Nonequilibrium Field Theories and Stochastic Dynamics

Zehong Liao

October 31, 2025

1 Preamble: A Review of Entropy Generation and the Signature of Irreversibility

We discussed the master equation that describes the time evolution of a system's state, and distinguished between two important states: Global Balance and Detailed Balance. We know that a system satisfying detailed balance must be in a steady state, but not all steady state systems satisfy detailed balance. This distinction is key to understanding non-equilibrium physics, as it leads us to the core concept of "irreversibility".

When a system is in a Non-Equilibrium Steady State (NESS), although its macroscopic properties (such as the probability distribution) do not change over time, there are continuous probability currents within the system. The existence of these net probability currents is precisely the manifestation of the system's irreversibility. To quantify this irreversibility, we need a physical quantity, which is the entropy production rate.

1.1 Quantifying Irreversibility: The Entropy Production Rate σ

The professor first gives the expression for the **entropy production rate (σ)** when reviewing the previous content:

$$\sigma = \sum_{i,j} \pi_j w_{ji} \ln \left(\frac{\pi_j w_{ji}}{\pi_i w_{ij}} \right) \tag{1}$$

Let's break down this formula:

- π_i: The steady-state probability of the system being in state i.
- w_{ij}: The transition rate from state j to state i.
- Therefore, $\pi_j w_{ji}$ represents the **probability flux** from state j to state i in the steady state.
- The argument of the logarithm, $\left(\frac{\pi_j w_{ji}}{\pi_i w_{ij}} \right)$, measures the **asymmetry** between the forward process ($j \to i$) and the reverse process ($i \to j$). If $\frac{\pi_j w_{ji}}{\pi_i w_{ij}} = 1$, it indicates the transition is reversible.

Thus, the entire expression σ is the sum of all possible flux-weighted asymmetries, and σ precisely quantifies the asymmetry or irreversibility of the system at the macroscopic level.

Next, the professor introduces an elegant transformation, yielding an alternative form for σ:

$$\sigma = \sum_{i,j} J_{ji} \ln \left(\frac{\pi_j w_{ji}}{\pi_i w_{ij}} \right) \tag{2}$$

where $J_{ji} = \pi_j w_{ji} - \pi_i w_{ij}$ is the net probability current.

This new form is conceptually profound. The term inside the logarithm becomes the ratio of the **forward flux** $\pi_j w_{ji}$ to the **reverse flux** $\pi_i w_{ij}$. This directly links the source of entropy production to the **net circulation** of probability flow inside the system.

Two Fundamental Properties of the Entropy Production Rate

Non-negativity: $\sigma \geq 0$. This is a consequence of **Jensen's inequality** in probability theory. Its profound physical meaning is that the entropy production rate of any thermodynamic system must be

greater than or equal to zero. This embodies the **Second Law of Thermodynamics** in the context of microscopic processes.

Connection to Equilibrium

- **When the system is exactly in Detailed Balance (DB), $\sigma = 0$.** In detailed balance, the forward flux and reverse flux between every pair of states are equal: $\pi_j w_{ji} = \pi_i w_{ij}$. This corresponds to **true thermodynamic equilibrium**.
- **When the system is in a Non-Equilibrium Steady State (NESS), $\sigma > 0$.** In this case, at least one **net probability current** exists: $J_{ji} \neq 0$.

Therefore, σ is not merely a "switch" indicating whether the system is in equilibrium; the **magnitude of the value** itself is a measure of degree. A larger σ means a stronger internal probability flow circulation, indicating that the system is further away from equilibrium (i.e., the irreversibility is stronger).

2 Ehrenfest Model: A Simple System for Exploring Complex Ideas

2.1 Introduction of the Classical Model

To further understand the nature of irreversibility, entropy, and equilibrium, the professor introduced a classical model from physics—the Ehrenfest Model. This model was first proposed by Paul and Tatyana Ehrenfest in the early 20th century to discuss the statistical nature of the Second Law of Thermodynamics, which is probabilistic, not absolute.

The Ehrenfest Model was developed to address a core paradox in statistical mechanics: How do the irreversible processes we observe in the macroscopic world (such as gas always diffusing from high to low pressure, consistent with the Second Law) arise from microscopic physical laws that are entirely reversible?

To do this, the physicists Ehrenfest introduced an extremely simplified "Urn Model" (or "Dog-Flea Model"). Imagine two boxes and N distinguishable particles. In each time step, one particle is randomly selected and moved to the other box. Even though every single transition is reversible at the microscopic level, the model clearly demonstrates that the system, with overwhelming probability, will evolve from a non-equilibrium initial state (like all particles in one box) to the most probable equilibrium state (particles roughly equally distributed in both boxes).

Therefore, this model is not intended to simulate a specific physical system, but serves as a classic thought experiment and teaching tool. It profoundly reveals that the nature of macroscopic irreversibility is fundamentally a statistical phenomenon, and provides the most intuitive and simplest mathematical example for understanding equilibrium, detailed balance, and how entropy drives a system toward this state.

This seemingly simple model, by simulating the random jumping of particles between two containers, helps us answer three fundamental questions:

- What does the system's equilibrium look like?
- How does the system approach equilibrium?
- What is the nature of equilibrium?

2.2 Formal Description of the "Dog-Flea" Model

The model has a vivid analogy: Imagine two dogs (A and B) and N fleas. The fleas jump randomly between the two dogs at certain rates. Our focus is on the number of fleas on dog A.

Formal Definition of the Model:

State: The system's macroscopic state is defined by the number of fleas on dog A, denoted by n, where n can take integer values $0, 1, \ldots, N$.

Transition Rates (Rates of Change):

- **Losing a Flea (A $\to$ B):** The rate of transition from state n to $n-1$ is $\omega_{n\to n-1} = \lambda n$.
- **Gaining a Flea (B $\to$ A):** The rate of transition from state n to $n+1$ is $\omega_{n\to n+1} = \lambda(N-n)$.

These rate settings are very intuitive:

- The rate at which dog A loses fleas should be proportional to the number of fleas already on it, n, because each of the n fleas has a probability λ to jump away in a unit of time.
- The rate at which dog A gains fleas should be proportional to the number of fleas on the other dog B, which is $(N-n)$.

3 The Essence of Balance: Careful Balance and the Binomial Distribution

3.1 Looking for the equilibrium state:

Now we turn to answer the first question: What is the **equilibrium state** (i.e., the steady-state probability distribution π_n) of this system?

Since there is **no external driving force** (such as a "flea pump") in this system to create continuous circulation, we expect the system will eventually settle into a **true thermodynamic equilibrium**. Such an equilibrium state must satisfy the stronger condition of **Detailed Balance**.

The Detailed Balance condition requires that, in the steady state, the forward probability flux and the reverse probability flux between any two adjacent states must be equal:

$$\pi_n \cdot \omega_{n\to n-1} = \pi_{n-1} \cdot \omega_{n-1\to n} \tag{3}$$

Substituting the specific transition rates of our model:

$$\pi_n \cdot (\lambda n) = \pi_{n-1} \cdot (\lambda(N-(n-1))) \tag{4}$$

3.2 Solving the recurrence relation

By solving the equation derived from the Detailed Balance condition, we can find the steady-state distribution π_n.

