UniTs - University of Trieste

Faculty of Scientific and Data Intensive Computing

Department of mathematics informatics and geosciences

Stocasthic Modeling and Simulation

Lecturer:
Prof. Alberto D'Onofrio

Author:
Andrea Spinelli

July 18, 2025

github.com/Spina02 andreaspinelli2002@gmail.com

Preface

As a student of Scientific and Data Intensive Computing, I've created these notes while attending the **Stochastic Modeling and Simulation** course.
The course covers a wide range of topics, including Stochastic Nonlinear Models in different application fields (as physics, biomedicine, mathematics, ...), Stochastic Differential Equations, and Stochastic Simulation.

The topics covered in these notes include:

- Recap of Deterministic Models
- Stochastic Differential Equations and White Noise
- Fokker Planck Equation
- Noise-induced Transitions
- Colored noises
- Bounded Stochastic Processes
- Spatio-temporal Stochastic Processes
- Parameter Estimation form Data
- Stochasticity …
- …
- Continuous state space-discrete time Stochastic Processes
- Discrete Time Markov Chains
- Continuous Time Markov Chains
- Mean Field Approximation

While these notes were primarily created for my personal study, they may serve as a valuable resource for fellow students and professionals interested in Stochastic Modeling and Simulation.

Contents

1 Introduction

A diverse array of fields, ranging from epidemiology and systems biology to finance and renewable energy, are fundamentally concerned with understanding systems that evolve over time. While simple, deterministic models can offer initial insights, many real-world phenomena are characterized by inherent randomness, uncertainty, and intricate interconnections. To capture this richness, we turn to the powerful framework of **Stochastic Modeling and Simulation**. Many systems we encounter are ***complex***: disease spread modeled by the classic SIR (Susceptible, Infected, Recovered) framework, oscillating chemical reaction concentrations, or predator-prey ecosystem dynamics. In each case, the system's behavior emerges from interactions between countless individual components, that cannot be predicted from studying the parts in isolation.

Definition: ***Complex System***

A **complex system** is a system composed of interconnected parts that, as a whole, exhibit one or more properties which are not obvious from the properties of individual parts alone.

A key feature of complex systems is **emergent behavior**, where macroscopic patterns arise from the local interactions of microscopic agents. This is closely related to the concepts of adaptation and self-organization. **Adaptation** refers to the process by which a system achieves a better fit with its environment. **Self-organization** is the spontaneous process through which a system's internal structure changes, often to facilitate this adaptation, without external control. A classic example is the formation of a flock of birds: each bird follows simple rules based on its neighbors, yet the flock as a whole displays coordinated, complex flight patterns that no single bird orchestrates.

Observation: ***The Constructive Role of Noise***

When studying complex systems, we frequently discover that randomness, or *noise*, serves a purpose beyond merely interfering with deterministic signals. Surprisingly, the presence of noise combined with nonlinear dynamics can actually promote the formation of ordered structures and novel, persistent patterns.

1.1 The Modeling Framework

To build quantitative models of complex systems, we require a clear mathematical framework. This involves defining several key ingredients that form the foundation of any dynamical model:

- **Entities:** The fundamental components of the system. These can be modeled as *discrete* objects (e.g. individuals in a population) or as *continuous* quantities (e.g. the density of a population).
- **State:** A complete description of the system at a specific moment. The *state space* is the set of all possible states the system can occupy.
- **Time:** The independent variable against which the system's state evolves. Time can be treated as *discrete* (advancing in steps) or *continuous*.
- **Evolution Rules:** The laws or functions that dictate how the system's state changes from one moment to the next. These rules are often expressed using mathematical equations.

The Role of Differential Equations

For systems where the state and time are treated as continuous, the evolution rules are most naturally expressed using differential equations.

Definition: *Differential Equation*

A **differential equation** is a mathematical equation that relates one or more unknown functions to their derivatives. In modeling, it describes the instantaneous rate of change of a system's state:

$$\frac{dx}{dt} = f(x,t)$$

where x is the state variable and f is a function that describes how x changes over time.

In essence, differential equations are the language we use to describe a changing world. Perhaps the most famous and foundational example is Newton's Second Law of Motion, which relates the force on an object to its acceleration (the second derivative of its position).

$$F = ma \quad \Longrightarrow \quad m\frac{d^2x}{dt^2} = F(x(t),\dot{x}(t))$$

This single equation forms the basis of classical mechanics and is a prime example of a **dynamical system**: a system whose state evolves over time according to a deterministic rule.

1.2 Bridging Models and Data in the Modern Era

In our era of big data and machine learning, an important question emerges: will data-driven methods make traditional mathematical modeling redundant? The answer is both more complex and more promising: the path forward involves **hybrid approaches** that harness the complementary strengths of both methodologies. Although real-world datasets are frequently noisy and incomplete, while models necessarily simplify reality, their synergistic integration can unlock insights that neither approach could achieve alone.

Hybrid strategies are rapidly becoming central to scientific discovery:

1. **Physics-Informed Neural Networks (PINNs):**
   This is a cutting-edge deep learning framework for solving problems involving differential equations. A neural network is trained not only to fit observed data but also to obey the underlying physical laws of the system, encoded as differential equations. This helps the model generalize better from sparse or noisy data.
2. **Machine Learning for Parameter Estimation:**
   While a mathematical model may capture the structure of a system (e.g., the SIR model), the specific parameters (infection rate, recovery rate) must be estimated from real-world data. Machine learning and statistical inference techniques are essential tools for this task, especially when dealing with high-dimensional and complex models.
3. **Hybrid Modeling:**
   In some cases, parts of a system may be well-understood and easily described by equations, while other parts may be too complex. A hybrid model might use a traditional differential equation for the well-understood component and a neural network, trained on data, to represent the more complex, "black-box" component.

2 Stochastic Dynamics

In the previous chapter, we introduced the concept of modeling systems using differential equations. We begin our deeper exploration with **linear systems**. Many sophisticated systems, when analyzed locally around a point of equilibrium, can be effectively approximated by a linear model; this linearization is a cornerstone of dynamical systems theory.

2.1 Linear Systems: The Foundation of Dynamics

Linear Ordinary Differential Equations (ODEs) are characterized by equations where the dependent variable and its derivatives appear only to the first power. They are invaluable because they can often be solved analytically, providing clear insight into the system's behavior.

Consider a simple electrical **RL circuit** consisting of a resistor (R) and an inductor (L) in series. Kirchhoff's voltage law gives:

$$L\frac{di}{dt} + Ri = 0$$

This first-order linear homogeneous ODE has the solution:

$$i(t) = i_0 e^{-\frac{R}{L}t}$$

where $\tau = L/R$ is the *time constant* characterizing the exponential decay.

Adding a capacitor creates the **RLC circuit**, a second-order system. Using the state vector $\mathbf{y} = [i, Q]^T$ where Q is the charge, we obtain:

$$L\frac{di}{dt} + Ri + \frac{1}{C}Q = 0$$

This is a second-order differential equation. To analyze it as a system, we can define a state vector $\mathbf{y} = [i, Q]^T$. The system of first-order equations is:

$$\begin{cases} \frac{di}{dt} = -\frac{R}{L}i - \frac{1}{LC}Q \\ \frac{dQ}{dt} = i \end{cases} \qquad \Rightarrow \qquad A = \begin{bmatrix} -R/L & -1/(LC) \\ 1 & 0 \end{bmatrix}$$

The behavior of this system (e.g., oscillations, decay) is entirely determined by the eigenvalues of the matrix A.

Observation: ***The Ubiquity of the Linear Model***

The same mathematical structure appears in completely different physical domains. Consider a damped harmonic oscillator, such as a mass on a spring moving through a viscous fluid. Newton's second law gives:

$$m\ddot{x} = -kx - \gamma\dot{x}$$

where k is the spring constant and γ is the damping coefficient. Rearranging gives:

$$m\ddot{x} + \gamma\dot{x} + kx = 0$$

This equation shares the same mathematical structure as the RLC circuit equation, illustrating how mathematical models can uncover similarities across entirely different physical systems.

2.2 The Need for Stochasticity: introducing noise

The deterministic models discussed so far are powerful idealizations, however, they assume that all parameters and forces are known and constant. In reality, systems are constantly subjected to small, unpredictable influences from their environment. The aggregate effect of these influences is termed ***noise***. Modeling this randomness is crucial for creating realistic descriptions of the world.

A classic illustration of this is the phenomenon of **Brownian motion**. In 1827, botanist Robert Brown observed the erratic, random motion of pollen grains suspended in water. A simple deterministic model for a particle in a fluid, incorporating only a drag force ($m\dot{v} = -kv$), predicts that the particle's velocity should decay to zero almost instantly. This prediction is in contradiction with Brown's empirical observations of perpetual, jittery motion.

The solution to this paradox came from understanding what friction really is at the microscopic level. The drag force $-kv$ represents only the *average* effect of countless molecular collisions with the particle. Einstein and Langevin realized that each individual collision is random and unpredictable, creating fluctuating forces that keep the particle in perpetual motion.

In general, a particle moving through a fluid can experience three types of forces: the damping force we've already discussed, random collisions from molecules, and possibly some external deterministic force $F_d(x)$:

$$m\ddot{x} = -k\dot{x} + F_d(x) + F_s(t)$$

Let's consider the case where the particle is very light compared to the damping strength, meaning $m \ll k$. To make our notation cleaner, we can write $F_s(t) = k\xi(t)$ and $F_d(x) = kf(x)$:

$$m\ddot{x} = -k\dot{x} + kf(x) + k\xi(t)$$

When $m/k \ll 1$ (the so-called overdamped limit), the inertial term becomes negligible, and we arrive at:

$$\dot{x} = f(x) + \xi(t)$$

This is the famous **Langevin equation**. It represents a fundamental shift in how we think about physical systems—moving from purely deterministic descriptions to ones that embrace randomness. The equation captures the idea that what we observe as smooth, deterministic forces are often just the averaged effects of chaotic microscopic processes.

Example: *Noise in a RLC circuit*

In the previous example of the RLC circuit

$$Ri = -\frac{d}{dt}\Phi(B_{self}) = -\frac{d}{dt}(Li)$$

the derivation was incomplete!

Indeed, a circuit where a current i is flowing is embedded not only in the magnetic field B_{self} generated by $\Phi(Li)$ but also in other external random magnetic fields.

$$B_{total} = B_{self} + B_{external}$$

$$\Phi(B_{total}) = Li + \Phi(B_{external}) = Li + KB_{external}$$

with stochastic external disturbances reds as follows (by the current):

$$Ri = -L\frac{d}{dt}i - K\frac{d}{dt}B_{external}$$

which we may rewrite as

$$L\frac{d}{dt}i = -Ri + K\zeta(t)$$

2.2.1 Modeling Impulsive Events: The Dirac Delta

Before we can properly define the stochastic force $F_s(t)$, we must first develop a mathematical tool to describe events that are instantaneous and intense. Consider the impact of a baseball bat on a ball. The force is immense but acts over a very short duration. We care about the net effect, the change in the ball's velocity, rather than the precise evolution of the force during the infinitesimal impact time. Let's model this using Newton's law, $m\dot{v} = F(t)$. The total change in momentum is the integral of the force over the impact interval, say from $t = 0$ to $t = a$.
The velocity after the impact is:

$$v_{\text{after}} = \frac{1}{m}\int_0^a F(t)\,dt$$

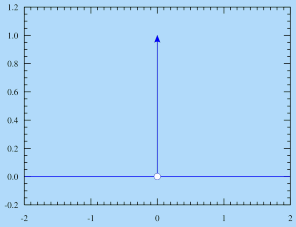
This integral, which represents the total impulse delivered, is finite. To model this, we introduce an idealized mathematical object known as the **Dirac delta function**, denoted $\delta(t)$. It is not a true function in the classical sense but a *distribution* defined by its effect under an integral.

Definition: ***The Dirac Delta Function***

The Dirac delta function, $\delta(t)$, is a generalized function that describes an idealized impulse. Rather than being a function in the classical sense, $\delta(t)$ is a mathematical object with the following properties:

1. There exists a "very small" interval $J = (-\varepsilon/2, \varepsilon/2)$ outside which $\delta(t) \approx 0$.
2. $\delta(t) > 0$ within this interval.
3. The integral of $\delta(t)$ over J is equal to one:

$$\int_J \delta(t)\,dt = 1$$



The crucial consequence of these properties is the ***sifting property***: for any function $f(t)$ that is continuous at $t = 0$, the Dirac delta "sifts out" its value at that point:

$$\int_{-\infty}^{\infty} f(t)\delta(t)\,dt = f(0)$$

To further illustrate the sifting property, consider a function $f(t)$ that is continuous and differentiable at $t = 0$, with $f(0) < \infty$ and $f'(0) < \infty$. Expanding $f(t)$ in a Taylor series around $t = 0$ gives:

$$f(t) = f(0) + f'(0)t + O(t^2)$$

Integrating $f(t)$ against the Dirac delta over the real line, we have:

$$\int_{-\infty}^{\infty} \delta(t)f(t)\,dt = \int_{-\infty}^{\infty} \delta(t)\left[f(0) + f'(0)t\right]dt$$

which can be split as:

$$= \int_{-\infty}^{\infty} \delta(t)f(0)\,dt + \int_{-\infty}^{\infty} \delta(t)f'(0)t\,dt$$

Now, recall the two fundamental properties of the Dirac delta: $\int_{-\infty}^{\infty} \delta(t)\,dt = 1$ and $\int_{-\infty}^{\infty} \delta(t)t\,dt = 0$ (since $t\delta(t)$ is an odd function). Applying these, we find:

$$\int_{-\infty}^{\infty} \delta(t)f(t)\,dt = f(0)\cdot 1 + f'(0)\cdot 0 = f(0)$$

This demonstrates that the Dirac delta "picks out" the value of f at $t = 0$, regardless of the behavior of f elsewhere.

Dirac Delta Classes of Functions

The Dirac delta function, being a mathematical idealization, can be approximated by various families of functions that become increasingly "spike-like" as a parameter grows large.
For $N \gg 1$, the family of rectangular pulse functions

$$\delta_N(t) = \begin{cases} N, & \text{if } t \in \left(-\frac{1}{2N}, \frac{1}{2N}\right) \\ 0, & \text{otherwise} \end{cases}$$

provides an example of Dirac delta approximation. As N increases, the function becomes taller and narrower while maintaining unit area, converging to the delta function in the distributional sense.
Similarly, the exponential family

$$\delta_\gamma(t) = \frac{\gamma}{2} e^{-\gamma|t|}$$

for $\gamma \gg 1$ offers a smooth approximation that avoids the discontinuities of the rectangular pulse. This family has several advantages: it is infinitely differentiable, has exponentially decaying tails, and provides a more realistic model for physical phenomena.

Observation: *Verification of the Sifting Property*

Let's verify that the rectangular pulse function $\delta_N(t)$ satisfies the sifting property in the limit as $N \to \infty$. For any continuous function $f(t)$:

$$\int_{-\infty}^{\infty} f(t)\delta_N(t)dt = \int_{-1/(2N)}^{1/(2N)} f(t) \cdot N\, dt = N \int_{-1/(2N)}^{1/(2N)} f(t)\, dt$$

Since $f(t)$ is continuous at $t = 0$, we can use the mean value theorem.
There exists some $c \in [-1/(2N), 1/(2N)]$ such that:

$$N \int_{-1/(2N)}^{1/(2N)} f(t)\, dt = N \cdot f(c) \cdot \frac{1}{N} = f(c)$$

As $N \to \infty$, the interval shrinks to zero and $c \to 0$. By continuity of f, we have $f(c) \to f(0)$.
Therefore:

$$\lim_{N\to\infty} \int_{-\infty}^{\infty} f(t)\delta_N(t)dt = f(0)$$

This confirms that $\delta_N(t)$ approaches the Dirac delta function in the distributional sense.

2.2.2 Defining the Stochastic Force: White Noise

The stochastic force in the Langevin equation, commonly denoted as $\xi(t)$ (or sometimes $F_s(t)$, after normalization by mass and other constants), models the cumulative effect of a huge number of independent, microscopic collisions. This motivates the following statistical properties, which define what is known as **Gaussian white noise**.

1. **White Noise (Temporal Uncorrelation):** The values of $\xi(t)$ at different times are uncorrelated, reflecting the physical assumption that molecular collisions at different instants are independent. This is mathematically expressed as:

   $$\langle \xi(t)\xi(q) \rangle = 0 \qquad \text{for } t \neq q$$

   More generally, the autocorrelation function is given by the Dirac delta:

   $$\langle \xi(t)\xi(q) \rangle = \delta(t-q)$$

where $\delta(t-q)$ is the Dirac delta function, indicating that the noise is only correlated with itself at the same instant.

2. **Gaussian Distribution:** By the central limit theorem, the sum of many independent random impacts leads to a Gaussian distribution for $\xi(t)$ at any fixed time t. This ensures that all finite-dimensional distributions of the process are Gaussian, making the process analytically tractable.
3. **Zero Mean:** The random molecular collisions are, on average, isotropic and unbiased, so the mean of the stochastic force vanishes:
$$\langle \xi(t) \rangle = 0$$
where $\langle \cdot \rangle$ denotes the ensemble average.
4. **Infinite Instantaneous Variance:** The idealized white noise process is so irregular that its variance at any fixed time is formally infinite:
$$\langle \xi^2(t) \rangle \gg 1$$
This reflects the mathematical abstraction of white noise, which is not a function in the usual sense but a generalized function (distribution).

Warning: *A Mathematical Abstraction*

The concept of white noise, with its delta-correlated structure, is a powerful but physically unrealizable abstraction. It implies infinite variance ($\langle \xi(t)^2 \rangle \to \infty$) and infinite power. In any real system, correlations exist over some small but non-zero timescale. However, if this correlation time is much shorter than any other characteristic timescale of the system, modeling the noise as "white" is an excellent and mathematically convenient approximation.

Multiplicative Noise: The Case of the SIS Epidemic Model

Multiplicative noise arises when the amplitude of the noise depends on the state of the system itself. In the context of stochastic differential equations, this means the noise term is not simply additive, but is multiplied by a function of the state variable:
$$\frac{dx}{dt} = f(x,t) + g(x,t)\,\xi(t)$$
where $g(x,t)$ is not constant.
A classic example is found in epidemiology, specifically in the **SIS epidemic model**. In this model, individuals can become infected, recover, and then become susceptible again (as opposed to the SIR model, where recovered individuals are removed from the susceptible pool). The deterministic SIS model is given by:
$$\frac{dS}{dt} = -\beta SI + \mu(N-S)$$
$$\frac{dI}{dt} = \beta SI - \mu I$$
where S and I are the numbers of susceptible and infected individuals, β is the infection rate, μ is the recovery rate, and N is the total population.

Incorporating Stochasticity

In reality, the contact rate β is not perfectly constant: it fluctuates due to random factors such as social behavior, environmental changes, or other sources of randomness. To model this, we

introduce a stochastic term:

$$\beta \to \beta + \sigma \xi(t)$$

where σ quantifies the strength of the noise and $\xi(t)$ is Gaussian white noise.
Substituting this into the SIS model, the equation for the number of infected individuals becomes:

$$\frac{dI}{dt} = [\beta SI - \mu I] + \sigma SI\,\xi(t)$$

Here, the noise term is ***multiplicative***: its amplitude depends on both S and I. This reflects the fact that random fluctuations in the infection rate have a larger effect when there are more susceptible and infected individuals interacting.

Observation: ***Multiplicative vs. Additive Noise***

Additive noise affects the system independently of its state, while multiplicative noise depends on the current state. In epidemic models, multiplicative noise is more realistic, as the impact of random fluctuations in transmission is naturally proportional to the number of possible contacts.

2.2.3 The Simplest Langevin Equation (SLAE)

To build a deeper intuition for the behavior of stochastic systems, we analyze the simplest possible Langevin equation, where the deterministic drift term is zero ($f(x) = 0$). This corresponds to a free particle subject only to random kicks from its environment:

$$\dot{x} = \omega \xi(t)$$

Here, ω is a constant representing the intensity of the noise $\xi(t)$. While this equation is ill-defined in its derivative form, we can formally integrate it to find the particle's position $x(t)$, assuming an initial position $x(0)$:

$$x(t) = x(0) + \omega \int_0^t \xi(s) ds$$

From this expression, we can derive the fundamental statistical properties of the process $x(t)$.

1. **Average Position**
   The average position, or the first moment of $x(t)$, is found by taking the ensemble average of the equation. Since the noise has zero mean, $\langle \xi(s) \rangle = 0$, the integral of the average noise vanishes:
   $$\langle x(t) \rangle = \langle x(0) \rangle + \omega \int_0^t \langle \xi(s) \rangle ds = \langle x(0) \rangle$$
   If the initial position is deterministic, $\langle x(0) \rangle = x_0$, the average position of the particle does not change over time, since the particle is equally likely to be pushed in any direction.
2. **Autocorrelation and "Memory"**
   To understand how the position at one time relates to the position at another, we compute the autocorrelation function, $\langle x(t)x(q) \rangle$.
   $$\langle x(t)x(q) \rangle = \left\langle \left(\omega \int_0^t \xi(s) ds \right) \left(\omega \int_0^q \xi(\theta) d\theta \right) \right\rangle = \omega^2 \int_0^t \int_0^q \langle \xi(s)\xi(\theta) \rangle ds d\theta$$
   Using the white noise property $\langle \xi(s)\xi(\theta) \rangle = \delta(s - \theta)$, we get:
   $$\langle x(t)x(q) \rangle = \omega^2 \int_0^t \left(\int_0^q \delta(s - \theta) d\theta \right) ds$$

The inner integral with respect to θ is 1 if s is within the interval $[0,q]$, and 0 otherwise. This simplifies the double integral. Assuming, without loss of generality, that $t \leq q$, the condition $s \in [0,q]$ is always met for the outer integral's range $s \in [0,t]$. The integral thus becomes:

$$\langle x(t)x(q)\rangle = \begin{cases} \omega^2(t-q) & \text{if } t > q \\ \omega^2(q-t) & \text{if } t < q \end{cases} \quad = \quad \omega^2|t-q|$$

We can combine these cases into a single elegant expression:

$$\langle x(t)x(q)\rangle = \omega^2 \min(t,q)$$

This result reveals that, unlike the driving noise $\xi(t)$, the position process $x(t)$ ***does*** have memory. Its position at time t is correlated with its position at all other times.

3. **Mean Squared Displacement and Variance**
   Setting $q = t$ gives the second moment, or ***mean squared displacement*** from the origin (assuming $x(0) = 0$):

$$\langle x^2(t)\rangle = \omega^2 t$$

   The ***variance*** of the process is then:

$$\mathrm{Var}[x(t)] = \langle x^2(t)\rangle - \langle x(t)\rangle^2 = \omega^2 t$$

   The variance grows linearly with time, a hallmark of diffusive processes. The particle, on average, wanders further and further from its starting point.

4. **Mean Squared Increment and Non-Differentiability**
   Let's examine the behavior of the process over a small time increment. The mean squared change in position over an interval of length $|t-q|$ is:

$$\begin{aligned} \langle (x(t)-x(q))^2\rangle &= \langle x^2(t)\rangle + \langle x^2(q)\rangle - 2\langle x(t)x(q)\rangle \\ &= \omega^2 t + \omega^2 q - 2\omega^2 \min(t,q) \\ &= \omega^2|t-q| \end{aligned}$$

   Now consider the incremental ratio, which approximates the derivative. Let $q = t+h$:

$$\left\langle \left(\frac{x(t+h)-x(t)}{h} \right)^2 \right\rangle = \frac{\langle (x(t+h)-x(t))^2\rangle}{h^2} = \frac{\omega^2 h}{h^2} = \frac{\omega^2}{h}$$

   In the limit as the interval h shrinks to zero, this quantity diverges:

$$\lim_{h\to 0^+} \left\langle \left(\frac{x(t+h)-x(t)}{h} \right)^2 \right\rangle = \lim_{h\to 0^+} \frac{\omega^2}{h} = +\infty$$

   The mean squared value of the derivative is infinite. This is a profound result: it is the first formal evidence that the path $x(t)$, while continuous, is ***nowhere differentiable***. This highly irregular, "jagged" nature is a fundamental property of processes driven by white noise.

We said that the measure of the memory of the white noise is represented by the following average:

$$\langle\, \xi(t)\xi(q)\,\rangle = \delta(qt)$$

and we showed some heuristic reasoning. However, for the sake of precision, the above formula is derived by a statistical concept that is very useful to verify the degree of similarity between two generic random variables: the **covariance**, which is defined as follows:

$$\mathrm{COV}(x,y) = \langle (x - \langle x\rangle)(y - \langle y\rangle)\rangle = \langle xy\rangle - \langle x\rangle\langle y\rangle$$

For the sake of precision, in the cases where y is a temporal shift of x (as in our case), we say that we are computing the **autocovariance** of x.

2.2.4 The Wiener Process and its Properties

The mathematical pathologies of white noise $\xi(t)$ and the resulting non-differentiability of its integral motivated the development of a more rigorous framework. Instead of focusing on the ill-defined derivative form $\dot{x} = \xi(t)$, we work directly with the integrated process. This process, which formalizes the concept of Brownian motion, is named in honor of Norbert Wiener.

Definition: *The Wiener Process*

The **Wiener process** $W(t)$ is a continuous-time stochastic process that is the solution to the SDE:

$$\begin{cases} \dfrac{dW}{dt} = \xi(t) \\ W(0) = 0 \end{cases}$$

where $\xi(t)$ is Gaussian white noise with unit variance.

Note: While this differential form is intuitive, it should be understood that $\xi(t)$ is not a function in the classical sense, and the equation is interpreted in the sense of stochastic integration.

The Wiener process is a Gaussian process with the following key properties:

1. **Zero Mean:** The expected value of the process is zero:

   $$\langle W(t) \rangle = 0$$

   This property reflects the symmetric nature of the random fluctuations around the origin.

2. **Gaussian Distribution:** For any fixed time t, $W(t)$ is normally distributed with mean zero and variance t:

   $$W(t) \sim \mathcal{N}(0,t)$$

   This means the probability density function is given by:

   $$p(W,t) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{W^2}{2t}\right)$$

3. **Autocorrelation:** The autocorrelation function is given by

   $$\langle W(t)W(s) \rangle = \min(t,s)$$

   This property captures the ”memory” of the process.

4. **Independent Increments:** The increments of the process are independent. In particular, for any collection of non-overlapping time intervals $[t_1,s_1],[t_2,s_2],\ldots,[t_n,s_n]$, the corresponding increments $W(s_1) - W(t_1), W(s_2) - W(t_2), \ldots, W(s_n) - W(t_n)$ are independent random variables. The mean of each increment is:

   $$\langle W(s) - W(t) \rangle = 0$$

   and these increments are also Gaussian.

5. **Increment Variance:**
   The variance of the increment over the interval $[t,s]$ is proportional to the time difference:

   $$\langle (W(s) - W(t))^2 \rangle = |s-t|$$

   This scaling with time difference is a fundamental property that distinguishes diffusive processes from other types of motion.

6. **Increment Distribution:** More precisely, for $s > t$, the increment is distributed as

$$W(s) - W(t) \sim \mathcal{N}\Big(0, |s-t|\Big)$$

This means the increment has a normal distribution with zero mean and variance equal to the time difference. In particular, the distribution of $W(t)$ itself (starting from $W(0) = 0$) is:

$$W(t) \sim \mathcal{N}(0,t) \qquad \Rightarrow \qquad p(W,t) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{W^2}{2t}\right)$$

These properties completely characterize the Wiener process. Any continuous-time stochastic process satisfying these conditions is a Wiener process, making it a fundamental building block in the theory of stochastic processes and stochastic differential equations.

Increment Analysis and Non-Differentiability

The properties of the Wiener process's increments lead to one of its most counter-intuitive and important features: its paths are continuous but almost surely nowhere differentiable. Let's analyze an infinitesimal increment $dW_t = W(t+dt) - W(t)$ over a small time interval $dt > 0$. From the properties above, this increment is a Gaussian random variable with mean 0 and variance dt:

$$dW_t \sim \mathcal{N}(0, dt)$$

This implies that the standard deviation of the increment is $\sigma = \sqrt{dt}$. Now, consider the finite difference quotient used to define a derivative:

$$\frac{dW_t}{dt} = \frac{W(t+dt) - W(t)}{dt}$$

This new random variable is also Gaussian, but its variance is $\mathrm{Var}\left(\frac{dW_t}{dt}\right) = \frac{1}{dt^2}\mathrm{Var}(dW_t) = \frac{dt}{dt^2} = \frac{1}{dt}$. As $dt \to 0$, this variance diverges to infinity. The notion of a derivative in the classical sense is therefore problematic.

We can investigate this further by considering the probability that the magnitude of the incremental ratio exceeds some large, arbitrary value M. Let h be a small, finite time step. We are interested in:

$$\Pr\left(\left|\frac{W(t+h) - W(t)}{h}\right| > M\right)$$

Since the increment $W(t+h) - W(t)$ follows a $\mathcal{N}(0,h)$ distribution, we can standardize it by dividing by its standard deviation, $\sqrt{h}$. Let $Z = \frac{W(t+h)-W(t)}{\sqrt{h}}$, where $Z \sim \mathcal{N}(0,1)$. The probability becomes:

$$\Pr\left(\left|\frac{Z\sqrt{h}}{h}\right| > M\right) = \Pr\left(|Z| > M\sqrt{h}\right)$$

As $h \to 0$, the threshold $M\sqrt{h}$ also goes to zero. The probability that a standard normal variable Z has a magnitude greater than an infinitesimally small number approaches 1.

$$\lim_{h\to 0} \Pr\left(|Z| > M\sqrt{h}\right) = \Pr(|Z| > 0) = 1$$

This means that for any large M, as we make the time interval smaller, it becomes a near certainty that the slope of the secant line will exceed M. This confirms that the path of a Wiener process is not differentiable at any point.

3 Stochastic Numerical Methods

The non-differentiability of the Wiener process path fundamentally breaks the foundation of classical calculus, rendering traditional analytical techniques inadequate for stochastic differential equations. This mathematical obstacle necessitates the development of an entirely new framework—stochastic calculus, with its own integration theory and differentiation rules.

