Markov Processes*

Thomas Walker

Autumn 2023

Contents

*Chapter 8 of these notes contains material from Chapter 9 of Xue-Mei Lie's notes for the same course taught in the autumn of 2021. It is not examinable for the course taught in the autumn of 2023, however, it supplements the existing material.

1 Introduction to Markov Processes

1.1 Probability Preliminaries

We will define randomness using Kolmogorov's framework. That is, we define an underlying probability space, which is a tuple $(\Omega, \mathcal{F}, \mathbb{P})$ with the following components.

- Ω is an abstract space.
- $\mathcal{F}$ a σ-algebra.
- $\mathbb{P}$ is a probability measure. That is a measure with unit mass.

Random quantities will take values in a state space $\mathcal{X}$. Which we assume to be a complete and separable metric space. We will use $\mathcal{B}(\mathcal{X})$ to denote its Borel σ-algebra. That is, the σ-algebra generated by its open sets. Recall that random quantities are represented as $\mathcal{X}$-valued random variables, which are measurable functions

$$X : (\Omega, \mathcal{F}) \to (\mathcal{X}, \mathcal{B}(\mathcal{X})).$$

For X a $\mathcal{X}$-valued random variable, $\mathrm{Law}(X)$ or $X_*\mathbb{P}$ is the push forward of $\mathbb{P}$ of X. Which is the measure where for each $A \subset \mathcal{B}(\mathcal{X})$ we have

$$X_*\mathbb{P}(A) = \mathbb{P}(X \in A) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in A\}).$$

A family of $\mathcal{X}$-valued random variables, X_i is denoted using an index set I as $(X_i)_{i \in I}$.

- $\mathbb{P}(X_i \in A, X_j \in B) := \mathbb{P}\left(\{\omega \in \Omega : X_i(\omega) \in A \text{ and } X_j(\omega) \in B\}\right).$
- For $I' \subset I$ the family $(X_i)_{i \in I'}$ is independent if for all $A_i \in \mathcal{B}(\mathcal{X})$ we have

$$\mathbb{P}\left(X_i \in A_i \text{ for all } i \in I'\right) = \prod_{i \in I'} \mathbb{P}(X_i \in A_i).$$

- For dependent random variables, when $\mathbb{P}(X_j \in B) > 0$ we have

$$\mathbb{P}(X_i \in A | X_j \in B) := \frac{\mathbb{P}(X_i \in A, X_j \in B)}{\mathbb{P}(X_j \in B)}.$$

For I totally ordered, $X = (X_i)_{i \in I}$ defines a stochastic process. For I discrete we have a discrete-time process.

1.2 Markov Processes

Definition 1.2.1 (Intuitive)**.** *A Markov process is a process that could be characterised by either of the following statements.*

1. *For any prediction of the future, knowledge of the present is just as good as knowledge of both the past and present.*
2. *Conditioned on the present, the past and present are independent.*

Definition 1.2.2 (Formal)**.** *A process $(X_i)_{i=0}^{\infty}$ is Markov if it satisfies either one of the following characterisations.*

1. *For every $j \in \mathbb{N}$ and $A, B_1, \ldots, B_j \in \mathcal{B}(\mathcal{X})$ such that $\mathbb{P}(X_1 \in B_1, \ldots, X_j \in B_j) > 0$ we have*

$$\mathbb{P}(X_{j+1} \in A | X_1 \in B_1, \ldots, X_j \in B_j) = \mathbb{P}(X_{j+1} \in A | X_j \in B_j).$$

2. *For any $i < j < k$ and $A_0, \ldots, A_i, B, C_{j+1}, \ldots, C_k \in \mathcal{B}(\mathcal{X})$ with $\mathbb{P}(X_j \in B) > 0$ we have*

$$\begin{aligned}\mathbb{P}(X_0 \in A_0, \ldots, X_i \in A_i, X_{j+1} \in C_{j+1}, \ldots, X_k \in C_k | X_j \in B)& \\ = \mathbb{P}(X_0 \in A_0, \ldots, X_i \in A_i | X_j \in B)& \\ \cdot \mathbb{P}(X_{j+1} \in C_{j+1}, \ldots, X_k \in C_k | X_j \in B).&\end{aligned}$$

Exercise 1.2.3. *Suppose that $\mathcal{X}$ is a finite set. Show that the characterisations of Definition 1.2.2 are equivalent.*

Exercise 1.2.4. *Let $(X_j)_{j=0}^{\infty}$ be defined by*

$$X_j = X_0 + \sum_{i=1}^{j} Y_i,$$

where

- *X_0 is a $\mathbb{Z}$-valued random variable, and*
- *$(Y_i)_{i=1}^{\infty}$ is an $\mathrm{i.i.d}$ family of $\{\pm 1\}$-valued random variables, independent of X_0.*

Show that $(X_j)_{j=0}^{\infty}$ is a Markov process.

Exercise 1.2.5. *Let $(X_j)_{j=0}^{\infty}$ be as in Exercise 1.2.4 and define*

$$M_j = \max_{0 \le i \le j} X_i.$$

Taking $X_0 = 0$, show that $(M_j)_{j=0}^{\infty}$ is not a Markov process.

1.3 Filtrations

Definition 1.3.1. *The σ-algebra generated by a family of random variables $(Y_j)_{j \in J}$, denoted $\sigma\left((Y_j)_{j \in J}\right)$, is the smallest σ-algebra such that*

$$\left\{\left\{Y_j^{-1}(A)\right\} : j \in J, A \in \mathcal{B}(\mathcal{X})\right\}$$

is contained within the σ-algebra. In other words, $\sigma\left((Y_j)_{j \in J}\right)$ is the smallest σ-algebra such that the family $(Y_j)_{j \in J}$ is measurable.

Remark 1.3.2. *σ-algebras can be thought of as encoding information. If Z is a random variable with respect to $\sigma\left((Y_j)_{j \in J}\right)$, then knowing the values of $(Y_j)_{j \in J}$ is sufficient for determining the value of Z. That is, $Z = g\left((Y_j)_{j \in J}\right)$ for a measurable function g.*

Exercise 1.3.3. *Let Y_1 be a $\{-1, 0, 1\}$-valued random variable and let Y_2 be a $\{-3, 3\}$-valued random variable. Then what is the maximum value of $|\sigma(Y_1, Y_2)|$?*

Definition 1.3.4. *A filtration is a sequence of σ-algebras $(\mathcal{F}_n)_{n=0}^{\infty}$ with $\mathcal{F}_n \subset \mathcal{F}$ for all n such that for $m \le n$ we have $\mathcal{F}_m \subset \mathcal{F}_n$.*

Definition 1.3.5. *A stochastic process X with state space $\mathcal{X}$ is a collection $(X_n)_{n=0}^{\infty}$ is a collection of $\mathcal{X}$-valued random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.*

Definition 1.3.6. *A stochastic process $X = (X_n)_{n=0}^{\infty}$ is adapted to $(\mathcal{F}_n)_{n=0}^{\infty}$ if for every n we have that X_n is measurable with respect to $\mathcal{F}_n$.*

Definition 1.3.7. *For a stochastic process X, the filtration generated by X, denoted $\mathcal{F}^0 = \left(\mathcal{F}_n^0\right)_{n=0}^{\infty}$ is defined as*

$$\mathcal{F}_n^0 = \sigma\left((X_i)_{i=0}^n\right)$$

for every n.

Remark 1.3.8.

- *$\mathcal{F}^0$ is the smallest filtration for which X is adapted.*
- *Intuitively, $\mathcal{F}_n^0$ is the information generated by the process up until time n.*

1.4 Conditional Expectation

Throughout, suppose we have a main underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1.4.1. *A random variable X is integrable if $\mathbb{E}(|X|) < \infty$.*

Definition 1.4.2. *Let X be an integrable random variable, and consider a sub-σ-algebra $\mathcal{F}' \subset \mathcal{F}$. The conditional expectation of X with respect to $\mathcal{F}'$ is a $\mathcal{F}'$-measurable random variable X' such that for every $A \in \mathcal{F}'$ we have*

$$\int_A X(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_A X'(\omega)\,\mathrm{d}\mathbb{P}(\omega).$$

Proposition 1.4.3. *In the setting of Definition 1.4.2, there exists, up to $\mathbb{P}$-null sets, a unique conditional expectation of X with respect to $\mathcal{F}'$.*

Proof. Let μ be a the restriction of $\mathbb{P}$ to $\mathcal{F}'$. Let ν be the measure on $\mathcal{F}'$ defined by

$$\nu(A) = \int_A X(\omega)\,\mathrm{d}\mathbb{P}(\omega)$$

for all $A \in \mathcal{F}'$. Then if $\mu(A) = 0$ it follows that $\nu(A) = 0$ which implies that $v \ll \mu$. Therefore, by the Random-Nikodyn theorem there exists a unique $\mathcal{F}'$-measurable function, X', up to sets of zero measure such that

$$\nu(A) = \int_A X'(\omega)\,\mathrm{d}\mathbb{P}(\omega)$$

for all $A \in \mathcal{F}'$. □

Remark 1.4.4.

- *If we have a candidate for the conditional expectation, by uniqueness, it suffices to check the conditions* *of Definition 1.4.2 are satisfied to conclude that it is indeed the conditional expectation.*

- *Note that $X(\omega)$ is not necessarily $\mathcal{F}'$ measurable. So we cannot say that X' equals X pointwise.*

We denote the conditional expectation of X with respect to $\mathcal{F}'$ by $\mathbb{E}\left(X|\mathcal{F}'\right)$.

- We also write $\mathbb{P}\left(A|\mathcal{F}'\right) = \mathbb{E}\left(\mathbf{1}_A|\mathcal{F}'\right)$.
- Given a random variable Y, we write $\mathbb{E}\left(\cdot|Y\right) = \mathbb{E}\left(\cdot|\sigma(Y)\right)$.

Remark 1.4.5. *Intuitively, we think of $\mathbb{E}(X|Y)$ as being an approximation of X when we only have the information of $\sigma(Y)$ available to us. As $\mathbb{E}(X|Y)$ is measurable with respect to $\sigma(Y)$ we can think of it as a measurable function $\phi(Y)$.*

Example 1.4.6. *Let $\mathcal{F}' = \{\emptyset, A, A^c, \Omega\}$ where $A \in \mathcal{F}$ and $\mathbb{P}(A) \in (0,1)$. Let X be an integrable random variable. Then*

$$X'(\omega) = \begin{cases} \frac{1}{\mathbb{P}(A)} \int_A X(\omega)\, \mathrm{d}\mathbb{P}(\omega) & \omega \in A \\ \frac{1}{\mathbb{P}(A^c)} \int_{A^c} X(\omega)\, \mathrm{d}\mathbb{P}(\omega) & \omega \in A^c. \end{cases}$$

To see this one just has to note that X' is $\mathcal{F}'$-measurable and

$$\int_B X'(\omega)\, \mathrm{d}\mathbb{P} = \begin{cases} 0 & B = \emptyset \\ \int_A X(\omega)\, \mathrm{d}\mathbb{P}(\omega) & B = A \\ \int_{A^c} X(\omega)\, \mathrm{d}\mathbb{P}(\omega) & B = A^c \\ 1 & B = \Omega, \end{cases}$$

where all set equalities hold almost everywhere.

Example 1.4.7. *Let Y be a $\mathbb{N}$-valued random variable. Let X be an integrable random variable. Then*

$$\mathbb{E}(X|Y) = \sum_{i:\mathbb{P}(\{Y=i\})>0} \mathbb{E}\left(X|\{Y=i\}\right) \mathbb{P}\left(\{Y=i\}\right).$$

Exercise 1.4.8. *Let $(\Omega, \mathcal{F}, \mathbb{P}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathcal{N}(0,1))$. Let Y be an integrable random variable given by $Y(\omega) = \omega^2$. Let X be an integrable random variable. Give a formula for $\mathbb{E}(X|Y)$.*

1.4.1 Properties

Proposition 1.4.9. *Let X be an integrable random variable defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose $\mathcal{B} \subset \mathcal{F}$ is a σ-algebra such that $\sigma(X)$ and $\mathcal{B}$ are independent. Then,*

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

Proof. Let $B \in \mathcal{B}$, then

$$\begin{aligned}
\int_B X(\omega)\,\mathrm{d}\mathbb{P}(\omega) &= \int_\Omega \mathbf{1}_B X(\omega)\,\mathrm{d}\mathbb{P}(\omega) \\
&= \mathbb{E}(X\mathbf{1}_B) \\
&\overset{(1)}{=} \mathbb{E}(X)\mathbb{E}(1_{\mathbf{B}}) \\
&= \mathbb{E}(X)\mathbb{P}(B) \\
&= \int_B \mathbb{E}(X)\,\mathrm{d}\mathbb{P}(\omega),
\end{aligned}$$

where in (1) we have used the independence assumptions. As $\mathbb{E}(X)$ is a constant, it is clearly $\mathcal{B}$-measurable. Therefore, we can conclude that $\mathbb{E}(X|\mathcal{B}) = \mathbb{E}(X)$. ☐

A property of the conditional expectation that will be used extensively is the tower property.

Proposition 1.4.10 (The Tower Property). *Let X be an integrable random variable and let $\mathcal{A} \subset \mathcal{B}$ be sub-σ-algebras of $\mathcal{F}$. Then*

$$\mathbb{E}\left(\mathbb{E}\left(X|\mathcal{A}\right)|\mathcal{B}\right) = \mathbb{E}(\mathbb{E}(X|\mathcal{B})|\mathcal{A}) = \mathbb{E}(X|\mathcal{A}).$$

Proof. Note that $\mathbb{E}(X|\mathcal{A})$ is $\mathcal{A}$-measurable by definition, hence it is also $\mathcal{B}$ measurable. Therefore, by Proposition 1.4.9 we can conclude that

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

Furthermore, for $C \in \mathcal{A} \subset \mathcal{B}$ we know from the definition of conditional expectations that

$$\int_C \mathbb{E}(X|\mathcal{A})(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_C X(\omega)\,\mathrm{d}\mathbb{P}(\omega)$$

and

$$\int_C \mathbb{E}(X|\mathcal{B})(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_C X(\omega)\,\mathrm{d}\mathbb{P}(\omega).$$

Therefore,

$$\int_C \mathbb{E}(X|\mathcal{A})(\omega)\,\mathrm{d}\mathbb{P}(\omega) = \int_C \mathbb{E}(X|\mathcal{B})(\omega)\,\mathrm{d}\mathbb{P}(\omega). \tag{1.4.1}$$

From equation (1.4.1) and the fact $\mathbb{E}(X|\mathcal{A})$ is $\mathcal{A}$-measurable, we observe that $\mathbb{E}(X|\mathcal{A})$ satisfies the required conditions to be the conditional expectation of $\mathbb{E}(X|\mathcal{B})$ with respect to $\mathcal{A}$. Therefore, by Proposition 1.4.3 we conclude that

$$\mathbb{E}(\mathbb{E}(X|\mathcal{B})|\mathcal{A}) = \mathbb{E}(X|\mathcal{A})$$

which completes the proof. ☐

Remark 1.4.11.

- *Intuitively Proposition 1.4.10 holds as $\mathcal{A}$ has less information than $\mathcal{B}$, and so the inner conditioning retains sufficient information for the outer conditioning to take full effect.*
- *For $\mathcal{A} \subset \mathcal{B}$ σ-algebras we have that*

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

  This can be intuitively justified by saying that the restriction of information imposed by the first conditioning already matches or exceeds the restriction of information imposed by the outer conditioning. Therefore, the outer conditioning has no effect.

Proposition 1.4.12. *Let $X, Y \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{B} \subset \mathcal{F}$ be a σ-algebra such that Y is $\mathcal{B}$-measurable. Then*

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

Proof. We first assume that $Y = \mathbf{1}_A$ for some $A \in \mathcal{B}$. Then for any $B \in \mathcal{B}$ it follows that

$$\begin{aligned}\int_B Y(\omega)X(\omega)\,\mathrm{d}\mathbb{P}(\omega) &= \int_{B\cap A} X(\omega)\,\mathrm{d}\mathbb{P}(\omega)\\ &\overset{(1)}{=} \int_{B\cap A} \mathbb{E}(X|\mathcal{B})\,\mathrm{d}\mathbb{P}(\omega)\\ &= \int_B \mathbf{1}_A\mathbb{E}(X|\mathcal{B})\,\mathrm{d}\mathbb{P}(\omega)\\ &= \int_B Y\mathbb{E}(X|\mathcal{B})\,\mathrm{d}\mathbb{P}(\omega).\end{aligned}$$

Where (1) is simply the definition of $\mathbb{E}(X|\mathcal{B})$ as $B \cap A \in \mathcal{B}$. Moreover, as $A \in \mathcal{B}$ we note that $\mathbf{1}_A\mathbb{E}(X|\mathcal{B})$ is the product of $\mathcal{B}$-measurable functions and is therefore also $\mathcal{B}$-measurable. Due to the linearity of the integral we can see that this implies that for Y a simple function, the result still holds. When $Y \geq 0$ almost surely, then there exists a sequence of positive simple functions $(Y_n)_{n\in\mathbb{N}}$ such that $Y_n \nearrow Y$. It follows that $Y_nX \to YX$ almost surely, moreover, we have that $|Y_nX| \leq |X|$. Using the fact that X and Y are in L^2 we can use the Cauchy-Schwarz inequality to deduce that

$$\mathbb{E}(|XY|)^2 \leq \mathbb{E}(|X|)^2\mathbb{E}(|Y|)^2 < \infty.$$

Therefore, we can apply the dominated convergence theorem to conclude that

$$\mathbb{E}(YX|\mathcal{B}) \overset{\mathsf{DCT}}{=} \lim_{n\to\infty} \mathbb{E}\left(Y_nX|\mathcal{B}\right) \overset{(1)}{=} \lim_{n\to\infty} Y_n\mathbb{E}(X|\mathcal{B}) = Y\mathbb{E}(X|\mathcal{B}),$$

where in (1) we have the result for simple functions. Consequently, when Y is an arbitrary $\mathcal{B}$-measurable function we can use the decomposition $Y = Y^+ - Y^-$ and the previous step to conclude the result. $\square$

Proposition 1.4.13. *Suppose that X and Y are bounded $\mathbb{R}$-valued random variables, and let $f, g \in \mathcal{B}_b(\mathbb{R})$. Then*

$$\mathbb{E}\left(\mathbb{E}\left(Y|f(g(X))\right)|g(X)\right) = \mathbb{E}\left(\mathbb{E}\left(Y|g(X)\right)|f(g(X)\right) = \mathbb{E}(Y|f(g(X)).$$

Proof. First note that $f(g(X))$ and $g(X)$ are both random variables as f and g are measurable functions. Specifically, we have that

$$\sigma(f(g(X))) := \mathcal{A}_{fg} = \{A : f(g(X(A))) \in \mathcal{B}(\mathbb{R})\}$$

and

$$\sigma(g(X)) := \mathcal{A}_g = \{A : g(X(\omega)) \in \mathcal{B}(\mathbb{R})\}.$$

Now consider $A \in \mathcal{A}_{fg}$, then $f(g(X(A))) = B$ for some $B \in \mathcal{B}(\mathbb{R})$. Moreover, as f is measurable it follows that $g(X(A)) = f^{-1}(B) \in \mathcal{B}(\mathbb{R})$ which implies that $A \in \mathcal{A}_g$. Therefore, $\mathcal{A}_{fg} \subset \mathcal{A}_g$. Hence, we can apply the tower property of the expectation to deduce that

$$\mathbb{E}\left(\mathbb{E}\left(Y|g(X)\right)|f(g(X))\right) = \mathbb{E}(Y|f(g(X)).$$

Similarly, we can apply the reverse tower property to deduce that

$$\mathbb{E}\left(\mathbb{E}(Y|f(g(X))|g(X)\right) = \mathbb{E}(Y|f(g(X)).$$

$\square$

Proposition 1.4.14. *Suppose that X and Y are bounded $\mathbb{R}$-valued random variables. Then for any $f \in \mathcal{B}_b(\mathbb{R})$ we have*
$$\mathbb{E}\left((Y-\mathbb{E}(Y|X))^2\right) \leq \mathbb{E}\left((Y-f(X)^2\right).$$

Proof. First not that

$$\begin{aligned}
\mathbb{E}\left((Y-\mathbb{E}(Y|X))^2|X\right) &= \mathbb{E}\left(Y^2|X\right) - 2\mathbb{E}\left(Y\mathbb{E}(Y|X)|X\right) + \mathbb{E}\left(\mathbb{E}(Y|X)^2|X\right) \\
&= \mathbb{E}\left(Y^2|X\right) - 2\mathbb{E}\left(Y|X\right)^2 + \mathbb{E}(Y|X)^2 \\
&= \mathbb{E}\left(Y^2|X\right) - \mathbb{E}\left(Y|X\right)^2
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\mathbb{E}\left((Y-f(X))^2|X\right) &= \mathbb{E}\left(Y^2|X\right) - 2\mathbb{E}\left(Yf(X)|X\right) + \mathbb{E}\left(f(X)^2|X\right) \\
&= \mathbb{E}\left(Y^2|X\right) - 2f(X)\mathbb{E}\left(Y|X\right) + f(X)^2.
\end{aligned}$$

Therefore,
$$\mathbb{E}\left((Y-f(X))^2|X\right) - \mathbb{E}\left((Y-\mathbb{E}(Y|X))^2|X\right) = \left(\mathbb{E}(Y|X)-f(X)\right)^2 \geq 0$$
and hence
$$\mathbb{E}\left((Y-f(X))^2|X\right) \geq \mathbb{E}\left((Y-\mathbb{E}(X|Y))^2|X\right).$$
Taking the expectations of both sides we can use the law of iterated probability to conclude that

$$\mathbb{E}\left((Y-f(X))^2\right) \geq \mathbb{E}\left((Y-\mathbb{E}(X|Y))^2\right).$$

□

Proposition 1.4.15. *Suppose that $\mathcal{A} \subset \mathcal{B}$ are σ-algebras contained in $\mathcal{F}$. For a bounded random variable X we have that*
$$\mathrm{Var}(\mathbb{E}(X|\mathcal{B})) \geq \mathrm{Var}(\mathbb{E}(X|\mathcal{A})).$$

Proof. Observe that,

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}(\mathrm{Var}(X|\mathcal{A})) + \mathrm{Var}(\mathbb{E}(X|\mathcal{A})) \\
&= \mathbb{E}(\mathrm{Var}(X|\mathcal{B})) + \mathrm{Var}(\mathbb{E}(X|\mathcal{B})).
\end{aligned}$$

Therefore,
$$\mathrm{Var}(\mathbb{E}(X|\mathcal{B})) - \mathrm{Var}(\mathbb{E}(X|\mathcal{A})) = \mathbb{E}\left(\mathrm{Var}(X|\mathcal{A})\right) - \mathbb{E}\left(\mathrm{Var}(X|\mathcal{B})\right).$$
Observe that,

$$\begin{aligned}
\mathrm{Var}(X|\mathcal{A}) &= \mathbb{E}\left((X-\mathbb{E}(X|\mathcal{A}))^2|\mathcal{A}\right) \\
&= \mathbb{E}\left((X-\mathbb{E}(\mathbb{E}(X|\mathcal{B})|\mathcal{A}))^2|\mathcal{A}\right) \\
&\overset{(1)}{\geq} \mathbb{E}\left((X-\mathbb{E}(X|\mathcal{B}))^2|\mathcal{A}\right) \\
&\overset{(2)}{=} \mathbb{E}\left(\mathbb{E}\left((X-\mathbb{E}(X|\mathcal{B}))^2|\mathcal{B}\right)|\mathcal{A}\right) \\
&= \mathbb{E}(\mathrm{Var}(X|\mathcal{B})|\mathcal{A}). \qquad (1.4.2)
\end{aligned}$$

Where in (1) we use Proposition 1.4.14 as $\mathbb{E}(\mathbb{E}(X|\mathcal{B})|\mathcal{A})$ is $\mathcal{A}$-measurable and so $\mathcal{B}$-measurable. In (2) the tower property for conditional expectations is used. As $\mathrm{Var}(X|\mathcal{A})$ is an $\mathcal{A}$-measurable function, we have that $\mathbb{E}(\mathrm{Var}(X|\mathcal{A})|\mathcal{A}) = \mathrm{Var}(X|\mathcal{A})$. Using this and (1.4.2) we deduce that $\mathrm{Var}(X|\mathcal{A}) \geq \mathrm{Var}(X|\mathcal{B})$, which in particular means that
$$\mathrm{Var}(\mathbb{E}(X|\mathcal{B})) - \mathrm{Var}(\mathbb{E}(X|\mathcal{A})) = \mathbb{E}\left(\mathrm{Var}(X|\mathcal{A})\right) - \mathbb{E}\left(\mathrm{Var}(X|\mathcal{B})\right) \geq 0$$
and hence $\mathrm{Var}(\mathbb{E}(X|\mathcal{B})) \geq \mathrm{Var}(\mathbb{E}(X|\mathcal{A}))$. □

1.5 Solution to Exercises

Exercise 1.2.3

Solution. Suppose that statement 1 of Definition 1.2.2 holds. Let

- $A = \{X_0 \in A_0, \dots, X_i \in A_i\}$,
- $B = \{X_j \in B\}$, (we abuse notation slightly with the B's) and
- $C = \{X_{j+1} \in C_{j+1}, \dots, X_k \in C_k\}$. We assume throughout that $\mathbb{P}(B), \mathbb{P}(A, B) > 0$. Then

$$\begin{aligned}\mathbb{P}(A, C|B) &= \frac{\mathbb{P}(A, C, B)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}(C|A, B)\mathbb{P}(A, B)}{\mathbb{P}(B)} \\ &= \mathbb{P}(C|A, B)\mathbb{P}(A|B).\end{aligned}$$

Applying 1. we get that $\mathbb{P}(C|A, B) = \mathbb{P}(C|B)$ and so

$$\mathbb{P}(A, C|B) = \mathbb{P}(A|B)\mathbb{P}(C|B)$$

as required. Now suppose statement 2. of Definition 1.2.2 holds. Then

$$\begin{aligned}\mathbb{P}(X_{j+1} \in A_{j+1}|X_1 \in B_1, \dots, X_j \in B_j) &= \frac{\mathbb{P}(X_{j+1} \in A_{j+1}, X_1 \in B_1, \dots, X_j \in B_j)}{\mathbb{P}(X_1 \in B_1, \dots, X_j \in B_j)} \\ &= \frac{\mathbb{P}(X_{j+1} \in A_{j+1}, X_1 \in B_1, \dots, X_{j-1} \in B_{j-1}|X_j \in B_j)\mathbb{P}(X_j \in B_j)}{\mathbb{P}(X_1 \in B_1, \dots, X_j \in B_j)} \\ &= \frac{\mathbb{P}(X_{j+1} \in A_{j+1}|X_j \in B_j)\mathbb{P}(X_1 \in B_1, \dots, X_{j-1} \in B_{j-1}|X_j \in B_j)\mathbb{P}(X_j \in B_j)}{\mathbb{P}(X_1 \in B_1, \dots, X_j \in B_j)} \\ &= \mathbb{P}(X_{j+1} \in A_{j+1}|X_j \in B_j).\end{aligned}$$

$\square$

Exercise 1.2.4

Solution. Note that $X_j = X_{j-1} + Y_j$. Furthermore, as $\mathcal{X} = \mathbb{Z}$ it suffices to consider the events $\{k\}$ for $k \in \mathbb{Z}$. Therefore,

$$\begin{aligned}\mathbb{P}(X_{j+1} = x_{j+1}|X_0 = x_0, \dots, X_j = x_j) &= \mathbb{P}(X_{j+1} = x_j + Y_j|X_0 = x_0, \dots, X_j = x_j) \\ &= \mathbb{P}(X_{j+1} = x_j + Y_j|X_j = x_j).\end{aligned}$$

Where the second equality follows from the fact that the Y_j are independent and independent from X_0. Therefore $(X_j)_{j=1}^{\infty}$ satisfies statement 1 of Definition 1.2.2 and hence is a Markov process. $\square$

Exercise 1.2.5

Solution. To do this we will check what happens when our present is $j = 3$ and we are trying to predict what happens at $j + 1 = 4$. Recall, that $X_k = X_0 + \sum_{i=1}^{k} Y_i = \sum_{i=1}^{k} Y_i$, where $(Y_i)_{i=1}^{\infty}$ are i.i.d $\{\pm 1\}$-valued random variables. For the sequence (M_0, M_1, M_2, M_3) consider the following possibilities.

1. $(0, 0, 0, 1)$ which arises from the sequence $(X_0, X_1, X_2, X_3) = (0, -1, 0, 1)$.
2. $(0, 1, 1, 1)$ which may arise from the sequence $(X_0, X_1, X_2, X_3) = (0, 1, 0, -1)$ or $(X_0, X_1, X_2, X_3) = (0, 1, 0, 1)$.

Note that these are the only possibilities for which $M_3 = 1$. We can then compute the possibilities of M_4 in each case.

1. $\mathbb{P}(M_4 = 2 | M_0 = 0, M_1 = 0, M_2 = 0, M_3 = 1) = \frac{1}{2}$.
2. $\mathbb{P}(M_4 = 2 | M_0 = 0, M_1 = 1, M_2 = 1, M_3 = 1) = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$,
   - 1. The first $\frac{1}{2}$ comes from the fact that we must observe the sequence $(X_0, X_1, X_2, X_3) = (0, 1, 0, 1)$ in order to be able to reach $M_4 = 2$. The second $\frac{1}{2}$ is the probability of reaching $M_4 = 2$ from that sequence.

Therefore, we see that the future $j + 1 = 4$ is dependent on more than just the present $j = 3$.

□

Exercise 1.3.3

Solution. There are $3 \times 2 = 6$ possibilities for the tuple (Y_1, Y_2). Taking the power set of these gives a σ-algebra of size 2^6, which is the largest σ-algebra as σ-algebra is simply a collection of subsets. □

Exercise 1.4.8

Solution. For any $B \in \sigma(Y)$ we have that $B = \{\omega : Y(\omega) \in A\}$ for some $A \in \mathcal{B}(\mathbb{R})$. Consequently, by the observation that Y is even, we get

$$\begin{aligned} B \cap \mathbb{R}^+ &= \{\omega : Y(\omega) \in A\} \cap \mathbb{R}^+ \\ &= \{\omega : Y(-\omega) \in A\} \cap \mathbb{R}^+ \\ &= \{-\omega : Y(\omega) \in A\} \cap \mathbb{R}^+ \\ &= -\left(\{\omega : Y(\omega) \in A\} \cap \mathbb{R}^-\right) \\ &= -\left(B \cap \mathbb{R}^-\right). \end{aligned}$$

Therefore, using the substitution $u = -\omega$, and the fact that the measure is even, we get

$$\int_{B \cap \mathbb{R}^-} X(u)\, \mathrm{d}\mathbb{P}(u) = \int_{-(B \cap \mathbb{R}^-)} X(-\omega)\, \mathrm{d}\mathbb{P}(-\omega) = \int_{B \cap \mathbb{R}^+} X(-\omega)\, \mathrm{d}\mathbb{P}(\omega).$$

It then follows that

$$\begin{aligned} \int_B X(\omega)\, \mathrm{d}\mathbb{P}(\omega) &= \int_{B \cap \mathbb{R}^+} X(\omega)\, \mathrm{d}\mathbb{P}(\omega) + \int_{B \cap \mathbb{R}^-} X(\omega)\, \mathrm{d}\mathbb{P}(\omega) \\ &= \int_{B \cap \mathbb{R}^+} X(\omega) + X(-\omega)\, \mathrm{d}\mathbb{P}(\omega). \end{aligned}$$

Doing a similarly computation we also find that

$$\begin{aligned} \int_B X(\omega)\, \mathrm{d}\mathbb{P}(\omega) &= \int_{B \cap \mathbb{R}^+} X(\omega)\, \mathrm{d}\mathbb{P}(\omega) + \int_{B \cap \mathbb{R}^-} X(\omega)\, \mathrm{d}\mathbb{P}(\omega) \\ &= \int_{B \cap \mathbb{R}^-} X(\omega) + X(-\omega)\, \mathrm{d}\mathbb{P}(\omega). \end{aligned}$$

Adding these together we deduce that

$$2 \int_B X(\omega)\, \mathrm{d}\mathbb{P} = \int_B X(\omega) + X(-\omega)\, \mathrm{d}\mathbb{P}(\omega)$$

so that

$$\mathbb{E}(X|Y) = \frac{1}{2}\left(X(\omega) + X(-\omega)\right).$$

□

2 The Markov Property

2.1 The Discrete Time Markov Property

We can now reformulate the Markov property with conditional expectations. The process $(X_n)_{n=0}^{\infty}$ has the Markov property if one of the following equivalent conditions holds.

1. For every $j \in \mathbb{N}$ and $A \in \mathcal{B}(\mathcal{X})$ we have

$$\mathbb{P}(X_{j+1} \in A | X_0, \dots, X_j) = \mathbb{P}(X_{j+1} \in A | X_j).$$

2. For any $j < k$, $A \in \mathcal{B}\left(\mathcal{X}^j\right)$ and $B \in \mathcal{B}\left(\mathcal{X}^{k-j}\right)$, we have

$$\begin{aligned}\mathbb{P}((X_0, \dots, X_{j-1}) \in A, (X_{j+1}, \dots, X_k) \in B | X_j) =& \mathbb{P}((X_0, \dots, X_{j-1}) \in A | X_j) \\ &\cdot \mathbb{P}((X_{j+1}, \dots, X_k) \in B | X_j).\end{aligned}$$

Theorem 2.1.1. *Suppose that $\mathcal{X}$ is a discrete space, and $\mathcal{Y}$ is another state space. Let $F_n : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ be a measurable map for each $n \in \mathbb{N}$. Furthermore, let X_0 be a $\mathcal{X}$-valued random variable and $(\zeta_n)_{n=1}^{\infty}$ be a family of independent $\mathcal{Y}$-valued random variables that are also independent of X_0. Then the process $X = (X_n)_{n=0}^{\infty}$ defined by*

$$X_{n+1} = F_n(X_n, \zeta_{n+1})$$

for $n \in \mathbb{N}$ is a Markov process.

