AI2613 随机过程 2020–2021 春季学期

Lecture 13 – Martingale (II), Brownian Motion (I)

2021 年 5 月 24 日

Lecturer: 张驰豪 *Scribe*: 杨宽

1 Review of the Optional Stopping Theorem

Suppose that a stochastic process $\{X_n\}_{n\geq 0}$ is defined on a filtration $\{\mathcal{F}_n\}_{n\geq 0}$. Then $\{X_n\}$ is called a

1. **Martingale** if $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n$;
2. **Supermartingale** if $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] < X_n$;
3. **Submartingale** if $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] > X_n$.

If $\{X_n\}$ is a martingale with respect to a filtration $\{\mathcal{F}_n\}$, applying the law of total expectation we have $\mathbb{E}[X_n] = \mathbb{E}[X_0]$ for all fixed $n \geq 0$. So we wonder what happens if n is a random variable.

Theorem 1 (Optional Stopping Theorem). *Suppose that $\{X_n\}$ is a martingale with respect to a filtration $\{\mathcal{F}_n\}$ and τ is a stopping time with respect to the same filtration. Then $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$ if at leat one of the following holds*

1. *τ is bounded;*
2. *$\mathbf{Pr}[\tau < \infty] = 1$ and $\exists M$ such that $|X_n| \leq M$ for all $n < \tau$;*
3. *$\mathbb{E}[\tau] < \infty$ and $\exists c$ such that $\mathbb{E}[|X_{n+1} - X_n| \mid \mathcal{F}_n] \leq c$ for all $n < \tau$.*

Remark. If $\{X_n\}$ is a supermartingale (or submartingale), and at least one of the conditions holds, then the result holds as well, namely, $\mathbb{E}[X_\tau] < \mathbb{E}[X_0]$ (or $\mathbb{E}[X_\tau] > \mathbb{E}[X_0]$).

2 Supermartingale Convergence

Today we are talking about supermartingales.

For a supermartingale, it always holds that $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] < \mathbb{E}[X_n]$. So intuitively, the trend of a supermartingale should be declining. Since every decreasing and bounded below sequence of real numbers is convergent, hopefully $\{X_n\}$ should also converge with probability 1 if $\{X_n\}$ is a nonnegative supermartingale.

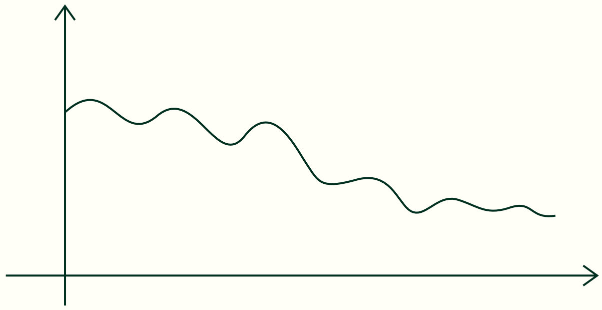

Figure 1: A possible figure of a supermartingale

Now we are going to formalize this idea.

Proposition 2. *Suppose that $\{X_n\}$ is a nonnegative supermartingale with $X_0 \le a$. For all $b > a$, define T_b by*

$$T_b \triangleq \inf\{t : X_t \ge b\}.$$

Then it holds that

$$\mathbf{Pr}[T_b < \infty] \le \frac{a}{b}.$$

Proof. Let

$$a \wedge b \triangleq \min\{a, b\},$$
$$a \vee b \triangleq \max\{a, b\}.$$

Fix $t > 0$. Clearly $T_b \wedge t$ is a stopping time. It is an exercise to verify that the condition of optional stopping theorem is satisfied. Therefore, applying the optional stopping theorem, it follows that

$$\mathbb{E}\left[X_{T_b \wedge t}\right] < \mathbb{E}[X_0] \le a.$$

On the other hand,

$$X_{T_b \wedge t} = \begin{cases} X_t, & \text{if } T_b > t; \\ \ge b, & \text{if } T_b \le t. \end{cases}$$

So we obtain that

$$X_{T_b \wedge t} \ge b \cdot \mathbb{1}_{[T_b \le t]},$$

which implies that

$$\mathbb{E}\left[X_{T_b \wedge t}\right] \ge b \cdot \mathbf{Pr}[T_b \le t].$$

Combining with $\mathbb{E}\left[X_{T_b \wedge t}\right] \le a$ we conclude that $\mathbf{Pr}[T_b \le t] \le a/b$ for all $t > 0$. □

We are ready to show the following theorem.