3.1 The Differential Form of SDEs: Itô Equation

Let's reconsider Newton's second law. In its most common form, it is written as $F = ma$. However, its more fundamental statement relates force to the change in momentum $p = mv$, expressed in differential form as:

$$dp = Fdt$$

This form is more general. For instance, consider the motion of a rocket, whose mass $m(t)$ changes as it consumes fuel. In this case, $F = ma$ is incorrect. The correct formulation is:

$$d(m(t)v(t)) = Fdt$$

This differential way of writing physical laws is powerful and provides the foundation for correctly interpreting stochastic equations. A Langevin equation written as $\dot{x} = f(x,t) + g(x,t)\xi(t)$ is mathematically problematic. The rigorous approach is to express it in its differential form using the Wiener process increment dW_t, which represents the integral of the white noise $\xi(t)$:

$$dx = f(x,t)dt + g(x,t)dW_t$$

This is a **Stochastic Differential Equation (SDE)**. Its solution is understood in an integral sense:

$$x(t) = x_0 + \int_0^t f(x(\tau),\tau)d\tau + \int_0^t g(x(\tau),\tau)dW_\tau$$

The second integral is a stochastic integral, an object whose properties are fundamentally different from the standard Riemann integral. The infinitesimal increment dW_t is defined as $dW_t = W(t + dt) - W(t)$. As we have established, it is a Gaussian random variable with mean 0 and variance dt, so $dW_t \sim \mathcal{N}(0, dt)$. This can be expressed as:

$$dW_t = G(t)\sqrt{dt}$$

where $G(t)$ is a random variable drawn from the standard normal distribution, $\mathcal{N}(0,1)$.
The stochastic integral $\int_0^t g(x(\tau),\tau)dW_\tau$ represents the cumulative effect of the random forcing over time. Unlike deterministic integrals, this integral cannot be evaluated using traditional calculus rules due to the irregular nature of the Wiener process paths. The integral must be understood in the sense of Itô or Stratonovich, with Itô integration being the more commonly used convention in stochastic differential equations.

Definition: ***Itô Equation***

Given an SLAE, we can always rewrite it as an **Itô equation**, by defining $dW = G(t)\sqrt{dt}$:

$$dx = f(x)dt + g(x)dW$$

3.1.1 The Euler-Maruyama Method

The Itô equation can be solved numerically using the Maruyama algorithm for stochastic differential equations. This is essentially the stochastic version of Euler's algorithm. Given a time interval $[0,T]$ and $h = T/N$, so that $t = jh$ for $j = 0,\ldots,N$. Suppose the Euler algorithm is:

$$dx = x(t_j + h) - x(t_j) = x(t_{j+1}) - x(t_j)$$

we can set $dt \approx h$ and write:

$$x(t_{j+1}) = x(t_j) + f(x(t_j))h$$

The Maruyama formula starts from this form by also considering the Gaussian effect, adding:

$$x(t_{j+1}) = x(t_j) + f(x(t_j))h + G_j\sqrt{h}$$

with $G_j \sim \mathcal{N}(0,1)$. Obviously, starting from this algorithm, more precise variants have been successively created.

Definition: ***The Euler-Maruyama Method***

For the stochastic differential equation $dX_t = a(X_t,t)dt + b(X_t,t)dW_t$ with initial condition $X(0) = x_0$ and uniform time step h, the Euler-Maruyama approximation is given by:

$$X_{j+1} = X_j + a(X_j,t_j)h + b(X_j,t_j)\sqrt{h}G_j$$

where $t_j = jh$ and $\{G_j\}_{j=0}^{N-1}$ is a sequence of independent standard normal random variables.

This numerical scheme provides a practical foundation for simulating stochastic processes, though more sophisticated methods have been developed to improve accuracy and stability for specific applications.

3.2 Basic Applications of SDEs

3.2.1 The Stochastic Malthusian Model and its Paradox

The simplest model of population growth is the Malthusian model, which assumes an unlimited environment and a constant per capita growth rate, r. The dynamics are described by the ODE:

$$\dot{x} = rx$$

where $x(t)$ represents the population size. The solution is simple exponential growth or decay, $x(t) = x(0)e^{rt}$. A more realistic model acknowledges that the growth rate is not constant but fluctuates randomly due to environmental variations. We can model this by making the growth rate a stochastic process, $r \to r + \omega\xi(t)$. This leads to the ***stochastic Malthusian model***, an SDE with multiplicative noise:

$$dx = (r + \omega\xi(t))x$$

where ω denotes the multiplicative noise amplitude. We can obtain the Ito formula:

$$dx = rx\mathrm{d}t + \omega x\mathrm{d}W$$

The solution of this stochastic equation can be obtained by applying Itô's formula. Setting $y = \ln x$, we find the drift and diffusion coefficients for the transformed process. Using Itô's rule:

$$dy = \left[\frac{1}{x}rx - \frac{1}{2x^2}\omega^2x^2\right]\mathrm{d}t + \frac{1}{x}\omega x\mathrm{d}W \quad \xrightarrow{\text{simplifying}} \quad dy = \left[r - \frac{\omega^2}{2}\right]dt + \omega dW$$

This simplifies to a linear SDE which can be solved directly. Integrating from 0 to t:

$$y(t) = y_0 + \left(r - \frac{\omega^2}{2}\right)t + \omega W(t)$$

Supposing $W_0 = 0$, we can calculate the moments of $y(t)$:

$$\langle y(t)\rangle = \langle y_0\rangle + \left(r - \frac{\omega^2}{2}\right)t$$

Therefore, if we have $\omega^2/2 > r$, then $y(t) \to -\infty$. We notice that, intuitively considering what $y(t)$ represents, if ω is large the population will tend to extinction regardless of r. In other words, a population highly subject to events, whether negative or positive, will tend to extinction.
We can then find the second moment of $y(t)$:

$$\langle y(t)^2\rangle = \left\langle \left[y_0 + \left(r - \frac{\omega^2}{2}\right)t\right]^2 + 2\omega\left[y_0 + \left(r - \frac{\omega^2}{2}\right)t\right]W(t) + \omega^2 W(t)^2 \right\rangle$$

which, upon solving, becomes:

$$... = \left[y_0 + \left(r - \frac{\omega^2}{2}\right)t\right]^2 + \omega^2 t$$

Assuming for convenience that y_0 is deterministic, then we have that the variance of $y(t)$ is:

$$\mathrm{Var}[y(t)] = \langle y(t)^2\rangle - \langle y(t)\rangle^2 = \omega^2 t$$

Again, we can note that the variance tends to diverge over time. Returning now to $x(t)$, we have:

$$x(t) = e^{y(t)} = e^{y_0 + \left(r - \frac{\omega^2}{2}\right)t} e^{\omega W(t)} = x(t) = x_0 e^{\left(r - \frac{\omega^2}{2}\right)t} e^{\omega W(t)}$$

Then its mean value will be:

$$\langle x(t)\rangle = x_0 e^{\left(r - \frac{\omega^2}{2}\right)t} \langle e^{\omega W(t)}\rangle$$

So, we must compute the mean of $\exp(\omega W(t))$. We know that $W(t)$ is distributed with the distribution:

$$W(t) \sim N(0,t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{W^2}{2t}}$$

so, to calculate the mean, we must compute:

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi t}} e^{\omega W} e^{-\frac{W^2}{2t}} dW$$

We can then combine the two exponentials and rewrite the resulting exponent as:

$$e^{\frac{-W^2 + 2t\omega W - t^2\omega^2 + t^2\omega^2}{2t}} = e^{-\frac{(W - t\omega)^2}{2t}} e^{\frac{\omega^2 t}{2}}$$

As a consequence, the integral rewritten this way becomes:

$$... = e^{\frac{\omega^2 t}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(W - t\omega)^2}{2t}} dW$$

This, however, is the integral of a translated Gaussian which is trivially equal to 1. Therefore:

$$\langle e^{\omega W(t)} \rangle = e^{\frac{\omega^2 t}{2}}$$

Thus,

$$\langle x(t) \rangle = \langle x_0 \rangle e^{\left(rt - \frac{\omega^2 t}{2}\right) + \frac{\omega^2 t}{2}}$$

In the Langevin hypothesis, it was the fact that the non-stochastic differential equation was nothing other than the differential equation of the mean, and here we find this result. Indeed, the two terms subtract, giving as a solution the solution of the non-stochastic case.

The Paradox

However, something strange emerges when we examine the long-term behavior. Returning to the definition of $y(t)$, we can rewrite it as:

$$y(t) = y_0 + \left(r - \frac{\omega^2}{2}\right) t + \omega \sqrt{t} G$$

As $t \to \infty$, the deterministic term $\left(r - \frac{\omega^2}{2}\right) t$ dominates, so:

$$\lim_{t \to +\infty} y(t) = \begin{cases} -\infty & \text{if } r < \frac{\omega^2}{2} \\ +\infty & \text{if } r > \frac{\omega^2}{2} \end{cases}$$

Since $x(t) = e^{y(t)}$, when $y(t) \to -\infty$, we have $x(t) \to 0$. This creates an apparent paradox: we showed that $\langle x(t) \rangle = x_0 e^{rt}$, which always grows exponentially regardless of ω, yet individual realizations $x(t)$ can tend to extinction when the noise is sufficiently large ($\omega^2/2 > r$).

The resolution lies in understanding the nature of the log-normal distribution. When $x(t) = e^{y(t)}$ where $y(t)$ is normally distributed, $x(t)$ follows a log-normal distribution, which is highly skewed. The mean and median can differ dramatically in such distributions.

Example: ***The Exponential Distribution***

Consider the simple exponential distribution with density ae^{-aU}. Its mean is $\langle U \rangle = 1/a$, but its median satisfies:

$$\int_0^{\text{MED}} ae^{-aU} dU = \frac{1}{2}$$

which gives $\text{MED} = \log(2)/a < 1/a$. The median is smaller than the mean because the distribution has a long tail that pulls the mean upward.

Similarly, in our stochastic population model, $x(t)$ has a log-normal distribution. While the mean grows exponentially due to rare but extremely large population bursts, the typical behavior (represented by the median) can show decline when noise is large. The mean becomes dominated by infrequent but massive population explosions, making it a poor representative of typical outcomes.

Calculation of the Median

To quantify the typical behavior of the population, we need to calculate the median of $x(t)$. Since $y(t) \sim \mathcal{N}(\mu, \omega^2)$ where $\mu = y_0 + (r - \omega^2/2)t$ and the variance parameter is $\omega^2 t$, the random variable $x(t) = e^{y(t)}$ follows a log-normal distribution with probability density function:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\omega\sqrt{t} \cdot x} \exp\left(-\frac{(\log x - \mu)^2}{2\omega^2 t}\right), \quad x > 0$$

The median m is defined as the value satisfying $P(X \leq m) = 1/2$, which gives us the integral equation:

$$\frac{1}{2} = \int_0^m \frac{1}{\sqrt{2\pi}\omega\sqrt{t}\cdot x} \exp\left(-\frac{(\log x - \mu)^2}{2\omega^2 t}\right) dx$$

To solve this integral, we employ the substitution $w = \log x$, which transforms $dx = e^w dw = x\,dw$. The limits of integration become $w \in (-\infty, \log m]$, and our integral becomes:

$$\frac{1}{2} = \int_{-\infty}^{\log m} \frac{1}{\sqrt{2\pi}\omega\sqrt{t}} \exp\left(-\frac{(w-\mu)^2}{2\omega^2 t}\right) dw$$

This is precisely the cumulative distribution function of a normal random variable $\mathcal{N}(\mu, \omega^2 t)$ evaluated at $\log m$. Since the median of any normal distribution equals its mean, we have:

$$\log m = \mu = y_0 + \left(r - \frac{\omega^2}{2}\right) t$$

Therefore, the median of $x(t)$ is:

$$\text{MED}[x(t)] = e^{\mu} = e^{y_0 + \left(r - \frac{\omega^2}{2}\right)t} = x_0 e^{\left(r - \frac{\omega^2}{2}\right)t}$$

This result elegantly resolves the paradox. While the mean $\langle x(t) \rangle = x_0 e^{rt}$ always grows exponentially, the median, which better represents typical population trajectories, follows the drift-corrected dynamics. The long-term behavior of the median is:

$$\lim_{t \to +\infty} \text{MED}[x(t)] = \begin{cases} 0 & \text{if } r < \frac{\omega^2}{2} \quad \text{(extinction regime)} \\ x_0 & \text{if } r = \frac{\omega^2}{2} \quad \text{(critical regime)} \\ +\infty & \text{if } r > \frac{\omega^2}{2} \quad \text{(growth regime)} \end{cases}$$

The threshold $r = \omega^2/2$ represents a critical noise level: below this threshold, typical populations grow indefinitely, while above it, typical populations face extinction despite the exponentially growing mean. This dichotomy between mean and median behavior is a hallmark of log-normal processes and highlights the importance of choosing appropriate summary statistics for highly skewed distributions.

3.2.2 The Perturbed Logistic Equation

We have discussed the fact that the Malthus model is not realistic because it assumes infinite resources. A much better model for population dynamics is the **Logistic model**, which incorporates a density-dependent growth rate. The deterministic logistic model can be written as follows:

$$\dot{x} = r(x)x$$

where $r(x)$ is a decreasing function of the population size x. The simplest choice for this function is linear:

$$r(x) = r_0 - \alpha x$$

Here, r_0 is the intrinsic growth rate at low densities, and α is a coefficient representing the strength of density-dependent regulation (e.g., competition for resources).

Introducing Stochasticity

In a realistic scenario, the intrinsic growth rate r_0 is not constant but fluctuates due to environmental variability. We can model its fast fluctuations as:

$$r_0 \to r_0 + \omega\xi(t)$$

where $\xi(t)$ is Gaussian white noise. This transforms the deterministic ODE into the stochastically perturbed logistic model:

$$dx = (r_0 - \alpha x)x\,dt + \omega x\,dW_t$$

To analyze this equation, we can again use the logarithmic transformation $y = \ln(x)$. Applying Itô's formula yields:

$$dy = \left(r_0 - \frac{\omega^2}{2} - \alpha e^y\right)dt + \omega dW$$

This transformed SDE can be formally integrated to give:

$$y(t) = y(0) + \left(r_0 - \frac{\omega^2}{2}\right)t + \omega W(t) - \alpha\int_0^t e^{y(s)}ds$$

We can now analyze the long-term behavior of the system. We know two key facts:

1. If $\omega^2 > 2r_0$, then the non-integral part will tend to $-\infty$ as $t \to \infty$.
2. The integral of the exponential is always positive since $x = e^y$ must be positive.

Slightly more precisely, the fact that $\int_0^t e^{y(s)}ds > 0$ gives us an upper bound on the process $y(t)$:

$$y(t) \leq y(0) + \left(r_0 - \frac{\omega^2}{2}\right)t + \omega W(t)$$

This leads us to a **sufficient condition** for population extinction. If the right-hand side of the inequality goes to $-\infty$, then $y(t)$ must also go to $-\infty$. This happens when the drift of the bounding process is negative. Thus, the condition for extinction is:

$$\frac{\omega^2}{2} > r_0 \quad \Longrightarrow \quad \lim_{t\to\infty} x(t) = 0$$

This powerful result shows that if the environmental noise is sufficiently strong, the population will go extinct regardless of the density-dependent term. The noise effectively suppresses the intrinsic growth.

4 Itô and Stratonovich Calculus

Having established the framework of Itô equations, we now turn our attention to the fundamental problem of change of variables in stochastic calculus. When dealing with deterministic differential equations, the chain rule provides a straightforward mechanism for transforming variables. However, in the stochastic setting, the irregular nature of Brownian motion necessitates a more sophisticated approach.

4.1 Itô's Lemma: The Stochastic Chain Rule

Consider an SDE in Itô form:

$$dx = a(x)dt + b(x)dW$$

Suppose we wish to perform a transformation from x to y defined by $y = \psi(x)$. Our objective is to derive an equation of similar form for the transformed variable y.
Following the classical approach, we expand dy in a Taylor series:

$$\begin{aligned} dy &= \psi'(x)dx + \tfrac{1}{2}\psi''(x)(dx)^2 + \dots \\ &= \Psi'(x)[a(x)dt + b(x)dw] + \frac{1}{2}\Psi''(x)\left[b^2(x)(dw)^2 + \underbrace{\cancel{a^2(x)(dt)^2}}_{O(dt^2)} + \underbrace{\cancel{2a(x)b(x)dtdw}}_{O(dt^{3/2})}\right] \end{aligned}$$

To derive a formula consistent with the Itô framework, we must carefully consider the order of magnitude of each term. Since $dW = O(\sqrt{dt})$, we can eliminate terms of order $(dt)^{3/2}$ and higher, including the mixed term $dt\,dW$ and $(dt)^2$. The remaining term $(dW)^2$ is of order dt, and by the fundamental property of Brownian motion, we have $(dW)^2 = dt$.
Substituting this, we get:

$$dy = \psi'(x)a(x)dt + \psi'(x)b(x)dW + \frac{b(x)^2}{2}\psi''(x)dt$$

Realigning the terms, we get:

$$dy = \left[\psi'(x)a(x) + \frac{b(x)^2}{2}\psi''(x)\right]dt + \psi'(x)b(x)dW$$

Definition: ***Itô's Lemma***

Let X_t be an Itô process that satisfies the SDE $dX_t = a(X_t,t)dt + b(X_t,t)dW_t$. Let $\psi(x,t)$ be a twice-differentiable scalar function. Then the process $Y_t = \psi(X_t,t)$ is also an Itô process, and its differential dY_t is given by:

$$dY_t = \left[\psi'(x)a(x) + \psi''(x)\frac{b(x)^2}{2}\right]dt + \psi'(x)b(x)dW_t$$

This fundamental result encompasses the classical chain rule terms, namely $\psi' a$ and $\psi' b$, augmented by an additional stochastic correction term, $\psi'' \frac{b^2}{2}$, known as the **Itô correction**. This correction term arises directly from the non-vanishing quadratic variation of the Wiener process and represents the fundamental distinction between stochastic and deterministic calculus.

4.1.1 Derivation of Ito's Formula

We recall that during the analysis of the change of variable for Ito's formula, we substituted $(dW)^2 \to dt$ into the equation without formally proving its correctness. Let us now investigate this further, starting from the general SDE:

$$dx = a(x)dt + b(x)dW$$

We know that the Wiener increment can be expressed as $dW = \sqrt{dt}G(t)$, where $G(t) \sim \mathcal{N}(0,1)$. Our goal was to apply a variable transformation $y = \psi(x)$ to this equation. We found that this transformation resulted in:

$$dy \equiv d\psi = dt\left[\psi'(x)a(x) + \psi''(x)\frac{b(x)^2}{2}\right] + \psi'(x)b(x)dW$$

This result was achieved through the "magical" substitution mentioned earlier. The objective was to obtain a new SDE for y that has the same form as the original one:

$$dy = q(y)dt + r(y)dW$$

This required having one term of order $O(dt)$ and another of order $O(\sqrt{dt})$, and to achieve this, we discarded all terms of higher order than dt. To verify this substitution, we must revisit a concept from earlier. Let us consider the increment of the Wiener process:

$$z = W(t+h) - W(t)$$

If we now consider the random variable $q = z^2$, we have $\langle q \rangle = h$, which is the variance of z. This was the rationale for the substitution we made in the derivation of dW, but in doing so, we were neglecting potentially important elements. Let us now evaluate the variance of q.
To formally establish this, we compute the fourth moment of a Gaussian random variable $z \sim \mathcal{N}(0,h)$, which represents the Wiener increment dW_t over a time step $h = dt$.

$$\langle z^4 \rangle = \int_{-\infty}^{+\infty} z^4 \frac{1}{\sqrt{2\pi h}} e^{-\frac{z^2}{2h}} dz.$$

Integration by parts, with $u = z^3$ and $dv = ze^{-z^2/2h}dz/\sqrt{2\pi h}$, yields:

$$\langle z^4 \rangle = \left[z^3\left(-\frac{h}{\sqrt{2\pi h}}e^{-z^2/2h}\right)\right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty}\left(-\frac{h}{\sqrt{2\pi h}}e^{-z^2/2h}\right)3z^2 dz.$$

The boundary term vanishes, leaving:

$$\langle z^4 \rangle = 3h\int_{-\infty}^{+\infty} z^2 \frac{1}{\sqrt{2\pi h}} e^{-\frac{z^2}{2h}} dz = 3h\langle z^2 \rangle = 3h(h) = 3h^2.$$

With $h = dt$, this gives $\langle (dW_t)^4 \rangle = 3(dt)^2$. The variance of $(dW_t)^2$ is therefore:

$$\mathrm{Var}[(dW_t)^2] = \langle (dW_t)^4 \rangle - \left(\langle (dW_t)^2 \rangle\right)^2 = 3(dt)^2 - (dt)^2 = 2(dt)^2.$$

Since the variance is of order $(dt)^2$, the fluctuations of $(dW_t)^2$ around its mean dt are of a higher order than dt itself. Consequently, in the limit $dt \to 0$, we can make the substitution $(dW_t)^2 = dt$. Substituting this result into the Taylor expansion for $d\Psi$ gives the final expression:

$$d\Psi = \left[\Psi'(x)a(x) + \frac{1}{2}\Psi''(x)b^2(x)\right]dt + \Psi'(x)b(x)dW,$$

which is the celebrated **Ito's formula**. The presence of the second-derivative term, arising from the non-zero quadratic variation of the Wiener process, is the fundamental feature that distinguishes stochastic from ordinary calculus.

4.1.2 Stochastic Equilibrium Points

The concept of an equilibrium point can be extended from deterministic to stochastic systems. For a deterministic system $\dot{x} = f(x)$, an equilibrium point x_e is defined by the condition $f(x_e) = 0$. For a stochastic differential equation (SDE) of the form

$$dx = f(x)dt + g(x)dW_t,$$

a **stochastic equilibrium point (SEP)**, denoted x_{se}, is a point where both the drift and diffusion terms vanish simultaneously:

$$f(x_{se}) = 0 \quad \text{and} \quad g(x_{se}) = 0.$$

This dual condition makes SEPs significantly rarer than their deterministic counterparts. As an example, consider the **stochastic Malthusian model**:

$$dx = r_0 x\,dt + \omega x dW_t.$$

In this case, $x_{se} = 0$ is a trivial SEP, as both $f(0) = r_0 \cdot 0$ and $g(0) = \omega \cdot 0$ are zero. The stability of this equilibrium is determined by the interplay between the growth rate r_0 and the noise intensity ω, as established previously:

- If $\omega^2/2 > r_0$, the median of the population size converges to zero, rendering the equilibrium point $x_{se} = 0$ **stochastically globally attractive**.
- If $\omega^2/2 < r_0$, the median grows exponentially, and the equilibrium point $x_{se} = 0$ acts as a **stochastic repulsor** (i.e., it is unstable).

This analysis extends to the **perturbed logistic model**, $dx = (r_0 x - \alpha x^2)dt + \omega x dW_t$, where $x_{se} = 0$ is also a SEP. The condition $\omega^2/2 > r_0$ remains sufficient for the global attractivity of the origin.

4.2 The Fokker-Planck Equation

While Ito's Lemma allows us to find the SDE for a transformed variable, its most powerful application is in deriving a deterministic equation for the evolution of the probability density function (PDF), $\rho(x,t)$. This bridge from the stochastic world of individual paths to the deterministic world of distributions is the celebrated **Fokker-Planck equation**.

The derivation is a beautiful piece of mathematical physics that relies on a "weak" formulation. Instead of tracking $\rho(x,t)$ directly, we analyze how the expected value of an arbitrary, well-behaved "test function" $\psi(x)$ evolves over time.

4.2.1 Probability Density Function and its Evolution

Given a continuous-time, continuous-state stochastic process, $x(t)$, we often want to describe its behavior not by a single trajectory, but by the probability of finding the process in a certain state at a certain time. This is accomplished using the **probability density function (PDF)**, denoted $\rho(x,t)$. The PDF is defined such that the probability of the random variable $x(t)$ being in an infinitesimal interval $[\hat{x}, \hat{x}+d\hat{x}]$ is given by:

$$\text{Prob}\left(x(t) \in [\hat{x}, \hat{x}+d\hat{x}]\right) = \rho(\hat{x},t)d\hat{x}$$

This implies that the probability of finding $x(t)$ in a finite interval $[a,b]$ is the integral of the PDF over that interval:

$$\text{Prob}\left(x(t) \in [a,b]\right) = \int_a^b \rho(x,t)dx$$

A central question in stochastic modeling is: if we know the initial distribution of the process, $\rho(x,0)$, how does this distribution evolve for $t > 0$?

In the most general case, the law of evolution for $\rho(x,t)$ could depend on the entire history of the process, often denoted as $\Omega(x(\theta), 0 \le \theta < t)$. This would mean that the future state depends on the full path taken to reach the present, leading to a complex law that could be described by an integro-differential equation.

However, for a large and very important class of processes, the situation is significantly simpler.

The Markov Property and Stochastic Differential Equations

Processes described by an Itô Stochastic Differential Equation (SDE) of the form

$$dx = f(x)dt + g(x)dW_t$$

have a special structure. The state of the system at an infinitesimal future time $t+dt$ is given by:

$$x(t+dt) = x(t) + f(x(t))dt + g(x(t))dW_t$$

Crucially, the statistical properties of $x(t+dt)$ depend only on the state $x(t)$ at the current time t, and not on the entire prior history. This is the hallmark of a **Markov process**.

Definition: ***Markov Process***

A stochastic process $x(t)$ is called a ***Markov process*** if its future probability distribution, conditioned on its past and present values, depends only on the present value. In other words, the past and the future are conditionally independent given the present.

For a process modeled by an Itô SDE, the following key properties hold:

1. **The process is Markovian.** The future state depends only on the present, not the path taken to get there.
2. **The process has continuous paths.** The Itô equation implies that increments are infinitesimal. Since the drift term $f(x)dt$ is of order $O(dt)$ and the diffusion term $g(x)dW_t$ is of order $O(\sqrt{dt})$, the process $x(t)$ does not have finite jumps.

These two properties together, being Markovian and having continuous paths, have a profound implication: the law governing the evolution of the PDF, $\rho(x,t)$, must be local in both time and space. It cannot depend on spatially distant values or past temporal values. This means the evolution operator must be a local differential operator. Therefore, the evolution equation for $\rho(x,t)$ must be a **Partial Differential Equation (PDE)**.

4.2.2 Derivation from Ito's Lemma

Before presenting the Fokker-Planck equation, we will derive it step by step using Ito's Lemma. The derivation proceeds by considering the time evolution of the expected value of an arbitrary smooth function of the process, and then using integration by parts to arrive at a local equation for the PDF itself.

Consider a stochastic process $x(t)$ described by the Itô SDE:

$$dx = a(x)\,dt + b(x)\,dW_t$$

To analyze the evolution of its probability distribution, we introduce a smooth "test function" $\psi(x)$, assumed to be twice continuously differentiable and to vanish (along with its first derivative) at the boundaries of the domain (for example, as $x \to \pm\infty$). This ensures that boundary terms will not contribute when integrating by parts.

We are now ready to systematically derive the celebrated **Fokker-Planck equation**, which governs the time evolution of the probability density $\rho(x,t)$ for the process $x(t)$:

1. **Apply Ito's Lemma to $\psi(x)$**
   From Ito's Lemma, we know the differential for $\psi(x(t))$ is:

$$d\psi(x) = \left[a(x)\psi'(x) + \frac{b(x)^2}{2}\psi''(x)\right]dt + b(x)\psi'(x)dW_t$$

2. **Take the Expectation**
   Next, we take the ensemble average (expectation) of this equation.

$$\langle d\psi(x)\rangle = \left\langle \left[a(x)\psi'(x) + \frac{b(x)^2}{2}\psi''(x)\right]dt \right\rangle + \langle b(x)\psi'(x)dW_t\rangle$$

   The expectation of the stochastic term vanishes, because dW_t has zero mean and is independent of the state $x(t)$ at the beginning of the infinitesimal step. This leaves us with an equation for the evolution of the mean of $\psi(x)$:

$$d\langle\psi(x)\rangle = \left\langle a(x)\psi'(x) + \frac{b(x)^2}{2}\psi''(x)\right\rangle dt$$

   Dividing by dt, we get the time derivative:

$$\frac{d}{dt}\langle\psi(x)\rangle = \left\langle a(x)\psi'(x) + \frac{b(x)^2}{2}\psi''(x)\right\rangle$$

3. **Introduce the PDF**
   The expectation of any function $h(x)$ can be written as an integral of that function against the PDF $\rho(x,t)$. Applying this to both sides of our equation:

$$\frac{d}{dt}\int_S \psi(x)\rho(x,t)dx = \int_S \left[a(x)\psi'(x) + \frac{b(x)^2}{2}\psi''(x)\right]\rho(x,t)dx$$

   where S is the state space. Since $\psi(x)$ does not depend on time, we can bring the time derivative inside the integral on the left:

$$\int_S \psi(x)\frac{\partial \rho(x,t)}{\partial t}dx = \int_S \psi'(x)a(x)\rho(x,t)dx + \int_S \psi''(x)\frac{b(x)^2}{2}\rho(x,t)dx$$

4. **Integration by Parts**
   The goal now is to remove the derivatives from the test function $\psi(x)$ and transfer them onto the other terms using integration by parts. The general formula is $\int udv = [uv] - \int vdu$.

   For the first term on the right, let $u = a(x)\rho(x,t)$ and $dv = \psi'(x)dx$.

$$\int_S \psi'(x)[a(x)\rho(x,t)]dx = \cancel{[\psi(x)a(x)\rho(x,t)]_{\partial S}} \quad \boxed{-\int_S \psi(x)\frac{\partial}{\partial x}[a(x)\rho(x,t)]dx}$$

   For the second term on the right, we must integrate by parts twice.

$$\int_S \psi''(x)\left[\frac{b(x)^2}{2}\rho(x,t)\right]dx = \left[\psi'(x)\frac{b^2\rho}{2}\right]_{\partial S} - \int_S \psi'(x)\frac{\partial}{\partial x}\left[\frac{b^2\rho}{2}\right]dx$$

   Due to our assumption that ψ and its derivatives are zero at the boundary ∂S, all the boundary terms vanish.

   We apply integration by parts again to the remaining integral:

$$-\int_S \psi'(x)\frac{\partial}{\partial x}\left[\frac{b^2\rho}{2}\right]dx = -\cancel{\left[\psi(x)\frac{\partial}{\partial x}\left[\frac{b^2\rho}{2}\right]\right]_{\partial S}} \quad \boxed{+\int_S \psi(x)\frac{\partial^2}{\partial x^2}\left[\frac{b(x)^2}{2}\rho(x,t)\right]dx}$$

   The boundary term again vanishes, leaving only the final integral.

5. **The Final Equation**
   Substituting these results back into our main equation, we have:

$$\int_S \psi(x)\frac{\partial \rho}{\partial t}dx = -\int_S \psi(x)\frac{\partial}{\partial x}[a\rho]dx + \int_S \psi(x)\frac{\partial^2}{\partial x^2}\left[\frac{b^2\rho}{2}\right]dx$$

   We can now group all terms under a single integral:

$$\int_S \psi(x)\left[\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}[a\rho] - \frac{\partial^2}{\partial x^2}\left[\frac{b^2\rho}{2}\right]\right]dx = 0$$

The crucial insight here is that this equation must be satisfied for ***any*** admissible test function $\psi(x)$. In mathematical analysis, the only way for an integral of the form

$$\int_S \psi(x)\,[\cdots]\,dx = 0$$

to hold for all sufficiently smooth and well-behaved $\psi(x)$ is for the integrand itself—the expression inside the square brackets—to vanish identically at every point in S. This is a direct consequence of the fundamental lemma of the calculus of variations.
By setting the integrand to zero, we arrive at one of the most important results in stochastic processes: the celebrated **Fokker-Planck equation**, which governs the time evolution of the probability density function associated with the stochastic differential equation.

