IB Mathematics HL Exploration

May 2018

Stochastic Processes with Applications in Finance

Table of Contents

Introduction

Rationale & Methodology

Is finance a gamble or a calculation? As someone striving to build a future in the world of finance, I have always been bugged by this question. It asks whether we can mathematically predict the outcomes of various commodity prices or should we just invest and pray it rises. How is finance not a calculation? From recent advancements in probability theory to almost every other Math HL question involving a fair coin, today we have at our disposal tools and techniques to approximate, if not predict, the outcomes of both gambles and carefully calculated investments.

In 1973, an investment banker, Burton Malkiel, wrote “A Random Walk down Wall Street,”[1] where he conceptualized the “Random walk theory”. The fact is that finance is way more complicated than a question about a fair coin as with millions of traders around the world, the behaviour and volatility of indicators are as close to random as we can get. Milkier argued that the outcome of any individual stock is not a suitable predictor of future movement, yet by understanding the general trends, we can probabilistically define an accurate range of future prices. It is reading about such theories that pushed me towards choose a topic that caters to my interest in Finance.

This exploration aims to gain an understanding of the fundamental tools and the thinking behind approximating stochastic processes. The first section introduces simple stochastic processes and the concepts slowly evolve in complexity as I start taking geometric trends, calculus and the use of simulations into consideration. Sometimes the structure, for example a discussion of the integer lattice, deviates from my initial aim of approximating stochastic processes but adds to a broader understanding of the assumptions used and is reflective of my exploratory journey.

Stochastic Process

A **stochastic process** is simply a collection of random variables indexed to time. As visualized through igure 1[2], different positions, y values, of random variables are arranged with respect to time on the x-axis. The span and development of these random positions is called the stochastic space. Therefore, an alternate and more appropriate definition is that a stochastic process is the probability distribution of paths outcomes over a space, the range of y values, at a particular time. Since a stochastic process is pegged to time, a peculiar property is that a position, even though intrinsically random, is always dependent on past behaviour. This conditional dependency allows us to predict, to a certain degree, the stochastic process’s next possible position. The more data we have, the more accurate the prediction.

Figure 1: Stochastic Process with various paths

1 Burton Gordon Malkiel, A Random Walk down Wall Street: The Time-tested Strategy for Successful Investing (New York: W.W. Norton & Company, 2007).

2 Asset Prices Simulated Using Brownian Motion, illustration, accessed December 12, 2017, http://www.turingfinance.com/wp-content/uploads/2015/04/Asset-Prices-Simulated-using-the-Brownian-Motion-Stochastic-Process.png.

Standard Brownian Processes

Simple Brownian Processes form the basis of perceiving a string of random variables as a probability distribution. In this section, I introduce way of thinking, investigate the probability of outcome through the Simple Random Walk and go beyond financial applications to explore the "Drunkard" Walk Scenario.

Brownian motion

Brownian motion is a random physical process named after Robert Brown who used it to describe zigzagged motion exhibited by a grain of pollen immersed in a liquid[3]. Since then, this abstract process has been used for modelling the constantly fluctuating prices of the stock market. A standard Brownian Process $W(t)$ is a stochastic process whereby each increment $W(t_n) - W(t_{n-1})$ where t_n is the present positon and t_{n-1} is the past position. The increment follows normal distribution about the mean and a standard deviation of $(t_n - t_{n-1})$. Furthermore, a simple random walk must commence at the origin $W(0) = 0$ and $W(t)$ is continuous. This last property is a mathematical abstraction because while financial trades are discrete from second to second, when expanded to an analysis of daily or annual fluctuations, it is reasonable to consider the increments continuous. It is appropriate to model financial securities using standard Brownian motion as long as sophisticated geometric processes do not offer greater accuracy.

Position- Time Simple Random Walk

A one-dimensional simple random walk is a specific Brownian stochastic scenario that is stationary about the origin of zero and has a constant random increment of ±1. Figure 2, can be used to describe a simple random walk. The process always starts at the origin or point Y_0. Effectively each iteration behaves like a coin toss whereby it can experience either a positive or a negative increment. In this case, the process moves to Y_1 whole the dotted lines represent other possible paths for the process to take. Therefore, Y_n can be looked at as the sum of all previous increments.

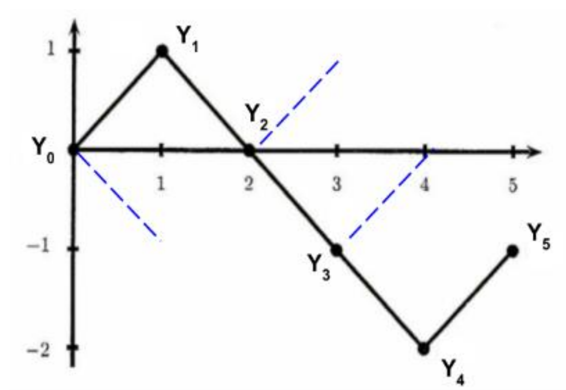


Figure 2: By the Author, A Simple Random Walk

$$Y_n = Y_0 + (Y_1 - Y_0) + (Y_2 - Y_1) + (Y_3 - Y_2) .... + (Y_n - Y_{n-1}), \qquad \text{where } (Y_n - Y_{n-1}) = Y_{increment} = Y_i = \pm 1$$

One crucial aspect of a simple random walk is that the future position Y_{n+1} does not require knowledge of the entire past but instead can be defined using the present location state Y_n and an increment. Therefore, in Figure 2, the probability of arriving at Y_5 is conditionally dependent on Y_4.

$$Y_{n+1} = Y_0 + (Y_1 - Y_0) .... + (Y_n - Y_{n-1}) + (Y_{n+1} - Y_n) \qquad \therefore Y_{n+1} = Y_n + (Y_{n+1} - Y_n)$$

3 Steven R. Dunbar, "Stochastic Processes and Advanced Mathematical Finance," reading, Department of Mathematics, University of Nebraska-Lincoln, https://pdfs.semanticscholar.org/f27e/fe3c1a7209a65565165ccb58d5d81ce69537.pdf.

