# Week #10: Stochastic Diffusion Processes

Notes by: Francisco Richter

May 15, 2025

## Overview

Stochastic diffusion processes represent a fundamental class of continuous-time stochastic models that describe the random movement of particles, information, prices, or other entities through time and space. Building upon our previous discussions of counting processes and Poisson processes, this week we explore the mathematical foundations and applications of diffusion processes. While counting processes track discrete events occurring at random times, diffusion processes model continuous, often erratic movements governed by stochastic differential equations.

The study of diffusion processes traces back to the pioneering work of Albert Einstein and Marian Smoluchowski in 1905, who provided a theoretical explanation for Brownian motion—the random movement of particles suspended in a fluid. Since then, the mathematical framework has evolved significantly, finding applications across diverse fields including physics, finance, biology, social sciences, and engineering. Modern approaches to diffusion processes incorporate sophisticated tools from stochastic calculus, particularly Itô calculus, which provides the mathematical machinery to handle the continuous but nowhere differentiable paths characteristic of these processes.

In this lecture, we will explore the theoretical foundations of diffusion processes, beginning with Brownian motion as the canonical example, and then extending to more general stochastic differential equations. We will examine the Fokker-Planck equation, which connects the microscopic stochastic description with macroscopic probability distributions. Finally, we will survey recent developments and applications across various domains, highlighting how these mathematical tools continue to advance our understanding of complex systems characterized by randomness and uncertainty.

---

## 1 Brownian Motion: The Foundation of Diffusion Processes

**Definition. Brownian Motion**

A stochastic process $\{W_t : t \geq 0\}$ is called a standard Brownian motion (or Wiener process) if:

(i) $W_0 = 0$ almost surely

(ii) The process has independent increments: for any $0 \leq s < t \leq u < v$, the random variables $W_t - W_s$ and $W_v - W_u$ are independent

(iii) For any $0 \leq s < t$, the increment $W_t - W_s$ follows a normal distribution with mean 0 and variance $t - s$: $W_t - W_s \sim \mathcal{N}(0, t - s)$

(iv) The sample paths $t \mapsto W_t$ are almost surely continuous

**Intuition: What is Brownian motion?**

Mathematically, Brownian motion represents the limit of a random walk as we take smaller and smaller steps. Picture a drunk person taking random steps left or right every second. If we make the steps smaller and more frequent (every millisecond, microsecond, etc.), eventually we get a continuous but extremely jagged path—this is Brownian motion.
The key insight: despite being completely random at each instant, Brownian motion follows statistical patterns. The expected displacement is zero (no preferred direction), but the variance of displacement grows linearly with time—this is why particles diffuse outward from their starting point at a predictable rate.

Brownian motion serves as the building block for more complex diffusion processes. Its mathematical properties make it both theoretically elegant and practically useful. The process was first rigorously constructed by Norbert Wiener in the 1920s, earning it the alternative name "Wiener process" in mathematical literature.

## 1.1 Key Properties of Brownian Motion

### 1.1.1 Martingale Property

Brownian motion is a martingale, meaning that the expected future value given the current state is equal to the current value:

$$\mathbb{E}[W_t|\mathcal{F}_s] = W_s \quad \text{for all } 0 \leq s \leq t$$

where $\mathcal{F}_s$ represents all information available up to time $s$.

**Intuition: Martingale property**

The martingale property means Brownian motion has no "drift" or bias in any direction. If you're currently at position $W_s = 5$, your best prediction for your future position $W_t$ is still 5, regardless of how you got there. This is like a fair game where your expected future wealth equals your current wealth—no systematic advantage or disadvantage.

### 1.1.2 Scaling Property

For any constant $c > 0$, the scaled process $\{c^{-1/2}W_{ct} : t \geq 0\}$ is also a standard Brownian motion. This self-similarity across time scales is a distinctive feature of Brownian motion.

**Example 1.1** (Scaling property in action)**.** *Consider a Brownian motion $W_t$ observed over 1 second, with typical fluctuations of magnitude 1. If we observe the same process over 4 seconds (i.e., $W_{4t}$), the typical fluctuations will be of magnitude 2 (i.e., $\sqrt{4} = 2$ times larger).*

*This is why if we rescale by dividing by $\sqrt{c}$, we get back a standard Brownian motion. For instance, $\frac{W_{4t}}{2}$ has the same statistical properties as $W_t$.*

*This scaling property is crucial in applications like option pricing, where we need to understand how volatility scales with time.*

### 1.1.3 Quadratic Variation

The quadratic variation of Brownian motion over an interval $[0, t]$ is equal to $t$:

$$\lim_{|\Pi|\to 0} \sum_{i=0}^{n-1} (W_{t_{i+1}} - W_{t_i})^2 = t$$

where $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = t\}$ is a partition of $[0, t]$ and $|\Pi|$ denotes the mesh size of the partition.

**Intuition: Quadratic variation**

While the total variation (sum of absolute changes) of Brownian motion is infinite over any interval, the quadratic variation (sum of squared changes) converges to the length of the time interval. This seemingly technical property is fundamental to Itô calculus.
Think of it this way: if you square the changes in Brownian motion and add them up, you get approximately the time elapsed, regardless of the specific path taken. This is why in stochastic calculus, $(dW_t)^2 = dt$, which has no analog in ordinary calculus.

### 1.1.4 Non-differentiability

Despite being continuous, Brownian motion is almost surely nowhere differentiable. This property reflects the highly irregular nature of its paths and necessitates the development of special mathematical tools, such as Itô calculus, to handle stochastic differential equations driven by Brownian motion.

**Example 1.2** (Visualizing non-differentiability)**.** *If you zoom in on any portion of a Brownian motion path, it looks just as jagged as the original path. Unlike smooth curves where zooming in eventually reveals a straight line (which is differentiable), Brownian motion maintains its roughness at all scales.*

*This is why we can't define the "velocity" of a Brownian particle at any instant—the derivative doesn't exist. Instead, we must work with the increments $dW_t$, which represent the infinitesimal changes in the process.*

**Practice Problems**

1. Let $W_t$ be a standard Brownian motion. Calculate $\mathbb{E}[W_5]$ and $\mathrm{Var}(W_5)$.
2. Using the properties of Brownian motion, determine whether the process $X_t = tW_{1/t}$ for $t > 0$ is a Brownian motion.

**Solution:**

For problem 1: Since $W_0 = 0$ and Brownian motion has independent increments with $W_t - W_s \sim \mathcal{N}(0, t - s)$, we have $W_5 \sim \mathcal{N}(0, 5)$. Therefore, $\mathbb{E}[W_5] = 0$ and $\mathrm{Var}(W_5) = 5$.

For problem 2: We need to check the four properties of Brownian motion. Clearly $X_0 = 0 \cdot W_\infty$ is undefined, so $X_t$ is not a Brownian motion. Additionally, using the scaling property, $X_t = tW_{1/t} = t \cdot \sqrt{1/t} \cdot W_1 = \sqrt{t} \cdot W_1$, which does not have independent increments.

---

## 2 Stochastic Differential Equations

Stochastic differential equations (SDEs) provide a framework for modeling systems where randomness plays a crucial role in the dynamics. They extend ordinary differential equations by incorporating a stochastic term, typically driven by Brownian motion.