Theorem 3. *Any nonnegative supermartingale converges with probability* 1.

Proof. Assume that a supermartingale is divergent. Then there are two cases. One is that there exists a subsequence goes to infinity, and another is that there exists an oscillating subsequence. If there exists a subsequence goes to infinity with probability > 0, then for every $b > 0$, the probability of $\inf\{t : X_t \geq b\} < \infty$ is greater than 0, which contradicts Proposition 2. So it suffices to show that oscillation does not exist.

Suppose that a supermartingale $\{X_n\}$ has two subsequence that converge to different values a and b. W.l.o.g. we further assume that $a < b$. Let $\varepsilon < (b-a)/2$. Then there exists a subsequence of $\{X_n\}$ bounded above by $a' \triangleq a + \varepsilon$ and exists a subsequence bounded below by $b' \triangleq b - \varepsilon$. We now show that this situation happens with probability 0.

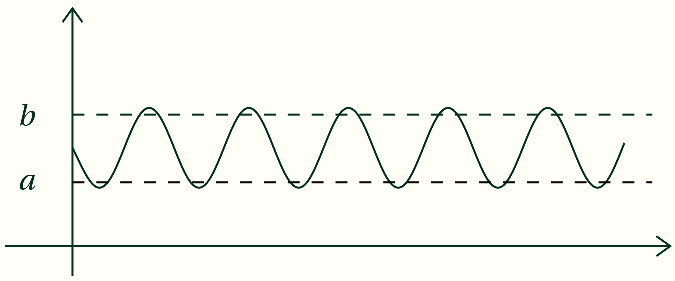

Figure 2: An oscillating sequence

Fix $a < b$ arbitrarily. We define the following stopping times. Let

$$
\begin{aligned}
&T_0 = 0, \\
&S_1 = \inf\{t : t > T_0 \wedge X_t \leq a\}, \\
&T_1 = \inf\{t : t > S_1 \wedge X_t \geq b\}, \\
&S_2 = \inf\{t : t > T_1 \wedge X_t \leq a\}, \\
&\quad \cdots\cdots
\end{aligned}
$$

Since $S_k \leq T_k$ for all $k \in \mathbb{N}$, the optional stopping theorem implies that for all $n \in \mathbb{N}$,

$$
\mathbb{E}\left[X_{S_k \wedge n}\right] \geq \mathbb{E}\left[X_{T_k \wedge n}\right]. \tag{1}
$$

Note that

$$
X_{T_k \wedge n} = \begin{cases} \geq b, & \text{if } T_k \leq n; \\ X_n, & \text{if } T_k > n. \end{cases}
$$

So it follows that

$$X_{T_k \wedge n} \geq b \cdot \mathbb{1}_{[T_k \leq n]} + X_n \cdot \left(1 - \mathbb{1}_{[T_k \leq n]}\right),$$

and hence

$$\mathbb{E}[T_k \wedge n] \geq b \cdot \mathbf{Pr}[T_k \leq n] + X_n \cdot (1 - \mathbf{Pr}[T_k \leq n]). \tag{2}$$

Using the same argument we obtain

$$\mathbb{E}[S_k \wedge n] \leq a \cdot \mathbf{Pr}[S_k \leq n] + X_n \cdot (1 - \mathbf{Pr}[S_k \leq n]). \tag{3}$$

