Nonequilibrium Field Theories and Stochastic Dynamics

Zehong Liao

October 9, 2025

In the previous lectures, we have already encountered several concrete examples of stochastic processes. From the "simple random walk" of a particle on a one-dimensional lattice in Lecture 1, to the "population dynamics" describing the birth and death of biological populations in Lecture 4, we have seen the central role of randomness in physical and biological systems. While these examples are specific, they lead us to a deeper question: Is there a universal mathematical framework, a common language, to describe all those stochastic evolutions that "do not remember the past, but only depend on the present"?

The goal of this lecture is precisely to establish and understand this universal language. Erwin Frey will guide us in deriving and explaining the two fundamental equations that govern all Markov processes. The first is the Chapman–Kolmogorov equation, which acts like a bridge in the macroscopic world, linking probability distributions at different points in time. Then, moving from the macroscopic to the microscopic level, we will derive the Master equation, which, in terms of instantaneous "rates," provides a powerful tool to describe how the probability of system states evolves continuously over time. Together, these two equations form the cornerstone of stochastic process theory.

1 Probabilistic description of stochastic processes

1.1 The Definition of stochastic processes

To establish a general theory, we first need an accurate definition of mathematics. A stochastic process can be regard as a collection of random variables indexed by time. In a most general case, it can be a multi-dimensional vector:

$$X(t) = (X_1(t), X_2(t), X_3(t) \cdots) \tag{1}$$

Each component $X_i(t)$ is itself a random variable, whose values can either be continuous real numbers (for example, the position of a particle in Brownian motion), or discrete integers (for example, the number of individuals in a population).

To simplify the discussion and focus on the core concepts, we follow the professor's approach in the lecture and consider, for the moment, only a single-component random variable.

1.2 Joint Probability Density: A Complete Description

How can a random process be completely described? The answer is that we need to know the **joint probability** that the random variables $X(t)$ take values $x_1, x_2, \ldots, x_n$ at any series of time points $t_1, t_2, \ldots, t_n$. This is characterized by the joint probability density function $P(x_n, t_n; x_{n-1}, t_{n-1}; \ldots; x_1, t_1)$. Its rigorous definition is as follows:

$$P(x_n, t_n; x_{n-1}, t_{n-1}; \ldots; x_1, t_1)dx_n \ldots dx_1 = \text{Prob}\{x_n \leq X(t_n) \leq x_n + dx_n; \ldots; x_1 \leq X(t_1) \leq x_1 + dx_1\} \tag{2}$$

The meaning of this expression is the joint probability of observing the system state near x_1 at time t_1, near x_2 at time t_2, ..., and finally near x_n at time t_n.

This joint probability density function is the most fundamental and complete way to describe a random process, as it encompasses all possible correlation information between the system at different times. However, this "completeness" also brings enormous complexity. Imagine, to determine this function for a real system, we would need to measure or compute the probability of all possible historical

paths. As the number of time points increases or the dimensionality of the state space expands, the complexity of this function grows explosively, making it virtually unusable in practice. This is a core challenge in theoretical modeling: we must introduce reasonable **physical simplifications** to extract useful, solvable mathematical models from this kind of overwhelming complexity. The Markov property is precisely the most important and powerful simplification we will introduce.

1.3 Stochastic Trajectories and Conditional Probability

To gain a more intuitive understanding of stochastic processes, we can introduce the concept of a **stochastic trajectory**. Each concrete realization of a stochastic process—that is, a specific path formed by its evolution over time—constitutes a stochastic trajectory. It depicts one particular history of how the system's state evolves with time.

To analyze the dependence relationship between different time points—that is, how the past influences the future—we need to use the mathematical tool of **Conditional Probability**. By definition, the probability of a future series of events (the state at times $t_{n+1}, \ldots, t_{n+m}$ is $x_{n+1}, \ldots, x_{n+m}$) occurring, given a past series of events (the state at times $t_1, \ldots, t_n$ is $x_1, \ldots, x_n$), is:

$$P(x_{n+m}, t_{n+m}; \ldots; x_{n+1}, t_{n+1} \mid x_n, t_n; \ldots; x_1, t_1) = \frac{P(x_{n+m}, t_{n+m}; \ldots; x_1, t_1)}{P(x_n, t_n; \ldots; x_1, t_1)} \tag{3}$$

This expression is our starting point for the subsequent discussion of the **Markov property**.

2 Core Assumption: The Markov Property

2.1 Definition of a "Memoryless" Process

The **Markov property** is a restriction on the "memory" of a stochastic process and forms the foundation for all the theory in the subsequent parts of this lecture. Its core idea is **memorylessness**. If a process possesses the Markov property, then its future state depends only on the current state and is independent of the entire history that led to the current state.

Expressed in the language of conditional probability, this property can be written as:

$$P(x_n, t_n \mid x_{n-1}, t_{n-1}; \ldots; x_1, t_1) = P(x_n, t_n \mid x_{n-1}, t_{n-1}) \tag{4}$$

This equality means that for predicting the probability of being in state x_n at time t_n, we only need to know the state x_{n-1} at the immediately preceding time t_{n-1}. All earlier historical information $(x_{n-2}, \ldots, x_1)$ is redundant for predicting the future. The current state x_{n-1} already contains all the necessary information for future prediction.

This seemingly simple mathematical assumption harbors profound physical insight. A physical system can often be approximated as a Markov process because there is a **separation of time scales**. For example, consider a macroscopic particle moving in a liquid (Brownian motion). Its position $X(t)$ is the random variable we are concerned with. Strictly speaking, its motion has memory because its velocity is maintained over short time periods. A complete description would need to consider both the position $X(t)$ and the velocity $V(t)$. However, the particle collides with a large number of tiny liquid molecules, causing the correlation (or "memory time") of its velocity to be extremely short. If our observation time scale is much larger than this velocity correlation time, then, at our scale of observation, the particle's next displacement is almost entirely dependent only on its current position, and independent of its past velocity history. Therefore, the **Markov property** is not an absolute physical law, but an extremely effective **approximation** applicable to systems where the "fast variables" that carry memory (which are not directly observed) can rapidly reach equilibrium on the time scale of interest, thereby not affecting the evolution of the "slow variables" we are concerned with.