First, by cancelling the rate constant λ, we obtain a **recurrence relation** for π_n:

$$\pi_n = \frac{N-n+1}{n}\pi_{n-1} \tag{5}$$

We can **"unfold"** this relation to express π_n in terms of π_0:

$$\begin{aligned}\pi_n &= \frac{N-n+1}{n} \cdot \frac{N-n+2}{n-1} \cdot \frac{N-n+3}{n-2} \cdots \frac{N-0}{1} \cdot \pi_0 \\ &= \frac{N!/(N-n)!}{n!}\pi_0 \\ &= \frac{N!}{n!(N-n)!}\pi_0\end{aligned}$$

We recognize that the fractional part on the right-hand side is exactly the definition of the **binomial coefficient**, $\binom{N}{n} = \frac{N!}{n!(N-n)!}$. Therefore:

$$\pi_n = \binom{N}{n}\pi_0 \tag{6}$$

Finally, we use the **normalization condition** for probability, $\sum_{n=0}^{N} \pi_n = 1$, to determine the constant π_0:

$$\sum_{n=0}^{N} \binom{N}{n}\pi_0 = \pi_0 \sum_{n=0}^{N} \binom{N}{n} = 1$$

According to the **Binomial Theorem**, we know that $\sum_{n=0}^{N} \binom{N}{n} = 2^N$. Thus, $\pi_0 = 1/2^N$.
Substituting π_0 back, we obtain the final steady-state distribution:

$$\pi_n = \frac{1}{2^N} \binom{N}{n} \tag{7}$$

This is a **binomial distribution**.

3.3 Physical interpretation of the results

This result is highly intuitive. We can understand it from a more fundamental level: the **microstate**. A microstate precisely specifies the position of **every single flea** on a dog. If we assume each flea has two possibilities (+1 and −1 representing dog A and dog B, respectively), then there are 2^N possible microstates in total.

If we assume the system is in equilibrium, all 2^N microstates are **equally probable**, so the probability of each microstate is $1/2^N$. Now, how many microstates correspond to the macroscopic state n (i.e., having n fleas on dog A)? This is equivalent to the number of combinations of choosing n fleas out of N to be on dog A, which is $\binom{N}{n}$.

Therefore, the probability of the macroscopic state n is:

$$P_n = (\text{Number of microstates for state } n) \times (\text{Probability of each microstate}) = \binom{N}{n} \frac{1}{2^N} \tag{8}$$

This exactly matches the result we derived through dynamics (**Detailed Balance**). This demonstrates that the **dynamics of the Ehrenfest Model naturally leads to the most probable state**, and does not persist in any unlikely state.

The mean (average) of this binomial distribution is $\langle n \rangle = N/2$, and the standard deviation is $\sigma_n = \frac{1}{2}\sqrt{N}$. This means that in equilibrium, we are most likely to find the fleas **evenly distributed** on both dogs, with a **fluctuation** around this mean.

When we consider a large system, i.e., N becomes very large (the **thermodynamic limit**), a profound phenomenon emerges. The **relative size of the fluctuation** is:

$$\frac{\sigma_n}{\langle n \rangle} = \frac{\frac{1}{2}\sqrt{N}}{N/2} = \frac{1}{\sqrt{N}} \tag{9}$$

As $N \to \infty$, this ratio approaches zero. This means that, although the absolute size of the fluctuation ($\propto \sqrt{N}$) is increasing, it becomes insignificant relative to the mean. This is how a macroscopic, "sharp," and predictable equilibrium state (such as the temperature or pressure of a gas in a container) emerges from a microscopic, chaotic, and fluctuating world. The Ehrenfest model beautifully illustrates the **Law of Large Numbers** underlying the physical background.

4 The statistical arrow of time: reconciling microscopic reversibility with macroscopic irreversibility

4.1 The Core Paradox

Now we address the most central question. The microscopic dynamics of the model are **time-reversible**: a flea jumping from A to B and from B to A are processes that are physically equally plausible. So, why do we always observe **irreversible behavior** at the macroscopic level?

For example, a system with an initial state where all fleas are on dog A ($n = N$) will always evolve toward the nearly evenly distributed state ($n \approx N/2$), yet we virtually **never see the reverse process**—a system in the evenly distributed state spontaneously evolving to a state where all fleas gather on one side.

4.2 The story of two trajectories

The professor reveals the answer to this paradox by comparing the probabilities of two trajectories:

- **Trajectory 1 (Forward):** The system is in a **highly improbable state** $n = N$ at time t, and evolves to the **most probable state** $n \approx N/2$ at time $t + \tau$.
- **Trajectory 2 (Reverse):** The system is in the **most probable state** $n \approx N/2$ at time t, and evolves to the **highly improbable state** $n = N$ at time $t + \tau$.

Let's calculate the ratio of the joint probabilities of these two complete paths:

$$\frac{P\left(n = \frac{N}{2} \text{ at } t + \tau; n = N \text{ at } t\right)}{P\left(n = N \text{ at } t + \tau; n = \frac{N}{2} \text{ at } t\right)} \tag{10}$$

Using the definition of conditional probability, $P(A, B) = P(A|B)P(B)$, where $P(B)$ is the probability of the initial state, given by the steady-state distribution π. The above equation can be rewritten as:

$$\frac{P\left(\frac{N}{2}, t + \tau | N, t\right) \cdot \pi_N}{P\left(N, t + \tau | \frac{N}{2}, t\right) \cdot \pi_{N/2}} \tag{11}$$

Here, the professor made a **key approximation**: for a system described by a master equation, the conditional probability of transition is symmetric in some sense, meaning the conditional probability of a transition is approximately equal to the conditional probability of the reverse process. Therefore, $P(N/2, t + \tau|N, t) \approx P(N, t + \tau|N/2, t)$. This simplifies the entire ratio to the ratio of the steady-state probabilities of the two end states:

$$\approx \frac{\pi_N}{\pi_{N/2}}$$

Substituting the binomial distribution result we obtained:

$$\frac{\pi_N}{\pi_{N/2}} = \frac{\frac{1}{2^N}\binom{N}{N}}{\frac{1}{2^N}\binom{N}{N/2}} = \frac{\binom{N}{N}}{\binom{N}{N/2}} = \frac{1}{\binom{N}{N/2}} = \frac{(N/2)!(N/2)!}{N!} \tag{12}$$

Using **Stirling's approximation** ($N! \approx \sqrt{2\pi N}\left(\frac{N}{e}\right)^N$), this ratio can be calculated to be approximately 2^{-N}. Even for a moderately sized system, such as $N = 50$, this ratio is an **astronomically small number** ($\approx 10^{-15}$).

This means that a path evolving from an **ordered state** ($n = N$) to a **disordered state** ($n = N/2$) is many orders of magnitude more likely to occur than the time-reversed path.

4.3 Resolution of the Paradox: Microstate vs. Macrostate

The key here is to distinguish between microstates and macrostates. Macroscopic irreversibility does not originate from a time asymmetry in the underlying physical laws, but from pure, overwhelming statistics.

- The macrostate $n = N$ corresponds to only 1 microstate (all eggs are in basket A).
- The macrostate $n = N/2$ corresponds to $\binom{N}{N/2}$ microstates, which is a huge number.

The system is not "pulled" toward the equilibrium state by some mysterious force; it is merely exploring all 2^N possible microstates equally. Because the number of microstates corresponding to the equilibrium macrostate is so vast, the system will spend most of its time in one of these microstates. Therefore, starting from a macrostate with only a single microstate, the system will almost inevitably evolve toward a macrostate possessing a huge number of microstates. The reverse process is not impossible, but its probability of occurrence is so small that it is virtually unobservable, even over the timescale of the age of the universe.

