Probability course notes (Duke MATH641)

概率论笔记

Huarui ZHOU

2024 Spring

This note provides a detailed overview of the graduate course Probability (MATH641) instructed by Prof. Quanjun Lang. The course was remarkably interesting, covering a wide range of advanced probability theory topics, including martingale, Markov chain, ergodic theory, and Brownian motion. I primarily use this summary note for review purposes after each lecture. The content is mainly sourced from Durrett's book [2] and Prof. Lang's lectures. I've reorganized many proofs myself to ensure a thorough examination of each detail, though some steps may appear trivial. I also added a few theorems I read from other books (like [5]) or online lecture notes. Thank you for taking the time to read this note if you happen to find it online.

Contents

1 Conditional expectation **4**

1.1 Definition of conditional expectation 4

1.2 Property of conditional expectation 6

2 Martingale **13**

2.1 Definition of martingale . . . 13
2.2 Martingale convergence theorem . . . 17
2.3 Doob's inequality . . . 26
2.4 Uniform integrability and convergence in L^1 . . . 33
2.4.1 Definition and examples . . . 33
2.4.2 UI and convergence . . . 38
2.5 Backward martingale . . . 44
2.6 Optional stopping theorem . . . 46

3 Markov Chain 52

3.1 Construction of Markov chain . . . 52
3.2 Properties of Markov chain . . . 55
3.3 Basic concepts of Markov chain on a countable state space . . . 65
3.3.1 Multistep transition probability . . . 65
3.3.2 Time of the k-th return . . . 68
3.4 Exit distribution and exit time . . . 72
3.5 Recurrence and transience . . . 76
3.6 Recurrence of simple random walk . . . 84
3.7 Periodicity . . . 85
3.8 Stationary Measures . . . 89
3.9 Asymptotic behavior . . . 103

4 Branching process 111

4.1 Model description and basic properties . . . 111
4.2 Generating function . . . 112
4.3 Moments . . . 114
4.4 Extinction probability . . . 117

4.5 Kesten-Stigum Theorem 119

5 Ergodic theory **121**

5.1 Measure-preserving map 121

5.2 Stationary sequence 121

5.3 Ergodicity 126

5.4 Birkhoff's Ergodic Theorem 127

5.5 Recurrence 131

5.6 Subadditive ergodic theorem 132

6 Brownian motion **133**

6.1 Definition and simple properties 133

6.2 Construction 137

6.3 Markov property and Blumenthal's 0-1 law 147

6.4 Continuous stopping time 156

6.5 Strong Markov property 157

6.6 Path properties 158

6.6.1 Zero set 158

6.6.2 Hitting time and maximum 159

6.6.3 Arcsine laws 166

6.7 p-variation and quadratic variation 169

6.8 Martingale 176

References **180**

1 Conditional expectation

1.1 Definition of conditional expectation

Definition 1.1. Suppose $X : \Omega \to \mathbb{R}$ is a random variable (r.v.) on the probability space $(\Omega, \mathcal{F}_0, \mathbb{P})$ and X is integrable (i.e. $\mathbb{E}(|X|) < \infty$). $\mathcal{F} \subseteq \mathcal{F}_0$ is a sub-σ-field. We call r.v. $Y : \Omega \to \mathbb{R}$ the conditional expectation of X given $\mathcal{F}$ if it satisfies two conditions:

(1) Y is $\mathcal{F}$-measurable (or $Y \in \mathcal{F}$ for short).

(2) $\mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(Y\mathbb{1}_A)$ for any $A \in \mathcal{F}$.

We denote Y by $\mathbb{E}(X|\mathcal{F})$.

Theorem 1.2. *Given the conditions in the above definition, such r.v. Y exists and is unique (in the sense of "almost sure").*

Proof. **Uniqueness.** Let Y, Y' be two r.v. that satisfy conditions (1) and (2). Then $Y, Y' \in \mathcal{F}$ and for any $A \in \mathcal{F}$, we have

$$\mathbb{E}(Y\mathbb{1}_A) = \mathbb{E}(Y'\mathbb{1}_A) = \mathbb{E}(X\mathbb{1}_A).$$

Taking $A = \{\omega \in \Omega : Y - Y' \geq \varepsilon > 0\}$ for any $\varepsilon > 0$, then $A \in \mathcal{F}$ (because $Y - Y' \in \mathcal{F}$ and $(Y - Y')^{-1}([\varepsilon, \infty)) \in \mathcal{F}$), and

$$0 = \mathbb{E}(Y\mathbb{1}_A) - \mathbb{E}(Y'\mathbb{1}_A) = \mathbb{E}[(Y - Y')\mathbb{1}_A] \geq \mathbb{E}(\varepsilon\mathbb{1}_A) = \varepsilon\mathbb{P}(A),$$

hence $\mathbb{P}(A) = 0$ for any $\varepsilon > 0$, in other word, $\mathbb{P}(Y - Y' \leq 0) = 1$. Similarly, $\mathbb{P}(Y - Y' \geq 0) = 1$, then

$$\mathbb{P}(Y = Y') = \mathbb{P}(Y - Y' \leq 0) - \mathbb{P}(Y - Y' < 0) = \mathbb{P}(Y - Y' \leq 0) - \mathbb{P}(\{Y - Y' \geq 0\}^c) = 1.$$

Therefore $Y = Y'$ a.s.

Existence. Recall Radon-Nikodym theorem:

Theorem(Radon-Nikodym). Suppose μ and ν are both σ-finite measure on $(\Omega, \mathcal{F})$, if $\nu \ll \mu$ (for any $A \in \mathcal{F}$, $\mu(A) = 0 \Longrightarrow \nu(A) = 0$), there exists a $\mathcal{F}$-measurable function $f : \Omega \to \mathbb{R}$ (called the density of ν over μ), s.t. for any $A \in \mathcal{F}$,

$$\nu(A) = \int_A f \, \mathrm{d}\mu.$$

First suppose $X \geq 0$. Define

$$\nu(A) = \mathbb{E}(X \mathbb{1}_A) = \int_A X \, \mathrm{d}\mathbb{P}, \quad \forall A \in \mathcal{F}.$$

Easy to verify $\nu : \mathcal{F} \to [0, \infty)$ is a finite measure on $\mathcal{F}$ and $\nu \ll \mathbb{P}$. By R-N thm, we can find a $\mathcal{F}$-measurable function f, s.t.

$$\nu(A) = \mathbb{E}(f \mathbb{1}_A), \quad \forall A \in \mathcal{F}.$$

Now f is the conditional expectation of X given $\mathcal{F}$. For general r.v. X, let $X^+ = \max\{X, 0\}$, $X^- = \max\{-X, 0\}$, then $X^+, X_- \geq 0$ and $X = X^+ - X^-$. By previous result, there exist $f^+, f^- \in \mathcal{F}$ s.t.

$$\mathbb{E}(f^+ \mathbb{1}_A) = \mathbb{E}(X^+ \mathbb{1}_A), \quad \mathbb{E}(f^- \mathbb{1}_A) = \mathbb{E}(X^- \mathbb{1}_A), \quad \forall A \in \mathcal{F}.$$

Define $f = f^+ - f^- \in \mathcal{F}$,

$$\mathbb{E}(f \mathbb{1}_A) = \mathbb{E}(X \mathbb{1}_A) \quad \forall A \in \mathcal{F}.$$

□

Example 1.3. (1) If $X = c$ is a constant, then $\mathbb{E}(c|\mathcal{F}) = c$, because the constant function is measurable on any σ-field).

(2) If $\mathcal{F} = \mathcal{F}_0$, then

$$\mathbb{E}(X|\mathcal{F}) = X.$$

(3) If $\mathcal{F} = \{\varnothing, \Omega\}$, then

$$\mathbb{E}(X|\mathcal{F}) = \mathbb{E}(X).$$

(4) Let $\Omega_1, \Omega_2, \cdots$ be a partition of Ω with $\mathbb{P}(\omega_i) > 0$, $\mathcal{F} = \sigma(\Omega_i; i \geq 1)$. Then

$$\mathbb{E}(X|\mathcal{F}) = \sum_i \frac{\mathbb{E}(X\mathbb{1}_{\Omega_i})\mathbb{1}_{\Omega_i}}{\mathbb{P}(\Omega_i)}$$

1.2 Property of conditional expectation

Proposition 1.4. *Let $(\Omega, \mathcal{F}_0, \mathbb{P})$ be a probability space, $\mathcal{F} \subseteq \mathcal{F}_0$ is a sub-σ-field. Let X and Y be r.v. with $\mathbb{E}(|X|) < 0$ and $\mathbb{E}(|Y|) < 0$.*

(1) *For any $a, b \in \mathbb{R}$, $\mathbb{E}(aX + bY|\mathcal{F}) = a\mathbb{E}(X|\mathcal{F}) + b\mathbb{E}(Y|\mathcal{F})$.*

(2) *If $X \leq Y$, then $\mathbb{E}(X|\mathcal{F}) \leq \mathbb{E}(Y|\mathcal{F})$ a.s.*

(3) *If $X_n \geq 0$ and $X_n \uparrow X$, then*

$$\mathbb{E}(X_n|\mathcal{F}) \uparrow \mathbb{E}(X|\mathcal{F}).$$

(4) *If $\sigma(X)$ and $\mathcal{F}$ are independent, then $\mathbb{E}(X|\mathcal{F}) = \mathbb{E}(X)$.*

(5) *If $X \in \mathcal{F}$, then $\mathbb{E}(X|\mathcal{F}) = X$.*

Proof. (1) verify the definition: first, since linear combination of measurable function is also measurable, $a\mathbb{E}(X|\mathcal{F}) + b\mathbb{E}(Y|\mathcal{F}) \in \mathcal{F}$; second, for any $A \in \mathcal{F}$,

$$\begin{aligned}\mathbb{E}[(a\mathbb{E}(X|\mathcal{F}) + b\mathbb{E}(Y|\mathcal{F}))\mathbb{1}_A] &= a\mathbb{E}[\mathbb{E}(X|\mathcal{F})\mathbb{1}_A] + b\mathbb{E}[\mathbb{E}(Y|\mathcal{F})\mathbb{1}_A] \\ &= a\mathbb{E}(X\mathbb{1}_A) + b\mathbb{E}(Y\mathbb{1}_A) \\ &= \mathbb{E}[(aX + bY)\mathbb{1}_A],\end{aligned}$$

thus $\mathbb{E}(aX+bY|\mathcal{F}) = a\mathbb{E}(X|\mathcal{F}) + b\mathbb{E}(Y|\mathcal{F})$.

(2) for any $\varepsilon > 0$, define $A = \{\omega \in \Omega : \mathbb{E}(X|\mathcal{F})(\omega) - \mathbb{E}(Y|\mathcal{F})(\omega) \geq \varepsilon > 0\}$, then $A \in \mathcal{F}$ since $\mathbb{E}(X|\mathcal{F}) \in \mathcal{F}$ and $\mathbb{E}(Y|\mathcal{F}) \in \mathcal{F}$. By the definition and $X \leq Y$, we have

$$\mathbb{E}[\mathbb{E}(X|\mathcal{F})\mathbb{1}_A] = \mathbb{E}(X\mathbb{1}_A) \leq \mathbb{E}(Y\mathbb{1}_A) = \mathbb{E}[\mathbb{E}(Y|\mathcal{F})\mathbb{1}_A],$$

then

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

we conclude $\mathbb{P}(A) = 0$ by $\varepsilon > 0$ and $\mathbb{P}(A) \geq 0$. In other word, $\mathbb{P}(\mathbb{E}(X|\mathcal{F}) \leq \mathbb{E}(Y|\mathcal{F})) = 1$, i.e. $\mathbb{E}(X|\mathcal{F}) \leq \mathbb{E}(Y|\mathcal{F})$ a.s.

(3) By $X_n \uparrow$ and the result from (2), we have $\mathbb{E}(X_n|\mathcal{F})$ is also increasing. Moreover, X_n is bounded, leading to $\mathbb{E}(X_n|\mathcal{F})$ is also bounded for all n. By the bounded convergence theorem, the limit of $\mathbb{E}(X_n|\mathcal{F})$ exists, denoted as Z. For any $A \in \mathcal{F}$, by the definition,

$$\mathbb{E}[\mathbb{E}(X_n|\mathcal{F})\mathbb{1}_A] = \mathbb{E}(X_n\mathbb{1}_A).$$

By the monotone convergence theorem and $\mathbb{E}(X_n|\mathcal{F}) \uparrow Z$, $X_n \uparrow X$, we have

$$\mathbb{E}(Z\mathbb{1}_A) = \lim_{n\to\infty} \mathbb{E}(\mathbb{E}(X_n|\mathcal{F})\mathbb{1}_A) = \lim_{n\to\infty} \mathbb{E}(X_n\mathbb{1}_A) = \mathbb{E}(X\mathbb{1}_A),$$

and $Z \in \mathcal{F}$ because $\mathbb{E}(X_n|\mathcal{F}) \in \mathcal{F}$ and the limit of measurable functions is also measurable. Therefore, Z satisfies the definition of $\mathbb{E}(X|\mathcal{F})$, i.e. $Z = \mathbb{E}(X|\mathcal{F})$.

(4) First $\mathbb{E}(X)$ is a constant so it is measurable for any σ-field, of course for $\mathcal{F}$. Second, for any $A \in \mathcal{F}$,

$$\mathbb{E}[\mathbb{E}(X)\mathbb{1}_A] = \mathbb{E}(X)\mathbb{E}(\mathbb{1}_A) = \mathbb{E}(X)\mathbb{P}(A),$$

and by independence,

$$\mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(X)\mathbb{E}(\mathbb{1}_A) = \mathbb{E}(X)\mathbb{P}(A).$$

(5) Obviously. □

Proposition 1.5 (Jensen's inequality)**.** *Let φ be a convex function on $\mathbb{R}$, and X be a r.v. with $\mathbb{E}(|X|) < \infty$ and $\mathbb{E}(|\varphi(X)|) < \infty$. Then*

$$\varphi(\mathbb{E}(X|\mathcal{F})) \leq \mathbb{E}(\varphi(X)|\mathcal{F}).$$

Proof. The proof will be much easier if we use the following property of convex function:

Theorem. Any convex function can be written as the supremum of some affine functions.[a]

[a] See https://proofwiki.org/wiki/Convex_Real_Function_is_Pointwise_Supremum_of_Affine_Functions

Let $S = \{(a,b) \in \mathbb{Q} \times \mathbb{Q} : ax + b \leq \varphi(x)\}$, from the above theorem, $\varphi(x) = \sup_{(a,b)\in S}(ax + b)$. For a fixed $(a,b) \in S$,

$$aX + b \leq \varphi(X),$$

by Proposition 1.4,

$$a\mathbb{E}(X|\mathcal{F}) + b \leq \mathbb{E}(\varphi(X)|\mathcal{F}), \quad \text{a.s.}$$

define

$$A_{(a,b)} := \{\omega \in \Omega : a\mathbb{E}(X|\mathcal{F}) + b > \mathbb{E}(\varphi(X)|\mathcal{F})\},$$

thus $A_{(a,b)}$ is a null set. Since the countable union of null sets is also a null set (Note: the uncountable union of null sets can generate an un-null set, that is why we make S countable!), we have

$$\begin{aligned}&\mathbb{P}[\omega \in \Omega : \sup_{(a,b)\in S}(a\mathbb{E}(X|\mathcal{F})+b)(\omega) > \mathbb{E}(\varphi(X)|\mathcal{F})(\omega)]\\ &= \mathbb{P}[\bigcup_{(a,b)\in S}\{\omega\in\Omega : a\mathbb{E}(X|\mathcal{F})(\omega)+b > \mathbb{E}(\varphi(X)|\mathcal{F})(\omega)\}]\\ &= \mathbb{P}(\bigcup_{(a,b)\in S} A_{(a,b)})\\ &= 0\end{aligned}$$

i.e.

$$\varphi(\mathbb{E}(X|\mathcal{F})) = \sup_{(a,b)\in S}(a\mathbb{E}(X|\mathcal{F})+b) \le \mathbb{E}(\varphi(X)|\mathcal{F}), \quad \text{a.s.}$$

□

Proposition 1.6 (Contraction in L^p)**.** *For any $p \ge 1$, we have*

$$\mathbb{E}(|X|^p|\mathcal{F}) \ge |\mathbb{E}(X|\mathcal{F})|^p.$$

Proposition 1.7 ("Fine enough")**.** *Let $\mathcal{F} \subseteq \mathcal{G}$ be two sub-σ-fields, and $\mathbb{E}(X|\mathcal{G}) \in \mathcal{F}$, then*

$$\mathbb{E}(X|\mathcal{F}) = \mathbb{E}(X|\mathcal{G}).$$

Proof. Since $\mathbb{E}(X|\mathcal{G}) \in \mathcal{F}$, for proving the equality, we only need to prove for any $A \in \mathcal{F}$,

$$\mathbb{E}[\mathbb{E}(X|\mathcal{G})\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A],$$

this is true from the definition of $\mathbb{E}(X|\mathcal{G})$, and above $A \in \mathcal{F} \subseteq \mathcal{G}$. □

Proposition 1.8 ("The smaller σ-field wins")**.** *Let $\mathcal{F} \subseteq \mathcal{G}$ be two sub-σ-field, then*

(1) $\mathbb{E}[\mathbb{E}(X|\mathcal{F})|\mathcal{G}] = \mathbb{E}(X|\mathcal{F})$

(2) $\mathbb{E}[\mathbb{E}(X|\mathcal{G})|\mathcal{F}] = \mathbb{E}(X|\mathcal{F})$

Proof. (1) First $\mathbb{E}(X|\mathcal{F}) \in \mathcal{G}$ because $\mathbb{E}(X|\mathcal{F}) \in \mathcal{F}$ and $\mathcal{F} \subseteq \mathcal{G}$. Second, for any $A \in \mathcal{G}$,

$$\mathbb{E}[\mathbb{E}(X|\mathcal{F})\mathbb{1}_A] = \mathbb{E}[\mathbb{E}(X|\mathcal{F})\mathbb{1}_A].$$

(2)First $\mathbb{E}(X|\mathcal{F}) \in \mathcal{F}$ by definition. Second, for any $A \in \mathcal{F} \subseteq \mathcal{G}$, by the definition of $\mathbb{E}(X|\mathcal{G})$ and $\mathbb{E}(X|\mathcal{F})$,

$$\mathbb{E}[\mathbb{E}(X|\mathcal{G})\mathbb{1}_A] = \mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[\mathbb{E}(X|\mathcal{F})\mathbb{1}_A]. \qquad \square$$

Corollary 1.9 (Law of total expectation)**.**

$$\mathbb{E}[\mathbb{E}(X|\mathcal{F})] = \mathbb{E}(X)$$

Proof. take $\mathcal{G} = \{\varnothing, \Omega\}$, obviously $\mathcal{G} \subseteq \mathcal{F}$, thus from Proposition 1.8,

$$\mathbb{E}[\mathbb{E}(X|\mathcal{F})] = \mathbb{E}[\mathbb{E}(X|\mathcal{F})|\mathcal{G}] = \mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X). \qquad \square$$

Proposition 1.10 ("Taking out what is known")**.** *Let* $X \in \mathcal{F}$*,* $\mathbb{E}(|Y|) < \infty$*,* $\mathbb{E}(|XY|) < \infty$*, then*

$$\mathbb{E}(XY|\mathcal{F}) = X\mathbb{E}(Y|\mathcal{F}).$$

Proof. It is obvious that $X\mathbb{E}(Y|\mathcal{F}) \in \mathcal{F}$, so we only need to prove (2) in the definition, i.e. for any $A \in \mathcal{F}$,

$$\mathbb{E}[X\mathbb{E}(Y|\mathcal{F})\mathbb{1}_A] = \mathbb{E}(XY\mathbb{1}_A). \tag{$*$}$$

We can prove it by performing the 4-step procedure.

1. Indicator. Suppose $X = \mathbb{1}_E \in \mathcal{F}$ with $E \in \mathcal{F}$, then

$$\mathbb{E}[\mathbb{1}_E\mathbb{E}(Y|\mathcal{F})\mathbb{1}_A] = \mathbb{E}[\mathbb{E}(Y|\mathcal{F})\mathbb{1}_{A\cap E}] = \mathbb{E}(Y\mathbb{1}_{A\cap E}) = \mathbb{E}(\mathbb{1}_E Y\mathbb{1}_A),$$

thus $(*)$ holds.

2. Simple function. Suppose $X = \sum_i a_i \mathbb{1}_{E_i}$ with $E_i \in \mathcal{F}$, then $(*)$ still holds by linearity.

3. Non-negative function. Suppose $X, Y \geq 0$. We can construct a series of simple functions X_n s.t. $X_n \uparrow X$. Since $Y \geq 0, \mathbb{E}(Y|\mathcal{F}) \geq 0$, we have $X_n\mathbb{E}(Y|\mathcal{F}) \uparrow X\mathbb{E}(Y|\mathcal{F})$ and $X_nY \uparrow XY$, by the monotone convergence theorem,

$$\mathbb{E}[X_n\mathbb{E}(Y|\mathcal{F})\mathbb{1}_A] \to \mathbb{E}[X\mathbb{E}(Y|\mathcal{F})\mathbb{1}_A], \quad \mathbb{E}(X_nY\mathbb{1}_A) \to \mathbb{E}(XY\mathbb{1}_A),$$

Hence $\mathbb{E}[X\mathbb{E}(Y|\mathcal{F})\mathbb{1}_A] = \mathbb{E}(XY\mathbb{1}_A)$.

4. General case. Let $X = X^+ - X^-$ and $Y = Y^+ - Y^-$, then

$$\begin{aligned}
&\mathbb{E}[X\mathbb{E}(Y|\mathcal{F})\mathbb{1}_A] \\
=&\mathbb{E}[(X^+ - X^-)\mathbb{E}(Y^+ - Y^-|\mathcal{F})\mathbb{1}_A] \\
=&\mathbb{E}[X^+\mathbb{E}(Y^+|\mathcal{F})\mathbb{1}_A] + \mathbb{E}[X^-\mathbb{E}(Y^-|\mathcal{F})\mathbb{1}_A] - \mathbb{E}[X^+\mathbb{E}(Y^-|\mathcal{F})\mathbb{1}_A] - \mathbb{E}[X^-\mathbb{E}(Y^+|\mathcal{F})\mathbb{1}_A] \\
=&\mathbb{E}(X^+Y^+\mathbb{1}_A) + \mathbb{E}(X^-Y^-\mathbb{1}_A) - \mathbb{E}(X^-Y^+\mathbb{1}_A) - \mathbb{E}(X^+Y^-\mathbb{1}_A) \\
=&\mathbb{E}[(X^+ - X^-)(Y^+ - Y^-)\mathbb{1}_A] = \mathbb{E}(XY\mathbb{1}_A).
\end{aligned}$$

□

Proposition 1.11 (Conditional Expectation as projections in L^2). *Let X be a r.v. with $\mathbb{E}(X^2) < \infty$, i.e. $X \in L^2(\mathcal{F}_0)$. And for any $Y \in \mathcal{F}$ with $\mathbb{E}(Y^2) < \infty$, i.e. $Y \in L^2(\mathcal{F})$, we have*

$$\mathbb{E}[(X - Y)^2] \geq \mathbb{E}[(X - \mathbb{E}(X|\mathcal{F}))^2],$$

the equality holds if and only if $Y = \mathbb{E}(X|\mathcal{F})$.

Proof. 1. First we have $\mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)} < \infty$, then by Proposition 1.10,

$$Y\mathbb{E}(X|\mathcal{F}) = \mathbb{E}(YX|\mathcal{F}).$$

Taking the expectation, we have

$$\mathbb{E}[Y\mathbb{E}(X|\mathcal{F})] = \mathbb{E}[\mathbb{E}(YX|\mathcal{F})] = \mathbb{E}(YX),$$

i.e.

$$\mathbb{E}[Y(X - \mathbb{E}(X|\mathcal{F}))] = 0, \quad \forall Y \in L^2(\mathcal{F})$$

this means any $Y \in L^2(\mathcal{F})$ is perpendicular to $X - \mathbb{E}(X|\mathcal{F})$.

2. By the Jensen's inequality (Proposition 1.5), $[\mathbb{E}(X|\mathcal{F})]^2 \leq \mathbb{E}(X^2|\mathcal{F})$, thus

$$\mathbb{E}[[\mathbb{E}(X|\mathcal{F})]^2] \leq \mathbb{E}[\mathbb{E}(X^2|\mathcal{F})] = \mathbb{E}(X^2) < \infty,$$

i.e. $\mathbb{E}(X|\mathcal{F}) \in L^2(\mathcal{F})$.

3. Let $Z = \mathbb{E}(X|\mathcal{F}) - Y \in L^2(\mathcal{F})$ (since both Y and $\mathbb{E}(X|\mathcal{F})$ are in the $L^2(\mathcal{F})$), we have

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

The equality holds if and only if $Z = 0$ i.e. $Y = \mathbb{E}(X|\mathcal{F})$. □

2 Martingale

2.1 Definition of martingale

Definition 2.1. Suppose $\{\mathcal{F}_n : n \geq 0\}$ is a sequence of σ-fields on Ω, $\{X_n : n \geq 0\}$ is a sequence of r.v. on Ω.

- We call $\{\mathcal{F}_n\}$ a **filtration** if

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \cdots$$

- We say $\{X_n\}$ is **adapted** to $\{\mathcal{F}_n\}$ if X_n is $\mathcal{F}_n$-measurable ($X_n \in \mathcal{F}_n$) for all $n \geq 0$.

- We call $\{X_n\}$ a **martingale** w.r.t. $\{\mathcal{F}_n\}$ if

  (1) $\mathbb{E}(|X_n|) < \infty$

  (2) $\{X_n\}$ is adapted to $\{\mathcal{F}_n\}$

  (3) $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$ for all $n \geq 0$.

  $\{X_n\}$ is called a submartingale if the equality in (3) is replaced by $\geq$, or a supermartingale if replaced by $\leq$.

Proposition 2.2 (easy property)**.** *Suppose $\{X_n\}$ is a martingale w.r.t. $\{\mathcal{F}_n\}$, then*

(1) *for any $a \in \mathbb{R}$, $\{X_n + a\}$ is also a martingale.*

(2) *for any $n \geq 0$, $\mathbb{E}(X_{n+1} - X_n|\mathcal{F}_n) = 0$.*

(3) *for any $n \geq 1$, $\mathbb{E}(X_0) = \mathbb{E}(X_n)$.*

The original meaning of the martingale is a set of strings on the horse neck to control its head up or down.

Example 2.3. Let $\{X_n : n \geq 1\}$ be i.i.d. r.v. $S_n = S_0 + X_1 + \cdots + X_n$, where S_0 is a constant. $\mathcal{F}_0 = \{\varnothing, \Omega\}$ and $\mathcal{F}_n = \sigma(X_1, X_2, \cdots, X_n)$.

Figure 1: Martingale (the purple string)

(1) If $\mathbb{E}(X_i) = 0$ for any $i \geq 1$,then $\{S_n, n \geq 0\}$ is a martingale.

Obviously $\mathbb{E}(S_n) = S_0 < \infty$ and $S_n \in \mathcal{F}_n$. For the third requirement,

$$\mathbb{E}(S_{n+1}|\mathcal{F}_n) = \mathbb{E}(S_n + X_{n+1}|\mathcal{F}_n) = \mathbb{E}(S_n|\mathcal{F}_n) + \mathbb{E}(X_{n+1}|\mathcal{F}_n) = S_n + \mathbb{E}(X_{n+1}) = S_n,$$

where we used the fact that $S_n \in \mathcal{F}_n$ and $\sigma(X_{n+1})$ is independent of $\mathcal{F}_n$.

(2) If $\mathbb{E}(X_i) = 0$ and $\sigma^2 = \mathbb{E}(X_i^2) < \infty$ for any $i \geq 1$, then $\{S_n^2 - n\sigma^2 : n \geq 0\}$ is a martingale.

First we have $\mathbb{E}(S_n^2 - n\sigma^2) = \mathbb{E}(S_n^2) - n\sigma^2 = n\mathbb{E}(X_i^2) + S_0^2 - n\sigma^2 = S_0^2 < \infty$ and $S_n^2 - n\sigma \in \mathcal{F}_n$.

Moreover,

$$\begin{aligned}
\mathbb{E}[S_{n+1}^2 - (n+1)\sigma^2|\mathcal{F}_n] &= \mathbb{E}[S_{n+1}^2|\mathcal{F}_n] - (n+1)\sigma^2 \\
&= \mathbb{E}[(S_n + X_{n+1})^2|\mathcal{F}_n] - (n+1)\sigma^2 \\
&= \mathbb{E}[S_n^2 + 2S_nX_{n+1} + X_{n+1}^2|\mathcal{F}_n] - (n+1)\sigma^2 \\
&= \mathbb{E}(S_n^2|\mathcal{F}_n) + 2\mathbb{E}(S_nX_{n+1}|\mathcal{F}_n) + \mathbb{E}(X_{n+1}^2|\mathcal{F}_n) - (n+1)\sigma^2 \\
&= S_n^2 + 2S_n\mathbb{E}(X_{n+1}|\mathcal{F}_n) + \mathbb{E}(X_{n+1}^2) - (n+1)\sigma^2 \\
&= S_n^2 + 2S_n\mathbb{E}(X_{n+1}) + \mathbb{E}(X_{n+1}^2) - (n+1)\sigma^2 \\
&= S_n^2 - n\sigma^2.
\end{aligned}$$

Example 2.4. Let $X \in L^1(\mathcal{F})$, define

$$M_n = \mathbb{E}(X|\mathcal{F}_n),$$

then $\{M_n\}$ is a martingale.

Proof. Obviously $M_n \in L^1(\mathcal{F}_n)$, and

$$\mathbb{E}(M_{n+1}|\mathcal{F}_n) = \mathbb{E}[\mathbb{E}(X|\mathcal{F}_{n+1})|\mathcal{F}_n] = \mathbb{E}(X|\mathcal{F}_n) = M_n.$$ □

Proposition 2.5. (1) *Suppose $\{X_n : n \geq 0\}$ is a supermartingale w.r.t. $\{\mathcal{F}_n\}$, then for any $n > m$,*

$$\mathbb{E}(X_n|\mathcal{F}_m) \leq X_m.$$

(2) *Suppose $\{X_n : n \geq 0\}$ is a submartingale w.r.t. $\{\mathcal{F}_n\}$, then for any $n > m$,*

$$\mathbb{E}(X_n|\mathcal{F}_m) \geq X_m.$$

(3) *Suppose* $\{X_n : n \geq 0\}$ *is a martingale w.r.t.* $\{\mathcal{F}_n\}$*, then for any* $n > m$*,*

$$\mathbb{E}(X_n|\mathcal{F}_m) = X_m.$$

Proof. (1) Fix $m \geq 0$, by definition, $\mathbb{E}(X_{m+1}|\mathcal{F}_m) \leq X_m$. Now suppose $\mathbb{E}(X_{m+k}|\mathcal{F}_m) \leq X_m$ for some $k \geq 1$. Then for $k+1$ we have

$$\mathbb{E}(X_{m+k+1}|\mathcal{F}_m) = \mathbb{E}[\mathbb{E}(X_{m+k+1}|\mathcal{F}_{m+k})|\mathcal{F}_m] \leq \mathbb{E}[X_{m+k}|\mathcal{F}_m] \leq X_m,$$

the first "=" is due to "The smaller wins", the following "≤" is by the definition and induction hypothesis. Hence by induction, $\mathbb{E}(X_n|\mathcal{F}_m) \leq X_m$ for all $n > m$.

(2) Notice that $\{-X_n\}$ is supermartingale.

(3) Using the fact that martingale is both supermartingale and submartingale. □

Proposition 2.6. (1) *Suppose* $\{X_n\}$ *is a martingale w.r.t.* $\mathcal{F}_n$*,* $\varphi : \mathbb{R} \to \mathbb{R}$ *is a convex function with* $\mathbb{E}(|\varphi(X_n)|) < \infty$ *for all* $n \geq 0$*. Then* $\{\varphi(X_n)\}$ *is a submartingale w.r.t.* $\mathcal{F}_n$*.*

(2) *Suppose* $\{X_n\}$ *is a submartingale w.r.t.* $\mathcal{F}_n$*,* $\varphi : \mathbb{R} \to \mathbb{R}$ *is an increasing convex function with* $\mathbb{E}(|\varphi(X_n)|) < \infty$ *for all* $n \geq 0$*. Then* $\{\varphi(X_n)\}$ *is a submartingale w.r.t.* $\mathcal{F}_n$*.*

(3) *If* $\{X_n\}$ *is a submartingale,* $a \in \mathbb{R}$*, then* $\{(X_n - a)^+\}$ *is a submartingale.*

(4) *If* $\{X_n\}$ *is a supermartingale,* $a \in \mathbb{R}$*, then* $\{X_n \wedge a\}$ *is a supermartingale.*

Proof. (1)First, since φ is convex, then it is measurable, thus $\varphi \circ X_n \in \mathcal{F}_n$. Second, by Jensen's inequality,

$$\mathbb{E}[\varphi(X_{n+1})|\mathcal{F}_n] \geq \varphi(\mathbb{E}(X_{n+1}|\mathcal{F}_n)) = \varphi(X_n).$$

(2)Submartingale means $\mathbb{E}(X_{n+1}|\mathcal{F}) \geq X_n$. Since φ is increasing, we have

$$\mathbb{E}[\varphi(X_{n+1})|\mathcal{F}_n] \geq \varphi(\mathbb{E}(X_{n+1}|\mathcal{F}_n)) \geq \varphi(X_n).$$

(3)Because $\varphi(x) = (x-a)^+ = \max\{0, x-a\}$ is increasing and convex (See Figure 2).

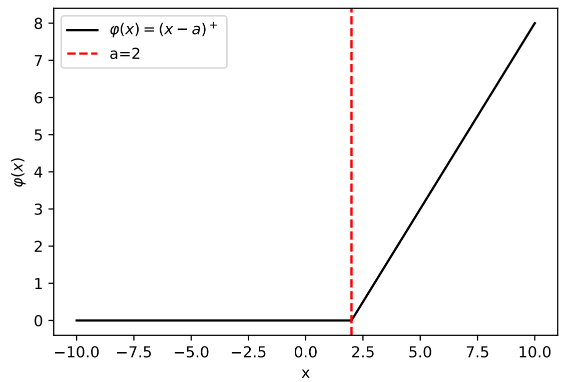


Figure 2: Plot of $\varphi(x) = (x-a)^+$

(4) Since

$$X_n \wedge a = \min\{X_n, a\} = \min\{X_n - a, 0\} + a = -\max\{-X_n + a, 0\} + a = -(-X_n + a)^+ + a,$$

where $\{-X_n\}$ is a submartingale. Then apply (3), $(-X_n + a)^+$ is a submartingale, thus $-(-X_n + a)^+ + a$ is a supermartingale. □

2.2 Martingale convergence theorem

We will prove the Martingale convergence theorem in this section.

Definition 2.7. Let $\{\mathcal{F}_n : n \geq 0\}$ be a filtration, r.v. the sequence $\{H_n : n \geq 1\}$ is said to be predictable if $H_n \in \mathcal{F}_{n-1}$ for all $n \geq 1$.

Consider a model of the stock market. Let X_n $(n \geq 0)$ be the value of one stock at time

n, and H_n be the total number of shares we hold between time $n-1$ and time n.[1] Then our total profit[2] from the stock market at time n ($n \geq 1$) is

$$(H \cdot X)_n = \sum_{m=1}^{n} H_m(X_m - X_{m-1}),$$

and define $(H \cdot X)_0 = 0$.

Example 2.8. Let $\{X_n = X_0 + \xi_1 + \cdots + \xi_n : n \geq 0\}$ be a random walk starting from $X_0 = 3$ with $\mathbb{P}(\xi_i = 1) = \mathbb{P}(\xi_i = -1) = 0.5$. Let $H_0 = 0$, for $n \geq 1$, define H_n as

$$H_n = \begin{cases} H_{n-1} + 1, & X_n \geq 3 \\ (H_{n-1} - 1)^+. & X_n < 3 \end{cases}$$

Figure 3 shows the simulation of this model.

Proposition 2.9 ("No profit for unfair game on average")**.** *Suppose* $\{X_n : n \geq 0\}$ *is a supermartingale,* $\{H_n : n \geq 1\}$ *is a predictable sequence with* $0 \leq H_n < \infty$*. Then* $(H \cdot X)_n$ *is a supermartingale.*

(This conclusion remains true if we replace all "supermartingale" with "submartingale" or "martingale".)

Proof. For $n \geq 0$,

$$\begin{aligned} \mathbb{E}((H \cdot X)_{n+1} | \mathcal{F}_n) &= \mathbb{E}[(H \cdot X)_n + H_{n+1}(X_{n+1} - X_n) | \mathcal{F}_n] \\ &= (H \cdot X)_n + H_{n+1}\mathbb{E}(X_{n+1} - X_n | \mathcal{F}_n) \\ &\leq (H \cdot X)_n, \end{aligned}$$

[1] We will buy or sell shares depending on the stock value at time $n-1$, then hold them until we know the updated value at time n, i.e. the update of H is always after the update of X, that is why $H_n \in \mathcal{F}_{n-1}$.

[2] To simplify the model, suppose we can get the shares without paying. So the profit is only affected by the fluctuation of the stock value and number of shares we hold.

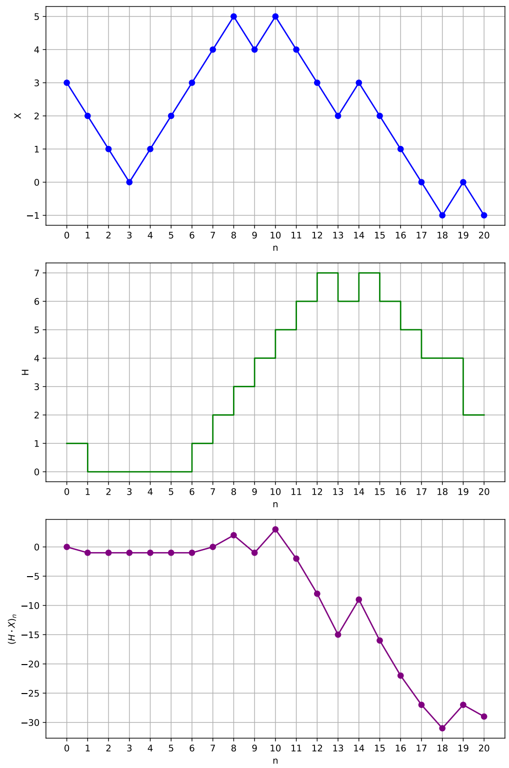


Figure 3: Simulation of the stock value, shares and profit

the last "≤" holds because $H_{n+1} \geq 0$ and $\mathbb{E}(X_{n+1} - X_n|\mathcal{F}_n) \leq 0$ for supermartingale. □

Remark. We immediately have $\mathbb{E}[(H\cdot X)_n] \leq \mathbb{E}[(H\cdot X)_0] = 0$ by the property of supermartingale, which means there is no profit on average for the supermartingale (unfair game).

Definition 2.10. We call r.v. N a stopping time, if for any $n \geq 0$,

$$\{N = n\} \in \mathcal{F}_n.$$

Proposition 2.11. *Suppose N is a stopping time, then for any $m \geq 0$,*

(1) $\{N < m+1\} = \{N \leq m\} \in \mathcal{F}_m$

(2) $\{N > m\} = \{N \geq m+1\} \in \mathcal{F}_m$

Proposition 2.12. *Suppose N is a stopping time, $\{X_n\}$ is a supermartingale, then $\{X_{N\wedge n}\}$ is a supermartingale.*

Proof. 1.For any $n \geq 1$, define $H_n(\omega) = \mathbb{1}_{\{N(\omega)\geq n\}}(\omega)$, then $\{H_n\}$ is predictable.
We only need to show $H_n \in \mathcal{F}_{n-1}$. This is true because

$$\{H_n = 1\} = \{N \geq n\} \in \mathcal{F}_{n-1}.$$

2. Show $(H \cdot X)_n = X_{N\wedge n} - X_0$.

$$\begin{aligned}(H \cdot X)_n &= \sum_{m=1}^{n} H_m(X_m - X_{m-1}) \\ &= \sum_{m=1}^{n} \mathbb{1}_{\{N\geq m\}}(X_m - X_{m-1}) \\ &= \sum_{m=1}^{N\wedge n}(X_m - X_{m-1}) \\ &= X_{N\wedge n} - X_0.\end{aligned}$$

3. Finally, applying Proposition 2.9, we have $X_{N\wedge n} = (H \cdot X)_n + X_0$ is a supermartingale. □

Next, we will prove the Martingale convergence theorem by constructing the "Crossing" model. Suppose $\{X_n : n \geq 0\}$ is a submartingale, and $a, b \in \mathbb{R}$ with $a < b$. Define $N_1 = \inf\{m : m \geq 0, X_m \leq a\}$, $N_2 = \inf\{m : m > N_1, X_m \geq b\}$, and for $k \geq 2$,

$$N_{2k-1} = \inf\{m : m > N_{2k-2}, X_m \leq a\}, \quad N_{2k} = \inf\{m : m > N_{2k-1}, X_m \geq b\},$$

in other word, N_{2k-1} is the the first time after N_{2k-2} that $X_m \leq a$ happens, N_{2k} is the the first time after N_{2k-1} that $X_m \geq b$ happens. During the time between N_{2k-1} and N_{2k}, X_m is upcrossing the interval $[a, b]$. Define $U_n = \sup\{k : N_{2k} \leq n\}$ is the total number of upcrossings by time n (See Figure 4).

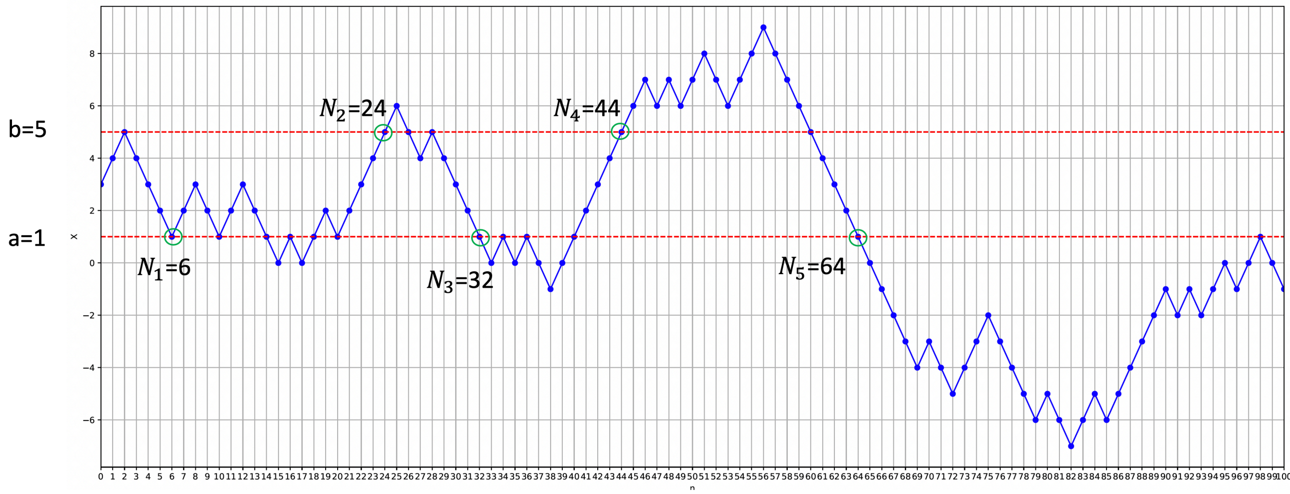


Figure 4: Example of upcrossings. In this case there are two upcrossings by time 100, i.e. $U_{100} = 2$.

Lemma 2.13 (Upcrossing inequality)**.** *Suppose* $\{X_m : m \geq 0\}$ *is a submartingale, then*

$$\mathbb{E}(U_n) \leq \frac{\mathbb{E}[(X_n - a)^+] - E[(X_0 - a)^+]}{b - a}.$$

Proof. 1. N_j are stopping time.

For $n \geq 0$,

$$\{N_1 = n\} = \{X_n \leq a\} \in \mathcal{F}_n,$$

$$\{N_2 = n\} = \{n > N_1, X_n \geq b\} = \{N_1 \leq n - 1\} \cap \{X_n \geq b\} \in \mathcal{F}_n,$$

then this claim is proved by induction.

2. For $m \geq 1$, define

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

then H_m is predictable[3].

[3] Actually, H_m is the strategy how we hold the shares: if the stock value is upcrossing the interval $[a, b]$, we always keep one (we only consider one or zero share for simplicity) share, otherwise, we sell all of them.

We need to show $H_m \in \mathcal{F}_{m-1}$. Notice

$$\{H_m = 1\} = \bigcup_{k=1}^{\infty} \{N_{2k-1} < m \le N_{2k}\} = \bigcup_{k=1}^{\infty} \{N_{2k-1} < m\} \cap \{N_{2k} \ge m\} \in \mathcal{F}_{m-1}$$

3.Define $Y_m = X_m \vee a = a + (X_m - a)^+$, then by Proposition 2.6, Y_m is a submartingale.

4.Claim: for all $n \ge 1$, $(b-a)U_n \le (H \cdot Y)_n$.

For $k \ge 1$ and $N_{2k} < \infty$,

$$(H \cdot Y)_{N_{2k}} = \sum_{i=1}^{k} \sum_{j=N_{2i-1}+1}^{N_{2i}} (Y_j - Y_{j-1}) = \sum_{i=1}^{k} (Y_{N_{2i}} - Y_{N_{2i-1}}) \ge k(b-a).$$

If $n \in \{N_{2k}, \cdots, N_{2k+1}\}$ (during the end of the k'th upcrossing to the beginning of the next upcrossing),

$$(H \cdot Y)_n = (H \cdot Y)_{N_{2k}} \ge k(b-a);$$

If $n \in \{N_{2k-1}+1, \cdots, N_{2k}\}$ (in the middle of the incomplete k'th upcrossing),

$$\begin{aligned}
(H \cdot Y)_n &= (H \cdot Y)_{N_{2k-1}} + \sum_{m=N_{2k-1}+1}^{n} (Y_m - Y_{m-1}) \\
&= (H \cdot Y)_{N_{2k-1}} + Y_n - Y_{N_{2k-1}} \\
&\ge (H \cdot Y)_{N_{2k-1}} \qquad \text{Since } Y_{N_{2k-1}} = a, \text{ and } Y_n \ge a \\
&= (H \cdot Y)_{N_{2k-2}} \\
&\ge (k-1)(b-a)
\end{aligned}$$

Above we have iterated all cases, hence $(H \cdot Y)_n \ge U_n(b-a)$ for all $n \ge 1$.

