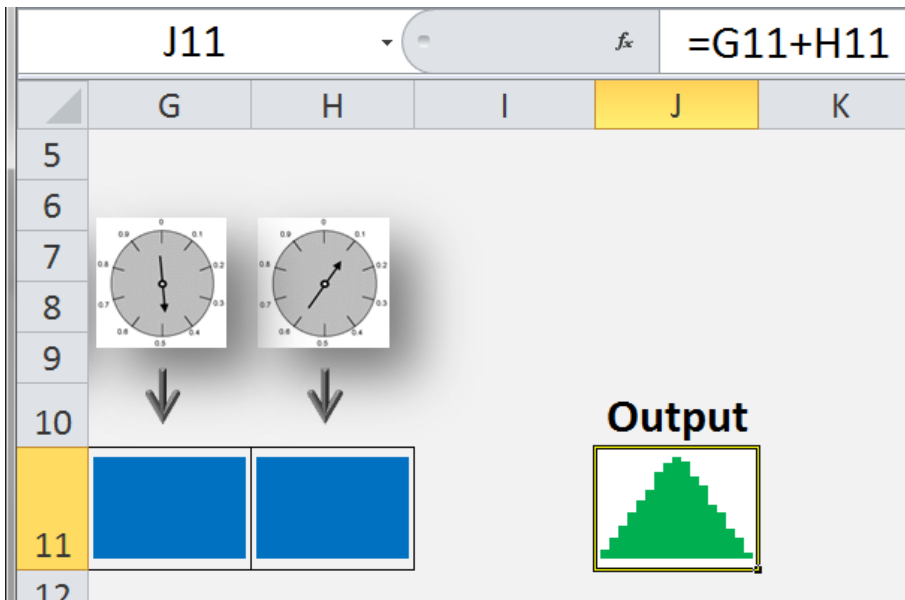


# Flat Plus Flat Equals Up in the Middle

## A Primer on the Arithmetic of Uncertainty

by

Sam L. Savage



## About this Primer

When people perform calculations with numbers they are not certain of, they often substitute “average” values for the uncertainties and then hope for the best. This leads to a set of systematic errors that I call the Flaw of Averages, which explains why so many projects are behind schedule, beyond budget, and below projection.

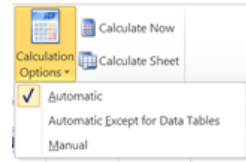
Today, laptop computers and common spreadsheet software are powerful enough to perform arithmetic with uncertainties. This arithmetic introduces three fundamental concepts not present in the arithmetic of numbers: *diversification*, the *impact of non-linear calculations*, and *interrelated uncertainties*.

This primer demonstrates these concepts from an experiential perspective using calculators created in native Microsoft Excel using the SIPmath™ Modeler Tools developed at nonprofit ProbabilityManagement.org.

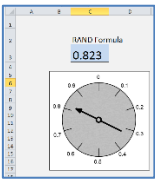
© Copyright 2015, 2016, 2017, 2018 Sam L. Savage  
All rights reserved. No portion of this work may be reproduced in any form without permission, except as permitted by U.S. copyright law. For permissions contact:  
Sam@ProbabilityManagement.org

# The Calculators

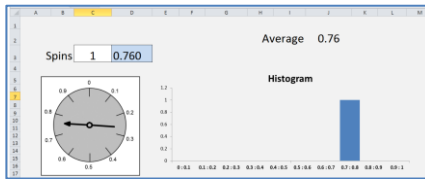
The Primer makes use of the six Excel files described below. For these files to function correctly, Excel's Calculation option must be set to automatic on the Formulas ribbon.



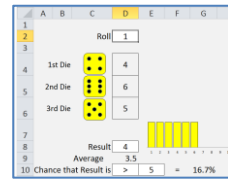
**Be sure to save copies of these files in case you end up overwriting a formula.**



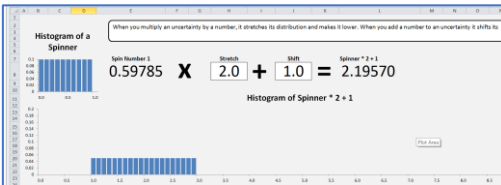
Spinner.xlsx p.4  
The RAND formula



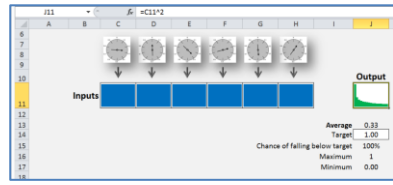
Histogram.xlsx p.4  
Histograms and convergence



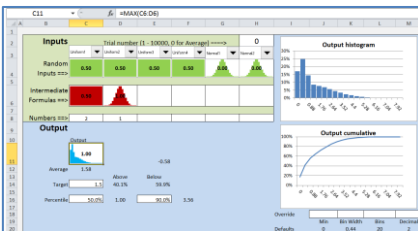
SIPmath Dice.xlsx p.9  
Simulating Dice



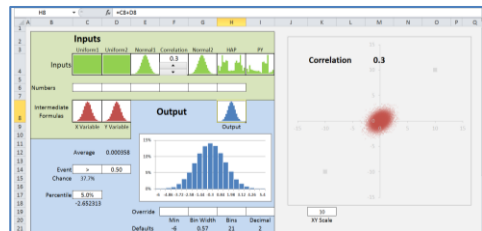
SIPmathCalc\_1 p. 6  
Combining numbers and distributions



SIPmathCalc\_2 p. 11  
Calculations with uniform distributions



SIPmathCalc\_3 p. 20  
General calculations with uniform and normal distributions



SIPmathCalc\_4 p. 32  
Calculations with interrelated uncertainties



# Contents

Flat Plus Flat Equals Up in the Middle .....	i
Introduction .....	1
Chapter 1: Communicating and Visualizing Uncertainties .....	3
Uncertainty as Data, the SIP .....	3
The “Number One” of Simulation.....	4
Histograms, the Shapes of Uncertainty .....	4
Idealized Uncertainty.....	6
Combinations of SIPs and Numbers.....	6
Chapter 2: The Basic Arithmetic of Uncertainty .....	8
Lesson 1: Addition.....	8
Lesson 2: Subtraction.....	14
Lesson 3: Multiplication.....	14
Lesson 4: Division.....	16
The Flaw of Averages .....	18
Chapter 3: Non-linear Formulas .....	20
Jensen’s Inequality.....	20
What is a non-Linear Formula?.....	20
SIPmathCalc_3 .....	20
Lesson 1: Why so Many things are Behind Schedule.....	22
Lesson 2: Why so Many Things Are Below Projection .....	24
Chapter 4: Interrelated Uncertainties and Modern Finance .....	30
4.1 Fire Insurance is a Ridiculous Investment .....	30
4.2 SIPmathCalc_4 .....	32
Examples.....	34
4.3 Portfolio Theory .....	37
4.4 Option Pricing .....	41
4.5 Summary.....	47



## Introduction

Somewhere in Italy, in 1199, Leonardo Fibonacci struggles to explain Arabic numbers to a group of medieval accountants. “Can’t you see we’re busy?” they respond impatiently. “We only have three months to add up this column of Roman numerals.” In spite of such encounters, the famous mathematician prevailed, and Western culture adopted his recommended Hindu-Arabic numbering system. The real revolution, however, was not just the symbols 0<sup>i</sup> through 9, but a new category of thought. We call it *arithmetic*.

There is a parallel category of thought that deals with uncertainties, not numbers. In the arithmetic of uncertainty, addition, subtraction, multiplication, and division are accompanied by such concepts as diversification, contingencies, and interrelationships. The result of such a calculation is not a number; it is a statement about chance. For example, instead of receiving an estimate for profit, one receives the **chance** that profit will achieve a specified target.

The theoretical foundations of this subject date back hundreds of years but are shrouded in a cloak of algebra. So, when faced with an uncertain future, the typical manager plugs in “average” assumptions and hopes for the best. This gives rise to a set of systematic errors I call the Flaw of Averages<sup>ii</sup>, and it explains why so many undertakings are behind schedule, beyond budget, and below projection.

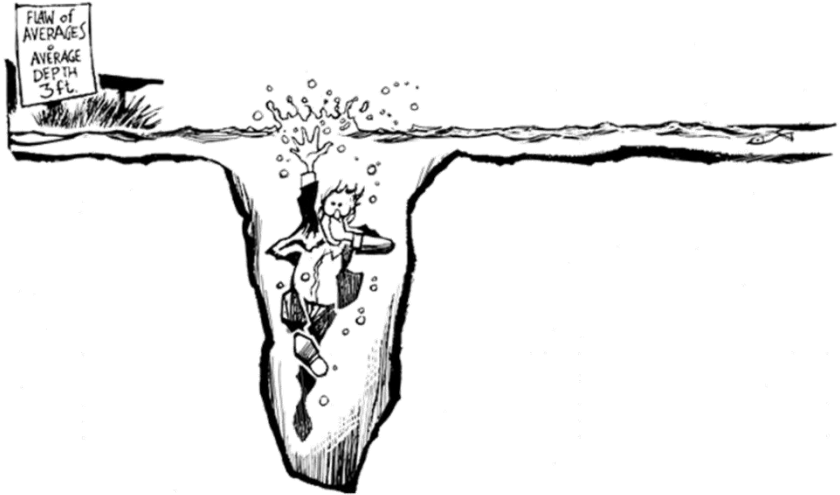


Figure 1 – A classic example of the Flaw of Averages; the statistician who drowns in the river that is on average 3 feet deep

The development of computers allowed us to simulate uncertainties, leading to computational statistics techniques such as a Monte Carlo simulation. These are now common tools for those with statistical training. Just as accounting software is meaningless without a background in standard arithmetic, simulation software cannot be intelligently used or interpreted without some knowledge of the arithmetic of uncertainty.

The common spreadsheet has recently become powerful enough to simulate uncertainty using the open [SIPmath](https://www.sipmath.org)<sup>™</sup> standard from [ProbabilityManagement.org](https://www.probabilitymanagement.org)<sup>iii</sup>. In this environment, you can use the arithmetic of uncertainty with the same keystrokes you use for the arithmetic of numbers. And because it is interactive, performing thousands of calculations with each of those keystrokes, you can learn it experientially much as you learned to ride a bike. As Steve Jobs suggested, you should use the computer as “a bicycle for the



mind,” and follow along with the included SIPmath calculators as you go through the lesson.