2.2 Simplification of the Joint Probability

The power of the **Markov property** lies in its ability to greatly simplify the extremely complex joint probability density we mentioned earlier. By using the chain rule for conditional probability, we can decompose the joint probability:

$$P(x_n, t_n; \ldots; x_1, t_1) = P(x_n, t_n \mid x_{n-1}, t_{n-1}; \ldots; x_1, t_1) \cdot P(x_{n-1}, t_{n-1}; \ldots; x_1, t_1) \tag{5}$$

Applying the Markov property, the conditional probability term in the above equation is greatly simplified. By repeating this process, we can decompose the entire joint probability into a product of a series of **two-point transition probabilities**:

$$P(x_n, t_n; \ldots; x_1, t_1) = P(x_n, t_n \mid x_{n-1}, t_{n-1}) P(x_{n-1}, t_{n-1} \mid x_{n-2}, t_{n-2}) \ldots P(x_2, t_2 \mid x_1, t_1) P(x_1, t_1) \tag{6}$$

This result is of extraordinary significance. It tells us that the complete statistical characteristics of a complex Markov process are entirely determined by just two elements:

1. **Initial Conditions**: The probability distribution of the system at the starting time t_1, denoted $P(x_1, t_1)$.
2. **Transition Rule**: The conditional probability that describes the system transitioning from one state to the next state,
$$P(x_j, t_j \mid x_{j-1}, t_{j-1})$$

We no longer need to deal with the intractable joint probability function that includes all historical information. The evolution of the entire system can now be viewed as the **transition probabilities** propagating the initial probability distribution forward step-by-step.

3 The Chapman–Kolmogorov Equation: A Bridge Connecting Different Times

Once we have the advantage of the **Markov property**, now we can derive the first basical equation: The Chapman–Kolmogorov Equatio.

3.1 Derivation: Marginalization

The derivation of the **Chapman-Kolmogorov (CK) equation** is based on a fundamental rule in probability theory—the **Law of Total Probability**, also known as **Marginalization**. Consider three time points $t_1 < t_2 < t_3$. We want to know the probability $P(x_3, t_3 \mid x_1, t_1)$ that the system transitions from state x_1 at t_1 to state x_3 at t_3.

To go from t_1 to t_3, the system must pass through some intermediate state x_2 at the intermediate time t_2. Since x_2 can be any possible state, we need to sum over (or integrate over) all possible intermediate states. This is the idea of marginalization:

$$P(x_3, t_3 \mid x_1, t_1) = \int dx_2 P(x_3, t_3; x_2, t_2 \mid x_1, t_1) \tag{7}$$

Now, we can use the definition of conditional probability and the Markov property to simplify the integrand:

$$P(x_3, t_3; x_2, t_2 \mid x_1, t_1) = \frac{P(x_3, t_3; x_2, t_2; x_1, t_1)}{P(x_1, t_1)} \tag{8}$$

Applying the chain rule and the Markov property to the numerator:

$$P(x_3, t_3; x_2, t_2; x_1, t_1) = P(x_3, t_3 \mid x_2, t_2; x_1, t_1) P(x_2, t_2 \mid x_1, t_1) P(x_1, t_1) \tag{9}$$

Substituting this result back, we get:

$$P(x_3, t_3; x_2, t_2 \mid x_1, t_1) = P(x_3, t_3 \mid x_2, t_2) P(x_2, t_2 \mid x_1, t_1) \tag{10}$$

Finally, substituting this simplified expression into the marginalization integral, we obtain the general **Chapman-Kolmogorov equation**.

3.2 Form and Physical Interpretation of the Equation

The **Chapman-Kolmogorov equation** has two forms, depending on whether the state variables are continuous or discrete:

- **Continuous State Variables**:

$$P(x_3, t_3 \mid x_1, t_1) = \int dx_2 P(x_3, t_3 \mid x_2, t_2) P(x_2, t_2 \mid x_1, t_1)$$

- **Discrete State Variables** (using n to denote the state, and setting $t_0 = t_1$, $t' = t_2$, $t = t_3$):

$$P(n, t \mid n_0, t_0) = \sum_{n'} P(n, t \mid n', t') P(n', t' \mid n_0, t_0)$$

Physical Meaning of the CK Equation: The physical meaning of the CK equation is very intuitive. It shows that the entire evolution process from state n_0 to the final state n can be broken down into two independent steps: first, evolving from n_0 to some intermediate state n', and then evolving from n' to the final state n. The core idea of the equation is the **"sum over all possible paths."** It expresses the transition probability over a large time interval as the sum of the probabilities of all possible intermediate paths, embodying the composition rule of Markov processes.

For discrete state systems, the form of the CK equation reveals a deep connection with linear algebra. If we define a transition probability matrix $\mathbf{T}(t_a \to t_b)$, where its matrix element $(\mathbf{T})_{ij} = P(j, t_b \mid i, t_a)$, then the discrete form of the CK equation can be written concisely in the form of **matrix multiplication**:

$$\mathbf{T}(t_1 \to t_3) = \mathbf{T}(t_2 \to t_3)\mathbf{T}(t_1 \to t_2)$$

This perspective is extremely important, as it transforms the problem of stochastic process evolution into the familiar and powerful problem of **linear algebra**. For example, the long-term behavior of the system (the **steady-state distribution**) is closely related to the eigenvectors and eigenvalues of the transition matrix.

4 From Macroscopic to Microscopic: The Master Equation

The Chapman–Kolmogorov equation describes the evolution of probabilities over finite time intervals. However, in many physical and chemical problems, we are more interested in the instantaneous rate of change of probabilities. To derive a differential equation that captures this instantaneous change, we start from the CK equation and consider an infinitesimally small time step, which leads us to the Master Equation.

