36106 Managerial Decision Modeling
Monte Carlo Simulation in Excel: Part III

Kipp Martin
University of Chicago
Booth School of Business

November 15, 2017
Reading and Excel Files

Reading:

- Powell and Baker: Chapter 14.6-14.8
- Simulation Elections Using @RISK – Handout link for Week Nine.

Files used in this lecture:

- distributions.xlsx
- confidence_interval.xlsx
- markowitzCorrelation.xlsx
- markowitzCorrelation_key.xlsx
Learning Objectives

1. Learn how to select among various probability distributions

2. Learn how to select a distribution based on data

3. Learn how to build a confidence interval on simulation outputs

4. Learn how to generate correlated random variables
Lecture Outline

Motivation

Probability Distributions

Selecting a Distribution From Data

How Many Trials?

Correlation
  Stock Correlations
  Politics – Blue or Red?

St. Bernard Case

Final Offer Arbitration
Errors

We would like to know $E(f(X))$ but there are pitfalls:

1. We must estimate $f(X)$ – the function is just an estimate for reality

2. We must estimate $X$ (the random variable)

3. We must run enough trials so that we can have confidence in $f(X)$
Probability Distributions

**THE KEY CONCEPT:** Don’t use averages, use distributions!

**A PROBLEM:** if you use a woefully miscast distribution you may get bad results.

*A bad distribution may be worse than no distribution at all.*

- maybe reality is very “skewed” and you use a symmetric distribution
- maybe outcomes are very correlated and you pick independent distributions

First, address the problem of selecting a distribution. Then address the problem of correlated random variables.
Probability Distributions

Some Options:

If you have some historical data you could:

1. build a histogram and use RiskDiscrete (more on this later)

2. fit a distribution to your data (more on this later) using @RISK

3. use your data to estimate certain parameters such as a mean and standard deviation and use these as inputs to a distribution (e.g. normal)

If you do not have any data you could subjectively pick a distribution. In this case it important to understand which distribution might be most appropriate.
Probability Distributions

©RISK Distributions (there are 71)
Probability Distributions

Characteristics:

▶ Discrete versus Continuous

▶ Symmetric versus skewed (measure of asymmetry – if you must know it is the third moment about the mean)

▶ Bounded versus unbounded

▶ Nonnegative versus positive and negative values
Probability Distributions

Key Distributions:

- **Discrete**
  1. RiskDiscrete \( \{x_1, x_2, \ldots, x_n\}, \{p_1, p_2, \ldots, p_n\} \) 
  2. RiskBinomial \( (N, p) \) 
  3. RiskBernoulli(p) 
  4. Poisson(Mean)

- **Continuous**
  1. RiskUniform \( (\text{Min}, \text{Max}) \) 
  2. RiskTriang(\( \text{Min}, \text{Most Likely}, \text{Max} \)) 
  3. RiskPert(\( \text{Min}, \text{Most Likely}, \text{Max} \)) 
  4. RiskNormal(\( \text{Mean}, \text{Std Dev} \)) 
  5. RiskLognorm(\( \text{Mean}, \text{Std Dev} \)) 
  6. Exponential(\( \text{Mean} \))
RiskDiscrete \(\{x_1, x_2, \ldots, x_n\}, \{p_1, p_2, \ldots, p_n\}\):

This is a **discrete** distribution that may be **skewed**. It is **bounded** and may have **negative** values.

If you have historical data and make a histogram, you can use the histogram to produce a RiskDiscrete distribution.

See distributions.xlsx
**Probability Distributions**

**RiskBinomial** \((N, p)\): where there are \(N\) “trials” and the probability of “success” is \(p\).

This is a **discrete** distribution that may be **skewed**. It is **bounded** and it is nonnegative.

The mean of this distribution is \(pN\).

This applies when the trials are independent events, for example flipping a fair coin.

You might use this as follows: you have sold 300 tickets for a flight. In the past the probability of a no-show is .1. The random variable for the number of people showing up is binomial with \(N = 300\) and \(p = .9\).

See distributions.xlsx
**RiskBernoulli (p):** The values of the random variable are either $x = 1$ with probability $p$, or 0 with probability $1 - p$.