This leads to a profound conclusion about the arrow of time. "An evolutionary direction determined by statistics stems from the assumption that a system in a highly improbable, ordered state is 'prepared' by an external influence." The only reason we perceive the arrow of time is that we (or nature) are capable of creating a low-entropy, highly ordered initial state. Starting from this special point, overwhelming probability dictates that the system must proceed toward disorder. If we observe a system already in equilibrium, we will see small, symmetric fluctuations around the mean value. At this

point, "a movie about these fluctuations, played forward or backward, is statistically indistinguishable," because "fluctuations near equilibrium have no sense of time direction." Therefore, the arrow of time is not an intrinsic property of the dynamics itself, but arises from a very special, non-equilibrium initial condition.

5 Simulating the Ehrenfest model with Python

We will use the Gillespie algorithm, which is the standard method for simulating such continuous-time Markov processes. Its core idea is as follows:

- Calculate the total exit rate from the current state.
- Use this rate to generate a random waiting time until the next jump occurs.
- Randomly decide which type of jump occurs (either losing or gaining a flea) based on the relative rates of the individual transition pathways."

5.1 Simulate a random trajectory

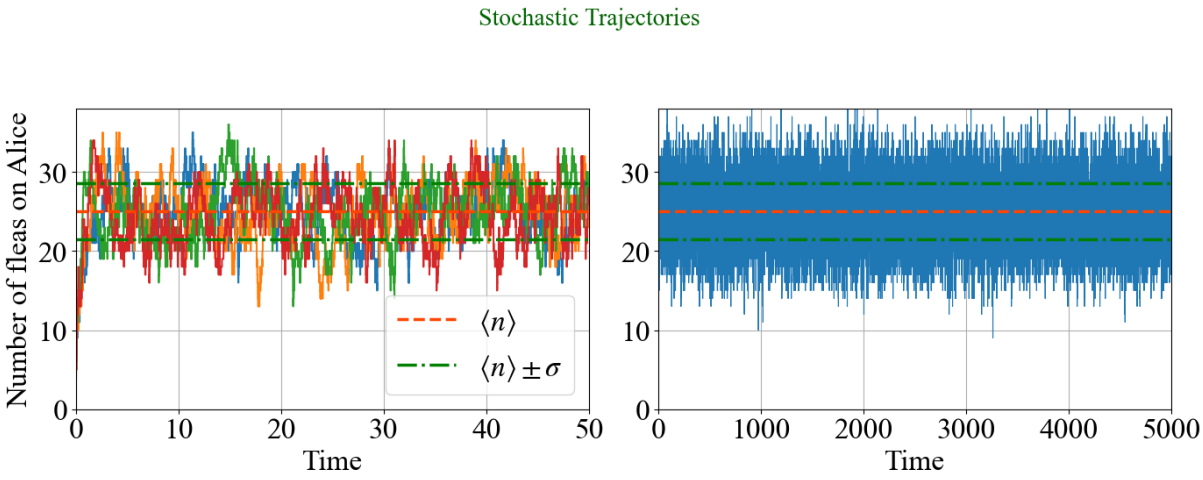

Figure 1:

The system rapidly relaxes from the initial highly ordered state ($n = 50$) towards the mean ($n = 25$), and subsequently undergoes random fluctuations around the mean, with the range of fluctuations roughly determined by the standard deviation.

5.2 Verify the equilibrium distribution

A single trajectory demonstrates the dynamic process, but to verify our statistical theory, we need to perform long-time simulations and calculate the proportion of time the system spends in each state (n). This proportion should converge to the binomial distribution π_n we derived theoretically

The resulting probability distribution matches the theoretical binomial distribution curve, verifying that our previous theoretical derivation is correct.

6 Quantifying the Journey Toward Equilibrium: Statistical Entropy and KL Divergence

6.1 Define time-dependent entropy

In statistical mechanics, the most fundamental definition of entropy is the Gibbs/Shannon entropy, which measures the degree of uncertainty or "spread" of the probability distribution over all possible

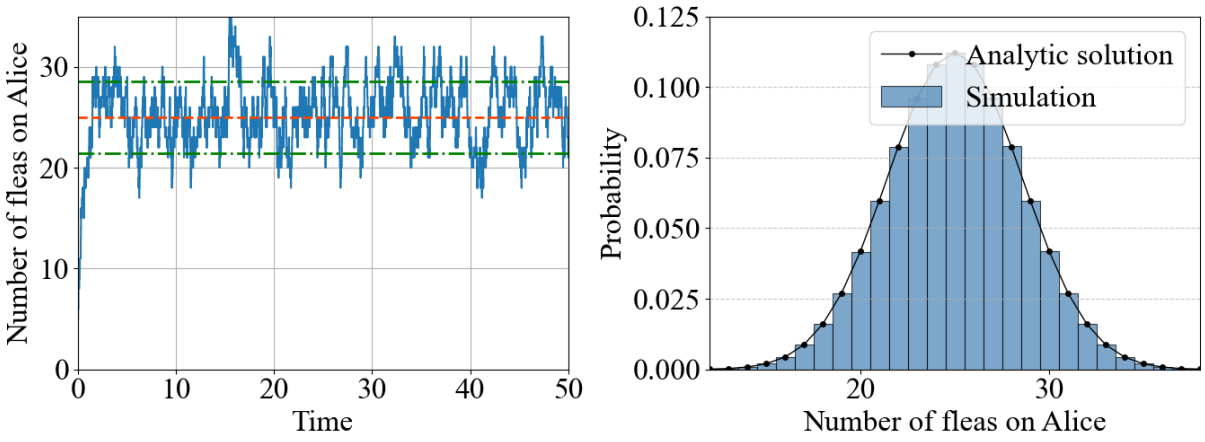

Figure 2:

microstates:

$$S(t) := -\sum_{\{\sigma_i\}} p(\{\sigma_i\}, t) \ln p(\{\sigma_i\}, t) \tag{13}$$

where the summation is over all 2^N microstates $\{\sigma_i\}$, and $p(\{\sigma_i\}, t)$ is the probability that the system is in a particular microstate at time t.

6.2 From Microscopic Entropy to Macroscopic Entropy

Directly dealing with the probability of microstates is often difficult. The professor demonstrated how to transition from the microscopic definition to the entropy expressed using the macroscopic state probability $p_n(t)$. The core assumption is: at any time t, all $\binom{N}{n}$ microstates belonging to the same macrostate n are equally probable. Therefore, the relationship between the macroscopic state probability and the microscopic state probability is:

$$p_n(t) = \binom{N}{n} p(\{\sigma_i\}, t) \tag{14}$$

Substituting this relationship into the definition of entropy and simplifying yields the expression for the macroscopic entropy:

$$S(t) = -\sum_{n=0}^{N} p_n(t) \ln p_n(t) + \sum_{n=0}^{N} p_n(t) \ln \binom{N}{n} \tag{15}$$

This formula reveals that the total entropy is composed of two parts:

[label=2., itemsep=1ex]

1. **First term** $-\sum p_n \ln p_n$: This is the entropy (uncertainty) concerning which **macrostate** the system is in.

2. **Second term** $\sum p_n \ln \binom{N}{n}$: This is the average of the **microstate** uncertainty, given the macroscopic state distribution $p_n(t)$. Each term $\ln \binom{N}{n}$ is the Boltzmann entropy of macrostate n (proportional to the logarithm of the number of microstates), and the entire sum is the expected value of this Boltzmann entropy under the current macroscopic state distribution.