Proof. As we are in a discrete space, it suffices to show the Markov property holds for singletons set. That is,

$$\begin{aligned}\mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n, \dots, X_0 = i_0) &= \mathbb{P}(F_n(X_n, \zeta_{n+1}) | X_n = i_n, \dots, X_0 = i_0) \\ &= \mathbb{P}(F_n(i_n, \zeta_{n+1})),\end{aligned}$$

where in the second equality we use the fact that ζ_{n+1} is independent of the $X_0, \dots, X_n$. Similarly,

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n) = \mathbb{P}(F_n(i_n, \zeta_{n+1})),$$

therefore, X has the Markov property. □

We can generalise Theorem 2.1.1 to arbitrary state spaces. To do so, we introduce the following result.

Proposition 2.1.2. *Let $\mathcal{X}$ and $\mathcal{Y}$ be state spaces and let $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a measurable function. Suppose that X is $\mathcal{X}$-valued random variable and Y is a $\mathcal{Y}$-valued random variable such that $\phi(X, Y)$ is integrable. Then for any σ-algebra $\mathcal{G} \subset \mathcal{F}$ such that X is $\mathcal{G}$-measurable and Y is independent of $\mathcal{G}$, it follows that*

$$\mathbb{E}(\phi(X, Y) | \mathcal{G})(\omega) = \mathbb{E}(\phi(X(\omega), Y)).$$

Proof. We can first show this for $\phi(X, Y) = \mathbf{1}_{\{X \in A\}} \mathbf{1}_{\{Y \in B\}}$ where $A \in \mathcal{B}(\mathcal{X})$ and $B \in \mathcal{B}(\mathcal{Y})$. In this case,

$$\begin{aligned}\mathbb{E}\left(\mathbf{1}_{\{X \in A\}} \mathbf{1}_{\{Y \in B\}} | \mathcal{G}\right) &\overset{(1)}{=} \mathbf{1}_{\{X \in A\}} \mathbb{E}\left(\mathbf{1}_{\{Y \in B\}} | \mathcal{G}\right) \\ &\overset{(2)}{=} \mathbf{1}_{\{X \in A\}} \mathbb{E}\left(\mathbf{1}_{\{Y \in B\}}\right),\end{aligned}$$

where (1) is because X is $\mathcal{G}$ measurable and (2) is because Y is independent of $\mathcal{G}$. Evaluating this at ω we get the desired conclusion. We can then use the linearity of expectation to extend this result to X and Y simple random variables. Consequently, we can apply the monotone convergence theorem to deduce the result for X and Y non-negative random variables. Then we generalise to arbitrary random variables by utilizing the decomposition $X = X^+ - X^-$ and similarly for Y. Using this result we extend our previous result to general state spaces. □

Theorem 2.1.3. *Let $\mathcal{X}$ and $\mathcal{Y}$ be state spaces. For each $n \in \mathbb{N}$, let $F_n : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ be a measurable map. Let X_0 be a $\mathcal{X}$-valued random variable and let $(\zeta_n)_{n=0}^{\infty}$ be a family of independent $\mathcal{Y}$-valued random variables that are also independent of X_0. Then the process $X = (X_n)_{n=0}^{\infty}$ defined by*

$$X_{n+1} = F_n(X_n, \zeta_{n+1})$$

for each $n \in \mathbb{N}$ is a Markov process.

Proof. Now we must consider general sets in the σ-algebras, rather than the singletons we considered in Theorem 2.1.1. Recall the Markov property as

$$\mathbb{P}(X_{n+1} \in A | X_n, \dots, X_0) = \mathbb{P}(X_{n+1} \in A | X_n)$$

for all $A \in \mathcal{F}$. In our setting

$$\mathbb{P}(X_{n+1} \in A | X_n, \dots, X_0) = \mathbb{P}(F_n(X_n, \zeta_{n+1}) \in A | X_n, \dots, X_0),$$

where

- X_n is clearly measurable with respect to $\sigma(X_0, \dots, X_n)$, and
- ζ_{n+1} is independent of $\sigma(X_0, \dots, X_n)$ as its a subset of $\sigma(X_0)$ for which ζ_{n+1} is independent of by assumption.

Therefore, by Proposition 2.1.2 it follows that

$$\begin{aligned}\mathbb{P}(X_{n+1} \in A | X_n, \dots, X_0) &= \mathbb{E}\left(\mathbf{1}_A\left(F_n(X_n, \zeta_{n+1}\right) | X_n, \dots, X_0\right) \\ &= \mathbb{E}\left(\mathbf{1}_A\left(F_n(X_n, \zeta_{n+1})\right)\right).\end{aligned}$$

A similar computation shows that $\mathbb{P}(X_{n+1} \in A | X_n) = \mathbb{E}\left(\mathbf{1}_A\left(F_n(X_n, \zeta_{n+1})\right)\right)$ which shows that the process is Markov. ☐

2.2 Continuous Time Processes

A continuous time process, $(X_s)_{s \in I}$, is index by a continuous well-ordered set I, such as $I = [0, \infty)$.

Definition 2.2.1. *A continuous time filtration is an uncountable family of σ-algebras $(\mathcal{F}_s : s \in I)$ with $\mathcal{F}_s \subset \mathcal{F}$ and for each $s \leq t$ we have*

$$\mathcal{F}_s \subset \mathcal{F}_t.$$

Definition 2.2.2. *Given a continuous time process $(X_s)_{s \in I}$, the natural filtration $\left(\mathcal{F}_t^0 : t \in I\right)$ is given by*

$$\mathcal{F}_t^0 = \sigma\left(\{X_s : s \leq t, s \in I\}\right).$$

Definition 2.2.3. *A stochastic process $(X_s)_{s \in I}$ has the Markov property with respect to the filtration $(\mathcal{F}_s : s \in I)$ if $(X_s)_{s \in I}$ is adapted to $(\mathcal{F}_s : s \in I)$ and for all $s, t \in I$ with $s < t$ and $A \in \mathcal{B}(\mathcal{X})$ we have*

$$\mathbb{P}(X_t \in A | \mathcal{F}_s) = \mathbb{P}(X_t \in A | X_s).$$

Remark 2.2.4.

1. *Definition 2.2.3 says that our prediction of the future using information from the past and the present is as good as our prediction of the future using only information from the present.*
2. *We can reformulate the condition Definition 2.2.3 as for all $s, t \in I$ with $s < t$, $A \in \mathcal{B}(\mathcal{X})$ and $C \in \mathcal{F}_s$*

we have

$$\mathbb{E}\left(\mathbf{1}_A(X_t)\mathbf{1}_C\right) = \mathbb{E}\left(\mathbb{P}(X_t \in A | X_s)\mathbf{1}_C\right).$$

Definition 2.2.5. *A non-empty collection of subsets, $\mathcal{D}$, is a π-system if for any $A, B \in \mathcal{D}$ it follows that $A \cap B \in \mathcal{D}$.*

Definition 2.2.6. *A non-empty collection of subsets, $\mathcal{G}$, is a λ-system if the following properties hold.*

1. *$\emptyset \in \mathcal{G}$.*
2. *For any $G \in \mathcal{G}$ it follows that $G^c \in \mathcal{G}$.*
3. *For $G_1 \subset G_2 \subset \cdots \in \mathcal{G}$ we have that $\bigcup_{i=1}^{\infty} G_i \in \mathcal{G}$.*

Theorem 2.2.7 (Dynkin π-λ)**.** *If $\mathcal{D}$ is a π-system and $\mathcal{G}$ is a λ-system with $\mathcal{D} \subset \mathcal{G}$. Then $\sigma(\mathcal{D}) \subset \mathcal{G}$.*

Proposition 2.2.8. *Suppose a π-system $\mathcal{D}$ generates $\mathcal{B}(\mathcal{X})$ and for each $s \in I$ we are given a π-system $\mathcal{D}_s$ that generates $\mathcal{F}_s$. Moreover, suppose that for every $s < t$, $A \in \mathcal{D}$ and $C \in \mathcal{D}_s$ we have*

$$\mathbb{E}(\mathbf{1}_A(X_t)\mathbf{1}_C) = \mathbb{E}\left(\mathbb{P}(X_t \in A | X_s)\mathbf{1}_C\right).$$

Then $(X_s)_{s \in I}$ has the Markov property.

Proof. For fixed $s, t \in I$ such that $s < t$ consider the set

$$\mathcal{C} := \{C \in \mathcal{F}_s : \mathbb{E}(\mathbf{1}_A(X_t)\mathbf{1}_C) = \mathbb{E}\left(\mathbb{P}(X_t \in A | X_s)\mathbf{1}_C\right) \text{ for all } A \in \mathcal{D}\}.$$

Then $\emptyset \in \mathcal{C}$ trivially and similarly $\mathcal{X} \in \mathcal{C}$. If $C \in \mathcal{C}$ then $\mathbf{1}_{C^c} = \mathbf{1} - \mathbf{1}_C$, therefore,

$$\begin{aligned}
\mathbb{E}(\mathbf{1}_A(X_t)\mathbf{1}_{C^c}) &= \mathbb{E}(\mathbf{1}_A(X_t)(\mathbf{1} - \mathbf{1}_C)) \\
&= \mathbb{E}(\mathbf{1}_A(X_t)) - \mathbb{E}(\mathbf{1}_A(X_t)\mathbf{1}_C) \\
&= \mathbb{E}(\mathbb{P}(X_t \in A | X_s)\mathbf{1}) - \mathbb{E}(\mathbb{P}(X_t \in A | X_s)\mathbf{1}_C) \\
&= \mathbb{E}(\mathbb{P}(X_t \in A | X_s)\mathbf{1}_{C^c}).
\end{aligned}$$

So that $C^c \in \mathcal{C}$. Next, if $C_1 \subset C_2 \subset \ldots \mathcal{C}$ is a sequence of increasing sets, then by application of the monotone convergence theorem we conclude that $\bigcup_{i=1}^{\infty} C_i \in \mathcal{C}$. Therefore, $\mathcal{C}$ defines a λ-system. As $\mathcal{D}_s \subset \mathcal{C}$ by assumption we can apply the Dynkin π-λ system to conclude that $\mathcal{F}_s = \sigma(\mathcal{D}_s) \subset \mathcal{C} \subset \mathcal{F}_s$ which implies that $\mathcal{F}_s = \mathcal{C}$. Similarly, we can consider

$$\mathcal{A} = \{A \in \mathcal{B}(\mathcal{X}) : \mathbb{E}(\mathbf{1}_A(X_t)\mathbf{1}_C) = \mathbb{E}\left(\mathbb{P}(X_t \in A | X_s)\mathbf{1}_C\right) \text{ for all } C \in \mathcal{F}_s\}.$$

Similarly, one can show that $\mathcal{A}$ defines a λ-system with $\mathcal{D} \subset \mathcal{A}$. Therefore, by Theorem 2.2.7 system we conclude that $\mathcal{B}(\mathcal{X}) = \sigma(\mathcal{D}) \subset \mathcal{A} \subset \mathcal{B}(\mathcal{X})$. Therefore, $A = \mathcal{B}(\mathcal{X})$ and so $(X_s)_{s \in I}$ has the Markov property. ☐

Proposition 2.2.9. *If $(X_s : s \in I)$ is a Markov process with respect to a filtration $(\mathcal{F}_s : s \in I)$, then it is a Markov process with respect to its natural filtration $\left(\mathcal{F}_s^0 : s \in I\right)$.*

Proof. The natural filtration is given by the σ-algebras $\mathcal{F}_s^0 = \sigma(X_r : 0 \leq r \leq s)$. Note that for any $s < t$ we have that $\sigma(X_s) \subset \mathcal{F}_s^0 \subset \mathcal{F}_s$. Therefore, for $t \in I$ such that $t > s$ it follows that

$$\begin{aligned}\mathbb{P}\left(X_t \in A|\mathcal{F}_s^0\right) &\overset{(1)}{=} \mathbb{E}\left(\mathbb{P}(X_t \in A|\mathcal{F}_s)|\mathcal{F}_s^0\right)\\ &\overset{(2)}{=} \mathbb{E}\left(\mathbb{P}(X_t \in A|X_s)|\mathcal{F}_s^0\right)\\ &\overset{(3)}{=} \mathbb{P}(X_t \in A|X_s),\end{aligned}$$

where (1) is an application of the tower property, (2) is applying the Markov property of $(X_s)_{s\in I}$ with respect to $(\mathcal{F}_s)_{s\in I}$ and (3) follows a reverse application of the tower property. □

Theorem 2.2.10. *$(X_s : s \in I)$ is a Markov process with filtration $(\mathcal{F}_s : s \in I)$ if and only if for any bounded measurable function $f : \mathcal{X} \to \mathbb{R}$ and $s < t$ we have that*

$$\mathbb{E}(f(X_t)|\mathcal{F}_s) = \mathbb{E}(f(X_t)|X_s).$$

Proof. $(\Rightarrow)$. Suppose that $(X_s : s \in I)$ is a Markov process and consider $f = \mathbf{1}_A$, for $A \in \mathcal{B}(\mathcal{X})$. Then

$$\begin{aligned}\mathbb{E}(f(X_t)|\mathcal{F}_s) &= \mathbb{P}(X_t \in A|\mathcal{F}_s)\\ &= \mathbb{P}(X_t \in A|X_s)\\ &= \mathbb{E}(f(X_t)|X_s).\end{aligned}$$

By the linearity of expectation it follows that the equality also holds for simple functions $f = \sum_{i=1}^n \mathbf{1}_{A_i}$ for $A_i \in \mathcal{B}(\mathcal{X})$. For f a non-negative bounded function, we can find a sequence of simple functions (f_n) such that $f_n \nearrow f$. Therefore, applying the monotone convergence theorem at the points $(\star)$ we deduce that

$$\begin{aligned}\mathbb{E}(f(X_t)|\mathcal{F}_s) &= \mathbb{E}\left(\lim_{n\to\infty} f_n(X_t)|\mathcal{F}_s\right)\\ &\overset{(\star)}{=} \lim_{n\to\infty}\left(\mathbb{E}(f_n(X_t)|\mathcal{F}_s\right)\\ &= \lim_{n\to\infty}\left(\mathbb{E}(f_n(X_t)|X_s\right)\\ &\overset{(\star)}{=} \mathbb{E}\left(\lim_{n\to\infty} f_n(X_t)|X_s\right)\\ &= \mathbb{E}(f(X_t)|X_s).\end{aligned}$$

Then for f an arbitrary measurable and bounded function, we can write $f = f^+ - f^-$ where f^+, f^- are both bounded non-negative measurable functions. Hence, we can extend the result to f using the linearity of expectation.
$(\Leftarrow)$. Conversely, if the equality holds for any bounded measurable function $f : \mathcal{X} \to \mathbb{R}$. Then for any $A \in \mathcal{B}(\mathcal{X})$ we can let $f = \mathbf{1}_A$, so that for $s < t$ we have that

$$\begin{aligned}\mathbb{P}(X_t \in A|\mathcal{F}_s) &= \mathbb{E}(f(X_t)|\mathcal{F}_s)\\ &= \mathbb{E}(f(X_t)|X_s)\\ &= \mathbb{P}(X_t \in A|X_s).\end{aligned}$$

Therefore, $(X_s)_{s\in I}$ is a Markov process. □

2.3 Discrete Time Processes

Throughout let I be a discrete well-ordered set. Furthermore, let $\mathcal{B}_b(\mathcal{X})$ be the set of bounded measurable functions on $\mathcal{X}$.

Theorem 2.3.1. *Suppose we have a process $(X_n)_{n=0}^\infty$ and $l < m < n$. Then the following are equivalent.*

1. *For every $f \in \mathcal{B}_b(\mathcal{X})$ we have*

$$\mathbb{E}\left(f(X_n)|X_l, X_m\right) = \mathbb{E}\left(f(X_n)|X_m\right).$$

2. *For every $g \in \mathcal{B}_b(\mathcal{X})$ we have*

$$\mathbb{E}\left(g(X_l)|X_m, X_n\right) = \mathbb{E}\left(g(X_l)|X_m\right).$$

3. *For every $f, g \in \mathcal{B}_b(\mathcal{X})$ we have*

$$\mathbb{E}\left(f(X_n)g(X_l)|X_m\right) = \mathbb{E}\left(f(X_n)|X_m\right)\mathbb{E}(g(X_l)|X_m).$$

Proof. Reversing the time of the process, we see that statement 3. remains the same whereas statements 1. and 2. switch. Therefore, it is sufficient to prove that statement 1. happens if and only if statement 3. happens.
$(1 \Rightarrow 3)$.

$$\begin{aligned}
\mathbb{E}(f(X_n)g(X_l)|X_m) &\overset{(1)}{=} \mathbb{E}\left(\mathbb{E}\left(f(X_n)g(X_l)|X_m, X_l\right)|X_m\right) \\
&= \mathbb{E}\left(g(X_l)\mathbb{E}\left(f(X_n)|X_m, X_l\right)|X_m\right) \\
&\overset{1.}{=} \mathbb{E}\left(g(X_l)\mathbb{E}\left(f(X_n)|X_m\right)|X_m\right) \\
&= \mathbb{E}\left(g(X_l)|X_m\right)\mathbb{E}\left(f(X_n)|X_m\right),
\end{aligned}$$

where (1) is an application of the tower property of conditional expectations.
$(3 \Rightarrow 1)$. To show this direction we use test functions of the form $g(X_m)h(X_l)$ for $g, h \in \mathcal{B}_b$. More specifically,

$$\begin{aligned}
\mathbb{E}\left(f(X_n)g(X_l)h(X_m)\right) &= \mathbb{E}\left(h(X_m)\mathbb{E}\left(f(X_n)g(X_l)|X_m\right)\right) \\
&\overset{3.}{=} \mathbb{E}\left(h(X_m)\mathbb{E}\left(f(X_n)|X_m\right)\mathbb{E}\left(g(X_l)|X_m\right)\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(g(X_l)h(X_m)\mathbb{E}\left(f(X_n)|X_m\right)|X_m\right)\right) \\
&= \mathbb{E}\left(g(X_l)h(X_m)\mathbb{E}\left(f(X_n)|X_m\right)\right).
\end{aligned}$$

Therefore, we have that

$$\int_A f(X_n)\,\mathrm{d}\mathbb{P} = \int_A \mathbb{E}\left(f(X_n)|X_m\right)\,\mathrm{d}\mathbb{P}$$

for $A = A_1 \cap A_2$ where $A_1 \in \sigma(X_l)$ and $A_2 \in \sigma(X_m)$. This proves statement 1. □

Remark 2.3.2.

- *In other words, statement 3. says that the future of the process is independent of its past, provided we know the present.*
- *Theorem 2.3.1 generic stochastic processes. The statements are weaker than the Markov property as they only consider three points in time.*

Lemma 2.3.3. *Let $(X_n)_{n=0}^{\infty}$ be a Markov process, and suppose we have times $t_1 < \cdots < t_m = k$. Let $f, h \in \mathcal{B}_b(\mathcal{X})$, then*

$$\mathbb{E}\left(f(X_{k+2})h(X_{k+1})|X_{t_1}, \ldots, X_{t_m}\right) = \mathbb{E}\left(f(X_{k+2})h(X_{k+1})|X_{t_m}\right).$$

Proof. Let $\mathcal{G} = \sigma(X_{t_1}, \dots, X_{t_m})$. Then

$$\begin{aligned}
\mathbb{E}(f(X_{k+2})h(X_{k+1})|\mathcal{G}) &\overset{(1)}{=} \mathbb{E}\left(\mathbb{E}\left(\mathbb{E}\left(f(X_{k+2})h(X_{k+1})|\mathcal{F}^0_{k+1}\right)|\mathcal{F}^0_k\right)|\mathcal{G}\right) \\
&\overset{(2)}{=} \mathbb{E}\left(\mathbb{E}\left(\mathbb{E}\left(f(X_{k+2})h(X_{k+1})|\mathcal{F}^0_{k+1}\right)|X_k\right)|\mathcal{G}\right) \\
&\overset{(3)}{=} \mathbb{E}\left(\mathbb{E}\left(f(X_{k+2})h(X_{k+1})|X_k\right)|\mathcal{G}\right) \\
&\overset{(4)}{=} \mathbb{E}\left(f(X_{k+2})h(X_{k+1})|X_k\right),
\end{aligned}$$

where

- (1) comes from the tower property for conditional expectations applied to $\mathcal{F}^0_{k+1} \supset \mathcal{F}^0_k \supset \mathcal{G}$.
- (2) comes from the Markov property, as $\mathbb{E}\left(f(X_{k+2})h(X_{k+1})|\mathcal{F}^0_{k+1}\right)$ is a bounded measurable function of X_{k+1}, say $g(X_{k+1})$ such that $\mathbb{E}\left(g(X_{k+1})|\mathcal{F}^0_k\right) = \mathbb{E}(g(X_{k+1})|X_k)$.
- (3) comes from the tower property for conditional expectations applied to $\mathcal{F}^0_{k+1} \supset \sigma(X_k)$.
- (4) comes from the tower property for conditional expectations applied to $\mathcal{G} \supset \sigma(X_k)$.

□

Corollary 2.3.4. *Let $(X_n)_{n=0}^\infty$ be a Markov process, and suppose we have times $t_1 < \cdots < t_m = k$. Then for $A \in \mathcal{B}(\mathcal{X})$ we have that*

$$\mathbb{P}(X_{k+2} \in A|X_{t_1}, \dots, X_{t_m}) = \mathbb{P}(X_{k+2} \in A|X_{t_m}).$$

Proof. This follows by taking $h = 1$ and $f = \mathbf{1}_A$ in the previous lemma. □

Through induction, Corollary 2.3.4 larger difference time steps.

Exercise 2.3.5. *Let $(X_n)_{n=0}^\infty$ be a Markov process, and suppose we have times $s_1 < \cdots < s_m < t_1 < \cdots < t_n$ and functions $f_i \in \mathcal{B}_b(\mathcal{X})$ for $1 \leq i \leq n$. Then*

$$\mathbb{E}\left(\prod_{i=1}^n f_i(X_{t_i})|X_{s_1}, \dots, X_{s_m}\right) = \mathbb{E}\left(\prod_{i=1}^n f_i(X_{t_i})|X_{s_m}\right).$$

Proposition 2.3.6. *A process $(X_n)_{n=0}^\infty$ is a Markov process with respect to its natural filtration if and only if one of the following conditions holds.*

1. *For any $A_i \in \mathcal{B}(\mathcal{X})$ we have*

$$\mathbb{P}(X_0 \in A_0, \dots, X_n \in A_n) = \int_\Omega \mathbb{P}(X_n \in A_n|X_{n-1})(\omega)\mathbf{1}\left(\left\{(X_i)_{i=0}^{n-1} \in \prod_{i=0}^{n-1} A_i\right\}\right)(\omega)\,\mathrm{d}\mathbb{P}(\omega).$$

2. *For every $n \in \mathbb{N}$ and $f \in \mathcal{B}_b(\mathcal{X})$ we have*

$$\mathbb{E}\left(f(X_n)|\mathcal{F}^0_{n-1}\right) = \mathbb{E}\left(f(X_n)|X_{n-1}\right).$$

3. *For any $n \in \mathbb{N}$ and $f_i \in \mathcal{B}_b(\mathcal{X})$ we have*

$$\mathbb{E}\left(\prod_{i=1}^n f_i(X_i)\right) = \mathbb{E}\left(\prod_{i=1}^{n-1} f_i(X_i)\mathbb{E}(f_n(X_n)|X_{n-1})\right).$$

Proof. To prove statement 1. we utilize the formulation of the Markov property as conditional expectations. Let

$$C = \{X_0 \in A_0, \dots, X_{n-1} \in A_{n-1}\},$$

then the left-hand side of statement 1. is $\mathbb{E}(\mathbf{1}_C \mathbf{1}_{A_n}(X_n))$ and so we deduce that the Markov property holds for sets of the form C. Such sets form a π-system that generates $\mathcal{F}_{n-1}$, and so by Proposition 2.2.8 we conclude the Markov property holds for all sets in $\mathcal{F}_{n-1}$. The equivalence of statement 2. is given in Theorem 2.2.10. Statement 3. implies statement 1. and so we achieve the equivalence this way. To show that Markov property leads to statement 3. we apply the tower property of conditional expectation in the following way,

$$\begin{aligned}
\mathbb{E}\left(\prod_{i=1}^{n} f_i(x_i)\right) &= \mathbb{E}\left(\mathbb{E}\left(\prod_{i=1}^{n} f_i(X_i)\middle|\mathcal{F}_{n-1}\right)\right) \\
&= \mathbb{E}\left(\prod_{i=1}^{n-1} f_i(X_i)\mathbb{E}\left(f(X_n)|\mathcal{F}_{n-1}\right)\right) \\
&= \mathbb{E}\left(\prod_{i=1}^{n-1} f_i(X_i)\mathbb{E}\left(f(X_n)|X_{n-1}\right)\right).
\end{aligned}$$

□

Exercise 2.3.7.

1. *Suppose $(X_n)_{n=0}^{\infty}$ is a bounded $\mathbb{R}$-valued process which is Markov with respect to its natural filtration. Let $g \in \mathcal{B}_b(\mathbb{R})$ be a injective function. Show that $(Y_n)_{n=0}^{\infty}$ defined by $Y_n = g(X_n)$ is Markov with respect to its own natural filtration.*

2. *Show that the above statement is not true if we remove the assumption that g is injective.*

2.4 Solution to Exercises

Exercise 2.3.5

Solution. First note that the statement of Lemma 2.3.3 can be extended to multiple time steps in the future. For instance, one can consider say that

$$\mathbb{E}\left(\prod_{p=1}^{P} f_p(X_{k+p})|X_{t_1}, \dots, X_{t_m}\right) = \mathbb{E}\left(\prod_{p=1}^{P} f_p(X_{k+p})|X_{t_m}\right)$$

for each f_p a bounded and measurable function. The proof follows in the same way as Lemma 2.3.3, except we condition on $\mathcal{F}_{k+p}^0$ for $p = 1, \dots, P-1$. Once one has this result we can set the $f_i = 1$ when $i \neq t_j$ for some $j = 1, \dots, n$. Doing so we achieve the result of the exercise. □

Exercise 2.3.7

Solution.

1. Let $A \in \mathcal{B}(\mathbb{R})$. As g is injective we know that

$$\{Y_n \in A\} = \left\{X_n \in g^{-1}(A)\right\}.$$

Therefore,

$$\begin{aligned}
\mathbb{P}(Y_{n+1} \in A|X_0, \dots, X_n) &= \mathbb{P}\left(X_{n+1} \in g^{-1}(A)|X_0, \dots, X_n\right) \\
&\overset{(1)}{=} \mathbb{P}\left(X_{n+1} \in g^{-1}(A)|X_n\right) \\
&= \mathbb{P}(Y_{n+1} \in A|X_n),
\end{aligned}$$

where (1) is an application of the Markov property of (X_n) with respect to its natural filtration. Consequently, (Y_n) is Markov with respect to the natural filtration of (X_n) so by a discrete-time analogue of Proposition 2.2.9 (Y_n) is Markov with respect to its own natural filtration.

2. Consider the symmetric random walk, $(X_n)_{n=0}^{\infty}$, on $\mathbb{Z}$, which is a Markov process. Let
$$Y_n = g(X_n) = \left\lfloor \frac{X_n}{3} \right\rfloor.$$
This is clearly, not an injective transformation of X_n. Moreover, the event $\mathbb{P}(Y_n = 0 | Y_{n-1} = 1) > 0$ as we could have $X_{n-1} = 3$ and then $X_n = 2$. However, $\mathbb{P}(Y_n = 0 | Y_{n-1} = 1, Y_{n-2} = 2) = 0$ as $Y_{n-2} = 2$ implies that $X_{n-2} \in \{6, 7, 8\}$ but $Y_n = 0$ implies that $X_n \in \{0, 1, 2\}$. We cannot transition between these sets with two steps and hence the probability is zero. Note that the conditioning is well-defined as $\{Y_{n-1} = 1, Y_{n-2} = 2\}$ is an event with non-zero probability. Therefore, the process $(Y_n)_{n=0}^{\infty}$ is not Markovian.

□

3 The Kolmogorov Extension Theorem

3.1 Stochastic Processes as Random Variables

We can view a stochastic process $(X_i)_{i=0}^{\infty}$ as a $\mathcal{X}^{\mathbb{N}}$-valued random variable, instead of as a sequence of $\mathcal{X}$-valued random variables. Using this interpretation we have a map $\Omega \to \mathcal{X}^{\mathbb{N}}$ given by

$$\omega \mapsto (X_i(\omega))_{i=0}^{\infty}.$$

For an index set Λ, each $m \in \Lambda$ has a projection map $\pi_m : \prod_{i\in\Lambda} \mathcal{X}_i \to \mathcal{X}_m$ given by

$$\prod_{i\in\Lambda} \mathcal{X}_i \ni a = (a_i : a \in \Lambda) \mapsto \pi_m(a) = a_m \in \mathcal{X}_m.$$

Definition 3.1.1. *Given an index set Λ and measurable spaces $\left(\mathcal{X}_i, \mathcal{F}^{(i)}\right)$ for each $i \in \Lambda$, the product σ-algebra, denoted $\bigotimes_{i\in\Lambda} \mathcal{F}^{(i)}$, is defined to be the smallest σ-algebra on $\prod_{i\in\Lambda} \mathcal{X}_i$ such that all projection maps π_m for $m \in \Lambda$ are measurable. That is,*

$$\bigotimes_{i\in\Lambda} \mathcal{F}^{(i)} = \sigma\left(\left\{\pi_m^{-1}(A_m) : A_m \in \mathcal{F}^{(m)},\, m \in \Lambda\right\}\right).$$

Remark 3.1.2.

1. *Sets that are finite intersections of sets of the form $\pi_m^{-1}(A_m)$ with $m \in \Lambda$ and $A_m \in \mathcal{F}^{(m)}$ are called cylinder sets.*
   (a) *Cylinder sets generate $\bigotimes_{i\in\Lambda} \mathcal{F}^{(i)}$.*
   (b) *Cylinder sets $A \subset \prod_{i\in\Lambda} \mathcal{X}_i$ are of the form $A = \prod_{i\in\Lambda} A_i$ where each $A_i \in \mathcal{F}^{(i)}$ and all but finitely many $A_i \neq \mathcal{X}_i$.*
2. *When Λ is countable, we know that $\prod_{i\in\Lambda} E_i$ of measurable sets $E_i \in \mathcal{F}^{(i)}$ is measurable.*
3. *When Λ is uncountable, we cannot assume that $\prod_{i\in\Lambda}$ for $E_i \in \mathcal{F}^{(i)}$ is measurable.*

3.2 Constructing Stochastic Processes

Throughout, we will be working with a discrete-time stochastic process, and usually have $\prod_{i\in\in\Lambda} \mathcal{X}_i = \mathcal{X}^{\mathbb{N}}$ with $\mathcal{X}$ a complete metric space.

Proposition 3.2.1. *Suppose that we have a countable product of measurable spaces $\prod_{i=0}^{\infty} \mathcal{X}_i$ where each $\mathcal{X}_i$ is equipped with the σ-algebra $\mathcal{F}^{(i)}$. Also suppose for each i we have $\mathcal{F}^{(i)} = \sigma(\mathcal{D}_i)$, then*

$$\bigotimes_{i=0}^{\infty} \mathcal{F}^{(i)} = \sigma\left(\prod_{i=0}^{\infty} E_i : E_i \in \mathcal{D}_i\right).$$

If the $\mathcal{X}_i$ are separable metric spaces, then the Borel σ-algebra of the product topological space $\prod_{i=0}^{\infty} \mathcal{X}_i$ is the product of Borel σ-algebras of the $\mathcal{X}_i$, that is

$$\mathcal{B}\left(\prod_{i=0}^{\infty} \mathcal{X}_i\right) = \bigotimes_{i=0}^{\infty} \mathcal{B}(\mathcal{X}_i).$$

Corollary 3.2.2. *If $(X_n)_{n=0}^{\infty}$ is a $\mathcal{X}$-valued stochastic process, the whole sequence $(X_n)_{n=0}^{\infty}$ is itself a $\left(\mathcal{X}^{\mathbb{N}}, \mathcal{B}\left(\mathcal{X}^{\mathbb{N}}\right)\right)$ random variable. Moreover, the process $(X_n)_{n=0}^{\infty}$ induces a probability measure*

$$\mathrm{Law}\left((X_n)_{n=0}^{\infty}\right)$$

on $\mathcal{B}\left(\mathcal{X}^{\mathbb{N}}, \mathcal{B}\left(\mathcal{X}^{\mathbb{N}}\right)\right)$.

Supposing we have a stochastic process $(X_n)_{n=0}^{\infty}$, we construct a canonical probability space for the process as

$$\left(\mathcal{X}^{\mathbb{N}}, \mathcal{B}\left(\mathcal{X}^{\mathbb{N}}\right), \mathrm{Law}\left((X_n)_{n=0}^{\infty}\right)\right).$$

Therefore, if we are only interested in the process $(X_n)_{n=0}^{\infty}$ we can forget about $(\Omega, \mathcal{F}, \mathbb{P})$ and focus on this canonical space instead. As a $(X_n)_{n=0}^{\infty}$ can be defined as the random variable corresponding to the identity map on this space. Supposing we do not have a stochastic process, we can only guess what its law should look like. With the equivalence observation we made above, we can construct this process by constructing its canonical probability space. More specifically, if we know the values that $(X_n)_{n=0}^{\infty}$ takes in $\mathcal{X}^{\mathbb{N}}$, then we just need to be build $\mathrm{Law}\left((X_n)_{n=0}^{\infty}\right)$ on $\mathcal{B}\left(\mathcal{X}^{\mathbb{N}}\right)$.

Definition 3.2.3. *A finite dimensional distribution on the first $n+1$ time steps, $\{0, \ldots, n\}$, is a measure μ_n defined on $\left(\mathcal{X}^{n+1}, \mathcal{B}\left(\mathcal{X}^{n+1}\right)\right)$.*

Finite dimensional distributions are going to constitute our guess on what the law of our process should look like.