This is like the binomial when $N = 1$.

This is a **discrete** distribution that may be **skewed**. It is **bounded** and it is nonnegative.

We use this distribution in our simulation of election outcomes.

See distributions.xlsx
Probability Distributions

**RiskUniform (Min, Max):** “Equally likely” events in an interval.

- A continuous distribution
- A bounded distribution
- May assume positive or negative values

See distributions.xlsx
Probability Distributions

RiskPert (Min, Most Likely, Max) and RiskTriang (Min, Most Likely, Max): where Min is the minimum possible value, Max is the maximum possible value, and Most Likely, is well, most likely.

Both are continuous distributions and may be asymmetric. They are bounded and may take on negative values.

The Pert distribution put less emphasis on extreme values.

See distributions.xlsx where the triangle and pert are superimposed.

These are good distributions to use when:

▶ You want an asymmetric distribution (although Triangle and Pert can be symmetric).

▶ You do not have historical data, but have subjective guesses as to minimum, maximum, and most likely values.
Probability Distributions

The mean of the triangle distribution is \( (a + b + c)/3 \). All values weighted equally.
The mean of the Pert distribution is \( (a + 4b + c)/6 \).
Probability Distributions

RiskNormal(Mean, Std. Dev.):

A continuous distribution that is symmetric. It is unbounded and may take on negative values.

Good to use when Central Limit Theorem applies.

Central Limit Theorem applies when you are summing independent random variables with well defined mean and variance.
Probability Distributions

RiskLognorm(Mean, Std. Dev.):

A continuous distribution that is asymmetric. It is unbounded but does not take on negative or zero values. The lognormal random variable is given by

\[ X = e^{(\mu + \sigma Z)} \]

where \( Z \) is a standard normal variable. For this random variable

\[ E(X) = e^{(\mu + \sigma^2 / 2)} \]

\[ Var(X) = (e^{\sigma^2} - 1)e^{(2\mu + \sigma^2)} \]
Probability Distributions

The lognormal is often used in finance to model stock prices.
If you use continuous compounding and then take the log of the ratio of
stock prices you get normally distributed returns.

\[ P_t = P_0 e^{\left((r-0.5\sigma^2)t + \sigma t^{0.5} Z\right)} \]

- \( P_0 \) is the stock price at time 0
- \( P_t \) is the estimated stock price at time \( t \)
- \( r \) is the mean continuous compounding growth rate (estimate from data)
- \( \sigma \) is the continuous compounding growth rate standard deviation (estimate from data)
- \( Z \) is the Normal(0,1)
Using the expression for Expected value of the lognormal, one can show:

\[ E(P_t) = P_0e^{rt} \]

See spreadsheet lognormal in the distributions.xlsx workbook. We estimate a stock price six periods out, two ways.

1. First way: generate values of N(0,1) and plug into the exponential formula

\[ P_t = P_0e^{((r-0.5\sigma^2)t+\sigma t^{0.5}Z)} \]

2. Second way: generate lognormal random values based on \( E(x) \) and \( Var(X) \) where

\[ X = e^{((r-0.5\sigma^2)t+\sigma t^{0.5}Z)} \]
\[ P_t = P_0X \]
Simulate the stock price two ways.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>rate--r</td>
<td>1.25%</td>
</tr>
<tr>
<td>sigma</td>
<td>10.00%</td>
</tr>
<tr>
<td>t</td>
<td>6</td>
</tr>
<tr>
<td>Initial Price</td>
<td>70</td>
</tr>
<tr>
<td>First Way</td>
<td></td>
</tr>
<tr>
<td>Normal (0, 1)</td>
<td>-0.50899719</td>
</tr>
<tr>
<td>exp argument</td>
<td>-0.07967834</td>
</tr>
<tr>
<td>new price normal</td>
<td>64.63893269</td>
</tr>
<tr>
<td>Second Way</td>
<td></td>
</tr>
<tr>
<td>lognormal mu</td>
<td>0.045</td>
</tr>
<tr>
<td>lognormal sigma</td>
<td>0.244948974</td>
</tr>
<tr>
<td>lognormal mean</td>
<td>1.077884151</td>
</tr>
<tr>
<td>lognormal std dev</td>
<td>0.26803697</td>
</tr>
<tr>
<td>new price lognormal</td>
<td>93.72709334</td>
</tr>
</tbody>
</table>
Probability Distributions

