Compiled Content
Module 5

MScFE 640

Portfolio Theory and Asset Pricing

Revised: 09/09/2019

Table of Contents

Module 5: Bayesian Portfolio Theory 3

Unit 1: Challenges Associated with Parameter Estimation in Mean-variance Optimization – Part 1 of 2 4

Unit 2: Challenges Associated with Parameter Estimation in Mean-variance Optimization – Part 2 of 2 8

Unit 3: Bayes' Theorem and Its Application to Mean-variance Optimization – Part 1 of 2 13

Unit 4: Bayes' Theorem and Its Application to Mean-variance Optimization – Part 2 of 2 18

Bibliography 22

Module 5: Bayesian Portfolio Theory

Module 5 begins by discussing the challenges associated with estimating the returns, volatilities, and correlations, and follows with the introduction of Bayesian methods for portfolio optimization that address these challenges. The module continues by discussing the Bayes' Theorem and how to use it to perform statistical inference. The module ends by showing an application of Bayes' Theorem to multi-modal distributions.

Unit 1: Challenges Associated with Parameter Estimation in Mean-variance Optimization – Part 1 of 2

Introduction

We will now turn our attention to an alternative method of estimating parameters for portfolio optimization, known as Bayesian statistics and named after an 18th century statistician. In this first section, we will address the challenges associated with estimating the returns, volatilities, and correlations required to run the mean-variance framework. In the next section, we will introduce Bayes' Theorem. The third section will cover how to perform inference in a Bayesian setting. The last section of the module covers how to perform inference in a specific portfolio example.

Parameters in mean-variance optimization

Recall that for portfolio optimization, our goal is to calculate an efficient frontier. Suppose we are given a risk tolerance, q. We then have a minimization problem that can be stated as follows:

$$w^T \Sigma w - q\ R^T w$$

$$\underbrace{(1 * n)(n * n)(n * 1)} - \underbrace{(k)(1 * n)(n * 1)}$$

Note that the dimensions of the matrices must be conformable – i.e. they must have values that match on the inside.

Let's define each of the terms, assuming a portfolio with n securities:

- w is a vector of portfolio weights. Its dimensions are $(n * 1)$.
- Σ is the covariance matrix of the portfolio returns. For n returns, this is a square $(n * n)$ matrix.
- w^T is simply the transpose of the weights. Its dimensions are $(1 * n)$.

When these three matrices are multiplied together they produce a scalar. Continuing to define the terms:

- q is a risk tolerance factor. The higher the value of q, the greater our tolerance for risk. It is a constant.

- R is a column matrix of expected returns. In the equation, note that we use its transpose.
- R^T is the transpose of expected returns. Its dimensions are $(1 * n)$.
- When q is multiplied by R^T and w, we get a scalar.

Therefore, the above equation simplifies to the difference of two numbers. The efficient frontier is the solution of the best set of weights that minimizes that difference. At this point, we should consider both what the key inputs to the equation are and how we determined them.

The key inputs to the equation are the covariance matrix and the expected returns. The most intuitive way to determine them would be to collect some historical data, calculate the sample mean and sample covariance of those observations, and use those estimates.

Let's formally notate this method.

- We estimate the population mean μ with our sample mean $\hat{\mu}$.
- We estimate the population variance matrix Σ with our sample covariance matrix $\hat{\Sigma}$.

How well does this method do? Surprisingly, not very well.

The problems associated with the estimation of means and variance have been known for more than 50 years. In 1971, Frankfurter, Phillips, and Seagle published a paper that stated that portfolios with equal weights generally performed better than portfolios whose weights were determined by sample estimates. Think about that for a moment. There is so much noise in using sample estimates that we could tend to do better if we just assumed equal weights within the portfolio. These findings are not as well-known as they should be. It is a disruptive finding, because it says while mean-variance optimization is great in theory, it does not work well in practice because of the difficulty in getting accurate estimates. To be fair, the problem is not with the optimization process; it is with getting quality inputs to feed into the procedure.

In 1980, Jobson and Korkie wrote in a paper that the estimators do not lend themselves to making good inferences. They showed there could be improvements made by using different methods of estimation.