### 2.1 General Form of SDEs

The general form of a one-dimensional SDE is:

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

where:

- $X_t$ is the stochastic process being modeled
- $\mu(X_t, t)$ is the drift coefficient, representing the deterministic component
- $\sigma(X_t, t)$ is the diffusion coefficient, scaling the random component
- $W_t$ is a standard Brownian motion

**Intuition: Understanding SDEs**

An SDE describes how a variable changes over time when affected by both deterministic and random factors:

- The $\mu(X_t, t)\,dt$ term represents the predictable part—how the variable would change in a small time interval $dt$ if there were no randomness.
- The $\sigma(X_t, t)\,dW_t$ term represents the unpredictable part—random fluctuations whose magnitude is controlled by $\sigma$.

Think of a boat on a river: the drift term $\mu$ is like the current pushing the boat in a predictable direction, while the diffusion term $\sigma$ is like random waves rocking the boat unpredictably. The larger $\sigma$ is relative to $\mu$, the more the random fluctuations dominate the boat's movement.

This equation is typically interpreted in the integral form:

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

where the first integral is a standard Riemann integral, and the second is an Itô stochastic integral.

**Example 2.1** (Step-by-step derivation of an SDE)**.** *Let's derive an SDE for a simple stock price model. Assume that:*

- *The expected return rate is a constant $\mu$ (e.g., 5% per year)*
- *The volatility is a constant $\sigma$ (e.g., 20% per year)*
- *Price changes are proportional to the current price*

*For a small time interval $\Delta t$:*

$$\begin{aligned}\Delta S &= \textit{Expected change} + \textit{Random fluctuation} \\ &= \mu S \Delta t + \sigma S \Delta W\end{aligned}$$

*As $\Delta t \to 0$, this becomes the SDE:*

$$dS_t = \mu S_t\,dt + \sigma S_t\,dW_t$$

*This is the famous Geometric Brownian Motion model used in the Black-Scholes option pricing formula.*

## 2.2 Itô's Lemma

Itô's lemma is the stochastic calculus counterpart of the chain rule in ordinary calculus. For a twice continuously differentiable function $f(x,t)$ and a process $X_t$ satisfying the SDE above, Itô's lemma states:

$$df(X_t,t) = \left(\frac{\partial f}{\partial t} + \mu\frac{\partial f}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2 f}{\partial x^2}\right)dt + \sigma\frac{\partial f}{\partial x}dW_t$$

**Intuition: Why Itô's Lemma has an extra term**

In ordinary calculus, if $y = f(x)$, then $dy = f'(x)dx$. But in stochastic calculus, we get an extra term: $df = f'(X_t)dX_t + \frac{1}{2}f''(X_t)(dX_t)^2$.
Why? Because $(dW_t)^2 = dt$ (from the quadratic variation property), not zero as in ordinary calculus. When we expand $(dX_t)^2 = (\mu dt + \sigma dW_t)^2$, the $(dW_t)^2$ term contributes a non-negligible $\sigma^2 dt$ term. This extra "second derivative" term is crucial—without it, option pricing formulas would be wrong, and many physical models would fail to match reality.

This formula is essential for solving SDEs and deriving important results in financial mathematics, such as the Black-Scholes equation.

**Example 2.2** (Applying Itô's Lemma)**.** *Let's use Itô's Lemma to solve the Geometric Brownian Motion SDE:*

$$dS_t = \mu S_t\,dt + \sigma S_t\,dW_t$$

*We want to find an expression for $S_t$ in terms of $S_0$. Let's try $f(S_t) = \ln(S_t)$. Then:*

$$\frac{\partial f}{\partial S} = \frac{1}{S}$$
$$\frac{\partial^2 f}{\partial S^2} = -\frac{1}{S^2}$$

*By Itô's Lemma:*

$$\begin{aligned} d(\ln S_t) &= \mu\frac{1}{S_t}S_t dt + \sigma\frac{1}{S_t}S_t dW_t + \frac{1}{2}\sigma^2 S_t^2\left(-\frac{1}{S_t^2}\right)dt \\ &= \mu dt + \sigma dW_t - \frac{1}{2}\sigma^2 dt \\ &= \left(\mu - \frac{1}{2}\sigma^2\right)dt + \sigma dW_t \end{aligned}$$

*Integrating from 0 to $t$:*

$$\ln S_t - \ln S_0 = \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t$$
$$\ln S_t = \ln S_0 + \left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t$$

*Taking exponentials:*

$$S_t = S_0\exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma W_t\right)$$

*This is the explicit solution to the GBM equation.*

## 2.3 Common Stochastic Differential Equations

### 2.3.1 Geometric Brownian Motion

Geometric Brownian motion (GBM) is widely used in financial mathematics to model stock prices:

$$dS_t = \mu S_t\, dt + \sigma S_t\, dW_t$$

where $\mu$ represents the expected return rate and $\sigma$ the volatility. The solution to this SDE is:

$$S_t = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t\right)$$

**Intuition: Why GBM for stock prices?**

GBM is popular for modeling stock prices for three key reasons:

- It ensures prices remain positive (since $\exp(x) > 0$ for any $x$)
- Returns are normally distributed, matching empirical observations reasonably well
- The model is simple enough to allow closed-form solutions for many derivatives

The $\frac{\sigma^2}{2}$ adjustment in the drift term isn't arbitrary—it ensures that the expected return is truly $\mu$. Without this adjustment, the expected return would be higher due to the convexity of the exponential function (Jensen's inequality).

### 2.3.2 Ornstein-Uhlenbeck Process

The Ornstein-Uhlenbeck process models mean-reverting behavior:

$$dX_t = \theta(\mu - X_t)\, dt + \sigma\, dW_t$$

where $\theta > 0$ is the rate of mean reversion, $\mu$ is the long-term mean, and $\sigma$ is the volatility parameter. This process is widely used in modeling interest rates, commodity prices, and other phenomena that tend to revert to a long-term average.

**Example 2.3** (Ornstein-Uhlenbeck in action)**.** *Consider an Ornstein-Uhlenbeck process with $\theta = 0.5$, $\mu = 100$, and $\sigma = 10$. If the current value is $X_0 = 80$:*

- *The drift term $\theta(\mu - X_t) = 0.5 \times (100 - 80) = 10$ is positive, pushing the process upward toward the mean.*
- *If instead $X_0 = 120$, the drift would be $0.5 \times (100 - 120) = -10$, pushing the process downward toward the mean.*
- *The strength of the "pull" toward $\mu$ is proportional to the distance from $\mu$.*
- *The parameter $\theta = 0.5$ means that, absent randomness, the process closes about 50% of the gap to the mean each unit of time.*

*This makes the Ornstein-Uhlenbeck process ideal for modeling variables like interest rates that tend to revert to historical averages.*

### 2.3.3 Cox-Ingersoll-Ross Process

The Cox-Ingersoll-Ross (CIR) process is commonly used to model interest rates:

$$dX_t = \kappa(\theta - X_t)\,dt + \sigma\sqrt{X_t}\,dW_t$$

The square root in the diffusion term ensures that the process remains positive, making it suitable for modeling quantities that cannot be negative, such as interest rates or volatility.

