UCSB, Physics 129L, Computational Physics
Lecture notes, Week 6

Zihang Wang (UCSB), zihangwang@ucsb.edu

February 16, 2025

Contents

1 Markov process to Stochastic Calculus

1.1 Chapman–Kolmogorov equation

The probability (or density) of transitioning from an initial state to a final state over a specified time interval can be decomposed into contributions from all possible intermediate states,

$$p(x,t \mid x_0,t_0) = \int_{-\infty}^{\infty} p(x,t \mid y,s)\, p(y,s \mid x_0,t_0)\, dy, \quad t_0 < s < t. \tag{1}$$

namely the **Chapman–Kolmogorov equation**.

1.2 Discrete-time Chapman-Kolmogorov Equation

For a continuous-time Markov process with discrete states, the Chapman-Kolmogorov equation states that the probability of being in state n at time $t + \Delta t$, given that we started in state m at time 0, is given by:

$$P_{nm}(t + \Delta t) = \sum_k P_{nk}(t) P_{km}(\Delta t). \tag{2}$$

The Δt means that This expresses how the probability of transition from m to n in time $t + \Delta t$ depends on an intermediate state k at time t.

1.3 Birth-Death Process

A **birth-death process** is a special type of continuous-time Markov process that models systems where entities (e.g., populations, molecules, or queues) undergo births (increases) and deaths (decreases) over time. It is characterized by transition rates λ_n (birth rate) and μ_n (death rate), which determine the probability of moving from state n to $n + 1$ or $n - 1$, respectively.

For small time interval Δt, the transition probabilities satisfy the following,

$$P_{kn}(\Delta t) = \begin{cases} \lambda_k \Delta t + o(\Delta t), & \text{if } k \to n \text{ (birth)} \\ \mu_k \Delta t + o(\Delta t), & \text{if } k \to n \text{ (death)} \\ 1 - (\lambda_k + \mu_k)\Delta t + o(\Delta t), & \text{if } k = n \text{ (no transition)} \end{cases} \tag{3}$$

We write the marginal probability at a $t + \Delta t$ by summing over all possible transitions from time t to time $t + \Delta t$ via the transition probabilities,

$$\pi_n(t + \Delta t) = \sum_k \pi_k(t) P_{kn}(\Delta t), \tag{4}$$

1.4 Forward Equation

Substituting the above marginal probability into the Chapman-Kolmogorov equation and taking the limit $\Delta t \to 0$, we obtain,

$$\frac{d}{dt} P_{nm}(t) = \lambda_{n+1} P_{n+1,m}(t) + \mu_{n-1} P_{n-1,m}(t) - (\lambda_n + \mu_n) P_{nm}(t). \tag{5}$$

On the other hand, if we sum over all index m, we get the evolution equation for marginal probability,

$$\frac{d\pi_n(t)}{dt} = \lambda_{n-1}\pi_{n-1}(t) + \mu_{n+1}\pi_{n+1}(t) - (\lambda_n + \mu_n)\pi_n(t). \tag{6}$$

This is the **Kolmogorov Forward Equation**, describing probability flow into and out of state n:

- $\lambda_{n-1}P_{n-1}(t)$: The probability of transitioning **into state** n from $n-1$.
- $\mu_{n+1}\pi_{n+1}(t)$: The probability of transitioning **into state** n from $n+1$.
- $-(\lambda_n+\mu_n)\pi_n(t)$: Probability leaving state n (either increasing or decreasing).

In other words, we are standing at time t and look at how a current state at time t makes transition to $t+\Delta t$, where states is at n. Therefore, the rates are given at current time t, associated with $n-1, n+1$ states.

1.5 Backward Equation

Rewriting the Chapman-Kolmogorov equation:

$$P_{nm}(t+\Delta t) = \sum_k P_{nk}(\Delta t)P_{km}(t). \tag{7}$$

For small time interval Δt, the transition probabilities (backward) satisfy the following,

$$P_{nk}(\Delta t) = \begin{cases} \lambda_n \Delta t + o(\Delta t), & \text{if } n \to k \text{ (birth)} \\ \mu_n \Delta t + o(\Delta t), & \text{if } n \to k \text{ (death)} \\ 1-(\lambda_n+\mu_n)\Delta t + o(\Delta t), & \text{if } n = k \text{ (no transition)} \end{cases} \tag{8}$$

You can see the rates are flipped.

