Probability Theory 2 (Wahrscheinlichkeitstheorie II)

Winter term 2019-20

Chiranjib Mukherjee
Institute of Mathematical Stochastics
Department of Mathematics and Computer Sciences
University of Münster

Chiranjib Mukherjee

Chapter 1

Martingales

1.1 Stochastic process.

A probability space consists of a triplet $(\Omega, \mathcal{F}, \mathbb{P})$ consisting of a set Ω, a σ-algebra $\mathcal{F}$ and a probability measure $\mathbb{P} : \mathcal{F} \to [0, 1]$.

Definition 1.1.1. *A (real-valued) random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a measurable function $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, i.e., for any Borel set $B \in \mathcal{B}(\mathbb{R})$, $X^{-1}(B) \in \mathcal{F}$.*

Remark 1 Although the same definition continues to hold when $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is replaced by another measure space, we will mostly content ourselves with real-valued random variables, unless otherwise specified.

Example 1.1.2. *A typical example of a probability space consists the outcomes of a sequence of independent coin tosses, which leads to the set up $\Omega = \{H, T\}^{\mathbb{N}}$, $\mathcal{F} = 2^{\Omega}$ and $\mathbb{P}[H] = \mathbb{P}[T] = 1/2$. A typical example of a random variable X is $X = \#\{\text{heads out of 10 tosses}\}$.*

Definition 1.1.3. *Let T be any non-empty set, typically denoting the index set of a stochastic process $\mathbb{X} = (X_t)_{t \in T}$, which is just a collection of random variables on some probability space. For any given $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is called the sample path of the stochastic process $\mathbb{X}$.*

Remark 2 Typically, we will work with $T = \mathbb{Z}$ or $T = \mathbb{N}$, or $T = \mathbb{N}_0$ or $T = [a, b]$ for $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$ and in these cases T can be interpreted as *time*. At times, we will also work with $T = A \subset \mathbb{R}^d$, whence T can be though of as a spatial location too (e.g. X_t could stand for the air temperature at location t on a given space).

1.2 Filtration and adaptability.

Definition 1.2.1. *An increasing family $\{\mathcal{F}_n\}_{n \in \mathbb{N}_0}$ of σ-algebras $\mathcal{F}_0 \subset \mathcal{F}_1 \subset F_2 \subset \ldots$ on a given probability space is called a filtration. A probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n \in \mathbb{N}_0}, \mathbb{P})$ equipped with a filtration is called a filtered probability space.*

The σ-algebra $\mathcal{F}_n$ in the above definition can be thought of as the "information gathered up until time n" (i.e., if $A \in \mathcal{F}_n$, then by time n we ought to know the (non)-occurrence of A).

Example 1.2.2. 07.10.2019 $= 280^{th}$*-day of the calendar year* 2019. *Then*

- *The event* $A = \big\{$*The average price of beer in August 2019 is below 50 cents per bottle*$\big\} \in \mathcal{F}_{280}$.
- *The event* $A = \big\{$*The average price of beer in 2019 is below 50 cents per bottle*$\big\} \notin \mathcal{F}_{280}$, *as we would not know the average price through out the year already in October.*

Definition 1.2.3. *A stochastic process* $(X_n)_{n\in\mathbb{N}_0}$ *is called adapted to the filtration* $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$ *if for each* $n \in \mathbb{N}_0$, X_n *is* $\mathcal{F}_n$*-measurable. Moreover, if* $\mathbb{X} = (X_n)_{n\in\mathbb{N}_0}$ *is a stochastic process, then with* $\mathcal{F}_n = \mathcal{F}_n^{(\mathbb{X})} = \sigma(X_0, X_1, \dots, X_n)$ *(i.e.* $\mathcal{F}_n$ *is the* σ*-algebra generated by* $(X_0, X_1, \dots, X_n)$*),* $\{\mathcal{F}_n\}_{n\in\mathbb{N}_0}$ *is called the natural or the canonical filtration of* $\mathbb{X}$.

1.3 Martingales in discrete time.

1.3.1 Definition and some examples.

Definition 1.3.1. *A stochastic process* $(X_n)_{n\in\mathbb{N}_0}$ *on a filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\in\mathbb{N}_0}, \mathbb{P})$ *is called a martingale if the following properties are satisfied for each* $n \in \mathbb{N}_0$*:*

- X_n *is* $\mathcal{F}_n$*-measurable.*
- $X_n \in L^1(\mathbb{P})$ *(i.e.,* $\int |X_n| \mathrm{d}\mathbb{P} < \infty$*).*
- $\mathbb{E}[X_{n+1}|\mathcal{F}_n] = X_n$ *almost surely.*

Remark 3 (i) Note that in the Definition 1.3.1, the third property is equivalent to $\mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n] = 0$ almost surely.

(ii) Note that if $(X_n)_{n\in\mathbb{N}_0}$ is a martingale, then $\mathbb{E}[X_n] = \mathbb{E}[X_0]$ for all n.

Example 1.3.2. *(i) (Simple random walk.) Let* $\{\xi_j\|_j$ *be an independent and identically distributed random variables with* $\mathbb{E}(\xi_j) = 0$ *for all* j*. Then with* $S_0 = 0$*, the partial sums* $S_n = \xi_1 + \cdots + \xi_n$ *is a martingale w.r.t. the natural filtration. Note taht thsi property holds irrespective of the distribution of* ξ_j*, but it is imperative that the increments are mean-zero for the martingale property to hold.*

(ii) (Geometric random walk.) Let $\{\xi_j\}_j$ *be an independent and identically distributed random variables with* $\mathbb{E}(\xi_j) = 1$ *for all* j*. Then* $M_n = \prod_{j=1}^n \xi_j$ *is a martingale w.r.t. the natural filtration. Again thsi property holds true regardless of the underlying distribution of* ξ_j*, but the martingale property relies on the normalization* $\mathbb{E}(\xi_j) = 1$ *for all* j*.*

(iii) (Cameron-Martin-Girsanov transformation.) Let $\{Z_j\}_j$ be a sequence of i.i.d. Gaussian random variables such that $Z_j \sim N(0,1)$ (that is, for any j, $\mathbb{P}[Z_j \in A] = \frac{1}{\sqrt{2\pi}} \int_A \mathrm{e}^{-\frac{|z|^2}{2}} \mathrm{d}z$). Then for any sequence of real number $(\beta_j)_j \subset \mathbb{R}$, we set

$$M_n \stackrel{\text{def}}{=} \exp\Big\{ \sum_{j=!}^n \beta_j Z_j - \frac{1}{2} \sum_{j=1}^n \beta_j^2 \Big\}.$$

It follows that $(M_n)_n$ is a martingale w.r.t. the canonical filtration. Indeed if we set $\xi_j = \exp[\beta_j Z_j - \frac{1}{2}\beta_j^2]$, then $\{\xi_j\}_j s$ are i.i.d. random variables with $\mathbb{E}[\xi_j] = 1$ (this is where we crucially use the Gaussian property that dictates if $Z \in N(0,1)$, then $\mathbb{E}[\mathrm{e}^{\beta Z}] = \mathrm{e}^{\beta^2/2}$). The martingale property for $(M_n)_n$ follows by the second part (geometric random walk example).

Definition 1.3.3. *A stochastic process $(X_n)_{n\in\mathbb{N}}$ on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\in\mathbb{N}_0}, \mathbb{P})$ is called a sub (resp. super)- martingale if the following properties are satisfied for each $n \in \mathbb{N}_0$:*

- *X_n is $\mathcal{F}_n$-measurable.*
- *$X_n \in L^1(\mathbb{P})$ (i.e., $\int |X_n| \mathrm{d}\mathbb{P} < \infty$).*
- $$\mathbb{E}[X_{n+1}|\mathcal{F}_n] \begin{cases} \geq X_n & \text{almost surely for sub-martingale property,} \\ \leq X_n & \text{almost surely, for super-martingale property.} \end{cases}$$

Lemma 1.3.4. *Let $(X_n)_{n\in\mathbb{N}_0}$ be a sub-martingale and $f : \mathbb{R} \to \mathbb{R}$ is such that $f(X_n) \in L^1$ for all $n \in \mathbb{N}$. Then $(f(X_n))_{n\in\mathbb{N}_0}$ is a sub-martingale if f is convex. Likewise, if $(X_n)_{n\in\mathbb{N}}$ is a super-martingale and $f : \mathbb{R} \to \mathbb{R}$ is such that $f(X_n) \in L^1$ for all $n \in \mathbb{N}$. Then $(f(X_n))_{n\in\mathbb{N}}$ is a super-martingale if f is concave.*

Proof. The proof is an immediate consequence of conditional Jensen's inequality, see Corollary 2.15.2 [Part (v)] □

1.3.2 Previsible processes.

Let us think of a gambling strategy where the increments of a stochastic process $(X_n)_{n\in\mathbb{N}_0}$ stands for the win (resp. loss) per unit stake of game. There is another stochastic process $(C_n)_{n\in\mathbb{N}}$ (starting at time 1) such that each C_n should be thought of as a particular "strategy" or the "bet" which needs to be decided upon in the time interval between game $n-1$ and n (and strictly before the n-th game). In other words, the value of C_n should be determined by the experience (or the information) that we would have gathered by the $n-1$-th game. In that case, it is conceivable that each C_n is $\mathcal{F}_{n-1}$-measurable and such a process $(C_n)_{n\in\mathbb{N}}$ is called *pre-visible.* Then $C_n(X_n - X_{n-1})$ defines the win in the n-th game, while

$$Y_n = \sum_{j=1}^n C_j (X_j - X_{j-1}) \tag{1.3.1}$$

defines the cumulative win process. Note that, $Y_n - Y_{n-1} = C_n(X_n - X_{n-1})$ for each $n \geq 1$.

Definition 1.3.5. *A stochastic process $(C_n)_{n\in\mathbb{N}}$ is called previsible if for each $n \geq 1$, C_n is $\mathcal{F}_{n-1}$-measurable.*

Lemma 1.3.6. *Let $(C_n)_{n\geq 1}$ be a previsible process such that for each $n \geq 1$, there is a constant K_n such that $|C_n(\cdot)| \leq K_n$ (i.e. each C_n is a bounded random variable). If $(X_n)_{n\in\mathbb{N}_0}$ is a martingale, then so is the process $(Y_n)_{nin\mathbb{N}}$ defined in* (1.3.1).

Proof. By definition Y_n depends on $C_1, \dots, C_n$ as well as $X_1, \dots, X_n$. Since C_n is $\mathcal{F}_{n-1}$ measurable and X_n is $\mathcal{F}_n$ measurable, it follows that Y_n is also $\mathcal{F}_n$ measurable. To show that $Y_n \in L^1$, note that

$$\mathbb{E}(|Y_n|) = \mathbb{E}[|\sum_{j=1}^{n} C_j(X_j - X_{j-1})|] \leq \mathbb{E}[\sum_{j=1}^{n} |C_j||X_j - X_{j-1}|] \leq \sum_{j=!}^{n} K_j \mathbb{E}[|X_j| + |X_{j-1}|] < \infty, \quad (1.3.2)$$

where the last estimate follows because $X_j \in L^1$ for each j. Hence, $Y_n \in L^1$ for each $n \geq 1$.

Finally, the martingale property of $(Y_n)_{n\in\mathbb{N}}$ follows from that of $(X_n)_{n\geq 0}$ because, using previsibility of $(C_n)_n$,

$$\mathbb{E}[Y_n - Y_{n-1}|\mathcal{F}_{n-1}] = \mathbb{E}[C_n(X_n - X_{n-1})|\mathcal{F}_{n-1}] = C_n \mathbb{E}[X_n - X_{n-1}|\mathcal{F}_{n-1}] = 0.$$

□

Lemma 1.3.7. *Let $(C_n)_{n\geq 1}$ be a non-negative and previsible process such that for each $n \geq 1$, C_n is a bounded random variable. If $(X_n)_{n\in\mathbb{N}_0}$ is a sub/super-martingale, then so is the process $(Y_n)_{nin\mathbb{N}}$ defined in* (1.3.1).

Proof. The proof is identical to that of Lemma 1.3.6 and is omitted. □

Remark 4 It follows (e.g. by Hölder's inequality) from (1.3.2) that the boundedness assumption on C_n in Lemma 1.3.6 and Lemma 1.3.7 can be relaxed by requiring that for each $n \in \mathbb{N}$, $C_n \in L^p$ and $X_n \in L^q$ where for any $p, q \geq 1$, $\frac{1}{p} + \frac{1}{q} = 1$. We will need this fact later for $p = q = 2$.

1.4 Doob's (almost sure) martingale convergence theorem.

The main goal of this section is to prove that any non-negative (super)-martingale converges almost surely to a random variable which is finite almost surely. Throughout this section $X = (X_n)_{n\in\mathbb{N}_0}$ stands for a process such that we think of

$$X_n - X_{n-1} = \text{ winnings per unit stake on game } n$$
$$Y_n = \sum_{j=1}^{n} C_j(X_j - X_{j-1}) \text{ stands for the total winning by time } .n$$

and C_n is $\mathcal{F}_{n-1}$ measurable.

We follow the following procedure and think of a gambling:

- Pick two real numbers $a < b$.
- Wait till the value of X gets below a.
- Start playing and play until the value of X gets above b and stop playing.
- While playing put black blobs for $C = 1$ and while waiting put open circles for $C = 0$. Phrased differently, define recursively,

$$\begin{aligned} &C_1 = 0 \\ &\dots \\ &\dots \\ &C_n = \mathbb{1}\{C_{n-1} = 1\mathbb{1}\{X_{n-1} \le b\} + \mathbb{1}\{C_{n-1} = 0\}\mathbb{1}|X_{n-1} < a\}. \end{aligned} \tag{1.4.1}$$

Definition 1.4.1. *Fix $N \in \mathbb{N}$. Then*

$$U_N = U_N^{(X.(\omega))}([a,b]) = \#\Big\{\text{upcrossings of } [a,b] \text{ made by } n \mapsto X_n(\omega) \text{ bytime } N\Big\} \tag{1.4.2}$$

In other words, U_N denotes the largest $k \in \mathbb{N}_0$ such that we can find a sequence $0 \le s_1 < t_1 < s_2 < t_2 < \dots < s_k < t_k \le N$ with $X_{s_j} < a$ and $X_{t_j} > b$ for $1 \le j \le k$.

Lemma 1.4.2. *The monotonic limit $U_\infty :=\uparrow U_N$ always exists as $N \to \infty$. Moreover, for any sequence of measurable functions $Z = \{Z_n\}_n$,*

$$\Big\{\omega\colon \liminf_{N\to\infty} U_N^{(Z.(\omega))}([a,b]) < a < b < \limsup_{N\to\infty} U_N^{(Z.(\omega))}([a,b])\Big\} \subset \Big\{\omega\colon U_\infty^{(Z.(\omega))}([a,b]) = \infty\Big\}$$

Proof. Note that for each fixed sample ω, $\big\{U_{(Z.(\omega))}([a,b])\big\}_N$ is an increasing sequence, implying the first assertion. The second assertion follows obviously from the definition of U_N. □

Lemma 1.4.3. *In the above set up,*

$$Y_N \ge (b-a)U_N - (X_N - a)^-.$$

Proof. Note that every upcrossing of $[a,b]$ increases the value of the process Y by at least $b-a$, while during the last interval of play $(X_N - a)^-$ overemphasizes the loss. □

Lemma 1.4.4. *Let X be a supermartingale. Then in the above set up,*

$$(b-a)\mathbb{E}(U_N) \le \mathbb{E}[(X_N - a)^-]$$

Proof. Note that by definition (1.4.1), C is non-negative, $|C_n| \le 2$ and C_n is previsible. By Lemma 1.3.7, $(Y_n)_{n\ge1}$ is a supermartingale. In particular, $\mathbb{E}(Y_N) \le \mathbb{E}(Y_1) = 0$. Then by Lemma 1.4.3 $(b-a)\mathbb{E}(U_N) \le \mathbb{E}(Y_n) + \mathbb{E}[(X_N - a)^-] \le \mathbb{E}[(X_N - a)^-]$. □

Definition 1.4.5. *Let $p > 0$. A sequence of measurable functions $\{Z_n\}_n$ on any measure space $(\Omega, \mathcal{F}, \mathbb{P})$ is called $L^p(\mathbb{P})$-bounded if $\sup_n \int |Z_n|^p \mathrm{d}\mathbb{P} < \infty$.*

Corollary 1.4.6. *Let X be a supermartingale which is bounded in $L^1(\mathbb{P})$. Then $(b-a)\mathbb{E}(U_\infty) \leq |a| + \sup_n \mathbb{E}(|X_n|) < \infty$. Consequently, with $U_\infty =\uparrow U_N$,*

$$\mathbb{P}[\omega\colon U_\infty(\omega) = \infty] = 0.$$

Proof. Since $(x-y)^- \leq |x| + |y|$, Lemma 1.4.4 dictates $(b-a)\mathbb{E}(U_N) \leq |a| + \mathbb{E}(|X_n|) \leq |a| + \sup_n \mathbb{E}(|X_n|)$. Since $\{U_N\}_N$ is a sequence of non-negative functions such taht $U_{N+1} \geq U_N$, by monotone convergence theorem $(b-a)\mathbb{E}(U_\infty) \leq |a| + \sup_n \mathbb{E}(|X_n|) < \infty$ which also implies that $U_\infty < \infty$ almost surely. $\square$

Theorem 1.4.7. *Let $X = (X_n)_{n\in\mathbb{N}_0}$ be a supermartingale which is bounded in $L^1(\mathbb{P})$. Then there exists a random variable X_∞ such that $X_n \to X_\infty$ almost surely. Moreover, $\mathbb{P}[\omega\colon X_\infty(\omega) \in \{-\infty, \infty\}] = 0$. Consequently, any non-negative supermartingale converges almost surely to a limit which is finite almost surely.*

Proof. Note that the second part follows from the first because if a supermartingale is non-negative, then $\mathbb{E}(|X_n|) = E(X_n) \leq \mathbb{E}(X_0)$, making the supermartingale $L^1(\mathbb{P})$ bounded.

