SC646 – Distributed Optimization and Machine Learning

Option Pricing as a Distributed Stochastic Optimization Problem: Black-Scholes and Monte Carlo Approaches

Submitted by:
Dhrumil Lotiya
21D070026

March 2025

Contents

1 Introduction **1**

2 Stochastic Processes **2**
2.1 Continuous-time stochastic process 2
2.2 Standard Brownian motion 2

3 Itō Calculus **4**
3.1 Itō's Calculus 4
3.2 Properties of Ito Calculus 5
3.3 Change of Measure 6

4 Black-Scholes Formula, Risk-Neutral Valuation **7**
4.1 Interest discounted asset prices as martingales 7
4.2 Black-Scholes(assumption and conclusion) 7
4.3 European call option(Black-Scholes example) 7
4.4 Determining risk-neutral probability 8
4.5 Implied Volatility 8
4.6 Perspective on Acurracy of Black-Scholes 8

1 Introduction

In the field of computational finance and distributed machine learning, efficiently estimating the value of financial derivatives is critical for informed decision-making, portfolio risk management, and algorithmic trading strategies. This project investigates two widely used methodologies—**Monte Carlo simulation** and the **Black-Scholes model**—from the lens of distributed optimization and scalable stochastic computation.

Monte Carlo methods offer a numerical optimization approach that estimates option values by simulating the evolution of the underlying asset across multiple trajectories. These paths are generated using stochastic differential equations (SDEs) under geometric Brownian motion dynamics, and the expected payoff is estimated via sample averaging, making it amenable to parallel and distributed implementations. The estimated payoff is then discounted to present value to approximate the option price. Due to their generality and ability to handle non-convex and path-dependent payoffs, such simulations are well-suited for large-scale optimization in distributed systems.

The underlying asset dynamics in these simulations are modeled using the **Black-Scholes framework**, which assumes geometric Brownian motion with constant drift and volatility. This framework also provides a closed-form analytical solution for European-style options, which we use as a **benchmark** for evaluating the accuracy of distributed Monte Carlo estimates. While the Black-Scholes formula is efficient for basic instruments, it lacks flexibility for real-world constraints, motivating the need for approximate, simulation-based, and parallelizable methods.

The key parameters influencing the system dynamics and optimization objective include:

- **Risk-free interest rate** (r) – acts as a discounting factor in expected reward computation.
- **Asset price** (S_0) – the initial condition for simulating asset dynamics.
- **Strike price** (K) – defines the non-linearity in the payoff function.
- **Time to maturity** (T) – determines the simulation horizon.
- **Volatility** (σ) – introduces stochasticity and uncertainty into the trajectory samples.

By integrating these parameters into a **distributed Monte Carlo pipeline**, and comparing results with the closed-form **Black-Scholes benchmark**, this project demonstrates how financial modeling problems can be framed as large-scale optimization tasks. It also highlights how simulation-driven inference and parallel computation—core ideas in distributed machine learning—can be leveraged for pricing financial derivatives efficiently.

2 Stochastic Processes

2.1 Continuous-time stochastic process

In probability theory, a continuous stochastic process is a type of stochastic process that may be said to be "continuous" as a function of its "time" or index parameter.

2.2 Standard Brownian motion

There exists a probability distribution over the set of continuous functions $B : R \to R$ satisfying the following conditions:

1. $B(0) = 0.$.
2. **(stationary)** for all $0 \leq s < t$, the distribution of $B(t)B(s)$ is the normal distribution with mean 0 and variance ts, and
3. **(independent increment)** the random variables $B(t_i)B(s_i)$ are mutually independent if the intervals $[t_i, s_i]$ are non-overlapping.

We refer to a particular instance of a path chosen according to the Brownian motion as a *sample Brownian path.*

One way to think of standard Brownian motion is as a limit of simple random walks. To make this more precise, consider a simple random walk $Y_0, Y_1, \dots,$ whose increments are of mean 0 and variance 1. Let Z be a piece-wise linear function from [0, 1] to R defined as

$$Z(\frac{t}{n}) = Y_t,$$

for $t = 0, \dots, n$, and is linear at other points. As we take larger values of n, the distribution of the path Z will get closer to that of the standard Brownian motion. Indeed, we can check that the distribution of $Z(1)$ converges to the distribution of $N(0, 1)$, by the central limit theorem. More generally, the distribution of $Z(t)$ converges to $N(0, t)$. Stock prices can also be modeled using standard Brownian motions. Here are some facts about the Brownian motion:

1. Crosses the x-axis infinitely often
2. Has a very close relation with the curve $x = y^2$ (it does not deviate from this curve too much).
3. Is nowhere differentiable. As a result, we cannot use calculus of it. But we do have a 'different calculus' to work with, it's called Ito's calculus.

