

Monte Carlo Methods and their Applications to Finance

Chin Zhe Ning

Ren Weiqing

Department of Mathematics

National University of Singapore

Capstone Project for Semester 1, AY25/26

Contents

1. Introduction	3
2. Option Pricing	4
2.1. Examples	4
3. Plain Monte Carlo	5
3.1. Examples	5
3.2. Limitations	6
4. Variance Reduction Techniques	8
4.1. Antithetic Sampling	8
4.1.1. Examples	8
4.1.2. Limitations	8
4.2. Control Variates	8
4.2.1. Examples	9
4.2.2. Limitations	12
4.3. Control Functionals	12
4.3.1. Stein Operator	13
4.3.2. Example	14
4.3.3. Limitations	16
4.4. Stratified Sampling	17
4.5. Importance Sampling	17
5. Applications in Finance	18
6. Conclusion	19
Bibliography	20
Appendix A: Derivations	21
Appendix B: Code	21

1. Introduction

In modern finance, a derivative is a security whose value depends on the price of an underlying asset, a group of assets, or a benchmark. Accurate pricing is essential, as even small mispricings can lead to significant financial consequences. Derivatives are often highly sensitive to fluctuations in their underlying assets, which makes it difficult to value. Prices are typically determined via risk-neutral valuation, in which the expected discounted payoff is computed under a risk-neutral measure. For derivatives with simple payoffs, analytical solutions are well-established. A canonical example is the Black-Scholes formula for a European call option, which provides a closed-form solution under the assumption that the underlying asset price follows a geometric Brownian motion. However, for more complex derivatives, e.g., those that are path-dependent or involve multiple underlying assets, analytical solutions are either difficult or impossible to obtain.

In such cases, Monte Carlo methods provide a powerful framework for estimating expected payoffs. These methods are particularly well-suited for path-dependent derivatives and derivatives involving multiple underlying assets, where conventional numerical techniques may be infeasible. Despite their versatility, Monte Carlo methods exhibit certain limitations. Their slow rate of convergence and the high computational cost required for precise estimates can be significant constraints. Nevertheless, due to their robustness and broad applicability, Monte Carlo simulations have become the workhorse of computational finance.

This report provides an overview of Monte Carlo methods in finance. Working knowledge of derivative pricing is assumed and the focus is on the implementation of Monte Carlo methods for pricing derivatives. We first introduce some basic concepts in option pricing, followed by the plain Monte Carlo method for estimating expected payoffs. The discussion then addresses the limitations of the plain method and also the standard variance reduction techniques.

2. Option Pricing

Suppose that the payoff of a derivative is given by a function $h(X)$ of a random variable X related to the underlying asset. The price of the derivative is given by the expected discounted payoff under a risk-neutral measure, i.e., $v = \mathbb{E}[e^{-rT}h(X)]$, where r is the risk-free interest rate and T is the time to maturity.

The random variable X can take various forms depending on the type of derivative. For example, for a European option, X could be the stock price at maturity S_T ; for an Asian option, X stock prices at multiple time points; for a basket option, X could be a vector of prices of multiple underlying assets. The underlying asset prices are often modeled using stochastic processes. A common model is the geometric Brownian motion

$$S_t = S_0 \exp\left\{\left(r - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right\}, \quad W_t \sim N(0, t)$$

where S_0 is the initial stock price, σ is the volatility, and W_t is a standard Brownian motion. Unless otherwise specified, we will assume that the underlying asset prices follow this model.

2.1. Examples

1. **European Call Option.** For a European call option with underlying stock price S_t the payoff of the option at time T is given by $h(S_T) = (S_T - K)^+$, where K is the strike price, and $z^+ = \max(z, 0)$. The price of the option is

$$v = \mathbb{E}[(S_T - K)^+].$$

It is possible to solve for v analytically from the Black-scholes formula[cite]. Given the parameters S_0, K, r, σ, T , the price of the option is

$$v = S_0 \Phi(\sigma\sqrt{T} - \theta) - Ke^{-rT} \Phi(-\theta)$$

where

$$\theta = \frac{1}{\sigma\sqrt{T}} \log \frac{K}{S_0} + \left(\frac{\sigma}{2} - \frac{r}{\sigma}\right)\sqrt{T}.$$

2. **Asian Call Option.** For a discretely monitored average price call option, the payoff is given by

$$h(S_{t_1}, S_{t_2}, \dots, S_{t_m}) = \left(\frac{1}{m} \sum_{i=1}^m S_{t_i} - K\right)^+,$$

where $0 < t_1 < \dots < t_m = T$ are a fixed set of dates. Under additional assumptions, it is possible to derive a closed-form solution[cite], but in general, there is no analytical solution. In order to compute v , we can simulate discrete paths of the stock price $\{S_{t_i}\}_{i=1}^m$, compute the discounted payoff and then average over N simulated paths.

To estimate the expected payoff v , we can use Monte Carlo methods, which rely on random sampling to approximate the expected value.

