Lecture 17 : Stochastic Processes II

1 Continuous-time stochastic process

So far we have studied discrete-time stochastic processes. We studied the concept of Makov chains and martingales, time series analysis, and regression analysis on discrete-time stochastic processes.

We now turn our focus to the study of continuous-time stochastic processes. In most cases, it is difficult to exactly describe the probability distribution for continuous-time stochastic processes. This was also difficult for discrete time stochastic processes, but for them, we described the distribution in terms of the increments $X_{k+1} - X_k$ instead; this is impossible for continuous time stochastic processes. An alternate way which is commonly used is to first describe the properties satisfied by the probability distribution, and then to show that there exists a probability distribution satisfying the given properties. Unfortunately, the second part above, the acutal construction, requires a non-trivial amount of work and is beyond the scope of this class. Hence here we provide a brief introduction to the framework, and mostly just state the properties of the stochastic processes of interest. Interested readers can take more advanced probability courses for deeper understanding.

To formally define a stochastic process, there needs to be an underlying probability space $(\Omega, \mathbf{P})$. A stochastic process X is then a map from the universe Ω to the space of real functions defined over $[0, \infty)$. Hence the probability of the stochastic process taking a particular path in some set A can be computed by computing the probability $\mathbf{P}(X^{-1}(A))$. When there is a single stochastic process, it is more convenient to just consider Ω as the space of all possible paths. Then $\mathbf{P}$ directly describes the probability distribution of the stochastic process. The more abstract view of taking an underlying abstract universe Ω is useful when there are several stochastic processes under consideration (for example, when changing measure). We use the letter ω to denote an element of Ω, or one possible path of the process (in most cases, the two describe the same object).

2 Standard Brownian motion

We first introduce a continuous-time analogue of the simple random walk, known as the standard Brownian motion. It is also refered to as the Wiener process, named after Norbert Wiener, who was a professor at MIT. The first person who actually considered this process is Bachelier, who used Brownian motion to evaluate stocks and options in his Ph.D thesis written in 1900 (see [3]).

Theorem 2.1. *There exists a probability distribution over the set of continuous functions $B : \mathbb{R} \to \mathbb{R}$ satisfying the following conditions:*

(i) $B(0) = 0$.

(ii) ***(stationary)*** *for all $0 \leq s < t$, the distribution of $B(t) - B(s)$ is the normal distribution with mean 0 and variance $t - s$, and*

(iii) ***(independent increment)*** *the random variables $B(t_i) - B(s_i)$ are mutually independent if the intervals $[s_i, t_i]$ are nonoverlapping.*

We refer to a particular instance of a path chosen according to the Brownian motion as a *sample Brownian path*.

One way to think of standard Brownian motion is as a limit of simple random walks. To make this more precise, consider a simple random walk $\{Y_0, Y_1, \cdots,\}$ whose increments are of mean 0 and variance 1. Let Z be a piecewise linear function from $[0, 1]$ to $\mathbb{R}$ defined as

$$Z\left(\frac{t}{n}\right) = Y_t,$$

for $t = 0, \cdots, n$, and is linear at other points. As we take larger values of n, the distribution of the path Z will get closer to that of the standard Brownian motion. Indeed, we can check that the distribution of $Z(1)$ converges to the distribution of $N(0, 1)$, by central limit theorem. More generally, the distribution of $Z(t)$ converges to $N(0, t)$.

Example 2.2. (i) [From wikipedia] In 1827, the botanist Robert Brown, looking through a microscope at particles found in pollen grains in water, noted that the particles moved through the water but was not able to determine the mechanisms that caused this motion. Atoms and molecules had long been theorized as the constituents of matter, and many decades later, Albert Einstein published a paper in 1905 that explained in precise detail how the motion that Brown had observed was a result of the pollen being moved by individual water molecules.

(ii) Stock prices can also be modelled using standard Brownian motions.