Akerlof and Yellen (1985) were one of the earliest researchers to publish that that small deviations from optimality resulted in significant differences in the performance of the portfolios. About 30 years, later, Janet Yellen later became the Chair of the Federal Reserve System in the United States.

In 1991, Best and Grauer also showed the extreme sensitivity of the mean-variance portfolios to small changes in the means of individual assets. They produced more specific results. Under budget constraints, their results showed that the mean-variance efficient portfolio's weights, mean, and variance are extremely sensitive to changes in asset means. If non-negativity constants are imposed, then the mean-variance efficient portfolio's weights are sensitive to changes in asset means, but the portfolio's returns are not sensitive. Surprisingly, a very small increase in the mean of just a single asset drives half the securities from the portfolio. However, the portfolio's expected return and standard deviation remain almost unchanged.

Around that time, another approach was throwing more computer power at problems. One of the statistical methods that had been around for decades, but was not feasible computationally, was bootstrapping. **Bootstrapping** is defined as sampling from a set of data with replacement. The idea behind it is that drawing different samples, which will invariably include replicates, will lead to a better estimate and better confidence intervals around the estimate than our classical method of sample means. In 1989, Michaud wrote that one could improve the efficient frontier by resampling it, bootstrapping, and running thousands of simulations, rather than using the **plug-in frontier**. This frontier is derived by simply using the sample estimates. He wrote a book, started a company, and patented this idea.

Vijay Chopra (1993) succinctly summarized the main pitfalls of portfolio optimization in a paper called "Improving Optimization". Portfolios that are close to the optimal portfolio in terms of risk and expected return can be very different in composition. He also noted that very small changes in the input parameters produced very different compositions, by large changes in the weights. In his paper, Chopra mentioned the use of Bayesian techniques to implement. He wrote that "it is more complex and is therefore more difficult to implement" (Chopra, 1993).

These research findings ultimately show that we have a statistical inference problem rather than a problem with the optimization. **Statistical inference** is the art and science of drawing inferences about a population based on statistics calculated from a data sample drawn from that population. It is a science because there is a deep, theoretical field that spans frequentist, Bayesian, and even computational approaches. It is an art because there are many varieties and subtleties in applying inferential tools such that there is no cookbook approach. Simply using the sample mean and sample covariance as the estimates for the population does not work as well as we would like.

Summary

Would it be possible to use our prior experience, insight, and expertise about these assets combined with information from the samples? This is exactly what the Bayesian framework allows us to do.

In this section, we discussed the main challenges with mean-variance optimization: using sample means and covariances leads to very different portfolios.

Unit 2: Challenges Associated with Parameter Estimation in Mean-variance Optimization – Part 2 of 2

Introduction

In this module, we will illustrate the importance of Bayesian statistics in the mean-variance portfolio optimization problem. In this first section, we'll begin by distinguishing a conditional probability from an unconditional probability. Thereafter, we'll explain Bayes' Theorem through an extended example called the Monty Hall problem. The objective is to understand the difference between a conditional probability and an unconditional probability.

Conditional probability versus unconditional probability

Consider a deck of cards. There are 52 cards in total, consisting of 4 aces, 4 kings, 4 queens, 4 jacks, and 36 other cards that are called numbered cards because they contain the numbers two through ten. We might pose ourselves a simple question: What is the probability of selecting a king? As there are four kings in the deck of 52 cards, the probability of selecting a king is 4/52 or, reducing, 1/13. This is known as an **unconditional probability.**

Given that a king was already selected, what is the probability of selecting another king? This is an example of a **conditional probability**, because it is conditioned on the first event occurring.

But what exactly does the first event do? It decreases the number of kings to choose from and the size of the deck by one. With only three kings in the deck of 51 cards, the probability of selecting a 2nd king is (4-1)/(52-1) = 1/17.

Let's now relate conditional probabilities and unconditional probabilities using Bayes' Theorem.

Bayes' Theorem

More than 250 years ago, Reverend Thomas Bayes wrote An Essay towards solving a Problem in the Doctrine of Chances. In that essay, he wrote an equation that would allow someone to update their beliefs in light of receiving new evidence. This equation came to be known as Bayes' Theorem.