4.1 From Finite Steps to Instantaneous Rates

We start with the discrete form of the CK equation and set the time interval $t - t'$ to be an infinitesimal small value Δt, i.e., $t' = t$, $t = t + \Delta t$. The equation becomes:

$$P(n, t + \Delta t \mid n_0, t_0) = \sum_m P(n, t + \Delta t \mid m, t) P(m, t \mid n_0, t_0) \tag{11}$$

We are interested in the **rate of change** of the probability with respect to time, i.e., the time derivative:

$$\frac{d}{dt} P(n, t \mid n_0, t_0) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left[P(n, t + \Delta t \mid n_0, t_0) - P(n, t \mid n_0, t_0) \right] \tag{12}$$

Substituting the CK equation into the above expression, and using the identity $P(n, t \mid n_0, t_0) = \sum_m \delta_{n,m} P(m, t \mid n_0, t_0)$ (where $\delta_{n,m}$ is the **Kronecker delta**), we obtain:

$$\frac{d}{dt} P(n, t \mid n_0, t_0) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \sum_m \left[P(n, t + \Delta t \mid m, t) - \delta_{n,m} \right] P(m, t \mid n_0, t_0) \tag{13}$$

The core of this expression lies in the term inside the square brackets, which describes the probability change of transitioning from state m to state n over an extremely short time Δt.

4.2 The Transition Rate Matrix Q

To simplify the above expression, we define a key physical quantity—the **Transition Rate Matrix Q** (sometimes written as Q_{nm}). The matrix element $Q(n, m)$ is defined as:

$$Q(n, m) := \lim_{\Delta t \to 0} \frac{1}{\Delta t} [P(n, t + \Delta t \mid m, t) - \delta_{n,m}] \tag{14}$$

The elements of this matrix have a clear physical meaning:

- **Off-Diagonal Elements (Gain)**: When $n \neq m$, $\delta_{n,m} = 0$. The matrix element is usually denoted as $w_{m \to n}$:

  $$w_{m \to n} := Q(n, m) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(n, t + \Delta t \mid m, t) \tag{15}$$

  $w_{m \to n}$ represents the **instantaneous rate** at which the system transitions from state m to state n. Its unit is $[\text{time}]^{-1}$. The infinitesimal probability of transitioning from m to n within the infinitesimal time $d\tau$ is $w_{m \to n} d\tau$. This represents the probability **"Gain"** term flowing into state n.

- **Diagonal Elements (Loss)**: When $n = m$, $\delta_{n,m} = 1$. The matrix element is:

  $$Q(m, m) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} [P(m, t + \Delta t \mid m, t) - 1] \tag{16}$$

  Since the probability of no transition occurring in a short time, $P(m, t + \Delta t \mid m, t)$, approaches 1, $Q(m, m)$ is a negative number. We typically define a positive **exit rate** w_m:

  $$w_m := -Q(m, m) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} [1 - P(m, t + \Delta t \mid m, t)] \tag{17}$$

  w_m represents the **total rate** of leaving state m (transitioning to any other state). This represents the probability **"Loss"** term flowing out of state m.

4.3 Final Form of the Master Equation

Substituting the definition of the **Transition Rate Matrix Q** into the previous time derivative expression, we obtain the **Master Equation** in matrix form:

$$\frac{d}{dt} P(n, t \mid n_0, t_0) = \sum_m Q(n, m) P(m, t \mid n_0, t_0) \tag{18}$$

To understand its physical meaning more intuitively, we can rewrite it in the **"Gain-Loss"** form. To do this, we need to use an important property.

4.4 Manifestation of Probability Conservation

Since the system must be in some state (including itself) at the next moment when starting from any state, the total probability must be conserved: $\sum_n P(n, t + \Delta t \mid m, t) = 1$. Utilizing this property, we can prove that the sum of every column of the **Q** matrix is zero:

$$\sum_n Q(n, m) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left[\sum_n P(n, t + \Delta t \mid m, t) - \sum_n \delta_{n,m} \right] = \lim_{\Delta t \to 0} \frac{1}{\Delta t} [1 - 1] = 0 \tag{19}$$

This mathematical property is a direct expression of the physical law of **probability conservation**. It means that the loss rate from the diagonal element is exactly equal to the sum of the gain rates to all other states from that state:

$$-Q(m, m) = \sum_{n \neq m} Q(n, m) \implies w_m = \sum_{n \neq m} w_{m \to n}$$

This relationship is key to the **"Gain-Loss"** structure of the master equation. The master equation can be viewed as an equation of **continuity**. The total probability is a conserved quantity; it does

not spontaneously appear or disappear but is merely redistributed among different states. The term $w_{m\to n}P(m,t)$ plays the role of the **"probability current"** flowing from state m to state n. The constraint that $\mathbf{Q}$'s column sums are zero ensures that the total probability current is zero, meaning the total probability does not change with time.

—

Now, we can write the master equation in its final and most intuitive form. For notational simplicity, let $P_n(t)$ represent $P(n,t \mid n_0,t_0)$:

$$\frac{d}{dt}P_n(t) = \sum_m Q_{nm}P_m(t) = \sum_{m\neq n} Q_{nm}P_m(t) + Q_{nn}P_n(t) \tag{20}$$

Using $w_{m\to n} = Q_{nm}$ and $w_n = -Q_{nn} = \sum_{m\neq n} w_{n\to m}$, the above equation becomes:

$$\frac{d}{dt}P_n(t) = \underbrace{\sum_{m\neq n} w_{m\to n}P_m(t)}_{\text{Gain: from all } m \text{ into } n} - \underbrace{\sum_{m\neq n} w_{n\to m}P_n(t)}_{\text{Loss: from } n \text{ to all } m} \tag{21}$$

This form perfectly explains the physical meaning of the master equation: the rate of change of the probability of state n over time is equal to the sum of all probability currents flowing **into** state n (**Gain** term), minus the sum of all probability currents flowing **out** of state n (**Loss** term).

