# Monte Carlo Simulation of SDEs and Variance Reduction via Control Variates

Aleksandr Mazovetskii

July 30, 2026

# Contents

# 1 Introduction

Since stochastic differential equations rarely admit closed-form solutions, expectations of functionals of their solutions are computed by discretization combined with Monte Carlo simulation, which incurs a discretization bias of order $\mathcal{O}(\Delta t)$, where $\Delta t$ is the time step of the discretization scheme, and a statistical error of order $\mathcal{O}(N^{-1/2})$, where $N$ is the number of simulated paths. The latter rate is dimension-independent but slow, so reducing the variance of the estimator is usually cheaper than simulating more paths. This paper studies the control variate method, with European call pricing under geometric Brownian motion as the main application.

# 2 Stochastic processes & SDEs

## 2.1 Stochastic processes

Firstly, we introduce the notion of a probability space. It consists of an outcome space $\Omega$, whose elements $\omega \in \Omega$ are called outcomes, a $\sigma$-algebra $\mathcal{F}$, and a probability measure $P : \mathcal{F} \to [0, 1]$.

**Definition 1 (Stochastic Process).** A stochastic process is a collection of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ and indexed by a totally ordered set $\mathbb{T} \subseteq [0, \infty)$, called the index set (here always interpreted as time):

$$(X_t)_{t\in\mathbb{T}}$$

If $t \in \mathbb{T}$ is fixed then a single random variable can be considered as a function:

$$\omega \mapsto X_t(\omega)$$

If $\omega$ is fixed then we consider a single path:

$$t \mapsto X_t(\omega), \qquad t \in \mathbb{T}$$

**Note:** Random variable $X$ is a measurable function defined on a probability space $(\Omega, \mathcal{F}, P)$: $X : \Omega \to \mathbb{R}$

Throughout this paper $\mathbb{T} = [0, T]$ for a fixed terminal time $T > 0$. As time passes, more events that were undecided at earlier times become decided. The filtration formalizes this: $\mathcal{F}_t$ collects the events whose occurrence is determined by what has happened up to time $t$.

**Definition 2 (Filtration and $\mathcal{F}_t$-adapted process).** Given a probability space $(\Omega, \mathcal{F}, P)$, we call a growing collection of $\sigma$-algebras filtration if:

$$\forall s, t \in \mathbb{T} : s \leq t \implies \mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}.$$

A stochastic process is said to be $\mathcal{F}_t$-adapted if

$$\forall t \in \mathbb{T} : X_t \text{ is } \mathcal{F}_t\text{-measurable}$$

This matters because at time $t$ we cannot base decisions on information that is not yet available; adaptedness is exactly the requirement that $X_t$ be determined by $\mathcal{F}_t$.

Another important concept is *Martingale*. One can think of it as a fair game process, meaning the expected value over time is equal to the amount you have at the moment.

**Definition 3 (Martingale).** [4] A stochastic process $(X_t)_{t\in\mathbb{T}}$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{T}}, P)$ is called a martingale if it satisfies the following conditions

1. $X_t$ is $\mathcal{F}_t$-measurable for all $t \in \mathbb{T}$.
2. $\mathbb{E}[|X_t|] < \infty$ for all $t \in \mathbb{T}$.
3. $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ for all $s, t \in \mathbb{T}$ with $s \leq t$.

## 2.2 Brownian motion (Wiener process)

An important example of a stochastic process is Brownian motion. It was discovered by Robert Brown, who observed random motion of pollen grains in liquid. It can be modeled as a stochastic process as follows: $(B_t)_{t\in\mathbb{T}}$. It has some properties that might be helpful for us:

1. The Brownian motion starts at zero: $B_0 = 0$
2. The Brownian motion is almost surely continuous.
3. The increments over disjoint intervals are independent, and are stationary: $\forall s, t \in \mathbb{T} : s < t \implies B_t - B_s \sim \mathcal{N}(0, t-s)$. This also implies that: $\mathbb{E}[B_t] = \mathbb{E}[B_t - B_0] = 0$

The third property is crucial for numerical simulations: the Brownian increments $\Delta B$ should be sampled from the normal distribution $\mathcal{N}(0, \Delta t)$.

It is important to mention that despite being continuous, Brownian motion is nowhere differentiable. More than that, if we zoom in on the motion we will see a fractal behavior, in contrast to normal smooth functions where it would eventually look like a line.

One more important property that is often used is called quadratic variation that is equal to the elapsed time. For a partition $0 = t_0 < \ldots < t_n = T$ of $[0, T]$, as the mesh $\max_i(t_i - t_{i-1})$ vanishes:

$$\sum_{i=1}^{n}(B_{t_i} - B_{t_{i-1}})^2 \xrightarrow{L^2} T$$

This is the cumulative discretized version of:

$$(dB_t)^2 = dt$$

Brownian motion is very important because its increments will allow us to model much more complex processes. If we add a drift term we will get a path that goes in some direction with a noisy behavior.

