CHAPTER 5: MARTINGALES

December 1, 2021

Contents

1 Conditional Expectation

1.1 Definition

Definition. X is a random variable on $(\Omega, \mathcal{F}, P)$ with $E|X| < \infty$. $\mathcal{A}$ is a σ-field and $\mathcal{A} \subset \mathcal{F}$. $E(X|\mathcal{A})$, called the **conditional expectation** of X given $\mathcal{A}$, is defined as a random variable Y satisfying

(i) $Y \in \mathcal{A}$, i.e., Y is $\mathcal{A}$-measurable.

(ii) $\forall A \in \mathcal{A}$: $E[X1_A] = E[Y1_A]$.

Remark. (1) Existence and uniqueness to be proved later.

(2) Let $A = \{Y > 0\}$ in (ii):

$$E(Y^+) = E[X1_A] \le E(|X|1_A).$$

Similarly, $E(T^-) \le E(|X|1_{A^c})$. Therefore,

$$E|Y| \le E|X| < \infty.$$

(3) Let $A = \Omega$ in (ii), we have

$$E[Y] = E[X].$$

(4) If $X \in \mathcal{A}$, then $E(X|\mathcal{A}) = X$.

(5) If X and $\mathcal{A}$ are independent, then $E(X|\mathcal{A}) = E(X)$.

Proof of a.s. Uniqueness. Suppose Y_1 and Y_2 both satisfy (i) and (ii) in the above definition, then by letting $A = \{Y_1 - Y_2 \ge \varepsilon > 0\}$ for some constant ε, we have $A \in \mathcal{A}$ and

$$\begin{aligned} 0 =& E(X1_A) - E(X1_A) \\ \overset{(ii)}{=}& E(Y_1 1_A) - E(Y_2 1_A) \\ =& E[(Y_1 - Y_2)1_A] && \text{(linearity of } E) \\ \ge& \varepsilon P(A). \end{aligned}$$

This means $P(A) = 0$ for any $\varepsilon > 0$. Therefore,

$$P(Y_1 > Y_2) = P(Y_1 - Y_2 > 0) = \lim_{\varepsilon \downarrow 0} P(Y_1 - Y_2 \ge \varepsilon) = 0.$$

Similarly, we can show that $P(Y_2 > Y_1) = 0$; hence $Y_1 = Y_2$ a.s. $\square$

Proof of Existence. We use the **Radon-Nikodym theorem** which we will not prove: Let μ and ν be σ-finite measures on $(\Omega, \mathcal{A})$. If $\nu \ll \mu$ (i.e., $\mu(A) = 0$ implies $\nu(A) = 0$), then there is a function $f \in \mathcal{A}$ s.t. $\forall A \in \mathcal{A}$,

$$\int_A f d\mu = \nu(A).$$

Such f, denoted by $d\nu/d\mu$, is called the Radon-Nikodym derivative.

To apply the Radon-Nikodym theorem, we first suppose $X \geq 0$. Let $\mu = P$, $\nu(A) = E(X1_A)$ for $A \in \mathcal{A}$. Then it is clear that $\nu \ll \mu$ and ν is a σ-finite measure. Let $Y = d\nu/d\mu$ be the Radon-Nikodym derivative. From the Radon-Nikodym theorem, we have $Y \in \mathcal{A}$ and $E(Y1_A) = \int_A Y d\mu = \nu(A) = E(X1_A)$. Therefore, Y satisfies (i) and (ii) in the definition of conditional expectation; hence proving the existence. □

1.2 Properties

Theorem 5.1.2. If $E|X|, E|Y| < \infty$, then

$$\begin{aligned} &(a) && E(aX+Y|\mathcal{A}) = aE(X|\mathcal{A}) + E(Y|\mathcal{A}). && \text{(linearity)} \\ &(b) && \text{If } X \leq Y, \text{ then } E(X|\mathcal{A}) \leq E(Y|\mathcal{A}). && \text{(monotonicity)} \\ &(c) && \text{If } X_n \geq 0, X_n \uparrow X, \text{ and } EX < \infty, \text{ then } E(X_n|\mathcal{A}) \uparrow E(X|\mathcal{A}). && \\ &&&&& \text{(monotone convergence theorem)} \end{aligned}$$

Proof. (a) We proceed to verify that the RHS satisfies (i) and (ii) in the definition of the conditional expectation. $\forall\ A \in \mathcal{A}$:

$$\begin{aligned} & E\left[aE(X|\mathcal{A})1_A + E(Y|\mathcal{A})1_A\right] \\ =& aE[E(X|\mathcal{A})1_A] + E[E(Y|\mathcal{A})1_A] && \text{(by linearity of E)} \\ =& aE(X1_A) + E(Y1_A) && \text{(from the definition of conditional expectation)} \\ =& E[(aX+Y)1_A]. && \text{(by linearity of E again)} \end{aligned}$$

This verifies (ii). (i) is obvious.