Plugging (2) and (3) into (1) and using the fact $S_k < T_k$, we have

$$\begin{aligned}
& a \cdot \mathbf{Pr}[S_k \leq n] + X_n \cdot (1 - \mathbf{Pr}[S_k \leq n]) \geq b \cdot \mathbf{Pr}[T_k \leq n] + X_n \cdot (1 - \mathbf{Pr}[T_k \leq n]) \\
\Longrightarrow \quad & a \cdot \mathbf{Pr}[S_k \leq n] \geq b \cdot \mathbf{Pr}[T_k \leq n] + X_n \cdot (\mathbf{Pr}[S_k \leq n] - \mathbf{Pr}[T_k \leq n]) \\
\Longrightarrow \quad & a \cdot \mathbf{Pr}[S_k \leq n] \geq b \cdot \mathbf{Pr}[T_k \leq n] \\
\Longrightarrow \quad & \mathbf{Pr}[T_k \leq n] \leq \frac{a}{b} \cdot \mathbf{Pr}[S_k \leq n] \\
\Longrightarrow \quad & \mathbf{Pr}[T_k \leq \infty] \leq \frac{a}{b} \cdot \mathbf{Pr}[S_k \leq \infty].
\end{aligned}$$

Since $S_k > T_{k-1}$, it implies that for all $k \in \mathbb{N}$,

$$\mathbf{Pr}[T_k \leq \infty] \leq \frac{a}{b} \cdot \mathbf{Pr}[T_{k-1} \leq \infty].$$

Taking the limit to the both sides, we conclude that $\forall\, \varepsilon > 0$, there exists $n \in \mathbb{N}$ s.t. $\mathbf{Pr}[T_n < \infty] < \varepsilon$.

□

3 Stochastic Approximation

We now consider an important application of supermartingales. This example is called *stochastic approximation.*

Suppose a function $f : \mathbb{R} \to \mathbb{R}$ has an unknown unique zero point (w.l.o.g. we further assume that $f(0) = 0$ but we do not know). To find the zero point we could use the binary search. However we do not know the exact value of f. Every time we ask an oracle for the value of f at some point x, it returns a number $\tilde{f}(x) = f(x) + \eta$ instead. Here η is a random variable with $\mathbb{E}[\eta] = 0$ and $\mathrm{Var}[\eta] = 1$.

Suppose that $f(x) > 0$ if $x >$ the zero point and $f(x) < 0$ if $x <$ the zero point. Then we guess the value of the zero point. Denote our guesses by a sequence $X_0, X_1, X_2, \ldots$. Given X_n, the oracle returns $Y_n = \tilde{f}(X_n) = f(X_n) + \eta_n$ and we let $X_{n+1} = X_n - a_n \cdot Y_n$ where a_n is to be determined. We hope that $X_n \to 0$ (the true zero point) with probability 1.

Now our goal is to determine a_n. Intuitively a necessary condition is that $a_n \to 0$ as $n \to \infty$ and a_n should not decrease too fast. Formally the following theorem tells us under which condition we will obtain $X_n \to 0$.

Theorem 4 (Stochastic Approximation). *Let $f\colon \mathbb{R} \to \mathbb{R}$ and $X_0, X_1, X_2, \ldots, Y_0, Y_1, Y_2, \ldots$ are two sequences of random variables such that $\mathbb{E}\left[(X_0)^2\right] < \infty$ and*

$$\begin{aligned} Y_n &= f(X_n) + \eta_n\,, \\ X_{n+1} &= X_n - a_n \cdot Y_n\,. \end{aligned}$$

Then $X_n \to 0$ as $n \to 0$ if the followings hold

1. *$X_0, \eta_1, \eta_2, \ldots$ are independent; $\mathbb{E}\left[\eta_i\right] = 0$, and $\operatorname{Var}\left[\eta_i\right] = 1$;*
2. *$\left|f(x)\right| < c \cdot |x|$ for some $c > 1$; (Lipschitz condition)*
3. *$\forall\, \delta > 0$, $\inf_{|x|>\delta} x \cdot f(x) > 0$;*
4. *$a_n \ge 0$, $\sum a_n = \infty$ but $\sum a_n^2 < \infty$.*