4.2.3 The Fokker-Planck Equation

Having derived the Fokker-Planck equation from Itô's Lemma, we are now ready to present its formal definition and explore its implications through illustrative examples. This powerful equation serves as the cornerstone for understanding how probability densities evolve under stochastic dynamics.

Definition: ***Fokker-Planck Equation***

Given a Stochastic Differential Equation of the type:

$$dx = a(x)dt + b(x)dW$$

its associated Probability Density Function, $\rho(x,t)$, will solve the Fokker-Planck equation:

$$\frac{\partial \rho}{\partial t} = -\frac{\partial}{\partial x}\left(a(x)\rho(x,t)\right) + \frac{\partial^2}{\partial x^2}\left(\frac{b(x)^2}{2}\rho(x,t)\right)$$

This must, of course, be supplemented with boundary conditions (e.g., Dirichlet) and an initial condition $\rho(x,0)$ to obtain a unique solution. Note that imposing a specific initial state $x(0)$ actually yields the conditional PDF, $\rho(x,t|x(0))$.

In summary, the probability density function $\rho(x,t)$ associated with the stochastic process governed by the SDE above satisfies the celebrated **Fokker-Planck equation** (also known as the forward Kolmogorov equation):

$$\frac{\partial}{\partial t}\rho(x,t) = -\frac{\partial}{\partial x}\left[a(x)\rho(x,t)\right] + \frac{\partial^2}{\partial x^2}\left(\frac{b^2(x)}{2}\rho(x,t)\right)$$

This partial differential equation describes the time evolution of the probability density $\rho(x,t)$ for the random variable x.

Since $\rho(x,t)$ is a probability density function, it must always satisfy the **normalization constraint**:

$$\int_{-\infty}^{\infty} \rho(x,t)\,dx = 1$$

This ensures that the total probability is conserved at all times.

To uniquely determine the solution, we must also specify the **initial distribution** of x:

$$\rho(x,0) = \theta(x)$$

where $\theta(x)$ is the given initial probability density function (for example, it could be a Dirac delta function if the initial state is known exactly, or a broader distribution if there is uncertainty).

Together, the Fokker-Planck equation, the normalization condition, and the initial condition fully characterize the time evolution of the probability density for the stochastic process.

Let's now consider some examples of the Fokker-Planck equation.

Example: *Particle in a Potential Well with Additive Noise*

Let us describe the motion of a particle in a conservative potential field $U(x)$, subject to random thermal fluctuations. In the overdamped limit ($m \ll 1$), where inertial effects are negligible, the system's dynamics can be described by:

$$dx = f(x)dt + \omega dW_t$$

where $f(x) = -\frac{\partial U}{\partial x}$, and ω represents the constant intensity of the noise. Our goal is to find the **steady-state probability distribution**. The Fokker-Planck equation for this system is:

$$\frac{\partial \rho(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[f(x)\rho(x,t)\right] + \frac{\omega^2}{2}\frac{\partial^2 \rho(x,t)}{\partial x^2}$$

At steady state, the distribution no longer changes with time, so $\frac{\partial \rho}{\partial t} = 0$. We denote the stationary solution as $P(x)$. The equation simplifies to an ordinary differential equation:

$$\frac{d}{dx}\left[-f(x)P(x) + \frac{\omega^2}{2}\frac{dP(x)}{dx}\right] = 0$$

The term in the square brackets is the ***probability current*** $J(x)$ (see following section). For a system with reflecting or natural boundaries at infinity, this current must be zero everywhere:

$$-f(x)P(x) + \frac{\omega^2}{2}\frac{dP(x)}{dx} = 0$$

Substituting $f(x) = -U'(x)$, this becomes a separable first-order ODE:

$$\frac{\omega^2}{2}\frac{dP}{dx} = f(x)P = -U'(x)P \quad \Longrightarrow \quad \frac{dP}{P} = -\frac{2}{\omega^2}U'(x)dx$$

Integrating both sides gives $\ln P = -\frac{2}{\omega^2}U(x) + \text{const}$. Exponentiating yields the solution:

$$P(x) = C\exp\left(-\frac{2}{\omega^2}U(x)\right)$$

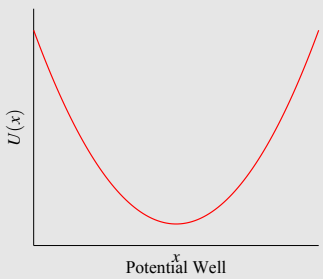
This is the famous **Boltzmann distribution**. C is determined by the condition $\int_{\mathbb{R}} P(x)dx = 1$:

$$C = \frac{1}{\displaystyle\int_{-\infty}^{+\infty} \exp\left(-\frac{2}{\omega^2}U(x)\right)dx}$$

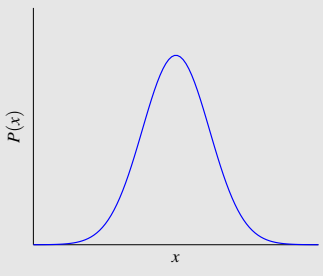
The particle is most likely to be found where $U(x)$ is lowest, though noise (ω) allows it to occasionally reach higher-energy states.
Let's consider the specific case of a harmonic potential well, $U(x) = \frac{1}{2}kx^2$. We have:

$$P(x) = C\exp\left(-\frac{k}{\omega^2}x^2\right)$$

This is a Gaussian distribution centered at $x = 0$; the variance is $\sigma^2 = \omega^2/(2k)$, showing that stronger noise (larger ω) leads to a wider distribution.



Potential Well



Stationary PDF

4.2.4 The Probability Current

The Fokker-Planck equation can also be interpreted from a different, physically intuitive perspective. Recall its general form:

$$\partial_t P = \partial_x \left[-a(x)P(x,t) - \partial_x \left(\frac{b(x)^2}{2} P(x,t) \right) \right]$$

If we factor out the outer derivative and define the expression inside the brackets as J, we obtain:

$$\partial_t P + \partial_x J = 0$$

This is immediately recognizable as a continuity equation (or conservation law) for probability. In higher dimensions, the equation generalizes to:

$$\partial_t \eta + \nabla \cdot J = 0$$

where η is the probability density and J is the probability current (or flux vector).
This formulation highlights that the probability behaves like a conserved "fluid". The change in its distribution is nothing more than a redistribution, so the total amount does not change.
In one dimension, the continuity equation is a direct reduction of the multi-dimensional case. For example, in the case of pure Brownian motion (diffusion), the probability current takes the form:

$$J = -K \nabla P$$

where K is the diffusion coefficient. Here, J describes the net flow of probability due to diffusion, always moving from regions of high to low probability density. For this reason, J is often referred to as the **probability current**.

Example: *Particle density*

We want to study how the spatial density $n(x,t)$ of a large number of particles evolves, where each particle follows the same stochastic dynamics:

$$\begin{cases} \dot{x}_j = f(x_j) & \rightarrow \ drift\ term \\ n(x,0) = \theta^*(x) & \rightarrow \ initial\ density \end{cases}$$

We assume that particles are neither created nor destroyed, so the total number N is conserved:

$$\int_{\mathbb{R}} n(x,t)\,dx = N = \int_{\mathbb{R}} n(x,0)\,dx = \int_{\mathbb{R}} \theta^*(x)\,dx$$

The motion takes place in $\mathbb{R}$. The probability that a particle is in $C = [a,b]$ at time t is:

$$\Pr[x(t) \in [a,b]] \simeq \frac{N_c(t)}{N} = \frac{\int_a^b n(x,t)\,dx}{N} = \frac{\#\text{ particles in } [a,b]}{\#\text{ particles in } \mathbb{R}}$$

Thus, the probability density is:

$$\rho(x,t) = \frac{n(x,t)}{N}$$

Evolution of $\mathbf{n(x,t)}$

Consider now an infinitesimal interval $[x, x+dx]$, and let $Q_x = N_{[x,x+dx]}$ be the number of particles in it. Since each particle evolves according to $\dot{x} = f(x)$, the velocity at position x is:

$$v(x) = f(x)$$

The particle fluxes are:

1. Number of ***particles entering*** $[x, x+dx]$ per unit time: $n(x,t)\,v(x)\,dt$
2. Number of ***particles leaving*** $[x, x+dx]$ per unit time: $n(x+dx,t)\,v(x+dx)\,dt$

The net change in particle number in $[x, x+dx]$ is:

$$dQ = n(x,t)v(x)\,dt - n(x+dx,t)v(x+dx)\,dt$$

Dividing by dt and using a Taylor expansion of $n(x+dx,t)$ around x:

$$\frac{dQ}{dt} = -\partial_x[n(x,t)v(x)]\,dx$$

but $Q_{[x,x+dx]} = n(x,t)\,dx$, so:

$$\frac{\partial}{\partial t}n(x,t)\,dx = -\partial_x[n(x,t)v(x)]\,dx$$

Dividing both sides by dx gives the following system:

$$\begin{cases} \dfrac{\partial n}{\partial t} + \dfrac{\partial}{\partial x}[v(x)n(x,t)] = 0 \\ n(x,0) = \theta^*(x) \\ \displaystyle\int n(x,t)\,dx = N \end{cases}$$

This is the ***Liouville equation*** for the normalized density $\frac{n(x,t)}{N}$.

The probability current

We now introduce two key quantities:

1. $n(x,t)\,dx$: # of particles in the interval $(x, x+dx)$ at time t.
2. $J(x,t) = n(x,t)\,v(x)$: # of particles crossing position x at time t, (***current density***).

The net change in the number of particles in $[x, x+dx]$ over time dt is:

$$dQ = J(x,t)\,dt - J(x+dx,t)\,dt$$

This leads to the continuity equation:

$$\partial_t n = -\partial_x J \qquad \Rightarrow \qquad \partial_t n + \partial_x J = 0 \qquad \xrightarrow{\text{in } \mathbb{R}^2} \qquad \partial_t n + \operatorname{div} J = 0$$

The FP equation can also be interpreted as describing the evolution of the spatial density of a large amount of particles, each subject to a white noise force and an initial density $\theta(x)$:

$$\partial_t n(x,t) = -\partial_x\big(f(x)\,n(x,t)\big) + \partial^2_{xx}\left(\frac{g^2(x)}{2}\,n(x,t)\right)$$

where n is normalized so that $\int_{\mathbb{R}} n(x,t)\,dx = N = \int_{\mathbb{R}} \theta(x)\,dx$.
If each particle evolves according to $\dot{x} = f(x) + g(x)\,\xi(t)$, the particle current is:

$$J = f(x)\,n(x,t) - \partial_x\left(\frac{g^2(x)}{2}\,n(x,t)\right)$$

The analogous expression for the ***probability current*** (using normalized density $\rho(x,t)$) is:

$$J_{\text{prob}} = f(x)\,\rho(x,t) - \partial_x\left(\frac{g^2(x)}{2}\,\rho(x,t)\right)$$

4.2.5 Special Cases of the Fokker-Planck Equation

Deterministic Systems with Random Initial Conditions

Suppose we have a deterministic system, $\dot{x} = a(x)$, but where the initial conditions are described by a probability distribution $\theta(x)$. Here, randomness comes only from the initial condition, not from noise in the dynamics: the diffusion coefficient is zero, $b(x) = 0$. The Fokker-Planck equation loses its second-order derivative term and becomes the **Liouville equation**:

$$\begin{cases} \dfrac{\partial P(x,t)}{\partial t} = -\dfrac{\partial}{\partial x}\left(a(x)P(x,t)\right) \\ P(x,0) = \theta(x) \end{cases}$$

This type of problem appears, for example, when studying the evolution of a population's distribution over time under deterministic laws, but starting from a known initial distribution.

Purely Diffusive Systems

The opposite case occurs when $a(x) = 0$, and the system is governed only by noise:

$$dx = b(x)dW$$

In population dynamics, this is useful in cases where the baseline growth rate is zero ($r = 0$) or constant. In this situation, the Fokker-Planck equation becomes a pure diffusion equation:

$$\partial_t P = \partial_x^2 \left(\frac{b(x)^2}{2} P(x,t) \right)$$

This is, in fact, a generalization of the process for the Wiener process, which I recall is $\dot{w} = \xi(t)$ (i.e., $a = 0, b = 1$). Indeed, using the Fokker-Planck equation, we can find the expression for the PDF of the Wiener process, which will be:

$$\partial_t P = \frac{1}{2}\partial_w^2 P$$

Or, analogously, for $\dot{x} = \omega\xi(t)$, which corresponds to the overdamped Brownian motion:

$$\partial_t P = \frac{\omega^2}{2}\partial_x^2 P$$

Kac's Lemma

Kac's Lemma establishes a fundamental relationship between stationary probabilities and expected return times in ergodic stochastic processes. Consider a stationary Markov process with stationary distribution π and let A be a measurable subset of the state space. Define T_A as the first return time to A when starting from a point in A. Kac's Lemma states:

$$\mathbb{E}[T_A | X_0 \in A] = \frac{1}{\pi(A)}$$

where $\pi(A) = \int_A \pi(x)dx$ is the stationary measure of set A.

This result provides a direct connection between the equilibrium properties of the system (captured by π) and its dynamical properties (captured by return times). States or regions with low stationary probability have correspondingly long expected return times.

The lemma follows from the ergodic theorem: for an ergodic process, the long-run proportion of time spent in any set equals its stationary probability. Since visits to A occur approximately every $\mathbb{E}[T_A]$ time units, we have $\pi(A) \approx 1/\mathbb{E}[T_A]$.

4.2.6 Multiple Independent Noises

Consider the following Itô stochastic differential equation (SDE) with two independent noise sources:

$$dx = a(x)\,dt + b_1(x)\,dW_1 + b_2(x)\,dW_2$$

where dW_1 and dW_2 are independent Wiener processes (i.e., uncorrelated Brownian motions). The associated Fokker-Planck equation (FPE) for the probability density $P(x,t)$ is obtained by extending the standard procedure for a single noise term. The result is:

$$\partial_t P(x,t) = -\partial_x\left(a(x)P(x,t)\right) + \partial_{xx}^2\left(\frac{b_1^2(x)+b_2^2(x)}{2}P(x,t)\right)$$

In other words, the diffusion coefficients from each independent noise source add together in the FPE. This generalizes immediately to any number of independent noises: the total diffusion is the sum of the squared amplitudes of each noise term.

Example: ***RL Circuit with Multiple Noise Sources***

Consider an RL circuit where both the external magnetic field and the resistance fluctuate independently. We start with a circuit perturbed by fluctuating external magnetic fields Φ_{ext} such that $(d/dt)\Phi_{ext}$ is white noise:

$$\frac{dI}{dt} = -\frac{R}{L}I + \omega_1\xi_1(t)$$

However, the resistance is not constant; it fluctuates due to thermal agitation, which we model as white noise:

$$R \to R + K\xi_2(t)$$

This yields the following model:

$$\frac{dI}{dt} = -\frac{R+K\xi_2(t)}{L}I + \omega_1\xi_1(t)$$

Expanding this expression:

$$\frac{dI}{dt} = -\frac{R}{L}I - \frac{K}{L}I\xi_2(t) + \omega_1\xi_1(t)$$

In this model, the noise is not only additive but also multiplicative. We have two independent noise sources: $\xi_1(t)$ (additive) from the magnetic field fluctuations and $\xi_2(t)$ (multiplicative) from the resistance fluctuations. The total noise contribution combines both effects according to the rules we have established for multiple independent noises.

4.3 Multiplicative Noise

While additive noise provides a good model for systems perturbed by external forces, many real-world systems feature fluctuations whose intensity depends on the state of the system itself. This leads to the concept of **multiplicative noise**, which can induce much richer and more complex behaviors than its additive counterpart.

Consider a system governed by a parameter p. The simplest case is when it is possible to write the field in the form

$$\dot{x} = b(x) + p\hat{g}(x)$$

that is, with linear dependence on the parameter considered. The parameter p is usually influenced by the external world, so in some cases it is possible for it to become stochastic. We can then describe a stochastic p as

$$p \to p + \alpha\xi(t)$$

Substituting, we obtain

$$\dot{x} = b(x) + p\hat{g}(x) + \hat{g}(x)\alpha\xi(t)$$

We can then group the stochastic terms, obtaining

$$\dot{x} = f(x) + g(x)\xi(t)$$

The **Fokker-Planck equation** for the general case is:

$$\frac{\partial \rho}{\partial t} = -\frac{\partial}{\partial x}[f(x)\rho] + \frac{\partial^2}{\partial x^2}\left[\frac{g(x)^2}{2}\rho\right]$$

So far, we have only considered cases where $g(x)$ is constant. Here, however, we are dealing with the most general version of what we have seen so far. The boundary conditions remain the same as before, and as usual, we analyze the stationary case. We have:

$$\frac{d}{dx}\left[\frac{g(x)^2}{2}P_s(x)\right] = f(x)P_s(x)$$

This is a first-order ODE that can be solved for $P_s(x)$.

Let's define an auxiliary function $Q(x) = \frac{g(x)^2}{2}P_s(x)$. The equation becomes:

$$\frac{dQ}{dx} = f(x)\frac{2}{g(x)^2}Q(x)$$

This is a standard linear ODE of the form $Q'(x) = a(x)Q(x)$, whose solution is $Q(x) = Ce^{A(x)}$, where $A(x) = \int a(s)ds$. Applying this, we find:

$$Q(x) = C\exp\left(\int \frac{2f(s)}{g(s)^2}ds\right)$$

Substituting back $P_s(x) = \frac{2}{g(x)^2}Q(x)$, we arrive at the general solution for the stationary PDF:

$$P_s(x) = \frac{C'}{g(x)^2}\exp\left(\int^x \frac{2f(z)}{g(z)^2}dz\right)$$

where C' is a normalization constant. Let's remark that, being a probability density, the integral of $P_s(x)$ over the entire state space must be equal to one.

We can distinguish two cases:

- If the integral exists (is finite), then everything is fine and we can calculate C to normalize $P_s(x)$.
- If instead the integral is infinite, we can only find a formal solution that has no physical meaning: in this case it is necessary to study the steady state directly starting from the SDE (Section 4.4).

4.4 Multiplicative Noise: Noise-Induced Transitions

In the case of additive noise ($g(x) = \text{const}$), the extrema of the stationary PDF $P_s(x)$ coincide with the equilibria of the deterministic system (where $f(x) = 0$). With multiplicative noise, this simple correspondence breaks down. Recalling the following identity:

$$\frac{1}{g^2(x)} = e^{-\ln g^2(x)} = e^{-2\ln g(x)}$$

We can rewrite the stationary probability density function in a more transparent exponentialform:

$$P_s(x) = 2C\exp\left\{-2\log(g(x)) + \int_\alpha^x \frac{2f(s)}{g^2(s)}ds\right\}$$

This highlights how both the noise amplitude $g(x)$ and the drift $f(x)$ contribute to shaping the stationary distribution. To investigate the location of the extrema of $P_s(x)$, we differentiate with respect to x. Applying the chain rule to the exponent, we obtain:

$$P_s'(x) = 2C\exp\left\{-2\frac{g'(x)}{g(x)} + \frac{2f(x)}{g^2(x)}\right\}$$

This expression will allow us to determine where the stationary distribution attains its maxima and minima, as these correspond to the points where $P_s'(x) = 0$.

To find the extrema of $P_s(x)$, we must find where its derivative is zero, $P_s'(x) = 0$. After some algebraic manipulation of the general solution, this condition simplifies to:

$$f(x) \geq g(x)g'(x)$$

Of course, if $g(x)$ is constant, then $g'(x) = 0$ and the condition reduces to $f(x) \geq 0$. This immediately recovers the familiar property that stable equilibria correspond to modes of the PDF, and unstable equilibria correspond to antimodes.

Instead, the most interesting case is when $g(x)$ is not constant. In this case, the condition $f(x) \geq g(x)g'(x)$ is a non-linear inequality that can have multiple solutions. This means that by simply varying a parameter that controls the noise intensity, new extrema can appear in the PDF that were completely absent in the unperturbed deterministic system. This phenomenon, discovered by Horsthemke and Lefever in the 1970s, is called a **Noise-Induced Transition (NIT)**.

Before the discovery of NITs, it was commonly believed that noise could only act in two ways:

- In unimodal systems, it creates a ”cloud” of fluctuations around a single deterministic equilibrium.
- In multimodal systems, it can cause ”jumps” between the existing deterministic equilibria.

NITs showed that noise can play a much more constructive role, inducing oscillations and creating entirely new states that are not present in the unperturbed deterministic system.

Observation: ***Extrema vs. Equilibria***

Even if solving

$$f(x) = g'(x,q)g(x,q)$$

one does find as extrema for the stochastic model as the equilibria for the deterministic system, the extrema of the stationary PDF do not correspond to equilibrium points of the deterministic equation, which solved a different equilibrium equation, namely:

$$f(x) = 0$$

Example: *Noise-Induced Bifurcation in Gene Expression*

Consider a simple model for protein production, where a protein X acts as its own transcription factor. The production rate is $Q(x)$ and the degradation rate is dx.

$$\dot{x} = Q(x) - dx$$

A common form for $Q(x)$ that models self-regulation is:

$$Q(x) = R + K\frac{x^2}{K_d + x^2}$$

where R is a basal production rate and the second term is a Hill function describing feedback. The deterministic system is:

$$\dot{x} = \underbrace{R + K\frac{x^2}{K_d + x^2} - dx}_{f(x)}$$

For small degradation rates d, this system has a single, globally stable equilibrium point. Now, suppose the degradation rate d is subject to stochastic fluctuations: $d \to d + \alpha\xi(t)$. The SDE becomes:

$$dx = \left(R + K\frac{x^2}{K_d + x^2} - dx\right)dt - \alpha x dW_t$$

Here, the noise is multiplicative, with $g(x) = -\alpha x$, so $g'(x) = -\alpha$. The condition for the extrema of the stationary PDF, $f(x) \geq g(x)g'(x)$, is:

$$R + K\frac{x^2}{K_d + x^2} - dx \geq (-\alpha x)(-\alpha) = \alpha^2 x$$

Rearranging gives:

$$R + K\frac{x^2}{K_d + x^2} \geq (d + \alpha^2)x$$

Depending on the value of the noise parameter α, this algebraic equation can have one or three solutions. This means that by changing the noise intensity, the stationary PDF can transition from being unimodal (with one peak) to bimodal (with two peaks), even though the underlying deterministic system only had one equilibrium point. This is a classic example of a noise-induced transition: the noise itself creates new, stable macroscopic states for the system.

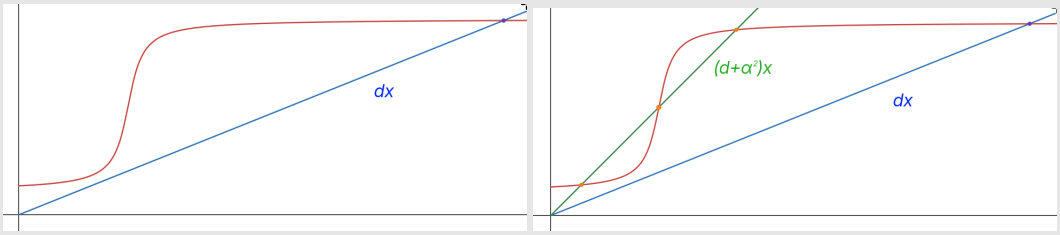


Figure 4.1: Left: Unimodal stationary PDF for null noise intensity $\alpha = 0$. Right: Bimodal stationary PDF emerging for higher noise intensity α, demonstrating a noise-induced transition where new stable states appear that are absent in the deterministic system.

4.4.1 Perturbation of the Logistic Growth Model

Let's revisit the logistic growth model, which serves as another prime example of how multiplicative noise can fundamentally alter a system's behavior. We start with the deterministic logistic equation:

$$\dot{z} = r_1 z - r_2 z^2$$

Through a linear change of variables $x = r_2 z$, we can adimensionalize the equation to a simpler form:

$$\dot{x} = rx - x^2$$

Now, we introduce stochasticity by perturbing the intrinsic growth rate, $r \to r + \omega\xi(t)$. This leads to the corresponding Itô equation with multiplicative noise:

$$dx = (rx - x^2)dt + \omega x dW_t$$

This defines our drift and diffusion coefficients as $f(x) = rx - x^2$ and $g(x) = \omega x$.
The stationary Fokker-Planck equation for this system is:

$$\frac{d}{dx}\left[\frac{g(x)^2}{2}P_s(x)\right] = f(x)P_s(x)$$

Substituting our specific $f(x)$ and $g(x)$ gives:

$$\frac{d}{dx}\left[\frac{\omega^2 x^2}{2}P_s(x)\right] = (rx - x^2)P_s(x)$$

To solve this differential equation, we can use the substitution method. Let $Q = x^2 P_s$, then:

$$Q' = \frac{2r - x}{\omega^2 x}Q$$

This can be rewritten as:

$$\frac{dQ}{Q} = \frac{2r - x}{\omega^2 x}dx = \left[\frac{2r}{\omega^2 x} - \frac{1}{\omega^2}\right]dx$$

Integrating both sides:

$$\ln Q = \frac{2r}{\omega^2}\ln x - \frac{x}{\omega^2} + C$$

Taking the exponential:

$$Q(x) = A\exp\left[\frac{2r}{\omega^2}\ln x - \frac{x}{\omega^2}\right] = Ax^{\frac{2r}{\omega^2}}e^{-\frac{x}{\omega^2}}$$

Since $Q = x^2 P_s$, we obtain:

$$P_s(x) = \frac{Q(x)}{x^2} = Ax^{\frac{2r}{\omega^2}-2}e^{-\frac{x}{\omega^2}}$$

where A is a normalization constant.
A crucial step is to check whether this candidate solution is a valid, normalizable probability distribution. The integral $\int_0^\infty P_s(x)dx$ must be finite.

$$\int_0^\infty x^{\frac{2r}{\omega^2}-2}e^{-\frac{2}{\omega^2}x}dx$$

This is a form of the Gamma function integral, $\int_0^\infty t^{k-1}e^{-t}dt$. For this integral to converge at the lower bound ($x \to 0$), the exponent of x must be greater than -1.

$$\frac{2r}{\omega^2} - 2 > -1 \quad \Longrightarrow \quad \frac{2r}{\omega^2} > 1 \quad \Longrightarrow \quad r > \frac{\omega^2}{2}$$

This is valid only if $r > \omega^2/2$ as we have already seen in the past. If this does not hold, we cannot integrate but what does this actually mean? To understand this, let's return to the SDE:

$$dx = x(r-x)dt + \omega x dW$$

Applying now the transformation $y = \psi(x) = \log x$, then $\psi'(x) = x^{-1}$ and $\psi''(x) = -x^{-2}$. Therefore:

$$dy = \left[\frac{1}{\cancel{x}}x(r-x) - \frac{1}{\cancel{x^2}}\frac{\omega^2\cancel{x^2}}{2}\right]dt + \omega dW$$

Therefore, substituting in y:

$$dy = \left[r - \frac{\omega^2}{2} - e^y\right]dt + \omega dW$$

Integrating then:

$$y(t) = y_0 + \left(r - \frac{\omega^2}{2}\right)t - \int_0^t e^{y(s)}ds + \omega[W(t) - W(0)]$$

Notice that the integral term, $-\int_0^t e^{y(s)}ds$, is always negative (since $e^{y(s)} > 0$ for all s).

- **Case 1:** $r < \omega^2/2$

  If $r < \omega^2/2$, the drift term $\left(r - \frac{\omega^2}{2}\right)t$ is negative and dominates for large t, while the stochastic term becomes negligible in comparison. As a result, $y(t) \to -\infty$ as $t \to \infty$, which means $x(t) = e^{y(t)} \to 0^+$. In this regime, the probability distribution $P(x,t)$ collapses to a delta function at zero:

  $$\lim_{t\to\infty} P(x,t) = \delta(x)$$

  This shows that, in this case, the Fokker-Planck equation predicts a stationary solution, but in reality, all probability accumulates at extinction ($x = 0$). Thus, the Fokker-Planck approach alone can be misleading for this parameter regime.

- **Case 2:** $r > \omega^2/2$

  On the other hand, when $r > \omega^2/2$, the stationary solution is normalizable and given by:

  $$P_s(x) = Ax^{\frac{2r}{\omega^2}-2}e^{-\frac{2}{\omega^2}x}$$

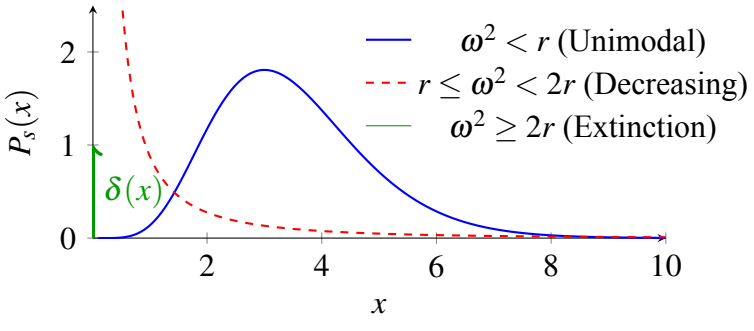


Figure 4.2: Noise-Induced Transition in the logistic model.

Let us also consider two special cases:

- If $\omega^2 = 2r$ (i.e., $2r/\omega^2 = 1$), the stationary distribution simplifies to:

$$P_s(x) = Ax^{-1}e^{-\frac{x}{r}}$$

- If there is no noise, $\omega^2 = 0$, the stationary distribution becomes:

$$P_s(x) = Ae^{-\frac{x}{r}}$$

  This change has a significant effect on the shape of the stationary distribution. Specifically, when the parameters satisfy

$$\frac{2r}{\omega^2} < 2 \qquad \text{or equivalently} \qquad r < \frac{\omega^2}{2}$$

  the distribution $P_s(x)$ is no longer unimodal (with a peak at some $x > 0$), but instead becomes a monotonically decreasing function of x. In this regime, the probability density is highest near $x = 0$ and decreases for larger x, so the system is most likely to be found close to $x = 0$.

Furthermore, it can be shown that the stationary distribution $P_s(x)$ is normalizable (i.e., its total probability integrates to 1) only if $\omega^2 < 2r$. This is consistent with the more general result discussed in the previous section.

4.5 The Stratonovich Integral

The Langevin equation, $\dot{x} = a(x,t) + b(x,t)\xi(t)$, provides an intuitive starting point, but its rigorous interpretation requires us to define the stochastic integral $\int b(x,t)dW_t$. This is best understood by rewriting the SDE in its formal integral form:

$$x(t) = x(0) + \int_0^t a(x(\theta),\theta)d\theta + \int_0^t b(x(\theta),\theta)dW(\theta)$$

As we will see, the very definition of the final term, the stochastic integral, is ambiguous and has led to two major, distinct formulations of stochastic calculus: Itô and Stratonovich.