(b) For an $\varepsilon > 0$, define

$$A = \{E(X|\mathcal{A}) - E(Y|\mathcal{A}) \geq \varepsilon\} \in \mathcal{A}.$$

We have

$$E[E(X|\mathcal{A})1_A] = E(X1_A) \leq E(Y1_A) = E[E(Y|\mathcal{A})1_A].$$

This implies

$$0 \geq E[(E(X|\mathcal{A}) - E(Y|\mathcal{A}))1_A] \geq \varepsilon P(A).$$

Hence $P(A) = 0$. (b) follows by sending ε to 0.

(c) From (a), we have

$$E(X|\mathcal{A}) - E(X_n|\mathcal{A}) = E(X - X_n|\mathcal{A}),$$

which is decreasing in n by (b), say to Z_∞, which is measurable with respect to $\mathcal{A}$. We are left to prove that $Z_\infty = 0$. We have, $\forall\ A \in \mathcal{A}$,

$$E(Z_\infty 1_A) \overset{DCT}{=} \lim_{n\to\infty} E[E(X - X_n|\mathcal{A})1_A] \overset{(ii)}{=} \lim_{n\to\infty} E[(X - X_n)1_A] \overset{DCT}{=} 0.$$

Hence $Z_\infty = 0$. $\square$

Jensen's Inequality. If φ is convex, $E|X| < \infty$, $E|\varphi(X)| < \infty$, then

$$\varphi(E(X|\mathcal{A})) \le E(\varphi(X)|\mathcal{A}).$$

Special case: If $p \ge 1$, then $|E(X|\mathcal{A})|^p \le E(|X|^p|\mathcal{A})$.

Chebyshev's Inequality. If $a > 0$, then

$$P(|X| \ge a|\mathcal{A}) \le a^{-2}E(X^2|\mathcal{A}).$$

Hölder's Inequality. If $p \ge 1$ and $\frac{1}{p} + \frac{1}{q} = 1$, then

$$|E(XY|\mathcal{A})| \le [E(|X|^p|\mathcal{A})]^{1/p}[E(|Y|^q|\mathcal{A})]^{1/q}.$$

The proofs are similar to that for the unconditional version of these inequalities and omitted here.

Theorem 5.1.7. If $X \in \mathcal{A}$, $E|XY| < \infty$, $E|Y| < \infty$, then

$$E(XY|\mathcal{A}) = XE(Y|\mathcal{A}).$$

Proof. We proceed to verify that $XE(Y|\mathcal{A})$ satisfies (i) and (ii) in the definition of conditional expectation. (i) follows from the assumption that $X \in \mathcal{A}$. To verify (ii), we N.T.S. for any $A \in \mathcal{A}$,

$$E(XY1_A) = E[XE(Y|\mathcal{A})1_A]. \tag{1.1}$$

To verify (1.1), we proceed by considering X being an 1. indicator variable, 2. simple variable, 3. positive variable and 4. general variable as before.

1. Suppose $X = 1_B$, $B \in \mathcal{A}$. Then

$$\begin{aligned} RHS(1.1) &= E[1_B 1_A E(Y|\mathcal{A})] = E[E(Y|\mathcal{A})1_{A\cap B}] \\ &\overset{(ii)}{=} E(Y1_{A\cap B}) = E(1_B Y 1_A) = E(XY1_A) = LHS(1.1). \end{aligned}$$

2. Suppose $X = \sum_{i=1}^n b_i 1_{B_i}$. Then

$$E(\sum_{i=1}^n b_i 1_{B_i} Y|\mathcal{A}) \overset{linearity}{=} \sum_{i=1}^n b_i E(1_{B_i} Y|\mathcal{A}) \overset{1.}{=} \sum_{i=1}^n b_i 1_{B_i} E(Y|\mathcal{A}) = XE(Y|\mathcal{A}).$$

3. Use monotone convergence theorem of the conditional expectation above and approximate X by simple variables from below.

4. Write general X as $X^+ - X^-$. $\square$

Theorem 5.1.6 (Tower Property). If $\mathcal{A}_1 \subset \mathcal{A}_2$, then

(a) $E(E(X|\mathcal{A}_1)|\mathcal{A}_2) = E(X|\mathcal{A}_1)$.