To prove the first part, let

$$\begin{aligned}
\Lambda &\overset{\text{(def)}}{=} \Big\{\omega\colon X_N(\omega)\text{does not converge to any limit in } [-\infty,\infty]\Big\} \\
&= \Big\{\omega\colon \liminf_{N\to\infty} X_N(\omega) < \limsup_{N\to\infty} X_N(\omega)\Big\} \\
&= \bigcup_{a,b\in\mathbb{Q},a<b} \big\{\omega\colon \liminf_{N\to\infty} X_N(\omega) < a < b < \limsup_{N\to\infty} \limsup_{N\to\infty} X_N(\omega)\big\} \\
&\overset{\text{(def)}}{=} \bigcup_{\substack{a,b\in\mathbb{Q}\\ a<b}} \Lambda_{a,b}.
\end{aligned}$$

By Lemma 1.4.2, $\Lambda_{a,b} \subset \big\{\omega : U_\infty = \infty\big\}$. But by Corollary 1.4.6, $\mathbb{P}[\omega\colon U_\infty(\omega) = \infty] = 0$. Therefore, $\mathbb{P}(\Lambda_{a,b}) = 0$. Hence, $\mathbb{P}(\Lambda) = 0$. Hence, $X_N \to X_\infty$ almost surely.

To conclude that $|X_\infty| < \infty$, note that

$$\mathbb{E}(|X_\infty|) = \mathbb{E}[\liminf_n |X_n|) \leq \liminf_n \mathbb{E}(|X_n|) \leq \sup_n \mathbb{E}(|X_n|) < \infty,$$

where the first upper bound follows from Fatou's lemma (which dictates that for a sequence $\{f_n\}_n$ of non-negative measurable functions $\liminf_n \int f_n \geq \int \liminf_n f_n$) and the third upper bound follows from our assumption. Hence, $\mathbb{P}[X_\infty \text{ is finite }] = 1$. $\square$

Exercise: Give an example of a martingale which is bounded in L^1 and therefore converges almost surely to a finite limit X_∞, but X_n does not converge in L^1 to X_∞.

1.5 Martingales bounded in L^2.

Let us first start with the case $p = 2$. The goal is to show

Theorem 1.5.1. *Let $(X_n)_n$ be a martingale which is bounded in L^2 (i.e. there exists $C \in (0, \infty)$ such that $\sup_n \mathbb{E}(X_n^2) \leq C$). Then there exists a random variable X_∞ such that $X_n \xrightarrow{L^2} X_\infty$.*

Proof. the proof proceeds in three steps.

Step 1: Since $(X_n)_n$ is a martingale, by induction $\mathbb{E}(X_n|\mathcal{F}_k) = X_k$ for all $k \leq n$.

Step 2: Let $Y_n := X_n - X_{n-1}$. Then Step 1 will imply that for $n \neq m$, $Y_n \perp Y_m$, i.e., $\langle Y_n, Y_m \rangle_{L^2} = \mathbb{E}(Y_n Y_m) = 0$.

Step 3: Given the above notation, $X_n = \sum_{j=1}^n Y_j$. Therefore, by Step 2, the parallelogram identity

$$Y_n^2 = \mathbb{E}(X_0^2) + \sum_{j=1}^n \mathbb{E}(Y_j^2) \tag{1.5.1}$$

holds. In particular,

$$\sup_n \mathbb{E}(X_n^2) \leq C \quad \Leftrightarrow \sum_{j=1}^\infty \mathbb{E}(Y_j^2) < \infty. \tag{1.5.2}$$

holds.

Step 4: Now assume that the martingale $(X_n)_n$ is L^2-*bounded*. Then it is in particular L^1 bounded too (e.g. by Jensen's inequality). Hence, by Doob's almost sure martingale convergence theorem, there exists a random variable X_∞ such that $X_n \to X_\infty$ almost surely. For any $r \in \mathbb{N}$, by (1.5.1) w ehave

$$\mathbb{E}[(X_{n+r} - X_n)^2] = \sum_{j=n+1}^{n+r} \mathbb{E}(Y_j^2) \tag{1.5.3}$$

For any fixed n, if we let $r \to \infty$, the left hand side in the above display converges (by Fatou's lemma or dominated convergence theorem, combined with the almost sure convergence $X_n \to X_\infty$) to $\mathbb{E}[(X_\infty - X_n)^2]$, while t he right hand side converges to the increasing limit $\sum_{j \geq n+1} \mathbb{E}(Y_j^2)$. By (1.5.2) and our assumption, $\sum_{j \geq n+1} \mathbb{E}(Y_j^2)$ is the tail of a convergent sum, which converges to zero. Hence, $\mathbb{E}((X_n - X_\infty)^2) \to 0$, and thus $X_n \to X_\infty$ in L^2. □

1.6 Doob's martingale inequalities.

The goal of this section is to prove two very powerful inequalities, both due to Doob:

Theorem 1.6.1. *Let $(X_n)_n$ be a sub-martingale. Then for any $\ell > 0$,*

$$\mathbb{P}\big[\max_{1 \leq j \leq n} |X_j| < \ell\big] \leq \frac{1}{\ell} \int_{\{\max_{j=1}^n |X_j| \geq \ell\}} |X_n| \mathrm{d}\mathbb{P} \leq \frac{1}{\ell} \mathbb{E}[|X_n|]. \tag{1.6.1}$$

Theorem 1.6.2. *Fix* $\mathbf{p > 1}$. *Let* $(X_n)_n$ *be a sub-martingale, and set* $X_n^\star = \max_{j=1}^n |X_j|$. *Then,*

$$\mathbb{E}\big[(X_n^\star)^p\big] \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\big[|X_n|^p\big]. \tag{1.6.2}$$

We will first prove a useful lemma.

Lemma 1.6.3. *Let* $p > 1$. *Let* X, Y *be non-negative random variables such that for all* $\ell > 0$,

$$\mathbb{P}[Y \geq \ell] \leq \frac{1}{\ell} \int_{\{Y \geq \ell\}} X \mathrm{d}\mathbb{P}. \tag{1.6.3}$$

Then,

$$\mathbb{E}[Y^p] \leq \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\big[X^p\big] \tag{1.6.4}$$

Proof of Lemma 1.6.3. We will carry out the proof in two steps.
Step 1: Assume that $Y \in L^p$. Then,

$$\begin{aligned}
\|Y\|_{L^p}^p = \mathbb{E}(Y^p) &= \int_\Omega \mathbb{P}(\mathrm{d}\omega) Y(\omega)^p \\
&= \int_\Omega \mathbb{P}(\mathrm{d}\omega) \Big(\int_0^{Y(\omega)} p\ell^{p-1} \mathrm{d}\ell \Big) \\
&= \int_0^\infty \mathrm{d}\ell p\ell^{p-1} \Big(\int_\Omega \mathbb{P}(\mathrm{d}\omega) \mathbb{1}_{Y(\omega) \geq \ell} \Big) \qquad \text{(by Fubini's theorem for non-negative functions)} \\
&= \int_0^\infty \mathrm{d}\ell p\ell^{p-1} \mathbb{P}[Y \geq \ell] \\
&\leq \int_0^\infty \mathrm{d}\ell p\ell^{p-2} \Big(\int_{Y \geq \ell} \mathbb{P}(\mathrm{d}\omega) X(\omega) \Big) \qquad \text{(by assumption (1.6.5))} \\
&= \int_\Omega \mathbb{P}(\mathrm{d}\omega) X(\omega) \int_0^{Y(\omega)} \mathrm{d}\ell p\ell^{p-2} \qquad \text{(again by Fubini's theorem for non-negative functions)} \\
&= \frac{p}{p-1} \int_\Omega \mathrm{d}\mathbb{P} X Y^{p-1} \\
&\leq \frac{p}{p-1} \|X\|_{L^p} \|Y\|_{L^p}^{p-1} \qquad \text{(by Hölder's inequality)},
\end{aligned}$$

which proves the claim in case $Y \in L^p$.

Step 2: Suppose $Y \notin L^p$. Then follow a routine approximation procedure by applying Step 1 above for $Y_M := Y \wedge M \in L^p$ to get uniform bounds and monotone convergence theorem. The details are left as exercise.

□

Proof of Theorem 1.6.1. Let E be the event

$$E = \{\omega : \max_{j=1}^{n} |X_j| \geq \ell|\}$$

so that E can be written as a *disjoint union* $E = \cup_{j=!}^{n} E_j$ with

$$E_j = \big\{|X_1| \leq \ell, |X_2| \leq \ell, \dots, |X_{j-1}| \leq \ell, |X_j| \geq \ell\big\} \in \mathcal{F}_j.$$

Then by Markov's inequality, we have

$$\mathbb{P}[E_j] \leq \frac{1}{\ell} \int_{E_j} |X_j| \mathrm{d}\mathbb{P}. \tag{1.6.5}$$

Since $(X_n)_n$ is a sub-martingale, so is $(|X_n|)_n$ by Jensen's inequality. Therefore, exploiting also that $E_j \in \mathcal{F}_j$, we have

$$\mathbb{E}\big[\mathbb{1}_{E_j}\big(|X_n| - |X_j|\big)\big|\mathcal{F}_j\big] = \mathbb{1}_{E_j}\mathbb{E}\big[\big(|X_n| - |X_j|\big)\big|\mathcal{F}_j\big] \geq 0 \quad \text{a.s.}$$

Applying expectations on both side we get

$$\mathbb{E}\big[\mathbb{1}_{E_j}\big(|X_n| - |X_j|\big)\big] \geq 0. \tag{1.6.6}$$

Now

$$\begin{aligned} \mathbb{P}(E) &= \sum_{j=1}^{n} \mathbb{P}(E_j) \\ &\overset{(1.6.5)}{\leq} \frac{1}{\ell}\bigg[\int_{E_1} |X_n| \mathrm{d}\mathbb{P} + \dots + \int_{E_n} |X_n| \mathrm{d}\mathbb{P}\bigg] \\ &= \frac{1}{\ell} \int_{E_1 \cup \dots \cup E_n} |X_n| = \frac{1}{\ell} \int_{E} |X_n| \mathrm{d}\mathbb{P}, \end{aligned}$$

which proves the desired assertion (1.6.1). □

Proof of Theorem 1.6.2. The proof is an immediate consequence of Theorem 1.6.1 and Lemma 1.6.3.

□

1.7 Uniform integrability.

Let us consider a very simple example: $\Omega = (0,1)$, $\mathcal{B}$ is the Borel σ-algebra, carrying the probability measure $\mathbb{P} = \text{Leb}$. Let $X_n = n\mathbb{1}_{(0,\frac{1}{n})}$. Then $X_n \to 0 =: X_\infty$ almost everywhere. On the other hand,

$$1 = E(X_n) = E(|X_n|) = \|X_n\|_{L^1} \nrightarrow 0 = \|X_\infty\|_{L^1}$$

and thus $X_n \nrightarrow X_\infty$ in L^1. What the sequence $(X_n)_n$ misses out on is known on *uniform integrability*, a very important property in probability theory.

Definition 1.7.1. *A family of random variables $(X_n)_n$ is called uniformly integrable (with the acronym $(X_n)_n$ is UI), if for any given $\varepsilon > 0$, there exists a $K \in (0,\infty)$ (which could possibly depend on ε but not on n) such that*

$$\int_{|X_n|\geq K} |X_n| < \varepsilon \qquad \textit{for all } n.$$

An immediate consequence of the above definition is

Lemma 1.7.2. *If $(X_n)_n$ is UI, then $(X_n)_n$ is L^1 bounded.*

Proof. It follows immediately from the definition (1.7.1) that that for a UI family $(X_n)_n$, $\sup_n \mathbb{E}[|X_n|] \leq K + \varepsilon$. □

Here is a couple of suitable criteria that guarantee uniform integrability.

Lemma 1.7.3. *Let $(X_n)_n$ be a sequence of random variables.*

- *If $(X_n)_n$ is bounded in L^p for $p > 1$, then $(X_n)_n$ is UI.*
- *If there exists a random variable $Y \in L^1$, then*

$$|X_n(\cdot)| \leq Y(\cdot) \quad \forall\, n \Rightarrow (X_n)_n \textit{ is UI.} \tag{1.7.1}$$

Proof. Let us prove the first part. Assume that $(X_n)_n$ is bounded in L^p for some $p > 1$. Fix any $\varepsilon > 0$.Then by Hölder's inequality, for $1/p + 1/q = 1$, for some K (to be chosen later)

$$\int_{|X_n|\geq K} |X_n| \leq \|X_n\|_{L^p} \mathbb{P}[|X_n| > K]^{1/q} \overset{\text{Markov's}\leq}{\leq} \|X_n\|_{L^p} K^{-q} \|X_n\|_{L^1}^{1/q}.$$

Since $\sup_n \|X_n\|_{L^p} \leq 1$ for $p > 1$, we can now choose K large enough to make the above quantity smaller than ε, which proves that $(X_n)_n$ is UI.

The proof of the second part depends on the following simple assertion which is of fundamental use in probability theory. In a sense the following lemma says that if the measure of a set is small, the integral of an L^1 function over that set is also small. The proof involves yet another beautiful application of Borel-Cantelli lemma.

Lemma 1.7.4. *Let $X \in L^1$.*

- *Given $\varepsilon > 0$, there exists $\delta > 0$ such that for all measurable set $F \in \mathcal{F}$,*

$$\mathbb{P}(F) < \delta \Rightarrow \int_F |X| \mathrm{d}\mathbb{P} < \varepsilon. \tag{1.7.2}$$

- *Given $\varepsilon > 0$ there exists $K \in (0,\infty)$ such that*

$$\int_{\{|X|>K\}} |X| < \varepsilon. \tag{1.7.3}$$

Proof of Lemma 1.7.4. Let us first prove (1.7.2). Assume that there is $\varepsilon > 0$ such that for all $\delta = 2^{-n}$ for $n = 1, 2, 3, \dots$ and some $F_n \in \mathcal{F}$,

$$\mathbb{P}(F_n) < 2^{-n}, \quad \text{and} \int_{F_n} |X| \geq \varepsilon \; \forall n \in \mathbb{N}.$$

Since $\sum_n \mathbb{P}(F_n) < \infty$, by Borel-Cantelli lemma,

$$\mathbb{P}(F) = 0, \qquad \text{where} \quad F := \limsup_n F_n := \cap_{n\geq 1} \cup_{m \geq n} F_m.$$

But

$$\int_F |X| \mathrm{d}\mathbb{P} = \int \overline{\lim}_{n\to\infty} (|X| \mathbb{1}_{F_n}) \geq \overline{\lim}_{n\to\infty} \int_{F_n} |X| \geq \varepsilon$$

which contradicts the fact that $\mathbb{P}(F) = 0$, proving the first assertion. In the above assertion, the lower bound follows from reverse Fatou's lemma. [1]

The second assertion follows from the first one since we can take $F = \{|X| > K\}$. Since $X \in L^1$, we can choose K large enough so that $\mathbb{P}(F) < \delta$, which, combined with the first part implies that $\int_{|X| \geq K} |X| < \varepsilon$. □

Proof of (1.7.2)**.** Since $Y \in L^1$, by the second part of Lemma 1.7.4, given any $\varepsilon > 0$, there exists K such that $\mathbb{E}(Y \mathbb{1}\{Y > K\|) < \varepsilon$. Now by our assumption $\mathbb{E}(|X_n| \mathbb{1}\{|X_n| \geq K\}) \leq \mathbb{E}(Y \mathbb{1}\{Y > K\}) < \varepsilon$. □

The following fact is also quite useful.