For small Δt, expanding $P_{nk}(\Delta t)$ and taking the limit $\Delta t \to 0$, we obtain,

$$\frac{d}{dt}P_{nm}(t) = \lambda_n P_{n+1,m}(t) + \mu_n P_{n-1,m}(t) - (\lambda_n+\mu_n)P_{nm}(t). \tag{9}$$

Again, if we sum over all index m, we get the evolution equation for marginal probability,

$$\frac{d\pi_n(t)}{dt} = \lambda_n \pi_{n-1}(t) + \mu_n \pi_{n+1}(t) - (\lambda_n+\mu_n)\pi_n(t). \tag{10}$$

This is the **Kolmogorov Backward Equation**, which describes how probabilities evolve based on the initial state.

- $\lambda_n\pi_{n+1}(t)$: If we start at state n, we might jump to $n+1$ (birth), in which case the marginal probability is π_{n+1}.
- $\mu_n\pi_{n-1}(t)$: If we start at state n, we might jump to $n-1$ (death), in which case the expectation is now $\pi_{n-1}(t)$.
- $-(\lambda_n+\mu_n)\pi_n(t)$: Accounts for the fact that, over an infinitesimal time dt, we might remain in n with probability $-(\lambda_n+\mu_n)\pi_n(t)$.

In other wards, it means that the backward equation has a different starting point: it starts at $t+\Delta t$. This gives the name of backward equation.

1.6 Poisson Process

A pure birth process describes a system where transitions occur sequentially from state j to $j+1$ at a constant rate λ. Given an initial state $j=0$, we look at the probability $P_{0j}(t)$ of being in state j at time t.

For $j=0$, since there is no state -1 ($P_{0,-1}(t)=0$), the governing equation simplifies to:

$$\frac{dP_{00}(t)}{dt} = -\lambda P_{00}(t). \tag{11}$$

With the initial condition $P_{00}(0)=1$, the solution is:

$$P_{00}(t) = e^{-\lambda t}. \tag{12}$$

For $j=1$, the probability satisfies:

$$\frac{dP_{01}(t)}{dt} = \lambda P_{00}(t) - \lambda P_{01}(t) = \lambda e^{-\lambda t} - \lambda P_{01}(t). \tag{13}$$

Using the integrating factor $\mu(t) = e^{\lambda t}$, we have,

$$\frac{d}{dt}(e^{\lambda t} P_{01}(t)) = \lambda. \tag{14}$$

Integrating and applying $P_{01}(0)=0$, we obtain,

$$P_{01}(t) = \lambda t e^{-\lambda t}. \tag{15}$$

For general j, assume the solution follows:

$$P_{0j}(t) = e^{-\lambda t}\frac{(\lambda t)^j}{j!}. \tag{16}$$

For $j+1$, the transition equation is:

$$\frac{dP_{0,j+1}(t)}{dt} = \lambda P_{0j}(t) - \lambda P_{0,j+1}(t). \tag{17}$$

Substituting the assumed form and differentiating:

$$\frac{dP_{0,j+1}(t)}{dt} = \lambda e^{-\lambda t}\frac{(\lambda t)^j}{j!} - \lambda e^{-\lambda t}\frac{(\lambda t)^{j+1}}{(j+1)!}, \tag{18}$$

If taking derivative on the left side, the above expression holds true. Therefore, by induction, we have,

$$P_{0j}(t) = e^{-\lambda t}\frac{(\lambda t)^j}{j!}, \quad j \geq 0. \tag{19}$$

This solution corresponds to the probability distribution of a **Poisson process** with rate λ, where $P_{0j}(t)$ gives the probability of exactly j transitions occurring by time t. This reflects an alternative interpretation from the limit of binomial distribution we discussed previously.

2 Martingales

A process is a martingale if the expected value of future outcomes is equal to the current value given all past information: when considering all information available up to time t, the expected future increment is zero—that is,

$$E[X_{t+1} - X_t \mid \mathcal{F}_t] = 0. \tag{20}$$

$\mathcal{F}_t$ is called the **filtration**, a collection of "information" prior to time t.