Definition 3.2.4. *Suppose we have a family of probability measures $(\mu_n)_{n=0}^{\infty}$ with each μ_n being a measure on $\left(\mathcal{X}^{n+1}, \mathcal{B}\left(\mathcal{X}^{n+1}\right)\right)$. We call the family $(\mu_n)_{n=0}^{\infty}$ consistent if for any $n \in \mathbb{N}$ and $A_0, \ldots, A_n \in \mathcal{B}(\mathcal{X})$ we have*

$$\mu_n(A_0 \times \cdots \times A_n) = \mu_{n+1}(A_0 \times \cdots \times A_n \times \mathcal{X}).$$

That is to say the measures are marginals of each other.

Example 3.2.5. *Let $(X_n)_{n=0}^{\infty}$ be a $\mathcal{X}$-valued stochastic process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If we set $\mu_n = \mathrm{Law}\left((X_j)_{j=0}^{n}\right)$, then $(\mu_n)_{n=0}^{\infty}$ is a consistent family of finite dimensional distributions.*

Remark 3.2.6. *Example 3.2.5 shows that in order for our guess of finite dimensional distributions to yield a valid stochastic process, we ought to ensure our guesses our consistent.*

Theorem 3.2.7. *Let $(\mu_n)_{n=0}^{\infty}$ be a consistent family of finite-dimensional distributions. Then there exists a unique probability measure μ on $\left(\mathcal{X}^{\mathbb{N}}, \mathcal{B}\left(\mathcal{X}^{\mathbb{N}}\right)\right)$ such that for any $n \in \mathbb{N}$ and $A \in \mathcal{B}\left(\mathcal{X}^{n+1}\right)$ we have*

$$\mu\left(A \times \left(\prod_{i=n+1}^{\infty} \mathcal{X}\right)\right) = \mu_n(A).$$

Corollary 3.2.8. *Given a stochastic process $(X_n)_{n=0}^{\infty}$, the family of laws*

$$\left(\mathrm{Law}\left(X_0, \ldots, X_n\right)\right)_{n=0}^{\infty}$$

uniquely characterises $\mathrm{Law}\left((X_n)_{n=0}^{\infty}\right)$.

From Theorem 3.2.7, we see that we can construct a stochastic process from a consistent family of finite-dimensional distributions.

Corollary 3.2.9. *Given any consistent family of finite dimensional distributions* $(\mu_n)_{n=0}^{\infty}$ *there exists a process* $(X_n)_{n=0}^{\infty}$ *with*

$$\mathrm{Law}(X_0, \dots, X_n) = \mu_n$$

for every $n \in \mathbb{N}$.

3.3 Stationarity

Working with sequences of random variables gives us a way to define stationarity.

Definition 3.3.1. *For each* $n \in \mathbb{N}$ *we define the shift map* $\theta_n : \mathcal{X}^{\mathbb{N}} \to \mathcal{X}^{\mathbb{N}}$ *by*

$$(a_0, a_1, \dots) \mapsto (a_n, a_{n+1}, \dots)$$

which is a $\mathcal{B}\left(\mathcal{X}^{\mathbb{N}}\right)$*-measurable function.*

Definition 3.3.2. *A stochastic process* $X = (X_n)_{n=0}^{\infty}$ *is stationary if for all* $n \in \mathbb{N}$ *the processes* $\theta_n X$ *and* X *have the same law.*

4 Transition Probabilities

Our method for constructing stochastic processes above does not have a way of guaranteeing the Markov property on the result sequence.

4.1 The Chapman-Kolmogorov Equation

Example 4.1.1. *The self-avoiding random walk can be shown to have consistent finite-dimensional distributions, however, it is not a Markov process.*

Definition 4.1.2. *A Markov process $(X_n)_{n=0}^{\infty}$ is time-homogeneous if for every $A \in \mathcal{B}(\mathcal{X})$ there is a choice of $P(\bullet, A) \in \mathcal{B}_b(\mathcal{X})$ such that*

$$\mathbb{P}(X_{n+1} \in A | X_n = \bullet) \sim P(\bullet, A)$$

modulo $\mathrm{Law}(X_n)$-null sets.

Remark 4.1.3.

1. *The main property of time-homogeneity is the independence of n. Note that many results still tend to hold without this assumption, however, the assumption simplifies notation.*
2. *The function $P(\bullet, A)$ describes the conditional probabilities and hence captures some information about the process. Eventually, we want to work back from this data and understand what properties $P(\bullet, A)$ needs to have to construct a Markov process.*

Definition 4.1.4. *A family $P = (P(x, A) : x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X}))$ is called a family of transition probabilities if the following two hold.*

1. *For each $x \in \mathcal{X}$, the function $P(x, \bullet)$ is a probability measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.*
2. *For each $A \in \mathcal{B}(\mathcal{X})$, the function $x \mapsto P(x, A)$ is Borel measurable.*

Remark 4.1.5. *Equivalently, we can say that there exists a measurable map $P : \mathcal{X} \to \mathcal{P}(\mathcal{X})$ such that for all $A \in \mathcal{B}(\mathcal{X})$ and $x \in \mathcal{X}$ we have*

$$(P(x))(A) = P(x, A).$$

Exercise 4.1.6. *Consider the random dynamical system as defined in Theorem 2.1.1 but with $F_n = F$ for all $n \in \mathbb{N}$. That is, $X_{n+1} = F(X_n, \zeta_{n+1})$ where the $(\zeta_n)_{n=0}^{\infty}$ are $\mathrm{i.i.d}$ random variables also independent of X_0. Let $\mathrm{Law}(X_0) = \nu$ and $\mathrm{Law}(\zeta_n) = \mu$. Show that $X = (X_n)_{n=0}^{\infty}$ is a time-homogeneous Markov process and compute its transition probabilities.*

Exercise 4.1.7. *Let P be the set of transition probability built from a time-homogeneous Markov process $X = (X_n)_{n=0}^{\infty}$. Show that $(X_{2n})_{n=0}^{\infty}$ and $(X_{3n})_{n=0}^{\infty}$ are both time-homogeneous Markov processes and compute their transition probabilities P^2 and P^3. Show that for all $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$ we have*

$$P^2(x, A) = \int_{\mathcal{X}} P(y, A) P(x, \mathrm{d}y) \tag{4.1.1}$$

and

$$P^3(x,A) = \int_{\mathcal{X}} P(y,A)P^2(x,\mathrm{d}y) = \int_{\mathcal{X}} P^2(y,A)P(x,\mathrm{d}y). \tag{4.1.2}$$

Equations (4.1.1) *and* (4.1.2) *are instances of the Chapman-Kolmogorov equation.*

Recall, that we think about $P(x,A)$ as encoding conditional probabilities. That is,

$$P(x,A) = \mathbb{P}(X_{n+1} \in A | X_n = x).$$

As we are operating in the time-homogeneous setting, we can consider $P(x,A)$ as a one-step conditional probability.

Definition 4.1.8. *A sequence of transition probabilities $(P^n)_{n=0}^{\infty}$ is called a transition function if the following statements hold.*

1. *$P^0(x,\bullet) = \delta_x$. Here δ_x is the Dirac measure at x. That is,*

$$\delta_x(A) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases}$$

2. *The family satisfies the Chapman-Kolmogorov equations, that is, for every $n,m \geq 0$, $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$ we have that*

$$P^{m+n}(x,A) = \int_{\mathcal{X}} P^n(y,A)P^m(x,\mathrm{d}y).$$

Remark 4.1.9.

1. *We can intuitively think of P^n as encoding an n-step conditional probability.*
2. *The Chapman-Kolmogorov equation says that a step into a set is consistent with taking smaller intermediary steps to get to the set.*

Due to the Chapman-Kolmogorov condition, we can build a family of transition probabilities by starting from a one-step conditional probability, P. More specifically, given any transition probability, P, we can construct a transition function $(P^n)_{n=0}^{\infty}$, by

1. setting $P^0(x,\bullet) = \delta_x$,
2. $P^1(x,\bullet) = P(x,\bullet)$, and
3. for $n > 1$ set

$$P^n(x,A) = \int_{\mathcal{X}} P(y,A)P^{n-1}(x,\mathrm{d}y).$$

4.2 Constructing Markov Processes

Theorem 4.2.1. *Let $(X_n)_{n=0}^{\infty}$ be a time-homogeneous Markov process with transition probability P. Let $(P^n)_{n=0}^{\infty}$ be the transition function built from P. Then the following statements hold.*

- *For any $n,m \geq 0$ and $f \in \mathcal{B}_b(\mathcal{X})$ we have*

$$\mathbb{E}\left(f(X_{n+m}) | X_m\right) = \int_{\mathcal{X}} f(y)P^n(X_m,\mathrm{d}y).$$

- *If $X_0 \sim \mu$, then for any $n \geq 0$ and $f \in \mathcal{B}_b(\mathcal{X})$ we have*

$$\mathbb{E}(f(X_n)) = \int_{\mathcal{X}} \int_{\mathcal{X}} f(y) P^n(x, \mathrm{d}y) \mu(\mathrm{d}x).$$

Proof. To prove the first statement we can induct on n.

- The base case $n = 0$ follows directly from

$$\begin{aligned} \mathbb{E}\left(f(X_m)|X_m\right) &= f(X_m) \\ &= \int_{\mathcal{X}} f(y) P^0(X_m, \mathrm{d}y). \end{aligned}$$

- For the inductive step we suppose the result holds for $n \leq k$. Then,

$$\begin{aligned} \mathbb{E}\left(f(X_{k+1+m})|X_m\right) &\overset{\text{Tow Prop.}}{=} \mathbb{E}\left(\mathbb{E}\left(f(X_{k+1+m})|\mathcal{F}^0_{k+m}\right)|X_m\right) \\ &\overset{\text{Markov.}}{=} \mathbb{E}\left(\mathbb{E}\left(f(X_{k+1+m})|X_{k+m}\right)|X_m\right) \\ &\overset{\text{Time Hom.}}{=} \mathbb{E}\left(\int_{\mathcal{X}} f(y) P(X_{m+k}, \mathrm{d}y)\Big|X_m\right) \\ &\overset{\text{Ind Hyp.}}{=} \int_{\mathcal{X}} \int_{\mathcal{X}} f(y) P(z, \mathrm{d}y) P^k(X_m, \mathrm{d}z) \\ &\overset{\text{Fubini.}}{=} \int_{\mathcal{X}} f(y) P^{k+1}(X_m, \mathrm{d}y). \end{aligned}$$

Therefore, by induction, we complete the proof of the first statement. For the second statement we proceed directly with,

$$\begin{aligned} \mathbb{E}(f(X_n)) &= \mathbb{E}\left(\mathbb{E}(f(X_n)|X_0)\right) \\ &= \mathbb{E}\left(\int_{\mathcal{X}} f(y) P^n(X_0, \mathrm{d}y)\right) \\ &= \int_{\mathcal{X}} \int_{\mathcal{X}} f(y) P^n(x, \mathrm{d}y) \mu(\mathrm{d}x). \end{aligned}$$

Where the last equality follows from the fact that X_0 is the only random variable in the expectation, and we know the law of X_0 to be μ. $\square$

Remark 4.2.2.

1. *For clarity we can take $f = \mathbf{1}_A$ with $A \in \mathcal{B}(\mathcal{X})$ so that the statements of Theorem 4.2.1 reduce to*
   - $\mathbb{P}(X_{n+m} \in A|X_m) = P^n(X_m, A)$*, and*
   - $\mathbb{P}(X_n \in A) = \int_{\mathcal{X}} P^n(x, A) \mu(\mathrm{d}x)$.
2. *X being a Markov process is a sufficient condition for the conclusions Theorem 4.2.1, but it is not a necessary condition.*

Proposition 4.2.3. *Let $X = (X_n)_{n=0}^{\infty}$ be a process with $\mu = \mathrm{Law}(X_0)$ and transition probability P, then X*

is Markov if and only if for all n and $f_i \in \mathcal{B}_b(\mathcal{X})$ we have

$$\mathbb{E}\left(\prod_{i=0}^{n} f_i(X_i)\right) = \underbrace{\int_{\mathcal{X}} \cdots \int_{\mathcal{X}}}_{n+1} \prod_{i=0}^{n} f_i(y_i) \prod_{i=0}^{n-1} P(y_i, \mathrm{d}y_{i+1})\mu(\mathrm{d}y_0).$$

Proof. We proceed by induction on n. For $n = 0$, as we know $\mathrm{Law}(X_0) = \mu$ it follows that

$$\mathbb{E}\left(f_0(X_0)\right) = \int_{\mathcal{X}} f(y)\mu(\mathrm{d}y_0).$$

Now suppose the result holds for $k \leq n-1$. Then

$$\begin{aligned}
\mathbb{E}\left(\prod_{i=1}^{n} f_i(X_i)\right) &\overset{\text{Tower.}}{=} \mathbb{E}\left(\mathbb{E}\left(\prod_{i=0}^{n} f_i(X_i)|\mathcal{F}_{n-1}\right)\right) \\
&= \mathbb{E}\left(\prod_{i=0}^{n-1} f_i(X_i)\mathbb{E}\left(f_n(X_n)|\mathcal{F}_{n-1}\right)\right) \\
&\overset{\text{Markov.}}{=} \mathbb{E}\left(\prod_{i=0}^{n-1} f_i(X_i)\mathbb{E}\left(f_n(X_n)|X_{n-1}\right)\right) \\
&\overset{(1)}{=} \mathbb{E}\left(\prod_{i=0}^{n-1} f_i(X_i)\int_{\mathcal{X}} f_n(y_n)P(X_{n-1}, \mathrm{d}y_n)\right) \\
&\overset{\text{Ind Hyp.}}{=} \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} \left(\prod_{i=0}^{n-1} f_i(y_i)\int_{\mathcal{X}} f_n(y_n)P(y_{n-1}, \mathrm{d}y_n)\right) \prod_{i=0}^{n-2} P(y_i, \mathrm{d}_{y_{i+1}})\mu(\mathrm{d}y_0) \\
&= \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} \prod_{i=0}^{n} f_i(y_i) \prod_{i=0}^{n-1} P(y_i, \mathrm{d}_{y_{i+1}})\mu(dy_0).
\end{aligned}$$

Where (1) follows as the law of the conditional one-step probabilities is given by P. For the converse one refers to Proposition 2.3.6. ☐

By taking each f_i to be an indicator function we arrive at the following.

Corollary 4.2.4. *Let $X = (X_n)_{n=0}^{\infty}$ be a process with $\mu = \mathrm{Law}(X_0)$ and transition probability P. Then X is Markov if and only if for all n and $A_i \in \mathcal{B}_b(\mathcal{X})$, we have*

$$\mathbb{P}(X_0 \in A_0, \ldots, X_n \in A_n) = \int_{A_0} \cdots \int_{A_n} \prod_{i=0}^{n-1} P(y_i, \mathrm{d}y_{i+1})\mu(\mathrm{d}y_0).$$

Remark 4.2.5. *As the transition probability determines all one-step conditional probabilities, the Markov process of Corollary 4.2.4 is time-homogeneous.*

Given P and μ for a Markov process X we can compute the finite-dimensional distributions of the process.

Proposition 4.2.6. *Given a transition probability P and measure μ on $\mathcal{X}$, there exists a unique (up to law) Markov process $X = (X_n)_{n=0}^{\infty}$ with transition probabilities P and $\mathrm{Law}(X_0) = \mu$.*

Proof. Consider the sequence of measures μ_n defined on $\mathcal{X}^n$ by

$$\mu_n(A_0 \times \cdots \times A_{n-1}) = \int_{A_0} \cdots \int_{A_{n-1}} \prod_{i=1}^{n} P(y_{i-1}, \mathrm{d}y_i)\mu(\mathrm{d}y_0).$$

From this definition it is clear that

$$
\begin{aligned}
\mu_{n+1}(A_0 \times \cdots \times A_{n-1} \times \mathcal{X}) &= \int_{A_0} \cdots \int_{A_{n-1}} \int_{\mathcal{X}} \prod_{i=1}^{n+1} P(y_{i-1}, \mathrm{d}y_i)\mu(\mathrm{d}y_0) \\
&= \int_{A_0} \cdots \int_{A_{n-1}} \int_{\mathcal{X}} P(y_n, \mathrm{d}y_{n+1}) \prod_{i=1}^{n-1} P(y_{i-1}, \mathrm{d}y_i)\mu(\mathrm{d}y_0) \\
&= \int_{A_0} \cdots \int_{A_{n-1}} \prod_{i=1}^{n} P(y_{i-1}, \mathrm{d}y_i)\mu(\mathrm{d}y_0).
\end{aligned}
$$

Where the last equality follows from the fact the $P(y_n, \mathrm{d}y_{n+1})$ defines a probability measure. Therefore, the $(\mu_n)_{n\in\mathbb{N}}$ form a consistent sequence of measure, and so we can apply Kolmogorov's extension theorem to conclude that there exists a unique measure $\mathbb{P}_\mu$ on $\mathcal{X}^{(\infty)}$ such that its restriction to $\mathcal{X}^n$ is μ_n. Now choosing $\Omega = \mathcal{X}^\infty$ and our probability measure to be $\mathbb{P}_\mu$, the canonical process $(\pi_n)_{n\in\mathbb{N}}$ with $\pi_n((\omega_0, \omega_1, \dots)) = \omega_n$ is such that

$$
\mathbb{P}(\pi_0 \in A_0, \dots, \pi_n \in A_{n+1}) = \mathbb{P}_\mu(A_0 \times \cdots \times A_n \times \mathcal{X}^\infty) = \mu_n(A_0 \times \cdots \times A_n).
$$

Therefore, $(\pi_n)_{n\in\mathbb{N}}$ has finite dimensional distributions μ_n. Therefore, by Corollary 4.2.4 the process $(\pi_n)_{n\in\mathbb{N}}$ is a Markov process with transition probability P and $\mathrm{Law}(X_0) = \mu$. □

Remark 4.2.7.

1. *When $\mu = \delta_x$ with $x \in \mathcal{X}$ we write $\mathbb{P}_x$ instead of $\mathbb{P}_\mu$. Similarly, we write $\mathbb{E}_x$ or $\mathbb{E}_\mu$ to denote the expectation over $\left(\mathcal{X}^{\mathbb{N}}, \mathcal{B}\left(\mathcal{X}^{\mathbb{N}}\right)\right)$ with respect to $\mathbb{P}_x$ and $\mathbb{P}_\mu$ respectively.*
2. *Note that by construction the Markov process of Proposition 4.2.6 is time-homogeneous.*

We can see P as propagating the law μ through time. Hence, we can view the transition probability as an operator on $\mathcal{P}(\mathcal{X})$.

Definition 4.2.8. *Given a transition probability P, we let $T^* : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$ be the operator where*

$$
\mu \mapsto (T^*\mu) = \int_{\mathcal{X}} P(x, \bullet)\mu(\mathrm{d}x)
$$

where

$$
(T^*\mu)(A) = \int_{\mathcal{X}} P(x, A)\mu(\mathrm{d}x).
$$

Exercise 4.2.9. *Formalise the intuition that P propagates the law of X_0 through time. That is, in the context of the Proposition 4.2.6 show that*

$$
\mathrm{Law}(X_n) = T\left(\dots(T(\mathrm{Law}(X_0)))\dots\right) = T^n\mathrm{Law}(X_0).
$$

Definition 4.2.10. *A measure μ is invariant for the transition probability P if $T\mu = \mu$.*

Remark 4.2.11. *Let π be an invariant measure of P, then $\mathbb{P}_\pi$, constructed by applying Proposition 4.2.6 to π, is the law of a stationary process.*

Above we have considered P as an operator on measures. We can also view P as an operator on $\mathcal{B}_b(\mathcal{X})$.

Definition 4.2.12. *Let P be a transition probability, we let $T_* : \mathcal{B}_b(\mathcal{X}) \to \mathcal{B}_b(\mathcal{X})$ be the operator where*

$$f \mapsto (T_* f) = \int_{\mathcal{X}} f(y) P(\bullet, \mathrm{d}y)$$

where

$$(T_* f)(x) = \int_{\mathcal{X}} f(y) P(x, \mathrm{d}y).$$

Exercise 4.2.13. *For a fixed transition probability P, the operators T^* and T_* are dual. That is, for any $f \in \mathcal{B}_b(\mathcal{X})$ and $\mu \in \mathcal{P}(\mathcal{X})$ we have*

$$\int_{\mathcal{X}} (T_* f)(x) \mu(\mathrm{d}x) = \int_{\mathcal{X}} f(x) \left(T^* \mu\right)(\mathrm{d}x).$$

Remark 4.2.14. *Henceforth, we will often simply write $T^* = T$.*

4.3 Solution to Exercises

Exercise 4.1.6

Solution. We observe that

$$\begin{aligned}
\mathbb{P}(X_{n+1} \in A | X_n) &= \mathbb{E}\left(\mathbf{1}_A(F(X_n, \zeta_{n+1}))\right) \\
&= \int_{\mathcal{Y}} \mathbf{1}_A F(X_n(\omega), y) \mathrm{d}\mu(dy) \\
&= \mu\left(\{y \in \mathcal{Y} : F(x, y) \in A\}\right).
\end{aligned}$$

Therefore, the process is a time-homogeneous Markov process. Moreover, $P(x, A) = \mu\left(\{y \in \mathcal{Y} : F(x, y) \in A\}\right)$ because by the measurability assumptions on F, one can show that $P(x, \bullet)$ satisfies the required measurability properties. ☐

Exercise 4.1.7

Solution. We first show that the process $(X_{2n})_{n=0}^\infty$ and $(X_{3n})_{n=0}^\infty$ are Markov processes.

$$\mathbb{P}(X_{2n+2} \in A | X_{2n}, X_{2n-2}, \dots, X_0) \overset{\text{Cor. } 2.3.4}{=} \mathbb{P}(X_{2n+2} \in A | X_{2n}).$$

Similarly, for $(X_{3n})_{n=0}^\infty$. Now we calculate the transition probability for the two-step process.

$$\begin{aligned}
\mathbb{P}(X_{2n+2} \in A | X_{2n}) &\overset{(1)}{=} \mathbb{E}(\mathbb{P}(X_{2n+2} | X_{2n+1}, X_{2n}) | X_{2n}) \\
&\overset{(2)}{=} \mathbb{E}(\mathbb{P}(X_{2n+2} | X_{2n+1}) | X_{2n}) \\
&= \mathbb{E}(P(X_{2n+1}, A) | X_{2n}) \\
&\overset{(3)}{=} \int_{\mathcal{X}} P(y, A) P(X_{2n}, \mathrm{d}y)
\end{aligned}$$

where (1) is an application of the tower rule and (2) is applying the Markov property of $(X_n)_{n=0}^\infty$. Therefore,

$$\mathbb{P}(X_{2n+2} \in A | X_{2n} = x) = \int_{\mathcal{X}} P(y, A) P(x, \mathrm{d}y).$$

In (3) we have used the fact that for f a measurable function we have

$$\mathbb{E}(f(X_{n+1}) | X_n = x) = \int_{\mathcal{X}} f(y) P(x, \mathrm{d}y).$$

The transition probability for the three-step process follows similarly, with the different formulations arising from how the tower rule is applied. $\square$

Exercise 4.2.9

Solution. Let μ_0 be the law of X_0. Then the law for X_1, denoted μ_1, is

$$\mu_1(A) = \int_{\mathcal{X}} P(x, A)\mu_0(\mathrm{d}x) = T(\mu_0)(A).$$

Similarly, the law for X_2, denoted μ_2, is

$$\begin{aligned}
\mu_2(A) &= \int_{\mathcal{X}} \int_{\mathcal{X}} P(y, A)P(x, \mathrm{d}y)\mu_0(\mathrm{d}x) \\
&= \int_{\mathcal{X}} P(y, A)\mu_1(\mathrm{d}y) \\
&= T(\mu_1)(A) \\
&= T^2(\mu_0)(A).
\end{aligned}$$

Continuing by induction completes the proof. $\square$

Exercise 4.2.13

Proof. We proceed directly, noting that the equivalence arises from changing the order of integration,

$$\begin{aligned}
\int_{x\in\mathcal{X}} f(x)(T\mu)(\mathrm{d}x) &= \int_{x\in\mathcal{X}} f(x) \int_{y\in\mathcal{X}} P(y, \mathrm{d}x)\mu(dy) \\
&= \int_{x\in\mathcal{X}} \int_{y\in\mathcal{X}} f(x)P(y, \mathrm{d}x)\mu(dy) \\
&= \int_{x\in\mathcal{X}} (T_\star f)(y)\mu(\mathrm{d}y).
\end{aligned}$$

$\square$

5 Discrete Time Markov Processes

5.1 Time-Homogeneous Processes

On a discrete space, a probability measure is just a, potentially infinite, vector

$$\mu(\mathrm{d}x) \longrightarrow \mu = (\mu_j : j \in \mathcal{X})$$

with $\mu_j \geq 0$ and $\sum_{j\in\mathcal{X}} \mu_j = 1$. In such a case, the transition probability P can be represented as a, potentially infinite, matrix

$$P(i, \{j\}) \longrightarrow P = (P_{ij} : i, j \in \mathcal{X})$$

where $P_{ij} \geq 0$ and $\sum_{j\in\mathcal{X}} P_{ij} = 1$. Note that if $X = (X_n)_{n=0}^{\infty}$ is a time-homogeneous Markov process with transition probability P then

$$\mathbb{P}(X_{n+1} = j | X_n = i) = P_{ij}.$$

Definition 5.1.1. *A matrix $P = (P_{ij} : i, j \in \mathcal{X})$ is called a stochastic matrix if*

1. *$P_{ij} \geq 0$, and*
2. *$\sum_{j\in\mathcal{X}} P_{ij} = 1$.*

For discrete state spaces $\mathcal{X}$, we have a direct correspondence between transition probabilities and stochastic matrices.

Exercise 5.1.2. *Previously it was for shown that from a transition probability P, by setting $P^0(x, \bullet) = \delta_x$, we can construct a transition function by letting*

$$P^n(x, A) = \int_{\mathcal{X}} P(y, A) P^{n-1}(x, \mathrm{d}y).$$

In discrete spaces, show that $P^0 = I$, where I is the matrix with entries $I_{ij} = \delta_{ij}$, and

$$P^n = \underbrace{P \times \cdots \times P}_{n}$$

where $\times$ is matrix multiplication. Moreover, verify that P^n is also a stochastic matrix.

With this, we can formulate our previous results in the specific context of a discrete state space.

Theorem 5.1.3. *Let $(X_n)_{n=0}^{\infty}$ be a time-homogeneous discrete Markov process with stochastic matrix P. Then the following statements hold.*

- *For any $n, m \geq 0$ and $f \in \mathcal{B}_b(\mathcal{X})$, we have*

$$\mathbb{E}(f(X_{n+m}) | X_m = i) = \sum_{j\in\mathcal{X}} P_{ij}^n f(j).$$

- *If $X_0 \sim \mu$, then for any $n \geq 0$ and $f \in \mathcal{B}_b(\mathcal{X})$ we have*

$$\mathbb{E}(f(X_n)) = \sum_{i,j\in\mathcal{X}} \mu_i P_{ij}^n f(j).$$

Again, we can contextualises these statements by letting $f = \mathbf{1}_A$ for $A \in \mathcal{B}(\mathcal{X})$. In this case

- $\mathbb{P}(X_{n+m} = j | X_m = i) = P_{ij}^n$, and
- $\mathbb{P}(X_n = j) = \sum_{i\in\mathcal{X}} \mu_i P_{ij}^n$.

Proposition 5.1.4. *Let $X = (X_n)_{n=0}^{\infty}$ be a discrete process with $\mu = \mathrm{Law}(X_0)$ and stochastic matrix P. Then X is Markov if and only if for all n we have*

$$\mathbb{E}\left(\prod_{m=0}^{n} f_m(X_m)\right) = \sum_{i_0,\dots,i_n \in \mathcal{X}} f_0(i_0)\dots f_n(i_n)\mu_{i_0}P_{i_0 i_1}\dots P_{i_{n-1}i_n},$$

where each $f_m \in \mathcal{B}_b(\mathcal{X})$.

Corollary 5.1.5. *Let $X = (X_n)_{n=0}^{\infty}$ be a discrete process with $\mu = \mathrm{Law}(X_0)$ and transition probability P. Then X is Markov if and only if for all n and $i_0, \dots, i_n \in \mathcal{X}$ we have*

$$\mathbb{P}(X_0 = i_0, \dots, X_n \in i_n) = \mu_{i_0}P_{i_0,i_1}\dots P_{i_{n-1}i_n}.$$

In the discrete settings the operators $T : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$ manifests as

$$(T\mu)(\{j\}) = \sum_{i\in\mathcal{X}} \mu_i P_{ij} = (\mu P)_j.$$

Similarly, the operator $T_* : \mathcal{B}_b(\mathcal{X}) \to \mathcal{B}_b(\mathcal{X})$ manifests as

$$(T_* f)(i) = \sum_{j\in\mathcal{X}} P_{ij} f_j = (Pf)_i.$$

Therefore, in the discrete setting, the behaviour of our process as $n \to \infty$ can be reduced to understanding the behaviour of P^n as $n \to \infty$.

5.2 Stopping Times

Definition 5.2.1. *Given a filtration $(\mathcal{F}_n)_{n=0}^{\infty}$, a $\mathbb{N}\cup\{\infty\}$-valued random variable T is called an $(\mathcal{F}_n)_{n=0}^{\infty}$-stopping time if for every $n \in \mathbb{N}$ we have that $\{T \le n\} \in \mathcal{F}_n$.*

Recalling that a filtration tells us what information we have at time n, the above says that by time n we have enough information to determine whether T has occurred yet. One can think of T as a random alarm clock that cannot look into the future. By induction, the above is equivalent to $\{T = n\} \in \mathcal{F}_n$ or all $n \in \mathbb{N}$. One can interchange between using $\{T = n\}$ and $\{T \le n\}$ when working with stopping times. Each form is useful in different contexts. We include ∞ in this definition to allow the possibility that the "alarm clock" T never rings.

Exercise 5.2.2. *Let $(X_n)_{n=0}^{\infty}$ be a stochastic process, let $A \in \mathcal{B}(\mathcal{X})$. Show that $\tau_A = \inf\{n \in \mathbb{N} : X_n \in A\}$ is a $\left(\mathcal{F}_n^0\right)_{n=0}^{\infty}$-stopping time. The random variable τ_A is also called the hitting time of A.*

Example 5.2.3.

1. *A deterministic time T is a stopping time, including $T = \infty$.*
2. *In general, for $A \in \mathcal{B}(\mathcal{X})$, the random variable $\ell_A = \sup\{n \ge 0 : X_n \in A\}$ is not a stopping time. Intuitively, the value of ℓ_A is dependent on the future.*

Definition 5.2.4. *Given a stochastic process $(X_n)_{n=0}^{\infty}$ and a $\mathbb{N}\cup\{\infty\}$-valued random variable, we define the* *stopped process* $\left(X_n^T\right)_{n=0}^{\infty}$ *by*

$$X_n^T(\omega) = X_{n\wedge T}(\omega) = \begin{cases} X_n(\omega) & n \le T(\omega) \\ X_{T(\omega)}(\omega) & \textit{otherwise.} \end{cases}$$

Exercise 5.2.5. *Let* $(\mathcal{F}_n)_{n=0}^{\infty}$ *be a filtration,* T *be a* $(\mathcal{F}_n)_{n=0}^{\infty}$*-stopping time, and* $(X_n)_{n=0}^{\infty}$ *be a* $(\mathcal{F}_n)_{n=0}^{\infty}$*-adapted process. Show that* $\left(X_n^T\right)_{n=0}^{\infty}$ *is adapted to* $(\mathcal{F}_n)_{n=0}^{\infty}$.

Remark 5.2.6. *When filtrations are not mentioned in a statement, just assume there is some fixed filtration operating in the background.*

Proposition 5.2.7. *Let* S *and* T *be stopping times, and let* $(T_n)_{n=0}^{\infty}$ *be a sequence of stopping times. Then the following hold.*

1. $S \vee T$ *and* $S \wedge T$ *are stopping times.*
2. $\sup_n T_n$, $\inf_n T_n$, $\liminf_{n\to\infty} T_n$ *and* $\limsup_{n\to\infty} T_n$ *are stopping times.*

Proof.

1. Note that
$$\{S \vee T \le n\} = \{S \le n\} \cap \{T \le n\} \in \mathcal{F}_n$$
and
$$\{S \wedge T \le n\} = \{S \le n\} \cup \{T \le n\} \in \mathcal{F}_n.$$
2. It suffices to show that $\sup_n T_n$ and $\inf_n T_n$ are stopping times as $\liminf_{n\to\infty} T_n = \sup_n \inf_{m\ge n} T_m$ and $\limsup_{n\to\infty} T_n = \inf_n \sup_{m\ge n} T_n$. We proceed in the same way as above
$$\left\{\sup_j T_j \le n\right\} = \bigcap_{j=0}^{\infty} \{T_j \le n\} \in \mathcal{F}_n$$
and
$$\left\{\inf_j T_j \le n\right\} = \bigcup_{j=0}^{\infty} \{T_j \le n\} \in \mathcal{F}^n.$$

□

Example 5.2.8. *If* S *and* T *are stopping times, then it is not necessarily the case that* $(S-T)_+$ *is a stopping time. Suppose that* $S = T = n$ *then* $(S-T)_+ = 0$*, but as* $n > 0$ *this implies that the random variable* $(S-T)_+$ *is dependent on future information.*

Given a $(\mathcal{F}_n)_{n=0}^{\infty}$-stopping time T we can define a σ-algebra that represents the information we have a time T. Let $\mathcal{F}_\infty = \bigvee_{n=0}^{\infty} \mathcal{F}_n$, that is the σ-algebra generated by $\bigcup_{n=0}^{\infty} \mathcal{F}_n$.

Exercise 5.2.9. *Let* T *be a* $(\mathcal{F}_n)_{n=0}^{\infty}$*-stopping time, show that* $\{T = \infty\} \in \mathcal{F}_\infty$.

Definition 5.2.10. *For a* $(\mathcal{F}_n)_{n=0}^{\infty}$*-stopping time* T*, define the stopped* σ*-algebra*

$$\mathcal{F}_T = \{A \in \mathcal{F}_\infty : \textit{for all } n \in \mathbb{N},\ A \cap \{T = n\} \in \mathcal{F}_n\}.$$

One can think of the above definition as saying that for an event A, conditioned on the event $\{T = n\}$ we could say whether A has occurred with the information we have up to time n. The event $\{T = n\}$ in the above definition can be replaced with $\{T \leq n\}$, as we operate in the discrete setting.