5. Define $K_m = 1 - H_m$, then K_m is predictable, and Y_m is a submartingale by claim 3, thus

both $(H \cdot Y)_n$ and $(K \cdot Y)_n$ are submartingales by Proposition 2.9. Then

$$\begin{aligned} Y_n - Y_0 &= \sum_{i=1}^{n}(Y_i - Y_{i-1}) = \sum_{i=1}^{n}(Y_i - Y_{i-1})\mathbb{1}_{i:\text{upcrossing}} + \sum_{i=1}^{n}(Y_i - Y_{i-1})\mathbb{1}_{i:\text{non-upcrossing}} \\ &= (H \cdot Y)_n + (K \cdot Y)_n. \end{aligned}$$

Therefore,

$$\mathbb{E}[(X_n - a)^+] - E[(X_0 - a)^+] = \mathbb{E}(Y_n - Y_0) = \mathbb{E}[(H \cdot Y)_n] + \mathbb{E}[(K \cdot Y)_n] \geq \mathbb{E}[(H \cdot Y)_n] = (b-a)\mathbb{E}(U_n),$$

where $\mathbb{E}[(K \cdot Y)_n] = \mathbb{E}[\mathbb{E}[(K \cdot Y)_n|\mathcal{F}_n]] \geq \mathbb{E}[(K \cdot Y)_0] = 0.$ □

Theorem 2.14 (Martingale convergence theorem)**.** *Suppose* $\{X_n : n \geq 0\}$ *is a submartingale with* $\sup \mathbb{E}(X_n^+) < \infty$*, then there exists a r.v.* X *with* $\mathbb{E}(|X|) < \infty$ *s.t.* $X_n \to X$ *a.s.*

Proof. 1. $\{U_n\}$ is increasing.

$\{k : N_{2k} \leq n\} \subseteq \{k : N_{2k} \leq n+1\}$, and

$$U_n = \sup\{k : N_{2k} \leq n\} \leq \sup\{k : N_{2k} \leq n+1\} = U_{n+1}.$$

Let $U_n \uparrow U$.

2. $\mathbb{E}(U_n)$ is uniformly bounded.

By Lemma 2.13, and $(X_n - a)^+ \leq X_n^+ + |a|$, $\mathbb{E}[(X_0 - a)^+] \geq 0$, we have

$$\mathbb{E}(U_n) \leq \frac{\mathbb{E}[(X_n - a)^+] - E[(X_0 - a)^+]}{b-a} \leq \frac{\mathbb{E}(X_n^+) + |a|}{b-a} \leq \frac{M + |a|}{b-a} < \infty,$$

where $M = \sup\{\mathbb{E}(X_n^+) : n \geq 0\}$.

3. By monotone convergence theorem,

$$\mathbb{E}(U) = \lim_{n\to\infty} \mathbb{E}(U_n) \leq \frac{M + |a|}{b-a} < \infty,$$

then $U < \infty$ a.s.

4.$\lim_{n\to\infty} X_n$ exists (finite or infinite) a.s.

Rewrite U as $U_{[a,b]}$ for any $a < b$. Notice

$$\{\omega : U_{[a,b]}(\omega) = \infty\} = \{\liminf X_n < a < b < \limsup X_n\}^4$$

then

$$\begin{aligned}
&\mathbb{P}(\lim X_n \text{ does not exist}) \\
&= \mathbb{P}(\liminf X_n < \limsup X_n) \\
&= \mathbb{P}(\bigcup_{a,b\in\mathbb{Q}} \{\liminf X_n < a < b < \limsup X_n\}) \\
&= \mathbb{P}(\bigcup_{a,b\in\mathbb{Q}} \{U_{[a,b]} = \infty\}) \\
&= 0,
\end{aligned}$$

therefore $\mathbb{P}(\lim X_n \text{ exists}) = 1$. Denote $X = \lim_{n\to\infty} X_n$ except the above null set, and $X = 0$ on the above null set, then $X_n \to X$ a.s.

5.$\mathbb{E}(|X|) < \infty$.

Notice that $|X_n| = X_n^+ + X_n^- = 2X_n^+ - X_n$, then

$$\mathbb{E}(|X_n|) = 2\mathbb{E}(X_n^+) - \mathbb{E}(X_n) \le 2\mathbb{E}(X_n^+) - \mathbb{E}(X_0) \le 2M + \mathbb{E}(|X_0|) < \infty,$$

where $\mathbb{E}(X_n) \ge \mathbb{E}(X_0)$ by the property of submartingale. By Fatou's Lemma,

$$\mathbb{E}(|X|) = \mathbb{E}(\liminf |X_n|) \le \liminf \mathbb{E}(|X_n|) \le 2M + \mathbb{E}(|X_0|) < \infty. \qquad \square$$

Corollary 2.15. *If $\{X_n : n \ge 0\}$ is a supermartingale, $X_n \ge 0$, then there exists a r.v. with*

[4] To make U (total number of upcrossings) infinite, X_n must be always oscillating to cross [a,b], i.e. its limit cannot exist.

$\mathbb{E}(X) \leq \mathbb{E}(X_0)$ *s.t.* $X_n \to X$ *a.s.*

Proof. $\{Y_n = -X_n : n \geq 0\}$ is a submartingale, and $\sup(Y_n^+) = 0 < \infty$, so by Theorem 2.14, there exists a r.v. Y with $\mathbb{E}(|Y|) < \infty$ s.t. $Y_n \to Y$ a.s. Let $X = -Y$, then $X_n \to X$ a.s. By Fatou's lemma,

$$\mathbb{E}(X) = \mathbb{E}(\liminf X_n) \leq \liminf \mathbb{E}(X_n) \leq \mathbb{E}(X_0),$$

because $\mathbb{E}(X_n) \leq \mathbb{E}(X_0)$ for all n by the definition of supermartingale. □

Application: Borel-Cantelli lemma for conditional probability

Lemma 2.16. *Suppose* $\{X_n : n \geq 0\}$ *is a martingale with* $|X_{n+1} - X_n| \leq M < \infty$ *for all* $n \geq 0$. *Let*

$$C = \{\lim_{n\to\infty} X_n < \infty\}, \quad D = \{\limsup X_n = +\infty \text{ and } \liminf X_n = -\infty\}.$$

Then $\mathbb{P}(C \cup D) = 1$.

Theorem 2.17 (Doob's decomposition)**.** *Any submartingale* $\{X_n : n \geq 0\}$ *can be written in a unique way as* $X_n = M_n + A_n$*, where* M_n *is a martingale and* A_n *is a predictable increasing sequence with* $A_0 = 0$.

Proof. 1. Define $A_0 = 0$. For $n \geq 1$, $A_n = \sum_{m=1}^{n}[\mathbb{E}(X_m|\mathcal{F}_{m-1}) - X_{m-1}] \in \mathcal{F}_{n-1}$, so A_n is predictable.
2. $A_n - A_{n-1} = \mathbb{E}(X_n|\mathcal{F}_{n-1}) - X_{n-1} \geq 0$, so A_n is increasing.
3.Let $M_n = X_n - A_n$.

$$\mathbb{E}(M_n|\mathcal{F}_{n-1}) = \mathbb{E}(X_n - A_n|\mathcal{F}_{n-1}) = \mathbb{E}(X_n|\mathcal{F}_{n-1}) - A_n = (A_n - A_{n-1} + X_{n-1}) - A_n = M_{n-1},$$

thus M_n is a martingale.

4. Suppose there is another pair M'_n and A'_n, s.t. $X_n = M'_n + A'_n$, then for $n \geq 1$, let

$$Y_n = M_n - M'_n = A'_n - A_n,$$

so Y_n is a martingale and $Y_n \in \mathcal{F}_{n-1}$. By Y_n is a martingale,

$$\mathbb{E}(Y_n|\mathcal{F}_{n-1}) = Y_{n-1}, \quad a.s.$$

By $Y_n \in \mathcal{F}_{n-1}$,

$$\mathbb{E}(Y_n|\mathcal{F}_{n-1}) = Y_n, \quad a.s.$$

thus $Y_n = Y_{n-1}$ a.s. And $Y_0 = A'_0 - A_0 = 0$ by definition, we have $Y_n = 0$ a.s. for all $n \geq 0$, i.e. $M_n = M'_n$ and $A_n = A'_n$ a.s. □

Theorem 2.18 (Borel-Cantelli lemma for conditional probability)**.** *Let*$\{\mathcal{F}_n : n \geq 0\}$ *be a filtration with* $\mathcal{F}_0 = \{\varnothing, \Omega\}$*, and* $\{B_n : n \geq 1\}$ *be a sequence of events with* $B_n \in \mathcal{F}_{n-1}$*, then*

$$\{B_n \ i.o.\} = \{\sum_{n=1}^{\infty} \mathbb{P}(B_n|\mathcal{F}_{n-1}) = \infty\}.$$

2.3 Doob's inequality

Lemma 2.19. *If* $X = Y$ *a.s., then* $\mathbb{E}(X) = \mathbb{E}(Y)$*.*

Proof. Denote $A = \{\omega : X(\omega) = Y(\omega)\}$,

$$\mathbb{E}(X - Y) = \mathbb{E}[(X_Y)\mathbb{1}_A] + \mathbb{E}[(X_Y)\mathbb{1}_{A^c}] = \mathbb{E}[(X_Y)\mathbb{1}_A] = 0,$$

where $\mathbb{E}[(X_Y)\mathbb{1}_{A^c}] = 0$ because the integral over a null set is zero. □

Lemma 2.20. *Let* $\{X_n : n \geq 0\}$ *be a submartingale, and* N *is a stopping time with*

$$\mathbb{P}(N \leq k) = 1$$

for some $k \in \mathbb{N}$[5], *then*

$$\mathbb{E}(X_0) \leq \mathbb{E}(X_N) \leq \mathbb{E}(X_k).$$

Proof. 1.By Proposition 2.12, $X_{N\wedge n}$ is a submartingale. And $X_{N\wedge k} = X_N$ a.s., then by the property of submartingale and Lemma 2.19,

$$\mathbb{E}(X_0) = \mathbb{E}(X_{N\wedge 0}) \leq \mathbb{E}(X_{N\wedge k}) = \mathbb{E}(X_N).$$

2. Define $K_n = \mathbb{1}_{\{N<n\}} = \mathbb{1}_{\{N\leq n-1\}}$, then $K_n \in \mathcal{F}_{n-1}$ thus predictable. By Proposition 2.9, $(K \cdot X)_n$ is also a submartingale, and

$$\begin{aligned}
(K \cdot X)_n &= \sum_{m=1}^{n} K_m(X_m - X_{m-1}) \\
&= \sum_{m=1}^{n} \mathbb{1}_{\{N\leq m-1\}}(X_m - X_{m-1}) \\
&= \begin{cases} \displaystyle\sum_{m=N+1}^{n} (X_m - X_{m-1}) = X_n - X_N, & N \leq n-1 \\ 0 & N \geq n \end{cases} \\
&= X_n - X_{N\wedge n}.
\end{aligned}$$

Taking $n = k$, we have (for $\omega \in \{N \leq k\}$)

$$(K \cdot X)_k = X_k - X_{N\wedge k} = X_k - X_N, \quad a.s.$$

then by the property of submartingale and Lemma 2.19,

$$0 = \mathbb{E}[(K \cdot X)_0] \leq \mathbb{E}[(K \cdot X)_k] = \mathbb{E}(X_k - X_N) = \mathbb{E}(X_k) - \mathbb{E}(X_N).$$ □

[5] $N \leq k$ for some k a.s. or N bounded a.s. is not the same as $N < \infty$ a.s. For example, suppose X has the normal distribution, then $X < \infty$ a.s. but X is not bounded a.s.

Theorem 2.21 (Doob's maximal inequality)**.** *Let* $\{X_m : m \geq 0\}$ *be a submartingale, for any* $\lambda > 0$*, denote*

$$\bar{X}_n = \max_{0 \leq m \leq n} X_m^+, \quad A = \{\bar{X}_n \geq \lambda\},$$

then

$$\lambda \mathbb{P}(A) \leq \mathbb{E}(X_n \mathbb{1}_A) \leq \mathbb{E}(X_n^+).$$

Proof. Let $N = \inf\{m : X_m \geq \lambda\} \wedge n$, then $X_N \geq \lambda$ on A, thus

$$\lambda \mathbb{P}(A) = \mathbb{E}(\lambda \mathbb{1}_A) \leq \mathbb{E}(X_N \mathbb{1}_A).$$

Since $N \leq n$ on Ω, then by Lemma 2.20, $\mathbb{E}(X_N) \leq \mathbb{E}(X_n)$. On $A^c = \{\bar{X}_n < \lambda\}$, $X_N = X_n$, i.e. $\mathbb{E}(X_N \mathbb{1}_{A^c}) = \mathbb{E}(X_n \mathbb{1}_{A^c})$, thus

$$\mathbb{E}(X_N \mathbb{1}_A) = \mathbb{E}(X_N) - \mathbb{E}(X_N \mathbb{1}_{A^c}) \leq \mathbb{E}(X_n) - \mathbb{E}(X_n \mathbb{1}_{A^c}) = \mathbb{E}(X_n \mathbb{1}_A) \leq \mathbb{E}(X_n^+ \mathbb{1}_A) \leq \mathbb{E}(X_n^+).$$

□

Below is an application of Doob's maximal inequality.

Theorem 2.22 (Kolmogorov's inequality)**.** *Suppose* $\{X_n : n \geq 1\}$ *are independent with* $\mathbb{E}(X_n) = 0$ *and* $\mathbb{E}(X_n^2) < \infty$*. Let* $S_n = \sum_{k=1}^n X_k$*, then for any* $a > 0$*,*

$$\mathbb{P}\left(\max_{1 \leq m \leq n} |S_m| \geq a\right) \leq \frac{\mathrm{Var}(S_n)}{a^2}.$$

Proof. S_n is a martingale because

$$\mathbb{E}(S_{n+1}|\mathcal{F}_n) = \mathbb{E}(S_n|\mathcal{F}_n) + \mathbb{E}(X_{n+1}|\mathcal{F}_n) = S_n + \mathbb{E}(X_{n+1}) = 0.$$

Then S_n^2 is a submartingale by Proposition 2.6. Applying Theorem 2.21 to S_n^2, and take

$\lambda = a^2$, we have

$$a^2\mathbb{P}\left(\max_{1\le m\le n} S_m^2 \ge a^2\right) \le \mathbb{E}(S_n^2) = \mathrm{Var}(S_n).$$

Notice that $\{\max_{1\le m\le n} S_m^2 \ge a^2\} = \{\max_{1\le m\le n} |S_m| \ge a\}$, which gives the desire result. □

Lemma 2.23. *If X is a r.v. with $X \ge 0$, then for any $p \in (1, +\infty)$,*

$$\int_0^{+\infty} pt^{p-1}\mathbb{P}(X \ge t)\,\mathrm{d}t = \mathbb{E}(X^p).$$

Proof.

$$\begin{aligned}\int_0^{+\infty} pt^{p-1}\mathbb{P}(X \ge t)\,\mathrm{d}t &= \int_0^{+\infty} pt^{p-1}\left[\int_\Omega \mathbb{1}_{\{X\ge t\}}\,\mathrm{d}\mathbb{P}\right]\mathrm{d}t \\ &= \int_\Omega \left[\int_0^{+\infty} pt^{p-1}\mathbb{1}_{\{X\ge t\}}\,\mathrm{d}t\right]\mathrm{d}\mathbb{P} \\ &= \int_\Omega \left[\int_0^{X} pt^{p-1}\,\mathrm{d}t\right]\mathrm{d}\mathbb{P} \\ &= \int_\Omega X^p\,\mathrm{d}\mathbb{P} = \mathbb{E}(X^p)\end{aligned}$$

□

Theorem 2.24 (Doob's L^p maximal inequality)**.** *If X_n is a submartingale, then for any $p \in (1, +\infty)$,*

$$\mathbb{E}(\bar{X}_n^p) \le \left(\frac{p}{p-1}\right)^p [\mathbb{E}(X_n^+)^p].$$

Morever, if Y_n is a martingale or a positive submartingale, and

$$Y_n^* = \max_{0\le m\le n} |Y_m|,$$

then for any $p \in (1, +\infty)$,

$$\mathbb{E}(|Y_n^*|^p) \le \left(\frac{p}{p-1}\right)^p \mathbb{E}(|Y_n|^p).$$

Proof. Take $M > 0$, we will work with $\bar{X}_n \wedge M$ first.

1. For any $t \geq 0$, if $M \geq t$, then

$$\{\omega : \bar{X}_n(\omega) \wedge M \geq t\} = \{\omega : \bar{X}_n(\omega) \geq t\};$$

if $M < t$, then

$$\{\omega : \bar{X}_n(\omega) \wedge M \geq t\} = \varnothing.$$

2. By Lemma 2.23, Theorem 2.21 and Fubini's theorem,

$$\begin{aligned}
\mathbb{E}[(\bar{X}_n \wedge M)^p] &= \int_0^{+\infty} pt^{p-1}\mathbb{P}(\bar{X}_n \wedge M \geq t)\,\mathrm{d}t \\
&= \int_0^{M} pt^{p-1}\mathbb{P}(\bar{X}_n \geq t)\,\mathrm{d}t \\
&\leq \int_0^{M} pt^{p-2}\mathbb{E}(X_n^+ \mathbb{1}_{\{\bar{X}_n \geq t\}})\,\mathrm{d}t \\
&= \int_0^{M} pt^{p-2}\mathbb{E}(X_n^+ \mathbb{1}_{\{\bar{X}_n \wedge M \geq t\}})\,\mathrm{d}t \\
&= p\mathbb{E}\left[X_n^+ \int_0^{\bar{X}_n \wedge M} t^{p-2}\,\mathrm{d}t\right] \\
&= \frac{p}{p-1}\mathbb{E}[X_n^+(\bar{X}_n \wedge M)^{p-1}] \\
&\leq \frac{p}{p-1}[\mathbb{E}(|X_n^+|^p)]^{1/p}[\mathbb{E}(|\bar{X}_n \wedge M|^p)]^{1/q}. \qquad q = p/(p-1)
\end{aligned}$$

Thus

$$\left(\mathbb{E}[|\bar{X}_n \wedge M|^p]\right)^{1/p} \leq \frac{p}{p-1}[\mathbb{E}(|X_n^+|^p)]^{1/p},$$

or

$$\mathbb{E}[|\bar{X}_n \wedge M|^p] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}(|X_n^+|^p),$$

let $M \to +\infty$, then we have

$$\mathbb{E}(\bar{X}_n^p) \le \left(\frac{p}{p-1}\right)^p [\mathbb{E}(X_n^+)^p].$$

3. Let $X_n = |Y_n|$, then X_n is a submartingale. Notice $X_n^+ = X_n = |Y_n|$, and

$$Y_n^* = \max_{0\le m\le n} |Y_m| = \max_{0\le m\le n} |Y_m|^+ = \bar{X}_n. \qquad \square$$

Theorem 2.25 (L^p convergence theorem)**.** *If X_n is a martingale or a positive submartingale with $\sup_n \mathbb{E}(|X_n|^p) < \infty$, $p \in (1,\infty)$, then $X_n \to X$ a.s. and in L^p.*

Proof. 1. By the property of martingale and positive submartingale,

$$\mathbb{E}(|X_{n+1}|^p|\mathcal{F}_n) \ge |\mathbb{E}(X_{n+1}|\mathcal{F}_n)|^p \ge |X_n|^p,$$

so $|X_n|^p$ is a submartingale. By the martingale convergence theorem, $|X_n|^p \to |X_\infty|^p$ a.s., then $X_n \to X_\infty$ a.s.

2. By Theorem 2.24,

$$\mathbb{E}\left[(\sup_{0\le m\le n} |X_m|)^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}(|X_n|^p).$$

3. Since $(\sup_{0\le m\le n} |X_m|)^p \uparrow (\sup_{n\ge 0} |X_n|)^p$, by monotone convergence theorem,

$$\mathbb{E}\left[(\sup_{n\ge 0} |X_n|)^p\right] = \lim_{n\to\infty} \mathbb{E}\left[(\sup_{0\le m\le n} |X_m|)^p\right] = \sup_n \mathbb{E}\left[(\sup_{0\le m\le n} |X_m|)^p\right] \le \left(\frac{p}{p-1}\right)^p \sup_n \mathbb{E}(|X_n|^p) < \infty,$$

thus $\sup_{n\ge 0} |X_n| \in L^p$.

4. Since $|X_\infty| = \limsup_n |X_n| \le \sup_n |X_n|$ a.s., we have $|X_n - X_\infty| \le (2\sup_n |X_n|)$ a.s., then by dominated convergence theorem, we have

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X_\infty|^p] = 0. \qquad \square$$

Lemma 2.26 (Orthogonality of martingale increments)**.** *Let X_n be a martingale with $\mathbb{E}(X_n^2) < \infty$ for all n. For $m \leq n$ and r.v. $Y \in \mathcal{F}_m$ with $\mathbb{E}(Y^2) < \infty$, we have*

$$\mathbb{E}[(X_n - X_m)Y] = 0.$$

Hence for $l < m < n$,

$$\mathbb{E}[(X_n - X_m)(X_m - X_l)] = 0.$$

Proof. 1. Cauchy-Schwarz: $\mathbb{E}[(X_n - X_m)Y] \leq \mathbb{E}(|X_nY|) + \mathbb{E}(|X_mY|) \leq \sqrt{\mathbb{E}(X_n^2)\mathbb{E}(Y^2)} + \sqrt{\mathbb{E}(X_m^2)\mathbb{E}(Y^2)} < \infty$.
2.$\mathbb{E}[(X_n - X_m)Y] = \mathbb{E}[\mathbb{E}[(X_n - X_m)Y|\mathcal{F}_m]] = \mathbb{E}[Y\mathbb{E}[(X_n - X_m)|\mathcal{F}_m]] = 0$. □

Lemma 2.27. *If X_n is a martingale with $\mathbb{E}(X_n^2) < \infty$ for all n, $m \leq n$, then*

$$\mathbb{E}[(X_n - X_m)^2|\mathcal{F}_m] = \mathbb{E}(X_n^2|\mathcal{F}_m) - X_m^2.$$

Proof. $\mathbb{E}[(X_n - X_m)^2|\mathcal{F}_m] = \mathbb{E}[X_n^2 - 2X_mX_n + X_m^2|\mathcal{F}_m] = \mathbb{E}(X_n^2|\mathcal{F}_m) - 2X_m\mathbb{E}(X_n|\mathcal{F}_m) + X_m^2$, conclude by $\mathbb{E}(X_n|\mathcal{F}_m) = X_m$. □

2.4 Uniform integrability and convergence in L^1

2.4.1 Definition and examples

Definition 2.28. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\{X_i : i \in I\}$ be a collection of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, we call they are uniformly integrable (UI) if

$$\lim_{M\to\infty}\left(\sup_{i\in I}\mathbb{E}(|X_i|\mathbb{1}_{\{|X_i|>M\}})\right) = 0$$

Proposition 2.29. *$\{X_i : i \in I\}$ are UI if and only if*

(1) *L^1 bounded:* $\sup_{i\in I}\mathbb{E}(|X_i|) < \infty$

(2) *uniform absolutely continuous: for any $\varepsilon > 0$, there exists $\delta > 0$ s.t. for any $A \in \mathcal{F}$ with $\mathbb{P}(A) < \delta$, we have*

$$\sup_{i\in I}\mathbb{E}(|X_i|\mathbb{1}_A) < \varepsilon.$$

Proof. $\Longrightarrow$: Suppose $\{X_i : i \in I\}$ is UI. We can find $M > 0$, s.t.

$$\sup_{i\in I}\mathbb{E}(|X_i|\mathbb{1}_{\{|X_i|>M\}}) < 1.$$

then for any $i \in I$

$$\mathbb{E}(|X_i|) = \mathbb{E}(|X_i|\mathbb{1}_{\{|X_i|\le M\}}) + \mathbb{E}(|X_i|\mathbb{1}_{\{|X_i|>M\}}) \le M + \mathbb{E}(|X_i|\mathbb{1}_{\{|X_i|>M\}}) < M + 1,$$

thus $\sup_{i\in I}\mathbb{E}(|X_i|) < M + 1 < \infty$. (1) is proved. Then we will prove (2). Take $\varepsilon > 0$, we can find $M > 0$, s.t.

$$\sup_{i\in I}\mathbb{E}(|X_i|\mathbb{1}_{\{|X_i|>M\}}) < \frac{\varepsilon}{2},$$

then take $\delta = \dfrac{\varepsilon}{2M}$, for any A with $\mathbb{P}(A) < \delta$, we have

$$\begin{aligned}\sup_{i\in I}\mathbb{E}(|X_i|\mathbb{1}_A) &= \sup_{i\in I}\mathbb{E}(|X_i|\mathbb{1}_A\mathbb{1}_{\{|X_i|\le M\}}) + \sup_{i\in I}\mathbb{E}(|X_i|\mathbb{1}_A\mathbb{1}_{\{|X_i|>M\}}) \\ &\le \sup_{i\in I}\mathbb{E}(M\mathbb{1}_A\mathbb{1}_{\{|X_i|\le M\}}) + \sup_{i\in I}\mathbb{E}(|X_i|\mathbb{1}_A\mathbb{1}_{\{|X_i|>M\}}) \\ &< M\mathbb{P}(A) + \frac{\varepsilon}{2} \\ &< M\cdot\frac{\varepsilon}{2M} + \frac{\varepsilon}{2} = \varepsilon.\end{aligned}$$

$\Longleftarrow$: Suppose (1) and (2) hold. Take $C > 0$ to satisfy $\sup_{i\in I}\mathbb{E}(|X_i|) < C < \infty$. For any $\varepsilon > 0$, take δ from (2). Let $N = \dfrac{\delta}{C}$, then

$$\mathbb{P}(|X_i| > M) \le \frac{\mathbb{E}(|X_i|)}{M} < \delta, \quad \forall i \in I,$$

by (2),

$$\sup_{i\in I}\mathbb{E}(|X_i|\mathbb{1}_{\{|X_i|>M\}}) < \varepsilon,$$

from the $\varepsilon - \delta$ definition, we have $\sup_{i\in I}\mathbb{E}(|X_i|\mathbb{1}_{\{|X_i|>M\}}) \to 0$ as $N \to \infty$. □

Lemma 2.30. *If $X \in L^1$, then*

1. *for any $\varepsilon > 0$, there exists $\delta > 0$ s.t. $A \in \mathcal{F}$ with $\mathbb{P}(A) < \delta$ implies $\mathbb{E}(|X|\mathbb{1}_A) < \varepsilon$.*

2. $$\lim_{M\to\infty}\mathbb{E}(|X|\mathbb{1}_{\{|X|>M\}}) = 0.$$

Example 2.31. Let $0 < C < \infty$ be a constant, then $\{X_n\}$ with $|X_n| \le C$ are UI.

Proof. Take $M = C$, then $\mathbb{E}(|X_n|\mathbb{1}_{\{X_n>M\}}) = 0$ for all n. □

Example 2.32. Suppose $\{X_1, X_2, \cdots, X_n\}$ are all in L^1, then they are also UI.

Proof. First they are L^1 bounded. Second, for any $\varepsilon > 0$, by Lemma 2.30, we can find $\delta_i > 0$, s.t. $\mathbb{P}(A) < \delta_i$ implies

$$\mathbb{E}(|X_i|\mathbb{1}_A) < \varepsilon.$$

Thus take $\delta = \min\{\delta_1, \cdots, \delta_n\}$, we have $\mathbb{P}(A) < \delta$ implies

$$\mathbb{E}(|X_i|\mathbb{1}_A) < \varepsilon, \quad \forall i \in \{1, \cdots n\},$$

s.t. $\sup_i \mathbb{E}(|X_i|\mathbb{1}_A) < \varepsilon$. By Proposition 2.29, $\{X_i\}$ is UI. □

Example 2.33. Let U be a r.v. with uniform distribution on $[0, 1]$, define

$$X_n = n\mathbb{1}_{\{U \le \frac{1}{n}\}},$$

then $\mathbb{E}(|X_n|) = 1$ for all n, thus they are L^1 bounded, but for any $M > 0$,

$$\mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|>M\}}) = 1, \quad \forall n \ge [M] + 1,$$

thus they are not UI.

Example 2.34. Let X be integrable r.v., then $\{Y_n\}$ with $|Y_n| \le |X|$ are UI.

Proof. Since $\{|Y_n| > M\} \subseteq \{|X| > M\}$, as $M \to \infty$,

$$\sup_n \mathbb{E}(|Y_n|\mathbb{1}_{\{|Y_n|>M\}}) \le \mathbb{E}(|X|\mathbb{1}_{\{|X|>M\}}) \to 0.$$ □

Example 2.35. Let $\{X_n\}$ be UI, then $\{Y_n\}$ with $|Y_n| \le |X_n|$ are also UI.

Proof. Since $\{|Y_n| > M\} \subseteq \{|X_n| > M\}$, $\mathbb{1}_{\{|Y_n|>M\}} \le \mathbb{1}_{\{|X_n|>M\}}$, then as $M \to \infty$,

$$\sup_n \mathbb{E}(|Y_n|\mathbb{1}_{\{|Y_n|>M\}}) \le \sup_n \mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|>M\}}) \to 0.$$ □

Example 2.36. Let $\{X_n\}$ and $\{Y_n\}$ both be UI, then $\{X_n + Y_n\}$ are also UI.

Proof. As $M \to \infty$,

$$\begin{aligned}
\sup_n \mathbb{E}(|X_n + Y_n|\mathbb{1}_{\{|X_n+Y_n|>M\}}) &\le \sup_n \mathbb{E}((|X_n| + |Y_n|)\mathbb{1}_{\{|X_n|+|Y_n|>M\}}) \\
&= \sup_n \mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|+|Y_n|>M\}}) + \sup_n \mathbb{E}(|Y_n|\mathbb{1}_{\{|X_n|+|Y_n|>M\}}) \\
&\le \sup_n \mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|+\sup_n |Y_n|>M\}}) + \sup_n \mathbb{E}(|Y_n|\mathbb{1}_{\{|Y_n|+\sup_n |X_n|>M\}}) \\
&= \sup_n \mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|>M-A\}}) + \sup_n \mathbb{E}(|Y_n|\mathbb{1}_{\{|Y_n|>M-B\}}) \to 0,
\end{aligned}$$

here we let $A = \sup_n |Y_n| < \infty$, $B = \sup_n |X_n| < \infty$. □

Example 2.37. Let $\mathcal{F}_n \subseteq \mathcal{F}$ be sub-σ-fields, $X \in L^1$, then $\{\mathbb{E}(X|\mathcal{F}_n) : n \ge 0\}$ is UI.

Proof. Let $Y_n = \mathbb{E}(X|\mathcal{F}_n) \in \mathcal{F}_n$. For any $\varepsilon > 0$, our goal is to find $M > 0$, s.t.

$$\mathbb{E}[|Y_n|\mathbb{1}_{\{|Y_n|>M\}}] < \varepsilon, \quad \forall n.$$

By Lemma 2.30, since $X \in L^1$, there exists $\delta > 0$, s.t. $\mathbb{P}(A) < \delta$ implies $\mathbb{E}(|X|\mathbb{1}_A) < \varepsilon$. By Jensen's inequality, $|Y_n| \le \mathbb{E}(|X||\mathcal{F}_n)$, then

$$\begin{aligned}
\mathbb{E}[|Y_n|\mathbb{1}_{\{|Y_n|>M\}}] &\le \mathbb{E}[\mathbb{E}(|X||\mathcal{F}_n)\mathbb{1}_{\{|Y_n|>M\}}] \\
&\le \mathbb{E}[\mathbb{E}(|X||\mathcal{F}_n)\mathbb{1}_{\{\mathbb{E}(|X||\mathcal{F}_n)>M\}}] \\
&= \mathbb{E}[|X|\mathbb{1}_{\{\mathbb{E}(|X||\mathcal{F}_n)>M\}}] \qquad (\text{definition of } \mathbb{E}(|X||\mathcal{F}_n))
\end{aligned}$$

where

$$\mathbb{P}(\mathbb{E}(|X||\mathcal{F}_n) > M) \le \frac{\mathbb{E}[\mathbb{E}(|X||\mathcal{F}_n)]}{M} = \frac{\mathbb{E}(|X|)}{M} < \delta,$$

if we take $M = [\mathbb{E}(|X|)/\delta] + 1$. Thus

$$\mathbb{E}[|Y_n|\mathbb{1}_{\{|Y_n|>M\}}] \le \mathbb{E}[|X|\mathbb{1}_{\{\mathbb{E}(|X||\mathcal{F}_n)>M\}}] < \varepsilon \quad \forall n.$$

□

Remark. $\{\mathcal{F}_n : n \geq 0\}$ does not need to be increasing or decreasing.

Proposition 2.38. *$\{X_n : n \in I\}$ are UI if and only if there exists a measurable function $\varphi : [0, \infty) \to [0, \infty)$ s.t.*

$$\lim_{x\to\infty} \frac{\varphi(x)}{x} = +\infty,$$

and

$$\sup_{n\in I} \mathbb{E}[\varphi(|X_n|)] < \infty.$$

Proof. $\Longleftarrow$: First there exists $M \in [0, \infty)$ s.t. for all i,

$$\mathbb{E}[\varphi(|X_n|)] \leq M.$$

Then by $\lim_{x\to\infty} \frac{\varphi(x)}{x} = +\infty$, for any $k \in \mathbb{Z}_+$, there exists $C_k > 0$, s.t.

$$\varphi(x) > kMx, \quad \forall x > C_k.$$

Therefore, for any $n \in I$,

$$M \geq \mathbb{E}[\varphi(|X_n|)] \geq \mathbb{E}[\varphi(|X_n|)\mathbb{1}_{\{|X_n|>C_k\}}] \geq kM\mathbb{E}[|X_n|\mathbb{1}_{|X_n|>C_k\}}],$$

then we have

$$\sup_{n\in I} \mathbb{E}[|X_n|\mathbb{1}_{|X_n|>C_k\}}] \leq \frac{1}{k}.$$

For any $\varepsilon > 0$, just choose $k = [1/\varepsilon] + 1$, take $N = C_k > 0$, we have

$$\sup_{n\in I} \mathbb{E}[|X_n|\mathbb{1}_{|X_n|>N\}}] = \sup_{n\in I} \mathbb{E}[|X_n|\mathbb{1}_{|X_n|>C_k\}}] \leq \frac{1}{k} < \varepsilon.$$

$\Longrightarrow$: Omitted. □

Corollary 2.39. *For $p \in (1, \infty)$, let $\{X_n : n \in I\}$ with $\sup_{n\in I} |X_n|^p < \infty$, then they are UI.*

Proof. $\varphi(x) = x^p$ satisfies

$$\lim_{x\to\infty} \frac{\varphi(x)}{x} = \lim_{x\to\infty} x^{p-1} = +\infty,$$

and $\sup_{n\in I} \varphi(|X_n|) < \infty$, thus it is proved by Proposition 2.38. □

2.4.2 UI and convergence

Lemma 2.40. *If $X_n \to X$ in probability, then there exists a subsequence $\{X_{n_k} : k \geq 0\}$ s.t.*

$$X_{n_k} \to X \quad a.s.$$

as $k \to \infty$.

Lemma 2.41. *Suppose $X_n \geq 0$ and $X_n \to X$ in probability, then*

$$\mathbb{E}(X) \leq \liminf_{n\to\infty} \mathbb{E}(X_n).$$

Lemma 2.42 (Bounded convergence theorem)**.** *Suppose $X_n \leq K < \infty$ for all $n \geq 0$, and $X_n \to X$ in probability, then $X_n \to X$ in L^1.*

Proof. Since $|X| = |X - X_n + X_n| \leq |X - X_n| + |X_n| \leq |X - X_n| + K$, then $|X| - K \geq \frac{1}{m}$ implies $|X - X_n| \geq \frac{1}{m}$, i.e.

$$\mathbb{P}(|X| \geq K + \frac{1}{m}) \leq \mathbb{P}(|X_n - X| \geq \frac{1}{m}) \to 0, \quad \text{as } n \to \infty.$$

Let $m \to \infty$, we have $\mathbb{P}(|X| > K) = 0$, i.e. $|X| \leq K$ a.s. For any $\varepsilon > 0$,

$$\begin{aligned}\mathbb{E}(|X_n - X|) &= \mathbb{E}(|X_n - X|\mathbb{1}_{\{|X_n - X| > \frac{\varepsilon}{2}\}}) + \mathbb{E}(|X_n - X|\mathbb{1}_{\{|X_n - X| \leq \frac{\varepsilon}{2}\}}) \\ &\leq 2K\mathbb{P}(|X_n - X| > \frac{\varepsilon}{2}) + \frac{\varepsilon}{2} \quad (\text{since } |X_n - X| \leq |X_n| + |X| \leq 2K \text{ a.s.}) \\ &\to \frac{\varepsilon}{2}, \quad \text{as } n \to \infty.\end{aligned}$$

Let $\varepsilon \to 0$, we have $\mathbb{E}(|X_n - X|) \to 0$. $\square$

Theorem 2.43. *Let $\{X_n : n \geq 0\}$ be a sequence of r.v. with $X_n \in L^1$, and $X_n \to X$ in probability, then TFAE:*

1. *$\{X_n : n \geq 0\}$ is UI;*

2. *$X_n \to X$ in L^1;*

3. *$\mathbb{E}(|X_n|) \to \mathbb{E}(|X|) < \infty$.*

Proof. $1 \Longrightarrow 2$. The idea is to truncate X_n at K and $-K$. For $K > 0$, define

$$\varphi_K(x) = x\mathbb{1}_{\{|x|\leq K\}} + K\mathbb{1}_{\{x>K\}} - K\mathbb{1}_{\{x<-K\}},$$

then $|\varphi_K(x)| \leq K$, $|\varphi_K(x) - x| \leq |x|\mathbb{1}_{\{|x|>K\}}$ and $|\varphi_K(x) - \varphi_K(y)| \leq |x - y|$. By triangle inequality, we have

$$\begin{aligned}\mathbb{E}(|X_n - X|) &\leq \mathbb{E}(|\varphi_K(X_n) - \varphi_K(X)|) + \mathbb{E}(|\varphi_K(X_n) - X_n|) + \mathbb{E}(|\varphi_K(X) - X|)\\ &\leq \mathbb{E}(|\varphi_K(X_n) - \varphi_K(X)|) + \mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|>K\}}) + \mathbb{E}(|X|\mathbb{1}_{\{|X|>K\}}).\end{aligned}$$

Take $\varepsilon > 0$. For the first term, since $|\varphi_K(X_n) - \varphi_K(X)| \leq |X_n - X|$, for any $\delta > 0$,

$$\mathbb{P}(|\varphi_K(X_n) - \varphi_K(X)| \geq \delta) \leq \mathbb{P}(|X_n - X| \geq \delta) \to 0,$$

which means $\varphi_K(X_n) \to \varphi_K(X)$ in probability, by Lemma 2.42, $\varphi_K(X_n) \to \varphi_K(X)$ in L^1, so there exists $N(\varepsilon, K) \in \mathbb{Z}_+$, s.t. for any $n \geq N$,

$$\mathbb{E}(|\varphi_K(X_n) - \varphi_K(X)|) < \frac{\varepsilon}{3}.$$

For the second term, since X_n is UI, then for any $\varepsilon > 0$, there exists $K_1 > 0$, s.t. for all n

$$\mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|>K_1\}}) < \frac{\varepsilon}{3}.$$

For the third term, by Lemma 2.41 and Proposition 2.29,

$$\mathbb{E}(|X|) \leq \liminf_{n\to\infty} \mathbb{E}(|X_n|) \leq \sup_n \mathbb{E}(|X_n|) < \infty,$$

therefore $X \in L^1$. By Lemma 2.30, there exists $K_1 > 0$, s.t.

$$\mathbb{E}(|X|\mathbb{1}_{\{|X|>K_1\}}) < \frac{\varepsilon}{3}.$$

Taken together, we choose $K_0 = \max\{K_1, K_2\}$ and $N = N(\varepsilon, K_0)$, then for all $n \geq N$,

$$\mathbb{E}(|X_n - X|) \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon,$$

i.e. $\mathbb{E}(|X_n - X|) \to 0$.

$2 \implies 3$.By Jensen's inequality and $X_n \to X$ in L^1, we have

$$|\mathbb{E}(|X_n|) - \mathbb{E}(|X|)| = |\mathbb{E}(|X_n| - |X|)| \leq \mathbb{E}(||X_n| - |X||) \leq \mathbb{E}(|X_n - X|) \to 0.$$

$3 \implies 1$. □

Theorem 2.44. *Let $\{X_n : n \geq 0\}$ be a submartingale, then TFAE:*

1. *it is UI;*
2. *it converges a.s. and in L^1;*
3. *it converges in L^1.*

Proof. 1 $\implies$ 2. UI implies $\sup_n |X_n| < \infty$, thus $\sup_n X_n^+ \leq \sup_n |X_n| < \infty$, by martingale convergence theorem (Theorem 2.14), there exists $X \in L^1$ s.t. $X_n \to X$ a.s., then $X_n \to X$ in probability. By Theorem 2.43, $X_n \to X$ in L^1.

2 $\implies$ 3. Trivial.

3 $\implies$ 1. Convergence in L^1 implies convergence in probability, then also by Theorem 2.43, $\{X_n : n \geq 0\}$ is UI. □

Lemma 2.45. *If $X_n \in L^1$, and $X_n \to X$ in L^1, then*

$$\mathbb{E}(X_n \mathbb{1}_A) \to \mathbb{E}(X \mathbb{1}_A).$$

Proof.

$$\begin{aligned}|\mathbb{E}(X_n \mathbb{1}_A) - \mathbb{E}(X \mathbb{1}_A)| &= |\mathbb{E}(X_n \mathbb{1}_A - X \mathbb{1}_A)| \leq \mathbb{E}(|X_n \mathbb{1}_A - X \mathbb{1}_A|) \\ &= \mathbb{E}(|X_n - X| \mathbb{1}_A) \leq \mathbb{E}(|X_n - X|) \to 0.\end{aligned}$$

□

Lemma 2.46. *If X_n is a martingale w.r.t. $\mathcal{F}_n$, and $X_n \to X$ in L^1, then $X_n = \mathbb{E}(X|\mathcal{F}_n)$.*

Proof. By the property of martingale, for any integer $m > n$, $\mathbb{E}(X_m|\mathcal{F}_n) = X_n$. By the definition of $\mathbb{E}(X_m|\mathcal{F}_n)$, for any $A \in \mathcal{F}_n$,

$$\mathbb{E}(X_m \mathbb{1}_A) = \mathbb{E}(X_n \mathbb{1}_A).$$

Since $X_m \to X$ in L^1, by Lemma 2.45,

$$\mathbb{E}(X \mathbb{1}_A) = \lim_{m \to \infty} \mathbb{E}(X_m \mathbb{1}_A) = E(X_n \mathbb{1}_A), \quad \forall A \in \mathcal{F}_n.$$

Since $X_n \in \mathcal{F}_n$, by the definition of $\mathbb{E}(X|\mathcal{F}_n)$, we conclude $X_n = \mathbb{E}(X|\mathcal{F}_n)$. □

Theorem 2.47. *Suppose X_n is a martingale w.r.t. $\mathcal{F}_n$. Then TFAE*

1. *It is UI*

2. *It converges a.s. and in L^1*

3. *It converges in L^1*

4. *There exists a r.v. $X \in L^1$ s.t. for any $n \geq 0$*

$$\mathbb{E}(X|\mathcal{F}_n) = X_n.$$

Proof. $1 \Longrightarrow 2 \Longrightarrow 3$ is copied from Theorem 2.44.

$3 \Longrightarrow 4$. From Lemma 2.46.

$4 \Longrightarrow 1$. From Example 2.37. □

Theorem 2.48. *Suppose $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, i.e. $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots$ are sub-σ-field, and $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$. If $X \in L^1$, then*

$$\mathbb{E}(X|\mathcal{F}_n) \to \mathbb{E}(X|\mathcal{F}_\infty) \quad \textit{a.s. and in } L^1.$$

Proof. By Example 2.4 and Example 2.37, $M_n = \mathbb{E}(X|\mathcal{F}_n)$ is a martingale and UI. Thus Theorem 2.47 implies there exists $M \in L^1$ s.t. $M_n \to M$ a.s. and in L^1. The only thing is to show $M = \mathbb{E}(X|\mathcal{F}_\infty)$. Lemma 2.46 implies

$$\mathbb{E}(X|\mathcal{F}_n) = M_n = \mathbb{E}(M|\mathcal{F}_n),$$

thus for any $A \in \mathcal{F}_n$,

$$\mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(M_n\mathbb{1}_A) = \mathbb{E}(M\mathbb{1}_A).$$

Therefore $\mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(M\mathbb{1}_A)$ for all $A \in \cup_n \mathcal{F}_n$. Define

$$\mathcal{C} = \{A \in \mathcal{F} : \mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(M\mathbb{1}_A)\},$$

then $\mathcal{C}$ is a λ-system and $\cup_n \mathcal{F}_n \subseteq \mathcal{C}$. By $\pi - \lambda$ theorem, we have

$$\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n) \subseteq \mathcal{C},$$

i.e. $\mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(M\mathbb{1}_A)$ for all $A \in \cup_n \mathcal{F}_\infty$. And $M \in \mathcal{F}_\infty$ (Since each $M_n \in \mathcal{F}_\infty$, thus their limit $M \in \mathcal{F}_\infty$), we have $M = \mathbb{E}(X|\mathcal{F}_\infty)$. □

Theorem 2.49 (Lévy's 0-1 law)**.** *Suppose $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, and $A \in \mathcal{F}_\infty$, then*

$$\mathbb{E}(\mathbb{1}_A|\mathcal{F}_n) \to \mathbb{1}_A \quad a.s.$$

Proof. Let $X = \mathbb{1}_A \in \mathcal{F}_\infty$ in Theorem 2.48, we have

$$\mathbb{E}(\mathbb{1}_A|\mathcal{F}_n) \to \mathbb{E}(\mathbb{1}_A|\mathcal{F}_\infty) = \mathbb{1}_A. \quad \text{a.s.}$$ □

Corollary 2.50 (Kolmogorov's 0-1 law)**.** *Suppose $\{X_n : n \geq 1\}$ are independent random variables, define tail σ-field by*

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(X_m, m \geq n),$$

then for any $A \in \mathcal{T}$, $\mathbb{P}(A) \in \{0,1\}$, i.e. $\mathcal{T}$ is trivial.

