O'REILLY

Probabilistic Machine Learning for Finance and Investing

A Primer to the Next Generation of AI with Python

Early Release

RAW & UNEDITED

Deepak Kanungo

Probabilistic Machine Learning for Finance and Investing

A Primer to the Next Generation of AI with Python

With Early Release ebooks, you get books in their earliest form—the author's raw and unedited content as they write—so you can take advantage of these technologies long before the official release of these titles.

Deepak Kanungo

O'REILLY®

Beijing • Boston • Farnham • Sebastopol • Tokyo

Probabilistic Machine Learning for Finance and Investing

by Deepak Kanungo

Printed in the United States of America.

Published by O'Reilly Media, Inc. , 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (http://oreilly.com). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com* .

Editors: Jeff Bleiel and Michelle Smith

Production Editor: Katherine Tozer

Interior Designer: David Futato

Cover Designer: Karen Montgomery

Illustrator: Kate Dullea

June 2023: First Edition

Revision History for the First Edition

- 2022-09-12: First Release

See http://oreilly.com/catalog/errata.csp?isbn=9781492097679 for release details.

978-1-492-09761-7

Chapter 1. The Need for Probabilistic Machine Learning

A NOTE FOR EARLY RELEASE READERS

With Early Release ebooks, you get books in their earliest form—the author's raw and unedited content as they write—so you can take advantage of these technologies long before the official release of these titles.

This will be the 1st chapter of the final book. Please note that the GitHub repo will be made active later on.

If you have comments about how we might improve the content and/or examples in this book, or if you notice missing material within this chapter, please reach out to the author at EMAIL.

A map will enable you to go from one geographic location to another. It is a very useful mathematical model for navigating the physical world. It becomes even more useful if you automate it into a GPS system using artificial intelligence (AI) technologies. However, neither the mathematical model nor the AI-powered GPS system will ever be able to capture the experience and richness of the terrain it represents. That's because all models have to simplify the complexities of the real world, thus enabling us to focus on some of the features of a phenomenon that interest us.

George Box, a statistician, famously quipped, "All models are wrong, but some are useful". This deeply insightful quip is our mantra. We accept that all models are wrong because they are inadequate and incomplete representations of reality. Our goal is to build financial systems based on models and supporting technologies that enable useful inferences and predictions for decision-making and risk management in the face of endemic uncertainty, incomplete information and inexact measurements.

All financial models, whether derived theoretically or discovered empirically by humans and machines, are not only wrong but are also at the mercy of three types of errors. In this chapter, we explain this trifecta of errors with an example from consumer credit and explore it using Python code. This exemplifies our claim that inaccuracies of financial models are features, not bugs. After all, we are dealing with people, not pendulums.

Finance is not an accurate physical science like physics dealing with precise estimates and predictions as academia will have us believe. It is an inexact social study grappling with a range of values with varying plausibilities that change continually, often abruptly.

We conclude the chapter by explaining why AI in general and probabilistic machine learning (ML) in particular offers the most useful and promising theoretical framework and technologies for developing the next generation of systems for finance and investing.

WHAT IS A MODEL?

AI systems are based on models. A model maps functional relationships among its inputs and outputs variables based on assumptions and constraints. In general, input variables are called independent variables and output variables are called dependent variables.

In high school, you learned that the equation of any line in the X-Y plane can be expressed as $y = mx + b$ where m is the slope and b is the y-intercept. For example, if you assume that consumer spending, the output/dependent variable y, has a linear relationship with personal income, the input/independent variable x, the equation for the line is called a model for consumer spending. Moreover, the slope m and the intercept b, are referred to as the model's parameters. They are treated as constants and their specific values define unique functional relationships or models.

Depending on the type of functional relationships, its parameters, the nature of inputs and outputs variables, models may be classified as deterministic or probabilistic. In a deterministic model, there are no uncertainties about the type of functional relationship, its parameters, the inputs or the outputs of the model. The exact opposite is true for probabilistic models discussed in this book.

Finance Is Not Physics

Adam Smith, generally recognized as the founder of modern economics, was in awe of Newton's laws of mechanics and gravitation [1]. Ever since then, economists have endeavored to make their discipline into a mathematical science like physics. They aspire to formulate theories that accurately explain and predict the economic activities of human beings at the micro and macro levels. This desire gathered momentum in the early 20th century with economists like Irving Fisher and culminated in the Econophysics movement of the late 20th century.

Despite all the complicated mathematics of modern finance, its theories are woefully inadequate, almost pitiful, especially when compared to those of physics. For instance, physics can predict the motion of the moon and the electrons in your computer with jaw-dropping precision. These predictions can be calculated by any physicist, at any time, anywhere on the planet. By contrast, market participants - traders, investors, analysts, finance executives - have trouble explaining the causes of daily market movements or predicting the price of an asset at any time, anywhere in the world.

THE POLITICAL ECONOMICS OF MISREPRESENTING A NOBEL PRIZE

In his will, Alfred Nobel did not create a prize in economics or mathematics or any other discipline besides physics, chemistry, medicine, literature and peace. The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel, now commonly and mistakenly referred to as the Nobel Prize in economics, was created by the Swedish Central Bank in 1968. The central bank funds the award in perpetuity and pays the Nobel Foundation to administer it like it does the Nobel prizes willed by its benefactor.

By elevating the status of economics to that of the physical sciences and by buying the ongoing support of the prestigious Nobel Foundation, the Swedish central bank was able to gain independence in its decision-making from the country’s politicians. Economic policy decisions were to be left to the economic scientists just as health policy decisions were left to medical scientists. However, by doing this, the prize disregards the will of Alfred Nobel and misrepresents the fundamental nature of economics as a social science.

Perhaps finance is harder than physics. Unlike atoms and pendulums, people are complex, emotional beings with free will and latent cognitive biases. They tend to behave inconsistently and continually react to the actions of others. Furthermore, market participants profit by beating or gaming the systems that they operate in.

After losing a fortune on his investment in the South Sea Company, Newton remarked, “I can calculate the movement of the stars, but not the madness of men.” Note that Newton was not “retail dumb money”. He served as the Warden of the Mint in England for almost 31 years, helping put the British pound on the gold standard where it would stay for over two centuries.

All Financial Models Are Wrong, Most Are Useless

Some academics have even argued that theoretical financial models are not only wrong but also dangerous; the veneer of a physical science lulls adherents of economic models into a false sense of certainty about the accuracy of their predictive powers. This blind faith has led to many disastrous consequences for their adherents and for society at large [1], [2]. Nothing exemplifies the dangerous consequences of academic arrogance and blind faith in analytical financial models than the spectacular disaster of LTCM discussed in the sidebar below.

THE DISASTER OF LONG TERM CAPITAL MANAGEMENT

LTCM was a hedge fund founded in 1994 by Wall Street veterans and academics, Merton Black and Myron Scholes, inventors of the famous Black-Scholes-Merton option pricing formula. The LTCM team was so confident in its investment models, overseen by two future 'Nobel laureates', that they leveraged their portfolios to dangerously high levels. They intended to magnify the tiny profits that LTCM was making on its various investments. In the first 4 years, LTCM had very impressive annual returns and had to turn away investor money.

However, the unpredictable complexity of social systems reared its ugly head in 1998 when the Russian government defaulted on its domestic local currency bonds. Such an event was not anticipated by LTCM's models since a government can always print more money rather than default on its debt. This shocked global markets and led to the rapid collapse of LTCM - leverage magnifies losses as it does gains. To prevent the crisis of LTCM from spreading and crashing the global financial markets, the Federal Reserve and a consortium of large banks bailed out LTCM. See figure 1-1 below which compares the value of $1000 invested separately in LTCM, Dow Jones (DJIA) and US Treasury bonds.

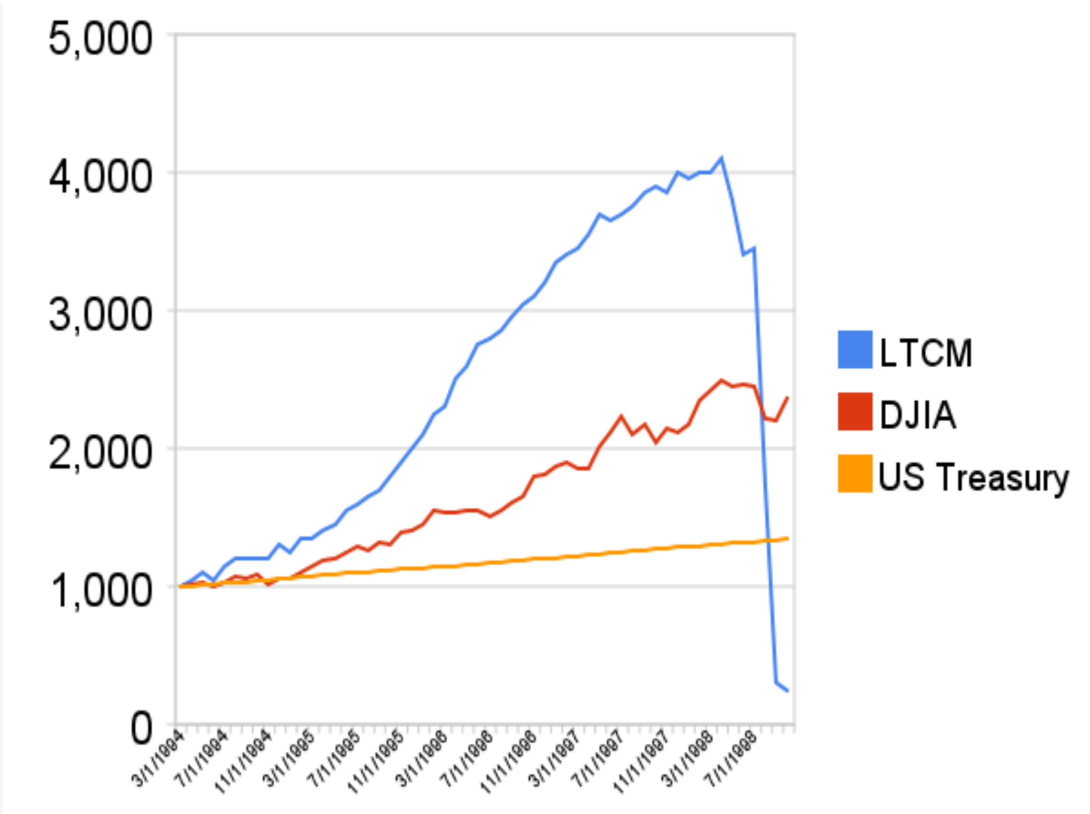

Figure 1-1. The disaster of LTCM. Image source: Wikimedia Commons

Unlike Newton, who was an intellectual giant and lost a personal fortune trading the markets, Scholes and Merton didn’t have the intellectual integrity and humility to admit to their mistakes even though they lost billions of dollars of other people’s money and nearly brought the global markets to its knees with their deeply flawed financial models.

The most successful hedge fund in history, Renaissance Technologies, has put its critical views of financial theories into practice. Instead of hiring people with a finance or Wall Street background, they prefer to hire physicists, mathematicians, statisticians and computer scientists. They trade the markets using quantitative models based on non-financial theories such as information theory, data science, and machine learning.

The Trinity of Modeling Errors

Whether financial models are based on academic theories or empirical data mining strategies, they are all subject to the trinity of modeling errors explained below. Errors in analysis and forecasting may arise from any of the following modeling issues [1],[2], [3],[4]: using an inappropriate functional form, inputting inaccurate parameters, or failing to adapt to structural changes in the market.

1. Errors in model specification

Almost all financial theories use the gaussian or normal distribution in their models. For instance, the normal distribution is the foundation upon which Markowitz's Modern Portfolio Theory and Black-Scholes-Merton Option Pricing Theory are built [1],[2],[3]. However, it is a well documented fact that stocks, bonds, currencies and commodities have fat-tailed distributions and are distinctly non-gaussian [1],[2],[3]. In other words, extreme events occur far more frequently than predicted by the normal distribution.

If asset price returns were normally distributed, none of the following financial disasters would occur within the age of the universe: Black Monday, the Mexican Peso Crisis, Asian Currency Crisis, the bankruptcy of LTCM, or the Flash Crash. "Mini flash crashes" of individual stocks occur with even higher frequency than these macro events.

Yet, finance textbooks, programs and professionals continue to use the normal distribution in their asset valuation and risk models because of its simplicity and analytical tractability. These reasons are no longer justifiable given today's advanced algorithms and computational resources. This reluctance in abandoning the normal distribution is a clear example of "the drunkard's search": a principle derived from a joke about a drunkard who loses his key in the darkness of a park but frantically searches for it under a lamppost because that's where the light is.

2. Errors in model parameter estimates

Errors of this type may arise because market participants have access to different levels of information with varying speeds of delivery. They also have different levels of sophistication in processing abilities and different cognitive biases. These factors lead to profound epistemic uncertainty about model parameters.

Let’s consider a specific example of interest rates. Fundamental to the valuation of any financial asset, interest rates are used to discount uncertain future cash flows of the asset and estimate its value in the present. At the consumer level, for example, credit cards have variable interest rates pegged to a benchmark called the prime rate. This rate generally changes in lock-step with the federal funds rate, an interest rate of seminal importance to the U.S. and the world economies.

Let’s imagine that you would like to estimate the interest rate on your credit card one year from now. Suppose the current prime rate is 2% and your credit card company charges you 10% plus prime. Given the strength of the current economy, you believe that the Federal Reserve is more likely to raise interest rates than not. Based on our current information, we know that the Fed will meet eight times in the next twelve months and will either raise the federal funds rate by 0.25% or leave it at the previous level.

In the following python code example, we use the binomial distribution to model your credit card’s interest rate at the end of the twelve-month period. Specifically, we’ll use the following parameters: fed_meetings = 8 (number of trials or meetings), prob_raises = [0.6, 0.7,0 .8, 0.9], for our range of estimates about the probability of the Fed raising the federal funds rate by 0.25% at each meeting.

```
#Import binomial distribution from scipy library
from scipy.stats import binom
#Import matplotlib library for drawing graphs
import matplotlib.pyplot as plt
#Total number of meetings of the Federal Open Market Committee
(FOMC) in any year
fed_meetings = 8
#Range of total interest rate increases at the end of the year
total_increases = list(range(0, fed_meetings + 1))
#Probability that the FOMC will raise rates at any given meeting
prob_raises = [0.6, 0.7, 0.8, 0.9]
for i in prob_raises:
  #Use the probability mass function to calculate probabilities of
total raises in eight meetings
  #Based on FOMC bias for raising rates at each meeting
```

```
  prob_dist = binom.pmf(k=total_increases, n=fed_meetings, p=i)
  #How each 25 basis point increase in the federal funds rate
affects your credit card interest rate
  cc_rate = [j * 0.25 + 12 for j in total_increases]
  #Plot the results for different FOMC probability
  plt.plot(cc_rate, prob_dist, label = i)
  plt.legend()
  plt.ylabel('Probability of credit card rate')
  plt.xlabel('Credit card interest rate')
  plt.title('Forecasts of credit card rates in 12 months based on
different probabilities')
```

In Figure 1-2, notice how the probability distribution for your credit card rate in twelve months depends critically on your estimate about the probability of the Fed raising rates at each of the eight meetings. You can see that for every increase of 0.1 in your estimate of the Fed raising rates at each meeting, the expected interest rate for your credit card in twelve months increases by about 0.2%.

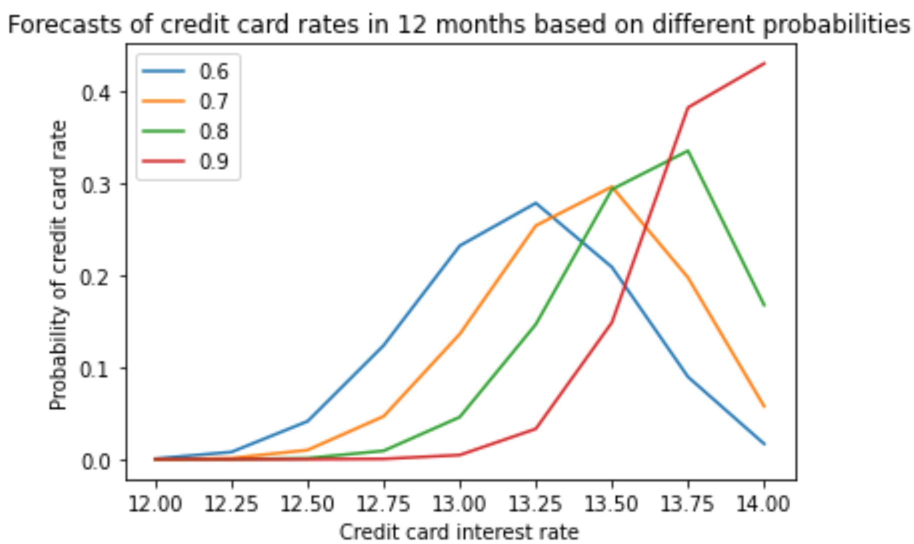

Figure 1-2. Probability distribution of credit card rates depends on your input estimates

Even if all market participants used the binomial distribution in their models, it’s easy to see how they could disagree about the future prime rate because of the differences in their estimate of Fed raising rates at each meeting. Indeed, this parameter is hard to estimate. Many institutions have dedicated analysts, including previous employees of the Fed, analyzing the Fed’s every document, speech and event to try to estimate this parameter.

Recall that we assumed that this parameter probs was constant in our model for each of the next eight Fed meetings. How realistic is that? Members of the Federal Open Market Committee (FOMC), the rate setting body, are not just a set of biased coins. They can and do change their individual biases based on how the economy changes over time. The assumption that the parameter prob_raises will be constant over the next twelve months is not only unrealistic but also risky.

3. Errors from the failure of a model to adapt to structural changes

The underlying data-generating stochastic process may vary over time — i.e. the process is not stationary ergodic. We live in a dynamic capitalist economy characterized by technological innovations and changing monetary and fiscal policies. Time-variant distributions for asset values and risks are the rule, not the exception. For such distributions, parameter values based on historical data are bound to introduce error into forecasts.

In our example above, if the economy were to show signs of slowing down, the Fed might decide to adopt a more neutral stance in its fourth meeting, making you change your probs parameter from 70% to 50% going forward. This change in your probs parameter will in turn change the forecast of your credit card interest rate.

Sometimes the time-variant distributions and their parameters change continuously or abruptly as in the Mexican Peso Crisis. For either continuous or abrupt changes, the models used will need to adapt to evolving market conditions. A new functional form with different parameters might be required to explain and predict asset values and risks in the new regime.

Suppose after the fifth meeting in our example, the US economy is hit by an external shock — say a new populist government in Greece decides to default on its debt obligations. Now the Fed may be more likely to cut interest rates than to raise them. Given this structural change in the Fed’s outlook, we will have to change the binomial probability distribution in our model to a trinomial distribution with appropriate parameters.

Probabilistic Financial Models

Inaccuracies of financial models are features, not bugs. It is intellectually dishonest and foolishly risky to represent financial estimates as scientifically precise values. All models should quantify the uncertainty inherent in financial inferences and predictions to be useful for sound decision-making and risk management in the business world. Financial data are noisy and have measurement errors. A model's functional form may be unknown or an approximation. Model parameters and outputs may have a range of values with associated plausibilities. In other words, we need mathematically sound probabilistic models because they accommodate inaccuracies and quantify uncertainties with logical consistency.

There are two ways uncertainty is quantified: forward propagation for output uncertainty and inverse propagation for input uncertainty. Figure 1-3 below shows the common types of probabilistic models used in finance today for quantifying both types of uncertainty.

Figure 1-3. Quantifying input and output uncertainty with probabilistic models

In forward uncertainty propagation, uncertainty arising from a model's inexact parameters and inputs are propagated forward throughout the model to generate the total uncertainty of the model's outputs. Most financial analysts use scenario and sensitivity analysis to quantify the uncertainty in their model's predictions. However, these are basic tools that only consider a few possibilities.

In scenario analysis, only three cases are built for consideration: best case, base case and worst case scenarios. Each case has a set value for all the inputs and parameters of a model. Similarly, in sensitivity analysis, only a few inputs or parameters are changed to assess their impact on the model's total output. For instance, a sensitivity analysis might be conducted on how the value of a company changes with interest rates or future earnings. In Chapter 3, we will learn how to do Monte Carlo Simulations (MCS) using python and apply it to common financial problems. MCS is the most powerful probabilistic numerical tool in all of the sciences and is used for analyzing both deterministic and probabilistic systems. It is a set of numerical methods that randomly samples input parameter distributions for computing the total expected uncertainty of a model, especially when its functional relationships are not analytically tractable.

In inverse uncertainty propagation, uncertainty of the model's input parameters is inferred from observed data. This is a much harder problem than forward propagation. Advanced statistical inference techniques or complex numerical computations are used to calculate confidence intervals or credible intervals of a model's input parameters. In Chapter 4, we explain the deep flaws and limitations of using p-values and confidence intervals, statistical techniques that are commonly used in financial data analysis today. Later we explain how to use advanced numerical techniques to compute credible intervals to quantify the uncertainty of a model's input parameters.

