MMAT 5340: Probability and Stochastic Analysis

Lecturer: Xiaolu Tan

March 25, 2025

Contents

1 Probability theory review

1.1 Basic probability theory

A probability space is a triple $(\Omega, \mathcal{F}, \mathbb{P})$, where

- Ω is the sample space, which is a (non-empty) set.
- $\mathcal{F}$ is a a σ-field, which is a space of subsets of Ω satisfying
  - $\Omega \in \mathcal{F}$,
  - $A \in \mathcal{F} \implies A^C \in \mathcal{F}$,
  - $A_n \in \mathcal{F},\ n \geq 1 \implies \cup_{n\geq 1} A_n \in \mathcal{F}$.

  A set $A \in \mathcal{F}$ is called an event.
- $\mathbb{P} : \mathcal{F} \longrightarrow [0,1]$ is a probability measure, i.e.
  - $\mathbb{P}[\Omega] = 1$,
  - If $\{A_n,\ n \geq 1\} \subset \mathcal{F}$ be such that $A_i \cap A_j = \emptyset$ for all $i \neq j$, then $\mathbb{P}[\cup_{n\geq 1} A_n] = \sum_{n\geq 1} \mathbb{P}[A_n]$.

Example 1.1. (i) *$\Omega = \{1, 2, \cdots, n\}$, $\mathcal{F} := \sigma(\{1\}, \cdots, \{n\})$, $\mathbb{P}[\{i\}] = \frac{1}{n}$, for each $i = 1, \cdots, n$. In above, $\sigma(\{1\}, \cdots, \{n\})$ means the smallest σ-field containing all events $\{1\}, \cdots, \{n\}$. In this case, it is the space of all subsets of Ω.*

(ii) *$\Omega = \mathbb{R}$, $\mathcal{F} := \mathcal{B}(\mathbb{R})$ is the Borel σ-field on $\mathbb{R}$, i.e. the smallest σ-field which contains all open set in $\mathbb{R}$. For some density function $\rho : \mathbb{R} \longrightarrow \mathbb{R}_+$, a probability measure $\mathbb{P}$ can be defined, first for all intervals (a, b) with $a \leq b$, by $\mathbb{P}[(a,b)] := \int_a^b \rho(x)dx$, and then extended on the Borel σ-field $\mathcal{F}$.*

A random variable is a map $X : \Omega \longrightarrow \mathbb{R}$ satisfying

$$X^{-1}(A) := \{\omega \in \Omega\ : X(\omega) \in A\} \in \mathcal{F},\ \text{for all } A \in \mathcal{B}(\mathbb{R}) \iff \{X \leq x\} \in \mathcal{F},\ \text{for all } x \in \mathbb{R}.$$

The distribution function of X is given by

$$F(x) \ := \ \mathbb{P}[X \leq x],\ x \in \mathbb{R}.$$

Example 1.2. (i) *A discrete random variable X:*

$$p_i = \mathbb{P}[X = x_i],\ i \in \mathbb{N}, \qquad \sum_{i\in\mathbb{N}} p_i = 1.$$

(ii) *A continuous random variable X (with continuous probability distribution), one has the density function*

$$\rho(x) = F'(x),\ x \in \mathbb{R}.$$

(iii) *There exists a some random variable, whoseis distribution neither discrete nor continuous.*

Expectation Let X be a (discrete or continuous) random variable, the expectation of $\mathbb{E}[f(X)]$ is defined as follows:

- When X is a discrete random variable such that $\mathbb{P}[X = x_i] = p_i$ for $i \in \mathbb{N}$. Then

$$\mathbb{E}[f(X)] \ := \ \sum_{i \in \mathbb{N}} f(x_i)\mathbb{P}[X = x_i] \ = \ \sum_{i \in \mathbb{N}} f(x_i) p_i.$$

- When X is a continuous random variable with density $\rho : \mathbb{R} \longrightarrow \mathbb{R}_+$. Then

$$\mathbb{E}[f(X)] \ := \ \int_{\mathbb{R}} f(x)\rho(x)dx, \quad \text{whenever the integral is well defined.}$$

Remark 1.3. *In general case, one defines the expectation as the following Lebesgue integration:*

$$\mathbb{E}[f(X)] \ := \ \int_{\Omega} f(X(\omega))d\mathbb{P}(\omega).$$

A rigorous definition of the above integral needs the measure theory, which is not required in this course.

For two (square integrable) random variables X and Y, their variance and co-variance are defined by

$$\mathrm{Var}[X] := \mathbb{E}\big[\big(X - \mathbb{E}[X]\big)^2\big], \quad \mathrm{Cov}[X, Y] := \mathbb{E}\big[\big(X - \mathbb{E}[X]\big)\big(Y - \mathbb{E}[Y]\big)\big].$$

The characteristic function of X is defined by $\Phi(\theta) := \mathbb{E}[e^{i\theta X}]$.

Independence The events $A_1, \cdots, A_n \in \mathcal{F}$ are said to be (mutually) independent if

$$\mathbb{P}\big[A_1 \cap \cdots \cap A_n\big] = \prod_{i=1}^n \mathbb{P}[A_i].$$

Next, we say that the σ-fields $\mathcal{F}_1, \cdots, \mathcal{F}_n$ are (mutually) independent if

$$\mathbb{P}\big[A_1 \cap \cdots \cap A_n\big] = \prod_{i=1}^n \mathbb{P}[A_i], \text{ for all } A_1 \in \mathcal{F}_1, \cdots, A_n \in \mathcal{F}_n.$$

Finally, we say that random variables $X_1, \cdots, X_n$ are (mutually) independent if

$$\sigma(X_1), \cdots, \sigma(X_n) \text{ are independent.}$$

Remark 1.4. (i) *The σ-field $\sigma(X_1)$ is defined as the smallest σ-field containing all events*

$$\{X_1 \le x\} \ := \ \{\omega \in \Omega \ : X_1(\omega) \le x\}, \ \textit{for all } x \in \mathbb{R}.$$

As X_1 is a random variable, it is clear that $\sigma(X_1) \subset \mathcal{F}$.

(ii) *We say that the a random variable X_1 is independent of $\mathcal{F}_2$ if $\sigma(X_1)$ and $\mathcal{F}_2$ are independent.*

Example 1.5. *Let us consider the case, where* $\Omega = \{0,1,2,3\}$, $\mathbb{P}[X=\omega] = \frac{1}{4}$, *define*

$$X_1(\omega) = \begin{cases} 0 & \omega \in \{0,2\}, \\ 1 & \omega \in \{1,3\}, \end{cases} \quad X_2(\omega) = \begin{cases} 0 & \omega \in \{0,1\}, \\ 1 & \omega \in \{2,3\}. \end{cases}$$

In this case, $\sigma(X_1) = \{\emptyset, \Omega, \{0,2\}, \{1,3\}\}$, *and* $\sigma(X_2) = \{\emptyset, \Omega, \{0,1\}, \{2,3\}\}$. *Moreover, it can be checked that* X_1 *is independent of* $\sigma(X_2)$. *For example, one can check that*

$$\mathbb{P}\big[\{X_1 = 0\} \cap \{X_2 = 0\}\big] \;=\; \mathbb{P}\big[\{0\}\big] \;=\; \mathbb{P}\big[\{0,2\}\big]\mathbb{P}\big[\{0,1\}\big] \;=\; \frac{1}{4},$$

which implies that the two events $\{X_1 = 0\}$ *and* $\{X_2 = 0\}$ *are independent. Similarly, one can check that* $\{X_1 = i\}$ *is independent of* $\{X_2 = j\}$ *for all* $i, j \in \{0,1\}$. *This is enough to show that* X_1 *and* X_2 *are independent.*

Lemma 1.6. *If* $X_1, \cdots, X_n$ *are independent,* f_i *are measurable functions. Then* $f_1(X_1), \cdots, f_n(X_n)$ *are independent.*

Proof. Let us consider the case $n = 2$. To prove that $f_1(X_1)$ is independent of $f_2(X_2)$, it is enough to check that the event $\{f_1(X_1) \le y_1\}$ is independent of the event $\{f_2(X_2) \le y_2\}$ for all real numbers $y_1, y_2 \in \mathbb{R}$. At the same time, we notice that $\{f_i(X_i) \le y_i\} = \{X_i \in f_i^{-1}((-\infty, y_i])\} \in \sigma(X_i)$. Since $\sigma(X_1)$ is independent of $\sigma(X_2)$, this is enough to conclude the proof. □

Lemma 1.7. *If* $X_1, \cdots, X_n$ *are independent, then*

$$\mathbb{E}[f_1(X_1) \cdots f_n(X_n)] \;=\; \mathbb{E}[f_1(X_1)] \cdots \mathbb{E}[f_n(X_n)].$$

Consequently,

$$\mathrm{Var}[X_1 + \cdots + X_n] = \mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n].$$
$$\mathrm{Cov}[f_i(X_i), f_j(X_j)] = 0, \;\; i \neq j.$$

Remark 1.8. *: The inverse may not be correct. Let us consider a random variable* $X_1 \sim \mathcal{U}[-1,1]$ *follows the uniform distribution on* $[-1,1]$, *whose density function is given by* $\rho(x) = \frac{1}{2}\mathbf{1}_{\{-1 \le x \le 1\}}$. *Let* $X_2 := X_1^2$. *By direct computation, one can check that*

$$\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2], \quad \textit{and hence} \;\; \mathrm{Cov}[X_1, X_2] = 0.$$

Nevertheless, it is clear that X_1 *and* X_2 *are not independent.*

We next provide some notions of convergence of random variables. Let $(X_n)_{n \ge 1}$ a sequence of random variables, ans X be a r.v.

- Almost sure convergence: We say X_n converges almost surely to X if

$$\mathbb{P}\big[\lim_{n \to \infty} X_n = X\big] = 1.$$

- Convergence in probability: We say X_n converges to X in probability if, for any $\varepsilon > 0$,

$$\lim_{n \to \infty} \mathbb{P}[|X_n - X| \ge \varepsilon] \;=\; 0.$$

- Convergence in distribution: We say X_n converges to X in distribution if, for any bounded continuous function f,

$$\lim_{n \to \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)].$$

- Convergence in L^p ($p \geq 1$) space: Assume $\mathbb{E}[|X_n|^p] < \infty$, we say X_n converges to X in L^p space if
$$\lim_{n\to\infty} \mathbb{E}\big[|X_n - X|^p\big] \;=\; 0.$$

Lemma 1.9 (Relations between the different notions of the convergence)**.** *One has*

$$\text{Cvg a.s.} \implies \text{Cvg in prob.} \implies \text{Cvg in dist.},$$

$$\text{Cvg in } L^p \implies \text{Cvg in prob.}$$

$$\text{Cvg in prob.} \implies \text{Cvg a.s. along a subsequence.}$$

Lemma 1.10 (Monotone convergence theorem)**.** *Assume that $0 \leq X_n \leq X_{n+1}$ for all $n \geq 1$, then*
$$\mathbb{E}\big[\lim_{n\to\infty} X_n\big] \;=\; \lim_{n\to\infty} \mathbb{E}[X_n].$$

Remark 1.11. *In practice, we may have $X_n := f_n(X)$ for a sequence $(f_n)_{n\geq 1}$ satisfying $0 \leq f_1 \leq f_2 \leq \cdots$. In this case, we have*

$$\mathbb{E}\big[\lim_{n\to\infty} f_n(X)\big] \;=\; \lim_{n\to\infty} \mathbb{E}[f_n(X)].$$

Theorem 1.1 (Law of Large Number)**.** *Assume that $(X_n)_{n\geq 1}$ is an i.i.d. sequence with the same distribution of X and such that $\mathbb{E}[|X|] < \infty$. Then*

$$\lim_{n\to\infty} \overline{X}_n \;:=\; \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k \;=\; \mathbb{E}[X], \;\; a.s.$$

Theorem 1.2 (Central Limit Theorem)**.** *Assume that $(X_n)_{n\geq 1}$ is an i.i.d. sequence with the same distribution of X and such that $\mathbb{E}[|X|^2] < \infty$. Then*

$$\frac{\sqrt{n}\big(\overline{X}_n - \mathbb{E}[X]\big)}{\sqrt{\mathrm{Var}[X]}} \;\; \textit{converges in distribution to } N(0,1).$$

We finally provide some useful inequalities.

Lemma 1.12 (Jensen inequality)**.** *Let X be a r.v., ϕ be a convex function. Assume that $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[|\phi(X)|] < \infty$. Then*
$$\phi(\mathbb{E}[X]) \;\leq\; \mathbb{E}\big[\phi(X)\big].$$

Proof. As ϕ is a convex function, there exists an affine function $g(x) = ax + b$ such that

$$\phi(\mathbb{E}[X]) = g(\mathbb{E}[X]), \quad \text{and } \phi(x) \geq g(x) \text{ for all } x \in \mathbb{R}.$$

Therefore,

$$\mathbb{E}[\phi(X)] \;\geq\; \mathbb{E}[g(X)] \;=\; \mathbb{E}[aX + b] \;=\; a\mathbb{E}[X] + b \;=\; g(\mathbb{E}[X]) \;=\; \phi(\mathbb{E}[X]).$$

□

Lemma 1.13 (Chebychev inequality)**.** *Let X be a r.v., $f : \mathbb{R} \to \mathbb{R}_+$ be an increasing function. Assume that $\mathbb{E}[f(X)] < \infty$ and $f(a) > 0$. Then*

$$\mathbb{P}[X \geq a] \;\leq\; \frac{\mathbb{E}[f(X)]}{f(a)]}.$$

Proof. We will prove this for continuous random variable X, and the proof for discrete random variable X is essentially the same, replacing integrals with sums. Let $\rho(x)$ be the probability density function of X. By definition, $\mathbb{E}[f(X)] = \int_{-\infty}^{\infty} f(x)\rho(x)dx$. By monotonicity of $f(x)$, and the fact that $f(x), \rho(x)$ are non-negative,

$$\begin{aligned}
\mathbb{E}[f(X)] &= \int_{-\infty}^{\infty} f(x)\rho(x)dx \\
&= \int_{-\infty}^{a} f(x)\rho(x)dx + \int_{a}^{\infty} f(x)\rho(x)dx \\
&\geq \int_{a}^{\infty} f(x)\rho(x)dx \\
&\geq \int_{a}^{\infty} f(a)\rho(x)dx
\end{aligned}$$

the result follows by taking out the constant $f(a)$ from the integral. □

Lemma 1.14 (Cauchy-Schwarz inequality)**.** *Let X and Y be two r.v. Assume that $\mathbb{E}[|X|^2] < \infty$ and $\mathbb{E}[|Y|^2] < \infty$. Then*

$$\mathbb{E}[XY] \leq \sqrt{\mathbb{E}[|X|^2]\mathbb{E}[|Y|^2]}.$$

1.2 Conditional expectation

Theorem 1.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$, X a random variable. Assume that $\mathbb{E}[|X|] < \infty$. Then there exists a random variable Z satisfying the following:*

- $\mathbb{E}[|Z|] < \infty$.
- *Z is $\mathcal{G}$-measurable.*
- *$\mathbb{E}[XY] = \mathbb{E}[ZY]$, for all $\mathcal{G}$-measurable bounded random variables Y.*

Moreover, the random Z is unique in the sense of almost sure.