Lemma 1.7.5. *Let $X \in L^1$. Then the family*

$$\mathscr{F} := \big\{ \mathbb{E}[X|\mathcal{G}] \colon \mathcal{G} \subset \mathcal{F} \big\}$$

is UI.

Proof. Fix $\varepsilon > 0$. Choose $\delta > 0$ such that for all $F \in \mathcal{F}$, $\mathbb{P}(F) < \delta$ implies $E(|X| \mathbb{1}_F) < \varepsilon$. Let $Y = \mathbb{E}(Y|\mathcal{G})$ so that

$$|Y| \leq \mathbb{E}(|X||\mathcal{G}) \Rightarrow \mathbb{E}(|Y|) \leq \mathbb{E}(|X|) \Rightarrow /P(|Y| > K) < \delta, \quad \text{for suitable } K.$$

By definition Y is $\mathcal{G}$-measurable, so that $\{Y > K\} \in \mathcal{G}$. Hence,

$$\mathbb{E}[|Y| \mathbb{1}\{|Y| \geq K\|] \leq \mathbb{E}\big[\mathbb{E}(|X||\mathcal{G}) \mathbb{1}\{|Y| \geq K\}\big] = \int_{\{|Y| > K\}} |X| < \varepsilon$$

since $\mathbb{P}(|Y| > K) < \delta$ and by the first part of Lemma 1.7.4. □

[1] Reverse Fatou's lemma asserts that if $f_n \geq 0$ and $f_n \leq g$ such that $g \in L^1$, then $\limsup_n \int f_n \geq \int \limsup_n f_n$. To see this, set $\widetilde{f}_n = g - f_n \geq 0$ and apply Fatou's lemma to $\widetilde{f}_n$. The argument will need that $g \in L^1$.

The next result provides an important result. Roughly speaking, it says

$$\text{Convergence in } L^1 \Leftrightarrow \text{ (Convergence in probability } + \text{ UI)}.$$

Theorem 1.7.6. *Let $(X_n)_n$ be a sequence of random variables such that $X_n \in L^1$ for all n. Let $X \in L^1$. Then the following conditions are equivalent:*

(i) $X_n \to X$ in L^1.

(ii) $X_n \to X$ in probability and $(X_n)_n$ is UI.

Proof. Let us first prove

(ii) $\Rightarrow$ (i): Fix some $K \in (0, \infty)$ to be chosen later. We define the following approximation of the identity function:

$$\varphi(x) = \varphi_K(x) = \begin{cases} K & \text{if } x > K \\ x & \text{if } |x| \le K \\ -K & \text{if } x < -K. \end{cases}$$

Then obviously φ is Lipschitz, i.e. $|\varphi(x) - \varphi(y)| \le |x - y|$ and in particular $\varphi(x) \le x$. Note also that φ differs from the identity function only when $|x| \ge K$. Using these properties,

$$\begin{aligned} &\mathbb{E}\big[|X_n - X|\big] \\ &\le \mathbb{E}\big[|X_n - \varphi(X_n)|\big] + \mathbb{E}\big[|\varphi(X) - X|\big] + \mathbb{E}\big[|\varphi(X_n) - \varphi(X)|\big] \\ &\le 2 \sup_n \mathbb{E}\big[|X_n| \mathbb{1}\{|X_n| \ge K\}\big] + 2\mathbb{E}\big[|X| \mathbb{1}\{|X| \ge K\}\big] + \mathbb{E}\big[|\varphi(X_n) - \varphi(X)|\big]. \end{aligned}$$

Since $(X_n)_n$ is UI, the first term on the last display above is less than ε for a suitable $K = K_1$, while the second term is also less than ε because of the second part of Lemma 1.7.4 for (possibly another) suitable $K = K_2$ (if K_1 and K_2 differ, we will choose $K = \max\{K_1, K_2\}$ at the end). To handle the third term, note that $X_n \to X$ in probability. Since $|\varphi(x) - \varphi(y)| \le |x - y|$, it follows that $\varphi(X_n) \to \varphi(X)$ in probability too. Moreover, since $|\varphi(X_n)| \le K$, bounded convergence theorem [2] allows us make the third term smaller than ε too.

We now prove the converse

(i) $\Rightarrow$ (ii): Obviously convergence in l^1 implies convergence in probability by Markov's inequality. We need to prove the UI property. Indeed, since $X_n \to X$ in L^1 and $X \in L^1$, it follows that

[2]Bounded convergence theorem says that if $X_n \to X$ in probability and $|X_n| \le K$ for some finite K, then $X_n \to X$ in L^1. It can be proved easily. Indeed, note that $\int |X_n - X| \le \int_{|X_n - X| \ge \varepsilon} |X_n - X| + \varepsilon$, which can be made smaller than $2K\mathbb{P}[|X_n - X| \ge \varepsilon] + \varepsilon$ if we can show that $|X| \le K$ almost surely, which will imply the desired claim. To see $|X| \le K$ a.s. note that $\mathbb{P}[|X| \ge K + \frac{1}{m}] \le \mathbb{P}[|X_n - X| > \frac{1}{m}]$ for all n. Since the last probability converges to zero, it follows that $\mathbb{P}[|X| \ge K + \frac{1}{m}] = 0$ for all $m \in \mathbb{N}$. We now take $m \uparrow \infty$ to conclude the proof of the bounded convergence theorem.

$\sup_n \mathbb{E}(|X_n|) \leq \delta K$ for all $\delta > 0$ and a suitable $K \in (0, \infty)$. Hence, $\mathbb{P}[|X_n| > K] < \delta$ by Markov's inequality. But we can estimate

$$\int_{\{|X_n|>K\}} |X_n| \leq \int_{\{|X_n|>K\}} |X - X_n| + \int_{\{|X_n|>K\}} |X|$$

so that the first term is smaller than ε since $X_n \to X$ in L^1 and the second term is smaller than ε because of the second part of Lemma 1.7.4 and the fact that we can make $\mathbb{P}[|X_n| > K] < \delta$.

□

1.8 Martingale convergence theorems in L^p for $p \geq 1$.

Throughout this section we will assume that we have a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_n, \mathbb{P})$ such that $\mathcal{F} = \sigma(\cup_n \mathcal{F}_n)$. The first observation is

Lemma 1.8.1. *Let $X \in L^p$ for some $p \geq 1$. Then $X_n := \mathbb{E}[X|\mathcal{F}_n]$ defines a martingale $(X_n)_n$.*

Proof. Obviously $X_n \in L^1$ and X_n is $\mathcal{F}_n$ measurable. Moreover

$$\begin{aligned}
\mathbb{E}[X_n - X_{n-1}|\mathcal{F}_{n-1}] &= \mathbb{E}\Big[\mathbb{E}\big[X|\mathcal{F}_n\big] - \mathbb{E}\big[X|\mathcal{F}_{n-1}\big]\big|\mathcal{F}_{n-1}\Big] \\
&= \mathbb{E}\big[\mathbb{E}(X|\mathcal{F}_n)|\mathcal{F}_{n-1}\big] - \mathbb{E}\big[\mathbb{E}(X|\mathcal{F}_{n-1})|\mathcal{F}_{n-1}\big] \\
&\text{(by part } (iv) + (ii) \text{ of Lemma2.15.2)} \\
&= \mathbb{E}[X|\mathcal{F}_{n-1}] - \mathbb{E}[X|\mathcal{F}_{n-1}] = 0.
\end{aligned}$$

□

The next theorem shows that martingales of the above form converge in L^p to the fixed random variable X.

Theorem 1.8.2. *Let $X \in L^p$ for some $p \geq 1$. Then $X_n := \mathbb{E}[X|\mathcal{F}_n]$ defines a martingale $(X_n)_n$ and $X_n \to X$ in L^p.*

Proof. We will split the proof into two cases.

Step 1: We assume that X is bounded. Then in particular,

$$|X_n| \leq \mathbb{E}[|X||\mathcal{F}_n] \Rightarrow \sup_{n,\omega} |X_n(\omega)| < \infty.$$

That is $(X_n)_n$ is bounded in L^∞. In particular, $(X_n)_n$ is a sequence of martingales which is bounded in L^2. By Theorem 1.5.1, there exists a random variable Y such that $X_n \to Y$ in L^2. Now fix $m \in \mathbb{N}_0$

and $A \in \mathcal{F}_m$. Then in particular,

$$\begin{aligned}
\int_A Y \mathrm{d}\mathbb{P} &= \int_A X_n \mathrm{d}\mathbb{P} \qquad \left(\text{Since } \int_A |X_n - Y| \mathrm{d}\mathbb{P} \leq C\left(\int_A |X_n - Y|^2 \mathrm{d}\mathbb{P}\right)^{1/2} \to 0\right) \\
&= \lim_n \int_A X_n \mathrm{d}\mathbb{P} = \lim_n \mathbb{E}[X_n \mathbb{1}_A] \\
&= \lim_n \mathbb{E}\big[\mathbb{E}(X|\mathcal{F}_n)\mathbb{1}_A\big] \\
&\overset{\forall m \geq n}{=} \int_A X \mathrm{d}\mathbb{P}.
\end{aligned}$$

Hence,

$$\int_A Y \mathrm{d}\mathbb{P} = \int_A X \mathrm{d}\mathbb{P} \qquad \forall A \in \mathcal{F}_m.$$

Since $\mathcal{F} = \sigma(\cup_n \mathcal{F}_n)$, it follows that $\int_A Y \mathrm{d}\mathbb{P} = \int_A X \mathrm{d}\mathbb{P} \quad \forall A \in \mathcal{F}$ and therefore $X = Y$ almost everywhere. Thus $X_n \to X$ in L^2. Since $(X_n)_n$ is bounded in L^∞, convergence in L^p are equivalent for all $p \geq 1$ [3]. Hence $X_n \to X$ in L^p for all $p \geq 1$.

Step 2: Now assume that $X \in L^p$ but $X \notin L^\infty$. Define

$$X' = \begin{cases} X & \text{if } \{|X| \leq M\}, \\ 0 & \text{else.} \end{cases}$$

Obviously $X' \in L^\infty$ and moreover $\int |X - X'|^p = \int_{|X|>M} |X|^p \mathrm{d}\mathbb{P} < \varepsilon$ by the second part of Lemma 1.7.4. Now let

$$X'_n = \mathbb{E}[X'|\mathcal{F}_n] \Rightarrow X'_n \xrightarrow{L^p} X'.$$

Now by definitions of X_n, X'_n and Jensen's inequality,

$$\|X_n - X'_n\|^p_{L^p} \leq \|X - X'\|^p_{L^p} < \varepsilon,$$

and therefore,

$$\|X_n - X\|^p_{L^p} \leq \|X_n - X'_n\|^p_{L^p} + \|X_n - X'\|^p_{L^p} + \|X - X'\|^p_{L^p} < 3\varepsilon,$$

which proves the claim. □

The next theorem shows that the martingales analyzed in Lemma 1.8.1 and Theorem 1.8.2 come up in a natural way, provided we assume that they are bounded in L^p for $p > 1$ (and thus, in particular they are UI).

Theorem 1.8.3. *Fix $p > 1$. Let $(X_n)_n$ be a martingale bounded din L^p. Then there exist $X \in L^p$ such that $X_n = \mathbb{E}[X|\mathcal{F}_n]$ for all n. In particular, $X_n \to X$ in L^p.*

[3]Indeed, convergence in L^2 implies convergence in L^p for $p \in [1, 2]$ and if $p > 2$, then $\int |X_n - X|^p = \int |X_n - X|^2 |X_n - X|^{p-2} \leq C(\|X_n\|^{p-2}_\infty + \|X\|^{p-2}_\infty) \int |X_n - X|^2 \leq C' \|X_n - X\|^2_{L^2} \to 0$.

Proof. Since $(X_n)_n$ is bounded in L^p for $p > 1$, by Banach-Alaoglu theorem (See Appendix), [4] we have some $X \in L^p$ and a subsequence $(X_{n_j})_j$ such that $\int X_{n_j} Y \to \int XY$ for all $Y \in L^q$ with $1/p + 1/q = 1$. Choose $A \in \mathcal{F}_m$ for some fixed m and take $Y = \mathbb{1}_A \in L^q$. Note that

$$\int_A X \mathrm{d}\mathbb{P} = \lim_j \int_A X_{n_j} \mathrm{d}\mathbb{P} = \lim_j \mathbb{E}[X_{n_j} \mathbb{1}_A] = \mathbb{E}[X_m \mathbb{1}_A],$$

where the last identity follows from the observation that eventually $n_j > m$ and $A \in \mathcal{F}_m$ and $(X_n)_n$ is a martingale. Thus,

$$\int_A X_n \mathrm{d}\mathbb{P} = \int_A X d\mathbb{P} \qquad \forall A \in \mathcal{F}_m.$$

Uniqueness of conditional expectation enforces $X_m = \mathbb{E}[X|\mathcal{F}_m]$. By Theorem 1.8.2 it also follows that $X_m \to X$ in L^p as $m \to \infty$. □

[4]This is exactly where we need reflexivity of the Banach space L^p which holds only if $p > 1$. The argument would not work for $p = 1$.

1.9 Stopping times

To be written down.

Chapter 2

Markov chains

2.1 Examples

Example 2.1.1 (Markov chain with two states)**.** *Consider a phone which can be in two states: "free"$=0$ and "busy"$=1$. The set of the states of the phone is*

$$E = \{0, 1\}.$$

We assume that the phone can randomly change its state in time (which is assumed to be discrete) according to the following rules.

case1. *If at some time n the phone is free, then at time $n+1$ it becomes busy with probability p or it stays free with probability $1-p$.*

case2. *If at some time n the phone is busy, then at time $n+1$ it becomes free with probability q or it stays busy with probability $1-q$.*

Denote by X_n the state of the phone at time $n = 0, 1, \dots$. Thus, $X_n : \Omega \to \{0, 1\}$ is a random variable and our assumptions can be written as follows:

$$p_{00} := \mathbb{P}[X_{n+1} = 0 | X_n = 0] = 1 - p, \qquad p_{01} := \mathbb{P}[X_{n+1} = 1 | X_n = 0] = p,$$
$$p_{10} := \mathbb{P}[X_{n+1} = 0 | X_n = 1] = q, \qquad p_{11} := \mathbb{P}[X_{n+1} = 1 | X_n = 1] = 1 - q.$$

We can write these probabilities in form of a transition matrix

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

Additionally, we will make the following assumption which is called the Markov property: Given that at some time n the phone is in state $i \in \{0, 1\}$, the behavior of the phone after time n does not depend on the way the phone reached state i in the past.

Problem 2.1.2. *Suppose that at time 0 the phone was free. What is the probability that the phone will be free at times $1, 2$ and then becomes busy at time 3?*

Solution 2.1.3. *This probability can be computed as follows:*

$$\mathbb{P}[X_1 = X_2 = 0, X_3 = 1] = p_{00} \cdot p_{00} \cdot p_{01} = (1-p)^2 p.$$

Problem 2.1.4. *Suppose that the phone was free at time* 0*. What is the probability that it will be busy at time* 3*?*

Solution 2.1.5. *We have to compute* $\mathbb{P}[X_3 = 1]$*. We know the values* $X_0 = 0$ *and* $X_3 = 1$*, but the values of* X_1 *and* X_2 *may be arbitrary. We have the following possibilities:*

(i) $X_0 = 0, X_1 = 0, X_2 = 0, X_3 = 1$. *Probability:* $(1-p) \cdot (1-p) \cdot p$.

(ii) $X_0 = 0, X_1 = 0, X_2 = 1, X_3 = 1$. *Probability:* $(1-p) \cdot p \cdot (1-q)$.

(iii) $X_0 = 0, X_1 = 1, X_2 = 0, X_3 = 1$. *Probability:* $p \cdot q \cdot p$.

(iv) $X_0 = 0, X_1 = 1, X_2 = 1, X_3 = 1$. *Probability:* $p \cdot (1-q) \cdot (1-q)$.

The probability we look for is the sum of these 4 *probabilities:*

$$\mathbb{P}[X_3 = 1] = (1-p)^2 p + (1-p)(1-q)p + p^2 q + p(1-q)^2.$$

Example 2.1.6 (Gambler's ruin)**.** *At each unit of time a gambler plays a game in which he can either win* 1 *Euro (which happens with probability* p*) or he can loose* 1 *Euro (which happens with probability* $1-p$*). Let* X_n *be the capital of the gambler at time* n*. Let us agree that if at some time* n *the gambler has no money (meaning that* $X_n = 0$*), then he stops to play (meaning that* $X_n = X_{n+1} = \ldots = 0$*). We can view this process as a Markov chain on the state space* $E = \{0, 1, 2, \ldots\}$ *with transition matrix*

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

2.2 Definition of Markov chains

Let us consider some system. Assume that the system can be in some states and that the system can change its state in time. The set of all states of the system will be denoted by E *and called the state space of the Markov chain. We always assume that the state space* E *is a finite or countable set. Usually, we will denote the states so that* $E = \{1, \ldots, N\}$*,* $E = \mathbb{N}$*, or* $E = \mathbb{Z}$*.*

Assume that if at some time the system is in state $i \in E$*, then in the next moment of time it can switch to state* $j \in E$ *with probability* p_{ij}*. We will call* p_{ij} *the transition probability from state* i *to state* j*. Clearly, the transition probabilities should be such that*

(i) $p_{ij} \geq 0$ *for all* $i, j \in E$*.*

(ii) $\sum_{j \in E} p_{ij} = 1$ *for all* $i \in E$.