## 2.3 Stochastic calculus

Many stochastic processes of interest arise as solutions of a Stochastic Differential Equation:

$$\forall t \in \mathbb{T} : \quad dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t$$

Assuming that we have an initial condition $X_0$, an equivalent integral equation would be:

$$X_t = X_0 + \int_0^t \mu(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s$$

Here $\int_0^t \mu(s, X_s)ds$ is a standard Riemann integral and $\int_0^t \sigma(s, X_s)dB_s$ is an Itô integral.

To see why the Riemann integral is not an appropriate method for a stochastic component, we consider the following random variable:

$$I_T(\omega) = \int_0^T B_s \, dB_s$$

We approximate the integral by summing areas of rectangles under/above the curve and split $[0, T]$ into $n$ subintervals: $0 = t_0 < t_1 < \ldots < t_n = T$. And let $\Delta B_i = B_{t_{i+1}} - B_{t_i}$.

The Left-Endpoint and Right-Endpoint Sums:

$$R_{\text{left}} = \sum_{i=0}^{n-1} B_{t_i}(B_{t_{i+1}} - B_{t_i}), \qquad R_{\text{right}} = \sum_{i=0}^{n-1} B_{t_{i+1}}(B_{t_{i+1}} - B_{t_i})$$

Under the Riemann logic this difference should vanish as $n \to \infty$, instead

$$R_{\text{right}} - R_{\text{left}} = \sum_{i=0}^{n-1}(B_{t_{i+1}} - B_{t_i})(B_{t_{i+1}} - B_{t_i}) = \sum_{i=0}^{n-1}(\Delta B_i)^2 \xrightarrow{L^2} T \quad \text{(as the mesh vanishes)}$$

The sums converge to different limits. The fundamental requirement of Riemann integration is violated, because the limits depend entirely on where we evaluate the function within the interval. This forces us to invent a new method for solving such equations.

The idea behind the Itô integral is to use the left-endpoint sum. This is especially helpful because it coincides with our $\mathcal{F}_t$-adaptedness requirement. In other words we are not looking in the future calculating the sum, but using the information we already have. For this we define a class of Itô integrable functions.

**Definition 4 (Class of Itô integrable functions).** [5] Let $U \in [0,T)$ be a fixed lower integration bound and let $\mathcal{V} = \mathcal{V}(U,T)$ be the class of functions

$$f(t,\omega) : [0,\infty) \times \Omega \to \mathbb{R}$$

such that

1. $(t,\omega) \to f(t,\omega)$ is $\mathcal{B} \times \mathcal{F}$-measurable, where $\mathcal{B}$ denotes the Borel $\sigma$-algebra on $[0,\infty)$
2. $f(t,\omega)$ is $\mathcal{F}_t$-adapted.
3. $\mathbb{E}[\int_U^T f(t,\omega)^2 dt] < \infty$

The next step is to approximate $f(t,\omega)$ by elementary functions.

**Definition 5 (Elementary function).** [5] A function $\phi \in \mathcal{V}(U,T)$ is called **elementary** if it has the form

$$\phi(t,\omega) = \sum_{i=0}^{n-1} e_i(\omega) \cdot 1_{[t_i,t_{i+1})}(t)$$

for a partition $U = t_0 < t_1 < \ldots < t_n = T$, where each $e_i$ is $\mathcal{F}_{t_i}$-measurable with $\mathbb{E}[e_i^2] < \infty$.

Since $e_i$ is $\mathcal{F}_{t_i}$-measurable, the value of $\phi$ on $[t_i, t_{i+1})$ is determined by the information available at the left endpoint $t_i$; this is precisely the $\mathcal{F}_t$-adaptedness discussed above. One can imagine $\phi$ as a sum of rectangles. We define the integral of an elementary function as

$$\int_U^T \phi(t,\omega) dB_t(\omega) = \sum_{i=0}^{n-1} e_i(\omega)[B_{t_{i+1}} - B_{t_i}](\omega)$$

The following isometry equates the mean square of the stochastic integral with the expected integral of the squared integrand. This is what makes the next definition work: it converts $L^2$-closeness of the integrands into $L^2$-closeness of their integrals, so the limit below exists and does not depend on the approximating sequence.

**Theorem 1 (The Itô isometry).** [5] If $\phi(t,\omega)$ is bounded and elementary then

$$\mathbb{E}\left[\left(\int_U^T \phi(t,\omega) dB_t(\omega)\right)^2\right] = \mathbb{E}\left[\int_U^T \phi(t,\omega)^2 dt\right]$$

Now we are ready to define the Itô integral as a limit of the sum of elementary functions.

**Definition 6 (The Itô integral).** [5] Let $f \in \mathcal{V}(U,T)$. Then the Itô integral of $f$ (from $U$ to $T$) is defined by

$$\int_U^T f(t,\omega) dB_t(\omega) = \lim_{n\to\infty} \int_U^T \phi_n(t,\omega) dB_t(\omega) \qquad \text{limit in } L^2(P)$$

where $\{\phi_n\}$ is a sequence of elementary functions such that

$$\mathbb{E}\left[\int_U^T (f(t,\omega) - \phi_n(t,\omega))^2 dt\right] \to 0 \qquad \text{as } n \to \infty$$

Here $n$ stands for the number of subintervals of the time partition, it says that as we divide the interval into more parts, the estimate of an integral converges to the true integral.

