()**] in Excel to simulate various discrete or continuous outcomes, without the use of add-ins such as @RISK or Crystal Ball, both of which are great programs. By the use of lookup tables, general discrete distributions or specific ones such as the Poisson are easy to simulate, so that when <**F9**> is pressed, the simulation repeats itself. For continuous distributions, the normal and exponential serve quite well to simulate outcomes. When not much is known about the distribution of the outcome, say, only the smallest and largest, it is possible to use the uniform distribution (the LaPlace conditions). But if the user also knows the most likely outcome, then the outcome might be simulated best by a triangle distribution. I was first shown this possibility at a conference at Dartmouth, the first conference on spreadsheets jointly sponsored by DSI and INFORMS. Sam Savage was one of the speakers, and his book *INSIGHT.xla* has some very nice add-ins and exercises (INSIGHT.xla, 1998). But in keeping with my theme of trying to make all templates work without add-ins, I decided to try to recreate the triangle distribution.

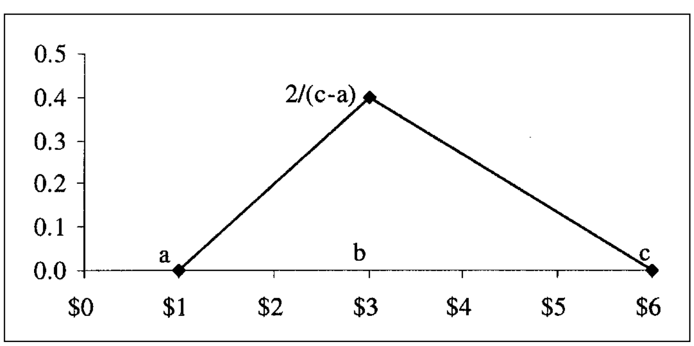

Figure 1: Triangle distribution.

The Algebra

This was not as easy as I first believed, because I thought that the height of the triangle distribution would just be 50%. However, this did not reproduce the results from the triangle distribution on Sam's templates nor was it available in the DAT tools in Excel. It became obvious that for this triangle to be a probability distribution, the area of the triangle must equal 1.000. Therefore, it was an easy matter of determining the height as 2/(c – a), shown in Figure 1.

Determining the mean and variance was not as easy, and I had to actually resort to algebra and integration for these definitions. To make a long story short, it was a fairly simple matter to show that the integration of the triangle area was 1.000 and that the mean of the distribution was (a+b+c)/3, but the variance was another matter. The algebra got so messy that I thought there would be no way to determine the variance in mathematical closed form.

Mathematica Link for Excel

It was at this point that Andrew Ruppel emailed me the news that Wolfram Research, developers of Mathematica, wanted Andrew to review Mathematica Link for Excel, and he thought that I might be in a better position to do so. I jumped at the opportunity because I could at least use Mathematica to confirm what I had already developed and maybe untie the Gordian knot of the variance. I was also interested in how easy it might

Rick Hesse
is professor of quantitative methods at Pepperdine University in the Graziadia Graduate School of Business. He received his BS, MS, and DSc at Washington University School of Engineering in applied math and computer science. Dr. Hesse is the author of Managerial Spreadsheet Modeling & Analysis *and* Applied Management Science: A Quick & Dirty Approach *(with Gene Woolsey), articles in numerous journals, and software for personal computers. Rick was the first professor to be awarded the Outstanding Civilian Service Medal by the Department of the Army at West Point in 1982, and was the winner of the Decision Sciences Institute's Innovative Instructional Award in 1981.*

Dr. Rick Hesse
Graziadia Graduate School of Business
Pepperdine University
Malibu, CA 90265
email: rickhesse@aol.com

be to use Excel as the platform or interface for this industrial-strength mathematical program. This is a program that not only can do definite integrals and differentiation, but also symbolic, algebraic integration and differentiation. (It also can do a multitude of other things, such as functions and distributions long since forgotten in the days of math and engineering schools, and then some I've never even heard of.)