**Intuition: State-dependent volatility in CIR**

The key insight of the CIR model is that volatility should depend on the level of the variable. When interest rates (or volatility) are low, the absolute size of fluctuations should also be low.
The $\sqrt{X_t}$ term in the diffusion coefficient accomplishes this: when $X_t$ is small, the random fluctuations are also small. This prevents the process from becoming negative, which would be economically meaningless for interest rates.
Additionally, this specification ensures that as $X_t$ approaches zero, the drift term $\kappa(\theta - X_t)$ becomes strongly positive, pushing the process away from zero.

**Practice Problems**

1. For a Geometric Brownian Motion with $\mu = 0.1$ and $\sigma = 0.3$, calculate the expected value and variance of $S_t$ given $S_0 = 100$.
2. Consider an Ornstein-Uhlenbeck process with $\theta = 0.2$, $\mu = 0$, and $\sigma = 1$. If $X_0 = 5$, what is the expected value of $X_t$ at $t = 10$?
3. For the CIR process, explain why the condition $2\kappa\theta > \sigma^2$ is important for ensuring the process remains positive.

**Solution:**

For problem 1: For GBM, we know that $S_t = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right)$. Since $W_t \sim \mathcal{N}(0, t)$, we have:

$$\begin{aligned}
\mathbb{E}[S_t] &= S_0 \mathbb{E}\left[\exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right)\right] \\
&= S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t\right)\mathbb{E}\left[\exp\left(\sigma W_t\right)\right] \\
&= S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t\right)\exp\left(\frac{\sigma^2 t}{2}\right) \\
&= S_0 \exp(\mu t)
\end{aligned}$$

So $\mathbb{E}[S_t] = 100\exp(0.1t)$. For the variance:

$$\mathrm{Var}(S_t) = S_0^2 \exp(2\mu t)\left(\exp(\sigma^2 t) - 1\right)$$

With the given values: $\mathrm{Var}(S_t) = 100^2 \exp(0.2t)\left(\exp(0.09t) - 1\right)$.
For problem 2: For the Ornstein-Uhlenbeck process, the expected value is:

$$\mathbb{E}[X_t | X_0] = \mu + (X_0 - \mu)e^{-\theta t}$$

With $\mu = 0$, $X_0 = 5$, $\theta = 0.2$, and $t = 10$:

$$\mathbb{E}[X_{10}] = 0 + 5e^{-0.2\times 10} = 5e^{-2} \approx 5 \times 0.135 \approx 0.677$$

For problem 3: In the CIR process, when $X_t$ approaches zero, the drift term becomes approximately $\kappa\theta$ (positive) while the diffusion term approaches zero. However, if $X_t$ is very close to zero, random fluctuations could still potentially push it negative. The condition $2\kappa\theta > \sigma^2$ ensures that the upward drift is strong enough compared to the potential downward random fluctuations, mathematically guaranteeing that the process remains positive.

---

# 3 The Fokker-Planck Equation

The Fokker-Planck equation (also known as the Kolmogorov forward equation) describes the time evolution of the probability density function of a stochastic process governed by an SDE.

## 3.1 Derivation and Interpretation

For a process $X_t$ satisfying the SDE $dX_t = \mu(X_t, t)\,dt + \sigma(X_t, t)\,dW_t$, let $p(x, t)$ denote the probability density function of $X_t$. The Fokker-Planck equation is:

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}[\mu(x,t)p(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma^2(x,t)p(x,t)]$$

**Intuition: The Fokker-Planck equation**

The Fokker-Planck equation describes how the probability distribution of a stochastic process evolves over time. It connects the microscopic description (the SDE) with the macroscopic behavior (the probability distribution).
The equation has two main terms:

- The first term ($-\frac{\partial}{\partial x}[\mu p]$) represents the deterministic drift, which shifts the probability mass in the direction of the drift.
- The second term ($\frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma^2 p]$) represents diffusion, which spreads out the probability mass due to random fluctuations.

Think of it like pouring dye into a flowing river: the current (drift) moves the center of the dye cloud, while diffusion causes the cloud to spread out over time.

**Example 3.1** (Deriving the Fokker-Planck equation)**.** *Let's derive the Fokker-Planck equation for a simple case with constant coefficients: $dX_t = \mu dt + \sigma dW_t$.*

*Consider the expected change in any smooth test function $f(x)$ over a small time interval:*

$$\begin{aligned}\mathbb{E}[f(X_{t+\delta t}) - f(X_t)|X_t = x] &= \mathbb{E}[f'(x)(X_{t+\delta t} - X_t) + \frac{1}{2}f''(x)(X_{t+\delta t} - X_t)^2 + \ldots |X_t = x] \\ &= f'(x)\mu\delta t + \frac{1}{2}f''(x)\sigma^2\delta t + o(\delta t)\end{aligned}$$

*Now, using the definition of the generator $\mathcal{L}$ of the process:*

$$\mathcal{L}f(x) = \mu f'(x) + \frac{1}{2}\sigma^2 f''(x)$$

*The adjoint of this operator acting on the probability density gives us the Fokker-Planck equation:*

$$\frac{\partial p}{\partial t} = \mathcal{L}^* p = -\frac{\partial}{\partial x}[\mu p] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma^2 p]$$

*This is the Fokker-Planck equation for constant drift and diffusion coefficients.*

## 3.2 Stationary Distributions

For time-homogeneous processes (where $\mu$ and $\sigma$ do not explicitly depend on time), a stationary distribution may exist, representing the long-term equilibrium distribution of the process. The stationary distribution $p_s(x)$ satisfies:

$$0 = -\frac{d}{dx}[\mu(x)p_s(x)] + \frac{1}{2}\frac{d^2}{dx^2}[\sigma^2(x)p_s(x)]$$

**Example 3.2** (Finding the stationary distribution)**.** *Let's find the stationary distribution for the Ornstein-Uhlenbeck process:*

$$dX_t = \theta(\mu - X_t)\,dt + \sigma\,dW_t$$

*The stationary distribution satisfies:*

$$\begin{aligned} 0 &= -\frac{d}{dx}[\theta(\mu - x)p_s(x)] + \frac{1}{2}\frac{d^2}{dx^2}[\sigma^2 p_s(x)] \\ &= -\theta(\mu - x)\frac{d}{dx}p_s(x) + \theta p_s(x) + \frac{\sigma^2}{2}\frac{d^2}{dx^2}p_s(x) \end{aligned}$$

*For a stationary Gaussian distribution with mean $\mu$ and variance $\frac{\sigma^2}{2\theta}$:*

$$p_s(x) = \sqrt{\frac{\theta}{\pi\sigma^2}}\exp\left(-\frac{\theta(x-\mu)^2}{\sigma^2}\right)$$

*We can verify this is the solution by substituting back into the equation.*

For the Ornstein-Uhlenbeck process, the stationary distribution is Gaussian:

$$p_s(x) = \sqrt{\frac{\theta}{\pi\sigma^2}}\exp\left(-\frac{\theta(x-\mu)^2}{\sigma^2}\right)$$

For the CIR process, the stationary distribution is a gamma distribution:

$$p_s(x) = \frac{c^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-cx}$$

where $\alpha = \frac{2\kappa\theta}{\sigma^2}$ and $c = \frac{2\kappa}{\sigma^2}$.

**Practice Problems**

1. For a process with $\mu(x) = -x$ and $\sigma(x) = 1$, write down the Fokker-Planck equation and find its stationary distribution.
2. Show that for a process with $\mu(x) = -U'(x)$ and $\sigma(x) = \sqrt{2}$, the stationary distribution is proportional to $e^{-U(x)}$.
3. For the CIR process, verify that the gamma distribution given above is indeed the stationary distribution by substituting it into the stationary Fokker-Planck equation.