Each turn’s probability is 0.5 and based on these characteristics, we know that the extreme bound scenario would be either all the way +1 or all the way -1. We can use this information to calculate the mean and the standard deviation. The probability is normally distributed. Using random increments, I computed a series of random paths on Microsoft Excel. Figure 3 is the computed paths superimposed on Desmos to show that since the variance is n, the standard deviation is $\sqrt{n}$ and hence 68% of the final position would be between $\pm\sqrt{x}$

$$E(Y_n) = E\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} E(Y_i) = 0$$

$$\sigma^2(Y_n) = \text{var}\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} \text{var}(Y_i) = n$$

$$\textit{where } Y_i \text{ is the increment of each iteration}$$

$$\therefore \text{Probability of } Y_n \sim N(0, n)$$

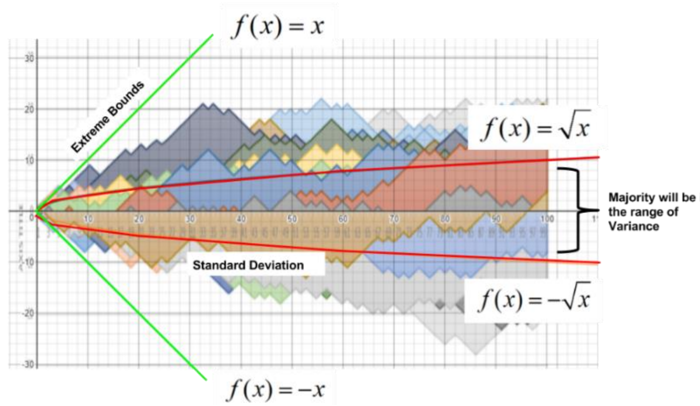
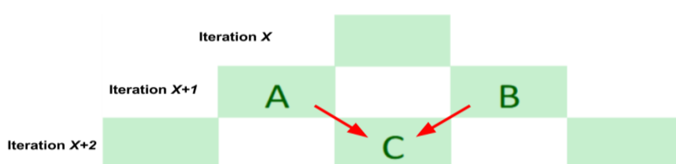


Figure 3: By the Author, Extreme Bounds and Standard Deviation

Normally Distributed Pattern – Pascal’s Triangle

After understanding the basics of simple random walks, I tried to, on paper, derive the probability of the first few iterations. The first iteration has a 50-50 chance of either positioning at +1 or -1. The next iteration has a conditional probability of either positioning at the extreme bounds with 0.25 probability each or returning to the origin with a probability of 0.5. What I noticed was that the probabilities followed a pattern, which was very “Pascaleque”, in that it was instead the reciprocal of the Pascal triangle sequence.



$$C_{x+2} = \frac{A_{x+1} + B_{x+1}}{2}$$

Figure 4: By the Author, Normally Distributed Pattern

Then, I coded this formula into Microsoft Excel, repeated for 256 iterations and then summed the individual columns since I wanted to know the overall probability of landing at a point irrespective of the iteration. A small snip-it of this operation is illustrated through Figure 5.

Zero Iteration									1								
First Iteration								0.5	0	0.5							
Second Iteration							0.25	0	0.5	0	0.25						
Third Iteration						0.125	0	0.375	0	0.375	0	0.125					
Fourth Iteration					0.0625	0	0.25	0	0.375	0	0.25	0	0.0625				
Fifth Iteration				0.03125	0	0.15625	0	0.3125	0	0.3125	0	0.15625	0	0.03125			
Sixth Iteration			0.015625	0	0.09375	0	0.234375	0	0.3125	0	0.234375	0	0.09375	0	0.015625		
Seventh Iteration		0.007813	0	0.054688	0	0.164063	0	0.273438	0	0.273438	0	0.164063	0	0.054688	0	0.007813	
	0.003906	0	0.03125	0	0.109375	0	0.21875	0	0.273438	0	0.21875	0	0.109375	0	0.03125	0	0.003906
SUM	0.003906	0.007813	0.046875	0.085938	0.265625	0.445313	0.953125	1.460938	2.460938	1.460938	0.953125	0.445313	0.265625	0.085938	0.046875	0.007813	0.003906

Figure 5: By the Author, Snip-it of Pascal Tringle Operation

As seen in Figure 6, probabilities of the 256 Iterations are summed and graphed to obtain a normally distributed curve. The differentiating factor of this graph is that since the probability through each iteration is added, we have obtained the overall probability of arriving at a given point across any iteration. Note that the curve looks pointed due to discrete nature of the points and as discussed previously, with infinite iterations, the shape will morph into a continuous bell curve.

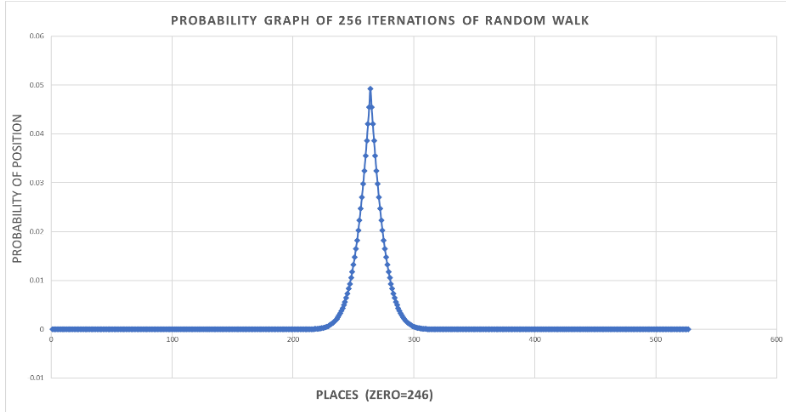


Figure 6: By the Author, Probability Graph of Pascal Triangle Pattern

This calculation of probability works for the perfect scenario whereby all possible paths are taken into account. While this is a fascinating discovery, in the real world, a perfect scenario is seldom, if ever, achieved. Therefore, I decided to check if this normal distribution of outcomes is observed when a series of random paths are graphed simultaneously. A random path formula was derived and programmed into Excel. As seen in Figure 7, I used it to plot the path of 1000 simple random walks.