Lemma 5.2.11. *If T is a $(\mathcal{F}_n)_{n=0}^{\infty}$-stopping time, then T is $\mathcal{F}_T$ measurable.*

Proof. Note that for any $m \in \mathbb{N}$ and all $n \in \mathbb{N}$ we have

$$\{T = m\} \cap \{T = n\} = \begin{cases} \emptyset & m \neq n \\ \{T = n\} & m = n. \end{cases}$$

In the first case $\{T = m\} \cap \{T = n\} \in \mathcal{F}_n$ trivially, as $\mathcal{F}_n$ is a σ-algebra. In the second case, as T is a $\mathcal{F}_n$ stopping time, it must be the case that $\{T = n\} \in \mathcal{F}_n$ which implies that $\{T = m\} \cap \{T = n\} \in \mathcal{F}_n$. Therefore, $\{T = m\} \in \mathcal{F}_T$ for all m, meaning that T is $\mathcal{F}_T$-measurable. $\square$

Exercise 5.2.12. *Let S and T be $(\mathcal{F}_n)_{n=0}^{\infty}$-stopping times.*

1. *Show that if $S \leq T$, then $\mathcal{F}_S \subset \mathcal{F}_T$.*
2. *Suppose $S \leq T$ and $A \in \mathcal{F}_S$. Show then that $S\mathbf{1}_A + T\mathbf{1}_{A^c}$ is a $(\mathcal{F}_n)_{n=0}^{\infty}$-stopping time.*
3. *Show that $\{A \cap \{S \leq T\} : A \in \mathcal{F}_S\} \subset \mathcal{F}_{S \wedge T}$.*
4. *For X a bounded random variable, show that*

$$\mathbb{E}\left(\mathbb{E}\left(X|\mathcal{F}_T\right)|\mathcal{F}_S\right) = \mathbb{E}\left(\mathbb{E}\left(X|\mathcal{F}_S\right)|\mathcal{F}_T\right) = \mathbb{E}\left(X|\mathcal{F}_{S\wedge T}\right).$$

Definition 5.2.13. *A stopping time T is finite if $\mathbb{P}(\{T < \infty\}) = 1$. When we have the stronger condition that $\{T < \infty\} = \emptyset$ we write this as $T < \infty$.*

If $T < \infty$, then for a stochastic process $X = (X_n)_{n=0}^{\infty}$ the random variable X_T is well-defined.

Lemma 5.2.14. *Let $(X_n)_{n=0}^{\infty}$ be adapted and let T be a stopping time, then for any $m \in \mathbb{N}$, the random variable $X_{T\wedge m}$ is $\mathcal{F}_T$-measurable. Moreover, suppose that $T < \infty$, then X_T is $\mathcal{F}_T$-measurable.*

Proof. Let $A \in \mathcal{B}(\mathcal{X})$, then for $m, n \in \mathbb{N}$ we have

$$\left\{X_{T(\cdot)\wedge m}(\cdot)\right\} \cap \{T = n\} = \{X_{n\wedge m} \in A\} \cap \{T = n\}.$$

The first set on the right-hand side is in $\mathcal{F}_{n\wedge m} \subset \mathcal{F}_n$. The second set on the left-hand side is in $\mathcal{F}_n$. Therefore, the intersection on the right-hand side is in $\mathcal{F}_n$ which implies that $X_{T\wedge m}$ is $\mathcal{F}_T$-measurable. One shows in a similar way that X_T is $\mathcal{F}_T$-measurable. $\square$

Recall, that $\left(\mathcal{F}_n^0\right)_{n=0}^{\infty}$ denotes the natural filtration of a stochastic process $X = (X_n)_{n=0}^{\infty}$.

Lemma 5.2.15. *Let $(X_n)_{n=0}^{\infty}$ be a stochastic process and let $T < \infty$ be a $\left(\mathcal{F}_n^0\right)_{n=0}^{\infty}$-stopping time. Then for any $k \in \mathbb{N}$ we have*

$$\{T = k\} \in \sigma(X_{T\wedge 0}, \ldots, X_{T\wedge k}).$$

Proof. We proceed by induction on k. For $k = 0$ it is clear that

$$\{T = 0\} \in \mathcal{F}_0^0 \in \sigma(X_0) = \sigma\left(X_{T\wedge 0}\right).$$

Suppose the result holds for $n \leq k-1$. Then

$$\begin{aligned}
\mathbf{1}_{\{T=k\}} &= \mathbf{1}_{\{T=k\}}\mathbf{1}_{\{T>k-1\}} \\
&\overset{(1)}{=} \tilde{\varphi}(X_0,\dots,X_k)\mathbf{1}_{\{T>k-1\}} \\
&\overset{(2)}{=} \tilde{\varphi}\left(X_{0\wedge T},\dots,X_{k\wedge T}\right)\mathbf{1}_{\{T>k-1\}} \\
&= \tilde{\varphi}\left(X_{0\wedge T},\dots,X_{k\wedge T}\right)\left(1-\mathbf{1}_{\{T\leq k-1\}}\right) \\
&\overset{\text{Ind Hyp.}}{=} \tilde{\varphi}\left(X_{0\wedge T},\dots,X_{k\wedge T}\right)\varphi\left(X_{0\wedge T},\dots,X_{(k-1)\wedge T}\right) \\
&\in \sigma\left(X_{0\wedge T},\dots,X_{k\wedge T}\right).
\end{aligned}$$

In (1) we have used the fact that T is a $\left(\mathcal{F}^0_n\right)_{n=0}^{\infty}$-stopping time, and so $\{T=k\} \in \mathcal{F}^0_k$. The equality of (2) follows from the fact that $i = i \wedge T$ for $i \in \{0,\dots,k\}$ holds on the domain of $\mathbf{1}_{\{T>k-1\}}$. □

Proposition 5.2.16. *Let $(X_n)_{n=0}^{\infty}$ be a stochastic process and $T<\infty$ be a $\left(\mathcal{F}^0_n\right)_{n=0}^{\infty}$-stopping time. Then*

$$\mathcal{F}_T = \sigma(X_{T\wedge n} : n \in \mathbb{N}).$$

Proof. As $X_{T\wedge n}$ is $\mathcal{F}_T$-measurable for any $n \in \mathbb{N}$ by Lemma 5.2.14, it is clear that $\sigma\left(X_{T\wedge n} : n \in \mathbb{N}\right) \subset \mathcal{F}_T$. On the other hand, let $A \in \mathcal{F}_T$. Then for any $n \in \mathbb{N}$ we know that $A \cap \{T=n\} \in \mathcal{F}_n$ which implies that

$$\mathbf{1}_{A\cap\{T=n\}} = \varphi(X_0,\dots,X_n)$$

for some $\varphi \in \mathcal{B}_b\left(\mathcal{X}^{n+1}\right)$. Note that

$$\varphi(X_0,\dots,X_n) = \varphi(X_0,\dots,X_n)\mathbf{1}_{\{T=n\}} = \varphi\left(X_{0\wedge T},\dots,X_{n\wedge T}\right)\mathbf{1}_{\{T=n\}}.$$

Due to Lemma 5.2.15 we know that $\mathbf{1}_{\{T=n\}} \in \sigma\left(X_{0\wedge T},\dots,X_{n\wedge T}\right)$, therefore, $\mathbf{1}_{A\cap\{T=n\}}$ is measurable with respect to $\sigma\left(X_{T\wedge n} : n \in \mathbb{N}\right)$. Which implies that $\mathcal{F}_T \subset \sigma\left(X_{T\wedge n} : n \in \mathbb{N}\right)$ which completes the proof. □

5.3 The Strong Markov Property

Recall, that we can interpret a stochastic process $X = (X_n)_{n=0}^{\infty}$ as a random element of the canonical probability space

$$\left(\mathcal{X}^{\mathbb{N}}, \mathcal{B}\left(\mathbb{X}^{\mathbb{N}}\right), \mathrm{Law}(X)\right).$$

Furthermore, for each $j \in \mathbb{N}$ we have a measurable map $\theta_j : \mathcal{X}^{\mathbb{N}} \to \mathcal{X}^{\mathbb{N}}$ given by

$$(a_0,\dots,a_j,a_{j+1},\dots) \mapsto (a_j,a_{j+1},\dots).$$

Suppose want a function $F \in \mathcal{B}_b\left(\mathcal{X}^{\mathbb{N}}\right)$ that only depends of times $n \geq j$. Then we can formulate it as a function $\Phi_F \in \mathcal{B}_b\left(\mathcal{X}^{\mathbb{N}}\right)$ where

$$F(\cdot) = \Phi_F(\theta_j \cdot).$$

Definition 5.3.1. *A process $X = (X_n)_{n=0}^{\infty}$ has the strong Markov property if for every finite stopping time T and every bounded measurable function $\Phi \in \mathcal{B}_b\left(\mathcal{X}^{\mathbb{N}}\right)$ we have*

$$\mathbb{E}\left(\Phi(\theta_T X)|\mathcal{F}_T\right) = \mathbb{E}(\Phi(\theta_T X)|X_T).$$

Remark 5.3.2. *In the setting of Definition 5.3.1, we can define another process $Y = (Y_n)_{n=0}^{\infty}$ by $Y_n = X_{T+n}$, that is, $Y = \theta_T X$. Consequently, we can define the filtration $(\mathcal{G}_n)_{n=0}^{\infty}$ where $\mathcal{G}_n = \mathcal{F}_{T+n}$.*

Exercise 5.3.3. *Show that Y is a Markov process on $(\mathcal{G}_n)_{n=0}^{\infty}$.*

Lemma 5.3.4. *Let X be a time-homogeneous Markov process with transition probability P. Then, for any finite stopping time T, fixed $n \in \mathbb{N}$, and $A \in \mathcal{B}(\mathcal{X})$ we have*

$$\mathbb{P}(X_{T+n} \in A | \mathcal{F}_T) = P^n(X_T, A).$$

Proof. It suffices to show the equivalent statement that for all $f \in \mathcal{B}_b(\mathcal{X})$ we have

$$\mathbb{E}(f(X_{T+n})|\mathcal{F}_T) = \int_{\mathcal{X}} f(y) P^n(X_T, \mathrm{d}y).$$

In one direction we set $f = \mathbf{1}_A$ and in the other we use an approximation argument to shows the equivalence between the statements. Note that $\int_{\mathcal{X}} f(y) P^n(X_T, \mathrm{d}y)$ is $\mathcal{F}_T$-measurable. Moreover, for $B \in \mathcal{F}_T$ we can write

$$B = \left(\bigcup_{m=0}^{\infty} B \cap \{T = m\} \right) \cup C$$

where $\mathbb{P}(C) = 0$. Let $B_m = B \cap \{T = m\}$, then

$$\begin{aligned}
\int_B f(X_{T+n}) \,\mathrm{d}\mathbb{P} &= \sum_{m=0}^{\infty} \int_{B_m} f(X_{T+n}) \,\mathrm{d}\mathbb{P} \\
&= \sum_{m=0}^{\infty} \int_{B_m} f(X_{m+n}) \,\mathrm{d}\mathbb{P} \\
&= \sum_{m=0}^{\infty} \mathbb{E}\left(\mathbf{1}_{B_m} f(X_{m+n})\right) \\
&= \sum_{m=0}^{\infty} \mathbb{E}\left(\mathbb{E}(\mathbf{1}_{B_m} f(X_{m+n})|\mathcal{F}_m)\right) \\
&= \sum_{m=0}^{\infty} \mathbb{E}(\mathbf{1}_{B_m} \mathbb{E}(f(X_{m+n})|\mathcal{F}_m)) \\
&= \sum_{m=0}^{\infty} \mathbb{E}\left(\mathbf{1}_{B_m} \int_{\mathcal{X}} f(y) P^n(X_m, \mathrm{d}y)\right) \\
&= \mathbb{E}\left(\left(\int_{\mathcal{X}} f(y) P^n(X_T, \mathrm{d}y) \sum_{m=0}^{\infty} \mathbf{1}_{B_m} \right) \right. \\
&= \int_{\mathcal{X}} f(y) P^n(X_T, \mathrm{d}y) \mathbb{E}(\mathbf{1}_B - \mathbf{1}_C) \\
&= \int_B \int_{\mathcal{X}} f(y) P^n(X_T, \mathrm{d}y) \mathrm{d}\mathbb{P}.
\end{aligned}$$

Therefore, $\int_{\mathcal{X}} f(y) P^n(X_T, \mathrm{d}y)$ satisfies both conditions to be the conditional expectation of $f(X_{T+n})$ with respect to $\mathcal{F}_T$ which completes the proof. ☐

Theorem 5.3.5. *Let $X = (X_n)_{n=0}^{\infty}$ be a time-homogeneous Markov process with transition probability P. If T is a finite stopping time, then the process $(\theta_T X_n)_{n \in \mathbb{N}}$ is also a time-homogeneous Markov process with transition probability P. In particular, for all $\Phi \in \mathcal{B}_b\left(\mathcal{X}^{\mathbb{N}}\right)$ we have*

$$\mathbb{E}\left(\Phi(\theta_T X)|\mathcal{F}_T\right) = \mathbb{E}(\Phi(\theta_T X)|X_T). \tag{5.3.1}$$

Moreover, for any $n > 0$ and $A \in \mathcal{B}(\mathcal{X})$ we have

$$\mathbb{P}(X_{n+T} \in A|\mathcal{F}_T) = P^n(X_T, A) \tag{5.3.2}$$

almost surely. It follows that $X = (X_n)_{n=0}^{\infty}$ has the strong Markov property.

Proof. For any $k \in \mathbb{N}$ and $A \in \mathcal{B}(\mathcal{X})$ one can consider the stopping time $\tilde{T} = T + k$ and apply Lemma 5.3.4 with $n = 1$ to deduce that

$$\mathbb{P}\left(X_{\tilde{T}+1} \in A|\mathcal{F}_{\tilde{T}}\right) = P\left(X_{\tilde{T}}, A\right)$$

which is exactly

$$\mathbb{P}\left((\theta_T X)_{k+1} \in A|\mathcal{G}_k\right) = P\left((\theta_T X)_k, A\right).$$

This saying that $(\theta_T X)_n$ is a time homogeneous Markov process with transition probability P. Moreover, Lemma 5.3.4 shows that

$$\mathbb{P}(X_{n+T} \in A|\mathcal{F}_T) = P^n(X_T, A)$$

holds almost surely for any $n > 0$ and $A \in \mathcal{B}(\mathcal{X})$. To show the equation (5.3.1) it suffices to show that the equation holds for all functions of the form $\Phi(a) = \prod_{i=0}^k f_i(a_i)$, where $a = (a_1, a_2, \dots)$, $f_i \in \mathcal{B}_b(\mathcal{X})$ and $k \in \mathbb{N}$. The reason why this is sufficient is because these functions approximate all functions in $\mathcal{B}_b(\mathcal{X})$. Let $\Phi(\theta_T X) = \prod_{i=0}^k f_i(X_{T+i})$. We proceed by induction on k. For $k = 0$ the result holds clearly as $f_0(X_T)$ is measurable with respect to both $\mathcal{F}_T$ and X_T. Suppose the result holds true for all $n \leq k-1$. Then

$$\begin{aligned}
\mathbb{E}\left(\prod_{i=0}^k f_i(X_{T+i})\middle|\mathcal{F}_T\right) &= \mathbb{E}\left(\mathbb{E}\left(\prod_{i=0}^k f_i(X_{T+i})\middle|\mathcal{F}_{T+k-1}\right)\middle|\mathcal{F}_T\right) \\
&= \mathbb{E}\left(\prod_{i=0}^{k-1} f_i(X_{T+i})\mathbb{E}\left(f_k(X_{T+k})|\mathcal{F}_{T+k-1}\right)\middle|\mathcal{F}_T\right) \\
&= \mathbb{E}\left(\prod_{i=0}^{k-1} f_i(X_{T+i})\int_{\mathcal{X}} f_k(y_k)P(X_{T+k-1}, \mathrm{d}y_k)\middle|\mathcal{F}_T\right).
\end{aligned}$$

Let

$$\tilde{f}_{k-1}(X_{T+k-1}) = f_{k-1}(X_{T+k-1})\int_{\mathcal{X}} f_k(y_k)P(X_{T+k-1}, \mathrm{d}y_k).$$

Then $\tilde{f}_{k-1} \in \mathcal{B}_b(\mathcal{X})$ and so we can apply our induction hypothesis to deduce that

$$\mathbb{E}\left(\tilde{f}_{k-1}(X_{T+k-1})\prod_{i=0}^{k-2} f_i(X_{T+i})\middle|\mathcal{F}_T\right) = \mathbb{E}\left(\tilde{f}_{k-1}(X_{T+k-1})\prod_{i=0}^{k-2} f_i(X_{T+i})\middle|X_T\right).$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left(\prod_{i=0}^k f_i(X_{T+i})\middle|\mathcal{F}_T\right) &= \mathbb{E}\left(\tilde{f}_{k-1}(X_{T+k-1})\prod_{i=0}^{k-2} f_i(X_{T+i})\middle|X_T\right) \\
&= \mathbb{E}\left(\prod_{i=0}^{k-1} f_i(X_{T+i})\int_{\mathcal{X}} f_k(y_k)P(X_{T+k-1}, \mathrm{d}y_k)\middle|X_T\right) \\
&= \mathbb{E}\left(\prod_{i=0}^{k-1} f_i(X_{T+i})\mathbb{E}\left(f_k(X_{T+k})|\mathcal{F}_{T+k-1}\right)\middle|X_T\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(\prod_{i=0}^k f_i(X_{T+i})\middle|\mathcal{F}_{T+k-1}\right)\middle|X_T\right) \\
&= \mathbb{E}\left(\prod_{i=0}^k f_i(X_{T+i})\middle|X_T\right).
\end{aligned}$$

$\square$

Remark 5.3.6.

- *There are continuous time Markov processes which are not strong Markov processes.*
- *The time-homogeneous condition for the discrete-time case is not necessary. We only state it here to simplify the proofs.*

If we do not have T finite, then we can condition on the event $\{T < \infty\}$.

Theorem 5.3.7. *Let $X = (X_n)_{n=0}^{\infty}$ be a time-homogeneous Markov process with transition probability P. Then for all $\Phi \in \mathcal{B}_b\left(\mathcal{X}^{\mathbb{N}}\right)$ we have*

$$\mathbb{E}\left(\Phi(\theta_T X)\mathbf{1}_{\{T<\infty\}}|\mathcal{F}_T\right) = \mathbb{E}_{X_T}\left(\Phi(X)\right)\mathbf{1}_{\{T<\infty\}}.$$

Proof. As before it suffices to show that the equation holds for all functions of the form $\Phi(a) = \prod_{i=0}^k f_i(a_i)$, where $a = (a_1, a_2, \dots)$, $f_i \in \mathcal{B}_b(\mathcal{X})$ and $k \in \mathbb{N}$. We consider a single fixed coordinate. Let $f \in \mathcal{B}_b(\mathcal{X})$ and $B \in \mathcal{F}_T$, then

$$\begin{aligned}
\int_{B\cap\{T<\infty\}} f(X_{T+n})\,\mathrm{d}\mathbb{P} &= \sum_{m=0}^{\infty}\int_{B\cap\{T=m\}} f(X_{T+n})\,\mathrm{d}\mathbb{P}\\
&= \sum_{m=0}^{\infty}\int_{B\cap\{T=m\}} f(X_{m+n})\,\mathrm{d}\mathbb{P}\\
&= \sum_{m=0}^{\infty}\int_{B\cap\{T=m\}} \mathbb{E}(f(X_{m+n})|\mathcal{F}_m)\,\mathrm{d}\mathbb{P}\\
&= \sum_{m=0}^{\infty}\int_{B\cap\{T=m\}} \mathbb{E}(f(X_{m+n})|X_m)\,\mathrm{d}\mathbb{P}\\
&= \int_{B\cap\{T<\infty\}} \mathbb{E}\left(X_{T+n}\mathbf{1}_{\{T<\infty\}}|X_T\right)\,\mathrm{d}\mathbb{P}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left(f\left(X_{T+n}\right)\mathbf{1}_{\{T<\infty\}}|\mathcal{F}_T\right) &= \mathbb{E}\left(f(X_{T+n})\mathbf{1}_{T<\infty}|X_T\right)\\
&= \mathbb{E}_{X_T}\left(f(X_{T+n})\right)\mathbf{1}_{\{T<\infty\}}.
\end{aligned}$$

One can extend this argument to a finite number of coordinates to complete the proof. ☐

Remark 5.3.8. *Recall, that $\mathbb{E}_{X_T}(\cdot)$ denotes the expectation when X_T is used to start the Markov process.*

5.4 Solution to Exercises

Exercise 5.1.2

Solution. We can show that $P^n = P$ through induction. The base case is clear. Therefore, suppose the result holds for $n \le k-1$. Then

$$\begin{aligned}
P^k(x,\{j\}) &= \sum_{i\in\mathcal{X}} P(i,\{j\})P^{k-1}(x,\{i\})\\
&= \sum_{i\in\mathcal{X}} P_{ji}\left(\underbrace{P\times\cdots\times P}_{k-1}\right)_{ix}\\
&= \left(\underbrace{P\times\cdots\times P}_{k}\right)_{jx}.
\end{aligned}$$

Similarly, note that P is a stochastic matrix then

$$\begin{aligned}\sum_{k\in\mathcal{X}}\left(P^2\right)_{ik} &= \sum_{k\in\mathcal{X}}\sum_{j\in\mathcal{X}}P_{ij}P_{jk}\\ &= \sum_{j\in\mathcal{X}}\sum_{k\in\mathcal{X}}P_{ij}P_{jk}\\ &= \sum_{j\in\mathcal{X}}P_{ij}\sum_{k\in\mathcal{X}}P_{jk}\\ &= \sum_{j\in\mathcal{X}}P_{ij}(1)\\ &= 1.\end{aligned}$$

Therefore, P^2 is a stochastic matrix and so through induction one can conclude that P^n is a stochastic matrix. □

Exercise 5.2.2

Solution. This follows directly from observing that

$$\{\tau_A = n\} = \left(\bigcap_{k=0}^{n-1}\{X_k \notin A\}\right) \cap \{X_n \in A\} \in \mathcal{F}_n^0.$$

□

Exercise 5.2.5
Note that for any $n \in \mathbb{N}$ we can write

$$X_n^T(\omega) = X_n(\omega)\mathbf{1}_{\{T(\omega)>n\}} + X_T\mathbf{1}_{\{T(\omega)\leq n\}}$$

where

- $X_n(\omega) \in \mathcal{F}_n$,
- $\mathbf{1}_{\{T(\omega)>n\}} = 1 - \mathbf{1}_{\{T\leq n\}} \in \mathcal{F}_n$ as T is a $(\mathcal{F}_n)_{n=0}^{\infty}$-stopping time,
- $X_{T(\omega)} \in \mathcal{F}_{T(\omega)} \subset \mathcal{F}_n$, and
- $\mathbf{1}_{\{T(\omega)\leq n\}}$ as T is a $(\mathcal{F}_n)_{n=0}^{\infty}$-stopping time.

Therefore, $X_n^T(\omega) \in \mathcal{F}_n$ which implies the stopped process is adapted to $(\mathcal{F}_n)_{n=0}^{\infty}$.
Exercise 5.2.9

Solution. This follows in the same way as previous statements

$$\{T = \infty\} = \bigcap_{n=1}^{\infty}\{T > n\} = \bigcap_{n=1}^{\infty}\{T \leq n\}^c,$$

as each $\{T \leq n\} \in \mathcal{F}_n$ by definition it follows that $\{T = \infty\} \in \mathcal{F}_\infty$. □

Exercise 5.2.12

Solution.

1. Let $A \in \mathcal{F}_S$. Then
$$A \cap \{T \leq n\} = (A \cap \{S \leq n\}) \cap \{T \leq n\}.$$
The bracketed term is in $\mathcal{F}_n$ by assumption, and $\{T \leq n\} \in \mathcal{F}_n$ by definition. Therefore, $A \cap \{T \leq n\} \in \mathcal{F}_n$ which implies that $A \in \mathcal{F}_T$.

2. Note that
$$\{S\mathbf{1}_A + T\mathbf{1}_{A^c} \le n\} = (\{S \le n\} \cap A) \cup (\{T \le n\} \cap A^c).$$
The first bracketed term is in $\mathcal{F}_n$ by construction of $\mathcal{F}_S$. By part 1. we know that $A \in \mathcal{F}_T$, therefore, by a similar argument we know that the second bracketed term is in $\mathcal{F}_n$. Therefore,
$$\{S\mathbf{1}_A + T\mathbf{1}_{A^c} \le n\} \in \mathcal{F}_n$$
which implies that $S\mathbf{1}_A + T\mathbf{1}_{A^c}$ is a $(\mathcal{F}_n)$-stopping time.

3. For $A \in \mathcal{F}_S$ it follows that $A \cap \{S \le n\} \in \mathcal{F}_n$. Recall, that $S \wedge T$ is itself a $(\mathcal{F}_n)$-stopping time, so that $\{S \wedge T \le n\} \in \mathcal{F}_n$. Moreover, if $\{S \le T\} \cap \{S \wedge T \le n\}$ it must be the case that $S \le n$. Therefore,
$$(A \cap \{S \le T\}) \cap \{S \wedge T \le n\} = A \cap \{S \le n\} \cap \{S \wedge T \le n\} \in \mathcal{F}_n.$$

4. Step 1: Show that $\mathcal{F}_{S\wedge T} = \mathcal{F}_S \cap \mathcal{F}_T$.
If $A \in \mathcal{F}_S \cap \mathcal{F}_T$, then
$$A \cap \{S \wedge T \le n\} = (A \cap \{S \le n\}) \cup (A \cap \{T \le n\}) \in \mathcal{F}_n.$$
Which implies that, $A \in \mathcal{F}_{S\wedge T}$. Suppose instead that $A \in \mathcal{F}_{S\wedge T}$. Then
$$\begin{aligned} A \cap \{S = n\} &= A \cap \{S = n\} \cap (\{S \le T\} \cup \{S > T\}) \\ &= A \cap \{S \wedge T = n\} \cup A \cap \{S \wedge T \le n-1\} \in \mathcal{F}_n. \end{aligned}$$
Therefore, $\mathcal{F}_S \cap \mathcal{F}_T = \mathcal{F}_{S\wedge T}$.
Step 2: Show that $\mathbb{E}\left(\mathbb{E}\left(X|\mathcal{F}_S\right)|\mathcal{F}_T\right) = \mathbb{E}\left(X|\mathcal{F}_{S\wedge T}\right)$.
Consider $\mathbb{E}\left(X|\mathcal{F}_{S\wedge T}\right)$. It is $\mathcal{F}_{S\wedge T}$-measurable and thus $\mathcal{F}_T$ measurable by Step 1. Let $G \in \mathcal{F}_T$. Then $G \cap \{T \le S\} \in \mathcal{F}_{S\wedge T} \subset \mathcal{F}_T$, again using Step 1. Hence, using the definition of conditional expectations we deduce that
$$\begin{aligned} \int_G \mathbf{1}_{\{T\le S\}} \mathbb{E}\left(X|\mathcal{F}_{S\wedge T}\right) \mathrm{d}\mathbb{P} &= \int_{G\cap\{T\le S\}} \mathbb{E}\left(X|\mathcal{F}_{S\wedge T}\right) \mathrm{d}\mathbb{P} \\ &= \int_{G\cap\{T\le S\}} X \,\mathrm{d}\mathbb{P} \\ &= \int_G \mathbf{1}_{\{T\le S\}} X \,\mathrm{d}\mathbb{P} \end{aligned}$$
and similarly
$$\int_G \mathbf{1}_{\{T\le S\}} \mathbb{E}(X|\mathcal{F}_S)\,\mathrm{d}\mathbb{P} = \int_G \mathbf{1}_{\{T\le S\}}.$$
Therefore,
$$\int_G \mathbf{1}_{\{T\le S\}} \mathbb{E}(X|\mathcal{F}_{S\wedge T})\,\mathrm{d}\mathbb{P} = \int_G \mathbf{1}_{\{T\le S\}} \mathbb{E}(X|\mathcal{F}_S)\,\mathrm{d}\mathbb{P}. \tag{5.4.1}$$
As $G \cap \{S < T\} \in \mathcal{F}_S \cap \mathcal{F}_T$, by the same reasoning we deduce that
$$\int_G \mathbf{1}_{\{T<S\}} \mathbb{E}(X|\mathcal{F}_{S\wedge T})\,d\mathbb{P} = \int_G \mathbf{1}_{\{S<T\}} \mathbb{E}(X|\mathcal{F}_S)\,\mathrm{d}\mathbb{P}. \tag{5.4.2}$$
Adding (5.4.1) and (5.4.2) together we deduce that
$$\int_G \mathbb{E}(X|\mathcal{F}_{S\wedge T})\,\mathrm{d}\mathbb{P} = \int_G \mathbb{E}(X|\mathcal{F}_S)\,\mathrm{d}\mathbb{P}$$
for all $G \in \mathcal{F}_T$ which implies that $\mathbb{E}\left(\mathbb{E}\left(X|\mathcal{F}_S\right)|\mathcal{F}_T\right) = \mathbb{E}\left(X|\mathcal{F}_{S\wedge T}\right)$.
Step 3: Show that $\mathbb{E}\left(\mathbb{E}\left(X|\mathcal{F}_T\right)|\mathcal{F}_S\right) = \mathbb{E}\left(X|\mathcal{F}_{S\wedge T}\right)$.
This follows by the similar arguments made in Step 2.

$\square$

Exercise 5.3.3

Solution. For every $f \in \mathcal{B}_b(\mathcal{X})$ and $n, m \in \mathbb{N}$ it follows that

$$\begin{aligned}\mathbb{E}(f(Y_{n+m})|\mathcal{G}_m) &= \mathbb{E}\left(f\left(X_{n+m+T}\right)|\mathcal{F}_{T+m}\right)\\ &\overset{(1)}{=} \mathbb{E}\left(f\left(X_{n+m+T}\right)|X_{T+m}\right)\\ &= \mathbb{E}(f(Y_{n+m})|Y_m),\end{aligned}$$

where (1) is an application of the strong Markov property of X. □

6 Discrete State Space Markov Processes

We will identify the setting of discrete state spaces by saying that $\mathcal{X}$ is countable. The notions developed here, do not necessarily have a direct analogue for continuous state space Markov processes. We will continue to consider discrete-time Markov processes.

6.1 Markov Chains as Graphs

Given a stochastic matrix P on $\mathcal{X}$, that is $P = (P_{ij} : i, j \in \mathcal{X})$, we can build an oriented graph on $\mathcal{X}$ where an edge is drawn from i to j if and only if $P_{ij} > 0$. As $P_{ij} = \mathbb{P}(X_{n+1} = j|X_n = i)$, we can think of this graph where the nodes are states and the edges represent paths between the states that are admissible in the process defined by P.

Example 6.1.1. *The stochastic matrix*

$$P = \frac{1}{10}\begin{pmatrix} 0 & 5 & 5 & 0 \\ 3 & 7 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 2 & 8 & 0 & 0 \end{pmatrix}$$

has the corresponding oriented graph depicted in Figure 1. We use this graph to help compute probabilities such as $\mathbb{P}(X_2 = 2|X_0 = 1)$. We see that the only paths which contribute to this probability are $1 \to 3 \to 2$ and $1 \to 2 \to 2$. Therefore,

$$\mathbb{P}(X_2 = 2|X_0 = 1) = \frac{1}{2}(1) + \frac{1}{2}\left(\frac{7}{10}\right).$$

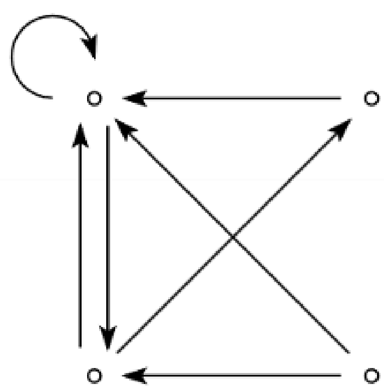


Figure 1: The oriented graph of Example 6.1.1

Definition 6.1.2. *Let $\mathcal{X}$ be countable and P be a stochastic matrix on $\mathcal{X}$.*

1. *We say j is accessible from i if $P^n_{ij} > 0$ for some $n \in \mathbb{N}$. We denote this $i \to j$.*
2. *States i and j are said to communicate if $i \to j$ and $j \to i$. We denote this $i \leftrightarrow j$.*
3. *Given a state i, we let $[i] = \{j \in \mathcal{X} : i \sim j\}$ denote the communication class of i.*
4. *P is said to be irreducible if $[i] = \mathcal{X}$ for some $i \in \mathcal{X}$. Otherwise, P is called reducible.*

Exercise 6.1.3. *Show that $\leftrightarrow$ is an equivalence class on $\mathcal{X}$.*

Example 6.1.4. *Consider Example 6.1.1. The set $\{1, 2, 3\}$ is a communicating class, and $\{4\}$ is another. Therefore, P is reducible.*

When $\mathcal{X}$ is infinite, only finite paths in the incidence graph guarantee accessibility, infinite paths may not.

Lemma 6.1.5. *Prove that $i \to j$, then for any $i' \in [i]$ and $j' \in [j]$ then $i' \to j'$.*

Proof. By assumption there exists n_1, n_2 and n_3 such that

- $(P^{n_1})_{i'i} > 0$,
- $(P^{n_2})_{ij} > 0$, and
- $(P^{n_3})_{jj'} > 0$.

Therefore,

$$\left(P^{n_1+n_2+n_3}\right)_{i'j'} \overset{(1)}{\geq} (P^{n_1})_{i'i} (P^{n_2})_{ij} (P^{n_3})_{jj'} > 0$$

where (1) follows from the fact that P^{n_1}, P^{n_2} and P^{n_3} are stochastic matrices and so have non-negative entries. Hence, $i' \to j'$. ☐

Figure 2: A graphical representation for the proof of Lemma 6.1.5.