What made all of this integration especially difficult was that it had to be done piece-wise, because there is one linear function from ***a*** to ***b*** with a positive slope [2(x – a)/((c – a)(b – a))], and then the other linear function from ***b*** to ***c*** with negative slope [2(c – x)/((c – a)(c – b))]. The 10-pound package of Mathematica soon arrived, its contents bulging with a 1,500-page manual, another 500-page manual for other add-ins, and a 100-page manual for the Excel link. I soon had Mathematica installed and was able to verify the first two integrations and also find the variance, shown in Figure 2.

Mathematica confirmed that the area under the curve was indeed 1.000 and that the mean of the triangle distribution is simply the mean of the parameters. Figure 2 shows the indefinite integral from a to b and its result, then the indefinite integral from b to c and its result. Finally, the command simplify does the algebraic simplification of the two parts minus the mean squared, which results in the two formulas shown at the bottom. I prefer the latter because it makes more intuitive sense.

This program is truly amazing, and my heart beat faster as I waited to see if this could be done in Excel with the Link. This is when things, admittedly, got stickier. It took some working with the examples to get the integration done, and then an email to Wolfram Research (with a 48-hour response) to find out how to add the two algebraic integrations together. This confirmed that only when the value ***b*** is the midpoint are the mode, mean, and median the same. The Excel spreadsheet is shown in Figure 3 and also has the commands necessary to do the integration, addition, and simplification. The functions in C12:C13 are simply labels, with ', ", or ^ for formatting.

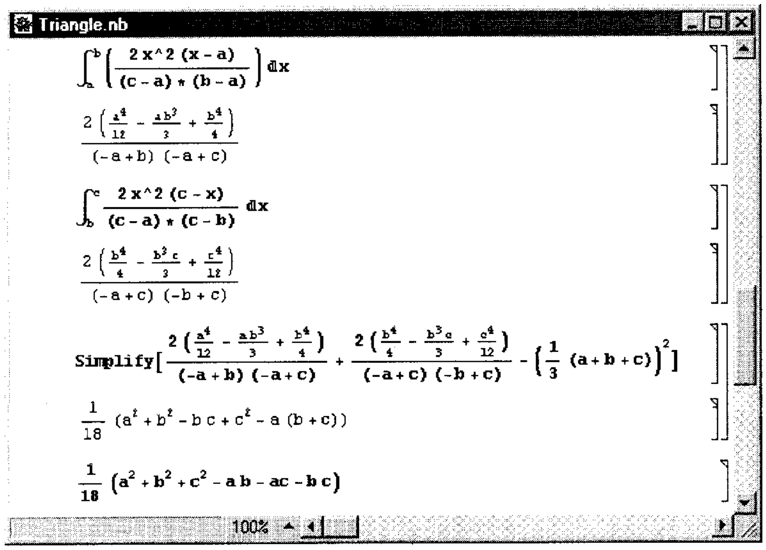

Figure 2. Mathematica integration for variance of triangle distribution.

Excel Simulation

Now I was ready to make up the Excel simulation spreadsheet to illustrate the triangle distribution. Consider an outcome, such as the shipping cost for an item ordered over the Internet, as shown in Figure 4. It may be that there is a $3.95 charge for shipping, and we want to simulate what the actual costs might be if we know that the lowest cost would be $1.00, the highest $6.00 and the most likely $3.00. B10:C10 uses the data in row 5 to determine the expected mean and standard deviation. The Runs Mean and Stdev in row 11 are computed from the single run of 50 random shipments in B16:B65 using the uniform random number in column A. The One-Way Data Table in C15:E65 tricks Excel into repeating (replicating) the runs 50 times and row 12 contains the Mean and Stdev of 2500 simulated shipping costs.

The most difficult formula is the one that determines the outcome of the triangle distribution in B16. Essentially the formula determines if the random number is below or above the mode (B7) and then does the inverse of the area in the proper triangle.

If you want a copy of this spreadsheet, just email me at rickhesse@aol.com and I'll be glad to email it to you.