(b) $E(E(X|\mathcal{A}_2)|\mathcal{A}_1) = E(X|\mathcal{A}_1)$.

The smaller σ-field always wins.

Proof. In verifying (i) and (ii) in the definition of conditional expectation, (i) is obvious, (ii) follows from

$$\forall\ A \in \mathcal{A}_1: \quad E(E(X|\mathcal{A}_1)1_A) = E(X1_A) = E(E(X|\mathcal{A}_2)1_A).$$

$\square$

Theorem 5.1.8 (Triangular Inequality). Suppose $EX^2 < \infty$. Then for any $Y \in \mathcal{A}$ with $EY^2 < \infty$, we have

$$E(X - E(X|\mathcal{A}))^2 \leq E(X - Y)^2.$$

Proof.

$$\begin{aligned} E(X-Y)^2 =& E(X - E(X|\mathcal{A}) + E(X|\mathcal{A}) - Y)^2 \\ =& E(X - E(X|\mathcal{A}))^2 + E(Y - E(X|\mathcal{A}))^2 - 2E[(X - E(X|\mathcal{A}))(Y - E(X|\mathcal{A}))] \\ \geq& E(X - E(X|\mathcal{A}))^2, \end{aligned}$$

where we note that the third term equals 0 because $Y - E(X|\mathcal{A}) = 0$. $\square$

Below, we clarify the connection to the conditional expectation learned in the elementary probability course.

Definition. For random variables X and Y, we write

$$E(X|Y) := E(X|\sigma(Y)),$$

which can be regarded as a function of Y, say $h(Y)$. We illustrate the computation of $h(\cdot)$ by the following examples.

Example 5.1.4. If X, Y have the joint pdf $f(x,y)$. Then the marginal density of Y is

$$f_Y(y) = \int f(x,y)dx,$$

and the conditional density of X given $Y = y$ is

$$\frac{f(x,y)}{f_Y(y)}.$$

If $E|X| < \infty$, the conditional expectation

$$E(X|Y=y) = \int x\frac{f(x,y)}{f_Y(y)}dx \overset{call\ it}{=}: h(y).$$

Then in this case,

$$E(X|Y) = h(Y). \tag{1.2}$$

Proof of (1.2). For any $A \in \sigma(Y)$, write it as $A = \{Y \in B\}$, we have

$$E[h(Y)1_A] = \int_B h(y) f_Y(y) dy = \int_B \int x \frac{f(x,y)}{f_Y(y)} dx f_Y(y) dy = E[X 1_{Y \in B}] = E[X 1_A].$$

□

Example 5.1.5. If X and Y are independent and $E|g(X,Y)| < \infty$. Then

$$E(g(X,Y)|Y) = h(Y), \tag{1.3}$$

where $h(y) = Eg(X,y)$.
Proof of (1.3). For any $A \in \sigma(Y)$, write it as $A = \{Y \in B\}$, we have

$$\begin{aligned} E[g(X,Y)1_A] =& E[g(X,Y)1_{Y \in B}] = \int \int g(x,y) 1_B(y) d\nu(y) d\mu(x) \\ =& \int h(y) 1_B(y) d\nu(y) = E[h(y) 1_{Y \in B}]. \end{aligned}$$

□

2 Martingales

2.1 Definition; Examples; Properties

Definition. Let $\{\mathcal{F}_n, n \geq 1\}$ be a sequence of increasing σ-fields, i.e., a filtration. Let $\{S_n, n \geq 1\}$ be a sequence of random variables (r.v.). $\{S_n\}$ is called a **martingale** with respect to (w.r.t.) $\{\mathcal{F}_n\}$ if:

(i) $E|S_n| < \infty, \ \forall n \geq 1$

(ii) $S_n \in \mathcal{F}_n, \ \forall n \geq 1$

(iii) $E(S_n|\mathcal{F}_{n-1}) = S_{n-1}, \ \forall n \geq 2.$

We often denote the martingale as $(S_n, \mathcal{F}_n)$ for simplicity.