Definition 1.15. *We say that the random variable Z given in Theorem 1.3 is the conditional expectation of X knowing $\mathcal{G}$, and denote*

$$\mathbb{E}[X|\mathcal{G}] := Z.$$

When $\mathcal{G} = \sigma(Y_1, \cdots, Y_n)$, for $Y = (Y_1, \cdots .Y_n)$, we also write

$$\mathbb{E}[X|Y_1, \cdots, Y_n] \ := \ \mathbb{E}[X|\mathcal{G}].$$

In this case, there exists a measurable function $f : \mathbb{R}^n \to \mathbb{R}$ such that $\mathbb{E}[X|Y] = f(Y)$. To compute $\mathbb{E}[X|Y]$, it is enough to compute the function:

$$\mathbb{E}[X|Y = y] \ := \ f(y), \text{ for all } y \in \mathbb{R}^n.$$

Example 1.16. (i) *Discrete case: $\mathbb{P}[X = x_i, Y = y_j] = p_{i,j}$ with $\sum_{i,j} p_{i,j} = 1$. Then*

$$\mathbb{E}[X|Y = y_j] \ = \ \frac{\mathbb{E}[X\mathbf{1}_{Y=y_j}]}{\mathbb{E}[\mathbf{1}_{Y=y_j}]} \ = \ \frac{\sum_{i\in\mathbb{N}} x_i p_{i,j}}{\sum_{i\in\mathbb{N}} p_{i,j}}.$$

Proof. Let us denote $f(y_j) := \frac{\sum_{i\in\mathbb{N}} x_i p_{i,j}}{\sum_{i\in\mathbb{N}} p_{i,j}}$, then it is enough to show that $\mathbb{E}[X|Y] = f(Y)$.

First, it is trivial that $f(Y)$ is $\sigma(Y)$-measurable.
Next, by direct computation,

$$\mathbb{E}[|f(Y)|] = \sum_{j\in\mathbb{N}} |f(y_j)|\mathbb{P}[Y = y_j] = \sum_{j\in\mathbb{N}} \frac{|\sum_{i\in\mathbb{N}} x_i p_{i,j}|}{\sum_{i\in\mathbb{N}} p_{i,j}} \sum_{i\in\mathbb{N}} p_{i,j} \le \sum_{i,j\in\mathbb{N}} |x_i| p_{i,j} = \mathbb{E}[|X|] < \infty.$$

Finally, for any $\sigma(Y)$-measurable bounded random variable Z, there exists a measurable function $g : \mathbb{R}^n \to \mathbb{R}$ such that $Z = g(Y)$, then we have

$$\mathbb{E}[f(Y)g(Y)] = \sum_{j\in\mathbb{N}} f(y_j)g(y_j)\mathbb{P}[Y = y_j] = \sum_{i,j\in\mathbb{N}} x_i g(y_j) p_{i,j} = \mathbb{E}[Xg(Y)].$$

This is enough to conclude the proof by the definition of conditional expectation. □

(ii) *Continuous case: Let $\rho(x, y)$ be the density function of (X, Y), and assume that $\int_{\mathbb{R}} \rho(x, y)dx > 0$ for all $y \in \mathbb{R}$. Then*

$$\mathbb{E}[X|Y = y] = \frac{\int_{\mathbb{R}} x\rho(x, y)dx}{\int_{\mathbb{R}} \rho(x, y)dx}. \tag{1}$$

Proof. Let us denote the r.h.s. of (1) as $f(y)$. Then it is enough to show that $\mathbb{E}[X|Y] = f(Y)$.

First, it is clear that $f(Y)$ is $\sigma(Y)$-measurable.
Next,

$$\begin{aligned}\mathbb{E}[|f(Y)|] &= \int_{\mathbb{R}}\int_{\mathbb{R}} |f(y)|\rho(x, y)dxdy = \int_{\mathbb{R}}\int_{\mathbb{R}} \left|\frac{\int_{\mathbb{R}} x\rho(x, y)dx}{\int_{\mathbb{R}} \rho(x, y)dx}\right|\rho(x, y)dxdy \\ &\le \int_{\mathbb{R}}\int_{\mathbb{R}} \frac{\int_{\mathbb{R}} |x|\rho(x, y)dx}{\int_{\mathbb{R}} \rho(x, y)dx}\rho(x, y)dxdy = \int_{\mathbb{R}}\int_{\mathbb{R}} |x|\rho(x, y)dxdy = \mathbb{E}[|X|] < \infty.\end{aligned}$$

Finally, for any $\sigma(Y)$-measurable bounded random variable Z, there exists a measurable function $g : \mathbb{R}^n \to \mathbb{R}$ such that $Z = g(Y)$, then we have

$$\begin{aligned}\mathbb{E}\big[f(Y)g(Y)\big] &= \int_{\mathbb{R}}\int_{\mathbb{R}} f(y)g(y)\rho(x, y)dxdy = \int_{\mathbb{R}}\int_{\mathbb{R}} \frac{\int_{\mathbb{R}} x\rho(x, y)dx}{\int_{\mathbb{R}} \rho(x, y)dx} g(y)\rho(x, y)dxdy \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} xg(y)\rho(x, y)dxdy = \mathbb{E}[Xg(Y)].\end{aligned}$$

This shows that $\mathbb{E}[X|Y] = f(Y)$ by the definition of conditional expectation. □

Example 1.17. *Let X and Y be two independent random variables with the same distribution, and $\mathbb{P}[X = \pm 1] = \mathbb{P}[X = \pm 1] = \frac{1}{2}$. One can compute that*

$$\mathbb{E}[X] = 0, \quad \textit{and } \mathbb{E}[X + Y|Y] = Y.$$

We finally provide some properties of the conditional expectation from its definition.

Lemma 1.18. *Let X and Y be two r.v. such that $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[|Y|] < \infty$, a, b be two real numbers. Then*

$$\mathbb{E}[aX + bY|\mathcal{G}] = a\mathbb{E}[X|\mathcal{G}] + b\mathbb{E}[Y|\mathcal{G}].$$

Proof. It is enough to verify that $a\mathbb{E}[X|\mathcal{G}]+b\mathbb{E}[Y|\mathcal{G}]$ satisfies the three properties in the definition of the conditional expectation $\mathbb{E}[aX+bY|\mathcal{G}]$.

First, $a\mathbb{E}[X|\mathcal{G}]+b\mathbb{E}[Y|\mathcal{G}]$ is obviously $\mathcal{G}$-measurable.

Next, from the definition of conditional expectation, we know $\mathbb{E}[|\mathbb{E}[X|\mathcal{G}]|], \mathbb{E}[|\mathbb{E}[Y|\mathcal{G}]|] < \infty$, then

$$\mathbb{E}[|a\mathbb{E}[X|\mathcal{G}]+b\mathbb{E}[Y|\mathcal{G}]|] \le |a|\mathbb{E}[|\mathbb{E}[X|\mathcal{G}]|]+|b|\mathbb{E}[|\mathbb{E}[Y|\mathcal{G}]|] < \infty.$$

Finally, for any $\mathcal{G}$-measurable bounded random variable Z, we know that

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]Z] = \mathbb{E}[XZ], \mathbb{E}[\mathbb{E}[Y|\mathcal{G}]Z] = \mathbb{E}[YZ].$$

Then by linearity of expectation, we have

$$\begin{aligned}\mathbb{E}[(a\mathbb{E}[X|\mathcal{G}]+b\mathbb{E}[Y|\mathcal{G}])Z] &= a\mathbb{E}[\mathbb{E}[X|\mathcal{G}]Z]+b\mathbb{E}[\mathbb{E}[Y|\mathcal{G}])Z] \\ &= a\mathbb{E}[XZ]+b\mathbb{E}[YZ] = \mathbb{E}[(aX+bY)Z].\end{aligned}$$

□

Lemma 1.19. *Let X, Y be r.v. such that $\mathbb{E}[|X|] < \infty$, Y is $\mathcal{G}$-measurable and $\mathbb{E}[|XY|] < \infty$, then*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X], \quad \textit{and} \quad \mathbb{E}[XY|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]Y.$$

If X is independent of $\mathcal{G}$, then

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X].$$

Proof. First, by taking $Y = \mathbb{1}_\Omega$ in the third property in Theorem 1.3, it follows immediately that $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$.

To prove $\mathbb{E}[XY|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]Y$, it is equivalent to verify that $\mathbb{E}[X|\mathcal{G}]Y$ satisfies the three properties in the definition of conditional expectation for $\mathbb{E}[XY|\mathcal{G}]$, by the uniqueness of the conditional expectation.

Let us first assume that X and Y are nonnegative. Then for any $k \in \mathbb{N}$, then $\mathbb{E}[X|\mathcal{G}]\,(Y \wedge k)$ is $\mathcal{G}$-measurable since both of $\mathbb{E}[X|\mathcal{G}]$ and $(Y \wedge k)$ are $\mathcal{G}$-measurable. Moreover, for the integrability, one has

$$\mathbb{E}[|\mathbb{E}[X|\mathcal{G}](Y\wedge k)|] \le k\mathbb{E}[|\mathbb{E}[X|\mathcal{G}]|] < \infty.$$

Finally, for any bounded $\mathcal{G}$-measurable r.v. Z, $(Y \wedge k)Z$ is bounded and $\mathcal{G}$-measurable, then one has

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}](Y\wedge k)Z] = \mathbb{E}[X(Y\wedge k)Z] = \mathbb{E}[\mathbb{E}[X(Y\wedge k)|\mathcal{G}]Z].$$

Hence it follows that

$$\mathbb{E}[X(Y\wedge k)|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}](Y\wedge k).$$

Then by monotone convergence theorem for conditional expectation (see Lemma 1.21 below), one obtains that

$$\mathbb{E}[X|\mathcal{G}]Y = \lim_{k\to+\infty}\mathbb{E}[X|\mathcal{G}](Y\wedge k) = \lim_{k\to+\infty}\mathbb{E}[X(Y\wedge k)|\mathcal{G}] = \mathbb{E}[\lim_{k\to+\infty}X(Y\wedge k)|\mathcal{G}] = \mathbb{E}[XY|\mathcal{G}].$$

When X, Y are not always nonnegative, one can write $X = X^+ - X^-$, $Y = Y^+ - Y^-$, where X^+, X^-, Y^+ and Y^- are all nonneagive random variables. Then

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

Moreover, $\mathbb{E}[X|\mathcal{G}]Y$ is $\mathcal{G}$-measurable since both of $\mathbb{E}[X|\mathcal{G}]$ and Y are $\mathcal{G}$-measurable. One can also check the integrability condition by

$$\mathbb{E}[|\mathbb{E}[X|\mathcal{G}]Y|] = \mathbb{E}[|\mathbb{E}[XY|\mathcal{G}]|] < \infty,$$

which proves that $\mathbb{E}[XY|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]Y$.

Finally, when X is independent of $\mathcal{G}$, we consider $\mathbb{E}[X]$ as a constant r.v., and check that it satisfies the properties in the definition of conditional expectation $\mathbb{E}[X|\mathcal{G}]$. As a constant r.v., $\mathbb{E}[X]$ is clearly $\mathcal{G}$-measurable and integrable. Moreover, for any bounded $\mathcal{G}$-measurability r.v. Z, we have by linearity of expectation

$$\mathbb{E}[\mathbb{E}[X]Z] = \mathbb{E}[XZ].$$

This proves that $\mathbb{E}[X]$ is the conditional expectation of X knowing $\mathcal{G}$. □

Lemma 1.20. *Let X be a random variable, φ be a convex function. Then*

$$\mathbb{E}[\varphi(X)|\mathcal{G}] \geq \varphi(\mathbb{E}[X|\mathcal{G}]), \ \textit{a.s.}$$

Proof. We first prove monotonicity for conditional expectation. Claim that if X, Y are r.v. such that $\mathbb{E}[|X|]$, $\mathbb{E}[|Y|] < \infty$ and $X \geq Y$, then $\mathbb{E}[X|\mathcal{G}] \geq \mathbb{E}[Y|\mathcal{G}]$ a.s. To see this, set $Z := \mathbb{E}[X - Y|\mathcal{G}]$ and $A := \{\omega : Z < 0\}$. Since $A \in \mathcal{G}$ by definition and $(X - Y) \geq 0$ a.s., $\mathbb{E}[Z\mathbf{1}_A] = E[(X-Y)\mathbf{1}_A] \geq 0$ so $\mathbb{P}[Z < 0]] = P[\mathbb{E}[X|\mathcal{G}] < \mathbb{E}[Y|\mathcal{G}]] = 0$ as claimed.