**Solution:**

For problem 1: The Fokker-Planck equation is:

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial x}[xp] + \frac{1}{2}\frac{\partial^2 p}{\partial x^2}$$

For the stationary distribution, we set $\frac{\partial p}{\partial t} = 0$:

$$\begin{aligned} 0 &= \frac{\partial}{\partial x}[xp_s] + \frac{1}{2}\frac{\partial^2 p_s}{\partial x^2} \\ \frac{\partial^2 p_s}{\partial x^2} &= -2\frac{\partial}{\partial x}[xp_s] \end{aligned}$$

This is satisfied by a Gaussian distribution:

$$p_s(x) = \frac{1}{\sqrt{\pi}}\exp(-x^2)$$

For problem 2: The Fokker-Planck equation for the stationary distribution is:

$$0 = \frac{\partial}{\partial x}[U'(x)p_s] + \frac{\partial^2 p_s}{\partial x^2}$$

Let's try $p_s(x) \propto e^{-U(x)}$. Then:

$$\begin{aligned} \frac{\partial p_s}{\partial x} &= -U'(x)e^{-U(x)} \\ \frac{\partial^2 p_s}{\partial x^2} &= -U''(x)e^{-U(x)} + (U'(x))^2 e^{-U(x)} \end{aligned}$$

Substituting:

$$\begin{aligned} 0 &= U'(x)e^{-U(x)} + \frac{\partial}{\partial x}[U'(x)e^{-U(x)}] \\ &= U'(x)e^{-U(x)} + U''(x)e^{-U(x)} - (U'(x))^2 e^{-U(x)} \\ &= e^{-U(x)}[U'(x) + U''(x) - (U'(x))^2] \end{aligned}$$

For problem 3: For the CIR process, the stationary Fokker-Planck equation is:

$$0 = -\frac{d}{dx}[\kappa(\theta - x)p_s(x)] + \frac{1}{2}\frac{d^2}{dx^2}[\sigma^2 x p_s(x)]$$

Substituting the gamma distribution and verifying it satisfies this equation is left as an exercise.

# 4 Recent Advances in Diffusion Processes

## 4.1 Fractional Brownian Motion and Long-Range Dependence

Standard Brownian motion assumes independent increments, but many real-world phenomena exhibit long-range dependence. Fractional Brownian motion (fBm) generalizes Brownian motion by introducing a parameter $H \in (0, 1)$, called the Hurst exponent:

$$\mathbb{E}[(B_t^H - B_s^H)^2] = |t - s|^{2H}$$

**Intuition: Fractional Brownian motion**

Fractional Brownian motion captures a crucial feature missing in standard Brownian motion: memory. Depending on the Hurst parameter $H$:

- When $H = 1/2$: We get standard Brownian motion with independent increments.
- When $H > 1/2$: The process exhibits "persistence"—if it was increasing in the past, it's more likely to continue increasing. This creates smoother, more trending paths.
- When $H < 1/2$: The process exhibits "anti-persistence"—if it was increasing, it's more likely to decrease next. This creates rougher, more mean-reverting paths.

This flexibility makes fBm valuable for modeling phenomena from network traffic (often $H > 1/2$) to volatility in financial markets (often $H < 1/2$).

When $H = 1/2$, fBm reduces to standard Brownian motion. For $H > 1/2$, the process exhibits positive autocorrelation (persistence), while for $H < 1/2$, it shows negative autocorrelation (anti-persistence). Fractional Brownian motion has found applications in modeling financial time series, network traffic, and various natural phenomena.

**Example 4.1** (Estimating the Hurst exponent)**.** *To estimate the Hurst exponent from data, we can use the fact that for fBm:*

$$\mathbb{E}[|B_t^H - B_s^H|^2] \propto |t - s|^{2H}$$

*Taking logarithms:*

$$\log(\mathbb{E}[|B_t^H - B_s^H|^2]) \approx \textit{constant} + 2H \log(|t - s|)$$

*So by plotting the log of mean squared increments against the log of time differences, we can estimate $H$ from the slope of the regression line.*

*For example, analyzing high-frequency financial data often yields $H \approx 0.1$ for volatility, indicating strong mean-reversion in volatility processes.*

## 4.2 Jump Diffusion Models

Pure diffusion processes have continuous sample paths, but many real-world processes exhibit sudden jumps. Jump diffusion models incorporate both continuous diffusion and discrete jumps:

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

where $J_t$ is a jump process, typically a compound Poisson process. These models are particularly relevant in finance for capturing market crashes, in seismology for modeling earthquakes, and in neuroscience for describing neuronal firing patterns.

**Intuition: Why we need jumps**

Continuous diffusion models assume that changes happen smoothly, but reality often includes sudden, discontinuous changes:

- In finance: Stock prices can jump on unexpected news (earnings surprises, geopolitical events)
- In insurance: Catastrophic events cause sudden large claims
- In physics: Quantum transitions occur instantaneously
- In biology: Neurons fire in discrete spikes

Jump diffusion models combine the continuous small fluctuations of diffusion with occasional large jumps, providing a more realistic model for many phenomena.

**Example 4.2** (Merton's jump-diffusion model)**.** *In Merton's jump-diffusion model for stock prices, the SDE is:*

$$\frac{dS_t}{S_t} = \mu\,dt + \sigma\,dW_t + (Y-1)\,dN_t$$

*where $N_t$ is a Poisson process with intensity $\lambda$, and $Y$ is a random variable representing the jump size (typically log-normal).*

*If a jump occurs (with probability $\lambda\,dt$ in a small interval), the price jumps from $S_t$ to $YS_t$. The term $(Y-1)$ represents the proportional change in price.*

*This model captures both the continuous day-to-day fluctuations and occasional large moves due to significant news events.*

## 4.3 Rough Volatility Models

Recent empirical studies in finance have suggested that volatility is rougher than previously thought, with a Hurst exponent $H \approx 0.1$ rather than the $H = 1/2$ implied by classical models. Rough volatility models use fractional processes with $H < 1/2$ to capture this roughness:

$$dS_t = \mu S_t\,dt + \sigma_t S_t\,dW_t$$
$$d\sigma_t = \kappa(\theta - \sigma_t)\,dt + \nu\sigma_t\,dB_t^H$$

where $B_t^H$ is a fractional Brownian motion with Hurst exponent $H < 1/2$.

**Intuition: Rough volatility**

Traditional stochastic volatility models assume volatility changes relatively smoothly ($H = 1/2$). However, empirical evidence suggests volatility is actually much rougher ($H \approx 0.1$).
This roughness explains several key features of options markets:

- The steep at-the-money skew in short-dated options
- The term structure of volatility
- The behavior of volatility of volatility

Rough volatility models provide a more parsimonious explanation of these features than traditional models, which often require many more parameters to fit market data.

**Practice Problems**

1. For fractional Brownian motion with Hurst exponent $H$, calculate the correlation between increments $B^H_{t+1} - B^H_t$ and $B^H_t - B^H_{t-1}$.
2. In Merton's jump-diffusion model, if jumps follow a log-normal distribution with parameters $\mu_J$ and $\sigma_J$, derive the expected return of the stock.
3. Explain how rough volatility models can better capture the term structure of implied volatility compared to traditional stochastic volatility models.