Result from using N(0,1) in exponential formula.
Probability Distributions

Result from using lognormal distribution.
Two important distributions often used in *queuing* (call centers, banks teller lines, etc.) are

1. Poisson – model arrival rates (may lead to fishy results (ugh!))

2. Exponential – model service times
You can even create your own.

We did this for the Mergers and Acquisitions case.

You can put them on a corporate SQL Server.
Selecting a Distribution From Data

You can use @RISK to pick a distribution for you based on your data.

It does a *goodness of fit test* based on the long laundry list of distributions supported by @RISK.

Best of all, it is an easy process. We illustrate with the **fitting** spreadsheet in the **distributions** Workbook.
Selecting a Distribution From Data

The numbers in E3:E30 were generated by a normal distribution with mean 0 and standard deviation 10. This is what we are trying to fit below.
Selecting a Distribution From Data

@RISK fits the normal distribution with a uniform with max 14.123 and min -17.292.
Selecting a Distribution From Data

This example illustrates the difficulty of using a small sample size for fitting a distribution.

Some sample sizes may not pick up the low probability events.

In this sample, there were no realizations more than two standard deviations below the mean, and two standard deviations above the mean.

@RISK will often fit a non-symmetric triangular distribution to the normal sample data – why do you think this might happen?
How Many Trials?

See the workbook `confidence_intervall.xlsx`.

Interval estimate of a population mean.

\[ \bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \]

- \( \bar{X} \) is the sample mean
- \( s \) is the sample standard deviation
- \( n \) is the sample size
- \( t_{\alpha/2} \) is the \( t \) value for a \( (1 - \alpha) \) confidence interval with \( n - 1 \) degrees of freedom (see next slide for appropriate Excel function)
How Many Trials?

Use Excel function $\text{T.INV.2T}$ to calculate $t_{\alpha/2}$.

See cell F32. It has the formula

$$=\text{T.INV.2T}(\text{alpha},$C32-1)$$
How Many Trials?

Consider the new product introduction model from Homework 6.

The expected profit $E(p(X, Y, Z))$ is 7,100,000.

We estimate this number through simulation.

Simulation is nothing more than sampling.

<table>
<thead>
<tr>
<th>n</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>7,084,302</td>
<td>2,473,681</td>
<td>$7,084,302 \pm 2.26 \times 2,473,681/\sqrt{10}$</td>
</tr>
<tr>
<td>100</td>
<td>7,135,080</td>
<td>2,617,859</td>
<td>$7,135,080 \pm 1.98 \times 2,617,859/\sqrt{100}$</td>
</tr>
<tr>
<td>1000</td>
<td>7,102,757</td>
<td>2,468,202</td>
<td>$7,102,757 \pm 1.96 \times 2,468,202/\sqrt{1000}$</td>
</tr>
<tr>
<td>10000</td>
<td>7,100,207</td>
<td>2,466,928</td>
<td>$7,100,207 \pm 1.96 \times 2,466,928/\sqrt{10000}$</td>
</tr>
</tbody>
</table>
How Many Trials?

<table>
<thead>
<tr>
<th>n</th>
<th>mean</th>
<th>std. dev</th>
<th>Student t</th>
<th>Lower Confidence Interval</th>
<th>Upper Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>7,084,302</td>
<td>2,473,681</td>
<td>2.26215716</td>
<td>5,314,737.21</td>
<td>8,853,866.79</td>
</tr>
<tr>
<td>100</td>
<td>7,135,080</td>
<td>2,617,859</td>
<td>1.98421695</td>
<td>6,615,639.98</td>
<td>7,654,520.02</td>
</tr>
<tr>
<td>1000</td>
<td>7,102,757</td>
<td>2,468,202</td>
<td>1.96234146</td>
<td>6,949,593.50</td>
<td>7,255,920.50</td>
</tr>
<tr>
<td>10000</td>
<td>7,100,207</td>
<td>2,466,928</td>
<td>1.96020126</td>
<td>7,051,850.25</td>
<td>7,148,563.75</td>
</tr>
</tbody>
</table>
How Many Trials?