3 Wiener process and stochastic differential equation

Let's consider a sequence of N independent, identically distributed random variables $\zeta_1, \zeta_2, \ldots$, with mean 0 and variance σ^2 over a time duration t. The sum of random variables ζ_i, define a stochastic process $\{W_N(t)\}_t$,

$$W_N(t) = \sum_{m=0}^{t/\delta t} \frac{1}{N} \sum_{i=1}^{N} \zeta_i(m\delta t) = \sum_{m=0}^{t/\delta t} \bar{\zeta}(m\delta t), \tag{21}$$

which can be understood as the following: with each discrete time interval $\delta t = t/M$, the normalized sum of such N random variables $\sum_{i=1}^{N} \zeta_i(m\delta t)/N$ at current time $m\delta t$ has variance σ^2. Those can be various noises contributed from unknown sources contributed over a time interval t. The above expression is in discrete-time, $m, i \in \mathbb{Z}_+$, and it can be extended into the continuous–time expression $t \in \mathbb{R}_+$ when taking the continuous-time limit $N, M \to \infty$.

The first limit, $N \to \infty$, follows the central limit theorem, at each time $m\delta t$, we have the sum,

$$\lim_{N\to\infty} \bar{\zeta}(m\delta t) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \zeta_i(m\delta t) \sim \mathcal{N}\left(0, \frac{\sigma^2}{N}\right). \tag{22}$$

The second limit, $M \to \infty$, the sum becomes,

$$\sum_{m=0}^{t/\delta t} \bar{\zeta}(m\delta t) \sim \mathcal{N}\left(0, \frac{\sigma^2}{N}\right) = \mathcal{N}\left(0, \frac{t}{\delta t}\frac{\sigma^2}{N}\right) = \mathcal{N}\left(0, t\sigma^2\right), \tag{23}$$

and in the last equality, we use the fact that $N \sim 1/\delta t$. The **Wiener process** finally has the form,

$$W_N(t) \sim \mathcal{N}\left(0, t\sigma^2\right), \tag{24}$$

and under scaling limit (continuous random variable), this is one of the most fundamental stochastic processes in physics: **Brownian motion**. While the expectation value of the Brownian motion remains zero, the variance of the Brownian motion grows linearly in time. The conditional expectation of the next value in a stochastic sequence is equal to the present value, regardless of all prior values, i.e.

$$W(t|t-\delta t, t-2\delta t, \dots) = W(t-\delta t). \tag{25}$$

Following the above Wiener process, we can define the following **stochastic differential** at a given time t,

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t, \tag{26}$$

where μ, σ are drift and diffusion coefficients, and the X_t is the state at time t. The above differential form can be integrated via stochastic integral,

$$\begin{aligned} X_t &= X_0 + \int_0^T \mu(X_t, t)dt + \int_0^T \sigma(X_t, t)dW_t \\ &\approx X_0 + \int_0^T \mu(X_t, t)dt + \sum_i^N \sigma(X_{t_i}, t_i)(W_{t_i+dt} - W_{t_i}), \end{aligned} \tag{27}$$

where $dt = T/N$, and we evaluate at the left point for $\sigma(X_{t_i}, t_i)$. Let's consider a test function, $f(X_t, t)$, with integrator dt and dX_t,

$$dX_t = \mu(X_t)dt + \sigma(X_t)(W_{t+dt} - W_t) = \mu(X_t)dt + \sigma(X_t)dW_t, \tag{28}$$

where,

$$W_{t+dt} - W_t \approx dW_t \sim \mathcal{N}(0, dt), \quad dW_t^2 = dt, \tag{29}$$

so that the expected order of dW_t is $\sqrt{dt}$ with $dW_t^2 = dt$.

Suppose that X_t satisfies an SDE and that we want to compute the increment of a smooth function $f(X_t)$. Over a small time interval we have the **left point approximation** (it is left because we evaluate the function at the time prior to $t + dt$),

$$X_{t+dt} = X_t + dX_t, \tag{30}$$

and a Taylor expansion of $f(X_{t+dt})$ about X_t gives:

$$f(X_{t+dt}) = f(X_t) + \partial_t f(X_t)dt + \partial_X f(X_t)dX_t + \frac{1}{2}[\partial_X^2 f(X_t)](dX_t)^2 + \cdots. \tag{31}$$

When taking $dt \to 0$, to the first order dt, the resulting differential is given by,

$$\begin{aligned} df(X_t) = f(X_{t+dt}) - f(X_t) &= \left(\partial_t + \mu\partial_X + \frac{\sigma^2}{2}\partial_X^2\right) f dt + \sigma(\partial_X f)dW_t \\ &= \left(\partial_t + \frac{\sigma^2}{2}\partial_X^2\right) f dt + (\partial_X f)dX_t, \end{aligned} \tag{32}$$

i.e. the **Ito's lemma**.

