[AI2613 Lecture 12] Brownian Bridge, Diffusion

May 31, 2022

1 Kolmogorov-Smirnov Test

In this section, we introduce an application of Brownian Bridge, the Kolmogorov-Smirnov test.

Suppose that $U_1, U_2, \dots, U_n$ are independently sampled from some distribution $[0,1]$ with CDF F. We would like to check if it is a uniform distribution, i.e., if the F satisfies $F(t) = t$ for every $t \in [0,1]$.

Let $\widehat{F}_n$ be the empirical cumulative distribution function, that is, for $t \in [0,1]$, $\widehat{F}_n(t) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}[U_i \le t]$. It then follows from the law of large numbers that

$$\widehat{F}_n(t) \overset{n\to\infty}{\longrightarrow} \mathbf{E}\left[\widehat{F}_n(t)\right] = \frac{1}{n}\sum_{i=1}^n \mathbf{Pr}\left[U_i \le t\right] = F(t).$$

The idea of Kolmogorov-Smirnov test is to monitor the variable $\widehat{F}_n(t) - t$ for every $t \in [0,1]$ and reject the uniformity hypothesis if there exists some t that $\left|\widehat{F}_n(t) - t\right|$ is large. Then our goal is to find a suitable rejection threshold b such that if F is indeed a uniform distribution, the failure probability $\lim_{n\to\infty} \mathbf{Pr}\left[\max_{t\in[0,1]} \left|\widehat{F}_n(t) - t\right| \ge b\right]$ is sufficiently small (i.e., $\le \frac{1}{100}$). If F is a uniform distribution, for a fixed t, we have

$$\begin{aligned}
\mathbf{E}\left[\widehat{F}_n(t)\right] &= F(t) = t;\\
\mathbf{Var}\left[\widehat{F}_n(t)\right] &= \frac{1}{n^2}\sum_{i=1}^n \mathbf{Var}\left[\mathbf{1}[U_i \le t]\right] = \frac{1}{n}\cdot t(1-t).
\end{aligned}$$

Let $X_n(t) \triangleq \sqrt{n}\cdot(\widehat{F}_n(t) - t)$ for $t \in [0,1]$. By the Central Limit Theorem, we have $X_n(t) \sim \mathcal{N}(0, t(1-t))$ when $n \to \infty$. For any $0 \le s \le t \le 1$,

$$\begin{aligned}
\mathrm{Cov}(X_n(s), X_n(t)) &= n \cdot \mathrm{Cov}\left(\widehat{F}_n(s) - s, \widehat{F}_n(t) - t\right)\\
&= \frac{1}{n}\mathrm{Cov}\left(\sum_{i=1}^n \mathbf{1}[U_i \le s], \sum_{i=1}^n \mathbf{1}[U_i \le t]\right)\\
&= \mathrm{Cov}\left(\mathbf{1}[U_1 \le s], \mathbf{1}[U_1 \le t]\right)\\
&= \mathbf{Pr}\left[U_1 \le s, U_1 \le t\right] - \mathbf{Pr}\left[U_1 \le s\right]\mathbf{Pr}\left[U_1 \le t\right]\\
&= s(1-t).
\end{aligned}$$

For any $0 \le t_1 \le t_2 \le \cdots \le t_k \le 1$, let $\Sigma = \left(\mathrm{Cov}\left(X_n(t_i), X_n(t_j)\right)\right)_{i,j}$. It follows from the high-dimensional Central Limit Theorem that

$$(X_n(t_1), X_n(t_2), \dots, X_n(t_k))^T \overset{D}{\longrightarrow} \mathcal{N}(\mathbf{0}, \Sigma) \sim (X(t_1), X(t_2), \dots, X(t_k))^T,$$

where $\{X(t)\}$ is a standard Brownian Bridge. Then using the result in Example 3 in Lecture 11, we have

$$\lim_{n\to\infty} \mathbf{Pr}\left[\max_{t\in[0,1]} \widehat{F}_n(t) - t \geq b\right] = \mathbf{Pr}\left[\max_{t\in[0,1]} X(t) \geq \sqrt{n}b\right]$$
$$= \mathbf{Pr}\left[\tau_{\sqrt{n}b} < 1 \,\middle|\, W(1) = 0\right] = \exp\left\{-2nb^2\right\}.$$

2 Diffusion

2.1 The Definition of Diffusion

A continuous stochastic process with Markov property is called a diffusion.[1] In other words, a diffusion can be viewed as a Markov process in continuous time with continuous sample paths.