## Chapter 1: Communicating and Visualizing Uncertainties

### *Uncertainty as Data, the SIP*

Since the late 1940s, computers have been used to simulate uncertainty. The discipline of Probability Management<sup>iv</sup> represents uncertain numbers as arrays of previously generated data called Stochastic Information Packets or SIPs. *Calculating* with such arrays is called SIPmath. For example, to create a SIP of the uncertainty of rolling a die, you could perform 1,000 rolls and store the results in a row or column of your spreadsheet as shown below.



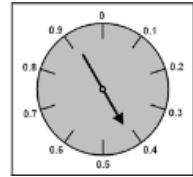
**= 6, 1, 2, 3, 6, 2, 5, 5, 4, 5, 2, ... 2, 4, 1**

#### The SIP of a die roll

Of course, this does not reflect all possible outcomes that could occur in real life. For example, your roll might bounce onto the floor, whereupon you would lean down to retrieve it and then crack your head on the table on the way back up. This problem is described by Nassim Nicholas Taleb as the Ludic Fallacy<sup>v</sup> in his book, *The Black Swan*. Although we should be on guard for such rare events, the SIP is nonetheless a great improvement over succumbing to the Flaw of Averages and expressing each roll of a die as the single average outcome of  $3 \frac{1}{2}$ . In fact, I refer to the belief that you can understand uncertainty without understanding dice as the Ludic Fallacy Fallacy.

## ***The “Number One” of Simulation***

Consider the uncertainty resulting from a game board spinner that goes between 0 and 1. Instead of the six discrete outcomes of a die, there are theoretically an infinite number of outcomes, continuous between 0 and 1. This simple uncertainty is the basic building



block of most computer simulations, and it describes the behavior of the RAND() formula in Excel. In the arithmetic of uncertainty, it plays a somewhat parallel role to the number “1” in ordinary arithmetic.

***Open Spinner.xlsx and observe the RAND() formula in cell C3. Press the <F9> key (<⌘=> on the mac) to get a feeling for how it relates to the arrow on the spinner. Type =RAND() into the blank cell of a worksheet and you should get similar but not identical results to those in C3.***

## ***Histograms, the Shapes of Uncertainty***

A visual description of an uncertain number is a bar graph, known as a *histogram* or *distribution*, which represents the relative likelihood of all possible outcomes. For the spinner, the bar graph is flat, because all numbers between 0 and 1 are equally (uniformly) likely. Therefore, the spinner is said to have a *uniform distribution*. Instead of thinking of an uncertainty as a number, it is better to think of it as a shape—the shape of its distribution.

*Open Histogram Intro.xlsx*

Conceptually, a histogram may be viewed as the result of recording multiple spins of the game board spinner across a number of intervals

or bins. Enter the number of desired spins between 1 and 10,000 in cell C3. For the first few spins, the bars appear to have no pattern. But as more spins are recorded, the graph becomes more uniform until at 10,000 it is virtually flat. Unlike the Spinner file, which generates random numbers with the RAND() formula, this file uses a pre-generated SIP of 10,000 trials so the results are repeatable.

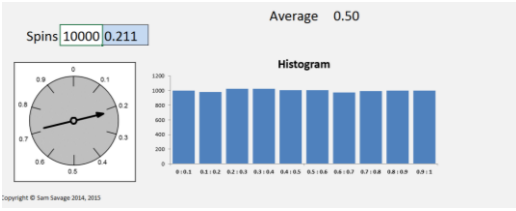
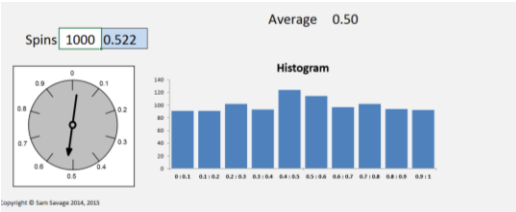
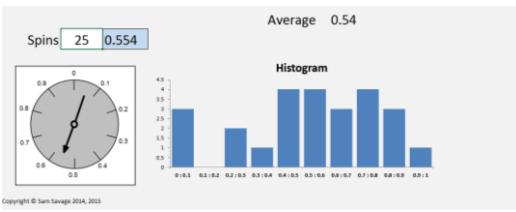
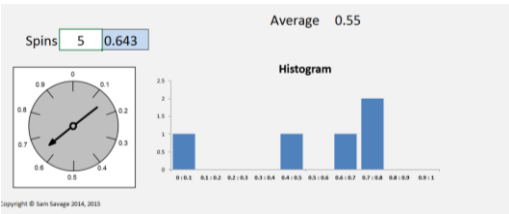
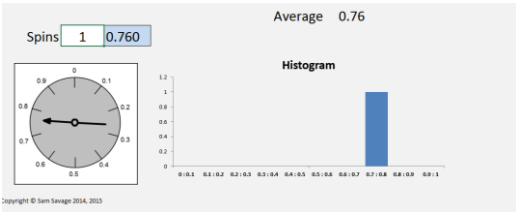
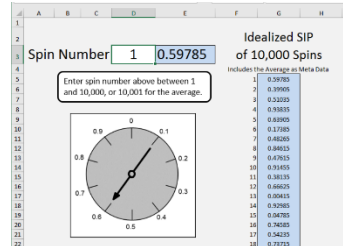


Figure 2 - The Generation of a Histogram

## Idealized Uncertainty

Open *SIPmathCalc\_1.xlsx*, as shown below.

With this file, we will perform some simple arithmetic calculations on an idealized version of the spinner used in the histogram example. Column G contains a SIP comprised of 10,000



outcomes of a spinner, which have been adjusted to yield a perfectly flat histogram. These are known as **stratified** trials by simulation experts. Unlike the trials in the histogram file, which were simulated to mimic the results of an actual spinner, the stratified samples are generated as follows. It starts with the 10,000 numbers 0.00005, 0.00015, up to 0.99995. These numbers are then randomly shuffled like playing cards.

It is sometimes useful to store metadata with a SIP. Here, the average of the 10,000 outcomes of the idealized trials, or 0.5, is stored just below the trials, in position 10,001. Specify the trial number you wish to view in cell D3, or enter 10,001 to display the average.

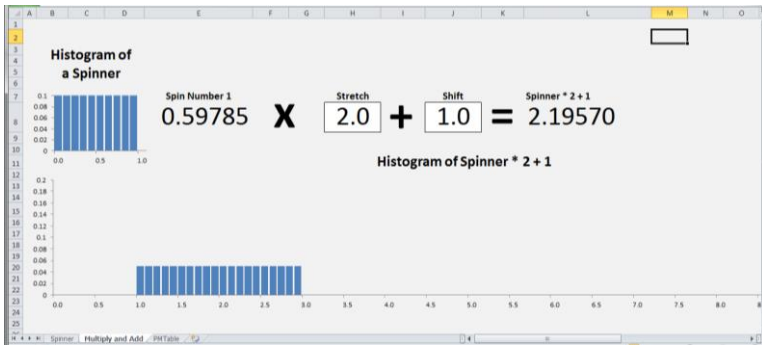
## Combinations of SIPs and Numbers

Click on the **Multiply and Add** tab of the *SIPmathCalc\_1* worksheet.

Uncertainties and numbers may be combined using standard arithmetic. This sheet allows you to multiply the spinner uncertainty by a number and then add another number to the result. This is the formula for a linear function:

$$Y=mX+b,$$

where  $X$  is the SIP of the spinner. The histograms on this sheet have bins of width 0.1. The vertical axis denotes the chance that the result appears in a given bin. Adding a number to a SIP ( $b$  in the formula above) shifts its histogram to the right, so I will call it “Shift” instead of  $b$  in the calculator. Multiplying a SIP by a number ( $m$  in the formula above) stretches the width of the histogram so I will call it “Stretch.” In the example below, we have multiplied a spinner by 2 and then added 1. The resulting histogram starts at 1 and has a width of 2, extending from 1 to 3. The bars of the resulting histogram have become shorter because there is less chance of the number ending up in any given bin.



***Change the numbers in cells H8 and J8 and observe the changes in the histogram at the bottom of the screen.***

Uncertainties with distributions of *any* shape are stretched and shifted in this manner under multiplication and addition by numbers.

## Chapter 2: The Basic Arithmetic of Uncertainty

### Lesson 1: Addition

What happens if you add the results of two separate spins or `RAND()` formulas together? This is the rough equivalent of “1+1” in the arithmetic of uncertainty. Before continuing, see if you can select the correct distribution from the figures below.



ANSWER: See footnote<sup>1</sup>

Why is that the answer?

Instead of spinners, think of dice.

When you roll one die (🎲), you can get any of the numbers from 1 to 6 with equal likelihood. A histogram of the results from many rolls of one die is flat across like that of one spinner.

What happens when you add the outcomes of two dice, then repeat the process many times? There is only one combination that results in 2, and one that results in 12. But there are six separate ways to get a 7, so the histogram of the results goes up in the middle.



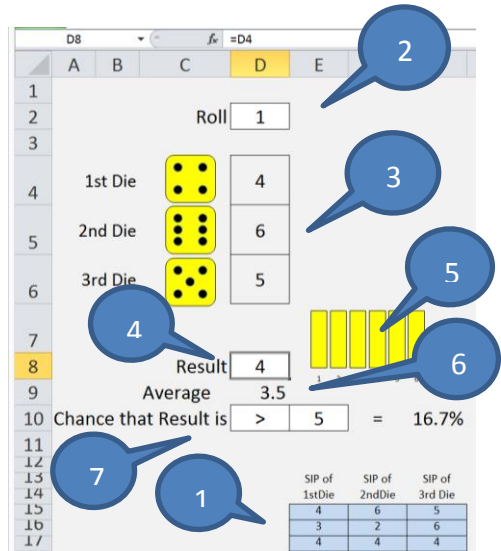
---

<sup>1</sup> Answer is the 6<sup>th</sup> letter in the alphabet

Similarly, with the sum of two spinners, the only way to get a number close to 0 is with *two* spins that are close to 0. The only way to get a sum close to 2 is with *two* spins close to 1. But a sum close to 1 may be achieved with many combinations such as 0 and 1, 1 and 0, 0.5 and 0.5, and so on.