You can see that it does not obey the traditional chain rule due to an extra second derivative term. Can we remove the term so that the stochastic calculus follows the same chain rule as the normal calculus? In **Stratonovich calculus**, the idea is to define the increment using the **midpoint approximation**.

Let's write a Taylor expansion about X_t for both $f(X_{t+dt/2})$ and $f(X_t)$,

For the forward increment, we have,

$$f\left(X_{t+dt}\right) = f(X_{t+dt/2}) + \frac{1}{2}\partial_t f dt + [\partial_X f] dX_t^+ + \frac{1}{2}[\partial_X^2 f](dX_t^+)^2 + \cdots , \quad (33)$$

and, for the backward increment, we have

$$f(X_t) = f\left(X_{t+dt/2}\right) - \frac{1}{2}\partial_t f dt - [\partial_X f] dX_t^- + \frac{1}{2}[\partial_X^2 f](dX_t^-)^2 + \cdots , \quad (34)$$

where $dX_t^\pm$ is the change from $t \pm dt$ to t. Taking the symmetric difference gives,

$$df(X_t) = f(X_{t+dt}) - f(X_t) = \partial_t f dt + [\partial_X f]\left(dX_t^+ + dX_t^-\right) + \frac{1}{2}[\partial_X^2 f]\left[(dX_t^+)^2 - (dX_t^-)^2\right] + \cdots , \quad (35)$$

where we define,

$$\begin{aligned} dX_t^+ &= \frac{1}{2}\mu\left(X_{t+dt/2}\right) dt + \sigma\left(X_{t+dt/2}\right)(W_{t+dt} - W_{t+dt/2}), \\ dX_t^- &= \frac{1}{2}\mu\left(X_{t+dt/2}\right) dt + \sigma\left(X_{t+dt/2}\right)(W_{t+dt/2} - W_t), \\ dX_t^M = dX_t^+ + dX_t^- &= \mu\left(X_{t+dt/2}\right) dt + \sigma\left(X_{t+dt/2}\right)(W_{t+dt} - W_t) \\ &= \mu\left(X_{t+dt/2}\right) dt + \sigma\left(X_{t+dt/2}\right) dW_t, \end{aligned} \quad (36)$$

and to the lowest order,

$$(dX_t^\pm)^2 = \sigma^2\left(X_{t+dt/2}\right)(dW_t)^2. \quad (37)$$

You should note that the $dW_{t\pm dt}$ is defined differently from the left-point approximation. Since Wiener process are symmetric, $(dW_{t+dt})^2 = (dW_{t-dt})^2 = (dt)^2$, such that when evaluating at mid points, the quadratic contribution cancels.

$$df(X_t) = f(X_{t+dt}) - f(X_t) = \partial_t f dt + [\partial_X f] dX_t^M + \cdots , \quad (38)$$

The above differential is called the **Stratonovich** form,

$$dX_t^M = \mu\left(X_{t+dt/2}\right) dt + \sigma\left(X_{t+dt/2}\right) dW_t = \mu(X_t) dt + \sigma(X_t) \circ dW_t, \quad (39)$$

such that the stochastic differential follows the usual chain rule given in regular calculus when evaluating at the mid point,

$$df = \partial_t f dt + (\partial_{X_t} f) \circ dX_t = \partial_t f dt + (\partial_{X_t} f) dX_t^M. \quad (40)$$

You should note that the circle $\circ$, representing the mid point evaluation of the differential. We also note that the following mid point approximation,

$$X_{t+dt/2} = \frac{X_{t+dt} + X_t}{2}. \quad (41)$$

If we want to convert between the **Stratonovich** and **Ito** formulation, we expand the expression with respect to the left point X_t. To the linear order, we have,

$$\begin{aligned}
\sigma\left(X_{t+dt/2}\right) &= \sigma\left(X_t\right) + \partial_X \sigma\left(X_t\right)\left(\frac{X_{t+dt}+X_t}{2} - X_t\right), \\
&= \sigma\left(X_t\right) + \partial_X \sigma\left(X_t\right)\left(\frac{X_{t+dt}-X_t}{2}\right), \\
&= \sigma\left(X_t\right) + \frac{1}{2}\partial_X \sigma\left(X_t\right) dX_t.
\end{aligned} \tag{42}$$