4.5 Forward vs. Backward Master Equations

At the end of the lecture, it was briefly mentioned that the Master Equation we derived has its time derivative taken with respect to the "final" (endpoint) time, and is called the Forward Master Equation. It answers the question: "Given an initial state, how does the probability distribution evolve in the future?"

Correspondingly, there also exists a Backward Master Equation, whose time derivative is taken with respect to the "initial" (starting) time. It addresses a different question: "In order to reach a given final state, how should the probability distribution over initial states be chosen?" The forward equation is more widely used in physics and chemistry, as it naturally describes the system's evolution forward in time.

5 Visualizing a Discrete Markov Process

In order to connect between the "rate" defined in the Master equation and the Actual Stochastic Trajectory, we can use numerical simulation method to do it. For a discrete-state, continuous-time Markov process described by the Master Equation, the most classical simulation algorithm is the Gillespie algorithm.

5.1 System Setup

We set up a simple three-state model. For example, a protein molecule may be in three conformations: **Open (O)**, **Closed (C)**, and **Inhibited (I)**. We use the numbers $0, 1, 2$ to represent these three states, respectively.

The transition rates between the states are as follows:

- $O \rightleftharpoons C$: The open and closed states are interconvertible, with rates $w_{0\to 1} = k_{OC}$ and $w_{1\to 0} = k_{CO}$.
- $C \to I$: The closed state can irreversibly transform into the inhibited state, with rate $w_{1\to 2} = k_{CI}$.

The corresponding **transition rate matrix Q** is:

$$\mathbf{Q} = \begin{pmatrix} -k_{OC} & k_{CO} & 0 \\ k_{OC} & -k_{CO} - k_{CI} & 0 \\ 0 & k_{CI} & 0 \end{pmatrix}$$

Note that the sum of each column is zero, which satisfies the conservation of probability.

5.2 Algorithm and Code Implementation

The core idea of the **Gillespie algorithm** is that at any given moment, both what event will happen next in the system and how long one must wait are **stochastic** (random). The algorithm determines these two aspects through a two-step random sampling process:

1. **Determine the waiting time:** The total escape rate W is the sum of the rates of all possible outgoing paths from the current state. The waiting time Δt follows an **exponential distribution** and can be generated using

$$\Delta t = \frac{1}{W} \ln\left(\frac{1}{r_1}\right) = -\frac{1}{W} \ln(r_1),$$

   where r_1 is a random number on the interval $(0, 1]$.

2. **Determine the next event:** Use a second random number r_2 to determine which specific transition occurs. The interval $(0, W]$ is partitioned according to the proportion of each transition rate. The transition corresponding to the subinterval where $r_2 \cdot W$ falls is the one that occurs.

   (Alternatively, r_2 can be compared directly to the normalized cumulative rates, checking where r_2 falls in the cumulative sum of a_i/W.)

In the dynamical structure of a protein molecular system, the transition from the closed state to the inactivated state is irreversible, whereas the open and closed states can interconvert with each other.

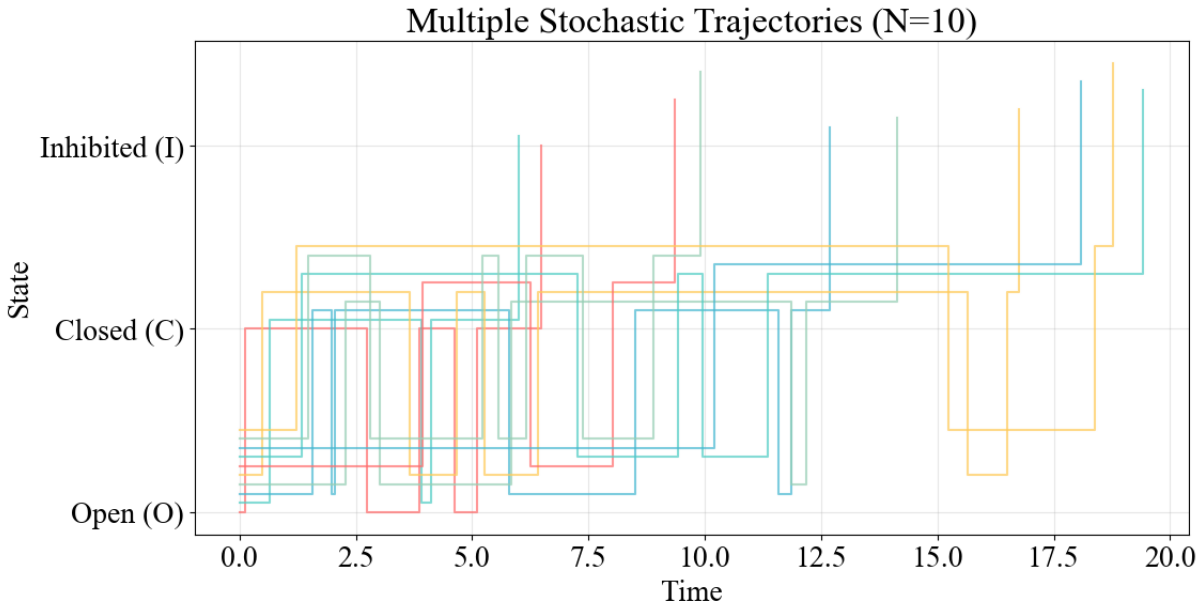

Figure 1:

The simulation of the system exhibits the following key characteristics:

- **Stochasticity (Randomness):** The states of different trajectories are not identical at the same time point, reflecting the random nature of the system.
- **State Jumps:** The system instantly jumps between different states, which conforms to the characteristics of a **Markov process**.
- **Absorbing State:** All trajectories eventually enter the **Inhibited (I)** state, because this is an absorbing state.
- **Dwell Time:** The time the system spends in each state is random and follows an **exponential distribution**.