In the world of both spinners and dice, *Flat Plus Flat Equals Up in the Middle*.

Let's explore the arithmetic of dice with SIPmath Dice.xlsx. This is also a good introduction to the rest of the SIPmathCalculators.



1. SIPs of ten thousand rolls of each three dice are stored starting in row 15.
2. Cell D2 contains the rolls displayed below. For example, the first roll is a 4, 6 and 5. Change D2 to 3, and you will get three 4s.
3. These cells contain Index formulas that point into the three SIPs.
4. Cell D8 is the Result of performing arithmetic on the three dice. All 10,000 trials will be run through this cell. Currently it is pointing to the first die. Try adding the first two dice together by entering the formula =D4+D5.
5. The histogram displays the results of the 10,000 rolls. When you add two dice together in cell D8, this will become a triangle.
6. Cell D9 contains the Average of the 10,000 rolls. When you add two dice this will go from 3.5 to 7.

7. You may specify conditions on the calculation in cells D10 and E10, whereupon cell G10 displays the chance that those conditions will be met.

Experiments:

- Add two dice together to confirm the triangular histogram.
- Multiply two dice together. Can you explain the histogram?
- Add three dice. How does the shape change from that of adding two dice?

### **So What?**

The principles demonstrated above with spinners and dice are at the heart of diversification, which plays a critical role in risk reduction in banking, investing, engineering, and many other applications. It underlies the saying that you shouldn't put all your eggs in one basket. **Open SIPmathCalc\_2.xlsx and we'll learn more.**

Each of the Input cells C11 through H11 are linked to the outcomes of 10,000 spins of an independent spinner. These are idealized like the ones in SIPmathCalc\_1, but each one has been permuted (scrambled in order) separately to assure that the number appearing on any spinner in a given trial has no relation to numbers appearing on the other spinners for that trial.

### **An Investment Example**

To provide some literary appeal to the spinner, imagine multiplying the results by \$1 million to model an uncertain investment.

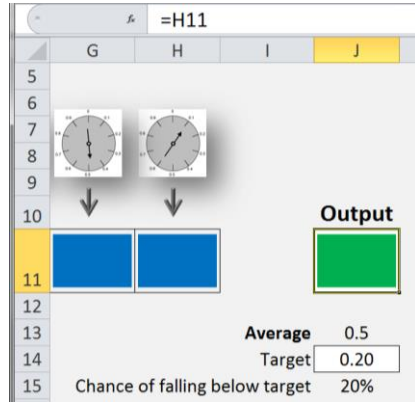
Intuitively, the return will range between 0 and \$1 million with an average of \$500,000. To thicken the plot, suppose that you owe \$200,000 to ruthless loan sharks, which means that a spin of 0.2 or less will be disastrous.



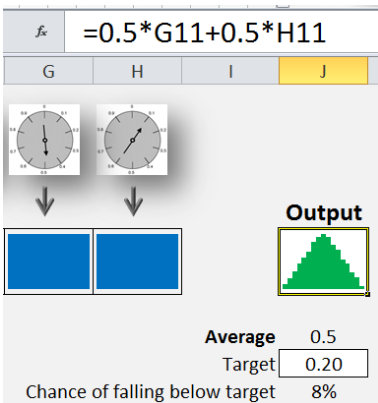
We will now use SIPmathCalc\_2.xlsx to calculate the chances of becoming shark bait.

With the cursor in the Output cell J11, type “=” and then click in one of the Input cells, and press <Enter>.

Next, change the Target value in J14 to 0.20.



This runs all 10,000 spins through the Output cell, resulting in the same flat histogram as the Input you selected. The formula in cell J15 shows us that there is a 20% chance of missing our Target of 0.2 in cell J14. That is, there is one chance in 5 we won't be able to pay back the loan.



What would happen if we split our investment over two independent spinners?

Change the Output cell formula to reflect putting half your money in each of two input cells as shown on the left.

The shape changes to a triangle, as predicted by the dice, and because it goes up in the middle, it goes down on the ends, implying lower risk of ruin. Cell J15 indicates that the risk is now only 8%, a big

drop from 20%. Furthermore, cell J13 indicates that we still have an average of \$500,000.

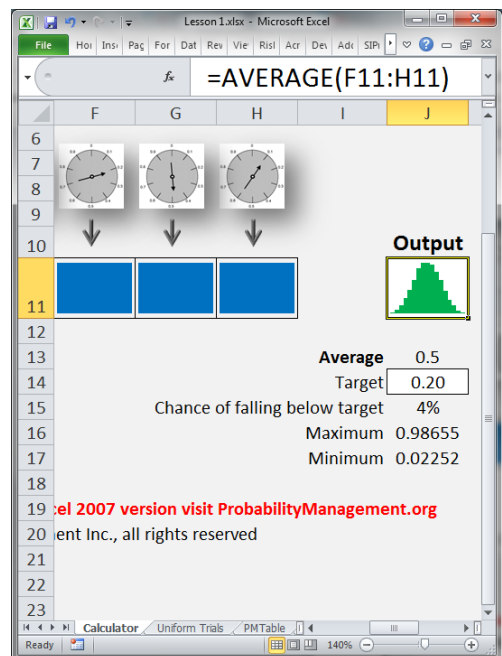
So, how did the computer do that so fast? The PMTable sheet contains the results of 10,000 trials for the Output cell calculated with an Excel function known as the Data Table. Whenever one of the Input spinners is referenced in Output cell J11, the stored results are used to calculate 10,000 possible outcomes for the Output cell.

In this setup, Output cell J11 has been named 'Output,' so you can refer to the cell this way instead of referring to it as J11 when you perform calculations with it. For example, in J16, we take the MAX(Output) instead of MAX(J11).

Now, how about diversifying across three spinners? In the Output cell, we could enter =SUM(F11:H11)/3, or equivalently, the =AVERAGE function.

**Enter the formula shown in cell J11.**

The risk of ruin has now dropped to 4% with the same average return. Notice that the histogram is now bell shaped.



## Now take the average over all six spinners.

The bell becomes narrower, and you have virtually nothing to fear from the loan sharks.

Each investment (entered in the Output cell formula), from the single spin to the one diversified across six spins, returned the same average of \$500,000. But after a few keystrokes, the calculator indicates very different levels of risk. This shows how erroneous it can be to represent uncertainties with single numbers.

## The Central Limit Theorem



The above example drives home the importance of diversification. But this is just a special case of a more profound phenomenon known as the *Central Limit Theorem*, which some have called the most important result in mathematics. It describes precisely how the sum or average of *independent* uncertainties become bell shaped as the degree of diversification is increased. This bell-shaped uncertainty is called a *Normal* or *Gaussian* distribution after one of its discoverers, Carl Friedrich Gauss (1777–1855), shown posing with his distribution on a German 10 mark note. It is important to mention that this result does not just apply to outcomes of spinners, but to independent samples of almost all uncertainties.

Normal distributions may be completely specified by two numbers, the *average* (this is where the distribution is centered), and the *Standard Deviation*, a measure of how wide the distribution is (or how much it has been stretched).

One more consequence of the Central Limit Theorem is that, when two Normal uncertainties are added together, the result is also Normal. The average of the sum is equal to the sum of the averages of the added uncertainties. The width of the sum will be discussed in Chapter 4.

### ***Lesson 2: Subtraction***

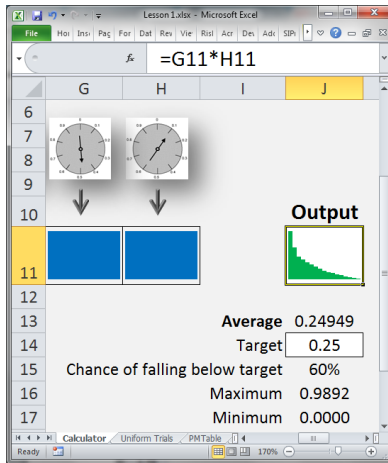
If you subtract one spinner from another (=G11-H11) you get the same triangular shape as when you add two spinners. But the average is now the average of one spinner minus the average of the other spinner, or zero.

You can determine the average derived from adding or subtracting uncertainties by merely adding or subtracting the averages of the uncertainties themselves. But this is **ONLY** true in general for addition and subtraction. As we will show, the average of *other* arithmetic calculations is *not* always equal to the calculation performed by using the average of each uncertainty.

### ***Lesson 3: Multiplication***

**Multiply two independent spinners together.**

And you will get the shape below.



The minimum is zero, and the maximum is nearly 1. But it is not flat or triangular. The only way to get a 1 is if both spinners result in 1. On the other hand, if either spinner results in zero, you get zero for the product. That's why the histogram is so much higher on the left. You will also note that the average of .25 is the product of the averages of the two spinners, but this is only the case if the underlying uncertainties are independent, as we will see further on.

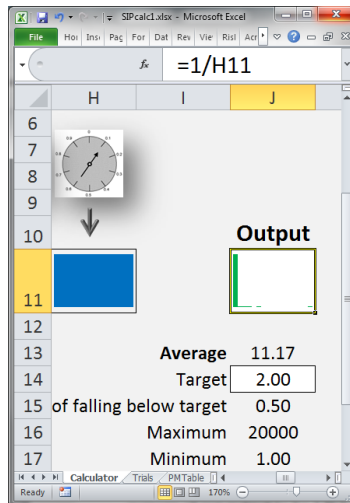
### So What?

Imagine that you are launching a new internet service with uncertain demand and price. The demand in millions of customers has the uncertainty of a spinner, while the price in dollars has the uncertainty of a second independent spinner. The revenue in millions is the product of demand and price, so the maximum is \$1 million, the minimum is \$0, and the average is \$250,000. But we can see from the calculator that it is virtually impossible to get a full \$1 million, and from cell J15 there is only a 40% chance of achieving at least the average revenue of \$250,000.

## Lesson 4: Division

Division is the only operation in standard arithmetic with a potential land mine—a zero in the denominator. With the arithmetic of uncertainty, all it takes is a chance of zero in the denominator to cause trouble. We will start by investigating the number 1 divided by a spinner (the reciprocal of the spinner).