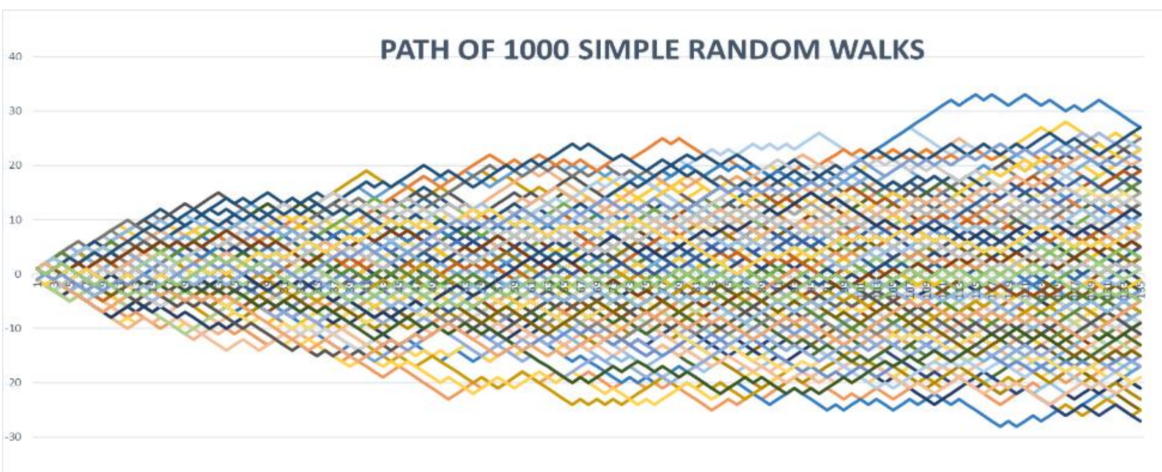


Figure 7: By the Author, Path of 1000 Simple Random Walks

Figure 8 portrays the frequency graph of the range of positions at the final iteration. Visually, the resultant resembles a bell curve with a few experimental discrepancies. I believe that with enough repeated iterations, these discrepancies would fix themselves and become more accurate.

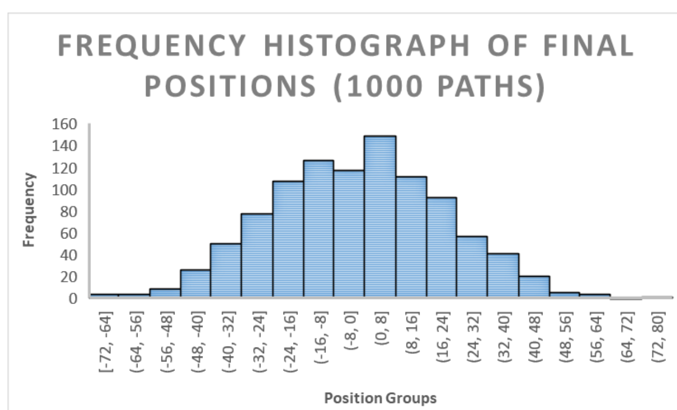


Figure 8: By the Author, Frequency Graph of 1000

Integer Lattice "Drunkard" Walk

While the Simple Random Walk previously examined was stochastic, time-bound in nature, we can also construct random walks in other dimensions. This was a tempting exploration for me as although not directly connected to my aim, it served as a catalyst for broader understanding of the subject. I use this insight and return to my direct aim in the next section.

There is a well-known mathematical joke that "If a drunk man is lost in New York, he will always find his way back… at least in infinite years". This notable scenario is referenced a simple random walk on the integer lattice. As seen in Figure 9[4], unlike the simple random walk, here time is not graphed, but instead, each step on the grid represents a time iteration. The setting is such that one can walk in either of the four directions one-step at a time and each junction, the drunk man makes a probabilistic decision of what direction to take next. Therefore, at the origin, there are four random possibilities with a probability of 0.25 each.

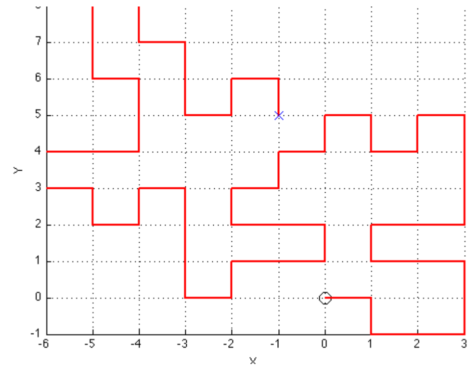


Figure 9: Integer Lattice Walk

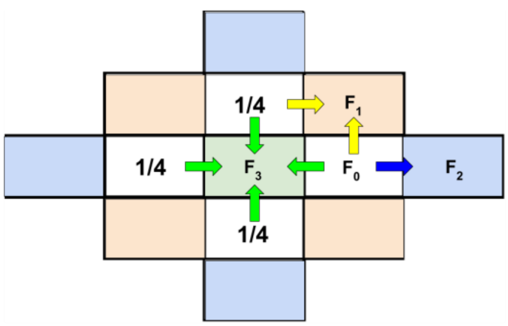


Figure 10: By the Author, Second Iteration

As shown in Figure 10, since the second iteration must take into account the first iteration's probability, we have three possible route combinations arising from the starting point F_0 . Since multiple routes converge, at F_2 and F_3 we add their probabilities.

$$\text{Random Walk}_{Green} = 4\times P(F_3 \cap F_0) = 4\times\frac{1}{16} = \frac{4}{16}$$

$$\text{Random Walk}_{Yellow} = 2\times P(F_1 \cap F_0) = 2\times\frac{1}{16} = \frac{2}{16}$$

$$\text{Random Walk}_{Blue} = 1\times P(F_2 \cap F_0) = 1\times\frac{1}{16} = \frac{1}{16}$$

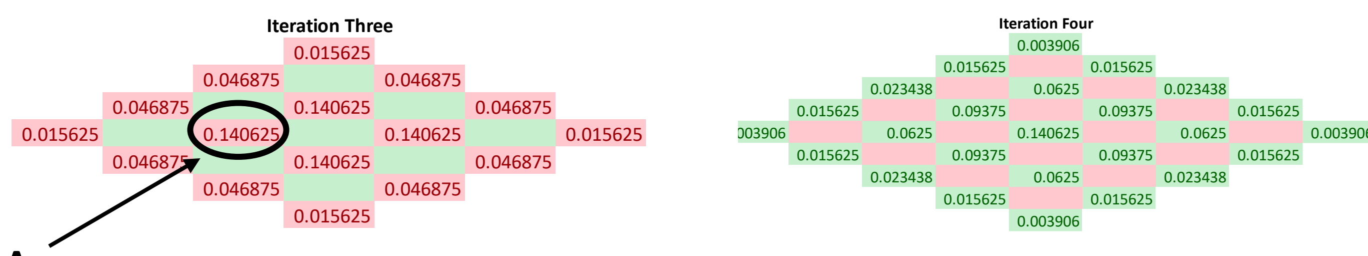


Figure 11: By the Author, Iteration Three and Four

Figure 11 shows the third and fourth iteration in the diamond diagram. This diagram is the birds eye view of the grid whereby the number, for example at A, represents the probability of arriving there at that iteration.

