Lecture 3

Improvements on and extensions of Monte Carlo Methods

Here we will extend our basic theory and concentrate on some simple techniques to improve the basic method. The techniques we will look at are using antithetic variables which is a very simple adjustment, as is the control variate technique. We will mention something about moment matching, importance sampling and low discrepancy sequences. Most of these are designed to reduce the variance of the error, except for low discrepancy sequences which are used to improve the convergence rate.

Monte Carlo is typically the simplest numerical scheme to implement but as you will see its accuracy and uncertain convergence is not ideal for accurate valuation. However, for complex problems with multiple Brownian motions it is often the only method that can be used which makes it important for us to improve the accuracy of the standard model. In addition to this, when there are multiple Brownian motions AND early exercise features it is crucial that there is an adaptation to the Monte Carlo method that allows us to compute an option value as it may well be the case that other numerical methods will not provide a reasonable estimate.

3.1 Antithetic variables

Antithetic variables or antithetic sampling is a simple adjustment to generating the ϕ_n ($1 \neq n \leq N$). Instead of making N independent draws, you draw the sample in pairs; if the ith Normally distributed variable is ϕ_i, then choose ϕ_{i+1} to be $-\phi_i$, then draw again for ϕ_{i+2}. Note here that $-\phi_i$ is also Normally distributed and most importantly the mean of the two draws is zero, and so this ensures that the mean of the sample paths will be correct and the distributions of draws will be symmetric.

Thus if $N = 500,000$, you only need to make $250,000$ random draws for ϕ and use the negative of the draw to complete the required 500,000 values. This should improve the convergence as the distribution of paths is better matched to the model, i.e. the mean of ϕ is zero. It should also be more efficient (i.e. take less computation time to produce the same accuracy).

Example 3.1 (Antithetic Variables)**.** Find the expectation of a set of antithetic random normally distributed numbers.

Solution 3.1.

Example 3.2 (Psuedo Code)**.** Write down a psuedo code algorithm to generate the option price with this method.

Solution 3.2.

3.2 Moment matching

One fairly simple strategy that works in a similar way to using antithetic variables is called moment matching. The typical way it is done is to ensure that the variance of the sample paths match the variance of the required distribution (antithetic variables will automatically ensure that the mean and skewness match). Our Brownian motion modelling requires the variance of ϕ to be 1, so we would like the variance of our random ϕ's to share this property. To do this we first sample our N ϕ values (requiring $N/2$ random numbers). Then calculate their variance, v say, now replace all of the ϕ values with $\phi \times v^{-\frac{1}{2}}$ and the variance of the new random draws is 1, as required.

Example 3.3 (Moment Matching)**.** Outline how to implement moment matching on a set of random numbers.

Solution 3.3.

3.3 Control variate technique

The *Control Variate* technique is best explained through an example. Consider the following situation:

- We want to compute $E^Q[V(T)]$
- And we can write
$$V(T) = V(T) - V_1(T) + V_1(T),$$
  where
  - $E^Q[V_1(T)]$ is known analytically
  - and error in estimating $E^Q[V(T) - V_1(T)]$ by simulation is less than error in estimating $E^Q[V(T)]$
- Then, a better estimate of $E^Q[V(T)]$ is the sum of
  - The known value of $E^Q[V_1(T)]$
  - Plus the estimate of $E^Q[V(T) - V_1(T)]$

Example 3.4 (Basket Option)**.** Consider a basket option with payoff of

$$V(T) = \max\left[\frac{1}{2}S_1(T) + \frac{1}{2}S_2(T) - X, 0\right]$$

priced under the Black-Scholes framework. Choose a control variate and outline how to calculate the value of the option.

Solution 3.4.

3.4 Importance sampling

The idea behind importance sampling is that if you know that the payoff function is zero outside of an interval $[a, b]$ then any draw which makes S_T lie outside of $[a, b]$ is wasted. Ideally, you would only to sample from distributions that cause S_T to lie in $[a, b]$ and then multiply the result by the actual probability of S_T being in this region. Recall that we have a function that turns a uniform random variable $[0, 1]$ into a realisation of S_T and so we can invert this map to find an interval $[x_1, x_2]$ that is mapped onto $[a, b]$, thus the probability of S_T being in $[a, b]$ is $x_2 - x_1$. Thus to compute the expectation at time T, draw variables from $[0, 1]$, multiply by $x_2 - x_1$ and then add to x_1 so that they are all in $[x_1, x_2]$ and then convert the x value into ϕ as before and get a value of S_T Determine the option value V_T from S_T and then average the option values to obtain an expectation. Then multiply this expectation by $(x_2 - x_1)$.