**Solution:**

For problem 1: For fractional Brownian motion, the covariance between increments is:

$$\begin{aligned}\text{Cov}(B^H_{t+1} - B^H_t, B^H_t - B^H_{t-1}) &= \frac{1}{2}(|2|^{2H} - 2|1|^{2H} + |0|^{2H}) \\ &= \frac{1}{2}(2^{2H} - 2)\end{aligned}$$

The variance of each increment is $|1|^{2H} = 1$, so the correlation is:

$$\rho = \frac{1}{2}(2^{2H} - 2)$$

For $H = 1/2$, $\rho = 0$ (independent increments). For $H > 1/2$, $\rho > 0$ (positive correlation). For $H < 1/2$, $\rho < 0$ (negative correlation).

For problem 2: In Merton's jump-diffusion model, the expected return must compensate for both diffusion risk and jump risk:

$$\begin{aligned}\mathbb{E}\left[\frac{dS_t}{S_t}\right] &= \mu dt + \mathbb{E}[(Y-1)dN_t] \\ &= \mu dt + \mathbb{E}[Y-1]\lambda dt \\ &= (\mu + \lambda(\mathbb{E}[Y]-1))dt\end{aligned}$$

For log-normal jumps with parameters $\mu_J$ and $\sigma_J$, $\mathbb{E}[Y] = e^{\mu_J + \sigma_J^2/2}$. Therefore:

$$\mathbb{E}\left[\frac{dS_t}{S_t}\right] = (\mu + \lambda(e^{\mu_J + \sigma_J^2/2} - 1))dt$$

For problem 3: Rough volatility models capture the observed fact that volatility exhibits stronger mean-reversion at shorter time scales. This naturally produces a steeper implied volatility skew for short-dated options that flattens for longer maturities, matching market observations. Traditional models with $H = 1/2$ struggle to simultaneously fit both short-term and long-term implied volatility surfaces without introducing additional parameters or time-dependent coefficients.

# 5 Applications of Diffusion Processes

## 5.1 Financial Mathematics

### 5.1.1 Option Pricing

The Black-Scholes model, based on geometric Brownian motion, revolutionized option pricing. The price of a European call option with strike price $K$ and maturity $T$ is given by:

$$C(S,t) = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2)$$

where $\Phi$ is the standard normal cumulative distribution function, and:

$$d_1 = \frac{\ln(S/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}, \quad d_2 = d_1 - \sigma\sqrt{T-t}$$

**Example 5.1** (Black-Scholes option pricing step-by-step)**.** *Let's price a European call option with:*

- *Current stock price* $S = \$100$
- *Strike price* $K = \$105$
- *Time to maturity* $T - t = 0.5$ *years*
- *Risk-free rate* $r = 0.05$ *(5%)*
- *Volatility* $\sigma = 0.2$ *(20%)*

*Step 1: Calculate* $d_1$ *and* $d_2$*:*

$$\begin{aligned} d_1 &= \frac{\ln(100/105) + (0.05 + 0.2^2/2) \times 0.5}{0.2 \times \sqrt{0.5}} \\ &= \frac{-0.0488 + 0.025 + 0.01}{0.2 \times 0.7071} \\ &= \frac{-0.0138}{0.1414} \\ &= -0.0976 \end{aligned}$$

$d_2 = d_1 - 0.2 \times \sqrt{0.5} = -0.0976 - 0.1414 = -0.239$
*Step 2: Calculate* $\Phi(d_1)$ *and* $\Phi(d_2)$*:* $\Phi(-0.0976) \approx 0.4611$ $\Phi(-0.239) \approx 0.4055$
*Step 3: Calculate the option price:*

$$\begin{aligned} C &= 100 \times 0.4611 - 105 \times e^{-0.05\times 0.5} \times 0.4055 \\ &= 46.11 - 105 \times 0.9753 \times 0.4055 \\ &= 46.11 - 41.54 \\ &= \$4.57 \end{aligned}$$

*This is the fair price of the call option according to the Black-Scholes model.*

Recent extensions include stochastic volatility models (e.g., Heston model), jump diffusion models (e.g., Merton model), and rough volatility models.

**Intuition: The genius of Black-Scholes**

The key insight of Black-Scholes is that option pricing doesn't depend on the expected return of the stock. This seems counterintuitive—how can the price of a derivative not depend on the expected return of the underlying?
The answer lies in the concept of risk-neutral pricing: by constructing a continuously rebalanced portfolio of the stock and a risk-free bond, one can perfectly replicate the option payoff. Since this replication is riskless, it must earn the risk-free rate to avoid arbitrage.
This insight transformed finance by providing a way to price complex derivatives without needing to estimate expected returns, which are notoriously difficult to estimate accurately.

### 5.1.2 Interest Rate Modeling

Short-term interest rates are commonly modeled using mean-reverting processes such as the Vasicek model (Ornstein-Uhlenbeck process) or the CIR model. These models form the basis for pricing interest rate derivatives and managing interest rate risk.

### 5.1.3 Portfolio Optimization

Diffusion processes underpin modern portfolio theory, particularly in continuous-time settings. The Merton portfolio problem, which seeks to maximize expected utility of terminal wealth, leads to explicit solutions for optimal investment strategies under various market models.

**Example 5.2** (Merton's portfolio problem)**.** *Consider an investor who can invest in a risk-free asset with return $r$ and a risky asset following GBM with expected return $\mu$ and volatility $\sigma$. The investor wants to maximize the expected utility of terminal wealth $\mathbb{E}[U(W_T)]$ where $U(W) = \frac{W^{1-\gamma}}{1-\gamma}$ (constant relative risk aversion).*

*The optimal proportion $\pi^*$ to invest in the risky asset is:*

$$\pi^* = \frac{\mu - r}{\gamma \sigma^2}$$

*For example, if $\mu = 0.1$, $r = 0.03$, $\sigma = 0.2$, and $\gamma = 2$:*

$$\pi^* = \frac{0.1 - 0.03}{2 \times 0.2^2} = \frac{0.07}{0.08} = 0.875$$

*So the investor should put 87.5% of their wealth in the risky asset and 12.5% in the risk-free asset.*

## 5.2 Physics and Engineering

### 5.2.1 Particle Diffusion

The original application of diffusion processes was in modeling the movement of particles in fluids. The diffusion equation (a special case of the Fokker-Planck equation) describes how the concentration of particles evolves over time:

$$\frac{\partial c}{\partial t} = D \nabla^2 c$$

where $D$ is the diffusion coefficient.

**Intuition: Diffusion in physics**

Diffusion is nature's way of evening things out. If you put a drop of ink in water, it spreads out over time, moving from regions of high concentration to low concentration.
The diffusion equation captures this behavior mathematically. The key insight is that the rate of change of concentration at a point is proportional to the "curvature" of the concentration ($\nabla^2 c$). Where the concentration curve is concave down, concentration decreases; where it's concave up, concentration increases.
This simple principle explains countless phenomena, from heat conduction to the spread of pollutants in the atmosphere.

### 5.2.2 Signal Processing and Filtering

The Kalman filter, which provides optimal estimates of the state of a linear dynamic system with Gaussian noise, is based on diffusion processes. Extensions like the extended Kalman filter and particle filters handle nonlinear systems and non-Gaussian noise.