We require a comprehensive probabilistic framework that combines both forward and inverse uncertainty propagation seamlessly. We don't want the piece-meal approach that is currently in practice today. That is, we want our probabilistic models to quantify the total uncertainty in the outputs of time-variant stochastic processes with its inexact input parameters estimated from sample data.

Our probabilistic framework will need to update continually the model outputs or its input parameters - or both - based on materially new datasets. Such models will have to be developed using small datasets, since the underlying environment may have changed too quickly to collect a sizable amount of relevant data. Most importantly, our probabilistic models need to know what they don't know. When extrapolating from datasets they have never encountered before, they need to provide answers with low confidence levels or wider margins of uncertainty.

Financial AI and ML

Probabilistic machine learning (ML) meets all of the above requirements for building state-of-the-art, next generation of financial systems [5]. But what is probabilistic ML? Before we answer that question, let's first make sure we understand what we mean by ML in particular and AI in general. It is common to see these terms bandied about as synonyms even though they are not. ML is a subfield of AI. See Figure 1-4 below.

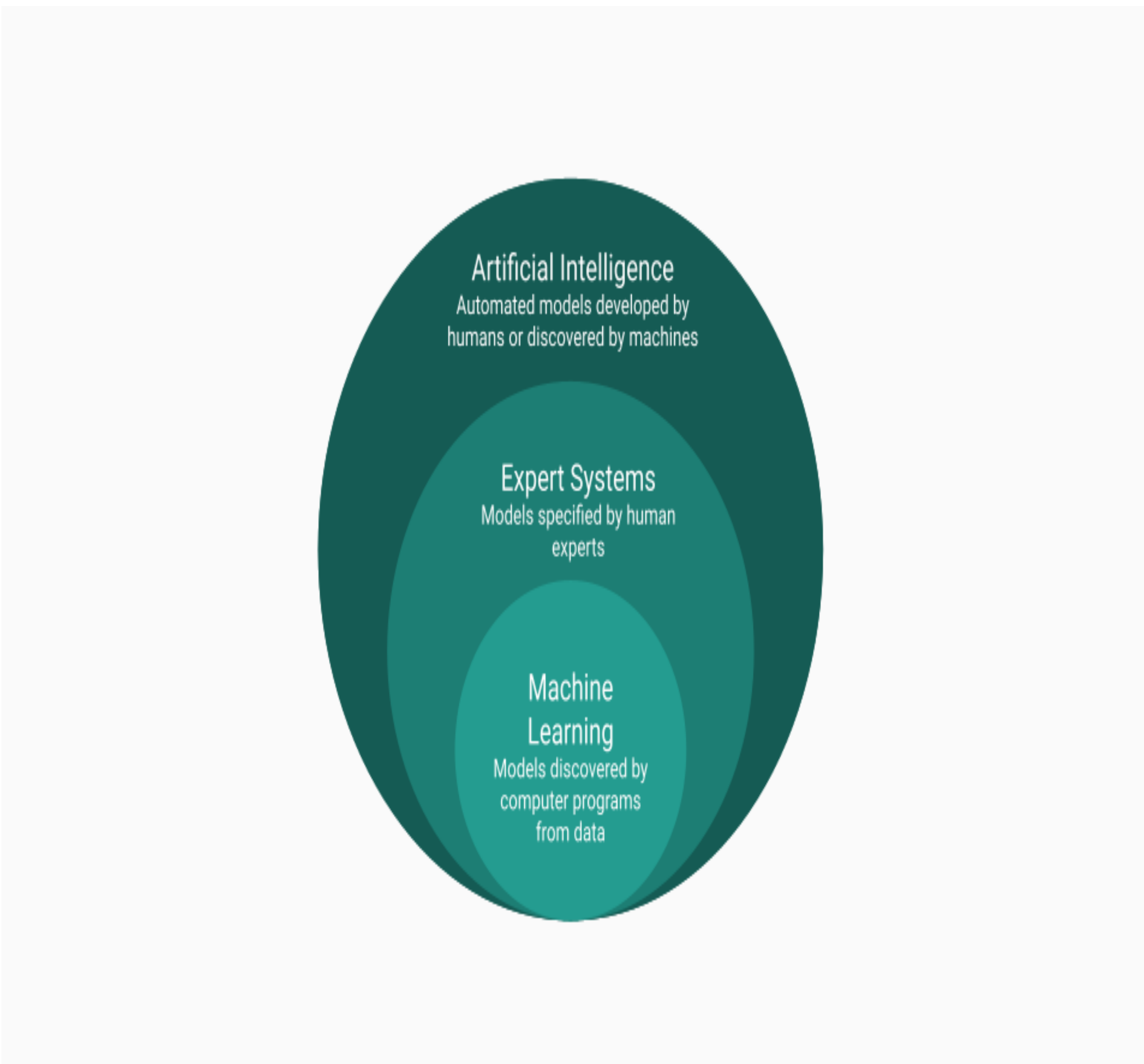

Figure 1-4. ML is a subfield of AI

AI is the general field that tries to automate the cognitive abilities of humans such as analytical thinking, decision-making and sensory perception. In the 20th century, computer scientists developed a subfield of AI called Symbolic AI (SAI) which included methodologies and tools to embed human knowledge in the form of well-defined rules or algorithms into computer systems.

SAI systems automate the models specified by domain experts and are aptly called expert systems. For instance, traders, finance executives and system developers work together to explicitly formulate all the rules and the model’s parameters that are to be automated by their financial and investment management systems. I have managed several such projects for marquee financial institutions in one of my previous companies.

However, SAI failed in automating complex tasks like image recognition and natural language processing, technologies used extensively in corporate finance and investing today. The rules for these types of expert systems are too complex and require constant updating for different situations. In the latter part of the 20th century, a new AI subfield of ML emerged from the confluence of improved algorithms, abundant data and cheap computing resources.

ML turns the SAI paradigm on its head. Instead of experts specifying models to process data, humans without domain expertise provide general purpose algorithms that discover a model’s parameters from data samples. More importantly, ML programs continually learn from new datasets and update their model’s parameters without any human intervention for code maintenance.

TRAINING A LINEAR ML SYSTEM TO LEARN

Recall the deterministic linear model discussed earlier and expressed by the equation $y = mx + b$. A unique line crosses at least two distinct points in the X-Y plane. The two points enable us to solve analytically for the exact values of parameters m and b using simple algebra. Once you have computed the parameters, you can use your model to make accurate predictions; given any point x, you can predict exactly what y will be.

However, financial models are not deterministic but probabilistic. For instance, if you were to plot a company's stock price returns on the y axis and the growth rate of its quarterly earnings on the x axis, you would see stock returns generally increase with earnings growth of a company. If you assume the relationship between stock price returns and quarterly earnings growth is **approximately** linear, you can use an analytical statistical technique to solve for the model's parameters m and b that gives you the line that best fits the company's sample data. If the linear approximation persists in the future, your model's predictions are not going to be precise but they are going to be better than making wild guesses or relying on luck.

Alternatively, you could use ML software to do similar calculations for you. In ML systems, the independent variable x is called a feature or predictor and the dependent variable y is called the target or response variable. Feeding sample data to the ML system is referred to as training the system. When our linear ML system computes the values of the parameters m and b, we say that the ML system has learned the model from the in-sample data. The objective in ML is to predict the target values on out-of-sample data which the system has not been trained on.

We will get into the details of modeling, training and testing probabilistic ML systems in part 2 of the book. Here is a useful definition of ML from Tom Mitchell, an ML pioneer: "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E." See figure 1-5 below.

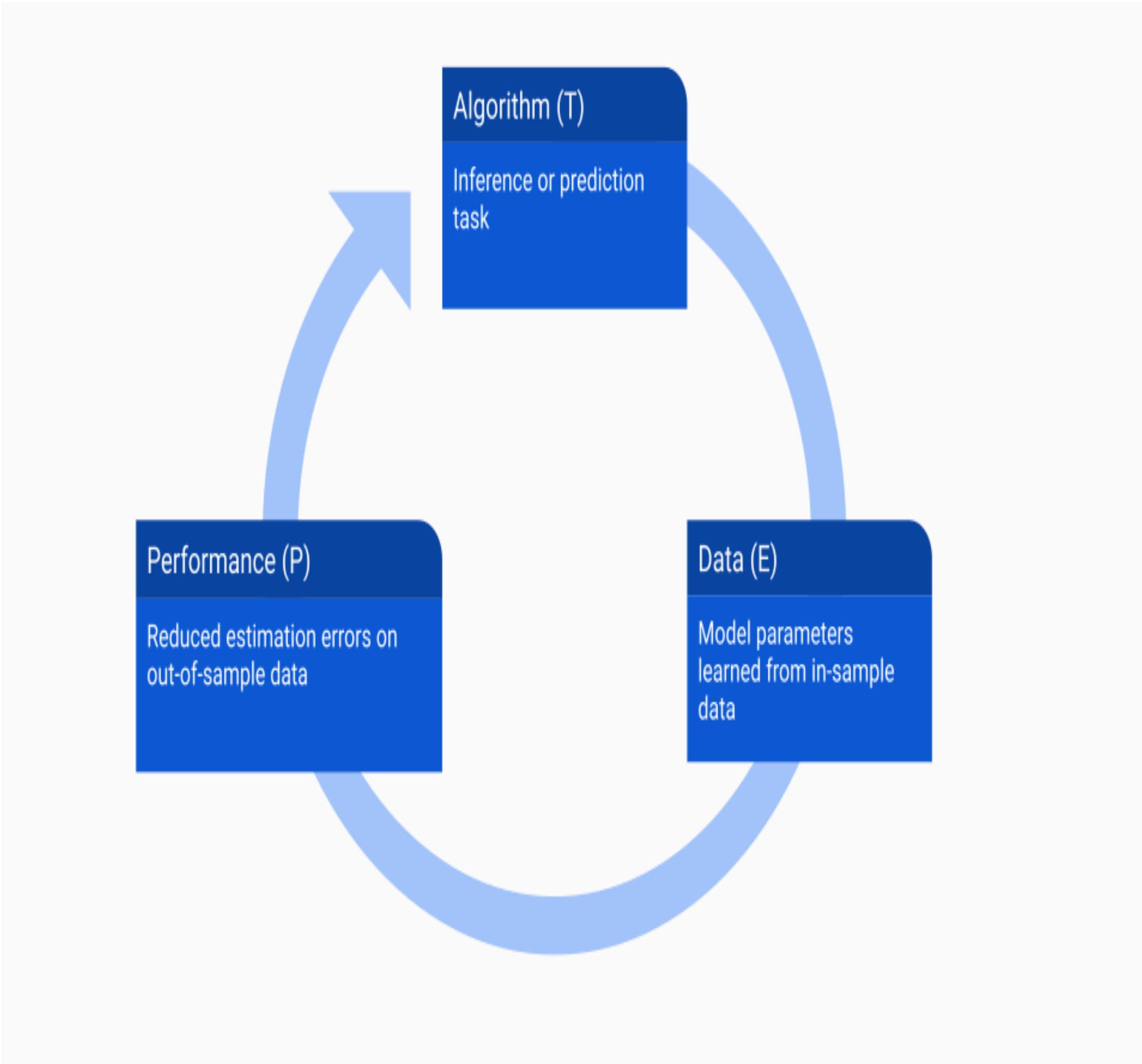

Figure 1-5. ML discovers model parameters from in-sample (training) data

ML systems are usually classified into three types based on how much assistance they need from their human teachers or supervisors:

Supervised learning

ML algorithms learn functional relationships from data which are provided in pairs of inputs and desired outputs. This is the most prevalent form of ML used in research and industry. Some examples of

ML systems include linear regression, logistic regression, random forests, gradient boosted machines and deep learning.

Unsupervised learning

ML algorithms are only given input data and learn structural relationships in the data on their own. K-means clustering algorithm is a commonly used data exploration algorithm used by investment analysts. Principal component analysis is a popular dimensionality reduction algorithm.

Reinforcement learning

ML algorithm continually updates a policy or strategy based on feedback from its environment with the goal of maximizing a cumulative reward. It’s different from supervised learning in that the feedback signal is not a desired output but a reward. Examples of algorithms are Q-learning and State-action-reward-state-action (SARSA). Reinforcement learning algorithms are being used in advanced trading applications.

In the 21st century, financial data scientists are training ML algorithms to discover complex functional relationships using data from multiple financial and non-financial sources. The newly discovered relationships may augment or replace the insights of finance and investment executives. ML programs are able to detect patterns in very high dimensional datasets, a feat that is difficult if not impossible for humans. They are also able to reduce the dimensions to enable visualizations for humans.

AI is used in all aspects of the finance and investment process, from idea generation to analysis, execution, portfolio and risk management. The leading AI-powered systems in finance and investing today use some combination of expert systems and ML based systems by leveraging the advantages of both types of approaches and expertise. Furthermore, AI-powered financial systems continue to leverage human intelligence (HI) for research, development and maintenance. Humans may also intervene in extreme market conditions where it may be difficult for AI systems to learn from abrupt changes. So you can think of modern financial systems as a complex combination of SAI + ML + HI.

Probabilistic ML

Probabilistic ML is the next generation ML framework and technology for AI-powered financial and investing systems. Leading technology companies clearly understand the limitations of conventional AI technologies and are developing their probabilistic versions to extend their applicability to more complex problems.

Google recently introduced TensorFlow Probability to extend their established TensorFlow platform. Similarly, Facebook and Uber have introduced Pyro to extend their Pytorch platform. Currently, the most popular open-source probabilistic ML technologies are PyMC3 and Stan. PyMC3 is written in Python and Stan is written in C++. In this book we will use the PyMC3 library.

Probabilistic ML is categorically different from conventional ML in use today, such as linear, nonlinear and deep learning systems, even though these other systems perform probabilistic computations. Below are the major differences between the two types of systems and are summarized in the figure 1-5 below:

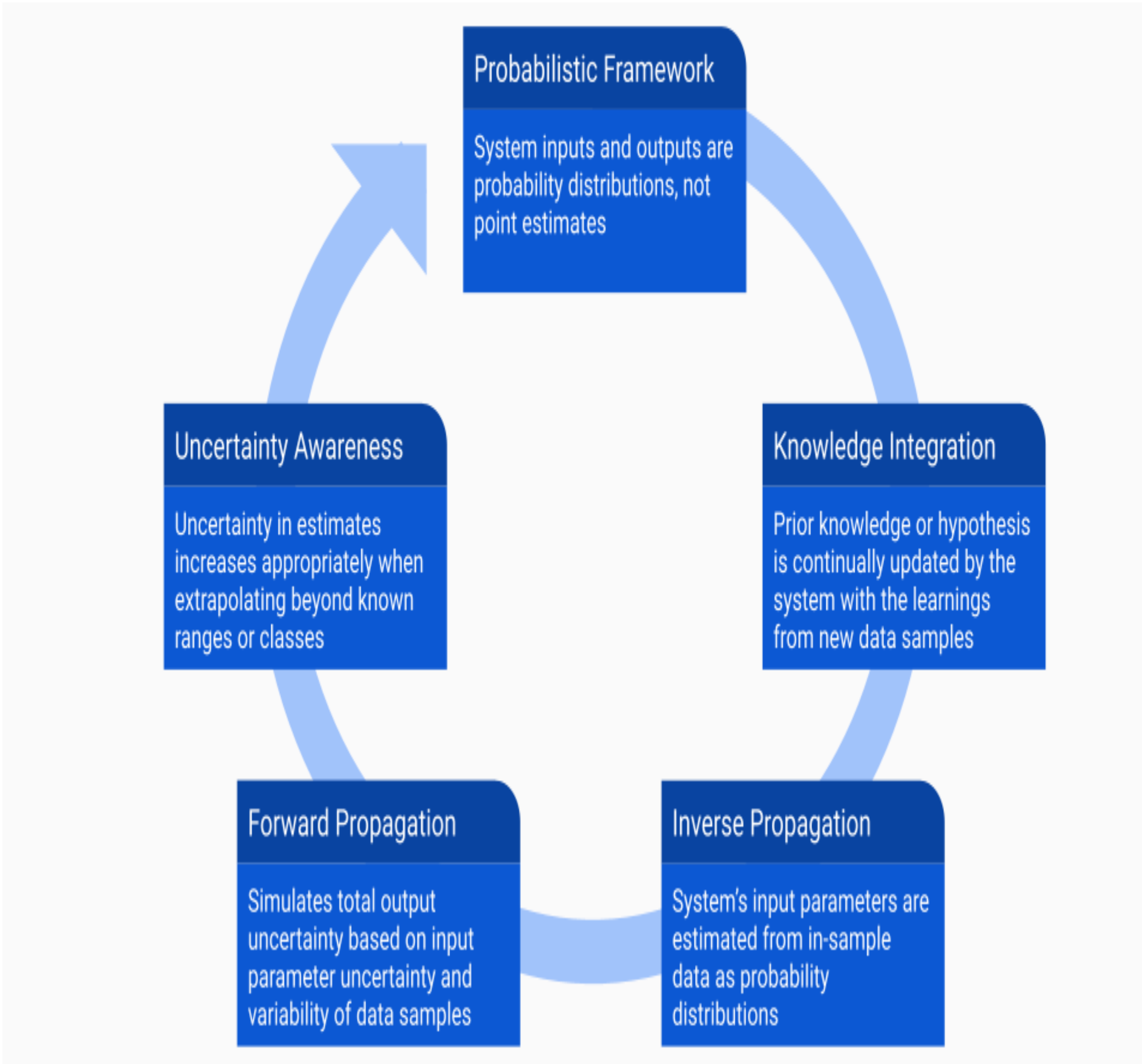

Figure 1-6. Summary of major characteristics of probabilistic ML systems

1. Only probability distributions

Even though conventional ML systems use a probabilistic framework, they only compute the most likely estimates and their associated probabilities as single values for inputs and outputs. This works well for domains, such as image recognition, where the data are plentiful and the signal to noise ratio is high. As was discussed and demonstrated in the previous sections, a point estimate is an inaccurate and misleading representation of financial reality where uncertainty is very high.

Probabilistic ML systems only deal in probability distributions in their computations of input parameters and model outputs. This is a realistic and honest representation of uncertainty of a financial model's variables. Furthermore, probability distributions leave the user considerable flexibility in picking the appropriate point estimate, if required, based on their business objectives.

2. Knowledge integration

Conventional ML systems do not have a theoretically sound framework for incorporating prior knowledge, whether it is well established scientific knowledge, institutional knowledge or personal insights. Later in the book, we will see that conventional statisticians sneak prior knowledge using ad hoc statistical methods, such as 95% significance level, and L2 regularizations while pounding the table about letting only 'the data speak for themselves'.

It is foolish not to integrate prior knowledge in our personal and professional lives. It is the antithesis of learning and vitiates against the nature of the scientific method. Yet this is the basis of null hypothesis significance testing (NHST), the prevailing statistical methodology in academia, research and industry since the 1960s. NHST prohibits the inclusion of prior knowledge in experiments based on the bogus claim that objectivity demands that we only let the data speak for themselves. By following this specious claim, NHST ends up committing the inverse fallacy as we will show in Chapter 4.

NHST's definition of objectivity would require us to touch fire everywhere and every time we find it because we cannot incorporate our prior knowledge of what it feels like in similar situations in the past. That is the definition of idiocy, not objectivity. In Chapter 4, we will discuss how and why several metastudies have shown that most published research based on NHST is false. Yes, you read that right and it has been an open secret since a seminal paper published in 2005.

Fortunately, we don't have to waste much ink on this specious argument about objectivity or the proliferation of junk science produced by NHST in this book. Probabilistic ML systems provide a mathematically rigorous framework for incorporating prior knowledge and updating it appropriately with learnings from new information. Representation of prior knowledge is done explicitly so that anyone can challenge it or change it. This is the essence of learning and the basis of the scientific method.

It is common knowledge that integration of accumulated institutional knowledge into a company's organization, process and systems leads to a sustainable competitive advantage in business. Moreover, personal insights and experiences of markets can lead to 'alpha' or the generation of exceptional returns in trading and investing for the fund manager who arrives at a subjectively different viewpoint from the rest of the crowd. That's how Warren Buffet, one of the greatest investors of all time, made his vast fortune. Markets mock dogmatic and unrealistic definitions of objectivity with lost profits and eventually with financial ruin.

3. Inverse uncertainty propagation

Almost all conventional ML systems use equally conventional statistical methodologies, such as p-values and confidence intervals, to estimate the uncertainty of a model's parameters. As will be explained in Chapter 4, these are deeply flawed, almost scandalous, statistical methodologies that plague the social sciences, including finance and economics. These methodologies adhere to a pious pretense to objectivity, implicit and unrealistic assumptions, obfuscated by inscrutable jargon, in order to generate solutions that are analytically tractable for a small set of scenarios.

Probabilistic ML is based on the calculus of probability theory in general and the inverse probability rule in particular. In the next chapter, we show how the inverse probability rule, mistakenly and mortifyingly known as Bayes' theorem, is a trivial reformulation of the product rule. It is a logical tautology that is embarrassingly easy to prove. It doesn't deserve to be called a theorem given how excruciatingly difficult it is to derive most mathematical theorems.