Recall that a function $f : \mathbb{R} \to \mathbb{R}$ is convex if and only if there exits a family $\{f_n\}$ of affine functions (i.e. $f_n(x) = a_n x + b_n$, for some $a_n, b_n \in \mathbb{R}$) such that

$$f(x) \ = \ \sup_n f_n(x), \quad \text{for all } x \in \mathbb{R}.$$

Thus,

$$\mathbb{E}[\varphi(X)|\mathcal{G}] \ \geq \ \mathbb{E}[a_n X + b_n|\mathcal{G}] \ = \ a_n \mathbb{E}[X|\mathcal{G}] + b_n.$$

By taking supremum over both sides, it follows that

$$\mathbb{E}[\varphi(X)|\mathcal{G}] \ \geq \ \sup_n \{a_n \mathbb{E}[X|\mathcal{G}] + b_n\} \ = \ \varphi(\mathbb{E}[X|\mathcal{G}]).$$

□

Lemma 1.21 (Monotone convergence theorem)**.** *Let $(X_n, n \geq 1)$ be a sequence of integrable random variable such that $0 \leq X_n \leq X_{n+1}$, a.s. Then*

$$\lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] = \mathbb{E}[\lim_{n\to\infty} X_n|\mathcal{G}].$$

Proof. Notice that by the increasing of $\{X_n\}_n$ for almost all ω, we have

$$\mathbb{E}[X_n|\mathcal{G}] \leq \mathbb{E}[\lim_{n\to\infty} X_n|\mathcal{G}] \text{ a.s.}$$

Then with the same procedure in the proof of conditional Jensen's Inequality, we can prove that $0 \leq \mathbb{E}[X_n|\mathcal{G}] \leq \mathbb{E}[X_{n+1}|\mathcal{G}]$ a.s. and we get the existence of $\lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}]$. Taking the limit in the above inequality, we have

$$\lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] \leq \mathbb{E}[\lim_{n\to\infty} X_n|\mathcal{G}] \text{ a.s.}$$

Then the monotone convergence theorem (Lemma 1.10) implies that

$$\mathbb{E}[\lim_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}]] \;=\; \lim_{n\to\infty} \mathbb{E}[\mathbb{E}[X_n|\mathcal{G}]] \;=\; \lim_{n\to\infty} \mathbb{E}[X_n] \;=\; \mathbb{E}[\lim_{n\to\infty} X_n] \;=\; \mathbb{E}[\mathbb{E}[\lim_{n\to\infty} X_n|\mathcal{G}]].$$

Hence we conclude the proof. □

Lemma 1.22. *Let X be an integrable random variable, and $\mathcal{G} := \{\emptyset, \Omega\}$. Then*

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X].$$

Proof. It is equivalent to prove that any $\mathcal{G}$-measurable random variable Z is a constant random variable a.s.

By contradiction, we assume that Z is not a constant random variable. Then there exist some constants $C_1, C_2 \in \mathbb{R}$ with $C_1 < C_2$ such that

$$\{Z = C_1\} \neq \phi, \; \{Z = C_2\} \neq \phi.$$

Hence we have $\{Z \leq C_1\} \notin \mathcal{G}$, which gives the fact that Z is not $\mathcal{G}$-measurable. Now since this is a contradiction, we complete the proof. □

Lemma 1.23. *Let X be an integrable random variable, and $\mathcal{G}_1 \subset \mathcal{G}_2$ be two sub-σ-field of $\mathcal{F}$. Then*

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1] = \mathbb{E}[X|\mathcal{G}_1].$$

Proof. Set $Z := \mathbb{E}[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1]$, it is enough to verify that Z satisfies the three properties in the definition of $\mathbb{E}[X|\mathcal{G}_1]$.

First, Z is obviously $\mathcal{G}_1$-measurable and integrable, as it is defined as the conditional expectation of some random variable knowing $\mathcal{G}_1$. Moreover, for any $\mathcal{G}_1$-measurable bounded random variable Y, we know by Lemma 1.19 that

$$\begin{aligned} \mathbb{E}[ZY] \;&=\; \mathbb{E}[\,\mathbb{E}[\,\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1]Y] \;=\; \mathbb{E}[\,\mathbb{E}[\,\mathbb{E}[X|\mathcal{G}_2]Y|\mathcal{G}_1]] \\ &=\; \mathbb{E}[\,\mathbb{E}[X|\mathcal{G}_2]Y] \;=\; \mathbb{E}[\,\mathbb{E}[XY|\mathcal{G}_2]] \;=\; \mathbb{E}[XY]. \end{aligned}$$

This concludes the proof. □

2 Discrete time martingale

Definition 2.1. *In a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*, a stochastic process is a family* $(X_n)_{n\geq 0}$ *of random variables indexed by time* $n \geq 0$ *(or* $t_n,\ n \geq 0$*). A filtration is family* $\mathbb{F} = (\mathcal{F}_n)_{n\geq 0}$ *of sub-*σ*-field of* $\mathcal{F}$ *such that* $\mathcal{F}_n \subset \mathcal{F}_{n+1}$ *for all* $n \geq 0$*.*

Example 2.2. *Let* $B = (B_n)_{n\geq 0}$ *be some stochastic process, then the following definition of* $\mathcal{F}_n$ *provides a filtration* $(\mathcal{F}_n)_{n\geq 0}$*:*

$$\mathcal{F}_n \ := \ \sigma(B_0, B_1, \cdots, B_n).$$

In particular, let $B_0 = 0$*,* $B_n = \sum_{k=1}^n \xi_k$ *where* $(\xi_k)_{k\geq 1}$ *is an i.i.d. sequence of random variables with distribution* $\mathbb{P}[\xi_k = \pm 1] = \frac{1}{2}$*. Then*

$$\mathcal{F}_0 = \{\emptyset, \Omega\}, \quad \mathcal{F}_1 = \mathcal{F}_0 \cup \{A, A^c\}, \ \text{with } A := \{\xi_1 = 1\}, \ A^c = \{\xi_1 = -1\}, \ \cdots$$

Definition 2.3. *Let* $X = (X_n)_{n\geq 0}$ *be a stochastic process,* $\mathbb{F} = (\mathcal{F}_n)_{n\geq 1}$ *be a filtration.*

We say X *is adapted to the filtration* $\mathbb{F}$ *if*

$$X_n \in \mathcal{F}_n \ (\text{i.e. } X_n \text{ is } \mathcal{F}_n\text{-measurable}), \text{ for all } n \geq 0.$$

We say X *is predictable w.r.t.* $\mathbb{F}$ *if*

$$X_n \in \mathcal{F}_{(n-1)\vee 0} \text{ for all } n \geq 0.$$

Remark 2.4. *Let* $\mathbb{F}$ *be the filtration generated by the process* B *as in the above example. If* X *is* $\mathbb{F}$*-adapted, then* $X_n \in \mathcal{F}_n = \sigma(B_0, \cdots, B_n)$ *so that*

$$X_n = g_n(B_0, \cdots, B_n), \ \text{for some measurable function } g_n.$$

Similarly, if X *is* $\mathbb{F}$*-predictable, then* $X_{n+1} \in \mathcal{F}_n$ *so that*

$$X_{n+1} = g'_{n+1}(B_0, \cdots, B_n), \ \text{for some measurable function } g'_{n+1}.$$

Example 2.5. *Let* $(\xi_k)_{k\geq 1}$ *be a sequence of i.i.d random variable, such that* $\mathbb{P}\big[\xi_k = \pm 1\big] = \frac{1}{2}$*. Then the process* $X = (X_n)_{n\geq 0}$ *defined as follows is called a random walk:*

$$X_0 = 0, \quad X_n = \sum_{k=1}^n \xi_k.$$

Remark 2.6. *In above examples, a stochastic process usually starts from time* 0*, but we can also consider stochastic process starting from some time* t_k*.*

Definition 2.7. *Let* $X = (X_n)_{n\geq 0}$ *be a stochastic process,* $\mathbb{F} = (\mathcal{F}_n)_{n\geq 1}$ *be a filtration.*

We say X *is a martingale (w.r.t.* $\mathbb{F}$*) if* X *is* $\mathbb{F}$*-adapted, each random variable* X_n *is integrable, and*

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

We say X *is a sub-martingale (w.r.t.* $\mathbb{F}$*) if* X *is* $\mathbb{F}$*-adapted, each random variable* X_n *is integrable, and*

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

We say X *is a super-martingale (w.r.t.* $\mathbb{F}$*) if* X *is* $\mathbb{F}$*-adapted, each random variable* X_n *is integrable, and*

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n] \leq X_n.$$

Notice that martingale X (w.r.t. to some filtration $\mathbb{F}$) is a sub-martingale, and at the same time a super-martingale.

Example 2.8. *Recall that the random walk $X = (X_n)_{n\geq 0}$ is defined as follows:*

$$X_0 = 0, \quad X_n = \sum_{k=1}^{n} \xi_k,$$

where $(\xi_k)_{k\geq 1}$ be a sequence of i.i.d. of random variable such that $\mathbb{P}[\xi = \pm 1] = \frac{1}{2}$.

Then

- *X is a martingale;*
- *$(X_n^2)_{n\geq 0}$ is a sub-martingale;*
- *$(X_n^2 - n)_{n\geq 0}$ is a martingale.*

Proof. First, it is clear that X is $\mathbb{F}$-adapted with respect to the natural filtration $\mathbb{F}$ generated by X, and X_n is integrable for all $n \geq 0$. Then by using Lemma 1.19,

$$\begin{aligned}
\mathbb{E}[X_{n+1}|\mathcal{F}_n] &= \mathbb{E}[X_n + \xi_{n+1}|\mathcal{F}_n] \\
&= \mathbb{E}[X_n|\mathcal{F}_n] + \mathbb{E}[\xi_{n+1}|\mathcal{F}_n] \\
&= X_n + \mathbb{E}[\xi_{n+1}] \\
&= X_n.
\end{aligned}$$

Next, as $(X_n^2)_{n\geq 0}$ is $\mathbb{F}$-adapted, and X_n^2 is integrable, for $\forall n \geq 0$, we compute that

$$\begin{aligned}
\mathbb{E}[X_{n+1}^2|\mathcal{F}_n] &= \mathbb{E}[(X_n + \xi_{n+1})^2|\mathcal{F}_n] \\
&= \mathbb{E}[X_n^2 + 2X_n\xi_{n+1} + \xi_{n+1}^2|\mathcal{F}_n] \\
&= \mathbb{E}[X_n^2|\mathcal{F}_n] + 2\mathbb{E}[X_n\xi_{n+1}|\mathcal{F}_n] + \mathbb{E}[\xi_{n+1}^2|\mathcal{F}_n] \\
&= X_n^2 + 2X_n\mathbb{E}[\xi_{n+1}|\mathcal{F}_n] + \mathbb{E}[\xi_{n+1}^2] \\
&= X_n^2 + 1.
\end{aligned}$$

Finally, $Y_n := X_n^2 - n$ is $\mathbb{F}$-adapted, and Y_n is integrable, then

$$\begin{aligned}
\mathbb{E}[Y_{n+1}|\mathcal{F}_n] &= \mathbb{E}[X_{n+1}^2 - (n+1)|\mathcal{F}_n] \\
&= X_n^2 + 1 - (n+1) \\
&= X_n^2 - n \\
&= Y_n.
\end{aligned}$$

□

Example 2.9. *Let $(Z_k)_{k\geq 1}$ be a sequence of random variable such that $Z_k \sim N(0,1)$, and $\sigma \in \mathbb{R}$, $X_0 \in \mathbb{R}$ be real constants. Let $\mathcal{F}_n := \sigma(Z_1, \cdots, Z_n)$, and*

$$X_n := X_0 \exp\Big(\sigma \sum_{k=1}^{n} Z_k - \frac{1}{2} n\sigma^2\Big).$$

Then $(X_n)_{n\geq 1}$ is a martingale (w.r.t. $\mathbb{F}$).

Example 2.10. *Let $\mathbb{F} = (\mathcal{F}_n)_{n\geq 1}$ be a filtration, Z be an integrable random variable, and*

$$X_n \ := \ \mathbb{E}\big[Z\big|\mathcal{F}_n\big].$$

Then $(X_n)_{n\geq 1}$ is a martingale (w.r.t. $\mathbb{F}$).

Lemma 2.11. *Let $\mathbb{F}$ be a filtration, and X be a martingale w.r.t. $\mathbb{F}$. Let $\mathbb{F}^X$ denote the natural filtration generated by X. Then X is also a martingale w.r.t. $\mathbb{F}^X$.*

Proof. Given that X is $\mathbb{F}$-adapted, we know that $X_s \in \mathcal{F}_n$ for $s \in \{0, 1, \cdots, n\}$. Define $\mathcal{F}_n^X$ as the σ-field generated by $X_0, X_1, \cdots, X_n$, i.e. $\mathcal{F}_n^X := \sigma(X_0, X_1, \cdots, X_n)$, then $\mathcal{F}_n^X \subset \mathcal{F}_n$. We know that X is $\mathbb{F}^X$-adapted, X_n is integrable for $\forall n \geq 0$, and

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n^X] = \mathbb{E}[\ \mathbb{E}[\ X_{n+1}|\mathcal{F}_n]|\mathcal{F}_n^X] = \mathbb{E}[X_n|\mathcal{F}_n^X] = X_n,$$

then it is clear that X is a martingale with respect to $\mathbb{F}^X$. □

Notice that a martingale X is associated to some filtration $\mathbb{F}$. However, when the filtration is not specified, we say X is a martingale means that X is a martingale w.r.t. the natural filtration generated by X. In this case, we can also write

$$\mathbb{E}\big[X_{n+1}\big|X_0, \cdots, X_n\big] = X_n, \quad \text{for all } n \geq 0.$$

Lemma 2.12. *Let X be a martingale w.r.t. the filtration $\mathbb{F}$, then*

$$\mathbb{E}\big[X_m\big|\mathcal{F}_n\big] \ = \ X_n, \ \textit{for all } m \geq n \geq 0.$$

Moreover,

$$\mathbb{E}\big[X_n\big] = \mathbb{E}\big[X_0\big], \ \textit{for all } n \geq 0.$$

Proof. As X is a martingale, we know that $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$ and $\mathcal{F}_n \subset \mathcal{F}_{n+1}$. Then by the tower property in Lemma 1.23,

$$\mathbb{E}[X_{n+2}|\mathcal{F}_n] = \mathbb{E}[\ \mathbb{E}[\ X_{n+2}|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n.$$

The result follows by using the above equation. □

2.1 Optional stopping theorem

Definition 2.13. *Let $\mathbb{F}$ be a filtration, a stopping time w.r.t. $\mathbb{F}$ is a random variable $\tau : \Omega \longrightarrow \{0, 1, \cdots\} \cup \{\infty\}$ such that*

$$\{\tau \leq n\} \in \mathcal{F}_n, \quad \textit{for all } n \geq 0. \tag{2}$$

Remark 2.14. *In place of (2), it is equivalent to define the stopping time by the property:*

$$\{\tau = n\} \in \mathcal{F}_n, \quad \textit{for all } n \geq 0.$$

Proof. We can write

$$\{\tau = n\} \ = \ \{\tau \leq n\} \setminus \{\tau \leq n-1\}, \tag{3}$$

$$\{\tau \leq n\} \ = \ \bigcup_{k=0}^{n} \{\tau = k\}. \tag{4}$$

Now if $\{\tau \leq n\} \in \mathcal{F}_n$ for any $n \geq 0$, then $\{\tau \leq n-1\} \in \mathcal{F}_{n-1} \subset \mathcal{F}_n$, hence we know from (3) that $\{\tau = n\} \in \mathcal{F}_n$.