**Enter =1/H11 in the output cell.**



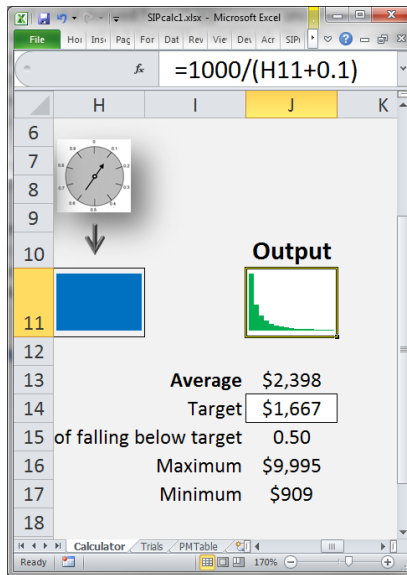
Because we have used pre-computed trials of an idealized spinner, of which the smallest value is 0.00005, the maximum output is  $1/0.00005$  or 20,000. But if we had generated the spinner trials in a different manner so that the smallest number was actually zero, then the maximum would have been infinite. So, this is an unstable calculation and NOT reliable. The danger of dividing by very small numbers transfers over to dividing by uncertainties that have a chance of being very small.

The histogram is puzzling at first, but may be explained as follows. If you adjust the target, you will see that 99.9% of all outputs are below 1,000. All those trials lie in the first bar of the histogram. The remaining tenth of a percent of the trials lie in the remaining bins, resulting in the long tail on the right, ending at 20,000 in this example.

### **So What?**

Consider estimating the cost per ton of a smelting process that extracts metal from ore. It costs \$1,000 for the energy and the ore itself to fill the smelter. Now, imagine that the quantity in tons of metal derived from this process is uncertain due to the varying purity of the ore. We will model this quantity as a spinner that goes between 0.1 and 1.1 (avoiding zero). That is, it is equally likely that the number of tons will be somewhere between a tenth of a ton and 1.1 tons, due to the uncertain purity. The average tonnage is  $0.5 + 0.1 = 0.6$ . So, you might guess that the average cost per ton would be  $\$1,000 / 0.6 = \$1,667$ . We will model this by adding 0.1 to a spinner as follows.

In SIPmathCalc\_2, enter the formula  $=1000/(H11+0.1)$  in the Output cell, and 1667 as the Target.



Note that there is a 50% chance of being over or under the target. However, the cost per ton given the average tonnage is \$1,667, but the average cost per ton is almost \$2,400. This is not close enough to keep the accountants happy.

### ***The Flaw of Averages***

As we have seen, if you represent uncertainties as ranges of possible outcomes in a spreadsheet, you can perform arithmetic on them just as if they were numbers. Why bother? Recall the various spinner investments. If you described them only in terms of their average returns of \$500,000, they would appear indistinguishable. But, of course, they have widely varying risks. I call this the *weak form* of the Flaw of Averages. If you perform the calculation with the averages of



the inputs, although you completely miss the risk implications, at least you get the average output right.

The *strong form* of the Flaw of Averages doesn't even get the average right, as demonstrated by the cost per ton example. This occurs when the calculations are non-linear, as we will discuss in Chapter 3.

## Chapter 3: Non-linear Formulas

### *Jensen's Inequality*

The technical term for the strong form of the Law of Averages is Jensen's Inequality, after the Danish mathematician and engineer Johan Ludvig William Valdemar Jensen (1859 – 1925). It applies to non-linear formulas, and helps explain why so many ventures are behind schedule, below projection, and beyond budget.



### *What is a non-Linear Formula?*

A linear formula only allows you to multiply variables by numbers, add the results together, and add a constant. All other formulas are non-linear.  $Y=mX+b$  is one of the simplest linear formulas.

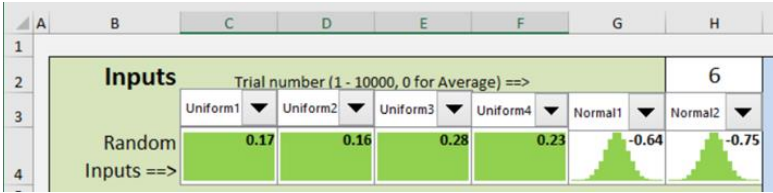
### *SIPmathCalc\_3*

Open SIPmathCalc\_3.xlsx. Note the following primary differences from SIPmathCalc\_2.

1. The Random Inputs may be set to any of 12 independent SIPs, each with distributions that are outcomes of spinners (Uniform) or normal (Gaussian).

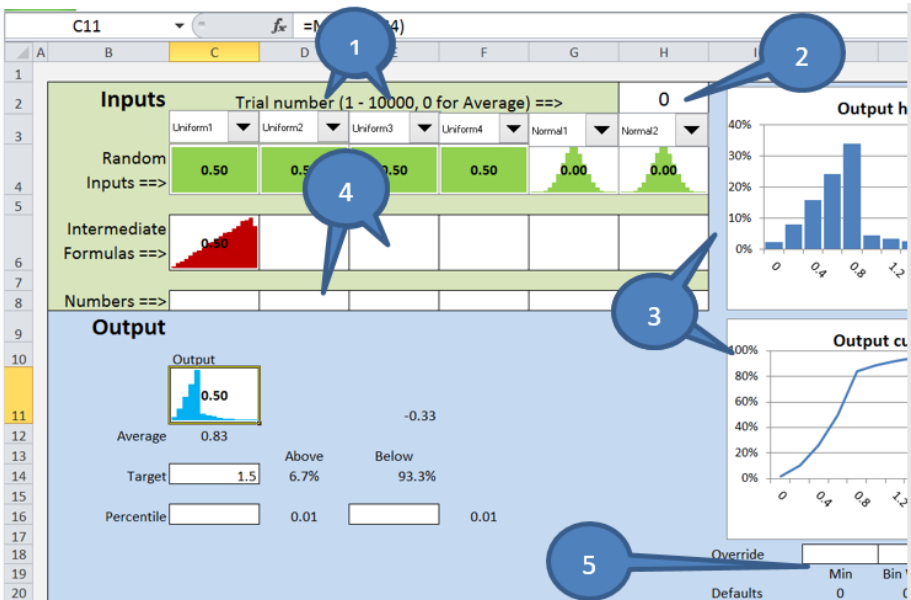
To change one of the Random Inputs, click and hold one of the drop-down arrows and select a distribution.

2. The sparkline graphs now contain a number, which may represent the value on any of the 10,000 trials as specified in cell H2, set to 6 below.



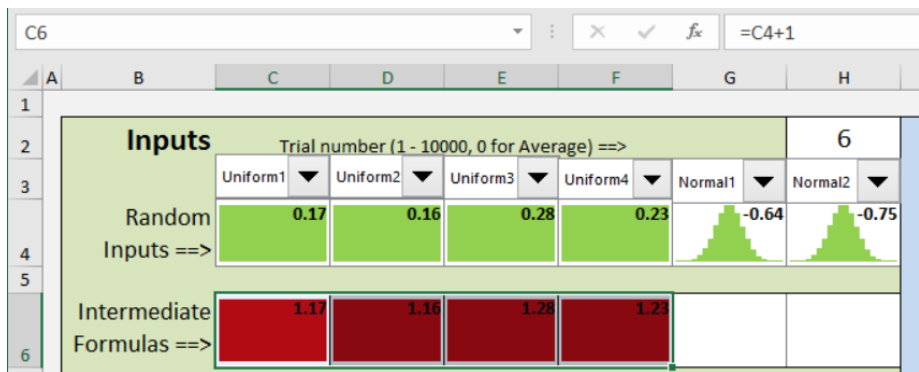
Setting H2 to 0 sets the input distributions to their average values.

3. Once the Output cell, C11 contains a formula that depends directly or indirectly on the Random inputs of row 4, all 10,000 trials of the Output distribution are displayed to the right on both a histogram and cumulative graph.
4. You may enter your own intermediate formulas and numbers to feed the Output cell in rows 6 and 8.
5. Note that the parameters of the histogram may be edited for better readability by values in cells J18 through M18. These cells turn yellow to remind you that they are not the defaults. You may return to the defaults by deleting the values.



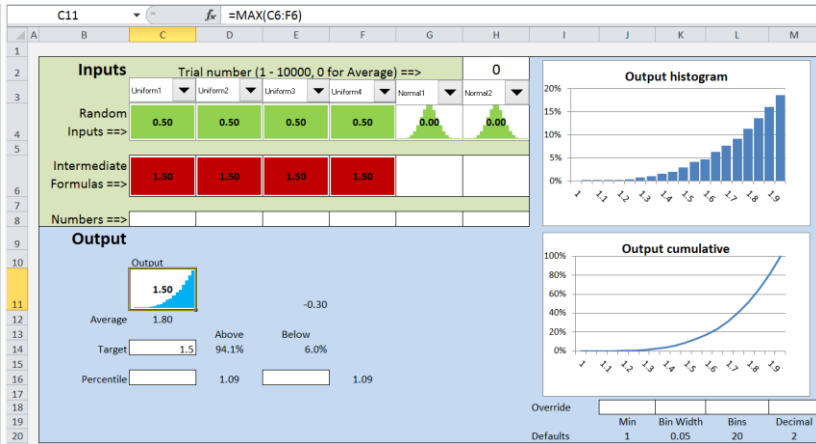
## Lesson 1: Why so Many things are Behind Schedule

Imagine a project consisting of four parallel tasks. Each task is uncertain in duration with a uniform distribution between 1 and 2 months. Thus, each task will take 1.5 months on average. Therefore, you might reasonably expect the project to take 1.5 months on average. Let's find out. Open SIPmathCalc\_3 and delete the contents of rows 6 and 8, and the Output cell.



We start by modeling the task duration in months by adding 1 to each of four uniform distributions. First, make sure the first four Random Inputs are set to Uniform1 through Uniform4 (cells C3 through F3). Then, enter the formula =C4+1 in cell C6, and then copy it over through column F. For example, the formula in E6 should be =E4+1.

The project won't be finished until the last of the four tasks is finished. So, in the Output cell (C11) enter =MAX(C6:F6). This formula causes the non-linearity. See the Output histogram below.



Set H2 to 0 to display the average value of the inputs. This sets the duration of each task to its average of 1.5 months as shown in C6 through F6, and results in the project also finishing in 1.5 months as shown in cell C11.