However, because of the normalizing constant in the inversion formula, it was previously impossible to invert probabilities analytically except for simple problems. With the recent advancement of state-of-the-art numerical algorithms, such as Hamiltonian Monte Carlo and Variational Inference, probabilistic ML systems are now able to invert probabilities to compute model parameter estimates from in-sample data for almost any real world problem. More importantly, they are able to quantify parameter uncertainties with mathematically sound credible intervals for any level of confidence.

4. Forward uncertainty propagation

Almost all conventional ML systems are based on discriminative models. This type of statistical model only learns a decision boundary from the sample data but not how the data are distributed statistically. Therefore, conventional discriminative ML systems cannot simulate and quantify total output uncertainty.

Probabilistic ML systems are based on generative models. This type of statistical model learns the statistical structure of data distribution and so can easily and seamlessly simulate new data, including generating data that might be missing or corrupted. Most importantly, these systems are able to simulate total uncertainty based on data variability and input parameter uncertainty, the probability distributions of which they have learned previously from in-sample data.

5. Uncertainty awareness

When computing probabilities, a conventional ML system uses the maximum likelihood estimation (MLE) method. This technique optimizes the parameters of an assumed probability distribution such that the data are most likely to be observed given the point estimates for the model's parameters. As we will see later in the book, MLE leads to wrong inferences and predictions when data are sparse, a common occurrence in finance and investing, especially when a market regime changes abruptly.

What makes it worse is that these MLE based ML systems attach horrifyingly high probabilities to these wrong estimates. We are automating the overconfidence of idiot savants. This makes conventional ML systems potentially risky and dangerous, especially when used in mission critical operations by personnel who either don't understand the fundamentals of these ML systems or have blind faith in them.

Probabilistic ML systems do not rely on a single point estimate, no matter how likely or optimal, but a weighted average of every possible estimate of a parameter's entire probability distribution. Moreover, the uncertainty of these estimates increases appropriately when systems deal with classes of data they have never seen before in training or are extrapolating beyond known data ranges. Unlike MLE based systems, probabilistic ML systems know what they don't know. This keeps the quantification of uncertainty honest and prevents overconfidence in estimates and predictions.

Conclusions

Finance is not a precise predictive science like physics. Not even close. So let's not pretend otherwise and treat academic theories and models of finance as if they were models of quantum physics, the obfuscating math and a counterfeited Nobel prize notwithstanding.

All financial models, whether based on academic theories or ML strategies, are at the mercy of the trinity of modeling errors. While this trifecta of errors can be mitigated with appropriate tools, it cannot be eliminated. There will always be asymmetry of information and cognitive biases. Models of asset values and risks will change over time due to the dynamic nature of capitalism, human behavior, and technological innovation.

Probabilistic ML technologies are based on a simple and intuitive definition of probability as logic and the rigorous calculus of probability theory. They enable the explicit and systematic integration of prior knowledge that is updated continually with new learnings.

These systems treat uncertainties and errors or financial and investing systems as features, not bugs. They quantify uncertainty generated from inexact inputs, parameters and outputs of finance and investing systems as probability distributions, not point estimates. This makes for realistic financial inferences and predictions that are useful for decision-making and risk management. Most importantly, these systems become capable of forewarning us when their inferences and predictions are no longer useful in the current market environment.

There are several reasons why probabilistic ML is the next generation ML framework and technology for AI-powered financial and investing systems. Its probabilistic framework moves away from flawed statistical methodologies (NHST, p-values, confidence intervals) and restrictive conventional view of probability as a limiting frequency. It moves us towards an intuitive view of probability as logic and a mathematically rigorous statistical framework that quantifies uncertainty holistically and successfully. Therefore it enables us to move away from wrong, idealistic, analytical models of the past towards less wrong, more realistic, numerical models of the future.

The algorithms used in probabilistic programming are among the most sophisticated algorithms in the AI world, which we will delve into in the second half of the book. In the next three chapters, we will take a deeper dive into why it is very risky to deploy your capital using conventional ML systems because they are based on orthodox probabilistic and statistical methods which are scandalously flawed and have brought shame and disrepute to the social sciences in general.

References

The Money Formula by David Orrell and Paul Wilmott, Wiley, 2017

Nobels For Nonsense by J.R. Thompson, L.S. Baggett, W.C. Wojciechowski, and E.E. Williams, Journal of Post Keynesian Economics, Fall 2006

Model Error by Katerina Simons, New England Economic Review, November 1997

Bayesian Risk Management by Matt Sekerke, Wiley, 2015

Probability Theory: The Logic of Science by E.T. Jaynes, Cambridge University Press, 2004

Chapter 2. Analyzing and Quantifying Uncertainty

A NOTE FOR EARLY RELEASE READERS

With Early Release ebooks, you get books in their earliest form—the author’s raw and unedited content as they write—so you can take advantage of these technologies long before the official release of these titles.

This will be the 2nd chapter of the final book. Please note that the GitHub repo will be made active later on.

If you have comments about how we might improve the content and/or examples in this book, or if you notice missing material within this chapter, please reach out to the author at EMAIL.

The Monty Hall problem, a probability brain-teaser, is an entertaining way to explore the complex and profound nature of uncertainty that we face in our personal and professional lives. In this chapter, we will solve the apparent paradox of the Monty Hall problem by developing two analytical solutions of differing complexity using the fundamental rules of probability theory.

We also derive the inverse probability rule that is pivotal to probabilistic machine learning. Later in this chapter, we confirm these analytical solutions with a Monte Carlo simulation, one of the most powerful numerical techniques that is used extensively in finance and investing. Note that throughout the chapter we explore how the Monty Hall problem and its solution, which is essentially a betting strategy, is related to some key concepts and pitfalls in finance and investing.

The Monty Hall problem

The famous Monty Hall problem was originally conceived and solved by an American statistician, Steve Selvin. The problem as we know it now is based on the popular 1970s game show *Let's Make a Deal* and named after its host, Monty Hall. Here are the rules of this brain-teaser:

1. There is a car behind one of three doors and goats behind the other two
2. The objective is to win the car (not a goat!)
3. Only Monty knows which door hides the car
4. Monty allows you to choose any one of the three doors
5. Depending on the door you choose, he opens one of the other two doors that has a goat behind it

So let's play the game. Monty asks you to pick a door you think the car is behind. It doesn't really matter which door you chose because the game plays out similarly regardless. Say you chose door 1. Based on your choice of door 1, Monty opens door 3 to show you a goat. See Figure 2-1 below.

Now Monty offers you a deal: he gives you the option of sticking with your original choice of door 1 or switching to door 2. Do you switch to door 2 or stay with your original decision of door 1? Try to solve this problem before you read ahead - it will be worth the trouble.

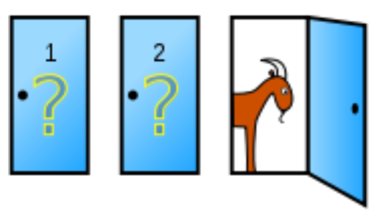

Figure 2-1. The Monty Hall Problem. Image source: Wikimedia Commons

I must admit that when I first came across this problem many years ago, my immediate response was that it didn't matter whether you stayed or switched doors since now it is equally likely that the car is behind either door 1 or door 2. So I stayed with my original choice. Turns out that my choice is wrong.

The optimal strategy is to switch doors because, by opening one of the doors, Monty has given you valuable new information which you can use to increase the odds of winning the car. After I worked through the solution and realized I was wrong, I took comfort in the fact that this problem had stumped thousands of PhD statisticians. It had even baffled the great mathematician Paul Erdos, who was only convinced that switching doors was a winning strategy after seeing a simulation of the solution. As the sidebar below explains, there are also psychological reasons why people don't switch doors.

THE PSYCHOLOGY OF FINANCIAL DECISION-MAKING

If the car is equally likely to be behind either of the two remaining doors, don't you think it is odd that most people don't switch doors? Behavioral Economics, which uses psychology to explain economic behavior, calls this the endowment effect. People tend to put a higher value on the things they own than they would if they did not possess it. They seem to form an emotional or irrational attachment to it.

There is also another psychological reason for the endowment effect - loss aversion. Losses cause investors more pain than gains give them pleasure. That explains why investors are generally reluctant to cut their losses but are quick to take their profits. Since I had chosen door 1, I felt like I owned it. It would have hurt a lot more if I switched doors and lost than if I switched doors and won.

Inertia also plays an important role in decision-making. People would rather be wrong through inaction, an error of omission, than be wrong through action, an error of commission.

Before we do a simulation of this problem, let's try to figure out a solution analytically by applying the axioms of probability.

Axioms of Probability

Here is a refresher on the axioms or fundamental rules of probability. It is simply astonishing that the calculus of probability can be derived entirely from the three axioms stated below and a few definitions.

Say S1 is the scenario (also known as an event) that the car is behind door 1. We define S2 and S3 similarly. The complement of S is S' (not S) and is the scenario in which there is a goat (or not a car) behind the door.

Scenarios S and S' are said to be mutually exclusive since there is either a goat or a car but not both behind any given door. Since those are the only possible scenarios in this game, S and S' are also said to be collectively exhaustive scenarios or events. The set of all possible scenarios is called the sample pace. Let's see how we can apply the rules of probability to the Monty Hall game.

Axiom 1: P(S) >= 0

> Probability of an event or scenario, P(S), is always assigned a non-negative real number. For instance, when Monty shows us that there is no car behind door 3, P(S3) = 0. An event probability of 0 means the event is impossible or didn't occur.

Axiom 2: P(S1) + P(S2) + P(S3) = 1

> What this axiom says is that we are absolutely certain that at least one of the scenarios in the sample space will occur. Note that this axiom implies that an event probability of 1 means the event will certainly occur or has already occurred. We know from the rules of the Monty Hall game that there is only one car and it is behind one of the three doors. This means that the scenarios S1, S2 and S3 are mutually exclusive and collectively exhaustive. Therefore P(S1) + P(S2) + P(S3) = 1. Also note that axioms 1 and 2 ensure that probabilities always have a value between 0 and 1 inclusive. Furthermore, P(S1) + P(not S1) = 1 implies P(S1) = 1 - P(not S1).

Axiom 3: P(S2 or S3) = P(S2) + P(S3)

> This axiom is known as the sum rule and enables us to compute probabilities of two scenarios that are mutually exclusive. Say we want to know the probability that the car is either behind door 2 or door 3 i.e we want to know P(S2 or S3). Since the car cannot be behind door 2 and door 3 simultaneously, S2 and S3 are mutually exclusive, i.e. P(S2 and S3) = 0. Therefore P(S2 or S3) = P(S2) + P(S3)

We will apply the axioms of probability to compute other probabilities in the Monty Hall problem. Since each scenario is mutually exclusive (there is either a goat or a car behind each door) and collectively exhaustive (those are all the possible scenarios), their probabilities must add up to 1 since at least one of the scenarios must occur.

P(S1) + P(S2) + P(S3) = 1

Before we make a choice, the most plausible assumption is that the car is equally likely to be behind any one of the three doors. There is nothing in the rules of the game to make us think otherwise and Monty Hall hasn't given us any hints to the contrary. So it is reasonable to assume that *P(S1)=P(S2)=P(S3)*. Using the equation 2.1 above we get

3 * (PS1) = 1 or P(S1)=$\frac{1}{3}$

Since *P(S1)=P(S2)=P(S3)*, equation 2.2 implies that it is logical to assume that there is a $\frac{1}{3}$ probability that the car is behind one of the three doors.

By the sum rule, the probability that the car is behind either door 2 or door 3 is

P(S2 or S3) = P(S2) + P(S3) = $\frac{1}{3}$ + $\frac{1}{3}$ = $\frac{2}{3}$

After you choose door 1 and Monty opens door 3 showing you a goat, P(S3) = 0. Substituting this value in equation 2.3 above and solving for P(S2), we get

P(S2) = P(S2 or S3) - P(S3) = $\frac{2}{3}$ - 0 = $\frac{2}{3}$

So switching your choice from door 1 to door 2 doubles your chances of winning the car - it goes from ⅓ to ⅔ . Switching doors is the optimal betting strategy in this game. See Figure 2-2 below.

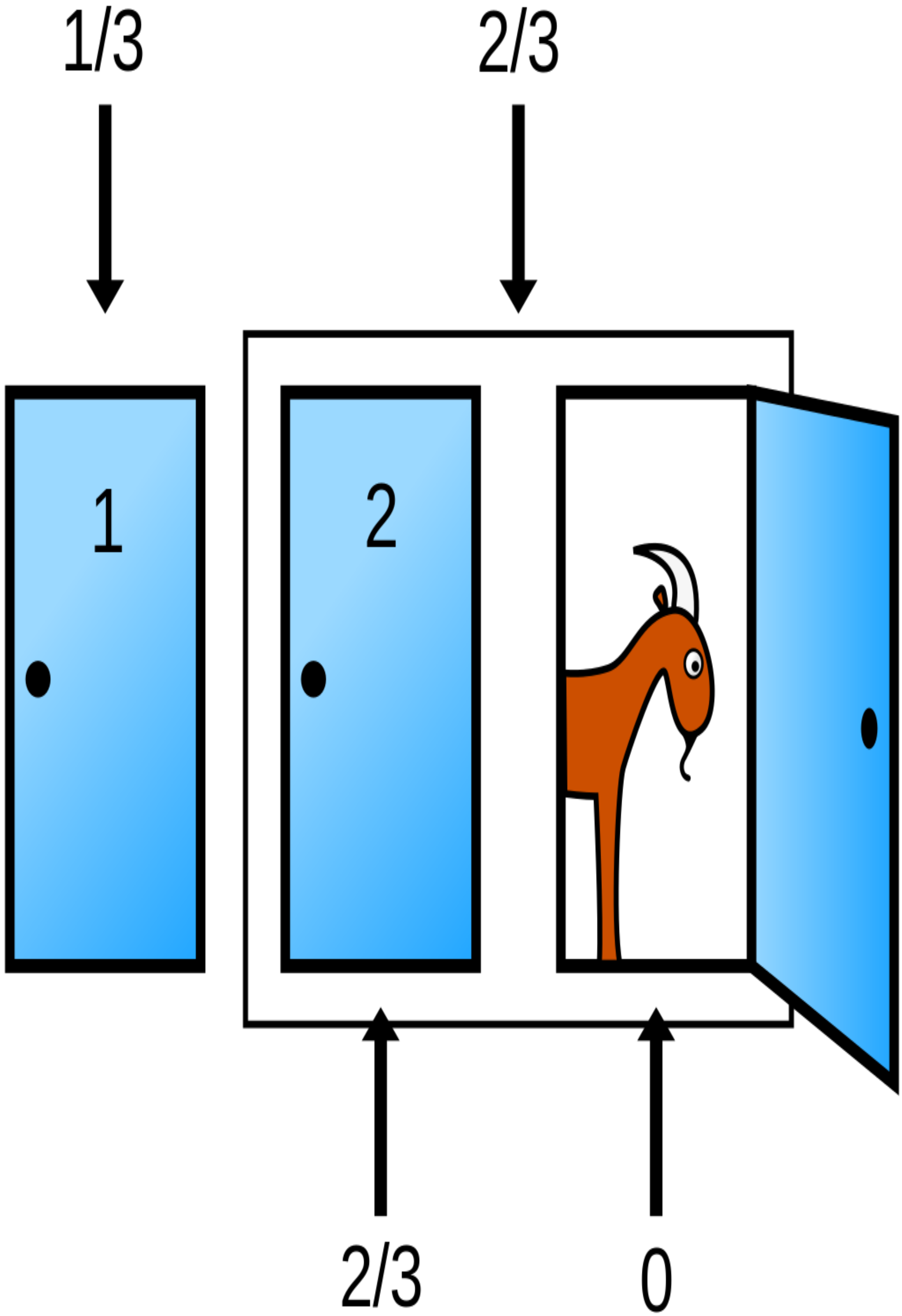
1/3
2/3
1
2
2/3
0

Figure 2-2. Figure 2-2. A simple analytical solution to the Monty Hall Problem. Image source: Wikimedia Commons

It is important to note that because of uncertainty, there is still a ⅓ chance that you could lose if you switch doors. In general, randomness of results makes it hard and frustrating to determine if your investment or trading strategy is a winning one or or a lucky one. It is much easier to determine a winning strategy in the Monty Hall problem because it can be determined analytically or by simulating the game many times.

Inverting Probabilities

Let's develop a more rigorous analytical solution to the Monty Hall problem. In order to do that we need to understand conditional probabilities and how to invert them. Recall that when we condition a probability, we revise the plausibility of a scenario or event by incorporating new information from the conditioning data. The conditional probability of a scenario H given a conditioning dataset D is represented as P(H|D) which reads as the probability of H given D and is defined as follows:

P(H|D) = P(H and D) / P(D) provided P(D) =\= 0 since division by 0 is undefined

The division by P(D) ensures that probabilities of all scenarios conditioned on D will add up to 1. Recall that if two events are independent, their joint probabilities are the product of their individual probabilities. That is, P(H and D) = P(H) * P(D) if knowledge of D does not improve our certainty of H and vice versa.

The definition of conditional probability P given H above also implies that P(H and D) = P(H|D) * P(D). This is called the product rule. We can now derive the inverse probability rule from the product rule. We know from the symmetry of the joint probability of two events:

P(H and D) = P(D and H)

P(H|D) * P(D) = P(D|H) * P(H)

$$P(H|D) = P(D|H) * P(H) / P(D)$$

And that, ladies and gentlemen, is the proof of the famous and wrongly named ‘Bayes theorem’. If only all mathematical proofs were so easy! As you can see, this alleged theorem is a trivial reformulation of the product rule. It’s as much a theorem as multiplying two numbers and solving for one of them in terms of their product (for example $H = H*D/D$). The hard part and the insightful bit is interpreting and applying the formula to invert probabilities and solve complex, real-world problems. Since the 1950s, the above formula is also wrongly known as Bayes’s theorem. See sidebar below.

BAYES DID NOT DISCOVER BAYES'S THEOREM AND WAS NOT A BAYESIAN

Thomas Bayes, an 18th century theologian and amateur mathematician, would probably be rolling over in his grave if he knew that his name was being used in an intellectual war that has raged for almost a century over the meaning and application of probability theory. Bayes' surreal confusion would stem from the fact that he was not responsible for developing the inverse probability rule or the foundations of the statistical theory and inference that wrongly bears his name. That distinction goes to the mathematician and physicist, Pierre-Simon Laplace who built the foundations of this statistical school of thought in the late 1700s. It is shameful that Laplace doesn't get much of the credit for discovering it independently and writing down the general formula in its modern form.

Bayes's paper on inverse probability was published posthumously by his friend, Richard Price, who wrote about half of that paper and corrected several of Bayes's errors over two years before he sent it for publication. It is equally shameful that Price has been ignored for his contributions for completing Bayes's paper and actually submitting it for publication without adding his name to it. The Bayes-Price paper did not cause much of a stir among mathematicians when it was finally published. That's because Daniel Bernoulli and A. de Moivre had already worked on the inverse probability problem much before Bayes and the paper was not breaking new ground. In fact, Bayes had read A. de Moivre's book, The Doctrine of Chances, and used it to solve his specific problem on inverting probabilities. Moreover, he never generalized the rule from his specific problem, a requirement for all rules and theorems, especially the eponymous ones.

In this book we will correct this blatant injustice and egregious misnomer by referring to Bayes' theorem as the inverse probability rule since the alternative of calling it the Laplace-Bernoulli-Moivre-Bayes-Price rule is way too long. Furthermore, we will refer to Bayesian inference as epistemic or probabilistic inference because that's what it is and Bayes had little or nothing to do it. In fact, there is no evidence to suggest that Bayes was even a Bayesian as the term is used today.

Epistemic inference in general and inverse probability rule in particular are the foundation of probabilistic machine learning and we will discuss it at length in the second part of this book. For now, let's see apply it to the Monty Hall problem. Let's continue with the same definitions of S1, S2 and S3 and their related probabilities. Now we define our dataset D which includes two observations: your choosing door 1 and based on choice of door 1, Monty opens door 3 to show you a goat. We want to solve for P(S2|D) i.e. the probability that the car is behind door 2 given the above dataset D.

We know from the inverse probability rule that this equals P(D|S2)*P(S2)/P(D). The challenging computation is P(D) which is the unconditional or marginal probability of seeing the dataset D regardless of which door the car is behind. The rule of total probability allows us to compute marginal probabilities from conditional probabilities. Specifically, the rule states that the marginal probability of D, P(D), is the weighted average probability of realizing D under different scenarios, with specific weights estimated for each scenario:

$$P(D) = P(D|S1)*P(S1) + P(D|S2) * P(S2) + P(D|S3)*P(S3)$$

We have estimated the three scenarios at the beginning of the Monty Hall game, namely P(S1) = P(S2) = P(S3) = ⅓ . These are going to be the weights for each possible scenario. Let's compute the conditional probabilities of observing our dataset D. Note that by P(D|S1) we mean the probability of seeing the dataset D given that the car is actually behind door 1 and so on.