We then plug in to the mid point differential,

$$\begin{aligned}
dX_t^M &= \mu\left(X_{t+dt/2}\right) dt + \sigma\left(X_{t+dt/2}\right) dW_t = \mu(X_t)dt + \sigma(X_t) \circ dW_t \\
&= \mu\left(X_{t+dt/2}\right) dt + \left[\sigma\left(X_t\right) + \frac{1}{2}\partial_X \sigma\left(X_t\right) dX_t\right] dW_t \\
&= \mu\left(X_t\right) dt + \left[\sigma\left(X_t\right) dW_t + \frac{1}{2}\partial_X \sigma\left(X_t\right)\left[\mu(X_t)dtdW_t + \sigma(X_t)(dW_t)^2\right]\right] \\
&= \left(\mu\left(X_t\right) + \frac{1}{2}\sigma\left(X_t\right)\partial_X \sigma\left(X_t\right)\right) dt + \sigma\left(X_t\right) dW_t,
\end{aligned} \tag{43}$$

where in the second from the last line, we use the Ito representation,

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t. \tag{44}$$

Finally, we have the conversion between the two,

$$dX_t^M = \mu(X_t)dt + \sigma(X_t) \circ dW_t = \left(\mu\left(X_t\right) + \frac{1}{2}\sigma\left(X_t\right)\partial_X \sigma\left(X_t\right)\right) dt + \sigma\left(X_t\right) dW_t. \tag{45}$$

Or in terms of the Ito differential,

$$dX_t^M = dX_t + \frac{1}{2}\sigma\left(X_t\right)\partial_X \sigma\left(X_t\right) dt. \tag{46}$$

Let W_t be a standard Brownian motion and consider the following quadratic function,

$$g(t, W_t) = W_t^2. \tag{47}$$

The differential is given by Ito Lemma,

$$dg = \frac{\partial g}{\partial t} dt + \frac{\partial g}{\partial W_t} dW_t + \frac{1}{2}\frac{\partial^2 g}{\partial W_t^2}(dW_t)^2. \tag{48}$$

Recall that the quadratic variation of Brownian motion satisfies $(dW_t)^2 = dt$. The derivatives are given by,

$$\frac{\partial g}{\partial t} = 0, \quad \frac{\partial g}{\partial W_t} = 2W_t, \quad \frac{\partial^2 g}{\partial W_t^2} = 2. \tag{49}$$

Plug the derivatives into Ito lemma:

$$d(W_t^2) = 0 \cdot dt + 2W_t\, dW_t + \frac{1}{2} \cdot 2 \cdot dt = 2W_t\, dW_t + dt \tag{50}$$

It shows that even though W_t^2 might appear differentiable, there is an extra dt term appears due to the quadratic variation of W_t. Integrating both sides from 0 to t gives the trajectory,

$$W_t^2 = 2\int_0^t W_s\, dW_s + t. \tag{51}$$

If we set $X = g = (W_t)^2$, under the Ito formulation, we have,

$$dX_t^I = 2W_t\, dW_t + dt. \tag{52}$$

The Stratonovich integrator,

$$dX_t^S = dX_t^I + \frac{1}{2}\sigma\,(X_t)\,\partial_X \sigma\,(X_t)\,dt. \tag{53}$$

In this case, from Ito formulation, we know $\sigma(X_t) = 2W_t = \pm 2\sqrt{X_t}$ and $\partial_X \sigma(X_t) = \pm 1/\sqrt{X_t}$. Plug in the above Stratonovich integrator, we have,

$$dX_t^S = dX_t^I + \frac{1}{2}(\pm 1/\sqrt{X_t})(\pm 2\sqrt{X_t})dt = 2W_t\, dW_t + 2dt. \tag{54}$$

We see that the correction term for the drift when converting from Ito to Stratonovich adds a constant offset to the drift term.