Bayes' Theorem relates conditional probabilities and unconditional probabilities in an equation, namely:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}.$$

Where:

- $P(A)$ is the probability that event A occurs. It is an unconditional probability.
- $P(B)$ is the probability that event B occurs. It is an unconditional probability.
- $P(A|B)$ is the probability that event A occurs given that event B has occurred. It is a conditional probability.
- $P(B|A)$ is the probability that event B occurs given that event A has occurred. It is a conditional probability.

Another way to think of these terms is as follows:

- $P(A)$ is the prior probability of, or an initial degree of belief in, event A occurring.
- $P(A|B)$ is the posterior probability of, or the degree of belief in, event A having accounted for event B occurring.

What we are doing is moving from prior probabilities to posterior probabilities by observing events occurring. Indeed, the Bayesian approach is a way of learning from data. It is typically easier to find unconditional probabilities than it is to find conditional probabilities. That is because conditional probabilities require you to incorporate more information into the system, which changes the sample space and, with it, the probability of an event occurring. Let's illustrate Bayes' Theorem with a famous example called the Monty Hall problem

The Monty Hall problem

Suppose you are a contestant in a game show. Monty Hall is the name of the game show host. (There was indeed a real Monty Hall and a game show called "Let's Make A Deal".) The game show lets you win a prize – typically a car – if you can correctly

select it from among three curtains. Behind two of the curtains are goats. Behind the other curtain is the car. Your objective is to select the curtain with the car and not the goats.

You begin by selecting one of the curtains – let's call it Curtain 1. Now, Monty knows which curtains hide the goats and which curtain hides the car. Once you have selected a curtain, Monty then picks one of the other curtains that you did not select that has a goat; let's call this curtain Curtain 2. Note that this is always possible since there are two goats. Monty then shows you the goat behind Curtain 2. He then asks you: "Do you want to stick with Curtain 1, or would you like to switch to Curtain 3?"

This mathematical puzzle involved huge amounts of debate and attracted a lot of attention because many mathematicians got this wrong (see [https://ima.org.uk/4552/dont-switch-mathematicians-answer-monty-hall-problem-wrong/ for the background](https://ima.org.uk/4552/dont-switch-mathematicians-answer-monty-hall-problem-wrong/)). We should pause and ask ourselves, however: What makes this problem so difficult?

First, let's think about the problem intuitively. Without having any information, when you select Curtain 1, you have a 1/3 chance of picking the car. If the game ended there, then you'd have a 1/3 chance of winning. Conversely, you would have a 2/3 chance of picking a goat and losing.

How does the game change when Monty shows us a curtain with a goat? We have new information. Bayes' Theorem is a way to process new information; it is a way for us to update our probabilities.

Let's apply Bayes' Theorem to the Monty Hall Problem.

Let's say that you pick Curtain 1.

- Let $P(A)$ be the probability that the car is behind Curtain 2.
- Let $P(B)$ be the probability that Monty opens Curtain 3.
- $P(A|B)$ is the probability that the car is behind Curtain 2, given that Monty opens Curtain 3.
- $P(B|A)$ is the probability that Monty opens Curtain 3, given that the car is
- behind Curtain 2.

Now let's determine which of these we know.

$P(A) = P(Curtain\ 2\ has\ the\ car) = \frac{1}{3}$ because there is a 1 in 3 chance of picking the right curtain without knowing any other information. In other words, the unconditional probability is 1/3.

$P(B) = P(Monty\ shows\ Curtain\ 3) = \frac{1}{2}$ because Monty only has two doors to choose from. Because Curtain B is just as likely as Curtain C to have the goat, each curtain has a 50% chance of being selected by Monty.

$P(B|A) = P(Monty\ shows\ Curtain\ 3|Curtain\ 2\ has\ the\ car) = 1$. Recall that Monty has at most two choices: Curtains 2 and 3. He cannot select Curtain 1 because that is what you selected. In this case, however, he also cannot select Curtain 2 because that curtain has the car. His only choice, then, is to select Curtain 3. Therefore, the probability of his showing Curtain 3 when Curtain 2 has the car is 100%.