If the car is behind door1 and you pick door 1, there are goats behind the other two doors. So Monty can open either door 2 or door 3 to show you a goat. Thus the probability of Monty opening door 3 to show you a goat given that you chose door 1 is ½ or P(D|S1) = ½ .

If the car is behind door 2 and you have chosen door 1, Monty has no choice but to open door 3 to show you a goat. So P(D|S2) = 1.

If the car were behind door 3, the probability of Monty opening door 3 is zero since he would ruin the game and you would get the car just for showing up. Therefore P(D|S3) = 0.

We plug the numbers into the rule of total probability to calculate the marginal or unconditional probability of seeing the dataset D in the game:

$$P(D) = P(D|S1)*P(S1) + P(D|S2) * P(S2) + P(D|S3)*P(S3)$$

$$P(D) = [\tfrac{1}{2} * \tfrac{1}{3}] + [1* \tfrac{1}{3}]+ [0*\tfrac{1}{3}]= \tfrac{1}{2}$$

Now we have all the probabilities we need to use in the inverse probability rule to calculate the probability the car is behind door 2 given our dataset D:

$$P(S2|D) = P(D|S2) * P(S2) / P(D)$$

$$P(S2|D) = [1*\tfrac{1}{3}] / \tfrac{1}{2} = \tfrac{2}{3}$$

We can similarly compute the probability the car is behind door 1 given our dataset D:

$$P(S1|D) = P(D|S1) * P(S1) / P(D)$$

$$P(S1|D) = [\tfrac{1}{2} * \tfrac{1}{3}] / \tfrac{1}{2} = \tfrac{1}{3}$$

Clearly, we double our chances by switching since $P(S2|D) = 2P(S1|D) = \tfrac{2}{3}$. Note that there is still a ⅓ chance that you can win by not switching. But like trading and investing, your betting strategy should always put the odds in your favor.

Simulating the Solution

Still not convinced? Let’s solve the Monty Hall problem by using a powerful numerical method called Monte Carlo Simulation (MCS) that we mentioned in the previous chapter. This powerful computational method is applied by theoreticians and practitioners in almost every field, including business and finance. Recall that MCS samples randomly from probability distributions to generate numerous probable scenarios of a system whose outcomes are uncertain. It is generally used to quantify the uncertainty of model outputs. The following MCS code shows how switching doors is the optimal betting strategy for this game if played many times.

```
import random
#Number of iterations in the simulation
N = 30
#List to store results of all iterations
stay_results = []
switch_results = []
#For loop for collecting results of each iteration
for i in range(N):
  doors = ['door 1', 'door 2', 'door 3']
  #Random selection of door to place the car
  car_door = random.choice(doors)
  #You select a door at random
  your_door = random.choice(doors)
  #Monty can only select the door that does not have the car and
one that you have not chosen
  monty_door = list(set(doors) - set([car_door, your_door]))[0]
  #The door that Monty does not open and the one you have not
chosen initially
  switch_door = list(set(doors) - set([monty_door, your_door]))[0]
  #Result if you stay with your original choice and it has the car
behind it
  stay_results.append(your_door == car_door)
  #Result if you switch doors and it has the car behind it
  switch_results.append(switch_door == car_door)
#Probability of winning the car if you stay with your original
choice of door
probability_staying = sum(stay_results)/N
#Probability of winning the car if you switch doors
probability_switching = sum(switch_results)/N
print("Probability of winning the car if you stay with your
original choice: {:.2f}".format(probability_staying))
print("Probability of winning the car if you switch your choice of
doors: {:.2f}".format(probability_switching))
```

As you can see from the results of the simulations, switching doors is the winning strategy over the long term. The probabilities are approximately the same as in the analytical solution if you play the game a thousand times. The probabilities become almost exactly the same as the analytical solution if you play the game over a hundred thousand times. We will explore the theoretical reasons for these results in particular and MCS in general in the next chapter.

The Trifecta of Uncertainty

Let’s now see how we can use the Monty Hall problem to understand the complexities of uncertainty. Uncertainty is generally classified into three types: aleatory, epistemic and ontological. We will apply these important concepts to the Monty Hall problem to understand these three types of uncertainty.

Aleatory Uncertainty

Aleatory means ‘of dice’ in Latin. Both the analytical and simulation solutions to the Monty Hall problem demonstrated that your strategy of staying or switching doors in this game does not guarantee you a win during a single play or even multiple plays of the game. You could stay with your original choice of door 1 and have a ⅓ chance of winning the car. You could switch to door 2 and have a ⅓ chance of losing the car. Whenever you play the game, you are indeed rolling the proverbial dice since the outcome is uncertain. Actually, it’s more uncertain than rolling dice since dice have no aleatory uncertainty, only epistemic uncertainty as explained in the next section.

PREDICTING COIN TOSSES WITH 100% ACCURACY

Tossing a coin is a canonical example of aleatory uncertainty in the current literature on probability and statistics. However, this shows an inexcusable ignorance of the laws of physics. It has been experimentally verified that if you know the initial conditions and other parameters of a coin toss, you can predict its outcome with 100% accuracy. That's because coin tossing is physics, not randomness [6].

Statistician and former magician, Persi Diaconis had engineers build him a mechanical coin flipper so that he could experiment and study coin tossing. Indeed, he and his colleagues verified that there is no randomness in a coin toss with the mechanical coin flipper [7]. The randomness of a coin toss arises from the inconsistency of initial conditions of human coin flipping and from the coin rolling on the ground.

There is no intrinsic aleatory uncertainty of coin tosses. The uncertainty we observe stems from our lack of precise information or knowledge of the physics of the tosses. It is a bad example of aleatory uncertainty. It also demonstrates that coins don't have any intrinsic limiting frequency. The physics of the toss can make a biased coin honest and vice-versa. Coin tossing and rolling dice are examples of epistemic uncertainty which we will discuss next.

In contrast to coins or dice, no amount of information about the physical characteristics about the doors or its motion will reduce the aleatory uncertainty of where the car is in the Monty Hall problem. It is a great example of aleatory uncertainty and why social systems are fundamentally different and much harder to predict than physical systems.

Epistemic Uncertainty

Episteme means knowledge in Greek. The uncertainty of any scenario depends on the state of knowledge or ignorance of the person confronting it.

Based on the information from Monty’s response to your choice, the probability of door 1 having a car behind it remained unchanged at ⅓, the probability for door 2 changed from ⅓ to ⅔ and the probability for door 3 changed from ⅓ to 0.

However, there is no uncertainty for Monty regarding which door the car is behind - his probability for each door is either 1 or 0 at all times. He is only uncertain about which door you are going to pick. Also, once you pick any door, he is certain what he is going to do. But he is uncertain what you will do next when offered the deal - will you stay with your original choice of door 1 or switch to door 2 and most likely win the car? Monty’s uncertainties are not epistemic but are of a fundamentally different nature which we will discuss in the next subsection.

So we can see from this game that the uncertainty of picking the right door for you is a function of one’s state of knowledge or ‘episteme’. It is important to note that this is not a subjective belief but a function of information or lack of it. Any participant and any host would have the same uncertainties calculated above and switching doors would still be the winning strategy.

NOTE

This game is also an example of **asymmetry of information** that characterizes financial deals and markets. Generally speaking, parties to a deal always have access to differing amounts of information about various aspects of a deal or asset which leads to uncertainty in their price estimates and deal outcomes. Different information processing capabilities and speeds further exacerbate those uncertainties.

Ontological Uncertainty

Ontology is the philosophical study of the nature of being and reality. Ontological uncertainty generally arises from the future of human affairs being essentially unknowable [7]. While the future generally resembles the past, there is no guarantee it will always continue to do so, especially in complex social systems consisting of emotional beings with free will and creativity. We discuss the problem of induction in the next subsection.

To make the Monty Hall game resemble a real world business deal or a trade, we have to dive deeper into the objective of the game, namely winning the car. From Monty's perspective, winning means keeping the car so he can reduce costs of the show. When the game is viewed in this way, Monty's knowledge of the car's placement behind any one of the doors does not decrease his ontological uncertainty about winning the game. This is because he doesn't know which door you're going to pick and if you will stay or switch doors when given the choice to do so. Since his odds are the complement of your odds, he has a ⅔ chance of keeping the car if you don't switch doors and ⅓ chance of losing the car to you if you do switch doors.

There are other possible ontological uncertainties for you. Say you show up to play the game a second time armed with the analysis of the game and the door switching strategy. Monty surprises you by changing the rules such that switching doors incurs costs and loses half the time. Moreover, he doesn't give you any time to analyze the game to discover an optimal betting strategy. Or he introduces another door and a second hand car as well.

Unexpected changes in business and financial markets are the rule, not the exception. Markets don't send out a memo to participants when undergoing structural changes. Companies, deals and trading strategies fail regularly and spectacularly because of these types of changes. It is similar to the way one of Hemmingway's characters described how he went bankrupt: 'Two ways. Gradually, then suddenly'.

Ontological Uncertainty and the Problem of Induction

Inductive reasoning synthesizes information from past data to formulate hypotheses that will continue to be plausible in the future. This method of reasoning is the foundation of the scientific method. But how can we be certain that just because our hypothesis has worked thus far, it will continue to work in the future? Well, because the past has resembled the future so far.

But that is exactly what we are trying to prove in the first place. This circular reasoning is generally referred to as the problem of induction and has been debated by philosophers for millennia everywhere in the world. [8]

Before your mind starts entertaining fantastical notions about the real world, you should consider that the success of the laws of physics, on which the technological wonders of the modern world has been built, is proof of the validity and success of inductive reasoning [9]. It is a crushing body blow to the philosophical problem of induction. That is why the philosopher C.D. Broad has aptly called induction ‘the glory of science and the scandal of philosophy’.

However, we cannot ignore the problem of induction in the social sciences, such as economics, where it is a clear and present danger to any simple extrapolation of the past. It’s because human beings have free will and creativity. Sometimes history repeats itself, sometimes it rhymes, and other times it makes no sense at all. That’s why I manage my risks carefully when trading the markets regardless of past successes.

But I would bet everything I have and more on the philosophically unjustifiable claim that the sun will rise every day in the east and set in the west for any period of time not exceeding 4 billion years. Any philosophers out there who want to take the other side of my bet? I’m willing to take the shortest odds. Of course, my bet has no ontological uncertainty in spite of the problem of induction and is a money making machine if I can find some fool to take the other side of it. This bet is only good for about 5 billion years after which time the sun is expected to become a red giant and swallow the earth. Notice that I have given myself a safety margin of 20% or a billion years.

Meaning of Probability

Probability is used to quantify and analyze uncertainty as we have done in the Monty Hall problem. It might be surprising for you to know that one can satisfy the axioms of probability theory but yet disagree on the meaning of probability. Two major schools of thought have been sparring over the fundamental meaning of probability - the very soul of statistics - for about a century. The two camps disagree on not only the fundamental meaning of probability in those axioms, but also on the methods for applying the axioms consistently to make inferences. These core differences have led to the development of divergent theories of statistical inference.

Frequentist Interpretation of Probability

Statisticians who believe that probability is a natural, immutable property of an event or physical object and is measured empirically as a long-run relative frequency are called frequentists. Frequentism is the dominant school of statistics of modern times, in both academic research and industrial applications. It is also known as orthodox, classical, or conventional statistics.

Orthodox statisticians claim that probability is a naturally occuring attribute of an event or physical phenomenon. The probability of an event should be measured empirically by repeating similar experiments ad nauseam—either in reality or hypothetically using simulations. For instance, if an experiment is conducted N times and an event E occurs with a frequency M times, the relative frequency M/N approximates the probability of E. As the number of experimental trials N approaches infinity, the probability of E equals M/N.

Frequentists consider any alternative interpretation of probability as anathema, almost blasphemous. As we will see later in the chapter, this definition is based on ideology and not on scientific fact or logic. In Chapter 4, we will examine how the frequentist philosophy of probability and statistics has had a profoundly damaging impact on the theory and practice of social sciences in general and finance in particular.

Epistemic Interpretation of Probability

The other important school of thought is popularly and mistakenly known as Bayesian. As mentioned earlier, this is an egregious misnomer and in this book we will refer to this interpretation as epistemic probability. Probabilities have a simpler, more intuitive meaning in the epistemic school: it is an extension of logic and quantifies the degree of plausibility of the event occurring based on the current state of knowledge (or ignorance). Probabilities are updated as more information is acquired using the inverse probability rule.

Most importantly, plausibility of an event is expressed as a probability distribution as opposed to a point estimate. This quantifies the degree of plausibility of various possibilities that can occur given the current state of knowledge. Point estimates are avoided as much as possible given the uncertainty endemic in life and business. Probabilistic ML is based on this school of thought.

It is important to note that the epistemic interpretation of probability is broad and encompasses the frequentist interpretation of probability as a special case. For instance, in the Monty Hall problem we assumed that it is equally likely that a car is behind one of three doors, the epistemic and frequentist probabilities are both ⅓. Furthermore, both schools of thought would come to the same conclusion that switching doors doubles your probability and is a winning strategy. Similarly, for simple games of chance such as dice and cards, both schools of probability give you the same results.

Relative Probabilities

Since the last century, frequentist ideologues have disparaged and tried to destroy the epistemic school of thought in covert and overt ways. Amongst other things, they labeled epistemic probabilities subjective, which in science is a pejorative term. The frequentist theory of statistics has been sold to academia and industry as the scientifically rigorous, efficient, robust and objective school of thought. Nothing could be further from the truth. We will see in Chapters 3 and 4 that those claims are bogus, bordering on statistical skulduggery. Despite the vigorous efforts of frequentists, epistemic inference has been proven to be theoretically sound and experimentally verified.

As a matter of fact, it is actually the frequentist version of probability that fails miserably exposing its frailties and *ad hoceries* when subjected to complex statistical phenomena. For instance, the frequentist approach cannot be logically applied to image processing and reconstruction, where the sampling distribution of any measurement is always constant. Probabilistic algorithms dominate the field of image processing, leveraging their broader, epistemic foundation. [9]

Are there any objective probabilities in the Monty Hall problem? The car is behind door 2 so isn't the 'true' and objective probability of S2 = 1, a constant? Yes, any host in the same position as Monty will assign S2=1 as would any participant assign S2=1/3. However, there is no 'true' probability of any event in the sense that it has an ontological existence in this game independent of the actions of humans. If Monty has the car placed behind door 1, S2 = 0 for him but remains constant at ⅓ for any participant.

Probabilities depend on the model used, phase of the game and the information available to the participant or the host. As noted above, probabilities of any participant or host are not subjective but a function of information since any host and any participant would arrive at the same probabilities using basic probability theory. Probability is a mental construct used to quantify uncertainties dynamically.

It is just like the physics of special relativity which have been experimentally verified since Albert Einstein published his monumental paper in 1905. The laws of physics are invariant across all frames of reference that are not accelerating. Two observers can still disagree on fundamental measurements of mass, length and time depending on how their frames of reference are moving relative to one another and compared to the speed of light. It is better to think in terms of relative probabilities based on an observer's frame of reference and information as opposed to objective or subjective probabilities.

We know for a fact that even a physical object like a coin or dice has no intrinsic probabilities based on long term frequencies. It depends on initial conditions and the physics of the toss. Probabilities are epistemic, not ontic

- they are a map, not the terrain. It’s about time frequentists stop fooling themselves and others with their mind projection fallacies and give up their pious pretense of objectivity and scientific rigor.

Summary

In this chapter we used the famous Monty Hall problem to review the fundamental rules of probability and apply them to solve the Monty Hall problem. We also realized how easy it was to derive the inverse probability rule that is pivotal to epistemic inference and probabilistic machine learning. Furthermore, we used the Monty Hall game to explore the profound complexities of aleatory, epistemic and ontological uncertainty that pervades our lives and businesses.

A better understanding of the three types of uncertainty and the meaning of probability will enable us to analyze and develop appropriate models for our ML systems to solve difficult problems we face in finance and investing. In the next chapter, we dive deeper into Monte Carlo methods and its applications to financial problems.

References

Misbehaving: The Making of Behavioral Economics by Richard Thaler, W. W. Norton & Company, 2014

Thinking, fast and slow by Daniel Kahneman, Farrar, Straus and Giroux, 2011

Who Discovered Bayes’s Theorem? by Stephen M. Stigler, The American Statistician, Volume 37, 1983 - Issue 4a

Probability Theory: The Logic of Science, Section 4.6.1, by E.T. Jaynes, Cambridge University Press, 2004

A History of Inverse Probability, Section 1.3, by A.I. Dale, Springer-Verlag, 1999

Probability Theory: The Logic of Science, Section 10.3, by E.T. Jaynes, Cambridge University Press, 2004

Dynamical Bias in the Coin Toss by Persi Diaconis, Susan Holmes and Richard Montgomery, Society for Industrial and Applied Mathematics (SIAM) Review, Vol.49, No.2, pp.211–235, 2007

Njå, O., Solberg, Ø., Braut, G.S. (2017). Uncertainty—Its Ontological Status and Relation to Safety. In: Motet, G., Bieder, C. (eds) The Illusion of Risk Control. SpringerBriefs in Applied Sciences and Technology(). Springer, Cham. https://doi.org/10.1007/978-3-319-32939-0_2

Wikipedia contributors, "Problem of induction," Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/w/index.php?title=Problem_of_induction&oldid=1093464035 (accessed July 4, 2022).

Statistical Rethinking, Chapter 1, by Richard McElreath, Taylor and Francis Group, LLC, 2016

Chapter 3. Quantifying Output Uncertainty with Monte Carlo Simulation

A NOTE FOR EARLY RELEASE READERS

With Early Release ebooks, you get books in their earliest form—the author's raw and unedited content as they write—so you can take advantage of these technologies long before the official release of these titles.

This will be the 3rd chapter of the final book. Please note that the GitHub repo will be made active later on.

If you have comments about how we might improve the content and/or examples in this book, or if you notice missing material within this chapter, please reach out to the author at EMAIL.

Monte Carlo Simulation (MCS), also known as the Monte Carlo Method, as we know it today was developed during the Second World War by some of the best mathematicians and physicists working on the nuclear weapons program in the US. Stanislaw Ulam, collaborating with Nicholas Metropolis and John von Neumann, invented the modern version of the method and implemented it using the ENIAC, the first programmable, electronic, general-purpose digital computer. Given the secretive nature of the weapons program, Metropolis code named the method Monte Carlo after the famous casino where Ulam's uncle would gamble away borrowed money. In the 18th century, Comte de Buffon had used a similar method as did the physicist Enrico Fermi in the 1930s.

The importance of MCS in finance cannot be overstated. It is used to value all types of assets, optimize diverse portfolios, estimate risks and evaluate complex trading strategies. It is especially used to solve problems that don't have an analytical solution[1]. Indeed, there are many types of financial derivatives—such as lookback options and Asian options—that cannot be valued using any other technique[2]. While the mathematics underpinning MCS is not simple, applying the method is actually quite easy, especially once you understand the key statistical concepts on which it is based.

MCS also pervades machine learning algorithms in general and probabilistic machine learning in particular. As discussed in Chapter 1 and demonstrated in Chapter 2, MCS enables you to quantify the uncertainty of a model's outputs in a process called forward propagation. It takes the traditional scenario and sensitivity analysis used by financial analysts to a completely different level.

You might be wondering how a method that uses random sampling can lead to a stable solution? Isn't that a contradiction in terms? In a sense it is. However, when you understand a couple of statistical theorems you will see that repetition of trials under certain circumstances tames randomness and makes it converge to a stable solution. In this chapter, we use MCS to provide a refresher on key statistical concepts and show you how to apply this powerful tool to solve real-world problems in finance and investing.

Monte Carlo Simulation: Proof of Concept

Before we begin down this path, how do we know that MCS actually works as described above? Let's do a simple proof of concept of MCS by computing the value of pi, a known constant. Figure 3-1 shows how we set up the simulation to estimate pi.

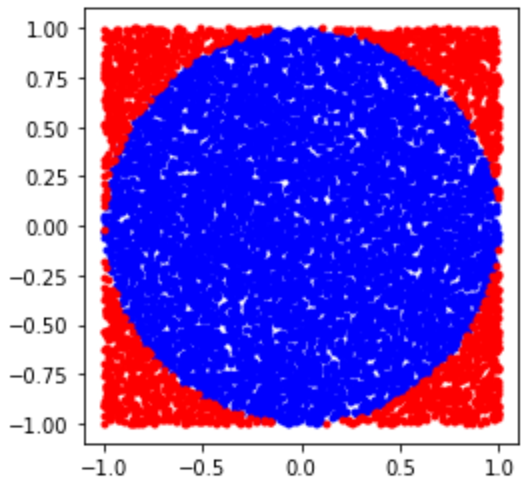

Figure 3-1. The blue circle of unit length in a red square with sides of 2 unit lengths is simulated to estimate the value of pi using MCS.