4.5.1 The Ambiguity of the Stochastic Integral

The integral of a function with respect to a Wiener process, $I = \int_0^t g(X(\theta))dW(\theta)$, cannot be defined in the ordinary Riemann sense because the path of $W(t)$ is nowhere differentiable and not of bounded variation. Instead, it is defined as the limit of a Riemann-Stieltjes sum over a partition of the interval $[0,t]$. Let us partition $[0,t]$ into N subintervals with points $0 = t_0 < t_1 < \ldots < t_N = t$, and let $J = \max_i(t_i - t_{i-1})$ denote the maximum interval length of the partition. For each subinterval, we choose an evaluation point $\tau_i \in [t_{i-1}, t_i]$. The stochastic integral is then defined as the limit (in probability) as $J \to 0$ of the sum:

$$S(J) = \sum_{i=1}^{N} g(X(\tau_i))\left[W(t_i) - W(t_{i-1})\right]$$

where $S(J)$ is a random variable for each partition.
If $W(t)$ were a deterministic, smooth function, the choice of τ_i would not matter in the limit $J \to 0$. However, because $W(t)$ is a random process with highly irregular paths, the value of the stochastic integral depends crucially on the choice of τ_i.

To illustrate this, consider the specific case $I = \int_0^t W(\theta)dW(\theta)$. The corresponding sum is:

$$S(J) = \sum_{i=1}^{N} W(\tau_i)\left[W(t_i) - W(t_{i-1})\right]$$

Let us compute the expected value of this sum:

$$\langle S(J)\rangle = \sum_{i=1}^{N} \langle W(\tau_i)\left[W(t_i) - W(t_{i-1})\right]\rangle = \sum_{i=1}^{N} \left(\langle W(\tau_i)W(t_i)\rangle - \langle W(\tau_i)W(t_{i-1})\rangle\right)$$

Using the autocorrelation property of the Wiener process, $\langle W(t)W(s)\rangle = \min(t,s)$, and noting that $t_{i-1} \leq \tau_i \leq t_i$, we have:

$$\langle S(J)\rangle = \sum_{i=1}^{N} \left(\min(\tau_i, t_i) - \min(\tau_i, t_{i-1})\right) = \sum_{i=1}^{N} (\tau_i - t_{i-1})$$

Thus, the expected value depends on the choice of τ_i within each interval. If we parameterize τ_i as

$$\tau_i = t_{i-1} + \alpha(t_i - t_{i-1}), \qquad \alpha \in [0,1],$$

and consider a uniform partition where $J = t_i - t_{i-1} = t/N$, then

$$\tau_i - t_{i-1} = \alpha J,$$

so

$$\langle S(J)\rangle = \sum_{i=1}^{N} \alpha J = \alpha N \frac{t}{N} = \alpha t.$$

Taking the limit as $J \to 0$ (i.e., $N \to \infty$ and $J \to 0$), we see that

$$\langle I\rangle = \lim_{J\to 0} \langle S(J)\rangle = \alpha t.$$

Therefore, the expectation of the stochastic integral depends directly on the convention used for the evaluation point α. This ambiguity in the definition of the stochastic integral is fundamental and necessitates a choice of interpretation.

- The **Itô Integral** convention chooses $\alpha = 0$, evaluating the integrand at the start of the interval ($\tau_i = t_{i-1}$). This ensures the integrand is independent of the future noise increment, making the resulting stochastic integral a martingale.
- The **Stratonovich Integral** convention chooses $\alpha = 1/2$, evaluating the integrand at the midpoint of the interval ($\tau_i = (t_{i-1} + t_i)/2$). This choice makes the rules of calculus for the integral resemble ordinary calculus.

4.5.2 Choosing a Convention: The Wong-Zakai Theorem

The choice between Itô and Stratonovich is not merely a mathematical preference; it is a modeling decision. The **Wong-Zakai theorem** provides crucial guidance.

Definition: ***Wong-Zakai Theorem (1965)***

Consider an ODE perturbed by "real" noise $\eta_\varepsilon(t)$ that is not perfectly white but has a very short correlation time ε:

$$\dot{x} = a(x) + b(x)\eta_\varepsilon(t)$$

As the correlation time of the noise approaches zero ($\varepsilon \to 0$), so that $\eta_\varepsilon(t)$ converges to a true white noise process, the solution $x(t)$ of the ODE converges to the solution of the **Stratonovich SDE**, denoted by:

$$dx = a(x)dt + b(x) \circ dW_t$$

Tip: ***When to use which integral?***

The Wong-Zakai theorem implies that if an SDE is derived as the limit of a physical system driven by noise with a finite (even if very small) correlation time, the Stratonovich interpretation is the more physically appropriate one. The Itô calculus is often preferred in mathematical finance and for problems where the noise is assumed to be idealized white noise from the outset, or for systems with intrinsically discrete state variables (like population counts).

4.5.3 Connecting the Itô and Stratonovich SDEs

We can find a direct relationship between the two formalisms by expanding the midpoint term in the Stratonovich definition. First, we approximate the state at the midpoint:

$$x\left(t + \frac{dt}{2}\right) \approx x(t) + a(x(t))\frac{dt}{2} + b(x(t))\left(W\left(t + \frac{dt}{2}\right) - W(t)\right)$$

Next, we expand the function b around $x(t)$:

$$b\left(x\left(t + \frac{dt}{2}\right)\right) \approx b(x(t)) + b'(x(t))\left(x\left(t + \frac{dt}{2}\right) - x(t)\right)$$

Substituting the first expression into the second gives:

$$b\left(x\left(t + \frac{dt}{2}\right)\right) \approx b(x(t)) + b'(x(t))\left[a(x(t))\frac{dt}{2} + b(x(t))d\hat{W}\right]$$

where $d\hat{W} = W(t + \frac{dt}{2}) - W(t)$. Now we substitute this back into the Stratonovich SDE definition:

$$dx = a(x)dt + \left(b(x) + b'(x)\left[a(x)\frac{dt}{2} + b(x)d\hat{W}\right]\right)dW_t$$

Expanding this, we get three terms involving dW_t:

$$dx = a(x)dt + b(x)dW_t + a(x)b'(x)\frac{dt}{2}dW_t + b(x)b'(x)d\hat{W}dW_t$$

The term $dt\,dW_t$ is of order $O(dt^{3/2})$ and can be neglected. The final term requires us to evaluate the expectation $\langle d\hat{W}dW_t\rangle$.

$$\langle d\hat{W}dW_t\rangle = \left\langle \left(W(t+\frac{dt}{2})-W(t)\right)(W(t+dt)-W(t))\right\rangle = \min\left(t+\frac{dt}{2},t+dt\right)-t = \frac{dt}{2}$$

where the minimum comes from the autocorrelation of the Wiener process $\langle W(t)W(s)\rangle = \min(t,s)$.

Therefore, the term $b(x)b'(x)d\hat{W}dW_t$ contributes a drift of $\frac{1}{2}b(x)b'(x)dt$. Combining all terms, the Stratonovich SDE is equivalent to the following Itô SDE:

$$dx = \left(a(x)+\frac{1}{2}b(x)b'(x)\right)dt + b(x)dW_t$$

Definition: ***Stratonovich-to-Itô Conversion***

A Stratonovich SDE

$$dx = a(x)dt + b(x)\circ dW_t$$

is equivalent to an Itô SDE with a modified drift term:

$$dx = \left(a(x)+\frac{1}{2}b(x)b'(x)\right)dt + b(x)dW_t$$

This conversion formula allows us to switch between the two calculi, leveraging the strengths of each. The extra drift term $\frac{1}{2}b(x)b'(x)$ is often called the "noise-induced drift" or "spurious drift."

4.5.4 Fokker-Planck Equation for Stratonovich SDEs

Given the conversion formula, we can easily find the Fokker-Planck equation for a Stratonovich SDE. We simply take the Fokker-Planck equation for an Itô SDE and replace the drift $a(x)$ with the effective drift $a(x)+\frac{1}{2}b(x)b'(x)$.

$$\frac{\partial\rho}{\partial t} = -\frac{\partial}{\partial x}\left[\left(a(x)+\frac{1}{2}b(x)b'(x)\right)\rho\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[b(x)^2\rho\right]$$

This equation can be rewritten in a more compact and elegant form. Noting that $b(x)b'(x) = \frac{1}{2}\frac{d}{dx}(b(x)^2)$, we can combine the terms:

$$\frac{\partial\rho}{\partial t} = -\frac{\partial}{\partial x}[a(x)\rho] + \frac{1}{2}\left(-\frac{\partial}{\partial x}\left[\frac{d(b^2)}{dx}\rho\right] + \frac{\partial^2}{\partial x^2}[b^2\rho]\right)$$

Let us use the product rule on the second derivative:

$$\frac{\partial^2}{\partial x^2}[b^2\rho] = \frac{\partial}{\partial x}\left(\frac{\partial(b^2)}{\partial x}\rho + b^2\frac{\partial\rho}{\partial x}\right) = \frac{\partial^2(b^2)}{\partial x^2}\rho + 2\frac{\partial(b^2)}{\partial x}\frac{\partial\rho}{\partial x} + b^2\frac{\partial^2\rho}{\partial x^2}$$

Now, consider the term

$$-\frac{\partial}{\partial x}\left[\frac{\partial(b^2)}{\partial x}\rho\right] + \frac{\partial^2}{\partial x^2}[b^2\rho]$$

Substituting the expanded form, we get:

$$\frac{\partial \rho}{\partial t} = -\left(\frac{\partial^2(b^2)}{\partial x^2}\rho + \frac{\partial(b^2)}{\partial x}\frac{\partial \rho}{\partial x}\right) + \left(\frac{\partial^2(b^2)}{\partial x^2}\rho + 2\frac{\partial(b^2)}{\partial x}\frac{\partial \rho}{\partial x} + b^2\frac{\partial^2 \rho}{\partial x^2}\right)$$

Canceling the common terms, we are left with:

$$\frac{\partial \rho}{\partial t} = \frac{\partial(b^2)}{\partial x}\frac{\partial \rho}{\partial x} + b^2\frac{\partial^2 \rho}{\partial x^2} = \frac{\partial}{\partial x}\left[b\frac{\partial}{\partial x}(b\rho)\right]$$

Therefore, the equation becomes:

$$\frac{\partial \rho}{\partial t} = -\frac{\partial}{\partial x}[a(x)\rho] + \frac{1}{2}\frac{\partial}{\partial x}\left[b(x)\frac{\partial}{\partial x}(b(x)\rho)\right]$$

This is the common form of the Fokker-Planck equation found in many physics textbooks. It highlights that in the Stratonovich interpretation, the diffusion term has a simpler structure.

The Stationary Distribution for Stratonovich SDEs

The Stratonovich form turns out to be quite convenient when dealing with the stationary distribution P_s, since

$$0 = -\frac{d}{dx}[a(x)P_s] + \frac{1}{2}\frac{d}{dx}\left[b(x)\frac{d}{dx}(b(x)P_s)\right]$$

By simplifying and rearranging, we easily obtain

$$\frac{b(x)}{2}\frac{d}{dx}[b(x)P_s] = a(x)P_s$$

Now, define $\mathscr{H} = b(x)P_s$ so that

$$\frac{d\mathscr{H}}{dx} = \frac{2a(x)}{b(x)^2}\mathscr{H} \implies \mathscr{H} = C\exp\left[\int_x \frac{2a(z)}{b(z)^2}dz\right]$$

Therefore,

$$P_s(x) = \frac{C}{b(x)}\exp\left[\int_x \frac{2a(z)}{b(z)^2}dz\right] = C\exp\left[-\log(b(x)) + \int_x \frac{2a(z)}{b(z)^2}dz\right]$$

Once again, we want a formula that allows us to calculate the stationary points, so we can verify that it is sufficient to consider the stationary points of the exponent:

$$\frac{2a(x)}{b(x)^2} - \frac{b'(x)}{b(x)} \geq 0 \implies a(x) \geq \frac{1}{2}b(x)b'(x)$$

The only difference compared to the Itô formulation is the factor $1/2$. The difference is therefore only quantitative, not qualitative. We have seen that the transitions due to noise can emerge in a new way that depends on the parameterization, even if the positions of the maxima and minima and their values change by different factors.

Applications of SSDEs

Let us revisit the Malthusian growth model, which is described by

$$\dot{x} = rx$$

where r is the growth rate.
Suppose now that the growth rate is subject to random fluctuations, so that $r \to r + \alpha\xi(t)$, where $\xi(t)$ is white noise.
Now, let us assume that the noise is such that it is described in the Stratonovich sense. The equation becomes

$$dx = rxdt + \alpha x \circ dW$$

where $\circ$ denotes the Stratonovich integral. The corresponding Ito version of this formula is

$$dx = \left(rx + \frac{\alpha^2}{2}x\right)dt + \alpha x dW$$

If we now apply the logarithmic transformation $y = \ln x$, we obtain

$$dy = \left(r + \frac{\alpha^2}{2} - \frac{\alpha^2}{2}\right)dt + \alpha dW = rdt + \alpha dW$$

However, if we instead use the Ito interpretation from the start, the equation for x is

$$dx = rxdt + \alpha x dW$$

which, under the transformation $y = \ln x$, gives

$$dy = \left(r - \frac{\alpha^2}{2}\right)dt + \alpha dW$$

This leads to a crucial difference between the two interpretations.

- **Itô case:** If $\frac{\alpha^2}{2} > r$, then the drift term in the equation for $y = \ln x$ becomes negative, so $y(t) \to -\infty$ as $t \to \infty$. This means $x(t) \to 0$, and the population will eventually go extinct even if $r > 0$.
- **Stratonovich case:** The noise-induced drift term cancels out, so extinction only occurs if $r < 0$. In other words, the population persists as long as the deterministic growth rate r is positive, regardless of the noise strength.

This difference is significant: in the Itô interpretation, strong enough noise can drive the population extinct even when the average growth rate is positive, while in the Stratonovich interpretation, only a negative growth rate leads to extinction.
Which interpretation should be used? The answer depends on the nature of the noise and the modeling context. The Itô calculus is appropriate when the noise is idealized as perfectly white and uncorrelated in time. The Stratonovich calculus is more suitable when the noise has a finite correlation time or comes from a smooth, physical process. In real-world applications, the correct choice depends on the details of the system being modeled.
In summary, the choice between Itô and Stratonovich calculus is not just a technicality: it can fundamentally change the predicted behavior of stochastic models, especially for questions like extinction thresholds and long-term outcomes.

5 Stochastic Systems, Applications

5.1 The Ornstein-Uhlenbeck Process

We now consider one of the most important stochastic processes in science and engineering: the **Ornstein-Uhlenbeck (OU) process**. It appears in numerous contexts, from describing the velocity of a particle in a fluid (the original Langevin model) to modeling the voltage in a noisy electronic circuit. In essence, it is the canonical model for any noisy, linear system that tends to return to an equilibrium state.

The general form of the OU process is given by the SDE:

$$dz = -\gamma z\,dt + \omega\,dW$$

where $\gamma > 0$ is the *drift* or *relaxation rate*, which constantly pulls the process back towards its mean (in this case, zero), and ω is the noise intensity. This linear restoring force is what distinguishes the OU process from the free-wandering Wiener process.

Solving the Ornstein-Uhlenbeck SDE

To solve this linear SDE, we can use a technique analogous to the integrating factor method for ODEs. We define a new variable $Q(t) = e^{\gamma t} z(t)$. Our goal is to find an SDE for $Q(t)$. Using Ito's product rule (or Ito's lemma on $\psi(z,t) = e^{\gamma t} z$), we find:

$$\begin{aligned} dQ &= (d(e^{\gamma t}))z + e^{\gamma t}(dz) \\ &= (\gamma e^{\gamma t} dt)z + e^{\gamma t}(-\gamma z\,dt + \omega dW(\tau)) \\ &= \gamma e^{\gamma t} z\,dt - \gamma e^{\gamma t} z\,dt + \omega e^{\gamma t} dW(\tau) \\ &= \omega e^{\gamma t} dW(\tau) \end{aligned}$$

The drift terms have canceled perfectly, leaving a very simple SDE for $Q(t)$. We can now integrate it from 0 to t:

$$Q(t) - Q(0) = \omega \int_0^t e^{\gamma\tau} dW(\tau) \quad \Longrightarrow \quad Q(t) = Q(0) + \omega \int_0^t e^{\gamma\tau} dW(\tau)$$

Substituting back $z(t) = e^{-\gamma t} Q(t)$ and $z(0) = Q(0)$, we arrive at the solution for the Ornstein-Uhlenbeck process:

$$z(t) = z(0)e^{-\gamma t} + \omega \int_0^t e^{\gamma(\tau - t)} dW(\tau)$$

The solution consists of two parts: a deterministic decay of the initial condition, $z(0)e^{-\gamma t}$, and a stochastic integral that represents the accumulated effect of the noise, with past noise contributions being exponentially "forgotten" over time due to the $e^{\gamma(\tau-t)}$ term.

Statistical Properties

From the solution, we can derive the key statistical properties of the OU process.

- **Mean**
  Since the stochastic integral $\int e^{\gamma(\tau-t)}dW(\tau)$ has zero mean, the mean of the process is simply the deterministic part:

$$\boxed{\langle z(t)\rangle = \langle z(0)\rangle e^{-\gamma t}}$$

  Assuming a deterministic starting point $z(0)$, the mean exponentially decays to zero, which is the equilibrium point of the system.
- **Variance**
  The variance calculation is more involved but highly instructive. Starting from our solution:

$$z(t) = z(0)e^{-\gamma t} + \omega \int_0^t e^{\gamma(\tau-t)}dW(\tau)$$

  we can calculate the variance by expanding $\mathrm{Var}[z(t)] = \langle z(t)^2\rangle - \langle z(t)\rangle^2$. Since $\langle z(t)\rangle = z(0)e^{-\gamma t}$ (assuming deterministic initial condition), we have:

$$\mathrm{Var}[z(t)] = \left\langle \left(z(0)e^{-\gamma t} + \omega \int_0^t e^{\gamma(\tau-t)}dW(\tau) \right)^2 \right\rangle - z(0)^2 e^{-2\gamma t}$$

  Expanding the square:

$$\mathrm{Var}[z(t)] = \left\langle z(0)^2 e^{-2\gamma t} + 2z(0)e^{-\gamma t}\omega \int_0^t e^{\gamma(\tau-t)}dW(\tau) + \omega^2 \left(\int_0^t e^{\gamma(\tau-t)}dW(\tau) \right)^2 \right\rangle - z(0)^2 e^{-2\gamma t}$$

  The cross term vanishes since $\langle dW(\tau)\rangle = 0$, leaving:

$$\mathrm{Var}[z(t)] = \omega^2 \left\langle \left(\int_0^t e^{\gamma(\tau-t)}dW(\tau) \right)^2 \right\rangle$$

  Applying the mean, we get:

$$\mathrm{Var}[z(t)] = \omega^2 \left\langle \left(\int_0^t e^{\gamma(\tau-t)}dW(\tau) \right)^2 \right\rangle = \omega^2 \int_0^t \int_0^t e^{\gamma(\tau-t)} e^{\gamma(s-t)} \langle dW(\tau)dW(s)\rangle$$

  Using the property that $\langle dW(\tau)dW(s)\rangle = \delta(\tau - s)d\tau ds$, the double integral becomes:

$$\mathrm{Var}[z(t)] = \omega^2 \int_0^t \int_0^t e^{\gamma(\tau+s-2t)} \delta(s-\tau)\, ds\, d\tau = \omega^2 \int_0^t e^{2\gamma(\tau-t)} d\tau$$

  Evaluating this integral:

$$\mathrm{Var}[z(t)] = \omega^2 e^{-2\gamma t} \int_0^t e^{2\gamma\tau} d\tau = \omega^2 e^{-2\gamma t} \left[\frac{e^{2\gamma\tau}}{2\gamma} \right]_0^t = \frac{\omega^2}{2\gamma}(1 - e^{-2\gamma t})$$

  The total variance becomes:

$$\boxed{\mathrm{Var}[z(t)] = \mathrm{Var}[z(0)]e^{-2\gamma t} + \frac{\omega^2}{2\gamma}(1 - e^{-2\gamma t})}$$

  As $t \to \infty$, the first term vanishes, and the variance approaches a constant stationary value:

$$\lim_{t\to\infty} \mathrm{Var}[z(t)] = \frac{\omega^2}{2\gamma}$$

  This is a crucial difference from the Wiener process, whose variance grows linearly and without bound. The restoring force in the OU process balances the diffusive effect of the noise, leading to a finite stationary variance.

- **Autocovariance**
  The autocovariance function for times t and q is defined as:

$$C(t,q) = \langle (z(t) - \langle z(t) \rangle)(z(q) - \langle z(q) \rangle) \rangle$$

  For simplicity, let's assume the process starts at $z(0) = 0$. In this case, the mean is $\langle z(t) \rangle = 0$ for all t, so the autocovariance is simply the autocorrelation: $C(t,q) = \langle z(t)z(q) \rangle$.
  Using the solution for $z(t)$ with $z(0) = 0$:

$$z(t) = \omega \int_0^t e^{\gamma(s-t)} dW(s)$$

  The autocorrelation is:

$$\begin{aligned} \langle z(t)z(q) \rangle &= \left\langle \left(\omega \int_0^t e^{\gamma(s-t)} dW(s) \right) \left(\omega \int_0^q e^{\gamma(\tau-q)} dW(\tau) \right) \right\rangle \\ &= \omega^2 e^{-\gamma(t+q)} \left\langle \int_0^t e^{\gamma s} dW(s) \int_0^q e^{\gamma \tau} dW(\tau) \right\rangle \\ &= \omega^2 e^{-\gamma(t+q)} \int_0^t \int_0^q e^{\gamma(s+\tau)} \langle dW(s) dW(\tau) \rangle \end{aligned}$$

  Using the property $\langle dW(s)dW(\tau) \rangle = \delta(s-\tau) ds\, d\tau$, the double integral collapses into a single integral. We must evaluate the integral over the domain $[0,t] \times [0,q]$.

$$\langle z(t)z(q) \rangle = \omega^2 e^{-\gamma(t+q)} \int_0^t \int_0^q e^{\gamma(\tau+s)} \delta(\tau - s) d\tau ds$$

  For the evaluation of this integral, we need to consider two cases:
  - **Case 1:** $t \leq q$
    When $t \leq q$, the domain of integration is restricted by the smaller upper limit. The delta function $\delta(\tau - s)$ forces $\tau = s$, so we integrate over the overlap region $[0,t]$:

$$\langle z(t)z(q) \rangle = \omega^2 e^{-\gamma(t+q)} \int_0^t e^{2\gamma s} ds$$

    Evaluating the integral:

$$\langle z(t)z(q) \rangle = \omega^2 e^{-\gamma(t+q)} \left[\frac{e^{2\gamma s}}{2\gamma} \right]_0^t = \frac{\omega^2}{2\gamma} e^{-\gamma(t+q)} (e^{2\gamma t} - 1) = \frac{\omega^2}{2\gamma} \left(e^{-\gamma(q-t)} - e^{-\gamma(q+t)} \right)$$

  - **Case 2:** $t > q$
    When $t > q$, we integrate over $[0,q]$:

$$\langle z(t)z(q) \rangle = \omega^2 e^{-\gamma(t+q)} \int_0^q e^{2\gamma s} ds = \frac{\omega^2}{2\gamma} \left(e^{-\gamma(t-q)} - e^{-\gamma(t+q)} \right)$$

  Combining both cases, we can write the autocorrelation function as:

$$\boxed{\langle z(t)z(q) \rangle = \frac{\omega^2}{2\gamma} \left(e^{-\gamma|t-q|} - e^{-\gamma(t+q)} \right)}$$

  For a stationary process (when $t, q \to \infty$), the second term vanishes, and we obtain the stationary autocorrelation function:

$$R_z(\tau) = \lim_{t \to \infty} \langle z(t)z(t+\tau) \rangle = \frac{\omega^2}{2\gamma} e^{-\gamma|\tau|}$$

  where $\tau = |t - q|$ is the time lag. This autocorrelation decays exponentially with a characteristic time $1/\gamma$. In the limit of rapid relaxation ($\gamma \to \infty$), the autocorrelation function becomes sharply peaked and tends toward a Dirac delta function.

We introduce now a scaling relation between the noise intensity and the relaxation rate, by setting:

$$\omega = c\,\gamma$$

with c a constant. Under this scaling, the SDE for z becomes

$$\dot{z} = -\gamma z + c\,\gamma\,\xi(t).$$

The autocorrelation function for z then reads

$$R_z(h) = \frac{c^2}{2}\,\gamma e^{-\gamma|h|}, \qquad R_z(0) = \frac{c^2}{2}\,\gamma.$$

Its total area is given by

$$\int_{-\infty}^{+\infty} R_z(h)\,dh = \int_{-\infty}^{+\infty} \frac{c^2}{2}\,\gamma e^{-\gamma|h|}\,dh. = \frac{c^2}{2}\,\gamma \cdot \frac{2}{\gamma} = c^2.$$

For small characteristic times ($\tau = 1/\gamma$ small), the autocorrelation function $R_z(h)$ approximates a white noise process:

$$\lim_{\gamma\to\infty} R_z(h;\gamma) = c^2\,\delta(h).$$

This derivation shows how, by choosing the scaling $\omega = c\,\gamma$, the fluctuations in the linearized dynamics around a stable equilibrium effectively become white noise in the fast relaxation limit.

5.1.1 The OU Process as a Low-Pass Filter

One of the most important interpretations of the Ornstein-Uhlenbeck process is as a **low-pass filter**. It takes a "noisy" input signal (white noise) and produces a smoother, more physically realistic output by attenuating high-frequency fluctuations. To understand this, we can analyze the process in the frequency domain using the Fourier transform.

The Fourier Transform and Linear Systems

Let $f(t)$ be a well-behaved function. Its Fourier transform, $\hat{f}(\omega)$, is defined as:

$$\mathscr{F}[f(t)] = \hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-i\omega t}dt$$

A key property of the Fourier transform, which can be shown using integration by parts, is how it acts on derivatives:

$$\mathscr{F}[f'(t)] = i\omega\mathscr{F}[f(t)] = i\omega\hat{f}(\omega)$$

This property turns differentiation in the time domain into multiplication in the frequency domain, which is extremely useful for solving linear differential equations. For example, consider a general first-order linear system with input $y(t)$ and output $f(t)$:

$$f'(t) + \alpha f(t) = y(t)$$

Taking the Fourier transform of the entire equation gives:

$$i\omega\hat{f}(\omega) + \alpha\hat{f}(\omega) = \hat{y}(\omega) \quad \Longrightarrow \quad \hat{f}(\omega) = \frac{\hat{y}(\omega)}{i\omega + \alpha}$$

The term $H(\omega) = 1/(i\omega + \alpha)$ is called the **transfer function** of the system. The power of the output signal is related to the power of the input signal by the squared magnitude of the transfer function:

$$|\hat{f}(\omega)|^2 = \frac{|\hat{y}(\omega)|^2}{\omega^2 + \alpha^2}$$

Power Spectrum of the OU Process

We can apply this framework to the OU process, which is described by the linear SDE:

$$\frac{dz}{dt} + \gamma z = \kappa \xi(t)$$

Taking the Fourier transform, we get:

$$(i\omega + \gamma)\hat{z}(\omega) = \kappa \mathscr{F}[\xi(t)] \quad \Longrightarrow \quad \hat{z}(\omega) = \frac{\kappa}{i\omega + \gamma}\mathscr{F}[\xi(t)]$$

To analyze the filtering effect, we compare the **power spectral density** of the input signal, $\xi(t)$, with that of the output signal, $z(t)$. The power spectrum is the Fourier transform of the autocorrelation function.

For the input signal (white noise), the autocorrelation is $R_\xi(h) = \delta(h)$ (assuming unit variance for simplicity). Its power spectrum is therefore constant:

$$\mathscr{F}[R_\xi(h)] = \mathscr{F}[\delta(h)] = \int_{-\infty}^{+\infty} \delta(h) e^{-i\omega h} dh = 1$$

This flat spectrum is why it's called "white" noise: it contains equal power at all frequencies ω.

For the output signal, $z(t)$, we previously found its stationary autocorrelation function to be:

$$R_z(h) = \frac{\kappa^2}{2\gamma} e^{-\gamma|h|}$$

The power spectrum of the output is the Fourier transform of this function:

$$\mathscr{F}[R_z(h)] = \frac{\kappa^2}{2\gamma} \int_{-\infty}^{+\infty} e^{-\gamma|h|} e^{-i\omega h} dh$$

We can split the integral over positive and negative values of h:

$$= \int_{-\infty}^{0} e^{(\gamma - i\omega)h} dh + \int_{0}^{+\infty} e^{-(\gamma + i\omega)h} dh = \frac{1}{\gamma - i\omega} + \frac{1}{\gamma + i\omega} = \frac{2\gamma}{\gamma^2 + \omega^2}$$

Substituting this back, we get the output power spectrum:

$$\mathscr{F}[R_z(h)] = \frac{\kappa^2}{2\gamma}\left(\frac{2\gamma}{\gamma^2 + \omega^2}\right) = \frac{\kappa^2}{\gamma^2 + \omega^2}$$

This function, known as a **Lorentzian**, is peaked at $\omega = 0$ and decays as $1/\omega^2$ for high frequencies. The OU process acts as a low-pass filter: it suppresses high-frequency components of the input white noise while allowing low-frequency components to pass through, resulting in a smoother, more physically plausible random process.

Input Spectrum (White Noise) Output Spectrum (OU Process)

Figure 5.1: Power spectra: input white noise (flat, left) and output OU process (Lorentzian, right).

5.1.2 Fokker-Planck equation for Ornstein-Uhlenbeck

Given the Ito equation for the OU process:

$$dz = -\gamma z dt + \omega dW$$

we can find the Fokker-Planck equation. Recalling the general form of the equation for $P(z,t)$:

$$\partial_t P = -\partial_z[a(z)P(z,t)] + \frac{1}{2}\partial_z^2\left[b(z)^2 P(z,t)\right]$$

In our case, with $a(z) = -\gamma z$ and $b(z) = \omega$, it becomes:

$$\partial_t P = \partial_z(\gamma z P) + \frac{\omega^2}{2}\partial_z^2 P$$

This equation is not simple to solve, even with the boundary conditions for a probability density:

$$P(\pm\infty,t) = 0 \quad \text{and} \quad \int_{\mathbb{R}} P(z,t)dz = 1$$

These are additional conditions compared to standard PDEs and are usually of great help.
The most tractable approach is to find the stationary solution $P_s(z)$, which is the long-time limit of $P(z,t)$:

$$P_s(z) = \lim_{t\to\infty} P(z,t)$$

In the stationary state, $\partial_t P = 0$, so the equation becomes an ODE:

$$0 = \frac{d}{dz}(\gamma z P_s) + \frac{\omega^2}{2}\frac{d^2 P_s}{dz^2}$$

which we can rearrange to:

$$\frac{d^2 P_s}{dz^2} = -\frac{2\gamma}{\omega^2}\frac{d}{dz}(zP_s)$$

Integrating once with respect to z, we obtain:

$$\frac{dP_s}{dz} = C - \frac{2\gamma}{\omega^2} z P_s(z)$$

The boundary conditions require that $P_s(z) \to 0$ and consequently $P_s'(z) \to 0$ as $z \to \pm\infty$. For this to hold, the integration constant C must be zero. This leaves us with a first-order linear ODE:

$$\frac{dP_s}{dz} = -\frac{2\gamma}{\omega^2} z P_s(z)$$

The solution is a Gaussian function, which is consistent with the properties of the Ornstein-Uhlenbeck process:

$$P_s(z) = Ae^{-\frac{\gamma}{\omega^2}z^2}$$

This is nothing other than a Gaussian with $\mu = 0$ and $\sigma^2 = \omega^2/2\gamma$, with A a constant required for normalization, namely

$$A = \frac{1}{\sqrt{2\pi\sigma^2}}$$

If we now consider the unperturbed system $\dot{z} = -\gamma z$, we see that the solution tends to 0 for long times and, since it is deterministic, we obviously have

$$P_s^{\text{DET}}(z) = \delta(z)$$

which is consistent with the perturbed case. In fact, the only equilibrium point of the deterministic case is 0, which turns out to be the center of the Gaussian we found. Therefore, for long times, in the perturbed case, the solution will oscillate around 0.