Exercise 6.1.6. *Show that the relation $[i] \leq [j]$ if and only if $j \to i$ is well-defined, and defines a partial order. That is, $\leq$ is reflexive, transitive, and anti-symmetric.*

Definition 6.1.7. *An equivalence class $[i]$ is said to be minimal, or closed, if there is no j such that $[j] \leq [i]$ and $[j] \neq [i]$.*

Example 6.1.8. *Consider a stochastic matrix identified by the graph in Figure 3.*

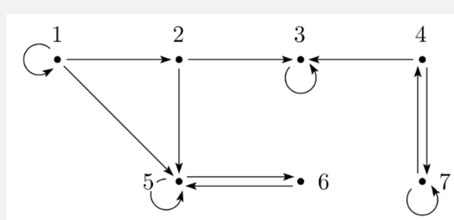


Figure 3: The oriented graph of the stochastic matrix referred to in Example 6.1.8.

In this case, the communication classes are the following.

- $[1] = \{1\}$.
- $[2] = \{2\}$.
- $[3] = \{3\}$.
- $[4] = \{4, 7\}$.
- $[5] = \{5, 6\}$.

Note that we have the relations

- $[5] \leq [2] \leq [1]$,
- $[3] \leq [4]$, *and*
- $[3] \leq [2] \leq [1]$.

However, note that we cannot build relations between $[4]$ *and* $[2]$. *Therefore,* $\leq$ *cannot be considered as a total order.*

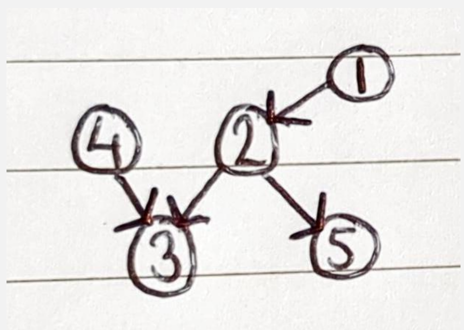


Figure 4: One can think of the equivalence classes as macro states, for which the process can only transition to a state that is less than or equal to it.

6.2 Recurrence and Transience

Definition 6.2.1. *Given a state* $i \in \mathcal{X}$, *let*

$$T_i = \inf\{n \geq 1 : X_n = i\}.$$

If $X_0 = i$ *then* T_i *is called the first return time to state* i.

For a state $i \in \mathcal{X}$ we adopt the notation

- $\mathbb{E}_i(\cdot) = \mathbb{E}(\cdot | X_0 = i)$, and
- $\mathbb{P}_i(\cdot) = \mathbb{P}(\cdot | X_0 = i)$.

Example 6.2.2. *Suppose* $\mathcal{X} = \{1, 2\}$ *and* $P_{12} = \alpha \in [0, 1]$ *and* $P_{21} = \beta \in [0, 1]$. *Then the stochastic matrix* P *is given by*

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

Furthermore,

$$\mathbb{P}_1(T_1 = n) = \begin{cases} 1-\alpha & n = 1 \\ \alpha(1-\beta)^{n-2}\beta & n \geq 2. \end{cases}$$

Hence,

$$\begin{aligned}\mathbb{P}_1(T_1 < \infty) &= \sum_{n=1}^{\infty} \mathbb{P}(T_1 = n) \\ &= (1-\alpha) + \alpha\beta \sum_{n=0}^{\infty} (1-\beta)^n \\ &= \begin{cases} 1 - \alpha + \alpha\beta \frac{1}{1-\beta} & \beta \neq 0 \\ 1 - \alpha & \beta = 0. \end{cases}\end{aligned}$$

Definition 6.2.3. *A state i is said to be recurrent if $\mathbb{P}_i(T_i < \infty) = 1$. If a state i is not recurrent, then it is said to be transient.*

Definition 6.2.4. *A Markov chain, that is a stochastic matrix P, on $\mathcal{X}$ is said to be recurrent if every state $i \in \mathcal{X}$ is recurrent and it is said to be transient if every state is transient.*

Remark 6.2.5. *We will see that if $i \in \mathcal{X}$ is recurrent, then every $i' \in [i]$ is also recurrent. This means recurrence and transience are properties of communication classes.*

Lemma 6.2.6. *For states $i, j \in \mathcal{X}$ we have that $i \to j$ if and only if $\mathbb{P}_i(T_j < \infty) > 0$. Moreover,*

$$\mathbb{P}_i(T_j < \infty) \leq \sum_{n=1}^{\infty} P_{ij}^n.$$

Proof. Observe that if $\mathbb{P}_i(T_j = n) > 0$ then

$$P_{ij}^n \overset{(\star)}{\geq} \mathbb{P}_i(T_j = n) > 0.$$

Where $(\star)$ is justified by the fact that P_{ij}^n is the probability of any path of length n between i and j, with $\mathbb{P}_i(T_j = n)$ only considering the subset of such paths that do not previously encounter j. Moreover, if $P_{ij}^n = 0$ then no path of length n exists from i to j with a non-zero probability and so $\mathbb{P}_i(T_j = n) = 0$. Hence, $\mathbb{P}_i(T_j = n) > 0$ if and only if $P_{ij}^n > 0$. Next, observe that

$$\mathbb{P}_i(T_j < \infty) = \mathbb{P}_i\left(\bigcup_{n=1}^{\infty} \{T_j = n\}\right). \tag{6.2.1}$$

So that $\mathbb{P}_i(T_j > \infty) > 0$ if and only if there exists an $n \geq 1$ such that $\mathbb{P}_i(\{T_j = n\}) > 0$ which happens if and only if $P_{ij}^n > 0$, and so by definition we have that $i \to j$. Applying a union bound to equation (6.2.1) we conclude that

$$\mathbb{P}_i(T_j < \infty) \leq \sum_{n=1}^{\infty} \mathbb{P}_i(T_j = n) \leq \sum_{n=1}^{\infty} P_{ij}^n.$$

□

Lemma 6.2.7. *Let $j \in \mathcal{X}$ be recurrent. For $i \in \mathcal{X}$ if $\mathbb{P}_j(T_i < \infty) > 0$ then $\mathbb{P}_i(T_j < \infty) = 1$.*

Proof. Assume that $\mathbb{P}_i(T_j < \infty) > 0$. This means that there is a set of infinite paths with non-zero probability that start at i and do not reach j. The condition that $\mathbb{P}_j(T_i < \infty) > 0$ says that i is accessible from j. In particular, $m := \min\left\{n : P_{ji}^n > 0\right\}$ is finite as the set is not empty. Note by construction that paths of length m

from j to i, reach i from j without returning to j beforehand. Otherwise, the Markov assumptions would imply that there exists a $k < m$ such that $\left(P^k\right)_{ij} > 0$. With this, it follows that

$$\begin{aligned}
\mathbb{P}_j(T_j = \infty) &\geq \mathbb{P}_j(T_j = \infty, T_i = m) \\
&= \mathbb{P}\left(\bigcap_{k=1}^{\infty}\{X_k \neq j\}, T_i = m \middle| X_0 = j\right) \\
&= \mathbb{P}\left(\bigcap_{k=1}^{\infty}\{X_{m+k} \neq j\}, T_i = m \middle| X_0 = j\right) \\
&= \mathbb{P}\left(\bigcap_{k=1}^{\infty}\{X_{m+k} \neq j\} \middle| T_i = m, X_0 = j\right)\mathbb{P}(T_i = m | X_0 = j) \\
&= \mathbb{P}\left(\bigcap_{k=1}^{\infty}\{X_k \neq j\} \middle| X_0 = i\right)\mathbb{P}_j(T_i = m) \\
&= \mathbb{P}_i(T_j = \infty)\mathbb{P}_j(T_i = m) \\
&> 0.
\end{aligned}$$

Which contradicts j being a recurrent state. $\square$

Exercise 6.2.8. *Show in the context of Lemma 6.2.7 that for any μ supported on $[j]$ it follows that $\mathbb{P}_\mu(T_j < \infty) = 1$.*

Definition 6.2.9. *We define passage times inductively from hitting times.*

- $T_j^0 = 0$.
- $T_j^1 = T_j$.
- $T_j^{n+1} = \inf\left\{k > T_j^n : X_k = j\right\}$ *for* $n \geq 1$.

Exercise 6.2.10. *Show that T_j^n is a $(\mathcal{F}_k)_{k=0}^{\infty}$-stopping time.*

Lemma 6.2.11. *Let X have an initial distribution μ, and suppose $\mathbb{P}_\mu(T_j < \infty) = 1$ for a recurrent state $j \in \mathcal{X}$. Then the random variables $\left\{T_j^n - T_j^{n-1}\right\}_{n=1}^{\infty}$ are independent with*

$$\mathbb{P}\left(T_j^k - T_j^{k-1} = m\right) = \mathbb{P}_j(T_j = m)$$

for any $m, k \in \mathbb{N}$.

Proof. By the strong Markov property for $k \geq 0$ we have that

$$\begin{aligned}
\mathbb{P}\left(T_j^{k+1} - T_j^k = m \middle| \mathcal{F}_{T_j^k}\right)(\omega) &= \mathbb{P}_{T_j^k(\omega)}(T_j = m) \\
&= \mathbb{P}_j(T_j = m).
\end{aligned}$$

Taking the expectation we conclude that

$$\mathbb{P}_j(T_j = m) = \mathbb{P}\left(T_j^{k+1} - T_j^k = m\right)$$

by the tower property and the fact that the left-hand side is a constant. Next observe that $\mathbb{P}\left(T_j^{k+1} - T_j^k = m \middle| \mathcal{F}_{T_j^k}\right)$ is constant in ω. That is, for any $A \in \mathcal{F}_{T_j^k}$, we have that

$$\begin{aligned}
\mathbb{E}\left(\mathbf{1}_A \mathbf{1}_{\{T_j^{k+1} - T_j^k = m\}}\right) &= \mathbb{E}\left(\mathbb{E}\left(\mathbf{1}_A \mathbf{1}_{\{T_j^{k+1} - T_j^k = m\}} \middle| \mathcal{F}_{T_j^k}\right)\right) \\
&= \mathbb{E}\left(\mathbf{1}_A \mathbb{E}\left(\mathbf{1}_{\{T_j^{k+1} - T_j^k = m\}} \middle| \mathcal{F}_{T_j^k}\right)\right) \\
&= \mathbb{P}(A)\mathbb{P}_j(T_j = m).
\end{aligned}$$

This shows the required independence. □

Remark 6.2.12. *The assumption on μ in Lemma 6.2.11 can be removed by conditioning. One can show that*

$$\mathbb{P}\left(T_j^k - T_j^{k-1} = m, T_j^k < \infty \middle| \mathcal{F}_{T_j^k}\right) = \mathbb{P}_j(T_j = m)\mathbf{1}_{T_j^k < \infty}$$

and an analogous statement of independence.

Note the following,

- $T_j^n = \sum_{k=1}^n T_j^k - T_j^{k-1}$, and
- $\{T_j^n < \infty\} = \bigcap_{k=1}^n \{T_j^k - T_j^{k-1} < \infty\}$.

Lemma 6.2.13. *For any $i, j \in \mathcal{X}$ and $k \in \mathbb{N}$ it follows that*

$$\mathbb{P}_i\left(T_j^{k+1} < \infty\right) = \mathbb{P}_i(T_j < \infty)\mathbb{P}_j\left(T_j^k < \infty\right).$$

Consequently,

$$\mathbb{P}_j\left(T_j^{k+1} < \infty\right) = \mathbb{P}_j\left(T_j < \infty\right)^{k+1}.$$

Proof. Let $\Phi \in \mathcal{B}_b\left(\mathcal{X}^{\mathbb{N}}\right)$ be given by $\Phi(X) = \mathbf{1}_A(X)$, where

$$A := \{X \in \mathcal{X}^n : |\{n \geq 1 : X_n = j\}| \geq k\}.$$

Note that

$$\mathbf{1}_{\{T_j^{k+1} < \infty\}}(\omega) = \Phi\left(\theta_{T_j(\omega)} X(\omega)\right) \mathbf{1}_{\{T_j < \infty\}}(\omega).$$

Taking the conditional expectation of both sides with respect to $\mathcal{F}_{T_j}$ we deduce using the strong Markov property that

$$\begin{aligned}
\mathbb{E}\left(\mathbf{1}_{\{T_j^{k+1} < \infty\}} \middle| \mathcal{F}_{T_j}\right) &= \mathbf{1}_{\{T_j < \infty\}} \mathbb{E}_{X_{T_j}}(\Phi(X)) \\
&= \mathbf{1}_{\{T_j < \infty\}} \mathbb{P}_j\left(T_j^k < \infty\right).
\end{aligned}$$

Taking $\mathbb{E}_i(\cdot)$ completes the proof. □

6.3 Recurrence Conditions

Definition 6.3.1. *The occupation time of a state $j \in \mathcal{X}$ is the random variable*

$$\eta_j = \sum_{n=1}^{\infty} \mathbf{1}_{\{X_n = j\}}.$$

Remark 6.3.2. *Occupation times are not stopping times.*

Theorem 6.3.3. *A state $j \in \mathcal{X}$ is transient if and only if $\sum_{n=1}^{\infty} P_{jj}^n < \infty$. Equivalently, $j \in \mathcal{X}$ is recurrent if and only if $\sum_{n=1}^{\infty} P_{jj}^n = \infty$.*

Proof. Note that $\mathbb{E}_j\left(\mathbf{1}_{\{X_n=j\}}\right) = P_{jj}^n$ so that

$$\begin{aligned}
\sum_{n=1}^{\infty} P_{jj}^n &= \mathbb{E}_j(\eta_j) \\
&= \sum_{n=1}^{\infty} n\mathbb{P}_j(\eta_j = n) \\
&= \sum_{n=1}^{\infty} \mathbb{P}_j(\eta_j \geq n) \\
&= \sum_{n=1}^{\infty} \mathbb{P}_j\left(T_j^n < \infty\right) \\
&= \sum_{n=1}^{\infty} \mathbb{P}_j(T_j < \infty)^n.
\end{aligned}$$

The geometric series is summable if and only if $\mathbb{P}_j(T_j < \infty) < 1$, which is to say that j is transient. $\square$

Corollary 6.3.4. *Suppose $j \in [i]$, then i is recurrent (transient) if and only if j is recurrent (transient).*

Proof. Since $j \to i$ and $i \to j$ there exists m_1 and m_2 such that $P_{ji}^{m_1}, P_{ij}^{m_2} > 0$. Therefore,

$$\begin{aligned}
\sum_{k=m_1+m_2}^{\infty} P_{jj}^k &\geq \sum_{n=1}^{\infty} P_{ji}^{m_1} P_{ii}^n P_{ij}^{m_2} \\
&= P_{ji}^{m_1} P_{ij}^{m_2} \sum_{n=1}^{\infty} P_{ii}^n,
\end{aligned}$$

so if i is recurrent then j is recurrent. By symmetry, the same holds for transience. $\square$

Lemma 6.3.5. *Let $k \in \mathcal{X}$, then one of the following holds.*

1. $\sum_{n=1}^{\infty} P_{ij}^n = \infty$*, for every* $i, j \in [k]$.
2. $\sum_{n=1}^{\infty} P_{ij}^n < \infty$ *for every* $i, j \in [k]$.

In particular, if $[k]$ has a finite number of elements and is a minimal class then the first case must hold. Which implies every element of $[k]$ is recurrent.

Proof. Suppose that $\sum_{n=1}^{\infty} P_{ij}^n = \infty$ for $i, j \in [k]$. Then for $i', j' \in [k]$ it follows that there exists m_1 and m_2 such that $P_{i'i}^{m_1} > 0$ and $P_{jj'}^{m_2} > 0$. Therefore,

$$\begin{aligned}
\sum_{n=1}^{\infty} P_{i'j'}^n &\geq \sum_{n=1}^{\infty} P_{ij}^n P_{i'i}^{m_1} P_{jj'}^{m_2} \\
&= P_{i'i}^{m_1} P_{jj'}^{m_2} \sum_{n=1}^{\infty} P_{ij}^n \\
&= \infty.
\end{aligned}$$

Suppose instead that $\sum_{n=1}^{\infty} P_{ij}^n < \infty$ for $i, j \in [k]$. Then for $i', j' \in [k]$ it follows that there exists m_1 and m_2 such that $P_{ii'}^{m_1} > 0$ and $P_{j'j}^{m_2} > 0$. Hence, as before, we deduce that

$$\sum_{n=1}^{\infty} P_{ij}^n \geq P_{ii'}^{m_1} P_{j'j}^{m_2} \sum_{n=1}^{\infty} P_{i'j'}^n,$$

which implies that $\sum_{n=1}^{\infty} P_{i'j'}^n < \infty$ as $P_{ii'}^{m_1} > 0$ and $P_{j'j}^{m_2} > 0$. Consequently, if $[k]$ is a finite minimal class with $i \in [k]$ observe that

$$\begin{aligned}\sum_{j\in[k]} \sum_{n=1}^{\infty} P_{ij}^n &\overset{(1)}{=} \sum_{n=1}^{\infty} \sum_{j\in[k]} P_{ij}^n \\ &\overset{(2)}{=} \sum_{n=1}^{\infty} 1 \\ &= \infty.\end{aligned}$$

Where (1) is just changing the ordering of a finite and infinite sum and (2) follows from the fact that $[k]$ is minimal and so $P_{ij}^n > 0$ if and only if $j \in [k]$. Therefore, we know that for at least one $j \in [k]$ the sum $\sum_{n=1}^{\infty} P_{ij}^n$ is infinite which implies that $\sum_{n=1}^{\infty} P_{ij}^n = \infty$ for all $j \in [k]$. Hence, every element of $[k]$ is recurrent. $\square$

Theorem 6.3.6.

- *A state j is recurrent if and only if $\mathbb{P}_j(X_n = j \text{ i.o.}) = 1$.*
- *A state j is transient if and only if $\mathbb{P}_j(X_n = j \text{ i.o.}) = 0$.*

Proof. Note that $\{X_n = j \text{ i.o.}\} = \{\eta_j = \infty\}$. Moreover,

$$\{\eta_j = \infty\} = \bigcap_{n=1}^{\infty} \{\eta_j \geq n\}.$$

The sets $\{\eta_j \geq n\}$ are decreasing in n and so

$$\begin{aligned}\mathbb{P}(\eta_j = \infty) &= \lim_{n\to\infty} \mathbb{P}(\eta_j \geq n) \\ &= \lim_{n\to\infty} \mathbb{P}_j\left(T_j^n < \infty\right) \\ &\overset{\text{Lem } 6.2.13}{=} \lim_{n\to\infty} \mathbb{P}_j\left(T_j < \infty\right)^n \\ &= \begin{cases} 1 & j \text{ recurrent} \\ 0 & j \text{ transient.}\end{cases}\end{aligned}$$

$\square$

Lemma 6.3.7. *Suppose $\mathcal{X}$ is finite, then a state is recurrent if and only if it is in a minimal class. In particular, there always exists a recurrent state.*

Proof. As $\mathcal{X}$ is finite, the partial order $\leq$ is defined on a finite number of communication classes, and so there must exist a minimal class. In particular, this minimal class must be finite and so using Lemma 6.3.5 the class must be recurrent, and so a recurrent state exists. Conversely, let $i \in \mathcal{X}$ be a recurrent state and suppose for contradiction that $i \in [k]$, where $[k]$ is not a minimal class. Then there exists a $j \in [k']$ such that $i \to j$ but $j \not\to i$. In particular, for $j \in [k']$ there exists an n_j such that $P_{ij}^{n_j} > 0$ and $P_{ji}^n = 0$ for all $n \in \mathbb{N}$. Therefore,

$$\mathbb{P}_i(T_i = \infty) \geq P_{ij}^{n_j} > 0.$$

This implies that $\mathbb{P}_i(T_i < \infty) < 1$ and so the state i cannot be recurrent, which is a contradiction. $\square$

Proposition 6.3.8. *Suppose that i and j are state such that $i \to j$ but $j \not\to i$, then i must be transient. In particular, if $[i]$ is not minimal, then it consists of transient states.*

Proof. Let m be the smallest number such that $P_{ij}^m > 0$. It follows that paths of length m from i to j never returns to i before time m. Moreover, as $j \not\to i$ it follows that such a path never returns to i after time m either. Consequently,

$$\mathbb{P}_i(T_i = \infty) \geq P_{ij}^m > 0.$$

Therefore, i is transient. In particular, if $[i]$ is not minimal then a j such that $i \to j$ but $j \not\to i$ exists by definition of not being minimal. Therefore, $[i]$ must contain transient states. □

Lemma 6.3.9. *For states $i, j \in \mathcal{X}$ we have that*

$$\sum_{n=1}^{\infty} P_{ij}^n = \frac{\mathbb{P}_i(T_j < \infty)}{1 - \mathbb{P}_j(T_j < \infty)}$$

with the understanding that the right-hand side is infinite if $\mathbb{P}_i(T_j < \infty)$ and $\mathbb{P}_j(T_j < \infty = 1)$.

Proof. Recall that for any $j \in \mathcal{X}$ the occupation time is given by $\eta_j = \sum_{n=1}^{\infty} \mathbf{1}_{\{X_n = j\}}$. Hence, we can write

$$\begin{aligned}
\sum_{n=1}^{\infty} P_{ij}^n &= \mathbb{E}_i(\eta_j) \\
&= \sum_{k=1}^{\infty} \mathbb{P}_i(\eta_j \geq k) \\
&= \sum_{k=1}^{\infty} \mathbb{P}_i\left(T_j^k < \infty\right) \\
&\overset{\text{Lem } 6.2.13}{=} \sum_{k=1}^{\infty} \mathbb{P}_i(T_j < \infty)\mathbb{P}_j\left(T_j^{k-1} < \infty\right) \\
&\overset{\text{Lem } 6.2.13}{=} \sum_{k=1}^{\infty} \mathbb{P}_i(T_j < \infty)\mathbb{P}_j\left(T_j < \infty\right)^{k-1} \\
&= \frac{\mathbb{P}_i(T_j < \infty)}{1 - \mathbb{P}_j(T_j < \infty)}.
\end{aligned}$$

□

Theorem 6.3.10. *If a state $j \in \mathcal{X}$ is transient, then for all $i \in \mathcal{X}$ we have that*

$$\sum_{n=1}^{\infty} P_{ij}^n < \infty.$$

Moreover, $\lim_{n\to\infty} P_{ij}^n = 0$.

Proof. If $j \in \mathcal{X}$ is transient then $\mathbb{P}_j(T_j < \infty) < 1$, so by Lemma 6.3.9 we have that

$$\sum_{n=1}^{\infty} P_{ij}^n = \frac{\mathbb{P}_i(T_j < \infty)}{1 - \mathbb{P}_j(T_j < \infty)} < \infty$$

which implies that $\lim_{n\to\infty} P_{ij}^n = 0$ for all $i \in \mathcal{X}$. □

Remark 6.3.11. *Intuitively, Theorem 6.3.10 says that transient states are difficult to reach.*

Theorem 6.3.12. *If P has an invariant probability measure π, then for any transient state $j \in \mathcal{X}$ we must have $\pi(j) = 0$.*

Proof. Without loss of generality, we can assume that $\mathcal{X} = \mathbb{N}$. Suppose for contradiction that a transient j is such that $\pi(j) > 0$. Then as $\sum_{k\in\mathbb{N}} \pi(k) = 1 < \infty$, there exists an $N \in \mathbb{N}$ such that

$$\sum_{k=N+1}^{\infty} \pi(k) < \frac{\pi(j)}{2}.$$

Moreover, we can an $n \in \mathbb{N}$ such that for all $0 \le k \le N$ we have $P^n_{kj} < \frac{\pi(j)}{2}$. As π is invariant we can write

$$\pi(j) = \sum_{k=0}^{\infty} \pi(k) P^n_{kj}.$$

Therefore,

$$\begin{aligned}
\pi(j) &= \sum_{k=0}^{N} \pi(k) P^n_{kj} + \sum_{k=N+1}^{\infty} \pi(k) P_{kj} \\
&\overset{(1)}{<} \frac{\pi(j)}{2} + \sum_{k=N+1}^{\infty} \pi(k) P_{kj} \\
&\overset{(2)}{<} \frac{\pi(j)}{2} + \sum_{k=N+1}^{\infty} P_{kj} \\
&< \frac{\pi(j)}{2} + \frac{\pi(j)}{2} \\
&= \pi(j).
\end{aligned}$$

Where in (1) we have used the fact that π is a probability measure, and so the sum $\sum_{k=0}^{N} \pi(k) P^n_{kj}$ can be thought of averaging the P^n_{kj}. However, it is not a full average as $\sum_{k=0}^{N} \pi(k) \le 1$. In (2) we have used the fact that P is a stochastic matrix and so $P_{kj} \le 1$. Thus, we get a contradiction. ☐

Corollary 6.3.13. *A transient Markov chain has no invariant probability measures.*

6.4 Constructing Invariant Probability Measures

Given any recurrent state $i \in \mathcal{X}$ we can define a measure μ^i on $\mathcal{X}$, where for $j \in \mathcal{X}$ we let

$$\begin{aligned}
\mu^i(j) &= \mathbb{E}_i\left(\sum_{n=0}^{T_i-1} \mathbf{1}_{\{X_n=j\}}\right) \\
&= \mathbb{E}_i\left(\sum_{n=0}^{\infty} \mathbf{1}_{\{n<T_i\}} \mathbf{1}_{\{X_n=j\}}\right) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = j, T_i > n).
\end{aligned}$$

Remark 6.4.1. *Note that μ^i need not be a finite measure. Even if $\mathbb{P}_i(T_i < \infty) = 1$ we have*

$$\begin{aligned}\sum_{j\in\mathcal{X}} \mu^i(j) &= \sum_{j\in\mathcal{X}} \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = j, T_i > n) \\ &= \sum_{n=0}^{\infty} \mathbb{P}_i(T_i > n) \\ &= \mathbb{E}_i(T_i).\end{aligned}$$

Which is not necessarily finite as a random variable can be finite almost everywhere without having a finite expectation. For example, the Cauchy distribution.

Theorem 6.4.2. *If $i \in \mathcal{X}$ is recurrent for P, then μ^i is invariant for P.*

Proof. Fix a recurrent state $i \in \mathcal{X}$ and let $\mu = \mu^i$. We want to show that $(\mu P)(j) = \mu(j)$ for all $j \in \mathcal{X}$.

1. For $j \neq i$ we have

$$\begin{aligned}\mu(j) &= \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = j, T_i > n) \\ &= \sum_{k\in\mathcal{X}} \sum_{n=1}^{\infty} \mathbb{P}_i(X_n = j, T_i > n, X_{n-1} = k, T_i > n-1) \\ &= \sum_{k\in\mathcal{X}} \sum_{n=1}^{\infty} \mathbb{P}_i(X_n = j, T_i > n | X_{n-1} = k, T_i > n-1)\mathbb{P}_i(X_{n-1} = k, T_i > n-1) \\ &= \sum_{k\in\mathcal{X}} \sum_{n=1}^{\infty} \mathbb{P}_i(X_n = j | X_{n-1} = k, T_i > n-1)\mathbb{P}_i(X_{n-1} = k, T_i > n-1) \\ &\overset{(1)}{=} \sum_{k\in\mathcal{X}} P_{kj}\mu(k) \\ &= (\mu P)(j).\end{aligned}$$

   Where (1) is an application of the Markov property.

2. For $j = i$, on the one hand,

$$\begin{aligned}\mu(i) &= \mathbb{E}_i\left(\sum_{n=0}^{T_i-1} \mathbf{1}_{\{X_n=i\}}\right) \\ &\overset{(1)}{=} \mathbb{E}_i\left(\mathbf{1}_{\{X_0=i\}}\right) \\ &= 1,\end{aligned}$$

where (1) follows as by definition of T_i it must be that $\{X_k \neq i\}$ for $1 \leq k \leq T_i - 1$. On the other hand,

$$\begin{aligned}
(\mu P)(i) &= \sum_{k \in \mathcal{X}} \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = k, T_i > n) P_{ki} \\
&= \sum_{k \in \mathcal{X} \setminus \{i\}} \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = k, T_i > n) P_{ki} \\
&= \sum_{k \in \mathcal{X} \setminus \{i\}} \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = k, T_j > n) \mathbb{P}(X_{n+1} = i | X_n = k) \\
&= \sum_{k \in \mathcal{X} \setminus \{i\}} \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = k, T_i > n) \mathbb{P}(X_{n+1} = i | X_n = k, T_i > n) \\
&= \sum_{k \in \mathcal{X} \setminus \{i\}} \sum_{n=0}^{\infty} \mathbb{P}_i(X_{n+1} = i, X_n = k, T_i > n) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_i(X_{n+1} = i, T_i > n) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_i(T_i = n+1) \\
&= \mathbb{P}_i(T_i < \infty) \\
&= 1.
\end{aligned}$$

Where the last equality follows from the fact that i is recurrent. Therefore, $\mu(i) = (\mu P)(i)$.

□

Definition 6.4.3.

- *A recurrent state $i \in \mathcal{X}$ is positive recurrent if $\mathbb{E}_i(T_i) < \infty$.*
- *A recurrent state $i \in \mathcal{X}$ is null recurrent if $\mathbb{E}_i(T_i) = \infty$.*

Corollary 6.4.4. *If $i \in \mathcal{X}$ is positive recurrent then μ^i is an invariant finite measure.*

Proof. The invariance of μ^i follows directly from Theorem 6.4.2. The finiteness of μ^i follows from the positive recurrence on i. More specifically,

$$\begin{aligned}
\sum_{j \in \mathcal{X}} \mu^i(j) &= \sum_{j \in \mathcal{X}} \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = j, T_i > n) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_i(T_i > n) \\
&= \mathbb{E}_i(T_i) \\
&< \infty.
\end{aligned}$$

□

Lemma 6.4.5. *Let $i \in \mathcal{X}$ be recurrent, then for any invariant measure ν and $k \in \mathcal{X}$ we have*

$$\nu(k) \geq \nu(i)\mu^i(k).$$

Proof.

1. When $k = i$ then $\mu^i(i) = 1$ and so the result holds clearly.
2. When $k \neq i$, note that

$$\begin{aligned} P^n_{jk} &= \mathbb{P}_j(X_n = k) \\ &\overset{(1)}{\geq} \sum_{m=0}^{n-1} \mathbb{P}_j\left(X_n = k, \{X_m = i\} \cup \bigcap_{l=m+1}^{n-1} \{X_j \neq i\}\right) \\ &= \sum_{m=0}^{n-1} \mathbb{P}\left(X_n = k, \bigcap_{l=m+1}^{n-1} \{X_j \neq i\} \Big| X_m = i\right) P^m_{ji} \\ &= \sum_{m=0}^{n-1} \mathbb{P}_i\left(X_{n-m} = k, T_i > n-m\right) P^m_{ji}. \end{aligned}$$

Where in (1) we are using the fact that

$$\{X_m = i\} \cup \bigcap_{l=m+1}^{n-1} \{X_j \neq i\}$$

are disjoint events for $m = 0, \ldots, n-1$, whose union is not necessarily the whole sample space. Intuitively, these sets are the events that m is last visit to i before n. It follows using the invariance of ν that

$$\begin{aligned} \nu(k) &= (\nu P^n)(k) \\ &= \sum_{j\in\mathcal{X}} \nu(j) P^n_{jk} \\ &\geq \sum_{j\in\mathcal{X}} \nu(j) \left(\sum_{m=0}^{n-1} \mathbb{P}_i(X_{n-m} = k, T_i > n-m) P^m_{ji}\right) \\ &\geq \sum_{m=0}^{n-1} \mathbb{P}_i(X_{n-m} = k, T_i > n-m) \sum_{j\in\mathcal{X}} \nu(j) P^m_{ji} \\ &\geq \nu(i) \sum_{m=0}^{n-1} \mathbb{P}_i(X_{n-m} = k, T_i > n-m). \end{aligned}$$

Now note that we can just re-index the sum to get

$$\sum_{m=0}^{n-1} \mathbb{P}_i(X_{n-m} = k, T_i > n-m) = \sum_{l=1}^{n} \mathbb{P}_i(X_l = k, T_i > l),$$

Moreover, for $k \neq i$ by the construction of μ^i we know that

$$\begin{aligned} \mu^i(k) &= \sum_{l=1}^{\infty} \mathbb{P}_i(X_l = k, T_i > l) \\ &= \lim_{n\to\infty} \sum_{l=1}^{n} \mathbb{P}_i(X_l = k, T_i > l) \\ &=: \lim_{n\to\infty} \mu^i_n(k). \end{aligned}$$

Consequently, we have shown that

$$\nu(k) \geq \nu(i)\mu_n^i(k)$$

and so taking the limit preserves the inequality and we conclude that

$$\nu(k) \geq \nu(i)\mu^i(k).$$

□

Theorem 6.4.6. *If a Markov chain is irreducible and recurrent, then its invariant measure is unique up to a multiplicative constant.*

Proof. Let ν be invariant and set $\mu = \mu^i$ for some $i \in \mathcal{X}$. We know by Theorem 6.4.2 that μ is invariant. Moreover, as $\mu(i) = 1$ we can write

$$\begin{aligned} 0 &= \nu(i) - \nu(i)\mu(i) \\ &= (\nu P^n)(i) - \nu(i)(\mu P^n)(i) \\ &= \sum_{k \in \mathcal{X}} (\nu(k) - \nu(i)\mu(k)) P_{ki}^n. \end{aligned}$$

Note that $\nu(k) - \nu(i)\mu(k) \geq 0$ which implies that all the individual terms of the sum are zero. As the chain is irreducible for any k there is an n such that $P_{ki}^n > 0$. Therefore, for all k we have that $\nu(k) = \nu(i)\mu(k)$. Thus we have shown invariant measures are the same up to some multiplicative constant, $\nu(i)$. □

Theorem 6.4.7. *Suppose we have an irreducible Markov chain.*

1. *If the chain has an invariant probability measure π, then all the states are positive recurrent and*

$$\pi(i) = \frac{1}{\mathbb{E}_i(T_i)}$$

for $i \in \mathcal{X}$.

2. *If there exists a positive recurrent state, then the chain has a unique invariant measure and every state is positive recurrent.*

Proof.