Definition. Suppose $\{S_n\}$ is a martingale w.r.t. $\{\mathcal{F}_n\}$. Let

$$X_1 := S_1, \ X_2 := S_2 - S_1, \ldots, X_n := S_n - S_{n-1}, \ldots$$

Then $\{X_n\}$ are called **martingale differences**.
Fact 1. Taking expectations on both sides of (iii), we have, for martingales,

$$E(S_1) = E(S_2) = \cdots = E(S_n) = \cdots$$

and, for martingale differences,

$$E(X_i) = 0, \ \forall i \geq 2.$$

Fact 2. Again from (iii), by taking expectations of the absolute value of both sides and then using Jensen's inequality for conditional expectations, we have

$$E|S_1| \leq E|S_2| \leq \cdots \leq E|S_n| \leq \cdots$$

Fact 3. For the above martingales, if $EX_n^2 < \infty \ \forall n$, then

$$E(S_n^2) = \sum_{i=1}^{n} E(X_i^2).$$

Proof of fact 3: Note that

$$\begin{aligned} E(S_{n-1}X_n) =& E(E(S_{n-1}X_n|\mathcal{F}_{n-1})) \\ =& E(S_{n-1}E(X_n|\mathcal{F}_{n-1})) \\ =& 0, \end{aligned}$$

where we used properties of the conditional expectation in the first two equations, and property (iii) of martingales in the last equation. Therefore,

$$E(S_n^2) = E(S_{n-1}^2 + 2S_{n-1}X_n + X_n^2) = E(S_{n-1}^2 + X_n^2) = \cdots = \sum_{i=1}^{n} E(X_i^2).$$

$\square$

Example 1. If $X_1, X_2, \dots$ are independent and $\mathbb{E}X_i = 0$, then $S_n = \sum_{i=1}^n X_i$ is a martingale w.r.t. $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$.
Proof: In verifying the definition of martingales, (i) and (ii) obviously hold. (iii) holds because

$$\begin{aligned} &E(S_n|\mathcal{F}_{n-1}) = E(S_{n-1} + X_n|\mathcal{F}_{n-1}) \\ =&E(S_{n-1}|\mathcal{F}_{n-1}) + E(X_n|\mathcal{F}_{n-1}) = S_{n-1} + E(X_n) = S_{n-1}. \end{aligned}$$

We used properties of the conditional expectation in the second and the third equality above. $\square$
Example 2. If $E|X| < \infty$ and $\{\mathcal{F}_n, n \geq 1\}$ is a filtration. Let $S_n = E(X|\mathcal{F}_n)$. Then $\{S_n\}$ is a martingale w.r.t. $\mathcal{F}_n$.
Proof: In verifying the definition of martingales, (i) holds because

$$E|S_n| = E|E(X|\mathcal{F}_n)| \leq E[E(|X||\mathcal{F}_n)] = E|X| < \infty$$

by Jensen's inequality for conditional expectations. (ii) holds because $E(X|\mathcal{F}_n)$ measurable w.r.t. $\mathcal{F}_n$ by the definition of the conditional expectation. (iii) holds because

$$E(S_n|\mathcal{F}_{n-1}) = E(E(X|\mathcal{F}_n)|\mathcal{F}_{n-1}) = E(X|\mathcal{F}_{n-1}) = S_{n-1}$$

recall the smaller sigma field always wins property. $\square$

Definition. $\{S_n, n \geq 1\}$ is called a **supermartingale** w.r.t. a filtration $\{\mathcal{F}_n, n \geq 1\}$ if:

(i) $E|S_n| < \infty, \ \forall n \geq 1$

(ii) $S_n \in \mathcal{F}_n, \ \forall n \geq 1$

(iii) $E(S_n|\mathcal{F}_{n-1}) \leq S_{n-1}, \ \forall n \geq 2.$

$\{S_n, n \geq 1\}$ is called a **submartingale** if (iii) is changed to

$$E(S_n|\mathcal{F}_{n-1}) \geq S_{n-1}.$$

Fact. It can be verified easily from the above definitions that if $\{S_n, \mathcal{F}_n\}$ is a supermartingale, then $\{-S_n, \mathcal{F}_n\}$ is a submartingale. [Therefore, results for supermartingales can be translated into results for submartingales and vice versa.]

A martingale is both a supermartingale and a submartingale.

Theorem 5.2.3 & 5.2.4. (Martingale transforms)

(1) If $(S_n, \mathcal{F}_n)$ is a martingale and φ is a convex function such that $E|\varphi(S_n)| < \infty\ \forall n$, then $\varphi(S_n)$ is a submartingale.

(2) Suppose $(S_n, \mathcal{F}_n)$ is a submartingale and φ is a convex function such that $E|\varphi(S_n)| < \infty\ \forall n$. Moreover, φ is an increasing function. Then $\varphi(S_n)$ is a submartingale.

Hereafter, when the filtration is obvious, we omit (w.r.t. $\{\mathcal{F}_n\}$).
Special case: If $(S_n, \mathcal{F}_n)$ is a martingale, then $(|S_n|, \mathcal{F}_n)$ is a submartingale.

Proof of Theorem 5.2.3 & 5.2.4. In the definition of submartingales, (i) and (ii) obviously hold. We only need to prove (iii).