5.2 Bistable Systems: The Ginzburg-Landau Model

We now move from the monostable Ornstein-Uhlenbeck process to a more complex and interesting class of systems: bistable systems. These are systems that possess two stable equilibrium states. The simplest and most iconic model for such a system is the **Ginzburg-Landau equation**.

The Deterministic Model

The deterministic Ginzburg-Landau model describes the dynamics of a particle in a "double-well" potential. The equation of motion is given by:

$$\dot{x} = x - x^3 = x(1-x)(1+x)$$

This system has three equilibrium points, found by setting $\dot{x} = 0$:

$$x_L = -1, \quad x_C = 0, \quad x_R = +1$$

A simple stability analysis reveals the nature of these points:

- $x_L = -1$ and $x_R = +1$ are **locally asymptotically stable**. If the system starts near one of these points, it will converge to it.
- $x_C = 0$ is an **unstable** equilibrium. If the system starts at this point, any infinitesimal perturbation will cause it to move away towards either x_L or x_R.

The force $F(x) = x - x^3$ can be derived from a potential $U(x)$ where $F = -dU/dx$:

$$U(x) = -\frac{x^2}{2} + \frac{x^4}{4}$$

This potential has a characteristic "double-well" shape, where the stable equilibria x_L and x_R correspond to the minima (the "wells") of the potential, and the unstable equilibrium $x_C = 0$ corresponds to the local maximum (the "barrier") between them.

If the initial condition $x(0)$ is drawn from a probability distribution $\theta(x)$, the long-term behavior of the system is deterministic. Any trajectory with $x(0) > 0$ will end up at $x_R = +1$, and any with $x(0) < 0$ will end up at $x_L = -1$. The stationary PDF will therefore consist of two Dirac delta peaks at the stable equilibria, with weights determined by the initial probability of being on either side of the unstable point:

$$\lim_{t\to\infty} \rho(x,t) = A_R \delta(x-1) + A_L \delta(x+1)$$

where $A_R = \Pr(x_0 > 0)$ and $A_L = \Pr(x_0 < 0)$.

Stochastic Perturbations of the Model

Now, let's add a stochastic white noise force to the system:

$$\dot{x} = x - x^3 + \omega\xi(t) \quad \text{or} \quad dx = (x - x^3)dt + \omega dW_t$$

This is a case of additive noise perturbing a conservative force. As we derived previously, the stationary probability distribution $P_s(x)$ for such a system is given by the Boltzmann distribution:

$$P_s(x) = C \exp\left(-\frac{2}{\omega^2} U(x)\right)$$

Substituting the Ginzburg-Landau potential $U(x) = -x^2/2 + x^4/4$, we get:

$$P_s(x) = C\exp\left(-\frac{2}{\omega^2}\left(-\frac{x^2}{2}+\frac{x^4}{4}\right)\right) = C\exp\left(\frac{x^2}{\omega^2}-\frac{x^4}{2\omega^2}\right)$$

where C is a normalization constant. The shape of this distribution is fundamentally controlled by the noise intensity ω.

The relationship between the potential $U(x)$ and the stationary PDF $P_s(x)$ is clear:

- The stable equilibria of the deterministic system (minima of $U(x)$) become the **modes** (peaks) of the stationary PDF.
- The unstable equilibrium (maximum of $U(x)$) becomes the **antimode** (trough) of the stationary PDF.

The noise intensity ω determines how closely the system is confined to these potential wells:

- **Weak Noise ($\omega \ll 1$):** The distribution $P_s(x)$ is sharply bimodal, with two distinct peaks centered around $x = -1$ and $x = +1$. The probability of finding the particle near the unstable point $x = 0$ is extremely low. The system will exhibit small stochastic fluctuations around one of the two stable states for very long periods.
- **Strong Noise ($\omega \gg 1$):** The noise provides enough energy for the particle to easily cross the potential barrier between the two wells. The two peaks in the PDF broaden and merge, and the distribution becomes unimodal, centered at $x = 0$. For very large x, the x^4 term still dominates, so the probability $P_s(x)$ must decay to zero.

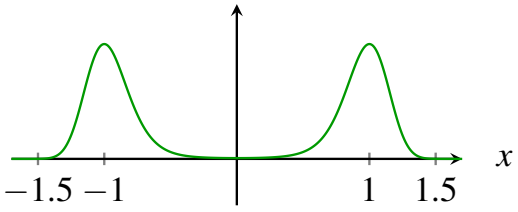


Figure 5.2: PDF (non normalized) for small ω

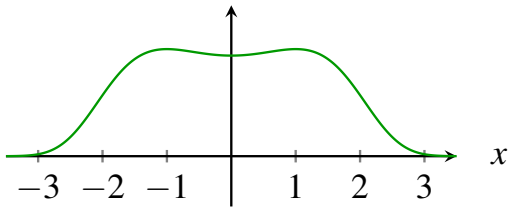


Figure 5.3: PDF (non normalized) for large ω

This ”inheritance” of stability properties from the deterministic potential to the modes of the stationary PDF is a key feature of systems with additive noise. As we will see, this direct correspondence breaks down in the presence of multiplicative noise.

Observation: ***Stochastic Miultistability***

It is possible to say that after a very long period of time, also the first case we could see that the system moves from one peak to the other. This is due to the fact that the system is not stable and we have a non-zero probability of moving from one peak to the other.

5.3 Bounded Noises

6 Missing notes

Warning: *Missing notes*

Here there should be the notes about:
- Bounded Noises - Euler-Maruyama correction for the Bounded Noises - NIT - Sparse examples

7 Introduction to Markov Chains

7.1 From Discrete Jumps to Continuous Diffusion

In the previous sections, we modeled stochastic systems using Stochastic Differential Equations (SDEs), which describe the infinitesimal evolution of processes with continuous paths, like the Ornstein-Uhlenbeck process. However, many systems in physics, chemistry, and biology are more naturally described by discrete "jumps" between states. Examples include a molecule switching between chemical species, an ion channel opening or closing, or a population changing by single-birth or death events.

These phenomena are governed by **jump processes**, where the system remains in a given state for a random period before instantaneously jumping to a new state. The evolution of the probability distribution for such a process is not described by a Fokker-Planck equation but by a more general framework: the **Master Equation**.

The Master Equation

Let's consider a process where the state variable x can take any value in a continuous state space. The core of a jump process is the **transition rate** (or jump kernel) $\Omega(y,x)$, which defines the probability per unit time of a jump occurring *from* state y *to* state x.

The probability of a system in state y jumping to any state within a small region around x in an infinitesimal time interval dt is given by:

$$\text{Prob}(\text{jump } y \to x \text{ in } dt) = \Omega(y,x)dt$$

Conversely, the probability that the system, currently in state x, does ***not*** jump to any other state during dt is one minus the total probability of jumping *out* of x:

$$\text{Prob}(\text{no jump from } x \text{ in } dt) = 1 - \left(\int \Omega(x,s)ds \right) dt$$

The Master Equation describes the time evolution of the probability density $P(x,t)$ by balancing the probability flows into and out of state x. The rate of change of $P(x,t)$ is the sum of all probability flowing *in* from other states s (the gain term) minus the total probability flowing *out* to all other states s (the loss term).

This balance gives the integro-differential Master Equation:

$$\boxed{\frac{\partial P(x,t)}{\partial t} = \underbrace{\int P(s,t)\Omega(s,x)ds}_{\text{Gain: Jumps from } s \text{ to } x} - \underbrace{P(x,t)\int \Omega(x,s)ds}_{\text{Loss: Jumps from } x \text{ to } s}}$$

Deriving the Diffusion Equation

A profound connection exists between discrete jump processes and continuous diffusion processes. Under the right conditions, a process consisting of many small, frequent jumps becomes indistinguishable from a continuous process described by a Fokker-Planck equation. We can demonstrate this by considering two simple jump models.

Example: ***The Uniform Jump Process***

Consider a particle that performs random jumps where the jump destination is uniformly distributed within a small interval of length 2ε around its current position. If the particle is at s, it can jump to any $x \in [s-\varepsilon, s+\varepsilon]$. The jump kernel can be written as:

$$\Omega(s,x) = C \cdot \mathbb{I}_{[s-\varepsilon,s+\varepsilon]}(x)$$

where C is a constant representing the overall jump frequency and $\mathbb{I}$ is the indicator function, therefore $\mathbb{I}$ will be 1 in the set and 0 outside. Thus I have the Master Equation:

$$\partial_t P = -2\varepsilon C P(x,y) + C \int_{x-\varepsilon}^{x+\varepsilon} P(y,t)dy$$

Now define $y = z + x$, then

$$\int_{x-\varepsilon}^{x+\varepsilon} P(y,t)dy = \int_{-\varepsilon}^{\varepsilon} P(x+z,t)dz$$

Since ε is small, I can Taylor expand $P(x+z,t)$ around x:

$$P(x+z,t) = P(x,t) + z\partial_x P(x,t) + \frac{z^2}{2}\partial_x^2 P(x,t) + \dots$$

Therefore

$$\begin{aligned} C\int_{-\varepsilon}^{\varepsilon} P(x+z,t)dz &= C\left[2\varepsilon + 0\cdot\partial_x + \frac{1}{2}\frac{1}{3}2\varepsilon^3\partial_x^2\right]P(x,t) \\ &= C\left[2\varepsilon + \frac{1}{3}\varepsilon^3\partial_x^2\right]P(x,t) \end{aligned}$$

Putting everything together, we obtain

$$\partial_t P(x,t) = \frac{C\varepsilon^3}{3}\partial_x^2 P(x,t)$$

To have $\frac{C\varepsilon^3}{3} = O(1)$, we need $C \propto 1/\varepsilon^3$, which is actially a huge jump rate.

Note that this is the formula for the PDF associated with a Wiener process, so a random walk process as described is described by a Wiener process.

Example: *The Fixed-Step Random Walk*

Now consider a different model where jumps are of a fixed size $\pm\varepsilon$. The transition rate is described by Dirac delta functions:

$$\Omega(y,x) = C \cdot \delta(|x-y| - \varepsilon)$$

Therefore,

$$\partial_t P = \int \left(P(y,t)\Omega(y,x) - P(x,t)\Omega(x,y)\right) dy$$

As before,

$$\partial_t P = C\left[P(x-\varepsilon,t) + P(x+\varepsilon,t) - 2P(x,t)\right]$$

and thus

$$\partial_t P = C\left(P(x+\varepsilon,t) - P(x,t) + P(x-\varepsilon,t) - P(x,t)\right)$$

We can now divide above and below by ε^2 and take the incremental ratio as $\varepsilon \to 0$, obtaining

$$\partial_t P = \varepsilon^2 C \frac{P(x+\varepsilon,t) - 2P(x,t) + P(x-\varepsilon,t)}{\varepsilon^2} = C\varepsilon^2 \partial_x^2 P(x,t)$$

Even for this discrete case, we obtain again the equation for the PDF of a Wiener process. However, there is something subtle: if C is finite, the process does not diffuse in the limit $\varepsilon \to 0$. To have a finite rate of change of probability, the diffusion coefficient must be defined as

$$\lim_{\varepsilon \to 0} C\varepsilon^2 = O(1)$$

So we must have $C \propto 1/\varepsilon^2$.

7.1.1 The Diffusion Limit and the Wiener Process

In both examples, we found that a process of discrete jumps can be described by a diffusion equation in the limit of small jump sizes ($\varepsilon \to 0$). However, for the diffusion coefficient D to remain finite and non-zero in this limit, the jump rate C must scale appropriately:

- For uniform jumps: $D = C\varepsilon^3/3 \implies C \propto 1/\varepsilon^3$.
- For fixed-step jumps: $D = C\varepsilon^2 \implies C \propto 1/\varepsilon^2$.

In both cases, the jump rate must diverge as the jump size shrinks. This is the **diffusion limit**: a macroscopic diffusion process emerges from an infinite number of infinitesimally small, infinitely frequent jumps.

The resulting equation,

$$\frac{\partial P}{\partial t} = D\frac{\partial^2 P}{\partial x^2}$$

is precisely the Fokker-Planck equation for a scaled **Wiener process** described by the SDE $dx = \sqrt{2D}dW_t$. This provides a powerful microscopic justification for the SDE framework: continuous diffusion models can be viewed as the large-scale limit of underlying discrete jump phenomena.

7.1.2 The Discrete-Space Limit: The CTMC

Let us now consider the case where the state space remains discrete. Specifically, take a random walk on the integers, $x(t) \in \mathbb{Z}$, where transitions (jumps) can only occur to neighboring sites, i.e., from n to $n+1$ or $n-1$. Each jump occurs with rate r.
Let us specify the ***jumping rates*** for this process:

$$\Omega(s,x) = r\delta(|x-s|-1) = r\delta(x-(s+1)) + r\delta(x-(s-1))$$

This expression indicates that transitions are only allowed between neighboring sites, with each jump occurring at rate r.
Next, we can write down the corresponding ***Master equation***. Here, time is continuous ($t \in \mathbb{R}$), and the state variable x takes integer values ($x \in \mathbb{Z} \subset \mathbb{R}$):

$$\begin{aligned}\frac{\partial P(x,t)}{\partial t} &= \int P(s,t)\Omega(s,x)ds - P(x,t)\int \Omega(x,s)ds \\ &= \int P(s,t)r\delta(|x-s|-1)ds - 2rP(x,t)\end{aligned}$$

In this setup, the probability $P(s,t)$ contributes only for the two possible jumps to x from its immediate neighbors, $x+1$ and $x-1$.

$$= \underbrace{rP(x+1,t)}_{backward} + \underbrace{rP(x-1,t)}_{forward} - \underbrace{2rP(x,t)}_{from\ x\ to\ others}$$

This is the master equation for a **Continuous-Time Markov Chain (CTMC)** on the integers:

$$\frac{dP_n(t)}{dt} = r\left[P_{n-1}(t) + P_{n+1}(t) - 2P_n(t)\right]$$

This equation describes a process that evolves in continuous time but jumps between discrete states.

General Case: Multiple Jump Sizes

The nearest-neighbor random walk can be generalized to allow jumps of arbitrary length. Consider a process where time remains continuous but the state space is discrete, $x \in \mathbb{Z}$. We allow transitions between any two states a and s, where the jump size is $j = s-a$. The transition rate can be written as:

$$\Omega(a,s) = \sum_{j\in\mathbb{Z}} K_j \delta(s-a-j)$$

where K_j represents the rate of jumps of size j. This formulation encompasses a wide range of jump processes: choosing $K_1 = K_{-1} = r$ and $K_j = 0$ for $|j| \neq 1$ recovers the nearest-neighbor case. Substituting this general form into the Master equation yields:

$$\begin{aligned}\frac{\partial P(x,t)}{\partial t} &= \sum_{j\in\mathbb{Z}} \int P(s,t)K_s\delta(s-x-j)ds - P(x,t)\sum_{j\in\mathbb{Z}}\int K_{x-a}\delta(x-a-j)da \\ &= \sum_{j\in\mathbb{Z}} K_j P(x-j,t) - P(x,t)\sum_{j\in\mathbb{Z}} K_j\end{aligned}$$

This equation describes the evolution of the probability distribution when the system can make jumps of various sizes, each occurring at its characteristic rate K_j.

7.1.3 Application: The Stochastic SIR Epidemic Model

The deterministic SIR (Susceptible-Infected-Recovered) model describes the dynamics of an epidemic through the coupled differential equations:

$$\begin{cases} S' = -\beta \frac{I}{N} S \\ I' = \beta \frac{I}{N} S - \gamma I \end{cases}$$

where $S(t)$ is the number of susceptibles, $I(t)$ is the number of infected, β is the transmission rate, γ is the recovery rate, and N is the total population size. The number of recovered individuals is $R(t) = N - S(t) - I(t)$.

While this deterministic model provides valuable insights for large populations, it fails to capture the inherent randomness of infection and recovery events. For finite populations, especially smaller communities, stochastic effects can be significant and may lead to qualitatively different behavior, such as stochastic extinction of the epidemic.

Stochastic Formulation

A more appropriate model is to represent the epidemic as a continuous-time Markov chain, with the state space defined as:

$$x(t) = (S(t), I(t)) \in C \subseteq \mathbb{N}^2$$

where $C = \{(S,I) \in \mathbb{N}^2 \mid S + I \leq N\}$ represents the constraint that the total number of susceptible and infected individuals cannot exceed the population size.

In any infinitesimal time interval $(t, t+dt)$, only two types of events can occur:

1. **Recovery Event**: The transition is $(S,I) \to (S, I-1)$, with probability:
$$\text{Prob}[x(t+dt) = (S, I-1) \mid x(t) = (S,I)] = \gamma I \, dt$$
2. **Infection Event**: The transition is $(S,I) \to (S-1, I+1)$, with probability:
$$\text{Prob}[x(t+dt) = (S-1, I+1) \mid x(t) = (S,I)] = \beta \frac{I}{N} S \, dt$$

Transition Rates and Master Equation

The transition rate function $\Omega(y,x)$ describes the rate of jumping from state y to state x. For the SIR model, this takes the form:

$$\Omega(y,x) = \gamma I_y \delta \left(y - x - \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) + \beta \frac{I_y}{N} S_y \delta \left(y - x - \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right)$$

where $y = (S_y, I_y)$ and $x = (S_x, I_x)$. The first term represents recovery events (decreasing I by 1), while the second term represents infection events (decreasing S by 1 and increasing I by 1).

To derive the Master equation for state (S,I), we must consider all possible transitions *into* it:

- From $(S+1, I-1)$ via an infection event, occurring at rate $\beta \frac{(I-1)}{N}(S+1)$
- From $(S, I+1)$ via a recovery event, occurring at rate $\gamma(I+1)$

The outflow from state (S,I) occurs at total rate $\gamma I + \beta \frac{IS}{N}$.

The resulting Master equation is:

$$\frac{\partial P(S,I,t)}{\partial t} = \underbrace{\beta \frac{(I-1)}{N}(S+1)P(S+1, I-1, t)}_{\text{infection}} + \underbrace{\gamma(I+1)P(S, I+1, t)}_{\text{recovery}} - \underbrace{\left(\gamma I + \beta \frac{IS}{N} \right) P(S,I,t)}_{\text{outflow}}$$

This equation governs the time evolution of the probability distribution over all possible epidemic states and captures the stochastic nature of individual infection and recovery events.

7.2 From Continuous to Discrete Time: Intro to DTMCs

In a discrete-time Markov chain (DTMC), both ***time and state space are discrete***: time steps t are non-negative integers ($t \in \mathbb{N} \cup \{0\}$), and states $x(t)$ take values in a discrete set ($x(t) \in \mathbb{N} \subset \mathbb{Z}^n$).

The **Transition probability** is the probability of jumping from a source state σ at time t to a target state α at time $t+1$ is defined as:

$$P\{x(t+1) = \alpha \mid x(t) = \sigma\} = \theta_{\sigma\alpha} \in [0,1]$$

The evolution of the probability of being in state α at time $t+1$ is given by:

$$P_\alpha(t+1) = \sum_\sigma P_\sigma(t)\theta_{\sigma\alpha}$$

where $P_\sigma(t)$ is the probability of being in state σ at time t. This can be rewritten by separating the contribution from staying in the same state ($\sigma = \alpha$) and from transitions from other states ($\sigma \neq \alpha$):

$$P_\alpha(t+1) = P_\alpha(t)\theta_{\alpha\alpha} + \sum_{\sigma\neq\alpha} P_\sigma(t)\theta_{\sigma\alpha}$$

Recall that the probability of remaining in state α is:

$$\theta_{\alpha\alpha} = 1 - \sum_{\beta\neq\alpha} \theta_{\alpha\beta}$$

since the total probability of leaving state α to any other state $\beta \neq \alpha$ is $\sum_{\beta\neq\alpha}\theta_{\alpha\beta}$.
Substituting this into the previous equation, we obtain the **Master equation** for a DTMC:

$$P_\alpha(t+1) - P_\alpha(t) = \sum_{\sigma\neq\alpha} P_\sigma(t)\theta_{\sigma\alpha} - \sum_{\beta\neq\alpha} P_\alpha(t)\theta_{\alpha\beta}$$

This equation describes the net change in probability for state α: the inflow from all other states minus the outflow from α to all other states.

7.2.1 Links between DTMCs and CTMCs

The Master equation for a DTMC provides insight into the relationship with continuous-time Markov chains (CTMCs). To see this, consider rescaling time by a small interval u:

$$\frac{P_\alpha(t+u) - P_\alpha(t)}{u} = \frac{1}{u}\sum_{\sigma\neq\alpha} P_\sigma(t)\theta_{\sigma\alpha} - \frac{1}{u}\sum_{\beta\neq\alpha} P_\alpha(t)\theta_{\alpha\beta}$$

In the limit as $u \to 0^+$, the left-hand side approaches the time derivative:

$$\frac{d}{dt}P_\alpha(t) = \lim_{u\to 0^+} \frac{P_\alpha(t+u) - P_\alpha(t)}{u}$$

For this limit to make sense, the transition probabilities must scale linearly with u:

$$\theta_{\sigma\alpha} = \Omega(\sigma,\alpha)u + O(u^2)$$

where $\Omega(\sigma,\alpha)$ is the transition rate from σ to α in the CTMC, and $u = dt$.
Thus, in the infinitesimal limit,

$$\theta_{\sigma\alpha} \approx \Omega(\sigma,\alpha)dt$$

This demonstrates two key points:

1. The probability of a transition in an infinitesimal interval $(t, t+dt)$ is proportional to dt, consistent with the definition of transition rates in CTMCs.
2. There is a direct and rigorous connection between the discrete-time and continuous-time Markov chain formulations.

7.2.2 Example: SIR Model with Vital Dynamics

Consider the SIR model with vital dynamics (birth & death), described by the system:

$$\begin{cases} S' = \mu(1-S) - \beta \frac{I}{N} S \\ I' = \beta \frac{I}{N} S - (\gamma + \mu) I \end{cases}$$

where $R(t) = N - S(t) - I(t)$. Here, we assume S and I are large enough to approximate the underlying CTMC by this deterministic system.

To incorporate stochasticity in the birth/death and contact rates, we let these parameters fluctuate:

$$\begin{cases} \mu \to \mu + \omega_\mu \xi_\mu(t) \\ \beta \to \beta + \omega_\beta \xi_\beta(t) \end{cases}$$

where $\xi_\mu(t)$ and $\xi_\beta(t)$ are white noise processes. Substituting these into the original system yields:

$$\begin{cases} \dot{S} = \mu(1-S) - \beta \frac{I}{N} S + (1-S)\omega_\mu \xi_\mu(t) - \beta \frac{I}{N} S \omega_\beta \xi_\beta(t) \\ \dot{I} = \beta \frac{I}{N} S - (\gamma + \mu) I + \beta \frac{I}{N} S \omega_\beta \xi_\beta(t) - \mu I \omega_\mu \xi_\mu(t) \end{cases}$$

To clarify the structure, define the state vector $x = \begin{pmatrix} S \\ I \end{pmatrix}$, the deterministic drift

$$f(x) = \begin{pmatrix} \mu(1-S) - \beta \frac{I}{N} S \\ \beta \frac{I}{N} S - (\gamma + \mu) I \end{pmatrix}$$

and the noise coefficient matrix

$$B(x) = \begin{pmatrix} (1-S)\omega_\mu & -\frac{\beta I}{N} S \omega_\beta \\ -I\omega_\mu & \frac{\beta I}{N} S \omega_\beta \end{pmatrix}$$

With these, the system can be compactly written as the 2D Itô SDE:

$$dx = f(x)dt + B(x)dW, \qquad \text{where } dW = \begin{pmatrix} dW_\mu \\ dW_\beta \end{pmatrix}$$

For simulation, discretize the time interval $[0,T]$ into N sub-intervals of length $h = T/N$, and denote $x_j = x(jh)$. The Euler-Maruyama scheme gives:

$$x_{j+1} = x_j + h f(x_j) + \sqrt{h} B(x_j) \begin{pmatrix} C_j \omega_\beta \\ G_j \omega_\mu \end{pmatrix}$$

where $C_j, G_j \sim \mathcal{N}(0,1)$ are independent standard normal random variables.

More generally, for $x \in \Omega \subset \mathbb{R}^n$, the Fokker-Planck equation associated to the Itô SDE

$$dx = a(x)dt + B(x)dW, \qquad \text{where } W \in \mathbb{R}^{n_W},\ B \in \mathbb{R}^{n \times n_W}$$

is

$$\begin{cases} \partial_t P(x,t) = -\text{div}(a(x)P(x,t)) + \frac{1}{2} \sum_{i,j} \partial^2_{x_i x_j} (C_{x_i x_j} P(x,t)) \\ \int_\Omega P(x,t)dx = 1 \end{cases}$$

where $C(x) = B(x)B(x)^T$ is the diffusion matrix. However, this equation is generally too complex to solve analytically.

8 Discrete-Time Markov Chains

We now transition from continuous-time stochastic processes to their discrete-time counterparts. While physical systems evolve continuously, many models are more naturally described at discrete time steps. This leads us to the study of **Discrete-Time Markov Chains (DTMCs)**.

8.1 Fundamental Properties

A DTMC is a stochastic process characterized by a discrete state space and the Markov property.

Definition: *Discrete-Time Markov Chain*

A stochastic process $\{X_t\}$, where time t is an integer ($t \in \mathbb{N}_0$), is a **DTMC** if it satisfies:

1. **Discrete State Space:** The process can only occupy states from a finite or countably infinite set $S = \{s_1, s_2, \ldots, s_N\}$.
2. **Markov Property:** The future state of the system depends only on its present state, not on the sequence of events that preceded it. Mathematically, we can write:

$$P(X_{t+1} = \sigma \mid X_t, X_{t-1}, \ldots, X_0) = P(X_{t+1} = \sigma \mid X_t)$$

This "memoryless" property is the cornerstone of Markov chain theory. It allows us to describe the entire system's dynamics using only the probabilities of one-step transitions.

Tip: *Is it really important that N is finite?*

No; in physics, it is common to work with infinite-dimensional (countable or even uncountable) state spaces and infinite matrices without difficulty.

The Evolution of the Probability Distribution and the Master Equation

Let $P_\sigma(t)$ be the probability that the system is in state $\sigma \in S$ at time t. To find the probability of being in state σ at the next time step, $t+1$, we can sum over all possible states δ the system could have been in at time t. Using the law of total probability:

$$P(X_{t+1} = \sigma) = \sum_{\delta \in S} P(X_{t+1} = \sigma \mid X_t = \delta) P(X_t = \delta)$$

Let's introduce more concise notation. Let $\theta_{\delta\sigma}(t) = P(X_{t+1} = \sigma \mid X_t = \delta)$ be the **transition probability** from state δ to state σ at time t. The evolution equation becomes:

$$P_\sigma(t+1) = \sum_{\delta \in S} P_\delta(t)\theta_{\delta\sigma}(t)$$

If we organize the state probabilities into a row vector $P(t) = [P_{s_1}(t), P_{s_2}(t), \ldots]$ and the transition probabilities into a matrix $\Theta(t)$, where $(\Theta)_{\delta\sigma} = \theta_{\delta\sigma}(t)$, we can write this system in a matrix form:

$$P(t+1) = P(t)\Theta(t)$$

We may rewrite the above equation as

$$P_\alpha(t+1) = P_\alpha(t)\theta_{\alpha\alpha} + \sum_{\sigma\neq\alpha} P_\sigma(t)\theta_{\sigma\alpha}$$

Since $\theta_{\alpha\alpha} = 1 - \sum_{\omega\neq\alpha}\theta_{\alpha\omega}$, we get the **Master Equation**:

$$\boxed{P_\alpha(t+1) = P_\alpha(t) + \sum_{\sigma\neq\alpha} P_\sigma(t)\theta_{\sigma\alpha} - \sum_{\omega\neq\alpha} P_\alpha(t)\theta_{\alpha\omega}}$$

8.1.1 The Time-Homogeneous Case

In many applications, the transition probabilities do not depend on time. Such a chain is called **time-homogeneous**. In this case, the transition matrix Θ is constant. The evolution of the probability distribution over multiple steps is then found by repeatedly applying the matrix:

$$\begin{aligned} P(1) &= P(0)\Theta \\ P(2) &= P(1)\Theta = P(0)\Theta^2 \\ &\vdots \\ P(t) &= P(0)\Theta^t \end{aligned}$$

The state of the system at any future time t is completely determined by the initial probability distribution $P(0)$ and the t-th power of the transition matrix.

Definition: *Stochastic Matrix*

The transition matrix Θ of a time-homogeneous DTMC has two fundamental properties and is known as a **(row) stochastic matrix**:

1. **Non-negativity:** All its elements are probabilities, so $0 \leq \theta_{\delta\sigma} \leq 1$.
2. **Rows sum to one:** From any given state δ, the system must transition to *some* state in S. Therefore, the sum of probabilities of transitioning out of state δ must be 1.

$$\sum_{\sigma\in S} \theta_{\delta\sigma} = 1 \quad \text{for all } \delta \in S$$

A consequence of the structure of the transition matrix Θ is that if $P(t)$ is a probability distribution:

$$\sum_{\sigma\in S} P_\sigma(t) = 1 \qquad \text{and} \qquad P_\sigma(t) \in [0,1]$$

then $P(t+1) = P(t)\Theta$ is also a probability distribution:

$$\sum_{\sigma\in S} P_\sigma(t+1) = 1 \qquad \text{and} \qquad P_\sigma(t+1) \in [0,1]$$

This guarantees that the probability distribution remains valid at each time step. Let's see why:

- **Preservation of normalization:** The sum of all probabilities at time $t+1$ is

$$\sum_{\sigma\in S} P_\sigma(t+1) = \sum_{\sigma\in S}\sum_{\delta\in S} P_\delta(t)\theta_{\delta\sigma}(t) = \sum_{\delta\in S} P_\delta(t) \sum_{\sigma\in S} \theta_{\delta\sigma}(t)$$

  Since each row of the transition matrix sums to 1 (i.e., $\sum_{\sigma\in S}\theta_{\delta\sigma}(t) = 1$ for all δ), this becomes:

$$= \sum_{\delta\in S} P_\delta(t)\cdot 1 = \sum_{\delta\in S} P_\delta(t)$$

  So, if the probabilities sum to 1 at time t, they will also sum to 1 at time $t+1$.
- **Non-negativity:** Since $P_\delta(t) \geq 0$ and $\theta_{\delta\sigma}(t) \geq 0$, it follows that $P_\sigma(t+1) \geq 0$ for all σ.