In real life we can only observe the value of a stochastic process up to some time resolution (in other words, we can only take finitely many sample points). The fact above implies that standard Brownian motion is a reasonable model, at least in this sense, since the real-life observation will converge to the underlying theoretical stochastic process as we take smaller time intervals, as long as the discrete-time observations behave like a simple random walk.

Suppose we use the Brownian motion as a model for daily price of a stock. What is the distribution of the days range?
Answer: the max value and min value over a day

$M(t) = \max_{0 \leq s \leq t} B(s)$, and note that $M(t)$ is well-defined since B is continuous and $[0, t]$ is compact. ($\Psi(t)$ is the cumulative distribution function of the normal random variable) The following holds: $\mathbf{P(M(t) \geq a) = 2p(B(t) > a) = 2 - 2\Psi(\frac{a}{\sqrt{t}})}$. For each $t \geq 0$, the Brownian motion is almost surely not differentiable at t.

Dvoretsky, Erdos, and Kakutani in fact proved a stronger statement asserting that the Brownian motion B is nowhere differentiable with probability 1. Hence a sample Brownian path is continuous but nowhere differentiable! [Quadratic variation] For a partition $\Pi = t_0, t_1, \ldots, t_j$ of an interval $[0, T]$, let $|\Pi| = max_i(t_{i+1}t_i)$. A Brownian motion B_t satisfies the following equation with probability 1:

$$\lim_{|\Pi| \to 0} \sum_i (B_{t_{i+1}} - B_{t_i})^2 = T.$$

Why is this theorem interesting? Suppose that instead of a Brownian motion, we took a function f that is continuously differentiable. Then

$$\sum_i (f(t_{i+1}) - f(t_i))^2 \leq \sum_i (t_{i+1} - t_i)^2 f'(s)^2 \leq \max_{s \in [0,T]} f'(s) \sum_i (t_{i+1} - t_i)^2$$
$$\leq \max_{s \in [0,T]} f'(s)^2 . \max_i \{t_{i+1} - t_i\} T.$$

As $\max\{t_{i+1}t_i\} \to 0$, we see that the above tends to zero. Hence this shows that Brownian motion fluctuates a lot. The above can be summarized by the differential equation $(dB)^2 = dt$.

[Brownian motion with drift] Let $B(t)$ be a Brownian motion, and let μ be a fixed real. The process $X(t) = B(t) + \mu t$ is called a Brownian motion with drift μ. By definition, it follows that $E[X(t)] = \mu t$.

Question : as time passes, which term will dominate? $B(t)$ or μt? It can be shown that μt dominates the behavior of $X(t)$. For example, for all fixed $\varepsilon > 0$, after a long enough time, the Brownian motion will always be between the lines $y = (\mu \varepsilon)t$ and $y = (\mu + \varepsilon)t)$.

When modeling the price of a stock, it is more reasonable to assume that the percentile change follows a normal distribution. This can be written in the following differential equation:

$$dS_t = \sigma S_t dB_t$$

Can we write the distribution of S_t in terms of the distribution of B_t? Is it$S = e^{\sigma B_t}$? Surprisingly, the answer is no. And this is because B_t is not differentiable. So how do we solve this? Here comes Ito's calculus.

3 Itō Calculus

3.1 Itō's Calculus

Here comes the motivation, Suppose we want to compute $f(B_t)$ for some smooth function f i.e. estimating infinitesimal differences. When using Brownian motion as a model, the situation of estimating the difference of a function of the type $f(B_t)$ over an infinitesimal time difference occurs quite frequently. If the differentiation $\frac{dB_t}{dt}$ existed, then we can easily do this by $df = (\frac{dB_t}{dt}.f'(B_t)dt$. We already know that the formula above makes no sense. One possible way to work around this problem is to try to describe the difference df in terms of the difference dB_t. In this case, the equation above becomes $df = f'(B_t).dB_t$.Our new formula at least makes sense, since there is no need to refer to the differentiation $\frac{dB_t}{dt}$ which does not exist. To compute further, we use Taylor series expansion of Δf and the fact that $E[(\Delta B_t)^2] = \Delta t$. Hence the equation above in terms of infinitesimals becomes

$$df(B_t) = f'(B_t).dB_t + \frac{1}{2}f''(B_t)dt.$$

This equation known as the Ito's lemma is the main equation of Ito's calculus.