3.1 Fokker-Planck equation, Forward and Backward Kolmogorov Equation

Let's consider the following,

$$\langle df \rangle = \langle \left(\partial_t + \mu \partial_{X_t} + \frac{\sigma^2}{2}\partial_{X_t}^2 \right) f \rangle\, dt = [\langle \partial_t f \rangle + \langle Af \rangle] dt, \tag{55}$$

where A is the **Markov infinitesimal generator** with expectation value (single integral because of the Markov property),

$$\langle f \rangle = \int f(X_t, t) \cdot p(X_t, t) dX_t. \tag{56}$$

The continuity equation for the test function and the probability density function must hold true at all times, i.e.,

$$\frac{d}{dt} f(X_t, t) = \frac{d}{dt} p(X_t, t) = 0. \tag{57}$$

Divide dt on both sides and assume f and p vanish at the boundaries, with integration by parts, we have,

$$\begin{aligned}
\langle \frac{df}{dt} \rangle &= \langle \left(\partial_t + \mu \partial_{X_t} + \frac{\sigma^2}{2} \partial^2_{X_t} \right) f \rangle = \langle \partial_t f \rangle + \langle A^\dagger f \rangle \\
-\langle \partial_t f \rangle &= \int [\mu \partial_{X_t} + \frac{\sigma^2}{2} \partial^2_{X_t}] f(X_t, t) \cdot p(X_t, t) dX_t \\
-\langle \partial_t f \rangle &= \int \left(-\partial_{X_t} [\mu p(X_t, t)] + \partial^2_{X_t} [\frac{\sigma^2}{2} p(X_t, t)] \right) f(X_t, t) dX_t.
\end{aligned} \tag{58}$$

Similarly, we can directly calculate the expectation value with an operator A,

$$\begin{aligned}
\int f(X_t, t) \cdot \frac{d}{dt} p(X_t, t) dX_t &= \int f(X_t, t) \cdot \partial_t p(X_t, t) dX_t + \int f(X_t, t) \cdot [Ap] dX_t \\
-\int f(X_t, t) \cdot \partial_t p(X_t, t) dX_t &= \int f(X_t, t) \cdot [Ap] dX_t.
\end{aligned} \tag{59}$$

From this equation, we must have,

$$A = \left(-\partial_{X_t} [\mu p(X_t, t)] + \partial^2_{X_t} [\frac{\sigma^2}{2} p(X_t, t)] \right). \tag{60}$$

the same expectation implies,

$$\langle A^\dagger f \rangle = \int A^\dagger f(X_t, t) \cdot p(X_t, t) dX_t = \int f(X_t, t) \cdot Ap(X_t, t) dX_t, \tag{61}$$

where $A^\dagger$ is the **adjoint operator** of A,

$$A^\dagger = \mu \partial_{X_t} + \frac{\sigma^2}{2} \partial^2_{X_t}. \tag{62}$$

To hold true for all test functions, the integrand of the above expression must be equal,

$$\partial_t p(X_t, t) = -\partial_{X_t} [\mu(X_t, t) p(X_t, t)] + \partial^2_{X_t} \left(\frac{\sigma^2(X_t, t)}{2} p(X_t, t) \right) = -\partial_{X_t} J(X_t, t), \tag{63}$$

where we condition on the initial condition $(X_{t'}, t')$, and we define the probability current $J(X_t, t)$, which is similar to the expression in the **continuity equation** with the total derivative for every possible state (source free), $\frac{d}{dt} p = 0$.

The above equation is the **Fokker-Planck equation** (also called **Forward Kolmogorov Equation**) for the probability density function.

The evolution also holds true for all test function, so that,

$$-\partial_t f(X_t, t) = \mu(X_t, t) \partial_{X_t} f(X_t, t) + \frac{\sigma^2(X_t, t)}{2} \partial^2_{X_t} f(X_t, t). \tag{64}$$

This is the **Backward Kolmogorov Equation**. You may notice that there is an extra minus sign that you did not see in the birth-death process. This is **conventional**: in practice, we would like the time to run backwards from $t_f \to t$. The above forward equation tracks how probability spreads out from an initial condition. The backward Kolmogorov equation tracks how probability contracts when viewed from a final condition. We can condition on a future time t_f and evolve backward. We should note the similarity between the forward and backward equation discussed in the birth-death process.