In summary, the evolution equation preserves both the normalization and non-negativity of the probability distribution at every step.

Eigenvalues and Steady State

The long-term (asymptotic) behavior of a discrete-time Markov chain (DTMC) is determined by the eigenvalues and eigenvectors of its transition matrix Θ.

A key property of any stochastic (row-stochastic) matrix is that it always has at least one eigenvalue equal to 1 (i.e., $\lambda = 1$). All other eigenvalues λ_i satisfy $|\lambda_i| \leq 1$; that is, their absolute values are less than or equal to 1.

The presence of the eigenvalue $\lambda = 1$ ensures that there exists a special probability distribution, called the **stationary distribution** (or steady state), usually denoted by π. This stationary distribution has the property that, if the system starts in π, it remains in π after each time step:

$$\pi = \pi\Theta$$

In other words, π is a **left eigenvector** of Θ with eigenvalue 1. If $\lambda = 1$ is a simple eigenvalue (i.e., its geometric multiplicity is one, which is typically the case for ergodic Markov chains), then the stationary distribution is unique.

Observation: *The Right Eigenvector*

The **right eigenvector** of Θ corresponding to $\lambda = 1$ is the column vector of all ones: $\mathbf{v} = [1, 1, \ldots, 1]^\top$. This is a direct consequence of the fact that each row of Θ sums to 1:

$$(\Theta\mathbf{v})_i = \sum_j \theta_{ij} v_j = \sum_j \theta_{ij} \cdot 1 = \sum_j \theta_{ij} = 1 = (1 \cdot \mathbf{v})_i$$

Alternative Formulation: Column Vector Evolution

In some texts, the probability distribution (PD) is represented as a column vector rather than a row vector. In this formalism, the evolution equation is written as

$$P(t+1) = WP(t)$$

where

$$P(t) = \begin{pmatrix} P_1(t) \\ \vdots \\ P_N(t) \end{pmatrix} \qquad \text{and} \qquad W = \Theta^\top$$

Here, W is the transpose of the original transition matrix Θ. The matrix W is also a (column-)stochastic matrix: each of its columns sums to 1, i.e.,

$$\sum_{R=1}^{N} W_{R,C}(t) = \sum_{R=1}^{N} \theta_{C,R}(t) = 1 \qquad \forall C.$$

A fundamental property of W is that the row vector $z = (1, \ldots, 1)$ is a left eigenvector of W associated with the eigenvalue $\lambda = 1$:

$$zW = z.$$

8.2 Asymptotic Behavior of the Probability Distribution

We have established that the evolution of the probability distribution is given by

$$P(t) = P(0)\Theta^t$$

To analyze the long-term behavior, it is useful to diagonalize the transition matrix Θ (when possible):

$$\Theta = H\,\mathrm{Diag}\{\lambda_1, \lambda_2, \ldots, \lambda_N\}H^{-1}$$

where λ_i are the eigenvalues of Θ and H is the matrix of corresponding eigenvectors. In the vast majority of cases, Θ is diagonalizable; in the rare cases where it is not, a Jordan decomposition can be used. Applying this decomposition, the evolution can be written as:

$$P(t) = P(0)H\,\mathrm{Diag}\{\lambda_1^t, \lambda_2^t, \ldots, \lambda_N^t\}H^{-1}$$

To interpret this result, recall the following key properties of stochastic (Markov) matrices:

- The transition matrix Θ always has an eigenvalue $\lambda_1 = 1$.
- All other eigenvalues satisfy $|\lambda_i| \leq 1$.
- If $P(0)$ is a valid probability distribution, then $P(t)$ remains a probability distribution for all t.

In the long-time limit ($t \to \infty$), all terms involving $|\lambda_i| < 1$ decay to zero, and only the contribution from $\lambda_1 = 1$ remains. This leads to the emergence of a stationary (steady-state) distribution, *except in the case where there is an eigenvalue* $\lambda = -1$. In that case, the probability distribution function (PDF) becomes periodic, oscillating between two (or more) values.

For completeness, let us also consider the evolution in terms of the right action of the transition matrix (i.e., treating $P(t)$ as a column vector and using the transpose $\Theta^\top$):

$$W = \Theta^\top$$

The t-step evolution is then given by

$$W^t = V\,\mathrm{Diag}\{\lambda_1^t, \ldots, \lambda_N^t\}V^{-1}$$

where V is the matrix of right eigenvectors of Θ.

Stationary Distribution: In the limit $t \to \infty$, the system approaches a stationary distribution P_∞ that satisfies

$$P_\infty = P_\infty\Theta$$

That is, P_∞ is a **left eigenvector** of Θ with eigenvalue 1. If the eigenvalue $\lambda = 1$ is simple (i.e., has geometric multiplicity one), then the stationary distribution is unique.

The stationary distribution must also satisfy the normalization condition:

$$\sum_\sigma (P_\infty)_\sigma = 1$$

In summary, the asymptotic behavior of a discrete-time Markov chain is governed by the spectral properties of its transition matrix. The existence and uniqueness of the stationary distribution are guaranteed under mild conditions, and the system will converge to this distribution regardless of the initial state, provided the chain is ergodic and aperiodic.

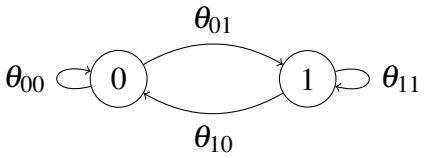
8.3 Simple DTMCs and Graphical Representation

A DTMC is often visualized as a directed graph where nodes represent the states and weighted edges represent the transition probabilities. This graphical representation is invaluable for building intuition about the system's dynamics.

For a general two-state system, $S = \{0, 1\}$, with a transition matrix:

$$\Theta = \begin{pmatrix} \theta_{00} & \theta_{01} \\ \theta_{10} & \theta_{11} \end{pmatrix}$$

the corresponding graph is:



The evolution of the probability of being in state 0, $P_0(t)$, is given by:

$$P_0(t+1) = \theta_{00} P_0(t) + \theta_{10} P_1(t)$$

Since we only have two states, $P_1(t) = 1 - P_0(t)$. Substituting this we get:

$$P_0(t+1) = \theta_{00} P_0(t) + \theta_{10}(1 - P_0(t)) = (\theta_{00} - \theta_{10}) P_0(t) + \theta_{10}$$

The stationary probability, P_0^{eq}, is found by setting $P_0(t+1) = P_0(t) = P_0^{eq}$, which yields:

$$P_0^{eq} = \frac{\theta_{10}}{1 - (\theta_{00} - \theta_{10})} = \frac{\theta_{10}}{(1 - \theta_{00}) + \theta_{10}} = \frac{\theta_{10}}{\theta_{01} + \theta_{10}}$$

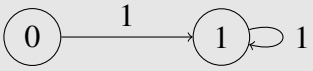
We can now calculate the stationary probability of being in state 1:

$$P_1^{eq} = 1 - P_0^{eq} = \frac{\theta_{01}}{\theta_{01} + \theta_{10}}$$

This general result provides the long-term probability of finding the system in state 0, provided the system is ergodic. Let's now explore some specific cases.

Example: *A Simple Absorbing Chain*

Consider a two-state Markov chain with states $S = \{0, 1\}$, where the transition probabilities are $\theta_{01} = 1$ and $\theta_{11} = 1$.



The transition matrix for this system is:

$$\Theta = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$$

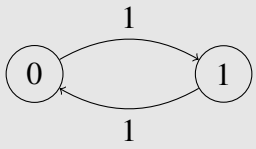
If the system starts in state 0 ($P(0) = [1, 0]$), it moves to state 1 in one step and stays there permanently. If it starts in state 1, it remains there. Therefore, regardless of the initial state, the system will always end up in state 1. The stationary (limiting) distribution is:

$$P(\infty) = [0, 1]$$

State 1 is called ***absorbing state*** because, once entered, the probability of leaving it is zero.

Example: *A Periodic Chain: The Flip-Flop*

Consider a system that **deterministically alternates** between two states, $S = \{0, 1\}$. In this Markov chain, the only possible transitions are from state 0 to state 1 and from state 1 to state 0, both with probability 1. That is, $\theta_{01} = 1$ and $\theta_{10} = 1$, while $\theta_{00} = \theta_{11} = 0$.



The transition matrix for this system is:

$$\Theta = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Let $P_0(t)$ denote the probability of being in state 0 at time t. The evolution equation is:

$$P_0(t+1) = \theta_{00}P_0(t) + \theta_{10}P_1(t) = 0 \cdot P_0(t) + 1 \cdot (1 - P_0(t)) = 1 - P_0(t)$$

or, equivalently,

$$P_0(t+1) = P_1(t) = 1 - P_0(t)$$

This is a simple linear recurrence relation. Its general solution is:

$$P_0(t) = \left(P_0(0) - \frac{1}{2}\right)(-1)^t + \frac{1}{2}$$

where $P_0(0)$ is the initial probability of being in state 0.

The probability $P_0(t)$ alternates between two values, depending on whether t is even or odd. The system never settles into a steady state; instead, it exhibits **periodic behavior** with period 2. This persistent oscillation is a direct consequence of the transition matrix Θ having eigenvalues $\lambda = 1$ and $\lambda = -1$.
The presence of an eigenvalue of -1 means that the system's state alternates in sign every time step, preventing convergence to a unique stationary distribution.

It is important to recognize that the discrete "time" variable t in a Discrete-Time Markov Chain (DTMC) does not always correspond to physical time. While in some applications t may indeed represent actual time steps (such as days, hours, or generations), in many contexts it serves as a convenient indexing parameter for the sequence of events or observations, regardless of whether these are temporally spaced.

For instance, in natural language processing, t might index the position of a letter or word in a string of text, allowing a DTMC to model the probabilistic structure of language. In clinical studies, t could denote the sequence number of a patient's follow-up visit, rather than a specific time interval. Similarly, in genetics, t might represent generations in a population, or in queueing theory, the order of customer arrivals.

A particularly illustrative example is the discrete-time SIR (Susceptible-Infected-Recovered) model in epidemiology. Although infection and recovery are inherently continuous-time processes, modeling them in discrete time, such as updating the state of the system day by day, can provide a practical and insightful approximation. This approach enables the use of DTMCs to capture the essential dynamics of the system while simplifying analysis and computation.

8.3.1 The Chapman-Kolmogorov Equations

The **Chapman-Kolmogorov equations** are a fundamental tool in the analysis of Markov chains, providing a recursive relationship for multi-step transition probabilities. They allow us to compute the probability of transitioning from one state to another in a given number of steps, by considering all possible intermediate states.

Suppose a Markov process starts in state S_i at time 0. We are interested in the probability that the process is in state S_j after n transitions, regardless of the specific path taken. This probability is denoted as:

$$P(X_n = S_j \mid X_0 = S_i) = (\Theta^{(n)})_{ij}$$

where $(\Theta^{(n)})_{ij}$ is the (i, j)-th entry of the n-step transition matrix $\Theta^{(n)}$.

To compute this probability, we can sum over all possible states S_k that the process could occupy at the previous time step $(n-1)$:

$$P(X_n = S_j \mid X_0 = S_i) = \sum_k P(X_n = S_j, X_{n-1} = S_k \mid X_0 = S_i)$$

By the definition of conditional probability and the Markov property, this can be rewritten as:

$$P(X_n = S_j \mid X_0 = S_i) = \sum_k P(X_n = S_j \mid X_{n-1} = S_k)\, P(X_{n-1} = S_k \mid X_0 = S_i)$$

Here, $P(X_n = S_j \mid X_{n-1} = S_k) = \theta_{kj}$ is the one-step transition probability from S_k to S_j, and $P(X_{n-1} = S_k \mid X_0 = S_i) = (\Theta^{(n-1)})_{ik}$ is the probability of being in S_k after $n-1$ steps starting from S_i.

Therefore, the Chapman-Kolmogorov equation for the n-step transition probability is:

$$(\Theta^{(n)})_{ij} = \sum_k (\Theta^{(n-1)})_{ik}\, \theta_{kj}$$

or, equivalently, in matrix notation:

$$\Theta^{(n)} = \Theta^{(n-1)} \cdot \Theta$$

By repeated application, we obtain:

$$\Theta^{(n)} = \Theta^n$$

That is, the n-step transition matrix is simply the n-th power of the one-step transition matrix.

Iterative Evolution of the State Distribution If $P(0)$ is the initial probability distribution over states (a row vector), then the distribution after n steps is given by:

$$P(n) = P(0) \cdot \Theta^n$$

Alternatively, the distribution evolves step by step as:

$$P(n) = P(n-1) \cdot \Theta$$

General Formulation More generally, the Chapman-Kolmogorov equations state that for any $0 \leq k \leq n$,

$$P(X_n = S_j \mid X_0 = S_i) = \sum_{S_k} P(X_n = S_j \mid X_k = S_k)\, P(X_k = S_k \mid X_0 = S_i)$$

It expresses the n-step transition probability as a sum over all possible intermediate states at time k.

8.4 Structural Properties of DTMCs

We have so far informally used graphs to represent DTMCs, as they provide valuable insight into the internal structure and dynamics of the chain. We now give a formal definition:

Definition: *Support Graph of a DTMC*

Given a DTMC with state space S and transition matrix $\Theta = (\theta_{ij})$, the **support graph** is the directed graph $G = (N, E)$ defined as follows:

- The set of nodes is $N = S$ (the set of states).
- The set of directed edges is $E = \{(i, j) \mid \theta_{ij} > 0\}$; that is, there is an edge from i to j if and only if the transition probability from i to j is positive.

Communicating Classes and Irreducibility

The strongly connected components of the support graph correspond to the **communicating classes** of the DTMC. Two states, i and j, are said to **communicate** (written $i \leftrightarrow j$) if it is possible to reach j from i and also to return from j to i, each in a finite number of steps. This communication relation is an equivalence relation, partitioning the state space S into disjoint communicating classes.

A Markov chain is called **irreducible** if there is only one communicating class; that is, every state can be reached from every other state. Such a chain is sometimes also called **ergodic**. If this is not the case, the chain is said to be **reducible**.

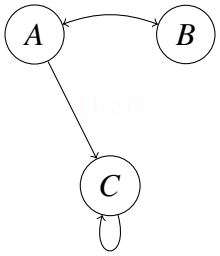


Figure 8.1: A reducible Markov chain with two communicating classes: $\{A, B\}$ and $\{C\}$.

Recurrence and Transience

A state i is called **recurrent** if, starting from i, the probability of eventually returning to i is 1. If this probability is less than 1, then i is called **transient**.

- In any finite Markov chain, at least one state must be recurrent.
- If i is recurrent and communicates with j, then j is also recurrent. So, recurrence is a property of the whole communicating class.
- If i is transient, the expected number of visits to i is finite; eventually, the process will leave i and not return.

A recurrent state i is called **positive recurrent** if the expected time to return to i (starting from i) is finite. In finite Markov chains, all recurrent states are automatically positive recurrent.

Periodicity

A state i has a **period** $d(i)$ if any return to i must occur in a number of steps that is a multiple of $d(i)$, where $d(i)$ is the greatest common divisor of all possible return times to i. If $d(i) = 1$, the state is called **aperiodic**. Like recurrence, periodicity is a class property: all states in a communicating class share the same period. A DTMC is called aperiodic if all its states are aperiodic.

The ”flip-flop” chain discussed earlier is a classic example of a periodic chain with period 2.

8.5 Absorbing States and Long-Term Behavior

A special and important case in DTMC analysis is the presence of absorbing states.

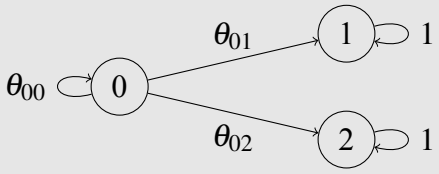
Definition: *Absorbing State and Subgraph*

A state s is an **absorbing state** if the probability of remaining in that state is 1, i.e., $\theta_{ss} = 1$. More generally, a communicating class C is an **absorbing subgraph** if there are no transitions leading from any state in C to any state outside of C.

In any DTMC with absorbing states, every non-absorbing state must be transient. This means the process will eventually leave the transient states and get "trapped" in one of the absorbing states or subgraphs forever. The long-term behavior of the system is therefore determined by which absorbing class it enters.

Example: *The General Case with Multiple Absorbing States*

The situation becomes more interesting when there are multiple absorbing states. Let's consider a system with three states, where states 1 and 2 are absorbing, and state 0 is transient. The corresponding graph is:



The evolution equations for the probabilities are:

$$\begin{aligned} P_0(t+1) &= \theta_{00} P_0(t) \\ P_1(t+1) &= \theta_{01} P_0(t) + P_1(t) \\ P_2(t+1) &= \theta_{02} P_0(t) + P_2(t) \end{aligned}$$

with the constraint that $\theta_{00} + \theta_{01} + \theta_{02} = 1$. The transition matrix for this system is:

$$\Theta = \begin{pmatrix} \theta_{00} & \theta_{01} & \theta_{02} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

From the first equation, we see that $P_0(t) = P_0(0)(\theta_{00})^t$. Since $\theta_{00} < 1$ (otherwise the other states would be unreachable), $P_0(t) \to 0$ as $t \to \infty$. This confirms that state 0 is transient. The probability mass that leaves state 0 at each step is distributed between states 1 and 2. To find the final distribution, we can sum the increments to $P_1(t)$ and $P_2(t)$ over all time:

$$P_1(\infty) = P_1(0) + \sum_{t=0}^{\infty} \theta_{01} P_0(t) = P_1(0) + \theta_{01} P_0(0) \sum_{t=0}^{\infty} (\theta_{00})^t$$

Using the formula for a geometric series, $\sum_{t=0}^{\infty} r^t = 1/(1-r)$, we get:

$$P_1(\infty) = P_1(0) + P_0(0) \frac{\theta_{01}}{1 - \theta_{00}}$$

Similarly for state 2:

$$P_2(\infty) = P_2(0) + P_0(0) \frac{\theta_{02}}{1 - \theta_{00}}$$

The final distribution is therefore $P(\infty) = [0, P_1(\infty), P_2(\infty)]$ and it depends on the initial distribution $P(0)$ and the transition probabilities out of the transient state. The eigenvalue $\lambda = 1$ has a geometric multiplicity of 2, corresponding to the two absorbing states.

8.5.1 Probability of Absorption

Let A be an absorbing subset of states in a DTMC. The **probability of absorption**, h_i^A, is the probability that the process will eventually enter the set A, given that it starts in state $i \notin A$.

$$h_i^A = P(\text{the process eventually enters } A \mid X_0 = i)$$

By conditioning on the first step, we can establish a system of linear equations for these probabilities. From state i, the process transitions to some state j with probability θ_{ij}. From there, the probability of being absorbed into A is h_j^A. Therefore:

$$h_i^A = \sum_{j \in S} \theta_{ij} h_j^A$$

This system must be solved with the boundary conditions:

- $h_i^A = 1$ if $i \in A$ (already absorbed).
- $h_i^A = 0$ if state i cannot reach any state in A.

Solving this system gives the absorption probabilities for all transient states.

8.5.2 Permanence Time

The **permanence time** is the number of consecutive steps a process stays in a given state before leaving it. Consider state $s \in S$, with the process starting at $X_0 = s$. Let T be the random variable for the number of steps the process remains in s before moving to any other state. The event $T = t$ means the process stays in s for t steps, then leaves to some $y \neq s$ at step $t+1$. The probability is:

$$P(T = t) = P(X_{t+1} \neq s, X_t = s, \ldots, X_1 = s \mid X_0 = s)$$

Due to the Markov property, we can factorize this conditional probability:

$$P(T = t) = P(X_1 = s \mid X_0 = s) \times \ldots \times P(X_t = s \mid X_{t-1} = s) \times P(X_{t+1} \neq s \mid X_t = s) = (\theta_{ss})^t \times \sum_{y \neq s} \theta_{sy}$$

Since the rows of the transition matrix must sum to one, it follows that the probability of leaving state s is $\sum_{y \neq s} \theta_{sy} = 1 - \theta_{ss}$. Therefore, the probability distribution of the permanence time is:

$$P(T = t) = (\theta_{ss})^t (1 - \theta_{ss})$$

This is the probability mass function of a **geometric distribution**. It is straightforward to verify that this is a valid probability distribution by summing over all possible times $t \geq 0$:

$$\sum_{t=0}^{\infty} P(T = t) = (1 - \theta_{ss}) \sum_{t=0}^{\infty} (\theta_{ss})^t = (1 - \theta_{ss}) \left(\frac{1}{1 - \theta_{ss}} \right) = 1$$

The **mean permanence time** in state s is the expected value of this geometric distribution:

$$\langle T \rangle = \frac{\theta_{ss}}{1 - \theta_{ss}}$$

Alternatively, if we start from state s, the mean permanence time is:

$$\langle T \rangle = \frac{1}{1 - \theta_{ss}}$$

The closer θ_{ss} is to 1, the longer the process is expected to remain in that state before transitioning. In the limit of an absorbing state ($\theta_{ss} = 1$), the mean permanence time is infinite.

8.6 Ergodicity

A DTMC is said to be **ergodic** if it is both irreducible (all states communicate) and aperiodic (no deterministic cycles). More formally, a chain is ergodic if for any two states i, j, there exists an integer n_0 such that the probability of transitioning from i to j in n steps is positive for all $n \geq n_0$:

$$(\Theta^n)_{ij} > 0 \quad \text{for all } n \geq n_0$$

Ergodicity is a crucial property because it guarantees a simple and predictable long-term behavior.

8.6.1 Properties of Ergodic Chains

If a DTMC is ergodic, it has several important properties:

1. **Unique Stationary Distribution:** There exists a unique equilibrium probability distribution, π, which is independent of the initial conditions.
2. **Convergence to Stationary Distribution:** For any initial distribution $P(0)$, the probability distribution at time t will converge to this unique stationary distribution π as $t \to \infty$.

   $$\lim_{t\to\infty} P(t) = \pi$$

   Furthermore, for sufficiently large t, this convergence is typically exponential, meaning $P(t) \approx \pi + O(e^{-qt})$ for some rate $q > 0$.
3. **Equivalence of Time and Ensemble Averages:** This is the most profound consequence of ergodicity. It states that the long-term time average of any function of the state is equal to the ensemble average (the expected value) of that function under the stationary distribution. For a function $A(x(t))$, the well-known law of large numbers for ergodic processes states:

   $$\underbrace{\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} A(x(t))}_{\text{Time Average}} = \underbrace{\sum_{s\in S} A(s)\pi_s}_{\text{Ensemble Average}}$$

8.6.2 Inferring the Stationary Distribution from Data

The property of ergodicity provides a powerful practical tool: we can infer the stationary distribution by analyzing a single, sufficiently long time series from the process.
Let's define a specific function $A(s) = \delta(s-r)$, which is 1 if the state s is our target state r, and 0 otherwise. The ensemble average is simply $\sum_s \delta(s-r)\pi_s = \pi_r$. The time average becomes the long-term frequency of visits to state r:

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \delta(x(t)-r) = \lim_{T\to\infty} \frac{\text{Number of visits to state } r}{T}$$

The ergodic property guarantees that these two are equal:

$$\pi_r = \lim_{T\to\infty} \frac{\text{Number of visits to state } r}{\text{Total number of observations}}$$

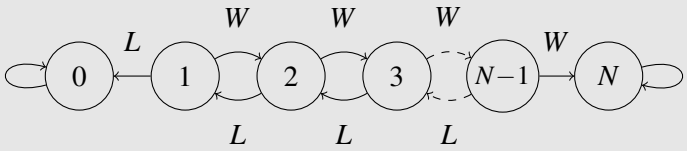
This means we can estimate the stationary probability of a state by simply measuring how often it occurs in a long simulation.

8.6.3 An Apparent Contradiction

In Markov chains with multiple absorbing states, an apparent paradox arises regarding the steady-state behavior. On one hand, the steady-state probability vector P_{eq} must be an eigenvector of the transition matrix W with eigenvalue 1, suggesting a unique equilibrium. On the other hand, when diagonalizing W, the general solution for $P(t)$ shows that the asymptotic distribution can depend on the initial condition $P(0)$. A classic example is the **Gambler's Ruin**.

Example: ***The Gambler's Ruin***

A gambler starts with i euros and, at each round, wins 1 euro with probability p or loses 1 euro with probability $q = 1 - p$. The game ends when the fortune reaches 0 or N euros (the absorbing states); states 1 to $N-1$ are transient.



For the specific case where $N = 4$, the transition matrix (using $W = \Theta^T$) is:

$$W = \begin{pmatrix} 1 & q & 0 & 0 & 0 \\ 0 & 0 & q & 0 & 0 \\ 0 & p & 0 & q & 0 \\ 0 & 0 & p & 0 & 0 \\ 0 & 0 & 0 & p & 1 \end{pmatrix}$$

The spectrum of W can be computed and is $\sigma(W) = \{1, 1, \sqrt{pq}, -\sqrt{pq}, 0\}$. The eigenvalue $\lambda = 1$ has a geometric multiplicity of 2, which corresponds to the two absorbing states. The eigenvectors for $\lambda = 1$ are $v_1 = (1,0,0,0,0)^T$ (ruin) and $v_2 = (0,0,0,0,1)^T$ (win). Since the system has two absorbing states, the long-term outcome depends on the initial state $P(0)$. *The final probability distribution will be a linear combination of the two absorbing states*:

$$P(t) \xrightarrow{t\to\infty} c_1 v_1 + c_2 v_2$$

where the coefficients c_1 and c_2 represent the probabilities of being absorbed at 0 and N, respectively, and depend on the initial capital $P(0)$.

This dependence on initial conditions seems to create a paradox. On one hand, the steady-state probability P_{eq} must be an eigenvector of W, satisfying $P_{eq} = WP_{eq}$. On the other hand, the general solution using diagonalization is:

$$P_{eq} = \lim_{t\to\infty} W^t P(0) = V\text{Diag}(\lambda_1^t, \ldots, \lambda_N^t)V^{-1}P(0) = V\text{Diag}(1,1,0,\ldots,0)V^{-1}P(0)$$

which explicitly depends on $P(0)$.

The resolution to this apparent paradox lies in the special properties of stochastic matrices. For a matrix with an eigenvalue $\lambda = 1$ of multiplicity one, the corresponding coefficient c_1 in the eigenvector expansion of $P(0)$ is independent of $P(0)$, guaranteeing a unique steady state.

However, in this case, the eigenvalue $\lambda = 1$ has a geometric multiplicity of two. This means the system is not ergodic, and the coefficients c_1 and c_2 that determine the final mixture of absorbing states *do* depend on the initial condition $P(0)$. There is no single steady state; rather, there is a different asymptotic outcome for each starting capital.

8.6.4 The Kac Lemma: Return Times

We have already encountered the **Kac Lemma** in Section 4.2.5, where it was discussed in the context of continuous-time Markov processes. The result applies equally to DTMCs: it provides a direct link between the stationary probability of a state and the expected time to return to it.

Let T_s be the random variable representing the **first return time** to state s, given that the process starts in state s. The Kac Lemma states that the expected value of this return time is the reciprocal of the stationary probability of that state:

$$\mathbb{E}[T_s] = \frac{1}{\pi_s}$$

This relationship is highly intuitive:

- If a state s is rarely visited, its stationary probability π_s is small. Consequently, the mean return time $\mathbb{E}[T_s]$ will be very long.
- If a state s is frequently visited, its probability π_s is close to one. The mean return time will be minimal, i.e., slightly larger than one.

9 Continuous-Time Markov Chains

Having explored discrete-time systems, we now turn our attention to processes that evolve continuously in time but still occupy a discrete set of states. This is a common scenario in many fields: a gene can be ”on” or ”off”, an ion channel can be ”open” or ”closed”, and an individual in a population can be ”susceptible”, ”infected”, or ”recovered”. These systems are not governed by the discrete-time transition matrix of a DTMC, but by a more dynamic framework known as the **Continuous-Time Markov Chain (CTMC)**.

9.1 Fundamental Properties

A Continuous-Time Markov Chain is a stochastic process that combines the discrete state space of a DTMC with a continuous timeline.

Definition: ***Continuous-Time Markov Chain***

A stochastic process $\{X_t\}$, where time t is continuous ($t \in \mathbb{R}_0^+$), is a **CTMC** if it satisfies:

1. **Discrete State Space:** The process can only occupy states from a finite or countably infinite set $S = \{s_1, s_2, \ldots, s_N\}$.
2. **Markov Property:** The future state of the system depends only on its present state, not on its history. For any infinitesimal time step $dt > 0$:

$$P(X(t+dt) = \alpha \mid X(\theta) \text{ for } \theta \leq t) = P(X(t+dt) = \alpha \mid X(t))$$

The core difference from DTMCs lies in how transitions are described. Instead of a matrix of one-step probabilities, CTMCs are characterized by **transition rates**. This concept is built on a set of assumptions about what can happen in an infinitesimally small time interval dt. These were first clarified by the physicist Wolfgang Pauli.

1. **At most one event:** In an infinitesimal time interval $(t, t+dt)$, the probability of two or more transitions occurring is negligible.

$$P(\text{Two or more events in } (t, t+dt)) = O(dt^2) \approx 0$$

2. **Infinitesimal transition probability:** The probability of a single transition from a state σ to a different state α is proportional to the duration of the interval, dt.

$$P(\text{One event in } (t, t+dt)) = O(dt)$$

This leads to the definition of the **transition rate** (or **propensity**), $W_{\alpha\sigma}$, from state σ to state α:

$$P(X(t+dt) = \alpha \mid X(t) = \sigma) = W_{\alpha\sigma}\, dt$$

Warning: *Rate, not Probability*

It is crucial to remember that $W_{\alpha\sigma}$ is a rate, not a probability. Its units are $1/\text{time}$. It can be greater than 1. For a transition to occur, this rate must be multiplied by the time interval dt.

From these principles, we can deduce the probability of remaining in the same state. Since the sum of probabilities of transitioning from state σ to any other state must be 1, we have:

$$\sum_{\alpha \in S} P(X(t+dt) = \alpha \mid X(t) = \sigma) = 1$$

Splitting the sum for the destination state $\alpha = \sigma$ and $\alpha \neq \sigma$:

$$P(X(t+dt) = \sigma \mid X(t) = \sigma) + \sum_{\alpha \neq \sigma} P(X(t+dt) = \alpha \mid X(t) = \sigma) = 1$$

Substituting the transition rates:

$$P(X(t+dt) = \sigma \mid X(t) = \sigma) + \sum_{\alpha \neq \sigma} W_{\alpha\sigma}\, dt = 1$$

Thus, the probability of no transition occurring is:

$$P(X(t+dt) = \sigma \mid X(t) = \sigma) = 1 - \left(\sum_{\alpha \neq \sigma} W_{\alpha\sigma} \right) dt$$

Example: *The SIR Epidemic Model*

The Susceptible-Infected-Recovered (SIR) model is a cornerstone of mathematical epidemiology and serves as a classic example of a system modeled by a CTMC. The state of the system at any time t is described by the number of susceptible and infected individuas:

$$X(t) = (S(t), I(t))$$

In an infinitesimal time interval dt, two types of events can occur, each with a specific transition rate:

- **Infection (con):** A susceptible individual becomes infected. The state change is:

  $$(S, I) \to (S-1, I+1)$$

  The rate of this event is proportional to the number of potential infectious contacts:

  $$W_{\text{con}} = \beta \frac{S \cdot I}{N}$$

  where β is the transmission rate and N is the total population size.
- **Recovery (rec):** An infected individual recovers and becomes immune. The state changs:

  $$(S, I) \to (S, I-1)$$

  The rate of recovery is proportional to the number of currently infected individuals:

  $$W_{\text{rec}} = \gamma I$$

  where γ is the recovery rate.