For case (1), we use Jensen's inequality to obtain

$$E(\varphi(S_n)|\mathcal{F}_{n-1}) \geq \varphi(E(S_n|\mathcal{F}_{n-1})) = \varphi(S_{n-1}),$$

verifying (iii).

For case (2), we use Jensen's inequality and the additional assumption that φ is increasing to obtain

$$E(\varphi(S_n)|\mathcal{F}_{n-1}) \geq \varphi(E(S_n|\mathcal{F}_{n-1})) \geq \varphi(S_{n-1}),$$

verifying (iii).

□

2.2 Martingale Convergence Theorem

Our main objective is to prove the following

Theorem 5.2.8: Martingale Convergence Theorem. Suppose $(S_n, \mathcal{F}_n)$ is a submartingale and $\liminf E(S_n^+) < \infty$. Then S_n converges almost surely (a.s.) to a limit S with $E|S| < \infty$.

As a consequence, we recover the following result we learnt before.

Corollary 1. Suppose $X_1, X_2, \ldots$ are independent and $E(X_i) = 0$. If $\sum_{i=1}^{\infty} E(X_i^2) < \infty$, then $S_n = \sum_{i=1}^{n} X_i$ converges a.s.

Proof of Corollary 1. Recall that mean zero independent sums S_n is a martingale (hence a submartingale). It suffices to verify the condition of Theorem 5.2.8. This follows from

$$E(S_n^+) \leq E|S_n| \leq \sqrt{E(S_n^2)} = \sqrt{\sum_{i=1}^{\infty} E(X_i^2)} < \infty.$$

□

As another consequence of the martingale convergence theorem, we have the following law of large numbers for martingales.

Corollary 2. Suppose $X_1, X_2, \ldots$ are identically distributed (not necessarily independent) and $E|X_1| < \infty$. Let $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Let $S_1 = X_1$ and $S_n = S_{n-1} + X_n - E(X_n|\mathcal{F}_{n-1}), n \geq 2$. Then $(S_n, \mathcal{F}_n)$ is a martingale and

$$\frac{S_n}{n} \to 0 \quad \text{in probability.}$$

Proof of Corollary 2. We omit the straightforward verification that $(S_n, \mathcal{F}_n)$ is a martingale. Let $Y_i = X_i I(|X_i| \leq i),\ i \geq 1$. Let

$$T_1 = Y_1, \quad T_n = T_{n-1} + Y_n - E(Y_n|\mathcal{F}_{n-1}), n \geq 2.$$

We have

$$\sum_{i=1}^{\infty} P(Y_i \neq X_i) = \sum_{i=1}^{\infty} P(|X_1| > i) < \infty \text{ (from } E|X_1| < \infty) \tag{2.1}$$

and

$$\sum_{i=2}^{\infty} \frac{E[Y_i - E(Y_i|\mathcal{F}_{i-1})]^2}{i^2} \leq \sum_{i=2}^{\infty} \frac{E(Y_i^2)}{i^2} \leq CE|X_1| < \infty, \tag{2.2}$$

where we used the triangle inequality for conditional expectations in the first inequality and some result we proved before when we studied law of large numbers for the second inequality.

From the martingale convergence theorem and (2.2) above, we have

$$\sum_{i=2}^{n} \frac{Y_i - E(Y_i|\mathcal{F}_{i-1})}{i} \text{ converges a.s.} \tag{2.3}$$

From (2.3) and the Kronecker's lemma, we have

$$\frac{1}{n} \sum_{i=2}^{n} (Y_i - E(Y_i|\mathcal{F}_{i-1})) \to 0 \ a.s. \tag{2.4}$$

(2.4) and (2.1) imply

$$\frac{1}{n} \sum_{i=2}^{n} (X_i - E(Y_i|\mathcal{F}_{i-1})) \to 0 \ a.s. \tag{2.5}$$

Corollary 2 follows from (2.5) and

$$\frac{1}{n} \sum_{i=2}^{n} E(Y_i|\mathcal{F}_{i-1}) - \frac{1}{n} \sum_{i=2}^{n} E(X_i|\mathcal{F}_{i-1}) \to 0 \text{ in probability.} \tag{2.6}$$