Proof. Define $\mathcal{F}_n = \sigma(X_m, 1 \leq m \leq n)$, then for any $A \in \mathcal{T}$ and any $n \in \mathbb{Z}_+$ A is independent of $\mathcal{F}_n$ because $A \in \sigma(X_m, m \geq n+1)$ and $\sigma(X_m, m \geq n+1)$ is independent of $\mathcal{F}_n$. Thus $\mathbb{E}(\mathbb{1}_A|\mathcal{F}_n) = \mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$. By Lévy's 0-1 law,

$$\mathbb{P}(A) = \mathbb{E}(\mathbb{1}_A|\mathcal{F}_n) \to \mathbb{1}_A \quad a.s.$$

therefore $\mathbb{P}(A) \in \{0,1\}$. □

2.5 Backward martingale

Definition 2.51. Suppose $\{X_{-n} : n \geq 0\}$ is a sequence of r.v. w.r.t. $\mathcal{F}_{-n}$ with $\mathcal{F}_{-n} \subseteq \mathcal{F}_{-n+1}$. We call $\{X_{-n} : n \geq 0\}$ a backward martingale if $X_0 \in L^1$ and for any $n \geq 1$,

$$\mathbb{E}(X_{-n+1}|\mathcal{F}_{-n}) = X_{-n}.$$

Lemma 2.52. *Suppose $\{X_{-n} : n \geq 0\}$ is a backward martingale. If $X_0 \in L^p$ for some $p \geq 1$, then $X_{-n} \in L^p$ for all $n \geq 1$.*

Proof. By Jensen's inequality,

$$|X_{-n}|^p = |\mathbb{E}(X_{-n+1}|\mathcal{F}_{-n})|^p \leq \mathbb{E}(|X_{-n+1}|^p|\mathcal{F}_{-n}),$$

thus

$$\mathbb{E}(|X_{-n}|^p) \leq \mathbb{E}(|X_{-n+1}|^p).$$

By induction, $\mathbb{E}(|X_{-n}^p| \leq \mathbb{E}(|X_0|^p) < \infty$ for all $n \geq 1$. □

Theorem 2.53. *There exists $X_{-\infty} \in L^1$ s.t.*

$$X_{-n} \to X_{-\infty} \quad a.s. \text{ and in } L^1,$$

as $n \to \infty$.

Proof. 1. Let U_n be the number of upcrossings of $[a, b]$ by $X_{-n}, \cdots, X_0$. Then upcrossing inequality 2.13 implies

$$(b - a)\mathbb{E}(U_n) \leq \mathbb{E}[(X_0 - a)^+] - \mathbb{E}[(X_{-n} - a)^+] \leq \mathbb{E}[(X_0 - a)^+].$$

Since $U_n \uparrow U_\infty$, by monotone convergence theorem,

$$\mathbb{E}(U_\infty) = \lim_{n\to\infty} \mathbb{E}(U_n) \le \mathbb{E}[(X_0 - a)^+] < \infty,$$

thus $U_\infty < \infty$ a.s. By the similar argument in the proof of Theorem 2.14, $X_{-\infty}$ exists a.s. hence also in probability.

2. For any $n \in \mathbb{Z}_+$, $X_{-n} = \mathbb{E}(X_0|\mathcal{F}_{-n})$. Since $X_0 \in L^1$, by Example 2.37, $\{X_{-n} : n \ge 0\}$ is UI. By Lemma 2.52, $X_{-n} \in L^1$ for all $n \ge 0$, then Theorem 2.43 implies $X_{-n} \to X_{-\infty}$ in L^1. □

Theorem 2.54. *If backward martingale $\{X_{-n} : n \ge 0\}$ has $X_0 \in L^p$, then as $n \to \infty$, $X_{-n} \to X_{-\infty}$ in L^p.*

Proof. 1. By Theorem 2.53, as $n \to \infty$, $X_{-n} \to X_{-\infty}$ a.s.

2. By the Theorem 2.24, for any $n \ge 0$, we have

$$\mathbb{E}\left[(\sup_{-n\le m\le 0} |X_m|)^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}(|X_0|^p) < \infty.$$

3. Since $(\sup_{-n\le m\le 0} |X_m|)^p \uparrow (\sup_{n\ge 0} |X_{-n}|)^p$, by monotone convergence theorem,

$$\mathbb{E}\left[(\sup_{n\ge 0} |X_{-n}|)^p\right] = \lim_{n\to\infty} \mathbb{E}\left[(\sup_{-n\le m\le 0} |X_m|)^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}(|X_0|^p) < \infty,$$

thus $\sup_{n\ge 0} |X_{-n}| \in L^p$.

4. Since

$$|X_{-\infty}| = \limsup_{n\ge 0} |X_{-n}| \le \sup_{n\ge 0} |X_{-n}| \quad a.s.$$

and $|X_{-n}| \le \sup_{n\ge 0} |X_{-n}|$, we have

$$|X_{-n} - X_{-\infty}| \le |X_{-n}| + |X_{-\infty}| \le 2\sup_{n\ge 0} |X_{-n}| \in L^p, \quad a.s.$$

Then since $|X_{-n} - X_{-\infty}| \to 0$ a.s., by the L^p dominated convergence theorem,

$$\lim_{n\to\infty} \mathbb{E}(|X_{-n} - X_{-\infty}|^p) = 0. \qquad \square$$

Theorem 2.55. *Let $\mathcal{F}_{-\infty} = \bigcap_{n=0}^{\infty} \mathcal{F}_{-n}$. Then*

1. $X_{-\infty} = \mathbb{E}(X_0|\mathcal{F}_{-\infty})$.

2. *For any $X \in L^1$, as $n \to \infty$,*

$$\mathbb{E}(X|\mathcal{F}_{-n}) \to \mathbb{E}(X|\mathcal{F}_{-\infty}).$$

Proof. 1. We only need to show (i) $X_{-\infty} \in \mathcal{F}_{-\infty}$ and (ii) for any $A \in \mathcal{F}_{-\infty}$,

$$\mathbb{E}(X_{-\infty}\mathbb{1}_A) = \mathbb{E}(X_0\mathbb{1}_A). \tag{1}$$

(i) can be checked by showing $\{X_{-\infty} < c\} \in \mathcal{F}_{-n}$ for all $n \geq 0$. For (ii), since $X_{-n} = \mathbb{E}(X_0\mathcal{F}_{-n})$, we have for any $A \in \mathcal{F}_{-\infty} \subseteq \mathcal{F}_{-n}$,

$$\mathbb{E}(X_{-n}\mathbb{1}_A) = \mathbb{E}(X_0\mathbb{1}_A).$$

Then (1) holds from Lemma 2.45 and $X_{-n} \to X_{-\infty}$ in L^1. $\square$

2.6 Optional stopping theorem

For submartingale X_n, it is obvious $\mathbb{E}(X_n) \geq \mathbb{E}(X_0)$, but this is not always true for X_N when N is a stopping time. Optional stopping theorems are talking about when $\mathbb{E}(X_N) \geq \mathbb{E}(X_0)$ holds.

Theorem 2.56. *Suppose X_n is a submartingale, N is a stopping time, and N is bounded i.e. $\mathbb{P}(N \leq k) = 1$ for some $k < \infty$. Then $\mathbb{E}(X_N) \geq \mathbb{E}(X_0)$.*

Proof. See Lemma 2.20. ☐

Lemma 2.57. *X_n is a submartingale, N is a stopping time. If $X_{N\wedge n}$ is UI, then $\mathbb{E}(X_N) \geq \mathbb{E}(X_0)$.*

Proof. By Proposition 2.12, $X_{N\wedge n}$ is a UI submartingale, by the property of submartingale,

$$\mathbb{E}(X_0) = \mathbb{E}(X_{N\wedge 0}) \leq \mathbb{E}(X_{N\wedge n}).$$

By Theorem 2.44, $X_{N\wedge n} \to X_N$ a.s. and in L^1, thus

$$\mathbb{E}(X_N) - \mathbb{E}(X_0) = \mathbb{E}(X_N - X_{N\wedge n}) + \mathbb{E}(X_{N\wedge n}) - \mathbb{E}(X_0) \geq \mathbb{E}(X_N - X_{N\wedge n}) \to 0.$$ ☐

Lemma 2.58. *X_n is a UI submartingale, N is a stopping time, then $X_{N\wedge n}$ is UI.*

Proof. 1. X_n^+ is a submartingale.
$\mathbb{E}(X_{n+1}^+|\mathcal{F}_n) \geq \mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n$.
2. $N \wedge n \leq n$, then by Lemma 2.20,

$$\mathbb{E}(X_{N\wedge n}^+) \leq \mathbb{E}(X_n^+).$$

3. $|X_n^+| \leq |X_n|$ and Example 2.35 implies that X_n^+ is also UI.
4. By the property of UI and Step 2,

$$\sup_n \mathbb{E}(X_{N\wedge n}^+) \leq \sup_n \mathbb{E}(X_n^+) < \infty.$$

5.By Martingale convergence theorem (2.14), $X_{N\wedge n} \to X_N$ a.s. and $\mathbb{E}(|X_N|) < \infty$.

6. prove $X_{N\wedge n}$ is UI. For any $M > 0$,

$$\begin{aligned}\mathbb{E}(|X_{N\wedge n}|\mathbb{1}_{\{|X_{N\wedge n}|>M\}}) &= \mathbb{E}(|X_N|\mathbb{1}_{\{|X_N|>M\}}\mathbb{1}_{\{N\leq n\}}) + \mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|>M\}}\mathbb{1}_{\{N>n\}}) \\ &\leq \mathbb{E}(|X_N|\mathbb{1}_{\{|X_N|>M\}}) + \mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|>M\}}).\end{aligned}$$

Take $\varepsilon > 0$. For the first term, since $\mathbb{E}(|X_N|) < \infty$, by Lemma 2.30, there exists $M_1 > 0$ s.t.

$$\mathbb{E}(|X_N|\mathbb{1}_{\{|X_N|>M_1\}}) < \frac{\varepsilon}{2}.$$

For the second term, since X_n is UI, then there exists $M_2 > 0$ s.t.

$$\sup_n \mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|>M_2\}}) < \frac{\varepsilon}{2}.$$

Therefore for $M \geq \max\{M_1, M_2\}$,

$$\sup_n \mathbb{E}(|X_{N\wedge n}|\mathbb{1}_{\{|X_{N\wedge n}|>M\}}) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

which implies $X_{N\wedge n}$ is UI. □

Theorem 2.59. *Suppose X_n is a UI submartingale, N is a stopping time. Then $\mathbb{E}(X_N) \geq \mathbb{E}(X_0)$.*

Proof. By Lemma 2.57 and 2.58. □

Actually, from Lemma 2.58, we can use a weaker assumption than the above UI condition.

Theorem 2.60. *Suppose X_n is a submartingale, N is a stopping time. If $\mathbb{E}(|X_N|) < \infty$ and $X_n\mathbb{1}_{\{N>n\}}$ is UI, then $X_{N\wedge n}$ is UI and hence $\mathbb{E}(X_N) \geq \mathbb{E}(X_0)$.*

Theorem 2.61. *Suppose X_n is a submartingale, N is a stopping time. If the following two conditions hold:*

1. there exists $B > 0$, s.t.

$$\mathbb{E}(|X_{n+1} - X_n||\mathcal{F}_n) \leq B \quad a.s.$$

2. $\mathbb{E}(N) < \infty$

then $X_{N\wedge n}$ is UI, hence $\mathbb{E}(X_N) \geq \mathbb{E}(X_0)$.

Proof. 1.By Proposition 2.12,

$$X_{N\wedge n} = X_0 + \sum_{m=1}^{n}(X_m - X_{m-1})\mathbb{1}_{\{N\geq m\}},$$

thus

$$|X_{N\wedge n}| \leq |X_0| + \sum_{m=1}^{n}|X_m - X_{m-1}|\mathbb{1}_{\{N\geq m\}} \leq |X_0| + \sum_{m=1}^{\infty}|X_m - X_{m-1}|\mathbb{1}_{\{N\geq m\}} =: Y.$$

2. We only need to prove $\mathbb{E}(|Y|) < \infty$, then by Example 2.34, $X_{N\wedge n}$ is UI. Notice that

$$\begin{aligned}
\mathbb{E}(|X_m - X_{m-1}|\mathbb{1}_{N\geq m}) &= \mathbb{E}[\mathbb{E}(|X_m - X_{m-1}||\mathcal{F}_{m-1})\mathbb{1}_{N\geq m}] \\
&\leq \mathbb{E}(B\mathbb{1}_{N\geq m}) \\
&= B\mathbb{P}(N \geq m),
\end{aligned}$$

then by monotone convergence theorem and tail sum formula,

$$\mathbb{E}\left[\sum_{m=1}^{\infty}|X_m - X_{m-1}|\mathbb{1}_{\{N\geq m\}}\right] \leq B\sum_{m=1}^{\infty}\mathbb{P}(N \geq m) = B\mathbb{E}(N) < \infty,$$

thus $\mathbb{E}(|Y|) < \infty$. □

Application

Theorem 2.62 (Wald's equation)**.** *Let $S_0 = 0$, $S_n = \xi_1 + \cdots + \xi_n$ where ξ_i are independent with $\mathbb{E}(\xi_i) = \mu$. If N is a stopping time with $\mathbb{E}(N) < \infty$, then $\mathbb{E}(S_N) = \mu\mathbb{E}(N)$.*

Proof. Let $X_n = S_n - n\mu$, then X_n is a martingale. Noticing that

$$\mathbb{E}(|X_{n+1} - X_n||\mathcal{F}_n) = \mathbb{E}(|\xi_{n+1} - \mu|\mathcal{F}_n) = \mathbb{E}(|\xi_{n+1} - \mu|),$$

then by Theorem 2.61,

$$0 = \mathbb{E}(X_0) = \mathbb{E}(X_N) = \mathbb{E}(S_N - N\mu) = \mathbb{E}(S_N) - \mu\mathbb{E}(N).$$ □

We also show Wald's second equation here although the proof doesn't apply any optional stopping theorem.

Theorem 2.63 (Wald's second equation)**.** *Let $S_0 = 0$, $S_n = \xi_1 + \cdots + \xi_n$ where ξ_i i.i.d. with $\mathbb{E}(\xi_i) = 0$ and $\mathrm{Var}(\xi_i) = \sigma^2$. If N is a stopping time with $\mathbb{E}(N) < \infty$, then $\mathbb{E}(S_N^2) = \sigma^2\mathbb{E}(N)$.*

Proof. Let $X_n = S_n^2 - n\sigma^2$, then X_n is a martingale and so is $X_{N\wedge n}$. Thus

$$0 = \mathbb{E}(X_{N\wedge 0}) = \mathbb{E}(X_{N\wedge n}) = \mathbb{E}(S_{N\wedge n}^2) - \sigma^2\mathbb{E}(N \wedge n). \tag{1}$$

Since $N \wedge n \uparrow N$, by monotone convergence theorem, we have $\mathbb{E}(N \wedge n) \to \mathbb{E}(N)$. By (1), we have $\mathbb{E}(S_{N\wedge n}^2) = \sigma^2\mathbb{E}(N \wedge n) \le \sigma^2\mathbb{E}(N) < \infty$, thus

$$\sup_n \mathbb{E}(S_{N\wedge n}^2) \le \sigma^2\mathbb{E}(N) < \infty.$$

Since S_n is a martingale, by L^2 convergence theorem (Theorem 2.25), $S_{N\wedge n} \to S_N$ a.s. and in L^2. Therefore

$$|\|S_{N\wedge n}\|_2 - \|S_N\|_2| \le \|S_{N\wedge n} - S_N\|_2 = \sqrt{\mathbb{E}[(S_{N\wedge n} - S_N)^2]} \to 0,$$

i.e. $\mathbb{E}(S_{N\wedge n}^2) \to \mathbb{E}(S_N^2)$. Taken together, we have

$$0 = \lim_{n\to\infty}[\mathbb{E}(S_{N\wedge n}^2) - \sigma^2\mathbb{E}(N \wedge n)] = \mathbb{E}(S_N^2) - \sigma^2\mathbb{E}(N). \qquad \square$$

3 Markov Chain

3.1 Construction of Markov chain

Definition 3.1 (Transition probability)**.** Let S be a non-empty set, $\mathcal{S}$ is a σ-field on S. We call $p : S \times \mathcal{S} \to [0, 1]$ is a transition probability if

(1) For any fixed point $x \in S$, $p(x, \cdot) : \mathcal{S} \to [0, 1]$ is a probability measure on $(S, \mathcal{S})$.

(2) For any fixed set $A \in \mathcal{S}$, $p(\cdot, A) : S \to [0, 1]$ is a $\mathcal{S}$-measurable function.

Definition 3.2. Suppose $\{X_n : n \geq 1\}$ is a r.v. sequence on $(S, \mathcal{S})$ w.r.t. $\mathcal{F}_n$, i.e. $X_n \in \mathcal{F}_n$. We call X_n is a Markov chain with transition probability p, if for any $B \in \mathcal{S}$,

$$\mathbb{P}(X_{n+1} \in B | \mathcal{F}_n) = p(X_n, B).$$

Theorem 3.3. *Suppose $(S, \mathcal{S})$ is a measurable space with $S \subseteq \mathbb{R}$, p is a transition probability on $(S, \mathcal{S})$, μ is the initial distribution on $(S, \mathcal{S})$. Then we can define the probability measure $\mathbb{P}_n$ on $(S^n, \mathcal{S}^n)$ by*

$$\begin{aligned}
\mathbb{P}_n(\boldsymbol{B}) &= \int_{B_0} \left(\int_{B_1} \cdots \left(\int_{B_n} p(x_{n-1}, \mathrm{d}x_n) \right) \cdots p(x_0, \mathrm{d}x_1) \right) \mu(\mathrm{d}x_0) \\
&= \int_{B_0} \mu(\mathrm{d}x_0) \int_{B_1} p(x_0, \mathrm{d}x_1) \cdots \int_{B_n} p(x_{n-1}, \mathrm{d}x_n),
\end{aligned}$$

where $\boldsymbol{B} = B_0 \times B_1 \times \cdots \times B_n \in \mathcal{S}^n$. It is easy to show that $\mathbb{P}_n$, $n \geq 0$ are consistent.
By Kolmogorov's extension theorem, there exists a unique probability measure $\mathbb{P}_\mu$ on $(\Omega, \mathcal{F}) = (S^{\mathbb{N}}, \mathcal{S}^{\mathbb{N}})$, s.t. for any $n \in \mathbb{N}$ and $\boldsymbol{B} = B_0 \times B_1 \times \cdots \times B_n \in \mathcal{S}^n$, we have

$$\mathbb{P}_\mu(\omega : (\omega_1, \cdots, \omega_n) \in \boldsymbol{B}) = \mathbb{P}_n(\boldsymbol{B}).$$

Now we extend $\mathbb{P}_n$ from the space of finite products to that of countable products. Define

$X_n : \Omega \to S$ by

$$\omega = (\omega_0, \cdots, \omega_n, \cdots) \mapsto \omega_n.$$

$\mathcal{F}_n = \sigma(X_i : 1 \leq i \leq n)$. Then X_n is a Markov chain w.r.t $\mathcal{F}_n$ with transition probability p.

Proof. 1.We only need to show for any $B \in \mathcal{S}$,

$$\mathbb{P}(X_{n+1} \in B|\mathcal{F}_n) = p(X_n, B). \tag{1}$$

From the construction, we have

$$\mathbb{P}(X_{n+1} \in B|\mathcal{F}_n) = \mathbb{P}_\mu(\omega_{n+1} \in B|\mathcal{F}_n) = \mathbb{E}_\mu(\mathbb{1}_{\{\omega\,:\,\omega_{n+1}\in B\}}|\mathcal{F}_n).$$

To show (1) holds, only need to prove

$$\mathbb{E}_\mu(\mathbb{1}_{\{\omega\,:\,\omega_{n+1}\in B\}}|\mathcal{F}_n) = p(\omega_n, B), \quad \forall \omega = (\omega_0, \cdots, \omega_n, \cdots) \in \Omega.$$

By definition of conditional expectation, only need to show for any $A \in \mathcal{F}_n$,

$$\mathbb{E}_\mu[\mathbb{1}_{\{\omega\,:\,\omega_{n+1}\in B\}}\mathbb{1}_A] = \mathbb{E}_\mu[p(\omega_n, B)\mathbb{1}_A]. \tag{2}$$

2. We will prove (2) holds for a weaker case first, then apply $\pi - \lambda$ theorem to prove it holds for all $A \in \mathcal{F}_n$. Let $A = \{X_0 \in B_0, \cdots, X_n \in B_n\} = \{\omega\,:\,\omega_0 \in B_0, \cdots, \omega_n \in B_n\} \in \mathcal{F}_n$, then

$$\begin{aligned}
\mathbb{E}_\mu[\mathbb{1}_{\{\omega\,:\,\omega_{n+1}\in B\}}\mathbb{1}_A] &= \mathbb{E}_\mu[\mathbb{1}_{\{\omega\,:\,\omega_0\in B_0,\cdots,\omega_n\in B_n,\omega_{n+1}\in B\}}] \\
&= \mathbb{P}_\mu(\omega\,:\,\omega_0 \in B_0, \cdots, \omega_n \in B_n, \omega_{n+1} \in B) \\
&= \mathbb{P}_{n+1}[B_0 \times B_1 \times \cdots \times B_n \times B] \\
&= \int_{B_0} \mu(\mathrm{d}x_0) \int_{B_1} p(x_0, \mathrm{d}x_1) \cdots \int_{B_n} p(x_{n-1}, \mathrm{d}x_n) \int_B p(x_n, \mathrm{d}x_{n+1}) \\
&= \int_{B_0} \mu(\mathrm{d}x_0) \int_{B_1} p(x_0, \mathrm{d}x_1) \cdots \int_{B_n} p(x_{n-1}, \mathrm{d}x_n) p(x_n, B)
\end{aligned}$$

And

$$\mathbb{E}_\mu[p(\omega_n, B)\mathbb{1}_A] = \int_\Omega p(\omega_n, B)\mathbb{1}_{\{\omega\,:\,\omega_0\in B_0,\cdots,\omega_n\in B_n\}}\,\mathrm{d}\mathbb{P}_\mu.$$

To show the above two items are equal, we can do it first for the indicator, then simple functions and finally any bounded measurable function. Let $C \in \mathcal{S}$, then

$$\begin{aligned}
\mathbb{E}_\mu[\mathbb{1}_C(\omega_n)\mathbb{1}_A] &= \mathbb{E}_\mu[\mathbb{1}_{\{\omega\,:\,\omega_0\in B_0,\cdots,\omega_n\in B_n\cap C\}}] \\
&= \mathbb{P}_\mu[\omega\,:\,\omega_0 \in B_0,\cdots,\omega_n \in B_n \cap C] \\
&= \mathbb{P}_n[B_0 \times B_1 \times \cdots \times B_n \cap C] \\
&= \int_{B_0} \mu(\mathrm{d}x_0)\int_{B_1} p(x_0, \mathrm{d}x_1)\cdots\int_{B_n\cap C} p(x_{n-1}, \mathrm{d}x_n) \\
&= \int_{B_0} \mu(\mathrm{d}x_0)\int_{B_1} p(x_0, \mathrm{d}x_1)\cdots\int_{B_n} p(x_{n-1}, \mathrm{d}x_n)\mathbb{1}_C(x_n)
\end{aligned}$$

Then by linearity, for any simple function f, we have

$$\mathbb{E}_\mu[f(\omega_n)\mathbb{1}_A] = \int_{B_0} \mu(\mathrm{d}x_0)\int_{B_1} p(x_0, \mathrm{d}x_1)\cdots\int_{B_n} p(x_{n-1}, \mathrm{d}x_n)f(x_n), \tag{3}$$

By the bounded convergence theorem, (3) also holds for any bounded $\mathcal{S}$-measurable function, particularly for $p(x, B)$.

3. Now we will prove (2) holds for any $A \in \mathcal{F}_n$. Define

$$\mathcal{C}_1 = \{A \in \mathcal{F}_n : (2) \text{ holds}\},$$

easy to verify $\mathcal{C}_1$ is a λ-system. Define the set of rectangles

$$\mathcal{C}_2 = \{\{\omega\,:\,\omega_0 \in B_0,\cdots,\omega_n \in B_n\} : B_i \in \mathcal{S}, 0 \le i \le n\},$$

$\mathcal{C}_2$ is a π-system, and $\mathcal{C}_2 \subseteq \mathcal{C}_1$ by Step 2. Then by $\pi - \lambda$ theorem,

$$\mathcal{F}_n = \sigma(\mathcal{C}_2) \subseteq \mathcal{C}_1. \qquad \square$$

3.2 Properties of Markov chain

We will keep using the notations in the last section. $(\Omega, \mathcal{F}, \mathbb{P}_\mu)$ is the probability space induced by the state space $(S, \mathcal{S})$, transition probability p and initial distribution μ. $X_n(\omega) = \omega_n$ is the Markov chain.

Theorem 3.4 (Monotone class theorem)**.** *Suppose $\mathcal{A}$ is a π-system containing Ω, $\mathcal{H} \subseteq \{f : \Omega \to \mathbb{R}\}$ and satisfies*

(1) *$A \in \mathcal{A}$ implies $\mathbb{1}_A \in \mathcal{H}$*

(2) *If $f, g \in \mathcal{H}$ then $f + g \in \mathcal{H}$; if $c \in \mathbb{R}$ and $f \in \mathcal{H}$, then $cf \in \mathcal{H}$*

(3) *If $f_n \in \mathcal{H}$, $f_n \geq 0$, and $f_n \uparrow f$, then $f \in \mathcal{H}$.*

Then $\{f : \Omega \to \mathbb{R} : f < \infty, f \in \sigma(\mathcal{A})\} \subseteq \mathcal{H}$.

Proof. 1. Claim: $\mathcal{G} = \{A \in 2^\Omega : \mathbb{1}_A \in \mathcal{H}\}$ is a λ-system.
By(1), $\Omega \in \mathcal{A} \subseteq \mathcal{G}$. Suppose $A, B \in \mathcal{G}$ and $A \subseteq B$, then $\mathbb{1}_A, \mathbb{1}_B \in \mathcal{H}$, by (2), $\mathbb{1}_{B \setminus A} = \mathbb{1}_B - \mathbb{1}_A \in \mathcal{H}$, so $B \setminus A \in \mathcal{G}$. Suppose $A_n \in \mathcal{G}$ and $A_n \uparrow A$, then $\mathbb{1}_{A_n} \in \mathcal{H}$ and $\mathbb{1}_{A_n} \uparrow A$, by (3), $\mathbb{1}_A \in \mathcal{H}$, thus $A \in \mathcal{G}$.
2.Since $\mathcal{A} \subseteq \mathcal{G}$ and $\mathcal{A}$ is a π-system, by $\pi - \lambda$ theorem, $\sigma(\mathcal{A}) \subseteq \mathcal{G}$.
3. Thus for any $A \in \sigma(\mathcal{A})$, $\mathbb{1}_A \in \mathcal{H}$, and by (2), any simple function $f \in \sigma(\mathcal{A})$ belongs to $\mathcal{H}$. For any bounded $\sigma(\mathcal{A})$-measurable function f, there is a non-negative f_n s.t. $f_n \uparrow f$, thus by (3), $f \in \mathcal{H}$. $\square$

Proposition 3.5. *Suppose X_n is a Markov chain with transition probability p.*

1. For any bounded $\mathcal{S}$-measurable f,

$$\mathbb{E}(f(X_{n+1})|\mathcal{F}_n) = \int_S p(X_n, \mathrm{d}y)f(y) \tag{1}$$

2. For any bounded $\mathcal{S}$-measurable f_m,

$$\mathbb{E}\left[\prod_{m=0}^{n} f_m(X_m)\right] = \int_S \mu(\mathrm{d}x_0)f_0(x_0)\int_S p(x_0, \mathrm{d}x_1)f_1(x_1)\cdots\int_S p(x_{n-1}, \mathrm{d}x_n)f_n(x_n). \tag{2}$$

Proof. 1. First $\mathcal{S}$ is a σ-field, thus a π-system. Define $\mathcal{H} = \{f : \text{Eq(1) holds for } \mathcal{A} := \mathcal{S}\}$, then $\mathcal{H}$ satisfies the three conditions in Theorem 3.4: i) for any $A \in \mathcal{S}$,

$$\mathbb{E}(\mathbb{1}_A(X_{n+1})|\mathcal{F}_n) = \mathbb{E}(X_{n+1} \in A|\mathcal{F}_n) = p(X_n, A) = \int_S p(X_n, \mathrm{d}y)\mathbb{1}_A;$$

ii) obviously iii) by monotone convergence theorem. Thus $\mathcal{H}$ contains all bounded $\mathcal{S}$-measurable function.

2. First, (2) holds for $n = 0$, since

$$\mathbb{E}(f_0(X_0)) = \int_S f_0(x_0)\mu(\mathrm{d}x_0).$$

Suppose (2) holds for $n-1$, then by the property of conditional expectation,

$$
\begin{aligned}
\mathbb{E}\left[\prod_{m=0}^{n} f_m(X_m)\right] &= \mathbb{E}\left[\mathbb{E}\left(\prod_{m=0}^{n} f_m(X_m)|\mathcal{F}_{n-1}\right)\right] \\
&= \mathbb{E}\left[\prod_{m=0}^{n-1} f_m(X_m)\mathbb{E}\left(f_n(X_n)|\mathcal{F}_{n-1}\right)\right] \qquad \text{(since } f(X_m) \in \mathcal{F}_{n-1} \text{ for } m \leq n-1\text{)} \\
&= \mathbb{E}\left[\prod_{m=0}^{n-1} f_m(X_m) \int_S p(X_{n-1}, \mathrm{d}y) f_n(y)\right] \\
&= \mathbb{E}\left[\prod_{m=0}^{n-2} f_m(X_m) f_{n-1}(X_{n-1}) g(X_{n-1})\right] \qquad \text{(let the integral be } g(X_{n-1})\text{)} \\
&= \int_S \mu(\mathrm{d}x_0) f_0(x_0) \int_S p(x_0, \mathrm{d}x_1) f_1(x_1) \cdots \int_S p(x_{n-2}, \mathrm{d}x_{n-1}) f_{n-1}(x_{n-1}) g(x_{n-1}) \\
&= \int_S \mu(\mathrm{d}x_0) f_0(x_0) \int_S p(x_0, \mathrm{d}x_1) f_1(x_1) \cdots \int_S p(x_{n-1}, \mathrm{d}x_n) f_n(x_n)
\end{aligned}
$$

Thus (2) holds for all $n \in \mathbb{N}$. □

Proposition 3.6. *If $f : S^{n+1} \to \mathbb{R}$ is bounded and $\mathcal{S}^{n+1}$-measurable, then*

$$
\mathbb{E}[f(X_0, X_1, \cdots, X_n)] = \int_S f(x_0, x_1, \cdots, x_n)\mu(\mathrm{d}x_0) \int_S p(x_0, \mathrm{d}x_1) \cdots \int_S p(x_{n-1}, \mathrm{d}x_n). \tag{1}
$$

Proof. Let $\mathcal{A} = \{\text{rectangles in } \mathcal{S}^{n+1}\}$, $\mathcal{H} = \{\text{bounded and } \mathcal{A}\text{-measurable } f \text{ s.t. (1) holds}\}$. We will show three conditions in monotone class theorem holds. i) for rectangle $A = A_0 \times$

$A_2 \times \cdots \times A_n \in \mathcal{A}$, we have

$$\begin{aligned}\mathbb{E}[\mathbb{1}_A(X_0, X_1, \cdots, X_n)] &= \mathbb{E}[\prod_{i=0}^{n} \mathbb{1}_{A_i}(X_i)] \\ &= \mathbb{P}(X_0 \in A_0, X_1 \in A_1, \cdots, X_n \in A_n) \\ &= \int_{A_0} \mu(\,\mathrm{d}x_0) \int_{A_1} p(x_0, \,\mathrm{d}x_1) \cdots \int_{A_n} p(x_{n-1}, \,\mathrm{d}x_n) \\ &= \int_S \mathbb{1}_{A_0}(x_0)\mu(\,\mathrm{d}x_0) \int_S \mathbb{1}_{A_1}(x_1) p(x_0, \,\mathrm{d}x_1) \cdots \int_S \mathbb{1}_{A_n}(x_n) p(x_{n-1}, \,\mathrm{d}x_n) \\ &= \int_S \mathbb{1}_A(x_0, x_1, \cdots, x_n)\mu(\,\mathrm{d}x_0) \int_S p(x_0, \,\mathrm{d}x_1) \cdots \int_S p(x_{n-1}, \,\mathrm{d}x_n)\end{aligned}$$

thus $\mathbb{1}_A \in \mathcal{H}$. ii) obviously iii) by monotone convergence theorem. Thus by monotone class theorem, $\mathcal{H}$ contains all bounded and $\sigma(\mathcal{A}) = \mathcal{S}^{n+1}$-measurable functions. □

Remark. The second result in Proposition 3.5 can be a special case of this proposition.

Definition 3.7. Suppose $\Omega = \mathcal{S}^{\mathbb{N}}$, $n \in \mathbb{N}$, we call $\theta_n : \Omega \to \Omega$ a shift operator if

$$\omega = (\omega_0, \omega_1, \cdots) \mapsto (\omega_n, \omega_{n+1}, \cdots).$$

Theorem 3.8 (Markov property)**.** *Let $Y : \Omega \to \mathbb{R}$ be bounded and measurable, then*

$$\mathbb{E}_\mu(Y \circ \theta_m | \mathcal{F}_m) = \mathbb{E}_{X_m} Y.$$

Remark. Here $\mathbb{E}_{X_m} Y$ is a r.v., if $X_m = x$, $\mathbb{E}_{X_m} Y = \mathbb{E}_x Y$ which takes $\mu = \delta(x)$ in $\mathbb{E}_\mu Y$.

Proof. 1.By the definition of conditional expectation, we only need to show for any $A \in \mathcal{F}_m$,

$$\mathbb{E}_\mu(Y \circ \theta_m \mathbb{1}_A) = \mathbb{E}_\mu(\mathbb{E}_{X_m} Y \mathbb{1}_A). \tag{1}$$

2. Consider A is a rectangle first, i.e. $A = \{\omega : \omega_0 \in A_0, \omega_1 \in A_1, \cdots, \omega_m \in A_m\}$. For

$k = 0, 1, \cdots, n$, let $g_k : S \to \mathbb{R}$ be bounded and measurable and

$$Y(\omega) = \prod_{k=0}^{n} g_k(\omega_k) = \prod_{k=0}^{n} g_k \circ X_k(\omega). \tag{2}$$

Define

$$f_k = \begin{cases} \mathbb{1}_{A_k} & 0 \le k < m \\ \mathbb{1}_{A_k} g_0 & k = m \\ g_{k-m} & m < k \le m+n, \end{cases}$$

by Proposition 3.5,

$$\mathbb{E}_\mu \left[\prod_{k=0}^{m+n} f_k(X_k) \right] = \int_S \mu(\mathrm{d}x_0) f_0(x_0) \int_S p(x_0, \mathrm{d}x_1) f_1(x_1) \cdots \int_S p(x_{m+n-1}, \mathrm{d}x_{m+n}) f_{m+n}(x_{m+n})$$

For the lefthand side,

$$\begin{aligned} LHS &= \mathbb{E} \left[\prod_{k=m}^{m+n} g_{k-m}(X_k) \prod_{k=0}^{m} \mathbb{1}_{A_k} \right] \\ &= \mathbb{E} \left[\prod_{k=m}^{m+n} g_{k-m}(X_k) \mathbb{1}_A \right] \\ &= \mathbb{E} \left[\prod_{k=0}^{n} g_k(X_{k+m}) \mathbb{1}_A \right] \\ &= \mathbb{E}_\mu(Y \circ \theta_m \mathbb{1}_A). \end{aligned}$$

For the righthand side,

$$\begin{aligned}
RHS &= \int_{A_0} \mu(\mathrm{d}x_0) \int_{A_1} p(x_0, \mathrm{d}x_1) \cdots \int_{A_m} p(x_{m-1}, \mathrm{d}x_m) g_0(x_m) \int_S p(x_m, \mathrm{d}x_{m+1}) g_1(x_{m+1}) \cdots \\
&\quad \int_S p(x_{m+n-1}, \mathrm{d}x_{m+n}) g_n(x_{m+n}) \\
&= \int_{A_0} \mu(\mathrm{d}x_0) \int_{A_1} p(x_0, \mathrm{d}x_1) \cdots \int_{A_m} p(x_{m-1}, \mathrm{d}x_m) \varphi(x_m) \\
&= \mathbb{E}_\mu(\varphi(X_m) \mathbb{1}_A), \qquad \text{(by Proposition 3.5)}
\end{aligned}$$

where

$$\begin{aligned}
\varphi(x_m) &= g_0(x_m) \int_S p(x_m, \mathrm{d}x_{m+1}) g_1(x_{m+1}) \cdots \int_S p(x_{m+n-1}, \mathrm{d}x_{m+n}) g_n(x_{m+n}) \\
&= g_0(x_m) \int_S p(x_m, \mathrm{d}x_1) g_1(x_1) \cdots \int_S p(x_{n-1}, \mathrm{d}x_n) g_n(x_n) \\
&= \mathbb{E}_{x_m} \left[\prod_{k=0}^{n} g_k(X_k) \right] \qquad \text{(by Proposition 3.5)} \\
&= \mathbb{E}_{x_m} Y
\end{aligned}$$

Replace x with r.v. X_m then we have

$$\varphi(X_m) = \mathbb{E}_{X_m} Y,$$

Thus $RHS = \mathbb{E}_\mu(\mathbb{E}_{X_m} Y \mathbb{1}_A)$. We obtain (1) holds in this case.

3. For Y defined by (2), we will prove (1) holds for any $A \in \mathcal{F}_m$. Let

$$\mathcal{C}_1 = \{A \in 2^\Omega : (1) \text{ holds on } A\},$$

easy to verify that $\mathcal{A}$ is a λ-system. Define

$$\mathcal{C}_2 = \{\text{rectangles} \in \mathcal{F}_m\},$$

$\mathcal{C}_2$ is a π-system and $\mathcal{C}_2 \subseteq \mathcal{C}_1$, then by $\pi - \lambda$ theorem, $\mathcal{F}_m = \sigma(\mathcal{C}_2) \subseteq \mathcal{C}_1$, thus (1) holds for any $A \in \mathcal{F}_m$.

4. The last step is to prove (1) holds for all bounded and measurable Y. Fix $A \in \mathcal{F}_m$, define

$$\mathcal{H} = \{\text{bounded and measurable } Y\text{: (1) holds}\},$$

by Step 3, any form of Y defined by (2) belongs to $\mathcal{H}$. And define

$$\mathcal{A} = \{\text{rectangles} \in \mathcal{F} = \mathcal{S}^{\mathbb{N}}\},$$

$\mathcal{A}$ is a π-system and $\Omega = S^{\mathbb{N}} \in \mathcal{A}$. Furthermore $\mathcal{H}$ satisfies all three conditions in Theorem 3.4: (i)for any $A = \{\omega : \omega_0 \in A_0, \omega_1 \in A_1, \cdots, \omega_k \in A_k\} \in \mathcal{A}$,

$$\mathbb{1}_A = \prod_{i=1}^{k} \mathbb{1}_{A_i} \in \mathcal{H},$$

(ii) obviously (iii) by monotone convergence theorem. Thus by Theorem 3.4, $\mathcal{H}$ contains all bounded and $\mathcal{F} = \sigma(\mathcal{A})$-measurable functions. □

Theorem 3.9. *For any bounded function $Y \in \sigma(X_k, k \geq n)$,*

$$\mathbb{E}_\mu(Y|\mathcal{F}_n) = \mathbb{E}_\mu(Y|X_n). \tag{1}$$

Proof. Since $Y \in \sigma(X_k, k \geq n)$, we have $Y \circ \theta_{-n}$ is bounded and $\mathcal{F}$-measurable, and

$$Y = (Y \circ \theta_{-n}) \circ \theta_n.$$

By Markov property (Theorem 3.8),

$$\mathbb{E}_\mu(Y|\mathcal{F}_n) = \mathbb{E}_\mu[(Y \circ \theta_{-n}) \circ \theta_n|\mathcal{F}_n] = \mathbb{E}_{X_n}(Y \circ \theta_{-n}),$$

take conditional expectation on X_n (i.e. $\sigma(X_n)$), we have

$$\mathbb{E}_\mu[\mathbb{E}_\mu(Y|\mathcal{F})|X_n] = \mathbb{E}_\mu[\mathbb{E}_{X_n}(Y\circ\theta_{-n})|X_n],$$

the left side is $\mathbb{E}_\mu(Y|X_n)$ since $\sigma(X_n)\subseteq\mathcal{F}_n$, the right side is $\mathbb{E}_{X_n}(Y\circ\theta_{-n}) = \mathbb{E}_\mu(Y|\mathcal{F}_n)$ since $\mathbb{E}_{X_n}(Y\circ\theta_{-n})\in\sigma(X_n)$, thus

$$\mathbb{E}_\mu(Y|\mathcal{F}_n) = \mathbb{E}_\mu(Y|X_n).$$

□

Corollary 3.10. *Let $A\in\mathcal{F}_n$, $B\in\sigma(X_k, k\geq n)$, then*

$$\mathbb{P}_\mu(A\cap B|X_n) = \mathbb{P}_\mu(A|X_n)\mathbb{P}_\mu(B|X_n).$$

Proof.

$$\begin{aligned}
\mathbb{P}_\mu(A\cap B|X_n) &= \mathbb{E}_\mu(\mathbb{1}_{A\cap B}|X_n)\\
&= \mathbb{E}_\mu[\mathbb{E}_\mu(\mathbb{1}_A\mathbb{1}_B|\mathcal{F}_n)|X_n] \qquad (\text{Since } \sigma(X_n)\subseteq\mathcal{F}_n)\\
&= \mathbb{E}_\mu[\mathbb{1}_A\mathbb{E}_\mu(\mathbb{1}_B|\mathcal{F}_n)|X_n] \qquad (\text{Since } \mathbb{1}_A\in\mathcal{F}_n)\\
&= \mathbb{E}_\mu[\mathbb{1}_A\mathbb{E}_\mu(\mathbb{1}_B|X_n)|X_n] \qquad (\text{By Theorem 3.9})\\
&= \mathbb{E}_\mu(\mathbb{1}_A|X_n)\mathbb{E}_\mu(\mathbb{1}_B|X_n)\\
&= \mathbb{P}_\mu(A|X_n)\mathbb{P}_\mu(B|X_n).
\end{aligned}$$

□

Remark. The above result shows that the past and future are conditionally independent given the present.

Theorem 3.11 (Strong Markov property)**.** *Suppose N is a stopping time, define*

$$\mathcal{F}_N = \{A : A\cap\{N=n\}\in\mathcal{F}_n\ \forall n\in\mathbb{N}\}.$$

For $n \in \mathbb{N}$, suppose $Y_n : \Omega \to \mathbb{R}$ is measurable and $\sup_n |Y_n| \le M$. Then on $\{N < \infty\}$,

$$\mathbb{E}_\mu(Y_N \circ \theta_N | \mathcal{F}_N) = \mathbb{E}_{X_N} Y_N.$$

Remark. $\mathbb{E}_{X_N} Y_N(\omega)$ is a r.v. and when $N(\omega) = n$, $X_N(\omega) = x$, it has value $\mathbb{E}_x Y_n$.