Next, if $\{\tau = n\} \in \mathcal{F}_n$ for any $n \geq 0$, then for any $0 \leq k \leq n$, $\{\tau = k\} \in \mathcal{F}_k \subset \mathcal{F}_n$, hence we know from (4) that $\{\tau \leq n\} \in \mathcal{F}_n$. □

Lemma 2.15. *Let X be a stochastic process adapted to the filtration $\mathbb{F}$, and B be a Borel set in $\mathbb{R}$. Then the hitting time τ defined below is a stopping w.r.t. $\mathbb{F}$:*

$$\tau \ := \ \inf\{n \geq 0 \ : X_n \in B\},$$

where $\inf \emptyset = +\infty$ *by convention.*

Proof. For any $n \in \mathbb{N}$, notice the facts that

$$\begin{aligned} \{\tau = n\} \ &= \ \{X_n \in B\} \bigcap \bigcap_{k=0}^{n-1} \{X_k \notin B\}, \\ \{\tau \leq n\} \ &= \ \bigcup_{k=0}^{n} \{X_k \in B\}, \\ \{X_k \in B\} \in \mathcal{F}_k &\subset \mathcal{F}_n \text{ for any } k = 0, 1, \cdots, n. \end{aligned}$$

It follows that $\{\tau \leq n\} \in \mathcal{F}_n$ for any $n \geq 0$. Then τ is a stopping time w.r.t. $\mathbb{F}$. □

Given a stochastic process X and a stopping time τ w.r.t. some filtration $\mathbb{F}$.

$$X_{\tau \wedge n}(\omega) \ := \ \begin{cases} X_n(\omega) & \text{if } \tau(\omega) \geq n, \\ X_{\tau(\omega)}(\omega) & \text{if } \tau(\omega) < n. \end{cases}$$

Theorem 2.1. *Let $\mathbb{F}$ be fixed filtration, X be a $\mathbb{F}$-martingale, and τ be a $\mathbb{F}$-stopping time. Then the process $(X_{\tau \wedge n})_{n \geq 0}$ is still a $\mathbb{F}$-martingale.*

Proof. Let us denote $Y_n := X_{\tau \wedge n}$ for any $n \in \mathbb{N}$, then we can write for any $n \geq 0$,

$$Y_n \ = \ \sum_{k=0}^{n-1} X_k \mathbb{1}_{\{\tau = k\}} + X_n \mathbb{1}_{\{\tau \geq n\}}, \tag{5}$$

$$\ = \ \sum_{k=0}^{n-1} X_k \mathbb{1}_{\{\tau = k\}} + X_n \mathbb{1}_{\{\tau > n-1\}}, \tag{6}$$

Now we verify the three conditions in the definition of martingale.

First, for any $n \in \mathbb{N}$, we have by (5)

$$|Y_n| \leq \sum_{k=0}^{n} |X_k|.$$

Then by the integrability of X, we know that

$$\mathbb{E}[|Y_n|] \leq \sum_{k=0}^{n} \mathbb{E}[|X_k|] < +\infty.$$

Next, since τ is a $\mathbb{F}$-stopping time, we have for any $k = 0, 1, \cdots, n$,

$$\{\tau = k\} \in \mathcal{F}_k \subset \mathcal{F}_n, \quad \{\tau > n-1\} = \{\tau \leq n-1\}^C \in \mathcal{F}_{n-1} \subset \mathcal{F}_n.$$

Then $X_k \mathbb{1}_{\{\tau = k\}}$ is $\mathcal{F}_k$-measurable, hence $\mathcal{F}_n$-measurable and $X_n \mathbb{1}_{\{\tau > n-1\}}$ is also $\mathcal{F}_n$-measurable. Thus by (5), we have Y_n is $\mathcal{F}_n$-measurable.

Finally, we prove that for any $n \in \mathbb{N}$

$$\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = Y_n \text{ a.s.}$$

By (5), we have

$$\begin{aligned}
\mathbb{E}[Y_{n+1}|\mathcal{F}_n] &= \mathbb{E}[\sum_{k=0}^{n} X_k \mathbb{1}_{\{\tau=k\}} + X_{n+1}\mathbb{1}_{\{\tau>n\}}|\mathcal{F}_n] = \sum_{k=0}^{n} X_k \mathbb{1}_{\{\tau=k\}} + \mathbb{E}[X_{n+1}|\mathcal{F}_n]\mathbb{1}_{\{\tau>n\}} \\
&= \sum_{k=0}^{n-1} X_k \mathbb{1}_{\{\tau=k\}} + X_n \mathbb{1}_{\{\tau\geq n\}} = Y_n \text{ a.s.}
\end{aligned}$$

□

When X is martingale and τ is a stopping w.r.t. the same filtration, it follows that

$$\mathbb{E}\big[X_{\tau\wedge n}\big] = \mathbb{E}\big[X_0\big].$$

The question is that whether one has $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$.

In order to answer the question, we introduce a version of the dominated convergence theorem below.

Lemma 2.16. *Let $\{Z_n\}_{n\geq 0}$ be a sequence of random variables with $\lim_{n\to\infty} Z_n = Z$ a.s. for some random variable Z and $\sup_{n\in\mathbb{N}} |Z_n| \leq M$ a.s. for some constant $M > 0$, then*

$$\lim_{n\to\infty} \mathbb{E}[Z_n] = \mathbb{E}[Z].$$

Proof. Let us denote that $X_n = \inf_{k\geq n}(2M - |Z_k - Z|)$ for any $n \in \mathbb{N}$, then it is clear that $0 \leq X_n \leq X_{n+1}$ for all $n \geq 1$ and $\lim_{n\to\infty} X_n = 2M$ a.s.

By Lemma 1.10, we have

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[\lim_{n\to\infty} X_n] = 2M,$$

Then we know that

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{E}[|Z_n - Z|] &\leq \lim_{n\to\infty} \mathbb{E}\Big[\sup_{k\geq n} |Z_k - Z|\Big] = -\lim_{n\to\infty} \mathbb{E}\Big[\inf_{k\geq n}(2M - |Z_k - Z|) - 2M\Big] \\
&= -\lim_{n\to\infty} \mathbb{E}\Big[\inf_{k\geq n}(2M - |Z_k - Z|)\Big] + 2M = -\lim_{n\to\infty} \mathbb{E}[X_n] + 2M \\
&= -\mathbb{E}\Big[\lim_{n\to\infty} X_n\Big] + 2M = -\mathbb{E}\Big[\lim_{n\to\infty} \inf_{k\geq n}(2M - |Z_k - Z|)\Big] + 2M \\
&= -\mathbb{E}[2M] + 2M = 0.
\end{aligned}$$

Hence, we have

$$\lim_{n\to\infty} \mathbb{E}[Z_n] = \mathbb{E}[Z].$$

□

Theorem 2.2. *Let $\mathbb{F}$ be a fixed filtration, X be a $\mathbb{F}$-martingale, and τ be a $\mathbb{F}$-stopping time. Assume that τ is bounded by some constant $m \geq 0$, or $\tau < \infty$ and the process $(X_{\tau\wedge n})_{n\geq 0}$ is uniformly bounded. Then*

$$\mathbb{E}\big[X_\tau\big] = \mathbb{E}\big[X_0\big].$$

Proof. First, we claim that

$$\lim_{n\to\infty} \mathbb{E}[X_{\tau\wedge n}] = \mathbb{E}[X_\tau]. \tag{7}$$

By Theorem 2.1, we have $X_{\tau\wedge\cdot}$ is a $\mathbb{F}$-martingale, then for any $n \in \mathbb{N}$,

$$\mathbb{E}[X_{\tau\wedge n}] = \mathbb{E}[X_0],$$

which combined with (7), implies that

$$\mathbb{E}[X_\tau] = \mathbb{E}[X_0].$$

Then it remains to prove the claim (7).

If τ is bounded by some constant $m \geq 0$, then for any $n \geq m$, we have $X_{\tau\wedge n} = X_\tau$, hence (7) remains true.

If $(X_{\tau\wedge n})_{n\geq 0}$ is uniformly bounded, by Lemma 2.16 and $\lim_{n\to\infty} X_{\tau\wedge n} = X_\tau$ a.s., (7) remains true. □

Example 2.17. *Let $(\xi_k)_{k\geq 1}$ be a sequence of i.i.d. random variables, $x \in \mathbb{N}$ be a positive integer, and*

$$X_n \;:=\; x + \sum_{k=1}^{n} \xi_k.$$

Let us define

$$\tau \;:=\; \inf\big\{n \geq 0 \;: X_n \leq 0 \text{ or } X_n \geq N\big\}.$$

Assume $\tau < \infty$, we can then compute the value of $\mathbb{E}\big[X_\tau\big]$ and $\mathbb{P}\big[X_\tau = 0\big]$.

2.2 Convergence of martingale

Theorem 2.3. *Let X be a submartingale or supermartingale such that $\sup_{n\geq 0} \mathbb{E}[|X_n|] < \infty$. Then*

$$\lim_{n\to\infty} X_n = X_\infty, \text{ for some r.v. } X_\infty \in L^1.$$

Proof. We will prove the case when X is a supermartingale, and the submartingale case follows by taking $-X$ as a supermartingale. Recall that the limit of a sequence of real numbers $(X_n)_{n\geq 1}$ does not exist if and only if one of the following holds:

1. $\lim_{n\to\infty} X_n = \infty$
2. $\lim_{n\to\infty} X_n = -\infty$
3. $\underline{\lim}_{n\to\infty} X_n < \overline{\lim}_{n\to\infty} X_n$.

Set $A_1 = \{\omega : \lim_{n\to\infty} X_n(\omega) = +\infty\}$, $A_2 = \{\omega : \lim_{n\to\infty} X_n(\omega) = -\infty\}$, $A_3 = \{\omega : \underline{\lim}_{n\to\infty} X_n(\omega) < +\overline{\lim}_{n\to\infty} X_n(\omega)\}$. If $\mathbb{P}[A_1] = \mathbb{P}[A_2] = \mathbb{P}[A_3] = 0$, then the result follows.

Given $\epsilon > 0$, we first assume that $\mathbb{P}[A_1] \geq \epsilon > 0$. Then $\forall M > 0, \exists N$ such that $X_n \geq M$ for $\forall n \geq N$. We know that $\mathbb{E}[|X_n|] \geq \mathbb{E}[|X_n|\mathbf{1}_{A_1}] \geq M\epsilon > C$ for large enough M, where $C = \sup_{n\geq 0} \mathbb{E}[|X_n|]$. This leads to a contradiction that $C = \sup_{n\geq 0} \mathbb{E}[|X_n|] < \infty$ and we can conclude that $\mathbb{P}[A_1] = 0$. Similarly, we can prove $\mathbb{P}[A_2] = 0$.

To show $P[A_3] = 0$, choose two rational numbers a and b such that $\underline{\lim}_{n\to\infty} X_n \le a < b \le \overline{\lim}_{n\to\infty} X_n$, we introduce two sequences of stopping times $(\sigma_n)_{n\ge1}, (\tau_n)_{n\ge1}$ by:

$$\begin{aligned}
\sigma_1 &:= \inf\{n \ge 1 : X_n \le a\}\\
\tau_1 &:= \inf\{n \ge \sigma_1 : X_n \ge b\}\\
\sigma_2 &:= \inf\{n \ge \tau_1 : X_n \le a\}\\
\tau_2 &:= \inf\{n \ge \sigma_2 : X_n \ge b\}.
\end{aligned}$$

It can be observed that at time τ_1, the process X has crossed $[a, b]$ once, and at time τ_2, the process X has crossed $[a, b]$ twice. Let $U_n(a, b) := \max\{k : \tau_k \le n\}$.

Claim that $\mathbb{E}[U_n(a,b)] \le \frac{\mathbb{E}[|X_n-a|]}{b-a}$. If this holds, then $\sup_{n\ge1} \mathbb{E}[U_n(a,b)] \le \sup_{n\ge1} \frac{\mathbb{E}[|X_n-a|]}{b-a}$. We know by Monotone Convergence Theorem that

$$\mathbb{E}[\lim_{n\to\infty} U_n(a,b)] = \lim_{n\to\infty} \mathbb{E}[U_n(a,b)] \le \sup_{n\ge1} \frac{\mathbb{E}[|X_n-a|]}{b-a} < \infty.$$

Thus $\lim_{n\to\infty} U_n(a,b) < \infty$ a.s., and $P[\ \underline{\lim}_{n\to\infty} X_n \le a < b \le \overline{\lim}_{n\to\infty} X_n\ ] = 0$. We then find from subadditivity that

$$\begin{aligned}
\mathbb{P}[A_3] &= \mathbb{P}[\ \underline{\lim}_{n\to\infty} X_n \le \overline{\lim}_{n\to\infty} X_n\ ]\\
&= \mathbb{P}[\cup_{\substack{a<b\\ a,b\in\mathbb{Q}}} \{\ \underline{\lim}_{n\to\infty} X_n \le a < b \le \overline{\lim}_{n\to\infty} X_n\}]\\
&\le \sum_{\substack{a<b\\ a,b\in\mathbb{Q}}} \mathbb{P}[\ \underline{\lim}_{n\to\infty} X_n \le a < b \le \overline{\lim}_{n\to\infty} X_n]\\
&= 0.
\end{aligned}$$

Finally, we prove $\mathbb{E}[U_n(a,b)] \le \frac{\mathbb{E}[|X_n-a|]}{b-a}$. Let $H_k := \sum_{i=1}^{\infty} \mathbf{1}_{\sigma_i \le k < \tau_i}$ and $V_n := \sum_{k=0}^{n-1} H_k(X_{k+1} - X_k)$. We claim that $V = (V_n)_{n\ge1}$ is a supermartingale. Indeed,

$$\mathbb{E}[V_{n+1} - V_n|\mathcal{F}_n] = H_n \mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n] \le 0.$$

Thus we know that $V_n \ge (b-a)\cdot U_n(a,b) - |X_n - a|$ by taking the first term and the second term as profit from the crossing event and loss of the last investment, respectively. Then

$$0 \ge \mathbb{E}[V_n] \ge \mathbb{E}[(b-a)U_n(a,b)] - \mathbb{E}[|X_n - a|].$$

We obtain the desired result. □

Theorem 2.4. *Let X be a martingale such that $\sup_{n\ge0} \mathbb{E}[|X_n|^2] < \infty$. Then*

$$\lim_{n\to\infty} X_n = X_\infty, \text{ for some r.v. } X_\infty \in L^2.$$

and

$$\lim_{n\to\infty} \mathbb{E}[|X_n - X_\infty|^2] = 0.$$

Proof. Recall from Cauchy-Schwarz inequality that $\sup_{n\ge1} \mathbb{E}[|X_n|] \le \sup_{n\ge1} \sqrt{\mathbb{E}[|X_n|^2]} \le \infty$. Then $\lim_{n\to\infty} X_n$ exists by 2.3.