Here are some facts about the Brownian motion:

1. Crosses the x-axis infinitely often.
2. Has a very close relation with the curve $x = y^2$ (it does not deviate from this curve too much).
3. Is nowhere differentiable.

Note that in real-life we can only observe the value of a stochastic process up to some time resolution (in other words, we can only take finitely many sample points). The fact above implies that standard Brownian motion is a reasonable model, at least in this sense, since the real-life observation will converge to the underlying theoretical stochastic process as we take smaller time intervals, as long as the discrete-time observations behave like a simple random walk.

Suppose we use the Brownian motion as a model for daily price of a stock. What is the distribution of the days range? (the max value and min value over a day)

Define $M(t) = \max_{0 \le s \le t} B(s)$, and note that $M(t)$ is well-defined since B is continuous and $[0, t]$ is compact. ($\Phi(t)$ is the cumulative distribution function of the normal random variable)

Proposition 2.3. *The following holds:*

$$\mathbf{P}(M(t) \ge a) = 2\mathbf{P}(B(t) > a) = 2 - 2\Phi(\frac{a}{\sqrt{t}}).$$

Proof. Let $\tau_a = \min_s\{s : B(s) = a\}$ and note that τ_a is a stopping time. Note that for all $0 \le s < t$, we have

$$\mathbf{P}(B(t) - B(s) > 0) = \mathbf{P}(B(t) - B(s) < 0).$$

Hence we see that

$$\mathbf{P}(B(t) - B(\tau_a) > 0 \,|\, \tau_a < t) = \mathbf{P}(B(t) - B(\tau_a) < 0 \,|\, \tau_a < t).$$

Here we assumed that the distribution of $B(t) - B(\tau_a)$ is not affected by the fact that we conditioned on $\tau_a < t$. This is called the Strong Markov Property of the Brownian motion.

This can be rewritten as

$$\mathbf{P}(B(t) > a \,|\, \tau_a < t) = \mathbf{P}(B(t) < a \,|\, \tau_a < t),$$

and is also known as the 'reflection principle'.

Now observe that

$$\begin{aligned}\mathbf{P}(M_t \geq a) &= \mathbf{P}(\tau_a < t) \\ &= \mathbf{P}(B(t) > a \,|\, \tau_a < t) + \mathbf{P}(B(t) < a \,|\, \tau_a < t). \\ &= 2\mathbf{P}(B(t) > a \,|\, \tau_a < t).\end{aligned}$$

Since

$$\mathbf{P}(B(t) > a \,|\, \tau_a < t) = \mathbf{P}(B(t) > a),$$

our claim follows. □

The proposition above also has very interesting theoretical implication. Using the proposition above, we can prove the following result.

Proposition 2.4. *For each $t \geq 0$, the Brownian motion is almost surely not differentiable at t.*

Proof. Fix a real t_0 and suppose that the Brownian motion B is differentiable at t_0. Then there exist constants A and ε_0 such that for all $0 < \varepsilon < \varepsilon_0$, $B(t) - B(t_0) \leq A\varepsilon$ holds for all $0 < t - t_0 \leq \varepsilon$. Let $E_{\varepsilon,A}$ denote this event, and $E_A = \bigcap_\varepsilon E_{\varepsilon,A}$. Note that

$$\begin{aligned}\mathbf{P}(E_{\varepsilon,A}) &= \mathbf{P}(E(t) - E(t_0) \leq A\varepsilon \quad \text{for all } 0 < t - t_0 \leq \varepsilon) \\ &= \mathbf{P}(M(\varepsilon) \leq A\varepsilon) = 2(1 - \Phi(A\sqrt{\varepsilon})),\end{aligned}$$

where the right hand side tends to zero as ε goes to zero. Therefore, $\mathbf{P}(E_A) = 0$. By countable additivity, we see that there can be no constant A satisfying above (it suffices to consider integer values of A). □

Dvoretsky, Erdős, and Kakutani in fact proved a stronger statment asserting that the Brownian motion B is nowhere differentiable with probability 1. Hence a sample Brownian path is continuous but nowhere differentiable! The proof is slightly more involved and requires a lemma from probability theory (Borel-Cantelli lemma).