Proof. We want to show for any $A \in \mathcal{F}_N$,

$$\mathbb{E}[Y_N \circ \theta_N \mathbb{1}_{A \cap \{N<\infty\}}] = \mathbb{E}[\mathbb{E}_{X_N} Y_N \mathbb{1}_{A\cap\{N<\infty\}}].$$

Since

$$\{N < \infty\} = \bigsqcup_{n=0}^{\infty} \{N = n\},$$

we have

$$\begin{aligned}
\mathbb{E}[Y_N \circ \theta_N \mathbb{1}_{A\cap\{N<\infty\}}] &= \mathbb{E}[Y_N \circ \theta_N \mathbb{1}_{A \cap \bigsqcup_{n=0}^{\infty}\{N=n\}}] \\
&= \mathbb{E}[Y_N \circ \theta_N \mathbb{1}_{\bigsqcup_{n=0}^{\infty} A\cap\{N=n\}}] \\
&= \sum_{n=0}^{\infty} \mathbb{E}[Y_n \circ \theta_n \mathbb{1}_{A\cap\{N=n\}}] \\
&= \sum_{n=0}^{\infty} \mathbb{E}[\mathbb{E}_{X_n} Y_n \mathbb{1}_{A\cap\{N=n\}}] \qquad \text{(by Theorem 3.8 and } A \cap \{N = n\} \in \mathcal{F}_n) \\
&= \mathbb{E}[\mathbb{E}_{X_N} Y_N \mathbb{1}_{A\cap\{N<\infty\}}].
\end{aligned}$$

□

Theorem 3.12 (Reflection principle)**.** *Let $\{X_k : k \ge 1\}$ be a sequence of i.i.d. r.v. with $\mathbb{P}(X_k > 0) = \mathbb{P}(X_k < 0)$. Let $S_0 = 0$, and for $n \ge 1$, $S_n = \sum_{k=1}^n X_k$. For any $a > 0$, we have*

$$\mathbb{P}\left(\sup_{1\le m\le n} S_m \ge a\right) \le 2\mathbb{P}(S_n \ge a).$$

Proof. Let $N = \inf\{m \le n : S_m \ge a\}$, define $\inf \varnothing = \infty$. Notice that

$$\{N \le n\} = \{S_m \ge a \text{ for some } m \le n\} = \{\sup_{m \le n} S_m \ge a\},$$

so

$$\mathbb{P}_0(N \le n) = \mathbb{P}_0(\sup_{m \le n} S_m \ge a).$$

For $m \le n$, define

$$Y_m = \mathbb{1}_{\{S_{n-m} \ge a\}},$$

then $Y_m \circ \theta_m = \mathbb{1}_{\{S_n \ge a\}}$. On $\{N < \infty\} = \{N \le n\}$,

$$Y_N \circ \theta_N(\omega) = \mathbb{1}_{\{S_n \ge a\}}, \tag{1}$$

and by the strong Markov property,

$$\mathbb{E}_0(Y_N \circ \theta_N | \mathcal{F}_N) = \mathbb{E}_{S_N}(Y_N). \tag{2}$$

If $x \ge a$, then for $m \le n$,

$$\mathbb{E}_x(Y_m) = \mathbb{P}_x(S_{n-m} \ge a) \ge \mathbb{P}_x(S_{n-m} \ge x) \ge \frac{1}{2},$$

thus on $\{N \le n\}$,

$$\mathbb{E}_{S_N}(Y_N) \ge \frac{1}{2}.$$

Since $\{N \le n\} \in \mathcal{F}_N$, applying the definition of conditional expectation to (2), we have

$$\mathbb{E}_0(Y_N \circ \theta_N \mathbb{1}_{\{N \le n\}}) = \mathbb{E}_0[\mathbb{E}_{S_N}(Y_N)\mathbb{1}_{\{N \le n\}}] \ge \mathbb{E}_0[\frac{1}{2}\mathbb{1}_{\{N \le n\}}] = \frac{1}{2}\mathbb{P}_0(N \le n),$$

and by (1),

$$\mathbb{E}_0(Y_N \circ \theta_N \mathbb{1}_{\{N \le n\}}) = \mathbb{E}_0(\mathbb{1}_{\{S_n \ge a\} \cap \{N \le n\}}) = \mathbb{P}_0(\{S_n \ge a\} \cap \{N \le n\}) = \mathbb{P}_0(S_n \ge a),$$

since $\{S_n \ge a\} \subseteq \{N \le n\}$. □

3.3 Basic concepts of Markov chain on a countable state space

Now consider the Markov chain X_n in a countable state space S.

3.3.1 Multistep transition probability

Lemma 3.13. *For any $i_0, i_1, \cdots, i_n \in S$,*

$$\mathbb{P}_\mu(X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n) = \mu(i_0) \prod_{m=1}^{n} p(i_{m-1}, i_m).$$

Proof. By definition, let $B_0 = \{i_0\}, B_1 = \{i_1\}, \cdots B_n = \{i_n\}$, then

$$\begin{aligned}\mathbb{P}_\mu(X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n) &= \int_{B_0} \mu(\mathrm{d}x_0) \int_{B_1} p(x_0, \mathrm{d}x_1) \cdots \int_{B_n} p(x_{n-1}, \mathrm{d}x_n) \\ &= \mu(i_0) \prod_{m=1}^{n} p(i_{m-1}, i_m).\end{aligned}$$

□

Definition 3.14. For any $x, y \in S$, $n \in \mathbb{Z}_+$, define $p^n(x, y)$ is the probability of starting from x and getting to y in time n, i.e.

$$p^n(x, y) = \mathbb{P}_x(X_n = y).$$

Lemma 3.15. *For any $x, y \in S$, $n \in \mathbb{Z}_+$,*

$$p^n(x,y) = \sum_{x_1,\cdots,x_{n-1}\in S} p(x,x_1)p(x_1,x_2)\cdots p(x_{n-1},y).$$

Proof. By definition, then

$$\begin{aligned}\mathbb{P}_x(X_n = y) &= \int_S \delta_x(\,\mathrm{d}x_0) \int_S p(x_0,\,\mathrm{d}x_1)\cdots\int_S p(x_{n-2},\,\mathrm{d}x_{n-1})\int_{\{y\}} p(x_{n-1},\,\mathrm{d}x_n)\\ &= \sum_{x_0\in S}\delta_x(x_0)\sum_{x_1\in S}p(x_0,x_1)\cdots\sum_{x_{n-1}\in S}p(x_{n-2},x_{n-1})p(x_{n-1},y)\\ &= \sum_{x_1\in S}p(x,x_1)\cdots\sum_{x_{n-1}\in S}p(x_{n-2},x_{n-1})p(x_{n-1},y)\\ &= \sum_{x_1,\cdots,x_{n-1}\in S}p(x,x_1)p(x_1,x_2)\cdots p(x_{n-1},y).\end{aligned}$$

□

Lemma 3.16. *For any $x, y \in S$, $k \in \mathbb{N}$, $n \in \mathbb{Z}_+$*

$$\mathbb{P}_\mu(X_{k+n} = y | X_k = x) = p^n(x,y).$$

Proof. By Markov property and Theorem 3.9,

$$\mathbb{P}_\mu(X_{k+n} = y|X_k) = \mathbb{P}_\mu(X_{k+n} = y|\mathcal{F}_k) = \mathbb{E}_\mu(\mathbb{1}_{\{X_n=y\}}\circ\theta_k|\mathcal{F}_k) = \mathbb{E}_{X_k}(\mathbb{1}_{\{X_n=y\}}) = \mathbb{P}_{X_k}(X_n = y),$$

when $X_k = x$, we have

$$\mathbb{P}_\mu(X_{k+n} = y|X_k = x) = \mathbb{P}_x(X_n = y) = p^n(x,y).$$ □

Lemma 3.17 (distribution at time n)**.** *For any $j \in S$,*

$$\mathbb{P}_\mu(X_n = j) = \sum_{i \in S} \mu(i) p^n(i, j).$$

Proof. By Proposition 3.5,

$$\begin{aligned}\mathbb{P}_\mu(X_n = j) &= \mathbb{E}_\mu[\mathbb{E}_\mu(\mathbb{1}_{\{X_n = j\}} | X_0)] \\ &= \mathbb{E}_\mu[p^n(X_0, j)] \\ &= \int_S p^n(x_0, j) \mu(\,\mathrm{d}x_0) \\ &= \sum_{x_0 \in S} p^n(x_0, j) \mu(x_0)\end{aligned}$$

□

Theorem 3.18 (Chapman-Kolmogorov equation)**.** *Suppose $x, z \in S$, then*

$$\mathbb{P}_x(X_{m+n} = z) = \sum_{y \in S} \mathbb{P}_x(X_m = y) \mathbb{P}_y(X_n = z),$$

i.e.

$$p^{m+n}(x, z) = \sum_{y \in S} p^m(x, y) p^n(y, z).$$

Proof.

$$\begin{aligned}
\mathbb{P}_x(X_{m+n} = z) &= \mathbb{E}_x(\mathbb{1}_{\{X_{m+n}=z\}}) \\
&= \mathbb{E}_x[\mathbb{E}_x(\mathbb{1}_{\{X_{m+n}=z\}}|\mathcal{F}_m)] \\
&= \mathbb{E}_x[\mathbb{E}_x(\mathbb{1}_{\{X_n=z\}} \circ \theta_m|\mathcal{F}_m)] \\
&= \mathbb{E}_x[\mathbb{E}_{X_m}(\mathbb{1}_{\{X_n=z\}})] \qquad \text{(by Theorem 3.8)} \\
&= \mathbb{E}_x[\mathbb{P}_{X_m}(X_n = z)] \\
&= \sum_y \mathbb{P}_x(X_m = y)\mathbb{P}_y(X_n = z)
\end{aligned}$$

□

3.3.2 Time of the k-th return

Definition 3.19. For any $x, y \in S$,

1. Define T_y^k to be the time of k-th visit to y, i.e. $T_y^0 = 0$,

$$T_y^k = \inf\{n \in \mathbb{N} : n > T_y^{k-1}, X_n = y\}$$

   and $\inf \varnothing = \infty$.

2. Denote $T_y = T_y^1 > 0$.

3. Define ρ_{xy} is the probability of starting from x and getting to y eventually, i.e.

$$\rho_{xy} = \mathbb{P}_x(T_y < \infty).$$

Lemma 3.20. *Let $x, y, z \in S$, then*

$$\rho_{xz} \geq \rho_{xy}\rho_{yz}.$$

Proof. We observe that if a chain can initiate from state x and eventually reach state y, and it can also initiate from state y and eventually reach state z, then it implies that the chain can initiate from state x and eventually reach state z. So on $\{X_0 = x\}$

$$\{T_y < \infty\} \cap \{T_z \circ \theta_{T_y} < \infty\} \subseteq \{T_z < \infty\},$$

thus

$$\begin{aligned}
\rho_{xz} &= \mathbb{P}_x(T_z < \infty) \\
&\geq \mathbb{P}_x(T_z \circ \theta_{T_y} < \infty, T_y < \infty) \\
&= \mathbb{E}_x[\mathbb{1}_{\{T_z<\infty\}} \circ \theta_{T_y} \mathbb{1}_{\{T_y<\infty\}}] \\
&= \mathbb{E}_x[\mathbb{E}_y(\mathbb{1}_{\{T_z<\infty\}})\mathbb{1}_{\{T_y<\infty\}}] \qquad (*) \\
&= \mathbb{E}_x[\mathbb{P}_y(T_z < \infty)\mathbb{1}_{\{T_y<\infty\}}] \\
&= \mathbb{P}_y(T_z < \infty)\mathbb{P}_x(T_y < \infty) \\
&= \rho_{yz}\rho_{xy},
\end{aligned}$$

where $(*)$ holds because we can apply the strong Markov property (Theorem 3.11) to get

$$\mathbb{E}_x(\mathbb{1}_{\{T_z<\infty\}} \circ \theta_{T_y} | \mathcal{F}_{T_y}) = \mathbb{E}_{X_{T_y}}(\mathbb{1}_{\{T_z<\infty\}}) = \mathbb{E}_y(\mathbb{1}_{\{T_z<\infty\}}),$$

then use $\{T_y < \infty\} \in \mathcal{F}_{T_y}$ and the definition of conditional expectation. □

Lemma 3.21. *For $x, y \in S$ and $x \neq y$, TFAE,*

1. $\rho_{xy} > 0$
2. $p^n(x, y) > 0$ *for some* $n \geq 1$.
3. *there exists* $i_0 = x, i_1, \cdots, i_n = y$ *s.t.* $p(i_{r-1}, i_r) > 0$ *for any* $r = 1, \cdots, n$.

Proof. 1 $\Longrightarrow$ 2.Suppose $\rho_{xy} > 0$. Then

$$0 < \rho_{xy} = \mathbb{P}_x(T_y < \infty) = \mathbb{P}_x(\bigsqcup_{n=1}^{\infty}\{T_y = n\}) = \sum_{n=1}^{\infty}\mathbb{P}(T_y = n) \leq \sum_{n=1}^{\infty}\mathbb{P}(X_n = y) = \sum_{n=1}^{\infty}p^n(x,y).$$

2 $\Longrightarrow$ 3. By Lemma 3.15,

$$p^n(x,y) = \sum_{x_1,\cdots,x_{n-1}\in S} p(x,x_1)p(x_1,x_2)\cdots p(x_{n-1},y),$$

thus $p^n(x,y) > 0$ implies $p(x,x_1)p(x_1,x_2)\cdots p(x_{n-1},y) > 0$ for some $x, x_1, \cdots, x_{n-1}, y$.

3 $\Longrightarrow$ 1. Since $\{X_n = y\} \subseteq \{T_y < \infty\}$,

$$p^n(x,y) = \mathbb{P}_x(X_n = y) \leq \mathbb{P}_x(T_y < \infty) = \rho_{xy},$$

and by Lemma 3.15,

$$p^n(x,y) \geq p(x,x_1)p(x_1,x_2)\cdots p(x_{n-1},y) > 0,$$

thus $\rho_{xy} > 0$. □

Lemma 3.22. *$T_y^k : \Omega \to \mathbb{N}$ is a stopping time.*

Proposition 3.23. $\mathbb{P}_x(T_y^k < \infty) = \rho_{xy}\rho_{yy}^{k-1}$.

Proof. 1.We will prove it by induction. For $k = 1$, $\mathbb{P}_x(T_y^1) = \mathbb{P}_x(T_y) = \rho_{xy}$. Suppose it holds for some $k \geq 2$. We will prove it also holds for $k+1$.

2.Define

$$Y(\omega) = \mathbb{1}_{\{T_y<\infty\}} = \begin{cases} 1 & \text{if } \omega_n = y \text{ for some } n \\ 0 & \text{otherwise} \end{cases}$$

Let $N = T_y^k$, then $Y \circ \theta_N = 1$ if and only if $T_y^{k+1} < \infty$ (because $\theta_N(\omega) = (\omega_N, \omega_{N+1}, \cdots)$ and if the $(k+1)$-th return to y happens after N, there must be a $n > N$ s.t. $\omega_n = y$). Thus

$$Y \circ \theta_N = \mathbb{1}_{\{T_y^{k+1} < \infty\}}.$$

3.By the Strong Markov property (Theorem 3.11), on $\{N < \infty\}$,

$$\mathbb{E}_x(Y \circ \theta_N | \mathcal{F}_N) = \mathbb{E}_{X_N} Y.$$

4.Since $N = T_y^k$, on $\{N < \infty\}$, $X_N = y$, then

$$\mathbb{E}_{X_N} Y = \mathbb{E}_y Y = \mathbb{E}(\mathbb{1}_{\{T_y < \infty\}}) = \mathbb{P}_y(T_y < \infty) = \rho_{yy}.$$

5.Therefore

$$\begin{aligned}
\mathbb{P}_x(T_y^{k+1} < \infty) &= \mathbb{P}_x(\{T_y^{k+1} < \infty\}\mathbb{1}_{\{N<\infty\}}) + \mathbb{P}_x(\{T_y^{k+1} < \infty\}\mathbb{1}_{\{N=\infty\}}) \\
&= \mathbb{P}_x(\{T_y^{k+1} < \infty\}\mathbb{1}_{\{N<\infty\}}) \qquad (\text{since } \{T_y^{k+1} < \infty\} \subseteq \{T_y^k < \infty\} = \{N = \infty\}^c) \\
&= \mathbb{E}_x[Y \circ \theta_N \mathbb{1}_{\{N<\infty\}}] \qquad (\text{by Step 2}) \\
&= \mathbb{E}_x[\mathbb{E}_{X_N} Y \mathbb{1}_{\{N<\infty\}}] \qquad (\text{by Step 3}) \\
&= \mathbb{E}_x[\rho_{yy} \mathbb{1}_{\{N<\infty\}}] \qquad (\text{by Step 4}) \\
&= \rho_{yy} \mathbb{P}_x(T_y^k < \infty) \\
&= \rho_{xy} \rho_{yy}^k. \qquad (\text{by the induction hypothesis in Step 1})
\end{aligned}$$

□

3.4 Exit distribution and exit time

Definition 3.24.
- For any $C \subseteq S$, define the hitting time on C as

$$V_C = \inf\{n \geq 0 : X_n \in C\}.$$

- For any $A, B \subseteq S$ with $A \cap B = \varnothing$, define the probability of exit at set A as $\mathbb{P}_x(V_A < V_B)$.

Lemma 3.25. *Suppose $C \subseteq S$, $S \setminus C$ is finite, and for any $x \in S \setminus C$, $\mathbb{P}_x(V_C < \infty) > 0$. Then*

1. *there exists $0 < N < \infty$ and $0 < \varepsilon \leq 1$, s.t. for any $x \in S \setminus C$ and $k \in \mathbb{Z}_+$,*

$$\mathbb{P}_x(T_C > kN) \leq (1 - \varepsilon)^k. \tag{1}$$

2. *$\mathbb{P}_x(T_C < \infty) = 1$ for any $x \in S - C$.*

Proof. 1. Since for any $x \in S - C$, $\mathbb{P}_x(T_C < \infty) > 0$, we can find $N_x > 0$, s.t.

$$\mathbb{P}_x(T_C \leq N_x) > 0,$$

otherwise

$$\mathbb{P}_x(T_C < \infty) > 0 = \mathbb{P}_x(\bigcup_{n=1}^{\infty}\{T_C \leq n\}) \leq \sum_{n=1}^{\infty}\mathbb{P}_x(T_C \leq n) = 0.$$

Let $N = \max_{x \in S-C} N_x$, then

$$\mathbb{P}_x(T_C \leq N) > 0, \quad \forall x \in S - C.$$

And let $\varepsilon = \min_{x \in S-C} \mathbb{P}_x(T_C \leq N)$, then

$$\mathbb{P}_x(T_C \leq N) \geq \varepsilon, \quad \forall x \in S - C.$$

Thus

$$\mathbb{P}_x(T_C > N) = \mathbb{P}_x(\{T_C \leq N\}^c) = 1 - \mathbb{P}_x(T_C \leq N) \leq 1 - \varepsilon, \quad \forall x \in S - C, \tag{2}$$

i.e. we find N and ε s.t. (1) holds for $k = 1$. Suppose that (1) also holds for k, we will prove the case $k+1$. By the Markov property,

$$\mathbb{E}_x(\mathbb{1}_{\{T_C > N\}} \circ \theta_{kN} | \mathcal{F}_{kN}) = \mathbb{E}_{X_{kN}} \mathbb{1}_{\{T_C > N\}} = \mathbb{P}_{X_{kN}}(T_C > N), \tag{3}$$

thus

$$\begin{aligned}
\mathbb{P}_x(T_C > (k+1)N) &= \mathbb{E}_x[(\mathbb{1}_{\{T_C > N\}} \circ \theta_{kN}) \cdot \mathbb{1}_{\{T_C > kN\}}] \\
&= \mathbb{E}_x[\mathbb{P}_{X_{kN}}(T_C > N) \cdot \mathbb{1}_{\{T_C > kN\}}] \qquad \text{(by (3) and } \{T_C > kN\} \in \mathcal{F}_{kN}) \\
&\leq (1-\varepsilon)\mathbb{E}_x(\mathbb{1}_{\{T_C > kN\}}) \qquad \text{(by (2) and } X_{kN} \in S - C) \\
&= (1-\varepsilon)\mathbb{P}_x(T_C > kN) \\
&\leq (1-\varepsilon)^{k+1}. \qquad \text{(by induction hypothesis)}
\end{aligned}$$

By induction, we have shown (1) holds for any $k \in \mathbb{Z}_+$.

2. Let $k \to \infty$ in (1), we have $\mathbb{P}_x(T_C = \infty) = 0$, i.e. $\mathbb{P}_x(T_C < \infty) = 1$. □

Theorem 3.26 (Exit distribution)**.** *Suppose $A, B \subseteq S$ with $A \cap B = \varnothing$, $S \setminus (A \cup B)$ is finite, $\mathbb{P}_x(V_A \wedge V_B < \infty) > 0$ for all $x \in S \setminus (A \cup B)$. Then $h(x) = \mathbb{P}_x(V_A < V_B)$ is the only solution of the equation*

$$\begin{cases}
h(x) = \sum_{y \in S} p(x,y)h(y), \quad \forall x \in S \setminus (A \cup B); \\
h(x) = 1, \quad \forall x \in A; \\
h(x) = 0, \quad \forall x \in B.
\end{cases} \tag{1}$$

Proof. 1. $h(x) = \mathbb{P}_x(V_A < V_B)$ satisfies (1).

For any $x \notin A \cup B$, V_A and V_B must ≥ 1, thus $\mathbb{1}_{\{V_A<V_B\}} \circ \theta_1 = \mathbb{1}_{\{V_A<V_B\}}$, then

$$\begin{aligned}
h(x) &= \mathbb{P}_x(V_A < V_B) = \mathbb{E}(\mathbb{1}_{\{V_A<V_B\}} \circ \theta_1) \\
&= \mathbb{E}_x[\mathbb{E}_x(\mathbb{1}_{\{V_A<V_B\}} \circ \theta_1 | \mathcal{F}_1)] \\
&= \mathbb{E}_x[\mathbb{P}_{X_1}(V_A < V_B)] \\
&= \mathbb{E}_x(h(X_1)) \\
&= \sum_{y \in S} p(x,y)h(y)
\end{aligned}$$

2. If $h(x)$ satisfies (1), then $Y_n = h(X_{n\wedge V_{A\cup B}})$ is a martingale.

First since $h(x) = \mathbb{P}_x(V_A < V_B) \in [0,1]$, $\mathbb{E}(|Y_n|) \leq 1 < \infty$. Second, since $n \wedge V_{A\cup B} \leq n$, $X_{n\wedge V_{A\cup B}} \in \mathcal{F}_n$, thus $Y_n \in \mathcal{F}_n$. Third, we will show $\mathbb{E}(Y_{n+1}|\mathcal{F}_n) = Y_n$ on both $\{V_{A\cup B} \geq n\}$ and $\{V_{A\cup B} < n\}$. On $\{V_{A\cup B} \geq n\}$, $Y_n = h(X_n)$, thus

$$\mathbb{E}(Y_{n+1}|\mathcal{F}_n) = \mathbb{E}(h(X_{n+1})|\mathcal{F}_n) = \mathbb{E}(h(X_1)\circ\theta_n|\mathcal{F}_n) = \mathbb{E}_{X_n}h(X_1) = \sum_{y\in S} p(X_n,y)h(y) = h(X_n) = Y_n.$$

On $\{V_{A\cup B} < n\} \in \mathcal{F}_n$, $Y_n = h(X_{V_{A\cup B}}) \in \mathcal{F}_n$ for any n, thus

$$\mathbb{E}(Y_{n+1}|\mathcal{F}_n) = \mathbb{E}[h(X_{V_{A\cup B}})|\mathcal{F}_n] = h(X_{V_{A\cup B}}) = Y_n.$$

Now we proved $\mathbb{E}(Y_{n+1}|\mathcal{F}_n) = Y_n$ and hence Y_n is a martingale.

3. $h(x) = \mathbb{P}_x(V_A < V_B)$ is the only solution of (1).

Suppose h satisfies (1), then $h(x) = \mathbb{E}_x(h(X_1))$. Since Y_n is a martingale, $\mathbb{E}_x(Y_n) = \mathbb{E}_x(Y_1)$ for any n. And since $x \notin A \cup B$, $V_{A\cup B} \geq 1$, we have $Y_1 = h(X_1)$. Thus

$$\mathbb{E}_x[h(X_{n\wedge V_{A\cup B}})] = \mathbb{E}_x(h(X_1)) = h(x). \tag{2}$$

By the Lemma 3.25, $V_{A\cup B}<\infty$ a.s., then let $n\to\infty$, (2) becomes

$$h(x)=\mathbb{E}_x[h(X_{V_{A\cup B}})].$$

Next, we will prove $\mathbb{1}_{\{V_A<V_B\}}=h(X_{V_{A\cup B}})$. If $\omega\in\{\omega:V_A(\omega)<V_B(\omega)\}$, then $X_{V_{A\cup B}}=X_{V_A}\in A$, thus

$$\mathbb{1}_{\{V_A<V_B\}}(\omega)=1=h(X_{V_A})=h(X_{V_{A\cup B}}).$$

If $\omega\in\{\omega:V_A(\omega)<V_B(\omega)\}$, $X_{V_{A\cup B}}=X_{V_B}\in B$, then

$$\mathbb{1}_{\{V_A<V_B\}}(\omega)=0=h(X_{V_B})=h(X_{V_{A\cup B}}).$$

Therefore $\mathbb{1}_{\{V_A<V_B\}}=h(X_{V_{A\cup B}})$, hence

$$\mathbb{P}_x(V_A<V_B)=\mathbb{E}_x(\mathbb{1}_{\{V_A<V_B\}})=\mathbb{E}_x(h(X_{V_{A\cup B}}))=h(x).$$ □

Example 3.27 (Wright-Fisher model)**.** Suppose state space is $S=\{0,1,2,\cdots\}$ and the transition probability is

$$p(i,j)=\binom{N}{j}\left(\frac{i}{N}\right)^j\left(1-\frac{i}{N}\right)^{N-j}.$$

Then for any $0\le x\le N$,

$$\mathbb{P}_x(V_N<V_0)=\frac{x}{N}$$

Proof. Let $h(x) = x/N$, then $h(N) = 1$, $h(0) = 0$, and

$$\begin{aligned}
\sum_{y\in S} p(x,y)h(y) &= \sum_{y\in S} \binom{N}{y} \left(\frac{x}{N}\right)^y \left(1-\frac{x}{N}\right)^{N-y} \cdot \frac{y}{N} \\
&= \sum_{y=1}^{N} \frac{N!}{y!(N-y)!} \left(\frac{x}{N}\right)^y \left(1-\frac{x}{N}\right)^{N-y} \cdot \frac{y}{N} \\
&= \frac{x}{N} \sum_{y=1}^{N} \frac{(N-1)!}{(y-1)![(N-1)-(y-1)]!} \left(\frac{x}{N}\right)^{y-1} \left(1-\frac{x}{N}\right)^{(N-1)-(y-1)} \\
&= \frac{x}{N} \sum_{y=0}^{N-1} \binom{N-1}{y} \left(\frac{x}{N}\right)^y \left(1-\frac{x}{N}\right)^{(N-1)-y} \\
&= \frac{x}{N} = h(x),
\end{aligned}$$

therefore by Theorem 3.26, $\mathbb{P}_x(V_N < V_0) = x/N$. □

Theorem 3.28 (Exit time)**.** *Let $C \subseteq S$ and $g(x) = \mathbb{E}_x(V_C)$. Suppose $S \setminus C$ is finite, $\mathbb{P}_x(V_C < \infty) > 0$ for all $x \in S \setminus C$. Then $g(x) = \mathbb{E}_x(V_C)$ is the only solution of the equation*

$$\begin{cases} g(x) = 1 + \sum_{y\in S} p(x,y)g(y), & \forall x \in S \setminus C; \\ g(x) = 0, & \forall x \in C. \end{cases} \tag{1}$$

3.5 Recurrence and transience

Definition 3.29. 1. We call $y \in S$

- transient if $\rho_{yy} < 1$.
- recurrent if $\rho_{yy} = 1$ or equivalently $T_y < \infty$ a.s.
- positive recurrent if $\mathbb{E}_y(T_y) < \infty$ (which implies $T_y < \infty$ a.s. thus recurrent)
- null recurrent if it is recurrent but not positive recurrent
- absorbing if $\{y\}$ is closed.

2. Define $N(y)$ is the number of returns to y in time $n \geq 1$, i.e.

$$N(y) = \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n = y\}}.$$

Lemma 3.30. *For any $y \in S$,*

$$N(y) = \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n = y\}} = \sum_{n=1}^{\infty} \mathbb{1}_{\{T_y^k < \infty\}}.$$

Corollary 3.31. *For any $x, y \in S$,*

$$\mathbb{E}_x[N(y)] = \sum_{n=1}^{\infty} p^n(x,y) = \sum_{n=1}^{\infty} \rho_{xy}\rho_{yy}^{n-1}.$$

Lemma 3.32. *Let $y \in S$, TFAE*

1. *y is recurrent*

2. $\mathbb{P}_y(X_n = y \text{ i.o.}) = 1$

3. $\mathbb{E}_y[N(y)] = \infty$

Proof. $1 \Longrightarrow 2$. Since $\rho_{yy} = 1$, by Proposition 3.23, $\mathbb{P}_y(T_y^k < \infty) = \rho_{yy}^k = 1$ for all $k \in \mathbb{Z}_+$. therefore

$$\mathbb{P}_y(X_n = y \text{ i.o.}) = \mathbb{P}_y(\bigcap_{k=1}^{\infty}\{T_y^k < \infty\}) = 1.$$

$2 \Longrightarrow 3$. $\mathbb{P}_y(X_n = y \text{ i.o.}) = 1$ implies $\mathbb{P}_y(T_y^k < \infty) = 1$ for all $k \in \mathbb{Z}_+$, thus by Corollary 3.31,

$$\mathbb{E}_y[N(y)] = \sum_{k=1}^{\infty} \mathbb{P}_y(T_y^k < \infty) = \infty.$$

$3 \Longrightarrow 1$. Suppose $\rho_{yy} < 1$, then

$$\mathbb{E}_y[N(y)] = \sum_{k=1}^{\infty} \mathbb{P}_y(T_y^k < \infty) = \sum_{n=1}^{\infty} \rho_{yy}^n = \frac{\rho_{yy}}{1-\rho_{yy}} < \infty,$$

leading to a contradiction! □

Example 3.33. Figure 5 shows a 4-state Markov chain. We have

$$\rho_{11} = 1 - \mathbb{P}_1(X_1 = 2, X_2 = 3) = 0.52, \quad \rho_{22} = 1 - \mathbb{P}_2(X_1 = 3) = 0.2, \quad \rho_{33} = \rho_{44} = 1,$$

thus state 1 and 2 are transient, state 3 and 4 are recurrent. Since

$$\rho_{12} = 1 - \mathbb{P}_1(X_n = 1, \forall n \in \mathbb{Z}_+) = 1,$$

we have

$$\mathbb{E}_1[N(2)] = \frac{\rho_{12}}{1-\rho_{22}} = \frac{1}{1-0.2} = \frac{5}{4}$$

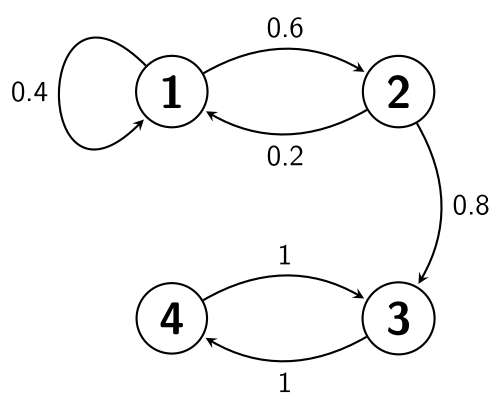


Figure 5: A 4-state Markov Chain

Proposition 3.34 (Recurrence is contagious)**.** *If x is recurrent and $\rho_{xy} > 0$, then y is recurrent and $\rho_{yx} = \rho_{xy} = 1$.*

Proof. 1.The case $y = x$ is trivial. Suppose $y \neq x$. Since $\rho_{xy} > 0$, by Lemma 3.21, there exists $n \in \mathbb{Z}_+$ s.t. $p^n(x, y) > 0$. Let $k = \inf\{n \in \mathbb{Z}_+ : p^n(x, y) > 0\}$, also by Lemma 3.21, there exists a state sequence $y_1, y_2, \cdots, y_{k-1}$ s.t.

$$p(x, y_1)p(y_1, y_2)\cdots p(y_{k-1}, y) > 0.$$

2.Define $h : \Omega \to \mathbb{R}$,

$$h(\omega) = \begin{cases} 1 & \text{if } \omega_k \neq x,\ \forall k \in \mathbb{Z}_+ \\ 0 & \text{else,} \end{cases}$$

obviously, $h = \mathbb{1}_{\{T_x=\infty\}}$. By Markov property, we have

$$\mathbb{E}_x(h(X_k, X_{k+1}, \cdots)|\mathcal{F}_k) = \mathbb{E}_{X_k}[h(X_0, X_1, \cdots)],$$

then for $A = \{X_1 = y_1, \cdots, X_{k-1} = y_{k-1}, X_k = y\} \in \mathcal{F}_k$,

$$\mathbb{E}_x(h(X_k, X_{k+1}, \cdots)\mathbb{1}_A) = \mathbb{E}_x[\mathbb{E}_{X_k}[h(X_0, X_1, \cdots)]\mathbb{1}_A],$$

the LHS is

$$\mathbb{E}_x[\mathbb{1}_{\{T_x=\infty\}}\mathbb{1}_A] = \mathbb{P}_x[X_1 = y_1, \cdots, X_{k-1} = y_{k-1}, X_k = y, T_x = \infty]$$

the RHS is

$$\mathbb{E}_x[\mathbb{E}_y[h(X_0, X_1, \cdots)]\mathbb{1}_A] = \mathbb{E}_y(h)\mathbb{E}_x(\mathbb{1}_A) = \mathbb{P}_y(T_x = \infty)\mathbb{P}_x(X_1 = y_1, \cdots, X_{k-1} = y_{k-1}, X_k = y).$$

Therefore,

$$\begin{aligned}1-\rho_{xx}&=\mathbb{P}_x(T_x=\infty)\\&\geq \mathbb{P}_x(X_1=y_1,\cdots,X_{k-1}=y_{k-1},X_k=y,T_x=\infty)\\&=\mathbb{P}_y(T_x=\infty)\mathbb{P}_x(X_1=y_1,\cdots,X_{k-1}=y_{k-1},X_k=y)\\&=(1-\rho_{yx})p(x,y_1)p(y_1,y_2)\cdots p(y_{k-1},y),\end{aligned}$$

thus $\rho_{xx}=1$ implies $\rho_{yx}=1$.

3. Since $\rho_{yx}=1>0$, by Lemma 3.21, there is an $l\in\mathbb{Z}_+$ s.t. $p^l(y,x)>0$, then by Theorem 3.18,

$$p^{l+n+k}(y,y)\geq p^l(y,x)p^n(x,x)p^k(x,y).$$

By Lemma 3.31, we have

$$\mathbb{E}_y[N(y)]=\sum_{n=1}^{\infty}p^n(y,y)\geq\sum_{n=1}^{\infty}p^{l+n+k}(y,y)\geq p^l(y,x)p^k(x,y)\sum_{n=1}^{\infty}p^n(x,x)=\infty,$$

however $\sum_{n=1}^{\infty}p^n(x,x)=\mathbb{E}_x[N(x)]=\infty$ by Proposition 3.32. Thus $\mathbb{E}_y[N(y)]=\infty$, and by Proposition 3.32 again, y is recurrent. □

Definition 3.35 (communication)**.** Suppose $x,y\in S$, $x\neq y$, we say x communicates with y if $\rho_{xy}>0$ and $\rho_{yx}>0$, denoted as $x\leftrightarrow y$. Define x always communicates with itself.

Definition 3.36. Let $C\subseteq S$ be a non-empty set. We call C

- closed if $x\in C$ and $\rho_{xy}>0$ implies $y\in C$, or equivalently, $x\in C$, $y\notin C$ implies $\rho_{xy}=0$.
- irreducible if $x,y\in C$ implies $x\leftrightarrow y$.

We call a Markov chain (or transition proposition p) to have some property (recurrent, transient, irreducible, closed,...) if S has such property.

Lemma 3.37. *For any $x,y\in S$, $x\neq y$, if $x\leftrightarrow y$, then $\rho_{xx}>0$.*

Proof. By Lemma 3.20,

$$\rho_{xx} \geq \rho_{xy}\rho_{yx} > 0.$$

□

Corollary 3.38. *Let $x \in S$. If $C_x = \{y \in S : \rho_{xy} > 0, \rho_{yx} > 0\}$ is not empty, then $C_x = \{y \in S : y \leftrightarrow x\}$*

Proposition 3.39. 1. *$\leftrightarrow$ is an equivalence relation.*

2. *S can be partitioned into equivalence classes of $\leftrightarrow$.*

3. *Each equivalence class is irreducible.*

Proposition 3.40. *Any equivalence class $C \subseteq S$ of $\leftrightarrow$ is either recurrent or transient.*

Proof. By Proposition 3.39, C is irreducible. If $|C| = 1$, there is only one state, so either recurrent or transient. Now assume $|C| \geq 2$. For any $x \in C$,
Case 1: x is recurrent. Then for any $y \in C$, $\rho_{xy} > 0$, by Proposition 3.34, y is also recurrent. Thus all states are recurrent.
Case 2: x is transient. If there exists $y \in C$, $y \neq x$ is recurrent, then by Case 1, x is also recurrent, which is a contradiction. So all states are transient. □

Remark. This shows recurrence and transience are class property, i.e. if one state in an equivalence class is recurrent (or transient), then all states in such class are recurrent (or transient).

Lemma 3.41. *If $C \subseteq S$ is closed, for any $x \in C$, we have*

$$\mathbb{P}_x(X_n \in C) = 1,$$

i.e.

$$\sum_{y \in C} p^n(x,y) = 1.$$

Proposition 3.42. *Suppose a non-empty set $C \subseteq S$ is finite and closed.*

1. *C contains a recurrent state.*

2. *All recurrent states in C are positive recurrent.*

3. *If C is irreducible then all states in C are recurrent.*

Proof. 1.Suppose no state in C is recurrent, i.e. for any $y \in C$, $\rho_{yy} < 1$, then by Proposition 3.32,

$$\mathbb{E}_x[N(y)] = \frac{\rho_{xy}}{1-\rho_{yy}} < \infty,$$

since C is finite, we have

$$\sum_{y\in C} \mathbb{E}_x[N(y)] < \infty.$$

However, by Fubini's theorem and Lemma 3.41,

$$\sum_{y\in C} \mathbb{E}_x[N(y)] = \sum_{y\in C}\sum_{n=1}^{\infty} p^n(x,y) = \sum_{n=1}^{\infty}\sum_{y\in C} p^n(x,y) = \sum_{n=1}^{\infty} 1 = \infty.$$

2.

3.If C only has one state, then by 1, it is recurrent. If C has more than one state, there exists a recurrent state$x \in C$. For any $y \in C$ and $y \neq x$, since X is irreducible, then $\rho_{xy} > 0$. By Proposition 3.34, y is also recurrent. □

Corollary 3.43. *If an irreducible Markov chain has finite states, then it is recurrent.*

Proof. Obviously, S is closed, thus this follows directly from Proposition 3.42. □

Proposition 3.44. *Suppose S is finite, $x \in S$.*

1. *If there is a $y \in S$, s.t. $\rho_{xy} > 0$ and $\rho_{yx} = 0$, then x is transient.*

2. *If any $y \in S$ with $\rho_{xy} > 0$ also has $\rho_{yx} > 0$, then x is recurrent.*

Proof. 1. Suppose x is recurrent, since $\rho_{xy} > 0$, then by Proposition 3.34, we must have $\rho_{yx} = 1$.

2. $C_x = \{y : \rho_{xy} > 0\} = \{y : x \leftrightarrow y\}$. Then C_x is the equivalence class containing x, thus C_x is irreducible by Proposition 3.39. C_x is also closed. If $C_x = S$, it is obviously closed; If $C_x \subsetneq S$, let $y \in C_x$, $z \notin C_x$ with $\rho_{yz} > 0$, then by Lemma 3.20, $\rho_{xz} \geq \rho_{xy}\rho_{yz} > 0$, which means $z \in C$. It is a contradiction, so $\rho_{yz} = 0$ and C_x is closed. By Proposition 3.42, C_x is recurrent. Thus $x \in C_x$ is recurrent. □

Lemma 3.45. *Suppose the equivalence class $C \subseteq S$ is recurrent, then it is closed.*

Proof. S is trivially closed, now suppose $C \subsetneq S$. Let $x \in C$ and $y \in S \setminus C$, if $\rho_{xy} > 0$, then by Proposition 3.34, $\rho_{yx} > 0$, thus $x \leftrightarrow y$, which implies $y \in C$, it is a contradiction. Therefore $\rho_{xy} = 0$, i.e. C is closed. □

Theorem 3.46 (Decomposition theorem)**.** *Let R be the set of all recurrent states. Then R can be written as the disjoint union of R_i where each R_i is irreducible and closed.*

Proof. By Proposition 3.39 and Lemma 3.45. □

Proposition 3.47. *Suppose p is irreducible and recurrent. μ is the initial distribution, then for any $y \in S$,*

$$\mathbb{P}_\mu(T_y < \infty) = 1.$$

Proof. For any $x \in S$, by irreducibility, $\rho_{xy>0}$, then by Proposition 3.34, $\rho_{xy} = 1$. Therefore,

$$\mathbb{P}_\mu(T_y < \infty) = \sum_{x\in S} \mu(x)\mathbb{P}_x(T_y < \infty) = \sum_{x\in S} \mu(x) = 1.$$

3.6 Recurrence of simple random walk

In this section, we consider the simple random walk on $\mathbb{Z}^d$. Define $\{X_i : i \geq 0\}$ are i.i.d. r.v. with

$$\mathbb{P}(X_i = e_j) = \mathbb{P}(X_i = -e_j) = \frac{1}{2d},$$

where e_j is unit vectors on $\mathbb{Z}^d$. Let $S_m = \sum_{i=1}^m X_i$, $S_0 = 0$, obviously $\{S_m : m \geq 0\}$ is a Markov chain on state space $\mathbb{Z}$ starting from 0.

Theorem 3.48 (Stirling's formula)**.**

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

Theorem 3.49. *0 is recurrent state for $\{S_m : m \geq 0\}$ in $d \leq 2$ and transient in $d \geq 3$.*

Proof. Let $p_d(m) = \mathbb{P}(S_m = 0)$, then $p_d(m) = 0$ if m is odd. And by Lemma 3.32, 0 is recurrent if $\sum_{m=1}^\infty p_d(m) = \infty$, and transient if $\sum_{m=1}^\infty p_d(m) < \infty$.

1. $d = 1$. For $2n$ steps, $S_{2n} = 0$ means there are n left steps and n right steps, so

$$p_1(2n) = \binom{2n}{n}(\frac{1}{2})^n(\frac{1}{2})^n = \frac{(2n)!}{n!n!2^{2n}} \sim \frac{1}{\sqrt{\pi n}},$$

thus

$$\sum_{n=1}^\infty p_1(2n) \sim \sum_{n=1}^\infty \frac{1}{\sqrt{\pi n}} = \infty.$$

2. $d = 2$. Similarly, to make $S_{2n} = 0$, there should be m up steps and m down steps, $n-m$ left steps and $n-m$ right steps for some $0 \leq m \leq n$. Then

$$p_2(2n) = \sum_{m=0}^n \frac{(2n)!}{m!m!(n-m)!(n-m)!}(\frac{1}{4})^m(\frac{1}{4})^m(\frac{1}{4})^{n-m}(\frac{1}{4})^{n-m} = [p_1(2n)]^2 \sim \frac{1}{\pi n},$$

so

$$\sum_{n=1}^{\infty} p_1(2n) \sim \sum_{n=1}^{\infty} \frac{1}{\pi n} = \infty.$$

3. $d = 3$.

3.7 Periodicity

Definition 3.50. Let $x \in S$ be a state.

1. I_x is the set of positive time n that makes $p^n(x,x) > 0$, i.e.

$$I_x = \{n \in \mathbb{Z}_+ : p^n(x,x) > 0\};$$

2. Let d_x be the greatest common divisor of I_x (If $I_x = \varnothing$ i.e. $p(x,x) = 0$, we define $d_x = 0$). We call d_x the period of x.

3. We call x is periodic if $d_x > 1$, aperiodic if $d_x = 1$.

4. We call the Markov chain aperiodic if all states are aperiodic.

Proposition 3.51 (period is a class property)**.** *Let $x \in S$ with $d_x > 0$, C_x is the equivalence class containing x, i.e. $C_x = \{y \in S : y \leftrightarrow x\}$. Then every state in C_x has period d_x.*

Proof. The trivial case $C_x = \{x\}$ is obvious. We can assume $|C_x| \geq 2$. Then for any $y \in C_x \setminus \{x\}$, $\rho_{xy} > 0$ and $\rho_{yx} > 0$. By Lemma 3.21, there exists $L, M \in \mathbb{Z}_+$, s.t. $p^L(x,y) > 0$ and $p^M(y,x) > 0$. Therefore, by Theorem 3.18,

$$p^{L+M}(y,y) \geq p^M(y,x)p^L(x,y) > 0,$$

which means $d_y \mid L + M$. For any $n \in I_x$, $p^n(x,x) > 0$, then

$$p^{L+n+M}(y,y) \geq p^M(y,x)p^n(x,x)p^L(x,y) > 0,$$

thus $d_y|L+n+M$. So $d_y|n$ for any $n \in I_x$, which implies $d_y|d_x$. By the same argument, we can show $d_x|d_y$, thus $d_y = d_x$. □

Corollary 3.52. *Suppose p is irreducible, then*

1. *All states have the same period.*

2. *If $p(x,x) > 0$ for some state x (a self loop), then $d_x = 1$, hence p is aperiodic.*

Example 3.53. For the Markov chain in Figure 6, $I_1 = \{4, 6, 8, 10, \cdots\}$, so $d_1 = 2$, the whole chain has period 2.

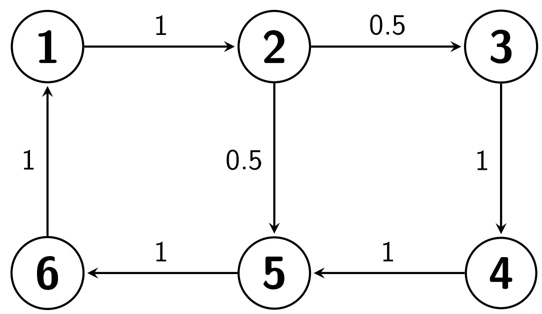


Figure 6: A 6-state Markov Chain

For the Markov chain in Figure 7, $I_1 = \{4, 5, 8, 9, 10, \cdots\}$, so $d_1 = 1$. The whole chain is aperiodic.

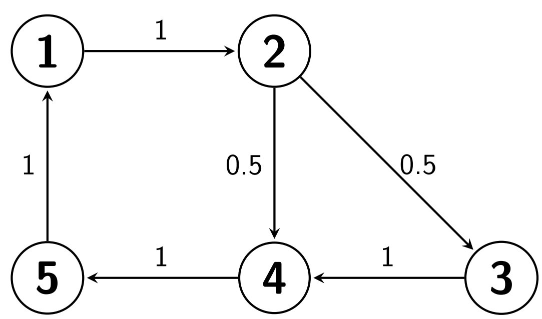


Figure 7: A 5-state Markov Chain

Figure 8 shows an irreducible chain with period 2, but it is transient.