Now we return to $I_T$ [5], for which we choose the elementary functions

$$\phi_n(t,\omega) = \sum_{i=0}^{n-1} B_{t_i} \cdot 1_{[t_i,t_{i+1})}(t), \qquad 0 = t_0 < \ldots < t_n = T$$

Since $\mathbb{E}[(B_s - B_{t_i})^2] = s - t_i$ on each subinterval, the approximation requirement of Definition 6 is satisfied: $\mathbb{E}\left[\int_0^T (B_s - \phi_n(s,\omega))^2 ds\right] = \sum_{i=0}^{n-1} \frac{1}{2}(t_{i+1} - t_i)^2 \to 0$ as the mesh size vanishes. Applying the identity $p(q-p) = \frac{1}{2}(q^2 - p^2) - \frac{1}{2}(q-p)^2$ termwise, we obtain

$$\int_0^T B_s dB_s = \lim_{n\to\infty} \sum_{i=0}^{n-1} B_{t_i}(B_{t_{i+1}} - B_{t_i}) = \frac{1}{2}B_T^2 - \frac{1}{2}T$$

The correction term $-T/2$ has no counterpart in classical calculus, where $\int_0^T f\, df = \frac{1}{2}\big(f(T)^2 - f(0)^2\big)$. It is the price of the non-vanishing quadratic variation of Brownian motion.

**Theorem 2 (The martingale property of the Itô integral).** [5] Let $f \in \mathcal{V}(0,T)$. Then the process

$$\int_0^t f(s,\omega)\, dB_s, \qquad t \in \mathbb{T},$$

is a martingale with respect to $(\mathcal{F}_t)_{t\in\mathbb{T}}$.

Often we are not only interested in a stochastic process, but in a function that depends on a stochastic process. It would be beneficial to be able to differentiate such functions. However, since a standard Wiener process (Brownian motion) is nowhere differentiable, the dependent function is also nowhere differentiable. Instead, we can introduce the differential of the function using Itô's formula. The idea behind Itô's formula is to use a Taylor expansion of the second order for a function $h \in C^2(\mathbb{R})$, because further expansion terms would vanish.

$$dh(x) \approx \frac{\partial h}{\partial x}(x)dx + \frac{1}{2}\frac{\partial^2 h}{\partial x^2}(x)(dx)^2$$

Plugging in $B_t$ instead of $x$ we obtain:

$$dh(B_t) \approx \frac{\partial h}{\partial x}(B_t)dB_t + \frac{1}{2}\frac{\partial^2 h}{\partial x^2}(B_t)(dB_t)^2 = \frac{1}{2}\frac{\partial^2 h}{\partial x^2}(B_t)dt + \frac{\partial h}{\partial x}(B_t)dB_t \quad \text{since } (dB_t)^2 = dt$$

Note that this function does not depend on time; for this reason, in the more general case $(h(t,x))$ a partial derivative with respect to time has to be added.

**Theorem 3 (Itô formula).** [4] Let $(X_t)_{t\in\mathbb{T}}$ be an Itô process with the differential

$$dX_t = \mu(t,X_t)dt + \sigma(t,X_t)dB_t$$

and assume a second process

$$Z_t = h(t,X_t)$$

with $h \in C^{1,2}(\mathbb{T}\times\mathbb{R})$, that is, continuously differentiable in $t$ and twice continuously differentiable in $x$. Then the Itô formula applied results in the differential:

$$dZ_t = \frac{\partial h}{\partial t}(t,X_t)dt + \frac{\partial h}{\partial x}(t,X_t)dX_t + \frac{1}{2}\frac{\partial^2 h}{\partial x^2}(t,X_t)(dX_t)^2$$

This is helpful because it allows for decomposition of the complex function into its deterministic and stochastic components. If the deterministic component vanishes, what remains is a pure Itô integral, so by Theorem 2 the transformed process is a martingale with a known expectation. This is exactly the construction we use later to build a control variate.

# 3 Numerical methods for solving SDEs

While Itô's calculus provides the rigorous theoretical machinery to define SDEs, finding exact, closed-form analytical solutions is rarely possible for complex systems. To extract meaningful insights such as expected values we are forced to rely on numerical approximations.

### 3.1 Euler–Maruyama method

The Euler method was proposed by Leonhard Euler to approximate ordinary differential equations. Later it was extended by Gisiro Maruyama to allow for a stochastic component. The approach is to discretize the time and Brownian increments.

**Definition 7 (Euler–Maruyama method).** [3] Consider a process $X = \{X_t\}_{t\in\mathbb{T}}$ that satisfies the stochastic differential equation:

$$\forall t \in \mathbb{T}: \quad dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t$$

and starts at $X_0$. The functions $\mu(t, x), \sigma(t, x)$ should satisfy the Lipschitz and linear growth conditions. For the uniform grid $t_n = n\,\Delta t$, $n = 0, \ldots, M$, with $\Delta t = T/M$ and $t_M = T$, an Euler approximation is a discrete-time stochastic process $Y = \{Y_n\}_{n=0}^{M}$ satisfying the iterative scheme

$$\forall n \in \{0, \ldots, M-1\}: \quad Y_{n+1} = Y_n + \mu(t_n, Y_n)\,\Delta t + \sigma(t_n, Y_n)\,\Delta B_n,$$

where $\Delta B_n = B_{t_{n+1}} - B_{t_n} \sim \mathcal{N}(0, \Delta t)$, and $Y_n$ approximates $X_{t_n}$. The initial value is $Y_0 = X_0$.