**Example 5.3** (Kalman filter in action)**.** *Consider tracking a moving object with noisy position measurements. The state evolves according to:*

$$x_k = Fx_{k-1} + w_{k-1}$$

*where $F$ is the state transition matrix and $w_{k-1}$ is process noise.*

*Measurements follow:*

$$z_k = Hx_k + v_k$$

*where $H$ is the measurement matrix and $v_k$ is measurement noise.*

*The Kalman filter recursively updates the state estimate $\hat{x}_k$ and error covariance $P_k$ in two steps:*
*Predict:*

$$\hat{x}_k^- = F\hat{x}_{k-1}$$
$$P_k^- = FP_{k-1}F^T + Q$$

*Update:*

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1}$$
$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-)$$
$$P_k = (I - K_k H)P_k^-$$

*This provides the optimal estimate given the noisy measurements.*

### 5.2.3 Control Theory

Stochastic control theory deals with optimizing systems subject to random disturbances, often modeled as diffusion processes. Applications include robotics, aerospace systems, and manufacturing processes.

## 5.3 Biology and Medicine

### 5.3.1 Population Genetics

The Wright-Fisher diffusion model describes the evolution of gene frequencies in a population under genetic drift:

$$dX_t = \mu(X_t)\,dt + \sqrt{\frac{X_t(1-X_t)}{2N}}\,dW_t$$

where $X_t$ represents the frequency of an allele, $\mu(X_t)$ captures selection effects, and $N$ is the effective population size.

**Intuition: Wright-Fisher diffusion**

The Wright-Fisher diffusion models how the frequency of a gene variant changes in a population due to random sampling in reproduction (genetic drift) and natural selection.
The diffusion term $\sqrt{\frac{X_t(1-X_t)}{2N}}$ captures a key insight: genetic drift is strongest when the allele frequency is around 0.5 (maximum diversity) and weakest when the frequency is near 0 or 1 (near fixation). The factor $\frac{1}{2N}$ shows that genetic drift is stronger in smaller populations.
This model helps explain why small, isolated populations often show reduced genetic diversity—random drift drives alleles to either fixation (frequency 1) or extinction (frequency 0) more quickly in small populations.

### 5.3.2 Neuroscience

The Ornstein-Uhlenbeck process is used in the leaky integrate-and-fire model of neuronal dynamics:

$$dV_t = -\frac{1}{\tau}(V_t - V_{\text{rest}})\,dt + \sigma\,dW_t$$

where $V_t$ is the membrane potential, $\tau$ is the membrane time constant, and $V_{\text{rest}}$ is the resting potential.

**Example 5.4** (Neuronal firing)**.** *In the leaky integrate-and-fire model, a neuron's membrane potential $V_t$ evolves according to an Ornstein-Uhlenbeck process until it reaches a threshold $V_{th}$, at which point the neuron "fires" and the potential resets to $V_{reset}$.*

*For a neuron with parameters $\tau = 20$ $ms$, $V_{rest} = -70$ $mV$, $V_{th} = -55$ $mV$, $V_{reset} = -75$ $mV$, and $\sigma = 5$ $mV/\sqrt{ms}$:*

- *Without input current, the membrane potential fluctuates around $V_{rest}$ due to noise, occasionally crossing threshold and generating spontaneous spikes.*
- *With a constant input current $I$, the effective resting potential shifts to $V_{rest} + I\tau$, increasing the firing rate.*
- *The mean time to first spike (from reset to threshold) can be calculated using the theory of first passage times for diffusion processes.*

*This simple model captures many features of real neuronal behavior while remaining mathematically tractable.*

### 5.3.3 Reaction-Diffusion Systems

Reaction-diffusion systems combine chemical reactions with diffusion to model pattern formation in biological systems:

$$\frac{\partial u_i}{\partial t} = D_i \nabla^2 u_i + f_i(u_1, \ldots, u_n)$$

where $u_i$ represents the concentration of the $i$-th species, $D_i$ is its diffusion coefficient, and $f_i$ captures the reaction kinetics. Stochastic versions incorporate randomness in both reaction and diffusion processes.

**Intuition: Pattern formation**

Reaction-diffusion systems explain how complex patterns can emerge from simple chemical interactions. The key insight, first proposed by Alan Turing, is that diffusion—which normally smooths out differences—can actually create patterns when multiple chemicals interact.
For example, if an activator chemical stimulates its own production but diffuses slowly, while an inhibitor suppresses the activator but diffuses quickly, stable patterns can form. This mechanism explains diverse biological patterns:

- Animal coat patterns (spots, stripes)
- Fingerprint formation
- Arrangement of leaves on a stem
- Patterns of ecological habitation

The stochastic version adds randomness to capture the inherent variability in biological systems, explaining why no two zebras have identical stripes.

## 5.4 Social Sciences and Network Analysis

### 5.4.1 Information Diffusion in Social Networks

Recent research has developed stochastic diffusion models for influence maximization in social networks. These models treat influence probabilities as random variables with unknown distributions, reflecting the uncertain and time-varying nature of social interactions. Learning automata-based algorithms can estimate these probabilities by sampling from the network, enabling more effective identification of influential nodes.

**Example 5.5** (Stochastic influence maximization)**.** *In a social network with $n$ users, the traditional Independent Cascade model assigns fixed influence probabilities $p_{ij}$ to each edge $(i, j)$. When user $i$ is activated, they have one chance to activate each neighbor $j$ with probability $p_{ij}$.*

*The stochastic diffusion model extends this by treating each $p_{ij}$ as a random variable with unknown distribution. For example, $p_{ij}$ might follow a Beta distribution whose parameters are updated based on observed interactions.*

*To find influential users, we can use a learning automata approach:*

1. *Initialize estimates of influence probabilities*
2. *Select a set of seed nodes based on current estimates*
3. *Observe the resulting cascade*
4. *Update probability estimates based on observations*
5. *Repeat until convergence*

*This approach outperforms traditional methods on real-world networks where influence probabilities vary over time.*

### 5.4.2 Opinion Dynamics

Continuous-time models of opinion formation often use diffusion processes to capture how individuals' opinions evolve through social interaction:

$$dX_i(t) = \sum_{j=1}^{n} a_{ij}(X_j(t) - X_i(t))\,dt + \sigma_i\,dW_i(t)$$

where $X_i(t)$ represents the opinion of individual $i$, $a_{ij}$ is the influence weight of individual $j$ on individual $i$, and the stochastic term captures random fluctuations in opinions.

**Intuition: Opinion dynamics**

Opinion dynamics models capture how people's views evolve through social interaction. The deterministic part of the equation represents social influence—people tend to adjust their opinions toward those of their connections, weighted by how much they trust or are influenced by each person.
The stochastic term captures the inherent randomness in opinion formation—exposure to media, random events, or mood fluctuations that cause opinions to change unpredictably.
These models help explain phenomena like:

- Opinion polarization in social networks
- The emergence of consensus in some groups but not others
- How the structure of social networks affects the spread of information and misinformation

### 5.4.3 Urban Growth and Spatial Economics

Diffusion processes model the spread of urban development and economic activity across space:

$$\frac{\partial p}{\partial t} = \nabla \cdot (D(x)\nabla p) + g(p, x, t)$$

where $p$ represents population density, $D(x)$ is a spatially varying diffusion coefficient, and $g$ captures growth factors.