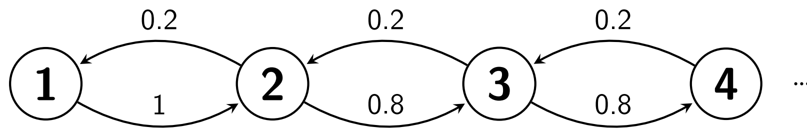


Figure 8: An irreducible, periodic but transient chain

In Figure 7, notice that $I_1 = \{4, 5, 8, 9, 10, 12, 13, 14, 15, 16, 17, \cdots\}$, i.e. $p^n(1,1) > 0$ for all $n \geq 12$, so we have the next result (Proposition 3.57), the following lemmas will be used to prove it.

Lemma 3.54. *If $m, n \in I_x$, then $m + n \in I_x$ and $km \in I_x$ for any $k \in \mathbb{Z}_+$.*

Proof. By Theorem 3.18. □

Lemma 3.55. *If $A \subseteq \mathbb{Z}_+$ is an infinite set, then there exists a finite subset $A' \subseteq A$ s.t.*

$$\gcd(A) = \gcd(A').$$

Lemma 3.56. *Suppose $A = \{a_1, a_2, \cdots, a_k\} \subseteq \mathbb{Z}_+$, then there exists $c_1, \cdots, c_k \in \mathbb{Z}$ s.t.*

$$c_1a_1 + c_2a_1 + \cdots + c_ka_k = \gcd(A).$$

Proposition 3.57. *Suppose $x \in S$ with $d_x = 1$, then there exists $m_x \in \mathbb{Z}_+$ s.t. $m \in I_x$ for all $m \geq m_x$.*

Proof. 1. We only need to show there are two consecutive integers n and $n+1$ in I_x. Then let $m_x = n(n-1)$, for any $m \geq m_x$, m can be written as $m = kn + r$ (divide m by n with remainder r), where $k \geq n-1$, $0 \leq r \leq n-1$, then by Lemma 3.54,

$$m = kn + r = (k - r)n + r(n + 1) \in I_x.$$

2. Since $\gcd(I_x) = 1$, by Lemma 3.55 and 3.56, there exists integers $i_1, i_2, \cdots, i_k \in I_x$ and

$c_1, c_2, \cdots, c_k \in \mathbb{Z}$, s.t.

$$c_1 i_1 + c_2 i_2 + \cdots + c_k i_k = d_x = 1,$$

let $a_j = c_j^+ = \max\{c_j, 0\}$, $b_j = c_j^- = \max\{-c_j, 0\}$[6], then $a_j, b_j \geq 0$ and $c_j = a_j - b_j$, we have

$$a_1 i_1 + \cdots + a_k i_k = b_1 i_1 + \cdots + b_k i_k + 1.$$

Let $n = b_1 i_1 + \cdots + b_k i_k \in I_x$, then $n + 1 = a_1 i_1 + \cdots + a_k i_k$ is also in I_x. □

Corollary 3.58. *Suppose $x \in S$ with $d_x \geq 1$, then there exists $m_x \in \mathbb{Z}_+$ s.t. $md_x \in I_x$ for all $m \geq m_x$.*

Remark. In Proposition 3.57, the problem of finding the minimal integer m_x s.t. $m \in I_x$ for all $m \geq m_x$ is called Frobenius problem.

Lemma 3.59. *Suppose p is irreducible with $d \geq 2$, if $p(x, y) > 0$ for some $x, y \in S$, then $p^N(y, x) > 0$ for some $N = d - 1 \pmod{d}$.*

Theorem 3.60 (decomposition theorem)**.** *Suppose p is irreducible and has period $d \geq 1$, then S can be written as the disjoint union of subsets $S_0, S_1, \cdots, S_{d-1}$ where for any $x \in S_i$*

$$p(x, y) > 0 \implies y \in S_{i+1 \bmod d}.$$

Moreover, this decomposition is unique up to the cyclic permutations.

Proof. Define relation $\sim$ on S: $x \sim y$ if $p^{nd}(x, y) > 0$ for some $n \in \mathbb{Z}_+$.

Claim. $\sim$ is indeed an equivalence relation.

i) $x \sim x$ by Corollary 3.58;

ii) if $x \sim y$, i.e. $p^{nd}(x, y) > 0$ for some $n \in \mathbb{Z}_+$, then suppose $p^L(y, x) > 0$ (by irreducibility),

[6] We can assume b_j are not all zero, otherwise $c_j \geq 0$ for all j, then there must be some $i_j = 1$ and $c_j = 1$, implying $I_x = \mathbb{Z}_+$, which is the trivial case. Thus $b_1 i_1 + \cdots + b_k i_k \in I_x$.

we have

$$p^{nd+L}(x,x) \geq p^{nd}(x,y)p^L(y,x) > 0,$$

then $d \,|\, nd+L$, thus $d \,|\, L$, which means $y \sim x$;

iii) Suppose $x \sim y$, $y \sim z$, i.e. $p^{md}(x,y) > 0, p^{nd}(y,z) > 0$ for some $m,n \in \mathbb{Z}_+$, then

$$p^{(m+n)d}(x,z) \geq p^{md}(x,y)p^{nd}(y,z) > 0,$$

thus $x \sim z$.

Therefore, the equivalence relation $\sim$ determines a unique partition on S. For any $x_0 \in S$, let $S_0 = [x_0]$, i.e. the equivalence class containing x_0. If $S_0 = S$ (i.e. $d = 1$), we are done; if $S_0 \subsetneq S$ (i.e. $d \geq 2$), then there must exist $x_1 \in S \setminus S_0$ s.t. $p(x_0,x_1) > 0$. Let $S_1 = [x_1]$, suppose $p(y,z) > 0$ for some $y \in S_0$, $z \in S$, we want to show $z \in S_1$. Since $p^{md}(x_0,y) > 0$ for some m and by Lemma 3.59, $p^{nd-1}(x_1,x_0) > 0$ for some n, then

$$p^{(m+n)d}(x_1,z) \geq p^{nd-1}(x_1,x_0)p^{md}(x_0,y)p(y,z) > 0,$$

so $x_1 \sim z$, $z \in S_1$. Repeating this procedure, we can find all desired $S_2, \cdots, S_{n-1}$. □

Remark. The above theorem actually shows such chain will visit S_i one after the other. Suppose $x \in S_i$, then $\mathbb{P}_x(X_n \in S_{n+i \bmod d}) = 1$ for any $n \in \mathbb{Z}_+$.

3.8 Stationary Measures

Here we still consider the countable state space S.

Definition 3.61. Suppose $\mu : \mathcal{S} \to [0,+\infty]$ is a measure on $(S,\mathcal{S})$. X_n is a Markov chain on S with transition probability p.

1. Denote
$$\mu p(y) = \sum_{x \in S} \mu(x)p(x,y).$$

2. μ is called a stationary measure if it is σ-finite[7] and for any $y \in S$,
$$\mu p(y) = \mu(y).$$

3. μ is called a stationary distribution, if μ is a stationary measure and $\mu(S) = 1$.

4. We say μ satisfies the detailed balanced condition or μ is reversible if for any $x, y \in S$
$$\mu(x)p(x,y) = \mu(y)p(y,x).$$

Proposition 3.62. $\mu \equiv 1$ *is a stationary measure if and only if for any* $y \in S$*,*
$$\sum_{x \in S} p(x,y) = 1.$$

Proof.
$$\sum_{x \in S} p(x,y) = \sum_{x \in S} \mu(x)p(x,y) = \mu(y) = 1.$$ □

Proposition 3.63. *If a measure* μ *is reversible, then it is a stationary measure.*

Proof. Suppose μ is reversible, then
$$\mu p(y) = \sum_{x \in S} \mu(x)p(x,y) = \sum_{x \in S} \mu(y)p(y,x) = \mu(y)\sum_{x \in S} p(y,x) = \mu(y).$$ □

Proposition 3.64. *Suppose* μ *is a stationary measure and* X_0 *has "distribution"* μ*. Let* $Y_m = X_{n-m}$*,* $0 \leq m \leq n$ *is a Markov chain with initial "distribution"* μ *and transition*

[7] This means for any $x \in S$, $\mu(x) := \mu(\{x\}) < \infty$.

probability

$$q(x,y) = \frac{\mu(y)p(y,x)}{\mu(x)}.$$

Furthermore, if μ is reversible, then $q = p$.

Theorem 3.65 (Kolmogorov's cycle condition)**.** *Suppose S is irreducible w.r.t. transition probability p. Then there exists a reversible measure if and only if the following two conditions hold,*

(i) $p(x,y) > 0$ implies $p(y,x) > 0$;

(ii) for any loop $x_0, x_1, \cdots, x_n = x_0$, if

$$\prod_{1\le i\le n} p(x_i, x_{i-1}) > 0, \tag{1}$$

then we have

$$\prod_{i=1}^{n} \frac{p(x_{i-1}, x_i)}{p(x_i, x_{i-1})} = 1. \tag{2}$$

Proof. $\Longrightarrow$:Suppose there is a reversible measure μ. Since S is irreducible, then for any $x, y \in S$, $\rho_{xy} = \mathbb{P}_x(T_y < \infty) > 0$, thus $\mu(x) > 0$ for any $x \in S$ (otherwise $\mathbb{P}_x \equiv 0$). By the definition of reversible measure,

$$\mu(x)p(x,y) = \mu(y)p(y,x),$$

therefore $p(x,y) > 0$ implies $p(y,x) > 0$. Next, suppose $x_0, x_1, \cdots, x_n = x_0$ is a loop, and (1) holds. Then by definition, for any $i = 1, \cdots, n$,

$$\mu(x_i)p(x_i, x_{i-1}) = \mu(x_{i-1})p(x_{i-1}, x_i)$$

multiply them together, we get

$$\prod_{i=1}^{n}\mu(x_i)p(x_i,x_{i-1})=\prod_{i=1}^{n}\mu(x_{i-1})p(x_{i-1},x_i),$$

i.e.

$$\prod_{i=1}^{n}\mu(x_i)\prod_{i=1}^{n}p(x_i,x_{i-1})=\prod_{i=0}^{i-1}\mu(x_i)\prod_{i=1}^{n}p(x_{i-1},x_i),$$

(2) is obtained since $\prod_{i=1}^{n}\mu(x_i)=\prod_{i=0}^{n-1}\mu(x_i)$.

$\Longleftarrow$:Suppose the two conditions hold. Fix $a\in S$, since S is irreducible, for any $x\in S$, $\rho_{ax}>0$, by Lemma 3.21, there exists a path $x_0=a,x_1,\cdots,x_n=x$, s.t. $\prod_{i=1}^{n}p(x_i-1,x_i)>0$. Define

$$\mu(x)=\prod_{i=1}^{n}\frac{p(x_{i-1},x_i)}{p(x_i,x_{i-1})}.$$

First, μ is well-defined, i.e. $\mu(x)$ is independent of the path from a to x. Let $\tilde{x}_0=a,\tilde{x}_1,\cdots,\tilde{x}_n=x$ be another path with $\prod_{i=1}^{n}p(\tilde{x}_i-1,\tilde{x}_i)>0$, then $x_0=a,x_1,\cdots,x_n=x=\tilde{x}_n,\tilde{x}_{n-1},\cdots,\tilde{x}_1,\tilde{x}_0=a$ is a loop, thus by (2), we have

$$1=\frac{p(x_0,x_1)}{p(x_1,x_0)}\cdots\frac{p(x_{n-1},x_n)}{p(x_n,x_{n-1})}\cdot\frac{p(\tilde{x}_n,\tilde{x}_{n-1})}{p(\tilde{x}_{n-1},\tilde{x}_n)}\cdots\frac{p(\tilde{x}_1,\tilde{x}_0)}{p(\tilde{x}_0,\tilde{x}_1)}=\prod_{i=1}^{n}\frac{p(x_{i-1},x_i)}{p(x_i,x_{i-1})}\prod_{i=1}^{n}\frac{p(\tilde{x}_i,\tilde{x}_{i-1})}{p(\tilde{x}_{i-1},\tilde{x}_i)},$$

therefore

$$\prod_{i=1}^{n}\frac{p(x_{i-1},x_i)}{p(x_i,x_{i-1})}=\prod_{i=1}^{n}\frac{p(\tilde{x}_{i-1},\tilde{x}_i)}{p(\tilde{x}_i,\tilde{x}_{i-1})},$$

i.e. two different paths give the same $\mu(x)$ value.

Second, we will show μ is reversible, i.e. for any $x,y\in S$

$$\mu(x)p(x,y)=\mu(y)p(y,x). \tag{3}$$

If $p(x,y)=0$, by (2), $p(y,x)=0$, then (3) holds. If $p(x,y)>0$, by (2), $p(y,x)>0$ and there exists a path from a to y, i.e. $x_0=a,x_1,\cdots,x_n=x,x_{n+1}=y$ with $\prod_{i=1}^{n+1}p(x_{i-1},x_i)>0$, then

$$\mu(y) = \prod_{i=1}^{n+1} \frac{p(x_{i-1}, x_i)}{p(x_i, x_{i-1})} = \left[\prod_{i=1}^{n} \frac{p(x_{i-1}, x_i)}{p(x_i, x_{i-1})}\right] \cdot \frac{p(x_n, x_{n+1})}{p(x_{n+1}, x_n)} = \mu(x) \cdot \frac{p(x, y)}{p(y, x)},$$

thus (3) follows immediately. □

Lemma 3.66. *If p is transient, then a stationary distribution does not exist.*

Proof. Suppose there is a stationary distribution π. By Lemma 3.31, p is transient implies for any $x, y \in S$,

$$\sum_{n=1}^{\infty} p^n(x, y) = \sum_{n=1}^{\infty} \rho_{xy}\rho_{yy}^{n-1} = \frac{\rho_{xy}}{1 - \rho_{yy}} < \infty,$$

thus as $n \to \infty$,

$$p^n(x, y) \to 0.$$

By the property of stationary distribution, for any $y \in S$,

$$\pi(y) = \pi p^n(y) = \sum_{x \in S} \pi(x) p^n(x, y) \to 0,$$

which contradicts that π is a distribution. □

Theorem 3.67 (construction of stationary measure)**.** *Suppose x is a recurrent state, then for any $y \in S$,*

$$\mu_x(y) = \mathbb{E}_x\left(\sum_{n=0}^{T_x - 1} \mathbb{1}_{\{X_n = y\}}\right) = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y, T_x > n)$$

defines a stationary measure.

Proof. Our goal is to show for any $z \in S$,

$$\mu_x p(z) = \mu_x(z).$$

First, we have

$$\begin{aligned}
\mu_x p(z) &= \sum_{y\in S} \mu_x(y)p(y,z) \\
&= \sum_{y\in S}\sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y, T_x > n)p(y,z) \\
&= \sum_{n=0}^{\infty}\sum_{y\in S} \mathbb{P}_x(X_n = y, T_x > n)p(y,z) \qquad \text{(by Fubini's theorem)} \\
&= \sum_{n=0}^{\infty}\sum_{y\in S} \mathbb{P}_x(X_n = y, T_x > n, X_{n+1} = z).
\end{aligned}$$

The last equality above holds because

$$p(X_n, z) = \mathbb{P}_x(X_{n+1} = z|\mathcal{F}_n) = \mathbb{E}_x(\mathbb{1}_{\{X_{n+1}=z\}}|\mathcal{F}_n),$$

then for $A = \{X_n = y, n < T_x\} \in \mathcal{F}_n$, we have

$$\mathbb{E}_x(\mathbb{1}_{\{X_{n+1}=z\}}\mathbb{1}_A) = \mathbb{E}_x(p(X_n, z)\mathbb{1}_A),$$

LHS is $\mathbb{P}_x(\{X_{n+1} = z\} \cap A) = \mathbb{P}_x(X_n = y, T_x > n, X_{n+1} = z)$, RHS is $p(y, z)\mathbb{P}_x(A) = \mathbb{P}_x(X_n = y, T_x > n)p(y, z)$.

Case 1. $z \neq x$.

Notice that $X_n \neq x$ on $T_x > n$, i.e. $\{X_n = x, T_x > n\} = \varnothing$, then

$$\begin{aligned}
\bigsqcup_{y\in S}\{X_n = y, T_x > n, X_{n+1} = z\} &= \{X_n \in S \setminus \{x\}, T_x > n, X_{n+1} = z\} \\
&= \{X_{n+1} = z, T_x > n+1\},
\end{aligned}$$

therefore,

$$
\begin{aligned}
\mu_x p(z) &= \sum_{n=0}^{\infty} \sum_{y \in S} \mathbb{P}_x(X_n = y, T_x > n, X_{n+1} = z) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_x \left(\bigsqcup_{y \in S} \{X_n = y, T_x > n, X_{n+1} = z\} \right) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_x(X_{n+1} = z, T_x > n+1) \\
&= \sum_{n=1}^{\infty} \mathbb{P}_x(X_n = z, T_x > n) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = z, T_x > n) \qquad (\text{since } \mathbb{P}_x(X_0 = z, T_x > 0) = 0) \\
&= \mu_x(z).
\end{aligned}
$$

Case 2. $z = x$.

In this case,

$$
\bigsqcup_{y \in S} \{X_n = y, T_x > n, X_{n+1} = x\} = \{X_n \in S \setminus \{x\}, T_x > n, X_{n+1} = x\} = \{T_x = n+1\},
$$

thus

$$
\begin{aligned}
\mu_x p(x) &= \sum_{n=0}^{\infty} \sum_{y \in S} \mathbb{P}_x(X_n = y, T_x > n, X_{n+1} = x) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_x(T_x = n+1) \\
&= \mathbb{P}_x(T_x < \infty) \\
&= \rho_{xx} = 1. \quad (\text{since } x \text{ is recurrent})
\end{aligned}
$$

On the other hand, $\{X_n = n, T_x > n\} = \varnothing$ for $n \geq 1$, so

$$\mu_x(x) = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = x, T_x > n) = \mathbb{P}_x(X_0 = x, T_x > n) = 1,$$

hence $\mu_x p(x) = \mu_x(x)$. □

Remark. 1. If x is transient, Case 1 still holds, but Case 2 will be different, because

$$\mu_x p(x) = \rho_{xx} < 1 = \mu_x(x).$$

2. μ_x is σ-finite, i.e. for any $y \in S$, $\mu_x(y) < \infty$.

   If $y = x$, clearly, $\mu_x(x) = 1 < \infty$. Suppose $y \neq x$. If $\rho_{xy} = 0$, since $\{X_n = y, T_x > n\} \subseteq \{X_n = y\} \subseteq \{T_y < \infty\}$, thus

$$\mu_x(y) = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = y, T_x > n) \leq \sum_{n=0}^{\infty} \mathbb{P}_x(T_y < \infty) = 0.$$

   If $\rho_{xy} > 0$, since x is recurrent, by Proposition 3.34, y is also recurrent, and $\rho_{yx} = 1 > 0$, by Lemma 3.21, $p^n(y, x) > 0$ for some $n \geq 1$. By the property of stationary measure, we have

$$1 = \mu_x(x) = \mu_x p^n(x) = \sum_{z \in S} \mu_x(z) p^n(z, x) \geq \mu_x(y) p^n(y, x),$$

   thus

$$\mu_x(y) \leq \frac{1}{p^n(y, x)} < \infty.$$

3. We have

$$\begin{aligned}\mu_x(S) &= \sum_{y\in S}\mu_x(y)\\ &= \sum_{y\in S}\sum_{n=0}^{\infty}\mathbb{P}_x(X_n=y,T_x>n)\\ &= \sum_{n=0}^{\infty}\sum_{y\in S}\mathbb{P}_x(X_n=y,T_x>n)\\ &= \sum_{n=0}^{\infty}\mathbb{P}_x(T_x>n)\\ &= \mathbb{E}_x(T_x), \qquad \text{(tail sum formula)}\end{aligned}$$

thus if $\mathbb{E}_x(T_x)<\infty$ (positive recurrent),

$$\pi := \frac{\mu_x}{\mathbb{E}_x(T_x)}$$

is a stationary distribution.

Theorem 3.68. *If p is irreducible and recurrent, then the stationary measure is unique up to constant multiples.*

Proof. Suppose ν is a stationary measure, let $a \in S$, then for any $z \in S$,

$$\begin{aligned}
\nu(z) &= \sum_{y\in S} \nu(y)p(y,z) \\
&= \nu(a)p(a,z) + \sum_{\substack{y\in S\\ y\neq a}} \nu(y)p(y,z) \\
&= \nu(a)p(a,z) + \sum_{\substack{y\in S\\ y\neq a}} \left[\sum_{x\in S} \nu(x)p(x,y)\right] p(y,z) \\
&= \nu(a)p(a,z) + \sum_{\substack{y\in S\\ y\neq a}} \left[\nu(a)p(a,y) + \sum_{\substack{x\in S\\ x\neq a}} \nu(x)p(x,y)\right] p(y,z) \\
&= \nu(a)p(a,z) + \sum_{\substack{y\in S\\ y\neq a}} \nu(a)p(a,y)p(y,z) + \sum_{\substack{y\in S\\ y\neq a}}\sum_{\substack{x\in S\\ y\neq a}} \nu(x)p(x,y)p(y,z) \\
&= \nu(a)\mathbb{P}_a(X_1 = z) + \nu(a)\mathbb{P}_a(X_1 \neq a, X_2 = z) + \mathbb{P}_\nu(X_0 \neq a, X_1 \neq a, X_2 = z) \\
&\vdots \\
&= \nu(a)\sum_{m=1}^{n} \mathbb{P}_a(X_k \neq a, 1 \leq k < m, X_m = z) + \mathbb{P}_\nu(X_k \neq a, 0 \leq k < n, X_n = z) \\
&= \nu(a)\sum_{m=0}^{n} \mathbb{P}_a(T_a > m, X_m = z) + \mathbb{P}_\nu(X_k \neq a, 0 \leq k < n, X_n = z) \\
&\geq \nu(a)\sum_{m=0}^{n} \mathbb{P}_a(T_a > m, X_m = z)
\end{aligned}$$

(it holds for both $z = a$ and $z \neq a$) let $n \to \infty$, we have

$$\nu(z) \geq \nu(a)\mu_a(z),$$

where μ_a is the stationary measure defined by Theorem 3.67. Next, we will prove it is actually an equality. Since p is irreducible, we have $\rho_{za} > 0$, thus by Lemma 3.21, $p^n(z,a) > 0$ for some $n \in Z_+$. Notice that

$$\nu(a) = \sum_{x \in S} \nu(x) p^n(x, a) \geq \nu(a) \sum_x \mu_a(x) p^n(x, a) = \nu(a)\mu_a(a) = \nu(a),$$

where we apply $\mu_a(a) = 1$ from Theorem 3.67. Therefore

$$\sum_{x \in S} [\nu(x) - \nu(a)\mu_a(x)] p^n(x, a) = 0,$$

where $[\nu(x) - \nu(a)\mu_a(x)]p^n(x, a) \geq 0$, thus

$$[\nu(x) - \nu(a)\mu_a(x)]p^n(x, a) = 0, \quad \forall x \in S.$$

When $x = z$, since $p^n(z, a) > 0$, it follows $\nu(z) - \nu(a)\mu_a(z) = 0$ i.e. $\nu(z) = \nu(a)\mu_a(z)$. Moreover, by the σ-finiteness of ν, we have $\nu(a) < \infty$. □

Now we have proved the existence (Theorem 3.67) and uniqueness (Theorem 3.68) of stationary measures.

Corollary 3.69. *If p is irreducible and recurrent, and there is a positive recurrent state $x \in S$, then there is a unique stationary distribution, which is*

$$\pi = \frac{\mu_x}{\mathbb{E}_x(T_x)}.$$

Proof. By the Remark 3 of Theorem 3.67, we can define a stationary distribution by

$$\pi = \frac{\mu_x}{\mathbb{E}_x(T_x)}.$$

By Theorem 3.68, such stationary distribution is unique. □

Lemma 3.70. *If there is a stationary distribution π, then any state y with $\pi(y) > 0$ is recurrent.*

Proof. For any $n \geq 1$, we have

$$\pi p^n = \pi,$$

then for any $y \in S$, by Fubini's theorem and $\pi(y) > 0$,

$$\sum_{x \in S} \pi(x) \sum_{n=1}^{\infty} p^n(x,y) = \sum_{n=1}^{\infty} \sum_{x \in S} \pi(x) p^n(x,y) = \sum_{n=1}^{\infty} \pi(y) = \infty.$$

By Lemma 3.31,

$$\sum_{n=1}^{\infty} p^n(x,y) = \sum_{n=1}^{\infty} \rho_{xy} \rho_{yy}^{n-1},$$

thus

$$\infty = \sum_{x \in S} \pi(x) \sum_{n=1}^{\infty} p^n(x,y) = \sum_{x \in S} \pi(x) \sum_{n=1}^{\infty} \rho_{xy} \rho_{yy}^{n-1} = \sum_{n=1}^{\infty} \rho_{yy}^{n-1} \sum_{x \in S} \pi(x) \rho_{xy} \leq \sum_{n=1}^{\infty} \rho_{yy}^{n-1},$$

which implies $\rho_{yy} = 1$, i.e. y is recurrent. □

Proposition 3.71. *If p is irreducible and there is a stationary distribution π, then*

1. $\pi(x) > 0$ *for any* $x \in S$*;*

2. p *is recurrent;*

3. π *is the unique stationary distribution;*

4. p *is positive recurrent;*

5. *for any* $x \in S$*,*

$$\pi(x) = \frac{1}{\mathbb{E}_x(T_x)}.$$

Proof. 1.Suppose there exists an $x \in S$ s.t. $\pi(x) = 0$. There must be some $y \in S$ s.t. $\pi(y) > 0$, otherwise π fails to be a distribution. Since p is irreducible, $\rho_{yx} > 0$, by Lemma

3.21, $p^n(y,x) > 0$ for some $n \in \mathbb{Z}_+$. But

$$0 = \pi(x) = \sum_{y \in S} \pi(y) p^n(y,x),$$

and $\pi(y) > 0$, suggesting $p^n(y,x) = 0$.

2. By Proposition 3.70, all states are recurrent.

3. By Theorem 3.68.

4. By Theorem 3.67, for any $x \in S$, μ_x is a stationary measure. And by Theorem 3.68, $\mu_x = c\pi$ for some $c < \infty$, thus by Remark 3 in Theorem 3.67,

$$\mathbb{E}_x(T_x) = \mu_x(S) = c\pi(S) = c < \infty,$$

i.e. all states are positive recurrent.

5. By Corollary 3.69, for any $x \in S$,

$$\pi = \frac{\mu_x}{\mathbb{E}_x(T_x)}$$

is a stationary distribution, and sicne $\mu_x(x) = 1$, we have

$$\pi(x) = \frac{\mu_x(x)}{\mathbb{E}_x(T_x)} = \frac{1}{\mathbb{E}_x(T_x)}.$$

□

Proposition 3.72. *Suppose S is irreducible, then TFAE,*

1. *Some x is positive recurrent;*
2. *There exists a stationary distribution;*
3. *p is positive recurrent.*

Proof. 1 $\implies$ 2: Suppose x is positive recurrent, then by Theorem 3.67

$$\pi = \frac{\mu_x}{\mathbb{E}_x(T_x)}$$

defines a stationary distribution.

2 $\implies$ 3: By Proposition 3.71.

3 $\implies$ 1: trivial. $\square$

3.9 Asymptotic behavior

In this section, we will consider the asymptotic behavior of $p^n(x,y)$.

Proposition 3.73. *If $y \in S$ is transient, then $p^n(x,y) \to 0$ as $n \to \infty$.*

Proof. By Lemma 3.31, y is transient implies for any $x \in S$,

$$\sum_{n=1}^{\infty} p^n(x,y) = \sum_{n=1}^{\infty} \rho_{xy}\rho_{yy}^{n-1} = \frac{\rho_{xy}}{1-\rho_{yy}} < \infty,$$

thus as $n \to \infty$,

$$p^n(x,y) \to 0.$$

□

How about the case when y is recurrent?

Definition 3.74. For any $y \in S$, $n \in \mathbb{Z}_+$, let $N_n(y)$ be the number of visits to y by time n, i.e.

$$N_n(y) = \sum_{m=1}^{n} \mathbb{1}_{\{X_m=y\}}.$$

Lemma 3.75. *Suppose y is recurrent and for any $k \geq 0$, let $R_k = T_y^k$ be the time of the k-th return to y. For $k \geq 1$, let $r_k = R_k - R_{k-1}$ be the k-th interarrival time. Then under $\mathbb{P}_y$, the vectors $v_k = (r_k, X_{R_{k-1}}, \cdots, X_{R_k - 1})$, $k \geq 1$ are i.i.d.*

Proof. Let's make some examples first, if $R_1 = 5$, $R_2 = 8$, $R_3 = 10$, then $v_1 = (5, X_0, \cdots, X_4)$, $v_2 = (3, X_5, X_6, X_7)$, $v_3 = (2, X_8, X_9)$. So we observe that $v_2(X_0, X_1, X_3, \cdots) = v_1(X_5, X_6, X_7, \cdots)$, in general,

$$v_k = v_1 \circ \theta_{R_{k-1}}.$$

i)First, v_k and v_1 have the same distribution.

Let $X = (X_0, X_1, \cdots)$, $X' = X \circ \theta_{R_{k-1}} = (X_{R_{k-1}}, X_{R_{k-1}+1}, \cdots)$, then for any $A \in \mathcal{F}$,

$$\begin{aligned}\mathbb{P}_y(X' \in A) &= \mathbb{E}_y(\mathbb{1}_{\{X\in A\}} \circ \theta_{R_{k-1}}) \\ &= \mathbb{E}_y[\mathbb{E}_y(\mathbb{1}_{\{X\in A\}} \circ \theta_{R_{k-1}} | \mathcal{F}_{R_{k-1}})] \\ &= \mathbb{E}_y[\mathbb{E}_{X_{R_{k-1}}}(\mathbb{1}_{\{X\in A\}})] \qquad (R_{k-1} < \infty \text{ a.s. and strong Markov pproperty}) \\ &= \mathbb{P}_y(X \in A),\end{aligned}$$

thus X and X' has the same distribution, then $v_k = v_1(X')$ and $v_1 = v_1(X)$ has the same distribution.

ii)Second $\sigma(v_k)$ is independent of $\mathcal{F}_{R_{k-1}}$.

Claim. For any $\{X \in A\} \in \sigma(X)$, if $\mathbb{P}(X \in A|\mathcal{F}) = \mathbb{P}(X \in A)$, then $\sigma(X)$ and $\mathcal{F}$ are independent.

Proof. For any $B \in \mathcal{F}$, by the definition of conditional expectation, we have

$$\mathbb{P}(\{X \in A\} \cap B) = \mathbb{E}(\mathbb{1}_{\{X\in A\}}\mathbb{1}_B) = \mathbb{E}(\mathbb{P}(X \in A)\mathbb{1}_B) = \mathbb{P}(X \in A)\mathbb{P}(B),$$

thus $\sigma(X)$ and $\mathcal{F}$ are independent.

Let $\{v_k \in V\} \in \sigma(v_k)$, by the strong Markov property, we have

$$\mathbb{P}_y(v_k \in V|\mathcal{F}_{R_{k-1}}) = \mathbb{E}_y(\mathbb{1}_{\{v_1\in V\}} \circ \theta_{R_{k-1}}|\mathcal{F}_{R_{k-1}}) = \mathbb{E}_{X_{R_{k-1}}}(\mathbb{1}_{\{v_1\in V\}}) = \mathbb{P}_y(v_1 \in V) = \mathbb{P}_y(v_k \in V).$$

Therefore, by the above claim, $\sigma(v_k)$ is independent of $\mathcal{F}_{R_{k-1}} \supseteq \sigma(v_1), \cdots, \sigma(v_{k-1})$, so v_k is independent of $v_1, \cdots, v_{k-1}$ and also has the same distribution as them. By induction, v_k, $k \geq 1$ are all i.i.d. $\square$

Theorem 3.76. *Suppose y is recurrent. Then for any $x \in S$,*

$$\frac{N_n(y)}{n} \to \frac{1}{\mathbb{E}_y(T_y)} \mathbb{1}_{\{T_y<\infty\}} \qquad \mathbb{P}_x\text{-}a.s.$$

as $n \to \infty$.

Proof. Case 1. Suppose the chain initiates at y. Let $r_k = T_y^k - T_y^{k-1}$, then by Lemma 3.75, $r_k, k \geq 1$ are i.i.d. and $\mathbb{E}_y(r_k) = \mathbb{E}_y(r_1) = \mathbb{E}_y(T_y)$ ($< \infty$ or $= \infty$). Therefore, by the strong law of large number,

$$\frac{\sum_{k=1}^n r_k}{n} = \frac{T_y^n}{n} \to \mathbb{E}_y(T_y) \qquad \mathbb{P}_y\text{-a.s.} \tag{1}$$

Since $T_y^{N_n(y)} \leq n < T_y^{N_n(y)+1}$ (where $T_y^{N_n(y)}$ means the time of the last return to y by time n, $T_y^{N_n(y)+1}$ means the time of the first return to y after time n),

$$\frac{T_y^{N_n(y)}}{N_n(y)} \leq \frac{n}{N_n(y)} < \frac{T_y^{N_n(y)+1}}{N_n(y)+1} \cdot \frac{N_n(y)+1}{N_n(y)}. \tag{2}$$

By Proposition 3.32, y recurrent implies $\mathbb{E}_y[N(y)] = \infty$, then $N(y) = \lim_{n\to\infty} N_n(y) = \infty$ a.s. Let $n \to \infty$ in (2), by equation (1), squeeze theorem of limit and subsequence convergence, we have

$$\frac{n}{N_n(y)} \to \mathbb{E}_y(T_y) \qquad \mathbb{P}_y\text{-a.s.}$$

Case 2. Suppose the chain initiates at x and $x \neq y$. Since ρ_{xy} may not be 1, we need to consider both $\{T_y = \infty\}$ and $\{T_y < \infty\}$. On $\{T_y = \infty\}$, $N_n(y) = 0$, for all $n \in \mathbb{Z}_+$, then

$$\frac{N_n(y)}{n} \to 0.$$

On $\{T_y < \infty\}$, by the same argument in Lemma 3.75, $r_k, k \geq 2$ are i.i.d, and for $k \geq 2$,

$\mathbb{P}_x(r_k = n) = \mathbb{P}_y(T_y = n)$, thus $\mathbb{E}_x(r_k) = \mathbb{E}_y(T_y)$, then by the strong law of large number,

$$\frac{T_y^n}{n} = \frac{T_y}{n} + \frac{\sum_{k=2}^n r_k}{n} \to 0 + \mathbb{E}_x(r_k) = \mathbb{E}_y(T_y) \qquad \mathbb{P}_x\text{-a.s.}$$

Repeating what we did in Case 1, we have

$$\frac{N_n(y)}{n} \to \frac{1}{\mathbb{E}_y(T_y)} \qquad \mathbb{P}_x\text{-a.s.}$$

Therefore in Case 2, we have

$$\frac{N_n(y)}{n} \to \frac{1}{\mathbb{E}_y(T_y)} \mathbb{1}_{\{T_y<\infty\}} \qquad \mathbb{P}_x\text{-a.s.}$$

□

Remark.

1. This theorem provides an interpretation of positive recurrent and null recurrent. If y is positive recurrent, then the asymptotic frequency of visits at y is positive; if y is null recurrent, then it is 0.

2. Since $\frac{N_n(y)}{n} \in [0,1]$, by bounded convergence theorem,

$$\mathbb{E}_x[\frac{N_n(y)}{n}] \to \mathbb{E}_x[\frac{1}{\mathbb{E}_y(T_y)} \mathbb{1}_{\{T_y<\infty\}}],$$

   i.e.

$$\frac{\mathbb{E}_x(N_n(y))}{n} \to \frac{\mathbb{P}_x(T_y < \infty)}{\mathbb{E}_y(T_y)} = \frac{\rho_{xy}}{\mathbb{E}_y(T_y)}.$$

   Notice that

$$\mathbb{E}_x(N_n(y)) = \mathbb{E}_x[\sum_{m=1}^n \mathbb{1}_{\{X_m=y\}}] = \sum_{m=1}^n \mathbb{P}_x(X_m = y) = \sum_{m=1}^n p^m(x,y),$$

   therefore

$$\frac{1}{n}\sum_{m=1}^n p^m(x,y) \to \frac{\rho_{xy}}{\mathbb{E}_y(T_y)}$$

as $n \to \infty$. This means $p^n(x, y)$ converges in the Cesaro sense if y is recurrent (Cesaro convergence is also true for transient state, because convergence implies Cesaro convergence).

Corollary 3.77. *Suppose p is irreducible. If p is transient or null-recurrent, then for any $x, y \in S$,*

$$\frac{1}{n}\sum_{m=1}^{n} p^m(x, y) \to 0,$$

as $n \to \infty$. If p is positive-recurrent, then for any $x, y \in S$,

$$\frac{1}{n}\sum_{m=1}^{n} p^m(x, y) \to \pi(y),$$

as $n \to \infty$, where π is the stationary distribution of p.

Theorem 3.78 (Convergence theorem). *Suppose Markov chain X_n has transition probability p and initial distribution μ. If p is irreducible, aperiodic, and has a stationary distribution π, then for all $y \in S$,*

$$\mathbb{P}_\mu(X_n = y) \to \pi(y),$$

as $n \to \infty$. In particular, for all $x, y \in S$,

$$\mathbb{P}_x(X_n = y) = p^n(x, y) \to \pi(y)$$

as $n \to \infty$.

Proof. We will use a technique called coupling. Let Y_n be a Markov chain with transition probability p and initial distribution π, and independent with X_n. Consider $Z_n = (X_n, Y_n)$.
1. Z_n is a Markov chain on $S^2 = S \times S$ with transition probability $\bar{p}$ and initial distribution λ, where

$$\bar{p}((x_1, y_2), (x_2, y_2)) = p(x_1, x_2)p(y_1, y_2), \quad \forall x_1, x_2, y_1, y_2 \in S,$$

and

$$\lambda(x,y) = \mu(x)\pi(y).$$

2. $\bar{p}$ is irreducible.

Since p is irreducible, there exists K and L, s.t. $p^K(x_1,x_2) > 0$ and $p^L(y_1,y_2) > 0$. And by Proposition 3.57, there exists $m(x_2), m(y_2) \in \mathbb{Z}_+$ s.t. for any $m \geq m(x_2)$ and $n \geq m(y_2)$,

$$p^m(x_2,x_2) > 0, \quad p^n(y_2,y_2) > 0.$$

Let $M = \max\{0, m(x_2) - K, m(y_2) - L\}$, then

$$p^{K+L+M}(x_1,x_2) \geq p^L(x_1,x_2)p^{K+M}(x_2,x_2) > 0, \quad p^{K+L+M}(y_1,y_2) \geq p^K(y_1,y_2)p^{L+M}(y_2,y_2) > 0,$$

thus

$$\bar{p}^{K+L+M}((x_1,y_2),(x_2,y_2)) = p^{K+L+M}(x_1,x_2)p^{K+L+M}(y_1,y_2) > 0,$$

as desired.

3. $\bar{\pi}$ defined by $\bar{\pi}(a,b) = \pi(a)\pi(b)$ is a stationary distribution for $\bar{p}$.

This is because for any $(x_1,y_1) \in S^2$,

$$\begin{aligned}
\bar{\pi}p(x_1,y_1) &= \sum_{(x,y)\in S^2} \bar{\pi}(x,y)p((x,y),(x_1,y_1)) \\
&= \sum_{(x,y)\in S^2} \pi(x)\pi(y)p(x,x_1)p(y,y_1) \\
&= \pi(x_1)\pi(y_1) \\
&= \bar{\pi}(x_1,y_1),
\end{aligned}$$

and $\sum_{(x,y)\in S^2} \bar{\pi}(x,y) = 1$.

4. Since $\bar{p}$ is irreducible and has a stationary distribution, by Proposition 3.71, $\bar{p}$ is positive recurrent.

5. For any $x \in S$, define $T = \inf\{n \geq 1 : X_n = Y_n\}$, $T_x = \inf\{n \geq 1 : X_n = Y_n = x\}$. Since $\bar{p}$ is irreducible and recurrent, by Proposition 3.47, $\mathbb{P}_\lambda(T_x < \infty) = 1$. Then we have $\mathbb{P}_\lambda(T < \infty) = 1$ because $\{T_x < \infty\} \subseteq \{T < \infty\}$.

6. On $\{T \leq n\}$ (after hitting the diagonal), X_n and Y_n have the same distribution.
Since $\{T \leq n\} = \bigsqcup_{m=1}^n \{T = m\}$, for any $y \in S$,

$$\begin{aligned}
\mathbb{P}_\lambda(X_n = y, T \leq n) &= \sum_{m=1}^n \mathbb{P}_\lambda(X_n = y, T = m) \\
&= \sum_{m=1}^n \sum_{x\in S} \mathbb{P}_\lambda(X_n = y, T = m, X_m = x) \\
&= \sum_{m=1}^n \sum_{x\in S} \mathbb{P}_\lambda(T = m, X_m = x)\mathbb{P}_\lambda(X_n = y | X_m = x, T = m) \\
&= \sum_{m=1}^n \sum_{x\in S} \mathbb{P}_\lambda(T = m, Y_m = x) p^{n-m}(x, y) \\
&= \sum_{m=1}^n \sum_{x\in S} \mathbb{P}_\lambda(T = m, Y_m = x)\mathbb{P}_\lambda(Y_n = y | Y_m = x, T = m) \\
&= \mathbb{P}_\lambda(Y_n = y, T \leq n).
\end{aligned}$$

7. Notice that

$$\begin{aligned}
\mathbb{P}_\lambda(X_n = y) &= \mathbb{P}_\lambda(X_n = y, T \leq n) + \mathbb{P}_\lambda(X_n = y, T > n) \\
&= \mathbb{P}_\lambda(Y_n = y, T \leq n) + \mathbb{P}_\lambda(X_n = y, T > n) \\
&\leq \mathbb{P}_\lambda(Y_n = y) + \mathbb{P}_\lambda(X_n = y, T > n),
\end{aligned}$$

and similarly,

$$\mathbb{P}_\lambda(Y_n = y) \leq \mathbb{P}_\lambda(X_n = y) + \mathbb{P}_\lambda(Y_n = y, T > n).$$

So $\mathbb{P}_\lambda(X_n = y) - \mathbb{P}_\lambda(Y_n = y) \leq \mathbb{P}_\lambda(X_n = y, T > n)$ and $\mathbb{P}_\lambda(Y_n = y) - \mathbb{P}_\lambda(X_n = y) \leq \mathbb{P}_\lambda(Y_n =$

$y, T > n$),

$$\begin{aligned}|\mathbb{P}_\mu(X_n = y) - \pi(y)| &= |\mathbb{P}_\lambda(X_n = y) - \mathbb{P}_\lambda(Y_n = y)| \\ &\leq \max\{\mathbb{P}_\lambda(X_n = y, T > n), \mathbb{P}_\lambda(Y_n = y, T > n)\} \\ &\leq \mathbb{P}_\lambda(X_n = y, T > n) + \mathbb{P}_\lambda(Y_n = y, T > n) \\ &\leq 2\mathbb{P}_\lambda(T > n) \to 0,\end{aligned}$$

because $T < \infty$ a.s. Therefore

$$\lim_{n\to\infty} \mathbb{P}_\mu(X_n = y) = \pi(y).$$

□

For the version of null-recurrent, we have the following theorem.

Theorem 3.79. *Suppose p is irreducible, aperiodic, and null-recurrence, then for any $y \in S$,*

$$\mathbb{P}_\mu(X_n = y) \to 0$$

as $n \to \infty$.

4 Branching process

4.1 Model description and basic properties

Galton-Watson tree or Branching process is a sequence of r.v. $\{Z_n : n \geq 0\}$ with $Z_0 = 1$ and for $n \geq 1$

$$Z_n = \begin{cases} \sum_{i=1}^{Z_{n-1}} \xi_i^{(n)} & Z_{n-1} \neq 0 \\ 0 & Z_{n-1} = 0 \end{cases}$$

where $\xi_i^{(m)} : \Omega \to \mathbb{N}$ for all m, i are i.i.d $\sim \xi$. In other word, $\{Z_n : n \geq 1\}$ can be viewed as a family starting from one ancestor (Z_0). Everyone can generate children following the distribution of ξ. And Z_n is the total number of people in the n-th generation.

Let $p_k = \mathbb{P}(\xi = k)$, $k \in \mathbb{N}$ be the probability that a person generates k children. Then $\sum_k p_k = 1$. Denote $\mu = \mathbb{E}(\xi)$. To avoid the trivial case, we always assume $p_0 > 0$ and $p_0 + p_1 < 1$.