You can create a confidence interval using RiskCIMean().

=RiskCIMean(profit,0.9,FALSE,1)
Correlation

Consider the Workbook distributions.xlsx pictured below.

<table>
<thead>
<tr>
<th>Correlation Examples</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncorrelated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal Distribution (0, 10)</td>
<td>9.6392147</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Output -- Risk Uniform (-10, 10)</td>
<td>3.8487698</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal Distribution (0, 10)</td>
<td>-3.650734</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk Output -- Risk Uniform (-10, 10)</td>
<td>1.3523502</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Correlation

**Objective:** generate a scatter diagram. Here are the steps:

**Step 1:** Insert a Risk Normal\((0,10)\) in cell B8 and a Risk Uniform\((-10,10)\) in cell B9.

**Step 2:** Generate a histogram for Cell B9 which is the Risk Output for the Risk Uniform distribution.

**Step 3:** Select a scatter plot. See next slide.
## Correlation

<table>
<thead>
<tr>
<th></th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Correlation Examples</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Uncorrelated</strong></td>
<td></td>
</tr>
<tr>
<td>Normal Distribution (0, 10)</td>
<td>1.0965948</td>
</tr>
<tr>
<td>Risk Output -- Risk Uniform (-10, 10)</td>
<td>8.2091065</td>
</tr>
<tr>
<td><strong>Correlated</strong></td>
<td></td>
</tr>
<tr>
<td>Normal Distribution (0, 10)</td>
<td>16.212806</td>
</tr>
<tr>
<td>Risk Output -- Risk Uniform (-10, 10)</td>
<td>7.3710068</td>
</tr>
</tbody>
</table>

![Generate Scatter Diagram]
**Correlation**

**Step 4:** Select Cell B8 with the Risk Normal as the second distribution for the scatter plot.
Correlation

Cell B8 is the output from a normal random variable with mean 0 and standard deviation 10.

Cell B9 is the output from a uniform random variable with minimum value -10 and maximum value 10.

These two random variables have zero correlation. A run of 1000 simulations gives the scatter plot below.
Important Takeaway: If you insert @RISK distributions in an Excel workbook, by default, they will have zero correlation.

They will have a scatter diagram like on the previous slide.

Zero correlation looks like a shotgun blast of bird shot.

On the next slide we illustrate scatter diagrams for positive and zero correlation. For now, don’t worry about how we generated these graphs.
Correlation

A correlation of 0.8 between a normal and uniform random variable. Note the **positive** slope.
Correlation

A correlation of -0.8 between a normal and uniform random variable. Note the **negative** slope.
Let \( X \) and \( Y \) be two random variables. The **covariance** between \( X \) and \( Y \) is

\[
\text{cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y))
\]

\[
= E((X \cdot Y - \mu_X \cdot Y - X \mu_Y + \mu_X \mu_Y)
\]

\[
= E(X \cdot Y) - \mu_X \cdot E(Y) - E(X) \cdot \mu_Y + E(\mu_X \cdot \mu_Y)
\]

\[
= E(X \cdot Y) - \mu_X \cdot \mu_Y - \mu_X \cdot \mu_Y + \mu_X \cdot \mu_Y
\]

\[
= E(X \cdot Y) - \mu_X \cdot \mu_Y
\]

**Note:** if \( X \) and \( Y \) have zero correlation, then

\[
E(X \cdot Y) - \mu_X \cdot \mu_Y = 0
\]

and this implies

\[
E(X \cdot Y) = \mu_X \cdot \mu_Y
\]
Correlation

Let $X$ and $Y$ be two random variables. The correlation between $X$ and $Y$ “scales” the covariance to be between -1 and 1.

$$\rho_{XY} = \text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

$$= \frac{E ((X - \mu_X)(Y - \mu_Y))}{\sqrt{E ((X - \mu_X)^2) \cdot E ((Y - \mu_Y)^2)}}$$

Given sample means $\overline{X}$ and $\overline{Y}$, that estimate $\mu_X$ and $\mu_Y$, respectively, the sample correlation coefficient for $X$ and $Y$ is

$$\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})$$

$$\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2 \cdot \sum_{i=1}^{n} (Y_i - \overline{Y})^2}$$
Correlation

Okay math nerds, how does one generate correlated random variables?