3. Plain Monte Carlo

To estimate the quantity

$$v = \mathbb{E}[h(X)]$$

where h is a function of X , we generate N independent and identically distributed (i.i.d.) samples $\{X_i\}_{i=1}^N$ from the distribution of X with density f_X . The plain Monte Carlo estimate of v is given by

$$\hat{v} = \frac{1}{N} \sum_{i=1}^N h(X_i).$$

This is an unbiased estimate \hat{v} of the true mean v , i.e., $\mathbb{E}[\hat{v}] = v$, and the variance of \hat{v} is given by

$$\text{Var}(\hat{v}) = \frac{1}{N^2} \sum_{i=1}^N \text{Var}(h(X_i)) = \frac{\text{Var}(h(X))}{N},$$

As the number of samples N increases, the variance of \hat{v} decreases proportionally to $1/N$. Consequently, the estimator \hat{v} converges to the true mean v in the long run.

Pseudocode:

for $i = 1, \dots, N$:

 generate X_i from f_X

 set $H_i = h(X_i)$

compute the estimate

$$\hat{v} = \frac{1}{N} \sum_{i=1}^N H_i$$

compute the standard error

$$\text{S.E.} = \sqrt{\frac{1}{N(N-1)} \left(\sum_{i=1}^N H_i^2 - N\hat{v}^2 \right)}$$

3.1. Examples

1. **European Call Option.** Consider the problem of estimating the price of a European call option described in Section 2.1 with strike price $K = 60$ and maturity $T = 1$ year. Assuming the parameters

$$S_0 = 50, r = 0.05, \sigma = 0.2,$$

we estimate the option price using the plain Monte Carlo method for various sample sizes N . The results are summarised in Table 1.

Sample Size	$N = 10^2$	$N = 10^3$	$N = 10^4$	$N = 10^5$	$N = 10^6$
Estimate	1.8824	1.6428	1.6586	1.6119	1.6181
S.E.	0.6072	0.1388	0.0447	0.0135	0.0043
R.E.	32.25%	8.45%	2.69%	0.84%	0.27%

Table 1: European Call Option: Plain Monte Carlo Estimate vs Sample Size

As N increases, the estimate \hat{v} converges to the true option price $v = 1.6237$, and the standard error decreases to zero, this demonstrates the consistency of the Monte Carlo estimate.

2. **Asian Call Option.** Consider the problem of estimating the price of a discretely monitored Asian call option described in Section 2.1. Assuming the parameters

$$S_0 = 50, r = 0.05, \sigma = 0.2, K = 60, T = 1$$

and monitoring at $m = 100$ equally spaced time points, we estimate the option price using the plain Monte Carlo method for various sample sizes N . The results are summarised in Table 2.

Sample Size	$N = 10^2$	$N = 10^3$	$N = 10^4$	$N = 10^5$	$N = 10^6$
Estimate	0.2823	0.2534	0.2650	0.2715	0.2700
S.E.	0.1224	0.0357	0.0122	0.0040	0.0013
R.E.	43.36%	14.09%	4.60%	1.47%	0.47%

Table 2: Asian Call Option: Plain Monte Carlo Estimate vs Sample Size

The standard error decreases as the sample size N increases, as observed in the European call option example.

3.2. Limitations

The numerical results in the previous section for the European call and Asian options demonstrate that, although the plain Monte Carlo method provides unbiased estimates, its convergence rate is relatively slow.

According to the *Central Limit Theorem (CLT)*, if $\text{Var}(h(X)) = \sigma^2 < \infty$, then as $N \rightarrow \infty$,

$$\sqrt{N}(\hat{v} - v) \rightarrow N(0, \sigma^2),$$

implying that for large N , the estimation error $\hat{v} - v$ is approximately normally distributed with mean 0 and standard deviation $\sigma N^{-\frac{1}{2}}$. So to be precise, the standard error (S.E.) of the estimate decays at rate $\mathcal{O}(N^{-\frac{1}{2}})$.

This is illustrated in Figure 1: the log-log plot of S.E. versus sample size N follows a straight line with slope approximately $-\frac{1}{2}$, so the empirical results match the theoretical rate.

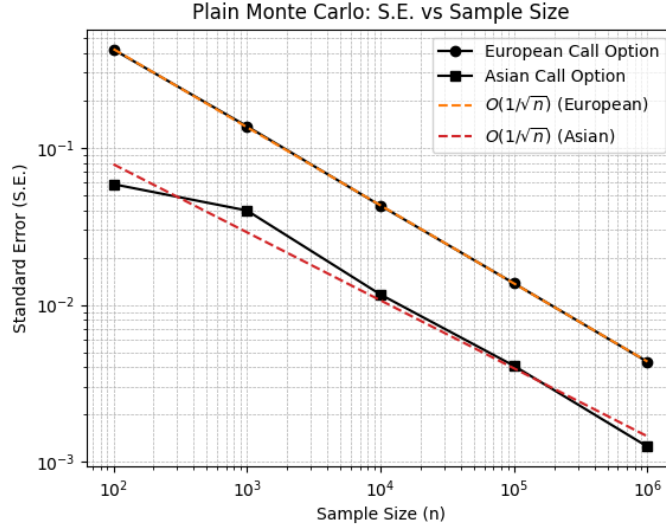


Figure 1: Plain Monte Carlo: Standard Error vs Sample Size

While this rate is independent of the problem dimension d , the variance σ^2 often increases with d or payoff complexity, and can bear a significant cost on estimation accuracy. For example, when the option becomes *deep out-of-the-money*, the payoff distribution becomes very skewed leading to large σ^2 . In Figure 2, we vary the strike price of the European call option from $K = 40$ to $K = 80$ and observe the R.E. for a fixed sample size of $N = 10,000$. As K increases, the R.E. also increases significantly.