These two transition rates fully define the stochastic dynamics of the SIR model. The evolution of the probability distribution $P(S, I, t)$ is then governed by the Master Equation.

9.2 The Master Equation

The rules governing infinitesimal transitions allow us to derive a deterministic differential equation for the evolution of the probability distribution, $P_\sigma(t) = P(X(t) = \sigma)$: the **Master Equation**.

From DTMC to CTMC: Pauli's Insight

Before deriving the equation, let's see how the CTMC framework naturally emerges from the limit of a DTMC. Recall the Master Equation for a DTMC with discrete time steps of size u:

$$P_i(t+u) = P_i(t) + \sum_{j \neq i} (P_j(t)\theta_{ji} - P_i(t)\theta_{ij})$$

Dividing by u, we get an expression resembling a time derivative:

$$\frac{P_i(t+u) - P_i(t)}{u} = \sum_{j \neq i} \left(P_j(t)\frac{\theta_{ji}}{u} - P_i(t)\frac{\theta_{ij}}{u} \right)$$

For this expression to converge to a meaningful limit as $u \to 0$, the transition probabilities θ_{ij} must themselves be infinitesimal and proportional to u. This forces us to define the transition probabilities in terms of a rate:

$$\theta_{ij} = W_{ij}u + O(u^2)$$

In the limit $u \to dt \to 0$, we recover the concept of a **transition rate** W_{ij}, and the equation becomes a differential equation.

9.2.1 Derivation and Properties

To find the probability of being in state σ at time $t + dt$, we consider all the ways the system could have arrived there:

1. The system was already in state σ at time t and did not transition out.
2. The system was in some other state $\alpha \neq \sigma$ at time t and transitioned into σ.

Using the law of total probability, we can write:

$$P_\sigma(t+dt) = P(X(t+dt) = \sigma \mid X(t) = \sigma)P_\sigma(t) + \sum_{\alpha \neq \sigma} P(X(t+dt) = \sigma \mid X(t) = \alpha)P_\alpha(t)$$

Substituting the expressions for the transition probabilities:

$$P_\sigma(t+dt) = \left(1 - dt \sum_{\alpha \neq \sigma} W_{\alpha\sigma}\right) P_\sigma(t) + \sum_{\alpha \neq \sigma} (W_{\sigma\alpha}dt)P_\alpha(t)$$

Rearranging the terms to find the change in probability, $P_\sigma(t+dt) - P_\sigma(t)$:

$$\frac{P_\sigma(t+dt) - P_\sigma(t)}{dt} = \sum_{\alpha \neq \sigma} W_{\sigma\alpha}P_\alpha(t) - \left(\sum_{\alpha \neq \sigma} W_{\alpha\sigma}\right) P_\sigma(t)$$

Taking the limit as $dt \to 0$, we arrive at the **Master Equation**:

$$\boxed{\dot{P}_\sigma(t) = \underbrace{\sum_{\alpha \neq \sigma} W_{\sigma\alpha}P_\alpha(t)}_{\text{Gain: Flow into } \sigma} - \underbrace{P_\sigma(t)\left(\sum_{\alpha \neq \sigma} W_{\alpha\sigma}\right)}_{\text{Loss: Flow out of } \sigma}}$$

This can also be written more compactly:

$$\dot{P}_\sigma(t) = \sum_{\alpha \neq \sigma} (W_{\sigma\alpha}P_\alpha(t) - W_{\alpha\sigma}P_\sigma(t))$$

The Fundamental Constraint: Conservation of Probability

The Master Equation must always be complemented by the normalization constraint:

$$\sum_{\sigma \in S} P_\sigma(t) = 1$$

The structure of the equation ensures that if the probability is normalized at $t = 0$, it remains normalized for all time. We can prove this by taking the time derivative of the total probability:

$$\frac{d}{dt} \sum_{\sigma \in S} P_\sigma(t) = \sum_{\sigma \in S} \dot{P}_\sigma(t) = \sum_{\sigma \in S} \sum_{\alpha \neq \sigma} \left(W_{\sigma\alpha} P_\alpha(t) - W_{\alpha\sigma} P_\sigma(t) \right)$$

The right-hand side is a double summation. We can split it into two parts:

$$\sum_{\sigma \in S} \sum_{\alpha \neq \sigma} W_{\sigma\alpha} P_\alpha(t) - \sum_{\sigma \in S} \sum_{\alpha \neq \sigma} W_{\alpha\sigma} P_\sigma(t)$$

In the first term, we can swap the dummy summation indices $\sigma \leftrightarrow \alpha$:

$$= \sum_{\alpha \in S} \sum_{\sigma \neq \alpha} W_{\alpha\sigma} P_\sigma(t) - \sum_{\sigma \in S} \sum_{\alpha \neq \sigma} W_{\alpha\sigma} P_\sigma(t) = 0$$

The two terms are identical and cancel out. Thus, the total probability is conserved.

Example: ***The Fundamental 2-State Irreversible CTMC***

The simplest, yet theoretically most important, CTMC is the irreversible two-state process. It is fundamental to understanding the nature of waiting times and forms the basis of the Gillespie simulation algorithm.

The system has a state space $S = \{A, B\}$, where state B is **absorbing**. The only possible transition is from A to B with a constant rate W. The transition probabilities in an infinitesimal interval dt are:

$$P(X(t+dt) = B \mid X(t) = A) = W\,dt$$
$$P(X(t+dt) = A \mid X(t) = A) = 1 - W\,dt$$

The Master Equation for the probability of being in state A, $P_A(t)$, is derived from the fact that the system can only be in state A at time $t + dt$ if it was already there at time t and did not transition:

$$P_A(t+dt) = P_A(t) \cdot (1 - W\,dt) \qquad \xrightarrow{dt \to 0} \qquad \dot{P}_A(t) = -W P_A(t)$$

With an initial condition $P_A(0)$, the solution is a simple exponential decay:

$$P_A(t) = P_A(0) e^{-Wt}$$

A crucial question is: if the system starts in state A, how long will it stay there before jumping to B? Let T be the random variable for this **permanence time**. We will see that the waiting time in any state of a CTMC is always exponentially distributed, thus:

$$\langle T \rangle = \frac{1}{W}$$

Intuitively, if the transition rate W is high, the average waiting time is short, and vice-versa.

9.2.2 Permanence Time and the Exponential Distribution

A fundamental question for CTMCs is: *Given that the system is in state σ at time $t = 0$, what is the probability distribution for the time T until it leaves this state?*

Let T be the random variable representing the **permanence time** (or waiting time) in state σ. The process can leave state σ by making a transition to any other state $\alpha \neq \sigma$, each with rate $W_{\alpha\sigma}$. The total rate of leaving state σ is

$$W_{\text{out}} = \sum_{\alpha \neq \sigma} W_{\alpha\sigma}.$$

The probability that the process remains in state σ up to time t is the **survival probability**:

$$P_\sigma(t) = \Pr(T > t \mid X(0) = \sigma).$$

This probability satisfies the differential equation

$$\frac{d}{dt} P_\sigma(t) = -W_{\text{out}} P_\sigma(t),$$

with initial condition $P_\sigma(0) = 1$. The solution is

$$P_\sigma(t) = e^{-W_{\text{out}} t}.$$

This is the survival function for the permanence time T. The CDF for T is:

$$F_T(t) = \Pr(T \leq t) = 1 - e^{-W_{\text{out}} t},$$

and the probability density function (PDF) is

$$\rho_T(t) = \frac{dF_T(t)}{dt} = W_{\text{out}} e^{-W_{\text{out}} t}.$$

Thus, the permanence time in any state of a CTMC is **exponentially distributed** with parameter W_{out}. The mean waiting time in state σ is

$$\mathbb{E}[T] = \int_0^\infty t\, \rho_T(t)\, dt = \frac{1}{W_{\text{out}}}.$$

A large W_{out} means the process leaves state σ quickly (short average waiting time), while a small W_{out} means it tends to remain in σ longer.

9.2.3 Matrix Formulation

The Master Equation for a continuous-time Markov process can be written as a system of linear ordinary differential equations (ODEs):

$$\dot{P} = AP$$

where $P = (P_1, \ldots, P_{|S|})^T$ is the column vector of state probabilities, and A is the **transition rate matrix** (also called the **generator matrix**). The formal solution to this ODE system is:

$$P(t) = e^{At} P(0) = \Pi(t) P(0)$$

where e^{At} is the matrix exponential and $\Pi(t)$ is the **transition probability matrix**, with entries:

$$\Pi_{ij}(t) = \Pr(X(t) = i \mid X(0) = j)$$

The Chapman-Kolmogorov Equation (CKE) states that for any $t_1, t_2 \geq 0$,

$$\Pi(t_1 + t_2) = \Pi(t_1)\Pi(t_2)$$

This property follows directly from the properties of the matrix exponential and encodes the Markov property: the probability of transitioning from j to i in time $t_1 + t_2$ is the sum over all possible intermediate states k of the probability of going from j to k in time t_2 and then from k to i in time t_1.

Structure of the Generator Matrix A

The elements of A are given by:

- **Off-diagonal elements** ($i \neq j$): $A_{ij} = W_{ij}$, the transition rate from state j to state i.
- **Diagonal elements**: $A_{ii} = -\sum_{k \neq i} W_{ki}$, the negative of the total rate of leaving state i.

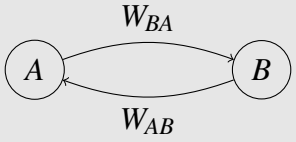
Thus, each column of A sums to zero:

$$\sum_i A_{ij} = 0 \qquad \forall j$$

This means A is singular and always has an eigenvalue $\lambda = 0$. The corresponding right eigenvector π (with $A\pi = 0$) gives the **stationary distribution** of the process, where $\dot{P} = 0$.

Example: ***Another Two-State CTMC***

Let us now consider a two-state continuous-time Markov chain (CTMC) where transitions can occur in both directions:



The system can transition from $A \to B$ with rate W_{BA} and from $B \to A$ with rate W_{AB}. The Master Equations for the probabilities $P_A(t)$ and $P_B(t)$ of being in states A and B at time t are:

$$\begin{aligned} \dot{P}_A(t) &= -W_{BA}P_A(t) + W_{AB}P_B(t) \\ \dot{P}_B(t) &= -W_{AB}P_B(t) + W_{BA}P_A(t) \end{aligned}$$

Since the system must be in either A or B at all times, $P_A(t) + P_B(t) = 1$ thus $P_B(t) = 1 - P_A(t)$. Substituting into the first equation we get:

$$\begin{aligned} \dot{P}_A(t) &= -W_{BA}P_A(t) + W_{AB}(1 - P_A(t)) \\ &= W_{AB} - (W_{AB} + W_{BA})P_A(t) \end{aligned}$$

At steady state ($\dot{P}_A = 0$), we solve for the stationary probability P_A^{eq}:

$$0 = W_{AB} - (W_{AB} + W_{BA})P_A^{\text{eq}} \quad \Longrightarrow \quad P_A^{\text{eq}} = \frac{W_{AB}}{W_{AB} + W_{BA}}$$

Similarly, for state B:

$$P_B^{\text{eq}} = 1 - P_A^{\text{eq}} = \frac{W_{BA}}{W_{AB} + W_{BA}}$$

The stationary distribution favors the state with the lower exit rate.
The transition rate matrix for this process is:

$$W = \begin{bmatrix} -W_{BA} & W_{AB} \\ W_{BA} & -W_{AB} \end{bmatrix}$$

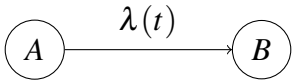
This matrix has rank 1, as can be verified by summing its rows. Its characteristic equation is:

$$\lambda^2 + (W_{AB} + W_{BA})\lambda = 0$$

which has a zero eigenvalue, guaranteeing the existence of a stationary probability distribution.

9.2.4 Time-Varying Transition Rates

So far, we have assumed that transition rates are constant in time, however in many cases rates can be time-dependent. Let us consider a two-state CTMC with states A and B, where transitions are only allowed from A to B, and the transition rate $\lambda(t)$ depends on t.



Let $P_A(t)$ be the probability that the system is in state A at time t. The Master Equation for $P_A(t)$ is:

$$\dot{P}_A(t) = -\lambda(t) P_A(t)$$

This is a first-order linear ODE with a time-dependent rate. Its solution is:

$$P_A(t) = P_A(0)\, e^{-L(t)} \qquad \text{where} \quad L(t) = \int_0^t \lambda(s) ds$$

If the system starts in state A at $t = 0$, then $P_A(t)$ is the probability that no transition has occurred up to time t. The PDF for the time T of the first transition (i.e., the **permanence time** in A) is:

$$\rho_T(t) = -\frac{d}{dt} P_A(t) = \lambda(t) e^{-L(t)}$$

and the cumulative distribution function (CDF) for the waiting time is:

$$\mathrm{CDF}_T(t) = 1 - e^{-L(t)}$$

Thus, the probability of remaining in state A up to time t is $e^{-L(t)}$.

The General Case: Time-Varying Matrix

The simple two-state example can be extended. In general, the Master Equation with a time-varying transition rate matrix is:

$$\dot{P}(t) = A(t) P(t)$$

where the entries of the matrix $A(t)$ are unrelated functions of time. This system does not have a general analytical solution. However, we can solve it in a few important special cases.

- **Case 1: Separable Rates** A solvable case occurs when all transition rates share the same time-dependent part, $\lambda(t)$. The transition matrix can be written as:

  $$A(t) = \lambda(t) \hat{A}$$

  where $\hat{A}$ is a constant matrix representing the fixed structure of the transitions. The Master Equation becomes $\dot{P}(t) = \lambda(t) \hat{A} P(t)$. By a change of time variable, this can be solved:

  $$P(t) = e^{L(t)\hat{A}} P(0) \qquad \text{where} \quad L(t) = \int_0^t \lambda(s) ds$$

- **Case 2: Piecewise-Constant Periodic Rates** Another important and tractable case is when the matrix $A(t)$ is piecewise constant and periodic. This is common in systems subject to seasonal or diurnal forcing. For example, consider a system with period Q where the dynamics are governed by matrix A_1 for a duration T_1 and then by matrix A_2 for a duration $T_2 = Q - T_1$.

  $$A(t) = \begin{cases} A_1 & \text{if } 0 < \mathrm{mod}(t, Q) \le T_1 \\ A_2 & \text{if } T_1 < \mathrm{mod}(t, Q) \le Q \end{cases}$$

  We can solve this system by integrating it over each constant interval.

– For the first interval $0 < t \leq T_1$:

$$\dot{P} = A_1 P \quad \Longrightarrow \quad P(t) = e^{A_1 t} P(0)$$

The state at the end of this interval is $P(T_1) = e^{A_1 T_1} P(0)$.

– For the second interval $T_1 < t \leq Q$:

$$\dot{P} = A_2 P \quad \Longrightarrow \quad P(t) = e^{A_2 (t - T_1)} P(T_1)$$

The state at the end of one full period is:

$$P(Q) = e^{A_2 (Q - T_1)} P(T_1) = e^{A_2 T_2} e^{A_1 T_1} P(0)$$

We can define a **one-period evolution matrix** $B = e^{A_2 T_2} e^{A_1 T_1}$. The state of the system at integer multiples of the period Q is then given by:

$$P(nQ) = B^n P(0)$$

This reduces the long-term continuous dynamics to a discrete-time evolution based on the matrix B. The stability and asymptotic behavior of the system are determined by the eigenvalues of B.

9.3 The Gillespie Algorithm

While the Master Equation describes the evolution of the probability distribution, it quickly becomes intractable for systems with many states. To study the behavior of a single realization of a CTMC, we use a simulation method that is exact, not an approximation: the **Gillespie Algorithm**.

The algorithm is based on the properties of permanence time we just derived. At any given time t_n, when the system is in state $X(t_n) = \sigma$, we need to answer two questions:

1. **When** will the next transition occur?
2. **Which** transition will it be?

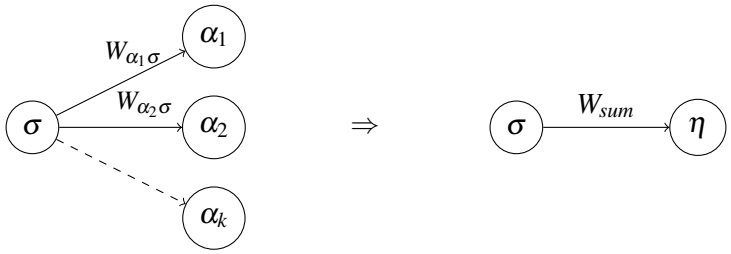


Figure 9.1: From state σ, the system can jump to multiple other states. This is equivalent to a single jump to a "rest of the world" state η with a total rate W_{sum}.

The key insight is that the waiting time T for the *next* event, regardless of which one it is, is exponentially distributed with a rate equal to the sum of all possible exit rates from the current state σ:

$$W_{sum} = \sum_{\alpha \neq \sigma} W_{\alpha \sigma}$$

Once we know that an event has occurred, the probability that it was a specific transition to state α_i is proportional to its individual rate:

$$P(\text{Jump to } \alpha_i \mid \text{A jump occurred}) = \frac{W_{\alpha_i \sigma}}{W_{sum}}$$

This leads to a simple and elegant simulation procedure.

Algorithm 1 The Gillespie Algorithm (Direct Method)

Initialize the system at time $t = 0$ with an initial state $X(0)$. Set a final time T_{final}.
while $t < T_{final}$ **do**
    Let the current state be $\sigma = X(t)$.
    Calculate rates: Compute all possible transition rates $W_{\alpha\sigma}$ out of state σ.
    Calculate total rate: Compute the sum of all rates, $W_{sum} = \sum_{\alpha} W_{\alpha\sigma}$.
    if $W_{sum} = 0$ **then**
        The system is in an absorbing state. **break** the loop.
    end if
    Sample waiting time: Draw a random number $u_1 \sim \mathrm{U}(0,1)$. Calculate the waiting time

$$T = -\frac{\ln(u_1)}{W_{sum}}$$

    Sample next event: Draw a second random number $u_2 \sim \mathrm{U}(0,1)$. Find the event α_j s.t.

$$\frac{\sum_{k=1}^{j-1} W_{\alpha_k\sigma}}{W_{sum}} < u_2 \leq \frac{\sum_{k=1}^{j} W_{\alpha_k\sigma}}{W_{sum}}$$

    Update: Set the new time $t \leftarrow t + T$ and the new state $X(t) \leftarrow \alpha_j$.
    Record the state and time if desired.
end while

9.3.1 Simulation with Time-Varying Rates

When rates $W_{\alpha\sigma}(t)$ depend on time, simulating the process becomes more complex because the waiting time for the next event is no longer exponentially distributed. If the system is in state σ at time t_n, we must find the time T to the next event.
The probability of remaining in state σ until time $t_n + T$ is:

$$P(\text{no event in } [t_n, t_n + T]) = \exp\left(-\int_{t_n}^{t_n+T} W_{sum}(t)dt\right)$$

where $W_{sum}(t) = \sum_{\alpha\neq\sigma} W_{\alpha\sigma}(t)$ is the time-dependent total exit rate. To sample the waiting time T using the inverse transform method, we draw a random number $u \sim \mathrm{U}(0,1)$ and solve the following integral equation for T:

$$\int_{t_n}^{t_n+T} W_{sum}(t)dt = -\ln(u)$$

This equation often has no analytical solution and must be solved numerically (e.g., using Newton's method) for T at each step of the simulation. Once T is found, the choice of the next event proceeds as before, using the rates evaluated at the time of the jump, $t_n + T$:

$$P(\text{Jump to } \alpha_i) = \frac{W_{\alpha_i\sigma}(t_n + T)}{W_{sum}(t_n + T)}$$

Example: *Gillespie Simulation of the SIR model*

Let's apply the Gillespie algorithm to the stochastic SIR model. The state of the system is given by the tuple (S,I). There are two possible events that can occur from any given state (S_n,I_n) at time t_n:

- **Infection (con):** An infected individual transmits the disease to a susceptible one:

$$(S,I) \to (S-1,I+1), \quad W_{\text{con}} = \beta \frac{S_n I_n}{N}$$

- **Recovery (rec):** An infected individual recovers:

$$(S,I) \to (S,I-1), \quad W_{\text{rec}} = \gamma I_n$$

The simulation proceeds step-by-step as follows:

1. At time t_n with state (S_n,I_n), calculate the two rates, W_{con} and W_{rec}.
2. Calculate the total exit rate: $W_{sum} = W_{\text{con}} + W_{\text{rec}}$.
3. Draw a random number $u_1 \sim \text{Unif}(0,1)$ and determine the time to the next event:

$$T_n = -\frac{\ln(u_1)}{W_{sum}}$$

   The next event will occur at time $t_{n+1} = t_n + T_n$.
4. Draw a second random number $u_2 \sim \text{Unif}(0,1)$ to determine which event occurs. The probability of a contagion event is:

$$P(\text{con}) = \frac{W_{\text{con}}}{W_{sum}} = \frac{\beta S_n I_n / N}{\beta S_n I_n / N + \gamma I_n}$$

5. If $u_2 < P(\text{con})$, the event is an infection, and the state is updated to $(S_n - 1, I_n + 1)$. Otherwise, the event is a recovery, and the state becomes $(S_n, I_n - 1)$.
6. Repeat the process until the epidemic ends ($I = 0$).

Extension: Periodic Contact Rate

Now consider a more realistic scenario where the contact rate varies periodically, for example, due to seasonality:

$$\beta(t) = \beta_m(1 + a\cos(\omega t))$$

In this case, sampling the waiting time T requires solving the integral equation:

$$L(T) = \int_{t_n}^{t_n+T} W_{sum}(s)ds = -\ln(u_1)$$

Given the state (S_n,I_n), the sum of rates is $W_{sum}(s) = \gamma I_n + \beta(s) S_n I_n / N$. The integral $L(T)$ becomes:

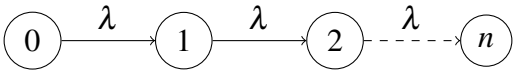
$$L(T) = T\gamma I_n + \frac{I_n S_n \beta_m}{N}\left(T + \frac{a}{\omega}(\sin(\omega(t_n+T)) - \sin(\omega t_n))\right)$$

This transcendental equation for T must be solved numerically at each step of the simulation. Once T is found, the choice of which event occurred is made by comparing u_2 to the ratio of the rates evaluated at the new time $t_n + T$.

9.4 Special Processes and Extensions

9.4.1 The Poisson Process

The Poisson process is a fundamental CTMC that models the counting of events occurring at a constant rate λ. Let $P_n(t)$ be the probability of having observed n events by time t. The system can only transition from state n to state $n+1$.



The transition rates are $W_{n+1,n} = \lambda$ for all $n \geq 0$. The Master Equation for state n is:

$$\begin{cases} \dot{P}_n(t) = \lambda P_{n-1}(t) - \lambda P_n(t) & n \geq 1 \\ \dot{P}_0(t) = -\lambda P_0(t) & n = 0 \end{cases}$$

With the initial condition $P_0(0) = 1$ (zero events at time zero), this system of ODEs can be solved recursively. The solution is the famous **Poisson distribution**:

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

Poisson Process with Time-Varying Rate

If the rate of events $\lambda(t)$ is time-dependent, the Master Equation becomes $\dot{P}_n(t) = \lambda(t)P_{n-1}(t) - \lambda(t)P_n(t)$. The solution is a generalization known as the non-homogeneous Poisson process, where the distribution at time t is:

$$P_n(t) = \frac{L(t)^n}{n!} e^{-L(t)}, \quad \text{where} \quad L(t) = \int_0^t \lambda(s)ds$$

Here, the number of events follows a Poisson distribution with a mean equal to $L(t)$.

9.4.2 Pure Death Process

Consider a system of N identical, independent particles at $t = 0$, where each particle has a constant rate of decay γ. Let $X(t) = n$ be the number of particles remaining at time t. This is a **pure death process**, a special case of the birth-death process where the birth rate is zero. The only possible transition is from state n to $n-1$. Since each of the n particles can decay independently, the total rate of leaving state n is:

$$d(n) = n\gamma$$

The Master Equation for the probability $P_n(t)$ is:

$$\dot{P}_n(t) = (n+1)\gamma P_{n+1}(t) - n\gamma P_n(t)$$

with the boundary condition $\dot{P}_N(t) = -N\gamma P_N(t)$. While solving this system for $P_n(t)$ is possible, it is often more insightful to study the evolution of the moments.

Evolution of the Mean

The mean number of particles is $\langle n(t) \rangle = \sum_{n=0}^{N} nP_n(t)$. Its time derivative is:

$$\frac{d\langle n \rangle}{dt} = \sum_{n=0}^{N} n\dot{P}_n(t) = \gamma \sum_{n=0}^{N} n\left[(n+1)P_{n+1} - nP_n\right]$$

This can be split into two sums. By re-indexing the first sum ($k = n+1$), we find that the terms nearly cancel, leading to a simple ODE for the mean:

$$\frac{d\langle n\rangle}{dt} = -\gamma \sum_{n=0}^{N} nP_n(t) = -\gamma\langle n(t)\rangle$$

With the initial condition $\langle n(0)\rangle = N$, the solution is:

$$\langle n(t)\rangle = Ne^{-\gamma t}$$

The average number of particles decays exponentially, just like in a deterministic model.

Evolution of the Second Moment

Similarly, we can find an equation for the second moment, $\langle n^2(t)\rangle$:

$$\frac{d\langle n^2\rangle}{dt} = \sum_{n=0}^{N} n^2 \dot{P}_n(t) = \gamma \sum_{n=0}^{N} n^2 \left[(n+1)P_{n+1} - nP_n\right]$$

After careful re-indexing and algebraic manipulation (using the identity $n^2 = (n+1)^2 - 2(n+1) + 1$), this simplifies to:

$$\frac{d\langle n^2\rangle}{dt} = -2\gamma\langle n^2(t)\rangle + \gamma\langle n(t)\rangle$$

Substituting the known solution for $\langle n(t)\rangle$:

$$\frac{d\langle n^2\rangle}{dt} = -2\gamma\langle n^2\rangle + \gamma N e^{-\gamma t}$$

This is a linear first-order ODE for $\langle n^2\rangle$. With the initial condition $\langle n^2(0)\rangle = N^2$, its solution is:

$$\langle n^2(t)\rangle = N(N-1)e^{-2\gamma t} + Ne^{-\gamma t}$$

From this, we can find the variance:

$$\mathrm{Var}[n(t)] = \langle n^2(t)\rangle - \langle n(t)\rangle^2 = Ne^{-\gamma t}(1 - e^{-\gamma t})$$

The variance starts at zero, increases to a maximum, and then decays back to zero as all particles eventually disappear.

9.4.3 Birth-Death Processes

A particularly important class of CTMCs is the **birth-death process**, which models populations where the state n (the population size) can only change by ± 1. These processes are defined by two state-dependent rates:

- **Birth rate** $b(n)$**:** The rate of transition from state n to $n+1$.
- **Death rate** $d(n)$**:** The rate of transition from state n to $n-1$. It is typically assumed that $d(0) = 0$.

The Master Equation for the probability $P_n(t)$ of being in state n is:

$$\dot{P}_n = b(n-1)P_{n-1} + d(n+1)P_{n+1} - (b(n) + d(n))P_n$$

Steady-State Analysis

At steady state ($\dot{P}_n = 0$), the probability distribution is constant. This implies that the net flow of probability between any two adjacent states must be zero. Let's define the **probability current** J_n as the net flow from state $n-1$ to n:

$$J_n = b(n-1)P_{n-1} - d(n)P_n$$

The steady-state Master Equation can be rewritten as $J_{n+1} - J_n = 0$, which means the current must be constant for all n. For systems with a reflective boundary at $n = 0$ (no negative population), this constant must be zero. Therefore, at steady state:

$$J_n = 0 \quad \Longrightarrow \quad d(n)P_n = b(n-1)P_{n-1}$$

This gives a simple recurrence relation for the stationary distribution P_n:

$$P_n = \frac{b(n-1)}{d(n)} P_{n-1}$$

By iterating this relation, we can express any P_n in terms of P_0:

$$P_n = P_0 \prod_{k=1}^{n} \frac{b(k-1)}{d(k)}$$

The value of P_0 is then found by imposing the normalization condition $\sum_{n=0}^{\infty} P_n = 1$.

Example: ***Immigration-Death Process***

Consider a population with a constant immigration rate (birth rate) $b(n) = b$ and a per-capita death rate $d(n) = kn$. The steady state recurrence is:

$$P_n = \frac{b}{kn} P_{n-1}$$

Let's define $R = b/k$. The recurrence becomes $P_n = \frac{R}{n} P_{n-1}$. Iterating this:

$$P_n = \frac{R}{n} \cdot \frac{R}{n-1} P_{n-2} = \ldots = \frac{R^n}{n!} P_0$$

To normalize, we sum over all n:

$$\sum_{n=0}^{\infty} P_n = P_0 \sum_{n=0}^{\infty} \frac{R^n}{n!} = P_0 e^R = 1 \quad \Longrightarrow \quad P_0 = e^{-R}$$

The stationary distribution is a Poisson distribution with mean R:

$$P_n = \frac{R^n}{n!} e^{-R}$$

If $R \leq 1$, the distribution is monotonically decreasing. If $R > 1$, it is unimodal, with its peak near R.

9.5 Population Continuous Time Markov Chains (PCTMC)

If the state variables represent a population (of humans, of animals, of molecules, of objects etc.) then it is possible to specialize the theoretical apparatus we have developed. To start:

$$\text{Prob}\left(x(t+dt)=\alpha|x(t)=\delta\right)=W_{\alpha\delta}dt$$

can be rewritten (thanks to the structure of $\mathbb{N}_0^M$) as:

$$\text{Prob}\left(x(t+dt)=\delta+(\alpha-\delta)|x(t)=\delta\right)=W_{\alpha\delta}dt$$

Setting

$$V_{\alpha\delta}=(\alpha-\delta),$$

which specifies the jump needed to switch state from δ to α

$$\text{Prob}\left(x(t+dt)=x(t)+V_{\alpha\delta}|x(t)=\delta\right)=W_{\alpha\delta}dt$$

In this context, we use the index η to label the different types of events that can occur in the system. Each event η is defined by a specific change (or "jump") in the system's state, and occurs at a rate that generally depends on the current state.