More generally, consider a smooth function $f(t,x)$ which depends on two variables, and suppose that we are interested in the differential of $f(t,B_t)$. In classical calculus, we will get

$$df = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial x}dx$$

but in Itō Calculus, we will have df(t,B$_t$) $= \frac{f}{\partial t}dt + \frac{\partial f}{\partial x dB_t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(dB_t)^2$ $= (\frac{\partial f}{\partial t}dt + \frac{1}{2}\frac{\partial^2 f}{\partial x^2})dt + \frac{\partial f}{\partial x}dB_t$. [Ito's lemma] Let $f(t,x)$ be a smooth function of two variables, and let X_t be a stochastic process satisfying $dX_t = \mu_t dt + \sigma_t dB_t$ for a Brownian motion $_t$. Then

$$df(t,X_t) = (\frac{\partial f}{\partial t}dt + \mu_t\frac{\partial f}{\partial x} + \frac{1}{2}\sigma_t^2\frac{\partial^2 f}{\partial x^2})dt + \frac{\partial f}{\partial x}dB_t$$

This all started from the quadratic variation.

1. Let $f(t,x) = e^{\mu t + \sigma x}$. Then

$$df(t,B_t) = (\mu + \frac{1}{2}\sigma^2)f(t,B_t)dt + f(t,B_t)dB_t$$

2. We can now answer the question of finding the stochastic process $X_t(t,B_t)$ such that $dX_t = \sigma X_t dB_t$. Suppose we want to model a stock price S_t, and we want the percentile change to behave like a Brownian motion with some variance. Then we can take X_t as S_t. To do this we can just set $\mu = \frac{1}{2}\sigma^2$ in the example above to get,

$$S_t(t,B_t) = e^{\frac{1}{2}\sigma_t^2 + \sigma B_t}$$

We define integration as an inverse of differentiation, i.e.,

$$F(t,B_t) = \int f(t,B_t)dB_t + \int g(t,B_t)dt,$$

if and only if

$$dF = f(t, B_t)dB_t + g(t, B_t)dt$$

. Now the question is, does there exist a Riemannian sum type description? We define Ito's integral such that, it is the limit of Riemannian sums when always a leftmost point of each interval is taken. **NOTE.** Time for some cool stuff, There exists an 'equivalent' of Ito's calculus such that $(dB_t)^2 = -dt$. This changes everything as now we will have,

$$df = f'(B_t)dB_t - \frac{1}{2}f''(B_t)dt.$$

3.2 Properties of Ito Calculus

Let X_t be a stochastic process. A process Δt is called an adapted process (with respect to X_t) if for all $t \geq 0$, the random variable Δ_t depends only on X_s for $s \leq t$. Let $B(t)$ be a Brownian motion, and let $\Delta(t)$ be a nonrandom function of time. Suppose that a stochastic process $I(t)$ satisfies

$$dI = \Delta(s)dB_s \quad i.e. \quad I(t) = \int \Delta(s)dB_s.$$

where $I(0) = 0$. Then for each $t \geq 0$, the random variable $I(t)$ is normally distributed. We generally consider $\Delta(t)$ as adapted process [Ito isometry)] Let $_t$ be a Brownian motion. Then for all adapted processes $\Delta(t)$, we have $\mathrm{E}\left[\left(\int_0^t \Delta(s)dB_s\right)^2\right] = E\left[\int_0^t \Delta(s)^2 ds\right]$ This is basically the quadratic variation, we can see that when we have $\Delta(t) = 1$, which results to $E[B_t^2] = E[t]$. Let B_t be a Brownian motion. Then for all adapted processes $g(t, B_t)$, the integral

$$\int g(t, B_t)dB_s$$

is a martingale, as long as g is a 'reasonable function'. Formally, if $g \in L_2$, i.e.,

$$\int \int_0^t g^2(t, B_t)dtdB_t < \infty.$$

So when we said $X_t = \mu(t, B_t)dt + \sigma(t, B_t)dB_t$, X_t is a martingale iff $\mu = 0$. When we look at a stochastic differential equation, it is a martingale if it doesn't have a drift term. If it has a drift term, it's not a martingale. $\mu(t, B_t)dt$ contributes towards the tendency, the slope of whatever is going to happen in the future. And $\sigma(t, B_t)dB_t$ is like the variance term. It adds some variance to your stochastic process. But still, it doesn't add or subtract value over time, it fairly adds variation. First we have a financial derivative, like the option of a stock. Then we have our portfolio strategy. Assume that we have some strategy that, at the expiration time, gives us the exact value of the option. Now we look at the difference between these two stochastic processes. Basically what the thing is, when your variance goes to 0, our drift also has to go to 0. So when we look at the difference, if we can somehow get rid of this variance term, that means no matter what we do, the drift term will govern the value of our portfolio. If it's positive, that means we can always make money, because there's no variance. Without the variance, we make money. That's called arbitrage, and we cannot have that.