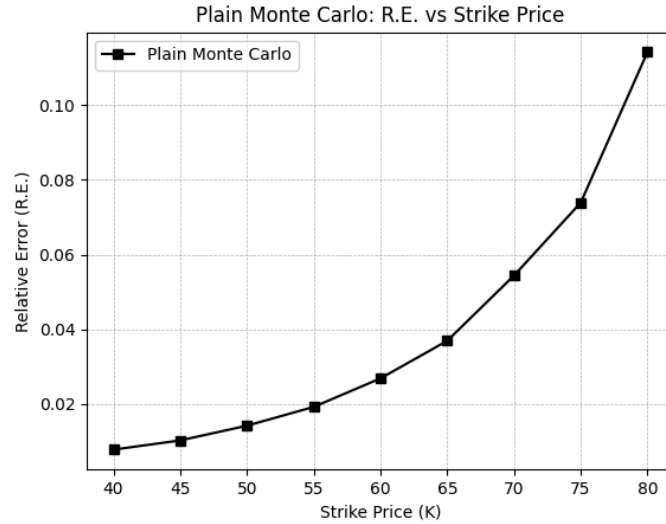


Figure 2: Plain Monte Carlo: Relative Error vs Strike Price

To get a more accurate estimate, increasing the sample size N can become very computationally expensive: reducing S.E. by a factor of ten requires roughly one hundred times more samples. This limitation motivates the development of *variance reduction techniques*, which aim to decrease the variance σ^2 of the estimator without increasing the sample size.

4. Variance Reduction Techniques

Although the plain Monte Carlo method provides unbiased estimates and is straightforward to implement, its slow convergence rate and potentially large variance often make it computationally inefficient for high-precision estimation. To address this, a range of variance reduction techniques have been developed to improve estimator accuracy without proportionally increasing the number of samples.

This section introduces several widely used methods: antithetic sampling, control variates, control functionals, stratified sampling, and importance sampling. Typically, these techniques construct a new Monte Carlo estimator \hat{v}_{new} that is still unbiased but has lower variance than the plain estimator \hat{v} for the same sample size N .

Other techniques, such as quasi-Monte Carlo methods and multi-level Monte Carlo, are often used to obtain further gains in efficiency but are beyond the scope of this report.

4.1. Antithetic Sampling

Antithetic sampling is a variance reduction technique that involves generating i.i.d. pairs of random variables (X_i, Y_i) such that

- i. X_i and Y_i have the same distribution.
- ii. X_i and Y_i are *negatively correlated*.

We can construct a new unbiased estimator for $\mathbb{E}[X]$ of the form

$$\hat{v}_{\text{AS}} = \frac{1}{N} \sum_{i=1}^N \frac{X_i + Y_i}{2}.$$

If β is the correlation between X and Y , and $\beta < 0$, then the variance of \hat{v}_{AS}

$$\text{Var}[\hat{v}_{\text{AS}}] = (1 + \beta) \frac{\text{Var}[X]}{2N}$$

is lower than the variance of the plain Monte Carlo estimator on $2N$ sample, that is,

$$\text{Var}[\hat{v}] = \frac{\text{Var}[X]}{2N}$$

Hence, the effectiveness of antithetic sampling depends on the strength of the negative correlation; stronger negative correlation leads to greater variance reduction¹.

4.1.1. Examples

4.1.2. Limitations

4.2. Control Variates

Suppose we can find another random variable Y such that:

- i. Y is correlated with X .

¹In practice, the antithetic pairs (X_i, Y_i) are often generated by transforming a common random variable. For example, if $X_i = h(U_i)$ where $U_i \sim \text{Uniform}(0, 1)$, then we can set $Y_i = h(1 - U_i)$. This construction ensures that X_i and Y_i are negatively correlated if h is monotonic.

- ii. Simulating Y_i alongside X_i is computationally inexpensive.
- iii. The expected value $\mathbb{E}[Y]$ is known exactly.

Then we can construct a new estimator for $\mathbb{E}[X]$ that leverages the information from Y to reduce variance. The form of the new estimator is

$$\hat{v}_{\text{CV}} = \frac{1}{N} \sum_{i=1}^N X_i - b \left(\frac{1}{N} \sum_{i=1}^N Y_i - \mathbb{E}[Y] \right)$$

and Y is known as the *control variate*. Note that \hat{v}_{CV} is still an unbiased estimator of $\mathbb{E}[X]$ for any choice of $b \in \mathbb{R}$. Then the variance of \hat{v}_{CV} is a quadratic in b :

$$\text{Var}[\hat{v}_{\text{CV}}] = \frac{1}{N} [\text{Var}[X] + b^2 \text{Var}[Y] - 2b \text{Cov}(X, Y)],$$

which is minimized when b is chosen as

$$b^* = \beta \frac{\sigma_X}{\sigma_Y} = \frac{\text{Cov}(X, Y)}{\text{Var}(Y)}$$

where β is the correlation between X and Y . Intuitively, we are regressing X against Y and removing the component of X that is explained by Y , resulting in lower variance². Under b^* , the variance of the new estimator is

$$\text{Var}[\hat{v}_{\text{CV}}] = (1 - \beta^2) \frac{\text{Var}[X]}{n}.$$

This is lower than the variance of the plain Monte Carlo estimator by a factor of $1 - \beta^2$. Therefore, the effectiveness of the control variate method depends on the strength of the correlation between X and Y ; stronger correlation leads to greater variance reduction.