6.3 KL Divergence: Measuring the "Distance" to Equilibrium

To describe the difference between the system's current state $p_n(t)$ and the final equilibrium state π_n, the professor introduced a powerful tool from information theory—the Kullback–Leibler Divergence

(KL Divergence):

$$H(t) = \sum_n p_n(t) \ln\left(\frac{p_n(t)}{\pi_n}\right) \tag{16}$$

The KL Divergence $H(t)$ is an asymmetric measure used to quantify the "degree of deviation" of the probability distribution $p_n(t)$ relative to the reference distribution π_n. It is non-negative ($H(t) \geq 0$), and equals zero if and only if the two distributions are completely identical (i.e., $p_n(t) = \pi_n$ for all n). In our physical context, $H(t)$ quantitatively describes the "information distance" of the system's current probability distribution from its final equilibrium distribution.

The deepest connection is that a simple relationship exists between the KL Divergence and the macroscopic entropy we defined:

$$H(t) = S_\infty - S(t) \tag{17}$$

Where S_∞ is the system's maximum entropy when it reaches equilibrium. This relationship shows that the process of $H(t)$ monotonically decreasing over time, eventually reaching zero, is entirely equivalent to the process of the system's entropy $S(t)$ monotonically increasing over time and finally reaching its maximum value S_∞. This provides a formal, quantitative description of the Second Law of Thermodynamics (the principle of entropy increase), known as the **H-Theorem**.

6.4 Simulating the Evolution of Entropy

To visualize the evolution of entropy and KL divergence, we need to simulate an ensemble, which is a large number of independent system replicas. By taking statistics on the states of the ensemble at different times, we can approximate the probability distribution that changes over time.

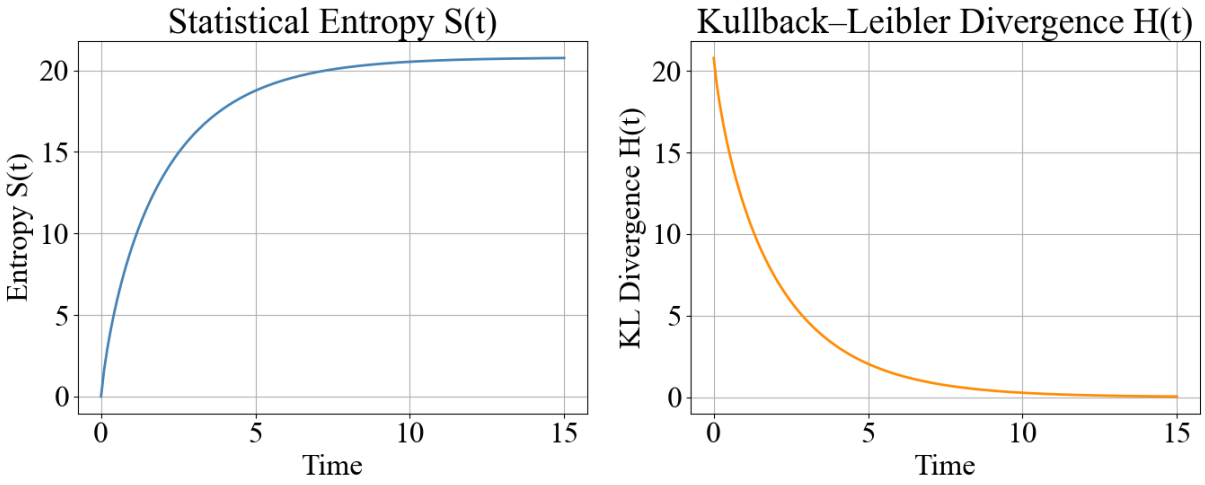

Figure 3:

One figure shows the entropy $S(t)$ monotonically increasing from a low initial value (corresponding to a highly ordered initial state) and eventually saturating at a maximum value; the other figure shows the KL divergence $H(t)$ starting from a positive value, monotonically decreasing, and eventually approaching zero. These two figures vividly visualize the operation of the Second Law of Thermodynamics within the Ehrenfest Model.

7 Conclusion: Fundamental Insights from Statistical Mechanics

A macroscopically imbalanced initial state relaxes rapidly to a less ordered state, which is close to equilibrium.This is precisely what we showed by calculating the probability ratio $\pi_{n/2}/\pi_n \sim 2^{-N}$. The system moves towards equilibrium not because of some "force," but because there are simply far more microstates corresponding to the equilibrium state.

In the equilibrium state, fluctuations have no time direction.This is closely related to the concept of detailed balance. Near the equilibrium state, the system jumps back and forth between states with similar probabilities, and the forward and reverse processes are statistically indistinguishable.

Large fluctuations away from the equilibrium state (i.e., states that are extremely improbable) are extremely rare.This is a direct inference from the nature of the binomial distribution. Since the probability distribution forms a sharp peak around the mean, the probability of states far from the mean drops exponentially, making the possibility of spontaneous macroscopic ordering practically zero.

The statistically defined arrow of time comes from the assumption that a system starting in an extremely improbable, ordered state was "prepared" by an external influence. Even for the small system we consider, using the words "improbable" and "extremely" is being very conservative.The existence of a time direction determined by statistics stems from our assumption that the system was "prepared" in a highly improbable, ordered state. This is the deepest insight into the origin of the "arrow of time." We observe irreversibility because we always start observing from a special, initial condition. The arrow of time is not inherent to the laws of physics but is a consequence of the initial condition of the universe (or our experiment).In conclusion, although the Ehrenfest Model is remarkably simple, it allows us to clearly observe and understand the cornerstones of statistical mechanics—the statistical origins of irreversibility, entropy, and the arrow of time.

8 Master Equation Review

8.1 Chapman-Kolmogorov Equation

The theoretical foundation of all Markov processes stems from the Markov property: the future state of a system depends only on its current state and is independent of its past history. The mathematical manifestation of this property is the Chapman-Kolmogorov equation:

$$p(n,t|n_0,t_0) = \sum_{n'} p(n,t|n',t')p(n',t'|n_0,t_0), \quad t_0 \le t' \le t \tag{18}$$

The physical meaning of this equation is: the total probability of starting from the initial state n_0 at time t_0 and arriving at state n at time t is equal to the sum of the probabilities of all possible intermediate states n'. It describes how probability propagates through intermediate times t'.

8.2 Mater equation

When we write the forward Kolmogorov equation (i.e., the Master Equation) in differential form, we obtain the master equation that describes the continuous evolution of the probability over time. Its general form is:

$$\frac{d}{dt}\mathbf{p}(t) = \mathbf{Q}\mathbf{p}(t) \tag{19}$$

where $\mathbf{p}(t)$ is a column vector whose element $p(n,t)$ is the probability that the system is in state n at time t, and $\mathbf{Q}$ is the transition rate matrix. More specifically, the master equation can be written in the following "gain-loss" form:

$$\frac{\partial p(n,t)}{\partial t} = \sum_m w(n,m)p(m,t) - w(n)p(n,t) \tag{20}$$

Here, $w(n,m)$ is the rate of transition from state m to state n. The physical meaning of the equation is very intuitive: the rate of change of the probability $p(n,t)$ of being in state n is equal to the total rate of flow *into* state n from all other states m (the gain term), minus the total rate of flow *out* of state n to all other states (the loss term). Where the total outflow rate is $w(n) = \sum_m w(m,n)$. By rearranging the terms, we can also obtain the "net-flow" form:

$$\frac{\partial p(n,t)}{\partial t} = \sum_m [w(n,m)p(m,t) - w(m,n)p(n,t)] \tag{21}$$

This form emphasizes the net probability flow between every pair of states (n,m).