Here is the basic idea.

1. generate uncorrelated random variables

2. generate new random variables that are linear combinations of the uncorrelated random variables

3. the new random variables will be correlated and we find the right coefficients in the linear combination to give the desired correlations

Here are some links with details:

http://www.numericalexpert.com/blog/correlated_random_variables/

http://math.stackexchange.com/questions/446093/generate-correlated-normal-random-variables
Correlation

In @RISK you can generate correlated random variables.

The user inputs the correlations using the RiskCorrmat function.

There are two methods to specify correlations using RiskCorrmat.

Regardless of which method you choose, you must first use Define Distributions to insert the random variables of interest.
Correlation

Method One:

Step 1: Decide which variables you think should be correlated. In our example, cells B17 and B18 contain the result of Define Distributions

\[=\text{RiskNormal}(0,10)\]
\[=\text{RiskUniform}(-10,10)\]

Step 2: Generate a correlation matrix for these variables. The correlations typically come from historical data. I created a correlation matrix in the range correlationMatrix.

Step 3: Manually edit cells B17 and B18 and insert the RiskCorrmat argument. Link the distributions back to the correlation matrix. See cells B17 and B18 with the edited formulas.

\[=\text{RiskNormal}(0,10,\text{RiskCorrmat}(\text{correlationMatrix},1))\]
\[=\text{RiskUniform}(-10,10,\text{RiskCorrmat}(\text{correlationMatrix},2))\]
Correlation

**Important:** In the formulas below

\[
=\text{RiskNormal}(0,10,\text{RiskCorrmatrix}(\text{correlationMatrix},1))
\]

\[
=\text{RiskUniform}(-10,10,\text{RiskCorrmatrix}(\text{correlationMatrix},2))
\]

It is important to “link” the random variable to the correlation matrix.

This is what the 1 and 2 do, respectively.
**Correlation**

**Step 4:** Run the simulation.

**Step 5:** Look at the simulation output and create a scatter diagram.

In the following slide I put -.8 into the correlation matrix.

I added an @RISK output to cell B18.

I then treated cell B17 as the input. The scatter diagram plots the output versus the input.
Correlation

Note the sample correlation is -.7798.
**Correlation**

**Method 2:** Use @RISK Define Correlations.

![Diagram of @RISK Define Correlations window](image)
Correlation

Select the random variables to correlate.

Then type in the correlations.
Creating correlated random variables is incredibly useful!

**Where we are headed:** take the Markowitz model, and for a given portfolio, simulate returns based on the stock correlations.
Stock Correlations

Open Workbook markowitzCorrelation.xlsx. In our the Markowitz optimization model there are three random variables:

- the monthly return on Apple stock – denote by $r_X$
- the monthly return on AMD stock – denote by $r_Y$
- the monthly return on Oracle stock – denote by $r_Z$

There is a constraint on expected return. Let $X$ denote the percentage of the portfolio invested in Apple, $Y$ denote the percentage of the portfolio invested in AMD, and $Z$ denote the percentage of the portfolio invested in Oracle.

The unity constraint is

$$X + Y + Z = 1.$$
The unity constraint is straightforward and does not involve random variables.

However, the required return constraint does involve stochastic parameters (random variables). The required return constraint is

\[ E(r_X X + r_Y Y + r_Z Z) \geq .12. \]

We replaced this constraint with

\[ E(r_X X) + E(r_Y Y) + E(r_Z Z) \geq .12. \]

Why is this a legitimate thing to do? What is \( f(r_X, r_Y, r_Z) \)?
Stock Correlations

Here is the Markowitz optimal solution: $X \approx .12$, $Y = 0$, and $Z \approx .88$. 