As the Python code shows, you simulate the random spraying of N points to fill up the entire square. Next we count M points in the circle of unit length r. The area of the circle is pi*r^2 = M. The length of the square is 2r so its area is 2r*2r = 4*r^2 = N. This implies the ratio of area of the circle to the area of the square is pi/4 = M/N. So pi = 4*M/N.

```
#Import modules
import numpy as np
from numpy import random as npr
import matplotlib.pyplot as plt
#Number of iterations in the simulation
n = 100000
#Draw random points from a uniform distribution in the X-Y plane to
fill the area of a square that has a side of 2 units
x = npr.uniform(low=-1, high=1, size=n)
y = npr.uniform(low=-1, high=1, size=n)
#Points with a distance less than or equal to one unit from the
origin will be inside the area of the unit circle. Using
Pythagoras's theorem c^2 = a^2 + b^2
inside = np.sqrt(x**2 + y**2) <=1
#We generate N random points within our square and count the number
of points that fall within the circle. Summing the points inside
the circle is equivalent to integrating over the area of the
circle.
#Note that the ratio of the area of the circle to the area of the
square is pi*r^2/(2*r)^2 = pi/4. So if we can calculate the areas
of the circle and the square, we can solve for pi
pi = 4.0*sum(inside)/n
#Estimate percentage error using the theoretical value of Pi
```

```
error = abs((pi-np.pi)/np.pi)*100
print("After {0} simulations, our estimate of Pi is {1} with an
error of {2}%".format(n, pi, round(error,2)))
#Points outside the circle are the negation of the boolean array
inside
outside = np.invert(inside)
#Plot the graph
plt.plot(x[inside], y[inside], 'b.')
plt.plot(x[outside], y[outside], 'r.')
plt.axis('square');
```

As in the Monty Hall simulation, you can see from the results of this simulation that the MCS approximation of pi is close to the theoretical value. Moreover, the difference between the estimate and the theoretical value gets closer as you increase the number of points N sprayed on the square. This makes the ratio of areas of the square and circle more accurate, giving you a better estimate of pi. Let's now explore the key statistical concepts that enable MCS to harness randomness to solve complex problems without analytical solutions.

Key Statistical Concepts

There are some very important statistical concepts that you need to understand so that you will have deeper insights into why MCS works and how to apply it to solve complex problems. These are also the concepts that provide the theoretical foundation of financial and other statistical models in general.

Mean and Variance

Figure 3-2 should refresh your memory of the basic descriptive statistical concepts you learned in high school.

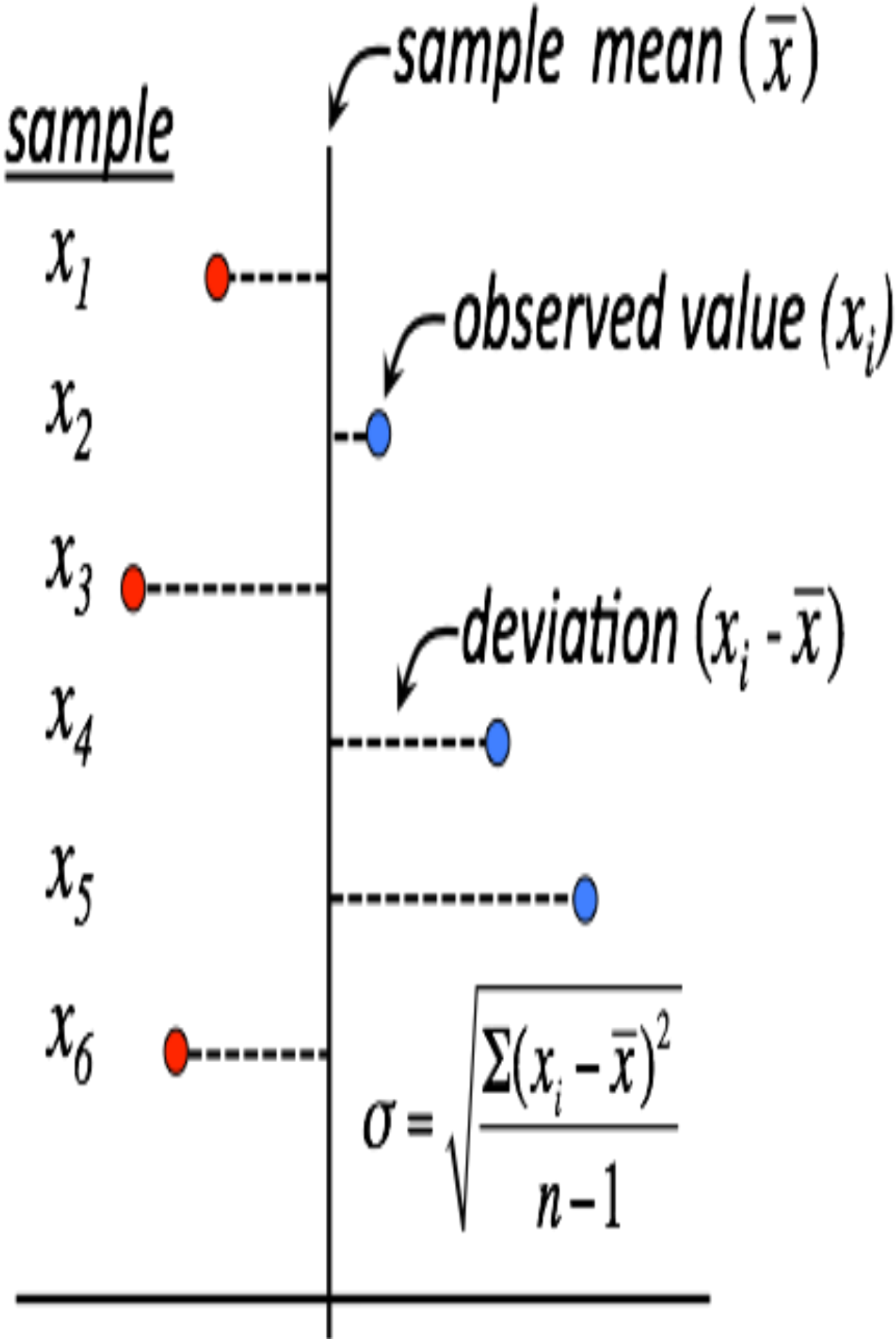
sample mean ($\bar{x}$)
sample
x_1
x_2
x_3
x_4
x_5
x_6
observed value (x_i)
deviation ($x_i - \bar{x}$)
$\sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}}$

Figure 3-2. Formulas for mean and standard deviation. Image source: Wikimedia Commons

The arithmetic mean is a measure of the central tendency of a sample of data points. It is simple to calculate: add up all the point values in a sample and divide the sum by the total number of points. It is important to note that the sum of all deviations from the mean of the values always equals zero. That is what makes the arithmetic mean a good measure of the central tendency of a sample. That is also why you have to square the deviations from the mean or use absolute values to get a sense of dispersion or spread of the data sample.

EXPECTED VALUE

An important type of arithmetic mean is the expected value of a trade or investment. Expected value is defined as a probability-weighted arithmetic mean of future payoffs:

- Expected value = (prob event 1 * event 1 payoff) + …… + (prob event N * event N payoff)

In finance, you should use expected value to estimate the returns of your trades and investments. Other measurements used for this purpose are incomplete or misleading. For instance, it is common to hear traders on financial news networks talk about the reward to risk ratio of their trades. That ratio is an incomplete metric to consider because it does not factor in the estimated probabilities of positive and negative payoffs. You can structure a trade to have any reward to risk ratio you want. It says nothing about how likely you think the payoffs are going to be. If reward to risk ratio is the key metric you're going to consider in an opportunity, don't waste your time with investing. Just buy a lottery ticket. The reward to risk ratio can go over 100 million to 1.

Note that the variance of a sample is calculated by adding the sum of the squared deviations and dividing by one less than the total number of points. The reason you use n-1 instead of n is that you have lost a degree of freedom by calculating the mean. Standard deviation, which is in the units of the mean, is obtained by taking the square root of variance. Volatility of asset price returns is calculated using the standard deviation of sample returns.

Gaussian or Normal Distribution

Gaussian distributions are found everywhere in nature. This is the most popular distribution in all of the sciences. That is why it is also called the normal distribution. Unfortunately, the data shows that normal distributions are not so common in financial markets. But that hasn’t stopped academics from using it for their theoretical models. Why? Because gaussian distributions lend themselves to elegant analytical formulas. If you know the mean and standard deviation of a gaussian distribution, you know everything about the distribution. For instance, in figure 3-3 below, you can see that about 68% of the data are within a standard deviation of the mean, 95% is within 2 standard deviations of the mean and almost all of the data are within 3 standard deviations of the mean.

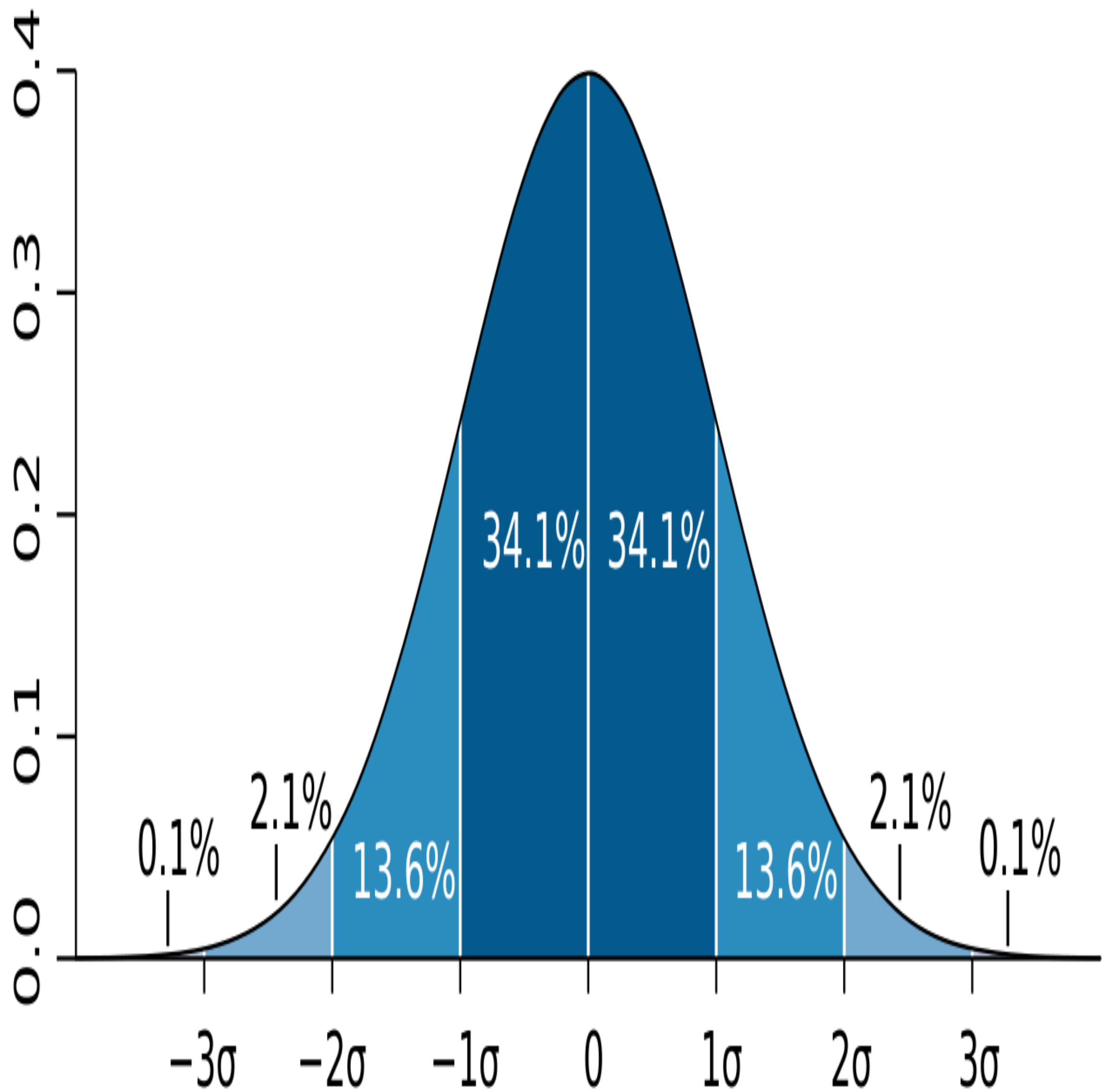

Figure 3-3. The gaussian or normal probability distribution. Image source: Wikimedia Commons

The S&P 500 is a global market index and is used by investors worldwide as a benchmark for the equity market. According to theoretical finance, asset price returns of the S&P 500 index should be normally or gaussian distributed. This implies that the skewness and kurtosis of its distribution should be approximately zero. Skewness tells you if the distribution is symmetrical and kurtosis tells you if the distribution has fat tails. Most financial time series are asymmetric and fat tailed. Fat-tailed distributions imply that low probability are more likely than would be expected if the distribution were gaussian.

These are not nice-to-know financial and statistical trivia. Asset price return distributions with negatively skewed, fat tails have the potential to bankrupt investors, corporations and entire economies if their modelers ignore them. The Great Financial Crisis is a recent reminder of the devastating consequences of building theoretical models using elegant mathematical equations that ignore the basic principles of the scientific method and the noisy, ugly, fat tailed realities of real world data.

The Law of Large Numbers

This is one of the most important statistical theorems. The law of large numbers (LLN) says that if samples are independent and drawn from the same distribution, the sample mean will almost surely converge to the theoretical mean as the sample size grows larger. In figure 3-4 below, the value of the sum of all the numbers that appear on each throw of a die divided by the total number of throws or trials approaches 3.5 as the number of trials increases.

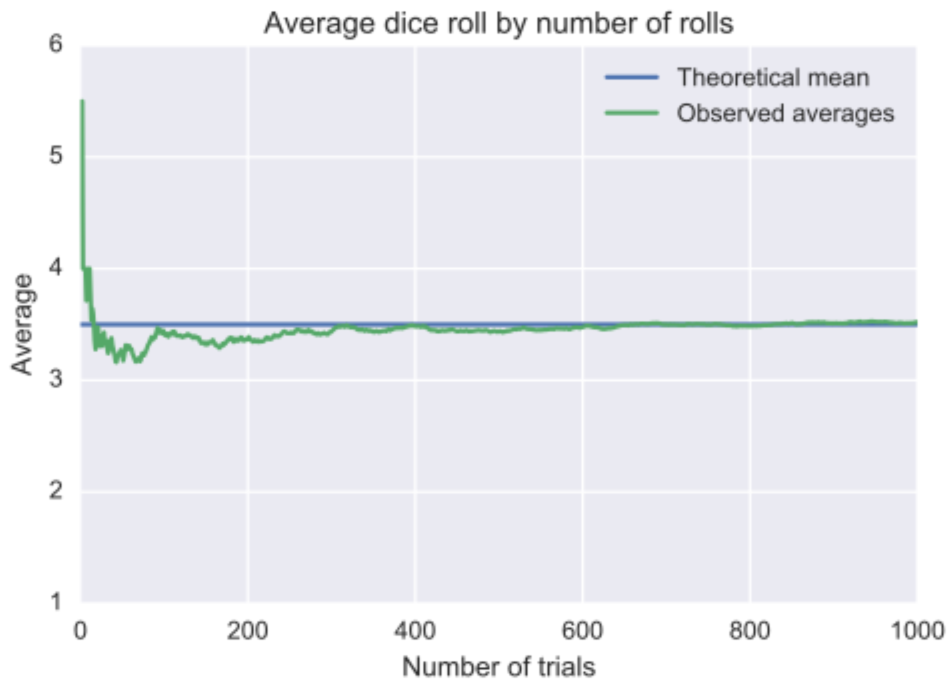

Figure 3-4. The sample mean of dice throws approaches its theoretical mean as the sample size gets larger. Image source: Wikimedia Commons

Note that the theoretical average does not have to be a physical outcome. There is no 3.5 on any fair die. Also, notice how the outcome of the first few trials vary widely about the mean. However, in the long run they converge inexorably to the theoretical mean. Of course, we assume that the die is fair and that we don't know the physics of the dice throws.

The Central Limit Theorem

The central limit theorem (CLT) says that if you keep taking samples from an unknown population of any shape and calculate the mean of each of the samples of size n, the distribution of these sample means will be normally distributed as shown in Figure 3-6.

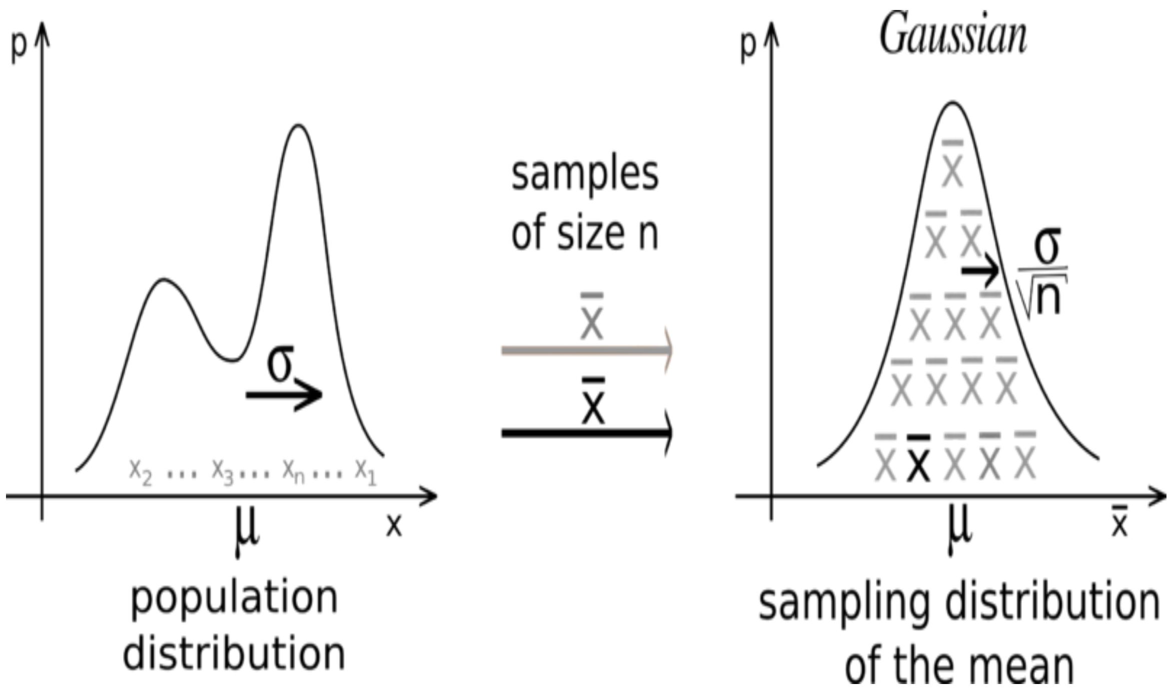

Figure 3-5. The sampling distribution of the sample mean is normally distributed. Image source: Wikimedia Commons

This is one of the most amazing statistical phenomena. To appreciate the power of the CLT consider a fair die which has a uniform distribution since each number on the die is equally likely at ⅙ . Figure 3-6 shows what happens when you roll two fair dice and add the numbers that show up on each throw and repeat the trials many times. Behold the magic of the CLT: horizontal lines made into a bell curve.