Now we can solve for the remaining conditional probability using Bayes' Theorem:

$$P(B|A) = P(Curtain\ 2\ has\ the\ car|Monty\ shows\ Curtain\ 3) = \frac{P(A)P(B|A)}{P(B)} = \frac{\frac{1}{3} * 1}{\frac{1}{2}} = \frac{2}{3}$$

Astonishingly, there is twice the probability of winning the car by switching from Curtain 1 to Curtain 2. In other words, it pays to switch.

What exactly did Bayes' Theorem do, though? Prior to any curtains being revealed, we have a 1/3 chance of winning the car. When new information is revealed, we know that the curtain to which we could switch has a 2/3 chance of having the car. What Bayes' theorem did was provide a way for us to update a probability we were interested in given new information.

People do not always find this intuitive, so let's extend the problem to make it easier to understand. Considering the same game again, let's increase the number of curtains from three to 100. As before, you select Curtain 1. Now Monty will show you 98 other curtains that do not have the car. Therefore, only one curtain remains that may contain the car. The question is: Do you switch?

Even if you got the first problem wrong and decided not to switch, here it is more intuitive to switch. Initially, you think the odds of selecting the car are small because you only have a 1% chance of picking it. That leaves a 99% chance that the car is somewhere in the other group. Therefore, when all but one of those curtains remain, it seems more reasonable to switch in that case. In this example, we have a 1% prior probability of winning if we do not switch, and a 99% probability of winning if we do switch.

Nevertheless, switching increases the chance of winning because the curtains that remained are simply more likely to have the car behind them, based on the new data –i.e. Monty revealing another curtain with the goat – that updates our uniform prior probability into a posterior probability that favors this new curtain.

Summary

In this section, we outlined Bayes' Theorem and applied it to an example of the classic Monty Hall problem. In the next section, we will discuss how to apply Bayes' Theorem to perform statistical inference.

Unit 3: Bayes' Theorem and Its Application to Mean-variance Optimization – Part 1 of 2

Introduction

In the previous section, we defined Bayes' Theorem. The Bayesian approach is a different way of thinking and doing statistics compared to the more traditional and commonly used frequentist approach. Surprisingly, the Bayesian approach precedes the frequentist approach by about 100 years. Without computer power, however, it is more difficult to employ because of the greater number of calculations involved. With the power of computers and the strength of statistical software, it is now easier than ever to harness the power of the Bayesian approach. In this section, we will illustrate that with a classic problem of Bayesian inference. This example will lay the foundation for the next section, where we apply this to the specific area of portfolio optimization.

An outline of Bayesian inference

Statistical inference means that you are inferring the value of particular parameters from a sample of data.

Bayesian inference can be divided into five steps:

1. *Identify* the observed data.
2. *Build* a probabilistic model to represent the data. This will specify a likelihood function.
3. *Hypothesize* different values for the parameters of your probabilistic model. This will specify a prior distribution.
4. *Collect* data. This comes from experiments, observations, or simulations. Given the outcomes, update the likelihood function.
5. *Update* the prior probability by using the collected data in Bayes' Theorem. This will output the posterior probability.

Note that Steps 4 and 5 can be repeated indefinitely.

To summarize, Bayesian statistics is an information processing technique that has you update your prior belief (the prior distribution) in light of new observations (data) through a likelihood function (likelihood) to produce a new set of beliefs (the posterior distribution). Let's illustrate this methodology using the classic example of testing if a coin is fair.

Suppose we find a coin on the street. We would like to know if it's a fair coin.

The first step is to identify the data. There are two possible outcomes: heads or tails. We can denote the random variable, X, as the outcome and assign it to 1 if it is heads and to 0 if it is tails.

The second step is to build a probabilistic model to represent the data. Since the data is bimodal (heads or tails), our data is essentially a binomial distribution. We can represent this as:

$$P(X = 1|\theta) = \theta$$

and

$$P(X = 0|\theta) = 1 - \theta$$

where θ is the probability of flipping a head. Let's translate these equations into words, assuming that the coin is fair. The probability of flipping heads is 50%, and the probability of flipping tails is 50%. Suppose the coin were biased towards heads 75%. Then the equation says the probability of heads (which we define as a "success" or 1) is 75%, and the probability of tails (which we define as "failing" or 0) is 25%.