[1] This is an informal definition of diffusions and it is enough for this course.

Actually, diffusions can be built up from local Brownian motions in the same way as differentiable functions being built up from local linear functions. Imagine that we want to draw the image of a function f with knowing $f'(t) = e^t$ and $f(0) = 1$. How to do this if you are not allowed to integrate $f'(t)$. A natural idea is to approximate f using segmented linear functions:

- Select a step length h;
- Draw a segment on $[0, h]$ which starts from $(0, f(0))$ with slope $f'(0) = 1$;
- Draw a segment on $[h, 2h]$ which starts from $(h, h + f(0))$ with slope $f'(h) = e^h$;
- ....

When $h \to 0$, our drawing is exactly the image of f. This gives an intuition that a differentiable function can be locally approximated as linear functions.

A diffusion $\{X(t)\}_{t\geq 0}$ is the stochastic analog of above process. That is, if we are currently at the position $X(t) = X_t$ and consider the small time interval $[t, t+h]$, the process acts as a $\left(\mu(X_t), \sigma^2(X_t)\right)$ Brownian motion where μ and σ^2 are functions of the position X_t. Let $Z \sim \mathcal{N}(0,1)$ be a standard Gaussian. We can break the process into segments and use these normal random variables to simulate the diffusion:

- $X_h = X_0 + \mu(X_0)h + \sigma(X_0)\sqrt{h} \cdot Z_1$;
- $X_{2h} = X_h + \mu(X_h)h + \sigma(X_h)\sqrt{h} \cdot Z_2$;
- $\dots$,

where each Z_i are independent standard Gaussian. Then for any $k \in \mathbb{N}$, $X_{(k+1)h} - X_{kh} \sim \mathcal{N}\left(\mu(X_{kh})h, \sigma^2(X_{kh})h\right)$.

Thus, when $h \to 0$, we can naturally develop a specification of diffusion: A time homogeneous diffusion can be specified by two functions $\mu(x)$ and $\sigma^2(x)$ which satisfies:

- $\forall t$, $\mathbf{E}[X(t+h) - X(t) \mid X(t) = x] = \mu(x)h + o(h)$;
- $\forall t$, $\mathbf{Var}[X(t+h) - X(t) \mid X(t) = x] = \sigma^2(x)h + o(h)$;
- $\forall t$, $\mathbf{E}\left[|X(t+h) - X(t)|^p \;\middle|\; X(t) = x\right] = o(h)$ for $p > 2$.

Note that

$$
\begin{aligned}
&\mathbf{Var}[X(t+h) - X(t) \mid X(t) = x] \\
&= \mathbf{E}\left[(X(t+h) - X(t))^2 \;\middle|\; X(t) = x\right] - (\mathbf{E}[X(t+h) - X(t) \mid X(t) = x])^2 \\
&= \mathbf{E}\left[(X(t+h) - X(t))^2 \;\middle|\; X(t) = x\right] - (\mu(x)h + o(h))^2.
\end{aligned}
$$

Thus

$$\mathbf{Var}[X(t+h) - X(t) \mid X(t) = x] = \sigma^2(x)h + o(h)$$

is equivalent to

$$\mathbf{E}\left[(X(t+h) - X(t))^2 \;\middle|\; X(t) = x\right] = \sigma^2(x)h + o(h).$$

Recall that in the analog of differentiable functions, we have $\mathrm{d}f(t) = g(t)\,\mathrm{d}t$ where $g(t)$ is the derivative of f. Similarly, for a diffusion $\{X(t)\}_{t\geq 0}$ specified by $\mu(x)$ and $\sigma^2(x)$, we can write it as

$$\mathrm{d}X(t) = \mu(X(t))\,\mathrm{d}t + \sigma(X(t))\,\mathrm{d}W(t),$$

where $\{W(t)\}$ is the standard Brownian motion and $\mathrm{d}W(t)$ can be understood as $\lim_{h\to 0} W(t+h) - W(t)$.