As a side note, if the system is **non-Markov**, we must incorporate the **memory kernel** when calculating the expectation,

$$\langle f \rangle(t) = \int_{-\infty}^{\infty} f(X_t, t) \left[\int_{-\infty}^{t} \int_{-\infty}^{\infty} K(t-s) p(X_t, t \mid X_s, s) p(X_s, s) \, dX_s \, ds \right] dX_t. \tag{65}$$

4 Formulation of the Langevin Equation

The classical **Langevin equation** describes the motion of a particle of mass m in a fluid medium with friction and random forces,

$$m \frac{d^2 \boldsymbol{x}(t)}{dt^2} = -\gamma \frac{d\boldsymbol{x}(t)}{dt} + \boldsymbol{F}(t), \tag{66}$$

where $\boldsymbol{x}(t)$ is the position vector of the particle, γ is the friction coefficient, and $\boldsymbol{F}(t)$ is a stochastic force vector representing the effect of collisions with the molecules of the medium.

This equation can be rewritten in the form of first-order differential equations by introducing the velocity vector $\boldsymbol{v}(t) = \frac{d\boldsymbol{x}(t)}{dt}$:

$$m \frac{d\boldsymbol{v}(t)}{dt} = -\gamma \boldsymbol{v}(t) + \boldsymbol{F}(t). \tag{67}$$

The stochastic force $\boldsymbol{F}(t)$ is often modeled as Gaussian white noise with the following isotropic conditions:

$$\langle \boldsymbol{F}(t) \rangle = 0, \quad \langle F_i(t) F_j(t') \rangle = D \delta(t-t') \delta_{ij}, \tag{68}$$

where D is the noise strength, $\delta(t-t')$ is the Dirac delta function, and δ_{ij} is the Kronecker delta, ensuring that different components of $\boldsymbol{F}(t)$ are uncorrelated. The noise strength D is related to the temperature T of the system via the **fluctuation-dissipation theorem**,

$$D = 2 k_B T \gamma, \tag{69}$$

where k_B is the Boltzmann constant.

4.1 Derivation of the fluctuation-dissipation theorem

To see how the diffusion and dissipation are related, let's consider a particle of mass m whose velocity $v(t)$ obeys the Langevin equation in one dimension,

$$m \, \dot{v} = -\gamma v + \xi(t), \tag{70}$$

where $\xi(t)$ is a stochastic force follows the above isotropic condition, and γ is the friction coefficient. Let's approximate the **force fluctuation** as $\xi(t) \sim \sqrt{D}\, dW_t/dt$ such that **the stochastic differential** dv from the Langevin equation is,

$$dv = -\frac{\gamma}{m} v\, dt + \frac{\sqrt{D}}{m}\, dW_t, \tag{71}$$

and to the lowest order, from Ito formulation, we have,

$$(dv)^2 = \frac{D}{m^2}\, dt, \tag{72}$$

Since we are interested in the kinetic energy dissipation $d(v^2)$ associated with the Langevin equation, we need to calculate the quadratic variation in veolcity $d(v^2)$. Rememver, v is a stochastic variable, and we must use the Ito lemma, such that $f(X_t = v) = v^2$, $dv = dW_t$, and $\sigma = \sqrt{D}/m$,

$$\begin{aligned} d(v^2) &= \left(\partial_t + \frac{\sigma^2}{2}\partial_v^2\right)(v^2)dt + (\partial_v[v^2])dv \\ &= \frac{D}{m^2} dt + 2v\, dv. \end{aligned} \tag{73}$$

dv is a stochastic term, unlike the classical case. The extra term $\frac{D}{m^2}\, dt$ is a direct consequence of the nonzero quadratic variation of the Weiner process dW_t and is essential for stochastic calculus.

We are interested in the time evolution of the mean square velocity $\langle v^2 \rangle$ that is directly related to the kinetic energy. In Ito calculus, the differential of v^2 is given by,

$$d(v^2) = 2v\left(-\frac{\gamma}{m} v\, dt + \frac{1}{m}\, dW\right) + \frac{D}{m^2}\, dt. \tag{74}$$

Taking the ensemble average and noting that $\langle v\, dW \rangle = 0$, it follows:

$$\frac{d}{dt}\langle v^2 \rangle = -\frac{2\gamma}{m}\langle v^2 \rangle + \frac{D}{m^2}. \tag{75}$$