4.2.1. Examples

1. **European Call Option.** Suppose we want to estimate the price of a European call option with strike price K and maturity T using Monte Carlo simulation. The payoff of the option is given by $h(S_T) = (S_T - K)^+$, where S_T is the stock price at maturity. We can use the stock price S_T itself as a control variate since its expected value under the risk-neutral measure is known.

$$\mathbb{E}[S_T] = S_0 e^{rT}$$

The control variate estimator is given by

$$H = \bar{X} - b(\bar{Y} - \mathbb{E}[Y])$$

where $X = e^{-rT} h(S_T)$ and $Y = S_T$. The optimal coefficient b^* can be estimated from a pilot simulation.

The plain Monte Carlo estimate is compared against the control variate estimate below.

[table of results]

²Usually, b^* is estimated from a pilot run independent to the main simulation, otherwise bias may be introduced. In this view, b^* is the estimated coefficient of Y in the linear model $X \sim Y$.

2. **Straddle Option.** A straddle option is a combination of a call and a put option with the same strike price and maturity. The payoff of a straddle option is

$$h(S_T) = (S_T - K)^+ + (K - S_T)^+.$$

Similar to the case of a European call option, the terminal stock price S_T can serve as a control variate to reduce the variance of Monte Carlo estimates. Consider the following parameter values:

$$K = 50, r = 0.02, \sigma = 0.2, \text{ and } T = 1.$$

We summarize the results for $S_0 = 40, 50, 60$ and sample size of $N = 10,000$ in Table 3.

Strike Price	$S_0 = 40$		$S_0 = 50$		$S_0 = 60$	
	Plain MC	CV	Plain MC	CV	Plain MC	CV
Estimate	10.3312	10.4381	7.8795	7.8737	12.7905	12.7598
S.E.	0.0614	0.0400	0.0628	0.0593	0.1018	0.0416
R.E.	0.59%	0.38%	0.80%	0.75%	0.83%	0.33%
$\hat{\beta}^2$	57.54%		10.89%		83.34%	

Table 3: Straddle Option: control variates versus plain Monte Carlo

A variance reduction of approximately 57.5% is observed when using S_T as a control variate for S_0 . However, as S_0 approaches the strike price $K = 50$, the efficiency of this control variate diminishes, as illustrated in Figure 3.

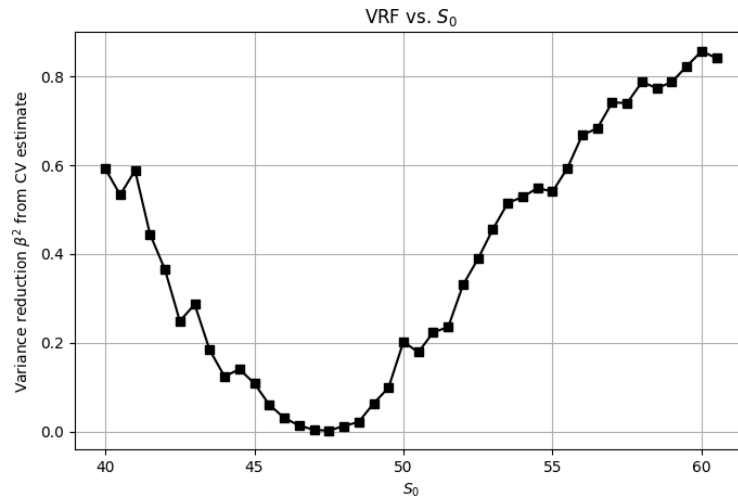


Figure 3: Variance reduction from the control variate as S_0 varies from 40 to 60, with $K = 50$.

When $S_0 = K = 50$, the variance reduction is approximately 10.9%. This decline in performance can be explained by the weak correlation between S_T and the straddle payoff $h(S_T)$. As shown in Figure 4, the relationship between these two quantities is highly non-linear when $S_0 \approx K$. Because the payoff function is symmetric and exhibits a kink at $S_T = K$, the linear CV S_T fails to capture much of the variation in $h(S_T)$, leading to suboptimal variance reduction.