[4]*Random Walk 2D Avoid Simulation*, November 10, 2009, illustration, accessed December 12, 2017, https://people.sc.fsu.edu/~jburkardt/m_src/random_walk_2d_avoid_simulation/random_walk_2d_avoid_simulation.html.

This process is computed through another nine iterations. The goal is to show probabilistically that the quote about the Drunk man holds true. Furthermore, instead of taking a snapshot in time by using just one iteration, we want to find the probability of landing on the origin is as a whole. The individual point probabilities are added across iterations and then divided by the sum count so that the results add up to 1. The results are shown in the Figure 12 diamond diagram.

$$\text{Total Unit Probability of Origin} = \frac{\sum_{n=0}^{9} P(0,0)}{\text{Total Count of Probability at Sum}}$$

$$P(0,0) \approx 0.136755$$

Total Unit Sum of All Probability

								1.01725E-06								
							8.13802E-06	8.13802E-06	8.13802E-06							
						2.84831E-05	5.69661E-05	9.76563E-05	5.69661E-05	2.84831E-05						
					5.69661E-05	0.000170898	0.000423177	0.000528971	0.000423177	0.000170898	5.69661E-05					
				7.12077E-05	0.000284831	0.00094401	0.001847331	0.002490234	0.001847331	0.00094401	0.000284831	7.12077E-05				
			5.69661E-05	0.000284831	0.001220703	0.003295898	0.006608073	0.008927409	0.006608073	0.003295898	0.001220703	0.000284831	5.69661E-05			
		2.84831E-05	0.000170898	0.00094401	0.003295898	0.009025065	0.018741862	0.02718099	0.018741862	0.009025065	0.003295898	0.00094401	0.000170898	2.84831E-05		
	8.13802E-06	5.69661E-05	0.000423177	0.001847331	0.006608073	0.018741862	0.042919922	0.058406576	0.042919922	0.018741862	0.006608073	0.001847331	0.000423177	5.69661E-05	8.13802E-06	
1.01725E-06	8.13802E-06	9.76563E-05	0.000528971	0.002490234	0.008927409	0.02718099	0.058406576	0.136755371	0.058406576	0.02718099	0.008927409	0.002490234	0.000528971	9.76563E-05	8.13802E-06	1.01725E-06
	8.13802E-06	5.69661E-05	0.000423177	0.001847331	0.006608073	0.018741862	0.042919922	0.058406576	0.042919922	0.018741862	0.006608073	0.001847331	0.000423177	5.69661E-05	8.13802E-06	
		2.84831E-05	0.000170898	0.00094401	0.003295898	0.009025065	0.018741862	0.02718099	0.018741862	0.009025065	0.003295898	0.00094401	0.000170898	2.84831E-05		
			5.69661E-05	0.000284831	0.001220703	0.003295898	0.006608073	0.008927409	0.006608073	0.003295898	0.001220703	0.000284831	5.69661E-05			
				7.12077E-05	0.000284831	0.00094401	0.001847331	0.002490234	0.001847331	0.00094401	0.000284831	7.12077E-05				
					5.69661E-05	0.000170898	0.000423177	0.000528971	0.000423177	0.000170898	5.69661E-05					
						2.84831E-05	5.69661E-05	9.76563E-05	5.69661E-05	2.84831E-05						
							8.13802E-06	8.13802E-06	8.13802E-06							
								1.01725E-06								

Figure 12: By the Author, Unit Sum of All Probabilities through nine Iterations

The probabilities were plotted, using Mathematica, as the height values of a series of points rising from the integer lattice. It is evident that, at any random moment of time, the probability of being at a point closer to the origin is relatively higher than away. In fact, since the probability at the origin itself is around 14%, we can observe a bell-shaped curve. It must be noted that this scenario is discrete and hence the continuous mesh graphed in Figure 13 is only for the sake of visualization. However, as we continue to plot points until the infinite iteration, we will get closer and closer to a continuous probability density function and hence achieve higher accuracy.

These forms of Brownian paths can observe everywhere in the real world. From Google's search algorithms, describing the complex biological movement to modelling the position of particles in a room. What is interesting is that although predicting the random future is impossible, we have tools such as probability graphs to make a more accurate approximation. I will be using this concept when investigating the Monte Carlo Simulations for financial analysis.

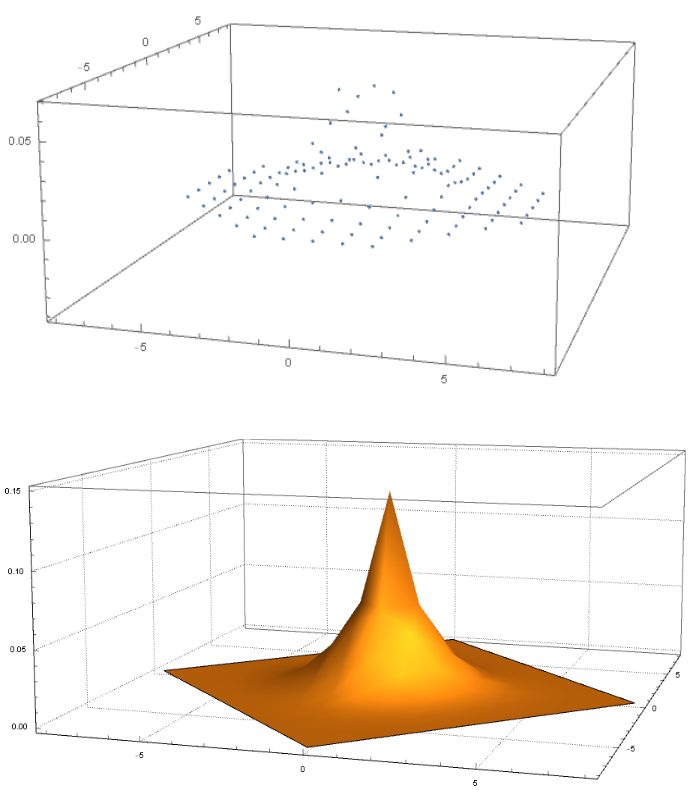


Figure 13: 3D Plot using Mathematica

Geometric Brownian Processes

Non Stationary and Biased Fluctuations

Figure 14[5] illustrates the long-term trends of the S&P 500 index, the correlation between the price of bitcoin and US dollars and the Amazon stock price. While exploring simple Brownian processes was interesting, none of the three adjacent financial indicators directly follows a simple random walk. Instead, they are irregular whereby their mean is non-stationary about the origin, and their fluctuations are inconsistent. They are positively exponential, follow long-term trends, founded on real economic factors and similar to most financial indexes are, in fact, not simple games of probability.