Combining these into a single term, we get:

$$P = \theta^x * (1 - \theta)^{(1-x)}$$

This is known as a Bernoulli distribution. This function can serve as our likelihood function for Bayes' Theorem. For example, suppose we have seven heads and three tails. We can figure out our likelihood for a given value of θ as θ raised to the number of heads, times (1 – θ) raised to the number of tails.

Now that we have established a likelihood function, we need to understand what the parameters in that likelihood function are. For the binomial distribution, the chief parameter is θ, which is the probability of success (here that means the probability of flipping heads). What can this probability of heads be? We can assume one hypothesis, that the probability of heads if 50%, but the advantage of the Bayesian approach is that it lets you treat this parameter as a random variable and entertain many possible values of θ. Let's assume there could be nine possible values for the probability of getting heads: 10%, 20%, 30%, and so on up to 90%. (In theory, the probability could be any value between 0 and 1, but using round numbers reduces the complexity and simplifies the calculations.) Any probability of heads that is not 50% results in a biased coin. If the probability of heads is 10%, then the coin is biased towards tails 90%. If the probability of heads is 80%, then the coin is heavily biased towards heads. However, we may believe that the coin is like most coins we have encountered, so we assume most likely that we have a probability of 50% of heads. Yet we will not discard other hypotheses, we will just weigh them less. The key idea is that we think of theta as a random variable. That means we wind up with a distribution of these values – our prior distribution.

Let's assume that our prior distribution is triangular. This is a symmetric distribution, with the center in this case being the chance of a fair coin. In order for this distribution to be a probability density function, the total probabilities must sum to 1. To quantify these probabilities, we will state that there is a 20% chance of the coin being fair. Then, as we move 10% to either side, we will drop 4% in probability. That means that the probability of θ = 0.60 is 16%. Likewise, the probability of θ = 0.40 is 16%. Continuing this trend, there is a 12% chance of being 70% biased towards heads, an 8% chance of being 80% biased towards heads, and a 4% chance of being 90% biased towards heads. Likewise, we have the same linear relationship on the left side of our distribution, as the probabilities drop from 20% at the center down to 4% for our most biased value.

There is no such thing as an incorrect prior probability distribution. The prior represents our knowledge of the system. If you have a very different prior distribution from me but we observe the same data, then our posterior distributions will look more similar. Indeed, the more data that you collect, the less important your prior distribution becomes.

Knowing how important it is to collect data, that's what we'll do in the next step. In the context of this example, this is done by assuming we get a certain number of heads and tails. In practice, it could be done by collecting historical data, for example, or by running Monte Carlo simulations. Indeed, it can be updated in real time if you use that data to calculate the likelihood, update a posterior distribution, and then set that as your new 'prior' distribution.

The final step is to apply Bayes' Theorem: multiply the prior by the likelihood, and divide by the unconditional probability, and that gives you the posterior probability. Sometimes the unconditional probability is called the **marginal probability**.

Let's look at some R code that illustrates this.

```
# Doing Bayesian Inference: Coin Flipping Example
# 1. Define outcomes ----
heads = 1
tails = 0
# 2. Define the probabilistic model for data (likelihood) ----
#Let's use the binomial distribution
# For example, what are the probabilities of 0, 1, or 2 heads
#  when flipping a fair coin twice?
dbinom(c(0, 1, 2), size=2, prob=0.5)
# What if the coin were biased 60% towards heads?
   dbinom(c(0, 1, 2), size=2, prob=0.6)
# How about if we flip the fair coin 10 times?
   dbinom(0:10, size=10, prob=0.5)
# 3. Define the prior distribution ----
# Define the Space of all theta values
   (theta <- seq(0, 1, length=11))
# Turn these values into a triangular function
   (probs = pmin(theta, 1 - theta))
# Normalize so that these "probs" sum to 1
   (prior = probs/sum(probs))
#4. Collect data & update the likelihood
   nHeads = 7
   nTails = 3
   (likelihood = theta^nHeads * (1-theta)^nTails)
#5. Apply Bayes' Theorem to update the posterior
   (posterior = likelihood * prior / sum(likelihood))
# Notice the likelihood changes the prior to the posterior
   (par(mfrow=c(3,1))
# Plot the prior, likelihood, and posterior
   plot(theta, prior, main="Prior Distribution", type="h",
        xlab=expression(theta),
        ylab=expression(P(theta)))
   points(theta, prior)
   plot(theta, likelihood, main="Likelihood Function", type="h",
         xlab=expression(theta),
         ylab=expression(paste("P(D|", theta, ")")))
   points(theta, likelihood, cex=2)
   plot(theta, posterior, main="Posterior", type="h",
        xlab=expression(theta),
        ylab=expression(paste("P(", theta, "|D)")))
   points(theta, posterior, cex=2)
```