We first denote that $\Delta X_n := X_n - X_{n-1}, n \geq 1$. We claim that

$$\mathbb{E}[X_n^2] = \mathbb{E}[X_0^2] + \sum_{k=1}^{n} \mathbb{E}[\Delta X_n^2].$$

Indeed, $X_n = X_0 + \Delta X_1 + \cdots + \Delta X_n$, then

$$X_n^2 = X_0^2 + \Delta X_1^2 + \cdots + \Delta X_n^2 + \sum_{\substack{i \neq j \\ 1 \leq i,j \leq n}} \Delta X_i \Delta X_j + \sum_{i=1}^{n} 2X_0 \Delta X_i$$

and

$$\begin{aligned} \mathbb{E}[X_0 \Delta X_i] &= \mathbb{E}[\mathbb{E}[X_0 \Delta X_i | \mathcal{F}_{i-1}]] \\ &= \mathbb{E}[X_0 \mathbb{E}[\Delta | \mathcal{F}_{i-1}]] \\ &= 0. \end{aligned}$$

Let $i < j$, we know that

$$\begin{aligned} \mathbb{E}[\Delta X_i \Delta X_j] &= \mathbb{E}[\mathbb{E}[\Delta X_i \Delta X_j | \mathcal{F}_{j-1}]] \\ &= \mathbb{E}[\Delta X_i \mathbb{E}[\Delta X_j | \mathcal{F}_{j-1}]] \\ &= 0. \end{aligned}$$

Thus,

$$\lim_{n \to \infty} \mathbb{E}[X_n^2] = \mathbb{E}[X_0^2] + \sum_{k=1}^{\infty} \mathbb{E}[\Delta X_k^2] \leq C \leq +\infty$$

where $C := \sup_{n \geq 1} \mathbb{E}[|X_n|^2] < \infty$. Therefore, for $m > n$,

$$\begin{aligned} \mathbb{E}[(X_m - X_n)^2] &= \mathbb{E}[(\sum_{k=n+1}^{m} \Delta X_k)^2] \\ &= \mathbb{E}[\sum_{k=n+1}^{m} \Delta X_k^2] + \mathbb{E}[\sum_{\substack{i \neq j \\ n+1 \leq i,j \leq m}} \Delta X_i \Delta X_j] \\ &= \sum_{k=n+1}^{m} \mathbb{E}[\Delta X_k^2] \to 0, \text{ as } m, n \to \infty. \end{aligned}$$

Then $(X_n)_{n \geq 1}$ is a Cauchy sequence in L^2 space. From the completeness of L^2, we know by 1.9 that X_n converges to X_∞ in L^2 space, i.e. $\lim_{n \to \infty} \mathbb{E}[|X_n - X_\infty|^2] = 0$. □

Application I: Law of large number

Theorem 2.5 (Law of large number). *Let $(\xi_k)_{k \geq 1}$ be a sequence of i.i.d. random variables, such that $\mathbb{E}[|\xi_i|] < \infty$. Then*

$$\frac{1}{n} \sum_{k=1}^{n} \xi_k \longrightarrow \mathbb{E}\big[X_1\big], \ a.s.$$

In the following, we will use the theorem of convergence of martingale to prove the above theorem (law of large number).

Lemma 2.18 (Kronecker)**.** *Let $(x_n)_{n\geq 1}$ be a sequence of real numbers such that*

$$\lim_{n\to\infty}\sum_{k=1}^{n} k^{-1}x_k \text{ exists.}$$

Then

$$\frac{1}{n}\sum_{k=1}^{n} x_k \longrightarrow 0.$$

Proof. Let $m_n := \sum_{k=1}^n k^{-1}x_k$ for all $n \geq 1$, let us denote $m_\infty := \lim_{n\to\infty} m_n$. Notice that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} m_k = m_\infty, \quad \text{and} \quad \sum_{k=1}^{n} m_k = (n+1)m_n - \sum_{k=1}^{n} x_k.$$

It follows immediately that $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n x_k = 0$. □

Proof of Theorem 2.5. In view of Kronecker's Lemma, it is enough to assume in addition that $E[X_1] = 0$ and then prove that

$$\lim_{n\to\infty}\sum_{k=1}^{\infty} k^{-1}X_k, \text{ exists a.s.} \tag{8}$$

Let us define

$$M_n := \sum_{k=1}^{n} k^{-1}X_k, \quad n \geq 1.$$

Since it is assumed that $\mathbb{E}[X_k] = 0$, we observe that $(M_n)_{n\geq 1}$ is a martingale.

(i) When X_1 is square integrable, i.e. $\mathbb{E}[|X_1|^2] < \infty$, we obtain that

$$\mathbb{E}\big[|M_n|^2\big] \;=\; \sum_{k=1}^{n}\frac{1}{k^2}\mathbb{E}\big[|X_k|^2\big] \;=\; \mathbb{E}[|X_1|^2]\sum_{k=1}^{n}\frac{1}{k^2}.$$

By the theorem of convergence of martingale, it follows that there exists a square-integrable random variable M_∞ such that

$$\lim_{n\to\infty} M_n = M_\infty, \text{ a.s.}$$

and we hence conclude of the proof of (8).

(ii) When we only have $\mathbb{E}[|X_1|] < \infty$, let us define

$$Y_n := X_n \mathbf{1}_{\{|X_n|\leq n\}}, \quad n \geq 1.$$

Then

$$\sum_{n\geq 1}\mathbb{P}[X_n \neq Y_n] = \sum_{n\geq 1}\mathbb{P}[|X_1| > n] = \mathbb{E}\Big[\sum_{n\geq 1}\mathbf{1}_{\{|X_1|>n\}} \leq \mathbb{E}\big[|X_1|\big] < \infty.$$

By Borel-Cantelli, il follows that there exists a random variable M such that

$$X_n = Y_n, \text{ for all } n \geq M, \text{ a.s.}$$

Therefore, whenever the last two limits below exist, one has

$$\lim_{n\infty}\frac{1}{n}\sum_{k=1}^{n} X_k \;=\; \lim_{n\infty}\frac{1}{n}\sum_{k=1}^{n} Y_k \;=\; \lim_{n\infty}\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[Y_k] \;+\; \lim_{n\infty}\frac{1}{n}\sum_{k=1}^{n}\big(Y_k - \mathbb{E}[Y_k]\big).$$

By the definition of Y_k, we notice that $\lim_{k\to\infty}\mathbb{E}[Y_k]=\mathbb{E}[X_1]=0$, so that

$$\lim_{n\infty}\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[Y_k] \;=\; 0.$$

To study the last limit, let us define $Z_n := n^{-1}(Y_n-\mathbb{E}[Y_n])$ and claim that

$$\sum_{n=1}^{\infty}\mathbb{E}\big[|Z_n|^2\big]<\infty. \tag{9}$$

Then by the arguments in Item (i), we conclude that

$$\lim_{n\infty}\frac{1}{n}\sum_{k=1}^{n}\big(Y_k-\mathbb{E}[Y_k]\big)=0,$$

which implies the requires result in the statement.

To finish the proof, it is enough to prove the claim in (9). In fact, we notice that

$$\sum_{n\geq 1}\mathbb{E}\big[|Z_n|^2\big]=\sum_{n\geq 1}n^{-2}\mathrm{Var}[Y_n]\leq\sum_{n\geq 1}n^{-2}\mathbb{E}[|Y_n|^2]=\mathbb{E}\Big[X_1^2\sum_{n\geq 1}n^{-2}\mathbf{1}_{\{|X_1|\leq n\}}\Big]=\mathbb{E}\big[X_1^2 f(|X_1|)\big],$$

where $f(x):=\sum_{x\leq n}n^{-2}$ satisfies that, for some constant $C>0$, $f(x)\leq Cx^{-1}$ for all $x\geq 0$. Therefore,

$$\sum_{n\geq 1}\mathbb{E}\big[|Z_n|^2\big]\leq\mathbb{E}\big[X_1^2 f(|X_1|)\big]\leq C\mathbb{E}\big[|X_1|\big]<\infty,$$

which proves (9) and hence concludes the proof. □

Application II: Stochastic Gradient Algorithm Let $(X_k)_{k\geq 1}$ be a sequence of i.i.d. random variables with the same law of X. Then we give the stochastic gradient algorithm

$$\theta_{k+1} \;=\; \theta_k-\gamma_{k+1}F(\theta_k,X_{k+1}),\ \forall k\in\mathbb{N}. \tag{10}$$

where $F:\mathbb{R}^d\times\mathbb{R}\to\mathbb{R}^d$ satisfies $\mathbb{E}[F(\theta,X)]=f(\theta)$.

To make the algorithm converges, we make the following assumptions:

Assumption 2.6.

- $\gamma_k>0$, $\sum_{k=1}^{\infty}\gamma_k=+\infty$, $\sum_{k=1}^{\infty}\gamma_k^2<+\infty$
- *There exists a point* $\theta^*\in\mathbb{R}^d$ *such that*

$$\langle\theta_k-\theta^*,f(\theta_k)\rangle>0,\ \forall\ \theta_k\neq\theta^*.$$

- F *is uniformly bounded by some constant* $C>0$.

Theorem 2.7. *Given* $F:\mathbb{R}^d\times\mathbb{R}\to\mathbb{R}^d$, $f:\mathbb{R}^d\to\mathbb{R}^d$, $\theta_0\in\mathbb{R}$ *and constants* $\{\gamma_k\}_{k\geq 1}$, *we define a sequence of random variables* $\{\theta_k\}_{k\geq 1}$ *by* (10) *iteratively, then under Assumption 2.6,* $\lim_{k\to\infty}\theta_k=\theta^*$ *a.s.*

Remark 2.19. *If* $g:\mathbb{R}^d\to\mathbb{R}$ *is strictly convex,* θ^* *is the minimizer of* $g(\theta)$, *then for any* $\theta\neq\theta^*$, $\langle\theta-\theta^*,\nabla g(\theta)\rangle>0$.

Proof. Let us define the $\mathbb{F}$-predictable process $(S_n)_{n\geq 0}$ by

$$S_n \ := \ \sum_{k=0}^{n-1} \gamma_{k+1}^2 \mathbb{E}\big[|F(\theta_k, X_{k+1})|^2 \big| \mathcal{F}_k\big],$$

where $\mathcal{F}_0 := \{\phi, \Omega\}$, $\mathcal{F}_k := \sigma(X_1, \cdots, X_k)$ for any $k \geq 1$ and $\mathbb{F} := (\mathcal{F}_k)_{k\geq 0}$. Then by the uniformly boundedness of F, we have

$$S_n \ \leq \ \sum_{k=0}^{n-1} \gamma_{k+1}^2 C^2 \ \leq \ C^2 \sum_{k=0}^{\infty} \gamma_{k+1}^2.$$

Hence by the martingale convergence theorem, we know the existence of $S_\infty := \lim_{n\to\infty} S_n$ and

$$S_\infty \ = \ \sum_{k=0}^{\infty} \gamma_{k+1}^2 \mathbb{E}\big[|F(\theta_k, X_{k+1})|^2 \big| \mathcal{F}_k\big] \ \leq \ C^2 \sum_{k=0}^{\infty} \gamma_{k+1}^2 \text{ a.s.}$$

Next, we define the adapted process $(Z_n)_{n\geq 0}$ by $Z_n := |\theta_n - \theta^*|^2 - S_n$ for any $n \in \mathbb{N}$ and we claim that $(Z_n)_{n\geq 0}$ is a $\mathbb{F}$-supermartingale. First, observe that

$$\begin{aligned}
\mathbb{E}[|Z_n|] \ &\leq \ \mathbb{E}[|S_n| + 2|\theta^*|^2 + 2|\theta_n|^2] \\
&\leq \ C^2 \sum_{k=0}^{\infty} \gamma_{k+1}^2 + 2|\theta^*|^2 + 2\mathbb{E}\left[\left|\theta_0 + \sum_{k=0}^{n-1} \gamma_{k+1} F(\theta_k, X_{k+1})\right|^2\right] \\
&\leq \ C^2 \sum_{k=0}^{\infty} \gamma_{k+1}^2 + 2|\theta^*|^2 + 4|\theta_0|^2 + 4n\mathbb{E}[|S_n|] \\
&\leq \ (4n+1) C^2 \sum_{k=0}^{\infty} \gamma_{k+1}^2 + 2|\theta^*|^2 + 4|\theta_0|^2 < \infty.
\end{aligned}$$

Next, for any $n \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{E}[Z_{n+1}|\mathcal{F}_n] \ &= \mathbb{E}[|\theta_{n+1} - \theta^*|^2 - S_{n+1}|\mathcal{F}_n]] \\
&= \ - S_{n+1} + |\theta_n - \theta^*|^2 + \mathbb{E}[|\gamma_{n+1} F(\theta_n, X_{n+1})|^2 |\mathcal{F}_n] \\
&\quad\ - 2\mathbb{E}[\langle \theta_n - \theta^*, \gamma_{n+1} F(\theta_n, X_{n+1})\rangle |\mathcal{F}_n] \\
&= \ - S_{n+1} + |\theta_n - \theta^*|^2 + \mathbb{E}[|\gamma_{n+1} F(\theta_n, X_{n+1})|^2 |\mathcal{F}_n] - 2\gamma_{n+1}\langle \theta_n - \theta^*, f(\theta_n)\rangle \\
&\leq \ - S_{n+1} + |\theta_n - \theta^*|^2 + \mathbb{E}[|\gamma_{n+1} F(\theta_n, X_{n+1})|^2 |\mathcal{F}_n] \\
&= Z_n \text{ a.s.}
\end{aligned}$$

Now let $K := C^2 \sum_{k=0}^{\infty} \gamma_{k+1}^2$, we have $(Z_n + K)_{n\geq 0}$ is a positive supermaringale and

$$\sup_{n\geq 0} \mathbb{E}[|Z_n + K|] \ = \ \sup_{n\geq 0} \mathbb{E}[Z_n + K] \ \leq \ \mathbb{E}[Z_0 + K] \ < \ \infty.$$

By the martingale convergence theorem, if follows that

$$\lim_{n\to\infty} Z_n + K = Z_\infty + K, \text{ for some r.v. } Z_\infty \in L^1.$$

Then let $L := S_\infty + Z_\infty$, we know that

$$\lim_{n\to\infty} |\theta_n - \theta^*|^2 = L \text{ a.s.}$$

and we claim that $L = 0$ a.s.