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Variance Covariance Matrix</td>
<td>0.0047</td>
<td>-0.0111</td>
<td>0.0058</td>
<td></td>
</tr>
<tr>
<td></td>
<td>AAPL</td>
<td>AMD</td>
<td>ORCL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>AAPL</td>
<td>0.0047</td>
<td>-0.0111</td>
<td>0.0058</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>AMD</td>
<td>-0.0111</td>
<td>0.4341</td>
<td>0.1120</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ORCL</td>
<td>0.0058</td>
<td>0.1120</td>
<td>0.2134</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Mean Returns</td>
<td>-0.0501</td>
<td>0.0359</td>
<td>0.1425</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Investment Level</td>
<td>0.116921314</td>
<td>0.0359</td>
<td>0.88307969</td>
<td></td>
</tr>
</tbody>
</table>
The expected return of the portfolio is indeed, 12%.

But what returns can we expect?

Let’s run a simulation of the portfolio based on the values of $X = .12$, $Y = 0$, and $Z = .88$.

In order to do this, we must generate realizations of stock returns that are correlated!

Let’s examine the simulation model starting in row 21 of the spreadsheet markowitzCorrelation.xlsx.
**Objective:** Simulate portfolio returns.

<table>
<thead>
<tr>
<th>Simulation Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Correlation Matrix</strong></td>
</tr>
<tr>
<td>AAPL</td>
</tr>
<tr>
<td>AMD</td>
</tr>
<tr>
<td>ORCL</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standard Deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AAPL</strong></td>
</tr>
<tr>
<td>0.068287379</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fitted</strong></td>
</tr>
<tr>
<td><strong>Normal</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Portfolio Return Fitted</th>
<th><strong>AAPL</strong></th>
<th><strong>AMD</strong></th>
<th><strong>ORCL</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Portfolio Return Normal</td>
<td><strong>AAPL</strong></td>
<td><strong>AMD</strong></td>
<td><strong>ORCL</strong></td>
</tr>
<tr>
<td>-0.180149272</td>
<td>-0.256432928</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Step 1: Calculate the correlation matrix based on the sample returns using the formula below for each pair of random variables.

\[
\frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}
\]

Use the Excel function \text{CORREL(Range1, Range2)}. 
Stock Correlations

The returns for the three stocks are in the returns spreadsheet.

For example, Cell D25 contains the correlation between Apple and Oracle. The formula is:

\[ =\text{CORREL(apple\_returns, orcl\_returns)} \]

For example, Cell C27 contains the correlation between Oracle and AMD. The formula is:

\[ =\text{CORREL(orcl\_returns, amd\_returns)} \]
Stock Correlations

Step 2: Now decide on the distribution for each of the random variables $r_X$, $r_Y$, and $r_Z$. We took two approaches:

- Use @RISK **Distribution Fitting** to fit a distribution to the return data in the spreadsheet **returns**. This gives
  
  1. For Apple the fitted distribution of returns is
     \[ \text{RiskUniform}(-0.19809, 0.082945) \]
  2. For AMD the fitted distribution of returns is
     \[ \text{RiskUniform}(-1.3596, 1.2825) \]
  3. For Oracle the fitted distribution of returns is
     \[ \text{RiskExtvalue}(-0.074098, 0.37482) \]

See Row 34

- Assume a normal distribution with mean and standard deviation set to the sample mean and standard deviation.

See Row 35
Stock Correlations

**Step 3:** Now add the RiskCorrmat function to each distribution. For example for the Apple fitted distribution we have

\[=\text{RiskUniform}(-0.19809, 0.082945, \text{RiskName("AAPL")}, \text{RiskCorrmat(correlationMatrix,1,"fitted")})\]

For the Oracle normal distribution we have

\[=\text{RiskNormal(orcl_mean, orcl_std_dev, RiskName("ORCL")}, \text{RiskCorrmat(correlationMatrix,3,"normal")})\]

**Important:** we are using the same correlation matrix to generate the realizations for two distinct sets of random variables.