So, the duration of the project, given the average assumptions, is 1.5 months, but the *average* duration of the project (cell C12) is 1.8 months. That is, if you sequentially entered all the numbers between 1 and 10,000 in cell H2, the average value of C11 (Output) would be 1.8. With linear calculations, there is no difference between these two results. That is, with a zero in cell H2, cells C11 and C12 would be equal.

Next, set the Target (C14) to 1.5, the duration you would have assumed based on averages. Had you done this, there is only about a 6% chance you would have come in on time. That is, the only way this could happen would be if all four tasks finished in less than 1.5 months, which is like flipping four heads in a row on a coin.

## **So What?**

You rarely hear about projects coming in early, but it is common to hear about projects that are late. There are several reasons for this, but the principles demonstrated above are a big contributor. To help avoid underestimating project completion times, managers of large industrial projects often use specialized simulation software to estimate the distribution of completion times.

## ***Lesson 2: Why so Many Things Are Below Projection***

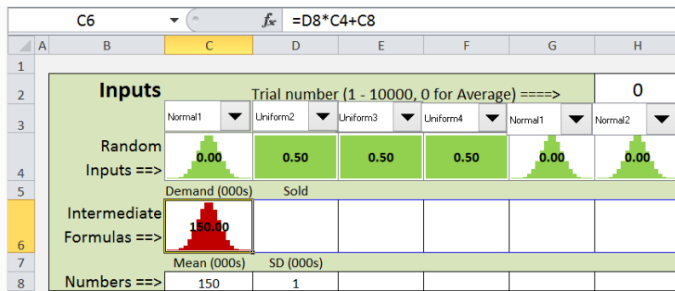
Imagine that you are ordering electronic equipment six months in advance of the Christmas selling season at a cost of \$8 per unit. The demand is uncertain, but your best guess is 150,000 units with a 90% confidence interval of between 100,000 and 200,000 units. The selling price is \$10, so that's what you will receive on each unit sold, yielding a profit margin of \$2 per unit. So, if you order 150,000 units, your expected revenue is \$300,000 ... or is it?

## **Modeling Demand**

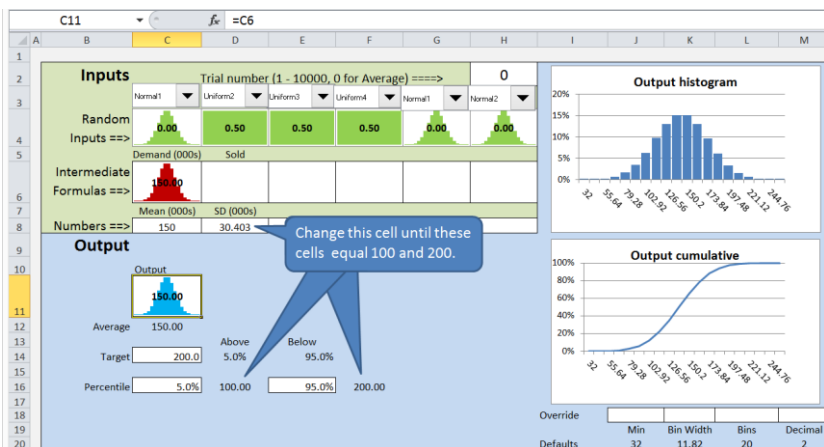
Because total demand is the sum of a large number of individual purchases, it is reasonable to assume that the uncertainty of total demand is normal. We will use the normal distributions in SIPmathCalc\_3, which are Standard Normals. Standard Normal distributions have an average of 0, and a *standard deviation* of 1. If you recall from Chapter 1, when you multiply an uncertainty you stretch the results, and when you add to it you shift the results. You may think of the standard deviation as the stretching distance to get the required amount of uncertainty, and we will use the abbreviation SD. *You* decide what you want it to stand for.

First, delete any formulas or numbers in rows 6 and 8 and the Output cell, then Select Normal 1 for cell C4. Next, set the percentiles in cells

C16 and E16 to 5% and 95% respectively. This will ensure that they encompass a 90% confidence interval. Because these numbers are large, it is recommended that they be expressed in thousands. That is, we will model the mean demand as 150 (000s) in cell C8, and a stretching distance (or standard deviation) of 1 (000s) in cell D8. Also create labels as shown in rows 5 and 7. Then, in C6 enter  $=D8*C4+C8$ . This will result in a normal distribution with a mean of 150 (000s) and  $SD=1$  (000s).



We will now interactively adjust SD until we achieve the desired confidence interval. First enter  $=C6$  in the Output cell C11 to temporarily set demand as the output. The histogram should appear normal. Next, look at the percentiles in D16 and F16. These contain the standard Excel Percentile formulas. The 90% interval should run from 100 to 200 in (000s). You will need to increase the stretching distance in cell D8 until the 5<sup>th</sup> and 95<sup>th</sup> percentiles are approximately 100 and 200. By experimenting with SD and observing cells D16 and F16, you will find that a value of 30.403 results in the desired percentiles. Also set the Target in C14 to 200. You should get the results below.



WARNING: It is important to note that, as in horseshoes and hand grenades, “close” counts with a simulation. It was satisfying to discover a value for SD that precisely matches the percentiles, but it imputes unwarranted accuracy on the results. A value of 30 is plenty close enough, so set it back to 30 to avoid this problem. As we will see, the errors resulting from representing demand as a single number, even if it is the *correct* average demand, swamp the inaccuracies in estimating the distribution of demand.

### The Empirical Rule of Normal Distributions

Before computers made it so easy to calculate probabilities, people made use of what is known as the empirical rule when using the Normal distribution. This rule is reflected in the table below.

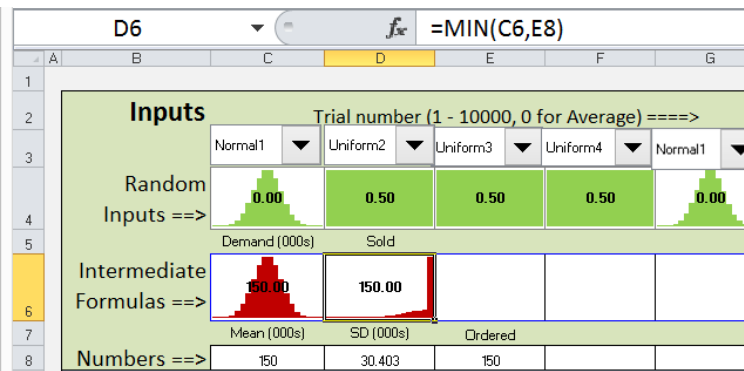
68% of the distribution lies between	1 Standard Deviation Above and Below the Average
95% of the distribution lies between	2 Standard Deviations Above and Below the Average
99.7% of the distribution lies between	3 Standard Deviations Above and Below the Average



This would not have helped us with a 90% confidence interval as specified in the problem above, and it would have been necessary to consult statistical tables. With computational statistics, rules such as these become less important.

### Modeling Quantity Sold

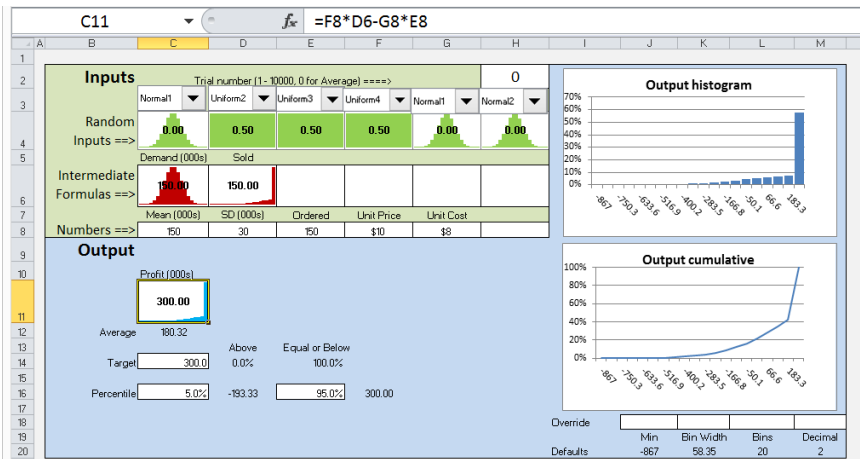
Now that we have modeled uncertain demand, we will model sales. This is where the non-linearity sneaks in. If actual demand is less than the quantity ordered, then we only sell the demand. But if demand is greater than the quantity ordered, we stock out, and can only sell the quantity ordered. Therefore, enter 150 in cell E8 and label it “Ordered.” Next, model the quantity sold in cell D6 as the minimum of demand and quantity ordered. Put =D6 in Output cell C11 to graph the result. MIN, like the MAX formula, creates a non-linearity. Notice that the histogram of the quantity sold has a big spike on the right. That’s because any demand greater than 150 results in sales of exactly 150.



### Modeling Profit

Create cells in row 8 for the price per unit of \$10 and the cost per unit of \$8. Next change the output cell to reflect profit in thousands. This is simply the *price per unit* times the *quantity sold* (total revenue)

minus the *cost per unit* times the *quantity ordered* (total cost), as shown below.

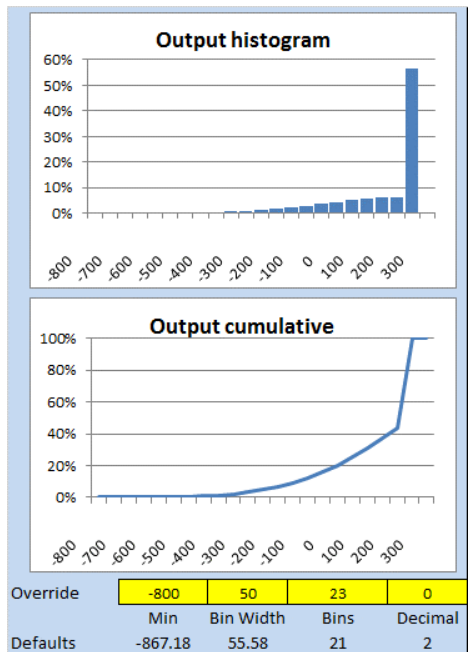


Make sure that cell H2 is set to zero to display the average inputs. The output cell shows that the profit of the *average* demand is \$300,000 while the true *average* profit is only \$180,000. If you forecast profit based on average demand, then on average you will be about \$120,000 below projection. The problem is that if demand is below average, you won't make your numbers, and if demand is above average, you stock out and can't take advantage of it. Set the target in C14 to 300 and you will see there is no chance of coming in above. Also, note that there is a 5% chance of losing almost \$200,000.