To measure how well the method converges, we need to distinguish between two types of convergence: strong and weak. Strong convergence describes the expected absolute difference between the true and the approximated value at the terminal time $T$: $\mathbb{E}[|X_T - Y_M|]$. The Euler–Maruyama scheme has been shown to converge strongly with order 0.5; however, this is not discussed here further. The reader is referred to [3].

### 3.2 Weak convergence

This type of convergence describes how similar statistical properties of the paths are. It is measured through a test function $g$ applied to the terminal value (e.g. the payoff of an option):

$$e_{\text{weak}} = \mathbb{E}[g(Y_M)] - \mathbb{E}[g(X_T)]$$

**Definition 8 (Weak convergence and its order, uniform grid).** [3] For the uniform grid of Definition 7, $Y_M$ is the approximation at time $T$; a subscript $\Delta$ marks quantities computed under this discretization. The scheme converges weakly to $X$ at time $T$ for a test function $g : \mathbb{R} \to \mathbb{R}$ if

$$\lim_{\Delta t \to 0} \big|\mathbb{E}[g(X_T)] - \mathbb{E}[g(Y_M)]\big| = 0,$$

and with order $\beta > 0$ if there exist constants $c > 0$ and $\Delta t_0 > 0$, both independent of $\Delta t$, such that

$$\big|\mathbb{E}[g(X_T)] - \mathbb{E}[g(Y_M)]\big| \le c\,(\Delta t)^{\beta} \qquad \forall \Delta t \in (0, \Delta t_0).$$

Here $g$ is taken to be $2(\beta + 1)$ times continuously differentiable with polynomially bounded derivatives.

The Euler–Maruyama scheme has been shown to converge weakly with an order of 1 [3]. The order guarantees that the approximation converges. However, $\mathbb{E}[g(Y_M)]$ is not known, since this quantity is an integral over the space of discrete paths and is in general not available in closed form. For this reason, it has to be approximated as well.

## 4 Monte Carlo simulation and error analysis

The Monte Carlo method provides a tool for estimating $\mathbb{E}[g(Y_M)]$ empirically. The idea behind this is to sample the paths and average them.

### 4.1 Estimating statistical properties using Monte Carlo simulation

Our goal is to estimate the expected value of a functional dependent on a discretized process: $\theta_\Delta = \mathbb{E}[g(Y_M)]$. Let $Y^{(i)}$ denote the $i$-th independently simulated trajectory from the previous section, so that $Y_M^{(i)}$ is its value at time $T$. Let $\hat{\theta}_\Delta$ be an empirical mean of the sampled paths. We

rely on the law of large numbers, which tells us that the sample mean converges almost surely to $\theta_\Delta$: $\frac{1}{N}\sum_{i=1}^{N} g(Y_M^{(i)}) \xrightarrow{\text{a.s.}} \theta_\Delta$ as $N \to \infty$.

For this reason we set the estimator for $\theta_\Delta$:

$$\hat{\theta}_\Delta = \frac{1}{N}\sum_{i=1}^{N} g(Y_M^{(i)})$$

Here $N$ denotes the sample size, that is, the number of simulated paths; it is not to be confused with the number of time steps $M$ per path introduced in Definition 7.

## 4.2 Error decomposition: Bias, Variance

When we are estimating a true expected value, we need to quantify the error of our approximation and understand where this error comes from. We define the error as a *mean-squared error* [2]:

$$\text{MSE}(\hat{\theta}_\Delta, \theta_\Delta) = \mathbb{E}[(\hat{\theta}_\Delta - \theta_\Delta)^2] = (\mathbb{E}[\hat{\theta}_\Delta] - \theta_\Delta)^2 + \mathbb{E}[(\hat{\theta}_\Delta - \mathbb{E}[\hat{\theta}_\Delta])^2] = \text{Bias}^2 + \text{Variance}$$

Now we can compute the bias of our estimate:

$$\text{Bias}(\hat{\theta}_\Delta, \theta_\Delta) = \mathbb{E}[\hat{\theta}_\Delta] - \theta_\Delta = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\big[g(Y_M^{(i)})\big] - \mathbb{E}[g(Y_M)] = \mathbb{E}[g(Y_M)] - \mathbb{E}[g(Y_M)] = 0$$

Hence, the estimator $\hat{\theta}_\Delta$ is unbiased. Note that we only considered a bias of a discretized version of $X_T$. Discretization comes with an unavoidable bias of a discrete estimator compared to the true expected value of the functional: $\theta = \mathbb{E}[g(X_T)]$. If we consider the bias of the discrete approximation with respect to the true continuous process $(X_T)$, we observe the following:

$$\text{Bias}(\hat{\theta}_\Delta, \theta) = \mathbb{E}[\hat{\theta}_\Delta] - \theta = \mathbb{E}[g(Y_M)] - \mathbb{E}[g(X_T)]$$