Let $A_\delta := \{\omega : L(\omega) > \delta\}$, then it is sufficient to prove that $\mathbb{P}[A_\delta] = 0$ for any $\delta > 0$.

We assume by contradiction that $\mathbb{P}[A_\delta] > 0$, then $\eta := \inf_{\delta \le |\theta_k - \theta^*|^2 \le 2L} \langle \theta_k - \theta^*, f(\theta_k) \rangle > 0$ on A_δ, and we have

$$\sum_{k=0}^{\infty} \gamma_{k+1} \langle \theta_k - \theta^*, f(\theta_k) \rangle \ \ge \ \sum_{k=0}^{\infty} \gamma_{k+1} \eta \ = \ +\infty, \ \text{on } A_\delta.$$

Then the monotone convergence theorem gives that

$$\sum_{k=0}^{\infty} \mathbb{E}[\gamma_{k+1} \langle \theta_k - \theta^*, f(\theta_k) \rangle] \ = \ +\infty.$$

However, by the definition of the algorithm, we have

$$\begin{aligned}
&\sum_{k=0}^{\infty} \mathbb{E}[\gamma_{k+1} \langle \theta_k - \theta^*, f(\theta_k) \rangle] \\
=\ &\sum_{k=0}^{\infty} \mathbb{E}[\langle \theta_k - \theta^*, \gamma_{k+1} F(\theta_k, X_{k+1}) \rangle] \\
=\ &\frac{1}{2} \sum_{k=0}^{\infty} \mathbb{E}\big[|\theta_{k+1} - \theta^*|^2 - |\theta_k - \theta^*|^2 - |\gamma_{k+1} F(\theta_k, X_{k+1})|^2\big] \\
=\ &\frac{1}{2} \Big(\lim_{n\to\infty} \mathbb{E}\big[|\theta_k - \theta^*|^2\big] - \mathbb{E}\big[|\theta_0 - \theta^*|^2\big] - \sum_{k=0}^{\infty} \gamma_{k+1}^2 \mathbb{E}\big[|F(\theta_k, X_{k+1})|^2\big] \Big) \\
=\ &\frac{1}{2} \mathbb{E}[S_\infty + Z_\infty - |\theta_0 - \theta^*|^2 - S_\infty] \\
=\ &\frac{1}{2} \mathbb{E}[Z_\infty - |\theta_0 - \theta^*|^2] < \infty.
\end{aligned}$$

Now we have a contradiction and complete the proof. $\square$

3 Discrete time Markov chain

3.1 Definition and examples

Let us recall that a stochastic process $X = (X_k)_{k\geq 0}$ is a family of random variables indexed by time $k \geq 0$. In this section, we consider the case that X takes value in a countable state space S.

Remark 3.1. *The state space S could be finite, e.g. $S = \{x_1, \cdots, x_n\}$, or infinite, e.g. $S = \mathbb{N} = \{0, 1, 2, \cdots\}$.*

Definition 3.2. *A stochastic process $X = (X_n)_{n\geq 0}$ taking value in a countable space S is called a Markov chain if, for all $x_0, x_1, \cdots, x_n, x_{n+1} \in S$, one has*

$$\mathbb{P}\big[X_{n+1} = x_{n+1} \big| X_n = x_n, \cdots, X_0 = x_0\big] \;=\; \mathbb{P}\big[X_{n+1} = x_{n+1} \big| X_n = x_n\big].$$

Example 3.3 (Random walk)**.** *Let $(\xi_k)_{k\geq 1}$ be a sequence of i.i.d. random variables such that*

$$\mathbb{P}[\xi_1 = 1] = p, \quad \mathbb{P}[\xi_1 = -1] = 1 - p.$$

Let

$$X_n := \sum_{k=1}^{n} \xi_k, \qquad n \geq 0.$$

One observes that X takes value in Z, and one can compute that

$$\mathbb{E}\big[f(X_{n+1}) \big| X_n = x_n, \cdots, X_1 = x_1\big] = pf(x_n + 1) + (1 - p)f(x_n - 1),$$

and

$$\mathbb{E}\big[f(X_{n+1}) \big| X_n = x_n\big] = pf(x_n + 1) + (1 - p)f(x_n - 1).$$

Thus, $(X_n)_{n\geq 0}$ is a Markov chain.

Notice also that, when $p = \frac{1}{2}$, $(X_n)_{n\geq 0}$ is a martingale.

Proposition 3.4. *A process X is a Markov chain if and only if*

$$\mathbb{E}\big[f(X_{n+1}) \big| \mathcal{F}_n\big] = \mathbb{E}\big[f(X_{n+1}) \big| X_n\big],$$

for all bounded function $f : S \longrightarrow \mathbb{R}$, where $\mathcal{F}_n := \sigma(X_0, \cdots, X_n)$.

Proof. to be completed. □

Definition 3.5. *A Markov chain X is called homogeneous if*

$$\mathbb{P}[X_{n+1} = y | X_n = x] = \mathbb{P}[X_1 = y | X_0 = x], \quad \text{for all } n \geq 0, \; x, y \in S.$$

In the following, we will only consider homogeneous Markov chain !

Definition 3.6. *Let X be a Markov chain.*

(i) *For all $x, y \in S$, $P(x, y) := \mathbb{P}[X_{n+1} = y | X_n = x]$ is called the transition probability from x to y.*

(ii) *The matrix $P = (P(x, y))_{x,y\in S}$ is then called the transition matrix.*

(iii) *The vector $\mu = (\mu(x))_{x\in S}$ defined by $\mu(x) := \mathbb{P}[X_0 = x]$ is the initial distribution of X.*

Example 3.7. (i) *Ranom walk.*

(ii) *Gambler's ruin.*

(iii) *Ehrenfest model.*

Remark 3.8. *Let us recall that*

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} \quad \Longleftrightarrow \quad \mathbb{P}[A \cap B] = \mathbb{P}[A|B]\mathbb{P}[B].$$

Proposition 3.9 (Chapman-Kolmogorov Equation)**.** *Let X be a Markov chain with transition matrix P. Then the joint law of $(X_0, X_1, \cdots, X_n)$ is given by*

$$\mathbb{P}[X_0 = x_0, \cdots, X_n = x_n] = \mathbb{P}[X_0 = x_0] P(x_0, x_1) \cdots P(x_{n-1}, x_n).$$

Proof. to be completed. □

Lemma 3.10. *One has*

$$\mathbb{P}[X_0 = x_0, X_n = x_n] = \mathbb{P}[X_0 = x_0] P^n(x_0, x_n)$$

and

$$\mathbb{P}[X_{m+n} = y | X_0 = x] = P^{m+n}(x, y).$$

Proof. to be completed. □

3.2 Recurrence, transience

Let us consider a Markov chain $X = (X_n)_{n \geq 0}$, with state space $S = \{x_1, x_2, \cdots\}$ and transition matrix P. Let us use the notation

$$\mathbb{P}_x[A] := \mathbb{P}[A | X_0 = x].$$

Definition 3.11. *A state $x \in S$ is communicate with state $y \in S$, denoted by $x \to y$, if*

$$\mathbb{P}_x[\tau_y < \infty] = \mathbb{P}[\tau_y < \infty | X_0 = x] > 0,$$

where $\tau_y := \min\{n \geq 0 \ : X_n = y\}$.

Notice that $\tau_y < \infty$ means that $X_n = y$ for some $n \geq 0$; and $\tau_y = \infty$ means that $X_n \neq y$ for all $n \geq 0$.

Proposition 3.12. *For $x, y \in S$, one has $x \to y$ if and only if $P^n(x, y) > 0$ for some $n \geq 0$.*

Proof. (i) If $x \to y$ so that $\mathbb{P}_x[\tau_y < \infty] > 0$, then

$$0 < \mathbb{P}_x[\tau_y < \infty] = \mathbb{P}_x\big[\cup_{n \geq 0} \{\tau_y \leq n\}\big] = \lim_{n \to \infty} \mathbb{P}_x[\tau_y \leq n],$$

since $\{\tau_y \leq n\} \subset \{\tau_y \leq n+1\}$. Thus, there exists some $n \geq 0$ such that

$$\mathbb{P}_x\big[\tau_y \leq n\big] > 0.$$

Further, as $\{\tau_y \leq n\} = \cup_{k=0}^n \{\tau_y = k\}$, then for some $k \geq 0$, one has

$$\mathbb{P}_x\big[\tau_y = k\big] > 0.$$

Therefore,

$$P^k(x,y) = \mathbb{P}_x\big[X_k = y\big] \geq \mathbb{P}_x\big[X_0 = x, X_1 \neq y, \cdots, X_{k-1} \neq y, X_k = y\big] = \mathbb{P}_x[\tau_y = k] > 0.$$

(ii) Next, if $P^n(x,y) > 0$ for some $n \geq 0$, then

$$\mathbb{P}_x[\tau_y < \infty] \geq \mathbb{P}_x[\tau_y \leq n] \geq \mathbb{P}_x[X_n = y] = P^n(x,y) > 0.$$

Hence $x \to y$. □

Proposition 3.13. *Let $x, y, z \in S$, then*

- *$x \to x$;*
- *$x \to y$ and $y \to z$ implies that $x \to z$.*

Proof. (i) By its definition, one has $\tau_x := \min\{n \geq 0 \ : X_0 = x\} = 0 < \infty$, $\mathbb{P}_x$-a.s. so that $x \to x$.

(ii) If $x \to y$ and $y \to z$, then there exist $m \geq 0$ and $n \geq 0$ such that $P^m(x,y) > 0$ and $P^n(y,z) > 0$. Then $P^{m+n}(x,z) \geq P^m(x,y)P^n(y,z) > 0$, and hence $x \to z$. □

Definition 3.14. (i) *Let $x, y \in S$, we say x and y are intercommunicate, denoted by $x \leftrightarrow y$, if $x \to y$ and $y \to x$.*

(ii) *A subset $B \subset S$ is called irreducible if $x \leftrightarrow y$ for all $x, y \in B$.*

(iii) *If S itself is irreducible, we say that the Markov chain is irreducible, or the transition matrix P is irreducible.*

Example 3.15. (i) *Ranom walk.*

(ii) *Gambler's ruin.*

(iii) *Ehrenfest model.*

Let us denote by N_x the number of times that X stays at point $x \in S$, i.e.

$$N_x \ := \ \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n = x\}}.$$

Further, let

$$\tau_x^1 \ := \ \min\{n \geq 1 \ : X_n = x\}.$$

Definition 3.16. (i) *We say $x \in S$ is recurrent if $\mathbb{P}_x[\tau_x^1 < \infty] = 1$.*

(ii) *We say $x \in S$ is transiant if $\mathbb{P}_x[\tau^1 < \infty] < 1$.*

Remark 3.17. *Notice that*

$$\tau_x^1 = \infty, \ \mathbb{P}_x\text{-a.s.} \iff N_x = 1, \ \mathbb{P}_x\text{-a.s.}$$

Theorem 3.1. (i) *If x is recurrent, i.e. $\mathbb{P}_x[\tau_x^1 < \infty] = 1$. Then $\mathbb{P}_x[N_x = \infty] = 1$.*

(ii) *If x is transient, i.e. $\alpha := \mathbb{P}_x[\tau_x^1 = \infty] = 1 - \mathbb{P}_x[\tau_x^1 < \infty] > 0$. Then $\mathbb{P}_x[N_x = n] = \alpha(1-\alpha)^{n-1}$, for all $n \geq 1$. Consequently, $\mathbb{E}_x[N_x] = 1/\alpha$.*

Lemma 3.18. *Let $\tau_x^{n+1} := \min\{k \geq \tau_x^n + 1 \ : X_k = x\}$, with $\tau_x^0 \equiv 0$. Then for any $k_1 < k_2 < \cdots < k_{n+1}$, one has*

$$\mathbb{P}_x\big[\tau_x^1 = k_1, \tau_x^2 = k_2, \cdots, \tau_x^{n+1} = k_{n+1}\big] = \mathbb{P}_x\big[\tau_x^1 = k_1, \cdots, \tau_x^n = k_n\big]\mathbb{P}_x\big[\tau_x^1 = k_{n+1} - k_n\big]. \quad (11)$$

Consequently, $(\tau_x^{n+1} - \tau_x^n)_{n\geq 0}$ is an i.i.d. sequence of random variables.

Proof. We only provide the proof for the case $n = 1$, where the proof for the general case $n > 1$ is almost the same, but with more heavy notations.

$$\begin{aligned}
&\mathbb{P}_x\big[\tau_x^1 = k_1, \tau_x^2 = k_2\big]\\
=\ &\mathbb{P}_x\big[X_0 = x, X_1 \neq x, \cdots, X_{k_1-1} \neq x, X_{k_1} = x, X_{k_1+1} \neq x, \cdots, X_{k_2-1} \neq x, X_{k_2} = x\big]\\
=\ &\mathbb{P}_x\big[X_0 = x, X_1 \neq x, \cdots, X_{k_1-1} \neq x, X_{k_1} = x\big]\\
&\cdot \mathbb{P}_x\big[X_{k_1} = x, X_{k_1+1} \neq x, \cdots, X_{k_2-1} \neq x, X_{k_2} = x \big| X_0 = x, X_1 \neq x, \cdots, X_{k_1-1} \neq x, X_{k_1} = x\big]\\
=\ &\mathbb{P}_x\big[X_0 = x, X_1 \neq x, \cdots, X_{k_1-1} \neq x, X_{k_1} = x\big]\\
&\cdot \mathbb{P}_x\big[X_{k_1} = x, X_{k_1+1} \neq x, \cdots, X_{k_2-1} \neq x, X_{k_2} = x \big| X_{k_1} = x\big]\\
=\ &\mathbb{P}_x\big[\tau_x^1 = k_1\big]\mathbb{P}_x\big[\tau_x^1 = k_2 - k_1\big].
\end{aligned}$$

This proves (11) for the case $n = 1$.