We will write the transition probabilities in form of a transition matrix

$$P = (p_{ij})_{i,j \in E}.$$

The rows and the columns of this matrix are indexed by the set E. The element in the i-th row and j-th column is the transition probability p_{ij}. The elements of the matrix P are non-negative and the sum of elements in any row is equal to 1. Such matrices are called stochastic.

Definition 2.2.1. *A Markov chain with state space E and transition matrix P is a stochastic process $\{X_n : n \in \mathbb{N}_0\}$ taking values in E such that for every $n \in \mathbb{N}_0$ and every states $i_0, i_1, \ldots, i_{n-1}, i, j$ we have*

$$\begin{aligned} \mathbb{P}[X_{n+1} = j | X_n = i] &= \mathbb{P}[X_{n+1} = j | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i] \\ &= p_{ij}, \end{aligned} \tag{2.2.1}$$

provided that $\mathbb{P}[X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i] \neq 0$ (which ensures that the conditional probabilities are well-defined).

Condition (2.2.1) *is called the Markov property.*

In the above definition it is not specified at which state the Markov chain starts at time 0. In fact, the initial state can be in general arbitrary and we call the probabilities

$$\alpha_i := \mathbb{P}[X_0 = i], \quad i \in E, \tag{2.2.2}$$

the initial probabilities. We will write the initial probabilities in form of a row vector $\alpha = (\alpha_i)_{i \in E}$. This vector should be such that $\alpha_i \geq 0$ for all $i \in E$ and $\sum_{i \in E} \alpha_i = 1$.

Theorem 2.2.2. *For all $n \in \mathbb{N}_0$ and for all $i_0, \ldots, i_n \in E$ it holds that*

$$\mathbb{P}[X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n] = \alpha_0 p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}. \tag{2.2.3}$$

Proof. We use the induction over n. The induction basis is the case $n = 0$. We have $\mathbb{P}[X_0 = i_0] = \alpha_{i_0}$ by the definition of initial probabilities, see (2.2.2). Hence, Equation (2.2.3) holds for $n = 0$.

Induction assumption: Assume that (2.2.3) holds for some n. We prove that (2.2.3) holds with n replaced by $n + 1$. Consider the event $A = \{X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\}$. By the induction assumption,

$$\mathbb{P}[A] = \alpha_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}.$$

By the Markov property,

$$\mathbb{P}[X_{n+1} = i_{n+1} | A] = p_{i_n i_{n+1}}.$$

It follows that

$$\begin{aligned}\mathbb{P}[X_0 = i_0, X_1 = i_1, \dots, X_n = i_n, X_{n+1} = i_{n+1}] &= \mathbb{P}[X_{n+1} = i_{n+1}|A] \cdot \mathbb{P}[A] \\ &= p_{i_n i_{n+1}} \cdot \alpha_{i_0} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n} \\ &= \alpha_{i_0} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n} p_{i_n i_{n+1}}.\end{aligned}$$

This completes the induction. □

Remark 5 If $\mathbb{P}[A] = 0$, then in the above proof we cannot use the Markov property. However, in the case $\mathbb{P}[A] = 0$ both sides of (2.2.3) are equal to 0 and (2.2.3) is trivially satisfied.

Theorem 2.2.3. *For every $n \in \mathbb{N}$ and every state $i_n \in E$ we have*

$$\mathbb{P}[X_n = i_n] = \sum_{i_0, \dots, i_{n-1} \in E} \alpha_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n}.$$

Proof. We have

$$\begin{aligned}\mathbb{P}[X_n = i_n] &= \sum_{i_0, \dots, i_{n-1} \in E} \mathbb{P}[X_0 = i_0, X_1 = i_1, \dots, X_n = i_n] \\ &= \sum_{i_0, \dots, i_{n-1} \in E} \alpha_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n},\end{aligned}$$

where the last step is by Theorem 2.2.2. □

2.3 General state spaces.

While we will mostly be concerned with the case when the state space of a Markov chain E is a discrete set, it is useful to consider state spaces E which is just assumed to be a complete separable metric space. For any Markov chain taking vales in E we will need a transition probability measure $P(x, \cdot)$ (i.e. for any $x \in E$, $P(x, \cdot)$ is a probability measure) and a starting distribution α which is assumed to be any given probability measure on E.

Definition 2.3.1. *Let $\mathcal{B}(E)$ be the set of all Borel-measurable and bounded functions $f : E \to \mathbb{R}$. For any transition probability measure $P(x, \cdot)$, we will write*

$$\mathbf{P} \colon \mathcal{B}(E) \to \mathcal{B}(E), \qquad (\mathbf{P}f)(x) = \int_E f(y) P(x, \mathrm{d}y).$$

We will also write

$$\mathscr{L} := \mathbf{P} - \mathrm{Id}$$

Definition 2.3.2. *A stochastic process $(X_n)_n$ taking values in a complete, separable metric space E is called a Markov chain with initial distribution α and transition probability measure $P(x, \cdot)$ if*

- *X_0 is distributed according to α.*
- *For any bounded and measurable function $f : E \to \mathbb{R}$, almost surely,*

$$\mathbb{E}_\alpha\big[f(X_{n+1})\big|\mathcal{F}_n\big] = \mathbb{E}_\alpha[f(X_{n+1}|X_n] = \mathbb{E}_{X_n}[f(X_{n+1})]$$
$$= \int f(y)P(X_n, \mathrm{d}y) =: (\mathbf{P}f)(X_n)$$

2.4 An important result: Martingale characterization of Markov chains.

The following result is extremely useful for various reasons.

Theorem 2.4.1. *Let $P(x,\cdot)$ be a transition probability measure on a space E. Then a stochastic process $(X_n)_n$ is a Markov chain with transition probability $P(x,\cdot)$ if and only if for any bounded measurable function $f : E \to \mathbb{R}$,*

$$M_n(f) = f(X_n) - f(X_0)\sum_{j=1}^{n-1}(Lf)(X_j), \qquad (Lf)(x) = (\mathbf{P} - I)f(x) = \int_E (f(y) - f(x))P(x, \mathrm{d}y)$$

defines a martingale $(M_n(f))_n$ w.r.t. the canonical filtration of $(X_n)_n$.

Proof. ($\Rightarrow$): Since f is bounded, obviously $M_n(f) \in L^1$ and by definition also $M_n(f) \in \mathcal{F}_n = \sigma(X_j : j \leq n)$. Now,

$$\mathbb{E}\big[M_{n+1}(f) - M_n(f)|\mathcal{F}_n\big] = \mathbb{E}[f(X_{n+1})|\mathcal{F}_n] - f(X_n) - (\mathbf{P} - I)f(X_n) = \underbrace{\mathbb{E}[f(X_{n+1})|\mathcal{F}_n]}_{\mathbb{E}[f(X_{n+1})|X_n)]} - (\mathbf{P}f)(X_n)$$
$$= (\mathbf{P}f)(X_n) - (\mathbf{P}f)(X_n) = 0.$$

($\Leftarrow$): We simply revert the argument. Indeed, the martingale property of $(M_n(f))_n$ dictates that, for every bounded measurable f,

$$\mathbb{E}[f(X_{n+1})|X_n] = (\mathbf{P}f)(X_n) = \int f(y)P(X_n, \mathrm{d}y),$$

which is the desired Markov property with transition probability $P(x,\cdot)$. $\square$

2.5 Application: Solution of the Dirichlet problem

A key goal in probability theory and analysis is the complete description of a harmonic function in terms of its boundary values. To illustrate a simple case, we consider a Markov chain $(X_n)_n$ taking values in a state space E. Let $A \subset E$ and

$$\tau_A = \inf\big\{n \geq 0 : X_n \in A\big\}, \qquad u(x) = u_A(x) :- = \mathbb{P}_x[\tau_A < \infty] \in [0,1].$$

It is an exercise to check that τ_A is a stopping time. Clearly, if $x \in A$, then $u(x) = 1$. Now if $x \notin A$, then by the Markov property, we have

$$u(x) = P(x, A) + \int_{A^c} u(y) P(x, \mathrm{d}y) = \int_E u(y) P(x, \mathrm{d}y) = (\mathbf{P}u)(x),$$

which implies that $Lu = 0$ (i.e. the function u is harmonic in A^c) and on A, $u = 1$, which is a solution of the Dirichlet problem

$$\begin{cases} Lv = 0 & \text{on } A^c \\ v = 1 & \text{on } \ A. \end{cases} \tag{2.5.1}$$

The following statement provides a stronger statement.

Theorem 2.5.1.
- *Among all non-negative solutions of (**??**), $u(x) = \mathbb{P}_x[\tau_A < \infty]$ is the sammlest.*

- *$u(x) = \mathbb{P}_x[\tau_A < \infty] = 1$, then any bounded solution of*

$$\begin{cases} Lv = 0 & \text{on } A^c \\ v = f & \text{on } \ A. \end{cases} \tag{2.5.2}$$

  is equal to

$$v(x) := \mathbb{E}_x\big[f(X_{\tau_A})\big]. \tag{2.5.3}$$

Proof. **Step 1:** Let us prove the first part, i.e. if v is any non-negative solution of (**??**), then we want to show that $v(x) \geq u(\cdot) = \mathbb{P}_x[\tau_A < \infty]$. Let

$$w(\cdot) = v(\cdot) \wedge 1 \in [0,1]$$

We claim that $\{W(X_n)\}_n$ is a *supermartingale*. Indeed, first note that $w(x) = 1$ on A and if $x \not{in} A$, then, since v solves $Lv = 0$,

$$(\mathbf{P}w)(x) \leq \int\int v(y) P(x, \mathrm{d}y) = v(x).$$

But obviously, $\mathbf{P}w(\cdot) \leq 1$ which, combined with the last upper bound, also implies $\mathbf{P}W(\cdot) \leq W(\cdot)$ on A^c. On the other hand, clearly, $\mathbf{P}W(\cdot) \leq 1 = W(\cdot)$ on A. Thus,

$$\mathbf{P}W(\cdot) \leq W(\cdot) \qquad \text{evereywhere}\,, \Rightarrow Lw(\cdot) \leq 0 \qquad \text{evereywhere.}$$

Hence, W is superharmonic w.r.t. the transition probability P, and thus by Theorem **??**, $\{W(X_n)\}_n$ is a *supermartingale*. Thus, by the optional stopping theorem, if τ is any *bounded stopping time*, then

$$\mathbb{E}_x[w(X_\tau)] \leq \mathbb{E}_x[w(X_0)] = w(x).$$

Since $\tau_A = \inf\{n \geq 0 : X_n \in A\}$ need not be bounded, we let $\tau_A^{(N)} = \tau_A \wedge N$. Then

$$\mathbb{E}_x[w(X_{\tau_A^{(N)}})] \leq \mathbb{E}_x[w(X_0)] = w(x) \tag{2.5.4}$$

the event $\{\tau_A < \infty\}$, $\tau_A^{(N)} \uparrow \tau_A$ and on this event, also $w(X_{\tau_A^{(N)}}) \to w(X_{\tau_A} = 1$, by definition of w. Thus,

$$\begin{aligned} u(x) = \mathbb{P}_x[\tau_A < \infty] = \int_{\tau_A<\infty} 1 \mathrm{d}\mathbb{P}_x = \int_{\tau_A<\infty} w(X_{\tau_A}) \mathrm{d}\mathbb{P}_x \\ \leq \limsup_{N\to\infty} \int_{\tau_A<\infty} w(X_{\tau_A^{(N)}}) \mathrm{d}\mathbb{P}_x \\ \leq \limsup_{N\to\infty} \mathbb{E}_x\left[w(X_{\tau_A^{(N)}})\right] \\ \leq w(x) \leq v(x), \end{aligned} \tag{2.5.5}$$

which proves the desired claim.

Step 2: Let $h := Lv$. Then we know that $h = 0$ on A^c, and

$$M_n(v) = v(X_n) - v(X_0) - \sum_{j=1}^{n-1} h(X_j)$$

defines a martingale. Note that $h(X_{j-1}) =$ for $j \leq \tau_A$. Thus again with $\tau_A^{(N)} = \tau_A \wedge N$, by the optional stopping theorem,

$$v(x) = \mathbb{E}_x[v(X_{\tau_A^{(N)}})].$$

Since we assume that $u(x) = \mathbb{P}_x[\tau_A < \infty] = 1$, by an argument similar to (2.5.5), we again have by bounded convergence theorem,

$$v(x) = \mathbb{E}_x[v(X_{\tau_A})],$$

which proves the second part of Theorem 2.5.1.

□

Remark 6 Such arguments are powerful tools for the study of qualitative proper- ties of Markov chains. If h is given, solutions v to equations of the type

$$Lv = (\mathbf{P} - I) = h$$

are often easily constructed. They can be used to produce martingales, sub- martingales or supermartingales that have certain behavior and that in turn implies certain qualitative behavior of the Markov chain. We will see several illustrations of this method as we move on.

2.6 n-step transition probabilities

If we want to indicate that the Markov chain starts at state $i \in E$ at time 0, we will write $\mathbb{P}_i$ instead of $\mathbb{P}$.

Definition 2.6.1. *The n-step transition probabilities of a Markov chain are defined as*

$$p_{ij}^{(n)} := \mathbb{P}_i[X_n = j].$$

We will write these probabilities in form of the n-step transition matrix $P^{(n)} = (p_{ij}^{(n)})_{i,j \in E}$.

By Theorem 2.2.3 we have the formula

$$p_{ij}^{(n)} = \sum_{i_1,\dots,i_{n-1} \in E} p_{ii_1} p_{i_1 i_2} \dots p_{i_{n-1} j}.$$

The next theorem is crucial. It states that the n-step transition matrix $P^{(n)}$ can be computed as the n-th power of the transition matrix P.

Theorem 2.6.2. *We have $P^{(n)} = P^n = P \cdot \ldots \cdot P$.*

Proof. We use induction over n. For $n = 1$ we have $p_{ij}^{(1)} = p_{ij}$ and hence, $P^{(1)} = P$. Thus, the statement of the theorem is true for $n = 1$.

Let us now assume that we already proved that $P^{(n)} = P^n$ for some $n \in \mathbb{N}$. We compute $P^{(n+1)}$. By the formula of total probability, we have

$$p_{ij}^{(n+1)} = \mathbb{P}_i[X_{n+1} = j] = \sum_{k \in E} \mathbb{P}_i[X_n = k] \mathbb{P}[X_{n+1} = j | X_n = k] = \sum_{k \in E} p_{ik}^{(n)} p_{kj}.$$

On the right hand-side we have the scalar product of the i-th row of the matrix $P^{(n)}$ and the j-th column of the matrix P. By definition of the matrix multiplication, this scalar product is exactly the entry of the matrix product $P^{(n)}P$ which is located in the i-th row and j-th column. We thus have the equality of matrices

$$P^{(n+1)} = P^{(n)} P.$$

But now we can apply the induction assumption $P^{(n)} = P^n$ to obtain

$$P^{(n+1)} = P^{(n)} P = P^n \cdot P = P^{n+1}.$$

This completes the induction. □

In the next theorem we consider a Markov chain with initial distribution $\alpha = (\alpha_i)_{i \in E}$ and transition matrix P. Let $\alpha^{(n)} = (\alpha_j^{(n)})_{j \in E}$ be the distribution of the position of this chain at time n, that is

$$\alpha_j^{(n)} = \mathbb{P}[X_n = j].$$

We write both $\alpha^{(n)}$ and α as row vectors. The next theorem states that we can compute $\alpha^{(n)}$ by taking α and multiplying it by the n-step transition matrix $P^{(n)} = P^n$ from the right.

Theorem 2.6.3. *We have*

$$\alpha^{(n)} = \alpha P^n.$$

Proof. By the formula of the total probability

$$\alpha_j^{(n)} = \mathbb{P}[X_n = j] = \sum_{i \in E} \alpha_i \mathbb{P}_i[X_n = j] = \sum_{i \in E} \alpha_i p_{ij}^{(n)}.$$

On the right-hand side we have the scalar product of the row α with the j-th column of $P^{(n)} = P^n$. By definition of matrix multiplication, this means that $\alpha^{(n)} = \alpha P^n$. □

2.7 Invariant measures

Consider a Markov chain on state space E with transition matrix P. Let $\mu : E \to \mathbb{R}$ be a function. To every state $i \in E$ the function assigns some value which will be denoted by $\mu_i := \mu(i)$. Also, it will be convenient to write the function μ as a row vector $\mu = (\mu_i)_{i \in E}$.