1. As π is invariant probability measure, there exists a state $i \in \mathcal{X}$ such that $\pi(i) > 0$. Therefore, the contrapositive of Theorem 6.3.12 tells us that i is a recurrent state. Therefore, as recurrence is a communication class property, and the chain is irreducible, we deduce that every state is recurrent. For a fixed state $i \in \mathcal{X}$ we know by Theorem 6.4.2 that μ^i is an invariant measure. By Theorem 6.4.6 we know that μ^i is equal to π up to a multiplicative constant. That is, $\pi(j) = k\mu^i(j)$ for all $j \in \mathcal{X}$ and some $k \in \mathbb{R}$. Hence, μ^i is also a finite measure. Consequently,

$$\infty > \sum_{j \in \mathcal{X}} \mu^i(j) = \mathbb{E}_i(T_i),$$

and so i is a positive recurrent state. Moreover,

$$1 = \sum_{j \in \mathcal{X}} \pi(j) = \sum_{j \in \mathcal{X}} k\mu^i(j) = k\mathbb{E}_i(T_i)$$

implies that $k = \frac{1}{\mathbb{E}_i(T_i)}$. As $\mu^i(i) = 1$ we deduce that

$$\pi(i) = \frac{1}{\mathbb{E}_i(T_i)}.$$

Repeating this for each $i \in \mathcal{X}$ we arrive at the same conclusion for each state of the chain.

2. Suppose $i \in \mathcal{X}$ is a positive recurrent state of the chain, then we know μ^i is an invariant finite measure by Corollary 6.4.4. Therefore, we can normalise μ^i and follow the arguments of the previous step to deduce that all states of the chain are positive recurrent.

□

6.5 Long Run Dynamics

In the setting of irreducible, positive recurrent chains we know there exists a unique invariant probability measure π. Now we want to answer the question as to whether measures will converge under the dynamics of the chain to this unique invariant probability measure. Formally, given a measure ν on $\mathcal{X}$ we would like to understand under what constraints (νP^n) converges to π. Of course, this requires an understanding of what we mean by convergence. For a positive result, we require added constraints on the structure of the chain. To see why we require additional constraints refer to Example 6.5.1.

Example 6.5.1. *Let $\mathcal{X} = \{1, 2\}$ and $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Clearly, P is irreducible and positive recurrent. Note that $P^{2n} = I$ and $P^{2n+1} = P$ so that*

$$P^n_{11} = \begin{cases} 1 & n \text{ even} \\ 0 & n \text{ odd} \end{cases}.$$

Hence, we see that $\lim_{n\to\infty} P^n_{11}$ does not exist. Therefore, we cannot make any conclusions about the convergence of (νP^n) when $\nu = \delta_1$.

For a state $i \in \mathcal{X}$ we define the return times to i as

$$R(i) := \{n > 0 : P^n_{ii} > 0\}.$$

Definition 6.5.2. *The period of a state $i \in \mathcal{X}$, denoted $d(i)$, is*

$$d(i) := \begin{cases} \gcd(R(i)) & R(i) \neq \emptyset \\ \infty & R(i) = \emptyset. \end{cases}$$

Definition 6.5.3. *For $i \in \mathcal{X}$, if $d(i) = 1$ then i is called aperiodic, while if $d(i) > 1$ then i is called periodic.*

Remark 6.5.4.

- *If $R(i) = \emptyset$, and so $d(i) = \infty$, it must be the case that i is in its own communication class.*
- *Note that if $d(i) < \infty$ then one can have $d(i) \notin R(i)$.*

Example 6.5.5. *Consider the chain depicted in Figure 5. The chain is irreducible, and positive recurrent. However, it is not aperiodic as the state 1, for example, has a period of 4.*

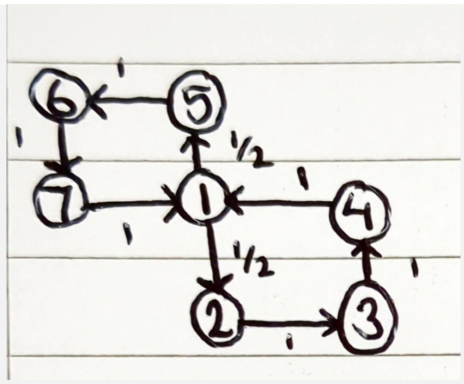


Figure 5: A chain that is irreducible, positive recurrent but not aperiodic.

Example 6.5.6. *Consider the chain depicted in Figure 6. Note that*

$$R(1) = \{3n + 4m : n, m \in \mathbb{N}\}.$$

In particular, $3, 4 \in R(1)$ *which implies that* $d(1) = 1$*. Moreover, this is an example of where* $d(i) \notin R(i)$*.*

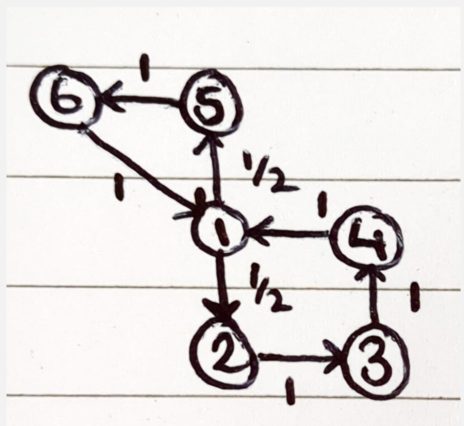


Figure 6: A chain that is irreducible, positive recurrent and aperiodic.

Proposition 6.5.7. *For* $i, j \in \mathcal{X}$ *in the same communication class. If* $i \neq j$ *then* $d(i) = d(j) < \infty$*.*

Proof. The condition that $i \neq j$ in the same communication class ensures that $d(i), d(j) < \infty$. As i and j are in the same communication class, we know there exists an $n \in \mathbb{N}$ such that $P_{ij}^n > 0$ and there exists an $m \in \mathbb{N}$ such that $P_{ji}^m > 0$. Consequently, by the Chapman-Kolmogorov equation we know that $P_{ii}^{m+n}, P_{jj}^{m+n} > 0$ which implies that $m+n \in R(i) \cap R(j)$. For $k \in R(i)$ we apply arguments involving the Chapman-Kolmogorov equation to note that $k + m + n \in R(j)$. Therefore, as $d(j) | n + m$ and $d(j) | k + n + m$ it follows that $d(j) | k$. As $k \in R(i)$ was arbitrary we conclude that $d(j)$ is a common divisor of $R(i)$. Hence, $d(j) \leq d(i)$. By symmetry, we also deduce that $d(i) \leq d(j)$ and so $d(i) = d(j)$. □

Definition 6.5.8. *A chain is aperiodic if every state is aperiodic, and it's periodic with period* d *if every state has period* d*.*

Corollary 6.5.9. *An irreducible chain is either periodic or aperiodic.*

Theorem 6.5.10. *Suppose P is irreducible, aperiodic, and positive recurrent. Let π denote its unique invariant probability measure. Then*

$$\lim_{n\to\infty}\left(\sum_{j\in\mathcal{X}}\left|P_{ij}^n-\pi(j)\right|\right)=0$$

for all $i\in\mathcal{X}$.

Proof. Let $(X_n)_{n\in\mathbb{N}}$ and $(X_n')_{n\in\mathbb{N}}$ be independent Markov process with transition probabilities P and initial distributions μ and ν respectively. Then by Lemma 9.2.8 we know that $Z_n=(X_n,X_n')$ is a time homogeneous Markov process on $\mathcal{X}\times\mathcal{X}$ with initial distribution $\mu\otimes\nu$ and transition probabilities

$$Q_{(i,i'),(j,j')}=P_{ij}P_{i'j'}$$

for all $i,j,i',j'\in\mathcal{X}$. Let $T=\inf\{n\geq: X_n=X_n'\}$. Using Lemma 9.2.6 we have the inequality

$$\sum_{j\in\mathcal{X}}|\mathbb{P}(X_n=j)-\mathbb{P}(X_n'=j)|\leq 2\mathbb{P}(T>n). \tag{6.5.1}$$

By Lemma 9.2.12 we know that $\mathbb{P}(T<\infty)=1$, which implies that $\mathbb{P}(T>n)\to 0$ as $n\to\infty$. Hence, by (6.5.1) we deduce that

$$\sum_{j\in\mathcal{X}}|\mathbb{P}(X_n=i)-\mathbb{P}(X_n'=i)|\overset{n\to\infty}{\longrightarrow}0.$$

In particular, we can take any $i\in\mathcal{X}$ and let $\mu=\delta_i$ and $\nu=\pi$ to see that

$$\lim_{n\to\infty}\left(\sum_{j\in\mathcal{X}}\left|P_{ij}^n-\pi(j)\right|\right)=0.$$

□

Remark 6.5.11. *With the notation used in the introduction to this section, Theorem 6.5.10 establishes the convergence of the measures $\nu=\delta_i$ for $i\in\mathcal{X}$ to π, in the sense outlined in the theorem, under the dynamics of the chain. This will be useful to use when we try to generalise to arbitrary probability measures.*

6.6 Total Variation

Definition 6.6.1. *The total variation distance between two probability measures μ and ν, on some measurable space $\mathcal{X}$, is given by*

$$\|\mu-\nu\|_{\mathrm{TV}}=2\sup_{A\subset\mathcal{X}}|\mu(A)-\nu(A)|,$$

where the supremum runs over all measurable subsets $A\subset\mathcal{X}$.

Remark 6.6.2.

1. *$\|\cdot\|_{\mathrm{TV}}$ has the dual formula,*

$$\|\mu-\nu\|_{\mathrm{TV}}=\sup_{f\in\mathcal{B}_b(\mathcal{X}),\|f\|_\infty\leq 1}\left|\int_{\mathcal{X}}f\,\mathrm{d}\mu-\int_{\mathcal{X}}f\,\mathrm{d}\nu\right|.$$

2. *Note that $\|\mu-\nu\|_{\mathrm{TV}}\in[0,2]$ with $\|\mu-\nu\|_{\mathrm{TV}}=0$ if and only $\mu=\nu$ and $\|\mu-\nu\|_{\mathrm{TV}}=2$ if and only if μ and ν are mutually singular. Where by mutually singular we mean that there exists a $A\subset\mathcal{X}$ such that*

$\mu(A) = 1$ and $\nu(A) = 0$.

Lemma 6.6.3. *Suppose that $\mathcal{X}$ is discrete, then*

$$\|\mu - \nu\|_{\mathrm{TV}} = \sum_{i \in \mathcal{X}} |\mu(i) - \nu(i)| = \|\mu - \nu\|_1.$$

Proof. Let $B = \{i \in \mathcal{X} : \mu(i) \geq \nu(i)\}$. Then since μ and ν are probability measures we note that

$$\begin{aligned}
0 &= 1 - 1 \\
&= \left(\sum_{i \in B} + \sum_{i \in B^c} \right) (\mu(i) - \nu(i)) \\
&= \sum_{i \in B} \mu(i) - \nu(i) + \sum_{i \in B^c} \mu(i) - \nu(i) \\
&= \sum_{i \in B} (\mu(i) - \nu(i)) - \sum_{i \in B^c} (\nu(i) - \mu(i)) .
\end{aligned}$$

Which implies that

$$\sum_{i \in B} \mu(i) - \nu(i) = \sum_{i \in B^c} \nu(i) - \mu(i).$$

Note that the terms of these sums are non-negative by construction of B, so that

$$\|\mu - \nu\|_1 = \sum_{i \in B} \mu(i) - \nu(i) + \sum_{i \in B^c} \nu(i) - \mu(i)$$

therefore,

$$\sum_{i \in B} \mu(i) - \nu(i) = \sum_{i \in B^c} \nu(i) - \mu(i) = \frac{1}{2} \|\mu - \nu\|_1.$$

Similarly,

$$\begin{aligned}
\|\mu - \nu\|_1 &= \sum_{i \in B} \mu(i) - \nu(i) + \sum_{i \in B^c} \nu(i) - \mu(i) \\
&= \mu(B) - \nu(B) + (\mu(B^c) - \nu(B^c)) \\
&= 2(\mu(B) - \nu(B)) \\
&\leq \|\mu - \nu\|_{\mathrm{TV}}.
\end{aligned}$$

For any $A \subset \mathcal{X}$ observe that

$$\begin{aligned}
|\mu(A) - \nu(A)| &= |\mu(A \cap B) - \nu(A \cap B) - (\mu(A \cap B^c) - \mu(A \cap B^c))| \\
&\leq 2 \max (|\mu(A \cap B) - \nu(A \cap B)|, |\mu(A \cap B^c) - \nu(A \cap B^c)|) \\
&\leq 2 \max (|\mu(B) - \nu(B)|, |\mu(B^c) - \nu(B^c)|) \\
&= 2|\mu(B) - \nu(B)| \\
&= \sum_{i \in \mathcal{X}} |\mu(i) - \nu(i)|.
\end{aligned}$$

Hence, taking the supremum of both sides we deduce that

$$\|\mu - \nu\|_{\mathrm{TV}} \leq \|\mu - \nu\|_1.$$

Combined with out previous observation that $\|\mu - \nu\|_1 \leq \|\mu - \nu\|_{\mathrm{TV}}$ we conclude the proof. $\square$

Definition 6.6.4. *Let $(\nu_n)_{n\in\mathbb{N}}$ be a sequence of measures and let ν be a measure.*

1. *We say $(\nu_n)_{n\in\mathbb{N}}$ converges in total variation to ν if $\|\nu_n - \nu\|_{\mathrm{TV}} \to 0$.*
2. *We say $(\nu_n)_{n\in\mathbb{N}}$ converges strongly to ν if $\nu_n(A) \to \nu(A)$ for every measurable A.*
3. *We say $(\nu_n)_{n\in\mathbb{N}}$ converges weakly if for every $f \in \mathcal{C}_b(\mathcal{X})$ we have*

$$\int_{\mathcal{X}} f \,\mathrm{d}\nu_n \to \int_{\mathcal{X}} f \,\mathrm{d}\nu.$$

Remark 6.6.5. *Note that convergence in total variation implies strong convergence which in turn implies weak convergence. However, the reverse implication in each of these cases does not hold.*

Example 6.6.6. *Consider the sequence of measures $(\nu_n)_{n\in\mathbb{N}}$ on $\mathbb{R}$ where $\nu_n = \delta_{\frac{1}{n}}$. Then for any $f \in \mathcal{C}_b(\mathcal{X})$ we have that*

$$\int_{\mathcal{X}} f\delta_{\frac{1}{n}} = f\left(\frac{1}{n}\right) \overset{n\to\infty}{\longrightarrow} f(0) = \int_{\mathcal{X}} f\delta_0.$$

Therefore, the sequence $(\nu_n)_{n\in\mathbb{N}}$ converges weakly to δ_0. However, $\|\nu_n - \delta_0\|_{\mathrm{TV}} = 2$ for all $n \in \mathbb{N}$, and so does not converge in total variation.

Theorem 6.6.7. *Suppose P is irreducible, aperiodic, and positive recurrent. Let π denote the unique invariant probability measure. Then for any probability measure ν on $\mathcal{X}$ we have that $\nu P^n \to \pi$ in total variation.*

Proof. Note that

$$\begin{aligned}
\|\mu P^n - \pi\|_{\mathrm{TV}} &= \sum_{j=1}^{\infty} \left| \sum_{i=1}^{\infty} \mu(i) P_{ij}^n - \pi(j) \right| \\
&= \sum_{j=1}^{\infty} \left| \sum_{i=1}^{\infty} \mu(i) P_{ij}^n - \sum_{i=1}^{\infty} \mu(i)\pi(j) \right| \\
&\le \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \mu(i) \left| P_{ij}^n - \pi(j) \right|.
\end{aligned}$$

Given an $\epsilon > 0$, we can choose N such that

$$\sum_{i=N+1}^{\infty} \mu(i) < \frac{\epsilon}{4}$$

as μ is a probability measure and so $\sum_{i\in\mathbb{N}} \mu(i) < \infty$. Consequently,

$$\sum_{i=N+1}^{\infty} \mu(i) \sum_{j=1}^{\infty} \left| P_{ij}^n - \pi(j) \right| < \frac{\epsilon}{2},$$

where we have just applied the triangle inequality to the inner sum, and the fact that P_{ij}^n and $\pi(j)$ are bounded by 1. On the other hand, by Theorem 6.5.10 we can choose M such that for all $i \le N$ we have that

$$\sum_{j=1}^{\infty} \left| P_{ij}^n - \pi(j) \right| \le \frac{\epsilon}{2}$$

for $n \ge M$. Therefore,

$$\sum_{j=1}^{\infty} \sum_{i=1}^{N} \mu(i) \left| P_{ij}^n - \pi(j) \right| < \sum_{i=1}^{N} \mu(i) \frac{\epsilon}{2} \le \frac{\epsilon}{2}.$$

Hence,

$$\begin{aligned}\|\mu P^n - \pi\|_{\mathrm{TV}} &\le \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}\mu(i)\left|P_{ij}^n - \pi(j)\right| \\ &= \sum_{j=1}^{\infty}\sum_{i=1}^{N}\mu(i)\left|P_{ij}^n - \pi(j)\right| + \sum_{j=1}^{\infty}\sum_{i=N+1}^{\infty}\mu(i)\left|P_{ij}^n - \pi(j)\right| \\ &\le \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon.\end{aligned}$$

□

Remark 6.6.8. *The proof of Theorem 6.6.7 utilises the result of Theorem 6.5.10, however, we could instead capitalise on the proof of Theorem 6.5.10. Note that Theorem 6.6.7 is just a stronger version of Theorem 6.5.10, which we have already proved. In fact, at the last step of the proof of Theorem 6.5.10 we can choose the initial distributions of our chains to be ν and π to arrive at Theorem 6.6.7.*

6.7 Periodic Chains

We now provide an alternative, but equivalent, definition of the period of an irreducible Markov chain.

Lemma 6.7.1. *The period of an irreducible stochastic matrix P is equal to the largest $d \in \mathbb{N}_{>0}$ such that one can partition the state space as*

$$\mathcal{X} = A_0 \sqcup \cdots \sqcup A_{d-1}$$

where if $i \in A_n$ then for j such that $P_{ij} > 0$ we have that $j \in A_{n+1 \bmod d}$.

Proof. Suppose P has period d. Fix a state i and define

$$A_n = \left\{j \in \mathcal{X} : P_{ij}^{kd+n} > 0 \text{ for some } k \in \mathbb{N}\right\}$$

for $n = 0, \ldots, d-1$. By irreducibility we know that $(A_n)_{n=0}^{d-1}$ forms a cover of $\mathcal{X}$. Suppose $j \in A_{n_1} \cap A_{n_2}$. This implies that there exists k_1, k_2 such that $P_{ij}^{k_1 d+n_1}, P_{ij}^{k_2 d+n_2} > 0$. However, by irreducibility, there exists a q such that $P_{ji}^q > 0$. So $k_1 d + n_1 + q$ and $k_2 d + n_2 + q$ are in $R(i)$ and so d divides $n_1 - n_2$ which implies $n_1 = n_2$. So we have shown the existence of such a decomposition of d disjoint sets. Now assume that p is the largest number for which such a decomposition of p disjoint sets exists. For $j \in \mathcal{X}$, if j returns to itself in q steps then we must have that $p|q$. Therefore, p is a divisor of the set of return times for j. Hence, by the definition of d we know that $p \le d$ and thus $p = d$ if it is the largest number where such a decomposition exists. □

Example 6.7.2. *For a d-periodic, irreducible stochastic matrix P, we note that by Lemma 6.7.1 the chain with the stochastic matrix P^d is restricted to one of the A^n for $n = 0, \ldots, d-1$. Consider the chain depicted in Figure 7. In this case, the chain has a period of 3 and a corresponding decomposition is given by*

- $A_0 = \{2\}$,
- $A_1 = \{4\}$, *and*
- $A_2 = \{1, 3\}$.

The corresponding stochastic matrix for such a chain has the form

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

for some $q \in (0,1)$. *We note that*

$$P^3 = \begin{pmatrix} q & 0 & 1-q & 0 \\ 0 & 1 & 0 & 0 \\ q & 0 & 1-q & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

which is no longer irreducible. The dynamics are constrained to the sets $\{2\}, \{4\}$ *and* $\{1,3\}$ *as expected.*

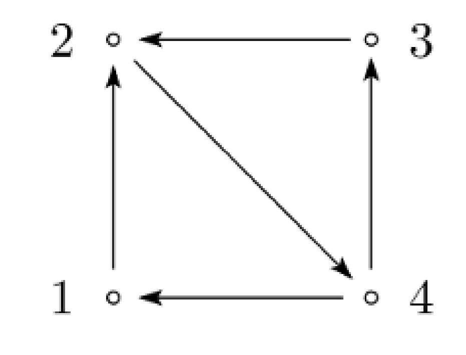


Figure 7: An example of a periodic, irreducible chain whose three-step dynamic is reducible.

Proposition 6.7.3. *Suppose* $T^n\mu = \mu$ *for some fixed* n, *and let* $\hat{\mu} = \frac{1}{n}\sum_{k=1}^n T^k\mu$. *Then* $T\hat{\mu} = \hat{\mu}$.

Proof. Let $A \in \mathcal{B}(\mathcal{X})$. Then

$$\begin{aligned} T\hat{\mu}(A) &= \frac{1}{n}\sum_{k=1}^{n} T^{k+1}\mu(A) \\ &= \frac{1}{n}\sum_{k=1}^{n-1} T^{k+1}\mu(A) + \frac{1}{n}T^{n+1}\mu(A) \\ &= \frac{1}{n}\sum_{k=2}^{n} T^{k}\mu(A) + \frac{1}{n}T\mu(A) \\ &= \hat{\mu}(A). \end{aligned}$$

□

Remark 6.7.4. *Suppose we have a period* d *chain on* $\mathcal{X}$ *with a decomposition*

$$\mathcal{X} = A_0 \sqcup \cdots \sqcup A_{d-1}$$

as in Lemma 6.7.1, and an invariant measure μ for P^d on A^n. Then

$$\hat{\mu} = \frac{1}{d}\sum_{k=1}^{d} \mu P^k$$

is an invariant measure for P on $\mathcal{X}$.

6.8 Ergodic Theorem

Theorem 6.8.1 (Strong Law of Large Numbers)**.** *Let $(\xi_n)_{n=1}^{\infty}$ be a sequence of independent and identically distributed real-valued random variables with $\mathbb{E}(|\xi_i|) < \infty$ and $\mathbb{E}(\xi_i) = a$. Then*

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right) = a$$

almost surely.

Exercise 6.8.2. *Suppose X is a time-homogeneous Markov process with initial distribution δ_i. Recall, that T^k is the k^{th} return time to state i. Show that the random variables*

$$\left\{\sum_{l=T^k+1}^{T^{k+1}} f(X_l) : k \in \mathbb{N}\right\}$$

are independent and identically distributed.

Theorem 6.8.3 (Ergodic Theorem)**.** *Let X be an irreducible, positive recurrent Markov chain on a discrete state space $\mathcal{X}$. Let π denote its unique invariant probability measure. Then for any π-integrable function $f : \mathcal{X} \to \mathbb{R}$ it follows that*

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{k=1}^{n} f(X_k)\right) = \sum_{j\in\mathcal{X}} f(j)\pi(j) = \int_{\mathcal{X}} f \, \mathrm{d}\pi \tag{6.8.1}$$

almost surely.

Proof. We first prove the statement for $f \geq 0$. Fix an arbitrary $i \in \mathcal{X}$ and write $T = T_i$ and $T^k = T_i^k$ for the passage times to i. Proving the statement for X initially distributed according to $\mathbb{P}_i$ will prove the statement for arbitrary initial distribution as i is arbitrary and

$$\mathbb{P}_\mu(\cdot) = \sum_{i\in\mathcal{X}} \mathbb{P}_i(\cdot)\mu(i).$$

We also write $\mu = \mu^i$ for the finite measure on $\mathcal{X}$ defined by

$$\mu(j) = \mathbb{E}\left(\sum_{k=1}^{T} \mathbf{1}_{\{X_k = j\}}\right) = \mathbb{E}_i(T)\pi(j).$$

Observe that,

$$
\begin{aligned}
\mathbb{E}_i\left(\sum_{l=1}^{T} f(X_l)\right) &= \mathbb{E}_i\left(\sum_{l=1}^{T}\sum_{j\in\mathcal{X}} \mathbf{1}_{\{X_l=j\}} f(j)\right) \\
&\overset{(1)}{=} \sum_{j\in\mathcal{X}} f(j)\mathbb{E}_i\left(\sum_{l=1}^{T}\mathbf{1}_{\{X_l=j\}}\right) \\
&= \mathbb{E}_i(T)\sum_{j\in\mathcal{X}} f(j)\pi(j) \\
&= \mathbb{E}_i(T)\int_{\mathcal{X}} f \,\mathrm{d}\pi \qquad (6.8.2) \\
&< \infty.
\end{aligned}
$$

Where we can exchange the order at of summation at (1) as the sum is absolutely convergent. Using Exercise 6.8.2 we can apply the strong law of large numbers to deduce that,

$$
\begin{aligned}
\lim_{n\to\infty}\left(\frac{1}{n}\sum_{k=1}^{n}\sum_{T^{k-1}+1}^{T^k} f(X_l)\right) &= \lim_{n\to\infty}\left(\frac{1}{n}\sum_{l=1}^{T^n} f(X_l)\right) \\
&\overset{(6.8.2)}{=} \mathbb{E}_i(T)\int_{\mathcal{X}} f\,\mathrm{d}\pi.
\end{aligned}
$$

almost surely. Since the differences of consecutive passage times are independent and identically distributed, Lemma 6.2.11, we can apply the strong law of large numbers to deduce that

$$
\lim_{n\to\infty}\left(\sum_{k=1}^{n}\left(T^k - T^{k-1}\right)\right) = \lim_{n\to\infty}\left(\frac{1}{n}T^n\right) = \mathbb{E}_i(T)
$$

almost surely. Now let

$$
\eta(n) = \sum_{k=1}^{n}\mathbf{1}_{\{X_k=i\}}
$$

for $n\in\mathbb{N}$. Observe that

$$
T^{\eta(n)} \le n < T^{\eta(n)+1}.
$$

It follows that

$$
\frac{1}{\eta(n)}\sum_{l=1}^{T^{\eta(n)}} f(X_l) \le \frac{1}{\eta(n)}\sum_{l=1}^{n} f(X_l) \le \frac{1}{\eta(n)}\sum_{l=1}^{T^{\eta(n)+1}} f(X_l). \qquad (6.8.3)
$$

As i is recurrent, by Theorem 6.3.6, we know the event $\{X_k = i\}$ occurs infinitely often with probability 1 and so $\eta(n)\to\infty$ almost surely. Therefore,

$$
\begin{aligned}
\lim_{n\to\infty}\frac{1}{\eta(n)}\sum_{l=1}^{T^{\eta(n)}} f(X_l) &= \lim_{n\to\infty}\frac{1}{\eta(n)}\sum_{l=1}^{T^{\eta(n)+1}} f(X_l) \\
&= \mathbb{E}_i(T)\int_{\mathcal{X}} f\,\mathrm{d}\pi
\end{aligned}
$$

almost surely. Hence, by (6.8.3) we have that

$$
\lim_{n\to\infty}\frac{1}{\eta(n)}\sum_{l=1}^{n} f(X_l) = \mathbb{E}_i(T)\int_{\mathcal{X}} f\,\mathrm{d}\pi \qquad (6.8.4)
$$

almost surely. Taking $f=1$ in (6.8.4) we note that

$$
\lim_{n\to\infty}\frac{n}{\eta(n)} = \mathbb{E}_i(T) > 0
$$

almost surely. Therefore,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^{n} f(X_l) &= \lim_{n\to\infty}\frac{\eta(n)}{n}\frac{1}{\eta(n)}\sum_{l=1}^{n} f(X_l)\\ &= \int_{\mathcal{X}} f\,\mathrm{d}\pi.\end{aligned}$$

To complete the proof we extend the result to general π-integrable function f by considering $f = f^{+} - f^{-}$. $\square$

Remark 6.8.4.

1. *Theorem 6.8.3 can be thought of as translating time averages into space averages.*
2. *Note that the statement of Theorem 6.8.3 holds for every initial distribution of X.*
3. *By taking $f = \mathbf{1}_j$ the equality $(\star)$ in our proof of Theorem 6.8.3 tells us that*

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{l=1}^{T^n}\mathbf{1}_j(X_l)\right) = \frac{\pi(j)}{\pi(i)}$$

almost surely. In other words, we have that $\frac{\pi(j)}{\pi(i)}$ is the average time spent at time j during one excursion starting and ending at i.

Corollary 6.8.5. *Suppose that X is an irreducible, positive recurrent and aperiodic on a discrete state space $\mathcal{X}$. Let π denote its unique invariant probability measure. Then for any $f : \mathcal{X} \to \mathbb{R}$ which is π-integrable, and probability measure μ on $\mathcal{X}$ we have that*

$$\mathbb{E}_\mu(f(X_k)) \to \int_{\mathcal{X}} f\,\mathrm{d}\pi$$

as $k \to \infty$. Moreover, it follows that

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}_\mu(f(X_k)) \to \int_{\mathcal{X}} f\,\mathrm{d}\pi \tag{6.8.5}$$

as $n \to \infty$.

Proof. Suppose $f = \mathbf{1}_j$ for some $j \in \mathcal{X}$, then

$$\mathbb{E}_\mu(f(X_k)) = \sum_{i\in\mathcal{X}} \mu P_{ij}^k.$$

We know by Theorem 6.5.10 that

$$\sum_{i\in\mathcal{X}}\left|\left(\mu P_{ij}^k\right) - \pi(i)\right| \to 0$$

as $k \to \infty$. Therefore,

$$\begin{aligned}\left|\mathbb{E}_\mu(f(X_k)) - \int_{\mathcal{X}} f\,\mathrm{d}\pi\right| &= \left|\sum_{i\in\mathcal{X}} \mu P_{ij}^k - \pi(j)\right| \\ &= \left|\sum_{i\in\mathcal{X}} \mu P_{ij}^k - \sum_{i\in\mathcal{X}} P_{ij}\pi(i)\right| \\ &\leq \sum_{i\in\mathcal{X}} \left|\mu P_{ij}^k - P_{ij}\pi(i)\right| \\ &\leq \sum_{i\in\mathcal{X}} \left|\mu P_{ij}^{k-1} - \pi(i)\right| \\ &\overset{k\to\infty}{\longrightarrow} 0.\end{aligned}$$

Now for $f : \mathcal{X} \to \mathbb{R}$ non-negative and π-integrable, we can write $f = \sum_{i\in\mathcal{X}} f(i)\mathbf{1}_i$. Let $f_n = \sum_{i=1}^n f(i)\mathbf{1}_i$. Then by the algebra of limits we know that

$$\mathbb{E}_\mu(f_n(X_k)) \to \int_{\mathcal{X}} f_n\,\mathrm{d}\pi$$

as $k \to \infty$. As f is non-negative we note that

$$\mathbb{E}_\mu(f_n(X_k)) \leq \mathbb{E}_\mu(f_{n+1}(X_k))$$

and so by monotone convergence we have that $\mathbb{E}_\mu(f_n(X_k)) \to \mathbb{E}_\mu(f(X_k))$ as $n \to \infty$. Similarly, we have that $\int_{\mathcal{X}} f_n\,\mathrm{d}\pi \to \int_{\mathcal{X}} f\,\mathrm{d}\pi$. Therefore,

$$\mathbb{E}_\mu(f(X_k)) \to \int_{\mathcal{X}} f\,\mathrm{d}\pi$$

as $k \to \infty$. We can then extend this result to general π-integrable functions $f : \mathcal{X} \to \mathbb{R}$ by considering $f = f^+ - f^-$. Note that (6.8.5) follows as for a converging sequence, the limit of the partial averages converges to the limit of the sequence. □

Remark 6.8.6. *Note that* (6.8.5) *is different from* (6.8.1) *as* (6.8.5) *contains an expectation.*

6.9 Reversible Markov Chains

Exercise 6.9.1. *For a, not necessarily discrete, state space $\mathcal{X}$ suppose we have a transition probability P and a probability measure μ on $\mathcal{X}$ such that $\mu P = \mu$. Show that one can construct a two-sided Markov process $(X_n : n \in \mathbb{Z})$ with*

- $\mathrm{Law}(X_n) = \mu$, *and*
- $\mathbb{P}(X_{n+1} \in A | X_n = x) = P(X, a)$

for all $n \in \mathbb{Z}$, $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$.

Definition 6.9.2. *For a two-sided stochastic process $X = (X_n)_{n\in\mathbb{Z}}$, we call the stochastic process $\hat{X} = \left(\hat{X}_m\right)_{m\in\mathbb{Z}}$ with $\hat{X}_m = X_{-m}$ the reversed stochastic process.*

Theorem 6.9.3. *Let P be an irreducible stochastic matrix with an invariant probability measure π. Let $X = (X_n)_{n\in\mathbb{Z}}$ be the two-sided Markov process constructed in Exercise 6.9.1 with transition probability P and $\mathrm{Law}(X_n) = \pi$. Let $\hat{X} = \left(\hat{X}_m\right)_{m\in\mathbb{Z}}$ be the reversed stochastic process of X. Then $\hat{X}$ is a time-homogeneous*

Markov chain with stochastic matrix

$$\hat{P}_{ji} = P_{ij}\frac{\pi(i)}{\pi(j)}.$$

Proof. As X is irreducible and there exists an invariant probability measure, every state $j \in \mathcal{X}$ is positive recurrent and so $\pi(j) > 0$. Given consecutive times $n_0 < n_1 < \cdots < n_l$, we have

$$\begin{aligned}
\mathbb{P}\left(\hat{X}_{n_0} = i_0, \ldots, \hat{X}_{n_l} = i_l\right) &= \mathbb{P}\left(X_{-n_0} = i_l, \ldots, X_{-n_l} = i_0\right) \\
&= \pi(i_l)P_{i_l,i_{l-1}} \ldots P_{i_1,i_0} \\
&= \left(\frac{\pi(i_l)}{\pi(i_{l-1})}P_{i_l,i_{l-1}}\right) \cdots \left(\frac{\pi(i_1)}{\pi(i_0)}P_{i_1,i_0}\right)\pi(i_0) \\
&= \hat{P}_{i_{l-1},i_1} \ldots \hat{P}_{i_0,i_l}\pi(i_0) \\
&= \pi(i_0)\hat{P}_{i_0,i_l} \ldots \hat{P}_{i_{l-1},i_1}.
\end{aligned}$$

Therefore, using Corollary 5.1.5 we deduce that $\hat{X}$ is Markov with transition probabilities $\hat{P}$. □

Definition 6.9.4. *A stochastic matrix P and a measure π satisfy detailed balance if*

$$\pi(i)P_{ij} = \pi(j)P_{ji}$$

for all $i, j \in \mathcal{X}$.