Visually, this looks like the following:

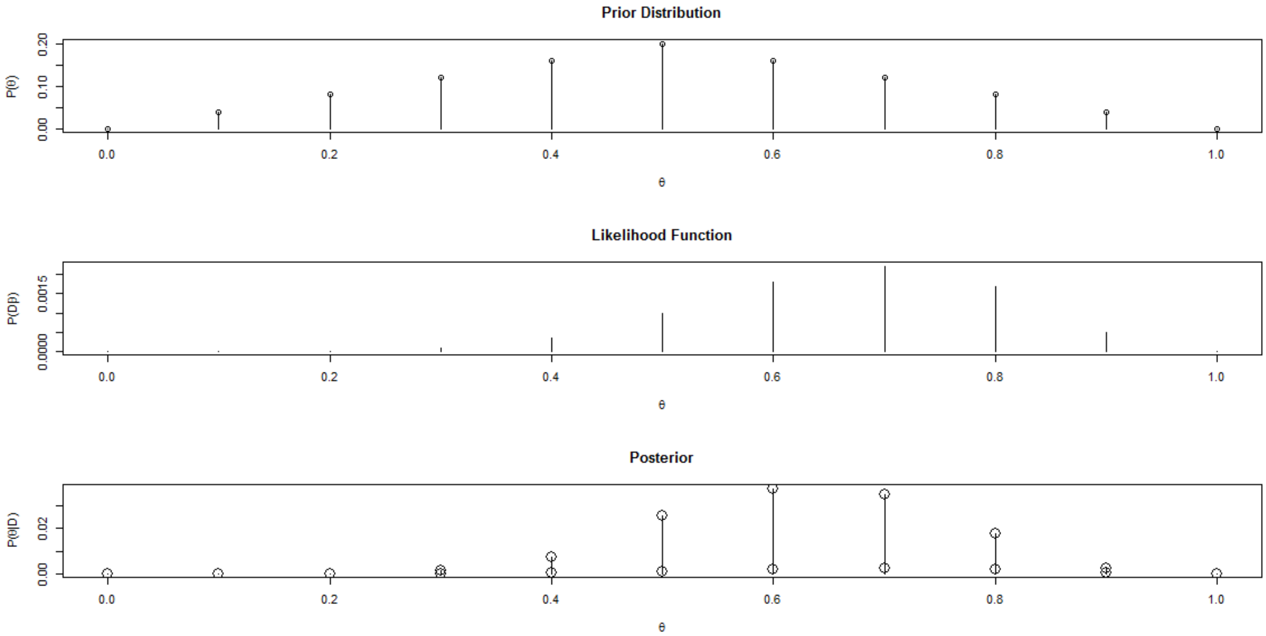

Figure 1: Application of Bayes' Theorem

In 50 lines of code, we were able to show how a basic Bayesian inference works. We started with a triangular distribution, thinking the coin is most likely biased. Then we observed 70% of the outcomes result in heads. Therefore, we shift our distribution (the posterior distribution) to have more probability for being biased (such as 60% or 70%). Note that we only recorded ten results; we would need significantly more data to make these conclusions really meaningful.

Summary

In this section, we examined how to do a Bayesian inference. In the next section, we will apply this to mean-variance portfolio optimization to address some of the challenges discussed earlier in this module.