Definition 2.7.1. *A function $\mu : E \to \mathbb{R}$ is called a measure on E if $\mu_i \geq 0$ for all $i \in E$.*

Definition 2.7.2. *A function $\mu : E \to \mathbb{R}$ is called a probability measure on E if $\mu_i \geq 0$ for all $i \in E$ and*

$$\sum_{i \in E} \mu_i = 1.$$

Definition 2.7.3. *A measure μ is called invariant if $\mu P = \mu$. That is, for every state $j \in E$ it should hold that*

$$\mu_j = \sum_{i \in E} \mu_i p_{ij}.$$

Remark 7 If the initial distribution α of a Markov chain is invariant, that is $\alpha P = \alpha$, then for every $n \in \mathbb{N}$ we have $\alpha P^n = \alpha$ which means that at every time n the position of the Markov chain has the same distribution as at time 0:

$$X_0 \stackrel{(\mathrm{d})}{=} X_1 \stackrel{(\mathrm{d})}{=} X_2 \stackrel{(\mathrm{d})}{=} \ldots.$$

Example 2.7.4. *Let us compute the invariant distribution for the Markov chain from Example 2.1.1. The transition matrix is*

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

The equation $\mu P = \mu$ for the invariant probability measure takes the following form:

$$(\mu_0, \mu_1) \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} = (\mu_0, \mu_1).$$

Multiplying the matrices we obtain the following two equations:

$$\mu_0(1-p) + \mu_1 q = \mu_0,$$
$$\mu_0 p + \mu_1(1-q) = \mu_1.$$

From the first equation we obtain that $\mu_1 q = \mu_0 p$. Solving the second equation we obtain the same relation which means that the second equation does not contain any information not contained in the first equation. However, since we are looking for invariant probability measures, we have an additional equation

$$\mu_0 + \mu_1 = 1.$$

Solving this equation together with $\mu_1 q = \mu_0 p$ we obtain the following result:

$$\mu_0 = \frac{q}{p+q}, \quad \mu_1 = \frac{p}{p+q}.$$

Problem 2.7.5. *Consider the phone from Example 2.1.1. Let the phone be free at time 0. What is (approximately) the probability that it is free at time $n = 1000$?*

Solution 2.7.6. *The number $n = 1000$ is large. For this reason it seems plausible that the probability that the phone is free (busy) at time $n = 1000$ should be approximately the same as the probability that it is free (busy) at time $n+1 = 1001$. Denoting the initial distribution by $\alpha = (1, 0)$ and the distribution of the position of the chain at time n by $\alpha^{(n)} = \alpha P^n$ we thus must have*

$$\alpha^{(n)} \approx \alpha^{(n+1)} = \alpha P^{n+1} = \alpha P^n \cdot P = \alpha^{(n)} P.$$

Recall that the equation for the invariant probability measure has the same form $\mu = \mu P$. It follows that $\alpha^{(n)}$ must be approximately the invariant probability measure:

$$\alpha^{(n)} \approx \mu.$$

For the probability that the phone is free (busy) at time $n = 1000$ we therefore obtain the approximations

$$p_{00}^{(n)} \approx \mu_0 = \frac{q}{p+q}, \quad p_{01}^{(n)} \approx \mu_1 = \frac{p}{p+q}.$$

Similar considerations apply to the case when the phone is busy at time 0 leading to the approximations

$$p_{10}^{(n)} \approx \mu_0 = \frac{q}{p+q}, \quad p_{11}^{(n)} \approx \mu_1 = \frac{p}{p+q}.$$

Note that $p_{00}^{(n)} \approx p_{10}^{(n)}$ and $p_{01}^{(n)} \approx p_{11}^{(n)}$ which can be interpreted by saying that the Markov chain almost forgets its initial state after many steps. For the n-step transition matrix we therefore may conjecture that

$$\lim_{n\to\infty} P^n = \lim_{n\to\infty} \begin{pmatrix} p_{00}^{(n)} & p_{01}^{(n)} \\ p_{10}^{(n)} & p_{11}^{(n)} \end{pmatrix} = \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_0 & \mu_1 \end{pmatrix}.$$

The above considerations are not rigorous. We will show below that if a general Markov chain satisfies appropriate conditions, then

(i) The invariant probability measure μ exists and is unique.

(ii) For every states $i, j \in E$ we have $\lim_{n\to\infty} p_{ij}^{(n)} = \mu_j$.

Example 2.7.7 (Ehrenfest model)**.** *We consider a box which is divided into 2 parts. Consider N balls (molecules) which are located in this box and can move from one part to the other according to the following rules. Assume that at any moment of time one of the N balls is chosen at random (all balls having the same probability $1/N$ to be chosen). This ball moves to the other part. Then, the procedure is repeated. Let X_n be the number of balls at time n in Part 1. Then, X_n takes values in $E = \{0, 1, \dots, N\}$ which is our state space. The transition probabilities are given by*

$$p_{0,1} = 1, \quad p_{N,N-1} = 1, \quad p_{i,i+1} = \frac{N-i}{N}, \quad p_{i,i-1} = \frac{i}{N}, \quad i = 1, \dots, N-1.$$

For the invariant probability measure we obtain the following system of equations

$$\mu_0 = \frac{\mu_1}{N}, \quad \mu_N = \frac{\mu_{N-1}}{N}, \quad \mu_j = \frac{N-j+1}{N}\mu_{j-1} + \frac{j+1}{N}\mu_{j+1}, \quad j = 1, \dots, N-1.$$

Additionally, we have the equation $\mu_0 + \dots + \mu_N = 1$. This system of equations can be solved directly, but one can also guess the solution without doing computations. Namely, it seems plausible that after a large number of steps every ball will be with probability $1/2$ in Part 1 and with probability $1/2$ in Part 2. Hence, one can guess that the invariant probability measure is the binomial distribution with parameter $1/2$:

$$\mu_j = \frac{1}{2^N}\binom{N}{j}.$$

One can check that this is indeed the unique invariant probability measure for this Markov chain.

Example 2.7.8. *Let $X_0, X_1, \dots$ be independent and identically distributed random variables with values $1, \dots, N$ and corresponding probabilities*

$$\mathbb{P}[X_n = i] = p_i, \quad p_1, \dots, p_N \geq 0, \quad \sum_{i=1}^{N} p_i = 1.$$

Then, $X_0, X_1, \dots$ is a Markov chain and the transition matrix is

$$P = \begin{pmatrix} p_1 & \dots & p_N \\ \dots & \dots & \dots \\ p_1 & \dots & p_N \end{pmatrix}.$$

The invariant probability measure is given by $\mu_1 = p_1, \dots, \mu_N = p_N$.

2.8 Class structure and irreducibility

Consider a Markov chain on a state space E with transition matrix P.

Definition 2.8.1. *We say that state $i \in E$ leads to state $j \in E$ if there exists $n \in \mathbb{N}_0$ such that $p_{ij}^{(n)} \neq 0$. We use the notation $i \rightsquigarrow j$.*

Remark 8 By convention, $p_{ii}^{(0)} = 1$ and hence, every state leads to itself: $i \rightsquigarrow i$.

Theorem 2.8.2. *For two states $i, j \in E$ with $i \neq j$, the following statements are equivalent:*

(i) $i \rightsquigarrow j$.

(ii) $\mathbb{P}_i[\exists n \in \mathbb{N} \colon X_n = j] \neq 0$.

(iii) There exist $n \in \mathbb{N}$ and states $i_1, \ldots, i_{n-1} \in E$ such that $p_{ii_1} \ldots p_{i_{n-1}j} > 0$.

Proof. We prove that Statements 1 and 2 are equivalent. We have the inequality

$$p_{ij}^{(n)} \leq \mathbb{P}_i[\exists n \in \mathbb{N} : X_n = j] \leq \sum_{n=1}^{\infty} \mathbb{P}_i[X_n = j] = \sum_{n=1}^{\infty} p_{ij}^{(n)}. \tag{2.8.1}$$

If Statement 1 holds, then for some $n \in \mathbb{N}$ we have $p_{ij}^{(n)} > 0$. Hence, by (2.8.1), we have $\mathbb{P}_i[\exists n \in \mathbb{N} : X_n = j] > 0$ and Statement 2 holds. If, conversely, Statement 2 holds, then $\mathbb{P}_i[\exists n \in \mathbb{N} : X_n = j] > 0$. Hence, by (2.8.1), $\sum_{n=1}^{\infty} p_{ij}^{(n)} > 0$, which implies that at least one summand $p_{ij}^{(n)}$ must be strictly positive. This proves Statement 1.

We prove the equivalence of Statements 1 and 3. We have the formula

$$p_{ij}^{(n)} = \sum_{i_1, \ldots, i_{n-1} \in E} p_{ii_1} \ldots p_{i_{n-1}j}. \tag{2.8.2}$$

If Statement 1 holds, then for some $n \in \mathbb{N}$ we have $p_{ij}^{(n)} > 0$ which implies that at least one summand on the right-hand side of (2.8.2) must be strictly positive. This implies Statement 3. If, conversely, Statement 3 holds, then the sum on the right-hand side of (2.8.2) is positive which implies that $p_{ij}^{(n)} > 0$. Hence, Statement 1 holds. □

Definition 2.8.3. *States $i, j \in E$ communicate if $i \rightsquigarrow j$ and $j \rightsquigarrow i$. Notation: $i \leftrightsquigarrow j$.*

Theorem 2.8.4. *$i \leftrightsquigarrow j$ is an equivalence relation, namely*

(i) $i \leftrightsquigarrow i$.

(ii) $i \leftrightsquigarrow j \iff j \leftrightsquigarrow i$.

(iii) $i \leftrightsquigarrow j, j \leftrightsquigarrow k \Rightarrow i \leftrightsquigarrow k$.

Proof. Statements 1 and 2 follow from the definition. We prove Statement 3. If $i \leftrightsquigarrow j$ and $j \leftrightsquigarrow k$, then, in particular, $i \rightsquigarrow j$ and $j \rightsquigarrow k$. By Theorem 2.8.2, Statement 3, we can find $r \in \mathbb{N}$, $s \in \mathbb{N}$ and states $u_1, \ldots, u_{r-1} \in E$ and $v_1, \ldots, v_{s-1} \in E$ such that $p_{iu_1}p_{u_1u_2} \ldots p_{u_{r-1}j} > 0$ and $p_{jv_1}p_{v_1v_2} \ldots p_{v_{s-1}k} > 0$. Multiplying both inequalities, we get

$$p_{iu_1}p_{u_1u_2} \ldots p_{u_{r-1}j}p_{jv_1}p_{v_1v_2} \ldots p_{v_{s-1}k} > 0.$$

By Theorem 2.8.2, Statement 3, we have $i \rightsquigarrow k$. In a similar way one shows that $k \rightsquigarrow i$. □

Definition 2.8.5. *The communication class of state $i \in E$ is the set $\{j \in E : i \leftrightsquigarrow j\}$. This set consists of all states j which communicate to i.*

Since communication of states is an equivalence relation, the state space E can be decomposed into a disjoint union of communication classes. Any two communication classes either coincide completely or are disjoint sets.

Definition 2.8.6. *A Markov chain is irreducible if every two states communicate. Hence, an irreducible Markov chain consists of just one communication class.*

Definition 2.8.7. *A communication class C is open if there exist a state $i \in C$ and a state $k \notin C$ such that $i \rightsquigarrow k$. Otherwise, a communication class is called closed.*

If a Markov chain once arrived in a closed communication class, it will stay in this class forever.

Example 2.8.8. *Show that a communication class C is open if and only if there exist a state $i \in C$ and a state $k \notin C$ such that $p_{ik} > 0$.*

Theorem 2.8.9. *If the state space E is a finite set, then there exists at least one closed communication class.*

Proof. We use a proof by contradiction. Assume that there is no closed communication class. Hence, all communication classes are open. Take some state and let C_1 be the communication class of this state. Since C_1 is open, there is a path from C_1 to some other communication class $C_2 \neq C_1$. Since C_2 is open, we can go from C_2 to some other communication class $C_3 \neq C_3$, and so on. Note that in the sequence $C_1, C_2, C_3, \ldots$ all classes are different. Indeed, if for some $l < m$ we would have $C_l = C_m$ (a "cycle"), this would mean that there is a path starting from C_l, going to C_{l+1} and then to $C_m = C_l$. But this is a contradiction since then C_l and C_{l+1} should be a single communication class, and not two different classes, as in the construction. So, the classes $C_1, C_2, \ldots$ are different (in fact, disjoint) and each class contains at least one element. But this is a contradiction since E is a finite set. $\square$

2.9 Aperiodicity

Definition 2.9.1. *The period of a state $i \in E$ is defined as*

$$\gcd\{n \in \mathbb{N} : p_{ii}^{(n)} > 0\}.$$

Here, gcd *states for the greatest common divisor. A state $i \in E$ is called aperiodic if its period is equal to* 1. *Otherwise, the state i is called periodic.*

Example 2.9.2. *Consider a knight on a chessboard moving according to the usual chess rules in a random way. For concreteness, assume that at each moment of time all moves of the knight allowed by the chess rules are counted and then one of these moves is chosen, all moves being equiprobable.*

This is a Markov chain on a state space consisting of 64 *squares. Assume that at time* 0 *the knight is in square* i*. Since the knight changes the color of its square after every move, it cannot return to the original square in an odd number of steps. On the other hand, it can return to* i *in an even number of steps with non-zero probability (for example by going to some other square and then back, many times). So,*

$$p_{ii}^{(2n+1)} = 0, \quad p_{ii}^{(2n)} > 0.$$

Hence, the period of any state in this Markov chain is 2*.*

Example 2.9.3. *Consider a Markov chain on a state space of two elements with transition matrix*

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

We have

$$p_{ii}^{(2n+1)} = 0, \quad p_{ii}^{(2n)} = 1.$$

Hence, the period of any state in this Markov chain is 2*.*

Example 2.9.4. *Show that in the Ehrenfest Markov chain (Example 2.7.7) every state is periodic with period* 2*.*

Lemma 2.9.5. *Let* $i \in E$ *be any state. The following conditions are equivalent:*

(i) *State* i *is aperiodic.*

(ii) *There is* $N \in \mathbb{N}$ *such that for every natural number* $n > N$ *we have* $p_{ii}^{(n)} > 0$*.*

Proof. If Statement 2 holds, then for some sufficiently large n we have $p_{ii}^{(n)} > 0$ and $p_{ii}^{(n+1)} > 0$. Since $\gcd(n, n+1) = 1$, the state i has period 1. Hence, Statement 1 holds.

Suppose, conversely, that Statement 1 holds. Then, we can find $n_1, \ldots, n_r \in \mathbb{N}$ such that $\gcd\{n_1, \ldots, n_r\} = 1$ and $p_{ii}^{(n_1)} > 0, \ldots, p_{ii}^{(n_r)} > 0$. By a result from number theory, the condition $\gcd\{n_1, \ldots, n_r\} = 1$ implies that there is $N \in \mathbb{N}$ such that we can represent any natural number $n > N$ in the form $n = l_1 n_1 + \ldots + l_r n_r$ for suitable $l_1, \ldots, l_r \in \mathbb{N}$. We obtain that

$$p_{ii}^{(l_1 n_1 + \ldots + l_r n_r)} \geq (p_{ii}^{(n_1)})^{l_1} \cdot \ldots \cdot (p_{ii}^{(n_r)})^{l_r} > 0.$$

This proves Statement 2. □

Lemma 2.9.6. *If state* $i \in E$ *is aperiodic and* $i \leftrightsquigarrow j$*, then* j *is also aperiodic.*

Remark 9 We can express this by saying that aperiodicity is a *class property*: If some state in a communication class is aperiodic, then all states in this communication class are aperiodic. Similarly, if some state in a communication class is periodic, then all states in this communication class must be periodic. We can thus divide all communication classes into two categories: the aperiodic communication classes (consisting of only aperiodic states) and the periodic communication classes (consisting only of periodic states).