This is precisely the weak error of the Euler–Maruyama scheme; since the scheme is of weak order 1, this bias is $\mathcal{O}(\Delta t)$ and can be reduced by decreasing the step size $\Delta t$. For the rest of the paper we will work on the discretized processes and mainly focus on the second part of the total error, namely the variance. The presence of variance in Monte Carlo simulations is a direct consequence of representing the infinite probabilistic space of an SDE with a finite set of discrete realizations. During the numerical integration, such as with the Euler–Maruyama scheme, each sample path is subjected to a unique sequence of stochastic increments. Consequently, the evaluated function scatters widely across different paths. The Monte Carlo estimator averages these scattered outcomes, but because the sample size $N$ is strictly finite, the resulting sample mean is itself a random variable subject to sampling error.

Since the estimator $\hat{\theta}_\Delta$ depends on random variables $g(Y_M^{(i)})$, $\hat{\theta}_\Delta$ itself is a random variable. For this reason, we can analyze how fast it converges to the true value $\theta_\Delta$. For this we rely on the *Lindeberg–Lévy central limit theorem*, which states that the sequence of independent and identically distributed random variables converges in distribution to the Normal distribution.

**Theorem 4 (Lindeberg–Lévy CLT).** [1] Suppose $\xi_1, \xi_2, \ldots$ is a sequence of i.i.d. random variables with $\mathbb{E}[\xi] = \nu$ and $\text{Var}[\xi] = \sigma_g^2 < \infty$. Then, as $N$ approaches infinity, the random variables $\sqrt{N}(\hat{\xi}_N - \nu)$ converge in distribution to a normal $\mathcal{N}(0, \sigma_g^2)$:

$$\sqrt{N}(\hat{\xi}_N - \nu) \xrightarrow{d} \mathcal{N}(0, \sigma_g^2),$$

where $\hat{\xi}_N = \frac{1}{N}\sum_{i=1}^{N} \xi_i$.

In the context of our simulation, $\xi_i = g(Y_M^{(i)})$, the mean $\nu$ is our target expectation $\theta_\Delta = \mathbb{E}[g(Y_M)]$, and $\hat{\xi}_N$ is $\hat{\theta}_\Delta$.

From that theorem we can conclude that the error is normally distributed:

$$(\hat{\theta}_\Delta - \theta_\Delta) \approx \mathcal{N}\Big(0, \tfrac{\sigma_g^2}{N}\Big) \qquad \text{for large } N, \text{ where } \sigma_g^2 = \text{Var}[g(Y_M)]$$

We come back to the formula of mean-squared error in order to find the convergence rate of the Monte Carlo estimate. We consider root mean squared error (RMSE), taking into account that the estimator $\hat{\theta}_\Delta$ is unbiased with respect to $\theta_\Delta$:

$$\text{RMSE}(\hat{\theta}_\Delta, \theta_\Delta) = \sqrt{\text{Variance}} = \sqrt{\mathbb{E}[(\hat{\theta}_\Delta - \mathbb{E}[\hat{\theta}_\Delta])^2]} = \sqrt{\frac{\sigma_g^2}{N}} = \frac{\sigma_g}{\sqrt{N}} = \mathcal{O}(N^{-1/2})$$

The convergence order means that reducing the error by a factor of 10 requires 100 times more paths. Since the cost per path is fixed, buying accuracy by brute force alone becomes prohibitive very quickly.

# 5 Control Variates

The RMSE derived above is $\sigma_g/\sqrt{N}$, so the sample size $N$ is not the only lever: the error can equally be reduced by shrinking the numerator. The control variate method does this by pairing the simulation output with a second quantity that is correlated with it and whose expectation is known in advance; the known expectation turns that correlation into a correction that cancels part of the noise without shifting the mean.

## 5.1 Definition

**Definition 9 (Control variates).** [2] Let $A_1, \ldots, A_N$ be outputs from $N$ replications of a simulation ($A_i = g(Y_M^{(i)})$). These variables $A_i$ are independent and identically distributed and our objective is $\mathbb{E}[A]$. Then the unbiased estimator is $\hat{A} = \frac{1}{N}\sum_{i=1}^N A_i$. Along with each $A_i$ we calculate another output variable $C_i$, whose expectation $\mathbb{E}[C]$ is known. Then for any fixed $b$ we can calculate

$$A_i(b) = A_i - b(C_i - \mathbb{E}[C])$$

This results in the sample mean (control variate estimator), with $\hat{C} = \frac{1}{N}\sum_{i=1}^N C_i$:

$$\hat{A}(b) = \frac{1}{N}\sum_{i=1}^N (A_i - b(C_i - \mathbb{E}[C])) = \hat{A} - b(\hat{C} - \mathbb{E}[C])$$

Here $\hat{C} - \mathbb{E}[C]$ serves as a control in estimating $\mathbb{E}[A]$.