Proof. We would like to construct a supermartingale. To achieve this goal, we first compute $\mathbb{E}\left[X_{n+1}^2 \mid \overline{X}_{0,n}\right]$. By $\mathbb{E}[\eta_n] = 0$, $\mathbb{E}[\eta_n^2] = 1$ and $\left|f(x)\right| < c|x|$, we have that

$$\begin{aligned} \mathbb{E}\left[X_{n+1}^2 \mid \overline{X}_{0,n}\right] &= \mathbb{E}\left[\left(X_n - a_n(f(X_n) + \eta_n)\right)^2 \mid \overline{X}_{0,n}\right] \\ &= \mathbb{E}\left[X_n^2 \mid \overline{X}_{0,n}\right] - \mathbb{E}\left[2a_n X_n(f(X_n) + \eta_n) \mid \overline{X}_{0,n}\right] + \mathbb{E}\left[a_n^2(f(X_n) + \eta_n)^2 \mid \overline{X}_{0,n}\right] \\ &= X_n^2 - 2a_n X_n f(X_n) + a_n^2\left(\mathbb{E}\left[f(X_n)^2 \mid \overline{X}_{0,n}\right] + 1\right) \\ &\le X_n^2 + a_n^2\left(c^2 X_n^2 + c^2\right) \\ &= \left(a_n^2 c^2 + 1\right) X_n^2 + a_n^2 c^2\,. \end{aligned}$$

Now, it is clear to justify that $\left\{W_n \triangleq b_n\left(X_n^2 + 1\right)\right\}_{n\ge 0}$ is supermartingale with respect to $\{X_n\}$, where b_n is given by

$$b_n \triangleq \prod_{k=1}^{n-1} \left(1 + a_k^2 c^2\right)^{-1}.$$

Applying Theorem 3, it follows that

$$\lim_{n\to\infty} W_n = \xi \qquad \text{for some } \xi > 0\,.$$

Note that

$$W_n = \frac{X_n^2 + 1}{\prod_{k=1}^{n-1}(1 + a_k^2 c^2)} .$$

Since $1 + a_k^2 c^2 \le e^{a_k^2 c^2}$, we obtain that

$$\prod_{k=1}^{n-1}(1 + a_k^2 c^2) \le e^{c^2 \sum a_k^2}$$

converge as $n \to \infty$. Therefore $\{X_n^2\}_{n\ge 0}$ is convergent.

We now show that $X_n \to \delta$ for some $\delta > 0$ with positive probability is impossible. Fix $\delta > 0$. $\forall\, m > 0$, we define the following *bad event*. Let

$$\mathcal{B}_m \triangleq \bigcap_{n \ge m} \{X_n \in D\}$$

where $D \triangleq \{x : |x| > \delta\}$. It is sufficient to show $\mathbf{Pr}[\mathcal{B}_m] = 0$ for all m.

By Condition 3, there exists $\varepsilon > 0$ s.t. $\inf_{x\in D} x \cdot f(x) \ge \varepsilon$. So we obtain that

$$X_n \cdot f(X_n) \ge \varepsilon \cdot \mathbb{1}_{[X_n \in D]} .$$

Taking the expectation to the both sides, it implies that

$$\mathbb{E}\big[X_n \cdot f(X_n)\big] \ge \varepsilon \cdot \mathbf{Pr}[X_n \in D] \ge \varepsilon \cdot \mathbf{Pr}[\mathcal{B}_m] .$$

Recall that $\{W_n\}$ is a supermartingale, and we further have (by the computation above) that

$$\mathbb{E}\Big[W_{n+1} \mid \overline{X}_{0,n}\Big] < W_n - 2a_n b_{n+1} X_n \cdot f(X_n) .$$

Taking the expectation to the both sides, it yields that

$$\begin{aligned}
\mathbb{E}[W_{n+1}] &< \mathbb{E}[W_n] - 2a_n b_{n+1} \mathbb{E}\big[X_n \cdot f(X_n)\big] \\
&< \mathbb{E}[W_n] - 2a_n b_{n+1} \varepsilon \cdot \mathbf{Pr}[\mathcal{B}_m] \\
&< \mathbb{E}[W_m] - 2\varepsilon \sum_{k=m}^{n} a_k b_{k+1} \cdot \mathbf{Pr}[\mathcal{B}_m] ,
\end{aligned}$$

which yields that

$$\mathbf{Pr}[\mathcal{B}_m] \le \frac{\mathbb{E}[W_m] - \mathbb{E}[W_{n+1}]}{2\varepsilon \sum_{k=m}^{n} a_k b_{k+1}} \longrightarrow 0 .$$