Definition 2.9.7. *An irreducible Markov chain is called aperiodic if some (and hence, all) states in this chain are aperiodic.*

Proof of Lemma 2.9.6. From $i \leftrightsquigarrow j$ it follows that $i \rightsquigarrow j$ and $j \rightsquigarrow i$. Hence, we can find $r, s \in \mathbb{N}_0$ such that $p_{ji}^{(r)} > 0$ and $p_{ij}^{(s)} > 0$. Since the state i is aperiodic, by Lemma 2.9.5 we can find $N \in \mathbb{N}$ such that for all $n > N$, we have $p_{ii}^{(n)} > 0$ and hence,

$$p_{jj}^{(n+r+s)} \geq p_{ji}^{(r)} \cdot p_{ii}^{(n)} \cdot p_{ij}^{(s)} > 0.$$

It follows that $p_{jj}^{(k)} > 0$ for all $k := n + r + s > N + r + s$. By Lemma 2.9.5, this implies that j is aperiodic. $\square$

2.10 Recurrence and transience

Consider a Markov chain $\{X_n : n \in \mathbb{N}_0\}$ on state space E with transition matrix P.

Definition 2.10.1. *A state $i \in E$ is called recurrent if*

$$\mathbb{P}_i[X_n = i \text{ for infinitely many } n] = 1.$$

Definition 2.10.2. *A state $i \in E$ is called transient if*

$$\mathbb{P}_i[X_n = i \text{ for infinitely many } n] = 0.$$

A recurrent state has the property that a Markov chain starting at this state returns to this state infinitely often, with probability 1. A transient state has the property that a Markov chain starting at this state returns to this state only finitely often, with probability 1.

The next theorem is a characterization of recurrent/transient states.

Theorem 2.10.3. *Let $i \in E$ be a state. Denote by f_i the probability that a Markov chain which starts at i returns to i at least once, that is*

$$f_i = \mathbb{P}_i[\exists n \in \mathbb{N} : X_n = i].$$

Then,

(i) The state i is recurrent if and only if $f_i = 1$.

(ii) The state i is transient if and only if $f_i < 1$.

Corollary 2.10.4. *Every state is either recurrent or transient.*

Proof. For $k \in \mathbb{N}$ consider the random event

$$B_k = \{X_n = i \text{ for at least } k \text{ different values of } n \in \mathbb{N}\}.$$

Then, $\mathbb{P}_i[B_k] = f_i^k$. Also, $B_1 \supset B_2 \supset \dots$. It follows that

$$\mathbb{P}_i[X_n = i \text{ for infinitely many } n] = \mathbb{P}_i[\cap_{k=1}^{\infty} B_k] = \lim_{k\to\infty} \mathbb{P}_i[B_k] = \lim_{k\to\infty} f_i^k = \begin{cases} 1, & \text{if } f_i = 1, \\ 0, & \text{if } f_i < 1. \end{cases}$$

It follows that state i is recurrent if $f_i = 1$ and transient if $f_i < 1$. □

Here is one more characterization of recurrence and transience.

Theorem 2.10.5. *Let $i \in E$ be a state. Recall that $p_{ii}^{(n)} = \mathbb{P}_i[X_n = i]$ denotes the probability that a Markov chain which started at state i visits state i at time n. Then,*

(i) The state i is recurrent if and only if $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$.

(ii) The state i is transient if and only if $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$.

Proof. Let the Markov chain start at state i. Consider the random variable

$$V_i := \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n = i\}}$$

which counts the number of returns of the Markov chain to state i. Note that the random variable V_i can take the value $+\infty$. Then,

$$\mathbb{P}_i[V_i \geq k] = \mathbb{P}[B_k] = f_i^k, \quad k \in \mathbb{N}.$$

Thus, the expectation of V_i can be computed as follows:

$$\mathbb{E}_i[V_i] = \sum_{k=1}^{\infty} \mathbb{P}_i[V_i \geq k] = \sum_{k=1}^{\infty} f_i^k. \tag{2.10.1}$$

On the other hand,

$$\mathbb{E}_i[V_i] = \mathbb{E}_i \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n = i\}} = \sum_{n=1}^{\infty} \mathbb{E}_i \mathbb{1}_{\{X_n = i\}} = \sum_{n=1}^{\infty} p_{ii}^{(n)}. \tag{2.10.2}$$

Case 1. Assume that state i is recurrent. Then, $f_i = 1$ by Theorem 2.10.3. It follows that $\mathbb{E}_i[V_i] = \infty$ by (2.10.1). (In fact, $\mathbb{P}_i[V_i = +\infty] = 1$ since $\mathbb{P}[V_i \geq k] = 1$ for every $k \in \mathbb{N}$). Hence, $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$ by (2.10.2)

Case 2. Assume that state i is transient. Then, $f_i < 1$ by Theorem 2.10.3. Thus, $\mathbb{E}_i V_i < \infty$ by (2.10.1) and hence, $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$ by (2.10.2). □

The next theorem shows that recurrence and transience are class properties: If some state in a communicating class is recurrent (resp. transient), then all states in this class are recurrent (resp. transient).

Theorem 2.10.6. case1. *If $i \in E$ be a recurrent state and $j \leftrightsquigarrow i$, then j is also recurrent.*
case2. *If $i \in E$ be a transient state and $j \leftrightsquigarrow i$, then j is also transient.*

Proof. It suffices to prove Part 2. Let i be a transient state and let $j \leftrightsquigarrow i$. It follows that there exist $s, r \in \mathbb{N}_0$ with $p_{ij}^{(s)} > 0$ and $p_{ji}^{(r)} > 0$. For all $n \in \mathbb{N}$ it holds that

$$p_{ii}^{(n+r+s)} \geq p_{ij}^{(s)} p_{jj}^{(n)} p_{ji}^{(r)}.$$

Therefore,

$$\sum_{n=1}^{\infty} p_{jj}^{(n)} \leq \frac{1}{p_{ij}^{(s)} p_{ji}^{(r)}} \sum_{n=1}^{\infty} p_{ii}^{(n+r+s)} \leq \frac{1}{p_{ij}^{(s)} p_{ji}^{(r)}} \sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty,$$

where the last step holds because i is transient. It follows that state j is also transient. □

Theorem 2.10.6 allows us to introduce the following definitions.

Definition 2.10.7. *A communicating class is called recurrent if at least one (equivalently, every) state in this class is recurrent. A communicating class is transient if at least one (equivalently, every) state in this class is transient.*

Definition 2.10.8. *An irreducible Markov chain is called recurrent if at least one (equivalently, every) state in this chain is recurrent. An irreducible Markov chain is called transient if at least one (equivalently, every) state in this chain is transient.*

The next theorem states that it is impossible to leave a recurrent class.

Theorem 2.10.9. *Every recurrent communicating class is closed.*

Proof. Let C be a non-closed class. We need to show that it is not recurrent. Since C is not closed, there exist states i, j so that $i \in C$, $j \notin C$ and $i \rightsquigarrow j$. This means that there exists $m \in \mathbb{N}$ so that $p_{ij}^{(m)} = \mathbb{P}_i[X_m = j] > 0$. If the event $\{X_m = j\}$ occurs, then after time m the chain cannot return to state i because otherwise i and j would be in the same communicating class. It follows that

$$\mathbb{P}_i[\{X_m = j\} \cap \{X_n = i \text{ for infinitely many } n\}] = 0.$$

This implies that

$$\mathbb{P}_i[X_n = i \text{ for infinitely many } n] < 1.$$

Therefore, state i is not recurrent. □

If some communicating class contains only finitely states and the chain cannot leave this class, then it looks very plausible that the chain which started in some state of this class will return to this state infinitely often (and, in fact, will visit any state of this class infinitely often), with probability 1. This is stated in the next theorem.

Theorem 2.10.10. *Every finite closed communicating class is recurrent.*

Proof. Let C be a closed communicating class with finitely many elements. Take some state $i \in C$. A chain starting in i stays in C forever and since C is finite, there must be at least one state $j \in C$ which is visited infinitely often with positive probability:

$$\mathbb{P}_i[X_n = j \text{ for infinitely many } n \in \mathbb{N}] > 0.$$

At the moment it is not clear whether we can take $i = j$. But since i and j are in the same communicating class, there exists $m \in \mathbb{N}_0$ so that $p_{ji}^{(m)} > 0$. From the inequality

$$\mathbb{P}_j[X_n = j \text{ for infinitely many } n] > p_{ji}^{(m)} \cdot \mathbb{P}_i[X_n = j \text{ for infinitely many } n] > 0$$

it follows that state j is recurrent. The class C is then recurrent because it contains at leats one recurrent state, namely j. □

So, in a Markov chain with finitely many states we have the following equivalencies

(i) A communicating class is recurrent if and only if it is closed.

(ii) A communicating class is transient if and only if it is not closed.

Lemma 2.10.11. *Consider an irreducible, recurrent Markov chain with an arbitrary initial distribution α. Then, for every state $j \in E$ the number of visits of the chain to j is infinite with probability 1.*

Proof. Exercise. □

2.11 Recurrence and transience of random walks

Example 2.11.1. *A simple random walk on $\mathbb{Z}$ is a Markov chain with state space $E = \mathbb{Z}$ and transition probabilities*

$$p_{i,i+1} = p, \quad p_{i,i-1} = 1 - p, \quad i \in \mathbb{Z}.$$

So, from every state the random walk goes one step to the right with probability p, or one step to the left with probability $1 - p$; see Figure 2.1. Here, $p \in [0, 1]$ is a parameter.

Theorem 2.11.2. *If $p = \frac{1}{2}$, then any state of the simple random walk is recurrent. If $p \neq \frac{1}{2}$, then any state is transient.*

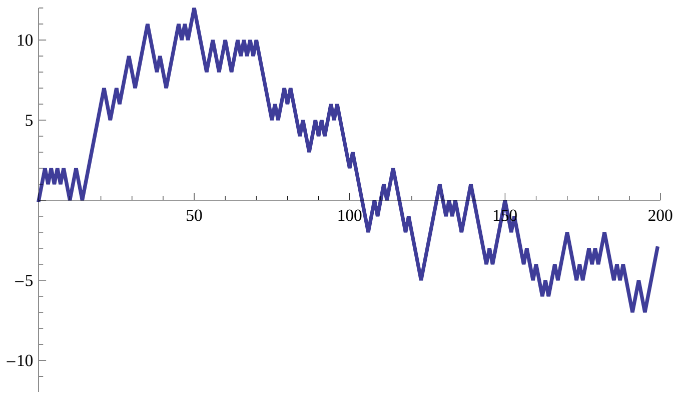

Figure 2.1: Sample path of a simple random walk on $\mathbb{Z}$ with $p = \frac{1}{2}$. The figure shows 200 steps of the walk.

Proof. By translation invariance, we can restrict our attention to state 0. We can represent our Markov chain as $X_n = \xi_1 + \ldots + \xi_n$, where $\xi_1, \xi_2, \ldots$ are independent and identically distributed random variables with Bernoulli distribution:

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

Case 1. Let $p \neq \frac{1}{2}$. Then, $\mathbb{E}\xi_k = p - (1-p) = 2p - 1 \neq 0$. By the strong law of large numbers,

$$\lim_{n\to\infty} \frac{1}{n} X_n = \lim_{n\to\infty} \frac{\xi_1 + \ldots + \xi_n}{n} = \mathbb{E}\xi_1 \neq 0 \quad \text{a.s.}$$

In the case $p > \frac{1}{2}$ we have $\mathbb{E}\xi_1 > 0$ and hence, $\lim_{n\to\infty} X_n = +\infty$ a.s. In the case $p < \frac{1}{2}$ we have $\mathbb{E}\xi_1 < 0$ and hence, $\lim_{n\to\infty} X_n = -\infty$ a.s. In both cases it follows that

$$\mathbb{P}[X_n = 0 \text{ for infinitely many } n] = 0.$$

Hence, state 0 is transient.

Case 2. Let $p = \frac{1}{2}$. In this case, $\mathbb{E}\xi_k = 0$ and the argument of Case 1 does not work. We will use Theorem 2.10.5. The n-step transition probability from 0 to 0 is given by

$$p_{00}^{(n)} = \begin{cases} 0, & \text{if } n = 2k+1 \text{ odd}, \\ \frac{1}{2^{2k}}\binom{2k}{k}, & \text{if } n = 2k \text{ even}. \end{cases}$$

The Stirling formula $n! \sim \sqrt{2\pi n}(\frac{n}{e})^n$, as $n \to \infty$, yields that

$$p_{00}^{(2k)} \sim \frac{1}{\sqrt{\pi k}}, \quad \text{as } k \to \infty.$$

Since the series $\sum_{k=1}^{\infty} \frac{1}{\sqrt{k}}$ diverges, it follows that $\sum_{n=1}^{\infty} p_{00}^{(n)} = \sum_{k=1}^{\infty} p_{00}^{(2k)} = \infty$. By Theorem 2.10.5, this implies that 0 is a recurrent state. $\square$

Example 2.11.3. *The simple, symmetric random walk on $\mathbb{Z}^d$ is a Markov chain defined as follows. The state space is the d-dimensional lattice*

$$\mathbb{Z}^d = \{(n_1, \dots, n_d) : n_1, \dots, n_d \in \mathbb{Z}\}.$$

Let $e_1, \dots, e_d$ be the standard basis of $\mathbb{R}^d$, that is

$$e_1 = (1, 0, 0 \dots, 0), \quad e_2 = (0, 1, 0, \dots, 0), \quad e_3 = (0, 0, 1, \dots, 0), \quad \dots, \quad e_d = (0, 0, 0, \dots, 1).$$

Let $\xi_1, \xi_2, \dots$ be independent and identically distributed d-dimensional random vectors such that

$$\mathbb{P}[\xi_i = e_k] = \mathbb{P}[\xi_i = -e_k] = \frac{1}{2d}, \quad k = 1, \dots, d, \quad i \in \mathbb{N}.$$

Define $S_n = \xi_1 + \dots + \xi_n$, $n \in \mathbb{N}$, and $S_0 = 0$. The sequence $S_0, S_1, S_2, \dots$ is called the simple symmetric random walk on $\mathbb{Z}^d$. It is a Markov chain with transition probabilities

$$p_{i,i+e_1} = p_{i,i-e_1} = \dots = p_{i,i+e_d} = p_{i,i-e_d} = \frac{1}{2d}, \quad i \in \mathbb{Z}^d.$$

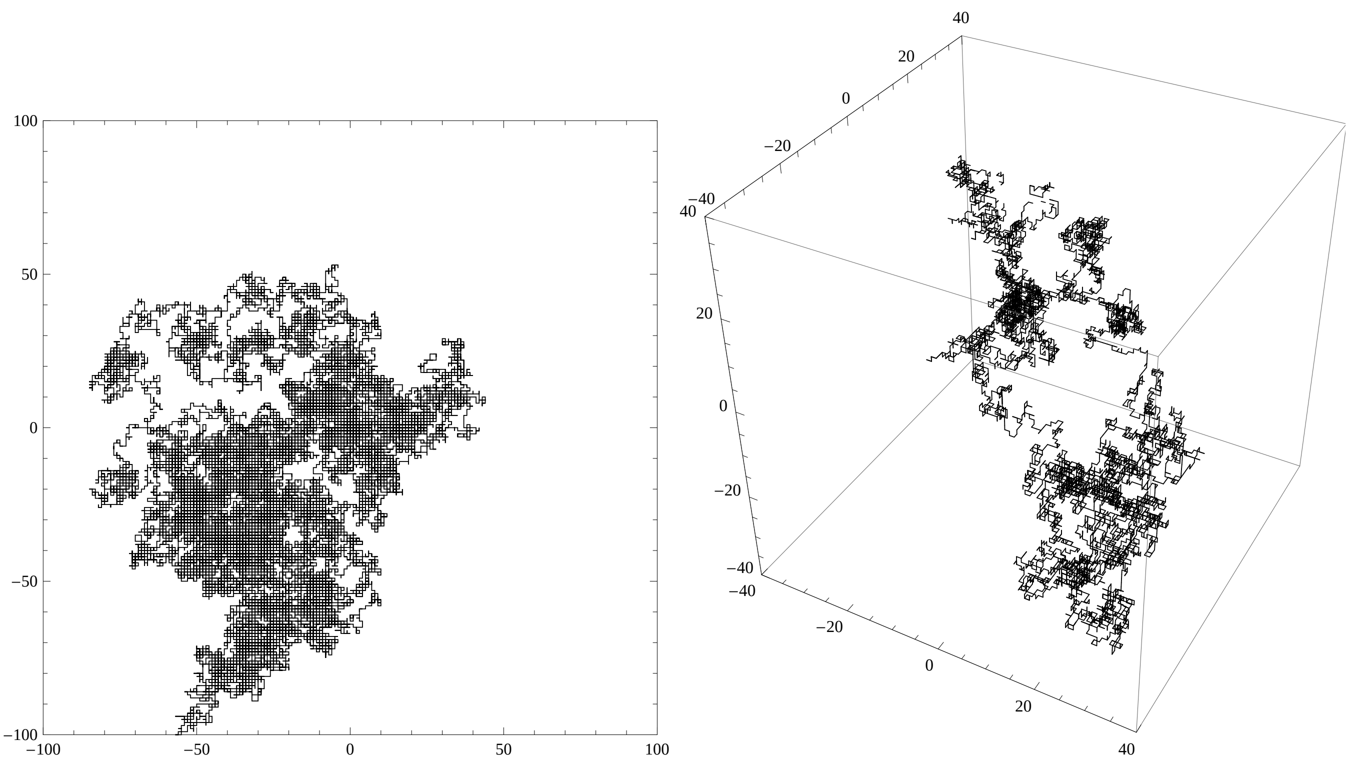

Figure 2.2: *Left:* Sample path of a simple symmetric random walk on $\mathbb{Z}^2$. *Right:* Sample path of a simple symmetric random walk on $\mathbb{Z}^3$. In both cases the random walk makes 50000 steps.