**Step 4:** Run the simulation and look at the distribution of returns that are calculated in cells B38 and B39.
Argle Bargle! 45% of the time we experience a negative return for the fitted random variables! What percentage of the returns are below .12?
Stock Correlations

Argle Bargle! 38.7% of the time we experience a negative return for the normal random variables! What percentage of the returns are below .12?
Stock Correlations

Next Week:

1. Use Risk Optimizer

2. Put a constraint on the fraction of time we can have negative returns.
Correlation in Senate Races

Larry Robinson correlation example. See:


St. Bernard Case

Please see corresponding Excel file stbernardData.xlsx.

This a capstone case combining many concepts used throughout the quarter.

The theme of this case is that a municipal bond underwriter is submitting bids to a municipality.
St. Bernard Case

Step 1: The municipal bond underwriter makes a bid to the municipality of St. Bernard. This bid includes the following cash flows to St. Bernard.

- The face value of all the bonds that are going to be sold. In this case the total is $5 million.

- A premium which is the difference between the total coupon interest payments and the underwriter’s profit (or spread). More on this later.

Step 2: St. Bernard accepts or rejects the bid. Assume for now that they accept the bid of the underwriter in our case.
Step 3: Two weeks elapse and the underwriter sells the bonds in the marketplace. Assume all bonds are sold. For purposes of this case, the sole function of the underwriter is acting as a bond salesman for St. Bernard. The revenue to the underwriter is price of the bonds times the number sold. Hence the underwriter’s profit is:

bond sales revenue - 5,000,000 - premium

Step 4: from 2009-2014 the municipality pays the coupon rates on the bonds and the total face value of bonds (which is $5 million). The total cost to St. Bernard is then the face value of the bonds plus the coupon interest payments minus the premium they are paid. Hence they want total cost to be as small as possible and accept the bid that minimizes total cost.
The sales price for the coupon rate - bond maturity

<table>
<thead>
<tr>
<th>Year</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.00%</td>
<td>$1,009.64</td>
<td>$1,000.00</td>
<td>$981.63</td>
<td>$966.38</td>
<td>$947.58</td>
<td>$925.66</td>
</tr>
<tr>
<td>3.25%</td>
<td>$1,014.46</td>
<td>$1,007.07</td>
<td>$990.82</td>
<td>$977.58</td>
<td>$960.68</td>
<td>$940.53</td>
</tr>
<tr>
<td>3.50%</td>
<td>$1,019.27</td>
<td>$1,014.14</td>
<td>$1,000.00</td>
<td>$986.79</td>
<td>$973.79</td>
<td>$955.00</td>
</tr>
<tr>
<td>3.75%</td>
<td>$1,024.09</td>
<td>$1,021.21</td>
<td>$1,009.18</td>
<td>$1,000.00</td>
<td>$986.89</td>
<td>$970.27</td>
</tr>
<tr>
<td>4.00%</td>
<td>$1,028.91</td>
<td>$1,028.29</td>
<td>$1,018.37</td>
<td>$1,011.21</td>
<td>$1,000.00</td>
<td>$985.13</td>
</tr>
<tr>
<td>4.25%</td>
<td>$1,033.73</td>
<td>$1,035.36</td>
<td>$1,027.55</td>
<td>$1,022.42</td>
<td>$1,013.11</td>
<td>$1,000.00</td>
</tr>
<tr>
<td>4.50%</td>
<td>$1,038.55</td>
<td>$1,042.43</td>
<td>$1,036.73</td>
<td>$1,033.62</td>
<td>$1,026.21</td>
<td>$1,014.87</td>
</tr>
<tr>
<td>4.75%</td>
<td>$1,043.37</td>
<td>$1,049.50</td>
<td>$1,045.91</td>
<td>$1,044.83</td>
<td>$1,039.32</td>
<td>$1,029.73</td>
</tr>
<tr>
<td>5.00%</td>
<td>$1,048.19</td>
<td>$1,056.57</td>
<td>$1,056.10</td>
<td>$1,056.04</td>
<td>$1,052.42</td>
<td>$1,044.60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of Bonds</th>
<th>250</th>
<th>425</th>
<th>1025</th>
<th>1050</th>
<th>1100</th>
<th>1150</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years to Maturity</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Yield to Maturity</td>
<td>0.025</td>
<td>0.03</td>
<td>0.035</td>
<td>0.0375</td>
<td>0.04</td>
<td>0.0425</td>
</tr>
</tbody>
</table>

bond face value = 1000
St. Bernard Case

Understand:

1. yield curve
2. yield to maturity
3. coupon rate
4. face value
5. sales price
An illustration of coupon rates assigned to maturities. Note the total revenue collected by the underwriter.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Coupon Rate (%)</th>
<th>Number Of Bonds</th>
<th>Face Value ($)</th>
<th>Maturity Payment ($)</th>
<th>Sales Price ($)</th>
<th>Revenue ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>3</td>
<td>250</td>
<td>1,000</td>
<td>250,000</td>
<td>1,009.64</td>
<td>252,409.28</td>
</tr>
<tr>
<td>2010</td>
<td>4 1/2</td>
<td>425</td>
<td>1,000</td>
<td>425,000</td>
<td>1,042.43</td>
<td>443,032.40</td>
</tr>
<tr>
<td>2011</td>
<td>4 3/4</td>
<td>1025</td>
<td>1,000</td>
<td>1,025,000</td>
<td>1,045.91</td>
<td>1,072,061.33</td>
</tr>
<tr>
<td>2012</td>
<td>4 1/2</td>
<td>1050</td>
<td>1,000</td>
<td>1,050,000</td>
<td>1,033.62</td>
<td>1,085,305.69</td>
</tr>
<tr>
<td>2013</td>
<td>4 1/2</td>
<td>1100</td>
<td>1,000</td>
<td>1,100,000</td>
<td>1,026.21</td>
<td>1,128,831.75</td>
</tr>
<tr>
<td>2014</td>
<td>4 1/2</td>
<td>1150</td>
<td>1,000</td>
<td>1,150,000</td>
<td>1,014.87</td>
<td>1,167,097.60</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>5000</td>
<td>5,000,000</td>
<td></td>
<td></td>
<td>5,148,738.05</td>
</tr>
</tbody>
</table>
For the given assignment, here are the interest payments by St. Bernard.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Coupon Rate (%)</th>
<th>Number Of Bonds</th>
<th>Face Value</th>
<th>Years To Maturity</th>
<th>Interest Payment ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>3</td>
<td>250</td>
<td>1000</td>
<td>2</td>
<td>15,000</td>
</tr>
<tr>
<td>2010</td>
<td>4 1/2</td>
<td>425</td>
<td>1000</td>
<td>3</td>
<td>57,375.00</td>
</tr>
<tr>
<td>2011</td>
<td>4 3/4</td>
<td>1025</td>
<td>1000</td>
<td>4</td>
<td>194,750.00</td>
</tr>
<tr>
<td>2012</td>
<td>4 1/2</td>
<td>1050</td>
<td>1000</td>
<td>5</td>
<td>236,250.00</td>
</tr>
<tr>
<td>2013</td>
<td>4 1/2</td>
<td>1100</td>
<td>1000</td>
<td>6</td>
<td>297,000.00</td>
</tr>
<tr>
<td>2014</td>
<td>4 1/2</td>
<td>1150</td>
<td>1000</td>
<td>7</td>
<td>362,250.00</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1,162,625.00</td>
</tr>
</tbody>
</table>
St. Bernard Case

Here are the cash flows:

1. From underwriter to St. Bernard:
   - $5,000,000 (face value of the bonds)
   - premium = $5,148,738.05 - $5,000,000 - $40,000 = $108,738.05

2. From investors to underwriter:
   - $5,148,738.05 (bond sales)

3. From St. Bernard to Investors:
   - $5,000,000 (bond face value at maturity)
   - $1,162,625 (bond face interest payments)
Objective: Minimize NIC (net interest charge) for St. Bernard in order to win the bid.

For this example, the NIC is

$$\text{NIC} = 1,162,625 - 108,738.05 = 1,053,886.95$$
1. We know \( f(E(X)) \) is basically meaningless.

2. We can estimate \( E(f(X)) \) using simulation. However:

   ▶ \( E(f(X)) \) may be totally irrelevant,

   ▶ Basing a decision on \( E(f(X)) \) may lead to terrible results.