	A	B	C	D	E	F	G	H	I	J	K
12	**Functions**	xF1:	2*x*(x-a)/((c-a)*(b-a))			a=>b		C14:	=Math("Integrate",C12,C2:E2)		
13		xF2:	2*x*(c-x)/((c-a)*(c-b))			b=>c		C15:	=Math("Integrate",C13,C3:E3)		
14	**Integration**	Int(xF1):	(2*(a^3/6 - (a*b^2)/2 + b^3/3))/((-a + b)*(-a + c)					C16:	=Math("Plus",C14,C15)		
15	**Mean**	Int(xF2):	(2*(b^3/3 - (b^2*c)/2 + c^3/6))/((-a + c)*(-b + c)					C17:	=Math("Simplify",C16)		
16		Add	(2*(a^3/6 - (a*b^2)/2 + b^3/3))/((-a + b)*(-a + c)) + (2*(b^3/3 - (b^2*c)/2 + c^3/6))/((-a + c)*(-b + c))								
17		Simplify	(a + b + c)/3								

Figure 3. Excel worksheet results.

Documentation:
D5: = A5
B6: = 2/(C5-A5)
B7: = (B5-A5)/(C5-A5)
B10: = **AVERAGE**(A5:C5)
C10: = **SQRT**((**SUMSQ**(A5:C5) - **SUMPRODUCT**(A5:C5,B5:D5))/18)
B11: = **AVERAGE**(B16:B65)
C11: = **STDEVP**(B16:B65)
B12: = **AVERAGE**(D16:D65)
C12: = **AVERAGE**(E16:E65)
D15: = B11
E15: = C11
A16: **RAND**()
B16: = **IF**(A16<B7,A5+**SQRT**((B5 A5)*(C5-A5)*A16), C5-**SQRT**((C5-A5)* (C5-B5)* (1-A16)))

Conclusion

This may have seemed like a convoluted way to make up a "simple" spreadsheet, but I did want to develop a "poor person's" way to simulate the triangle distribution in Excel. I also wanted the opportunity to try Mathematica and the Excel Link. I was never successful in being able to verify the variance in Excel because I could find no way to add two cells and subtract a third—and I tried many combinations. I'm sure it can be done, I just didn't have the time. All of that said, both Mathematica and Excel Link for Mathematica are terrific products, and can do an incredible amount of mathematics and engineering math. However, you will really need some special uses to justify the expense and learning curve. There is no way I could show all the functionality of these two programs, but if you have some high-powered mathematical needs, it certainly deserves your attention.

Reference

INSIGHT.xla: Business Analysis Software for Microsoft Excel, Savage, Duxbury, 1998. ■

	A	B	C	D	E
1	Triangular Distribution				
2	Shipping Costs				
3	low	likely	high		
4	a	b	c	a	
5	$1.00	$3.00	$6.00	$1.00	
6	0.00	0.40	0.00	Height	
7	0%	40%	100%	Cum Prob	
8					
9		Mean	Stdev		
10	**Expected**	$3.33	$1.0274		
11	**Runs**	$3.12	$1.0352		
12	**Reps**	$3.30	$1.0186		
13					
14			DATA TABLE		
15	RN	X		$3.12	$1.0352
16	40.1%	$3.00	**1**	$3.26	$1.0633
17	52.7%	$3.34	**2**	$3.29	$0.9694
18	26.0%	$2.61	**3**	$3.35	$1.1635
19	19.3%	$2.39	**4**	$3.35	$1.0325
20	9.4%	$1.97	**5**	$3.31	$0.9616
21	43.0%	$3.08	**6**	$3.33	$1.2006
60	87.1%	$4.61	**45**	$2.91	$0.9206
61	66.9%	$3.77	**46**	$3.31	$1.0011
62	24.0%	$2.55	**47**	$3.37	$0.9853
63	61.6%	$3.60	**48**	$3.40	$1.1322
64	31.4%	$2.77	**49**	$3.37	$1.1073
65	57.6%	$3.48	**50**	$3.40	$1.0119

Figure 4. Excel simulation worksheet.