To prove (2.6), note that

$$\begin{aligned} E|E(X_i - Y_i|\mathcal{F}_{i-1})| \le & E|E(X_i I(|X_i| > i)|\mathcal{F}_{i-1})| \\ \le & E|X_i I(|X_i| > i)| = E|X_1 I(|X_1| > i)| \to 0 \end{aligned}$$

by the dominated convergence theorem. □

Definition. $\{\mathcal{F}_n, n \ge 1\}$ is a filtration. $H_n, n \ge 2$ is called a **predictable sequence** if

$$H_n \in \mathcal{F}_{n-1},\ n \ge 2.$$

Theorem 5.2.5. $(S_n, \mathcal{F}_n)$ is a supermartingale. $H_n, n \ge 2$ is predictable and $H_n \ge 0$. Moreover, suppose H_n are bounded. Then

$$T_1 := 0, T_n := \sum_{i=2}^{n} H_i(S_i - S_{i-1}),\ n \ge 2$$

is a supermartingale w.r.t. $\{\mathcal{F}_n\}$.
Proof is by

$$\begin{aligned} E(T_n|\mathcal{F}_{n-1}) = & E(T_{n-1} + H_n(S_n - S_{n-1})|\mathcal{F}_{n-1}) \\ = & T_{n-1} + H_n E(S_n - S_{n-1}|\mathcal{F}_{n-1}) \le T_{n-1}. \end{aligned}$$

□

$(S_n, \mathcal{F}_n)$ is a submartingale. $a < b$ are two constants. Define

$$N_1 = \inf\{m \ge 1 : S_m \le a\},$$

$$N_2 = \inf\{m > N_1 : S_m \ge b\},$$

$$N_3 = \inf\{m > N_2 : S_m \le a\},$$

$$N_4 = \inf\{m > N_3 : S_m \ge b\},$$

and so on. They are all stopping times w.r.t. $\{\mathcal{F}_n\}$. Let

$$U_n := \sup\{k : N_{2k} \le n\}$$

be the number of **upcrossings** by time n.

[Draw a picture to see why it is called upcrossings.]

Theorem 5.2.7 (Upcrossing inequality) $(S_n, \mathcal{F}_n)$ is a submartingale. $a < b$ are two constants. For U_n defined above, we have

$$E(U_n) \le \frac{1}{b-a}\Big[E(S_n - a)^+ - E(S_1 - a)^+\Big]$$

Proof For $m \geq 2$, let

$$H_m = \begin{cases} 1 & \text{if } N_{2k-1} < m \leq N_{2k} \\ 0 & \text{otherwise} \end{cases}$$

Draw a picture to see that

(1) H_m is predictable, i.e., $H_m \in \mathcal{F}_{m-1}$.

(2) $(b-a)U_n \leq \sum_{m=2}^n H_m(S_m - S_{m-1})$.

From (1) and (2) above, we have

$$\begin{aligned}(b-a)EU_n \leq& \sum_{m=2}^n E[H_m(S_m - S_{m-1})] \\ =& \sum_{m=2}^n E\{E[H_m(S_m - S_{m-1})|\mathcal{F}_{m-1}]\} \\ =& \sum_{m=2}^n E\{H_m E[(S_m - S_{m-1})|\mathcal{F}_{m-1}]\} \\ \leq& \sum_{m=2}^n E(S_m - S_{m-1}) \\ =& E(S_n - S_1).\end{aligned}$$

Let $S'_i = a + (S_i - a)^+$. From the martingale transform theorem, it is also a submartingale. Define U'_n be the number of upcrossing of $\{S'_i, i \leq n\}$ similarly as U_n. Then $U'_n = U_n$ (this is because S'_i only moves those points below a to the level a, and does not change the number of upcrossings from a to b). Therefore, from the result on the last page (applied to S'_n), we have

$$(b-a)EU_n = (b-a)EU'_n \leq E(S'_n - S'_1) = E(S_n - a)^- - E(S_1 - a)^+.$$

This proves the theorem. □

Now we are ready to prove Theorem 5.2.8, which we recall as

Suppose $(S_n, \mathcal{F}_n)$ is a submartingale and $\liminf E(S_n^+) < \infty$. Then S_n converges almost surely (a.s.) to a limit S with $E|S| < \infty$.

Proof of Theorem 5.2.8. For any two constants $a < b$, let U_n be the number of upcrossings as defined above. U_n is obviously increasing and suppose they increase to a limit U. Because $U_n \geq 0$, we have EU_n increases to EU. From the upcrossing inequality, we have

$$EU = \liminf EU_n \leq \liminf \frac{E(S_n - a)^+}{b-a} \leq \liminf \frac{|a| + ES_n^+}{b-a} < \infty.$$

This means U is finite almost surely and $P(\liminf S_n < a < b < \limsup S_n) = 0$ (because the event in the brackets implies infinite number of upcrossings.)