Next, notice that

$$\begin{aligned}
\mathbb{P}_x\big[\tau_x^1 = k_1, \tau_x^2 = k_2\big] &= \mathbb{P}_x\big[\tau_x^1 = k_1, \tau_x^2 - \tau_x^1 = k_2 - k_1\big]\\
&= \mathbb{P}_x\big[\tau_x^1 = k_1\big]\mathbb{P}_x\big[\tau_x^2 - \tau_x^1 = k_2 - k_1 \big| \tau_x^1 = k_1\big].
\end{aligned}$$

This implies that, for all $k_1 \geq 1$,

$$\mathbb{P}_x\big[\tau_x^1 = n_1\big] = \mathbb{P}_x\big[\tau_x^2 - \tau_x^1 = n_1 \big| \tau_x^1 = k_1\big],\ \mathbb{P}_x\text{-a.s.}$$

Hence $\tau_x^2 - \tau_x^1$ is independent of τ_x^1 and has the same distribution as τ_x^1. □

Proof of Theorem 3.1. Let $\alpha := \mathbb{P}_x\big[\tau_x^1 = \infty\big]$, we claim that

$$\mathbb{P}_x\big[N_x > n\big] = \mathbb{P}_x\big[\tau_x^1 < \infty\big]^2 = (1-\alpha)^n.$$

Indeed, as $\{N_x > n\} = \{\tau_x^n < \infty\}$, one then has

$$\mathbb{P}_x\big[N_x > n\big] = \mathbb{P}_x\big[\tau_x^n < \infty\big] = \mathbb{P}_x\big[\tau_x^1 < \infty, \tau_x^2 - \tau_x^1 < \infty, \cdots, \tau_x^n - \tau_x^{n-1} < \infty\big].$$

Applying Lemma 3.18, it follows that

$$\mathbb{P}_x\big[N_x > n\big] = \mathbb{P}_x\big[\tau_x^1 < \infty\big]^2 = (1-\alpha)^n.$$

When x is recurrent, i.e. $\mathbb{P}_x[\tau_x^1 < \infty] = 1$, and hence $\alpha = 0$, one has $\mathbb{P}_x\big[N_x > n\big] = 1$ for all $n \geq 1$. Thus $\mathbb{P}_x\big[N_x = \infty\big] = 1$.

When x is transient so that $\alpha > 0$, one has

$$\mathbb{P}_x\big[N_x = n\big] = \mathbb{P}_x\big[N_x > n-1\big] - \mathbb{P}_x\big[N_x > n\big] = (1-\alpha)^{n-1} - (1-\alpha)^n = \alpha(1-\alpha)^{n-1}.$$

We hence conclude the proof. □

Proposition 3.19. *The state $x \in S$ is recurrent if and only if*

$$\sum_{n=0}^{\infty} P^n(x,x) = \infty.$$

Proof. If x is recurrent, then $\mathbb{P}_x\big[N_x = \infty\big] = 1$ and hence $\mathbb{E}_x\big[N_x\big] = \infty$. If x is transient, then $\mathbb{E}_x[N_x] = 1/\alpha$ with $\alpha := \mathbb{P}_x[\tau_x^1 = \infty] > 0$. Therefore, one has x is recurrent if and only if $\mathbb{E}_x\big[N_x\big] = \infty$.

By direct computation, one has

$$\mathbb{E}_x\big[N_x\big] = \mathbb{E}_x\Big[\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=x\}}\Big] = \sum_{n=0}^{\infty} \mathbb{E}_x\big[\mathbf{1}_{\{X_n=x\}}\big] = \sum_{n=0}^{\infty} \mathbb{P}_x\big[X_n = x\big] = \sum_{n=0}^{\infty} P^n(x,x).$$

Therefore, x is recurrent if and only if $\sum_{n=0}^{\infty} P^n(x,x) = \infty$. □

Example 3.20. *Let us consider the random walk $(X_n)_{n\geq 0}$, with $X_n := \sum_{k=1}^{n} \xi_k$, where $(\xi_k)_{k\geq 1}$ is an i.i.d. sequence of random variable such that $\mathbb{P}[\xi_1 = 1] = p$ and $\mathbb{P}[\xi_1 = -1] = 1-p$, for some $p \in [0,1]$.*

(i) *When $p = \frac{1}{2}$, one has*

$$\mathbb{P}[X_{2n} = 0] \;=\; C_{2n}^{n} 2^{-2n} \;=\; \frac{(2n)!}{n!n!} 2^{-2n}.$$

By Stirling formula: $n! \approx \sqrt{2\pi n}(\frac{n}{e})^n$, it follows that

$$\mathbb{P}[X_{2n} = 0] \approx \frac{1}{\sqrt{\pi n}}, \quad \textit{and hence} \quad \sum_{n=0}^{\infty} P^n(0,0) = \infty.$$

Therefore, X is recurrent when $p = \frac{1}{2}$.

(ii) *When $p \neq \frac{1}{2}$, we compute that*

$$\mathbb{P}_0[X_{2n} = 0] = C_{2n}^{n} p^n (1-p)^n \approx \frac{(4p(1-p))^n}{\sqrt{\pi n}} \approx \frac{1}{\sqrt{\pi}} n^{-1/2} \alpha^n,$$

where $\alpha := 4p(1-p) < 1$. Therefore, X is transient when $p \neq \frac{1}{2}$.

Definition 3.21. (i) *A set $B \subset S$ is called a class if it is irreducible and there does not exist a couple (x,y) such that $x \in B$, $y \notin B$ and $x \leftrightarrow y$.*

(ii) *A set $B \subset S$ is closed if there is no (x,y) such that $x \in B$, $y \notin B$ and $x \to y$.*

(iii) *A state $x \in S$ is absorbing if $\{x\}$ is closed.*

(iv) *Let $x \in S$, the period of x, denoted by $d(x)$, is the greatest common denominator of the return time set*

$$R(x) \;:=\; \{n \in \mathbb{N} \;: P^n(x,x) > 0\}.$$

We use the convention that $d(x) = 1$ if $R(x) = \emptyset$.

We say that the state $x \in S$ is aperiodic if $d(x) = 1$.

Proposition 3.22. *Let $x \leftrightarrow y$. Then x and y are both recurrent or both transient.*

Proof. As $x \leftrightarrow y$, there exists $k, \ell \geq 0$ such that $P^k(x,y) > 0$ and $P^\ell(y,x) > 0$, so that $\alpha := P^k(x,y)P^\ell(y,x) > 0$. Then

$$P^{k+n+\ell}(x,x) \geq P^k(x,y)P^n(y,y)P^\ell(y,x) = \alpha P^n(y,y).$$

Assume that x is transient so that $\sum_{n=0}^{\infty} P^n(x,x) < \infty$. Then

$$\sum_{n\geq 0} P^n(y,y) \leq \frac{1}{\alpha} \sum_{n\geq 0} P^{n+k+\ell}(x,x) < \infty,$$

and hence y is also transient.

If x is recurrent, then y cannot be transient. Otherwise, if y is transient then x must also be transient, which contradicts the fact that x is recurrent. Therefore, y must also be recurrent. □

Remark 3.23. *Let $x \leftrightarrow y$. By the same arguments,*

$$x \text{ and } y \text{ are transient} \iff \sum_{n=0}^{\infty} P^n(y,x) < \infty.$$

Proposition 3.24. *Let X be a Markov chain with a finite state space X. Then there exists a state $x \in S$ which is recurrent.*

Consequently, if X is in addition irreducible, then every state is recurrent.

Proof. Let us fix $y \in S$, then

$$\sum_{x\in S}\sum_{n\geq 0} P^n(y,x) = \sum_{n\geq 0}\sum_{x\in S} \mathbb{P}_y[X_n = x] = \sum_{n\geq 0} \mathbb{P}_y[X_n \in S] = \infty.$$

When S is finite, there must be some $x \in S$ such that

$$\sum_{n\geq 0} P^n(y,x) = \infty.$$

Next, let us denote

$$Q^m(y,x) := \mathbb{P}_y\big[X_0 = y, X_1 \neq x, \cdots, X_{m-1} \neq x, X_m = x\big] = \mathbb{P}_y[\tau_x^1 = m].$$

Then

$$\begin{aligned}\sum_{n\geq 0} P^n(y,x) &= \sum_{n\geq 0}\sum_{m=1}^{n} Q^m(y,x)P^{n-m}(x,x) = \sum_{m\geq 0}\sum_{n=m}^{\infty} Q^m(y,x)P^{n-m}(x,x) \\ &= \sum_{m\geq 0}\sum_{n\geq 0} Q^m(y,x)P^n(x,x) = \Big(\sum_{m\geq 0} Q^m(y,x)\Big)\Big(\sum_{n\geq 0} P^n(x,x)\Big).\end{aligned}$$

As $\sum_{n\geq 0} P^n(y,x) = \infty$ and $\sum_{m\geq 0} Q^m(y,x) \leq 1$, we must have $\sum_{n\geq 0} P^n(x,x) = \infty$. Hence x is recurrent. □

Remark 3.25. *For a class $B \subset S$, either all states in B are recurrent, or all states in B are transient.*

Proposition 3.26. *Let $B \subset S$ be a recurrent class, then B is closed.*

Proof. If B is not closed, then there exists a couple $(x, y) \in S \times S$ such that

$$x \in B,\ x \notin B,\ x \to y \text{ and } y \nrightarrow x.$$

Since $x \to y$, one has $\alpha := \mathbb{P}_x\big[\tau_y^1 = \infty\big] < 1$. Further, as $x \in B$ is recurrent, then

$$\begin{aligned}
1 &= \mathbb{P}_x\Big[\sum_{m=0}^{\infty} \mathbf{1}_{\{X_m = x\}} = \infty\Big] \\
&= \sum_{n \ge 0} \mathbb{P}_x\Big[\sum_{m=0}^{\infty} \mathbf{1}_{\{X_m = x\}} = \infty \Big| \tau_y^1 = n\Big] \mathbb{P}_x\big[\tau_y^1 = n\big] \\
&\quad + \mathbb{P}_x\Big[\sum_{m=0}^{\infty} \mathbf{1}_{\{X_m = x\}} = \infty \Big| \tau_y^1 = \infty\Big] \mathbb{P}_x\big[\tau_y^1 = \infty\big] \\
&= 0 + \alpha < 1.
\end{aligned}$$

In above, we use the computation that

$$\mathbb{P}_x\Big[\sum_{m=0}^{\infty} \mathbf{1}_{\{X_m = x\}} = \infty \Big| \tau_y^1 = n\Big] = \mathbb{P}_y\Big[\sum_{m \ge n} \mathbf{1}_{\{X_m = x\}} = \infty\Big] = 0,$$

as $y \nrightarrow x$. We notice that $1 < 1$ is a contradiction, hence B must be closed. □

Proposition 3.27. *Let $x \leftrightarrow y$, then $d(x) = d(y)$.*

Proof. Since $x \leftrightarrow y$, and hence there exists $m, n > 0$ such that

$$P^m(x, y) > 0, \quad P^n(y, x) > 0.$$

In particular, one has $P^{m+n}(x, x) > 0$ and hence $m + n \in R(x)$.

If $k \in R(y)$, then $P^k(y, y) > 0$, and hence

$$P^{m+n+k}(x, x) \ge P^m(x, y) P^k(y, y) P^n(y, x) > 0.$$

Therefore, $m + n + k \in R(x)$. This implies that

$$\frac{m+n}{d(x)} \in \mathbb{Z}, \quad \text{and} \quad \frac{m+n+k}{d(x)} \in \mathbb{Z} \quad \text{and hence} \quad \frac{k}{d(x)} \in \mathbb{Z}.$$

In particular, $d(x)$ divides k for all $k \in R(y)$, and hence $d(x) \le d(y)$.

Similarly, one has $d(y) \le d(x)$ and hence one must have $d(x) = d(y)$.

□

3.3 Stationary measure

Definition 3.28. (i) *We say $\mu = (\mu(x))_{x \in S}$ is a measure on S if $\mu(x) \ge 0$ for all $x \in S$. A measure μ is a distribution on S if $\sum_{x \in S} \mu(x) = 1$.*

(ii) *A measure μ on S is called a stationary measure if*

$$\mu P = \mu, \ \textit{i.e.} \ \sum_{x \in S} \mu(x) P(x, y) = \mu(y), \ \textit{for all } y \in S.$$

Remark 3.29. *Let μ a stationary distribution and $X_0 \sim \mu$. Then one can deduce that $X_1 \sim \mu$, $\cdots$, $X_n \sim \mu$.*

Example 3.30. (i) *$P = I_n$, then every distribution is a stationary distribution.*

(ii) *Let*

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Then $\mu = (\frac{1}{2}, \frac{1}{2})$ is a stationary distribution.

(iii) *Let*

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then both $(\frac{1}{2}, \frac{1}{2}, 0)$ and $(0, 0, 1)$ are stationary distributions.

Lemma 3.31. *Let X be an irreducible Markov chain and μ be a stationary measure. Assume that there exists $x \in S$ such that $\mu(x) \in (0, \infty)$. Then $\mu(y) \in (0, \infty)$ for all $y \in S$.*

Proof. Since the Markov chain is irreducible, then for any $y \in S$, there exists $m, n \geq 1$ such that $P^m(x, y) > 0$ and $P^n(y, x) > 0$. Therefore, when $\mu(x) > 0$, one has

$$\mu(y) = \mu P^m(y) = \sum_{z \in S} \mu(z) P^m(z, y) \geq \mu(x) P^m(x, y) > 0.$$

Similarly, when $\mu(x) < \infty$, one has

$$\infty > \mu(x) = \mu P^n(x) \geq \mu(y) P^n(y, x) \quad \Longrightarrow \quad \mu(y) < \infty.$$

This concludes the proof. □

Lemma 3.32. *Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be an affine function, i.e. $f(\lambda_1 x_1 + \cdots + \lambda_m x_m) = \lambda_1 f(x_1) + \cdots + \lambda_m f(x_m)$ for all $\lambda_1, \cdots, \lambda_m \geq 0$ such that $\sum_k \lambda_k = 1$. Let $K \subset \mathbb{R}^n$ be a convex compacty set such that $f(K) \subset K$. Then there exists a fixed point $x \in K$ of f, i.e. $f(x) = x$.*

Proof. Let us take a arbitrary point $x_1 \in K$, and defines $(x_n)_{n \geq 1}$ as follows:

$$x_n := \frac{1}{n} \sum_{k=0}^{n-1} f^{(k)}(x_1), \quad \text{where} \;\; f^{(k)} = f \circ \cdots \circ f \text{ with } k \text{ times composition.}$$

Notice that $f(K) \subset K$ and K is convex, one has $x_n \in K$.