Figure 14: Long Term Trends of S&P 500, BTC-USD and Amazon Stock (From Above to Below)

Each complexity I added to my mental toolkit, the more I thought I was getting closer to forecasting financial indicators. This section is dedicated to an investigation into two of those theories. **Geometric Brownian motion** is defined as a continuous time stochastic process[6] that follows long-term return rates while experiences short term Brownian fluctuations.

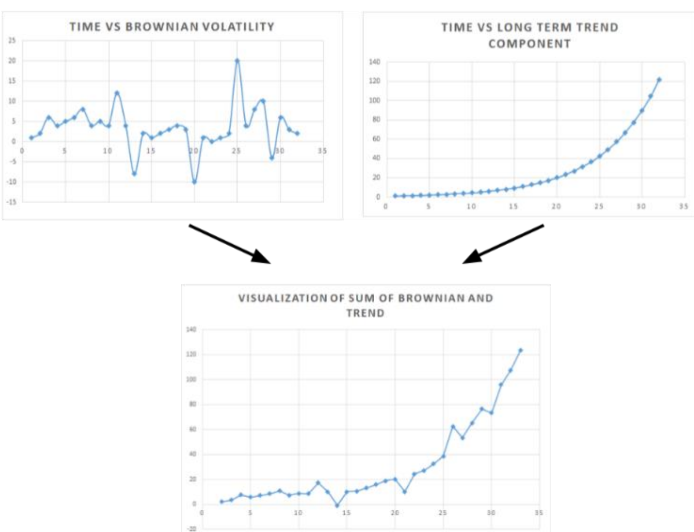


Figure 15: By the Author, Brownian and Trend Components of Geometric Brownian motion

Figure 15 and the expression below visualises the two components: A deterministically evolving mean and random fluctuations about the trend line. Here dS_T is the rate of change of a stochastic index that can be quantified using a constant percentage increase in Mean, μ, and a Brownian motion S_{Bt} in respect to time. Furthermore, the work of Kiyosi Itô, a Japanese mathematician, fascinated me as he conceptualised and later proved that using stochastic calculus and normal density functions; we can further approximate the Brownian component.

Deterministic Component that predicts the **long term trend** of the Geometric Brownian Process.

The Deviation about the trend mean

$$dS_T = \mu \overbrace{S_T dt} + \sigma \underbrace{S_{Bt} dB_t}$$

The Mean Value determined by Trend

Brownian Ito Lemma Component that cannot be directly predicted but was conceptualized by Ito as a way of approximating volatility

5 *Market Index Screeners*, https://finance.yahoo.com/

6 Dunbar, "*Stochastic Processes*," reading, Department of Mathematics, University of Nebraska-Lincoln.

Introduction to Itô Calculus

Calculus is the study of rates of change. It is useful when modelling deterministic functions where the discrete daily prices can be approximated as continuous where $\int_{Min}^{Max} xp(x)dx$ can be used to define the mean and $\int_{Min}^{Max} p(x)dx$ gives us the volume of stocks traded[7]. However, the problem comes when, in most situations, we have to consider the volatile fluctuations. Using Newtonian calculus, we can infinitely zoom into a curve to derive its slope, but a Brownian process by definition is continuously random. This intrinsic randomness means that irrespective of how small the change, the fluctuations behave like fractals and cannot be determined. As a tangential exploration, I decided to derive the Îto Calculus method which was used by Kiyosi Itô to pioneer a new form of calculus. Instead of approximating linear slopes within curves, he probabilistically modelled the next position based on historical trends.

Step One: The First fundamental theorem of calculus, as visualized in Figure 16[8], states that for any continuously differentiable function:

$$f(x) = f(0) + \int_0^x f'(x)dx$$

$$df(x) = f'(x)dx$$

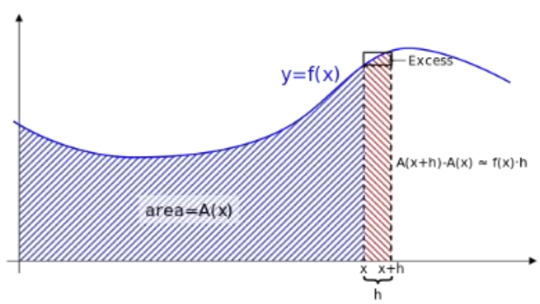


Figure 16: Visualizing the First Fundamental Theorem

One can think of dB_t as "increments of Brownian motion". Volatility is a term used to describe the frequency and magnitude of random fluctuations. While we know that the volatility is not continuously differentiable, the extension to Brownian motion is assumed valid for approximation.

Therefore, $\Delta f(B_t) = f(B_{t+\Delta t}) - f(B_s)$ where B_t represents the fluctuating variable with respect to time and B_s denotes the observed fluctuations over time.

$$\text{Equation One: } f(B_t) - f(B_0) = \int_0^t f'(B_s)dB_s$$

Step Two: The Taylor's formula can be extended to include the Brownian variables[9]

$$\Delta f(x) = f'(x)\Delta x + \frac{1}{2} f''(x)(\Delta x)^2 + \frac{1}{3!} f'''(x)(\Delta x)^3 ...$$

$$\therefore \Delta f(B_t) = f'(B_t)\Delta B_t + \frac{1}{2} f''(B_t)(\Delta B_t)^2 + \frac{1}{3!} f'''(B_t)(\Delta B_t)^3 ...$$

Similar to the assumption in Step one, Itô took the liberty of making the assumption that since the higher order derivatives get smaller over iterations, they can be neglected for approximation. Therefore,

$$\text{Equation Two: } df(B_t) = f'(B_t)dB_t + \frac{1}{2} f''(B_t)dt$$

7 "*Probability, Mean and Median*," in University of California, Department of Mathematics, accessed December 12, 2017,http://www.math.ucsd.edu/~mradclif/teaching/Math10C/LectureNotes/probability_mean_and_median.pdf.
8 "*Newton Leibniz formula and Fundamental Theorem of Calculus*," in Encyclopedia of Mathematics, last modified 2010, accessed December 12, 2017, https://www.encyclopediaofmath.org/index.php/Newton-Leibniz_formula.
9 Privault, "*Brownian Motion*," reading, Division of Mathematical Sciences, Nanyang Technological University.