In fact, we can prove that the new estimator is unbiased.

$$\mathbb{E}[\hat{A}(b)] = \mathbb{E}[\hat{A} - b(\hat{C} - \mathbb{E}[C])] = \mathbb{E}[\hat{A}] - b(\mathbb{E}[\hat{C}] - \mathbb{E}[C]) = \mathbb{E}[A] - b \cdot 0 = \mathbb{E}[A]$$

We can also find the optimal correction coefficient $b$, which results from minimization of the variance of the control variate estimator. Let us denote $\text{Var}[A] = \sigma_A^2$, $\text{Var}[C] = \sigma_C^2$, and let $\rho_{CA}$ be the correlation coefficient between $C$ and $A$.

$$\text{Var}[A_i(b)] = \text{Var}[A_i - b(C_i - \mathbb{E}[C])] = \sigma_A^2 - 2b\sigma_C\sigma_A\rho_{CA} + b^2\sigma_C^2 \equiv \sigma^2(b)$$

$$\text{Var}[\hat{A}(b)] = \frac{\sigma^2(b)}{N} \qquad \text{Var}[\hat{A}] = \frac{\sigma^2(0)}{N} = \frac{\sigma_A^2}{N}$$

The variance of the control variate estimator is smaller than the variance of the standard estimator if $b^2\sigma_C^2 < 2b\sigma_C\sigma_A\rho_{CA}$. The coefficient that minimizes $\sigma^2(b)$ is

$$b^* = \frac{\sigma_A}{\sigma_C}\rho_{CA} = \frac{\text{Cov}[C, A]}{\text{Var}[C]}$$

Now, if we compare the variances of the control variate estimate and the standard estimate, we obtain the following ratio

$$\frac{\text{Var}[\hat{A}(b^*)]}{\text{Var}[\hat{A}]} = 1 - \rho_{CA}^2$$

From this it follows that we need to prefer the controls with the highest absolute correlation. The variance reduction factor $1/(1-\rho_{CA}^2)$ increases sharply as $|\rho_{CA}|$ approaches 1, but drops off as $|\rho_{CA}|$ decreases [2].

One can notice that $b^*$ looks exactly the same as the slope coefficient in the linear regression. And this is precisely it, the following example shows the connection between these two terms.

**Example: (Approximating the value of $\pi$)**

Here we approximate the value of $\pi$ using Monte Carlo sampling. A point drawn uniformly on $[0,1]^2$ falls into the quarter disc with probability $\pi/4$, so we consider the scaled indicator:

$$\psi(x,y) = \begin{cases} 4 & \text{if } x^2+y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $m$ be the number of points we throw randomly in the unit square $[0,1]\times[0,1]$ in one trial. Let us denote the number of trials by $n$ (playing the role of $N$ in Definition 9). We set the estimator to $A_i = \frac{1}{m}\sum_{j=1}^{m}\psi(x_{ij},y_{ij})$. As a control variate $(C_i)$ we choose the average squared Euclidean distance from the point to the origin, that is, how far our points are on average from the origin: $C_i = \frac{1}{m}\sum_{j=1}^{m}[x_{ij}^2+y_{ij}^2]$. The expected value of the control variate is known: $\mathbb{E}[C] = \mathbb{E}[x^2+y^2] = \mathbb{E}[x^2]+\mathbb{E}[y^2] = \frac{1}{3}+\frac{1}{3} = \frac{2}{3}$, where $x, y$ are independent and uniformly distributed on $[0,1]$.

The coefficient $b$ is estimated empirically as $\hat{b} = \frac{\widehat{\mathrm{Cov}}[C,A]}{\widehat{\mathrm{Var}}[C]}$. This is exactly the slope coefficient for the linear regression that we would fit on the data. However, the empirical estimate also introduces a small bias $(\mathcal{O}(1/n))$, as it is derived from a sample. In practice it is usually ignored [2]. Our control variate estimate for a single trial is the following:

$$A_i(\hat{b}) = A_i - \hat{b}(C_i - \mathbb{E}[C]) = A_i - \hat{b}\left(C_i - \frac{2}{3}\right)$$

Then the control variate estimator for $\pi$ over $n$ trials:

$$\hat{A}(\hat{b}) = \hat{A} - \hat{b}\left(\hat{C} - \frac{2}{3}\right)$$

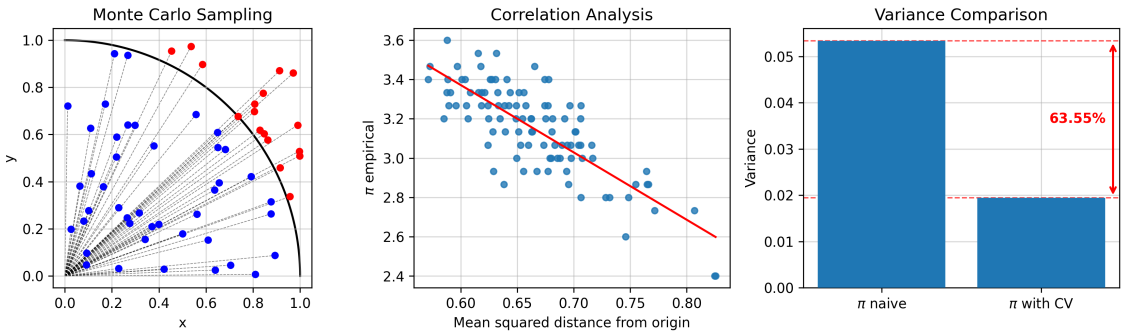


Figure 1: The plot shows Monte Carlo sampling algorithm, scatter plot of control variate and empirical $\pi$ with a linear regression fitted on the data. 63.55% reduction in variance is observed. $m = 100, n = 60$

## 5.2 Martingale as a control variate

For simulations of stochastic differential equations, control variates are also very useful. We want to construct a random variable that has a known expected value and is inexpensive to calculate.