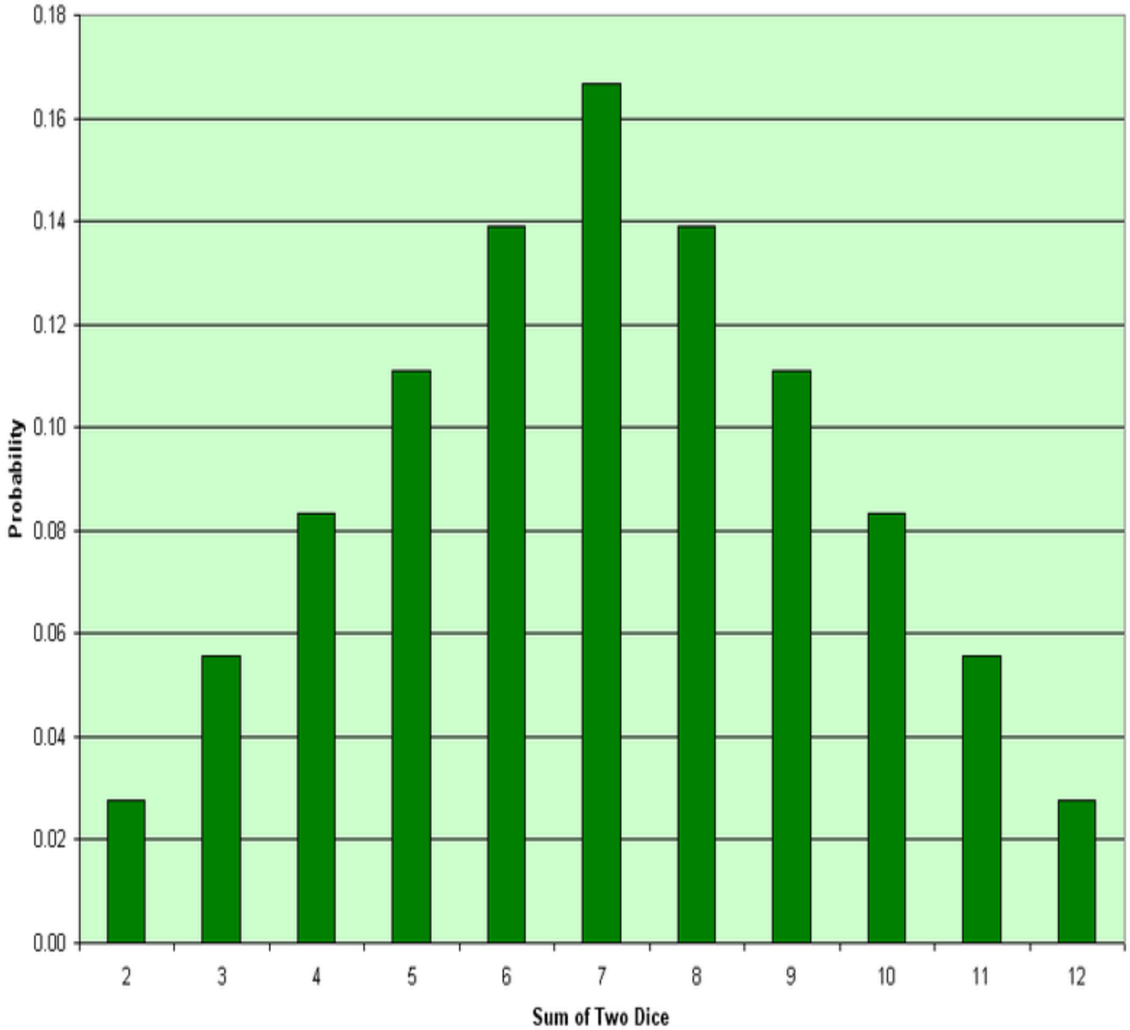

Figure 3-6. The CLT shows us how uniform distributions of two dice are transformed into a gaussian distribution. Image source: Wikimedia Commons

Theoretical Underpinnings of MCS

MCS is based on the two most important theorems in statistics mentioned above: the law of large numbers (LLN) and the central limit theorem (CLT) [3]. Recall that LLN ensures that as the number of trials increases, the sample mean almost surely converges to the theoretical or population mean. The CLT ensures that the sampling errors or fluctuations of the sample averages from the theoretical mean become normally distributed as sample sizes get larger.

One of the reasons MCS works and is scalable to multidimensional problems is that the sampling error is independent of the dimension of the variable.This sampling error approaches zero asymptotically as the square root of the sample size and not the dimension of the variable. This is very important. It implies that sampling error in an MCS is the same for a single variable as it for a 100 dimensional variable.

However, the error decreases as the square root of the sample size n. So you have to increase the MCS iterations by a factor of 100 to increase the accuracy of its estimate by a single digit. But with computing power becoming cheaper by the day, this is not as big an issue now as it was in the last century.

Valuing a Software Project

Let's increase our understanding of MCS by applying it to a real-life financial problem such as valuing a project. The discounted cash flow (DCF) model is used extensively in corporations worldwide for valuing projects and other investments like bonds and equities. A discounted cash flow (DCF) model forecasts expected free cash flows (FCF) over N periods, typically measured in years. FCF in a time period equals cash from operations minus capital expenditures. The model also needs an estimate of the rate of return (R) per period required by the firm's investors. This rate is called the discount rate because it is used to discount each of the N period FCFs of the project to the present. The reason the FCFs are discounted is that we need to account for an investor's opportunity cost of capital for undertaking the project instead of another investment of similar risk. The model is setup in four steps:

1. Forecast the free cash flows (FCF) of the project for each of the N periods
2. Estimate the appropriate discount rate (R) per period
3. Discount each period's FCFs back to the present

4. Add of the discounted FCFs above to get the net present value (NPV):

   NPV of project = FCF0 + FCF1/(1+R) + FCF2/(1+R)^2 + ...+ FCFn/(1+R)^N

The NPV decision rule says that you should accept any investment whose expected NPV is greater than zero. This is because an investment with a positive NPV gives investors a higher rate of return than an alternative investment of similar risk.

In order to create our DCF model, we need to focus on the main drivers of costs and revenues for our software project. We also need to make sure that these variables are not strongly correlated with one another. Ideally, all FCFs of the model should be formulated using very few, non-correlated variables or risk factors.

As you know, software development is labor intensive and so our main cost driver will be salaries and wages. Also, some developers will be working part-time and some full-time on the project. However, for developing the cost of labor, we only need the full-time equivalent (FTE) of the effort involved in producing the software i.e. we estimate the effort as if all required developers will work full-time. Scheduling is where we will figure out how much time and when we will need each developer.

```
#Import key Python libraries and packages that we need to process
and analyze our data
import pandas as pd
from datetime import datetime
import numpy as np
from numpy import random as npr
import matplotlib.pyplot as plt
plt.style.use('seaborn')
#Specify model constants per full-time equivalent (fte)
daily_rate = 400
technology_charges = 500
overhead_charges = 200
#Specify other constants
tax_rate = 0.15
#Specify model risk factors that have little or no correlation
among them.
#Number of trials/simulations
n = 10000
```

```
#Number of full-time equivalent persons on the team
fte = npr.uniform(low=1, high=5, size=n)
#In person days and driven independently by the scope of the
project
effort = npr.uniform(low=240, high=480, size=n)
#Based on market research or expert judgment or both
price = npr.uniform(low=100, high=200, size=n)
#Independent of price in the price range considered
units = npr.normal(loc=1000, scale=500, size=n)
#Discount rate for the project period based on risk of similar
efforts
discount_rate = npr.uniform(low=0.06, high=0.10, size=n)
#Specify how risk factors affect the project model
labor_costs = effort * daily_rate
technology_costs = fte * technology_charges
overhead_costs = fte * overhead_charges
revenues = price * units
#Duration determines the number of days the project will take to
complete assuming no interruption. Different from the elapsed time
of the project.
duration = effort/fte
#Specify target_value
free_cash_flow = (revenues - labor_costs - technology_costs -
overhead_costs) * (1 - tax_rate)
#Simulate project NPV assuming initial FCF=0
npv = free_cash_flow/(1 + discount_rate)
#Convert numpy array to pandas dataframe for easier analysis
NPV = pd.DataFrame(npv, columns=['NPV'])
#Estimate project duration in days
Duration = pd.DataFrame(duration, columns=['Days'])
#Plot histogram of NPV distribution
plt.hist(NPV['NPV'], bins=50), plt.title ('Distribution of Project
NPV'), plt.xlabel('Project NPV'), plt.ylabel('Frequency'),
plt.show();
print(NPV.describe().round())
success_probability = sum(NPV['NPV'] > 0)/n *100
print('There is a {0}% probability that the project will have a
positive NPV.'.format(round(success_probability)))
#Plot histogram of project duration distribution
plt.hist(Duration['Days'], bins=50), plt.title ('Distribution of
Project Duration'), plt.xlabel('Days'), plt.ylabel('Frequency'),
plt.show();
print(Duration.describe().round())
```

Note that we did not discount the FCF distributions at the risk free rate. The risk free rate is the interest rate on a government security such as the US 10 year note. It is a common mistake in NPV simulations but is incorrect since each simulation is estimating the expected value of the FCF. Each FCF needs to be discounted at the risk-adjusted discount rate to account for total risk of the project.

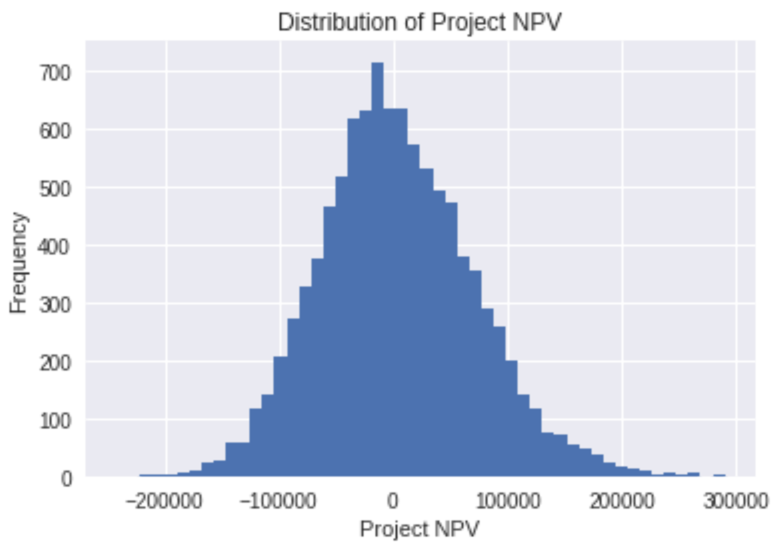

Figure 3-7. Distribution of risk-adjusted NPVs in the code output needs to be interpreted with caution. Using the dispersion of NPVs to make decisions would double count project risk. Using dispersion of NPVs adjusted at the risk-free rate to account for total risk has no sound theoretical basis in corporate finance.

Building a Sound MCS

In order to harness the power of MCS to solve complex financial and investment related problems, it is important that you follow a sound and replicable process. Here is a 0 step process for doing just that:

1. Formulate how dependent variables of your model are affected by independent variables, also called risk factors in finance.

2. Specify the probability distribution of each risk factor. Some common ones include gaussian, student-t, cauchy, binomial probability distributions.

3. Specify initial values and how time is discretized, such as seconds, minutes, days, weeks or years.
4. Specify how each risk factor changes over time if at all.
5. Specify how each risk factor is affected by other risk factors. This is important since correlation among risk factors can incorrectly amplify or dampen effects. This phenomenon is also called multicollinearity.
6. Let the computer draw a random value from the probability distribution of each independent risk factor
7. Compute the value of each risk factor based on that random value
8. Compute target/dependent variables based on the computed value of all risk factors
9. Iterate steps 6-8 as many times as necessary
10. Record and analyze descriptive statistics of all iterations.

The power of MCS is that it transforms a complex, intractable problem that involves integral calculus into a simple one of descriptive statistics with its sampling algorithms.

There are many challenges to building a sound MCS. The most important ones are:

- Specifying how each independent variable changes over time. Serial correlation (also known as auto correlation) is the correlation of a variable with an instance of itself in the past. This correlation is not constant and usually changes over time, especially in financial markets.
- Specifying how each variable is affected by other variables of the model. Correlations among input variables/risk factors usually change over time.
- Fitting a theoretical probability distribution to the actual outcomes. Probability distributions of variables usually change over time

- Convergence to the best estimate is non-linear making it slow and costly. It may not occur quickly enough to be of any practical value to trading or investing

These challenges can be met as follows:

- Rigorous data analysis, domain knowledge and industry expertise. You need to balance rigorous financial modeling with time, cost and effectiveness of the models that you produce.
- Treat all financial models as flawed and imperfect guides. Don't let the mathematical jargon intimidate or lull you into a false sense of security. Remember the adage 'All models are wrong, but some are useful'.
- Managing risk is of paramount importance. Always size capital positions appropriately, have wide error margins and fall back plans if models fail.
- Clearly, there is no substitute for managerial experience and business judgment. Rely on your common sense, be skeptical and ask difficult questions of a model's assumptions, inputs and outputs.

Summary

Fundamentally, MCS is a set of numerical techniques that uses random sampling of probability distributions for computing approximate estimates or for simulating uncertainties of outcomes of a model. The central idea is to harness the statistical properties of randomness to develop approximate solutions to complex deterministic models and analytically intractable problems. It transforms a complex, often intractable, multi-dimensional problem in integral calculus into a much easier problem of descriptive statistics that any practitioner can use.

MCS is especially useful when there is no analytically tractable solution to a problem you are trying to solve. It enables you to quantify the probability and impact of all possible outcomes given your assumptions. It should be used when the traditional analysis of best, worst and base case scenarios may be inadequate for your decision-making and risk management. MCS gives you a better understanding of the risk of complex financial models.

References

Handbook in Monte Carlo Simulation: Applications in Financial Engineering, Risk Management, and Economics by Paolo Brandimarte, John Wiley & Sons, Inc., 2014

Finance with Monte Carlo by Ronald W. Shonkwiler, Springer Science+Business Media, New York, 2013

A Tutorial Introduction to Monte Carlo Methods, Markov Chain Monte Carlo and Particle Filtering by A. Taylan Cemgil in Signal Processing Theory and Machine Learning, Volume 1, First Edition, Academic Press, Elsevier Ltd, 2014.

Quantitative Methods, CFA Program Curriculum, Volume 1, CFA Institute, 2007

Chapter 4. Errors in Quantifying Uncertainty with Conventional Statistical Methods

A NOTE FOR EARLY RELEASE READERS

With Early Release ebooks, you get books in their earliest form—the author's raw and unedited content as they write—so you can take advantage of these technologies long before the official release of these titles.

This will be the 4th chapter of the final book. Please note that the GitHub repo will be made active later on.

If you have comments about how we might improve the content and/or examples in this book, or if you notice missing material within this chapter, please reach out to the author at EMAIL.

Recall from Chapter 1 that all financial models are at the mercy of the Trinity of Errors, namely: errors in model specifications, errors in model parameter estimates, and errors resulting from the failure of a model to adapt to structural changes in its environment. Because of these errors, we need dynamic models that quantify the uncertainty inherent in our financial estimates and predictions.

A statistical inference methodology known as Null Hypothesis Significance Testing (NHST) almost completely dominates the research and practice of social and economic sciences. In this chapter, we examine how NHST and its p-value statistic is used for hypothesis testing and quantifying uncertainty of model parameters. The deep logical flaws of NHST methodology is primarily responsible for the reproducibility crisis in all of the social and economic sciences, where almost all published research findings are false [1][2]. In the next couple of sections, we expose the statistical skullduggery of NHST and its p-value statistic and show you how it is guilty of the prosecutor's fallacy. This fallacy is another version of the inverse fallacy, where a conditional statement is falsely equated with its inverse, thereby violating the inverse probability rule.

Given the deep flaws and abuses of p-values for quantifying parameter uncertainty [3][4], another methodology known as confidence intervals (CIs) is touted by orthodox statisticians as its mathematically rigorous replacement. Unfortunately, CIs are also the wrong tool for data analysis since it was not designed to make statistical inferences from a single experiment [5]. Most importantly, the application of CIs in finance generally violates the assumptions of the Central Limit Theorem (CLT) making CIs invalid. In this chapter, we explore the trifecta of errors in applying CIs that are common in financial research and practice. We develop an ordinary least squares (OLS) linear regression model of equity returns using Statsmodels, a Python statistical package, to illustrate these three error types. We use the diagnostic test results of our regression model to support our reasons why CIs should not be used in financial data analyses.

The Inverse Fallacy

Recall the proof of the inverse probability rule, a trivial reformulation of the product rule. For any non-zero probability events H and D:

P(H and D) = P(D and H) (product of probabilities commute)

P(H|D) * P(D) = P(D|H) * P(H) (applying product rule to both sides)

P(H|D) = P(D|H) * P(H) / P(D) (the inverse probability rule)

Note that joint probabilities, the product of two probabilities, commute i.e. the order of the individual probabilities does not change the result of their product.

P(H and D) = P(D and H)

As you can see from the last equation above, conditional probabilities do not commute:

P(H|D) =\= P(D|H)

This is a common logical mistake that people make and scientists continue to make when using NHST and p-values. This is called the inverse fallacy because you are incorrectly equating a conditional probability, P(H|D), with its inverse, P(D|H) and violating the inverse probability rule. The inverse fallacy is also known as transposed conditional fallacy or the prosecutor's fallacy. As a simple example, consider how the inverse fallacy incorrectly infers statement B from statement A below:

Given that someone is a programmer, it is likely that they are analytical

Given that someone is analytical, it is likely that they are a programmer

But P(analytical | programmer) =\= P(programmer | analytical). As you know, there are many, many analytical people who are not programmers and such an inference seems absurd when framed in this manner. However, you will see that humans are generally not very good at processing conditional statements and their inverses, especially in complex situations. Indeed, prosecutors have ruined people's lives by using this flawed logic in arguments that have led judges and juries to make terrible inferences and decisions [6]. The prosecutor's fallacy goes something like this:

About 0.1% of the 100,000 adults in your city have your blood type

A blood stain with your blood type is found on the murder victim

Therefore, claims the city prosecutor, there is a 99.9% probability that you are the murderer

That's clearly absurd. What is truly horrifying, and we should all be screaming bloody murder, is that researchers and practitioners are unknowingly using the prosecutor's fallacious logic when applying NHST and p-values for their statistical inferences. More on NHST in the next section. Let's expose the prosecutor's flawed reasoning in this section so that you can see how it is used in the NHST methodology.

The probability of your guilt (G) before the blood stain evidence (E) was discovered is $P(G) = 1/100{,}000$ since every adult in the city is an equally likely suspect. Therefore, the probability of your innocence (I) is $P(I) = 99{,}999/100{,}000$. The probability that the blood stain would match your blood type given you are actually guilty is a certainty, implying $P(E \mid G) = 1$. However, even if you are actually innocent, there is still a 0.1% probability that the blood stain would match your blood type merely by its prevalence in the city's adult population implying $P(E \mid I) = 0.001$. The prosecutor needs to estimate $P(G \mid E)$, the probability of your guilt given the evidence, with the above probabilities. Instead of using the inverse probability rule, the prosecutor uses a specious argument as follows:

(A) Given the evidence, you are either guilty or innocent so $P(G \mid E) + P(I \mid E) = 1$

(B) Now the prosecutor commits the inverse fallacy by making $P(I \mid E) = P(E \mid I)$

(C) Thus $P(G \mid E) = 1 - P(I \mid E) = 1 - P(E \mid I)$ which is termed the prosecutor's fallacy

(D) Plugging in the numbers, $P(G \mid E) = 1 - 0.001 = 0.999$ or 99.9%

Without explicitly using the inverse probability rule, your lawyer could use some common sense and correctly argue that there are 100 adults (0.1% * 100,000) in the city who have the same blood type as you do. Therefore, given evidence of the blood stain alone, there is only a 1 in 100 chance or 1% probability that you are guilty and 99% probability that you are innocent. This is approximately the same probability you would get when applying the inverse probability rule because its just formulates a commonsensical way of counting the possibilities. Let's do that now and calculate the probability of your innocence given the evidence $P(I \mid E)$:

(A) The inverse probability rule states $P(I \mid E) = P(E \mid I) * P(I) / P(E)$

(B) We use the law of total probability to get $P(E) = P(E \mid I)*P(I) + P(E \mid G) * P(G)$

(C) Therefore, $P(I \mid E) = 0.001 * 0.99999 / [(0.001 * 0.99999) + (1 * 0.00001)] = 0.99$

Before the prosecutor strikes you off the suspect list, it is important to note that it would also be fallacious for your lawyer to now ask the jury to disregard the blood stain as weak evidence of your guilt based on the 1% conditional probability calculated above. This flawed line of defense is called the defender's fallacy and was used in the notorious O.J. Simpson murder trial. The evidence is not weak because before the blood stain was found, you had a 1 in 100,000 chance of being the murderer. But after the blood stain was discovered, your chance of being guilty has gone up a thousand times to 1 in 100. That's very strong evidence indeed and nobody should disregard it. However, it is completely inadequate for a conviction if that is the only piece of evidence presented to the jury. The prosecutor will need additional incriminating evidence to make a valid case against you.

Now let's look at a realistic financial situation where the inverse fallacy might be harder to spot. Economic recessions are notoriously hard to recognize in the early stages of their development. As I write this chapter, there is a debate raging among economists and investors about whether the US economy is currently in a recession or about to enter one. Economists at the National Bureau of Economic Research (NBER), the organization responsible for making the recession official, can only confirm the fact in retrospect. Sometimes the NBER takes over a year to declare when the recession actually started as they did in the Great Recession of 2007-09. Of course traders and investors cannot wait that long and they develop their own indicators for predicting recessions in real time.

Say you have developed a proprietary economic indicator that correctly signals a recession 95% of the time when the US economy is actually in one or about to enter one. You also note that about 10% of the time your indicator signals a recession incorrectly. Assume R is the scenario that the US economy is in a recession and S is the event that your indicator signals a recession. The conditional probability of your indicator giving you a recession signal if we are actually are in one, its true positive rate, is $P(S|R) = 0.95$. Therefore the probability that your indicator fails to detect a recession when we are in one, $P(\text{not } S|R) = 1 - P(S|R) = 0.05$. This is the false negative rate of your indicator. Your indicator's false alarm rate or false positive rate is $P(S|\text{not } R) = 0.1$ because it incorrectly tells you there is a recession when there isn't one. Similarly, the probability that your indicator successfully detects that the economy is not in a recession, $P(\text{not } S| \text{ not } R) = 1 - P(S|\text{not } R) = 0.90$. This is the true negative rate of your indicator.

Say you just found out that your proprietary indicator is flashing a recession signal. What is the probability that the US economy has actually entered into a recession? If you answered that the probability was 95%, as many people instinctively do, you would have committed the inverse fallacy since P(R|S) =\= P(S|R).

To calculate the inverse probability $P(R|S)$, you can't only let the data speak about one specific scenario. Why? Because your economic indicator does not have 100% accuracy. It gives you a recession signal 10% of the time (false positive rate) when the economy is not in one and 5% of time it fails to detect a recession when the economy is actually in one (false negative rate). Could this scenario be 1 of the 10 when it is wrong about the economy being in a recession? Or maybe we are already in a recession and your indicator didn't flicker in the earlier month? How would you know?