The last quantity converges to zero because $\mathbb{E}[W_m] - \mathbb{E}[W_{n+1}]$ is bounded for all $n > m$, and

$$2\varepsilon \sum_{k=m}^{n} a_k b_{k+1} > 2\varepsilon b_n \sum_{k=m}^{n} a_k > 2\varepsilon \cdot e^{-c^2 \sum a_k^2} \sum_{k=m}^{n} a_k \to \infty$$

as $n \to \infty$ (since $\sum a_k^2$ converges and $\sum a_k \to \infty$). That completes our proof. □

4 Introduction to Brownian Motion

Brownian motion describes the random motion of small particles suspended in a liquid or in a gas. This process was named after the botanist Robert Brown, who observed and studied a jittery motion of pollen grains suspended in water under a microscope. Later, Albert Einstein gave a physical explanation of this phenomenon.

In mathematics, Brownian motion is characterized by the *Wiener process*, named after Norbert Wiener, a famous mathematician and the originator of cybernetics. Consider a uniform one-dimensional random walk starting from 0. Denote by X_i the i-th step. Then X_i is a uniform random variable in $\{-1,+1\}$. Now suppose that each time unit Δt we take a step of length δ. Let $X(t)$ be our position at time t. So it holds that

$$X(t) = \delta \cdot \left(X_1 + X_2 + \cdots + X_{t/\Delta t}\right).$$

Now we are interested in what happens if Δt and $\delta \to 0$. Since $\mathbb{E}[X_i] = 0$ and $\mathsf{Var}[X_i] = 1$, we have that

$$\mathbb{E}[X(t)] = 0,$$
$$\mathsf{Var}[X(t)] = \delta^2 \cdot \frac{t}{\Delta t}.$$

We would like to obtain a non-trivial stochastic process, so a natural idea is to fix $\delta^2/\Delta t$ to be a constant. Let

$$\delta = \sigma \cdot \sqrt{\Delta t}$$

for some constant $\sigma > 0$. Thus $\mathsf{Var}[X(t)] = \sigma^2 t$. The next question is what the distribution of $X(t)$ is? The central limit theorem tells us $X(t)$ has a normal distribution $\mathcal{N}(0, \sigma^t)$.

Theorem 5 (Central Limit Theorem). *Suppose that $X_1, X_2 \ldots,$ is a sequence of i.i.d. random variables with mean μ and variance σ^2. Then for $n \to \infty$, it holds that*

$$\frac{\sum X_i - n\mu}{\sigma\sqrt{n}} \sim \mathcal{N}(0,1).$$

Or equivalently,

$$\lim_{n\to\infty} \mathbf{Pr}\left[\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \le a\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} \mathrm{e}^{-x^2/2}\,\mathrm{d}x.$$

Let $Y_i = \delta X_i$ in our setting. Then $\mathsf{Var}[Y_i] = \sigma^2 \Delta t$. Applying the central limit theorem, it follows that

$$X(t) = \sum_{k=1}^{t/\Delta t} Y_k \sim \sigma^2 \Delta t \cdot \sqrt{\frac{t}{\Delta t}} \cdot \mathcal{N}(0,1) = \mathcal{N}(0, \sigma^2 t)$$

as $\Delta t \to 0$. This argument gives an intuition of Wiener process. We now formalize the definition.

Definition 1 (Brownian Motion (Wiener Process)). A stochastic process $\{X(t)\}_{t\geq 0}$ is said to be a *Brownian motion* if

- $X(0) = 0$.
- **Independent increments.** $\forall\, 0 = t_0 < t_1 < \cdots < t_n$,

$$X(t_n) - X(t_{n-1}), \quad X(t_{n-1}) - X(t_{n-2}), \quad \ldots, \quad X(t_1) - X(t_0)$$

  are independent.
- **Stationary increments.** $\forall\, t, s > 0$, $X(t+s) - X(t)$ only depends on s, and has a normal distribution $\mathcal{N}(0, \sigma^2 s)$ for some constant σ.
- $X(t)$ is continuous.