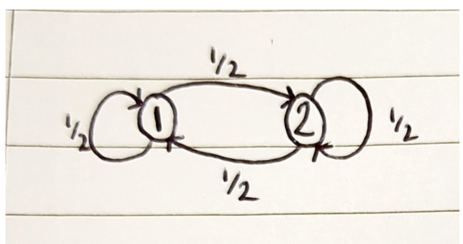


Figure 8: An example of a detailed balance chain.

Proposition 6.9.5. *Let π be a probability measure that satisfies detailed balance with respect to P. Then π is P-invariant.*

Proof. This is clear as

$$\begin{aligned}
\sum_{j\in\mathcal{X}} \pi(j)P_{ji} &= \sum_{j\in\mathcal{X}} \pi(i)P_{ij} \\
&= \pi(i)\sum_{j\in\mathcal{X}} P_{ij} \\
&= \pi(i).
\end{aligned}$$

□

Theorem 6.9.6. *In the setting of Theorem 6.9.3, if π and P satisfy detailed balance, then $\hat{P} = P$ and* $\mathrm{Law}\left(\hat{X}\right) = \mathrm{Law}(X)$.

A process satisfying the statement of Theorem 6.9.6 is called reversible. Intuitively, in an irreducible chain that satisfies detailed balance, one cannot tell whether the chain is being propagated forward or backward in time.

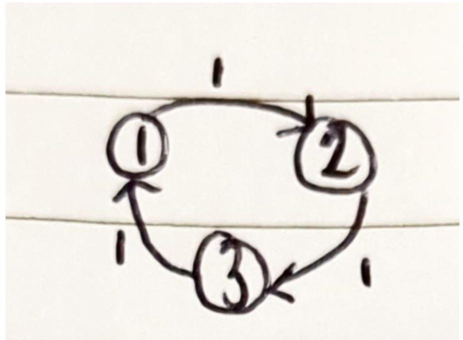


Figure 9: We know this is not a detailed balanced chain, as we can distinguish whether the chain is moving forward or backwards in time.

6.9.1 Markov Chain Monte Carlo

Suppose we want to simulate a probability measure π on a large but finite state space $\mathcal{X}$. That is, we want to calculate $\sum_{i\in\mathcal{X}} f(i)\pi(i)$ for observables $f : \mathcal{X} \to \mathbb{R}$. One approach is to construct a Markov chain for which π is its invariant distribution and use the ergodic law of large numbers to get that

$$\sum_{i\in\mathcal{X}} f(i)\pi(i) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k).$$

Despite not knowing the specific values of $\pi(i)$ for $i \in \mathcal{X}$, one can often model them up to constants of proportionality. We note that the ratios

$$\delta(j,i) = \frac{\pi(j)}{\pi(i)}$$

eliminate those constants. Ideally, one would sum these ratios over $j \in \mathcal{X}$ to determine $\frac{1}{\pi(i)}$, however, in practice this sum is expensive. Moreover, $\delta(i,j)$ may be intractable to compute for certain (i,j).

Example 6.9.7. *Let $\mathcal{X} = \{-1,1\}^{\Lambda}$ with $\Lambda = [-N,N]^d \cap \mathbb{Z}^d$. For a configuration $\sigma = (\sigma_x : x \in \Lambda)$, let*

$$\pi(\sigma) \propto \exp\left(\frac{\beta}{2} \sum_{x,y\in\Lambda,|x-y|=1} \sigma_x\sigma_y\right)$$

for some $\beta > 0$. This model is called the Ising model. An application of the Ising model is to approximate the dynamics of molecular spin within a material. In this setting calculating the normalisation constant to determine $\pi(\sigma)$ is difficult as materials contain a large number of molecules. However, if σ and σ' differ at exactly one $\bar{x} \in \Lambda$ then

$$\delta(\sigma,\sigma') = \exp\left(\beta \sum_{y\in\Lambda,|x-y|=1} (\sigma_x - \sigma'_x)\,\sigma_y\right).$$

Using these ideas we can now consider constructing a suitable Markov chain. Start with some irreducible Markov chain with transition probability Q and then set

1. $P_{ij} = Q_{ij} \wedge \delta(j,i)Q_{ji}$ for $i \neq j$, and
2. $P_{ii} = 1 - \sum_{j\neq i} P_{ij}$.

The chain with transition matrix P on $\mathcal{X}$ is not necessarily irreducible. However, showing P and π satisfy the detailed balance tells us that π is an invariant measure of the chain,

$$\begin{aligned}\pi(i)P_{ij} &= \pi(i)Q_{ij} \wedge \pi(j)Q_{ji} \\ &= \pi(j)Q_{ji} \wedge \pi(i)Q_{ij} \\ &= \pi(j)\left(Q_{ji} \wedge \delta(i,j)Q_{ij}\right) \\ &= \pi(j)P_{ji}.\end{aligned}$$

As multiple invariant measures may exist, due to P potentially not being irreducible, the chain may converge to a different invariant measure. However, supposing that the chain is irreducible, or that it will converge to π, we know by the ergodic law that

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \approx \sum_{i\in\mathcal{X}} f(i)\pi(i)$$

for large n.

6.10 Finite State Space Markov Chains

Throughout we will suppose that $\mathcal{X} = \{1,\dots,N\}$. It will useful to let

$$\delta_n = \min_{1\le i,j\le N} P^n_{ij}$$

for $n \in \mathbb{N}$.

Proposition 6.10.1. *Let P be a stochastic matrix on $\mathcal{X}$. Then δ_n is increasing in n. Moreover, the following are equivalent.*

1. *P is irreducible and aperiodic.*
2. *P^n is irreducible for every $n \ge 1$.*
3. *There exists some n_0 such that $\delta_{n_0} > 0$.*

Proof. The Chapman-Kolmogorov equation tells us that

$$\begin{aligned}P^{m+n}_{ij} &\ge \sum_{k\in\mathcal{X}} P^m_{ik}P^n_{kj} \\ &\ge \delta_n \sum_{k\in\mathcal{X}} P^m_{ik} \\ &= \delta_n\end{aligned}$$

for all $i,j \in \mathcal{X}$. Hence, $\delta_{m+n} = \min_{1\le i,j\le n} P^{m+n}_{ij} \ge \delta_n$ and so δ_n is increasing in n
$(2) \Rightarrow (1)$. If P were not aperiodic, then as the chain is irreducible and finite it follows that it must be periodic. Let P be d-periodic, then using the reasoning of Example 6.7.2 we see that P^d is reducible, which is a contradiction.
$(3) \Rightarrow (1)$. By assumption, P^{n_0} has strictly positive entries, which implies that P is irreducible. As δ_n is increasing we know that $\delta_n > \delta_{n_0} > 0$ for all $n \ge n_0$. Therefore, P^n also has strictly positive entries for $n \ge n_0$. Consequently, for any $i \in \mathcal{X}$ we have that $\{n_0, n_0+1, \dots\} \subseteq R(i)$ which implies that $d(i) = 1$ and so P is aperiodic.
$(3) \Rightarrow (2)$. By similar arguments to above, if $n < n_0$ there exists a k for which $\delta_{nk} > 0$. Which implies that P^n is irreducible.
$(1) \Rightarrow (3)$. By Lemma 9.2.9 we know that for all $1 \le i \le N$ there exists a k_i such that $kd(i) \in R(i)$ for $k \ge k_i$. As P is aperiodic $d(i) = 1$ for all i and so for all $n \ge N_0$ it follows that $n \in R(i)$ for all $i \in \mathcal{X}$. That is,

$$P^n_{ii} > 0$$

for all $i \in \mathcal{X}$ and $n \geq N_0$. As P is irreducible it follows that for $i, j \in \mathcal{X}$ there exists a $m(i,j) \in \mathbb{N}$ such that $P_{ij}^{m(i,j)} > 0$. Using the Chapman-Kolmogorov equation we deduce that

$$P_{ij}^{n+m(i,j)} \geq P_{ii}^{n} P_{ij}^{m(i,j)} > 0.$$

Hence, $n_0 = N_0 + \max_{i,j \in \mathcal{X}} m(i,j)$ is such that $\delta_{n_0} > 0$. □

Exercise 6.10.2. *Let P be a stochastic matrix. Then for a recurrent state $i \in \mathcal{X}$ and $j \in [i]$, show that $\mathbb{P}_j(T_i < \infty) = 1$.*

Lemma 6.10.3. *Let P be an irreducible, aperiodic stochastic matrix on a finite state space. Then for any $i, j \in \mathcal{X}$ and $\alpha > 0$ we have that*

$$\mathbb{E}_j\left(T_i^{\alpha}\right) < \infty.$$

Proof. Using Exercise 6.10.2 we know that $\mathbb{P}_j\left(T_i^{\alpha} = \infty\right) = 0$ and so we can write

$$\mathbb{E}_j\left(T_i^{\alpha}\right) = \sum_{n=0}^{\infty} n^{\alpha} \mathbb{P}_j(T_i = n).$$

Hence,

$$\mathbb{E}_j\left(T_i^{\alpha}\right) \leq \sum_{n=0}^{\infty} n^{\alpha} \mathbb{P}_j(T_i > n-1).$$

By Proposition 6.10.1 we know that there exists a n_0 such that $\delta_{n_0} > 0$, consequently,

$$\begin{aligned}
\mathbb{P}\left(X_{n_0(k+1)} \neq i | X_{n_0 k} \neq i\right) &= \sum_{l \neq i} \mathbb{P}\left(X_{n_0(k+1)} \neq i | X_{n_0 k} = l\right) \frac{\mathbb{P}(X_{n_0 k} = l)}{\mathbb{P}(X_{n_0 k} \neq i)} \\
&\leq \sum_{l \neq i} \left(1 - \delta_{n_0}\right) \frac{\mathbb{P}(X_{n_0 k} = l)}{\mathbb{P}(X_{n_0 k} \neq i)} \\
&\leq 1 - \delta_{n_0}
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathbb{P}_j(T_i > n_0(k+1)) &\leq \mathbb{P}_j\left(X_{n_0(k+1)} \neq i, T_i > n_0 k\right) \\
&= \mathbb{P}(X_{n_0(k+1)} \neq i | T_i > n_0 k) \mathbb{P}_j(T_i > n_0 k) \\
&\overset{(1)}{=} \mathbb{P}\left(X_{n_0(k+1)} \neq i | X_{n_0 k} \neq i\right) \mathbb{P}_j(T_i > n_0 k) \\
&\leq \left(1 - \delta_{n_0}\right) \mathbb{P}_j(T_i > n_0 k) \\
&\overset{(2)}{\leq} \left(1 - \delta_{n_0}\right)^{k+1}
\end{aligned}$$

where in (1) we used the fact that $\{T_i > n_0 k\} \in \mathcal{F}_{n_0 k}$ and so we can apply the Markov property to condition with respect to $\{X_{n_0 k} \neq i\} \, \sigma(X_{n_0 k})$ instead. In (2), we are just iterating the previous computations. Therefore,

$$\begin{aligned}
\sum_{n=2n_0}^{\infty} n^{\alpha} \mathbb{P}_j(T_i > n-1) &= \sum_{k=2}^{\infty} \sum_{j=n_0 k}^{n_0(k+1)-1} j^{\alpha} \mathbb{P}_j(T_i > j-1) \\
&\leq \sum_{k=2}^{\infty} \sum_{j=n_0 k}^{n_0(k+1)-1} (n_0(k+1))^{\alpha} \, \mathbb{P}_j(T_i > n_0 k - 1) \\
&\leq \sum_{k=2}^{\infty} \sum_{j=n_0 k}^{n_0(k+1)-1} (n_0(k+1))^{\alpha} \, (1-\delta_0)^{k+1} \\
&\leq \sum_{k=2}^{\infty} n_0^{\alpha+1} (k+1)^{\alpha} (1-\delta_0)^{k-1} \\
&< \infty.
\end{aligned}$$

□

6.10.1 Perron-Frobenius

Let $\mathbb{R}_+^N = \left\{ \eta \in \mathbb{R}^N : \eta(i) \geq 0 \text{ for all } 1 \leq i \leq N \right\}$.

Lemma 6.10.4. *Let P be irreducible and aperiodic on $\mathcal{X} = \{1, \ldots, N\}$. Then there exists some $n \in \mathbb{N}$ and $\delta > 0$ such that for every $\eta \in \mathbb{R}_+^N$ we have*

$$(\eta P^n)(i) \geq \delta \|\eta\|_1$$

for all $i \in \mathcal{X}$, where $\|\eta\|_1 = \sum_{i \in \mathcal{X}} \eta(i)$.

Proof. Take $n = n_0$ and $\delta = \delta_{n_0}$ as in Proposition 6.10.1, then

$$\begin{aligned}
(\eta P^n)(i) &= \sum_{j=1}^{N} \eta(j) P_{ji}^n \\
&\geq \delta_n \sum_{j=1}^{N} \eta(j) \\
&= \delta_n \|\eta\|_1.
\end{aligned}$$

□

Lemma 6.10.5. *Suppose that P is an irreducible stochastic matrix on $\mathcal{X} = \{1, \ldots, N\}$. Then there exists a $n \in \mathbb{N}$ and $\delta > 0$ such that $T^n = \frac{1}{n} \sum_{j=1}^{n} P^j$ satisfies*

$$\min_{1 \leq i,j \leq N} T_{ij}^n \geq \delta.$$

Proof. If P is aperiodic then this result follows from Lemma 6.10.4. So let P not be aperiodic, then as we are operating on a finite state space, and P is irreducible, it follows that P is periodic. Suppose that P has period d, then we can write

$$\mathcal{X} = A_0 \sqcup \cdots \sqcup A_{d-1}$$

where P^d is irreducible and aperiodic on A_j for $0 \leq j \leq d-1$. In particular, for $0 \leq l \leq d-1$, there exists a m_l such that

$$\min_{i,j \in A_l} \left(P^d\right)_{ij}^{m_l} > 0.$$

Now let $m = \max_{0 \le l \le d-1} m_l > 0$ and $n = 2dm + d$. Suppose $i \in A_l$ and $j \in A_{l'}$. If $l = l'$ then $P_{ij}^d > 0$ and $d \le 2dm + d$. Suppose instead that $|l - l'| = r > 0$. By irreducibility we know that there exists an $i' \in A_l$ and $j' \in A_{l'}$ such that $P_{i'j'}^r > 0$. Therefore,

$$P^{dm_l + dm_{l'} + r} \ge P_{ii}^{dm_l} P_{i'j'}^r P_{j'j}^{dm_{l'}} > 0$$

where $dm_l + dm_{l'} + r \le 2dm + d$. Thus, $T^n = \frac{1}{n} \sum_{j=1}^n P^j$ has the property that

$$\min_{1 \le i,j \le N} T_{ij}^n > 0,$$

where we maintain the strict inequality as our state space is finite. □

Theorem 6.10.6 (Perron-Frobenius)**.** *Let P be a $N \times N$ irreducible stochastic matrix on a finite state space $\mathcal{X}$. Then all the eigenvalues of P satisfy $|\lambda| \le 1$. Moreover, the real number 1 is a left-eigenvalue with a unique real left-eigenvector π, up to multiplication by a constant, that is $\pi P = \pi$. In particular, π can be chosen so that $\pi(i) > 0$ for every i and $\sum_{i=1}^N \pi(i) = 1$.*

Proof. As P is a stochastic matrix it follows that

$$\|\eta P\|_1 \le \|\eta\|_1$$

for every $\eta \in \mathbb{C}^N$ which shows that $|\lambda| \le 1$ for any eigenvalue λ of P. Observe that $\frac{1}{N}(1, \ldots, 1)$ is a right-eigenvector of P with eigenvalue 1, and so there must be a left-eigenvector π of P with eigenvalue 1. Since P is real we can take π to be real as well, moreover, we can normalize π such that $\|\pi\|_1 = 1$. Now suppose that π_+ and π_- are both non-zero, where $\pi_\pm(i) = \max(\pm\pi(i), 0)$ for $1 \le i \le N$. That is, π contains both positive and negative entries. Let $\alpha = \min(\|\pi_+\|_1, \|\pi_-\|_1) > 0$. By the irreducibility of P we can consider a T^n and $\delta > 0$ satisfying the statement of Lemma 6.10.5. Note that π is an eigenvector of T^n with eigenvalue 1. For $\eta \in \mathbb{R}^N$, we can write $\eta = \eta_+ - \eta_-$. Thus

$$\begin{aligned}
\|\pi\|_1 &= \|\pi T^n\|_1 \\
&= \|\pi_+ T^n - \pi_- T^n\|_1 \\
&\le \|\pi_+ T^n - \delta\alpha\mathbf{1}\|_1 + \|\pi_- T^n - \delta\alpha\mathbf{1}\|_1 \\
&\overset{(1)}{\le} \|\pi_+ T^n\|_1 + \|\pi_- T^n\|_1 - 2\|\alpha\delta\mathbf{1}\|_1 \\
&= \|\pi_+ T^n\|_1 + \|\pi_- T^n\|_1 - 2\alpha\delta N \\
&\overset{(2)}{\le} \|\pi_+\|_1 + \|\pi_-\|_1 - 2\alpha\delta N \\
&\overset{(3)}{=} \|\pi\|_1 - 2\alpha\delta N,
\end{aligned}$$

which is a contradiction for $\alpha > 0$. In (1) we have used Lemma 6.10.5 to deduce that

$$\|\pi_\pm T^n - \delta\alpha\mathbf{1}\|_1 = \|\pi_\pm T^n\|_1 - \|\delta\alpha\mathbf{1}\|_1 .$$

In (2) we have used that $\|\eta T^n\| \le \|\eta\|_1$. In (3), we have used the fact that $\|\pi\|_1 = \|\pi_+\|_1 + \|\pi_-\|_1$. As we have established that $\pi \in \mathbb{R}_+^N$, by construction of the T^n it follows that

$$\pi(i) = (\pi T^n)(i) \ge \|\pi\|_1,$$

that is, all its entries are positive. For uniqueness, we suppose that $\tilde{\pi} \in \mathbb{R}_+^N$ is another real left-eigenvector with eigenvalue 1. Moreover, we can assume that $\tilde{\pi}$ has positive entries and $\|\tilde{\pi}\|_1 = 1$. Then $\theta = \pi - \tilde{\pi}$ is another real left-eigenvector with eigenvalues 1, and so its entries must all have the same sign by our previous arguments. However, note that

$$\sum_{i \in \mathcal{X}} \theta(i) = \sum_{i \in \mathcal{X}} \pi(i) - \sum_{i \in \mathcal{X}} \tilde{\pi}(i) = \|\pi\|_1 - \|\tilde{\pi}\|_1 = 0.$$

Hence, $\theta(i) = 0$ for all i, which implies that $\pi = \tilde{\pi}$. □

Remark 6.10.7.

- *One of the consequences of Theorem 6.10.6, is that any irreducible Markov chain on a finite state space has an invariant probability measure.*
- *The specific π outlined in Theorem 6.10.6 is known as the Perron-Frobenius vector of P.*

6.11 Solution to Exercises

Exercise 6.1.3

Solution. As $P^0_{ii} = 1$ for all $i \in \mathcal{X}$ it follows that $i \leftrightarrow i$.
If $i \leftrightarrow j$ then $i \to j$ and $j \to i$, hence $j \leftrightarrow j$.
If $i \leftrightarrow j$ and $j \leftrightarrow k$. Then there exists an n_1 such that $(P^{n_1})_{ij} > 0$ and a n_2 such that $(P^{n_2})_{jk} > 0$. Therefore,

$$\left(P^{n_1+n_2}\right)_{ik} \geq (P^{n_1}))_{ij} \left(P^{n_2}\right)_{jk} > 0$$

and so $i \to k$. Similarly, $k \to i$ so that $i \leftrightarrow k$. □

Exercise 6.1.6

Solution. Let i and j be such that $[i] \leq [j]$. Consider $i' \in [i]$ and $j' \in [j]$. Then by Lemma 6.1.5 we have $j' \to i'$ and thus $[i'] \leq [j']$. Therefore, $\leq$ is well-defined.
It is clear that $\leq$ is reflexive as $i \to i$ trivially and so $[i] \leq [i]$.
If $[i] \leq [j]$ and $[j] \leq [k]$ then $j \to i$ and $j \to k$. Therefore, as seen previously we have $k \to i$ which implies that $[i] \leq [k]$.
If $[i] \leq [j]$ and $[j] \leq [i]$ then $i \leftrightarrow j$ so that $[i] = [j]$, as $\leftrightarrow$ is an equivalence relation. □

Exercise 6.2.8

Solution. Observe that

$$\begin{aligned}
\mathbb{P}_\mu(T_j < \infty) &= \sum_{i \in [j]} \mathbb{P}_i(T_j < \infty)\mu(\{i\}) \\
&\overset{\text{Lem. 6.2.7}}{=} \sum_{i \in [j]} (1)\mu(\{i\}) \\
&= 1.
\end{aligned}$$

□

Exercise 6.2.10

Solution. We can proceed by induction on n. Note that $T^0_j = 0$ is clearly a stopping time as it is a constant. Assume that T^m_j is a $(\mathcal{F}_k)$-stopping time for $m \leq n-1$. Observe that $\left\{T^n_j \leq k\right\} = \emptyset \in \mathcal{F}_k$ for $k \leq n-1$. Therefore, suppose that $k \geq n$ then

$$\left\{T^n_j \leq k\right\} = \bigcup_{l=n-1}^{k} \left(\left\{T^{n-1}_j = l\right\} \cap \bigcap_{p=l+1}^{k} \{X_p = j\} \right).$$

We know that $\{X_p = j\} \in \mathcal{F}_p \subseteq \mathcal{F}_k$ and by the inductive assumption we know that $\left\{T^{n-1}_j = l\right\} \in \mathcal{F}_l \subseteq \mathcal{F}_k$. Therefore, $\left\{T^n_j \leq k\right\} \in \mathcal{F}_k$, which means that T^n_j is a $(\mathcal{F}_k)$-stopping time. □

Exercise 6.8.2

Solution. Let $Y_k = \sum_{j=T^k+1}^{T^{k+1}} f(X_j)$, and consider $g, h \in \mathcal{B}(\mathbb{R})$. Without loss of generality suppose that $k' = k+n$ for $n \geq 1$. Then,

$$\begin{aligned}
\mathbb{E}(g(Y_k)h(Y_{k'})) &= \mathbb{E}\left(\mathbb{E}\left(g(Y_k)h(Y_{k'})|\mathcal{F}_{T^k}\right)\right) \\
&\overset{\text{SMP}}{=} \mathbb{E}_i(g(Y_0)h(Y_n)) \\
&= \mathbb{E}_i\left(\mathbb{E}_i\left(g(Y_0)h(Y_n)|\mathcal{F}_{T^n}\right)\right) \\
&\overset{(1)}{=} \mathbb{E}_i\left(g(Y_0)\mathbb{E}_i\left(h(Y_n)|\mathcal{F}_{T^n}\right)\right) \\
&\overset{\text{SMP}}{=} \mathbb{E}_i\left(g(Y_0)\mathbb{E}_i\left(h(Y_0)\right)\right) \\
&= \mathbb{E}_i(g(Y_0))\mathbb{E}_i(h(Y_n)) \\
&\overset{\text{SMP}}{=} \mathbb{E}(g(Y_k))\mathbb{E}(h(Y_k)),
\end{aligned}$$

where SMP denotes an application of the strong Markov property, and (1) follows as $g(Y_0)$ is $\mathcal{F}_{T^n}$ measurable. As the $g, h \in \mathcal{B}(\mathbb{R})$ were arbitrary this shows the mutual independence of Y_k and $Y_{k'}$. One easily extends the above argument to show that

$$\mathbb{E}\left(\prod_{j=1}^{m} g_i(Y_i)\right) = \prod_{j=1}^{m} \mathbb{E}(g_i(Y_i))$$

for all $g_i \in \mathcal{B}(\mathbb{R})$ and $m \in \mathbb{N}$. This then shows the independence of the random variables $\{Y_k\}_{k\in\mathbb{N}}$. Moreover, for any $k \in \mathbb{N}$ we have that

$$\begin{aligned}
\mathbb{P}(Y_k \in A) &= \mathbb{E}(\mathbf{1}_A(Y_k)) \\
&= \mathbb{E}\left(\mathbb{E}\left(\mathbf{1}_A(Y_k)|\mathcal{F}_{T^k}\right)\right) \\
&\overset{\text{SMP}}{=} \mathbb{E}_i\left(\mathbf{1}_A(Y_0)\right) \\
&= \mathbb{P}_i(Y_0 \in A)
\end{aligned}$$

where SMP denotes an application of the strong Markov property. Therefore, each Y_k is identically distributed. Hence, we have shown that the random variables

$$\left\{\sum_{l=T^k+1}^{T^{k+1}} f(X_l) : k \in \mathbb{N}\right\}$$

are independent and identically distributed. □

Exercise 6.9.1

Solution. For any $m \in \mathbb{Z}$, as μ is stationary with respect to P we note that

$$\mathbb{P}(X_m \in A_0, \ldots X_{m+n} \in A_n) = \mathbb{P}\left(X_0 \in A_0, \ldots, X_n \in A_n\right) \tag{6.11.1}$$

for any $n \in \mathbb{N}$. Note that the the family of measures $(\mu_n)_{n\in\mathbb{N}}$ given by $\mu_n = \mathrm{Law}(X_0, \ldots, X_n)$ is consistent. Consequently, we can apply Theorem 3.2.7 to construct a stochastic process such that

- $\mathrm{Law}(X_n) = \mu$, and
- $\mathbb{P}(X_{n+1}|X_n = x) = P(X, a)$

for $n \in \mathbb{N}$, $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$. Using (6.11.1) it is clear that we can extend this stochastic process to $\mathbb{Z}$, with the properties detailed above now holding for $n \in \mathbb{Z}$, $x \in \mathcal{X}$ and $A \in \mathcal{B}$ as required. □

Exercise 6.10.2

Solution. Let $\eta_i = \sum_{n=1}^{\infty} \mathbf{1}_{\{X_n = i\}}$. Then as i is a recurrent state it follows that $\mathbb{P}_i(\eta_i = \infty) = 1$. Note that by the Markov property it follows that

$$\{\eta_i = \infty\} = \{X_n = i, \text{ for some } n \geq m\}$$

for any $m \in \mathbb{N}$. Observe that

$$\begin{aligned}
1 &= \mathbb{P}_i(X_n = i, \text{ for some } n \geq m) \\
&= \sum_{k \in \mathcal{X}} P_{ik}^m \mathbb{P}_i(X_n = i, \text{ for some } n \geq m | X_m = k) \\
&= \sum_{k \in \mathcal{X}} P_{ik}^m \mathbb{P}_k(T_i < \infty) \\
&\leq \sum_{k \in \mathcal{X}} P_{ik}^m \\
&= 1.
\end{aligned}$$

Therefore, for the $k \in \mathcal{X}$ for which $P_{ij}^m \neq 0$ it follows that $\mathbb{P}_k(T_i < \infty) = 1$. Hence, for $j \in [i]$ we can let m be such that $P_{ij}^m > 0$ and deduce that $\mathbb{P}_j(T_i < \infty) = 1$. $\square$

7 Continuous State Space Markov Processes

Now we generalise to the state space $\mathcal{X}$ being a separable complete metric space. For instance, $\mathcal{X} = \mathbb{R}^n$. Moreover, we can let $\mathcal{X}$ be an infinite dimensional Banach space, for instance $\mathcal{X} = \mathcal{C}([0,1])$.

7.1 Weak Convergence

Recall that in this setting we work with transition probabilities $P(\cdot, \bullet)$ where for $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$ we have

$$P(x, A) = \mathbb{P}(X_{n+1} \in A | X_n = x).$$

Moreover, we can use P to act on functions and measures using the following.

- On functions we use $T_* : \mathcal{B}_b(\mathcal{X}) \to \mathcal{B}_b(\mathcal{X})$ given by $(T_* f)(\cdot) = \int_{\mathcal{X}} f(y) P(\cdot, \mathrm{d}y)$. Equivalently,

$$(T_* f)(x) = \mathbb{E}(f(X_{n+1}) | X_n = x).$$

- On measures we use $T^* : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$ given by $(T^* \mu)(\bullet) = \int_{\mathcal{X}} P(y, \bullet) \mu(\mathrm{d}y)$. Equivalently,

$$(T^* \mu)(A) = \int_{\mathcal{X}} \mathbb{P}(X_{n+1} \in A | X_n = y) \mu(\mathrm{d}y).$$

Remark 7.1.1. *In most cases $P(\bullet, A) = 0$ for $A = \{y\}$ a singleton. Therefore, notions developed previously regarding irreducibility, recurrence and transience do not generalise to this setting.*

Henceforth, we will denote the set of bounded and continuous functions on $\mathcal{X}$ as $\mathcal{C}_b(\mathcal{X})$.

Lemma 7.1.2. *Let $\mu, \mu' \in \mathcal{P}(\mathcal{X})$ be such that*

$$\int_{\mathcal{X}} f \, \mathrm{d}\mu = \int_{\mathcal{X}} f \, \mathrm{d}\mu'$$

for every $f \in \mathcal{C}_b(\mathcal{X})$. Then $\mu = \mu'$.

As $\mathcal{C}_b(\mathcal{X})$ can distinguish probability measures, it is natural to use them to define a notion of convergence.

Definition 7.1.3. *A sequence of probability measures $(\mu_n)_{n \in \mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ converges weakly to $\mu \in \mathcal{P}(\mathcal{X})$ if*

$$\int_{\mathcal{X}} f(x) \, \mu_n(\mathrm{d}x) \overset{n \to \infty}{\longrightarrow} \int_{\mathcal{X}} f(x) \, \mu(\mathrm{d}x)$$

for every $f \in \mathcal{C}_b(\mathcal{X})$.

Remark 7.1.4.

1. *On $\mathcal{P}(\mathcal{X})$ there exists a metric $d(\cdot, \cdot)$, known as the Levy-Prokhorov metric, that topologizes weak convergence. That is, μ_n converges weakly to μ if and only if $d(\mu_n, \mu) \to 0$.*
2. *For random variables we have that $Z_n \to Z$ almost every implies $Z_n \to Z$ in probability which implies that $\mathrm{Law}(Z_n) \to \mathrm{Law}(Z)$ weakly.*

7.2 The Feller Property

Definition 7.2.1. *A transition probability P, or T, is Feller if T^* maps $\mathcal{C}_b(\mathcal{X}) \subset \mathcal{B}_b(\mathcal{X})$ into itself. That is, $T^*(\mathcal{C}_b(\mathcal{X})) \subseteq \mathcal{C}_b(\mathcal{X})$. Moreover, P, or T, is strong Feller if $T^*(\mathcal{B}_b(\mathcal{X})) \subseteq \mathcal{C}_b(\mathcal{X})$.*

Remark 7.2.2. *A transition probability P being Feller is equivalent to the map $\mathcal{X} \to \mathcal{P}(\mathcal{X})$ given by $x \mapsto P(x, \bullet)$ being continuous, where $\mathcal{P}(\mathcal{X})$ is equipped with the topology of weak convergence.*

Example 7.2.3.

- *For $\mathcal{X} = \mathbb{R}$ consider the transition probability*

$$P(x, \cdot) = \begin{cases} \delta_1 & x > 0 \\ \delta_0 & x \leq 0. \end{cases}$$

  Then

$$(T_* f)(x) = \begin{cases} f(1) & x > 0 \\ f(0) & x \leq 0. \end{cases}$$

  Hence, P is not Feller.