You don't even know how common or uncommon recessions are in the US. Why is that relevant? Because you don't know if your false positive rate is too high or low enough compared to the rate at which recessions tend to occur in the US for your indicator to be any useful. You need to estimate the probability that the US could be in a recession in any given month $P(R)$ based on past occurrences before looking at the signal from your indicator. In general, $P(R)$ is also called the base rate of the particular event/scenario R or the prior probability of R.

Ignoring the base rate is called base rate neglect and leads to a violation of the inverse probability rule and invalid inferences as we will demonstrate. Once you have the estimate of P(R), you would plug it into the law of total probability to get the unconditional probability, or marginal probability P(S), of getting a recession signal from your indicator regardless of the state of the economy.

Let’s compute it using actual economic data. The NBER has a time series for every month since 1982 that the US was in an economic recession which we can download from Federal Reserve Economic Data (FRED). Let’s use the following code to calculate the monthly base rate of economic recessions in the US.

```
#Import libraries and FRED datareader
import numpy as np
import pandas as pd
import pandas_datareader.data as pdr
from datetime import datetime
start = datetime(1982, 1, 1)
end = datetime(2022, 7, 30)
#NBER business cycle classification
recession = pdr.DataReader('USREC', 'fred', start, end)
#Percentage of time the US economy was in recession since 1982
round(recession['USREC'].sum()/recession['USREC'].count()*100, 2)
```

From the data above, the US has been in an economic recession only 9.67% of the time in any given month since 1982. Let’s plug in the numbers into the law of total probability and the inverse probability rule to calculate the probability the US economy is in recession given that your proprietary indicator is signaling a recession with a 95% true positive rate and a 10% false positive rate:

$P(S) = P(S|R)*P(R) + P(S|\text{not } R)*P(\text{not } R) = (0.95*0.0967) + (0.1*0.92) = 0.184$

$P(R|S) = P(S|R) * P(R) / P(S) = (0.95*0.0967)/0.184 = 0.499$ or 49.9%

The calculation for P(S) above says that you can expect your indicator to generate a signal 18.4% of the time regardless of whether the US economy is in a recession or not. Of the times you do see it flicker, P(R|S) says that in only 49.9% of those situations will the signal be correct about the economy being in a recession. Your signal will give you a false positive P(not R|S) about 50.1% of the time. That's a poor indicator - you're better off tossing a fair coin (without knowing the initial conditions of the toss of course).

This result seems counterintuitive since your indicator has a 95% true positive rate P(S|R). That's because you cannot shove your indicator's false positive rate under the rug and blithely ignore the base rate of US economic recessions with some rubbish of being objective and not including subjective prior knowledge. That would be idiocy because you would be denying the inverse probability rule and ignoring objective prior data about US economic cycles. Such fallacious inferences and decision-making will almost surely see you go broke or be out of a job sooner rather than later.

In the real world of finance and investing, you will need a signal with a false positive rate lower than the base rate to give you a signal with probability greater than 50% of being correct. To see this, let's redo the calculation for a revised false positive of 9%, which is slightly less than the 9.67% base rate at which the US economy has been in a recession in any given month since 1982:

$P(S) = P(S|R)*P(R) + P(S|\text{not } R)*P(\text{not } R) = (0.95*0.0967) + (0.09*0.91) = 0.174$

$P(R|S) = P(S|R) * P(R) / P(S) = (0.95*0.0967)/0.174 = 0.529$ or 52.9%

Now your indicator will have an edge or positive expectation for better decision-making.

To summarize, the true positive rate of your indicator is important. However, what is equally important is that the false positive rate of the indicator needs to be less than the base rate of the underlying feature in the population you are sampling from. So if you ignore the fact that your indicator is generating false positives P(S|not R) at a 10% rate while the US economy is generating a recessionary month at a 9.67% rate, its base rate P(R), you would be falling into the inverse fallacy trap. It doesn't seem so far-fetched now to think that unscrupulous prosecutors, snake oil salesmen and pseudo-scientists could fool you (and themselves) with the inverse fallacy.

Randomness could also fool you too by granting you a lucky guess and the US economy could end up being in a recession. However, your probability estimate of 95% would be way off and your reasoning would be fallacious. A trading or investing strategy based on luck, incorrect reasoning and poor probability estimates will lead to financial ruin sooner rather than later. Far worse, a statistical methodology like NHST based on the inverse fallacy will overwhelm us with false positive studies and ruin the scientific enterprise that we cherish and value so much.

NHST Is Guilty of the Prosecutor's Fallacy

Ronald Fisher, the head architect of modern statistics, introduced NHST in the 1920s. He also included Karl Pearson's p-value into his methodology for quantifying uncertainty. This was a post-data methodology and was meant to enable researchers to make statistical inferences from a single experiment based on a null hypothesis that was the negation of the hypothesis that the researcher is trying to prove.

In 1925, Fisher made the absurd and unsubstantiated claim that 'the theory of inverse probability is founded upon error, and must be wholly rejected' [7]. Of course Fisher didn't provide any proof for this claim and he expected everyone to take his word for it. By rejecting the inverse probability rule, Fisher was able to use the prosecutor's fallacy to promote his flawed discriminatory ideas under the guise of objectivity and 'letting the data speak for themselves' [8]. Fisher's fawning cohorts in industry and slavish acolytes in academia merely repeated the lie about the inverse probability theory and banished it from their practice and curricula which continues to this day.

NHST is built behind the facade of a valid form of propositional logic known as proof by contrapositive. The logic is as follows: suppose we have two propositions H and D such that if H is true, then D is true. Now if we can prove that D is false, then we can validly conclude that H must be false.

Following the above logic, researchers using NHST formulate a hypothesis, called the null hypothesis (H0), that they want to disprove before observing any data. H0 is viewed as the negation of an alternative hypothesis (H1) that they want to establish but is not explicitly specified i.e. H1 = not H0 and P(H1) = P(not H0) = 1 - P(H0). In this regard, they play the devil's advocate for the null hypothesis.

The null hypothesis is generally formulated as a summary statistic, such as the difference in the sample means of the data distribution of two groups that need to be compared. It is important to note that researchers do not predict the data that their research hypothesis H1 is expected to generate, assuming that H1 is true.

Before starting their experiment, researchers also choose a significance level, denoted by alpha, which works as a decision threshold to accept or reject the null hypothesis after observing the data. The convention is to set alpha to 5%. The alpha level is claimed to be the long run probability that the researcher might incorrectly reject a true null hypothesis, thereby committing a Type I error and generating false positive results (a result that is claimed to be true when it is actually false). The alpha level is the most critical element of the experiment since it determines if the experiment is considered statistically significant or not.

It is important to note that any significance level is entirely subjective as it is not based on the observed data or the null hypothesis or a scientific reason or any mathematical rule. The conventional use of the 5% alpha level is a totally arbitrary and self-fulling ritual. Since Fisher used a 5% alpha significance level, researchers and academics blindly follow his example. So much for the vaunted objectivity and scientific rigor of frequentists.

Assuming the null hypothesis is true, the researcher computes a statistic called the p-value to quantify the probability of observing the summary statistic of the sample data (D) or something more extreme than it:

p-value = P(D|H0)

If the p-value <= alpha, H0 is rejected as false at the alpha significance level and the alternative hypothesis (H1) is accepted as true.

But this logic of NHST is absurd. By rejecting the null hypothesis (H0) given the p-value of the test statistic (D), the researcher is committing the inverse fallacy because P(H0 | D) =\= P(D | H0). See Figure 4-1 below.

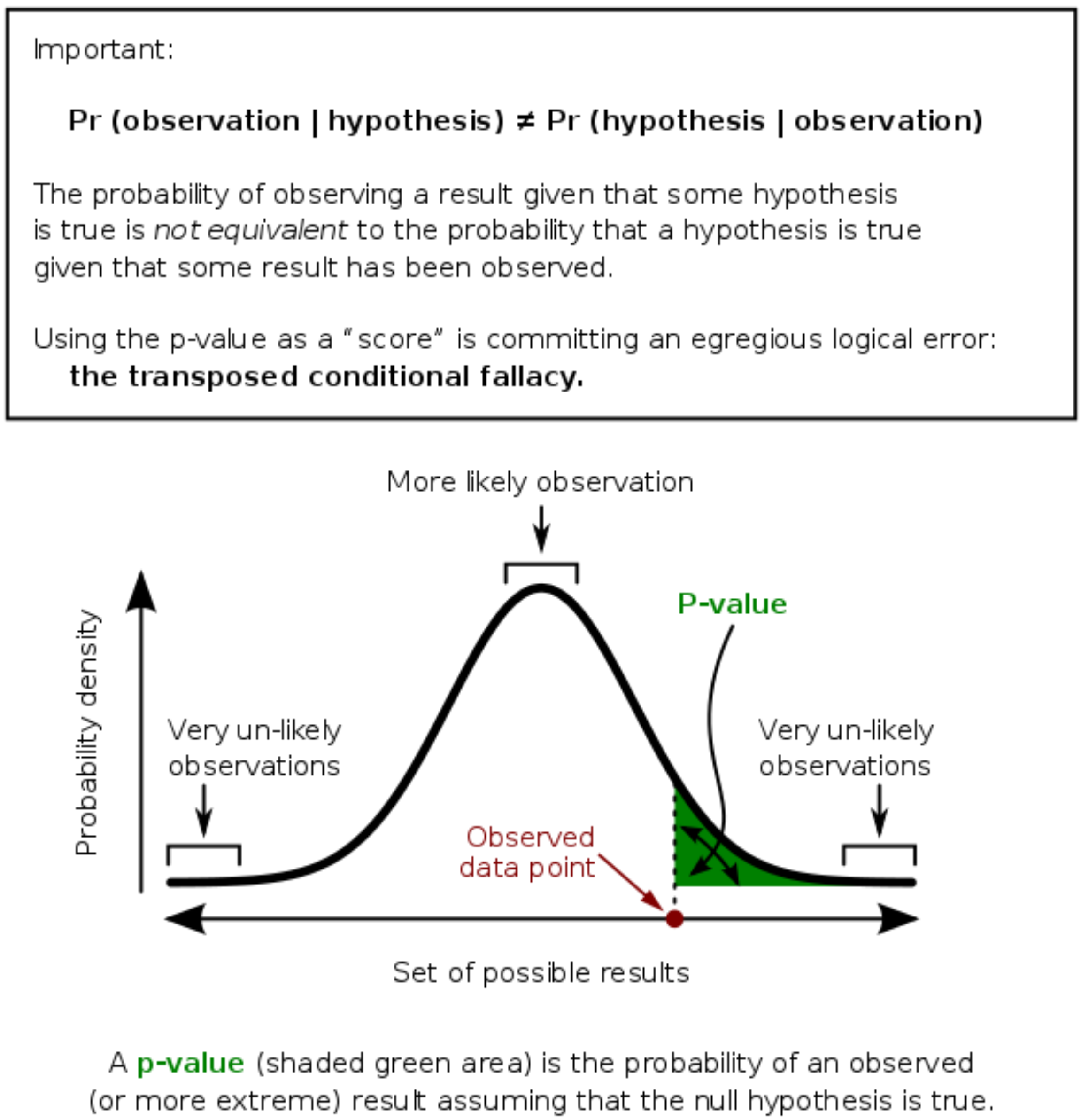

Figure 4-1. How p-values are used in NHST. Image Source: Wikimedia Commons.

NHST makes an even more absurd leap of logic by allowing researchers to accept the unspecified, complementary alternative hypothesis which the data was not modeling in the first place. The researcher wants to determine P(H1|D). But NHST only computes P(D|H0). So researchers following the NHST methodology commit the prosecutor's fallacy as follows:

$P(H1|D) = 1 - P(H0|D)$ (true statement)

$P(H0|D) = P(D|H0)$ (the inverse fallacy)

$P(H1|D) = 1 - P(D|H0)$ (the prosecutor's fallacy)

How should we validly calculate $P(H1|D)$? The binary logic of proof by contrapositive in a deterministic world needs to be translated into the calculus of conditional probabilities in an uncertain world. This translation is enabled by the inverse probability rule and the law of total probability as was applied in the previous section:

$P(H1|D) = 1 - P(H0|D)$

$P(H1|D) = 1 - [P(D|H0)P(H0)/P(D)]$

$P(H1|D) = 1 - [P(D|H0)P(H0) / (P(D|H0)P(H0) + P(D|H1)P(H1))]$

As the equation above shows, the researcher needs to examine $P(D|H1)$, the probability of observing the data assuming their research hypothesis H1 is true. This will also enable the researcher to compute the experiment's false negative errors. Most importantly, the researchers need to estimate the prior probability or base rate of at least one of their complementary hypotheses, $P(H0)$ or $P(H1)$. That's because without the base rates, p-values are useless and cannot make valid inferences about either the null hypothesis or the alternative, research hypothesis. It is for very good reason that Jerzy Neyman, an eminent statistician and Fisher's peer, called Fisher's work on statistical inference 'worse than useless' [9].

It is clear that NHST, the cornerstone of social science education, research and practice, is committing the prosecutor's fallacy. No wonder most of the published research using NHST is false. NHST has wasted billions of research dollars, defamed science and done a great disservice to humanity with its false positive research studies. All this while professing the farce of rigor and objectivity. NHST continues to wreak havoc on the social and economic sciences producing too many false research claims to this day despite many failed attempts to abolish it or reform it for almost a century [10]. It's about time we reject the NHST because it 'is founded upon error and must be wholly rejected'.

HOW NHST BECAME CHILDISH, HORRIFYING, AND WORSE THAN USELESS

Soon after Fisher introduced NHST, Jerzy Neyman and Egon Pearson, the principal architects of modern statistics, introduced their methodology of statistical hypothesis testing as a decision framework for industrial quality control.

There was a bitter rivalry between Fisher and Neyman as they advocated for their respective methodologies. Neyman called Fisher's statistical work 'worse than useless' and Fisher called Neyman's work 'childish' and 'horrifying'. Unfortunately, both Neyman and Fisher were right. Fisher's work on statistical inference is actually worse than useless because he rejected the inverse probability rule and unabashedly committed the prosecutor's fallacy.

Fisher was right in that Neyman's work is not applicable to social sciences but to industrial quality control. It would be childish and horrifying to apply it to social systems of creative and free willed humans instead of factory widgets. Fisher also didn't understand that Neyman's theory was a pre-data theory and it would be absurd if it were applied as a post-data theory of statistical Inference.

As the two competing methodologies spread throughout the social and economic sciences, researchers tried to reconcile the ideas of the two bitter rivals. Unfortunately, these social scientists and statisticians did not have a deep understanding of either methodology [11]. This inept fusion of the two methodologies is how NHST became childish, horrifying and worse than useless.

Many in the social and economic sciences recommend replacing p-values with CI theory which is touted as a mathematically more rigorous way of quantifying uncertainty. So let's examine CI theory to see if it is useful.

CI Theory Is a Pre-Data Theory

As mentioned in the sidebar above, Jerzy Neyman developed his statistical decision theory designed to support industrial quality control. His statistical theory provides a decision framework that seeks to control Type I (false positive) and Type II (false negative) errors in order to balance costs versus benefits over the long run based on many experiments. Neyman intentionally left out p-values because it was a nonsensical concept violating basic probabilistic logic.

In 1937, Neyman developed CI theory to be a *pre-data theory* of statistical inference, intended to inform statistical procedures that have long-run average properties *before* data are sampled from a population distribution. Neyman made it very clear that his CI theory was not intended to support inferences *after* data are sampled in a single scientific experiment. CI theory is not a *post-data theory of* statistical inference despite how it is applied today in research and practice in social and economic sciences.

CI theory quantifies uncertainty of population parameter estimates. For example, a 90% confidence interval (CI), as shown in Figure 2, is generally understood to imply that there is a 90% probability that the true value of a parameter of interest is in the interval [-a, a].

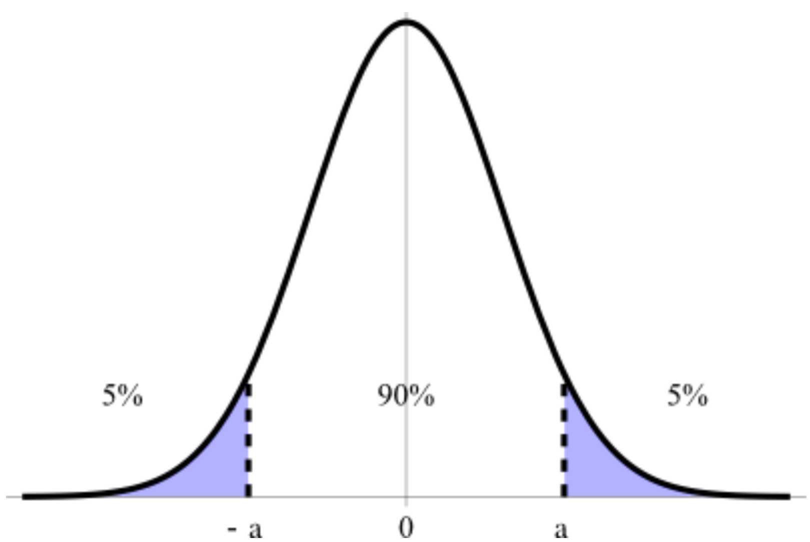

Figure 4-2. The interval [-a, a] is called a 90% confidence interval. Image Source: Wikimedia Commons.

Fisher attacked Neyman's CI theory by claiming it did not serve the needs of scientists and potentially would lead to mutually contradictory inferences from data. Fisher’s criticisms of CI theory have proved to be justified—but not because Neyman’s CI theory is logically or mathematically flawed, as Fisher claimed.

Let’s examine the trifecta of errors that arises from the common practice of misusing Neyman’s CI theory as a post-data theory—i.e., for making inferences about population parameters based on a specific data sample. The three types of errors using CIs are:

- making probabilistic claims about population parameters
- making probabilistic claims about a specific confidence interval
- making probabilistic claims about sampling distributions

In order to understand this trifecta of errors, we need to understand probability and statistical inference from the perspective of a modern statistician.

Frequentists, such as Fisher and Neyman, claim that probability is a natural property of an event and is measured empirically as a long-run relative frequency. Frequentists postulate that the underlying stochastic process that generates data has statistical properties that do not change in the long-run: the probability distribution is time invariant. Even though the parameters of this underlying process may be unknown or unknowable, frequentists believe that these parameters are constant and have “true” values. Population parameters may be estimated from random samples of data. It is the randomness of data that creates uncertainty in the estimates of the true, fixed population parameters. This frequentist philosophy of probability and statistical inference has had a profound impact on the theory and practice of financial economics in general and CIs in particular.

Pricing Stocks Regressively with Statsmodels

Modern portfolio theory assumes that rational, risk-averse investors demand a risk premium, a return in excess of a risk-free asset such as a treasury bill, for investing in risky assets such as equities. A stock’s single factor market model (MM) is basically a linear regression model of the realized excess returns of a stock (outcome or dependent variable) regressed against the realized excess returns of a single risk factor (predictor or independent variable) such as the overall market as formulated below:

$$(R - F) = \alpha + \beta * (M - F) + \epsilon$$

Where *R* is the realized return of a stock, *F* is the return on a risk-free asset such as a US treasury bill, *M* is the realized return of a market portfolio such as the S&P 500, α (alpha) is the expected stock-specific return, β (beta) is the level of systematic risk exposure to the market and ε (epsilon) is the unexpected stock-specific return. The beta of a stock gives the average percentage return response to a 1% change in return of the overall market portfolio. For example, if a stock has a beta of 1.4 and the S&P 500 falls by 1%, the stock is expected to fall by -1.4% *on average*. See Figure 4-3.

Figure 4-3. Market Model showing the excess returns of Apple Inc (AAPL) regressed against the excess returns of the S&P 500.

Note that the MM of an asset is different from its Capital Asset Pricing Model (CAPM). The CAPM is the pivotal economic equilibrium model of modern finance that predicts expected returns of an asset based on its β or systematic risk exposure to the overall market. Unlike the CAPM, an asset's MM is a statistical model that has both an idiosyncratic risk term α and an error term ε in its formulation. According to the CAPM, alpha of an asset's MM has an expected value of zero because market participants are assumed to hold efficient portfolios that diversify the idiosyncratic risks of any specific asset. Market participants are only rewarded for bearing systematic risk since it cannot be diversified away. In keeping with the general assumptions of an OLS regression model, both CAPM and MM assume that the expected value of the residuals ε to be normally distributed with a zero mean and a constant, finite variance.

A financial analyst, relying on modern portfolio theory and practice, assumes there is an underlying, time-invariant, stochastic process generating the price data of Apple Inc., which can be modeled as an OLS linear regression MM. This MM will have population parameters, alpha and beta, which have true, fixed values that can be estimated from random samples of Apple's closing price data.