For example, $S_0 = 100$, $X = 100$, $r = 0.05$, $\sigma = 0.2$, $T = 1$. Consider a call option and so for $[0, 100]$ the payoff contribution is zero. Thus we only need to sample in $[100, \infty)$. $S_T = 100$ corresponds to a ϕ value of:

$$\begin{aligned}\phi &= \frac{\log(100/100) - (r - \frac{1}{2}\sigma^2)}{\sigma} \\ &= -0.1\end{aligned}$$

but $P(\phi = -0.1) = 0.461$. So $[0.461, 1]$ is the range of x values required.

So every draw of x is multiplied by 0.539 and then added to 0.461 to obtain a ϕ value and this only gives S_T values greater than 100.

3.5 Low discrepancy sequences

One of the most useful practical techniques for improving the accuracy of Monte Carlo methods is by using Low discrepancy sequences (also known as Quasi Monte Carlo methods). The theory behind Monte Carlo techniques is that as you take more and more sample paths then they will eventually cover the entire distribution of S_T in the correct manner. Another way to think of this is that our random numbers drawn from $[0, 1]$ will eventually cover this interval in a uniform manner.

Unfortunately, we cannot actually draw an infinite number of paths and for any size of N it may well be the case that our S_T values all cluster around particular values while missing out other regions of S space entirely. This problem becomes more pronounced as we increase the number of dimensions, d

To overcome this problem we throw away the idea of using 'random' numbers at all and instead choose a deterministic sequence of numbers that does a very good job of covering the $[0, 1]$ interval. Note that as we have already discussed, most random number generators are deterministic to some extent and so this approach isn't as odd as you would imagine.

The most interesting thing about a low discrepancy series is that it can improve the convergence of the Monte Carlo method from $1/N^{\frac{1}{2}}$ to $1/N$, making the Monte Carlo method fully competitive with binomial lattices. We will deal with a simple sequence here but there is an extensive literature on low discrepancy series, see Jäckel, *Monte Carlo Methods in Finance* for an excellent summary.

3.6 The Halton Sequence

Sobol sequences are the most common of the low discrepancy sequences, but for ease of explanation we will consider the Halton sequence. The Halton sequence is a sequence of numbers $h(i; b)$ for $i = 1, 2, ...$ where b is the base and all of the numbers in the sequence are in $[0, 1]$. You can choose the base, let us select base 2. The Halton sequence is the reflection of the positive integers in the decimal point and is best observed through an example:

Integers base 10	Integers base 2	Halton sequence base 2	Halton number base 10
1	1	$1 \times \frac{1}{2}$	0.5
2	10	$0 \times \frac{1}{2} + 1 \times \frac{1}{4}$	0.25
3	11	$1 \times \frac{1}{2} + 1 \times \frac{1}{4}$	0.75
4	100	$0 \times \frac{1}{2} + 0 \times \frac{1}{4} + 1 \times \frac{1}{8}$	0.125

- So $10 \rightarrow 0.01$ in base 2 terms.
- In general the ith integer, i, can be expressed as:

$$i = \sum_{j=1}^{m} a_j b^{j-1}$$

where b is the base and $0 \leq a_j < b$. The Halton numbers are given by:

$$h(i; b) = \sum_{j=1}^{m} a_j b^{-j}$$

What is nice about the Halton sequence is that it fills the range $[0, 1]$ gradually When extending to d multiple dimensions you should choose d prime bases. Note that this

Example 3.5 (Halton Sequence)**.** Outline how the Halton Sequence can be used to generate normally distributed numbers.

What problems might arise if you want multiple sequences of random numbers (for multi-factor model)?

Solution 3.5.

3.7 American options

One of the key unanswered questions in finance is how to value options with early exercise features by using a Monte Carlo method. Recall that the American option value V_0 is

$$V_0 = \max_{\tau}[E^Q_{\tau}[e^{-r\tau}\max(S_{\tau} - X, 0)].$$

The problem comes from the fact that Monte Carlo is a forward looking method. To use the sample paths we would have to test early exercise at each point in time on each sample path in order to determine what the optimal exercise strategy would be.

Example 3.6 (Optimisation with Monte Carlo)**.** Outline why this is so difficult to do.

Solution 3.6.

This is incredibly time consuming and not practical, and simplistic approaches, such as the perfect foresight method where you simply choose the highest early exercise value during the lifetime of the option do not give acceptable approximations to the option value.