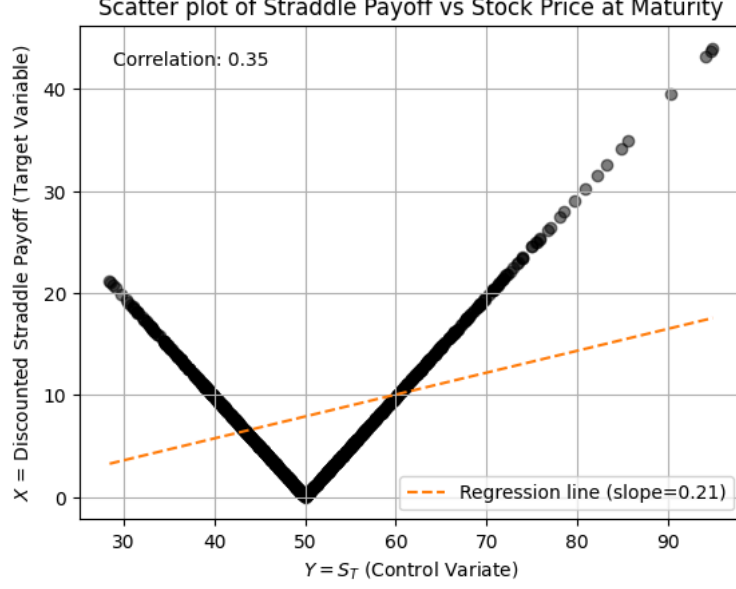


Figure 4: $h(S_T)$ vs. S_T for straddle Option, $S_0 = K = 50$, using S_T as a control variate.

To improve upon this, additional control variates can be introduced. In particular, using both S_T and S_T^2 may result in a better fit. Under the risk-neutral measure, the expected value of S_T^2 is known analytically as

$$\mathbb{E}[S_T^2] = \mathbb{E}[S_T]^2 + \text{Var}(S_T) = S_0^2 e^{2rT + \sigma^2 T}.$$

The optimal coefficients b_1^* and b_2^* can then be estimated via multiple linear regression of $h(S_T)$ on S_T and S_T^2 . The new control variate estimator is given by

$$\hat{H} = \bar{X} - b_1^*(\bar{Y}_1 - \mathbb{E}[Y_1]) - b_2^*(\bar{Y}_2 - \mathbb{E}[Y_2])$$

where $Y_1 = S_T$ and $Y_2 = S_T^2$. Incorporating both control variates yields a variance reduction of approximately 86.2%, a substantial improvement over the single control case.

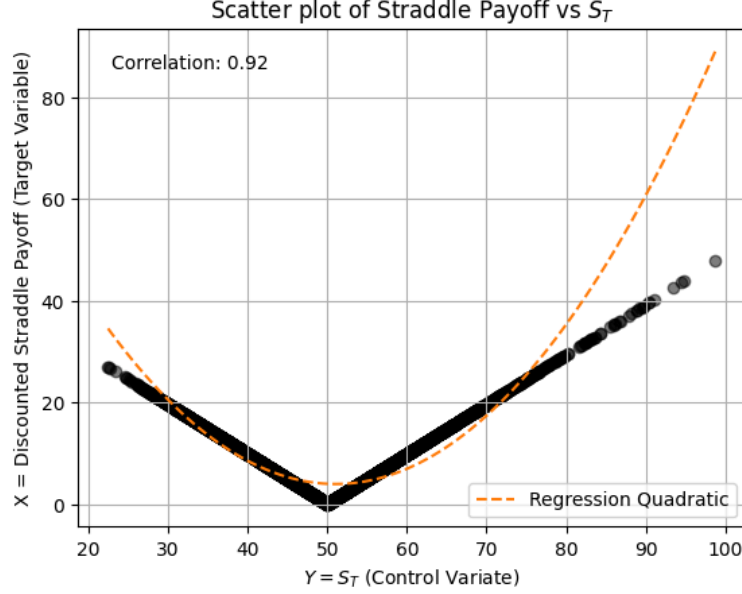


Figure 5: $h(S_T)$ vs. S_T for straddle Option, $S_0 = K = 50$, using S_T and S_T^2 as control variates.

This result highlights that for payoffs exhibiting strong non-linearity such as the straddle option, augmenting the control variate basis with higher-order terms could significantly enhance variance reduction efficiency.

4.2.2. Limitations

The requirement that $\mathbb{E}[Y]$ is known exactly is often prohibitive – in many practical scenarios, it may not be feasible to find a control variate with a known expected value. There have been some attempts to get around this by designing a finite basis of control variates

$$\Phi(Y) = \{\psi_1(Y), \psi_2(Y), \dots, \psi_m(Y)\}$$

each with a known expectation $\mathbb{E}[\psi_i(Y)]$. This approach is powerful but inherently limited by the expressiveness of the finite basis. The space of functions that can be represented in this basis may not be rich enough to approximate the relationship between X and Y well. If the dependence is highly non-linear or complex, capturing it accurately would require a very large number of basis functions, making the method computationally infeasible.

To overcome these limitations, the control functional (CF) framework introduced in [1] places the payoff function within a *richer function space*, enabling flexible, non-parametric approximation of complex relationships.

4.3. Control Functionals

The key idea of control functionals is to replace the *finite linear basis* with a **function space** that can approximate a much broader class of relationships between random variables.

Suppose we wish to estimate $\mathbb{E}[h(Z)]$ where Z is a random variable with known density $p(z)$ and h is a given integrable function. In this setting, the classical control variate idea can be seen as constructing a function $g(Z)$ (the control variate) that approximates $h(Z)$ from basis functions $\psi_i(Z)$. That is,

$$h(Z) \approx g(Z) = b_1\psi_1(Z) + b_2\psi_2(Z) + \dots + b_m\psi_m(Z)$$

where $\mathbb{E}[g(Z)]$ is known from $\mathbb{E}[\psi_i(Z)]$ and the coefficients b_i are chosen to minimize the mean squared error between $h(Z)$ and $g(Z)$ over the samples Z_i .