Unit 4: Bayes' Theorem and Its Application to Mean-variance Optimization – Part 2 of 2

Introduction

In the previous section, we examined how to do Bayesian inference: we derive a likelihood model, start with a prior distribution, collect data, update the likelihood function, and then apply Bayes' Theorem to transform the prior probability to produce a posterior distribution. If we observe additional data, then our posterior distribution from Round 1 becomes the prior distribution for Round 2, and this process continues. Indeed, every time new data is introduced, a posterior distribution from a previous iteration becomes a prior distribution for the next one. Think of the prior and posterior distributions relative to a set of data that has yet to be processed.

There are several strategies for applying a Bayesian framework to mean variance portfolio optimization. These include Bayesian shrinkage estimation and Black-Litterman. Due to course constraints, we do not have enough time to provide the foundation, explanation, and illustration of these topics. The approach we will use is just one example that fits within the context and scope of the course.

In this section, we will first discuss the advantage of what we did in the previous sections when we used data to update our prior probability. Thereafter, we will apply a Bayesian inference in the mean-variance optimization using a real-world example by an asset manager in the international investment industry.

Coining it

Let's think back to our coin example. Suppose flipping a coin produces heads 45 times in 100 trials. Would we then say that the probability of getting heads is 45%? Maybe, but our advantage in this example is unlimited repetition. If we can repeatedly flip the coin, the estimator will converge towards its true parameter. Thus, if the coin in indeed fair, we will get close to 50% with enough tosses.

The researchers at Finaltis give the following scenario. Suppose you want to select the coin that has the greatest probability of producing a head in a future coin toss, and you have three coins to choose from.

- Two of the coins are fair

- One coin is biased, which truly has a probability of getting heads 45% of the time.

During testing, the biased coin produced a score of 52%. This means that 52% of the tosses resulted in heads. The two fair coins produced 51% and 47% respectively. Selecting a coin from one experiment, we would pick the biased coin because of the greatest frequency of heads. However, our goal should be to run more experiments, because with more samples the biased coin will not produce as many heads as the fair coins. This will be more and more likely as the number of samples increases.

Now consider a history of stock returns. Suppose that when we compute the standard deviation of this sample, we get 20%. Therefore, we state that the historical volatility is 20%. We then put this number into the mean-variance portfolio optimization equation, treating it as if it were known. Nevertheless, it has uncertainty.

The problem with the uncertainty is that it can appear in both the mean and the variance. It results in extremes: weights become too large or too small, favoring assets that have even a slight advantage in the estimate without taking into account the uncertainty in those estimates. Using a Bayesian approach can improve this process. Let's see how Finaltis does it.

A multi-modal distribution

They create a distribution of volatilities by examining 20 years' worth of stock returns. They notice that when they study the statistical distribution of these stock volatilities, five distinct peaks were evident. This 5-modal distribution becomes the prior distribution in their model. They can then recreate an empirical density function.

Likewise, you can think of the coin flipping example as a multi-modal distribution. Recall that our prior distribution established 9 possible, distinct probabilities of getting heads. Because we thought most coins were fair, we created a triangular distribution to weight the probability of having a fair coin most heavily, and to weight coins that are less biased with a greater weight than coins that are more biased. Based on observing 70% of heads, our first posterior distribution decreased the weights of the probabilities of 10% heads, 20% heads, and so, and it increased the weights of the probabilities of 60% heads, 70% heads, and so on.

Returning to the Finaltis case study (Croisille et al., 2016), the authors produced a table that shows the average volatility of a bucket and the relative frequency of that bucket. Here is a table of their findings back from the fourth quarter of 2015.

Probability	Average volatility
30.0%	23.3% volatility
45.2%	29.7 % volatility
13.1%	36.9% volatility
7.9%	43.3% volatility
3.8%	67.6% volatility
100.0%	

These values are specific to a given quarter and year. Therefore, the values of both the average volatilities and the proportion of the population in each bucket get updated quarterly. However, what did not change was the overall shape of the distribution containing five peaks. They now have a classic Bayesian inference problem, just like our coin problem. Instead of having two different outcomes, like we do in the coin toss, they have five potential outcomes: Buckets 1, 2, 3, 4, and 5.