**Example 5.6** (Urban growth modeling)**.** *Consider a simplified model of urban growth where population density $p(x, y, t)$ evolves according to:*

$$\frac{\partial p}{\partial t} = D\nabla^2 p + rp\left(1 - \frac{p}{K(x, y)}\right)$$

*This combines:*

- *Diffusion ($D\nabla^2 p$): People move from high-density to low-density areas*
- *Logistic growth ($rp(1 - p/K)$): Population grows until reaching the local carrying capacity $K(x, y)$*

*The carrying capacity $K(x, y)$ can incorporate geographical features (rivers, mountains), infrastructure (roads, utilities), and economic factors (job opportunities).*

*Simulating this model can reproduce observed patterns of urban sprawl, the formation of suburbs, and the impact of transportation networks on development patterns.*

**Practice Problems**

1. In the Black-Scholes model, calculate the delta ($\Delta = \frac{\partial C}{\partial S}$) of a European call option and explain its significance for hedging.
2. For the Wright-Fisher diffusion model, calculate the expected time for a neutral allele (no selection) with initial frequency $p$ to reach fixation or extinction.
3. In a social network with 5 users, if influence probabilities follow Beta distributions with parameters $\alpha_{ij}$ and $\beta_{ij}$, describe how you would update these parameters after observing that user 1 successfully influenced user 2 but failed to influence user 3.

**Solution:**

For problem 1: The delta of a European call option in the Black-Scholes model is:

$$\Delta = \frac{\partial C}{\partial S} = \Phi(d_1)$$

where $\Phi$ is the standard normal CDF and $d_1$ is as defined in the Black-Scholes formula.

Delta represents the rate of change of the option price with respect to changes in the underlying stock price. It's significant for hedging because holding $\Delta$ shares of the stock for each option sold creates a locally risk-neutral portfolio. For example, if $\Delta = 0.7$, holding 70 shares of stock hedges 100 call options against small price movements.

For problem 2: For the Wright-Fisher diffusion with no selection, the expected time to fixation or extinction for an allele with initial frequency $p$ is:

$$\mathbb{E}[T] = -2N[p\ln(p) + (1-p)\ln(1-p)]$$

where $N$ is the effective population size. This formula shows that alleles starting at frequency 0.5 take the longest to reach fixation or extinction, while those starting near 0 or 1 resolve quickly.

For problem 3: In a Bayesian framework with Beta priors, after observing that user 1 successfully influenced user 2 but failed to influence user 3:

$$\begin{aligned}
&\alpha_{12} \to \alpha_{12} + 1 \quad \text{(add one success)} \\
&\beta_{12} \to \beta_{12} \quad \text{(no change in failures)} \\
&\alpha_{13} \to \alpha_{13} \quad \text{(no change in successes)} \\
&\beta_{13} \to \beta_{13} + 1 \quad \text{(add one failure)}
\end{aligned}$$

The updated Beta distributions reflect our revised beliefs about the influence probabilities based on the observed interactions.

---

# 6 Computational Methods for Diffusion Processes

## 6.1 Numerical Solutions of SDEs

### 6.1.1 Euler–Maruyama Method

The simplest explicit time-stepping scheme for an SDE

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

is the Euler–Maruyama update

$$X_{n+1} = X_n + \mu(X_n, t_n)\,\Delta t \;+\; \sigma(X_n, t_n)\,\Delta W_n,$$

where $\Delta W_n \sim \mathcal{N}(0, \Delta t)$.

Euler–Maruyama Simulation [1] Initial value $X_0$, drift $\mu(\cdot,\cdot)$, diffusion $\sigma(\cdot,\cdot)$, time step $\Delta t$, number of steps $N$ Approximate path $\{X_n\}_{n=0}^{N}$ $X \leftarrow X_0$ $n = 0, \ldots, N-1$ Draw $\Delta W \sim \mathcal{N}(0, \Delta t)$ $X \leftarrow X + \mu(X, t_n)\,\Delta t \;+\; \sigma(X, t_n)\,\Delta W$ Record $X_{n+1} \leftarrow X$ $\{X_n\}$

**Example 6.1** (Ornstein–Uhlenbeck Process)**.** *Simulate*

$$dX_t = \theta(\mu - X_t)\,dt \;+\; \sigma\,dW_t$$

*with parameters $\theta = 0.5$, $\mu = 1.0$, $\sigma = 0.2$, $X_0 = 0$ over $[0, 10]$. Use the Euler–Maruyama algorithm above with $N$ steps to observe fluctuations around the mean $\mu = 1.0$.*

### 6.1.2 Milstein Method

To achieve strong order 1, the Milstein scheme adds the Itô correction:

$$X_{n+1} = X_n + \mu(X_n)\,\Delta t + \sigma(X_n)\,\Delta W_n + \tfrac{1}{2}\,\sigma(X_n)\,\sigma'(X_n)\big[(\Delta W_n)^2 - \Delta t\big],$$

where $\sigma' = \partial\sigma/\partial X$.

Milstein Simulation [1] $X_0$, $\mu(\cdot)$, $\sigma(\cdot)$, $\sigma'(\cdot)$, $\Delta t$, $N$ $\{X_n\}_{n=0}^{N}$ $X \leftarrow X_0$ $n = 0, \ldots, N-1$ Draw $\Delta W \sim \mathcal{N}(0, \Delta t)$ $X \leftarrow X + \mu(X)\,\Delta t + \sigma(X)\,\Delta W + \tfrac{1}{2}\,\sigma(X)\,\sigma'(X)\big[(\Delta W)^2 - \Delta t\big]$ Record $X_{n+1} \leftarrow X$ $\{X_n\}$

**Intuition: Why Milstein is more accurate**

Euler–Maruyama is a first-order Itô–Taylor truncation. Milstein includes the quadratic variation correction $[(\Delta W)^2 - \Delta t]$, crucial when $\sigma(X)$ varies with $X$, reducing bias in the diffusion term.

### 6.1.3 Implicit Methods

For stiff or mean-reverting SDEs, one may use an implicit drift:

$$X_{n+1} = X_n + \mu(X_{n+1}, t_{n+1})\,\Delta t + \sigma(X_n, t_n)\,\Delta W_n,$$

solving a (small) nonlinear equation for $X_{n+1}$ at each step to enhance stability.

## 6.2 Monte Carlo Simulation

Monte Carlo methods approximate expectations by averaging many sample paths. For example, to price a European call with payoff $\max(S_T - K, 0)$ under risk-neutral drift $r$, we simulate

$$dS_t = r\,S_t\,dt + \sigma\,S_t\,dW_t, \quad S_0 = \text{given},$$

then estimate

$$C_0 = e^{-rT}\,\frac{1}{M}\sum_{m=1}^{M} \max(S_T^{(m)} - K, 0).$$

Monte Carlo Option Pricing [1] $S_0, K, r, \sigma, T, M$ (paths), $N$ (steps per path) Estimate of call price $C_0$ sum $\leftarrow 0$ $m = 1, \ldots, M$ $S \leftarrow S_0$ $n = 1, \ldots, N$ Draw $\Delta W \sim \mathcal{N}(0, \frac{T}{N})$ $S \leftarrow S + rS\,\frac{T}{N} + \sigma S\,\Delta W$ sum $\leftarrow$ sum $+ \max(S - K, 0)$ $C_0 \leftarrow e^{-rT}\,\text{sum}/M$ $C_0$

**Example 6.2** (European Call Pricing)**.** *Using the algorithm above, one samples $M$ terminal prices $S_T^{(m)}$, computes payoffs, discounts, and averages to obtain $C_0 = e^{-rT}\,\mathbb{E}[\max(S_T - K, 0)]$ under the risk-neutral measure.*

## 6.3 Machine Learning Approaches

### 6.3.1 Neural Networks for SDEs

Physics-Informed Neural Networks (PINNs) and Deep BSDE solvers approximate SDE solutions or Fokker–Planck densities, overcoming the curse of dimensionality by learning high-dimensional mappings from $(x, t)$ to $X_t$ or $p(x, t)$.

