

Probabilistic Modeling

How to Incorporate Probability into Financial Models

Nick DeRobertis¹

¹University of Florida
Department of Finance, Insurance, and Real Estate

September 20, 2020

Table of Contents

- 1 Motivation for Probability Modeling
- 2 Mathematical Tools for Probability Modeling
- 3 Scenario Modeling
- 4 Internal Randomness

Why Model Probability

- So far everything in our models has been deterministic
- Further, we have not explored any scenarios in our models, we have taken the base case as the only case
- Unfortunately, the real world is very random. Many possible scenarios could occur.



How to Bring Probability In

- There are a few ways we can gain a richer understanding of the modeled situation by incorporating probability
- The simplest is **scenario modeling**, in which different situations are defined with probabilities, and the result of the model is the expected value across the cases.
- Another is **internal randomness** where randomness is incorporated directly within the model logic
- Finally, **Monte Carlo simulation** treats the model as deterministic but externally varies the inputs to get a distribution of outputs.

Table of Contents

- 1 Motivation for Probability Modeling
- 2 Mathematical Tools for Probability Modeling
- 3 Scenario Modeling
- 4 Internal Randomness

Math Review: Discrete and Continuous Variables

- When something is measured numerically, it can be either a **discrete** variable, or a **continuous** variable.

Discrete Variables

$$x \in \{x_1, x_2, \dots, x_n\} \quad (1)$$

- $\{x_1, x_2, \dots, x_n\}$: A specific set of values

Continuous Variables

$$x \in \mathbb{R} \text{ or } [a, b] \quad (2)$$

- \mathbb{R} : All real numbers
- $[a, b]$: Some interval between two values or infinity

Math Review: Expected Value

- **Expected value** is the average outcome over repeated trials
- It is generally useful to get a single output from multiple possible cases

Discrete Variables

$$E[x] = \sum_{i=1}^N p_i x_i \quad (3)$$

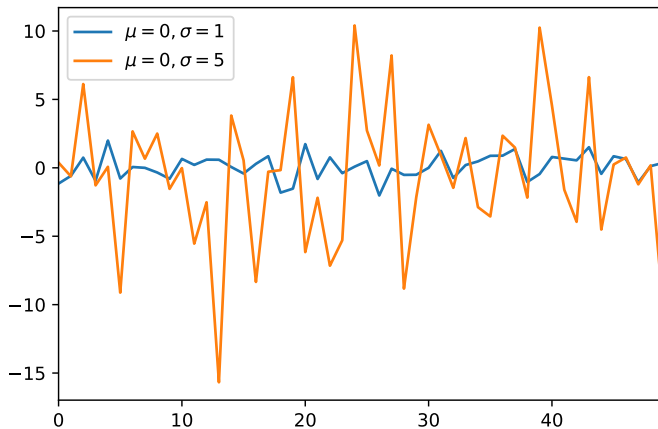
- $E[x]$: Expected value for x
- x_i : A specific value for x
- p_i : The probability associated with value x_i
- N : The total number of possible values of x

Continuous Variables

$$E[x] = \frac{1}{N} \sum_{i=1}^N x_i \quad (4)$$

- N : The number of samples collected for x

Math Review: Variance in One Picture



Math Review: Variance and Standard Deviation

- **Variance** and **standard deviation** are measures of the dispersion of values of a random variable.
- Variance is the real quantity of interest, but standard deviation is easier to understand because it has the same units as the variable, while variance has units squared

Variance of a Continuous Variable

$$\text{Var}[x] = \sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu)^2 \quad (5)$$

- N : Number of samples of x
- μ : Sample mean

Standard Deviation

$$\sigma = \sqrt{\text{Var}[x]} \quad (6)$$

- σ : Standard deviation

Math Review: Probability Distributions

- A **probability distribution** represents the probabilities of different values of a variable
- For discrete variables, this is simply a mapping of possible values to probabilities, e.g. for a coin toss, heads = 50% and tails = 50%
- For continuous variables, a continuous distribution is needed, such as the normal distribution



Math Review: Normal Distribution

- You've probably heard of the **normal distribution** as it is very commonly used because it occurs a lot in nature
- It is so common because of the **central limit theorem** which says that averages of variables will follow a normal distribution, regardless of the distribution of the variable itself
- This has many applications. For example, we can view the investment rate as an average across individual investment returns, and so it will be normally distributed.