Now experiment with SD. If you change it back to 30.403, you will see it makes almost no difference in your average profit. In fact, if you lower it all the way to 5 you will still be \$20,000 below projection. Set SD back to 30 when you are done.

## Histogram Setting

Histograms are most readable when the parameters are adjusted by hand. Note: When the settings are changed in cells J18:M18 the background turns yellow to remind you they are no longer automatic. When you are done, the defaults may be reset deleting the Override numbers. In the histogram shown, the number of decimals has been reduced to 0, and minimum and bin width have been set to make the X axis more understandable.



Sometimes, setting the minimum and bin width will force the upper end of the histogram off the right side of the graph. In this case, you must increase the number of bins as shown on the right.

### Deciding How Much to Order

It would make sense that ordering more up front would let us take advantage of higher demand if it occurred. On the other hand, if demand was lower than expected, you would then have more unsold inventory lying around and maybe you should lower the quantity ordered. Adjust the quantity ordered in cell E8, attempting to maximize the average profit in cell C12. These sorts of problems are non-intuitive, and really require a model to help us think through the

decision. When I made up a similar exercise for my first textbook, I expected the opposite result.

### **So What?**

Every business opportunity that requires an up-front investment in capacity to meet uncertain demand faces some version of the problem above. That is, there is a serious downside if demand is low, and limited upside if demand exceeds capacity. No wonder so many ventures are below projection.

## **Chapter 4: Interrelated Uncertainties and Modern Finance**

This chapter provides experiential insight into portfolio theory and option pricing. The models are not intended to provide numerically valid results, but rather to build directionally correct intuition through interactive simulations. When used in conjunction with a traditional book on finance, and a good source of financial data, these models will help connect the seat of the intellect to the seat of the pants.

### ***4.1 Fire Insurance is a Ridiculous Investment ...***

unless you own a house to go along with it.

Whether your house will burn down or not next year is uncertain. It is also uncertain whether your fire insurance policy will pay you money next year. But these two uncertainties are not independent, like two spinners. They are directly related; your insurance will only pay you *if* your house burns down.

Many uncertainties in life are interrelated with other uncertainties, but usually less directly than the value of your house and your insurance policy. Statistical interrelationships became widely

recognized in the area of investments through the Nobel Prize-winning *Modern Portfolio Theory* of Markowitz and Sharpe in the 1950s and 1960s. This was followed by another Nobel Prize involving interrelated investments: *The Theory of Options*, developed by Black, Scholes and Merton in the 1970s. In fact, modern finance is fundamentally based on the arithmetic of interrelated uncertainties.

But first, let's explore how to model a house and fire insurance using SIPs.

Imagine that your house is worth \$200,000 and has one chance in 1,000 of going up in flames next year. Suppose that your insurance costs you \$200 per year, and will pay the full value of your house if it burns. A SIP of the value of your house over the next year appears on the right. A fire reduces its value to zero on trial 998, but otherwise it's worth \$200,000.

	House
1	\$200,000
2	\$200,000
3	\$200,000
4	\$200,000
:	:
:	:
:	:
997	\$200,000
998	\$0
999	\$200,000
1000	\$200,000

	House	Insurance
1	\$200,000	-\$200
2	\$200,000	-\$200
3	\$200,000	-\$200
4	\$200,000	-\$200
:	:	:
:	:	:
:	:	:
997	\$200,000	-\$200
998	\$0	\$199,800
999	\$200,000	-\$200
1000	\$200,000	-\$200

A SIP describing your insurance policy will have 999 trials with a value of -\$200 when your house doesn't burn down, and one trial with a value of \$200,000 - \$200 = \$199,800, when it does burn down (note you still need to pay your annual premium of \$200). But, unlike previous examples in this book, the order of the Insurance SIP relative to

the house SIP is critical. The insurance payout must line up with the fire loss or it is like buying fire insurance on a stranger's house, which is ridiculous.

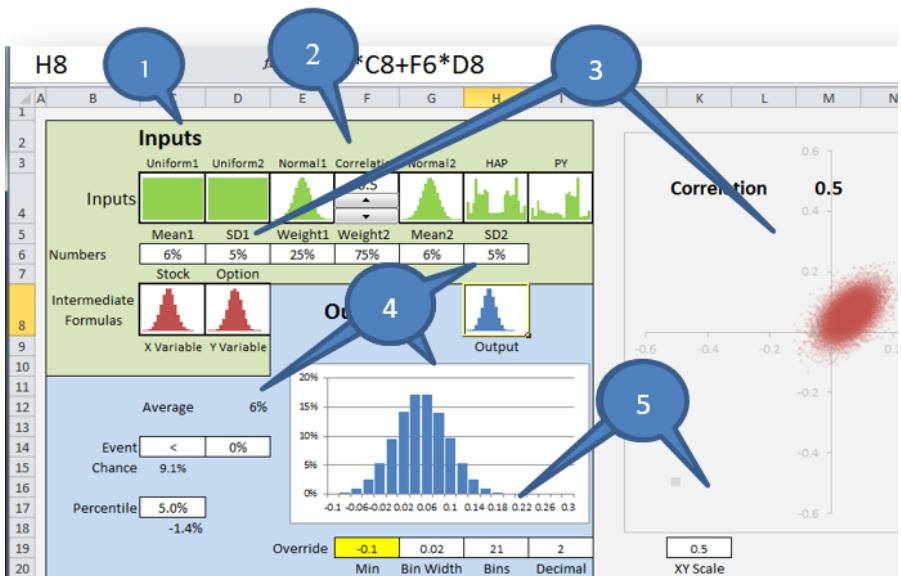
A set of SIPs such as these, which embody interrelationships, are said to be **Coherent**, and are referred to as a **Stochastic Library Unit with Relationships Preserved**, or a **SLURP**.

As we will see in this chapter, interrelationships between the variables can have a profound effect on the resulting arithmetic. For example, when you add the SIPs of your house and your insurance policy above, you get \$199,800 on every trial. **So, in this special case, the sum of these two uncertainties is a certainty!**

## 4.2 SIPmathCalc\_4

The SIPmathCalc\_4 calculator has the following features:

1. The inputs include two uniforms (spinners), two standard normals (mean=0, SD=1) whose interrelationship may be adjusted, and two other distributions, called HAP, and PY.
2. A control to adjust the degree of interrelationship (correlation) between the two normal.
3. A scatter plot of the distributions in cells C8 and D8.
4. An output cell with associated statistics and histogram.
5. Scale factors to control the limits of the graph; note that the



parameters of the histogram are automatically set, but may be overridden in row 19 for better readability. To return to the defaults, simply delete the override values.

## **Scatter Plots**

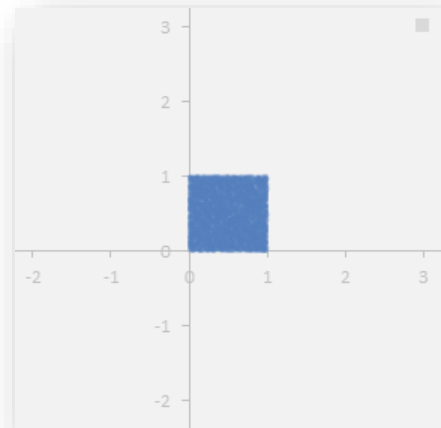
Before computers, the primary measurement of statistical interrelationships was correlation as discussed below, but the best way to visualize the interrelationship between two SIPs is to create a scatterplot in which the X axis reflects the values that make up the first SIP and the Y axis reflects the corresponding values of the second SIP. Each data point in the scatter plot represents the value of the first SIP and the value of the second SIP for a given trial.

## **Correlation**

The dictionary definition of correlation is the mutual relation of two or more things. There is also a mathematical definition of correlation used in statistics. The Pearson Correlation, as it is called, measures the degree to which the scatter plot of two uncertain variables lie along a straight line. Like many concepts from classical statistics, it is most applicable to normal distributions. As we will see, there are some distributions for which this sort of correlation is misleading. A major benefit of SIPs is that they can preserve statistical relationships within the data, so the correlation is handled automatically. In addition, a much broader set of relationships may be expressed than those captured in the Pearson Correlation.

## Examples

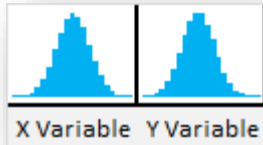
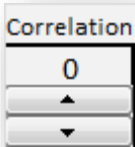
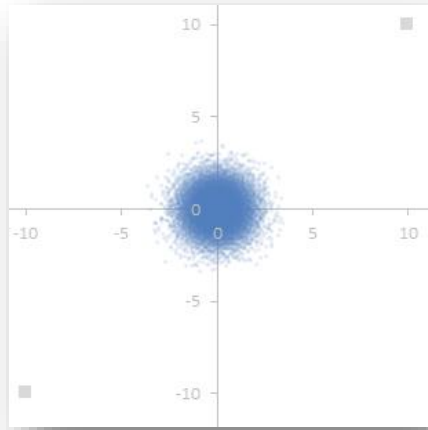
The scatter plot of the SIPs of two independent spinners (uniform uncertainties) looks like a square. The Pearson Correlation is 0. To get this



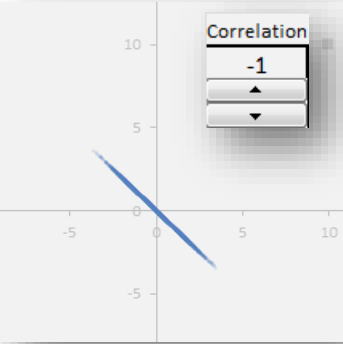
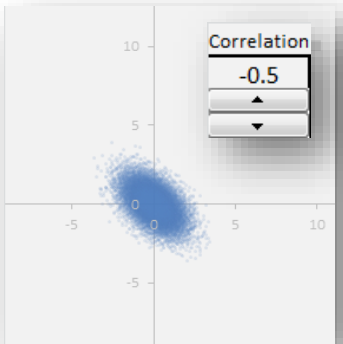
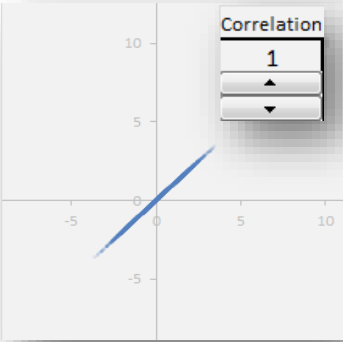
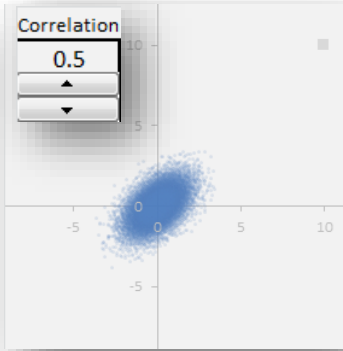
plot with SIPmathCalc\_4, set the X Variable to =C4, the Y Variable to =D4, and the XY Scale of the scatter plot to 3.