8.3 Irreducible Markov processes

For a Markov process with a finite and irreducible state space (i.e., any state can be reached from any other state in a finite number of steps), we have the following important conclusions:

1. There exists a unique stationary distribution $\mathbf{p}^{ss}$, where all its elements are positive.
2. If the process is reversible, it satisfies the detailed balance condition:

$$w(n,m)p^{ss}(m) = w(m,n)p^{ss}(n)$$

   In the stationary state, the net probability flow between any two states is zero.
3. If the process is irreversible (non-reversible), there will be persistent probability currents in the stationary state, forming cyclic flows.
4. Irreversibility can be quantified through entropy production, which reflects the asymmetry between the probabilities of the forward and reverse trajectories.

These concepts, particularly detailed balance, are not only key to understanding equilibrium physics but also form the theoretical foundation for a powerful computational method we will introduce next: Markov Chain Monte Carlo (MCMC).

9 Markov Chain Monte Carlo (MCMC)

The Markov Chain Monte Carlo (MCMC) method was pioneered by Nicholas Metropolis and others in 1953 for solving nuclear physics problems (the Metropolis algorithm), and was later generalized by statisticians such as W.K. Hastings.

The origin of the MCMC method lies in a fundamental, thorny problem in physics and statistics: how to compute the average value of a physical quantity or sample from a complex probability distribution in an extremely huge or infinite state space. Directly seeking all states and calculating the sum or integral is literally impossible. The core idea for solving this difficulty is to "trade time for space": it gives up exhaustive exploration of the space and instead designs a clever random process, a **Markov chain**.

This chain is precisely constructed so that it eventually "forgets" its initial state and converges to a unique stationary distribution, which is precisely the target probability distribution we want to study (e.g., the Boltzmann distribution in physics). Therefore, we only need to let this random process run for a sufficient amount of time, then collect the visited state samples. These samples naturally represent the target distribution, allowing us to approximate the true system average using a simple average of these samples.

Its specific applications are vast and have long surpassed its physics origins:

1. **Statistical Physics:** This is the classical application, used for simulating the Ising model, crystal field theory, etc., to calculate quantities like magnetization rate or macroscopic heat capacity.
2. **Bayesian Statistics and Machine Learning:** MCMC is the cornerstone of modern Bayesian inference. When the posterior probability distribution of a model is complex and cannot be solved analytically, MCMC methods (such as Gibbs sampling and Metropolis-Hastings) are used to draw samples, estimate model parameters, and perform model selection and prediction.
3. **Computational Biology:** Used for modeling the evolutionary process of species (e.g., generating phylogenetic trees), performing sequence alignment, and simulating protein folding.
4. **Financial Engineering:** Used for modeling complex random processes in financial derivatives and risk value (VaR) calculations.

9.1 Core idea: Sampling using Markov chains

In statistical physics, we frequently need to calculate the system average value of a physical quantity A, such as magnetization or energy. Its definition is:

$$\langle A(X)\rangle = \sum_i A(X_i)p_i \tag{22}$$

where X_i is a microstate of the system (e.g., a specific spin configuration), and p_i is the probability of that state occurring. For a system in thermal equilibrium, this probability is given by the Boltzmann distribution in the canonical ensemble:

$$p_i = \frac{1}{Z}e^{-E_i/k_BT} \tag{23}$$

Here E_i is the energy of state i, T is the temperature, k_B is the Boltzmann constant, and Z is the partition function.

The problem is that for a system involving a large number of particles (such as a 50×50 Ising model, with 2^{2500} states), directly computing this sum is absolutely impossible.

The **Markov Chain Monte Carlo (MCMC) method** provides an elegant solution. Its core idea is: we no longer try to exhaust all states, but instead design an auxiliary Markov process whose stationary distribution is exactly the target distribution we wish to sample. Then, we let this Markov process evolve for a sufficient amount of time, and the sequence of visited states $X_1, X_2, \ldots, X_N$ will be samples drawn as if from p_i. Thus, the complex ensemble average calculation is transformed into a simple arithmetic average:

$$\langle A(X)\rangle \approx \frac{1}{N}\sum_{k=1}^{N} A(X_k) \tag{24}$$

9.2 Metropolis algorithm

How to ingeniously design a Markov process such that its stationary distribution is exactly the Boltzmann distribution? The Metropolis algorithm provides a very elegant and universal answer.

1. **Goal:** Construct a Markov process with transition rates $w_{i\to j}$ such that the stationary distribution p_i^{therm} is precisely the Boltzmann distribution:

$$p_i^{therm} \propto e^{-E_i/k_BT} \tag{25}$$

2. **Sufficient Condition:** Based on our previous discussion, detailed balance is a sufficient condition that guarantees the system reaches the required stationary state. We only need the transition rates to satisfy:

$$w_{i\to j}p_i^{therm} = w_{j\to i}p_j^{therm} \tag{26}$$

3. **Derivation:** Substitute the expression for p_i^{therm} into the detailed balance condition, and we obtain a requirement for the ratio of the transition rates:

$$\frac{w_{i\to j}}{w_{j\to i}} = \frac{p_j^{therm}}{p_i^{therm}} = \frac{e^{-E_j/k_BT}}{e^{-E_i/k_BT}} = e^{-(E_j-E_i)/k_BT} \tag{27}$$

4. **Significance:** This relationship is the core of the Metropolis algorithm. It links the required dynamic rate (transition rate) to the energy landscape of the system (the energy difference ΔE).

5. **Metropolis's Choice:** There are many ways to satisfy the ratio $\frac{w_{i\to j}}{w_{j\to i}}$. Metropolis proposed a very simple and effective method: the **Propose / Accept-Reject (propose/accept/reject)** algorithm.

   The transition rate $w_{i\to j}$ is decomposed into two steps: a proposal step and an acceptance step:

$$w_{i\to j} = Q_{i\to j}P_{\text{accept}}(i \to j)$$

- **Proposal:** From the current state i, randomly propose a new state j. The simplest method is a small random perturbation on state i (e.g., flipping one spin in the Ising model). Assume the probability of proposing j from i and i from j are equal (i.e., the proposal distribution is symmetric): $Q_{i\to j} = Q_{j\to i}$.
- **Accept-Reject Criterion:** The acceptance probability $P_{\text{accept}}(i \to j)$ is calculated based on the energy change $\Delta E = E_j - E_i$.

  · If $\Delta E \leq 0$ (the new state is lower or equal in energy), the proposal is **always accepted**, and the system transitions to state j.

  · If $\Delta E > 0$ (the new state is higher in energy), the proposal is accepted with a certain probability $P_{\text{accept}} = e^{-\Delta E/k_B T}$. The specific operation is: generate a random number r between $[0, 1]$. If $r \leq P_{\text{accept}}$, the proposal is accepted; if $r > P_{\text{accept}}$, the proposal is rejected, and the system **remains in its original state i**.

This acceptance probability, $P_{\text{accept}} = \min(1, e^{-\Delta E/k_B T})$, happens to satisfy the detailed balance requirement. It intuitively reflects the physics: the system favors transitioning to lower energy states, but due to the presence of thermal fluctuations, it also has a probability of jumping to higher energy states, thus allowing it to explore the entire state space instead of merely falling into local energy minima.

9.3 Python simulation: Two-dimensional Ising model

To specifically feel the power of the Metropolis algorithm, we use it to simulate a classic physical system: the **Two-Dimensional Ising Model (2D Ising Model)**. This is a simplified model describing ferromagnetism, consisting of a two-dimensional lattice where each lattice site has a spin, with states being either "up" $(+1)$ or "down" (-1).