At thermal equilibrium, the system's statistical properties are stationary, such that:

$$\frac{d}{dt}\langle v^2 \rangle = 0. \tag{76}$$

Setting the derivative to zero yields:

$$0 = -\frac{2\gamma}{m}\langle v^2 \rangle + \frac{D}{m^2}, \tag{77}$$

and solving for $\langle v^2 \rangle$, we obtain:

$$\langle v^2 \rangle = \frac{D}{2\gamma m}. \tag{78}$$

From statistical mechanics, at equilibrium, the equipartition theorem suggests that,

$$\frac{1}{2}m\langle v^2\rangle = \frac{1}{2}k_BT, \quad \rightarrow \quad \langle v^2\rangle = \frac{k_BT}{m}. \tag{79}$$

We then equate the two expressions for $\langle v^2\rangle$,

$$\frac{D}{2\gamma m} = \frac{k_BT}{m}. \tag{80}$$

Canceling the mass m and solving for D, we have

$$\frac{D}{2\gamma} = k_BT \quad \rightarrow \quad D = 2\gamma k_BT. \tag{81}$$

This result is a direct consequence of the fluctuation-dissipation theorem, which connects the strength of force field fluctuations dW_t to the dissipation γ (friction) in the system.

4.2 Solution of the Langevin Equation

We start with the stochastic differential equation (SDE):

$$dv(t) = -\gamma v(t)\,dt + \sqrt{D}\,dW(t), \tag{82}$$

with the initial condition, $v(0) = v_0$. Let's first multiply both sides by the integrating factor $e^{\gamma t}$:

$$e^{\gamma t}\,dv(t) = -\gamma e^{\gamma t}\,v(t)\,dt + \sqrt{D}\,e^{\gamma t}\,dW(t). \tag{83}$$

Notice that the left-hand side is the differential of $e^{\gamma t}v(t)$,

$$d\left(e^{\gamma t}v(t)\right) = e^{\gamma t}\,dv(t) + \gamma e^{\gamma t}\,v(t)\,dt. \tag{84}$$

Therefore, we have:

$$d\left(e^{\gamma t}v(t)\right) = \sqrt{D}\,e^{\gamma t}\,dW(t). \tag{85}$$

Integrate both sides from 0 to t:

$$e^{\gamma t}v(t) - v(0) = \sqrt{D}\int_0^t e^{\gamma s}\,dW(s). \tag{86}$$

Since $v(0) = v_0$, this gives:

$$v(t) = v_0e^{-\gamma t} + \sqrt{D}e^{-\gamma t}\int_0^t e^{\gamma s}\,dW(s). \tag{87}$$

Since the Wiener process has zero mean, $\langle dW(s)\rangle = 0$, the expectation value of the velocity is,

$$\langle v(t)\rangle = v_0e^{-\gamma t}. \tag{88}$$

The second moment contributed by the stochastic term is:

$$(v - v_0)^2 = De^{-2\gamma t} \int_0^t e^{2\gamma s}\, ds. \tag{89}$$

Evaluating the integral:

$$e^{-2\gamma t} \int_0^t e^{2\gamma s}\, ds = \frac{1 - e^{-2\gamma t}}{2\gamma}, \tag{90}$$

and the variance is:

$$\mathrm{Var}[v] = \frac{D}{2\gamma} \left(1 - e^{-2\gamma t}\right). \tag{91}$$

In the long-time limit ($t \to \infty$), the process reaches a stationary state with:

$$\mathrm{Var}[v(\infty)] = \frac{D}{2\gamma} = k_B T, \tag{92}$$

as the ratio between the diffusion and dissipation coefficients, directly related to the thermal energy $k_B T$.

5 Correlation function and autocorrelation

Correlation quantifies the degree to which two random variables, X and Y, are related. One fundamental measure is the covariance, defined as:

$$\mathrm{cov}(X, Y) = \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle, \tag{93}$$

where $\langle \cdot \rangle$ denotes the ensemble or time average. A normalized version, the correlation coefficient, is given by:

$$\rho_{XY} = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}, \tag{94}$$

which lies in the interval $[-1, 1]$ and measures the strength and direction of a linear relationship between X and Y.

For a time-dependent variable $X(t)$, the **autocorrelation function** describes how the value of X at one time is correlated with its value at another time. It is defined as:

$$C(\tau) = \langle X(t) X(t + \tau) \rangle, \tag{95}$$

where τ is the time lag. For a stationary process, the autocorrelation depends only on τ (and not on the absolute time t), with the normalization condition:

$$C(0) = \langle X^2(t) \rangle. \tag{96}$$

A rapidly decaying $C(\tau)$ indicates that the process loses memory quickly, while a slow decay implies long-range correlations.