Lemma 4.1. *$\{Z_n : n \geq 0\}$ is a Markov chain on $S = \mathbb{N}$ with transition probability*

$$p(i, j) = \mathbb{P}(\sum_{m=1}^{i} \xi_m = j).$$

Proposition 4.2. *All states $k \geq 1$ are transient. State 0 is recurrent and absorbing.*

Proof. First we have

$$\rho_{k,0} = \mathbb{P}_k(T_0 < \infty) \geq p(k, 0) = [\mathbb{P}(\xi = 0)]^k > 0,$$

and $\rho_{0,k} = 0$. If k is recurrent, then by Proposition 3.34, $\rho_{0,k} = 1$ which leads to a contradiction, thus state $k \geq 1$ is transient. Second, $\rho_{0,0} = 1$, so 0 is recurrent by definition. Moreover, x is also an absorbing state since $\rho_{0,k} = 0$ for any $k \geq 1$. □

Lemma 4.3. *Let $\mathcal{F}_n = \sigma(\xi_i^{(m)}, i \geq 1, 1 \leq m \leq n)$, $\mu \in (0, +\infty)$, then $\{W_n = Z_n/\mu^n : n \geq 0\}$ is a non-negative martingale w.r.t. $\{\mathcal{F}_n\}$.*

Proof. $W_n \in \mathcal{F}_n$. And Since $Z_{n+1} = Z_{n+1}\mathbb{1}_{\sqcup_{k=1}^{\infty}\{Z_n=k\}} = Z_{n+1}\sum_{k=1}^{\infty}\mathbb{1}_{\{Z_n=k\}}$, we have

$$\begin{aligned}
\mathbb{E}(Z_{n+1}|\mathcal{F}_n) &= \sum_{k=1}^{\infty}\mathbb{E}(Z_{n+1}\mathbb{1}_{\{Z_n=k\}}|\mathcal{F}_n) \\
&= \sum_{k=1}^{\infty}\mathbb{E}[(\xi_1^{(n+1)} + \xi_2^{(n+1)} + \cdots + \xi_k^{(n+1)})\mathbb{1}_{\{Z_n=k\}}|\mathcal{F}_n] \\
&= \sum_{k=1}^{\infty}\mathbb{1}_{\{Z_n=k\}}\mathbb{E}(\xi_1^{(n+1)} + \xi_2^{(n+1)} + \cdots + \xi_k^{(n+1)}) \\
&= \sum_{k=1}^{\infty}\mathbb{1}_{\{Z_n=k\}}k\mu \\
&= \mu Z_n,
\end{aligned}$$

thus

$$\mathbb{E}(W_{n+1}|\mathcal{F}_n) = \mathbb{E}(\frac{Z_{n+1}}{\mu^{n+1}}|\mathcal{F}_n) = \frac{Z_n}{\mu^n} = W_n.$$

□

Corollary 4.4. *$W_n \to W_\infty$ a.s. and $\mathbb{E}(W_\infty) \leq 1$.*

Proof. Direct from Corollary 2.15. □

4.2 Generating function

Definition 4.5. Define generating function $\varphi : [0, 1] \to \mathbb{R}$ by

$$\varphi(s) = \mathbb{E}(s^\xi) = \sum_{k=0}^{\infty} p_k s^k.$$

Lemma 4.6. *The generating function φ has the following properties:*

1. $\varphi(0) = p_0$, $\varphi(1) = 1$

2. $\varphi'(0) = p_1$, $\varphi'(1) = \mu$

3. *$\varphi'(s) > 0$ for all $s \in (0,1)$, i.e. φ is strictly increasing on $(0,1)$.*

4. *$\varphi''(s) \geq 0$ for all $s \in (0,1)$, i.e. φ is convex on $(0,1)$.*

Proof. 1. $\varphi(0) = p_0$ is obvious,

$$\varphi(1) = \sum_{k=0}^{\infty} p_k = 1.$$

2. Since $\varphi(s)$ is absolutely convergent on $[0,1]$, We have

$$\varphi'(s) = \sum_{k=0}^{\infty} (p_k s^k)' = \sum_{k=1}^{\infty} k p_k s^{k-1} = p_1 + 2p_2 s + 3p_3 s^2 + \cdots,$$

thus $\varphi'(0) = p_1$, $\varphi'(1) = \sum_{k=1}^{\infty} k p_k = \mathbb{E}(\xi) = \mu$.

3. By assumption, $p_1 > 0$, so $\varphi'(s) \geq p_1 > 0$ on $(0,1)$.

4. Since

$$\varphi''(s) = \sum_{k=2}^{\infty} k(k-1) p_k s^{k-2} = 2p_2 + 6p_3 s + \cdots \geq 0.$$

□

Proposition 4.7. *Suppose p is the transition probability, then*

$$\varphi(s) = \sum_{k=0}^{\infty} p(1,k) s^k, \quad [\varphi(s)]^j = \sum_{k=0}^{\infty} p(j,k) s^k.$$

Proof. The first equality is the definition of $\varphi(s)$. For the second one, consider the expansion of

$$[\varphi(s)]^j = \left[\sum_{k=0}^{\infty} p(1,k) s^k\right]^j,$$

the coefficient of s^n equals

$$\sum_{\substack{k_1,k_2,\cdots,k_j \\ k_1+k_2+\cdots+k_j=n}} \prod_{i=1}^{j} p(1,k_i) = \mathbb{P}(\xi_1 + \xi_2 + \cdots + \xi_j = n) = p(j,n).$$

□

Proposition 4.8. *Let $\varphi^{(n)}(s) = \mathbb{E}(s^{Z_n}) = \sum_k p^n(1,k)s^k$, where $p^n(1,k)$ is the n-step transition probability. Let φ^n be the n-th iteration of φ, i.e. $\varphi^{n+1}(s) = \varphi(\varphi^n(s))$ for all $n \geq 1$. Then*

1. $\varphi^{(n)}(s) = \varphi^n(s)$

2. *For any $j \geq 0$,*
$$[\varphi^n(s)]^j = \sum_{k=0}^{\infty} p^n(j,k)s^k.$$

3. $p^n(j,0) = [\varphi^n(0)]^j$.

Proof. 1. On $\{Z_{n-1} = k\}$, we have

$$\mathbb{E}(s^{Z_n}\mathbb{1}_{\{Z_{n-1}=k\}}|\mathcal{F}_{n-1}) = \mathbb{E}(\prod_{i=1}^{k} s^{\xi_i}\mathbb{1}_{\{Z_{n-1}=k\}}|\mathcal{F}_{n-1}) = \prod_{i=1}^{k}\mathbb{E}(s^{\xi_i})\mathbb{1}_{\{Z_{n-1}=k\}} = [\varphi(s)]^k\mathbb{1}_{\{Z_{n-1}=k\}},$$

take expectation, we have
$$\mathbb{E}(s^{Z_n}) = \mathbb{E}[[\varphi(s)]^{Z_{n-1}}],$$

since $\mathbb{E}(s^{Z_1}) = \varphi(s)$, $\mathbb{E}(s^{Z_2}) = \mathbb{E}[[\varphi(s)]^{Z_1}] = \varphi(\varphi(s))$, we finish the proof by induction.

2. See proof in Proposition 4.7.

3. Let $s = 0$ in 2). □

4.3 Moments

Let $\mu = \mathbb{E}(\xi)$, $\sigma^2 = \mathrm{Var}(\xi^2)$.

Proposition 4.9. $\mathbb{E}(Z_n) = \mu^n$.

Proof. Since $W_n = Z_n/\mu^n$ is a martingale,

$$1 = \mathbb{E}(W_1) = \mathbb{E}(W_n) = \frac{\mathbb{E}(Z_n)}{\mu^n}.$$ □

Proposition 4.10.

$$\mathrm{Var}(Z_n) = \begin{cases} \dfrac{\sigma^2\mu^{n-1}(\mu^n - 1)}{\mu - 1} & \textit{if } \mu \neq 1 \\ n\sigma^2 & \textit{if } \mu = 1 \end{cases}$$

Proof. Observe that

$$[\varphi^n(s)]' = \sum_{k=1}^{\infty} kp^n(1,k)s^{k-1}, \quad [\varphi^n(s)]'' = \sum_{k=2}^{\infty} k(k-1)p^n(1,k)s^{k-2},$$

so

$$\mathbb{E}(Z_n^2) = \sum_{k=0}^{\infty} k^2 p^n(1,k) = [\varphi^n(1)]' + [\varphi^n(1)]''.$$

$[\varphi^n(1)]' = \sum_{k=1}^{\infty} kp^n(1,k) = \mathbb{E}(Z_n) = \mu^n$. For $[\varphi^n(1)]''$, we note that

$$\begin{aligned} [\varphi^n(s)]'' &= [\varphi(\varphi^{n-1}(s))]'' \\ &= [\varphi'(\varphi^{n-1}(s))[\varphi^{n-1}(s)]']' \\ &= \varphi''(\varphi^{n-1}(s)) \cdot [[\varphi^{n-1}(s)]']^2 + \varphi'(\varphi^{n-1}(s)) \cdot [\varphi^{n-1}(s)]'' \end{aligned},$$

let $s = 1$, since $\varphi^n(1) = 1$, $\varphi'(1) = \mu$, $\varphi''(1) = \mathbb{E}(\xi^2) - \varphi'(1) = \sigma^2 + \mu^2 - \mu$, we have

$$[\varphi^n(1)]'' = \varphi''(1) \cdot [[\varphi^{n-1}(1)]']^2 + \varphi'(1) \cdot [\varphi^{n-1}(1)]'' = (\sigma^2 + \mu^2 - \mu)\mu^{2n-2} + \mu[\varphi^{n-1}(1)]''.$$

By induction, we have

$$[\varphi^n(1)]'' = (\sigma^2 + \mu^2 - \mu)(\mu^{2n-2} + \mu^{2n-3} + \cdots + \mu^{n-1}),$$

therefore

$$\begin{aligned}
\mathrm{Var}(Z_n) &= \mathbb{E}(Z_n^2) - [\mathbb{E}(Z_n)]^2 \\
&= \mu^n + (\sigma^2 + \mu^2 - \mu)(\mu^{2n-2} + \mu^{2n-3} + \cdots + \mu^{n-1}) - \mu^{2n} \\
&= \sigma^2\mu^{n-1}(1 + \mu + \mu^2 + \cdots + \mu^{n-1}) + \mu^n(\mu - 1)(1 + \mu + \cdots + \mu^{n-1}) + \mu^n - \mu^{2n} \\
&= \sigma^2\mu^{n-1}(1 + \mu + \mu^2 + \cdots + \mu^{n-1}) \\
&= \begin{cases} \dfrac{\sigma^2\mu^{n-1}(\mu^n - 1)}{\mu - 1} & \text{if } \mu \neq 1 \\ n\sigma^2 & \text{if } \mu = 1 \end{cases}
\end{aligned}$$

□

Corollary 4.11.

$$\mathrm{Var}(W_n) = \frac{\mathrm{Var}(Z_n)}{\mu^{2n}} = \begin{cases} \dfrac{\sigma^2(\mu^n - 1)}{\mu^{n+1}(\mu - 1)} & \textit{if } \mu \neq 1 \\ \dfrac{n\sigma^2}{\mu^{2n}} & \textit{if } \mu = 1 \end{cases}$$

Proposition 4.12. *If $\mu > 1$, $\sigma^2 < \infty$, then*

1. *$W_n \to W_\infty$ in L^2 and L^1*

2. *$\mathbb{E}(W_\infty) = 1$,*

$$\mathbb{E}(W_\infty^2) = 1 + \frac{\sigma^2}{\mu(\mu - 1)}.$$

Proof. By Corollary 4.11, for all $n \geq 0$,

$$\mathbb{E}(W_n^2) = \mathrm{Var}(W_n) + [\mathbb{E}(W_n)]^2 = \frac{\sigma^2(\mu^n - 1)}{\mu^{n+1}(\mu - 1)} + 1 = \frac{\sigma^2(1 - \frac{1}{\mu^n})}{\mu(\mu - 1)} + 1 < \frac{\sigma^2}{\mu(\mu - 1)} + 1 < \infty,$$

thus $\sup_n \mathbb{E}(W_n^2) < \infty$, by Theorem 2.25, $W_n \to W_\infty$ in L^2 hence also in L^1. Then

$$\mathbb{E}(W_\infty) = \lim_{n\to\infty} \mathbb{E}(W_n) = 1, \quad \mathbb{E}(W_\infty^2) = \lim_{n\to\infty} \mathbb{E}(W_n^2) = \frac{\sigma^2}{\mu(\mu - 1)} + 1.$$ □

4.4 Extinction probability

Definition 4.13. We say the population goes extinct if $Z_n \to 0$, denoted as $Z_\infty = 0$. We say the population does not go extinct if $Z_n \not\to 0$, denoted as $Z_\infty > 0$. (Z_n may not have a limit in case of non-extinction, here Z_∞ is just a notation).

Lemma 4.14. *For any $\omega \in \Omega$, $Z_n(\omega)$ goes extinct if and only if $Z_n(\omega) = 0$ for some n.*

Proposition 4.15. *If $\mu < 1$, then $Z_\infty = 0$ a.s. Hence $W_\infty = 0$ a.s.*

Proof. Since Z_n takes integers, $\{Z_n \geq 1\} = \{Z_n > 0\}$, therefore when $\mu < 1$,

$$\mathbb{P}(Z_n > 0) = \mathbb{E}(\mathbb{1}_{\{Z_n>0\}}) \leq \mathbb{E}(Z_n \mathbb{1}_{\{Z_n>0\}}) \leq \mathbb{E}(Z_n) = \mu^n \to 0,$$

which implies $Z_n \to 0$ in probability. Since

$$\sum_{n=1}^{\infty} \mathbb{P}(Z_n > 0) \leq \sum_{n=1}^{\infty} \mu^n = \frac{\mu}{1-\mu} < \infty,$$

by Borel-Cantelli lemma, $\mathbb{P}(Z_n > 0,\ i.o.) = 0$, thus $Z_n \to 0$ a.s. □

Proposition 4.16. *If $\mu = 1$, then $Z_\infty = 0$ a.s.*

Proof. When $\mu = 1$, $W_n = Z_n \to Z_\infty$ a.s. and $Z_\infty < \infty$ a.s. For any $k > 0$, since k is transient, by Proposition 3.73,

$$\mathbb{P}(Z_\infty = k) = \lim_{n\to\infty} \mathbb{P}(Z_n = k) = \lim_{n\to\infty} p^n(1,k) = 0.$$

Combining $Z_\infty < \infty$ a.s., we conclude that $Z_\infty = 0$ a.s. Another way to illustrate: if $Z_\infty = k$ for some $k > 0$ if and only if there exists $N > 0$ s.t. $Z_n = k$ for all $n \geq N$. However, since $p(k,k) \leq 1 - p(k,0) = 1 - p_0^k < 1$,

$$\mathbb{P}(Z_n = k \text{ for all } n \geq N) = \lim_{n\to\infty} [p(k,k)]^n = 0,$$

so w.p.1., $Z_\infty \neq k$, for any $k > 0$. □

Lemma 4.17. $\mathbb{P}(Z_\infty = 0) = \lim_{n\to\infty} \mathbb{P}(Z_n = 0) = \lim_{n\to\infty} \varphi^n(0)$.

Proof. Since $\{Z_n = 0\} \uparrow \{Z_\infty = 0\}$, we obtain the result by the continuity of probability. □

Lemma 4.18. *Let* $\rho = \inf\{s \in (0, 1] : \varphi(s) = s\}$*, then* $\lim_{n\to\infty} \varphi^n(0) = \rho$.

Proof. Let $\theta_n = \varphi^n(0)$, then $\theta_1 = \varphi(0) = p_0 > 0$. First, we have θ_n is an increasing sequence, because φ is strictly increasing, then $\theta_2 = \varphi(\theta_1) > \varphi(0) = \theta_1$, $\theta_3 = \varphi(\theta_2) > \varphi(\theta_1) = \theta_2$ and so on. Second, $\theta_n \leq \rho$ for all n, because $0 < \rho$, then $\theta_1 = \varphi(0) < \varphi(\rho) = \rho$, $\theta_2 = \varphi(\theta_1) < \varphi(\rho) = \rho$ and so on. By monotone convergence theorem, there is a limit for θ_n, denoted as θ_∞. Take limit on both side of $\theta_{n+1} = \varphi(\theta_n)$, we have $\theta_\infty = \varphi(\theta_\infty)$, since $\theta_\infty \leq \rho$, θ_∞ cannot be other solution of $\varphi(s) = s$ that is larger than ρ, therefore $\theta_\infty = \rho$. □

Proposition 4.19. *If* $0 < \mu \leq 1$*, then* 1 *is the only solution for* $\varphi(s) = s$ *on* $[0, 1]$*. Hence* $\mathbb{P}(Z_\infty = 0) = 1$.

Proposition 4.20. *If* $\mu > 1$*, there is a unique* $\rho \in (0, 1)$ *s.t.* $\varphi(\rho) = \rho$*. Moreover,* $\mathbb{P}(Z_\infty = 0) = \rho$.

Proof. 1.Since φ is increasing and $\varphi'(1) = \mu > 1$, there must be $h \in (0, 1)$ s.t. $\varphi(h) < h$. And $\varphi(0) = p_0 > 0$, so there exists $\rho \in (0, h)$ s.t. $\varphi(\rho) = \rho$.
2. Since $\mu = \mathbb{E}(\xi) > 1$, then $p_k > 0$ for some $k \geq 2$, otherwise $\mu = p_1 < 1$. So $\varphi''(s) > 0$ on $(0, 1)$, i.e. strictly convex.
3. Let $\rho = \inf\{s \in (0, 1) : \varphi(s) = s\}$, then by the property of strictly convex function, for any $s \in (\rho, 1)$, we have $s = \lambda\rho + (1 - \lambda) \cdot 1$ where $\lambda = (1 - s)/(1 - \rho) \in (0, 1)$ and

$$\varphi(s) = \varphi(\lambda\rho + (1 - \lambda) \cdot 1) < \lambda\varphi(\rho) + (1 - \lambda)\varphi(1) = \lambda\rho + (1 - \lambda) \cdot 1 = s,$$

so ρ is the unique solution of $\varphi(s) = s$. □

4.5 Kesten-Stigum Theorem

If $Z_\infty(\omega) = 0$, then $W_\infty(\omega) = 0$. How about the case of non-extinction? What's the probability of $\{\omega : W_\infty(\omega) > 0\}$ if $Z_\infty(\omega) > 0$?

Theorem 4.21 (Kesten and Stigum)**.** *Let $m > 1$, TFAE*

1. $\mathbb{E}(W_\infty) = 1$
2. $\mathbb{P}(W_\infty > 0 | Z_\infty > 0) = 1$
3. $\mathbb{P}(W_\infty = 0) = \rho$
4. $\mathbb{E}(\xi \ln_+ \xi) < \infty$

Here $\ln_+(x) = \ln \max\{1, x\}$, $\rho = \mathbb{P}(Z_\infty = 0)$.

Lemma 4.22. *If* $\mathbb{P}(W_\infty = 0) < 1$, *then* $\mathbb{P}(W_\infty = 0) = \mathbb{P}(Z_\infty = 0)$ *and hence*

$$\{W_\infty > 0\} = \{Z_\infty > 0\} \quad a.s.$$

Proof. Let $\rho = \mathbb{P}(W_\infty = 0)$, conditioning on Z_1, we have

$$\rho = \mathbb{P}(W_\infty = 0) = \sum_{k=0}^{\infty} \mathbb{P}(W_\infty = 0 | Z_1 = k) p_k = \sum_{k=0}^{\infty} p_k [\mathbb{P}(W_\infty = 0)]^k = \varphi(\rho),$$

thus ρ is a root of $\varphi(s) = s$. If $\rho < 1$, by Proposition 4.20, ρ is the only root in $(0, 1)$ and we have

$$\mathbb{P}(W_\infty = 0) = \rho = \mathbb{P}(Z_\infty = 0).$$

Immediately,

$$\mathbb{P}(W_\infty > 0) = \mathbb{P}(Z_\infty > 0).$$

And $\{W_\infty > 0\} \subseteq \{Z_\infty > 0\}$ because for any $\omega \in \{W_\infty > 0\}$, $Z_n(\omega)$ cannot be 0 for some n, otherwise $Z_\infty(\omega) = 0$. We conclude that $\{W_\infty > 0\} = \{Z_\infty > 0\}$ a.s. □

Proposition 4.23. *Let* $m > 1$*, if* $\mathbb{E}(\xi^2) = \sum_{k=1}^{\infty} k^2 p_k < \infty$*, then* $\mathbb{P}(W_\infty = 0) = \rho$*.*

Proof. By Proposition 4.12, $\mathbb{E}(W_\infty) = 1$, which implies $\mathbb{P}(W_\infty = 0) < 1$. Then by Lemma 4.22, $\mathbb{P}(W_\infty = 0) = \rho$ and $\{W_\infty > 0\} = \{Z_\infty > 0\}$ a.s. □

Remark. This is a weaker result than Theorem 4.21.

Now we start to prove Theorem 4.21.

Lemma 4.24. *Define* $f(x) = \mathbb{E}(e^{-xW_\infty})$*. Then* f *satisfies Abel's equation, i.e.*

$$f(x) = \varphi(f(\frac{x}{\mu})).$$

Lemma 4.25. *Let* X *be a r.v. with* $X \geq 0$ *and* $0 < \mathbb{E}(X) = m < \infty$*. Then for any* $a > 0$*,*

$$\int_0^a \frac{1}{u^2}\left[\mathbb{E}(e^{-uX/m}) - e^{-u}\right] \mathrm{d}u < \infty$$

if and only if

$$\mathbb{E}(X|\log X|) < \infty.$$

5 Ergodic theory

5.1 Measure-preserving map

Definition 5.1. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and $\varphi : \Omega \to \Omega$ is a measurable map. We call φ a measure-preserving map, if for any $A \in \mathcal{F}$,

$$\mathbb{P}[\varphi^{-1}(A)] = \mathbb{P}(A).$$

Lemma 5.2. *φ is measure-preserving if and only if for any bounded r.v. X,*

$$\mathbb{E}(X \circ \varphi) = \mathbb{E}(X). \tag{1}$$

If φ is measure-preserving, then (1) also holds for any $X \in L^1$.

Proof. $\Leftarrow$. Take $X = \mathbb{1}_A$ where $A \in \mathcal{F}$, then

$$\mathbb{P}(A) = \mathbb{E}(\mathbb{1}_A) = \mathbb{E}[\mathbb{1}_A(\varphi)] = \mathbb{P}(\omega : \varphi(\omega) \in A) = \mathbb{P}[\varphi^{-1}(A)].$$

$\Rightarrow$. If φ preserves the measure, by the above argument, (1) holds for all indicators $\mathbb{1}_A$, also all simple functions. By approximation of simple functions, (1) holds for all $X \in L^1$. □

5.2 Stationary sequence

Definition 5.3 (stationary sequence)**.** Let $\{X_i : i \in I\}$ be a sequence of random variables where the index set I is closed under addition (e.g. $\mathbb{N}, \mathbb{Z}, \mathbb{R}$). We call it a stationary sequence if for any $k \in I$, $\{X_i : i \in I\}$ and $\{X_{i+k} : i \in I\}$ have the same joint distribution (finite terms have the same distribution).

Lemma 5.4. *Suppose $I = \mathbb{N}$ or $\mathbb{Z}$, then $\{X_i : i \in I\}$ is stationary if and only if $\{X_i : i \in I\}$ and $\{X_{i+1} : i \in I\}$ have the same distribution.*

Example 5.5. Suppose $X = \{X_n : i \geq 0\}$ is a sequence of i.i.d. r.v., then X is stationary.

Proof. For any $m \in \mathbb{N}$, suppose $A_i \in \mathcal{B}(\mathbb{R})$, $0 \leq i \leq m$, then

$$\mathbb{P}(X_0 \in A_0, \cdots, X_m \in A_m) = \prod_{i=0}^{m} \mathbb{P}(X_i \in A_i) = \prod_{i=1}^{m+1} \mathbb{P}(X_i \in A_i)\mathbb{P}(X_1 \in A_0, \cdots, X_{m+1} \in A_m).$$

By π-λ theorem, we have for any $A \in \mathcal{B}(\mathbb{R}^m)$,

$$\mathbb{P}[(X_0, \cdots, X_m) \in A] = \mathbb{P}[(X_1, \cdots, X_{m+1}) \in A].$$

□

Example 5.6. Suppose $X = \{X_n : n \geq 0\}$ is a Markov chain with a unique stationary distribution π. If X_0 has distribution π, then X is stationary.

Proof. For any bounded and $\mathcal{S}^{m+1}$-measurable function f, by Proposition 3.6,

$$\begin{aligned}
\mathbb{E}_\pi[f(X_1, X_2, \cdots, X_{m+1})] &= \int_S f(x_1, x_2, \cdots, x_{m+1})\pi(\,\mathrm{d}x_0) \int_S p(x_0, \mathrm{d}x_1) \cdots \int_S p(x_m, \mathrm{d}x_{m+1}) \\
&= \int_S f(x_1, x_2, \cdots, x_{m+1})\pi(\,\mathrm{d}x_1) \cdots \int_S p(x_m, \mathrm{d}x_{m+1}) \\
&= \int_S f(x_0, x_1, \cdots, x_m)\pi(\,\mathrm{d}x_0) \cdots \int_S p(x_{m-1}, \mathrm{d}x_m) \\
&= \mathbb{E}_\pi[f(X_0, X_1, \cdots, X_m)],
\end{aligned}$$

so $(X_0, \cdots, X_m)$ and $(X_1, \cdots, X_{m+1})$ have the same distribution. □

Proposition 5.7. *Suppose $X = \{X_i : i \geq 0\}$ is stationary and $g : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ is measurable. Define*

$$Y_k = g(\{X_{k+n} : n \geq 0\}),$$

then $Y = \{Y_k : k \geq 0\}$ is a stationary sequence.

Proof. Define $G : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ by

$$G(X_0, X_1, \cdots) = (Y_0, Y_1, \cdots) = (g(X_0, X_1, \cdots), g(X_1, X_2, \cdots), \cdots),$$

obviously, for any $k \geq 0$,

$$G(X_k, X_{k+1}, \cdots) = (Y_k, Y_{k+1}, \cdots).$$

For any bounded and measurable function $f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$, we have

$$\begin{aligned}\mathbb{E}[f(Y_0, Y_1, \cdots, Y_m)] &= \mathbb{E}[f \circ G(X_0, X_1, \cdots)] \\ &= \mathbb{E}[f \circ G(X_1, X_2, \cdots)] \qquad \text{(By } X_n \text{ is stationary)} \\ &= \mathbb{E}[f(Y_1, Y_2, \cdots)],\end{aligned}$$

thus $\{Y_n : n \geq 0\}$ and $\{Y_n : n \geq 1\}$ has the same distribution. □

Proposition 5.8. *Suppose $X = \{X_i : i \geq 0\}$, then X can be extended to a stationary sequence on $\mathbb{Z}$, i.e. there exists a stationary sequence $\tilde{X} = \{\tilde{X}_i : i \in \mathbb{Z}\}$ s.t. $\{\tilde{X}_i : i \geq 0\}$ and $\{X_i : i \geq 0\}$ have the same distribution.*

Proof. For any $n \geq 0$, define

$$\mathbb{P}_n(\tilde{X}_{-n} \in A_{-n}, \tilde{X}_{-n+1} \in A_{-n+1}, \cdots) = \mathbb{P}(X_0 \in A_{-n}, X_1 \in A_{-n+1}, \cdots),$$

then $\mathbb{P}_n, n \geq 0$ is consistent because

$$\begin{aligned}&\mathbb{P}_{n+1}(\tilde{X}_{-n-1} \in \mathbb{R}, \tilde{X}_{-n} \in A_{-n}, \tilde{X}_{-n+1} \in A_{-n+1}, \cdots) \\ =\,& \mathbb{P}(X_0 \in \mathbb{R}, X_1 \in A_{-n}, X_2 \in A_{-n+1}, \cdots) \\ =\,& \mathbb{P}(X_0 \in A_{-n}, X_1 \in A_{-n+1}, \cdots) \\ =\,& \mathbb{P}_n(\tilde{X}_{-n} \in A_{-n}, \tilde{X}_{-n+1} \in A_{-n+1}, \cdots).\end{aligned}$$

By Kolmogorov's extension theorem, there exists probability measure $\tilde{\mathbb{P}}$ s.t.

$$\tilde{\mathbb{P}}(\tilde{X}_{-n} \in A_{-n}, \tilde{X}_{-n+1} \in A_{-n+1}, \cdots) = \mathbb{P}_n(\tilde{X}_{-n} \in A_{-n}, \tilde{X}_{-n+1} \in A_{-n+1}, \cdots).$$

By the construction, $\{\tilde{X}_i : i \geq 0\}$ and $\{X_i : i \geq 0\}$ have the same distribution. To show $\{\tilde{X}_i : i \geq 0\}$ is stationary, we only need to show the negative integer part, for any $m, n \geq 0$

$$\begin{aligned}
&\tilde{\mathbb{P}}(\tilde{X}_{-m+1} \in A_{-m}, \tilde{X}_{-m+2} \in A_{-m+1}, \cdots, \tilde{X}_{n+1} \in A_n) \\
&= \mathbb{P}(X_0 \in A_{-m}, X_1 \in A_{-m+1}, \cdots, X_{m+n} \in A_n) \\
&= \tilde{\mathbb{P}}(\tilde{X}_{-m} \in A_{-m}, \tilde{X}_{-m+1} \in A_{-m+1}, \cdots, \tilde{X}_n \in A_n).
\end{aligned}$$

□

Proposition 5.9. *Suppose $\varphi : \Omega \to \Omega$ is a measure-preserving map on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\varphi^0 := \mathrm{id}$, $\varphi^n = \varphi \circ \varphi^{n-1}$. For any $X \in \mathcal{F}$, define $X_n := X \circ \varphi^n$, then $\{X_n : n \geq 0\}$ is stationary.*

Proof. For any bounded and measurable function $f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}[f(X_0, X_1, \cdots)] &= \mathbb{E}[f(X(\omega), X(\varphi(\omega)), X(\varphi^2(\omega)), \cdots)] \\
&= \mathbb{E}[F_X(\omega))] \qquad \text{here we define } F_X(\omega) := f(X(\omega), X \circ \varphi(\omega), X \circ \varphi^2(\omega), \cdots) \\
&= \mathbb{E}[F_X \circ \varphi(\omega)] \qquad \text{(By Lemma 5.2)} \\
&= \mathbb{E}[f(X(\varphi(\omega)), X(\varphi^2(\omega)), X(\varphi^3(\omega)), \cdots)] \\
&= \mathbb{E}[f(X_1, X_2, \cdots)],
\end{aligned}$$

therefore $\{X_n : n \geq 0\}$ and $\{X_n : n \geq 1\}$ have the same distribution. □

Proposition 5.10. *Suppose $\{Y_n : n \geq 0\}$ is a stationary real-valued r.v. sequence, then there exists a measure-preserving map $\varphi : \Omega \to \Omega$ and $X \in \mathcal{F}$ s.t. $\{X_n : n \geq 0\}$ and $\{Y_n : n \geq 0\}$ have the same distribution where $X_n = X \circ \varphi$.*

Proof. First, $(Y_0, Y_1, \cdots, Y_m)$ defines a probability measure $\mathbb{P}_m$ on $\mathcal{B}(\mathbb{R}^{m+1})$ by

$$\mathbb{P}_m(A) = \mathbb{P}((Y_0, Y_1, \cdots, Y_m) \in A),$$

and $\mathbb{P}_m, m \geq 0$ is obviously consistent, then by Kolmogorov's extension theorem, there exists a probability measure $\tilde{\mathbb{P}}$ on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ s.t. for any $A \in \mathcal{B}(\mathbb{R}^{m+1})$,

$$\tilde{\mathbb{P}}(A) = \mathbb{P}_m(A).$$

For any $\omega = (\omega_0, \omega_1, \cdots) \in \mathbb{R}^{\mathbb{N}}$, define $X(\omega) = \omega_0$, and shift operator

$$\varphi = \theta_1 : (\omega_0, \omega_1, \cdots) \mapsto (\omega_1, \omega_2, \cdots),$$

then we have $X_n(\omega) = X \circ \varphi^n(\omega) = \omega_n$.

φ is measure-preserving because for any $A \in \mathcal{B}(\mathbb{R}^{m+1})$,

$$\begin{aligned}\tilde{\mathbb{P}}(\varphi^{-1}(A)) &= \mathbb{P}_m[(Y_0, Y_1, \cdots, Y_m) \in \varphi^{-1}(A)] \\ &= \mathbb{P}_m[(Y_1, \cdots, Y_{m+1}) \in A] \\ &= \mathbb{P}_m[(Y_0, \cdots, Y_m) \in A] \\ &= \tilde{\mathbb{P}}(A).\end{aligned}$$

$\{X_n : n \geq 0\}$ and $\{Y_n : n \geq 0\}$ have the same distribution because for any $A \in \mathcal{B}(\mathbb{R}^{m+1})$,

$$\begin{aligned}\tilde{\mathbb{P}}((X_0, X_1, \cdots, X_m) \in A) &= \mathbb{P}((\omega_0, \omega_1, \cdots, \omega_m) \in A) \\ &= \tilde{\mathbb{P}}(A) \\ &= \mathbb{P}((Y_0, Y_1, \cdots, Y_m) \in A).\end{aligned}$$

□

5.3 Ergodicity

Definition 5.11. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and $\varphi : \Omega \to \Omega$ is a measure-preserving map. We call event $A \in \mathcal{F}$ invariant if $\varphi^{-1}(A) = A$. We call φ ergodic if for any invariant event A, we have $\mathbb{P}(A) \in \{0, 1\}$.

Definition 5.12. Suppose $\{X_n : n \geq 0\}$ is a stationary sequence, we call it ergodic if the induced measure-preserving map (shift operator) in Proposition 5.10 is ergodic.

Lemma 5.13. *Set of invariant events $\mathcal{I} := \{A \in \mathcal{F} : \varphi^{-1}(A) = A\}$ is a σ-field. X is $\mathcal{I}$-measurable if and only if $X \circ \varphi = X$ a.s.*

Proposition 5.14. *Suppose $\varphi : \Omega \to \Omega$ is a measure-preserving map on $(\Omega, \mathcal{F}, \mathbb{P})$, TFAE*

1. *φ is ergodic;*
2. *For any $A \in \mathcal{F}$, $\mathbb{P}(A \triangle \varphi^{-1}(A)) = 0$ implies $\mathbb{P}(A) \in \{0, 1\}$;*
3. *For any $A \in \mathcal{F}$, $\mathbb{P}(A) > 0$ implies*
$$\mathbb{P}(\bigcup_{n=1}^{\infty} \varphi^{-n}(A)) = 1;$$
4. *(mixing) For any $A, B \in \mathcal{F}$,*
$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{P}(\varphi^{-1}(A) \cap B) = \mathbb{P}(A)\mathbb{P}(B);$$
5. *For any $A, B \in \mathcal{F}$ with $\mathbb{P}(A) > 0$ and $\mathbb{P}(B) > 0$, there exists $n \geq 1$ s.t. $\mathbb{P}(\varphi^{-1}(A) \cap B) > 0$;*
6. *For any $X \in L^2$, $X \circ \varphi = X$ a.s. implies $f = C$ a.s. where C is a constant.*

Example 5.15. Suppose $\{X_n : n \geq 0\}$ is a sequence of i.i.d. r.v. Let $(\Omega = \mathbb{R}^{\mathbb{N}}, \mathcal{F}, \mathbb{P})$ be the probability space s.t. for any $\omega \in \Omega$, $X_n(\omega) = \omega_n$. Then the shift operator φ on Ω is ergodic.

Proof. Suppose $A \in \mathcal{F}$ is invariant, then $A = \varphi^{-1}(A)$, i.e.

$$A = \{\omega : \omega \in A\} = \{\omega : \varphi(\omega) \in A\} \in \sigma(X_1, X_2, \cdots),$$

By iteration, we have

$$A = \{\omega : \varphi^n(\omega) \in A\} \in \sigma(X_n, X_{n+1}, \cdots),$$

thus

$$A \in \mathcal{T} = \bigcap_{k=0}^{\infty} \sigma(X_n : n \geq k).$$

By Kolmogorov's 0-1 law, we have $\mathbb{P}(A) \in \{0, 1\}$, therefore φ is ergodic. □

Example 5.16. Suppose $\{X_n : n \geq 0\}$ is a Markov chain on a countable state space S with a stationary distribution π ($\pi(x) > 0$ for all $x \in S$). Then the induced shift operator φ is ergodic if and only if X_n is irreducible.

5.4 Birkhoff's Ergodic Theorem

In this section, we always suppose φ is a measure preserving map on $(\Omega, \mathcal{F}, \mathbb{P})$.

Theorem 5.17 (Birkhoff's Ergodic Theorem)**.** *For any $X \in L^1$,*

$$\frac{1}{n} \sum_{k=0}^{n-1} X(\varphi^k) \to \mathbb{E}(X|\mathcal{I}) \quad a.s. \text{ and in } L^1.$$

Lemma 5.18 (Maximal ergodic lemma)**.** *Let $X_k(\omega) = X(\varphi^k(\omega))$ for $k \in \mathbb{N}$ and $\omega \in \Omega$. Define*

$$S_n(\omega) = \sum_{k=0}^{n-1} X_k(\omega),$$

and

$$M_n(\omega) = \max\{0, S_1(\omega), \cdots, S_n(\omega)\}.$$

Then $\mathbb{E}(X\mathbb{1}_{\{M_n>0\}}) \geq 0$ *for all* $n \in \mathbb{Z}_+$.

Proof. For $1 \leq k \leq n$,

$$M_n \circ \varphi(\omega) \geq S_k \circ \varphi(\omega),$$

then

$$X(\omega) + M_n \circ \varphi(\omega) \geq X(\omega) + S_k \circ \varphi(\omega) = S_{k+1}(\omega),$$

thus

$$X(\omega) \geq S_{k+1}(\omega) - M_n \circ \varphi(\omega), \quad \forall 1 \leq k \leq n. \tag{1}$$

Since $M_n \circ \varphi(\omega) \geq 0$, we have

$$X(\omega) + M_n \circ \varphi(\omega) \geq X(\omega) = X_0(\omega) = S_1(\omega),$$

i.e. $X(\omega) \geq S_1(\omega) - M_n \circ \varphi(\omega)$. Therefore,

$$\mathbb{E}(X\mathbb{1}_{\{M_n>0\}}) \geq \mathbb{E}[(S_k - M_n \circ \varphi)\mathbb{1}_{\{M_n>0\}}], \quad \forall\, 1 \leq k \leq n,$$

then

$$\begin{aligned}\mathbb{E}(X\mathbb{1}_{\{M_n>0\}}) &\geq \mathbb{E}\left[(\max_{1\leq k\leq n} S_k - M_n \circ \varphi)\mathbb{1}_{\{M_n>0\}}\right] \\ &= \mathbb{E}\left[(M_n - M_n \circ \varphi)\mathbb{1}_{\{M_n>0\}}\right] \\ &\geq \mathbb{E}\left[M_n - M_n \circ \varphi\right],\end{aligned}$$

the last inequality holds because

$$\mathbb{E}(M_n) = \mathbb{E}(M_n\mathbb{1}_{\{M_n>0\}}) + \mathbb{E}(M_n\mathbb{1}_{\{M_n\leq 0\}}) = \mathbb{E}(M_n\mathbb{1}_{\{M_n>0\}}) + 0 = \mathbb{E}(M_n\mathbb{1}_{\{M_n>0\}}),$$

and

$$\mathbb{E}(M_n \circ \varphi) = \mathbb{E}(M_n \circ \varphi \mathbb{1}_{\{M_n > 0\}}) + \mathbb{E}(M_n \circ \varphi \mathbb{1}_{\{M_n \le 0\}}) \ge \mathbb{E}(M_n \circ \varphi \mathbb{1}_{\{M_n > 0\}}).$$

Finally, since φ is measure preserving, by Lemma 5.2,

$$\mathbb{E}\left[M_n - M_n \circ \varphi\right] = 0.$$

□

Proof of Theorem 5.17. 1. We only need to prove the case when $\mathbb{E}(X|\mathcal{I}) = 0$, i.e.

$$\frac{S_n}{n} \to 0, \quad \text{a.s. and in } L^1.$$

2. Define

$$\bar{X} = \limsup \frac{S_n}{n},$$

and let $\varepsilon > 0$, define $D = \{\omega : \bar{X}(\omega) > \varepsilon\}$. Our goal is to prove $\mathbb{P}(D) = 0$.

3.Since $\bar{X}(\varphi(\omega)) = \bar{X}(\omega)$, we have

$$\varphi^{-1}(D) = \{\varphi^{-1}(\omega) : \bar{X}(\omega) > \varepsilon\} = \{\omega : \bar{X}(\varphi(\omega)) > \varepsilon\} = D,$$

thus $D \in \mathcal{I}$.

4. Let $X^*(\omega) = (X(\omega) - \varepsilon)\mathbb{1}_D(\omega)$,

$$S_n^*(\omega) = X^*(\omega) + \cdots + X^*(\varphi^{n-1}(\omega)),$$

$$M_n^*(\omega) = \max\{0, S_1^*(\omega), \cdots, S_n^*(\omega)\},$$

$F_n = \{\omega : M_n^*(\omega) > 0\}$, and

$$F = \bigcup_{n=1}^{\infty} F_n = \{\sup_{k\ge 1} \frac{S_k^*}{k} > 0\}.$$

Then $F = D$.

5.$\mathbb{E}(X^* \mathbb{1}_F) \geq 0$.

6. From Step 5,

$$0 \leq \mathbb{E}(X^* \mathbb{1}_D) = \mathbb{E}((X - \varepsilon)\mathbb{1}_D) = \mathbb{E}(X \mathbb{1}_D) - \varepsilon \mathbb{P}(D) = \mathbb{E}(\mathbb{E}(X|\mathcal{I})\mathbb{1}_D) - \varepsilon \mathbb{P}(D) = -\varepsilon \mathbb{P}(D),$$

then $\mathbb{P}(D) = 0$. Therefore,

$$\limsup \frac{S_n}{n} \leq 0, \quad a.s.$$

Similarly,

$$\liminf \frac{S_n}{n} \geq 0, \quad a.s.$$

thus

$$\frac{S_n}{n} \to 0, \quad a.s.$$

7. L^p $(p \geq 1)$ convergence.

Take $M > 0$, let $X'_M = X \mathbb{1}_{\{|X| \leq M\}}, \quad X''_M = X - X'_M$. For X'_M, by the above proof,

$$\frac{1}{n} \sum_{m=0}^{n-1} X'_M(\varphi^m \omega) - \mathbb{E}(X'_M|\mathcal{I}) \to 0 \quad a.s.$$

and

$$\begin{aligned}\left| \frac{1}{n} \sum_{m=0}^{n-1} X'_M(\varphi^m \omega) - \mathbb{E}(X'_M|\mathcal{I}) \right|^p &\leq \left(\frac{1}{n} \sum_{m=0}^{n-1} \left| X'_M(\varphi^m \omega) \right| + \mathbb{E}(|X'_M||\mathcal{I}) \right)^p \\ &\leq \left(\left| \frac{1}{n} \sum_{m=0}^{n-1} M \right| + |M| \right)^p \\ &= (2M)^p,\end{aligned}$$

then by the bounded convergence theorem,

$$\mathbb{E}\left|\frac{1}{n}\sum_{m=0}^{n-1} X'_M(\varphi^m\omega) - \mathbb{E}(X'_M|\mathcal{I})\right|^p \to 0.$$

For X''_M, we have

$$\begin{aligned}\left(\mathbb{E}\left|\frac{1}{n}\sum_{m=0}^{n-1} X''_M(\varphi^m\omega) - \mathbb{E}(X''_M|\mathcal{I})\right|^p\right)^{1/p} &\le \left(\mathbb{E}\left|\frac{1}{n}\sum_{m=0}^{n-1} X''_M(\varphi^m\omega)\right|^p\right)^{1/p} + \left(\mathbb{E}\left|\mathbb{E}(X''_M|\mathcal{I})\right|^p\right)^{1/p} \\ &\le \left(\frac{1}{n}\sum_{m=0}^{n-1}\mathbb{E}\left|X''_M(\varphi^m\omega)\right|^p\right)^{1/p} + \left(\mathbb{E}[\mathbb{E}(|X''_M|^p|\mathcal{I})]\right)^{1/p} \\ &= 2(\mathbb{E}|X''_M|^p)^{1/p}.\end{aligned}$$

Therefore

$$\begin{aligned}&\limsup_{n\to\infty}\left(\mathbb{E}\left|\frac{1}{n}\sum_{m=0}^{n-1} X(\varphi^m\omega) - \mathbb{E}(X|\mathcal{I})\right|^p\right)^{1/p} \\ &\le \limsup_{n\to\infty}\left(\mathbb{E}\left|\frac{1}{n}\sum_{m=0}^{n-1} X'_M(\varphi^m\omega) - \mathbb{E}(X'_M|\mathcal{I})\right|^p\right)^{1/p} + \limsup_{n\to\infty}\left(\mathbb{E}\left|\frac{1}{n}\sum_{m=0}^{n-1} X''_M(\varphi^m\omega) - \mathbb{E}(X''_M|\mathcal{I})\right|^p\right)^{1/p} \\ &\le 2(\mathbb{E}|X''_M|^p)^{1/p},\end{aligned}$$

since M is arbitrary, let $M\to\infty$, the above limit then goes to 0, now L^p convergence is proved. □

5.5 Recurrence

Theorem 5.19. *Let $\{X_n : n\ge 1\}$ be a stationary sequence with $X_i:\Omega\to\mathbb{R}^d$. Let*

$$S_n = \sum_{k=1}^{n} X_k,$$

$$A = \{\omega : S_n(\omega)\neq 0 \quad \forall n\ge 1\},$$

i.e. the set of trajectories that never hit 0*. Let* $R_n = \#\{S_1, \cdots, S_n\}$ *be the number of points (without repeat) visited by time* n*. Then*

$$\frac{R_n}{n} \to \mathbb{E}(\mathbb{1}_A|\mathcal{I}) \quad a.s.$$

as $n \to \infty$*.*

5.6 Subadditive ergodic theorem

Theorem 5.20. *Suppose* $X_{m,n}$, $0 \le m < n$*, is a r.v. series satisfying*

(i) $X_{0,m} + X_{m,n} \ge X_{0,n}$

(ii) $\{X_{nk,(n+1)k}, n \ge 1\}$ *is a stationary sequence for each* $k \ge 1$

(iii) The distribution of $\{X_{m,m+k} : k \ge 1\}$ *does not depend on* m

(iv) $\mathbb{E}(X_{0,1}^+) < \infty$ *and for each* n*,* $\mathbb{E}(X_{0,n}) \ge \gamma_0 n$ *for some* $\gamma_0 > -\infty$

Then there exists $\gamma \in \mathbb{R}$ *and r.v.* $X \in L^1$ *s.t.*

(a)

$$\lim_{n\to\infty} \frac{\mathbb{E}(X_{0,n})}{n} = \inf_n \frac{\mathbb{E}(X_{0,n})}{n} = \gamma$$

(b)

$$\frac{X_{0,n}}{n} \to X \quad a.s. \text{ and in } L^1,$$

and $\mathbb{E}(X) = \gamma$

(c) if all the stationary sequences in (ii) are ergodic, then

$$X = \gamma \quad a.s.$$

6 Brownian motion

Brownian motion is a Gaussian Markov process with stationary independent increments.

6.1 Definition and simple properties

Definition 6.1 (First definition of Brownian motion)**.** A real-valued process B_t, or written as $B(t)$, $t \in [0, \infty)$ is called a Brownian motion if

(1) (Independent increment) For any $0 \leq t_0 < t_1 < \cdots < t_n$,

$$B(t_0),\ B(t_1) - B(t_0),\ \cdots, B(t_n) - B(t_{n-1})$$

are independent;

(2) For any $s, t \in [0, \infty)$,

$$B_{s+t} - B_s \sim \mathcal{N}(0, t);$$

(3) With probability 1, $t \to B_t$ is continuous.