Recall the martingales. By definition, a martingale is a stochastic process whose conditional expectation of future values, given all prior information, is simply its current value. This property makes them exceptionally powerful candidates for control variates. To successfully implement the control variate method as defined in Section 5.1, two conditions must be met: the control variable must be highly correlated with our target simulation, and its exact theoretical expectation must be known in advance.

Martingales satisfy both requirements in the context of SDE simulations. Because they can be constructed using the exact same Brownian increments ($\Delta B_n$) that drive the primary numerical integration scheme (such as the Euler–Maruyama method), they naturally capture the underlying stochastic noise of the system, ensuring a strong correlation. Furthermore, their defining characteristic guarantees that their theoretical expectation is known exactly.

To demonstrate it, let us consider an example of a European call option pricing. A call option is a contract that gives you a right to buy an underlying asset on an expiration date for a fixed predefined price (strike $K$). This is convenient because it allows for avoiding the higher losses if the asset falls and earning if the price grows. Then the payoff of the option on the expiration date $T$ is equal to $\max(S_T - K, 0)$. However, it is often discounted, since there is always a possibility to invest money without the risk. By discounting the payoff by a factor of $e^{-rT}$, where $r$ stands for the risk-free rate, the resulting function is obtained:

$$A(S_T) = e^{-rT} \max(S_T - K, 0)$$

where $S_T$ stands for a price of an underlying asset, which can be modeled by geometric Brownian motion under a risk-neutral pricing framework, with constant volatility $\sigma > 0$:

$$dS_t = rS_t dt + \sigma S_t dB_t$$

Our goal is to find the expected payoff $\theta = \mathbb{E}[A]$. Although there exists a closed form solution, and the paths could be sampled without introduction of the discretization bias, we assume that the solution is not known and solve this by sampling the paths of the underlying asset using the Euler–Maruyama scheme with $\mu(t, x) = rx$ and $\sigma(t, x) = \sigma x$:

$$Y_{n+1} = Y_n + rY_n \Delta t + \sigma Y_n \Delta B_n, \qquad Y_0 = S_0$$

Hence, we approximate $\theta_\Delta = \mathbb{E}[A_\Delta] = \mathbb{E}[A(Y_M)]$ by empirical mean $\hat{\theta}_\Delta = \frac{1}{N} \sum_{i=1}^{N} A_{\Delta,i}$, where $A_{\Delta,i} = e^{-rT} \max(Y_M^{(i)} - K, 0)$. In simulation, the true terminal price $S_T$ is replaced by its Euler–Maruyama approximation $Y_M$; for this reason, we write $A_{\Delta,i}$ for the discounted payoff evaluated on the $i$-th path. We also note that such estimator is unbiased with respect to $\theta_\Delta$, but biased relative to $\theta$. The bias between $\theta_\Delta$ and $\theta$ is governed by the discretization step and is equivalent to the weak error of the Euler–Maruyama scheme.

In order to reduce the variance of the estimator $\hat{\theta}_\Delta$ using the control variate method, a highly correlated random variable with a known expected value is desirable. A martingale can be a good choice because its expected value is known and is equal to its initial value. Let us look at the formula of the discounted asset price.

$$Z_t = h(t, S_t) = e^{-rt} S_t$$

Recall Itô's formula, which allows us to decompose the function into its deterministic and stochastic components. Applying it, we obtain:

$$\begin{aligned} dZ_t &= \frac{\partial h}{\partial t}(t, S_t)dt + \frac{\partial h}{\partial x}(t, S_t)dS_t + \frac{1}{2}\frac{\partial^2 h}{\partial x^2}(t, S_t)(dS_t)^2 \\ &= -re^{-rt} S_t dt + e^{-rt} dS_t \qquad (\text{since } \frac{\partial^2 h}{\partial x^2} = 0) \\ &= -re^{-rt} S_t dt + e^{-rt} \left(rS_t dt + \sigma S_t dB_t\right) = e^{-rt} \sigma S_t dB_t \end{aligned}$$

From the martingale property of the Itô integral (Theorem 2) we know that $Z_t$ is a martingale, since $e^{-rt}\sigma S_t$ is in the class of Itô integrable functions.