The scatter plot of the SIPs of two independent normals with the same mean and SD looks like a circle that gets denser toward the center. The Pearson Correlation is 0. To get this plot, set the Correlation control to 0, the X Variable to =E4, the Y Variable to =G4, and the XY Scale of the scatter plot to 10.



When the correlation control is increased, or decreased, the circle turns into an ellipse sloping up or down to the right. The Pearson Correlation matches the number on the control as shown below. At  $\pm 1$ , the ellipse turns into a straight line.



One might conclude from this that if the Pearson Correlation is zero, then the variables are uncorrelated in the dictionary sense of the word. Set the X variable to HAP (=H4), and the Y variable to PY (=I4) to find out just how wrong this assumption can be.

### ***4.3 Portfolio Theory***

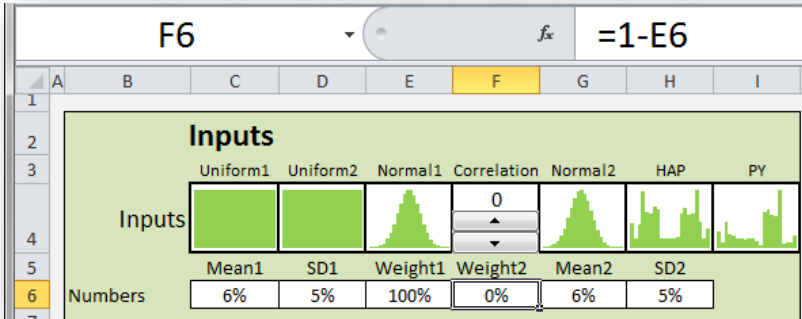
There are two basic ideas underlying Modern Portfolio Theory.

First, when you ***diversify*** across multiple investments, you are less likely to get extreme events. This is just a generalization of the fact that flat plus flat equals up in the middle. The second basic idea is that some stocks move strongly up or down together, while some move less strongly together or even in opposite directions. This is measured by the ***correlation***. If you split your investment between two stocks that go up and down together, it is like owning a half share in two houses so close to each other that if one burns down, the other is likely to burn down as well. On the other hand, if the two stocks are independent, or better yet, move in opposite directions, they play the role of each other's insurance policies to some extent, and the combined investment of both stocks is less risky than the investment in only one stock.

We will model the returns of two stocks with normal distributions having identical mean annual growth rates of 6% and SD of 5%. This implies that, on average, each of the two investments will grow 6% a year, but there is 95% chance that the growth rates will be plus or minus  $2 \times \text{SD} = 10\%$ , or from earning 16% to losing 4%. We will start with the assumption that the returns of the two stocks are independent, and then experiment with correlated returns to see

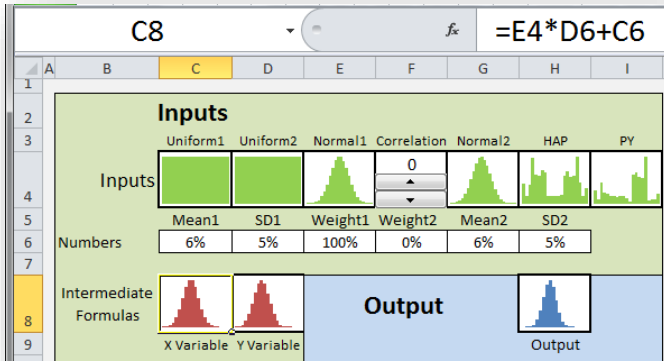
how the risk and return of the portfolio changes. Set up SIPmathCalc\_4 as described below.

1. Fill out rows 5 and 6 of SIPmathCalc\_4 as shown.



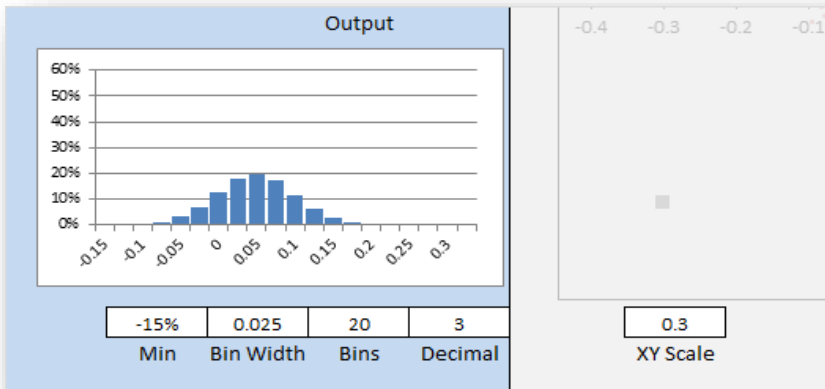
Start with mean and SD of the return of stock 1. Then enter 100% for the weight (percentage) of stock 1 in your portfolio. Enter the formula shown in cell F6 for the weight of the second stock. This will ensure that the sum of the two weights will always be 100%. Finally enter the mean and SD for stock 2.

2. The X and Y variables in C8 and D8 will model the returns of the two stocks. We start with cell E4, which contains a standard normal, that is, a normal with a mean of zero, and SD of 1. Recall from chapter 1 that to get a distribution with SD of 5% we can simply multiply E4 by 5%. Because the mean was zero to begin with, multiplying it by any number leaves it at zero. Adding 6% to E4 brings us to the desired mean return while leaving SD at 5%. The formula in C8 is shown below.



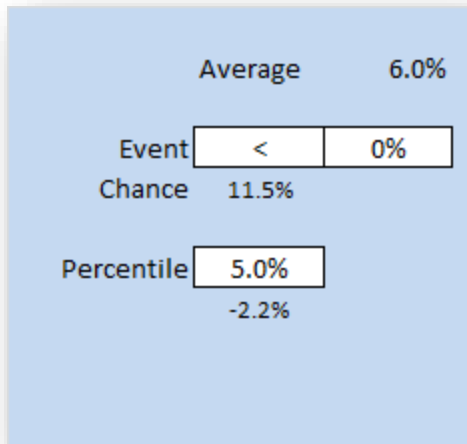
Similarly, the Y variable in D8 will model the second stock as  $=G4*H6+G6$ .

3. Now specify the output in cell H8 as the weighted sum of the two returns, that is,  $=E6*C8+F6*D8$ .
4. Set the graph scale parameters as shown below.
5. Note the average return is 6% as we would expect with 100% of our portfolio in stock 1. Finally, specify statistics as shown below.



Enter a “less than” sign into C14, and 0 into D14. This shows that the chance the stock loses money (a return of less than 0) is 11.5%.

Enter 5% in cell C17. The resulting -2.2% indicates that there is a 5% probability (or 1 chance in 20) that a dollar invested in stock 1 will become a loss of \$0.022 or more.



### The Diversification Effect

Now repeatedly reduce the percentage of stock 1 in the portfolio (E6) by 10% until you reach 0%. Cell F6 should simultaneously go from 0% to 100%. Observe the average return and chance of losing money with the resulting portfolio (C15) along with the 5<sup>th</sup> percentile (C18). Also see how the histogram of the output changes. For which portfolios (combinations of weights) is the chance of losing money 5% or less?

We see that when uncorrelated investments with the same average return are blended, that the average return of the portfolio stays the same, but the chance of loss goes down. This confirms the proverb about not putting all your eggs in one basket.

### **The Correlation Effect**

Set weight 1 to 50%, which minimizes the chance of loss at 4.5%. Now incrementally change the correlation between +1 and -1, while observing the chance of losing money and the histogram. Positive correlation increases the risk of loss, while negative correlation decreases the risk of loss. All the while, the average portfolio return remains at 6%. Thus, portfolio theory teaches that we reduce the risk of a portfolio by reducing the correlation among its constituent assets.

## ***4.4 Option Pricing***

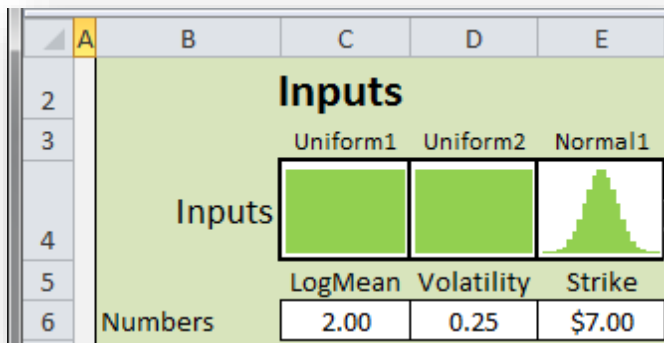
An option is a financial contract that gives the owner the right, but not the obligation, to buy or sell a share of stock at a pre-specified price known as the **Strike Price** on a pre-specified date known as the **Maturity**. The right to buy is a **Call Option**, while the right to sell is a **Put Option**. Until 1973, the fair market price for options was not well understood. Then a paper by Fischer Black and Myron Scholes and another by Robert Merton provided a theoretical approach for pricing options. This unleashed tremendous growth in the market for financial derivatives.

The interrelationships involved in option pricing are not between the returns of different stocks, as they are in portfolio theory, but rather between the value of an underlying stock and an option on that stock. We will use *SIPmathCalc\_4* to examine some important aspects of the

arithmetic of option pricing. Note that you will need to reformat certain cells such as C6 and E6 to display decimal numbers or dollar values rather than percentages. A full explanation of option theory requires an understanding of the time value of money or NPV, which is not presented here. Here we will focus on the interrelationships between the payoff of an option and its underlying stock to show how the value of the option is effected by the strike price and volatility.