Step Three: The above formula needs to be simplified. While exploring the simple random walk, I noted that $\sigma_{\text{Stochastic Process}} = \sqrt{n}$ where σ is the standard deviation and n is the number of iterations. Here, the volatility, as previous defined, is another way of expressing the standard deviation and so a change in volatility or motion is approximately equal to a change in standard deviation.

$$\Delta B_t \approx \pm\sqrt{\Delta t} \qquad\qquad (\Delta B_t)^2 \approx \Delta t$$

$$\therefore dB_t \approx \sqrt{dt}$$

Furthermore, $\dfrac{dB_t}{dt} = \dfrac{\sqrt{dt}}{dt} = \dfrac{1}{\sqrt{dt}} \approx Infinity$, this is because, $\sqrt{dt}$ is by definition infinitesimally small.

Step Four: By connecting Equation One and Equation Two previously derived and simplifying using relationships in Step Three, we have derived the Itô Formula.

$$f(B_t) = f(B_0) + \int_0^t f'(B_s)dB_s + \frac{1}{2}\int_0^t f''(B_s)ds$$

On reflection, I would consider the derivation of this Îto Calculus formula as the most satisfying component of this exploration because to grasp an understanding of the theory and connect the steps, I had to delve into lectures and notes of second-year mathematical finance at Massachusetts Institute of Technology[10]. If the formula is carefully examined, Îto was able to remove the future unpredictability from the calculations. The result is significant because it broke the misconception that financial trading was like gambling and provided a systematic tool for building new stochastic processes by just looking at the past. After verification through sophisticated simulations, this form of calculus has become useful in theory and practice and this connection between the past and the present is regarded as one of the most significant innovations of modern probability theory[11].

Furthermore, I was drawn into this exploration as I had read a book about the Black Scholes method. The method, summarized in Figure 17[12], involves the use of the Îto formula for quantifying volatility. It is widely used for deciding the prices of options in the finance world, received the Nobel Prize in 1997 and was fascinatingly employed by short sellers to predict the 2008 financial crisis.

$$d_1 = \frac{\overbrace{\ln \frac{S_0}{K}}^{\substack{x\text{ – the value} \\ \text{In this case, the value} \\ \text{above the strike price}}} + \overbrace{(r + \frac{\sigma^2}{2})(T-t)}^{\substack{\mu\text{ – mean of our normal distribution} \\ \text{In this case, a function of volatility, it reflects the} \\ \text{risk-free rate + spread of possible options over time}}}}{\underbrace{\sigma\sqrt{T-t}}_{\substack{\sigma\text{ – standard deviation} \\ \text{In this case, a function of volatility over time}}}}$$

Figure 17: The Black Scholes Method for Call Options

10 Choongbum Lee, Dr, "*Ito Calculus, Topics in Mathematics with Applications in Finance*," lecture, October 2013, MIT Open Courseware, https://ocw.mit.edu/courses/mathematics/18-s096-topics-in-mathematics-with-applications-in-finance-fall-2013/.

11 J. Michael Steele, "*Ito Calculus*," lecture, Wharton Statistics, University of Pennsylvania, accessed December 12, 2017, http://www-stat.wharton.upenn.edu/~steele/Publications/PDF/EASItoCalculus.pdf.

12 *Black-Scholes-Merton*, illustration, accessed December 12, 2017, https://brilliant.org/wiki/black-scholes-merton/.

Investigative Forecast of S&P 500 Index

In the two previous sections, we introduced the concept of stochastic processes and worked to build the complexity by incorporating geometric deterministic components and investigating the use of calculus to quantify the random fluctuations about a trend line. Now, I directly relate and apply that gained insight to archive this exploration's goal of approximating stochastic processes. I will do this by trying to forecast the future outcomes of the S&P 500 Index and evaluate the results.

Monte Carlo Simulations

Monte Carlo Simulations[13] is a method of calculation that utilizes large numbers of random samples to obtain probabilistically valuable results. They are often used in Finance and are most useful when it is impossible to use other mathematical methods. For example, in the case of forecasting Brownian components, Monte Carlo is used to calculating the expected value of future outcomes. Furthermore, although computationally time-consuming, due to the law of large numbers, with enough data, the results get closer and to the real value.

The best way to investigate a practical method like this is to practice. Therefore, I am using historical trends of the Standard and Poor's 500 Index, which is a stock index of some of the most significant companies as per market capitalisation, to construct a forecast of a week into the future. The reason I chose this is that, while individual stocks may be drastically impacted by qualitative factors like a press release by an executive, the S&P 500 as an aggregate remains mostly unaffected. Therefore, we can utilise the deterministic plus Brownian function developed by Itô and briefly discussed in the previous section.

Since time change is 1 Day, t=1

$$S_t = S_0 e^{\overbrace{(\mu - \frac{1}{2}\sigma^2)t}^{\text{Positively Exponential Deterministic Trend}} + \underbrace{(\sigma \sqrt{t} z_t)}_{\text{Brownian Component with Random Variable for Z Score whereby } 0<Z_t<1}}$$

Drift in respect to time

Here, S_t is the future price of the index while S_0 is the current. The S&P, similar to most stock indicators, is currently[14] undergoing a positively exponential trend alongside constant random fluctuations. Therefore the two values are related by an exponential constant that takes into account both the deterministic trend, as well as the volatility, Brownian fluctuations. By using historical data from the last five years to calculate this daily rate of change, we can isolate the components of the exponential constant. This will allow us to calculate values for each of the variables in the formula above and by inserting random Z scores; we will be able to forecast different random paths. This is useful because, similar to the Simple Random Walks previously discussed, by aligning the random paths together, the probabilities form a normally distributed graph.