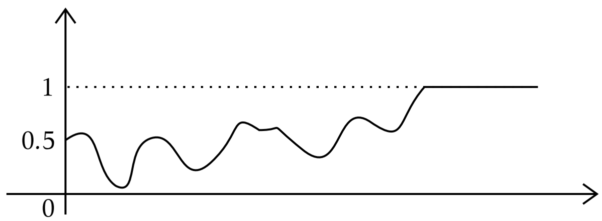
This expression is informal as we haven't give a mathematical meaning to the notation $dW(t)$. We will do this in the next lecture, but for now, we can sloppily understand it as a Brownian motion in an infinitisimal time.

Example 1 (Ornstein-Uhlenbeck Process) *Consider a diffusion $\{X(t)\}_{t\geq 0}$ specified by $\sigma(x) = -x$ and $\sigma^2(x) = 2$ with $X(0) = 0$. This diffusion always has a tendency to 0 since if $X(t)$ is large, $\mu(X(t))$ is also large towards the reverse direction which acts as a spring intuitively. We can write this process as*

$$\mathrm{d}X(t) = -X(t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}W(t).$$

The process can be used to model the discrete Ehrenfest chain. Suppose we have two boxes with a balls in the first box and b balls in the second box in the initial state. In each round, we choose a ball uniformly at random among the $a+b$ 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 $\frac{a+b}{2}$ balls.

Example 2 (Wright-Fisher Process) *Next we consider a stochastic 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 non-negative random variable. Assume that there is no interracial marriage and the child's 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.

2.2 Diffusions with a Stopping Time

Let $\{X(t)\}$ be a diffusion specified by $\mu(x)$ and $\sigma^2(x)$, or equivalently,

$$\mathrm{d}X(t) = \mu(X(t))\,\mathrm{d}t + \sigma(X(t))\,\mathrm{d}W(t).$$

The diffusion ends at a random stopping time T. Let $g : \mathbb{R} \to \mathbb{R}$ be a cost function that the diffusion pays at rate $g(x)$ when it arrives at x. We are interested in the average cost of the diffusion, that is, $\mathbf{E}\left[\int_0^T g(X(t))\,\mathrm{d}t\right]$.

Let $w(x) \triangleq \mathbf{E}_x\left[\int_0^T g(X(t))\,\mathrm{d}t\right] = \mathbf{E}\left[\int_0^T g(X(t))\,\mathrm{d}t \,\middle|\, X(0) = x\right]$. We have the following theorem.

Theorem 1 *The function w satisfies the differential equation:*

$$\mu(x)w'(x) + \frac{1}{2}\sigma^2(x)w''(x) = -g(x).$$

Before proving this theorem, we see an application of it. Recall the 1-D random walk with two absorbing barriers $-a$ and b which we discussed in Lecture 7. We used the tool of martingale to show that the average stopping time is ab. If we consider the continuous process, that is, a diffusion specified by $\mu(x) = 0$ and $\sigma^2(x) = 1$ with $X(0) = 0$ and $g(x) \equiv 1$, we have $\mathbf{E}[T] = \mathbf{E}\left[\int_0^T 1 \, \mathrm{d}t\right] = w(0)$ where $T = \min_t \{X(t) = -a \vee X(t) = b\}$. With Theorem 1, we have

$$\frac{1}{2} w''(x) = -1.$$

Combining the fact that $w(-a) = w(b) = 0$, we have $w(x) = -(x+a)(x-b)$ and consequentlly $\mathbf{E}[T] = w(0) = ab$.

Then we prove Theorem 1.

Proof. [Proof of Theorem 1] Pick a sufficiently small h such that $\mathbf{Pr}[T > h] = 1 - o(h)$. Then we have

Whether such h exists requires justification and depends on T. Nevertheless, we assume so.

$$\begin{aligned} w(x) &= \mathbf{E}_x\left[\int_0^T g(X(t))\,\mathrm{d}t\right] \\ &= \mathbf{E}_x\left[\int_0^h g(X(t))\,\mathrm{d}t\right] + \mathbf{E}_x\left[\int_h^T g(X(t))\,\mathrm{d}t\right] \\ &= \mathbf{E}_x\left[\int_h^T g(X(t))\,\mathrm{d}t\right] + h \cdot g(x) + o(h). \end{aligned} \tag{1}$$