Theorem 2.11.4 (Pólya, 1921)**.** *The simple symmetric random walk on $\mathbb{Z}^d$ is recurrent if and only if $d = 1, 2$ and transient if and only if $d \geq 3$.*

Proof. For $d = 1$ we already proved the statement in Theorem 2.11.2.

Consider the case $d = 2$. We compute the n-step transition probability $p_{00}^{(n)}$. For an odd n this probability is 0. For an even $n = 2k$ we have

$$p_{00}^{(2k)} = \frac{1}{4^{2k}} \sum_{i=0}^{k} \binom{2k}{i, i, k-i, k-i} = \frac{1}{4^{2k}} \binom{2k}{k} \sum_{i=0}^{k} \binom{k}{i} \binom{k}{k-i} = \left(\frac{1}{2^{2k}} \binom{2k}{k} \right)^2 \sim \frac{1}{\pi k},$$

as $k \to \infty$, where the last step is by the Stirling formula. The harmonic series $\sum_{k=1}^{\infty} \frac{1}{k}$ diverges. Therefore, $\sum_{n=1}^{\infty} p_{00}^{(n)} = \infty$ and the random walk is recurrent in $d = 2$ dimensions.

Generalizing the cases $d = 1, 2$ one can show that for an arbitrary dimension $d \in \mathbb{N}$ we have, as $k \to \infty$,

$$p_{00}^{(2k)} \sim \frac{1}{(\pi k)^{d/2}}.$$

Since the series $\sum_{k=1}^{\infty} k^{-d/2}$ is convergent for $d \geq 3$ it holds that $\sum_{n=1}^{\infty} p_{00}^{(n)} < \infty$ and the random walk is transient in $d = 3$ dimensions. □

2.12 Existence and uniqueness of the invariant measure

The next two theorems state that any irreducible and recurrent Markov chain has a unique invariant measure μ, up to a multiplication by a constant. This measure may be finite (that is, $\sum_{i \in E} \mu_i < +\infty$) or infinite (that is, $\sum_{i \in E} \mu_i = +\infty$).

First we provide an explicit construction of an invariant measure for an irreducible and recurrent Markov chain. Consider a Markov chain starting at state $k \in E$. Denote the time of the first return to k by

$$T_k = \min\{n \in \mathbb{N} : X_n = k\} \in \mathbb{N} \cup \{+\infty\}.$$

The minimum of an empty set is by convention $+\infty$. For a state $i \in E$ denote the expected number of visits to i before the first return to k by

$$\gamma_i = \gamma_i^{(k)} = \mathbb{E}_k \sum_{n=0}^{T_k - 1} \mathbb{1}_{\{X_n = i\}} \in [0, +\infty].$$

Theorem 2.12.1. *For an irreducible and recurrent Markov chain starting at state $k \in E$ we have*

(i) $\gamma_k = 1$.

(ii) *For all $i \in E$ it holds that $0 < \gamma_i < \infty$.*

(iii) $\gamma = (\gamma_i)_{i \in E}$ *is an invariant measure.*

Proof. We proceed in three steps.

Step 1. We show that $\gamma_k = 1$. By definition of T_k, we have $\sum_{n=0}^{T_k-1} \mathbb{1}_{\{X_n=k\}} = 1$, if the chain starts at k. It follows that $\gamma_k = \mathbb{E}_k 1 = 1$.

Step 2. We show that for every state $j \in E$,

$$\gamma_j = \sum_{i \in E} p_{ij} \gamma_i. \tag{2.12.1}$$

(At this moment, both sides of (2.12.1) are allowed to be infinite, but in Step 3 we will show that both sides are actually finite). The Markov chain is recurrent, thus $T_k < \infty$ almost surely. By definition, $X_{T_k} = k = X_0$. We have

$$\gamma_j = \mathbb{E}_k \sum_{n=1}^{T_k} \mathbb{1}_{\{X_n=j\}} = \mathbb{E}_k \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n=j, n \le T_k\}} = \sum_{n=1}^{\infty} \mathbb{P}_k[X_n = j, T_k \ge n].$$

Before visiting state j at time n the chain must have been in some state i at time $n-1$, where $i \in E$ can be, in general, arbitrary. We obtain that

$$\gamma_j = \sum_{i \in E} \sum_{n=1}^{\infty} \mathbb{P}_k[X_n = j, X_{n-1} = i, T_k \ge n] = \sum_{i \in E} \sum_{n=1}^{\infty} p_{ij} \mathbb{P}_k[X_{n-1} = i, T_k \ge n].$$

Introducing the new summation variable $m = n - 1$, we obtain that

$$\gamma_j = \sum_{i \in E} p_{ij} \sum_{m=0}^{\infty} \mathbb{E}_k \mathbb{1}_{\{X_m=i, T_k \ge m+1\}} = \sum_{i \in E} p_{ij}\, \mathbb{E}_k \sum_{m=0}^{T_k-1} \mathbb{1}_{\{X_m=i\}} = \sum_{i \in E} p_{ij} \gamma_i.$$

This proves that (2.12.1) holds.

Step 3. Let $i \in E$ be an arbitrary state. We show that $0 < \gamma_i < \infty$. Since the chain is irreducible, there exist $n, m \in \mathbb{N}_0$ such that $p_{ik}^{(m)} > 0$ and $p_{ki}^{(n)} > 0$. From (2.12.1) it follows that

$$\gamma_i = \sum_{l \in E} p_{li}^{(n)} \gamma_l \ge p_{ki}^{(n)} \gamma_k = p_{ki}^{(n)} > 0.$$

On the other hand, again using (2.12.1), we obtain that

$$1 = \gamma_k = \sum_{l \in E} p_{lk}^{(m)} \gamma_l \ge p_{ik}^{(m)} \gamma_i.$$

This implies that $\gamma_i \le 1/p_{ik}^{(m)} < \infty$. □

The next theorem states the uniqueness of the invariant measure, up to multiplication by a constant.

Theorem 2.12.2. *Consider an irreducible and recurrent Markov chain and fix some state $k \in E$. Then, every invariant measure μ can be represented in the form*

$$\mu_j = c\gamma_j^{(k)} \quad \forall j \in E,$$

where c is a constant (not depending on j). In fact, $c = \mu_k$.

Remark 10 Hence, the invariant measure is unique up to a multiplication by a constant. In particular, the invariant measures $(\gamma_i^{(k_1)})_{i\in E}$ and $(\gamma_i^{(k_2)})_{i\in E}$, for different states $k_1, k_2 \in E$, differ by a multiplicative constant.

Proof. Let μ be an invariant measure.

Step 1. We show that $\mu_j \geq \mu_k \gamma_j^{(k)}$ for all $j \in E$. We will *not* use the irreducibility and the recurrence of the chain in this step. The invariance of the measure μ implies that

$$\mu_j = \sum_{i_0 \in E} \mu_{i_0} p_{i_0 j} = \sum_{i_0 \neq k} \mu_{i_0} p_{i_0 j} + \mu_k p_{kj}.$$

Applying the same procedure to μ_{i_0}, we obtain

$$\begin{aligned} \mu_j &= \sum_{i_0 \neq k} \left(\sum_{i_1 \neq k} \mu_{i_1} p_{i_1 i_0} + \mu_k p_{k i_0} \right) p_{i_0 j} + \mu_k p_{kj} \\ &= \sum_{i_0 \neq k} \sum_{i_1 \neq k} \mu_{i_1} p_{i_1 i_0} p_{i_0 j} + \left(\mu_k p_{kj} + \mu_k \sum_{i_0 \neq k} p_{k i_0} p_{i_0 j} \right). \end{aligned}$$

Applying the procedure to μ_{i_1} and repeating it over and over again we obtain that for every $n \in \mathbb{N}$,

$$\mu_j = \sum_{i_0, i_1, \ldots, i_n \neq k} \mu_{i_n} p_{i_n i_{n-1}} \ldots p_{i_1 i_0} p_{i_0 j} + \mu_k \left(p_{kj} + \sum_{i_0 \neq k} p_{k i_0} p_{i_0 j} + \ldots + \sum_{i_0, \ldots, i_{n-1} \neq k} p_{k i_0} p_{i_0 i_1} \ldots p_{i_{n-1} j} \right).$$

Noting that the first term is non-negative, we obtain that

$$\mu_j \geq 0 + \mu_k \mathbb{P}_k[X_1 = j, T_k \geq 1] + \mu_k \mathbb{P}_k[X_2 = j, T_k \geq 2] + \ldots + \mu_k \mathbb{P}_k[X_n = j, T_k \geq n].$$

Since this holds for every $n \in \mathbb{N}$, we can pass to the limit as $n \to \infty$:

$$\mu_j \geq \mu_k \sum_{n=1}^{\infty} \mathbb{P}_k[X_n = j, T_k \geq n] = \mu_k \gamma_j^{(k)}.$$

It follows that $\mu_j \geq \mu_k \gamma_j^{(k)}$.

Step 2. We prove the converse inequality. Consider $\mu_j := \mu_j - \mu_k \gamma_j^{(k)}$, $j \in E$. By the above, $\mu_j \geq 0$ for all $j \geq 0$ so that $\mu = (\mu_j)_{j \in E}$ is a measure. Moreover, this measure is invariant because it is a linear combination of two invariant measures. Finally, note that by definition, $\mu_k = 0$. We will prove that this implies that $\mu_j = 0$ for all $j \in E$. By the irreducibility of our Markov chain, for every $j \in E$ we can find $n \in \mathbb{N}_0$ such that $p_{jk}^{(n)} > 0$. By the invariance property of μ,

$$0 = \mu_k = \sum_{i \in E} \mu_i p_{ik}^{(n)} \geq \mu_j p_{jk}^{(n)}.$$

It follows that $\mu_j p_{jk}^{(n)} = 0$ but since $p_{jk}^{(n)} > 0$, we must have $\mu_j = 0$. By the definition of μ_j this implies that $\mu_j = \mu_k \gamma_j^{(k)}$. $\square$

We can now summarize Theorems 2.12.1 and 2.12.2 as follows:

Theorem 2.12.3. *A recurrent, irreducible Markov chain has unique (up to a constant multiple) invariant measure.*

This invariant measure may be finite or infinite. However, if the Markov chain has only finitely many states, then the measure must be finite and we can even normalize it to be a probability measure.

Corollary 2.12.4. *A finite and irreducible Markov chain has a unique invariant probability measure.*

Proof. A finite and irreducible Markov chain is recurrent by Theorem 2.10.10. By Theorem 2.12.1, there exists an invariant measure $\mu = (\mu_i)_{i \in E}$. Since the number of states in E is finite by assumption and $\mu_i < \infty$ by Theorem 2.12.1, we have $M := \sum_{i \in E} \mu_i < \infty$ and hence, the measure μ is finite. To obtain an invariant *probability* measure, consider the measure $\mu_i' = \mu_i / M$.

To show that the invariant probability measure is unique, assume that we have two invariant probability measures $\nu' = (\nu_i')_{i \in E}$ and $\nu'' = (\nu_i'')_{i \in E}$. Take an arbitrary state $k \in E$. By Theorem 2.12.2, there are constants c' and c'' such that $\nu_i' = c' \gamma_i^{(k)}$ and $\nu_i'' = c'' \gamma_i^{(k)}$, for all $i \in E$. But since both ν' and ν'' are *probability* measures, we have

$$1 = \sum_{i \in E} \nu_i' = c' \sum_{i \in E} \gamma_i^{(k)}, \quad 1 = \sum_{i \in E} \nu_i'' = c'' \sum_{i \in E} \gamma_i^{(k)}.$$

This implies that $c' = c''$ and hence, the measures ν' and ν'' are equal. $\square$

Above, we considered only irreducible, recurrent chains. What happens if the chain is irreducible and transient? It turns out that in this case everything is possible:

(i) It is possible that there is no invariant measure at all (except the zero measure).

(ii) It is possible that there is a unique (up to multiplication by a constant) invariant measure.

(iii) It is possible that there are at least two invariant measures which are not constant multiples of each other.

Exercise 2.12.5. *Consider a Markov chain on $\mathbb{N}$ with transition probabilities $p_{i,i+1} = 1$, for all $i \in \mathbb{N}$. Show that the only invariant measure is $\mu_i = 0$, $i \in \mathbb{N}$.*

Exercise 2.12.6. *Consider a Markov chain on $\mathbb{Z}$ with transition probabilities $p_{i,i+1} = 1$, for all $i \in \mathbb{Z}$. Show that the invariant measures have the form $\mu_i = c$, $i \in \mathbb{Z}$, where $c \geq 0$ is constant.*

Exercise 2.12.7. *Consider a simple random walk on $\mathbb{Z}$ with $p \neq \frac{1}{2}$. Show that any invariant measure has the form*

$$\mu_i = c_1 + c_2 \left(\frac{p}{1-p} \right)^i, \quad i \in \mathbb{Z},$$

for some constants $c_1 \geq 0$, $c_2 \geq 0$.

2.13 Positive recurrence and null recurrence

The set of recurrent states of a Markov chain can be further subdivided into the set of positive recurrent states and the set of negative recurrent states. Let us define the notions of positive recurrence and null recurrence.

Consider a Markov chain on state space E. Take some state $i \in E$, assume that the Markov chain starts at state i and denote by T_i the time of the first return of the chain to state i:

$$T_i = \min\{n \in \mathbb{N} : X_n = i\} \in \mathbb{N} \cup \{+\infty\}.$$

Denote by m_i the expected return time of the chain to state i, that is

$$m_i = \mathbb{E}_i T_i \in (0, \infty]$$

Note that for a transient state i we always have $m_i = +\infty$ because the random variable T_i takes the value $+\infty$ with strictly positive probability $1 - f_i > 0$, see Theorem 2.10.3. However, for a recurrent state i the value of m_i may be both finite and infinite, as we shall see later.

Definition 2.13.1. *A state $i \in E$ as called positive recurrent if $m_i < \infty$.*

Definition 2.13.2. *A state $i \in E$ is called null recurrent if it is recurrent and $m_i = +\infty$.*

Remark 11 Both null recurrent states and positive recurrent states are recurrent. For null recurrent states this is required by definition. For a positive recurrent state we have $m_i < \infty$ which means that T_i cannot attain the value $+\infty$ with strictly positive probability and hence, state i is recurrent.

Theorem 2.13.3. *Consider an irreducible Markov chain. Then the following statements are equivalent:*

(i) Some state is positive recurrent.

(ii) All states are positive recurrent.

(iii) The chain has invariant **probability** *measure $\mu = (\mu_i)_{i\in E}$.*

Also, if these statements hold, then $m_i = \frac{1}{\mu_i}$ for all $i \in E$.

Proof. The implication $2 \Rightarrow 1$ is evident.

Proof of $1 \Rightarrow 3$**.** Let $k \in E$ be a positive recurrent state. Then, k is recurrent and all states of the chain are recurrent by irreducibility. By Theorem 2.12.1, $(\gamma_i^{(k)})_{i\in E}$ is an invariant measure. However, we need an invariant *probability* measure. To construct it, note that

$$\sum_{j\in E} \gamma_j^{(k)} = m_k < \infty$$

(since k is positive recurrent). We can therefore define $\mu_i = \gamma_i^{(k)}/m_k$, $i \in E$. Then, $\sum_{i\in E}\mu_i = 1$, and $(\mu_i)_{i\in E}$ is an invariant probability measure.

Proof of $3 \Rightarrow 2$**.** Let $(\mu_i)_{i\in E}$ be an invariant probability measure. First we show that $\mu_k > 0$ for every state $k \in E$. Since μ is a probability measure, we have $\mu_l > 0$ for at least one $l \in E$. By irreducibility, we have $p_{lk}^{(n)} > 0$ for some $n \in \mathbb{N}_0$ and by invariance of μ, we have

$$\mu_k = \sum_{i\in E} p_{ik}^{(n)}\mu_i \geq p_{lk}^{(n)}\mu_l > 0.$$

This proves that $\mu_k > 0$ for every $k \in E$.