Set up *SIPmathCalc\_4* as described below.

1. Fill out rows 5 and 6 of *SIPmathCalc\_4* as shown.

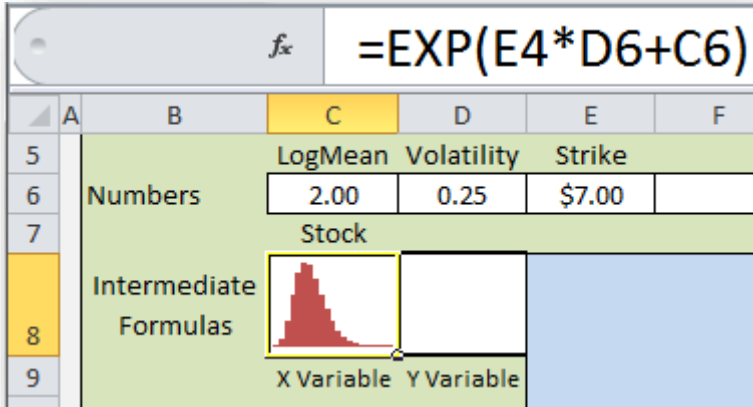


The LogMean and Volatility determine the distribution of the future price of the underlying stock as shown in the next step, while the Strike is the price at which the option holder may purchase the stock.

2. Cell C8 will contain the underlying stock price at maturity, specified as a “lognormal” distribution. You should recognize  $E4 * D6 + C6$  to be a normal distribution with a mean of C6 and SD of D6, as in the portfolio example above. If you are not



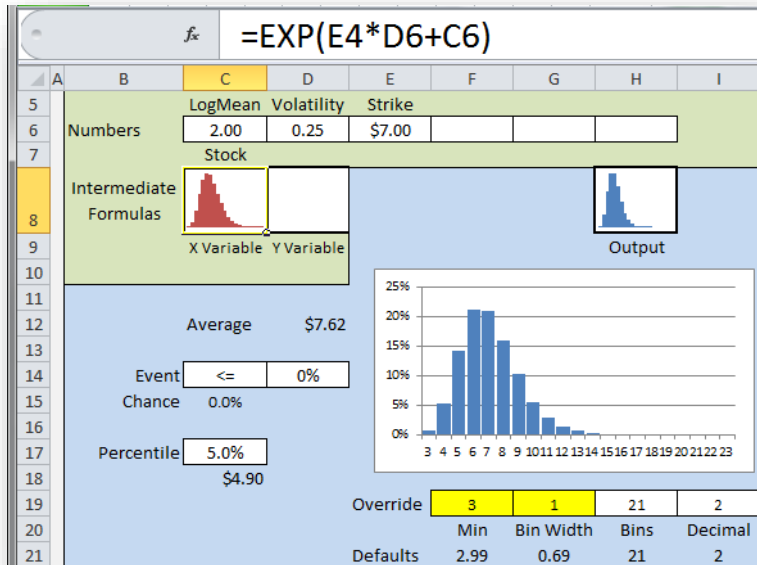
familiar with the EXP (exponentiation) formula, you may simply refer to the graph in C8, which is known as a lognormal distribution.





For those who do understand the concept of exponentiation, this formula raises a number with special mathematical properties, called  $e$ , to a power drawn from a normal distribution. Because the log of the resulting variable is normal, it is called lognormal. I find this nomenclature confusing. It would be like calling me dognormal because I have a normal dog!

In any event, just as normal distributions arise when uncertainties are added together, lognormal distributions typically arise when uncertainties are multiplied together. In the case of stock prices, it is the daily percentage increases or decreases in value that are multiplied through time to get the current price.

Set the Output cell to =C8 and adjust the override values to display the full histogram and average stock price as shown below.



Be sure to format the Average in D12 and the Percentile in C18 as \$0.00 (Currency). Experiment with the LogMean and Volatility parameters to see how the shape of the distribution and average stock price respond. For example, change Volatility from .25 to .3, and C8 should go

from  with an Average of \$7.62 to  with an Average of \$7.73.

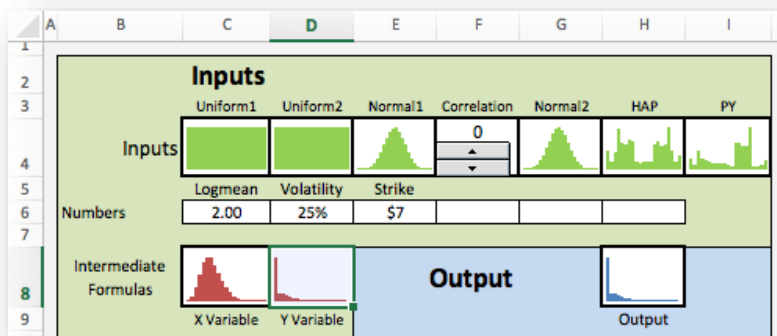
These display the characteristic “skewed” shape of the lognormal distribution.

The Minimum and Bin Width of the histogram can be overridden temporarily with the numbers 3 and 1 in cells F19 and G19 to improve the readability of the graph. Delete these values before proceeding to the next step.

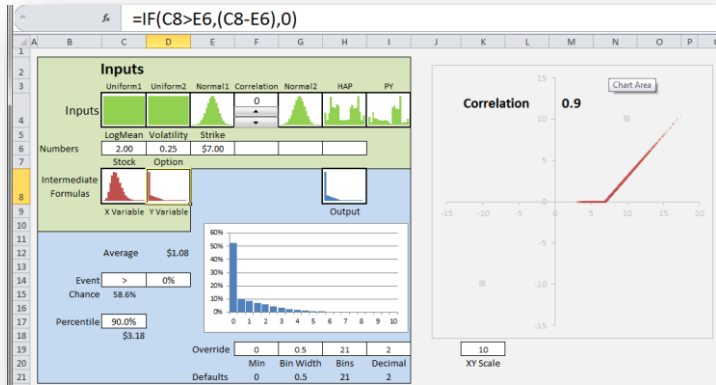
3. Now we will enter the formula for the payoff of the option in cell D8. This is an “IF” formula as follows.

$$=IF(C8>E6, (C8-E6), 0)$$

- a. If the stock price in C8 is greater than the strike price in E6, then the option will be exercised. That means the option holder will purchase the stock at the Strike price in E6 and sell it at the price in C8, pocketing the difference (C8-E6).
- b. If the stock price is not greater than the Strike price, then the option does not pay off, and it is worth nothing.



4. Set the Output cell equal to the Option payoff and set the XY Scale to 10.



## Histogram

The histogram reflects the fact that there is a high chance that the option will have a very low value, often zero, but never below zero. There is also a small chance of quite a large value.

## Statistics

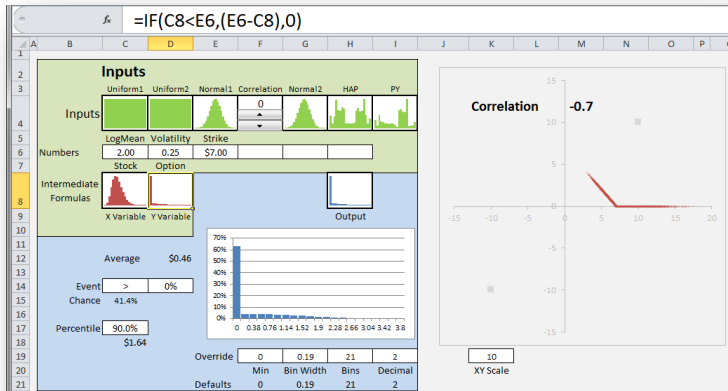
Set cell C14 to ">" and D14 to 0. The average value of the option is \$1.08 as shown in D12. The option pays off (has value >0) if stock ends up greater than the Strike price. The chance of this is nearly 60% as shown in C15.

Set C17 to 90%. C18 shows there is a 90% chance that the payoff is less than or equal to \$3.18.

## XY Chart

The scatter plot of the option vs. the stock looks like a hockey stick. This is because when the stock price (on the X axis) is below the Strike price, the option is worth zero, and for every dollar the stock moves above the strike price, the value of the option increases by a dollar.

5. Play with the model. Adjust the Strike price and observe how the Histogram, Average payoff, Chance of payoff, and XY Chart change. Can you explain the changes you see?
  
6. We will now convert the model to a Put, that is, an option that provides the holder with the right but not the obligation to *sell* the stock at the Strike price. Exercising this option only makes sense when the stock price is less than the Strike. In fact, a Put option may be viewed as sort of a fire insurance policy on a stock, which protects against loss if the stock drops in value. Change the formula in D8 as shown below, and again play with the inputs to gain an experiential understanding of the mathematical relationships.



## 4.5 Summary

Although this chapter focused on two aspects of finance, there are many other areas of commerce where interrelated uncertainties play an important role.

For example, in planning for natural disasters, it is important to make sure that multiple adverse events are not interrelated through a common root cause. A classic example involved the disaster at the Fukushima nuclear reactor in Japan following the earthquake in 2011. The resulting tsunami breached the sea wall and took out the electricity powering the cooling system. Planning ahead, the plant had backup diesel generators for just such power interruptions. It is too bad they put them in a place that would flood when they were needed most. Thus, the chances of both primary and backup power systems were positively interrelated.

---

<sup>i</sup> Seife, Charles (2000) *Zero: The Biography of a Dangerous Idea*. Penguin Books.

<sup>ii</sup> Savage, Sam L. (2009). *The Flaw of Averages – Why We Underestimate Risk in the Face of Uncertainty*. John Wiley.

<sup>iii</sup> Distribution Processing and the Arithmetic of Uncertainty, Sam L. Savage, *Analytics Magazine*, November/December 2012, <http://viewer.zmags.com/publication/90ffcc6b#/90ffcc6b/29>

<sup>iv</sup> Sam Savage, Stefan Scholtes, and Daniel Zweidler, *Probability Management, OR/MS Today*, February 2006, Volume 33 Number 1

<sup>v</sup> Taleb, Nassim N. (2007). *The Black Swan*. Random House.