3.3 Change of Measure

Can a stochastic process with drift also be viewed as a process without drift? Recall that a stochastic process is a probability distribution over a set of paths. A change of a measure of a stochastic process is a method of shifting the probability distribution into another probability distribution. We say that two probability distributions $\mathbf{P}$ and $\mathbf{P}$ are equivalent if

$$\mathbf{P}(A) > 0 \iff \mathbf{P}(A) > 0 \quad \forall A$$

When changing measures, we fix a set Σ of possible paths, change a probability distribution $\mathbf{P}$ over Σ, and make it into another equivalent probability distribution over Σ. Thus the underlying space is the same but we only change our point of view. This is not true for general transformations. For example, consider the square of a path. The probability distribution of the square of a Brownian motion $B(t)^2$ is not equivalent to the probability distribution of $B(t)$.

Changing measures is of theoretical importance since it provides a tool to understand the relation between two different but equivalent stochastic processes. It is also of practical importance since converting one probability distribution into another can reveal hidden insights. For example, in finance, we can convert a non-martingale stochastic process into a martingale by changing measures, and this gives a method of pricing financial derivatives.

We now come back to the original question that we posted: 'Can a stochastic process with drift also be viewed as a process without drift?'. We now see that this is the same question as asking whether the two stochastic processes are equivalent. Our first theorem asserts that this indeed is the case for Brownian motions. [Girsanov's theorem, simple version] Let $(\Omega, \mathbf{P})$ be a probability space, and let $X : \Omega \to [0, T]^\infty$ be a stochastic process which is a Brownian motion with drift μ under the probability distribution induced by $(\Omega, \mathbf{P})$. Consider the probability distribution $\mathbf{P}$ over Ω defined as

$$\frac{d\mathbf{P}}{d\mathbf{P}}(\omega) = e^{-\mu\omega(T) - \mu^2 T/2}.$$

Then X is a Brownian motion with no drift under the probability distribution induced by $(\Omega, \mathbf{P})$. We can define a $Z(\omega)$ such that $\mathbf{P}(\omega) = Z(\omega)\mathbf{P}(\omega)$. The function Z is known as the Radon-Nikodym derivative of $\mathbf{P}$ with respect to $\mathbf{P}$, and is denoted as

$$Z(\omega) = \frac{d\mathbf{P}}{d\mathbf{P}}(\omega)$$

We can extend this result to compute expectations, instead of computing $E[X_t]$ in probability space $(\Omega, \mathbf{P})$, we can instead compute the expectation as $E[Z_t X_t]$ in a different probability space i.e. $(\Omega, \mathbf{P})$.

$$E[X_t] = E[Z_t X_t]$$

4 Black-Scholes Formula, Risk-Neutral Valuation

4.1 Interest discounted asset prices as martingales

If r is a risk-free interest rate, then by definition, the price of a contract paying dollar at time T if A occurs is $P_{RN}(A)e^{-rT}$. If A and B are disjoint, what is the price of a contract that pays 2 dollars if A occurs, 3 if B occurs, 0 otherwise?
Answer: $(2P_{RN}(A)+3P_{RN}(B))e^{-rT}$. Generally, in the absence of arbitrage, the price of the contract that pays X at time T should be $E_{RN}(X))e^{-rT}$ T where E_{RN} denotes expectation with respect to the risk neutral probability. If a non-divided paying stock will be worth X at time T, then its price today should be $E_{RN}(X))e^{-rT}$. Risk-neutral probability basically defined so the price of an asset today is e^{-rT} times the risk-neutral expectation of time T price. In particular, the risk-neutral expectation of tomorrow's (interest-discounted) stock price is today's stock price. Implies fundamental theorem of asset pricing, which says discounted price $\frac{X(n)}{A(n)}$ (where A is a risk-free asset) is a martingale with respect to risk-neutral probability.

4.2 Black-Scholes(assumption and conclusion)

Assumption: the log of an asset price X at fixed future time T is a normal random variable (call it N) with some known variance (call it $T\sigma^2$) and some mean (call it μ) with respect to risk neutral probability.