Graphical explanation of perfect hindsight

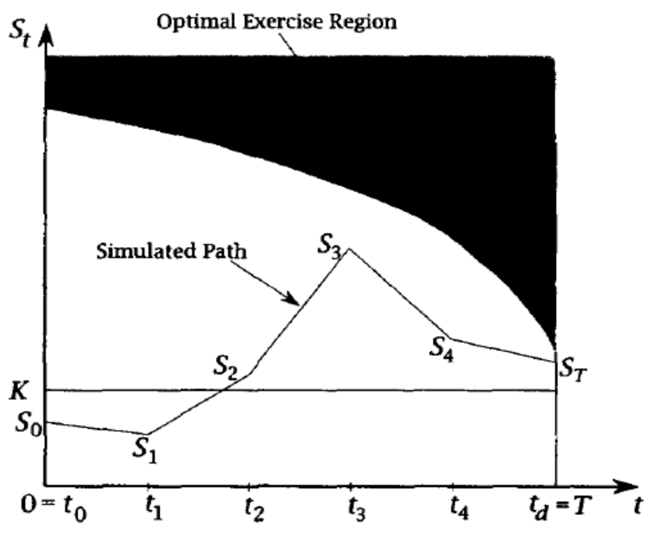

Fig. 1. Sample simulated path.

If you follow the optimal strategy (shaded) then you should only exercise at expiry, but you can get a better payoff from exercising at t_3. This is perfect hindsight and does not value the option correctly.

3.8 Literature Review

Tilley (1993) made the first effort to adapt Monte Carlo methods to cope with early exercise features. The method involves a technique known as "bundling" where the paths are constructed as usual and, at each timestep, they are bracketed into regions of asset price. At expiry the option price is the average of the payoff from all the paths in the bundle. Working backwards through time, as the paths are known at each point it can be checked whether or not it was, on average, worth exercising in each bundle by comparing this to the discounted option price from the future time step. There are evident shortcomings with this approach:

- first, all the paths have to be stored, which can cause computer memory problems
- more importantly the process overvalues the options.
- Crucially, it is also very difficult to extend to options on multiple underlyings (usually the Monte Carlo method's advantage over other numerical techniques).

Realising the main drawback in Tilley's method, Barraquand and Martineau (1995) adapted his approach so that the bundling was in terms of payoff value rather than underlying asset value. Payoff value only has one dimension and so extension to many underlyings does not create any undue problems. However, although not requiring as much memory as Tilley's approach, the approach still does not converge to the correct value and always underestimates the option value (see Boyle et al., 1997) and this estimation error can be serious. This approach is extended by Raymar and Zwecher (1998) but is still less effective than the following two methods. Broadie and Glasserman (1997) approach the Monte Carlo method as by creating upper and lower bounds. To create these bounds they use a 'bushy tree effect' to pursue sub and superoptimal strategies. The superoptimal strategy is obtained by creating a tree whose possible next states are determined, by simulation, all the way to expiry.

Then, in a similar vein to Tilley (1993) the option value at the previous time, t, is the maximum of the average of the values at $t + \Delta t$ discounted and the value from early exercise at t. This strategy does assume the investor has some foresight and so overvalues the option. The suboptimal procedure entails using the b possible paths at each time and, for each path,using the remaining $b - 1$ paths to determine whether the option is continued or exercised. This exercise choice is then applied the initial path one was focusing on. All the possible combinations are then averaged at each timestep. This method is shown to be suboptimal, for more details see Broadie and Glasserman (1997). These can then be combined to provide bounds for the put option value.

As usual, the computational effort increases only linearly as more underlying assets are added. However, as the number of observation times is increased the calculations increase exponentially, i.e with n paths, d observation dates and b branches in the tree the effort is nbd (b here is typically quite large, e.g. 50) Thus, to estimate the value of a continuously, or even daily, observed option involves the use of extrapolation which is somewhat ad hoc if, as is the case here, the initial results are not converging at a known rate. There has been more recent work by Fu et al. (2001) who parametrise the early exercise curve by using Monte Carlo simulations and Rogers (2002) who uses a Lagrangian Martingale to achieve a close upper bound on the option value.

The most popular method for incorporating early exercise features in the Monte Carlo methods is by Longstaff and Schwartz (2001) which we will see in detail in later. Its appeal is that it is simple to implement although there remain questions about its accuracy and efficiency. Another, more academically rigorous approach is the dual approach Haugh and Kogan (2004) which expresses the option pricing problem as a minimisation problem, from which a tight upper bound on the price is calculated. Unfortunately, its calculation is often problematic, although there are practical approaches to circumvent this see Andersen and Broadie (2004) for more details.

The Carriere (1996) valuation of early exercise price for options using simulations and non-parametric regression is also highly regarded.

3.9 Overview

We have looked through a variety of extensions to the standard Monte Carlo in an effort to reduce the variance of the error or to improve the convergence. Most of the improvements are simple to apply such as antithetic variables and moment matching, others are more complex such as low discrepancy sequences. Finally, we looked at some of the early attempts to use Monte Carlo methods to value American style options. This is a precursor to the Longstaff and Schwartz approach.