The total energy (Hamiltonian) of the system is given by the following equation:

$$H = -J \sum_{\langle i,j \rangle} S_i S_j \tag{28}$$

where $\sum_{\langle i,j \rangle}$ denotes the sum over all nearest-neighbor pairs; S_i and S_j are the spin values on neighboring lattice sites. We set the coupling constant $J = 1$. When the neighboring spins are in the same direction ($S_i S_j = 1$), the energy contribution is $-J$, and the system energy is lowered; when the directions are opposite ($S_i S_j = -1$), the energy contribution is $+J$, and the system energy is raised. Therefore, at low temperatures, the system tends towards all spins aligning, forming macroscopic magnetization, i.e., the ferromagnetic state.

Below is the Python code that implements the Metropolis algorithm to simulate the 2D Ising model.

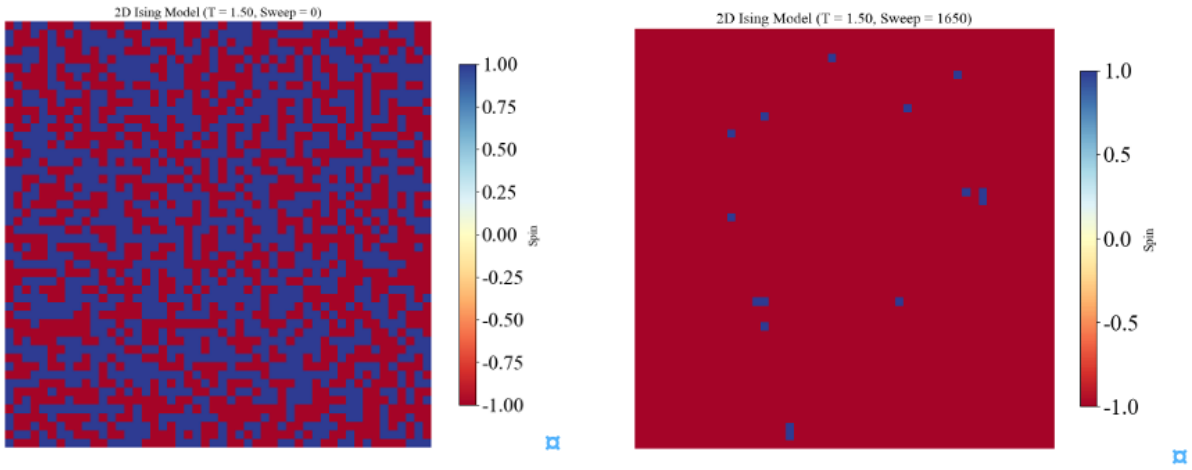

Figure 4:

When the temperature T is set below the critical temperature $T_C \approx 2.269$ (for example, $T = 1.5$), you will observe the system gradually evolving from an initial random, "checkerboard" state into large domains of aligned spins (or "clumps"). Eventually, the entire system will converge to one of the two almost perfectly ordered ferromagnetic states (either all blue or all red). This process vividly demonstrates how the Metropolis algorithm, by following simple **local rules** (which satisfy the detailed balance acceptance/rejection criterion), successfully guides the system to find its **global low-energy equilibrium**, resulting in the emergence of macroscopic ordered structures. This is precisely the essence of statistical mechanics: deriving macroscopic behavior from microscopic rules.

10 Markov processes in continuous state space

Now, we are ready to make the core leap of this lesson: generalizing the Markov process description from discrete states (like spin states ± 1) to a continuous state space. This means that the system's state variable, such as $\vec{X}(t) \in R^d$, can take values in a continuous domain. To build an intuitive understanding, the lecturer presented several classic examples from physics and biology.

1. **Brownian Motion**

   A pollen grain suspended in a liquid undergoes perpetual, irregular motion due to numerous small, random collisions with water molecules. The position $\vec{X}(t) \in R^3$ of this particle is a continuous state variable. Its trajectory is continuous but nowhere differentiable, full of jagged, zigzag random steps.

2. **Molecular Motors** Take the motor protein **Kinesin** as an example. It acts like a tiny mover, walking along the microtubule track inside the cell to transport materials. Each of its steps is a fixed size (about 8 nm), and therefore, on a macroscopic scale, its movement can be coarsely described by a position variable $x(t)$ on the microtubule. Although the underlying chemical reaction (ATP hydrolysis) is a discrete event, its position trajectory mixes discrete jumps with thermal fluctuations near each binding site.

3. **Bacterial Chemotaxis**

   When the E. coli bacterium searches for food in its environment, it employs a motion strategy called "Run & Tumble." During the "Run" phase, it moves straight ahead along one direction; during the "Tumble" phase, it randomly changes direction in place. The state variables required to describe this process are at least the position $x(t)$ and the orientation $\theta(t)$, both of which are continuous. Its trajectory is composed of smooth straight segments and instantaneous changes in direction.

These three examples were not chosen at random; they ingeniously foreshadow the different mathematical models we are about to establish. Brownian motion is the archetype of a **pure diffusion process**, whose sample paths are continuous. The movement of Kinesin is representative of a **jump process**, whose sample paths are piecewise constant and discontinuous. The bacterium's motion, however, is a **mixed process**, containing both continuous motion and discrete events. These physical pictures provide concrete physical scenarios for the abstract mathematical equations we will derive next.

11 Sample path and evolution equation

The fundamental change from discrete to continuous states lies in how we describe the "transition" between states. For a continuous space, we no longer use the probability $P_n(t)$, but instead use the **Probability Density Function (PDF)** $P(x,t)$. The quantity $P(x,t)dx$ represents the probability of finding the system within an infinitesimal volume element dx centered at x at time t.

11.1 Two properties of a path

Based on the nature of the system's state trajectory (sample path) evolution over time, we can divide Markov processes in continuous space into two major categories. The mathematical criterion for this

classification (the **Kramers-Moyal conditions**) is to look at how the probability of the system making a displacement larger than an arbitrarily small amount ϵ changes over an infinitesimal time interval Δt.

1. **Continuous Sample Path**

   For this class of processes, the change in the system's state is smooth, and instantaneous, finite-sized jumps do not occur. Mathematically, this means the probability of a large jump occurring in a small time interval Δt approaches zero faster than Δt itself:

   $$\lim_{\Delta t \to 0} \frac{1}{\Delta t} \int_{|x-x'|>\epsilon} dx\, p(x, t+\Delta t \mid x', t) = 0 \tag{29}$$

   Brownian motion is the classic example. Its transition probability density is a Gaussian distribution (Normal distribution), with the variance proportional to Δt:

   $$p(x, t+\Delta t \mid x', t) = \frac{1}{\sqrt{4\pi D \Delta t}} \exp\left(-\frac{(x-x')^2}{4D\Delta t}\right) \tag{30}$$

   As $\Delta t \to 0$, this distribution concentrates infinitely at the starting point x', which indicates that the probability of a large jump occurring is extremely low.

2. **Discontinuous Sample Path**

   For this class of processes, the system can experience instantaneous, finite-sized jumps. In this case, in the limit $\Delta t \to 0$, the transition probability density, when divided by Δt, converges to a finite **transition rate density** $w(x \mid x')$:

   $$w(x \mid x') = \lim_{\Delta t \to 0} \frac{1}{\Delta t} p(x, t+\Delta t \mid x', t) \tag{31}$$

   $w(x \mid x')dx$ can be understood as the rate at which the system jumps from the current state x' per unit time into a small volume element dx centered at x. The stepwise movement of the Kinesin motor protein belongs to this category.