13 "*Monte Carlo Simulation*," Finance Training Course, last modified September 1, 2011, accessed December 12, 2017, https://financetrainingcourse.com/education/2011/09/monte-carlo-simulation-simulating-returns-by-replacing-the-normal-distribution-with-historical-returns/.

14 As of December 10th, 2017.

Calculating Variables for Simulations

Days Since 2012	Closing Price	The Rate of Change
1		
2		
3		
4		
5		
6		
7	...	
8	1432.675	
9	1423.68	
10	1415.665	
11	1422.105	
12	1439.235	
13	1441.775	
14	1438.26	
15	1433.125	
16	1427.405	...
17	1422.925	0.007495
18	1412.3	0.001743
19	1409.84	-0.00183
20	1412.425	-0.02232
21	1444.31	-0.01115
...	...	...
1249	2565.145	-0.00636
1250	2581.52	0.000283
1251	2580.79	-0.00025
1252	2581.44	-0.00531
1253	2595.18	-0.00112
1254	2598.085	-0.00163
1255	2602.315	-0.00012
1256	2602.64	-0.00534
1257	2616.565	-0.00421

Step One: We use the historic data of S&P 500 to derive the exponential rate of change. The table alongside illustrates a small snipit of the initial closing price and subsequently calculated rate of change.

$$S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + (\sigma\sqrt{t}z_t)}$$

$$\ln(\frac{S_t}{S_0}) = (\mu - \frac{1}{2}\sigma^2)t + (\sigma\sqrt{t}z_t)$$

$$\ln(\frac{S_t}{S_0}) = \text{Drift} + \text{Volatility}$$

$$\Delta\text{Drift} + \Delta\text{Volatility} = \text{Rate of Change}$$

$$\text{Rate of Change} = \ln(\frac{S_t}{S_0})$$

Step Two: This rate of change can thus be used, through the expression previously discussed to break down the two components: Deterministic "Drift" and Brownian Volatility. We can therefore simplify to find the mean, standard deviation, Variance and Drift by analysing the table of rate of change. The simplified calculations for the results is presented alongside.

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} = 0.000049$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} = 3.492 \times 10^{-5}$$

$$\sigma = \sqrt{\sigma^2} = 0.00591$$

$$\text{Drift } = (\mu - \frac{\sigma^2}{2}) = 0.000474$$

Step Three: Now that we have values for all the variables we need, we can substitute them into the first equation. Note that, the Brownian component also has a Z score variable, which is still intentionally left undefined. We will assign a series of random values for this variable to forecast a range of random possibilities. For this, I used Excel's *NormsINV(Rand())* formula which in turn provided me with a random Z score between 0 and 1.

$$S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + (\sigma\sqrt{t}z_t)}$$

$$\text{Since t} = 1,$$

$$S_t = S_0 e^{Drift + (\sigma z_t)}$$

Results and Verification of Forecasts

The derived formula and the defined variables are used to generate a string of random forecasted paths. Interestingly, the paths follow the general trends of the historical data yet are each different due to the Brownian component. This volatility was mimicked by using a random Z score for each path. Such a process is reiterated a thousand times using Mathematica and graphed through Microsoft Excel. The resultant graph is illustrated in Figure 18.

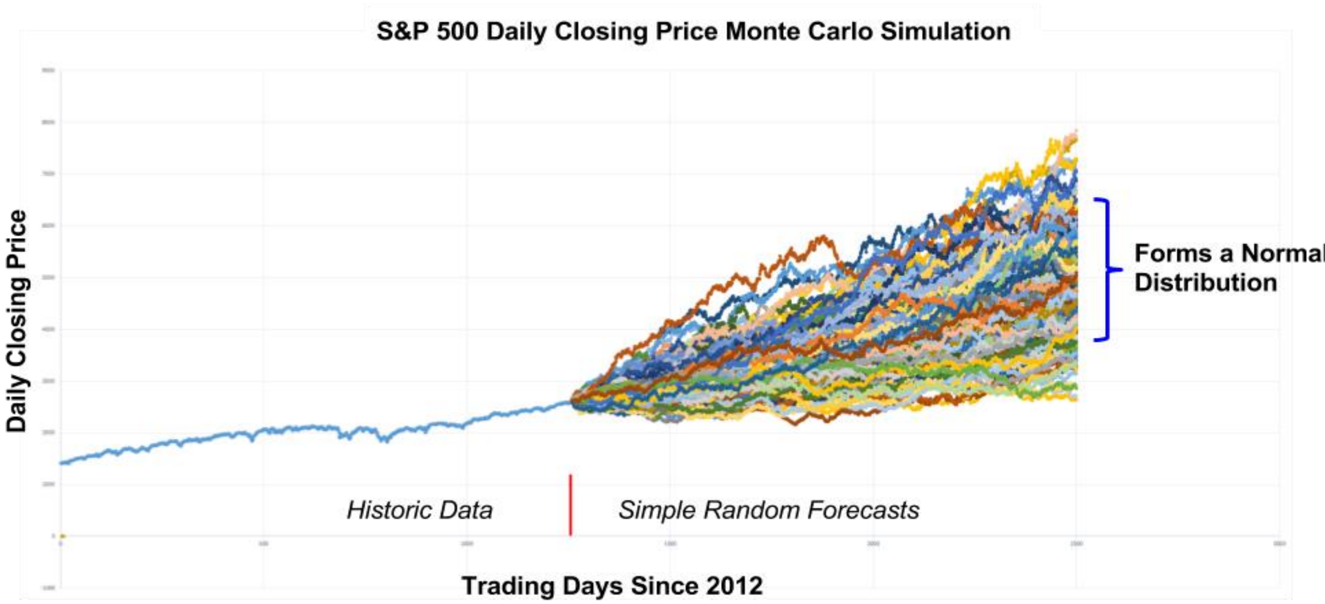


Figure 18: By the Author, Monte Carlo Simulation for S&P 500

The range of outcomes on the n^{th} day can be plotted to derive a probability distribution graph. In this case, the increments are not constant on a daily basis, and there is a general deterministic skew yet, just like the simple random walk, a normal distribution is observed. The expected value will provide us with the most likely price closing of the S&P 500 index at the n^{th} day. The more the paths used, the more accurate our result. 28th of November 2017 was the date of separation between the Historical Data and Simple Random Forecasts on the graph. To verify the accuracy of this process, I compared the predicted closing price of the following 3 days to the real life closing price[15]. The results are tabulated below and show that, over the course of the 3 days, the predicted price was very similar to the actual closing price. There was a degree of the discrepancy but within the range of expected inaccuracy.