Let's run our Python code to estimate alpha and beta based on a sample of 5 years of daily closing prices of Apple. We can use any holding period return as long as it is used consistently throughout the formula. Using a daily holding period is convenient because it makes price return calculations much easier using Pandas dataframes.

```
#Install Yahoo finance package
!pip install yfinance
#Import relevant Python packages
import statsmodels.api as sm
import pandas as pd
import yfinance as yf
import matplotlib.pyplot as plt
plt.style.use('seaborn')
from datetime import datetime
#Import financial data
start = datetime(2017, 8, 3)
end = datetime(2022, 8, 6)
#S&P 500 index is a proxy for the market
market = yf.Ticker('SPY').history(start=start, end=end)
#Ticker symbol for Apple, the most liquid stock in the world
stock = yf.Ticker('AAPL').history(start=start, end=end)
#10 year US treasury note is the proxy for risk free rate
```

```
riskfree_rate = yf.Ticker('^TNX').history(start=start, end=end)
#Create dataframe to hold daily returns of securities
daily_returns = pd.DataFrame()
daily_returns['market'] = market['Close'].pct_change(1)*100
daily_returns['stock'] = stock['Close'].pct_change(1)*100
#Compounded daily rate based on 360 days for the calendar year used
in the bond market
daily_returns['riskfree'] = (1 + riskfree_rate['Close']) ** (1/360)
- 1
#Plot and summarize the distribution of daily returns
plt.hist(daily_returns['market']), plt.title('Distribution of
Market (SPY) Daily Returns'), plt.xlabel('Daily Percentage
Returns'), plt.ylabel('Frequency'), plt.show()
#Analyze descriptive statistics
print("Descriptive Statistics of the Market's daily percentage
returns:\n{}".format(daily_returns['market'].describe()))
plt.hist(daily_returns['stock']), plt.title('Distribution of Apple
Inc. (AAPL) Daily Returns'), plt.xlabel('Daily Percentage
Returns'), plt.ylabel('Frequency'), plt.show()
#Analyze descriptive statistics
print("Descriptive Statistics of the Apple's daily percentage
returns:\n{}".format(daily_returns['stock'].describe()))
plt.hist(daily_returns['riskfree']), plt.title('Distribution of the
riskfree rate (TNX) Daily Returns'), plt.xlabel('Daily Percentage
Returns'), plt.ylabel('Frequency'), plt.show()
#Analyze descriptive statistics
print("Descriptive Statistics of the 10 year note daily percentage
returns:\n{}".format(daily_returns['riskfree'].describe()))
#Examine missing rows in the dataframe
market.index.difference(riskfree_rate.index)
#Fill rows with previous day's risk-free rate since daily rates are
generally stable
daily_returns = daily_returns.ffill()
#Drop NaNs in first row because of percentage calculations
daily_returns = daily_returns.dropna()
#Check dataframe for null values
daily_returns.isnull().sum()
#Check first five rows of dataframe
daily_returns.head()
#AAPL's Market Model based on daily excess returns
#Daily excess returns of AAPL
y = daily_returns['stock'] - daily_returns['riskfree']
#Daily excess returns of the market
x = daily_returns['market'] - daily_returns['riskfree']
#Plot the data
plt.scatter(x,y)
#Add the constant vector to obtain the intecept
x = sm.add_constant(x)
```

```
#Use Ordinary Least Squares algorithm to find the line of best fit
market_model = sm.OLS(y, x).fit()
#Plot the line of best fit
plt.plot(x, x*market_model.params[0]+market_model.params['const'])
plt.title('Market Model of AAPL'), plt.xlabel('SPY Daily Excess
Returns'), plt.ylabel('AAPL Daily Excess Returns'), plt.show();
#Display the values of alpha and beta of AAPL's market model
print("According to AAPL's Market Model, the security had a
realized Alpha of {0}% and Beta of
{1}".format(round(market_model.params['const'],2),
round(market_model.params[0],2)))
#Summarize and analyze the statistics of your linear regression
print("The Market Model of AAPL is summarized
below:\n{}".format(market_model.summary()));
```

After running our Python code, a financial analyst would estimate that alpha is 0.071% and beta is 1.2385, as shown in the Statsmodels summary output in Figure 4-4.

```
The Market Model of AAPL is summarized below:
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.624
Model:                            OLS   Adj. R-squared:                  0.624
Method:                 Least Squares   F-statistic:                     2087.
Date:                Sun, 07 Aug 2022   Prob (F-statistic):          2.02e-269
Time:                        06:28:33   Log-Likelihood:                -2059.8
No. Observations:                1260   AIC:                             4124.
Df Residuals:                    1258   BIC:                             4134.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0710      0.035      2.028      0.043       0.002       0.140
0              1.2385      0.027     45.684      0.000       1.185       1.292
==============================================================================
Omnibus:                      202.982   Durbin-Watson:                   1.848
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1785.931
Skew:                           0.459   Prob(JB):                         0.00
Kurtosis:                       8.760   Cond. No.                         1.30
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 4-4. Output of Statsmodels summarizing the linear regression results of Apple's MM from 8/4/2017 to 8/5/2022.

Clearly, these point estimates of alpha and beta will vary depending on the sample size, start and end dates used in our random samples, with each estimate reflecting Apple's idiosyncratic price fluctuations during that specific time period. Even though the population parameters alpha and beta are unknown, and possibly unknowable, the financial analyst considers them to be true constants of a stochastic process. It is the random sampling of Apple's price data that introduces uncertainty in the estimates of constant population parameters. It is the data, and every statistic derived from the data such as CIs, that are treated as random by frequentists. Financial analysts calculate CIs from random samples in order to express the uncertainty around point estimates of constant population parameters.

CIs provide a range of values with a probability value or significance level attached to that range. For instance, in Apple's MM above, a financial analyst could calculate the 95% confidence interval by calculating the standard error (SE) of alpha and beta. Since the residuals ε are assumed to be normally distributed with an unknown, constant variance, the t-statistic would to be used in computing CIs. However, because the sample size is greater than 30, the t-distribution converges to the standard normal distribution and the t-statistic values are the same as Z-scores of a standard normal distribution. So, the analyst would multiply each SE by +/- the Z-score for a 95% CI and then add the result to the point estimate of alpha and beta to obtain its CI. In the Statsmodels results above, the 95% CI for alpha and beta are computed as follows:

α+/- (SE * t-statistic/Z-Score for 95% CI) = 0.0710 % +/- (0.035 % * 1.96) = [0.002%, 0.140%]

β+/- (SE * t-statistic/Z-Score for 95% CI) = 1.2385 +/- (0.027 * 1.96) = [1.185, 1.292])

The Confidence Game

What most people think they are getting from a 95% CI is a 95% probability that the true population parameter is within the limits of the *specific* interval calculated from a specific data sample. For instance, based on the Statsmodels results above, you would think there is a 95% probability that the true value of beta of Apple is in the range [1.185, 1.292]. Strictly speaking, your interpretation of such a CI would be wrong.

According to Neyman's CI theory, what a 95% CI actually means is that if we were to draw 100 random samples from Apple's underlying stock return distribution, we would end up with 100 different confidence intervals, and we can be confident that 95 of them will contain the true population parameter within their limits. However, we won't know which specific 95 CIs of the 100 CIs include the true value of the population parameter and which five CIs do not. We are assured that only the long-run ratio of the CIs that include the population parameter to the ones that do not will approach 95% as we draw random samples *ad nauseam*.

Winston Churchill could just as well have been talking about CIs instead of Russia's world war strategy when he said, "It is a riddle, wrapped in a mystery, inside an enigma; but perhaps there is a key." Indeed, we do present a key in this chapter. Let's investigate the trifecta of fallacies that arise from misusing CI as a post-data theory in financial data analysis.

Errors in Making Probabilistic Claims About Population Parameters

Recall that a frequentist statistician considers a population parameter to be a constant with a "true" value. This value may be unknown or even unknowable. But that does not change the fact that its value is fixed. Therefore, a population parameter is either in a CI or it is not. For instance, if you believe the theory that capital markets are highly efficient, you would also believe that the true value of alpha is 0. Now 0 is definitely not in the interval [0.002%, 0.14%] calculated above. Therefore, the probability that alpha is in our CI is 0% and not 95% or any other value.

Because population parameters are believed to be constants by frequentists, there can be absolutely no ambiguity about them: the probability that the true value of a population parameter is within *any* CI is either 0% or 100%. So, it is erroneous to make probabilistic claims about any population parameter under a frequentist interpretation of probability.

Errors in Making Probabilistic Claims About a Specific Confidence Interval

A more sophisticated interpretation of the above CIs found in the literature and textbooks goes as follows: hypothetically speaking, if we were to repeat our linear regression many times, the interval [1.185, 1.292] would contain the true value of beta within its limits about 95% of the time.

Recall that probabilities in the frequentist world apply only to long-run frequencies of *repeatable* events. By definition, the probability of a unique event, such as a specific CI, is undefined and makes no sense to a frequentist. Therefore, a frequentist cannot assign a 95% probability to either of the specific intervals for alpha and beta that we have calculated above. In other words, we can't really infer much from a specific CI.

But that is the main objective of our exercise! This limitation of CIs is not very helpful for data scientists who want to make inferences about population parameters from their specific data samples: i.e., they want to make post-data inferences. But, as was mentioned earlier, Neyman intended his CI theory to be used for only pre-data inferences based on long term frequencies.

Errors in Making Probabilistic Claims About Sampling Distributions

How do financial analysts justify making these probabilistic claims about CIs in research and practice? How do they square the circle? What is the key to applying CIs? Statisticians can, in theory or in practice, repeatedly sample data from a population distribution. The point estimates of sample means computed from many different random samples create a pattern called the sampling distribution of the sample mean. Sampling distributions enable frequentists to invoke the Central Limit Theorem (CLT) in calculating the uncertainty around sample point estimates of population parameters. In particular, the CLT states that if many samples are drawn randomly from a population with a finite mean and variance, the sampling distribution of the sample mean approaches a normal distribution asymptotically. The shape of the underlying population distribution is immaterial and can only affect the speed of this inexorable convergence to normality.

The frequentist definition of probability as a long-run relative frequency of repeatable events resonates with the CLT's repeated drawing of random samples from a population distribution to generate its sampling distributions. So, statisticians square the circle by invoking the CLT and claiming that their sampling distributions almost surely converge to a normal distribution, regardless of the shape of the underlying population distribution. This also enables them to compute CIs using the Z-scores of the standard normal distribution as shown above. This is the key to the enigmatic use of CI as a post-data theory.

However, as financial executives and investors putting our capital at risk, we need to read the fine print of the CLT: specifically, we need to note its assumption that the underlying population distribution needs to have a finite mean and variance. While most distributions satisfy these two conditions, there are many that do not, especially in finance and economics. For these types of population distributions, the CLT cannot be invoked to save CIs. The key does not work on these doors - it is not a magic key. For instance, the Cauchy and Pareto distributions are fat-tailed distributions that do not have finite mean or variances. In fact, a Cauchy (or Lorentzian) distribution looks deceptively similar to a normal distribution but has very fat tails because of its infinite variance. See Figure 4-5.

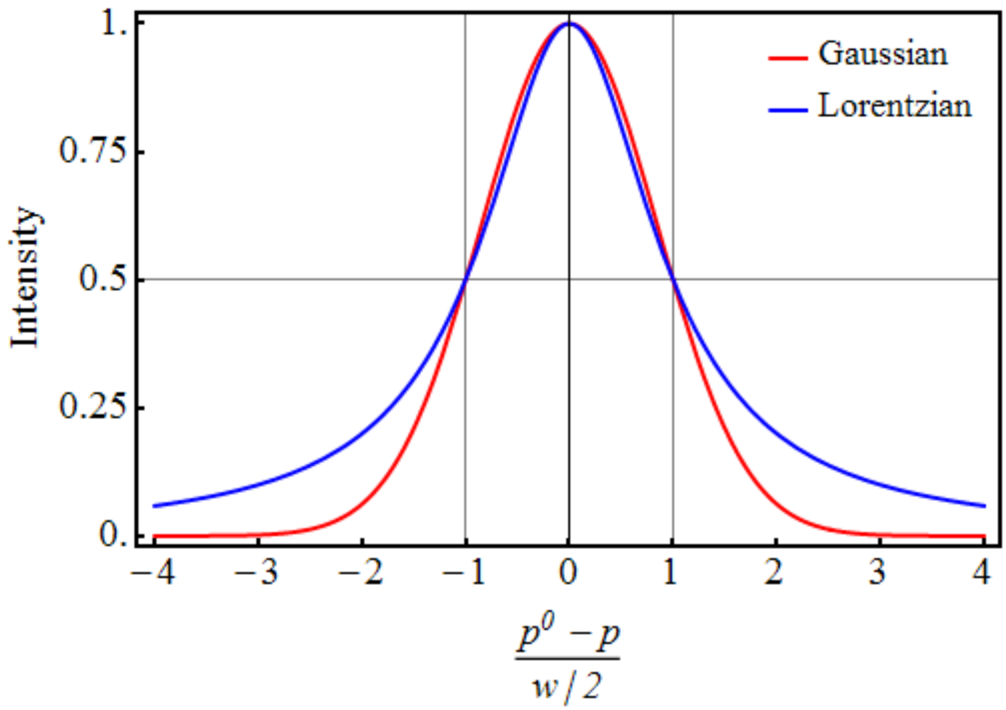

Figure 4-5. Compare Cauchy/Lorentzian distribution with the Normal distribution. Image Source: Wikimedia Commons.

The diagnostic tests computed by Statsmodels in Figure 4-4 show us that the equity market has wrecked the key assumptions of our MM. Specifically, the Bera-Jarque and Omnibus normality tests show the probability that the residuals $\boldsymbol{\varepsilon}$ are normally distributed is almost surely zero. This distribution is positively skewed and has very fat tails—a kurtosis that is about three times that of a standard normal distribution.

How about making the sample size even larger? Won't the distribution of the residuals get more normal with a much larger sample size as claimed by financial theory? Let's run our MM using 25 years of Apple's daily closing prices — a quarter of a century worth of data. The results are shown in Figure 4-6:

```
The Market Model of AAPL is summarized below:
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.270
Model:                            OLS   Adj. R-squared:                  0.270
Method:                 Least Squares   F-statistic:                     2331.
Date:                Sun, 07 Aug 2022   Prob (F-statistic):               0.00
Time:                        07:03:34   Log-Likelihood:                -14187.
No. Observations:                6293   AIC:                         2.838e+04
Df Residuals:                    6291   BIC:                         2.839e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.1063      0.029      3.656      0.000       0.049       0.163
0              1.1208      0.023     48.281      0.000       1.075       1.166
==============================================================================
Omnibus:                     2566.940   Durbin-Watson:                   2.020
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           662984.825
Skew:                          -0.736   Prob(JB):                         0.00
Kurtosis:                      53.262   Cond. No.                         1.25
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure 4-6. Output of Statsmodels summarizing the linear regression results of AAPL's MM from 8/5/1997 to 8/5/2022.

All the diagnostic test results above make it clear that the equity market has savaged the "Nobel-prize" winning CAPM (and related MM) theory. Even with a sample size that includes a quarter of a century of daily closing prices, the distribution of our model's residuals is grossly more non-normal than before. It is now very negatively skewed with an absurdly high kurtosis—almost 18 times that of a standard normal distribution. Most notably, the CI of our 20-year beta is [1.075, 1.166], which is outside the range of the CI of our 5-year beta [1.185,1.292]. In fact, the beta of AAPL seems to be regressing toward 1, the beta value of the S&P 500.

Invoking some version of the CLT and claiming asymptotic normality for the *sampling distributions* of the residuals or the coefficients of our regression model seem futile, if not invalid. There is a compelling body of economic research claiming that the underlying distributions of all financial asset price returns do not have finite variances. Financial analysts should not be so certain that they can summon the powers of the CLT and assert asymptotic normality in their CI computations. Furthermore, they need to be sure that convergence to asymptotic normality is reasonably fast because, as the eminent economist Maynard Keynes found out the hard way with his personal equity investments, "Markets can stay irrational longer than you can stay solvent." For an equity trade, a quarter of a century is an eternity.

Conclusions

Because of the errors detailed in this chapter with NHST, p-values and CIs, I have no confidence in them (or the CAPM) and do not use it in my financial data analyses. I would not waste a penny trading or investing based on the estimated CIs of alpha and beta of any frequentist MM computed by Statsmodels or any other software application. I would also throw away any social or economic study that uses NHST, p-values or confidence intervals in the trash where junk belongs.

Statistical hypothesis testing developed by Neyman and Pearson only makes sense as a pre-data decision theory for mechanical processes like industrial quality control. The mish-mash of the competing statistical theories of Fisher and Neyman was created by non-statisticians (or incompetent statisticians) to please two bitter rivals and ended up creating a nonsensical, confusing blend of the two. Of course, this has not stopped data scientists from using NHST, p-values and CIs blindly or academics from teaching it as a mathematically rigorous post-data theory of statistical inference.

CIs are not designed for making post-data inferences about population parameters from a single experiment. The use of CIs as a post-data theory is epistemologically flawed. It flagrantly violates the frail philosophical foundation of frequentist probability on which it rests. Yet, orthodox statisticians have concocted a mind-bending, spurious rationale for doing exactly that. You might get away with misusing Neyman's CI theory if the CLT applies to your data analysis—i.e., the underlying population distribution has a finite mean and variance resulting in asymptotic normality of its sampling distributions.

However, it is common knowledge among academics and practitioners that price returns of all financial assets are not normally distributed. It is likely that these fat tails are a consequence of infinite variances of their underlying population distributions. So, the theoretical powers of the CLT cannot be utilized by analysts to rescue CIs from the non-normal, fat-tailed, ugly realities of financial markets. Even if asymptotic normality is theoretically possible in some situations, the desired convergence may not be quick enough for it to be of any practical value for trading and investing. Financial analysts should heed another of Keynes's warnings when hoping for asymptotic normality of their sampling distributions: "In the long run, we are all dead" (and almost surely broke).

Regardless, financial data analysts using CIs as a post-data theory are making invalid inferences and grossly misestimating the uncertainties in their point estimates. Unorthodox statistical thinking, ground-breaking numerical algorithms, and modern computing technology make the use of 'worse than useless' NHST, p-values and CI theory in financial data analysis unnecessary. The second half of this book is dedicated to exploring and applying epistemic inference and probabilistic machine learning to finance and investing.

References

Ioannidis J. P. (2005). Why most published research findings are false. PLoS medicine, 2(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Harvey, Campbell R. and Liu, Yan and Zhu, Caroline, …and the Cross-Section of Expected Returns (February 3, 2015). Available at SSRN: https://ssrn.com/abstract=2249314 or http://dx.doi.org/10.2139/ssrn.2249314

Colquhoun David (2014). An investigation of the false discovery rate and the misinterpretation of p-values R. Soc. open sci.1140216140216 http://doi.org/10.1098/rsos.140216

Lambdin, C. (2012). Significance tests as sorcery: Science is empirical—significance tests are not. Theory & Psychology, 22(1), 67–90. https://doi.org/10.1177/0959354311429854

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E. J. (2016). The fallacy of placing confidence in confidence intervals. Psychonomic bulletin & review, 23(1), 103–123. https://doi.org/10.3758/s13423-015-0947-8

Thompson, W. C., & Schumann, E. L. (1987). Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy. Law and Human Behavior, 11(3), 167–187. http://www.jstor.org/stable/1393631

John Aldrich (2008), R. A. Fisher on Bayes and Bayes' Theorem. International Society for Bayesian Analysis, 3, Number 1, pp. 161 - 170

LOUÇÃ, F. (2009). Emancipation Through Interaction — How Eugenics and Statistics Converged and Diverged. Journal of the History of Biology, 42(4), 649–684. http://www.jstor.org/stable/25650625

Lenhard, J. (2006). Models and Statistical Inference: The Controversy between Fisher and Neyman-Pearson. The British Journal for the Philosophy of Science, 57(1), 69–91. http://www.jstor.org/stable/3541653

Szucs D and Ioannidis JPA (2017) When Null Hypothesis Significance Testing Is Unsuitable for Research: A Reassessment. Front. Hum. Neurosci. 11:390. doi: 10.3389/fnhum.2017.00390

Gigerenzer, G. (2018). Statistical Rituals: The Replication Delusion and How We Got There. Advances in Methods and Practices in Psychological Science, 198–218. https://doi.org/10.1177/2515245918771329

Probability Theory: The Logic of Science, by E.T. Jaynes, Cambridge University Press, 2003

Quantitative Methods, CFA Program Curriculum, Volume 1, CFA Institute, 2007

Alternative Investments, CAIA, Level I, John Wiley & Sons, 2015

About the Author

Deepak Kanungo is the founder and CEO of Hedged Capital LLC, an AI-powered, proprietary trading and analytics firm built around probabilistic machine learning technologies. In 2005, long before machine learning was an industry buzzword, Deepak invented a probabilistic machine learning method and software system for managing the risks and returns of project portfolios. It is a unique framework that has been cited by IBM and Accenture, among others.

Previously, Deepak was a financial advisor at Morgan Stanley, a Silicon Valley fintech entrepreneur, and a director in the Global Planning Department at Mastercard International. He was educated at Princeton University (astrophysics) and the London School of Economics (finance and information systems).