Suppose they found a stock that a had a volatility of 23.9%. They would use their understanding of the historical data and conclude that this has:

- an 89.9% chance of being in Bucket 1;
- a 10.1% chance of being in Bucket 2; and
- a 0.0% chance of being in Bucket 3, 4, or 5.

It should be noted that the authors do not show any code, nor provide the specific assumptions they made about the distribution. In light of this, we will assume that these distributions are normally distributed with the same standard deviation, say 2.8%. (Please note that the 2.8% number is the standard deviation of the volatilities within a bucket, not the volatilities themselves. These categories are based on average volatilities not average returns.)

Let's look at some R code, because in about 20 lines we can replicate their results.

```
# Build the prior distribution
   bktMeans = c(0.233, 0.297, 0.369, 0.433, 0.676)
   bktSigma = .028   # assume same vol for each bucket
# Observe a data point
   data = 0.239
```

```
# Let's compute the z-scores
   (zScores = (data-bktMeans)/bktSigma)
# What is the height of the pdf at that z-score?
   (prob = round(dnorm(zScores),4))
# Let's normalize these probabilities
   (round (prob/ sum(prob),4))
# There's 89.3% chance we're in Bkt1; 10.7% in Bkt2, 0 elsewhere
# Rather than use our sample mean 23.9, we use 0.239848
   .893 * bktMeans[1] + .107 * bktMeans[2]
```

What have we done differently? Previously, we would have used the sample estimate, 23.9%, as the volatility. Now we use 0.239848. Why is the number higher? To accommodate some 10% chance that we are actually in the next higher volatility bucket.

Here, we exploit the advantages of a Bayesian approach. How? First, by realizing that we know something about this stock's return distribution. It comes from an industry that we have studies on (or at least Finaltis has studied), where we know stocks belong to one of five categories or clusters of volatility. We take advantage of our experience and insights. They become part of the prior distribution. When data is introduced into the system, we can attempt to assign a probability of that stock belonging to each and every category.

The first category has an average volatility close to the observed. Not surprisingly, it has about a 90% chance of being in that group. The second category has an average volatility about 6% away. That is about 2 standard deviations away. On a relative basis, we determine there is about a 10% chance of being in Bucket 2. All the other buckets are multiple standard deviations away and, after normalizing, have virtually no chance of coming from Buckets 3, 4, or 5. This is not an absolute truth per se, but it is true relative to the prior distribution with which we draw inferences.

This example was intentionally chosen to be easy so that it clearly illustrates the methodology at play. In general, you can use the market understanding of your investment management team, combine it with observed data, and make better estimates with the Bayesian approach.

Bibliography

Akerlof, G. and Yellen, J. (1985). Can Small Deviations from Rationality Make Significant Differences to Economic Equilibria? *American Economic Review*, 75(4), pp. 708-20.

Best M. and Grauger R. (1991) On the sensitivity of Mean-Variance-Efficient portfolios to changes in asset means: Some analytical and computational results. *The Review of Financial Studies*, pp. 315-342.

Britten-Jones, M. (1999). The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights. *Journal of Finance*, 14(2), pp. 655-671.

Chopra, V. (1993). Improving Optimization. *Journal of Investing*, 2(3), pp. 51-59.

Croisille, R., Olivier, C., and Renaud, N. (2016) *Bayes to the Rescue of Markowitz* [Online] Available at: http://www.finaltis.com/downloads/finaltisefficientbetaeuro/lettrerecherche/201602ResearchLetter.pdf [Accessed December 2018].

Frankfurter, G., Phillips, H., and Seagle, J. (1971) Portfolio selection: The effects of uncertain means, variances, and co-variances. *Journal of Financial and Quantitative Analysis*, 6, pp. 1251-1262.

Jobson, J.D. and Korkie, B. (1980) Estimation for Markowitz Efficient Portfolios. *Journal of the American Statistical Association*, 75(371), pp. 544-554.

Michaud, R. (1989). *Efficient Asset Management*. Harvard Business School Press: Boston, MA.