Theorem 2.5. *(Quadratic variation) For a partition $\Pi = \{t_0, t_1, \cdots, t_j\}$ of an interval $[0, T]$, let $|\Pi| = \max_i(t_{i+1} - t_i)$. A Brownian motion B_t satisfies the following equation with probability 1:*

$$\lim_{|\Pi| \to 0} \sum_i (B_{t_{i+1}} - B_{t_i})^2 = T.$$

Proof. For simplicity, here we only consider partitions where the gaps $t_{i+1} - t_i$ are uniform. In this case, the sum

$$\sum_i (B_{t_{i+1}} - B_{t_i})^2$$

is a sum of i.i.d. random variables with mean $t_{i+1} - t_i$, and finite second moment. Therefore, by the law of large numbers, as $\max\{t_{i+1} - t_i\} \to 0$, we have

$$\sum_i (B_{t_{i+1}} - B_{t_i})^2 = T$$

with probability 1. □

Why is this theorem interesting? Suppose that instead of a Brownian motion, we took a function f that is continuously differentiable. Then

$$\sum_i \Big(f(t_{i+1}) - f(t_i)\Big)^2 \leq \sum_i (t_{i+1} - t_i)^2 f'(s_i)^2 \leq \max_{s\in[0,T]} f'(s)^2 \cdot \sum_i (t_{i+1} - t_i)^2.$$
$$\leq \max_{s\in[0,T]} f'(s)^2 \cdot \max_i \{t_{i+1} - t_i\} \cdot T.$$

As $\max\{t_{i+1} - t_i\} \to 0$, we see that the above tends to zero. Hence this shows that Brownian motion fluctuates a lot. The above can be summarized by the differential equation $(dB)^2 = dt$. As we will see in the next lecture, this fact will have very interesting implications.

Example 2.6. (Brownian motion with drift) Let $B(t)$ be a Brownian motion, and let μ be a fixed real. The process $X(t) = B(t) + \mu t$ is called a *Brownian motion with drift* μ. By definition, it follows that $\mathbb{E}[X(t)] = \mu t$. Question : as time passes, which term will dominate? $B(t)$ or μt? It can be shown that μt dominates the behavior of $X(t)$. For example, for all fixed $\varepsilon > 0$, after long enough time, the Brownian motion will always be between the lines $y = (\mu - \varepsilon)t$ and $y = (\mu + \varepsilon)t$.

What is the main advantage of the continuous-world against the discrete world? The beauty, of course, is one advantage. A more practical advantage is the powerful toolbox of calculus. Unfortunately, we saw that it is impossible to differentiate Brownian motion. Surprisingly, there exists a theory of generalized calculus that can handle Brownian motions, and other continuous-time stochastic processes. This will be the topic of the remaining lectures.

Suppose we want to go further. As discussed in previous lecture, when modelling the price of a stock, it is more reasonable to assume that the percentile change follows a normal distribution. This can be written in the following differential equation:

$$dS_t = \sigma S_t dB_t.$$

Can we write the distribution of S_t in terms of the distribution of B_t? Is it $S_t = e^{\sigma B_t}$? Surprisingly, the answer is no.

References

[1] S. Ross, A first course in probability

[2] D. Bertsekas, J. Tsitsiklis, Introduction to probability

[3] L. Bachelier, Théorie de la spéculation, Annales Scientifiques de l'École Normale Supérieure, **3**, 21-86.

[4] R. Durrett, Probability: Theory and Examples, 3rd edition.

[5] S.R.Srinivasa Varadhan, Lecture Note (http://www.math.nyu.edu/faculty/varadhan/spring06/spring06.1.pdf)