From that also follows that the expected value of $Z_T$ is equal to its initial value $Z_0 = e^0 S_0 = S_0$. Since the simulation produces Euler–Maruyama paths rather than the continuous process, we take the discretized martingale as the control: $C_\Delta = e^{-rT} Y_M$. Its expectation is still known in closed form: taking expectations in the scheme and using $\mathbb{E}[\Delta B_n] = 0$ gives $\mathbb{E}[Y_{n+1}] = (1 + r\Delta t)\mathbb{E}[Y_n]$, hence

$$\mathbb{E}[C_\Delta] = S_0 \, e^{-rT} (1 + r\Delta t)^M.$$

The estimator therefore remains unbiased with respect to $\theta_\Delta$ for any fixed $b$; as in Section 5.1, estimating $b$ by $\hat{b} = \frac{\widehat{\mathrm{Cov}}[C_\Delta, A_\Delta]}{\widehat{\mathrm{Var}}[C_\Delta]}$ from the same sample introduces the usual $\mathcal{O}(1/N)$ bias, because $\hat{b}$ and the correction term it multiplies are computed from the same $N$ paths, so their product no longer has zero mean. It is dominated by the standard error $\mathcal{O}(N^{-1/2})$ and is usually ignored [2].

For a single sampled path the following control variate is constructed:

$$A_{\Delta,i}(\hat{b}) = A_{\Delta,i} - \hat{b}(C_{\Delta,i} - \mathbb{E}[C_\Delta]) = A_{\Delta,i} - \hat{b}(e^{-rT}Y_M^{(i)} - \mathbb{E}[C_\Delta])$$

And the goal estimator is equal to:

$$\hat{\theta}_\Delta(\hat{b}) = \hat{\theta}_\Delta - \hat{b}\left(\hat{C}_\Delta - \mathbb{E}[C_\Delta]\right), \quad \text{where } \hat{C}_\Delta = \frac{1}{N}\sum_{i=1}^{N} e^{-rT}\, Y_M^{(i)}$$

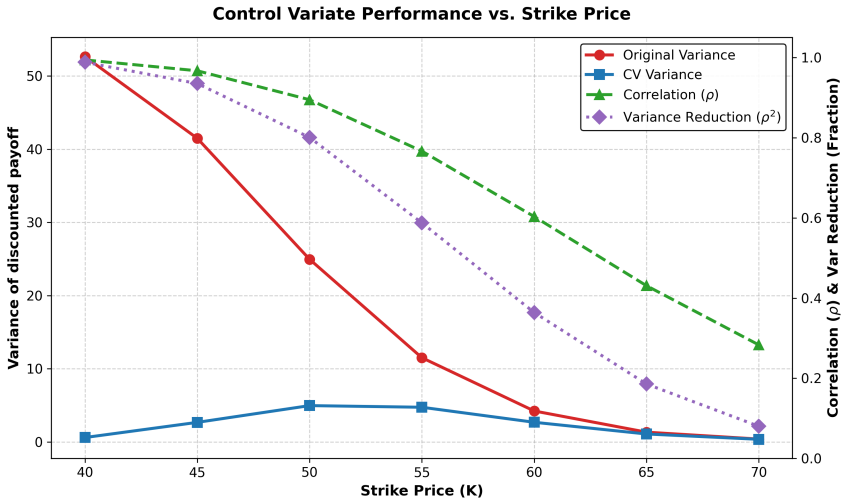


Figure 2: Variance of the discounted payoff, with and without the control variate, as a function of the strike. The variance of the estimator $\hat{\theta}_\Delta$ is obtained by dividing by $N$. Parameters: $S_0 = 50, r = 0.05, T = 0.25, \sigma = 0.3, N = 10\,000, M = 100$

As illustrated above, the martingale provides a significant variance reduction without introducing a discretization-related bias, since its expectation under the discretization is known exactly. From the graph, it follows that the effectiveness of the control variate decreases drastically as the strike price increases. When the strike price is lower than the initial price, the option is most likely to pay out; as the strike price grows, the option is less likely to pay out and the payoff tends to be zero; it leads to decreased correlation and worsens the performance of a control variate.

# 6 Conclusion

We studied the control variate method for the Monte Carlo estimation of expectations of functionals of SDEs. Taking the discounted asset price under the Euler–Maruyama scheme, $e^{-rT}Y_M$, as a control for the discounted call payoff, the variance of the estimator is reduced, at the optimal coefficient, to a fraction $1 - \rho_{CA}^2$ of its original value. Its expectation is known in closed form, so the estimator stays unbiased with respect to the discretized target $\theta_\Delta$. The experiments confirm this reduction and its deterioration as the strike rises: the option becomes less likely to pay out, the payoff decouples from the underlying, and the correlation $\rho_{CA}$ collapses. A natural next step is to replace the hand-crafted control with a learned one: since the optimal coefficient is precisely a regression slope, control variates extend naturally to machine-learning approaches, where a neural network can be trained to produce a highly correlated control of known expectation for payoffs too complex to treat analytically.

# References


[1] Patrick Billingsley. *Probability and Measure*. John Wiley & Sons, New York, 3rd edition, 1995.

[2] Paul Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer New York, NY, 2013.

[3] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Springer Berlin, Heidelberg, 1999.

[4] Matthias Wagner. *On the simulation of stochastic differential equations with applications in mechatronical engineering*. PhD thesis, Johannes Kepler University Linz, 2020.

[5] Bernt Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Heidelberg, New York: Springer-Verlag, 2000.