Proposition 6.2 (Translation invariance)**.** $\{B_t - B_0, t \geq 0\}$ *is independent of* B_0 *and has the same distribution as Brownian motion* $\{\tilde{B}_t, t \geq 0\}$ *with* $\tilde{B}_0 = 0$*.*

Proof. 1. Let $\mathcal{A}_1 = \sigma(B_0) = \sigma(\{B_0 \in A_0\}, A_0 \in \mathcal{B}(\mathbb{R}))$, and $\mathcal{A}_2$ be the set of events of the following form

$$\{B_{t_1} - B_0 \in A_1, \cdots, B_{t_n} - B_{t_{n-1}} \in A_n\},$$

where $A_i \in \mathcal{B}(\mathbb{R})$. Then $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent by the property of independent increment. They are also both π-system. Then $\sigma(\mathcal{A}_1)$ and $\sigma(\mathcal{A}_2)$ are independent.

2. Claim: $\sigma(\mathcal{A}_2) = \sigma(\{B_t - B_0 : t \geq 0\})$.

We can show $\sigma(B_{t_1} - B_0, B_{t_2} - B_{t_1}, \cdots, B_{t_n} - B_{t_{n-1}}) = \sigma(B_{t_1} - B_0, B_{t_2} - B_0, \cdots, B_{t_n} - B_0)$.

Take the union over all $0 < t_1 < \cdots < t_n$, we have $\sigma(\mathcal{A}_2) = \sigma(\{B_t - B_0 : t \geq 0\})$. This Claim is proved. Therefore $\{B_t - B_0, t \geq 0\}$ is independent of B_0.

3. For $0 < t_1 < \cdots < t_n$, we have

$$(B_{t_1} - B_0, B_{t_2} - B_{t_1}, \cdots, B_{t_n} - B_{t_{n-1}})$$

has the same distribution as

$$(\tilde{B}_{t_1}, \tilde{B}_{t_2} - \tilde{B}_{t_1}, \cdots, \tilde{B}_{t_n} - \tilde{B}_{t_{n-1}}),$$

therefore,

$$\begin{aligned}\sigma(B_{t_1} - B_0, B_{t_2} - B_0, \cdots, B_{t_n} - B_0) &= \sigma(B_{t_1} - B_0, B_{t_2} - B_{t_1}, \cdots, B_{t_n} - B_{t_{n-1}}) \\ &= \sigma(\tilde{B}_{t_1}, \tilde{B}_{t_2} - \tilde{B}_{t_1}, \cdots, \tilde{B}_{t_n} - \tilde{B}_{t_{n-1}}) \\ &= \sigma(\tilde{B}_{t_1}, \tilde{B}_{t_2}, \cdots, \tilde{B}_{t_n}),\end{aligned}$$

which means $\{B_t - B_0 : t \geq 0\}$ and $\{\tilde{B}_t : t \geq 0\}$ have the same finite dimensional distribution, thus they have the same distribution. □

Proposition 6.3 (Scaling relation)**.** *Suppose $\{B_t : t \geq 0\}$ is a Brownian motion with $B_0 = 0$, then for any $t > 0$, $\{B_{st} : s \geq 0\}$ and $\{t^{1/2} B_s : s \geq 0\}$ have the same distribution.*

Proof. We need to show they have the same finite dimensional distribution. Let $s_1 > 0$, then

$$B_{s_1 t} \sim \mathcal{N}(0, s_1 t)$$

and

$$t^{1/2} B_{s_1} \sim t^{1/2} \mathcal{N}(0, s_1) = \mathcal{N}(0, s_1 t),$$

so $B_{s_1 t}$ and $t^{1/2} B_{s_1}$ has the same distribution. Let $0 < s_1 < s_2$, then $X = (B_{s_1 t}, B_{s_2 t} - B_{s_1 t})^T$

is multivariant Gaussian with

$$\mathbb{E}(X) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \Sigma(X) = \begin{pmatrix} s_1 t & 0 \\ 0 & (s_2 - s_1)t \end{pmatrix},$$

and $Y = (t^{1/2}B_{s_1}, t^{1/2}B_{s_2} - t^{1/2}B_{s_1})^T$ is also multivariant Gaussian with the same mean and covariance matrix. By the property of multivariant Gaussian distribution, X and Y have the same distribution. Thus

$$\begin{pmatrix} B_{s_1 t} \\ B_{s_2 t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} X, \quad \begin{pmatrix} t^{1/2}B_{s_1} \\ t^{1/2}B_{s_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} Y$$

has the same distribution. □

Definition 6.4 (Second definition of Brownian motion)**.** A real-valued process $\{B_t, t \in [0,\infty)\}$ with $B_0 = 0$ is called Brownian motion if

(1′) B_t is a Gaussian process, i.e. for any $t_0, t_1, \cdots, t_n$,

$$(B(t_0), B(t_1), \cdots, B(t_n))$$

is a multivariant Gaussian distribution.

(2′) For any $s, t \in [0,\infty)$, $\mathbb{E}(B_s) = 0$, and

$$\mathbb{E}(B_s B_t) = s \wedge t;$$

(3′) With probability 1, $t \to B_t$ is continuous.

Proposition 6.5. *The second definition is equivalent to the first definition with* $B_0 = 0$.

Proof. (1)(2) $\Longrightarrow$ (1′). Notice that

$$B(t_i) = B(t_0) + B(t_1) - B(t_0) + \cdots + B(t_i) - B(t_{i-1}),$$

then

$$\begin{pmatrix} B(t_0) \\ B(t_1) \\ B(t_2) \\ \vdots \\ B(t_n) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{pmatrix} \begin{pmatrix} B(t_0) \\ B(t_1) - B(t_0) \\ B(t_2) - B(t_1) \\ \vdots \\ B(t_n) - B(t_{n-1}) \end{pmatrix}$$

where $(B(t_0), B(t_1) - B(t_0), \cdots, B(t_n) - B(t_{n-1}))^T$ is multivariant Gaussian, thus its linear transformation $(B(t_0), B(t_1), \cdots, B(t_n))^T$ is also multivariant Gaussian.

(1)(2)$\Longrightarrow$ (2′). First,

$$\mathbb{E}(B_s) = \mathbb{E}(B_s - B_0) + \mathbb{E}(B_0) = 0,$$

second, suppose $s < t$,

$$\mathbb{E}(B_s B_t) = \mathbb{E}(B_s(B_t - B_s)) + \mathbb{E}(B_s^2) = s.$$

(1′)(2′)$\Longrightarrow$(1). For any $t_0 < t_1 < \cdots < t_n$, since

$$\begin{pmatrix} B(t_0) \\ B(t_1) - B(t_0) \\ B(t_2) - B(t_1) \\ \vdots \\ B(t_n) - B(t_{n-1}) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix} \begin{pmatrix} B(t_0) \\ B(t_1) \\ B(t_2) \\ \vdots \\ B(t_n) \end{pmatrix}$$

$(B(t_0), B(t_1) - B(t_0), \cdots, B(t_n) - B(t_{n-1}))^T$ is multivariant Gaussian. For $k < j$,

$$\begin{aligned}\mathrm{Cov}(B(t_k) - B(t_{k-1}), B(t_j) - B(t_{j-1})) &= \mathbb{E}[(B_{t_k} - B_{t_{k-1}})(B_{t_j} - B_{t_{j-1}})] \\ &= \mathbb{E}[B_{t_k}B_{t_j} + B_{t_{k-1}}B_{t_{j-1}} - B_{t_k}B_{t_{j-1}} - B_{t_{k-1}}B_{t_j}] \\ &= t_k + t_{k-1} - t_k - t_{k-1} = 0,\end{aligned}$$

thus the covariance matrix of Gaussian $(B(t_0), B(t_1) - B(t_0), \cdots, B(t_n) - B(t_{n-1}))^T$ is diagonal, which implies $B(t_0), B(t_1) - B(t_0), \cdots, B(t_n) - B(t_{n-1})$ are independent.
$(1')(2')\Longrightarrow(2)$. For any $s, t \geq 0$, $B_{s+t} - B_s$ is the linear combination of two Gaussian distributions, thus it is also Gaussian,

$$\mathbb{E}(B_{s+t} - B_s) = 0, \quad \mathrm{Var}(B_{s+t} - B_s) = \mathbb{E}[(B_{s+t} - B_s)^2] = t,$$

so $B_{s+t} - B_s \sim \mathcal{N}(0, t)$. □

6.2 Construction

Theorem 6.6. *Define*

$$\Omega_0 = \{functions\ \omega : [0, \infty) \to \mathbb{R}\},$$

and

$$\mathcal{F}_0 = \sigma(\{\omega : \omega(t_i) \in A_i, 1 \leq i \leq n, \}),$$

where $A_i \in \mathcal{B}$. Then for any $x \in \mathbb{R}$, there exists a unique probability measure ν_x on $(\Omega_0, \mathcal{F}_0)$, s.t.

- $\nu_x(\{\omega : \omega(0) = x\}) = 1$;
- $\nu_x(\{\omega : \omega(t_1) \in A_1, \cdots, \omega(t_n) \in A_n\}) = \mu_{x,t_1,\cdots,t_n}(A_1 \times A_2 \times \cdots \times A_n)$,

where

$$\mu_{x,t_1,\cdots,t_n}(A_1\times A_2\times\cdots\times A_n)=\int_{A_1}p_{t_1}(x,x_1)\,\mathrm{d}x_1\int_{A_2}p_{t_2-t_1}(x_1,x_2)\,\mathrm{d}x_2\cdots\int_{A_n}p_{t_n-t_{n-1}}(x_{n-1},x_n)\,\mathrm{d}x_n,$$

and

$$p_t(x,y)=\frac{1}{\sqrt{2\pi t}}e^{-\frac{(x-y)^2}{2t}}.$$

Proof. Check consistency and apply Kolmogorov's extension theorem. □

Remark. Although the construction in Theorem 6.6 satisfies Definition (1)(2) and (3), it fails to satisfy (4). Specifically, if $C=\{\omega:[0,\infty)\to\mathbb{R}\text{ is continuous}\}$, then $C\notin\mathcal{F}_0$. Actually, Ω_0 is too large and $\mathcal{F}_0$ is too coarse.

Lemma 6.7. *For any $\Gamma\in\mathcal{F}_0$, there is a countable set $S=S_\Gamma\subseteq[0,\infty)$ s.t. for any $\omega\in\Omega_0$, $\gamma\in\Gamma$ satisfying*

$$\omega\big|_S=\gamma\big|_S\ ,$$

we have $\omega\in\Gamma$.

Proof. Let $\Sigma=\{\Gamma\subseteq\Omega_0:\text{the above property holds for some countable set }S\subseteq[0,\infty)\}$.
Claim: Σ is a σ-algebra.
For any $t\in[0,\infty)$ and $A\in\mathcal{B}(\mathbb{R})$, $B_t^{-1}(A)\subseteq\Sigma$ because we can choose $S=\{t\}$. Let

$$\mathcal{A}=\{B_t^{-1}(A):A\in\mathcal{B}(\mathbb{R}),t\in[0,\infty)\},$$

which is a π-system, and Σ is a λ-system, then by π-λ theorem,

$$\mathcal{F}_0=\sigma(\mathcal{A})\subseteq\Sigma.$$

Corollary 6.8. $C\notin\mathcal{F}_0$.

Proof. Suppose $C \in \mathcal{F}_0$, , let $S \subseteq [0,\infty)$ be the countable set in the Lemma, for a fixed continuous function $f \in \Omega_0$, choose $t_0 \in [0,\infty) \setminus S$, define $\omega \in \Omega$ by

$$\omega(t) = \begin{cases} f(t) & \text{if } t \neq t_0 \\ f(t_0)+1 & \text{if } t = t_0 \end{cases}$$

then t_0 is a removable discontinuity point for ω, thus $\omega \notin C$, which contradicts the above Lemma! □

In order to construct the Brownian motion that satisfies all properties in the definition, we need some preparation. The basic idea is to construct the path on the dense set $\mathbb{Q}_2$ first, then extend it to $[0,\infty)$.

Theorem 6.9 (Kolmogorov's continuity theorem)**.** *Suppose* $\{X_t, t \in [0,1]\}$ *is a process defined on* $(\Omega, \mathcal{F}, \mathbb{P})$*, s.t. for any* $s, t \in [0,1]$*,*

$$\mathbb{E}(|X_t - X_s|^\beta) \leq K|t-s|^{1+\alpha},$$

where $\alpha, \beta > 0$*. If* $0 < \gamma < \dfrac{\alpha}{\beta}$*, then with probability* 1*, there exists a constant* C*, s.t. for any* $q, r \in \mathbb{Q}_2 \cap [0,1]$*,*

$$|X(q) - X(r)| \leq C|q-r|^\gamma.$$

Proof. 1. Let

$$G_n = \{\omega : \left| X(\frac{i}{2^n}) - X(\frac{i-1}{2^n}) \right| \leq 2^{-\gamma n},\ \forall\, 0 < i \leq 2^n\}.$$

We want to show G_n holds for any large n with probability 1. Notice that

$$G_n^c = \{\left| X(\frac{i}{2^n}) - X(\frac{i-1}{2^n}) \right| > 2^{-\gamma n}, \text{ for some } 0 < i \leq 2^n\} \subseteq \bigcup_{i=1}^{2^n} \{\left| X(\frac{i}{2^n}) - X(\frac{i-1}{2^n}) \right| > 2^{-\gamma n}\},$$

so

$$\begin{aligned}
\mathbb{P}(G_n^c) &= \sum_{i=1}^{2^n} \mathbb{P}\left(\left|X(\frac{i}{2^n}) - X(\frac{i-1}{2^n})\right| > 2^{-\gamma n}\right) \\
&\le \sum_{i=1}^{2^n} (2^{-\gamma n})^{-\beta} \mathbb{E}\left(\left|X(\frac{i}{2^n}) - X(\frac{i-1}{2^n})\right|^{\beta}\right) \qquad \text{(Chebyshev's inequality)} \\
&\le \sum_{i=1}^{2^n} 2^{\beta\gamma n} \cdot K \left|\frac{i}{2^n} - \frac{i-1}{2^n}\right|^{1+\alpha} \\
&\le \sum_{i=1}^{2^n} 2^{\beta\gamma n} \cdot K 2^{-n(1+\alpha)} \\
&= 2^n \cdot 2^{\beta\gamma n} \cdot K 2^{-n(1+\alpha)} \\
&= K \cdot 2^{-n\lambda},
\end{aligned}$$

where $\lambda = \alpha - \beta\gamma > 0$.

Let $H_N = \bigcap_{n=N}^{\infty} G_n$, then $H_N^c = \bigcup_{n=N}^{\infty} G_n^c$,

$$\mathbb{P}(H_N^c) \le \sum_{n=N}^{\infty} \mathbb{P}(G_n^c) \le \sum_{n=N}^{\infty} K \cdot 2^{-n\lambda} = \frac{K 2^{-N\lambda}}{1 - 2^{-\lambda}},$$

thus

$$\sum_{N=1}^{\infty} \mathbb{P}(H_N^c) \le \frac{K}{1 - 2^{-\lambda}} \cdot \frac{2^{-\lambda}}{1 - 2^{-\lambda}} < \infty,$$

by Borel-Cantelli lemma, $\mathbb{P}(H_N^c, \, i.o.) = 0$. Hence for almost sure $\omega \in \Omega$, ω is only in finitely many H_N^c, in other words, there exists $N_0(\omega)$ s.t. whenever $N \ge N_0$, $\omega \notin H_N^c$, i.e. $\omega \in H_N = \bigcap_{n=N}^{\infty} G_n$.

2. On H_N, we have for all $q, r \in \mathbb{Q}_2 \cap [0,1]$ with $|q - r| < 2^{-N}$,

$$|X(q) - X(r)| \le \frac{3}{1 - 2^{-\gamma}} |q - r|^{\gamma}.$$

3. From Step 1, for almost sure ω, for $q, r \in \mathbb{Q}_2 \cap [0,1]$, we have $|q - r| < \delta(\omega) = 2^{-N_0(\omega)}$,

then from Step 2,

$$|X(q) - X(r)| \leq A|q - r|^{\gamma}.$$

4. We want to extend the above equation to all $q, r \in \mathbb{Q}_2 \cap [0,1]$. Suppose $r - q > \delta(\omega)$, let $S_0 = q < s_1 < \cdots < s_k = r$ with $|s_i - s_{i+1}| = \dfrac{r-q}{k} \leq \delta(\omega)$ (thus $k \geq \dfrac{r-q}{\delta} > 1$), then

$$|X(q) - X(r)| \leq \sum_{i=1}^{k} |X(s_i) - X(s_{i-1})| \leq A \sum_{i=1}^{k} |s_i - s_{i-1}|^{\gamma} = A \sum_{i=1}^{k} \left|\frac{q-r}{k}\right|^{\gamma} = C(\omega)|q - r|^{\gamma},$$

where $C(\omega) = Ak^{1-\gamma} \leq A$. □

Now we can start to construct the desired Brownian motion.

Theorem 6.10. *Define* $\mathbb{Q}_2 = \{\frac{m}{2^n} : m, n \in \mathbb{N}\}$, *and*

$$\Omega_q = \{functions\ \omega : \mathbb{Q}_2 \to \mathbb{R}\},$$

and

$$\mathcal{F}_q = \sigma(\{\omega \in \Omega_q : \omega(t_i) \in A_i, 1 \leq i \leq n\}),$$

where $A_i \in \mathcal{B}$. *Then for any* $x \in \mathbb{R}$, *there exists a unique probability measure* ν_x *on* $(\Omega_q, \mathcal{F}_q)$, *s.t.*

- $\nu_x(\{\omega : \omega(0) = x\}) = 1$;
- *for any* $0 < t_1 < \cdots < t_n$ *and* $t_i \in \mathbb{Q}_2$,

$$\nu_x(\{\omega : \omega(t_1) \in A_1, \cdots, \omega(t_n) \in A_n\}) = \mu_{x,t_1,\cdots,t_n}(A_1 \times A_2 \times \cdots \times A_n).$$

The Brownian motion in this construction is continuous by the following Lemma.

Lemma 6.11. *Let $T < \infty$ and $x \in \mathbb{R}$, define*

$$A = \{\omega \in \mathbb{Q}_q : \omega \text{ uniformly continuous on } \mathbb{Q}_2 \cap [0, T]\},$$

then $\nu_x(A) = 1$.

Proof. By scaling and translation invariance, B_t with $B_0 = x$ has the same distribution as $T^{1/2} B_{t/T}$ with $B_0 = 0$, we can assume $x = 0$ and $T = 1$. Then

$$\mathbb{E}_0(|B_t - B_s|^4) = \mathbb{E}_0(|B_{t-s} - B_0|^4) = \mathbb{E}_0(|B_{t-s}|^4) = \mathbb{E}_0(|(t-s)^{1/2} B_1|^4) = (t-s)^2 \mathbb{E}_0(|B_1|^4).$$

Apply Kolmogorov's continuity theorem and let $\alpha = 1, \beta = 4$, let $\gamma < 1/4$, then for almost sure $\omega \in \Omega_q$, there exists a constant C, s.t. for any $q, r \in \mathbb{Q}_2 \cap [0, 1]$,

$$|B(q) - B(r)| \leq C|q - r|^\gamma.$$

For any $\varepsilon > 0$, let $\delta = (\varepsilon/C)^{1/\gamma}$, then for any $q, r \in \mathbb{Q}_2 \cap [0, 1]$ with $|q - r| < \delta$,

$$|B(q) - B(r)| \leq C|q - r|^\gamma < \varepsilon,$$

i.e. such path ω is uniformly continuous. □

Therefore the Brownian paths constructed in Theorem 6.10 are continuous on $\mathbb{Q}_2$. Moreover, thanks to the uniform continuity, we can actually extend the continuity from $\mathbb{Q}_2$ to $[0, +\infty)$.

Lemma 6.12. *If $f : \mathbb{Q}_2 \to \mathbb{R}$ is uniformly continuous, then there exists a unique continuous function $g : [0, \infty) \to \mathbb{R}$ s.t. $f = g$ on $\mathbb{Q}_2$.*

Now define $C = \{\text{continous functions } \omega : [0,\infty) \to \mathbb{R}\}$,

$$\mathcal{C} = \sigma(\{\omega \in C : \omega(t_i) \in A_i, 1 \le i \le n\}),$$

where $A_i \in \mathcal{B}$. Let Ω_q' is the set of the uniformly continuous functions in Ω_q, by Lemma 6.12, there exists a unique map $\psi : \Omega_q' \to C$ s.t. for any $\omega \in \Omega_q'$, $\psi(\omega)$ is ω's unique continuous extension on $[0,\infty)$.

Lemma 6.13. *ψ defined above is invertible and measurable.*

By Lemma 6.13, we can define measure $\mathbb{P}_x$ on $(C, \mathcal{C})$ by

$$\mathbb{P}_x = \nu_x \circ \psi^{-1}.$$

Now the Brownian motion defined on $(\Omega, \mathcal{F}, \mathbb{P}) := (C, \mathcal{C}, \mathbb{P}_x)$ satisfies all properties in the definition. We have finished the construction.

Below are two important properties related to the continuity of Brownian paths.

Definition 6.14. For $\Gamma > 0$, a function $f : [0,\infty) \to \mathbb{R}$ is called (locally) γ-Hölder continuous if for every interval $[a,b]$, there is a constant $C = C(f, \gamma, [a,b]) > 0$, s.t.

$$|f(s) - f(t)| \le C|s-t|^\gamma, \quad \forall s,t \in [a,b].$$

If $\gamma = 1$, we say f is (locally) Lipschitz continuous.

Theorem 6.15 (Wiener,1923)**.** *For any $0 < \gamma < 1/2$, with probability 1, Brownian paths are γ-Hölder continuous.*

Proof. For any $m \in \mathbb{Z}_+$, let $s > t$, then

$$\mathbb{E}(|B_s - B_t|^{2m}) = \mathbb{E}[((s-t)^{1/2})^{2m}|B_1|^{2m}] = C_m|s-t|^m,$$

where $C_m = \mathbb{E}(|B_1|^{2m})$. Apply Kolmogorov's continuity theorem and take $\alpha = m-1$, $\beta = 2m$, we have with probability 1, for all $s, t \in \mathbb{Q}_2 \cap [0,1]$,

$$|B_s - B_t| \leq C|s-t|^{\gamma},$$

where

$$\gamma < \frac{\alpha}{\beta} = \frac{m-1}{2m}.$$

Let $m \to \infty$, $\gamma < \dfrac{1}{2}$. □

Theorem 6.16. *With probability* 1*, Brownian paths are nowhere Lipschitz continuous.*

Proof. 1. By translation invariance, we only need to show Brownian path is nowhere Lipschitz continuous on interval $[0,1]$.
2. Suppose $t \mapsto B_t$ is locally Lipschitz continuous at $s \in [0,1]$, then there exists $C > 0$ and $\delta > 0$ s.t. for all t with $|t-s| < \delta$, we have

$$|B(s) - B(t)| \leq C|s-t|. \tag{1}$$

Define

$$E = \{\omega : \exists s \in [0,1] \text{ s.t.} B_t \text{ is locally Lipschitz continuous at } s\},$$

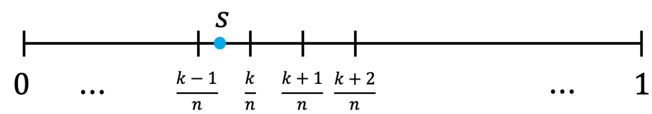
$$A_{n,C} = \{\omega : \exists s \in [0,1]\} \text{ s.t. } |B(t) - B(s)| \leq C|t-s| \text{ for all } |t-s| \leq \frac{3}{n}\}.$$

then $E \subseteq \bigcup_{C=1}^{\infty} \bigcup_{n=1}^{\infty} A_{n,C}$. For $1 \leq k \leq n-2$, let

$$Y_{k,n} = \max\{\left|B(\frac{k+j}{n}) - B(\frac{k+j-1}{n})\right| : j = 0, 1, 2\},$$

$$B_{n,C} = \{at\ least\ one\ 1 \le k \le n-2\ s.t.\ Y_{k,n} \le \frac{5C}{n}\}.$$

3. $A_{n,C} \subseteq B_{n,C}$.



Suppose a path $\omega \in A_{n,C}$. If $0 \le s \le \frac{n-2}{n}$, there exists $1 \le k \le n-2$, s.t. $s \in [\frac{k-1}{n}, \frac{k}{n}]$, then

$$\left|B(\frac{k}{n}) - B(\frac{k-1}{n})\right| \le \left|B(\frac{k}{n}) - B(s)\right| + \left|B(s) - B(\frac{k-1}{n})\right| \le C|s - \frac{k-1}{n}| + C|s - \frac{k}{n}| \le \frac{C}{n},$$

$$\left|B(\frac{k+1}{n}) - B(\frac{k}{n})\right| \le \frac{3C}{n},$$

$$\left|B(\frac{k+2}{n}) - B(\frac{k+1}{n})\right| \le \frac{5C}{n},$$

so $Y_{k,n} \le \frac{5C}{n}$, $\omega \in B_{n,C}$. If $\frac{n-2}{n} \le s \le 1$, same argument can show $\omega \in B_{n,C}$. Therefore $A_{n,C} \subseteq B_{n,C}$.

4. $\mathbb{P}(\bigcup_{n=1}^{\infty} A_{n,C}) = 0$.

Notice that

$$\begin{aligned}
\mathbb{P}(A_{n,C}) &\leq \mathbb{P}(B_{n,C}) \\
&\leq \mathbb{P}\left(\bigcup_{k=1}^{n-2}\{Y_{k,n} \leq \frac{5C}{n}\}\right) \\
&\leq n\mathbb{P}\left(Y_{k,n} \leq \frac{5C}{n}\right) \\
&\leq n\mathbb{P}\left(\left|B(\frac{k+j}{n}) - B(\frac{k+j-1}{n})\right| \leq \frac{5C}{n}, j = 0, 1, 2\right) \\
&\leq n\left[\mathbb{P}\left(\left|B(\frac{1}{n})\right| \leq \frac{5C}{n}\right)\right]^3 \\
&= n\left[\mathbb{P}\left(|B_1| \leq \frac{5C}{\sqrt{n}}\right)\right]^3 \\
&= n\left[2\int_0^{5C/\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-\frac{-x^2}{2}}\,\mathrm{d}x\right]^3 \\
&\leq n\left[\frac{10C}{\sqrt{2\pi n}}\right]^3 \to 0,
\end{aligned}$$

as $n \to \infty$. Since $A_{n,C} \subseteq A_{n+1,C}$,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_{n,C}\right) = \lim_{n\to\infty} \mathbb{P}(A_{n,C}) \leq \lim_{n\to\infty} \mathbb{P}(B_{n,C}) = 0.$$

Therefore E is contained in a null set.

$$E^c = \{\omega : B_t \text{ is nowhere Lipschitz continuous}\}$$

contains a set w.p.1., although we don't know whether E^c is measurable. □

6.3 Markov property and Blumenthal's 0-1 law

Definition 6.17. Suppose $\{B_t : t \geq 0\}$ is a Brownian motion, define

$$\mathcal{F}_s^0 = \sigma(B_t : t \leq s),$$

and

$$\mathcal{F}_s^+ = \bigcap_{t>s} \mathcal{F}_t^0.$$

Proposition 6.18. $\mathcal{F}_s^0 \subseteq \mathcal{F}_s^+$.

Proof. For any $t > s$, $\mathcal{F}_s^0 \subseteq \mathcal{F}_t^0$, thus

$$\mathcal{F}_s^0 \subseteq \bigcap_{t>s} \mathcal{F}_t^0 = \mathcal{F}_s^+.$$

□

Proposition 6.19. $\mathcal{F}_s^+$ *is right continuous, i.e.*

$$\bigcap_{t>s} \mathcal{F}_t^+ = \mathcal{F}_s^+.$$

Proof. By definition,

$$\bigcap_{t>s} \mathcal{F}_t^+ = \bigcap_{t>s} \bigcap_{u>t} \mathcal{F}_u^0 = \bigcap_{u>s} \mathcal{F}_u^0 = \mathcal{F}_s^+.$$

□

However, $\mathcal{F}_s^0$ is not right continuous.

Definition 6.20. For $x \in \mathbb{R}^d$, suppose $B_t(\omega) = \omega(t)$ is a Brownian motion on $(C, \mathcal{C}, \mathbb{P}_x)$. For $s \geq 0$, define the shift transformation $\theta_s : C \to C$ by

$$\theta_s(\omega(t)) = \omega(s + t), \quad t \in [0, \infty).$$

Theorem 6.21 (Markov property)**.** *Suppose $s \geq 0$, $Y : C \to \mathbb{R}$ is bounded and $\mathcal{C}$-measurable,* *then for any* $x \in \mathbb{R}^d$*,*

$$\mathbb{E}_x(Y \circ \theta_s | \mathcal{F}_s^+) = \mathbb{E}_{B_s}(Y)$$

Proof. This proof is very similar to the proof of Theorem 3.8.

1. By the definition of conditional expectation, we only need to show for any $A \in \mathcal{F}_s^+$,

$$\mathbb{E}[(Y \circ \theta_s)\mathbb{1}_A] = \mathbb{E}[\mathbb{E}_{B_s}(Y)\mathbb{1}_A].$$

Corollary 6.22. $\mathbb{E}_x(Y \circ \theta_s | \mathcal{F}_s^+) = \mathbb{E}_x(Y \circ \theta_s | \mathcal{F}_s^0)$.

Proof. By Theorem 6.21,

$$\mathbb{E}_x(Y \circ \theta_s | \mathcal{F}_s^+) = \mathbb{E}_{B(s)}(Y) \in \mathcal{F}_s^0 \subseteq \mathcal{F}_s^+,$$

then Proposition 1.7 implies

$$\mathbb{E}_x(Y \circ \theta_s | \mathcal{F}_s^+) = \mathbb{E}_x(Y \circ \theta_s | \mathcal{F}_s^0).$$ □

Proposition 6.23. *If* Z *is bounded and* $\mathcal{C}$*-measurable, then for any* $s \geq 0$ *and* $x \in \mathbb{R}^d$*,*

$$\mathbb{E}_x(Z | \mathcal{F}_s^+) = \mathbb{E}_x(Z | \mathcal{F}_s^0). \tag{1}$$

Proof. We only need to prove the case when

$$Z = \prod_{m=1}^{n} f_m(B(t_m)),$$

where $t_1 < t_1 < \cdots < t_n$ and f_m are bounded and measurable. Suppose $t_k \leq s$, let

$$Z_1 = \prod_{m=1}^{k} f_m(B(t_m)) \in \mathcal{F}_s^0 \subseteq \mathcal{F}_s^+,$$

and

$$Z_2 = \prod_{m=k+1}^{n} f_m(B(t_m)) = Y \circ \theta_s,$$

for some $\mathcal{C}$-measurable Y, then $Z = Z_1 Z_2 = Z_1(Y \circ \theta_s)$. Therefore

$$\mathbb{E}_x(Z|\mathcal{F}_s^+) = \mathbb{E}_x[Z_1(Y\circ\theta_s)|\mathcal{F}_s^+] = Z_1\mathbb{E}_x[Y\circ\theta_s|\mathcal{F}_s^+] = Z_1\mathbb{E}_x[Y\circ\theta_s|\mathcal{F}_s^0] = \mathbb{E}_x[Z_1(Y\circ\theta_s)|\mathcal{F}_s^0] = \mathbb{E}_x[Z|\mathcal{F}_s^0].$$

□

Corollary 6.24. *$\mathcal{F}_s^+$ and $\mathcal{F}_s^0$ are the same up to null sets.*

Proof. First $\mathcal{F}_s^0 \subseteq \mathcal{F}_s^+$. Let Z is $\mathcal{F}_s^+$-measurable, then by Proposition 6.23,

$$Z = \mathbb{E}(Z|\mathcal{F}_s^+) = \mathbb{E}(Z|\mathcal{F}_s^0) \quad a.s.,$$

so Z is $\mathcal{F}_s^0$-measurable except for some null sets. Thus $\mathcal{F}_s^+ \subseteq \mathcal{F}_s^0$ except for some null sets. □

Theorem 6.25 (Blumenthal's 0-1 law)**.** *If $A \in \mathcal{F}_0^+$, then for any $x \in \mathbb{R}^d$,*

$$\mathbb{P}_x(A) \in \{0, 1\}.$$

Proof. Since $\mathbb{1}_A \in \mathcal{F}_0^0$ and $\mathcal{F}_0^0 = \sigma(B_0) = \{\varnothing, \Omega\}$ is trivial, thus

$$\mathbb{1}_A = \mathbb{E}_x(\mathbb{1}_A|\mathcal{F}_0^0) = \mathbb{E}_x(\mathbb{1}_A) = \mathbb{P}_x(A), \quad a.s.$$

therefore almost surely $\mathbb{P}_x(A) \in \{0, 1\}$. □

Remark. We call $\mathcal{F}_0^+$ germ field, and Blumenthal's 0-1 law implies germ field is trivial.

Proposition 6.26. *If $\tau = \inf\{t \geq 0 : B_t > 0\}$, then $\mathbb{P}_0(\tau = 0) = 1$.*

Proof. 1. By $\{B_t > 0\} \subseteq \{\tau \le t\}$ and $B_t \sim \mathcal{N}(0,t)$,

$$\mathbb{P}_0(\tau \le t) \ge \mathbb{P}_0(B_t > 0) = \frac{1}{2}.$$

2. Since $\{\tau < \frac{1}{n}\} \downarrow \{\tau = 0\}$, by the continuity of measure, we have

$$\mathbb{P}_0(\tau = 0) = \mathbb{P}_0\left(\bigcap_{n=0}^{\infty}\{\tau \le \frac{1}{n}\}\right) = \lim_{n\to\infty} \mathbb{P}_0(\tau \le \frac{1}{n}) \ge \frac{1}{2}.$$

3. $\{\tau \le t\} \subseteq \{B_t > 0\} \in \mathcal{F}_t^0$ implies

$$\{\tau = 0\} = \bigcap_{t>0}\{\tau \le t\} \in \mathcal{F}_0^+,$$

thus by Blumenthal's 0-1 law (Theorem 6.25), $\mathbb{P}_0(\tau = 0) = 1$. □

Remark. This result says Brownian path starting from 0 must immediately hit $(0, +\infty)$, also immediately hit $(-\infty, 0)$ by symmetry.

Proposition 6.27. *Suppose $\{B_s : s \ge 0\}$ starts from 0. Let $T_0 = \inf\{t > 0 : B_t = 0\}$, $\mathcal{Z} = \{t \ge 0 : B_t = 0\}$. Then with probability 1,*

1. *Brownian path changes its sign infinitely many times in any interval $[0, \varepsilon]$ ($\varepsilon > 0$).*

2. *$T_0 = 0$.*

3. *0 is an accumulation point of $\mathcal{Z}$.*

Proof. 1. Let $\tau' = \inf\{t \ge 0 : B_t < 0\}$. By Proposition 6.26, for each path $\omega \in \{\tau = 0\} \cap \{\tau' = 0\}$ (w.p.1.), we have $\inf\{t \ge 0 : B_t > 0\} = 0$, i.e. for any $\varepsilon > 0$, $B_{t_0} > 0$ for some $t_0 \in (0, \varepsilon)$. Thus there is a sequence $t_n \downarrow 0$ with $t_n \in (0, t_{n-1})$ (so all different), s.t. $B_{t_n} > 0$ for all $n \in \mathbb{N}$. Similarly, there is a sequence $s_n \downarrow 0$, s.t. $B_{s_n} < 0$ for all $n \in \mathbb{N}$. Therefore the path ω changes sign infinitely many times.

2 and 3. For each path $\omega \in \{\tau = 0\} \cap \{\tau' = 0\} \cap \{\text{continuous paths}\}$ (w.p.1.), by continuity and $B_{s_n} < 0$, $B_{t_n} > 0$, we can find u_n between s_n and t_n s.t. $B_{u_n} = 0$. Moreover, the sequence $u_n \downarrow 0$, which implies $T_0 = 0$ and 0 is an accumulation point of $\mathcal{Z}$. □

Lemma 6.28 (Law of large number for Brownian motion)**.** *Suppose* $\{B_t : t \geq 0\}$ *starts from* 0*, then*

$$\lim_{t\to\infty} \frac{B_t}{t} = 0, \quad a.s.$$

Proof. For integer case, Since $B_{n+1} - B_n \sim \mathcal{N}(0, 1)$, by the strong law of large number,

$$\frac{B_n}{n} = \frac{\sum_0^{n-1}(B_{n+1} - B_n)}{n} \to 0, \quad a.s.$$

For real values between integers, we will use Kolmogorov's inequality (Theorem 2.22). For $m \in \mathbb{Z}_+$, let

$$X_i = B(n + \frac{i}{2^m}) - B(n + \frac{i-1}{2^m}),$$

then $X_i \sim_{i.i.d.} \mathcal{N}(0, \frac{1}{2^m})$. Let

$$S_k = \sum_{i=1}^{k} X_i = B(n + \frac{k}{2^m}) - B(n),$$

we have $\text{Var}(S_k) = \sum_{i=1}^{k} \text{Var}(X_i) = \frac{k}{2^m}$. By Kolmogorov's inequality,

$$\mathbb{P}\left(\sup_{1\leq k\leq 2^m} |B(n + \frac{k}{2^m}) - B(n)| > n^{2/3}\right) = \mathbb{P}\left(\sup_{1\leq k\leq 2^m} |S_k| > n^{2/3}\right) \leq \frac{\text{Var}(S_{2^m})}{n^{4/3}} = \frac{2^m}{2^m n^{4/3}} = \frac{1}{n^{4/3}},$$

let $m \to \infty$, we have

$$\mathbb{P}\left(\sup_{t\in[n,n+1]} |B(u) - B(n)| > n^{2/3}\right) \leq \frac{1}{n^{4/3}}.$$

Since $\sum_{n=1}^{\infty} \frac{1}{n^{4/3}} < \infty$, by Borel-Catelli lemma,

$$\mathbb{P}\left(\sup_{t\in[n,n+1]} |B(u) - B(n)| > n^{2/3},\ i.o. \right) = 0,$$

which means for almost sure ω,

$$\sup_{t\in[n,n+1]} |B(t) - B(n)| > n^{2/3}$$

holds for only finitely many n, i.e. for all large enough n,

$$\sup_{t\in[n,n+1]} |B(t) - B(n)| \le n^{2/3}.$$

Therefore for any large enough t, let $[t]$ be the integer part of t,

$$\begin{aligned}\left|\frac{B_t}{t}\right| &\le \frac{|B_t|}{[t]} = \frac{1}{[t]}|B_t - B_{[t]} + B_{[t]}| \\ &\le \frac{1}{[t]}|B_t - B_{[t]}| + \frac{|B_{[t]}|}{[t]} \\ &\le \frac{[t]^{2/3}}{[t]} + \frac{|B_{[t]}|}{[t]} \to 0\end{aligned}$$

□

Proposition 6.29. *Suppose B_t is a Brownian motion with $B_0 = 0$. Define*

$$X_t = \begin{cases} 0 & t = 0 \\ tB(\frac{1}{t}) & t > 0 \end{cases}$$

then $\{X_t : t \ge 0\}$ is also a Brownian motion starting from 0.

Proof. Check the definition of Brownian motion.

(i) For $0 < t_1 < t_2 < \cdots < t_n$,

$$(X(t_1), \cdots, X(t_n)) = (t_1 B(\frac{1}{t_1}), \cdots t_n B(\frac{1}{t_n}))$$

is multivariant Gaussian.

(ii) For any $t > 0$, $\mathbb{E}(X_t) = \mathbb{E}(tB_{1/t}) = 0$.

(iii) For any $0 < t < s$,

$$\mathbb{E}(X_t X_s) = \mathbb{E}(tsB_{1/s}B_{1/t}) = ts \cdot \frac{1}{s} = t.$$

(iv) For $t > 0$, since B_t and $1/t$ are continuous, their composition $B(1/t)$ is also continuous, thus X_t is continuous on $(0, \infty)$. For $t = 0$, by Lemma 6.28,

$$\lim_{t \to 0^+} X(t) = \lim_{t \to 0^+} tB(\frac{1}{t}) = \lim_{s \to +\infty} \frac{B(s)}{s} = 0 = X(0),$$

thus $X(t)$ is also continuous at 0.

□

Theorem 6.30 (Kolmogorov's 0-1 law)**.** *If $A \in \mathcal{T} = \bigcap_{t \geq 0} \sigma(B_s : s \geq t)$, then $\mathbb{P}_x(A) \in \{0, 1\}$.*

Proposition 6.31. *Suppose B_t starting from 0 is a Brownian motion in $\mathbb{R}$, then almost surely,*

$$\limsup_{t \to \infty} \frac{B_t}{\sqrt{t}} = \infty, \quad \liminf_{t \to \infty} \frac{B_t}{\sqrt{t}} = -\infty.$$

Proof. Notice

$$\limsup_{t \to \infty} \frac{B_t}{\sqrt{t}} \geq \limsup_{n \to \infty} \frac{B_n}{\sqrt{n}},$$

so we only need to show the integer case. Let $K < \infty$, then by scaling invariance

$$\mathbb{P}_0(\frac{B_n}{\sqrt{n}} \geq K \ i.o.) \geq \limsup_{n \to \infty} \mathbb{P}_0(B_n \geq K\sqrt{n}) = \mathbb{P}_0(B_1 \geq K) > 0.$$

And

$$\{\frac{B_n}{\sqrt{n}} \geq K\ i.o.\} = \bigcap_{m\geq 1}\bigcup_{n\geq m}\{\frac{B_n}{\sqrt{n}} \geq K\} \in \mathcal{T},$$

thus by Kolmogorov's 0-1 law (Theorem 6.30),

$$\mathbb{P}_0(\frac{B_n}{\sqrt{n}} \geq K\ i.o.) = 1.$$

Since $A_K = \{\frac{B_n}{\sqrt{n}} \geq K\ i.o.\} \downarrow \{\frac{B_n}{\sqrt{n}} = \infty\ i.o.\} = \{\limsup_{n\to\infty}\frac{B_n}{\sqrt{n}} = \infty\}$, by the continuity of probability, we have

$$\mathbb{P}_0(\limsup_{n\to\infty}\frac{B_n}{\sqrt{n}} = \infty) = \lim_{K\to\infty}\mathbb{P}_0(A_K) = 1.$$

The $\liminf$ case is also true by symmetry. □

Proposition 6.32 (one-dimensional Brownian motion is recurrent)**.** *Suppose B_t is a Brownian motion in $\mathbb{R}$, let*

$$A = \bigcap_n\{\textit{there exists some } t \geq n \textit{ s.t. } B_t = 0\},$$

then $\mathbb{P}_x(A) = 1$ for any $x \in \mathbb{R}$.

Proof. For any continuous Brownian path B_t (w.p.1.), by Proposition 6.31 and translation invariance ($B_m - B_0$ is a Brownian motion starting from 0), there are infinitely many $m, n \in \mathbb{Z}_+$ s.t. $\frac{B_m}{\sqrt{m}} = -\infty$ and $\frac{B_n}{\sqrt{n}} = \infty$, so $B_m < 0$ and $B_n > 0$ i.o. By continuity, $B_k = 0$ for i.o. $k \in \mathbb{Z}_+$. (Take $N_1 > 0$, we have $B_{n_1} > 0$ and $B_{m_1} < 0$ for some $m_1, n_1 > N_1$, then there must be some k_1 between m_1 and n_1 s.t. $B_{k_1} = 0$. Take $N_2 = k_1$, repeat this step, we can construct a sequence $k_i \uparrow \infty$ s.t. $B_{k_i} = 0$). Therefore

$$A = \bigcap_n\bigcup_{m\geq n}\{B_m = 0\} = \{B_n = 0\ i.o.\}$$

has probability 1. □

Based on the above discussions, we can improve our filtration by adding all null sets.

Definition 6.33 (Filtration of Brownian motion)**.** Let

$$\begin{aligned}\mathcal{N}_x &= \{A \in \mathcal{F} : A \subseteq D, \mathbb{P}_x(D) = 0\} \\ \mathcal{F}_s^x &= \sigma(\mathcal{F}_s^+ \cup \mathcal{N}_x) \\ \mathcal{F}_s &= \bigcap_x \mathcal{F}_s^x.\end{aligned}$$

$\mathcal{F}_s$ is called the filtration of Brownian motion.

Remark. $\mathcal{F}_s$ does not depend on the initial state and is right-continuous.

At the end, we introduce two alternative forms of Markov property.

Theorem 6.34. *For $t \geq 0$, suppose Y is a bounded and $\sigma(B_s, s \geq t)$-measurable, then*

$$\mathbb{E}_x(Y|\mathcal{F}_t) = \mathbb{E}_x(Y|B_t).$$

Proof. $Y \circ \theta_{-t}$ is bounded and $\mathcal{C}$-measurable, applying Markov property (Theorem 6.21), we have

$$\mathbb{E}_x(Y|\mathcal{F}_t) = \mathbb{E}_x[(Y \circ \theta_{-t}) \circ \theta_t|\mathcal{F}_t] = \mathbb{E}_{B_t}(Y \circ \theta_{-t}) \in \sigma(B_t),$$

Taking conditional expectation on B_t, we have

$$\mathbb{E}_x[\mathbb{E}_x(Y|\mathcal{F}_t)|B_t] = \mathbb{E}_x[\mathbb{E}_{B_t}(Y \circ \theta_{-t})|B_t],$$

the left side above is $\mathbb{E}_x(Y|B_t)$ since $\sigma(B_t) \subseteq \mathcal{F}_t$, the right side is $\mathbb{E}_{B_t}(Y \circ \theta_{-t}) = \mathbb{E}_x(Y|\mathcal{F}_t)$, therefore

$$\mathbb{E}_x(Y|\mathcal{F}_t) = \mathbb{E}_x(Y|B_t).$$ □

Theorem 6.35. *Suppose $\{B_t : t \geq t\}$ is a Brownian motion with $B_0 = x$, for any $s \geq 0$, $\{B_{t+s} - B_s : t \geq 0\}$ is a Brownian motion starting from 0 and independent of $\mathcal{F}_s$.*

Proof. For any bounded and measurable function f, g, let $Y \in \mathcal{F}_s$, then for any $t \geq 0$,

$$\begin{aligned}\mathbb{E}_x[f(B_{t+s} - B_s)g(Y)|\mathcal{F}_s] &= g(Y)\mathbb{E}_x[f(B_{t+s} - B_s)|\mathcal{F}_s] \\ &= g(Y)\mathbb{E}_x[f(B_{t+s} - B_s)|B_s] \\ &= g(Y)\mathbb{E}_x[f(B_{t+s} - B_s)],\end{aligned}$$

take expectation on both sides, we have

$$\mathbb{E}_x[f(B_{t+s} - B_s)g(Y)] = \mathbb{E}_x[g(Y)]\mathbb{E}_x[f(B_{t+s} - B_s)],$$

therefore $B_{t+s} - B_s$ and $\mathcal{F}_s$ are independent, hence $\sigma(B_{t+s} - B_s : t \geq 0)$ and $\mathcal{F}_s$ are independent. □

6.4 Continuous stopping time

Definition 6.36. We call r.v. S a stopping time if for all $t \geq 0$, $\{S < t\} \in \mathcal{F}_t$.