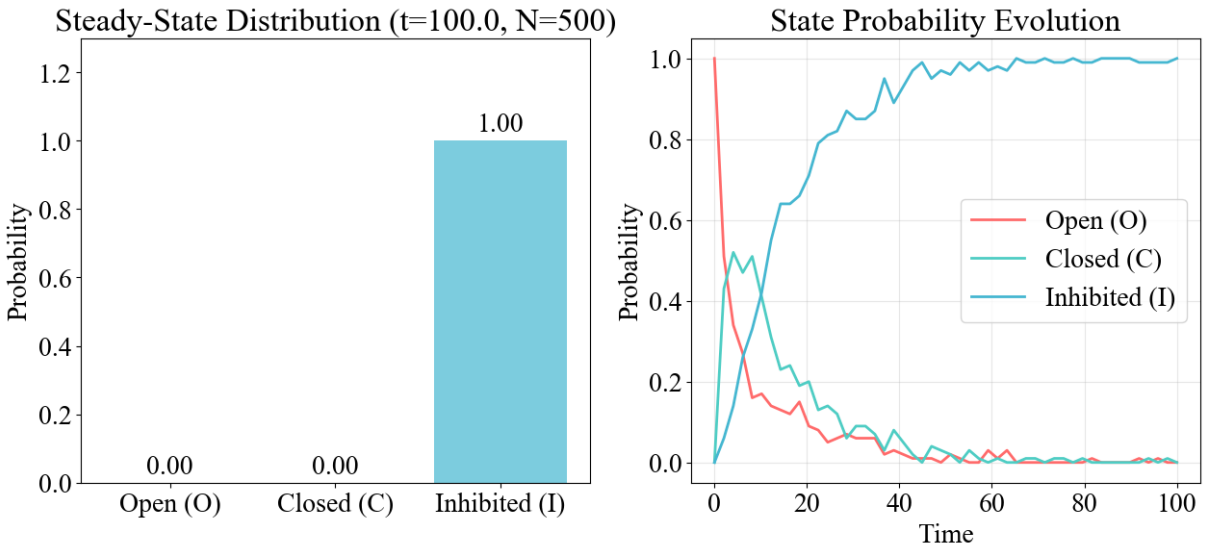

Figure 2:

It shows the final state distribution of the system after a long-time evolution ((t = 100)): all samples have entered the **inactivated state** (Inhibited), with probability approaching 1.0. The probabilities of the open and closed states are close to zero, indicating that the system will eventually be "trapped" in the inactivated state.

6 Case Study: The Linear Birth–Death Process

Now, we will apply the Master Equation and the Transition Rate Matrix Q to analysis a highly important in population dynamics, chemical reactions, and queueing theory: the linear birth–death process.

6.1 The definition of the system

We consider a population composed of many identical individuals (e.g., molecules, cells, or organisms), with a population size of n. This system allows for only two types of stochastic events:

- **Birth:** An individual replicates itself at a rate μ, leading to an increase of the population size by one. This process can be represented as $A \xrightarrow{\mu} 2A$. If there are n individuals in the population, the total birth rate is proportional to n. Therefore, the rate for the transition from state n to $n+1$ is
$$w(n+1, n) = \mu n.$$
- **Death:** An individual disappears from the system at a rate λ. This process can be represented as $A \xrightarrow{\lambda} \emptyset$. Similarly, the total death rate is proportional to the population size n. Therefore, the rate for the transition from state n to $n-1$ is
$$w(n-1, n) = \lambda n.$$

In this model, the state $n = 0$ is a very special state. When the population size is zero, there are no individuals to undergo birth or death, so the total birth and death rates of the system are zero. This means that once the system enters the state $n = 0$, it will remain there permanently. This state is called an **absorbing state**.

6.2 Constructing the Transition Rate Matrix Q for the Birth–Death Process

Now, we construct the **Q** matrix for this process step-by-step according to the rules described above. The rows and columns of the matrix are indexed by the population size $n = 0, 1, 2, \ldots$.

Off-Diagonal Elements

- **Transition from n to $n+1$ (Birth):** We have $Q_{n+1,n} = w(n+1, n) = \mu n$. These elements are located on the line **immediately above the main diagonal** (upper off-diagonal).
- **Transition from n to $n-1$ (Death):** We have $Q_{n-1,n} = w(n-1, n) = \lambda n$. These elements are located on the line **immediately below the main diagonal** (lower off-diagonal).
- Since only one individual can be born or die at a time, there are no transitions that skip a single unit, such as from n to $n+2$. Therefore, **all other off-diagonal elements are zero**.

Diagonal Elements

- The diagonal element $Q_{n,n}$ is the negative of the **total escape rate**. Starting from state n, the system can only transition to $n+1$ or $n-1$. Therefore, the total escape rate $W(n)$ is the sum of the rates of these two processes:

$$W(n) = w(n+1, n) + w(n-1, n) = \mu n + \lambda n = (\mu + \lambda)n.$$

- Thus, the diagonal element is $Q_{n,n} = -W(n) = -(\mu + \lambda)n$.
- For the **absorbing state** $n = 0$, the escape rate is zero, so $Q_{0,0} = 0$.

The matrix **Q** (showing indices for clarity) looks like this for the first few states:

$$\mathbf{Q} = \begin{pmatrix} Q_{0,0} & Q_{0,1} & Q_{0,2} & \cdots \\ Q_{1,0} & Q_{1,1} & Q_{1,2} & \cdots \\ Q_{2,0} & Q_{2,1} & Q_{2,2} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} 0 & \lambda & 0 & 0 & \cdots \\ 0 & -(\mu+\lambda) & 2\lambda & 0 & \cdots \\ 0 & \mu & -2(\mu+\lambda) & 3\lambda & \cdots \\ 0 & 0 & 2\mu & -3(\mu+\lambda) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

This example illustrates a profound principle: the structure of the Q matrix is a direct "fingerprint" of the system's microscopic dynamical rules. In the birth–death process, the population can change by only ± 1, which corresponds to "nearest-neighbor" jumps in the state space. When we translate these rules into a Q matrix, we naturally obtain a sparse matrix with nonzero elements only on the main diagonal and the two immediately adjacent diagonals.This is a general principle: the sparsity pattern of the Q matrix directly reflects the connectivity of the state space graph. A tridiagonal matrix indicates that the states form a one-dimensional chain, whereas a dense Q matrix implies that any state can directly transition to any other state.