**Intuition: Neural SDEs**

Traditional grids explode in high dimensions. Neural networks, trained to satisfy the differential equations and boundary conditions, can generalize across the domain and handle many state variables efficiently.

### 6.3.2 Graph Neural Networks for Reaction–Diffusion

Encoding spatial cells as nodes and diffusion pathways as edges, GNNs propagate concentrations and apply reaction updates via learnable message-passing layers, achieving fast, accurate predictions for complex spatial networks.

**Example 6.3** (GNNs for Reaction–Diffusion)**.**

1. *Model each cell as a graph node with state features (protein concentrations).*
2. *Use graph convolutions to diffuse information along edges.*
3. *Apply node-wise MLPs to model local reactions.*
4. *Train on simulation or experimental data to predict spatiotemporal dynamics.*

**Practice Problems**

1. Implement the Euler–Maruyama and Milstein algorithms above to simulate geometric Brownian motion with $\mu = 0.1$, $\sigma = 0.3$, $S_0 = 100$ over one year.
2. Compare local error of Euler–Maruyama vs. Milstein on $dX_t = X_t\,dt + X_t^2\,dW_t$ with $X_0 = 1$ over $[0, 1]$.
3. Outline a PINN setup to solve the Fokker–Planck equation for an Ornstein–Uhlenbeck process.

**Solution:**

For problem 1, apply the two algorithms in their pseudocode form, noting that Milstein's extra term $\frac{1}{2}\,\sigma\sigma'[(\Delta W)^2 - \Delta t]$ uses the known derivative of $\sigma(X) = X$ or $X^2$ as required.

For problem 2, the strong order of Euler–Maruyama is 0.5, while Milstein attains order 1, so Milstein's local errors shrink approximately linearly in $\Delta t$.

For problem 3, define a network $p_\theta(x, t)$, impose the PDE residual

$$\partial_t p + \partial_x[\theta(\mu - x)p] - \tfrac{\sigma^2}{2}\partial_{xx} p = 0$$

as a loss term, plus initial and normalization constraints, and train $\theta$ via stochastic gradient descent.

## 7 Some comments

Stochastic diffusion processes provide a powerful framework for modeling systems with randomness across various disciplines. From the foundational Brownian motion to sophisticated jump diffusion and rough volatility models, the field continues to evolve to address new challenges and applications.

As computational power increases and interdisciplinary collaboration grows, we can expect stochastic diffusion processes to play an increasingly important role in understanding and predicting complex systems characterized by randomness and uncertainty.

**Intuition: The power of stochastic modeling**

The true power of stochastic diffusion models lies in their ability to capture the essential features of complex systems while remaining mathematically tractable. By embracing randomness rather than trying to eliminate it, these models often provide more realistic and useful descriptions of the world. As the mathematician Henri Poincaré noted: "It is far better to foresee even without certainty than not to foresee at all." Stochastic models allow us to make probabilistic forecasts in situations where deterministic predictions would fail completely.
The future of the field lies in combining the mathematical rigor of stochastic analysis with the flexibility of machine learning and the computational power of modern computing—creating models that are both theoretically sound and practically useful across an ever-widening range of applications.

---

## 8 References

1. Einstein, A. (1905). "On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat." Annalen der Physik, 17, 549-560.

2. Øksendal, B. (2003). Stochastic Differential Equations: An Introduction with Applications. Springer.

3. Black, F., & Scholes, M. (1973). "The Pricing of Options and Corporate Liabilities." Journal of Political Economy, 81(3), 637-654.

4. Heston, S. L. (1993). "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options." The Review of Financial Studies, 6(2), 327-343.

5. Gatheral, J., Jaisson, T., & Rosenbaum, M. (2018). "Volatility is rough." Quantitative Finance, 18(6), 933-949.

6. Rezvanian, A., Vahidipour, S. M., & Meybodi, M. R. (2023). "A new stochastic diffusion model for influence maximization in social networks." Scientific Reports, 13, 6122.

7. Cao, Z., Chen, R., Xu, L., Zhou, X., Fu, X., Zhong, W., & Grima, R. (2024). "Efficient and scalable prediction of stochastic reaction-diffusion processes using graph neural networks." Mathematical Biosciences, 375, 109248.

8. Merton, R. C. (1976). "Option pricing when underlying stock returns are discontinuous." Journal of Financial Economics, 3(1-2), 125-144.

9. Karatzas, I., & Shreve, S. E. (1998). Methods of Mathematical Finance. Springer.

10. Gardiner, C. W. (2009). Stochastic Methods: A Handbook for the Natural and Social Sciences. Springer.

## Exercises

**Exercise 1:** Prove that for any constant $c > 0$, the rescaled process

$$\widetilde{W}_t = c^{-\frac{1}{2}} W_{ct}, \quad t \geq 0,$$

of a standard Brownian motion $(W_t)_{t\geq 0}$ is itself a standard Brownian motion.

**Exercise 2: Non-Differentiability.** Explain why almost all sample paths of Brownian motion are nowhere differentiable. In particular, argue that zooming in on any portion of a Brownian path preserves its jagged appearance, preventing the existence of a well-defined derivative at any point.

**Exercise 3: Derivation of Geometric Brownian Motion.** Starting from the discrete-time approximation

$$\Delta S = \mu\, S\, \Delta t + \sigma\, S\, \Delta W,$$

where $\Delta W \sim \mathcal{N}(0, \Delta t)$, derive the continuous-time SDE

$$dS_t = \mu\, S_t\, dt + \sigma\, S_t\, dW_t.$$

**Exercise 4: Applying Itô's Lemma.** Let $(S_t)$ satisfy

$$dS_t = \mu\, S_t\, dt + \sigma\, S_t\, dW_t,$$

and set $Y_t = \ln S_t$. Use Itô's lemma to show that

$$dY_t = \left(\mu - \tfrac{1}{2}\sigma^2\right) dt + \sigma\, dW_t,$$

and hence derive the explicit solution

$$S_t = S_0 \exp\Big((\mu - \tfrac{1}{2}\sigma^2)t + \sigma W_t\Big).$$

**Exercise 5: Ornstein–Uhlenbeck in Action.** Consider the Ornstein–Uhlenbeck process

$$dX_t = \theta(\mu - X_t)\, dt + \sigma\, dW_t,$$

with $\theta = 0.5$, $\mu = 100$, $\sigma = 10$.

(a) If $X_0 = 80$, compute the drift term $\theta(\mu - X_0)$ and describe its effect.

(b) If $X_0 = 120$, compute the drift term and explain how it pushes $X_t$ toward $\mu$.