The control functional framework extends this idea by allowing g to come from a *reproducing kernel hilbert space* denoted \mathcal{H}_+ . This hilbert space is associated with a positive-definite kernel $k_+(\cdot, \cdot)$ that implicitly defines an infinite-dimensional set of basis functions $\psi_i(z) = k_+(z, z_i)$ for any $z_i \in \Omega$. So we can write

$$h(Z) \approx g(Z) = \sum_{i=1}^{\infty} b_i \psi_i(Z)$$

where $c \in \mathbb{R}$ is a constant term. The kernel $k_+(\cdot, \cdot)$ is constructed such that the corresponding functions have analytically known expectations (in particular, $\mathbb{E}[\psi_i(Z)] = 1$ for all i).

Now the control functional $g \in \mathcal{H}_+$ is the function that best approximates h in \mathcal{H}_+ . This can be realized as the solution to the following regularized least squares problem:

$$g = \arg \min_{g \in \mathcal{H}_+} \frac{1}{N} \sum_{i=1}^N (h(z_i) - g(z_i))^2 + \lambda \|g\|_{\mathcal{H}_+}^2 \quad (1)$$

where $\lambda > 0$. After obtaining g , the control functional estimator for $\mathbb{E}[h(Z)]$ is given by

$$\hat{H}_{\text{cf}} = \bar{h} - \bar{g} + \mathbb{E}[g(Z)]$$

where \bar{h} and \bar{g} denote the sample means of $h(Z_i)$ and $g(Z_i)$ respectively, and $\mathbb{E}[g(Z)]$ is computed as the sum of b_i 's. Because g resembles h , the residuals $h(Z_i) - g(Z_i)$ have lower variance than $h(Z_i)$ alone, leading to a more efficient estimator.

4.3.1. Stein Operator

The construction of the desired RKHS \mathcal{H}_+ , we rely on the *Stein operator* \mathcal{S}_p , which encodes information about the target distribution p . For a sufficiently smooth function φ , the Stein operator is defined as

$$\mathcal{S}_p[\varphi](z) = \nabla_z \cdot \varphi(z) + \varphi(z) \cdot \nabla_z \log p(z)$$

where ∇_z denotes the gradient with respect to z .

1. $\mathcal{H} \rightarrow \mathcal{H}_0$. A key property of this operator is that, for any φ belonging to an appropriate function class \mathcal{H} , $\mathbb{E}[\mathcal{S}_p[\varphi](Z)] = 0$.

Under mild regularity conditions, we can define the hilbert space

$$\mathcal{H}_0 = \{\mathcal{S}_p[\varphi] : \varphi \in \mathcal{H}\},$$

where \mathcal{H} is a base RKHS with kernel $k(\cdot, \cdot)$. The induced kernel on \mathcal{H}_0 is given by

$$k_0(z, z') = \nabla_z \cdot \nabla_{z'} k(z, z') + u(z) \cdot \nabla_{z'} k(z, z') + u(z') \cdot \nabla_z k(z, z') + u(z) \cdot u(z') k(z, z')$$

where $u(z) = \nabla_z \log p(z)$ is the *score function* of the distribution.

Equivalently, k_0 can be expressed more compactly as

$$k_0(z, z') = \mathcal{S}_p^z \mathcal{S}_p^{z'} [k(z, z')],$$

where \mathcal{S}_p^z and $\mathcal{S}_p^{z'}$ denote the Stein operator acting on the first and second arguments of k respectively. This kernel satisfies the zero-mean property

$$\int k_0(z, z')p(z') \, dz' = 0$$

for almost all $z \in \Omega$, which ensures that any function $g \in \mathcal{H}_0$ has zero expectations under p , i.e., $\mathbb{E}_p[g(Z)] = 0$. This property is fundamental to control functionals as it guarantees that all functions in \mathcal{H}_0 can serve as valid, analytically tractable control variates.

2. $\mathcal{H}_0 \rightarrow \mathcal{H}_+$. As established, the stein operator induces a RKHS \mathcal{H}_0 of zero-mean functions. This is ideal for constructing mean-zero control variates, but to approximate a general target function h , we also need to represent its mean component.

To achieve this, we augment \mathcal{H}_0 with the one-dimensional RKHS of constant functions:

$$\mathcal{C} = \{c : c \in \mathbb{R}\},$$

which has the reproducing kernel $k_C(z, z') = 1$ for all $z, z' \in \Omega$. The augmented RKHS is then defined as

$$\mathcal{H}_+ = \mathcal{C} + \mathcal{H}_0 = \{c + g : c \in \mathbb{R}, g \in \mathcal{H}_0\}.$$

By the properties of RKHS addition, the corresponding kernel is simply the sum of the component kernels:

$$k_+(z, z') = k_C(z, z') + k_0(z, z') = 1 + k_0(z, z').$$

This construction gives us a rich function space \mathcal{H}_+ that can approximate a wide variety of target functions h , while still allowing for analytically tractable expectations via the Stein operator.