11.2 Derivation of the evolution equation

Our goal is to find an equation that describes the time evolution of the probability density $P(\vec{x}, t)$, which will be the continuous-space version of the "Master Equation." The starting point for the derivation is to examine the rate of change of the expected value of any smooth observable $f(\vec{x})$ over time, $\partial_t \langle f \rangle (t)$:

$$\partial_t \langle f \rangle (t) = \partial_t \int d\vec{x}\, f(\vec{x}) P(\vec{x}, t) \tag{32}$$

According to the definition of the derivative, we have:

$$\partial_t \langle f \rangle (t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left[\int d\vec{x}\, f(\vec{x}) P(\vec{x}, t+\Delta t) - \int d\vec{x}\, f(\vec{x}) P(\vec{x}, t) \right] \tag{33}$$

The key step here is to use the forward Kolmogorov equation to express $P(\vec{x}, t+\Delta t)$:

$$P(\vec{x}, t+\Delta t) = \int d\vec{x}'\, p(\vec{x}, t+\Delta t \mid \vec{x}', t) P(\vec{x}', t) \tag{34}$$

Substituting this identity, and utilizing the probability conservation $\int d\vec{x}'\, p(\vec{x}', t+\Delta t \mid \vec{x}, t) = 1$, and after some skillful transformation, we can obtain a single, unified starting point. Next, we will treat this expression differently based on whether the sample paths are "jump" or "diffuse" in nature.

11.3 Case 1: Pure jump process

For discontinuous sample paths (jump processes), we use the transition rate density $w(\vec{x} \mid \vec{x}')$ defined earlier. Substituting it into the expression for $\partial_t \langle f \rangle (t)$ and exchanging the order of integration variables, we ultimately obtain:

$$\partial_t \langle f \rangle (t) = \int d\vec{x}\, f(\vec{x}) \left[\int d\vec{x}'\, w(\vec{x} \mid \vec{x}') P(\vec{x}', t) - \int d\vec{x}'\, w(\vec{x}' \mid \vec{x}) P(\vec{x}, t) \right] \tag{35}$$

Since this equation must hold for any arbitrarily chosen function $f(\vec{x})$, we can directly remove the outer layer of $\int d\vec{x}\, f(\vec{x})$, thereby obtaining the evolution equation for the probability density $P(\vec{x}, t)$ itself:

$$\partial_t P(\vec{x}, t) = \int d\vec{x}' \left[w(\vec{x} \mid \vec{x}') P(\vec{x}', t) - w(\vec{x}' \mid \vec{x}) P(\vec{x}, t) \right] \tag{36}$$

This is the **Master Equation in Continuous Space**. Its structure is exactly the same as the discrete Master Equation: the first term is the **gain term** (jumping into $\vec{x}$ from all other positions $\vec{x}'$), and the second term is the **loss term** (jumping out of $\vec{x}$ to all other positions $\vec{x}'$). The only difference is that the summation $\sum_m$ in the discrete case is replaced by the integral $\int d\vec{x}'$ in the continuous case.

11.4 Case 2: Pure diffusion process and Fock-Planck equation

For continuous sample paths, the key is that the system's state only shifts locally in Δt. Therefore, we can return to the expectation value $\partial_t \langle f \rangle (t)$ and perform a Taylor series expansion of the function $f(\vec{x})$ around $\vec{x}'$:

$$f(\vec{x}) = f(\vec{x}') + (\vec{x} - \vec{x}') \cdot \partial_{\vec{x}'} f(\vec{x}') + \frac{1}{2} (\vec{x} - \vec{x}')(\vec{x} - \vec{x}') : \partial_{\vec{x}'} \partial_{\vec{x}'} f(\vec{x}') + \dots \tag{37}$$

Substituting this into the integral expression and realizing that the limit of terms corresponding to displacements $\Delta\vec{x} = \vec{x} - \vec{x}'$ that are higher than the second order are negligible in the $\Delta t \to 0$ limit, we define two crucial physical quantities:

1. **Drift Vector** $A(\vec{x}, t)$:

$$A(\vec{x}, t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int d\Delta\vec{x}\, (\Delta\vec{x}) p(\vec{x} + \Delta\vec{x}, t + \Delta t \mid \vec{x}, t) \tag{38}$$

   This describes the average velocity or the systematic "drift" experienced by the particle starting at $\vec{x}$ due to systematic forces.

2. **Diffusion Tensor** $B(\vec{x}, t)$:

$$B(\vec{x}, t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \int d\Delta\vec{x}\, (\Delta\vec{x} \Delta\vec{x}^{\mathrm{T}}) p(\vec{x} + \Delta\vec{x}, t + \Delta t \mid \vec{x}, t) \tag{39}$$

   This describes the magnitude and correlation of the random displacement, giving rise to the diffusion equation. It represents the random force (noise) that pushes the particle in different directions.

By substituting the expansion and utilizing integration by parts on the derivatives of the test function $f(\vec{x})$, we finally arrive at the time evolution equation for the probability density $P(\vec{x}, t)$: the **Fokker-Planck Equation (FPE)**:

$$\partial_t P(\vec{x}, t) = -\partial_{\vec{x}} \cdot [A(\vec{x}, t) P(\vec{x}, t)] + \frac{1}{2} \partial_{\vec{x}} \partial_{\vec{x}} : [B(\vec{x}, t) P(\vec{x}, t)] \tag{40}$$

The physical meaning of this equation can be conceptually expressed by a continuity equation for the probability current $\vec{J}_P$: $\partial_t P = -\nabla \cdot \vec{J}_P$. The current $\vec{J}_P$ is composed of two parts:

- **Drift Current:** $\vec{J}_{\mathrm{Drift}} = AP$. This is the flow driven by the systematic forces.

- **Diffusion Current (Flow from High to Low Probability):** $\vec{J}_{\text{Diffusion}} = -\frac{1}{2}B\partial_{\vec{x}}P$. This is the flow caused by random fluctuations, always from regions of high probability density to low probability density.

The elegant derivation of the Fokker-Planck equation provides a deep representation of continuous stochastic dynamics. The equation itself is **deterministic** (in the evolution of the PDF $P(\vec{x}, t)$). The two terms, the Drift term (A term) and the Diffusion term (B term), are linked to the properties of the underlying microscopic random process (the mean and variance of the random steps). This demonstrates the core principle that even though the microscopic trajectory (the single sample path) is stochastic, the macroscopic behavior of the ensemble (the average and variance) follows a fixed deterministic evolution equation.

The Fokker-Planck equation is often referred to as the continuous version of the Kolmogorov forward equation. It maps the time evolution of the probability density from the random motion of individual particles (drift and random fluctuations). This elegant equation perfectly captures the competition and balance between the systematic forces (Drift Term) and random noise (Diffusion Term).

Its specific applications extend far beyond classical and modern physics, into quantitative finance, where it forms the basis for the Black-Scholes model and other complex pricing models; into computational biology, where it models the dynamics of complex molecules and gene regulatory networks; into neurophysics, where it describes the firing processes of model neurons; and into atmospheric science and fluid mechanics, where it is used to study the diffusion and transport of substances in flows. The Fokker-Planck equation provides us with a powerful mathematical tool to model and understand the "physics of collective randomness" in systems composed of numerous particles without memory.