Therefore,

$$\begin{aligned} & P(\liminf S_n < \limsup S_n) \\ \leq & \sum_{\substack{a<b \\ a,b \text{ are rational numbers}}} P(\liminf S_n < a < b < \limsup S_n) = 0; \end{aligned}$$

and hence S_n has a limit, say S. Now we are left to show that $E|S| < \infty$. This follows from

(a) (Fatou) $ES^+ \leq \liminf ES_n^+ < \infty$ and

(b) $ES^- \leq \liminf ES_n^- = \liminf(ES_n^+ - ES_n) \leq \liminf(ES_n^+ - ES_1) < \infty$. □

Theorem 5.2.9. $(S_n, \mathcal{F}_n)$ is a supermartingale. If $S_n \geq 0, \forall n \geq 1$, then $S_n \to S$ a.s. and $ES < \infty$.

Proof. $-S_n \leq 0$ is a submartingale. $E((-S_n)^+) = 0$. By the martingale convergence theorem, we have $S_n \to S$ a.s. Moreover, by Fatou and the definition of supermartingales,

$$ES = E(\lim S_n) \leq \liminf ES_n \leq ES_1.$$

□

2.3 Doob's Inequality; L^p Convergence; CLT

In the following, we state and prove some results regarding the maximum of submartingales. These are generalization of the Kolmogorov's maximal inequality for sums of independent random variables.

Theorem 5.2.6. Let $\{\mathcal{F}_n, n \geq 1\}$ be a filtration. Recall that $N \geq 1$ is a **stopping time** if $\{N = n\} \in \mathcal{F}_n,\ \forall n \geq 1$. If $(S_n, \mathcal{F}_n)$ is a submartingale, then

$$T_n := S_{n \wedge N}$$

is a submartingale.

Proof. Let $X_1 = S_1$ and $X_i = S_i - S_{i-1},\ i \geq 2$. Rewrite

$$T_n = S_{n \wedge N} = \sum_{i=1}^{n \wedge N} X_i = \sum_{i=1}^{n} X_i I(N \geq i).$$

(i) and (ii) in the definition of submartingales follow easily from the above expression. To prove (iii), we have

$$\begin{aligned} E(T_n|\mathcal{F}_{n-1}) =& T_{n-1} + E(X_n I(N \geq n)|\mathcal{F}_{n-1}) \\ =& T_{n-1} + I(N \geq n) E(X_n|\mathcal{F}_{n-1}) \geq T_{n-1}. \end{aligned}$$

□

Note that under the above setting, we have

$$
\begin{aligned}
ET_n =& \sum_{i=1}^{n} EX_i I(N \geq i) \\
=& EX_1 + \sum_{i=2}^{n} E[E(X_i I(N \geq i)|\mathcal{F}_{i-1})] \\
=& EX_1 + \sum_{i=2}^{n} E[I(N \geq i)E(X_i|\mathcal{F}_{i-1})] \\
\leq& EX_1 + \sum_{i=2}^{n} E[E(X_i|\mathcal{F}_{i-1})] \\
=& EX_1 + \sum_{i=2}^{n} E(X_i) = ES_n.
\end{aligned}
\tag{2.7}
$$

Used below.

Theorem 5.4.2 (Doob's inequality). $(S_n, \mathcal{F}_n)$ is a submartingale. Then for any $x > 0$, we have

$$
P(\max_{1 \leq k \leq n} S_k \geq x) \leq \frac{1}{x} E\Big[S_n I(\max_{1 \leq k \leq n} S_k \geq x)\Big] \leq \frac{E(S_n^+)}{x}.
$$

Exercise: Show that Kolmogorov's maximal inequality follows from the above inequality.

Proof of Theorem 5.4.2. Let $N = \inf\{k : S_k \geq x\}$. N is a stopping time. Let

$$
A = \{\max_{1 \leq k \leq n} S_k \geq x\} = \{N \leq n\}.
$$

Then

$$
1_A \leq \frac{S_{n \wedge N}}{x} 1_A.
$$

We have, from (2.7),

$$
\begin{aligned}
P(A) =& E(1_A) \leq \frac{1}{x} E(S_{n \wedge N} 1_A) \\
=& \frac{1}{x}\Big[E(S_{n \wedge N}) - E(S_{n \wedge N}) 1_{A^c}\Big] \\
\leq& \frac{1}{x}\Big[E(S_n) - ES_n 1_{A^c}\Big] \\
=& \frac{1}{x} E(S_n 1_A).
\end{aligned}
$$