Lemma 6.37. *S is a stopping time if and only if for all $t \geq 0$, $\{S \leq t\} \in \mathcal{F}_t$.*

Proof. Suppose $\{S \leq t\} \in \mathcal{F}_t$, then since $\mathcal{F}_t$ is right continuous,

$$\{S \leq t\} = \bigcap_{n=1}^{\infty}\{S < t + \frac{1}{n}\} \in \mathcal{F}_t$$

Proposition 6.38. *Let S, T be stopping times. Then*

- $S \wedge T$
- $S \vee T$
- $S + T$

are all stopping times.

Proposition 6.39. *Suppose $\{T_n : n \geq 1\}$ is a sequence of stopping times. We have*

1. *If $T_n \uparrow T$, then T is a stopping time.*

2. *If $T_n \downarrow T$, then T is a stopping time.*

3. $\sup_n T_n$ *and* $\inf_n T_n$ *are stopping times*

4. $\limsup_n T_n$ *and* $\liminf_n T_n$ *are stopping times*

Proposition 6.40. *Let $A \subseteq \mathbb{R}$ be a set. Define $T_A = \inf\{t \geq 0 : B_t \in A\}$. Then*

1. *If A is an open set, T_A is a stopping time*

2. *If A is a closed set, T_A is a stopping time*

3. *If A is a countable union of closed sets, T_A is a stopping time.*

Proposition 6.41. *If $S \leq T$ are both stopping times, then $\mathcal{F}_S \subseteq \mathcal{F}_T$.*

Proposition 6.42. *If $T_n \downarrow T$ are stopping times, then*

$$\mathcal{F}_T = \bigcap_{n=1}^{\infty} \mathcal{F}_{T_n}.$$

Proposition 6.43. *If S is a stopping time, then $B_S \in \mathcal{F}_S$.*

6.5 Strong Markov property

Theorem 6.44. *Let $(s, \omega) \mapsto Y_s(\omega)$ be bounded and $\mathcal{B}(\mathbb{R}) \times \mathcal{C}$ measurable. If S is a stopping time, then for any $x \in \mathbb{R}$, on $\{S < \infty\}$,*

$$\mathbb{E}_x(Y_S \circ \theta_S | \mathcal{F}_S) = \mathbb{E}_{B_S} Y_S.$$

6.6 Path properties

6.6.1 Zero set

Definition 6.45. Suppose ω is a path of Brownian motion, define the zero set as $\mathcal{Z}_\omega = \{t \geq 0 : B_t = 0\}$.

Proposition 6.46. *For a.s. path $\omega \in \Omega$, $\mathcal{Z}_\omega$*

1. *has Lebesgue measure* 0*,*
2. *is closed and unbounded,*
3. *has no isolated point,*
4. *is dense in itself (perfect set),*
5. *is uncountable,*
6. *has Hausdorff dimension* $\frac{1}{2}$.

Proof. 1.For any $t > 0$, $B_t \sim \mathcal{N}(x, t)$ under $\mathbb{P}_x$, then

$$\mathbb{E}_x(\mathbb{1}_{\{t\in\mathcal{Z}_\omega\}}) = \mathbb{P}_x(t \in \mathcal{Z}_\omega) = \mathbb{P}_x(B_t = 0) = 0,$$

therefore by Fubini's theorem,

$$\mathbb{E}_x[m(\mathcal{Z}_\omega)] = \mathbb{E}_x[\int_0^\infty \mathbb{1}_{\{t\in\mathcal{Z}_\omega\}}\,\mathrm{d}t] = \int_0^\infty \mathbb{E}_x[\mathbb{1}_{\{t\in\mathcal{Z}_\omega\}}]\,\mathrm{d}t = 0.$$

2. To prove a set is closed, we only need to show it contains all its limits. Let ω be a continuous path (w.p.1.). For any sequence $t_n \in \mathcal{Z}_\omega$, if $t_n \to t$, then by the continuity,

$$B(t) = \lim_{n\to\infty} B(t_n) = 0,$$

thus $t \in \mathcal{Z}_\omega$. $\mathcal{Z}$ unbounded is proved in Proposition 6.32.

3. Let $T_0 = \inf\{t > 0 : B_t = 0\}$. By Proposition 6.27, $\mathbb{P}_0(T_0 = 0) = 1$. For any $t > 0$, let $R_t = \inf\{u > t : B_u = 0\}$, by Propostion 6.32, there exists $n \geq t$ (w.p.1.) s.t. $B_n = 0$, thus $R_t \leq n < \infty$ a.s. By the definition of inf, there is a sequence t_n in $\mathcal{Z} \cap (t, \infty)$ s.t. $t_n \to R_t$, then by continuity, $R_t \in \mathcal{Z}$. Now applying the strong Markov property, we have

$$\mathbb{E}_x[\mathbb{1}_{\{T_0=0\}} \circ \theta_{R_t} | \mathcal{F}_{R_t}] = \mathbb{E}_{B(R_t)}(\mathbb{1}_{\{T_0=0\}}) = \mathbb{P}_0(T_0 = 0) = 1,$$

take expectation, we have for any $t > 0$,

$$\mathbb{P}_x(T_0 \circ \theta_{R_t} = 0) = 1.$$

Let $A_t = \{\omega : T_0 \circ \theta_{R_t} > 0\}$, then A_t is null, thus the union over all rational numbers

$$A := \bigcup_{t \in \mathbb{Q}} A_t$$

is also null, which implies on $\Omega \setminus A$ (w.p.1.), $T_0 \circ \theta_{R_t} = 0$ for all rational t. For path $\omega \in \Omega \setminus A$, take $u \in \mathcal{Z}_\omega$: if $u = R_t$ for some rational t, u is obviously not isolated from the right; if $u \neq R_t$ for any rational t, there is a rational sequence t_n s.t. $t_n \uparrow u$. Since $t_n \leq R_{t_n} < u$, we obtain a sequence R_{t_n} in $\mathcal{Z}_\omega$ s.t. $R_{t_n} \to u$. Therefore $\mathcal{Z}_\omega$ is not isolated w.p.1.

4. Closed set without isolated points is dense in itself.

5. Perfect set is uncountable (See [6]).

6. □

6.6.2 Hitting time and maximum

Definition 6.47. We say $\{X_t : t \geq 0\}$ has stationary increments if for any $t, h \geq 0$, the distribution of $X_{t+h} - X_t$ only depends on h not t.

Proposition 6.48. *Let* $T_a = \inf\{t > 0 : B_t = a\}$*, then under* $\mathbb{P}_0$*,* $\{T_a, a \geq 0\}$ *has stationary independent increments.*

Proof. 1. (stationary increments). If $0 < a < b$, then

$$T_b \circ \theta_{T_a} = T_b - T_a.$$

Then for any bounded and measurable f, by strong Markov property and translation invariance, we have

$$\begin{aligned} \mathbb{E}_0[f(T_b - T_a)|\mathcal{F}_{T_a}] &= \mathbb{E}[f(T_b \circ \theta_{T_a})|\mathcal{F}_{T_a}] \\ &= \mathbb{E}[f(T_b) \circ \theta_{T_a}|\mathcal{F}_{T_a}] \\ &= \mathbb{E}_{B(T_a)}[f(T_b)] \\ &= \mathbb{E}_a[f(T_b)] \\ &= \mathbb{E}_0[f(T_{b-a})], \end{aligned}$$

thus

$$\mathbb{E}_0[f(T_b - T_a)] = \mathbb{E}_0[f(T_{b-a})],$$

which implies $T_b - T_a$ has the same distribution as T_{b-a}.

2. (independent increments) Let $a_0 < a_1 < \cdots < a_n$, for any bounded and measurable

functions $f_1, \cdots, f_n$,

$$\begin{aligned}\mathbb{E}_0\left[\prod_{i=1}^n f_i(T_{a_i}-T_{a_{i-1}})\right] &= \mathbb{E}_0\left[\mathbb{E}_0\left[\prod_{i=1}^n f_i(T_{a_i}-T_{a_{i-1}})\,\Big|\mathcal{F}_{T_{a_{n-1}}}\right]\right]\\ &= \mathbb{E}_0\left[\prod_{i=1}^{n-1} f_i(T_{a_i}-T_{a_{i-1}})\mathbb{E}_0\left[f_n(T_{a_n}-T_{a_{n-1}})\,\Big|\mathcal{F}_{T_{a_{n-1}}}\right]\right]\\ &= \mathbb{E}_0\left[\prod_{i=1}^{n-1} f_i(T_{a_i}-T_{a_{i-1}})\mathbb{E}_0\left[f_n(T_{a_n}-T_{a_{n-1}})\right]\right]\\ &= \mathbb{E}_0\left[\prod_{i=1}^{n-1} f_i(T_{a_i}-T_{a_{i-1}})\right]\mathbb{E}_0\left[f_n(T_{a_n}-T_{a_{n-1}})\right],\end{aligned}$$

by induction, we have

$$\mathbb{E}_0\left[\prod_{i=1}^n f_i(T_{a_i}-T_{a_{i-1}})\right] = \prod_{i=1}^n \mathbb{E}_0\left[f_i(T_{a_i}-T_{a_{i-1}})\right],$$

thus $T_{a_i}-T_{i-1}$, $1\le i\le n$ are independent. □

Theorem 6.49 (Reflection principle)**.** *Let $a>0$, then*

$$\mathbb{P}_0(T_a\le t) = 2\mathbb{P}_0(B_t\ge a).$$

Proof. We can just modify the proof of Theorem 3.12. Fix $t\ge 0$. Let $S=\inf\{s\le t: B_s=a\}$, define $\inf\varnothing=\infty$. Notice that

$$\{S\le t\}=\{S<\infty\}=\{T_a\le t\}.$$

For $s\le t$, define

$$Y_s = \mathbb{1}_{\{B_{t-s}\ge a\}},$$

then $Y_s \circ \theta_s = \mathbb{1}_{\{B_t \geq a\}}$. On $\{S < \infty\} = \{S \leq t\}$,

$$Y_S \circ \theta_S(\omega) = \mathbb{1}_{\{B_t \geq a\}}, \tag{1}$$

and by the strong Markov property,

$$\mathbb{E}_0(Y_S \circ \theta_S | \mathcal{F}_S) = \mathbb{E}_{B_S}(Y_S). \tag{2}$$

For $s \leq t$,

$$\mathbb{E}_a(Y_s) = \mathbb{P}_a(B_{t-s} \geq a) = \frac{1}{2},$$

thus on $\{S \leq t\}$, $B_S = a$,

$$\mathbb{E}_{B_S}(Y_S) = \frac{1}{2}.$$

Since $\{S \leq t\} \in \mathcal{F}_S$, applying the definition of conditional expectation to (2), we have

$$\mathbb{E}_0(Y_S \circ \theta_S \mathbb{1}_{\{S \leq t\}}) = \mathbb{E}_0[\mathbb{E}_{B_S}(Y_S)\mathbb{1}_{\{S \leq t\}}] = \mathbb{E}_0[\frac{1}{2}\mathbb{1}_{\{S \leq t\}}] = \frac{1}{2}\mathbb{P}_0(S \leq t),$$

and by (1),

$$\mathbb{E}_0(Y_S \circ \theta_S \mathbb{1}_{\{S \leq t\}}) = \mathbb{E}_0(\mathbb{1}_{\{B_t \geq a\} \cap \{S \leq t\}}) = \mathbb{P}_0(\{B_t \geq a\} \cap \{S \leq t\}) = \mathbb{P}_0(B_t \geq a),$$

since $\{B_t \geq a\} \subseteq \{S \leq t\}$. □

Theorem 6.50 (Generalized reflection principle)**.** *Let $a > 0$, $x \leq a$, then*

$$\mathbb{P}_0(T_a \leq t, B_t \leq x) = \mathbb{P}_0(B_t \geq 2a - x).$$

Proof. Let $S = \inf\{s \leq t : B_s = a\}$. Define $\inf \varnothing = \infty$. Let $Y_s = \mathbb{1}_{\{s \leq t\}}\mathbb{1}_{\{B_{t-s} \leq x\}}$, $Z_s =$

$\mathbb{1}_{\{s\le t\}}\mathbb{1}_{\{B_{t-s}\ge 2a-x\}}$. By symmetry, we have

$$\mathbb{E}_a(Y_s)=\mathbb{E}_a(Z_s).$$

And

$$Y_S\circ\theta_S=\mathbb{1}_{\{S\le t\}}\mathbb{1}_{\{B_t\le x\}},\quad Z_S\circ\theta_S=\mathbb{1}_{\{S\le t\}}\mathbb{1}_{\{B_t\ge 2a-x\}}.$$

By the strong Markov property, on $\{S\le t\}$,

$$\mathbb{E}_0(Y_S\circ\theta_S|\mathcal{F}_S)=\mathbb{E}_{B_S}Y_S=\mathbb{E}_{B_S}Z_S=\mathbb{E}_0(Z_S\circ\theta_S|\mathcal{F}_S),$$

thus

$$\mathbb{E}_0(Y_S\circ\theta_S)=\mathbb{E}_0(Z_S\circ\theta_S),$$

which is

$$\mathbb{P}_0(S\le t,B_t\le x)=\mathbb{P}_0(S\le t,B_t\ge 2a-x).\tag{1}$$

Since $\{S\le t\}=\{T_a\le t\}$ and $\{B_t\ge 2a-x)\}\subseteq\{S\le t\}$, (1) becomes

$$\mathbb{P}_0(T_a\le t,B_t\le x)=\mathbb{P}_0(B_t\ge 2a-x).$$ □

Proposition 6.51 (Density of T_a)**.** *Let $a>0$, then*

$$\mathbb{P}_0(T_a\in \mathrm{d}t)=\frac{a}{\sqrt{2\pi t^3}}e^{-\frac{a^2}{2t}}\mathbb{1}_{\{t\ge 0\}}\,\mathrm{d}t.$$

Proof. By Theorem 6.49,

$$\begin{aligned}
\mathbb{P}_0(T_a \le t) &= 2\mathbb{P}_0(B_t \ge a) \\
&= \frac{2}{\sqrt{2\pi t}} \int_a^\infty e^{-x^2/2t} \, \mathrm{d}x \\
&= \frac{2}{\sqrt{2\pi t}} \int_{1/a^2}^0 e^{-1/2ut} \cdot \left(-\frac{u^{-3/2}}{2}\right) \mathrm{d}u \quad (\text{let } x = u^{-1/2}) \\
&= \frac{1}{\sqrt{2\pi t}} \int_0^{1/a^2} e^{-1/2ut} u^{-3/2} \, \mathrm{d}u \\
&= (2\pi t)^{-1/2} \int_0^t e^{-a^2/2s} \left(\frac{s}{ta^2}\right)^{-3/2} \frac{1}{ta^2} \, \mathrm{d}s \quad (\text{let } u = \frac{s}{ta^2}) \\
&= \int_0^t e^{-a^2/2s} \left(2\pi t \cdot \frac{s^3}{t^3}\right)^{-1/2} \cdot \frac{a}{t} \, \mathrm{d}s \\
&= \int_0^t \frac{a}{\sqrt{2\pi s^3}} \exp\left(-\frac{a^2}{2s}\right) \mathrm{d}s.
\end{aligned}$$

□

Remark. We have $\mathbb{E}_0(T_a) = \infty$.

Corollary 6.52 (Density of $T_b - T_a$)**.** *Let $0 \le a < b < \infty$, we have*

$$\mathbb{P}_0(T_b - T_a \in \mathrm{d}t) = \frac{b-a}{\sqrt{2\pi t^3}} e^{-\frac{(b-a)^2}{2t}} \mathbb{1}_{\{t \ge 0\}} \, \mathrm{d}t.$$

Proof. From Proposition 6.48, $T_b - T_a$ has the same distribution as T_{b-a}. □

Definition 6.53. Define the maximum process of Brownian motion as $M_t = \max_{0 \le s \le t} B_s$.

Remark. M_t has some simple properties:

1. From Proposition 6.26, $M_t > 0$ for any $t > 0$.

2. $t \mapsto M_t$ is increasing.

3. $\{M_t \ge a\} = \{T_a \le t\}$.

Corollary 6.54 (Density of M_t)**.** *For any $t > 0$,*

$$\mathbb{P}_0(M_t \in \mathrm{d}a) = \frac{2}{\sqrt{2\pi t}} e^{-\frac{a^2}{2t}} \mathbb{1}_{\{a \geq 0\}} \,\mathrm{d}a.$$

Proof.

$$\mathbb{P}_0(M_t \leq a) = 1 - \mathbb{P}_0(M_t \geq a) = 1 - \mathbb{P}_0(T_a \leq t) = \frac{2}{\sqrt{2\pi t}} \int_0^a e^{-x^2/2t} \,\mathrm{d}x.$$

□

Proposition 6.55. *For any $t > 0$,*

$$\mathbb{E}_0(M_t) = \sqrt{\frac{2t}{\pi}}.$$

Proposition 6.56 (joint distribution of M_t and B_t)**.**

$$f_{(M_t, B_t)}(a, x) = \frac{2(2a - x)}{\sqrt{2\pi t^3}} e^{-\frac{(2a-x)^2}{2t}} \mathbb{1}_{\{a \geq 0\}} \mathbb{1}_{\{a \geq x\}}.$$

Proof. From Theorem 6.50. □

Proposition 6.57. *For a fixed $t \geq 0$, M_t, $M_t - B_t$, and $|B_t|$ have the same distribution.*

Proof. 1. From Theorem 6.49,

$$\mathbb{P}_0(M_t \geq a) = \mathbb{P}_0(T_a \leq t) = 2\mathbb{P}_0(B_t \geq a) = \mathbb{P}_0(|B_t| \geq a),$$

so M_t and $|B_t|$ has the same distribution.

2. Let $U = M_t - B_t$, $V = B_t$, i.e. $M_t = U + V$, $B_t = V$. We will compute the joint distribution of (U, V) from Proposition 6.56. The Jocobian is

$$J(a, x) = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix},$$

$|J(a,x)| = 1$. Therefore the joint density of (U,V) is

$$f_{(U,V)}(u,v) = \frac{f_{(M_t,B_t)}(a,x)}{|J(a,x)|} = \frac{2(2u+v)}{\sqrt{2\pi t^3}} e^{-\frac{(2u+v)^2}{2t}} \mathbb{1}_{\{u+v\geq 0\}} \mathbb{1}_{\{u\geq 0\}},$$

then the density of U is

$$\begin{aligned} f_U(u) &= \int_{-u}^{\infty} \frac{2(2u+v)}{\sqrt{2\pi t^3}} e^{-\frac{(2u+v)^2}{2t}} \mathbb{1}_{\{u\geq 0\}} \, \mathrm{d}v \\ &= \mathbb{1}_{\{u\geq 0\}} \int_{u^2/2t}^{\infty} \frac{2(2u+v)}{\sqrt{2\pi t^3}} e^{-z} \frac{t}{2u+v} \, \mathrm{d}z \quad (\text{let } z = (2u+v)^2/2t) \\ &= \mathbb{1}_{\{u\geq 0\}} \int_{u^2/2t}^{\infty} \frac{2}{\sqrt{2\pi t}} e^{-z} \, \mathrm{d}z \\ &= \frac{2}{\sqrt{2\pi t}} e^{-\frac{u^2}{2t}} \mathbb{1}_{\{u\geq 0\}}, \end{aligned}$$

which means $U = M_t - B_t$ has the same distribution as M_t (Corollary 6.54). □

6.6.3 Arcsine laws

There are three arcsine laws in Brownian motion. Based on previous results, we are already able to prove two of them!

Lemma 6.58. *Let* $T_0 = \inf\{t > 0 : B_t = 0\}$ *and* $L = \sup\{t \leq 1 : B_t = 0\}$. *Then*

$$\mathbb{P}_0(L \leq t) = \int_{-\infty}^{\infty} p_t(0,y) \mathbb{P}_y(T_0 > 1-t) \, \mathrm{d}y.$$

Proof. Let $R_t = \{u > t : B_u = 0\}$, then $\{L \leq t\} = \{R_t > 1\}$. Notice that $T_0 \circ \theta_t + t = R_t$, thus

$$\mathbb{1}_{\{T_0 > 1-t\}} \circ \theta_t = \mathbb{1}_{\{T_0 \circ \theta_t > 1-t\}} = \mathbb{1}_{\{R_t - t > 1-t\}} = \mathbb{1}_{\{L \leq t\}}.$$

By Markov property, we have

$$\mathbb{E}_0[\mathbb{1}_{\{L\leq t\}} | \mathcal{F}_t] = \mathbb{E}_0[\mathbb{1}_{\{T_0 > 1-t\}} \circ \theta_t | \mathcal{F}_t] = \mathbb{E}_{B_t}(\mathbb{1}_{\{T_0 > 1-t\}}) = \mathbb{P}_{B_t}(T_0 > 1-t),$$

take expectation on each side, we have

$$\begin{aligned}\mathbb{P}_0(L \le t) &= \mathbb{E}_0[\mathbb{P}_{B_t}(T_0 > 1-t)] \\ &= \int \mathbb{P}_y(T_0 > 1-t)\mathbb{P}_0(B_t \in \mathrm{d}y) \\ &= \int \mathbb{P}_y(T_0 > 1-t)p_t(0,y)\,\mathrm{d}y.\end{aligned}$$

□

Theorem 6.59 (Arcsine law)**.** *Let* $L = \sup\{t \in [0,1] : B_t = 0\}$, *then*

$$\mathbb{P}_0(L \le t) = \frac{2}{\pi}\arcsin(\sqrt{t}).$$

Proof. By Lemma 6.58, we have

$$
\begin{aligned}
\mathbb{P}_0(L \le t) &= \int_{-\infty}^{\infty} p_t(0,y)\mathbb{P}_y(T_0 > 1-t)\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} p_t(0,y)\mathbb{P}_0(T_y > 1-t)\,\mathrm{d}y \\
&= 2\int_0^{\infty} p_t(0,y)[1-\mathbb{P}_0(T_y \le 1-t)]\,\mathrm{d}y \\
&= 2\int_0^{\infty} \frac{e^{-y^2/2t}}{\sqrt{2\pi t}} \cdot \int_{1-t}^{\infty} \frac{ye^{-y^2/2s}}{\sqrt{2\pi s^3}}\,\mathrm{d}s\,\mathrm{d}y \qquad \text{(by Proposition 6.51)} \\
&= \frac{1}{\pi}\int_0^{\infty}\int_{1-t}^{\infty} (ts^3)^{-1/2} y \exp\left(-\frac{(t+s)y^2}{2ts}\right)\mathrm{d}s\,\mathrm{d}y \\
&= \frac{1}{\pi}\int_{1-t}^{\infty} (ts^3)^{-1/2}\,\mathrm{d}s \int_0^{\infty} \exp\left(-\frac{(t+s)y^2}{2ts}\right) y\,\mathrm{d}y \qquad \text{(Fubini's theorem)} \\
&= \frac{1}{2\pi}\int_{1-t}^{\infty} (ts^3)^{-1/2}\,\mathrm{d}s \int_0^{\infty} \exp\left(-\frac{(t+s)u}{2ts}\right)\mathrm{d}u \qquad (\text{let } u = y^2) \\
&= \frac{1}{\pi}\int_{1-t}^{\infty} (ts^3)^{-1/2}\frac{ts}{t+s}\,\mathrm{d}s \\
&= \frac{1}{\pi}\int_{1-t}^{\infty} \frac{t^{1/2}s^{-1/2}}{t+s}\,\mathrm{d}s \\
&= \frac{1}{\pi}\int_{1/\sqrt{1-t}}^{0} \frac{t^{1/2}x}{t+1/x^2}\cdot\frac{-2}{x^3}\,\mathrm{d}x \qquad (\text{let } x = s^{-1/2}) \\
&= \frac{2t^{1/2}}{\pi}\int_0^{1/\sqrt{1-t}} \frac{1}{1+tx^2}\,\mathrm{d}x \\
&= \frac{2}{\pi}\arctan\Big(\sqrt{\frac{t}{1-t}}\Big) \qquad \Big(\text{since } \int \frac{1}{1+tx^2}\,\mathrm{d}x = \frac{1}{\sqrt{t}}\arctan(\sqrt{t}x) + C\Big) \\
&= \frac{2}{\pi}\arcsin(\sqrt{t}).
\end{aligned}
$$

□

Theorem 6.60 (Another arcsine law)**.** *Let* $M = \arg\max_{t\in[0,1]} B_t = \inf\{t \ge 0 : B_t = M_1\}$, *then*

$$\mathbb{P}_0(M \le t) = \frac{2}{\pi}\arcsin(\sqrt{t}).$$

Proof. By Proposition 6.57, $M_t - B_t$ and $|B_t|$ has the same distribution,

$$\begin{aligned} &\mathbb{P}_0(L = \sup\{t \in [0,1] : B_t = 0\} \leq s) \\ = &\mathbb{P}_0(\sup\{t \in [0,1] : |B_t| = 0\} \leq s) \\ = &\mathbb{P}_0(\sup\{t \in [0,1] : M_t - B_t = 0\} \leq s) \\ = &\mathbb{P}_0(\inf\{t \geq 0 : B_t = M_1\} \leq s) \\ = &\mathbb{P}_0(M \leq s), \end{aligned}$$

therefore M and L have the same distribution. □

6.7 p-variation and quadratic variation

Definition 6.61. 1. We say $\Pi = \{t_0, t_1, \cdots, t_n\}$ is a partition of the interval $[0, T]$ if

$$0 = t_0 < t_1 < \cdots < t_n = T.$$

2. The mesh (maximal interval length) of the partition Π is

$$|\Pi| = \max_{0 \leq k \leq n-1} |t_{k+1} - t_k|.$$

3. The p-variation of function $f : [0, T] \to \mathbb{R}$ over partition Π is

$$V^p(f, [0,T], \Pi) = \sum_{k=1}^{n} |f(t_k) - f(t_{k-1})|^p.$$

4. The p-variation of function $f : [0, T] \to \mathbb{R}$ is

$$V^p(f, [0,T]) = \sup_{\Pi} \sum_{k=0}^{n-1} |f(t_{k+1}) - f(t_k)|^p$$

5. The quadratic variation of $f : [0, T]$ is

$$QV(f, [0, T]) = \lim_{|\Pi_m| \to 0} V^2(f, [0, T], \Pi_m),$$

where $\{\Pi_m\}_{m=1}^{\infty}$ is a sequence of partition s.t. their mesh shrinks to 0. Sometime we denote quadratic variation as $[f, f](T)$.

Remark. 1. Quadratic variation is a different concept from 2-variation, and

$$QV(f, [0, T]) \leq V^2(f, [0, T]).$$

2. For the random process $\{X_t : t \geq 0\}$, we can define p-variation over the partition Π for each path $\omega \in \Omega$ (the partition does not depend on ω), which is a random variable. And the quadratic variation of X_t can be viewed as a limit (convergence in probability, L^2 or a.s.) of a sequence of random variables.

Theorem 6.62. *$QV(B_t, [0, T]) = T$ in the sense of L^2 limit.*

Proof. Suppose $\Pi = \{0 = t_0 < t_1 < \cdots < t_n = T\}$. First, we have

$$\mathbb{E}[V^2(B_t, [0, T], \Pi)] = \mathbb{E}\left[\sum_{k=1}^{n-1} |B_{t_{k+1}} - B_{t_k}|^2\right] = \sum_{k=1}^{n-1} \mathbb{E}[|B_{t_{k+1}} - B_{t_k}|^2] = \sum_{k=1}^{n-1} (t_{k+1} - t_k) = T.$$

And

$$\begin{aligned}
\mathrm{Var}[V^2(B_t,[0,T],\Pi)] &= \mathrm{Var}\left[\sum_{k=1}^{n-1}|B_{t_{k+1}}-B_{t_k}|^2\right] \\
&= \sum_{k=1}^{n-1}\mathrm{Var}[|B_{t_{k+1}}-B_{t_k}|^2] \qquad \text{(by independence)} \\
&= \sum_{k=1}^{n-1}\mathrm{Var}[(t_{k+1}-t_k)\xi_k^2] \qquad (\text{here } \xi_k\sim\mathcal{N}(0,1)) \\
&= 2\sum_{k=1}^{n-1}(t_{k+1}-t_k)^2 \qquad (\mathrm{Var}(\xi_k^2)=2) \\
&\le 2\max_k(t_{k+1}-t_k)\sum_{k=1}^{n-1}(t_{k+1}-t_k) \\
&= 2|\Pi|T\to 0,
\end{aligned}$$

as $|\Pi|\to 0$. Therefore

$$\mathbb{E}[|V^2(B_t,[0,T],\Pi)-T|^2]=\mathrm{Var}[V^2(B_t,[0,T],\Pi)]\to 0, \quad \text{as } |\Pi|\to 0. \qquad \square$$

We will see two corollaries of Theorem 6.62.

Lemma 6.63. *If $f:[0,T]\to\mathbb{R}$ is continuous, and $\Pi=\{0=t_0<\cdots<t_n=T\}$ is a partition of $[0,T]$. Then*

$$\max_{0\le k\le n-1}|f(t_{k+1})-f(t_k)|\to 0,$$

as $|\Pi|\to 0$.

Proof. Any continuous function is uniformly continuous on a closed interval, so for any $\varepsilon>0$, there is $\delta>0$ s.t. for any $s,t\in[0,T]$ with $|s-t|<\delta$,

$$|f(s)-f(t)|<\varepsilon.$$

So we can choose $|\Pi| < \delta$, then $|f(t_{k+1}) - f(t_k)| < \varepsilon$ for all $k = 0, \cdots, n-1$ and thus

$$\max_k |f(t_{k+1}) - f(t_k)| < \varepsilon,$$

i.e. its limit is 0 as $|\Pi| \to 0$ □

Corollary 6.64. *Let $V^p(B_t, [0,T])$ be the p-variation for the Brownian motion.*

(1) *If $p > 2$, $V^p(B_t, [0,T]) < \infty$ a.s.*

(2) *If $0 < p < 2$, $V^p(B_t, [0,T]) = \infty$ a.s.*

Proof. (1)By the γ-Hölder continuity of Brownian motion for any $\gamma \in (0, 1/2)$, there is $C > 0$ s.t.

$$|B_t - B_s| \le C|t-s|^\gamma, \quad \forall t, s \in [0,T], \qquad a.s.$$

Therefore choose $\gamma = 1/p < 1/2$, we have w.p.1.

$$\begin{aligned} V^p(B, [0,T]) &= \sup_\Pi \sum_{k=0}^{n-1} |B_{t_{k+1}} - B_{t_k}|^p \\ &\le \sup_\Pi \sum_{k=0}^{n-1} C^p |t_{k+1} - t_k|^{\gamma \cdot p} \\ &= C^p \sup_\Pi \sum_{k=0}^{n-1} (t_{k+1} - t_k) \\ &= C^p T < \infty. \end{aligned}$$

(2)Suppose there is an event $A \in \mathcal{F}$ s.t. $\mathbb{P}(A) > 0$ and $V^p(B(\omega), [0,1]) < \infty$ for all path

$\omega \in A$. For partition $\Pi = \{0 = t_0 < \cdots < t_n = T\}$,

$$\begin{aligned} V^2(B_t, [0,T], \Pi) &= \sum_{k=0}^{n-1} |B_{t_{k+1}} - B_{t_k}|^2 \\ &= \sum_{k=0}^{n-1} |B_{t_{k+1}} - B_{t_k}|^p |B_{t_{k+1}} - B_{t_k}|^{2-p} \\ &\le \max_k |B_{t_{k+1}} - B_{t_k}|^{2-p} \sum_{k=0}^{n-1} |B_{t_{k+1}} - B_{t_k}|^p \\ &\le \left(\max_k |B_{t_{k+1}} - B_{t_k}| \right)^{2-p} \cdot V^p(B, [0,T]) \end{aligned}$$

Since B_t is continuous for all paths in some set A_0 with $\mathbb{P}(A_0) = 1$, by Lemma 6.63, we have for any path $\omega \in A \cap A_0$ (easy to check $\mathbb{P}(A \cap A_0) > 0$)

$$\lim_{|\Pi| \to 0} V^2(B(\omega), [0,T], \Pi) = 0,$$

then

$$\mathbb{E}[|V^2(B(\omega), [0,T], \Pi) - T|^2] \ge \mathbb{E}[|V^2(B(\omega), [0,T], \Pi) - T|^2 \mathbb{1}_{A \cap A_0}] \to \mathbb{E}[\mathbb{1}_{A \cap A_0}] = \mathbb{P}(A \cap A_0) > 0,$$

thus

$$\lim_{|\Pi| \to 0} \mathbb{E}[|V^2(B(\omega), [0,T], \Pi) - T|^2] \ge \mathbb{P}(A \cap A_0) > 0,$$

which contradicts Theorem 6.62. □

Corollary 6.65. *Brownian motion is nowhere γ-Hölder continuous for $\gamma > 1/2$.*

Proof. For a fixed interval $[a,b]$ with $a, b \in \mathbb{Q} \cap [0, \infty)$, suppose on a non-null set $A \in \mathcal{F}$, B_t is γ-Hölder continuous on $[a,b]$ for $\gamma > 1/2$. Then for any partition $\Pi = \{a = t_0 < \cdots < t_n = b\}$

and for any path $\omega \in A$,

$$\begin{aligned} V^2(B(\omega),[a,b],\Pi) &= \sum_{k=0}^{n-1} |B_{t_{k+1}}(\omega) - B_{t_k}(\omega)|^2 \\ &\le C^2(\omega) \sum_{k=0}^{n-1} |t_{k+1} - t_k|^{2\gamma} \\ &\le C^2(\omega) \max_k |t_{k+1} - t_k|^{2\gamma-1} \sum_{k=0}^{n-1} |t_{k+1} - t_k| \\ &= C^2(\omega)|\Pi|^{2\gamma-1}(b-a) \to 0, \end{aligned}$$

as $|\Pi| \to 0$, which contradicts Theorem 6.62 (By the same argument in Corollary 6.64). Therefore for $\gamma > 1/2$ and any rational interval $[a,b] \subseteq [0,\infty)$,

$$\mathbb{P}\left(B_t \text{ is } \gamma\text{-Hölder continuous on } [a,b]\right) = 0,$$

then by the fact that countable union of null sets are still null, we have

$$\begin{aligned} &\mathbb{P}\left(B_t \text{ is } \gamma\text{-Hölder continuous on some rational interval } [a,b] \subseteq [0,\infty)\right) \\ &= \mathbb{P}\left(\bigcup_{a,b\in\mathbb{Q}\cap[0,\infty),a<b} \{B_t \text{ is } \gamma\text{-Hölder continuous on } [a,b]\}\right) = 0. \end{aligned}$$

If B_t is γ-Hölder continuous on some interval $[a,b]$, then B_t is also γ-Hölder continuous on any closed interval $[c,d] \subseteq [a,b]$. Therefore,

$$\begin{aligned} &\mathbb{P}\left(\text{there is no interval } [a,b] \subseteq [0,\infty) \text{ s.t. } B_t \text{ is } \gamma\text{-Hölder continuous on it}\right) \\ &= \mathbb{P}\left(\{B_t \text{ is } \gamma\text{-Hölder continuous on some real interval } [a,b] \subseteq [0,\infty)\}^c\right) \\ &= \mathbb{P}\left(\{B_t \text{ is } \gamma\text{-Hölder continuous on some rational interval } [a,b] \subseteq [0,\infty)\}^c\right) = 1. \end{aligned}$$

□

Remark. This corollary provides a second proof of Theorem 6.16.

Almost sure convergence

In what condition, the quadratic variation of Brownian motion will be the almost sure limit?

Theorem 6.66. *Let* $\{\Pi_m\}_{m=1}^{\infty}$ *be a sequence of partition of* $[0,T]$ *with* $|\Pi_m| \to 0$. *If*

$$\sum_{m=1}^{\infty} |\Pi_m| < \infty,$$

then

$$V^2(B, [0,T], \Pi_m) \to T, \quad a.s. \qquad as\ |\Pi_m| \to 0.$$

Proof. Let $V_m^2 := V^2(B_t, [0,T], \Pi_m)$, by the proof of Theorem 6.62, we have

$$\mathrm{Var}(V_m^2) \leq 2|\Pi_m|T,$$

then by Chebyshev's inequality, for any $\varepsilon > 0$,

$$\mathbb{P}(|V_m^2 - T| > \varepsilon) \leq \frac{\mathrm{Var}(V_m^2)}{\varepsilon^2} = \frac{2|\Pi_m|T}{\varepsilon^2}.$$

Summing over m, we have

$$\sum_{m=1}^{\infty} \mathbb{P}(|V_m^2 - T| > \varepsilon) < \infty,$$

by Borel-Cantelli Lemma, we have

$$\mathbb{P}(|V_m^2 - T| > \varepsilon,\ i.o.) = 0.$$

Let $A := \{|V_m^2 - T| > \varepsilon,\ i.o.\}$, then for $\omega \in A^c$ ($\mathbb{P}(A^c) = 1$), $|V_m^2(\omega) - T| > \varepsilon$ holds only for finitely many m, i.e. there is $N > 0$, s.t.

$$|V_m^2(\omega) - T| \le \varepsilon, \quad \forall m \ge N,$$

which means $V_m^2(\omega) \to T$. □

Theorem 6.67. *Let $\{\Pi_m\}_{m=1}^\infty$ be a sequence of partition of $[0,T]$ with $|\Pi_m| \to 0$. If $\{\Pi_m\}_{m=1}^\infty$ is nested, i.e. $\Pi_1 \subseteq \Pi_2 \subseteq \cdots \subseteq \Pi_m \subseteq \cdots$, then*

$$V^2(B, [0,T], \Pi_m) \to T, \quad a.s. \qquad as\ |\Pi_m| \to 0.$$

6.8 Martingale

Theorem 6.68. *Suppose X_t is a right continuous martingale w.r.t. a right continuous filtration, T is a stopping time. If $\mathbb{P}(T \le k) = 1$ for some k, then $\mathbb{E}(X_T) = \mathbb{E}(X_0)$.*

Proposition 6.69. *Suppose $\{B_t : t \ge 0\}$ is a Brownian motion starting from x, then*

1. B_t

2. $B_t^2 - t$

3. $e^{\theta B_t - t\theta^2/2}$

are martingales w.r.t. $\mathcal{F}_t$.

Proof. 1. By Markov property, for any $0 \le s \le t$,

$$\mathbb{E}_x(B_t|\mathcal{F}_s) = \mathbb{E}_x(B_{t-s} \circ \theta_s|\mathcal{F}_s) = \mathbb{E}_{B_s}(B_{t-s}) = B_s,$$

the last equality holds because B_{t-s}(starting from B_s)$\sim \mathcal{N}(B_s, t-s)$.

2. $\mathbb{E}_x(B_t^2|\mathcal{F}_s) = \mathbb{E}_{B_s}(B_{t-s}^2) = \mathrm{Var}_{B_s}(B_{t-s}) + [\mathbb{E}_{B_s}(B_{t-s})]^2 = t - s + B_s^2$, so $\mathbb{E}_x(B_t^2 - t|\mathcal{F}_s) = B_s^2 - t$.

3. Similarly, we have

$$\mathbb{E}_x(e^{\theta B_t}|\mathcal{F}_s) = \mathbb{E}_{B_s}(e^{\theta B_{t-s}}) = e^{\theta B_s + (t-s)\theta^2/2},$$

since for $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\mathbb{E}(e^{\theta X}) = \int_{-\infty}^{\infty} e^{\theta x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, \mathrm{d}x = e^{\mu\theta + \sigma^2\theta^2/2}.$$

□

Theorem 6.70. *If $a < x < b$, then*

$$\mathbb{P}_x(T_a < T_b) = \frac{b-x}{b-a}.$$

Proof. Let $T = T_a \wedge T_b$, then by Proposition 6.31, $T < \infty$ a.s. (Because w.p.1. $B_m = \infty$ and $B_n = \infty$ for i.o. $m, n \in \mathbb{Z}_+$). Thus for any $t \in [0, \infty]$, $T \wedge t < \infty$. Thus by Theorem 6.68 and Proposition 6.69,

$$\mathbb{E}_x(B_{T\wedge t}) = \mathbb{E}_x(B_0) = x.$$

Since $|B_{T\wedge t}| \le |B_T| \le |a| + |b| < \infty$, by bounded convergence theorem,

$$\mathbb{E}_x(B_T) = \lim_{t\to\infty} \mathbb{E}_x(B_{T\wedge t}) = x,$$

then

$$x = \mathbb{E}_x(B_T) = a\mathbb{P}_x(B_T = a) + b\mathbb{P}_x(B_T = b) = a\mathbb{P}_x(T_a < T_b) + b[1 - \mathbb{P}_x(T_a < T_b)],$$

i.e.

$$\mathbb{P}_x(T_a < T_b) = \frac{b-x}{b-a}.$$

□

Proposition 6.71. *Let $a < 0 < b$, $T = \inf\{t \geq 0 : B_t \notin (a, b)\}$. Then*

$$\mathbb{E}_0(T) = -ab.$$

Proof. Consider bounded stopping time $T \wedge t$, since $B_t^2 - t$ is a martingale,

$$\mathbb{E}_0(B_{T\wedge t}^2 - T \wedge t) = \mathbb{E}_0(B_0^2 - 0^2) = 0,$$

i.e.

$$\mathbb{E}_0(B_{T\wedge t}^2) = \mathbb{E}_0(T \wedge t).$$

Since $T \wedge t \uparrow T$, by the monotone convergence theorem,

$$\lim_{t\to\infty} \mathbb{E}_0(T \wedge t) = \mathbb{E}_0(T).$$

By $|B_{T\wedge t}^2| \leq a^2 \vee b^2 < \infty$ and bounded convergence theorem, we have

$$\lim_{t\to\infty} \mathbb{E}_0(B_{T\wedge t}^2) = \mathbb{E}_0(B_T^2),$$

thus

$$\begin{aligned}
\mathbb{E}_0(T) &= \mathbb{E}_0(B_T^2) \\
&= a^2\mathbb{P}_0(B_T = a) + b^2\mathbb{P}_0(B_T = b) \\
&= a^2\mathbb{P}_0(T_a < T_b) + b^2[1 - \mathbb{P}_0(T_a < T_b)] \\
&= a^2 \cdot \frac{b}{b-a} + b^2(1 - \frac{b}{b-a}) \\
&= -ab.
\end{aligned}$$

□

Proposition 6.72. *Let $a, \lambda > 0$, $T_a = \inf\{t \geq 0 : B_t = a\}$, then*

$$\mathbb{E}_0(e^{-\lambda T_a}) = e^{-a\sqrt{2\lambda}}.$$

Proof. Since $\varphi(t) = e^{\theta B_t - t\theta^2/2}$ is a martingale,

$$\mathbb{E}_0(\varphi(T_a \wedge t)) = \mathbb{E}_0(\varphi(0)) = 1.$$

Bounded convergence theorem ($B_{T_a \wedge t} \leq a$, so $e^{B_{T_a \wedge t}} \leq e^a$) and monotone convergence theorem ($e^{(T_a \wedge t)\theta^2/2} \uparrow e^{T_a\theta^2/2}$) give

$$\mathbb{E}_0(\varphi(T_a \wedge t)) = \mathbb{E}_0[e^{\theta B_{T_a \wedge t} - (T_a \wedge t)\theta^2/2}] \to \mathbb{E}_0[e^{\theta B_{T_a} - T_a\theta^2/2}] = e^{\theta a}\mathbb{E}_0[e^{-T_a\theta^2/2}],$$

therefore

$$\mathbb{E}_0(e^{-T_a\theta^2/2}) = e^{-\theta a},$$

taking $\theta = -\sqrt{2\lambda}$ gives the desired result. □

Theorem 6.73. *If $u(t, x)$ is a polynomial in t and x satisfying*

$$\frac{\partial u}{\partial t} + \frac{1}{2}\frac{\partial^2 u}{\partial x^2} = 0,$$

then $u(t, B_t)$ is a martingale.

Proposition 6.74. *For $a > 0$, let $T = \inf\{t \geq 0 : B_t \notin (-a, a)\}$. Then*

1. *B_T and T are independent.*

2. $\mathbb{E}_0(T) = a^2$.

3. $\mathbb{E}_0(T^2) = \dfrac{5a^4}{3}$.

4. $\mathbb{E}_0(T^3) = \dfrac{61a^6}{15}$.

References

[1] Kai-Lai Chung. *A course in Probability Theory*. China Machine Press, third edition, 2010.

[2] Rick Durrett. *Probability: Theory and Examples*. Cambradge University Press, fifth edition, 2019.

[3] Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes*. Oxford University Press, third edition, 2001.

[4] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer New York, second edition, 1998.

[5] Achim Klenke. *Probability Theory: A Comprehensive Course*. Springer London, second edition, 2014.

[6] Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill Publishing Company, third edition, 1976.

[7] Zhan Shi. *Branching Random Walks*. Springer Cham, first edition, 2015.