A Non-Constant Interest Rate

- We want to extend our retirement model to say that the investment return is not constant.
- We can treat the interest rate as either a discrete (specific values) or a continuous (range of values, more realistic) variable

As a Discrete Variable

Interest Rate	Probability
2%	30%
5%	50%
7%	20%

As a Continuous Variable

$$r_i \sim N(\mu, \sigma^2) \quad (7)$$

- N : Normal distribution
- μ : Interest rate mean
- σ : Interest rate standard deviation

Table of Contents

- 1 Motivation for Probability Modeling
- 2 Mathematical Tools for Probability Modeling
- 3 Scenario Modeling**
- 4 Internal Randomness

Scenario Modeling

Interest Rate Scenarios

State of Economy	Interest Rate	Savings Rate	Probability
Recession	2%	35%	30%
Normal	5%	30%	50%
Expansion	7%	25%	20%

- In scenario modeling, different cases for model parameters are chosen. Several parameters may be altered at once in a given case.
- Here we are making the different cases the state of the economy. When the economy is doing poorly, the individual earns a lower return, but also saves more because they don't want to overspend at a bad time
- When the economy does well, the individual earns a higher return, but also spends more

Implementing Scenario Modeling

- We can implement scenario modeling **internal** or **external** to our model
- With an internal implementation, the cases are built into the model logic itself, and model logic also takes the expected value of the case outputs. The inputs of the model are now the cases and probabilities.
- With an external implementation, the model logic is left unchanged, instead the model is run separately with each case, then the expected value is calculated across the outputs from the multiple model runs.

Internal or External Scenario Analysis?

Internal	External
Original model is now an old version	Original model can still be used normally
Model runs exactly as before	Getting full results of model requires running the model multiple times and aggregating output
Model complexity has increased	Model complexity unchanged
Complexity to run model is unchanged	Complexity to run model has increased

Scenario Analysis in Excel

- For internal scenario analysis, set up a table of the cases and probabilities. Then calculate the expected value of these cases for each model parameter. Then use the expected value as the new model parameter.
- For external scenario analysis, a data table is useful. Create the data table of outputs for each case and another table of case probabilities, then combine them to produce the expected value of the output.
- If you are trying to change more than two inputs at once in external scenario analysis, this becomes more challenging but you can assign a number to each set of inputs and have the model look up the inputs based on the case number, using the case number as the data table input.

Scenario Analysis in Excel

Adding Scenario Analysis to the Dynamic Retirement Excel Model

- I will now go through adding external scenario analysis to the Dynamic Salary Retirement Model in Excel
- The completed exercise is in Examples > Scenario Analysis > Excel

Scenario Analysis Excel Lab

Adding Scenario Analysis to Project 1 - Excel

- 1 Add external scenario analysis to your Excel model from Project 1
- 2 Create three cases, for a bad, normal, and good economy. Change the initial demand and price per phone in each of the cases. Both demand and price should be higher in better economic situations.

Scenario Analysis in Python

- For internal scenario analysis, set up a DataFrame or dictionary of the cases and probabilities. Then calculate the expected value of these cases for each model parameter. Then use the expected value as the new model parameter.
- For external scenario analysis, just call your model function with each input case, collect the results, and combine them to produce the expected value of the output.

Scenario Analysis in Python

Adding Scenario Analysis to the Dynamic Retirement Python Model

- I will now go through adding external scenario analysis to the Dynamic Salary Retirement Model in Python
- The completed exercises are in Examples > Scenario Analysis > Python

Scenario Analysis Python Lab

Adding Scenario Analysis to Project 1 - Python

- 1 Add external scenario analysis to your Python model from Project 1
- 2 Create three cases, for a bad, normal, and good economy. Change the initial demand and price per phone in each of the cases. Both demand and price should be higher in better economic situations.

Table of Contents

- 1 Motivation for Probability Modeling
- 2 Mathematical Tools for Probability Modeling
- 3 Scenario Modeling
- 4 Internal Randomness**

What is Internal Randomness?

- Using the technique of **internal randomness**, something random is added internally to the model
- Instead of taking a fixed input, random values for that variable are drawn
- This technique can be used with both discrete and continuous variables

Internal Randomness in One Picture



Should I Use Internal Randomness?

- Internal randomness makes sense when the random behavior is integral to your model
- If you are just trying to see how changing inputs affects outputs, or trying to get confidence intervals for outputs, an external method such as sensitivity analysis or Monte Carlo simulation would make more sense.
- For example, if we want to allow investment returns to vary in our retirement model, an external method fits well because the core model itself is deterministic
- If instead we were modeling a portfolio, we might use internal randomness to get the returns for each asset.

Internal Randomness Advantages and Pitfalls

- Similarly to our discussion of internal vs. external sensitivity analysis, internal randomness keeps operational complexity (how to run the model) low, but increases model complexity.
- The main drawback of internal randomness is that the same set of inputs will give different outputs each time the model is run
- While this is the desired behavior, it can make it difficult to determine whether everything is working.

Internal Randomness with Continuous Variables

- Instead of taking the input as fixed, draw it from a distribution
- We need to define a distribution for each input we want to randomize. This will typically be a normal distribution, and then we just need to give it a reasonable mean and standard deviation
- Put the most reasonable or usual value as the mean. Then think about the probabilities of the normal distribution relative to standard deviation to set it



Internal Randomness with Continuous Variables in Excel

- The main functions for randomness in Excel are `=RAND` and `=RANDBETWEEN`
- The latter gives a random number between two numbers, while the former gives a random number between 0 and 1. Both of these draw from a uniform distribution (every number equally likely)
- Meanwhile, the `=NORM.INV` function gives the value for a certain normal distribution at a certain probability (it is not random)
- We can combine these two functions to draw random numbers from a normal distribution
- `=NORM.INV(RAND(), 10, 1)` would draw a number from a normal distribution with mean 10 and standard deviation 1