A key feature of Population Continuous Time Markov Chains (PCTMC) is that there are only a finite number of possible events, each associated with a fixed jump vector (independent of the current state), but with a rate that depends on the current state $x(t)$:

$$\text{Prob}\left(\text{Event}_\eta \mid x(t)\right)=\text{Prob}\left(x(t+dt)=x(t)+V^{(\eta)} \mid x(t)\right)=r_\eta(x(t))\,dt$$

This is a restatement of the general CTMC rule, but here the focus is on events, and the fact that, in population models, event rates depend on the current state.

Let's formally define a PCTMC as:

Definition: ***Population Continuous Time Markov Chain (PCTMC)***

A PCTMC is defined by a tuple:

$$X=(X,X_0,S,T)$$

where:

- $S\subseteq\mathbb{N}_0^M$ is the state space (e.g., all possible population counts),
- $X,X_0\in S$ are the current and initial states,
- $T=\{(\text{Name}_\eta,v_\eta,r_\eta(X;t))\}_{\eta=1}^{Z}$ is the set of transitions, where:
  - Name_η is the name of event η,
  - $v_\eta\in S$ is the jump vector for event η,
  - $r_\eta(X;t)\geq 0$ is the rate of event η, which may depend on the current state X and time t.

The rate $r_\eta(X;t)$ must be zero if the resulting state would fall outside the state space:

$$(X+v_\eta)\notin S \implies r_\eta(X;t)=0.$$

With this structure, the Master Equation for the probability $P_x(t)$ of being in state x at time t becomes:

$$\dot{P}_x(t)=\sum_\eta\left[r_\eta(x-v_\eta)P_{x-v_\eta}(t)-r_\eta(x)P_x(t)\right]$$

This form of the Master Equation is commonly known as the *Chemical Master Equation*.

9.6 Approximating PCTMC Dynamics with ODEs

A stochastic model described by a PCTMC can often be approximated by a deterministic ODE system, provided the event rates and state variables scale appropriately with a large parameter N (such as population size, volume, or area). Specifically, we require that the rates satisfy:

$$r_\eta(X;t) = N\, r_\eta\left(\frac{X}{N};t\right)$$

where X is the state and r_η is the rate of event η.
Define the normalized state variable $y = X/N$. The average dynamics of the system are then given by the cauchy problem:

$$\begin{cases} \dot{y}(t) = \sum_\eta v_\eta\, r_\eta(y;t) \\ y(0) = \frac{X(0)}{N} \end{cases}$$

Kurtz's theorem formalizes this approximation: for any $\delta > 0$,

$$\lim_{N\to\infty} \Pr\left(\left| y(t) - \frac{X(t)}{N} \right| > \delta\right) = 0$$

That is, as N becomes large, the normalized stochastic process $\frac{X(t)}{N}$ converges in probability to the deterministic solution $y(t)$.
We can express the deviation as

$$\frac{X(t)}{N} = y(t) + \varepsilon^{(N)}(t)$$

where, for large N, the fluctuations $\varepsilon^{(N)}(t)$ are small:

$$|\varepsilon^{(N)}(t)| = O(1/\sqrt{N}) \ll 1$$

Example: ***ODE Approximation of the SIR model***

Recall that the deterministic SIR model is governed by the following ODEs:

$$\dot{S} = -\beta\frac{SI}{N}, \quad \dot{I} = \beta\frac{SI}{N} - \gamma I, \quad \dot{R} = \gamma I$$

where S, I, and R represent the numbers of susceptible, infected, and recovered individuals. As we've seen, the SIR model involves two types of events: infection (contagion) and recovery. The rates at which these events occur both scale with the total population size N:

$$r_{\text{con}}(S,I) = \beta\frac{1}{N}SI = N\beta\frac{S}{N}\frac{I}{N} = N r_{\text{con}}\left(\frac{S}{N},\frac{I}{N}\right)$$
$$r_{\text{rec}}(S,I) = \gamma I = N\gamma\frac{I}{N} = N r_{\text{rec}}\left(\frac{I}{N}\right)$$

Defining the normalized variables

$$(y_S, y_I) = \left(\frac{S}{N},\frac{I}{N}\right),$$

for large N the PCTMC is very well approximated by the ODE system

$$\dot{y}_S = -\beta y_S y_I, \qquad \dot{y}_I = \beta y_S y_I - \gamma y_I.$$

9.6.1 The Second Gillespie Algorithm (First Reaction Method)

Gillespie also proposed a second exact simulation algorithm, often called the **First Reaction Method**. This approach is particularly natural for Population Continuous-Time Markov Chains (PCTMCs) and is based on the idea of independently sampling the waiting time for each possible event.

1. For each possible event η, sample an independent uniform random variable:

$$u_\eta \sim \text{Unif}([0,1])$$

2. For each event, compute its tentative firing time:

$$\tau_\eta = \frac{-\log(1-u_\eta)}{r_\eta(x(t_n))}$$

   where $r_\eta(x(t_n))$ is the propensity (rate) of event η given the current state $x(t_n)$.
3. The next event to occur is the one with the smallest τ_η:

$$\eta^* = \text{ArgMin}_\eta\{\tau_\eta\}$$

4. The time of the next event is:

$$t_{n+1} = t_n + \tau_{\eta^*}$$

5. Update the system state according to event η^*, and repeat the procedure.

This method is mathematically equivalent to the Direct Method, but can be more intuitive in some settings, as it treats each possible event as a competing exponential clock and simply executes the one that rings first.

9.6.2 The Tau-Leaping Algorithm and Other Approximations

A notable result of the PCTMC framework is that the state at time t can be written as:

$$x(t) = x(0) + \sum_\eta V_\eta \cdot K_\eta(t,0)$$

where $K_\eta(t,0)$ represents the number of occurrences of event type η in the time interval $(0,t)$. More compactly, we define:

$$K_\eta(t,dt) = (\#\text{ events of type } \eta \text{ in } (t,t+dt))$$

Since the probability of occurrence of an event of type η in the interval $(t,t+dt)$ is $r_\eta(x(t))dt$, and this probability depends on the state variable $x(t)$, the formula above is practically unusable in its current form because $x(t)$ will stochastically depend on the history of events itself.
Indeed, following the formal proof of Theorem 1.22 in Chapter 1 of Anderson and Kurtz's book *Stochastic Analysis of Biochemical Systems* (Springer, 2015), we have:

$$K_\eta(t,0) = \text{Poisson}\left(\int_0^t r_\eta(x(s))ds\right)$$

Daniel Gillespie, observing that:

$$x(t+\tau) = x(t) + \sum_\eta V_\eta K_\eta(t+\tau,t)$$

and that for small τ, the equation above can be straightforwardly approximated as:

$$K_\eta(t+\tau,t) \approx \text{Poisson}\left(\tau r_\eta(x(t))\right)$$

proposed the following approximated simulation algorithm:

$$x(t+\tau) \approx x(t) + \sum_\eta V_\eta \text{Poisson}\left(\tau r_\eta(x(t))\right)$$

This is termed the **tau-leaping algorithm**. This approximation is appropriate only when all state variables are sufficiently large. Otherwise, some populations may become negative, violating the requirement that they remain non-negative. The method also assumes that the rates $r_\eta(x(t))$ stay nearly constant over the interval τ, which is only true if small changes in the state do not significantly affect the rates.

Recovering the ODE Approximation

If we deterministically approximate the Poisson random variable with its mean value, we obtain:

$$x(t+\tau) \approx x(t) + \tau \sum_\eta V_\eta r_\eta(x(t))$$

Taking the limit as $\tau \to 0$, we recover:

$$\dot{x} = \sum_\eta V_\eta r_\eta(x(t))$$

This is precisely the ODE approximation derived from Kurtz's theorem, providing an alternative derivation of the deterministic limit of PCTMCs.

The Chemical Langevin Equation

If instead of using the mean approximation, we approximate the Poisson distribution using a Gaussian distribution with matching mean and variance:

$$\text{Poisson}\left(\tau r_\eta(x(t))\right) \approx \mathcal{N}\left(\mu = \tau r_\eta(x(t)), \sigma^2 = \tau r_\eta(x(t))\right)$$

We can write this as:

$$\tau r_\eta(x(t)) + \sqrt{\tau r_\eta(x(t))}\mathcal{N}(0,1)$$

Setting $\tau = dt$ yields:

$$x(t+dt) = x(t) + dt \sum_\eta V_\eta r_\eta(x(t)) + \sum_\eta V_\eta \sqrt{\tau r_\eta(x(t))}\mathcal{N}(0,1)$$

This leads to the following stochastic differential equation, known as the **Chemical Langevin Equation**:

$$\dot{x} = \sum_\eta V_\eta r_\eta(x(t)) + \sum_\eta V_\eta \sqrt{r_\eta(x(t))}\xi_\eta(t)$$

where $\xi_\eta(t)$ are independent white noise processes satisfying:

$$\langle \xi_\eta(t) \rangle = 0, \quad \langle \xi_\eta(t)\xi_{\eta'}(s) \rangle = \delta_{\eta\eta'}\delta(t-s)$$

Similar to the tau-leaping algorithm, the Chemical Langevin Equation also suffers from the risk that some state variables may become negative due to the stochastic fluctuations. This approximation should only be used when all populations are sufficiently large that the probability of negative values is negligible.

10 Noisy DTMC, DTCSS, CTRW

11 Raw Lecture Notes

Warning: *Raw Lecture Notes*

The following chapters contain unstructured notes taken during lectures and may require further organization and refinement.

11.1 Lecture 11/04/2025

—

$$\dot{x} = a(x) + b(x)\eta_h(t)$$

$$\langle \eta_h(t) \rangle = 0$$

$\mathbb{R}_h$ is a function with a peak at $|z| < h$ and $\mathbb{R}_h(z) = 0$ for $|z| > h$.

If the limit $\lim_{h\to 0^+} \mathbb{R}_h(z) = \delta(\tau)$, then:

$$\dot{x}_h = a(x_h) + b(x_h)\eta_h(t) \quad \to \quad \dot{x} = a(x) + b(x) \circ \xi(t)$$

—

Let's consider a population and two opinions:

$$x + y = 1$$

People can change their opinion with a rate R and the ratio of changing from x to y is θ, and from y to x is k.

$$\begin{cases} \dot{x} = +\theta xy - kyx - \varepsilon x + \varepsilon y \\ \dot{y} = -\theta xy + kyx + \varepsilon x - \varepsilon y \end{cases}$$

where ε is the rate of changing opinion. (?)
Substituting $y = 1 - x$ we can rewrite the first equation as:

$$\dot{x} = x(1-x)(\theta - k) + \varepsilon(1-x) - \varepsilon x$$

$$\frac{dx}{dt} = \lambda x(1-x) + 1 - 2x, \quad \lambda \to \lambda + \alpha\xi(t)$$

$$dx = \underbrace{\{\lambda x(1-x) + 1 - 2x\}}_{a(x)} dt + \underbrace{\alpha x(1-x)}_{b(x)} \circ dW$$

…

$$dx = \{1-2x\}\,dt + \alpha x(1-x)\circ dW$$

$$a(x) = \tfrac{1}{2}b'(x)b(x), \qquad b(x) = \alpha(x-x^2), \qquad b'(x) = \alpha(1-2x)$$

so

$$1-2x = \frac{\alpha^2}{2}x(1-x)(1-2x)$$

which has two equilibrium points:

$$x_1 = \frac{1}{2}, \qquad x_2 : 1 = \frac{\alpha^2}{2}x(1-x) \;\; (?)$$

…

—

$$m\ddot{x} = -\gamma_T\dot{x} + F_T(x) + \omega_T\xi(t)$$

$$\begin{cases} \dot{x} = v \\ \dot{v} = -\dfrac{\gamma_T}{m}v + \dfrac{F_T(x)}{m} + \dfrac{\omega_T}{m}\xi(t) \end{cases}$$

to simplify the notation we can write:

$$\gamma = \frac{\gamma_T}{m}, \qquad F = \frac{F_T(x)}{m}, \qquad \omega = \frac{\omega_T}{m}, \qquad \underbrace{U = \int F_T(x)dx, \qquad U' = \frac{dU}{dx} = F_T(x)}_{?}$$

we get:

$$\begin{cases} \dot{x} = v \\ \dot{v} = -\gamma v + F(x) + \omega\xi(t) \end{cases}$$

$$\frac{\partial p}{\partial t} = -v\frac{\partial p}{\partial x} - \frac{\partial}{\partial v}\left[(F(x)-\gamma v)p\right] + \frac{\partial^2}{\partial v^2}\frac{\omega^2}{2}$$

$$\frac{\partial p}{\partial t} = -v\frac{\partial p}{\partial x} - F(x)\frac{\partial p}{\partial v} + \gamma\frac{\partial}{\partial v}(vp) + \frac{\partial^2}{\partial v^2}\frac{\omega^2}{2}$$

$$0 = -v\frac{\partial p}{\partial x} + U'(x)\frac{\partial p}{\partial v} + \gamma p + \gamma v\frac{\partial p}{\partial v} + \frac{\partial^2}{\partial v^2}\frac{\omega^2}{2}$$

$$p(x,v) = A(x)B(v)$$

$$-vA'(x)B(v) + U'(x)A(x)B'(v) + \gamma A(x)B(v) + \gamma vA(x)B'(v) + \frac{\omega^2}{2}B''(v) = 0$$

$$\underbrace{\overbrace{-v\frac{A'(x)}{A(x)} + U'(x)\frac{B'(v)}{B(v)}}^{1^{st}\text{ term}}}_{=0} + 1\,\overbrace{\gamma + \gamma v\frac{B'(v)}{B(v)} + \frac{\omega^2}{2}\frac{B''(v)}{B(v)}}^{2^{nd}\text{ term}} = 0$$

We have to options:

1. set the second term to zero
2. boh

Let's consider the first option and let's define some "test" variables B_T and B'_T. We have:

$$\frac{B'_T(v)}{B_T(v)} = -\eta v \quad \Rightarrow \quad B'_T(v) = -\eta v B_T(v) \quad \Rightarrow \quad B(v) = Ce^{-\eta v^2/2}$$

we have

$$B'(v) = -\eta v B(v), \qquad B''(v) = -\eta v B'(v) = -\eta v(-\eta v B(v))$$

MISSING: boh

$$P_s = \frac{1}{z} e^{-\frac{\gamma}{\omega^2}v^2 - \frac{2\gamma}{\omega^2}U(x)} = \frac{1}{z} e^{-\frac{2\gamma}{\omega^2}\left[\frac{v^2}{2}+U(x)\right]}$$

Applying back the transformation we have:

$$p(x,v) = \frac{1}{z} e^{-\frac{2\gamma_T}{\omega_T}\left[\frac{mv^2}{2}+U_T(x)\right]}$$

so:

$$\iint p_s(x,v)dxdv = 1, \qquad \frac{1}{z}\iint e^{-\frac{2\gamma_T}{\omega_T^2}E_T(x,v)}dxdv = 1$$

TODO: check if this is correct

Draft

MISSING: fishes example ?

$$dx = f(x)dt - \underbrace{(cxdt + \omega x dW)}_{\text{\#fishes killed in } (t,\, t+dt)}$$

We want the number of fishes to be positive.

MISSING: end of the lecture

11.2 Lecture: 05/05/2025

…

... if there is no linearity, ...

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}\left\{(\theta \int zP(z,t)\mathrm{d}z + (1-\theta)x - x^3)P\right\} + \frac{\omega^2}{2}\frac{\partial^2 P}{\partial x^2} \qquad N \gg 1$$

$$M(t) = \int_{\mathbb{R}} xP(x,t)dx$$

$$P_s(x,M_s) = C(M_s)\exp\{\theta M_s x + ...\}$$

MISSING: end of the formula above

This solution is not actually so "usable"

$$M_s = \int_{\mathbb{R}} xP_s(x;M_s)dx \quad \Rightarrow \quad M_s = \Psi(M_s)$$

$$M_s = \Psi(M_s) \quad \rightarrow \quad \text{"unique solution"}$$

There are more interesting cases, for instance when $\Psi(M_s)$ has more than one solution:
In this case, our system has more than one steady states. It means that we loose the unicity of the solution (so there is no more global actractiveness)

E.g:

$$\dot{X}_i = f(x_i, < x >) + g(x_i)\xi_i \quad N \gg 1$$

$$\dot{x} = f(x, M(t)) + g(x)\xi(t)$$

$$M(t) = \int zP(z,t)dt$$

The Fokker-Plank equation will be:

$$\frac{\partial P}{\partial t} = -\frac{\partial}{\partial x}[f(x, M(t))] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[g^2(x)P]$$

The steady state solutions will be the solutions of the following equation:

$$\begin{cases} 0 = -\dfrac{\mathrm{d}}{\mathrm{d}x}[t(x, M_s)P] + \dfrac{\mathrm{d}^2}{\mathrm{d}x^2}\left[\dfrac{g^2(x)}{2}P\right] \\ M_s = \int zP_s(z; M_s)dx \end{cases}$$

$$\boxed{M_s = \Psi(M_s)}$$

TODO: add linking sentence

$$P(x, M_s, \theta) = C(M, \theta)\exp\left[\frac{2}{\omega}\left(\theta M_s x + (1-\theta)\frac{x^2}{2} - \frac{x^4}{4}\right)\right]$$

$$\boxed{M_s = 0}$$

…(?)
$0 < \theta < \theta_c$
…(?)
So we have two solutions:

$$M_s = a$$
$$M_s = -a$$

Example:

$$M_s = \Psi(M_s; \theta)$$

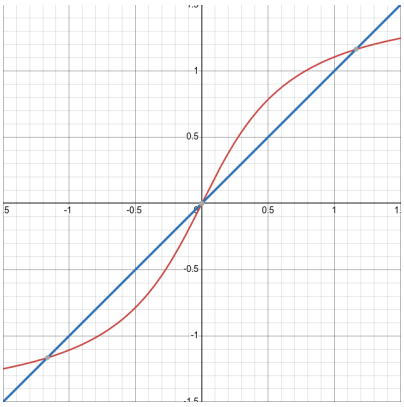
$$\begin{cases} y = M_s \\ y = \Psi(M_s; \theta) \end{cases} \quad \Rightarrow \quad \begin{matrix} P_s(x; M_1) \\ P_s(x; M_2) \\ P_s(x; M_3) \end{matrix}$$

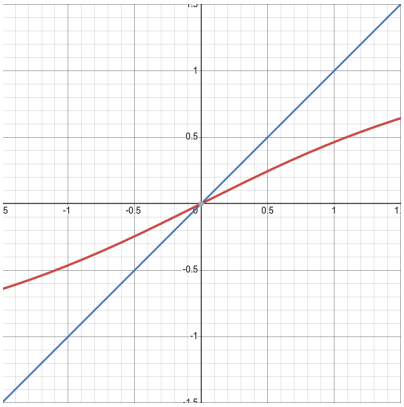
so for $\theta = \theta_1$ we have ***multistability***, while for $\theta = \theta_2$ we have ***monostability***.

Theorem 11.2.0.1.

$$\left| \frac{\mathrm{d}\Psi}{\mathrm{d}M_s}\Big|_{M_s=M_c} \right| < 1 \quad \Rightarrow \quad P_s(x; M_1, \theta^1) \text{ is locally stable}$$

$$\left| \frac{\mathrm{d}\Psi}{\mathrm{d}M_s}\Big|_{M_s=M_c} \right| > 1 \quad \Rightarrow \quad P_s(x; M_2, \theta^2) \text{ is locally unstable}$$





...

$$\dot{x} = (ax + x^3 - x^5) - D(x - M(t)) + \alpha(1 + x^2) \odot \xi(t)$$

$$M = \Psi(M; D, \alpha)$$

we have that for small D and α we have a unique solution, while for large D and α we have 5 different solutions.

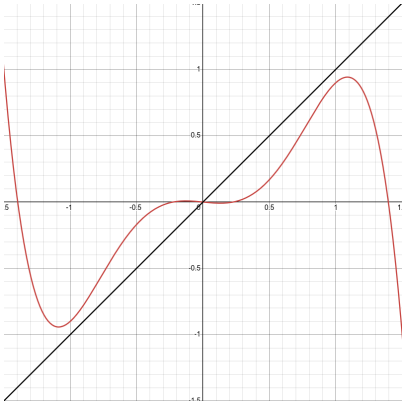


Figure 11.1: 1 solution

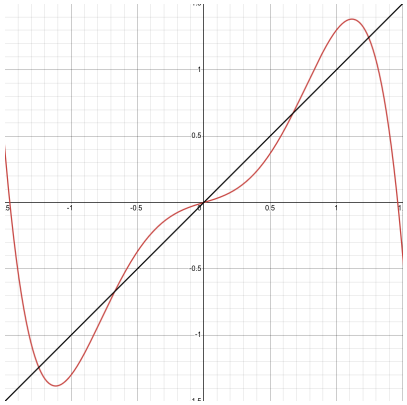


Figure 11.2: 5 solutions

So 0 is always a solution, and from a certain value of D we have 5 solutions, 3 of which are stable and 2 are unstable.

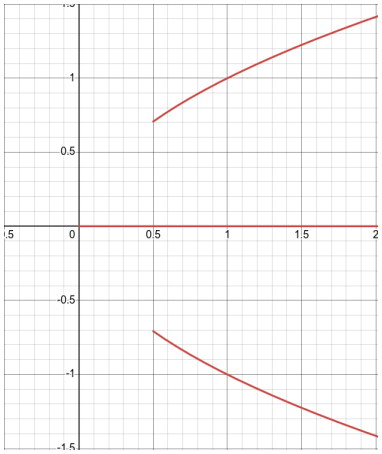


Figure 11.3: Stable solutions of the system

…

$$\dot{x} = F(x_i, < x >) + g(x_i, < x >)\xi_i(t)$$

An example iis the movement of a guitar string that vibrates.

$$m_i \ddot{x}_i = -\gamma \dot{x}_i$$

If the deviation is big, we have a function in the copmplex space, but if the deviation is small, we can use the linear approximation.

$$m_j \ddot{z}_j = -\gamma \dot{z}_j - k(z_j - z_{j-1}) - k(z_j - z_{j+1}) = -\gamma \dot{z}_j - k\,(z_{j-1} - 2z_j + z_{j+1})$$

we have now a discretization of the position of the string $z(t,x)$:

$$m_j \frac{\mathrm{d}^2 z}{\mathrm{d}t^2}(t,x_j) = -\gamma \frac{\mathrm{d}z}{\mathrm{d}t}(t,x_j) + k[z(t,x_j+D) - 2z(t,x_j) + z(t,x_j-D)]$$

$$\mu \frac{\partial^2 z}{\partial t^2} = -\gamma \frac{\partial z}{\partial t} + c\frac{\partial^2 z}{\partial x^2} + \hat{\omega}\xi(x,t)$$

We can see the stochastic term as the wind that moves the string.

11.3 Lecture: 09/05/2025

11.3.1 Spatiotemporal noisy model

$$\frac{\partial \phi}{\partial t} = f(\phi) + g(\phi)\xi_m(r,t) + D\mathscr{L}[\phi] + h(\phi)F(t) + \xi_a(r,t)$$

- $f(\phi)$: deterministic part
- $g(\phi)$: multiplicative noise
- $D\mathscr{L}[\phi]$: linear part
- $h(\phi)$: additive noise

L[ϕ] is a Laplacian or a integral operator
Examples:

- $$\mathscr{L}[\phi] = \nabla^2\phi$$
- $$\mathscr{L}[\phi] = -a_0\nabla^2\phi - \nabla^4\phi$$
- $$\mathscr{L}[\phi] = -(\nabla^2 + k_0^2)^2\phi = -(K_0^2 + 2k_0\nabla^2 + \nabla^4)\phi$$

Observation:

If we apply the fourier transform of:

$$\mathscr{L}[\phi] = -(\nabla^2 + k_0^2)^2\phi = -(K_0^2 + 2k_0\nabla^2 + \nabla^4)\phi$$

we get:

$$...F(\phi)$$

$$\mathscr{L}[\phi(r)] = \int \phi(r')\omega(r-r')\mathrm{d}r'$$

How do we simulate this equation? We can simply obtain the domain and discretize it.
Lattice.based Approximation:

...

Field coupling approximation:

$$l(\phi_i,\phi_j) = w_i\phi_i + \sum_{j\in nn(i)} w_j\phi_j$$

For example:

$$\mathscr{L}[\phi] = \nabla^2\phi \approx l(\phi_i,\phi_j) = \frac{1}{\Delta^2}\sum_{j\in nn(i)}(\phi_j-\phi_i)$$

If we have a stochastic process which is discrete in time and space, we have:

$$\langle \xi(r,t)\xi(r',t')\rangle = sC\left(\frac{|r-r'|}{d},\frac{|t-t'|}{\tau_c}\right)$$

As in the purely temporal noise, τ_c is a measure of the temporal memory of the noise, d is the spatial memory of the noise.
The spatiotemporal brother of the ornstein-uhlenbeck noise is "Ojalvo e al" noise.

$$\frac{\partial\phi}{\partial t} = a\phi + D\nabla^2\phi + \xi_{gn}$$

Observation: ***Ojalvo e al and the Ornstein-Uhlenbeck process***

If we set $D = 0$ we have a series of Ornstein-Uhlenbeck processes at each point of the domain.

$$\frac{\partial\phi}{\partial t} = a\phi + \xi_{gn}$$

Noise induced pattenrs §

$$\frac{\partial\phi}{\partial t} = f(\phi) + g(\phi)\xi_m(r,t) + D\mathscr{L}[\phi] + ...$$

…
Perturbed Swift-Hohenberg model:

$$\frac{\partial\phi}{\partial t} = a\phi + D\mathscr{L}[\phi] + \xi_{gn}...$$

$$\frac{\partial\phi}{\partial t} = f(\phi) + D\mathscr{L}[\phi] = a\phi - D(\nabla^2+k_0^2)^2\phi$$

Transitory pattern that disappear.
Additive noise generate pattenrs

$$\frac{\partial\phi}{\partial t} = a\phi + D\mathscr{L}[\phi] + \xi_{gn}, \qquad \mathscr{L}[\phi] = -(\nabla^2+k_0^2)^2\phi$$

Permanent patterning: details change in time.
we can distinguish two cases:

- $a < 0$
- $a > 0$

with multiplicative noise it can induce bimodality in the pdf of ϕ:

$$\frac{\partial \phi}{\partial t} = a\phi - phi^3 + \phi \xi_{gn} + D\mathscr{L}[\phi]$$

—

A bad model of glaciations

$$\mathrm{d}x = [x(a - x^2) + A\cos\Omega t]\mathrm{d}t$$

where

- x: is the (normalized) Earth's temperature
- $A\cos(\Omega t)$: small periodic variations of the solar irradiation.

We have that if A is small, $x(t)$ fluctuates around $+\sqrt{(a)}$.
The model fails.
Including stochastic noise:

$$\mathrm{d}x = [x(a - x^2) + A\cos\Omega t]\mathrm{d}t + \varepsilon \mathrm{d}W$$

where ε is the noise intensity.
This time we have a white noise, according to:

- ε is small: the noise is negligible
- ε is large: the noise is dominant

they finally managed to model the glaciations.

—

Spatial Stochastic Resonance

$$\frac{\partial \phi}{\partial t} = a\phi - \phi^3 + D\frac{\partial^2 \phi}{\partial x^2} + F(t) + \varepsilon \xi_{gn}$$

TODO: check the formula

…

12 Lecture 30/05/2025

12.1 Random Walks

Let's consider a particle that moves on a line.

$$\rho(x,t) = \Pr[\text{Particle is at location } (x, x+\mathrm{d}x) \text{ at time } t]$$

we have:

$$\frac{\partial \rho}{\partial t} = \int [\rho(y,t)\Omega(y,x) - \rho(x,t)\Omega(x,y)]\,\mathrm{d}y$$

Let's consider now the probability that the particle just jumped in the location $(x, x+\mathrm{d}x)$ at time t:

$$p(x,t) = \Pr[\text{Particle just jumped in the location } (x, x+\mathrm{d}x) \text{ at time } t]$$

We define a jump weight as:

$$\eta(y)$$

so:

$$p(x,t) = \int_0^t \int_{-\infty}^{+\infty} \mathrm{d}y \eta(y) p(x-y, t-T)\omega(t)\mathrm{d}t + \delta(x)\delta(t)$$

Definition: ***Fourier and Laplace Transforms***

- **Fourier transform**
  Let's recall the Fourier transform:

  $$\mathscr{F}_x[f(x)] = \int_{-\infty}^{+\infty} f(x) e^{-ikx} \mathrm{d}x$$

  we have:

  $$\mathscr{F}_x[f(x) \star g(x)] = \hat{f}(k)\hat{g}(k)$$

- **Laplace transform**
  Let's recall the Laplace transform:

  $$\mathscr{L}_x[f(x)] = \int_0^{\infty} f(x) e^{-st} \mathrm{d}t$$

  we have:

  $$\mathscr{L}_x[f(t) \star g(t)] = \tilde{f}(s)\tilde{g}(s)$$

$$p(x,t) = [p(x,t) \star_s \eta(x)] \star_t \omega(t)$$

Let's call $\bar{p}(k,s) = \hat{\bar{p}}(x,t)$, then we can write:

$$\bar{p}(k,s) = \hat{\eta}(k)\tilde{\omega}(s)\bar{p}(k,s) + 1 \qquad \Rightarrow \qquad \bar{p}(k,s)\{1 - \hat{\eta}(k)\tilde{\omega}(s)\} = 1$$

$$\bar{p}(k,s) = \frac{1}{1 - \hat{\eta}(k)\tilde{\omega}(s)}$$

$$\Psi(T) = \int_T^{+\infty} \omega(q)\mathrm{d}q = 1 - \int_0^T \omega(q)\mathrm{d}q$$

and we can rewrite $\rho(x,t)$ as:

$$\rho(x,t) = \int_0^t p(x,t-T)\Psi(T)\mathrm{d}T$$

The Laplace transform of $\Psi(T)$ is:

$$\mathscr{L}[\Psi(T)] = \mathscr{L}[1 - \int_0^T \omega(q)\mathrm{d}q] = \frac{1}{s} - \frac{1}{s}\tilde{\omega}(s)$$

$$\bar{\rho}(k,s) = \left(\frac{1 - \tilde{\omega}(s)}{s}\right)\left(\frac{1}{1 - \hat{\eta}(k)\tilde{\omega}(s)}\right)$$

Cauchy distribution (???)

$$\omega(T) \sim \frac{1}{T^{1+\alpha}} \qquad T \gg 1$$

$$\omega(T) = \frac{A}{1 + B\,T^2} \qquad T \gg 1 \qquad \Rightarrow \qquad \omega(T) \sim \frac{A}{B}\frac{1}{T^2}$$

Then its integral is infinite (cause compares a logarithm).

$$\frac{\partial \rho}{\partial t} = D\frac{\partial^2 \rho}{\partial x^2} \qquad \dot{x} = c\xi(t)$$

$$\langle x^2(t)\rangle \equiv t^{\alpha}$$

for $0 < \alpha < 1$ we have subdiffusion, for $1 < \alpha < 2$ we have superdiffusion.