6.3 Python Simulation and Visualization

To gain a more intuitive understanding of the structure of the Q matrix, we can use Python code to generate and visualize it

- The diagonal element $Q_{n,n}$ (Blue): They are the negative leaving rate $Q_{n,n} = -(\mu + \lambda)n$. The absolute value is linearly increasing with the n.
- The upper diagonal element $Q_{n,n+1}$ (red): These elements are born rate $Q_{n+1,n} = \mu n$.
- The lower diagonal element $Q_{n+1,n}$ (red): These elements are born rate $Q_{n+1,n} = \lambda n$.

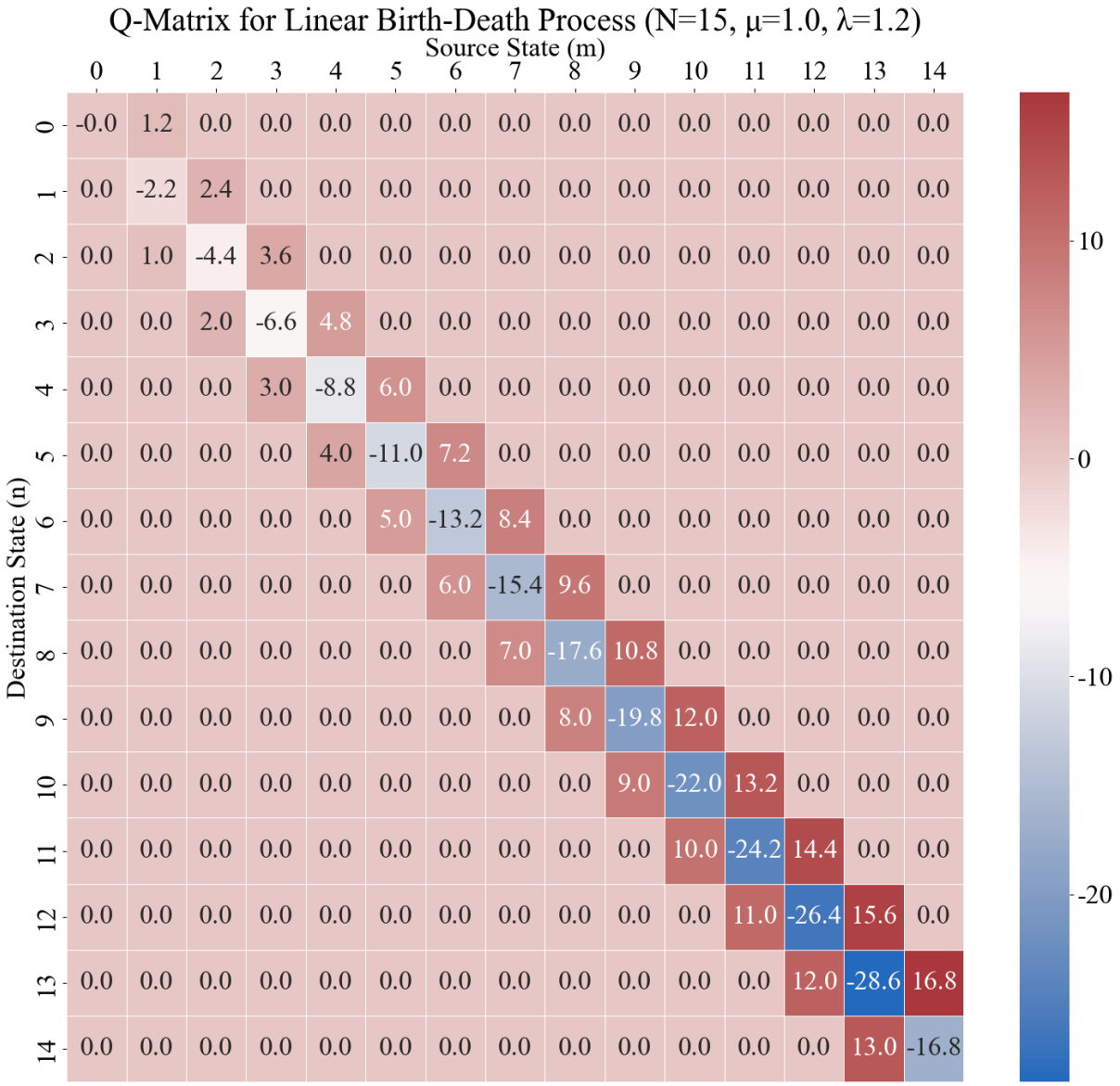

Figure 3:

6.4 The evolution of the probability and Phase space

With the **Q** matrix as our "engine," we can now numerically solve the Master Equation

$$\partial_t \vec{P} = \mathbf{Q}\vec{P}$$

to observe how the probability distribution evolves from a determined initial state toward a steady-state distribution.

Theory tells us what the steady-state is, but it does not describe how the system reaches that state. The transient behavior during the evolution process is equally important, as it reflects the **relaxation dynamics** of the system. An animation can visually demonstrate the concrete manifestation of the abstract concept of "probability current."

We now solve the Master Equation to observe how the probability distribution dynamically evolves over time.

The top figure shows the probabilities of all states at each time step, with the color intensity dynamically reflecting the magnitude of the probability (darker = higher probability). It illustrates how the probability distribution gradually "spreads out" from the initial sharp peak.