□

Theorem 5.4.3 (L^p maximum inequality). For a submartingale S_n and any constant $p > 1$, we have

$$E\Big[(\max_{1\le k\le n} S_k^+)^p\Big] \le \Big(\frac{p}{p-1}\Big)^p E\Big[(S_n^+)^p\Big].$$

Proof. Let $M > 0$ be a constant. We have

$$\begin{aligned} & E\Big[(\max_{1\le k\le n} S_k^+ \wedge M)^p\Big] \\ =& \int_0^\infty p\lambda^{p-1} P(\max_{1\le k\le n} S_k^+ \wedge M \ge \lambda) d\lambda \\ \le& \int_0^\infty p\lambda^{p-1} \frac{1}{\lambda} E\Big[S_n^+ I(\max_{1\le k\le n} S_k^+ \wedge M \ge \lambda)\Big] d\lambda, \end{aligned} \tag{2.8}$$

where the equality is from the alternative expression of moments of positive random variables and the last inequality is trivial for $M < \lambda$ and from Doob's inequality applied to the submartingale S_n^+ for $M \ge \lambda$.

From (2.8) and direct integration, we have

$$\begin{aligned} & E\Big[(\max_{1\le k\le n} S_k^+ \wedge M)^p\Big] \\ \le& \frac{p}{p-1} E S_n^+ (\max_{1\le k\le n} S_k^+ \wedge M)^{p-1} \\ \le& \frac{p}{p-1} (E(S_n^+)^p)^{1/p} (E(\max_{1\le k\le n} S_k^+ \wedge M)^p)^{1/q}, \end{aligned}$$

where $1/p + 1/q = 1$. The theorem is proved by moving the last factor on the right-hand side to the left. □

Theorem 5.4.5 (L^p convergence theorem). If (S_n) is a martingale and for some constant $p > 1$, $\sup_k E|S_k|^p < \infty$, then S_n converges almost surely to a limit S and $E|S_n - S|^p \to 0$.

Proof. The a.s. convergence follows from the martingale convergence theorem. We only need to show $E|S_n - S|^p \to 0$.

Because (S_n) is a martingale, we have $(|S_n|)$ is a submartingale. From the previous theorem, we have

$$E(\max_{1\le k\le n} |S_k|)^p \le \Big(\frac{p}{p-1}\Big)^p E(|S_n|^p).$$

This means

$$E(\sup_{k\ge 1} |S_k|)^p \le \Big(\frac{p}{p-1}\Big)^p \sup E(|S_n|^p),$$

which is finite by the assumption.

Because

$$|S_n - S|^p \le (2\sup_{k\ge 1} |S_k|)^p,$$

we have, by the dominated convergence theorem,

$$E|S_n - S|^p \to 0.$$

For more results on martingales, we refer to the book 'Martingale Limit Theory and its Applications' by P. Hall and C.C. Heyde (1981). Here we mention two useful results without proof.

Theorem (Martingale CLT). Let $(S_n, \mathcal{F}_n)$ be a zero-mean, square integrable martingale. Let $X_1 = S_1$ and $X_i = S_i - S_{i-1}$ for $i \geq 2$. Define

$$\sigma_n^2 = E(X_n^2 | \mathcal{F}_{n-1}), \quad B_n^2 = \sum_{i=1}^n E\sigma_i^2.$$

Suppose that $E|X_i|^4 < \infty$ for all $i \geq 1$. Then we have

$$\sup_{x\in\mathbb{R}} \Big| P(\frac{S_n}{B_n} \leq x) - \Phi(x) \Big| \leq C \Big\{ \frac{1}{B_n^4} \big[\sum_{i=1}^n E|X_i|^4 + E((\sum_{i=1}^n \sigma_i^2) - B_n^2)^2 \big] \Big\}^{1/5}$$

where $\Phi(\cdot)$ denotes the standard normal distribution function and C is a constant.

Theorem (Skorokhod representation). Suppose $(S_n, \mathcal{F}_n)$ is a zero-mean, square integrable martingale. Then, we can define it and a standard Brownian motion $\{B_t, t \geq 0\}$ and $\tau_1, \tau_2, \ldots, \geq 0$ on the same probability space such that, with $T_n = \tau_1 + \cdots + \tau_n$ and $S'_n = B_{T_n}, X'_n = S'_n - S'_{n-1}$, $\mathcal{G}_n = \sigma(S'_1, \ldots, S'_n, B_{[0,T_n]})$, we have

(i) $\{S_n, n \geq 1\}$ has the same joint distribution as $\{S'_n, n \geq 1\}$,

(ii) $E(\tau_n | \mathcal{G}_{n-1}) = E(X_n'^2 | \mathcal{G}_{n-1})$.