1. N normal $(\mu, T\sigma^2)$ implies $E[e^N] = e^{\mu+T\sigma^2/2}$.
2. If X_0 is the current price then
$$X_0 = E_{RN}[X]e^{-rT} = E_{RN}[e^N]e^{-rT} = e^{\mu+T(\sigma^2/2-r)}$$
3. This implies $\mu = \log X_0 + (r\sigma^2/2)T$.

General Black-Scholes conclusion: If g is any function then the price of a contract that pays $g(X)$ at time T is
$$E_{RN}[g(X)]e^{-rT} = E_{RN}[g(e^N)]e^{-rT}$$
where N is normal with mean μ and variance $T\sigma^2$.

4.3 European call option(Black-Scholes example)

- A **European call option** on stock at **maturity date** T, **strike price** K, gives the holder the right (but not obligation) to purchase a share of stock for K dollars at time T.
  Example, The document gives the bearer the right to purchase one share of MSFT from me on May 31 for 35 dollars. $\mathcal{SS}$. If X is the value of the stock at T, then the value of thE option at time T is given by $g(X) = max0, XK$.
- **Black-Scholes:** price of contract paying $g(X)$ at time T is $E_{RN}[g(X)]e^{-rT} = E_{RN}[g(e^N)]e^{-rT}$ where N is normal with variance $T\sigma^2$, mean $\mu = \log X_0 + (r - \sigma^2/2)T$.

- Wite it as,

$$e^{-rT}E_{RN}[max0, XK] = e^{-rT}E_{RN}[(e^N - K)]_{N\geq \log k} = \frac{e^{-rT}}{\sigma\sqrt{2\pi T}}\int_{\log K}^{\infty} e^{-\frac{(x-\mu)^2}{2T\sigma^2}}(e^x - K)dx$$

- Let T be time to maturity, X_0 current price of underlying asset, K strike price, r risk free interest rate, σ the volatility. We need to compute $e^{-rT}\int_{\log K}^{\infty} e^{-\frac{(x-\mu)^2}{2T\sigma^2}}(e^x - K)dx$ where $\mu = rT + \log X_0 - T\sigma^2/2$.

- Can use complete-the-square tricks to compute the two terms explicitly in terms of standard normal cumulative distribution function ϕ.

- Price of European call is $\phi(d_1)X_0 - \phi(d_2)Ke^(- rT)$ where $d_1 = \frac{\ln\frac{X_0}{K}+(r+\frac{\sigma^2}{2})(T)}{\sigma\sqrt{T}}$ and $d_2 = \frac{\ln\frac{X_0}{K}+(r-\frac{\sigma^2}{2})(T)}{\sigma\sqrt{T}}$

4.4 Determining risk-neutral probability

- If $C(K)$ is price of European call with strike price K and $f = f_X$ is risk neutral probability density function for X at time T, then $C(k) = e^{-rT}\int_{\infty}^{\infty} f(x)max0, x - Kdx$.

- Differentiating under the integral, we find that

$$e^{rT}C'(K) = \int f(x)(-1_{x>K})dx = -P_{RN}X > K = F_X(K) - 1, e^{rT}C''(K) = f(K)$$

.

4.5 Implied Volatility

- Risk-neutral probability densities derived from call quotes are not quite lognormal in practice. Tails are too fat. Main Black-Scholes assumption is only approximately correct.

- "Implied volatility" is the value of σ that (when plugged into the Black-Scholes formula along with known parameters) predicts the current market price.

- If Black-Scholes were completely correct, then given a stock and an expiration date, the implied volatility would be the same for all strike prices. In practice, when the implied volatility is viewed as a function of strike price (sometimes called the "volatility smile"), it is not constant.

4.6 Perspective on Acurracy of Black-Scholes

- **Main Black-Scholes assumption:** risk neutral probability densities are lognormal.

- **Heuristic support for this assumption:** If the price goes up 1 % or down 1 % each day (with no interest) then the risk-neutral probability must be 0.5 for each (independently of previous days). The central limit theorem gives log normality for large T.
- **Replicating portfolio point of view:** in the simple binary tree models (or continuum Brownian models), we can transfer money back and forth between the stock and the risk-free asset to ensure our wealth at time Tequals the option payout. The option price is required initial investment, which is the risk-neutral expectation of payout. "True probabilities" are irrelevant.
- **Where arguments for assumption break down:** Fluctuation sizes vary from day to day. Prices can have big jumps.
- **Fixes:** variable volatility, random interest rates, Levy jumps...