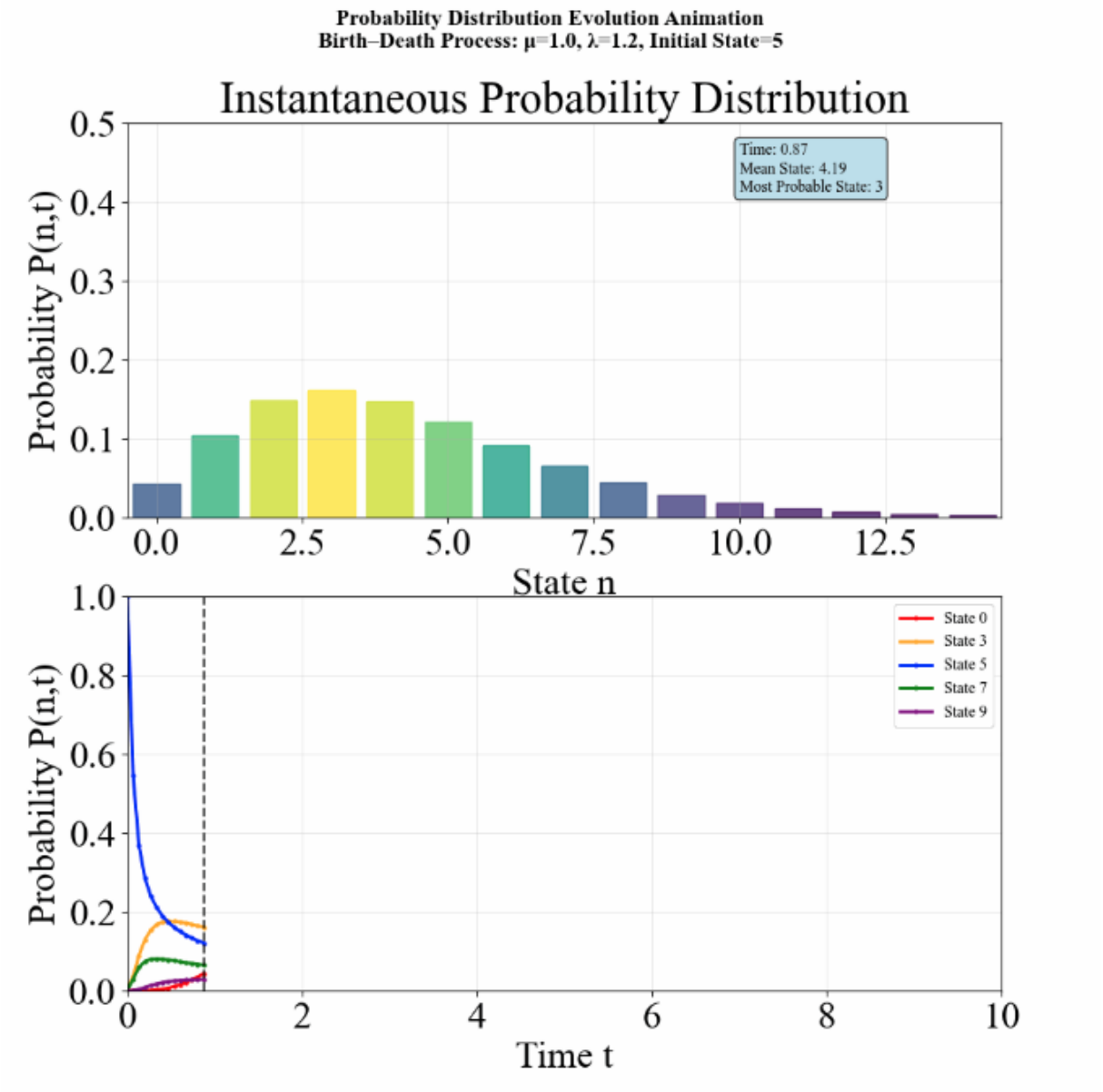

Figure 4:

The bottom figure tracks the probability evolution of five representative states over time.

We can clearly observe the net probability flow from higher states toward lower states. In the space of mean, variance, and skewness, the system's evolutionary trajectory eventually converges to the fixed point corresponding to the steady state, reflecting the system's relaxation dynamics.

7 Long-term Dynamics and General Properties

After building the Master equation and Q matrix, a natural question is: As time progresses, what state will the system evolve into? Is there a stable equilibrium state?

7.1 Reaching Equilibrium: Steady State

When the system reaches a **steady state** or a **stationary distribution**, the probability distribution no longer changes with time. We use the vector $\vec{\pi}$ to denote this steady-state distribution. Mathemat-

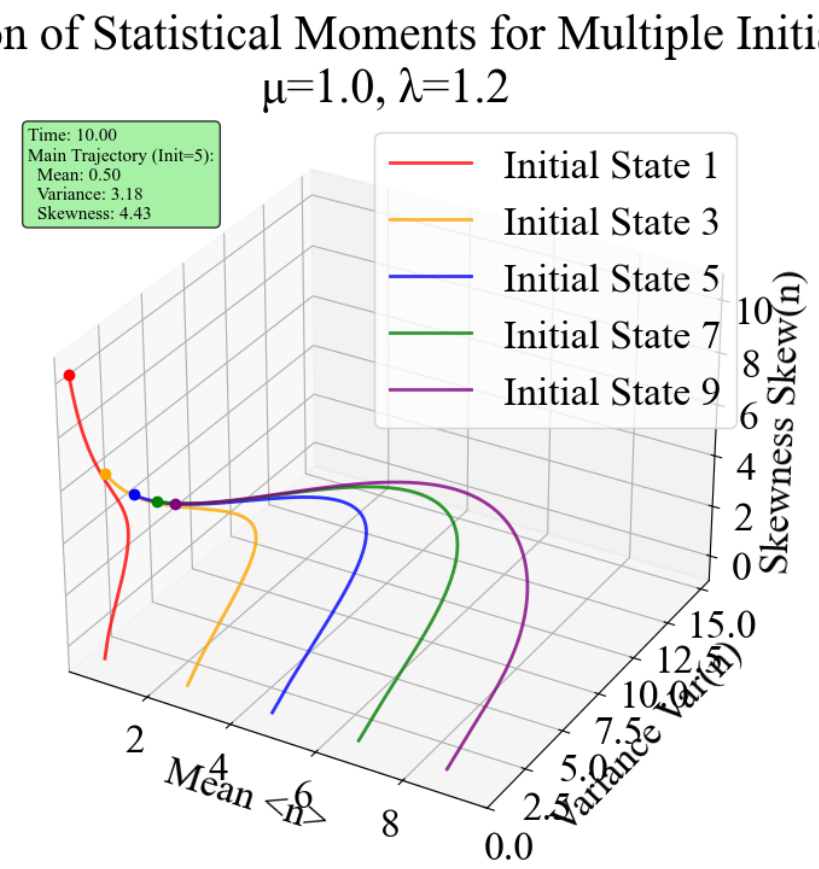

Figure 5:

ically, this means the time derivative of the probability vector is zero:

$$\partial_t \vec{P}(\tau) = 0 \tag{22}$$

Substituting this condition into the Master Equation $\partial_t \vec{P} = \mathbf{Q}\vec{P}$, we immediately obtain the core equation for finding the steady-state distribution:

$$\mathbf{Q}\vec{\pi} = \vec{0} \tag{23}$$

The meaning of this equation is profound: the steady-state distribution $\vec{\pi}$ is the **right eigenvector** of the $\mathbf{Q}$ matrix corresponding to the **eigenvalue** $\lambda = 0$. This is a very deep conclusion, as it transforms a dynamics problem (finding the distribution that does not change with time) into a problem in linear algebra (finding the **null space** of the $\mathbf{Q}$ matrix).

Furthermore, we already know that due to the conservation of probability, the sum of each column of the $\mathbf{Q}$ matrix is zero. This can be written in vector form as

$$(1, 1, \ldots, 1)\mathbf{Q} = \vec{0},$$

where $(1, 1, \ldots, 1)$ is a row vector of all ones, denoted as $\vec{1}^T$. This means that the **all-ones vector** $\vec{1}^T$ is the **left eigenvector** of the $\mathbf{Q}$ matrix corresponding to the **eigenvalue** $\lambda = 0$.

7.2 Structure of the State Space

The long-term behavior of a system not only depends on the transfer rate, but also depends on the **topological structure** of its state space, that is, how are the states interconnected?

- **Absorbing States:** As stated before, it is called as the Absorbing State if it enters a states then can not leave. In the Q matrix, That means the leaving rate is 0 from the states i, $Q_{ii} = 0$.

Because the row sum is zero, this also implies that all transition rates Q_{ji} from state i to other states j must also vanish.

- **Transient States:** A state is a **transient state** if the system, once having left it, has a non-zero probability of never returning to it. In systems with absorbing states, all non-absorbing states are transient because the system will eventually be "captured" by the absorbing state.
- **Irreducibility:** A system (or a subset of its states) is **irreducible** if any state can be reached from any other state through a finite number of transitions. This means the entire state space is "**fully connected**" and does not break down into several mutually non-communicating parts. The block-diagonal form of the $\mathbf{Q}$ matrix (as shown below), which the professor drew on the board, represents a **irreducible** system because it splits into two independent subsystems, A and B, which cannot communicate with each other.

$$\mathbf{Q} = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

7.3 Perron–Frobenius Theorem: Guarantee of a Unique Physical Reality

For an irreducible system, we care about whether a physically meaningful steady state (where all probability components are non-negative) exists and whether it is unique. The powerful **Perron-Frobenius Theorem** provides a definitive answer.

This theorem is a profound result in linear algebra with many conclusions for matrices with non-negative off-diagonal elements (such as the transpose of the $\mathbf{Q}$ matrix). For the irreducible $\mathbf{Q}$ matrices we are concerned with, the two most important consequences can be stated generally as follows:

1. **Uniqueness Property:** The eigenvalue $\lambda = 0$ is **simple** (non-degenerate), meaning its algebraic and geometric multiplicity is one. This implies that the eigenspace corresponding to the eigenvalue $\lambda = 0$ is one-dimensional. Therefore, the solution $\vec{\pi}$ to $\mathbf{Q}\vec{\pi} = \vec{0}$ is unique after normalization ($\sum_n \pi_n = 1$).
2. **Physical Probability Property:** All components of this unique steady-state vector $\vec{\pi}$ can be chosen to be **strictly positive** ($\pi_n > 0$ for all n). This guarantees a physically meaningful probability distribution where every state has a non-zero probability of being occupied.

The significance of this theorem is enormous. If we were to write down an arbitrary $\mathbf{Q}$ matrix and attempt to solve $\mathbf{Q}\vec{\pi} = \vec{0}$, we might obtain multiple solutions or solutions containing negative numbers, leading to a physically ambiguous situation.

The Perron-Frobenius Theorem tells us that as long as the system satisfies an intuitive physical condition—**irreducibility** (i.e., the system is "well-mixed" internally)—then mathematics guarantees the system will eventually evolve to a single, physically sensible equilibrium state. The final distribution of a reducible system (e.g., one with absorbing states) may depend on its initial starting point. However, if the system is irreducible, it can reach any state from any other state, avoiding ambiguity.

Mathematically, the Perron-Frobenius Theorem proves that this "well-mixed" property (**irreducibility**) is what ensures the system will ultimately settle into a definite equilibrium state. In this equilibrium state, every state can be reached ($\pi_n > 0$). This connects a topological property (the **connectedness of the state space**) with a spectral property (the **simplicity of the $\lambda = 0$ eigenvalue**), with the latter guaranteeing the crucial physical property of **unique existence**.

8 Summary

The derivation proceeds through the following four steps:

1. We start from the most complete and most complex description of a stochastic process: the **Joint Probability Density Function**.
2. To make the problem tractable, we introduce a core physical simplification: the **Markov property**, also known as the "memoryless" property.

3. Based on the Markov property, we derive the first fundamental equation: the **Chapman-Kolmogorov equation** (CK equation). This is an integro-difference equation that describes the probability of the system transitioning from one state to another within a finite time step. Its core idea is "summing over all intermediate paths."

4. Finally, by taking the **infinitesimal time limit** ($\Delta t \to 0$) of the CK equation, we obtain the more powerful **Master Equation**. This is a differential equation that precisely describes the instantaneous change in the system's state probability by defining the **transition rates**. Its core idea is the "gain-loss balance" of probability between different states.

References