AI2613: Informal Definition of Diffusion and Examples

May 17, 2024

1 The Definition of Diffusion

Informally, a continuous stochastic process with Markov property is called a diffusion. In other words, a diffusion can be viewed as a Markov process in continuous time with continuous sample paths.

Example 1. *Examples and non-example of diffusions.*

- *The standard Brownian motion $\{ W_t \}_{t\geq 0}$ is a diffusion.*
- *$X_t = \mu t + \sigma W_t$ is a diffusion. Since X_t is obtained by scaling and translating W_t, it is easy to verify that $X_t \sim N\left(\mu t, \sigma^2 t\right)$. Thus, we call X_t a (μ, σ^2)-Brownian motion.*
- *Poisson process is* not *a diffusion since its sample path is not continuous. It is a* jump process*, another important family of stochastic processes.*

We have studied discrete Markov chains over the past few weeks. From now on, we will mainly focus on Markov chains over continuous time, which are usually called Markov processes. We will learn tools to work with them. It would soon be clear that the continuity property of time allows us to apply powerful tools in the analysis to manipulate the processes.

The first question is how to describe a Markov process. Recalling our treatment for the familiar discrete Markov chains, we usually have the following two ways to describe a chain:

1. by giving the transition matrix $P_t(i, j)$, or
2. by locally describing how to get X_{t+1} from X_t.

We will develop the analogue of the transition matrix in future lectures. Today, we try to describe how the process evolves locally. Let us start with the simplest case: a continuous function such as $X_t = e^t$ is obviously a diffusion. Another way to describe it is through an *ordinary differential equation* (ODE):

$$\begin{cases} \mathrm{d}X_t = e^t \,\mathrm{d}t \\ X_0 = 1 \end{cases},$$

This tells us where we are and in which direction to go at each time t. Similarly, we can use a *stochastic differential equation* (SDE) to describe a stochastic process. Consider SBM $\{ W_t \}_{t\geq 0}$. We know that it starts from $W_0 = 0$ and at time t, its movement in a very small amount of time $\mathrm{d}t$ is $\mathrm{d}W_t := W_{t+dt} - W_t \sim \mathcal{N}(0, \mathrm{d}t)$. We denote this random variable with the law $\mathcal{N}(0, \mathrm{d}t)$ as $\mathrm{d}B_t$. So the standard Brownian motion can be described by the SDE:

$$\begin{cases} \mathrm{d}W_t = \mathrm{d}B_t \\ W_0 = 0 \end{cases}$$

For the (μ, σ^2)-Brownian motion, the corresponding SDE is

$$\begin{cases} \mathrm{d}X_t = \mu \,\mathrm{d}t + \sigma \,\mathrm{d}B_t \\ X_0 = 0 \end{cases}$$

From the above example, we see that we can use an SDE of the following form to describe a diffusion:

$$\mathrm{d}X_t = \mu(t, X_t) \,\mathrm{d}t + \sigma(t, X_t) \,\mathrm{d}B_t,$$

which means that at time t, in a small amount of time dt, the process proceeds by moving some distance following the law $\mathcal{N}(\mu(t, X_t), \sigma(t, X_t)^2)$. We call a solution of the above SDE an Itô diffusion. We will make everything rigorous in the next lecture.

2 Some Important Diffusions

2.1 Ornstein-Uhlenbeck Process

Consider a diffusion with SDE

$$\mathrm{d}X_t = -X_t \,\mathrm{d}t + \mathrm{d}B_t.$$

If we ignore the term $\mathrm{d}B_t$, then it is easy to see that X_t tends to 0. In fact, $\mathrm{d}B_t$ acts as a disturbance, and X_t will tend to a normal distribution (we will prove this in the future lectures).

The process can be used to model the discrete Ehrenfest process. Suppose we have N balls distributed into two boxes. In each round, we choose a ball uniformly at random among the N balls and put the chosen ball into the other box. It is more likely to choose the balls in the box with more balls. Thus, this discrete Markov process tends to the equilibrium state where each box has $\frac{N}{2}$ balls.

Let Y_t denote the number of balls in one box, and we will choose to move a ball every $\Delta t = \frac{1}{N}$ time interval. We claim that

$$Z_t = \frac{Y_t - \frac{N}{2}}{\sqrt{N}}$$

is exactly an Ornstein-Uhlenbeck process as N approaches infinity.

First we calculate the expectation of ΔZ_t.

$$\begin{aligned} \mathbf{E}\left[\Delta Z_t\right] &= \left(1 - \frac{Y_t}{N}\right) \frac{1}{\sqrt{N}} - \frac{Y_t}{N} \frac{1}{\sqrt{N}} \\ &= \frac{1}{\sqrt{N}} \frac{N - 2Y_t}{N} = -2\frac{Z_t}{N} = -2Z_t \Delta t. \end{aligned}$$

Thus the variance is

$$\mathrm{Var}\left(\Delta Z_t\right) = \frac{1}{N} - 4\frac{Z_t^2}{N^2} \approx \frac{c}{N},$$

which is equal to the variance of $\sqrt{c} \cdot \mathcal{N}\left(0, \Delta t\right)$. Hence the SDE is

$$\mathrm{d}Z_t = -2Z_t \,\mathrm{d}t + \sqrt{c} \,\mathrm{d}B_t$$

Therefore, it is difficult to directly analyze the discrete process, but we can analyze the Ornstein-Uhlenbeck process (which is easier) when N is large.

2.2 Wright-Fisher Process

Next consider a random walk with absorbing boundaries. Let $\mu(x) = 0$, $\sigma^2(x) = x(1-x)$ and $X(0) = \frac{1}{2}$. Then this diffusion is jittery around $\frac{1}{2}$ and is more steady around the boundaries.

The process can be used to model the following model of racial reproduction. Assume the total population is N which is invariant over time. At the t-th generation, there is X_t black people and $N - X_t$ white people where X_t is a nonnegative random variable. Assume that there is no interracial marriage and the childs race is the same with his or her parents. At the $t+1$-th generation, each person is white w.p. $1 - \frac{X_t}{N}$ and is black w.p. $\frac{X_t}{N}$. Assume the race of each individual is independent with other people. If it starts with half white and half black, then we want to ask: Will there be genocide after a long period of time or will the two races tend to keep a balance?

The continuous version of the model is the Wright-Fisher process we just introduced. It is equivalent to ask whether the process tends to keep jittery or be absorbed. Since it seems to be "lazier" when it comes closer to the boundary, the answer of this question is not obvious. In fact, however, after a sufficiently long time, it does reach the boundary.

2.3 Langevin Dynamics

Recall that if we want to perform gradient descent to minimize a function $f(x)$, we iteratively update the following equation:

$$X_{t+1} - X_t = -\eta \nabla f(X_t).$$

It is in fact the *Euler* discretization of the following ODE, called *gradient flow*:

$$\mathrm{d}X_t = -\nabla f(X_t)\,\mathrm{d}t.$$

We can add some *white noises* to the process to obtain the Langevin dynamics:

$$\mathrm{d}X_t = -\nabla f(X_t)\,\mathrm{d}t + \mathrm{d}B_t.$$

We will show in future lectures that the stationary distribution of this Markov process is

$$p(x) \propto e^{-f(x)}.$$

This indicates that if we want to sample from $p(x)$, we only need to know the derivative of f and then update the discretized Langevin dynamics:

$$X_{t+1} - X_t = -\eta \nabla f(X_t) + \sqrt{2\eta} B_t.$$

For strongly convex f, we will further show that it converges exponentially fast, in the way similar to the convergence of gradient descent.