	CLOSING DAY ONE	CLOSING DAY TWO	CLOSING DAY THREE
Expected Value of the Predicted Probability Distribution Graph ($)	2629.25	2630.70	2630.85
Actual Closing Price ($)	2,627.04	2,626.07	2,633.93
Discrepancy % $\frac{(\text{Actual - Predicted})}{Actual}\times 100$	-0.08%	-0.18%	0.12%

15 *Market Index Screeners*, https://finance.yahoo.com/.

Conclusion

Evaluation

This exploration aimed to gain an understanding of the fundamental tools and the thinking behind approximating stochastic processes. Through an inquiry into simple random walks that in turn led me to a more in-depth exploration into geometric Brownian motion, Itô calculus and conducted my financial analysis using Monte Carlo Simulations, I believe I was successful in reaching my aim. At the same time, I enjoyed the process of realising that what may seem as random may encode information that can help narrow the range of future possibilities. I would consider my investigation in the work of Itô to be the most interesting as it serves as a gateway to many of the more sophisticated theories, such as the Black Scholes method, that will be used by me in the future.

Personal Insight

Over the course of researching and writing my Math HL exploration, I have had the opportunity to reflect on specific skills and gain insight. The most important is that my previous perception of the finance being a dichotomy of gambling and calculation has evolved and led a new way of looking at the stock market. Secondly, finding ways to go to delve into Itô Calculus or expanding my finance-centric investigation to random walks that have implications in other fields such as Physics. These tangents allowed me to examine the same concept from an interdisciplinary perspective. Finally, whether it be Microsoft Excel continually crashing when trying to simultaneously graph a thousand forecasts, going through Russian Mathematica tutorials or initially visualising geometric Brownian motion, this exploration has forced me to think outside the box and exhibit problem-solving. For example, the use of Mathematica for computing the Monte Carlo Simulations served as an inspiration for expanding my random walk exploration to include a 3D plotted model.

Limitations to Accuracy

It is impossible to predict the random future, but through the tools discussed in the exploration, we can approximate the future values of stochastic processes especially in Finance. I think one of the characteristics of this investigation was that it was entirely quantitative whereby only the statistical data was used to derive forecasts. This characteristic is genuinely fascinating because finance is the interdisciplinary link between business, economics and mathematics yet even when ignoring the qualitative components, we have sound results. Therefore, it is understandable that once economic indicators such as unemployment, mergers and acquisitions or geo-political events are taken into account, the accuracy would significantly increase.

Bibliography

Asset Prices Simulated Using Brownian Motion. Illustration. Accessed December 12, 2017. http://www.turingfinance.com/wp-content/uploads/2015/04/Asset-Prices-Simulated-using-the-Brownian-Motion-Stochastic-Process.png.

Black-Scholes-Merton. Illustration. Accessed December 12, 2017. https://brilliant.org/wiki/black-scholes-merton/.

Dunbar, Steven R. "Stochastic Processes and Advanced Mathematical Finance." Reading. Department of Mathematics, University of Nebraska-Lincoln. Last modified July 15, 2016. Accessed December 12, 2017. https://pdfs.semanticscholar.org/f27e/fe3c1a7209a65565165ccb58d5d81ce69537.pdf.

Halls-Moore, Michael. "Introduction to Stochastic Calculus." QuantStart. Last modified September 4, 2012. Accessed December 12, 2017. https://www.quantstart.com/articles/Introduction-to-Stochastic-Calculus.

Lee, Choongbum, Dr. "Ito Calculus, Topics in Mathematics with Applications in Finance." Lecture, October 2013. Video file. MIT Open Courseware. https://ocw.mit.edu/courses/mathematics/18-s096-topics-in-mathematics-with-applications-in-finance-fall-2013/.

Majumdar, Satya N. "The Wonderful World of Brownian Functionals." Lecture. Laboratoire de Physique Theorique et Modeles Statistiques. Last modified July 11, 2011. Accessed December 12, 2017. https://www.pks.mpg.de/~lafnes11/Slides/Satya_Majumdar_LAFNES11.pdf.

Malkiel, Burton Gordon. *A Random Walk down Wall Street: The Time-tested Strategy for Successful Investing*. New York: W.W. Norton & Company, 2007.

Market Index Screeners. https://finance.yahoo.com/.

"Monte Carlo Simulation." Finance Training Course. Last modified September 1, 2011. Accessed December 12, 2017. https://financetrainingcourse.com/education/2011/09/monte-carlo-simulation-simulating-returns-by-replacing-the-normal-distribution-with-historical-returns/.

"Newton Leibniz Formula and Fundamental Theorem of Calculus." In *Encyclopedia of Mathematics*. Last modified 2010. Accessed December 12, 2017. https://www.encyclopediaofmath.org/index.php/Newton-Leibniz_formula.

Privault, Nicolas. "Brownian Motion and Stochastic Calculus." Reading. Division of Mathematical Sciences, Nanyang Technological University. Last modified September 20, 2017. Accessed December 12, 2017. https://www.ntu.edu.sg/home/nprivault/MA5182/brownian-motion-stochastic-calculus.pdf.

"Probability, Mean and Median." In *University of California, Department of Mathematics*. Accessed December 12, 2017. http://www.math.ucsd.edu/~mradclif/teaching/Math10C/LectureNotes/probability_mean_and_median.pdf.

Random Walk 2D Avoid Simulation. November 10, 2009. Illustration. Accessed December 12, 2017. https://people .sc.fsu.edu/jburkardt/ m_src/random_walk_2d_avoid_simulation/random_walk_2d_avoid_simulation.html.

Steele, J. Michael. "Ito Calculus." Lecture. Wharton Statistics, University of Pennsylvania. Accessed December 12, 2017. http://www-stat.wharton.upenn.edu/~steele/Publications/PDF/EASItoCalculus.pdf.

Wang, Chong. "Rare Events in Brownian Motion: A Path Integral Approach." Lecture. Department of Physics, Massachusetts Institute of Technology. Last modified May 17, 2012. Accessed December 12, 2017. http://web.mit.edu/8.334/www/grades/projects/projects12/Chong%20Wang.pdf.