Note that

$$\begin{aligned} \mathbf{E}_x\left[\int_h^T g(X(t))\,\mathrm{d}t\right] &= \mathbf{E}_x\left[\mathbf{E}_x\left[\int_h^T g(X(t))\,\mathrm{d}t \,\middle|\, X(h)\right]\right] \\ &= \mathbf{E}_x\left[\mathbf{E}_x\left[\int_h^T g(X(t))\,\mathrm{d}t \,\middle|\, X(h), T > h\right]\right] + o(h) \\ &= \mathbf{E}_x[w(X(h))] + o(h). \end{aligned} \tag{2}$$

Using Taylor's expansion, we have

$$\begin{aligned} \mathbf{E}_x[w(X(h))] &= \mathbf{E}_x\left[w(x) + w'(x)(X(h)-x) + \frac{1}{2}w''(x)(X(h)-x)^2\right] + o(h) \\ &= w(x) + h \cdot w'(x)\mu(x) + \frac{h}{2} \cdot w''(x)\sigma^2(x) + o(h). \end{aligned} \tag{3}$$

Combining Equation (1), Equation (2) and Equation (3), we have

$$h \cdot g(x) + h \cdot w'(x)\mu(x) + \frac{h}{2} w''(x)\sigma^2(x) + o(h) = 0.$$

Thus,

$$\begin{aligned} 0 &= \lim_{h \to 0} \frac{h \cdot g(x) + h \cdot w'(x)\mu(x) + \frac{h}{2} w''(x)\sigma^2(x) + o(h)}{h} \\ &= g(x) + \mu(x)w'(x) + \frac{1}{2}\sigma^2(x)w''(x). \end{aligned}$$

□

2.3 Geometric Brownian Motion

Let $\{X(t)\}$ be a (μ, σ^2) Brownian motion, that is, $\mathrm{d}X(t) = \mu\,\mathrm{d}t + \sigma\,dW(t)$. Define $Y(t) = e^{X(t)}$. Then $\{Y(t)\}$ is called a geometric Brownian motion. Geometric Brownian motion is widely applied to model the stock prices in finance. In fact, we can consider a more generalized situation that $\{Y(t)\}$ is defined by $Y_t = f(X_t)$ where f is strictly monotone and twice differentiable. Then we have the following proposition.

Proposition 2 *Suppose $\{X(t)\}$ is a diffusion specified by $\mu_X(x)$ and $\sigma_X^2(x)$. Let f be a strictly monotone and twice differentiable function. Define $Y(t) = f(X(t))$. Then $\{Y(t)\}$ is a diffusion specified by $\mu_Y(y)$ and $\sigma_Y^2(y)$ which satisfy*

$$\mu_Y(y) = \mu_X(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x) \quad \text{and} \quad \sigma_Y^2(y) = (f'(x))^2\,\sigma_X^2(x)$$

where $x = f^{-1}(y)$.

Proof. For a small h, we have

$$\begin{aligned}\mathbf{E}\left[Y(t+h) - Y(t) \,\middle|\, Y(t) = y\right] &= \mathbf{E}\left[f(X(t+h)) - f(X(t)) \mid X(t) = x\right] \\ &= \mathbf{E}\left[f'(X(t))(X(t+h) - X(t)) + \frac{1}{2}f''(X(t))(X(t+h) - X(t))^2 \,\middle|\, X(t) = x\right] + o(h) \\ &= \mu_X(x)f'(x)h + \frac{1}{2}\sigma^2(x)f''(x)h + o(h),\end{aligned}$$

so that

$$\mu_Y(y) = \lim_{h\to 0}\frac{\mathbf{E}\left[Y(t+h) - Y(t) \,\middle|\, Y(t) = y\right]}{h} = \mu_X(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x).$$

Similarly, we have

$$\begin{aligned}\mathbf{E}\left[(Y(t+h) - Y(t))^2 \,\middle|\, Y(t) = y\right] &= \mathbf{E}\left[(f(X(t+h)) - f(X(t)))^2 \,\middle|\, X(t) = x\right] \\ &= \mathbf{E}\left[(f'(X(t))(X(t+h)) - X(t))^2 \,\middle|\, X(t) = x\right] + o(h) \\ &= (f'(x))^2\,\sigma_X^2(x)h + o(h),\end{aligned}$$

so that

$$\sigma_Y^2(y) = \lim_{h\to 0}\frac{\mathbf{E}\left[(Y(t+h) - Y(t))^2 \,\middle|\, Y(t) = y\right]}{h} = (f'(x))^2\,\sigma_X^2(x).$$

□