4.3.2. Example

Let $Z \sim N(0, 1)$, so $p(z) = (2\pi)^{-\frac{1}{2}}e^{-z^2/2}$, and consider the standard Radial Basis Function (RBF) kernel:

$$k(z, z') = \exp(-(z - z')^2/2\alpha^2)$$

that defines an RKHS \mathcal{H} of smooth functions, with basis functions $\{k(z, z_i)\}_{i=1}^n$ for any set of points $\{z_i\}_{i=1}^n$. The standard normal density has score function $u(z) = \nabla_z \log p(z) = -z$, so the corresponding *Stein operator* is

$$\mathcal{S}_p[f](z) = f'(z) - zf(z),$$

where f is differentiable. Applying the Stein operator to the RBF kernel, we obtain

$$k_0(z, z') = \mathcal{S}_p^z \mathcal{S}_p^{z'}[k(z, z')] = e^{-(z-z')^2/2\alpha^2} \left[\frac{1}{\alpha^2} - (z - z')^2 \left(\frac{1}{\alpha^4} + \frac{1}{\alpha^2} \right) + zz' \right].$$

See Figure 6 for a visualization of elements from \mathcal{H} and \mathcal{H}_0 .

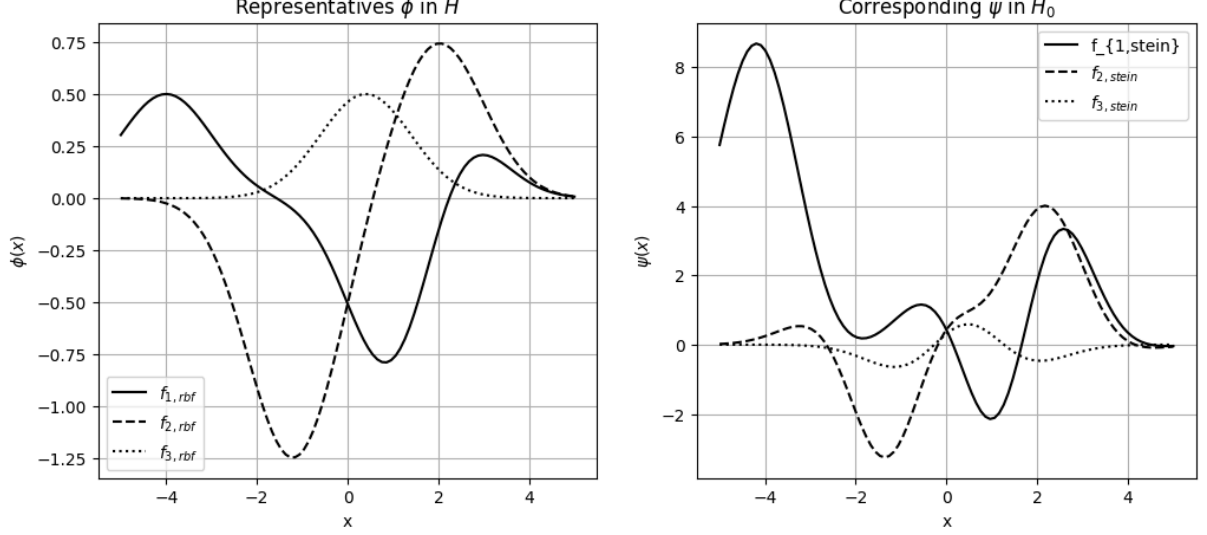


Figure 6: Elements of \mathcal{H} and \mathcal{H}_0 for the standard normal distribution and RBF kernel with lengthscale $\alpha = 1.0$

1. **Butterfly Multistrike Option.** Consider a deep out-of-the-money butterfly option with strike prices $K_1 = 90, K_2 = 100$, and $K_3 = 120$. The payoff of the option at maturity T is given by

$$h(S_T) = (S_T - K_1)^+ - 2(S_T - K_2)^+ + (S_T - K_3)^+.$$

We compare the performance of the plain Monte Carlo estimator against the control functional estimator using the standard normal distribution and RBF kernel with $\alpha = 1.0$. The parameters used are:

$$r = 0.05, \sigma = 0.3, T = 1.$$

For $S_0 = 60$ and a sample size of $N = 10,000$, we report the results in Table 4 below:

	Estimate	S.E.	R.E.
Plain MC	0.1585	0.0179	11.28%
Control Variate	0.1608	0.0181	11.27%
Control Functional	0.1628	0.0023	1.39%
Theoretical	0.1628		

Table 4: Butterfly Option: control functionals versus control variates and plain Monte Carlo

The control functional estimator achieves a significant variance reduction compared to both the plain Monte Carlo and control variate methods. Specifically, the relative error is reduced from 11.28% to 1.39%. This demonstrates the effectiveness of control functionals in capturing complex relationships in the payoff function that linear control variates may miss.