Example for Continuous Random Variables in Excel

Generating Random Numbers from Normal Distributions in Excel

- I will now go through generating random continuous variables in Excel
- The completed exercise is in Examples > Internal Randomness > Excel
- We will focus only on the "Continuous" sheet for now

Randomness Excel Lab

Generating and Visualizing Random Numbers - Excel

- 1 Complete the following exercise in Excel for $n=10$ and $n=1000$
- 2 Generate n values between 0 and 1 with a uniform distribution
- 3 Generate n values from a normal distribution with a 0.5 mean and 10 standard deviation
- 4 Visualize each of the two outputs with a histogram
- 5 Calculate the mean and standard deviation of each of the two sets of generated numbers
- 6 Re-calculate it a few times, take note of how much the mean and standard deviation change

Internal Randomness with Continuous Variables in Python

- In Python, we have the built-in `random` module
- It has functions analogous to those in Excel: `random.random` works like `=RAND` and `random.uniform` works like `=RANDBETWEEN`
- Drawing numbers from a normal distribution ([and other distributions](#)) is easier: just one function `random.normalvariate`
- `random.normalvariate(10, 1)` would draw a number from a normal distribution with mean 10 and standard deviation 1

Example for Continuous Random Variables in Python

Generating Random Numbers from Normal Distributions in Python

- I will now go through generating random continuous variables in Python
- The completed exercise is in Examples > Internal Randomness > Python
- We will focus only on the "Continuous" section for now

Randomness Python Lab

Generating and Visualizing Random Numbers - Python

- 1 Complete the following exercise in Python for $n=10$ and $n=1000$
- 2 Generate n values between 0 and 1 with a uniform distribution
- 3 Generate n values from a normal distribution with a 0.5 mean and 10 standard deviation
- 4 Visualize each of the two outputs with a histogram
- 5 Calculate the mean and standard deviation of each of the two sets of generated numbers
- 6 Re-calculate it a few times, take note of how much the mean and standard deviation change

Internal Randomness with Discrete Variables

- We can also build randomness into the model for discrete variables
- With discrete variables, our distribution is just a table of probabilities for the different values
- To pick a random value for a discrete variable, first add another column to your table which has the cumulative sum of the prior probabilities, and then another column which is that column plus the current probability
- Then generate a random number between 0 and 1 from a uniform distribution
- If the generated number is between the probability and the cumulative sum of prior probabilities, choose that case

An Example of Internal Randomness with Discrete Variables

Interest Rate Scenarios

State of Economy	Interest Rate	Probability	Begin Range	End Range
Recession	2%	30%	0%	30%
Normal	5%	50%	30%	80%
Expansion	7%	20%	80%	100%

- The Begin Range column is calculated as the cumulative sum of prior probabilities
- The End Range column is calculated as Begin Range + Probability
- Generate a random number between 0 and 1. If it is between the begin and end range, that is the selected value
- If it's 0.15, it's a recession. If it's 0.45, it's a normal period. If it's 0.94, it's an expansion period.

Random Discrete Variables in Python

- The steps in the preceding slides need to be carried out manually in Excel
- In Python, there is a built-in function which is doing all of this in the background, `random.choices`
- Simply do `random.choices(['Recession', 'Normal', 'Expansion'], [0.3, 0.5, 0.2])` to yield the exact same result for the prior example

Example for Discrete Random Variables in Excel and Python

Generating Random Numbers from Discrete Distributions in Excel and Python

- I will now go through generating random discrete variables in both Excel and Python
- We will be continuing with the same Excel workbook and Jupyter notebook from before which are in Examples > Internal Randomness
- We will focus only on the "Discrete" sheet/section now

Example for Discrete Random Variables in Excel and Python

Generating Random Numbers from Discrete Distributions in Excel and Python

- I will now add internal randomness with discrete variables to both the Excel and Python Dynamic Salary Retirement models to simulate economic conditions changing year by year
- The completed models are in Examples > Internal Randomness
- We will focus only on the "Discrete" sheet/section now

Internal Randomness Simple Model Lab

Building a Simple Model of Stock Returns

- 1 Create the following model in both Excel and Python
- 2 A stock starts out priced at 100. Each period, it can either go up or down.
- 3 When it goes up, it will grow by 1%. When it goes down, it will decrease by 1%.
- 4 The likelihood of the stock going up is 60%, and down 40%.
- 5 Build a model which shows how the stock price changes throughout time. Visualize it up to 100 periods and show the final price.

Internal Randomness Model Lab

Extending the Dynamic Salary Retirement Model with Internal Randomness

- 1 Add internal randomness to your Project 1 Excel and Python models
- 2 We will relax the assumption that there is a single investment return which holds for every period
- 3 Instead we will now assume that investment returns are drawn from a normal distribution
- 4 For baseline values of the inputs, you can use a 4% mean and 3% standard deviation
- 5 You should be able to run the model repeatedly and see different years to retirement each time