By Step 1 from the proof of Theorem 2.12.2 (note that this step does not use recurrence), we have for all $j \in E$,

$$\mu_i \geq \mu_k \gamma_i^{(k)}.$$

Hence,

$$m_k = \sum_{i\in E}\gamma_i^{(k)} \leq \sum_{i\in E}\frac{\mu_i}{\mu_k} = \frac{1}{\mu_k} < \infty.$$

It follows that k is positive recurrent, thus establishing statement 2.

Proof that $m_k = \frac{1}{\mu_k}$**.** Assume that statements 1,2,3 hold. In particular, the chain is recurrent and by Theorem 2.12.2, we must have $\mu_i = \mu_k\gamma_i^{(k)}$ for all $i \in E$. It follows that

$$m_k = \sum_{i\in E}\gamma_i^{(k)} = \sum_{i\in E}\frac{\mu_i}{\mu_k} = \frac{1}{\mu_k},$$

thus proving the required formula. □

Example 2.13.4. *Any state in a finite irreducible Markov chain is positive recurrent. Indeed, such a chain has an invariant probability measure by Corollary 2.12.4.*

Example 2.13.5. *Consider a simple symmetric random walk on $\mathbb{Z}$ or on $\mathbb{Z}^2$. This chain is irreducible. Any state is recurrent by Pólya's Theorem 2.11.4. We show that in fact, any state is null recurrent. To see this, note that the measure assigning the value 1 to every state $i \in E$ is invariant by the definition of the chain. By Theorem 2.12.2, any other invariant measure must be of the form $\mu_i = c$, $i \in E$, for some constant $c \geq 0$. However, no measure of this form is a probability measure. So, there is no invariant probability measure and by Theorem 2.13.3, all states must be null recurrent.*

2.14 Convergence to the invariant probability measure

We are going to state and prove a "strong law of large numbers" for Markov chains. First recall that the usual strong law of large numbers states that if $\xi_1, \xi_2, \dots$ are i.i.d. random variables with $\mathbb{E}|\xi_1| < \infty$, then

$$\frac{\xi_1 + \dots + \xi_n}{n} \overset{\text{a.s.}}{\to} \mathbb{E}\xi_1. \tag{2.14.1}$$

The statement is not applicable if $\mathbb{E}|\xi_1| = \infty$. However, it follows (e.g. from the monotone convergence theorem) that if $\xi_1, \xi_2, \dots$ are i.i.d. random variables which are a.s. nonnegative with $\mathbb{E}\xi_1 = +\infty$, then

$$\frac{\xi_1 + \dots + \xi_n}{n} \overset{\text{a.s.}}{\to} +\infty. \tag{2.14.2}$$

Consider a Markov chain $\{X_n : n \in \mathbb{N}_0\}$ with initial distribution $\alpha = (\alpha_i)_{i \in E}$. Given a state $i \in E$, denote the number of visits to state i in the first n steps by

$$V_n(i) = \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k = i\}}.$$

Theorem 2.14.1. *Consider an irreducible Markov chain $\{X_n : n \in \mathbb{N}_0\}$ with an arbitrary initial distribution $\alpha = (\alpha_i)_{i \in E}$.*

case 1. *If the Markov chain is transient or null recurrent, then for all $i \in E$ it holds that*

$$\frac{V_n(i)}{n} \to 0 \quad \text{a.s.} \tag{2.14.3}$$

case 2. *If the Markov chain is positive recurrent with invariant probability measure μ, then for all $i \in E$ it holds that*

$$\frac{V_n(i)}{n} \to \mu_i \quad \text{a.s.} \tag{2.14.4}$$

Proof. If the chain is transient, then $V_i(n)$ stays bounded as a function of n, with probability 1. This implies (2.14.3). In the sequel, let the chain be recurrent.

For simplicity, we will assume in this proof that the chain starts in state i. Denote the time of the k-th visit of the chain to i by S_k, that is

$$\begin{aligned} S_1 &= \min\{n \in \mathbb{N} : X_n = i\}, \\ S_2 &= \min\{n > S_1 : X_n = i\}, \\ S_3 &= \min\{n > S_2 : X_n = i\}, \end{aligned}$$

and so on. Note that $S_1, S_2, S_3, \dots$ are a.s. finite by the recurrence of the chain. Let also $\xi_1, \xi_2, \xi_3, \dots$ be the excursion times between the returns to i, that is

$$\xi_1 = S_1, \quad \xi_2 = S_2 - S_1, \quad \xi_3 = S_3 - S_2, \quad \dots.$$

Then, $\xi_1, \xi_2, \xi_3, \ldots$ are i.i.d. random variables by the Markov property.

By definition of $V_i(n)$ we have

$$\xi_1 + \xi_2 + \ldots + \xi_{V_i(n)-1} \leq n \leq \xi_1 + \xi_2 + \ldots + \xi_{V_n(i)}.$$

Dividing this by $V_i(n)$ we get

$$\frac{\xi_1 + \xi_2 + \ldots + \xi_{V_n(i)-1}}{V_n(i)} \leq \frac{n}{V_i(n)} \leq \frac{\xi_1 + \xi_2 + \ldots + \xi_{V_n(i)}}{V_n(i)}. \tag{2.14.5}$$

Note that by recurrence, $V_n(i) \to \infty$ a.s.

Case 1. Let the chain be *null* recurrent. It follows that $\mathbb{E}\xi_1 = \infty$. By using (2.14.2) and (2.14.5), we obtain that

$$\frac{n}{V_n(i)} \to \infty \qquad \text{a.s.}$$

This proves (2.14.3).

Case 2. Let the chain be *positive* recurrent. Then, by Theorem 2.13.3, $\mathbb{E}\xi_1 = m_i = \frac{1}{\mu_i} < \infty$. Using (2.14.1) and (2.14.5) we obtain that

$$\frac{n}{V_n(i)} \to \frac{1}{\mu_i} \qquad \text{a.s.}$$

This proves (2.14.4). □

In the next theorem we prove that the n-step transition probabilities converge, as $n \to \infty$, to the invariant probability measure.

Theorem 2.14.2. *Consider an irreducible, aperiodic, positive recurrent Markov chain $\{X_n : n \in \mathbb{N}_0\}$ with transition matrix P and invariant probability measure $\mu = (\mu_i)_{i\in E}$. The initial distribution $\alpha = (\alpha_i)_{i\in E}$ may be arbitrary. Then, for all $j \in E$ it holds that*

$$\lim_{n\to\infty} \mathbb{P}[X_n = j] = \mu_j.$$

In particular, $\lim_{n\to\infty} p_{ij}^{(n)} = \mu_j$ for all $i, j \in E$.

Remark 12 In particular, the theorem applies to any irreducible and aperiodic Markov chain with finite state space.

For the proof we need the following lemma.

Lemma 2.14.3. *Consider an irreducible and aperiodic Markov chain. Then, for every states $i, j \in E$ we can find $N = N(i, j) \in \mathbb{N}$ such that for all $n > N$ we have $p_{ij}^{(n)} > 0$.*

Proof. The chain is irreducible, hence we can find $r \in \mathbb{N}_0$ such that $p_{ij}^{(r)} > 0$. Also, the chain is aperiodic, hence we can find $N_0 \in \mathbb{N}$ such that for all $k > N_0$ we have $p_{ii}^{(k)} > 0$. It follows that for all $k > N_0$,

$$p_{ij}^{(k+r)} > p_{ii}^{(k)} p_{ij}^{(r)} > 0.$$

It follows that for every $n := k + r$ such that $n > N_0 + r$, we have $p_{ij}^{(n)} > 0$. □

Proof of Theorem 2.14.2. We use the "coupling method".

Step 1. Consider *two* Markov chains called $\{X_n : n \in \mathbb{N}_0\}$ and $\{Y_n : n \in \mathbb{N}_0\}$ such that

(i) X_n is a Markov chain with initial distribution α and transition matrix P.

(ii) Y_n is a Markov chain with initial distribution μ (the invariant probability measure) and the same transition matrix P.

(iii) The process $\{X_n : n \in \mathbb{N}_0\}$ is independent of the process $\{Y_n : n \in \mathbb{N}_0\}$.

Note that both Markov chains have the same transition matrix but different initial distributions. Fix an arbitrary state $b \in E$. Denote by T be the time at which the chains meet at state b:

$$T = \min\{n \in \mathbb{N} : X_n = Y_n = b\} \in \mathbb{N} \cup \{+\infty\}.$$

If the chains do not meet at b, we set $T = +\infty$.

Step 2. We show that $\mathbb{P}[T < \infty] = 1$. Consider the stochastic process $W_n = (X_n, Y_n)$ taking values in $E \times E$. It is a Markov chain on $E \times E$ with transition probabilities given by

$$\tilde{p}_{(i,k),(j,l)} = p_{ij} p_{kl}, \quad (i,k) \in E \times E, \quad (j,l) \in E \times E.$$

The initial distribution of W_0 is given by

$$\mu_{(i,k)} = \alpha_i \mu_k, \quad (i,k) \in E \times E.$$

Since the chains X_n and Y_n are aperiodic and irreducible by assumption of the theorem, we can apply Lemma 2.14.3 to obtain for every $i, j, k, l \in E$ a number $N = N(i, j, k, l) \in \mathbb{N}$ such that for all $n > N$ we have

$$\tilde{p}_{(i,k),(j,e)}^{(n)} = p_{ij}^{(n)} p_{ke}^{(n)} > 0.$$

Thus, the chain W_n is irreducible. Also, it is an exercise to check that the probability measure $\tilde{\mu}_{(i,k)} := \mu_i \mu_k$ is invariant for W_n. Thus, by Theorem 2.13.3, the Markov chain W_n is positive recurrent and thereby recurrent. Therefore, $T < \infty$ a.s. by Lemma 2.10.11.

Step 3. Define the stochastic process $\{Z_n : n \in \mathbb{N}_0\}$ by

$$Z_n = \begin{cases} X_n, & \text{if } n \le T, \\ Y_n, & \text{if } n \ge T. \end{cases}$$

Then, Z_n is a Markov chain with initial distribution α and the same transition matrix P as X_n and Y_n. (The Markov chain Z_n is called the coupling of X_n and Y_n). The chain Y_n starts with the invariant probability measure μ and hence, at every time n, Y_n is distributed according to μ. Also, the chain Z_n has the same initial distribution α and the same transition matrix P as the chain X_n, so that in particular, the random elements X_n and Z_n have the same distribution at every time n. Using these facts, we obtain that

$$|\mathbb{P}[X_n = j] - \mu_j| = |\mathbb{P}[X_n = j] - \mathbb{P}[Y_n = j]| = |\mathbb{P}[Z_n = j] - \mathbb{P}[Y_n = j]|.$$

By definition of Z_n, we can rewrite this as

$$\begin{aligned}
|\mathbb{P}[X_n = j] - \mu_j| &= |\mathbb{P}[X_n = j, n < T] + \mathbb{P}[Y_n = j, n \geq T] - \mathbb{P}[Y_n = j]| \\
&= |\mathbb{P}[X_n = j, n < T] - \mathbb{P}[Y_n = j, n < T]| \\
&\leq \mathbb{P}[T > n].
\end{aligned}$$

But we have shown in Step 2 that $\mathbb{P}[T = \infty] = 0$, hence $\lim_{n\to\infty} \mathbb{P}[T > n] = 0$. It follows that

$$\lim_{n\to\infty} \mathbb{P}[X_n = j] = \mu_j,$$

thus establishing the theorem. □

Appendix

2.15 Conditional expectation.

Let us recall the following consequence of the Radon-Nikodym theorem.

Theorem 2.15.1 (Existence and uniqueness). *Let $X \in L^1(\mathbb{P})$ be defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. Then there exists a random variable Y which is written as $Y = \mathbb{E}(X|\mathcal{G})$ such that the following two properties hold:*

- *Y is $\mathcal{G}$-measurable.*
- *For any $G \in \mathcal{G}$, $\int_G X \mathrm{d}\mathbb{P} = \int_G Y \mathrm{d}\mathbb{P}$ almost surely.*

Such a random variable Y, which is called the conditional expectation of X w.r.t. $\mathcal{G}$, is unique up to a set of measure zero.

It is useful to collect the following properties of conditional expectation:

Corollary 2.15.2 (Properties of conditional expectation). *In what follows $X, X', X_n \in L^1(\mathbb{P})$ for each $n \in \mathbb{N}_0$ and $\mathcal{G} \subset \mathcal{F}$ is a sub-σ-algebra of $\mathcal{F}$ in the ambient probability space $(\Omega, \mathcal{F}, \mathbb{P})$.*

(i) (Linearity and monotonicity.) For any $c_1, c_2 \in \mathbb{R}$, we have $\mathbb{E}[c_1 X_1 + c_2 X_2|\mathcal{G}] = c_1\mathbb{E}[X_1|\mathcal{G}] + c_2\mathbb{E}[X_2|\mathcal{G}]$ almost surely. Moreover, if $X \geq X'$ almost surely, then $\mathbb{E}[X|\mathcal{G}] \geq \mathbb{E}[X'|\mathcal{G}]$ almost surely.

(ii) (Taking out what is known.) If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X|\mathcal{G}] = X$ almost surely.

(iii) (The law of total expectation) We always have $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]]$.

(iv) (Tower property) If $\mathcal{G}_1 \subset \mathcal{G}_2$, both being sub-$\sigma$ algebras of $\mathcal{F}$, then $\mathbb{E}\big[\mathbb{E}[X|\mathcal{G}_1]|\mathcal{G}_2\big] = \mathbb{E}\big[\mathbb{E}[X|\mathcal{G}_2]|\mathcal{G}_1\big] = \mathbb{E}[X|\mathcal{G}_1]$ almost surely.

(v) (Jensen's inequality) If $f : \mathbb{R} \to \mathbb{R}$ is a convex function, then $\mathbb{E}\big[f(X)|\mathcal{G}\big] \geq f\big(\mathbb{E}(X|\mathcal{G})\big)$ almost surely. In particular, $X \mapsto \mathbb{E}[X|\mathcal{G}]$ is a contraction in $L^p(\mathbb{P})$ for any $p \geq 1$ (i.e., $\mathbb{E}\big[\big|\mathbb{E}(X|\mathcal{G})\big|^p\big] \leq \mathbb{E}(|X|^p)$).

(vi) *(Hölder's inequality) For any $p, q \geq 1$ with $\frac{1}{p} + \frac{1}{q} = 1$, we have $\mathbb{E}[|X_1 X_2||\mathcal{G}] \leq \mathbb{E}[|X_1|^p|\mathcal{G}]^{1/p}.\mathbb{E}[|X_2|^q|\mathcal{G}]^{1/q}$.*

(vii) *(Fatou's lemma) If $X_n \geq 0$ for each n then $\liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}] \geq \mathbb{E}[\liminf_{n\to\infty} X_n|\mathcal{G}]$ almost everywhere.*

(viii) *(Monotone convergence theorem) If $X_n \geq 0$ for each n and $X_n \uparrow X$ almost everywhere (i.e., for each n, $X_{n+1} \geq X_n$ and X_n converges pointwise to a limit X), then $\mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}]$ almost everywhere.*

(ix) *(Dominated convergence theorem) If $\sup_n |X_n(\cdot)| \leq Y(\cdot)$ almost everywhere so that $Y \in L^1(\mathbb{P})$ and X_n converges pointwise to a limit X, then $\mathbb{E}[X_n|\mathcal{G}] \to \mathbb{E}[X|\mathcal{G}]$ almost everywhere.*

Theorem 2.15.3 (Orthogonal projection of Hilbert spaces)**.** *Let H be a Hilbert space (i.e., H is a vector space equipped with an inner-product $\langle \cdot, \cdot \rangle_H$ that defines a norm $\|x\|_H^2 := \langle x, x \rangle_H$ making H a complete metric space). Let $K \subset H$ be a closed subspace of H. Then for any $x \in H$, there exists $y \in K$ such that one of the equivalent properties hold:*

(i) *For any $z \in K$, $\langle x - y, z \rangle_H = 0$.*

(ii) *For any $z \in K$, $\|y - x\|_H \leq \|z - x\|_H$.*

Such y is unique, it is written as $y = \pi_K(x)$ and is called the orthogonal projection of x onto K. □

The following result is a direct consequence of the above result. For any σ-algebra $\mathcal{A}$, we denote by $L^2(\mathcal{A})$ to be the space of square integrable functions which are measurable w.r.t. $\mathcal{A}$.

Theorem 2.15.4 (Conditional expectation as a projection)**.** *Let $H = L^2(\mathcal{F})$ equipped with an inner product $\langle X, Y \rangle_H = \mathbb{E}[XY]$. Fix $X \in L^2(\mathcal{F})$ and let $K = L^2(\mathcal{G})$ where $\mathcal{G} \subset \mathcal{F}$ is a sub-σ-algebra. Then $K \subset H$ is a closed subspace of H and the orthogonal projection of X onto K is uniquely identified as*

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

References