11.5 Python simulation: Two-dimensional Brownian motion

To intuitively understand the diffusion process and the solution to the Fokker-Planck equation, we can simulate the simplest case: a particle undergoing free diffusion on a two-dimensional plane. In this case, there is no drift ($A = 0$), and the diffusion is isotropic ($B_{ij} = 2D\delta_{ij}$), where D is the diffusion coefficient. The corresponding Fokker-Planck equation is the famous **Diffusion Equation**:

$$\partial_t P = D\nabla^2 P \tag{41}$$

The solution to this equation is a Gaussian distribution (Normal distribution) with a mean of zero, and whose variance increases linearly with time.

We can approximate this process by simulating a **2D Random Walk**. In each discrete time step Δt, the particle's position undergoes a random displacement:

$$\vec{X}(t + \Delta t) = \vec{X}(t) + \sqrt{2D\Delta t} \cdot \vec{\eta} \tag{42}$$

where $\vec{\eta}$ is a two-dimensional standard normal random vector (i.e., its components are independent Gaussian random numbers with a mean of 0 and a variance of 1).

The Python code below simulates the trajectories of multiple 2D random walks starting from the origin.

Ten independent "random walkers" start from the same central point (red dot) and trace their paths after 1000 time steps. Each colored curve represents a unique historical path that a particle could potentially traverse. This diagram illustrates the microscopic foundation described by the Focke-Planck equations. The equations themselves are not concerned with a specific, tortuous path, but rather with describing how the "probability cloud," statistically formed by a large number of such paths, evolves over time.

To glean statistical patterns from the chaotic individual trajectories, we increased the number of simulated particles to 10,000 and focused solely on their final positions. This image, presented as a "heatmap," illustrates the distribution density of these endpoints in a two-dimensional plane. The overall distribution exhibits a circular pattern, dense at the center and gradually thinning outwards. This is precisely the solution to the Falk-Planck equation, without external forces (drift term is zero), evolving from a point source (initial condition of the delta function)—a two-dimensional Gaussian distribution. The superimposed blue concentric circles in the figure are the contour lines of the theoretically predicted two-dimensional Gaussian distribution. We can see that the simulated data (heatmap) matches the theoretical prediction (contour lines), and the Falk-Planck equation accurately captures the macroscopic statistical behavior of the diffusion process.

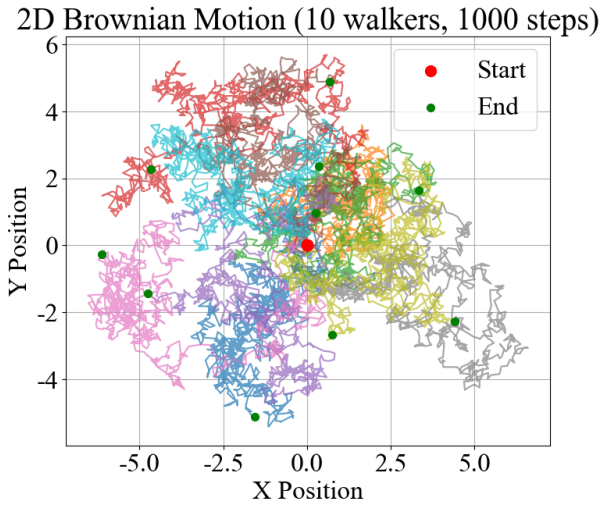

Figure 5:

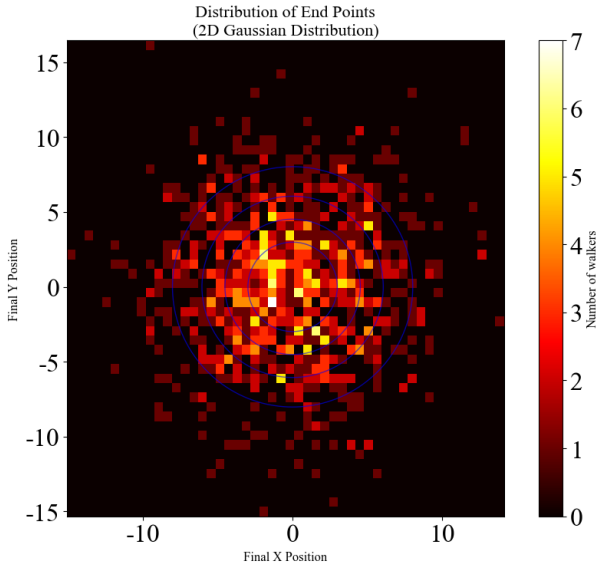

Figure 6:

This diagram serves as a bridge between microscopic randomness and macroscopic determinism. While the endpoint of a single particle is unpredictable, the endpoint distribution of a large number of particles follows a very deterministic mathematical law.

To analyze this two-dimensional Gaussian distribution more quantitatively, we can project it onto the X and Y axes respectively, obtaining two one-dimensional probability density distribution plots. The left plot (blue) shows the distribution of the final X-coordinates of all particles, and the right plot (pink) shows the distribution of the final Y-coordinates. Both projection plots clearly present the classic one-dimensional Gaussian distribution ("bell curve") shape. This indicates that although the particle moves in a two-dimensional plane, its displacement components in any orthogonal direction follow independent one-dimensional Brownian motion laws.

12 Mixing process and summary

We have discussed the pure jump process and the pure diffusion process separately. However, as demonstrated by the "Run & Tumble" motion of the E. coli bacterium, many real-world systems

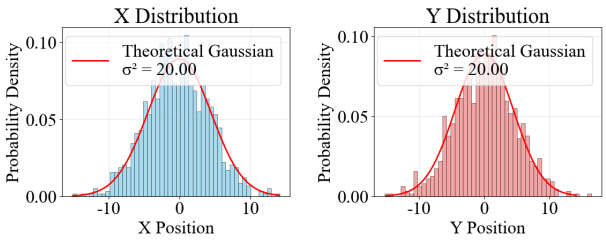

Figure 7:

simultaneously incorporate both dynamic features. To describe this **mixed process**, we can simply combine the two types of equations derived earlier.

The change in the probability density $P(\vec{x}, t)$ of a system in state $\vec{x}$ can arise from local drift and diffusion, and also from non-local jumps to and from other positions. Therefore, its evolution equation is the simple summation of the Fokker-Planck terms and the Master Equation terms:

$$\partial_t P(\vec{x}, t) = \underbrace{-\partial_{\vec{x}} \cdot [A(\vec{x}, t)P] + \frac{1}{2}\partial_{\vec{x}}\partial_{\vec{x}} : [B(\vec{x}, t)P]}_{\text{Drift and Diffusion (Fokker-Planck)}} + \underbrace{\int d\vec{x}' \left[w(\vec{x} \mid \vec{x}')P(\vec{x}', t) - w(\vec{x}' \mid \vec{x})P(\vec{x}, t)\right]}_{\text{Jump (Master Equation)}} \quad (43)$$

This generalized universal equation provides a powerful, unified framework for describing complex stochastic processes that have a continuous state space but whose sample paths contain both continuous and discontinuous parts.

In this lesson, we completed the generalization of the Markov process from discrete to continuous space, which is a crucial conceptual leap. We found that, based on the nature of the sample paths, continuous-space stochastic processes can be divided into two major categories:

Reference

- Nonequilibrium Field Theories and Stochastic Dynamics, Prof. Erwin Frey, LMU Munich, Summer Semester 2025
- Non-equilibrium States: Irreversibility and the Implications of Entropy Production — Non-equilibrium Field Theory and Stochastic Dynamics, Interesting Methods PhD
- https://mp.weixin.qq.com/s/lP1PQQTHFmK5SaMI3Rh3Lw

References