Further, as f is affine, one has

$$f(x_n) \;=\; f\Big(\frac{1}{n} \sum_{k=0}^{n-1} f^{(k)}(x_1)\Big) \;=\; \frac{1}{n} \sum_{k=0}^{n-1} f^{(k+1)}(x_1) \;=\; x_n + \frac{1}{n}\big(f^{(n)}(x_1) - x_1\big).$$

Hence

$$\big|f(x_n) - x_n\big| \;\longrightarrow\; 0, \text{ as } n \longrightarrow \infty.$$

Moreover, as K is compact, along a possible subsequence $(n_k)_{k \geq 1}$, one has $x_{n_k} \to x_\infty \in K$ so that $f(x_{n_k}) \to f(x_\infty)$ by continuity of f. Therefore, one must have $f(x_\infty) = x_\infty$. □

Theorem 3.2. *Let X be a Markov chain with a finite state space S. Then there exists a stationary distribution.*

Assume in addition that X is irreducible, then there exists a unique stationary distribution.

Proof. (i) Assume that $S = \{1, 2, \cdots, n\}$ so that we denote a distribution by $\mu = (\mu(1), \cdots, \mu(n))$. Then the space of all distribution

$$K \ := \ \{x \in \mathbb{R}^n \ : x_k \geq 0, \ \forall k, \ \text{and} \ \sum_{k=1}^n x_k = 1\}$$

is a compact and convex subset of $\mathbb{R}^n$. Further $f : K \longrightarrow K$ defined by

$$f(\mu) \ := \ \mu P$$

is clearly an affine function. Then we can apply Lemma 3.32 to find a stationary distribution.

(ii) Assume in addition that X is irreducible, and μ and π be two stationary distribution. Then by Lemma 3.31, one has $\mu(i) > 0$ and $\pi(i) > 0$ for all $i \in S$. Let $k \in S$ be such that

$$\frac{\mu(k)}{\pi(k)} \ = \ \min_{i \in S} \frac{\mu(i)}{\pi(i)},$$

so that

$$\mu(i) \ \geq \ \frac{\mu(k)}{\pi(k)} \pi(i), \quad \text{for all } i \in S.$$

Then

$$\mu(k) = \big(\mu P\big)(k) = \sum_{i \in S} \mu(i) P(i,k) \geq \sum_{i \in S} \frac{\mu(k)}{\pi(k)} \pi(i) P(i,k) = \frac{\mu(k)}{\pi(k)} \big(\pi P\big)(k) = \mu(k).$$

This implies that the inequality "$\geq$" in above should be an equality, so that

$$\mu(i) \ = \ \frac{\mu(k)}{\pi(k)} \pi(i), \quad \text{for all } i \in S.$$

Equivalently,

$$\frac{\mu(i)}{\pi(i)} = \frac{\mu(k)}{\pi(k)}, \quad \text{for all } i \in S.$$

Notice that both μ and π are distributions, hence their total mass are both 1. Then $\mu = \pi$. □

Theorem 3.3. *Let X be a Markov chain, recall that $\tau_x^1 := \inf\{n \geq 1 \ : X_n = x\}$. Let $x \in S$ be a fixed recurrent state, we define*

$$\mu_x(y) := \mathbb{E}_x \Big[\sum_{n=0}^{\tau_x^1 - 1} \mathbf{1}_{\{X_n = y\}} \Big], \quad \textit{for each } y \in S.$$

Then μ_x is a stationary measure such that $\mu_x(x) = 1$ and $\mu_x(y) \in (0, \infty)$ for all $y \in S$.

Proof. (i) Since the fixed state $x \in S$ is recurrent, one has $\tau_x^1 < \infty$, $\mathbb{P}_x$-a.s. Then

$$\sum_{n=0}^{\tau_x^1 - 1} \mathbf{1}_{\{X_n = y\}} \ = \ \sum_{n=1}^{\tau_x^1} \mathbf{1}_{\{X_n = y\}} + \mathbf{1}_{\{X_0 = y\}} - \mathbf{1}_{\{X_{\tau_x^1} = y\}}.$$

Notice that $X_0 = x$ and $X_{\tau_x^1} = x$, $\mathbb{P}_x$-a.s. Then

$$\begin{aligned}\mu_x(y) = \ & \mathbb{E}_x\Big[\sum_{n=0}^{\tau_x^1-1}\mathbf{1}_{\{X_n=y\}}\Big] = \mathbb{E}_x\Big[\sum_{n=1}^{\tau_x^1}\mathbf{1}_{\{X_n=y\}}\Big] \\ = \ & \mathbb{E}_x\Big[\sum_{n=1}^{\infty}\mathbf{1}_{\{X_n=y\}}\ \mathbf{1}_{\{n\le\tau_x^1\}}\Big] = \sum_{n=1}^{\infty}\mathbb{E}_x\Big[\mathbf{1}_{\{X_n=y\}}\ \mathbf{1}_{\{n\le\tau_x^1\}}\Big].\end{aligned}$$

Next, notice that $\{n \le \tau_x^1\} = \{\tau_x^1 \le n-1\}^c \in \mathcal{F}_{n-1}^X$, it follows that

$$\mathbb{E}_x\Big[\mathbf{1}_{\{X_n=y\}}\ \mathbf{1}_{\{n\le\tau_x^1\}}\Big] \ = \ \mathbb{E}_x\Big[\mathbf{1}_{\{n\le\tau_x^1\}}\mathbb{E}_x\Big[\mathbf{1}_{\{X_n=y\}}\Big|\mathcal{F}_{n-1}^X\Big]\Big] \ = \ \mathbb{E}_x\Big[\mathbf{1}_{\{n\le\tau_x^1\}}P(X_{n-1},y)\Big].$$

Therefore,

$$\begin{aligned}\mu_x(y) = \ & \sum_{n=1}^{\infty}\mathbb{E}_x\Big[\mathbf{1}_{\{X_n=y\}}\ P(X_{n-1},y)\Big] = \mathbb{E}_x\Big[\sum_{n=1}^{\tau_x^1}P(X_{n-1},y)\Big] \\ = \ & \mathbb{E}_x\Big[\sum_{n=0}^{\tau_x^1-1}P(X_n,y)\Big] = \mathbb{E}_x\Big[\sum_{n=0}^{\tau_x^1-1}\sum_{z\in S}P(z,y)\mathbf{1}_{\{X_n=z\}}\Big] \\ = \ & \sum_{z\in S}\mathbb{E}_x\Big[\sum_{n=0}^{\tau_x^1-1}\mathbf{1}_{\{X_n=z\}}\Big]P(z,y) = \sum_{z\in S}\mu_x(z)P(z,y).\end{aligned}$$

This proves that μ_x is a stationary measure.

Finally, notice that $X_0 = x$, $X_n \neq x$ for all $n = 1, \cdots, \tau_x^1 - 1$. Then $\mu_x(x) = 1$ by its definition. We can then use Lemma 3.31 to conclude that $\mu_x(y) \in (0,\infty)$ for all $y \in S$. □

Remark 3.33. *Notice that μ_x is only a stationary measure, but not a stationary distribution, in Theorem 3.3.*

Proposition 3.34. *Let X be a recurrent and irreducible Markov chain. Let us fix $x \in S$ so that μ_x defined in Theorem 3.3 is a stationary measure. Let ν be another stationary measure such that $\nu(y) \in (0,\infty)$ for all $y \in S$. Then there exists a constant $C > 0$ such that $\nu(y) = C\mu_x(y)$ for all $y \in S$.*

Proof. First, let us recall that, for $y \neq x$,

$$\mu_x(y) \ := \ \mathbb{E}_x\Big[\sum_{n=0}^{\tau_x^1-1}\mathbf{1}_{\{X_n=y\}}\Big] \ = \ \sum_{n=1}^{\infty}\mathbb{E}_x\Big[\mathbf{1}_{\{X_n=y;\ n<\tau_x^1\}}\Big] \ = \ \sum_{n=1}^{\infty}\mathbb{P}_x\big[X_n=y;\ n<\tau_x^1\big].$$

Next, multiplying $\nu(y)$ by the same constant $C > 0$ for all y, one obtains again a stationary measure. One can then assume without loss of generality that

$$\nu(x) = \mu_x(x) = 1.$$

We next claim that, for all $y \neq x$ and all $N \ge 1$,

$$\nu(y) \ \ge \ \sum_{n=1}^{N}\mathbb{P}_x\big[X_n=y;\ n<\tau_x^1\big]. \qquad (12)$$

Taking $N \longrightarrow \infty$, it follows that

$$\nu(y) \ \ge \ \sum_{n=1}^{\infty} \mathbb{P}_x\big[X_n = y;\ n < \tau_x^1\big] = \mu_x(y), \quad \text{for all } y \neq x.$$

Therefore, one has

$$\frac{\nu(x)}{\mu_x(x)} \ = \ 1 \ \le \ \min_{y\in S} \frac{\nu(y)}{\mu_x(y)}.$$

One can then conclude by exactly the same arguments as in Part (ii) in the proof of Theorem 3.2 to conclude that

$$\nu(y) = \mu_x(y), \quad \text{for all } y \in S.$$

To conclude, it is then enough to prove the claim in (12). First, it holds true for $N = 1$ since for $y \neq x$,

$$\nu(y) \ = \ (\nu P)(y) \ge \nu(x)P(x,y) \ = \ P(x,y) \ = \ \mathbb{P}_x\big[X_1 = y;\ 2 < \tau_x^1\big].$$

Next, assume that (12) holds true for $N \ge 1$, i.e.

$$\nu(y) \ \ge \ \sum_{n=1}^{N} \mathbb{P}_x\big[X_n = y;\ n < \tau_x^1\big],$$

we then consider the case $N+1$. Recall that ν is a stationary measure such that $\nu(x) = 1$, then for $y \neq x$,

$$\nu(y) = \sum_{z\in S} \nu(z)P(z,y) = P(x,y) + \sum_{z\neq x} \nu(z)P(z,y) \ge P(x,y) + \sum_{n=1}^{N}\sum_{z\neq x} \mathbb{P}_x\big[X_n = z;\ n < \tau_x^1\big]P(z,y).$$

By direct computation,

$$\begin{aligned}
\sum_{z\neq x} \mathbb{P}_x\big[X_n = z;\ n < \tau_x^1\big]P(z,y) = \ & \sum_{z\neq x} \mathbb{P}_x\big[X_1 \neq x, \cdots, X_{n-1} \neq x, X_n = z\big]P(z,y) \\
= \ & \mathbb{P}_x\big[X_1 \neq x, \cdots, X_{n-1} \neq x, X_n \neq x, X_{n+1} = y\big] \\
= \ & \mathbb{P}_x\big[X_{n+1} = y;\ n+1 < \tau_x^1\big].
\end{aligned}$$

Therefore,

$$\nu(y) \ge P(x,y) + \sum_{n=1}^{N} \mathbb{P}_x\big[X_{n+1} = y;\ n+1 < \tau_x^1\big] = \sum_{n=1}^{N+1} \mathbb{P}_x\big[X_n = y;\ n < \tau_x^1\big],$$

i.e. (12) holds true for the case $N+1$. We can then finish the proof of claim (12) for all $N \ge 1$ by induction, which concludes the proof of the proposition. □

Proposition 3.35. *Let X be a recurrent and irreducible Markov chain. Assume that $\mathbb{E}_x[\tau_x^1] < \infty$ for some $x \in S$. Then $\mathbb{E}_y[\tau_y^1] < \infty$ for all $y \in S$. Moreover,*

$$\pi(y) := \frac{1}{\mathbb{E}_y[\tau_y^1]}, \ y \in S, \ \textit{defines the unique stationary distribution.}$$

Proof. (i) Given the fixed state $x \in S$ such that $\mathbb{E}_x[\tau_x^1] < \infty$, we recall that μ_x defined in Theorem 3.3 is a stationary measure. In particular, one has $\mu_x(x) = 1$ and $\mu_x(y) \in (0, \infty)$ for all $y \in S$.

Further, by direct computation

$$\sum_{y \in S} \mu_x(y) \;=\; \sum_{y \in S} \mathbb{E}_x\Big[\sum_{n=0}^{\tau_x^1 - 1} \mathbf{1}_{\{X_n = y\}}\Big] \;=\; \mathbb{E}_x\Big[\sum_{n=0}^{\tau_x^1 - 1} \sum_{y \in S} \mathbf{1}_{\{X_n = y\}}\Big] \;=\; \mathbb{E}_x\big[\tau_x^1\big] < \infty.$$

Then by renormalization, $\pi_x(y) := \frac{\mu_x(y)}{\mathbb{E}_x\left[\tau_x^1\right]}$ for all $y \in S$ defines a stationary distribution $\pi_x = (\pi_x(y))_{y \in S}$. In particular, one has

$$\pi_x(x) = \frac{1}{\mathbb{E}_x[\tau_x^1]}.$$

(ii) Let us consider an arbitrary $z \in S$, which is also recurrent, so that one obtains a stationary measure $\mu_z = (\mu_z(y))_{y \in S}$. By Proposition 3.34, there exists a constant $C > 0$ such that $\mu_z(y) = C\mu_x(y)$ for all $y \in S$. Therefore, one has

$$\mathbb{E}_z\big[\tau_z^1\big] = \sum_{y \in S} \mu_z(y) = C \sum_{y \in S} \mu_x(y) = C\mathbb{E}_x\big[\tau_x^1\big] < \infty.$$

One can then obtain a stationary measure π_z defined by $\pi_z(y) := \frac{\mu_z(y)}{\mathbb{E}_z\left[\tau_z^1\right]}$ for all $y \in S$. Similarly, one has

$$\pi_z(z) = \frac{1}{\mathbb{E}_z[\tau_z^1]}.$$

Finally, in view of Proposition 3.34, there exists at most one stationary distribution. Therefore, $\pi_x = \pi_z$ for all $z \in S$, which concludes the proof. □

Example 3.36. (i) *Random walk on Z.*

(ii) *Random walk on graph.*

(iii) *Ehrenfest model.*