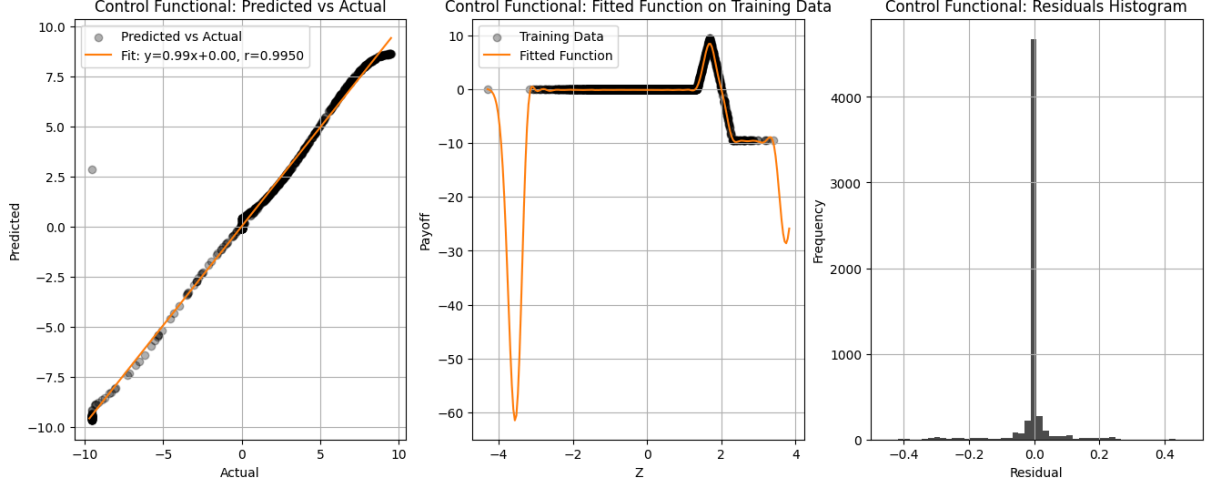


Figure 7: Butterfly Multistrike Option Pricing with $S_0 = 60$ and $N = 10,000$.

4.3.3. Limitations

While control functionals provide a powerful framework for variance reduction, they come with certain limitations. This section discusses several key limitations, focusing on computational cost, kernel selection, and theoretical assumptions.

1. **Computational Cost.** A principal drawback of control functional methods lies in their computational complexity. Solving the RLS problem in equation (1) involves inverting an $N \times N$ kernel matrix K , which incurs a computational cost of $\mathcal{O}(N^3)$ in time and $\mathcal{O}(N^2)$ cost in memory. For large sample sizes N , the variance reduction benefits must be weighed against the increased computational burden. In contexts where the cost of function evaluation is low, the overhead of control functionals may not be justified.
2. **Kernel Selection.** The performance of control functionals is sensitive to the choice of kernel and its hyperparameters (e.g., lengthscale in RBF kernels). Selecting an appropriate kernel may require domain knowledge or empirical tuning, which can be non-trivial.
3. **Theoretical Assumptions.** The theoretical appeal of control functionals stem from their super-root- N rate of convergence, which, under ideal conditions, can outperform other variance reduction techniques. However, these rates are guaranteed only under specific regularity assumptions about the kernel and target function. In particular, this convergence property assumes that

$$\sup_{z \in \Omega} k_0(z, z) < \infty,$$

ensuring that the reproducing kernel k_0 is bounded. Commonly used kernels such as the RBF kernel

$$k(z, z') = \exp(-(z - z')^2 / 2\alpha^2)$$

examined in the previous example, the derived kernel k_0 does not satisfy this assumption. To address this, alternative kernels with bounded k_0 can be employed. For

instnace, Oates *et al.* [1] modifies the RBF kernel by introducing polynomial decay terms to ensure boundedness:

$$k'(z, z') = (1 + \alpha_1(z^2 + z'^2))^{-1} \exp(-(z - z')^2 / 2\alpha_2^2)$$

where $\alpha_1, \alpha_2 > 0$.

Despite these limitations, ongoing research continues to refine and extend the control functional framework, aiming to mitigate these challenges and broaden its applicability.

4.4. Stratified Sampling

4.5. Importance Sampling

Importance sampling (IS) reduces variance by sampling more frequently from regions of the domain that contribute most to the expectation. Instead of drawing samples from the original distribution $f(x)$, samples are taken from a proposal distribution $g(x)$ and reweighted to correct for the change in measure.

To estimate $v = \mathbb{E}[h(X)]$, the standard Monte Carlo estimator

$$\hat{v} = \frac{1}{N} \sum_{i=1}^N h(X_i), \quad X_i \sim f(x)$$

is replaced by the importance sampling estimator

$$\hat{v}_{\text{is}} = \frac{1}{N} \sum_{i=1}^N h(X_i) \frac{f(X_i)}{g(X_i)}, \quad X_i \sim g(x).$$

The choice of $g(x)$ is critical: a good proposal distribution closely matches the shape of the function $|h(x)|f(x)$, leading to lower variance in the weighted estimator.

5. Applications in Finance

6. Conclusion

Bibliography

- [1] C. J. Oates, M. Girolami, and N. Chopin, “Control Functionals for Monte Carlo Integration,” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 79, no. 3, pp. 695–718, 2016, doi: 10.1111/rssb.12185.
- [2] P. Glasserman and J. Li, “Importance Sampling for Portfolio Credit Risk,” *Management science*, vol. 51, no. 11, pp. 1643–1656, 2005.

Appendix A: Derivations

Appendix B: Code