- *Let X be a homogeneous Markov process given by $X_n = Y_n + X_{n-1}$ where $(Y_n)_{n=1}^{\infty}$ are $\mathrm{i.i.d}$ with law μ.*

  1. *If μ is the law taking values ± 1 with mean 0, then P is Feller but not strong Feller. To see this note that*

$$(T_* f)(x) = \frac{f(x+1) + f(x-1)}{2}$$

  2. *If $\mu = \mathcal{N}(0,1)$, then P is strong Feller. To see this note that*

$$(T_* f)(x) = \frac{1}{\sqrt{2\pi}} \int f(y) e^{-\frac{(x-y)^2}{2}} \, \mathrm{d}y.$$

Definition 7.2.4. *Let $\mathcal{X}$ be a separable metric space. Given $\mu \in \mathcal{P}(\mathcal{X})$ let $\operatorname{supp}(\mu)$ be the intersection of all closed sets $C \subset \mathcal{X}$ with $\mu(C) = 1$.*

Lemma 7.2.5. *Given a separable metric space $\mathcal{X}$ and $\mu \in \mathcal{P}(\mathcal{X})$ one has*

$$\operatorname{supp}(\mu) = \{x \in \mathcal{X} : \mu(B(x,\epsilon)) > 0, \text{ for every } \epsilon > 0\}.$$

Proof. Let $x \in E := \{x \in \mathcal{X} : \mu(B(x,\epsilon)) > 0, \text{ for every } \epsilon > 0\}$. Let C be a closed set with $\mu(C) = 1$. Suppose that $x \notin C$, so that there exists an $\epsilon > 0$ such that $B(x,\epsilon) \not\subseteq C$. It follows that

$$\mu(C \cup B(x,\epsilon)) = \mu(C) + \mu(B(x,\epsilon)) > 1,$$

which contradicts μ being a probability measure. Hence, $x \in C$ which implies that $E \subseteq \operatorname{supp}(\mu)$. Now consider $x \in \operatorname{supp}(\mu)$ and suppose that there exists an $\epsilon > 0$ for which $\mu(B(x,\epsilon)) > 0$. For any closed set C with $\mu(C)$ we know that $x \in C$. In particular, $C \cap B(x,\epsilon)^c$ is also closed and such that

$$\mu(C \cap B(x,\epsilon)^c) = \mu(C) + \mu\left(B(x,\epsilon)^c\right) - \mu\left(C \cup B(x,\epsilon)^c\right) \geq 1 + 1 - 1 = 1.$$

Hence, $C \cap B(x,\epsilon)^c$ is a closed set with $\mu\left(C \cap B(x,\epsilon)^c\right) = 1$, but $x \notin C \cap B(x,\epsilon)^c$ which contradicts $x \in \operatorname{supp}(\mu)$. Therefore, $\mu(B(x,\epsilon)) > 0$ for all $\epsilon > 0$ and so $\operatorname{supp}(\mu) \subseteq E$. $\square$

Proposition 7.2.6. *Show that $\mu(\operatorname{supp}(\mu)) = 1$, and so $\operatorname{supp}(\mu)$ is the smallest closed set of $\mathcal{X}$ with full μ-measure.*

Proof. The set $V := \mathcal{X} \setminus \operatorname{supp}(\mu)$ is separable as $\mathcal{X}$ is separable. Let $Q \subset V$ be countably dense in V. By Lemma 7.2.5 we know that for all $q \in Q$ there exists a $\tilde{\epsilon}_q > 0$ such that $\mu(B(q, \tilde{\epsilon}_q)) = 0$. In particular, $\mu(B(q, \epsilon)) = 0$ for all $\epsilon < \tilde{\epsilon}_q$. Hence, as V is open we can choose $\epsilon_q > 0$ such that $\mu(B(q, \epsilon_q)) = 0$ and $B(q, \epsilon_q) \subset V$. As Q is dense we have $V \subset \bigcup_{q \in Q} B(q, \epsilon_q)$, hence, as Q is countable it follows by countable additivity that

$$\mu(V) \leq \sum_{q \in Q} B(q, \epsilon_q) = 0.$$

Therefore, $\mu(\operatorname{supp}(\mu)) = 1$. Now suppose that $C \subset \operatorname{supp}(\mu)$ is a closed subset such that $\operatorname{supp}(\mu) \setminus C \neq \emptyset$ and $\mu(C) = 1$. As $\operatorname{supp}(\mu) \setminus C$ is non-empty and open there exists an $x \in \mathcal{X}$ and $\epsilon > 0$ such that $B(x, \epsilon) \in \operatorname{supp}(\mu) \setminus C$. This implies that $x \in \operatorname{supp}(\mu)$ and so by Lemma 7.2.5 we must have $\mu(B(x, \epsilon)) > 0$. However, this would mean that

$$\mu(C \cup B(x, \epsilon)) = \mu(C) + \mu(B(x, \epsilon)) > 1,$$

which is a contradiction as μ is a probability measure. Therefore, $\operatorname{supp}(\mu) \setminus C = \emptyset$ and $\operatorname{supp}(\mu)$ is the smallest closed subset with full measure. □

Exercise 7.2.7. *Recall that $\mu, \nu \in \mathcal{P}(\mathcal{X})$ are mutually singular, denoted $\mu \perp \nu$, if there exists $A \in \mathcal{B}(\mathcal{X})$ such that $\mu(A) = 1$ and $\nu(A) = 0$. Show that $\operatorname{supp}(\mu) \cap \operatorname{supp}(\nu) = \emptyset$ implies that $\mu \perp \nu$. However, show that μ and ν being mutually singular does not guarantee that $\operatorname{supp}(\mu) \cap \operatorname{supp}(\nu) = \emptyset$.*

Theorem 7.2.8. *Let $\mu, \nu \in \mathcal{P}(\mathcal{X})$ be invariant for a transition operator T. Suppose that T has the strong Feller property, then $\mu \perp \mu$ implies that $\operatorname{supp}(\mu) \cap \operatorname{supp}(\nu) = \emptyset$.*

Proof. Due to mutual singularity there exists a $F \in \mathcal{B}(\mathcal{X})$ such that $\mu(F) = 1$ and $\nu(F) = 0$. Let $\psi = T\mathbf{1}_F$, then $0 \leq \psi \leq 1$ and $\psi \in \mathcal{C}_b(\mathcal{X})$ by the strong Feller property of T. By the invariance of ν it follows that

$$\int_{\mathcal{X}} \psi(y)\, \nu(\mathrm{d}y) = \int_{\mathcal{X}} \mathbf{1}_F\, \nu(\mathrm{d}y) = \nu(F) = 0. \tag{7.2.1}$$

Similarly,

$$\int_{\mathcal{X}} \psi(y)\, \mu(\mathrm{d}y) = \int_{\mathcal{X}} \mathbf{1}_F(y)\, \mu(\mathrm{d}y) = \mu(F) = 1. \tag{7.2.2}$$

Consider the disjoint closed sets $A = \psi^{-1}(\{0\})$ and $B = \psi^{-1}(\{1\})$. As $0 \leq \psi \leq 1$ by (7.2.1) we must have $\nu(A) = 1$ and similarly by (7.2.2) $\mu(B) = 1$. Consequently, $\operatorname{supp}(\nu) \subset A$ and $\operatorname{supp}(\mu) \subset B$. As A and B are disjoint it follows that $\operatorname{supp}(\nu) \cap \operatorname{supp}(\mu) = \emptyset$. □

With the following we develop a criterion for showing a transition operator is strong Feller.

Lemma 7.2.9. *Let $f \in \mathcal{B}_b(\mathbb{R}^n)$ and $g \in L^1(\mathbb{R}^n)$, then $f \star g \in \mathcal{C}_b(\mathbb{R}^n)$, where*

$$(f \star g)(x) = \int_{\mathbb{R}^n} f(y) g(x-y)\, \mathrm{d}y.$$

Proof. <u>Step 1: Show that $\|f \star g\|_{L^\infty} \leq \|f\|_{L^\infty} \|g\|_{L^1}$.</u>
Observe that

$$\begin{aligned} |(f \star g)(x)| &\overset{\text{T.I}}{\leq} \int_{\mathbb{R}^n} |f(y)||g(x-y)|\, \mathrm{d}y \\ &\leq \|f\|_{L^\infty} \int_{\mathbb{R}^n} |g(x-y)|\, \mathrm{d}y \\ &\overset{(1)}{=} \|f\|_\infty \|g\|_{L^1}, \end{aligned}$$

where the translational invariance of the Lebesgue measure is used in (1). Therefore, $f \star g$ is indeed bounded.
Step 2: For $g \in \mathcal{C}_c^\infty$ show that $f \star g \in \mathcal{C}_b(\mathbb{R}^n)$.
Let $g \in \mathcal{C}_c^\infty$, then for any $x, x' \in \mathbb{R}^n$ we have

$$(f \star g)(x) - (f \star g)(x') = \int_{\mathbb{R}^n} f(y)\left(g(x-y) - g(x'-y)\right) \mathrm{d}y.$$

By the continuity of g we know that $g(x-y) - g(x'-y) \to 0$ as $x \to x'$. Moreover, we know that

$$|f(y)\left(g(x-y) - g(x'-y)\right)| \leq 2\|f\|_{L^\infty}\|g\|_{L^\infty} < \infty,$$

where the finiteness follows from the fact that f is bounded and g is continuous with compact support and so is also bounded. Therefore, using the dominated convergence theorem,

$$|(f \star g)(x) - (f \star g)(x')| \to 0$$

as $x \to x'$.
Step 3: Given $g \in L^1(\mathbb{R}^n)$, find a sequence $(g_n)_{n\in\mathbb{N}} \subseteq \mathcal{C}_c^\infty(\mathbb{R}^n)$ such that $g_n \to g$ in $L^1(\mathbb{R}^n)$.
Recall that $\mathcal{C}_c^\infty(\mathbb{R}^n)$ is dense in $L^1(\mathbb{R}^n)$. So for $g \in L^1(\mathbb{R}^n)$ there exists a sequence $(g_n)_{n\in\mathbb{N}} \subseteq \mathcal{C}_c^\infty(\mathbb{R}^n)$ such that $g_n \to g$ in $L^1(\mathbb{R}^n)$.
Step 4: Argue that $f \star g_n \to f \star g$ in L^∞.
Let $x \in \mathbb{R}^n$ with $(x_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}^n$ such that $x_n \to x$. Note that

$$\begin{aligned}|(f \star g)(x) - (f \star g)(x_n)| \leq& |(f \star g)(x) - (f \star g_k)(x)| + |(f \star g_k)(x) - (f \star g_k)(x_n)| \\ &+ |(f \star g_k)(x_n) - (f \star g)(x_n)|.\end{aligned} \tag{7.2.3}$$

Where

$$\begin{aligned}|(f \star g)(x) - (f \star g_k)(x)| &\leq \|f\|_{L^\infty} \int_{\mathbb{R}^n} |g(x) - g_n(x)| \,\mathrm{d}x \\ &= \|f\|_{L^\infty}\|g - g_n\|_{L^1}.\end{aligned}$$

Similarly,

$$|(f \star g_k)(x_n) - (f \star g)(x_n)| \leq \|f\|_{L^\infty}\|g - g_n\|_{L^1}.$$

Therefore, given an $\epsilon > 0$ there exists an $N_1 \in \mathbb{N}$ such that

$$|(f \star g)(x) - (f \star g_k)(x)| + |(f \star g_k)(x_n) - (f \star g)(x_n)| < \frac{2\epsilon}{3}$$

for all $n \geq N_1$. Using step 3, there exists a $\delta > 0$ such that for $|x - x'| < \delta$ we have

$$|(f \star g_k)(x) - (f \star g_k)(x_n)| < \frac{\epsilon}{3}.$$

As $x_n \to x$ we can choose an $N_2 \in \mathbb{N}$ such that $|x - x_n| < \delta$ for all $n \geq N_2$. Therefore, for all $n \geq \max(N_1, N_2)$ we have

$$|(f \star g)(x) - (f \star g)(x_n)| \leq \frac{2\epsilon}{3} + \frac{\epsilon}{3} = \epsilon,$$

which means that $f \star g$ is continuous. Recall that $f \star g$ is bounded from step 1 and so $f \star g \in \mathcal{C}_b(\mathbb{R}^n)$. ☐

Consequently, we obtain a criterion for the strong Feller property.

> **Corollary 7.2.10.** *Suppose $\mathcal{X} = \mathbb{R}^n$. If there exists some $g \in L^1(\mathbb{R}^n)$ such that for every $f \in \mathcal{B}_b(\mathcal{X})$ we have $Tf = f \star g$, then T is strong Feller.*

Proof. Follows directly from Lemma 7.2.9. ☐

7.3 Existence of Invariant Probability Measures

In the setting of continuous state spaces, the Krylov-Bogoliubov Theorem is an argument for the existence of invariant measures. Consequently, it will be useful to investigate the existence of convergent subsequences.

Definition 7.3.1. *Let A be a topological space.*

1. *A subset $K \subseteq A$ is (sequentially) compact if every sequence $(a_n)_{n\in\mathbb{N}} \subseteq K$ has a subsequence convergent in K.*
2. *A subset $J \subseteq A$ is relatively compact if its closure is compact.*

As we are only interested in the existence of invariant measures, it will be sufficient to consider relative compactness.

Definition 7.3.2. *A subset $\mathcal{M} \subset \mathcal{P}(\mathcal{X})$ is tight if for every $\epsilon > 0$ there exists a compact set $K_\epsilon \subset \mathcal{X}$ such that*

$$\mu\left(K_\epsilon\right) > 1 - \epsilon$$

for all $\mu \in \mathcal{M}$.

Example 7.3.3. *Let $\mathcal{M} = (\delta_n)_{n\in\mathbb{N}}$ for δ_n the delta measure at $x = n \in \mathbb{N}$. Then for every compact set $K \subseteq \mathbb{R}$ there exists an $n \in \mathbb{N}$ such that $n \notin K$. Therefore, $\sup_n(\delta_n(\mathbb{R} \setminus L)) = 1$. Therefore, $\mathcal{M}$ is not tight.*

Lemma 7.3.4. *If $\mathcal{M} \subset \mathcal{P}(\mathcal{X})$ is finite then it is tight.*

Proof. As the finite union of compact sets is compact, it suffices to consider $\mathcal{M} = \{\mu\}$. Since $\mathcal{X}$ is separable it has a countably dense subset, which we enumerate as $(r_k)_{k=1}^\infty \subset \mathcal{X}$. Let $B(x,\delta) = \{y \in \mathcal{X} : d(x,y) < \delta\}$, then

$$\bigcup_{k=1}^{\infty} B\left(r_k, \frac{1}{n}\right) = \mathcal{X}$$

for every $n \geq 1$. Which means that

$$\lim_{N\nearrow\infty} \mu\left(\bigcup_{k=1}^{N} B\left(r_k, \frac{1}{n}\right)\right) = 1.$$

Let $\epsilon > 0$, then for each $m \geq 1$ we can find an N_m such that

$$\mu\left(\bigcup_{k=1}^{N_m} B\left(r_k, \frac{1}{m}\right)\right) > 1 - 2^{-m}\epsilon.$$

If $\mathcal{X}$ is locally compact, then one can take $K = \bar{J}$ with

$$J = \bigcup_{k=1}^{N_1} B(r_k, 1).$$

If $\mathcal{X}$ is not locally compact, then take

$$J = \bigcap_{m=1}^{\infty} \bigcup_{k=1}^{N_m} B\left(r_k, \frac{1}{m}\right).$$

It follows that J is totally bounded and so $\bar{J}$ is compact. We also have

$$\mu\left(J^c\right) < \sum_{m=1}^{\infty} 2^{-m}\epsilon = \epsilon.$$

Therefore, we set $K = \bar{J}$ to deduce that

$$\mu(K) \geq \mu(J) > 1 - \epsilon$$

for all $\mu \in \mathcal{M}$. ∎

It turns out that we can use tightness to show relative compactness.

Theorem 7.3.5 (Prokhorov). *Let $\mathcal{X}$ be a complete separable metric space. Then $\mathcal{M} \subset \mathcal{P}(\mathcal{X})$ is relatively compact if and only if $\mathcal{M}$ is tight.*

In practice, to obtain subsequential limits one often tries to show tightness. Moreover, as $\mathcal{P}(\mathcal{X})$ with weak convergence is metrizable, we can go from convergent subsequences to the convergence of the full sequence if the limits of subsequential limits are unique.

Exercise 7.3.6. *Suppose $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\mathcal{X})$ is tight, and every subsequence has the same limit μ. Show that $\mu_n \to \mu$ weakly.*

Theorem 7.3.7 (Krylov-Bogiulobov). *Let P be a Feller transition probability, and suppose that there exists an x_0 such that sequence of measures $(P^n(x_0, \bullet))_{n=1}^{\infty}$ is tight. Then there exists an invariant probability measure for P.*

Proof. Let us define $\mu_N \in \mathcal{P}(\mathcal{X})$ to be

$$\mu_N(\bullet) = \frac{1}{N}\sum_{n=1}^{N} P^n(x_0, \bullet).$$

Note that by the tightness of $(P^n(x_0, \bullet))_{n=1}^{\infty}$, given an $\epsilon > 0$ there exists a compact set $K_\epsilon \subset \mathbb{R}$ such that

$$\sup_{n\geq 1} \left(P^n\left(x_0, \mathbb{R} \setminus K_\epsilon\right)\right) < \epsilon.$$

Consequently, we have that

$$\begin{aligned}
\sup_{N\geq 1} \left(\mu_N\left(\mathbb{R} \setminus K_\epsilon\right)\right) &= \sup_{N\geq 1} \left(\frac{1}{N}\sum_{n=1}^{N} P^n(x_0, \mathbb{R} \setminus K_\epsilon)\right) \\
&\leq \sup_{N\geq 1} \left(\frac{1}{N}\sum_{n=1}^{N} \epsilon\right) \\
&= \sup_{N\geq 1} (\epsilon) \\
&= \epsilon.
\end{aligned}$$

Hence, the set of measures $(\mu_N)_{N\in\mathbb{N}}$ is also tight. Therefore, there exists a subsequence $(\mu_{N_k})_{k\in\mathbb{N}}$ and $\mu \in \mathcal{P}(\mathcal{X})$ such that $\mu_{N_k} \to \mu$ weakly. Since P is Feller $T_*\phi \in \mathcal{C}_b$, hence,

$$\begin{aligned}
\int_{\mathcal{X}} T_*\phi \,\mathrm{d}\mu &= \lim_{k\to\infty} \int_{\mathcal{X}} T_*\phi \,\mathrm{d}\mu_{N_k} \\
&= \lim_{k\to\infty} \frac{1}{N_k}\sum_{n=1}^{N_k} \int_{\mathcal{X}} \phi(y) P^{n+1}(x_0, \mathrm{d}y) \\
&= \lim_{k\to\infty} \left(\frac{1}{N_k}\sum_{n=1}^{N_k} \int_{\mathcal{X}} \phi(y) P^n(x_0, \mathrm{d}y) + \frac{1}{N_k}\int_{\mathcal{X}} \phi(y) P^{N_k+1}(x_0, \mathrm{d}y) - \frac{1}{N_k}\int_{\mathcal{X}} \phi(y) P(x_0, \mathrm{d}y)\right) \\
&= \lim_{k\to\infty} \left(\int_{\mathcal{X}} \phi \,\mathrm{d}\mu_{N_k} + \frac{1}{N_k}\int_{\mathcal{X}} \phi(y)\left(P^{N_k+1}(x_0, \mathrm{d}y) - P(x_0, \mathrm{d}y)\right)\right) \\
&= \lim_{k\to\infty} \int_{\mathcal{X}} \phi \,\mathrm{d}\mu_{N_k} \\
&\stackrel{(1)}{=} \int_{\mathcal{X}} \phi \, \mu
\end{aligned}$$

where (1) follows by the weak convergence of the μ_{N_k}. Therefore,

$$\int_{\mathcal{X}} \phi \, \mathrm{d}\left(T^* \mu\right) \overset{\text{Ex } 4.2.13}{=} \int_{\mathcal{X}} T_* \phi \, \mathrm{d}\mu$$
$$= \int_{\mathcal{X}} \phi \, \mathrm{d}\mu$$

for all $\phi \in \mathcal{C}_b(\mathcal{X})$. Hence, μ is an invariant measure for P. □

Corollary 7.3.8. *Let $\mathcal{X}$ be a compact state space, it follows that every Feller transition function P has an invariant probability measure.*

Proof. Note that $\mathcal{P}(\mathcal{X})$ is compact as $\mathcal{X}$ is compact. Hence, the tightness of measures follows. Therefore, we can apply Theorem 7.3.7 to conclude. □

Example 7.3.9. *When $\mathcal{X}$ is discrete we note that $\mathcal{B}_b(\mathcal{X}) = \mathcal{C}_b(\mathcal{X})$, which means that every transition probability is strong Feller. Moreover, $\mathcal{X}$ being finite means that it is compact and discrete. So applying Corollary 7.3.8 we deduce that every transition function P has an invariant probability measure.*

Exercise 7.3.10. *Suppose X is a Markov process on $\mathbb{R}^n$ with Feller transition function P. Moreover, suppose that there exists $G : \mathbb{R}_+ \to (0, \infty)$ an increasing function with $\lim_{r\to\infty} G(r) = \infty$ and that*

$$\sup_{n\geq 0} \mathbb{E}_x \left(G(|X_n|)\right) < \infty$$

for some $x \in \mathcal{X}$. Then there exists an invariant probability measure for P.

We can systematize the ideas of Exercise 7.3.10 using Lyapunov functions.

Lemma 7.3.11. *Let P be a transition function on $\mathcal{X}$ and let $V : \mathcal{X} \to [0, \infty]$ be a measurable function. Suppose that there exists a $\gamma \in (0, 1)$ and $C > 0$ such that*

$$(TV)(x) \leq \gamma V(x) + C. \tag{7.3.1}$$

Then

$$T^n V(x) \leq \gamma^n V(x) + \frac{C}{1-\gamma}.$$

Proof. Iterating (7.3.1) gives

$$\begin{aligned} T^n V(x) &= \left(T \circ T^{n-1} V\right)(x) \\ &\leq \gamma T^{n-1} V(x) + c \\ &\leq \gamma^2 T^{n-2} V(x) + C\gamma_C \\ &\vdots \\ &\leq \gamma^n V(x) + C \sum_{j=0}^{n-1} \gamma^j \\ &\leq \gamma^n V(x) + C \sum_{j=0}^{\infty} \gamma^j \\ &= \gamma^n V(x) + \frac{C}{1-\gamma}. \end{aligned}$$

□

Definition 7.3.12. *Let $\mathcal{X}$ be a complete separable metric space and let P be a transition probability on $\mathcal{X}$. A Borel measurable function $V : \mathcal{X} \to [0, \infty]$ is called a Lyapunov function for P if it satisfies the following.*

1. $V^{-1}([0, \infty)) \neq \emptyset$.
2. *For every $a \in [0, \infty)$ the set $K_a = \{y : V(y) \leq a\}$ is compact.*
3. *There exists a $\gamma \in (0, 1)$ and $C > 0$ such that*

$$(TV)(x) \leq \gamma V(x) + C.$$

Theorem 7.3.13. *If a transition function P is Feller and has a Lyapunov function, then it has an invariant measure.*

Proof. Fix $x_0 \in \mathcal{X}$ such that $V(x_0) \neq \infty$. Then for $a > 0$ we have

$$\begin{aligned}
\sup_{n\geq 0} P^n(x_0, \mathbb{R} \setminus K_a) &\leq \sup_{n\geq 0} \int_{\mathcal{X}} \frac{V(y)}{a} P^n(x_0, \mathrm{d}y) \\
&= \sup_{n\geq 0} (T^n V(x_0)) \\
&\overset{\text{Lem 7.3.11.}}{\leq} \frac{1}{a}\left(V(x_0) + \frac{C}{1-\gamma}\right) \\
&\overset{a\to\infty}{\longrightarrow} 0.
\end{aligned}$$

Hence, the family of measures $(P^n(x_0, \bullet))_{n\in\mathbb{N}}$ is tight and so we can conclude by applying Theorem 7.3.7. $\square$

Proposition 7.3.14. *Let P be a transition probability on $\mathcal{X}$, and let $V : \mathcal{X} \to [0, \infty)$ be a measurable function which satisfies*

$$(TV)(x) \leq \gamma V(x) + C$$

for $\gamma \in (0, 1)$ and $C > 0$. Then for every P-invariant probability measure π, we have

$$\int_{\mathcal{X}} V(x)\pi(\mathrm{d}x) \leq \frac{C}{1-\gamma}.$$

Proof. Let $M \geq 0$ and let $V_M(x) = V(x) \wedge M$. Note that $r \mapsto r \wedge M$ is concave, so we can apply Jensen's inequality to get

$$\begin{aligned}
\int_{\mathcal{X}} V_M(y)\, P(x, \mathrm{d}y) &\leq \left(\int_{\mathcal{X}} V(y) P(x, \mathrm{d}y)\right) \wedge M \\
&= (\gamma V(x) + C) \wedge M.
\end{aligned}$$

Iterating this bound it follows that

$$\begin{aligned}
\int_{\mathcal{X}} V_M(y)\, \pi(\mathrm{d}x) &\overset{(1)}{=} \int_{\mathcal{X}} V_M(y) T^n \pi(\mathrm{d}x) \\
&= \int_{\mathcal{X}} (T^n V_m)(y)\, \pi(\mathrm{d}x) \\
&\leq \int \left(\gamma^n V(x) + \frac{C}{1-\gamma}\right) \wedge M\, \pi(\mathrm{d}x)
\end{aligned}$$

for $n \geq 1$. In (1) we use the fact that π is P-invariant. By taking $n \to \infty$ we get

$$\int_{\mathcal{X}} V_M(y)\, \pi(\mathrm{d}x) \leq \frac{C}{1-\gamma} \wedge M \leq \frac{C}{1-\gamma}.$$

Then taking $M \to \infty$ we can apply the monotone convergence theorem to deduce that

$$\int_{\mathcal{X}} V(y)\, \pi(\mathrm{d}x) \leq \frac{C}{1-\gamma}.$$

□

Remark 7.3.15. *Theorem 7.3.13 gives us sufficient conditions a test function with respect to a transition probability P ought to have to ensure the existence of a P-invariant probability measure. Proposition 7.3.14 considers weaker test functions, and hence, cannot guarantee the existence of P-invariant probability measures. We note that the test functions in Proposition 7.3.14 are not necessarily Lyapunov functions as their image is not the extended non-negative real line, and there are no conditions on the compactness of sub-level sets. However, if we assume that such P-invariant probability measures exist then the weaker test functions considered in Proposition 7.3.14 can still provide information on the behaviour of this P-invariant probability measure.*

7.4 Random Dynamical Systems

Recall that a random dynamical system is a Markov processes $X = (X_n)_{n=0}^{\infty}$ where $X_0 = \xi_0$ and

$$X_{n+1} = F(X_n, \xi_n)$$

for $(\xi_j)_{j=0}^{\infty}$ a sequence of independent and identically distributed random variables, with $\nu = \mathrm{Law}(\xi_i)$, on a measurable space $\mathcal{Y}$, and $F : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ a measurable function.

Remark 7.4.1. *It turns out that any discrete-time Markov process can be formulated as a random dynamical system. However, in practice, the functions and noises involved may be complicated.*

7.4.1 Existence of Invariant Measures

To determine the existence of an invariant measure of a random dynamical system we need to investigate the Feller property and the tightness of the n-step transition probabilities. For the tightness, we will utilise Lyapunov functions.

Exercise 7.4.2. *Suppose that $X = (X_n)_{n=0}^{\infty}$ is a random dynamical system. Show that*

$$(Tf)(x) = \int_{\mathcal{Y}} f(F(x,y))\nu(\mathrm{d}y)$$

for any $f \in \mathcal{B}_b(\mathcal{X})$.

Theorem 7.4.3. *Suppose that $X = (X_n)_{n=0}^{\infty}$ is a random dynamical system and that there is a measurable set $A \subset \mathcal{Y}$ with $\nu(A) = 1$ and $x \mapsto F(x,y)$ continuous for every $y \in A$. Then T is Feller.*

Proof. Let $\phi \in \mathcal{C}_b(\mathcal{X})$ and suppose that $(x_n)_{n\in\mathbb{N}} \subset \mathcal{X}$ is convergent to $x \in \mathcal{X}$. Then the continuity of ϕ, and the almost everywhere continuity of $F(\cdot, y)$, we have that

$$\phi(F(x_n, y)) \to \phi(F(x,y))$$

for $y \in \mathcal{Y}$, ν almost everywhere. So by Exercise 7.4.2 it follows that

$$\begin{aligned}
\lim_{n\to\infty} (T\phi)(x_n) &= \lim_{n\to\infty} \int \phi(F(x_n, y))\, \nu(\mathrm{d}y) \\
&\overset{(1)}{=} \int \lim_{n\to\infty} \phi(F(x_n, y))\, \nu(\mathrm{d}y) \\
&= \int \phi(F(x,y))\, \nu(\mathrm{d}y) \\
&= (T\phi)(x),
\end{aligned}$$

where (1) follows as ϕ is bounded. This implies that $(T\phi) \in \mathcal{C}_b(\mathcal{X})$ and T is Feller. $\square$

Theorem 7.4.4. *Suppose that $X = (X_n)_{n=0}^{\infty}$ is a random dynamical system. Suppose that there is a measurable set $A \subset \mathcal{Y}$ with $\nu(A) = 1$ such that $x \mapsto F(x,y)$ continuous for every $y \in A$. Furthermore, suppose that there is some Borel measurable function $V : \mathcal{X} \to [0,\infty]$, with compact sub-level sets, V is finite at some point, and*

$$\int_{\mathcal{Y}} V(F(x,y))\, \nu(\mathrm{d}y) \le \gamma V(x) + C$$

for all $x \in \mathcal{X}$, for some $\gamma \in (0,1)$ and $C > 0$. Then X has at least one invariant probability measure.

Proof. From Theorem 7.4.3 we have that T is Feller, and we note that V is a Lyapunov function. Therefore, we can conclude by applying Theorem 7.3.13. $\square$

We have here established conditions for the existence of invariant measures for random dynamical systems. We will now investigate the uniqueness of these invariant measures. For this, we will use the contraction properties of the function F.

7.4.2 Uniqueness of Invariant Probability Measures via Contraction

Definition 7.4.5. *For $i = 1,2$ let $p_i : \mathcal{X}^2 \to \mathcal{X}$ be the projection maps given by $(x_1,x_2) \mapsto x_i$. For $\pi_1, \pi_2 \in \mathcal{P}(\mathcal{X})$, the measure $\mu \in \mathcal{P}\left(\mathcal{X}^2\right)$ is a coupling of π_1 and π_2 if*

$$(p_i)_*\mu = \pi_i$$

for $i = 1,2$. That is, if $Z = (X,Y) \sim \mu$, then $X \sim \pi_1$ and $Y \sim \pi_2$.

Lemma 7.4.6. *Let $\Delta = \{(x,x) \in \mathcal{X} \times \mathcal{X} : x \in \mathcal{X}\}$. If there exists a coupling $\mu \in \mathcal{P}\left(\mathcal{X}^2\right)$ of $\pi_1, \pi_2 \in \mathcal{P}(\mathcal{X})$ with $\mu(\Delta) = 1$, then $\pi_1 = \pi_2$. In particular, if*

$$\int_{\mathcal{X}\times\mathcal{X}} 1 \wedge d(x,y)\, \mu(\mathrm{d}x,\mathrm{d}y) = 0 \tag{7.4.1}$$

for d the metric on $\mathcal{X}$, then $\pi_1 = \pi_2$.

Proof. Let $A \in \mathcal{B}(\mathcal{X})$, then

$$\begin{aligned}
\pi_1(A) &= \mu(A \times \mathcal{X}) \\
&\overset{(\star)}{=} \mu((A \times \mathcal{X}) \cap \Delta) \\
&\overset{(\star\star)}{=} \mu((\mathcal{X} \times A) \cap \Delta) \\
&\overset{(\star)}{=} \mu(\mathcal{X} \times A) \\
&= \pi_2(A),
\end{aligned}$$

where $(\star)$ follows from the fact that $\mu(\Delta) = 1$, and $(\star\star)$ follows from how Δ is constructed. In particular, note that $1 \wedge d(x,y) \ge 0$ and

$$\{(x,y) \in \mathcal{X} \times \mathcal{X} : 1 \wedge d(x,y) = 0\} = \Delta.$$

Hence, by (7.4.1) it follows that $\mu(\Delta) = 1$. Therefore, by the above arguments, it follows that $\pi_1 = \pi_2$. $\square$

Lemma 7.4.7. *Let $(\mu_n)_{n\in\mathbb{N}}$ be a family of couplings of $\pi_1, \pi_2 \in \mathcal{P}(\mathcal{X})$, then $(\mu_n)_{n\in\mathbb{N}}$ is tight.*

Proof. Given an $\epsilon > 0$, there exists compact sets $K_1, K_2 \subset \mathcal{X}$ such that

$$\pi_i(K_i) > 1 - \frac{\epsilon}{2}$$

for $i = 1, 2$. Hence,

$$\begin{aligned}\mu_n\left(\mathcal{X}^2 \setminus K_1 \times K_2\right) &\leq \mu_n\left((\mathcal{X} \setminus K_1) \times \mathcal{X}\right) + \mu_n\left(\mathcal{X} \times (\mathcal{X} \setminus K_2)\right) \\ &= \pi_1\left(K_1^c\right) + \pi_2\left(K_2^c\right) \\ &< \epsilon,\end{aligned}$$

which shows that $(\mu_n)_{n\in\mathbb{N}}$ is tight. □

Lemma 7.4.8. *Let $(\nu_n)_{n\in\mathbb{N}}$ be a sequence of measures on $\mathcal{X}$ which converges weakly to a measure ν. Then for any continuous map $\phi : \mathcal{X} \to \mathcal{Y}$, it follows that $\phi_*\nu_n$ converges weakly to $\phi_*\nu$.*

Proof. Let $f \in \mathcal{C}_b(\mathcal{Y})$. Then

$$\begin{aligned}\int_{\mathcal{Y}} f\left(\mathrm{d}\phi_*\nu_n\right) &= \int_{\mathcal{X}} (f \circ \phi)\,\mathrm{d}\nu_n \\ &\overset{n\to\infty}{\longrightarrow} \int (f \circ \phi)\,\mathrm{d}\nu \\ &= \int f\,\mathrm{d}(\phi_*\nu)\end{aligned}$$

where the convergence follows as $f \circ \phi \in \mathcal{C}_b(\mathcal{X})$ and ν_n converges weakly to ν. □

Lemma 7.4.9. *Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of couplings of $\pi_1, \pi_2 \in \mathcal{P}(\mathcal{X})$ converging weakly to μ. Then μ is a coupling of π_1 and π_2.*

Proof. The projection maps $p_i : \mathcal{X}^2 \to \mathcal{X}$ for $i = 1, 2$ are continuous functions. So by Lemma 7.4.8 it follows that $(p_i)_*\mu_n$ converges weakly to $(p_i)_*\mu$. As $(p_i)_*\mu_n = \pi_i$ by construction it follows that

$$\int_{\mathcal{X}} f\,\mathrm{d}\pi_i = \int_{\mathcal{X}} f\,\mathrm{d}\left((p_i)_*\mu\right)$$

for all $f \in \mathcal{C}_b(\mathcal{X})$. Hence, by Lemma 7.1.2 it follows that $(p_i)_*\mu = \pi_i$, which implies that μ is a coupling. □

Remark 7.4.10. *Lemma 7.4.7 along with Theorem 7.3.5 shows that the set of all couplings is relatively compact. Lemma 7.4.9 shows that the set of couplings is closed, and hence the set of all couplings for $\pi_1, \pi_2 \in \mathcal{P}(\mathcal{X})$ is compact.*

Definition 7.4.11. *Let X_0 and X_0' be random dynamical systems driven by the independent and identically distributed random variables $(\xi_n)_{n=0}^{\infty}$. Then the synchronized coupling $Z = (X, X')$ is defined to be the Markov process on $\mathcal{X} \times \mathcal{X}$ with*

$$Z_{n+1} = \left(X_{n+1}, X_{n+1}'\right) = \left(F(X_n, \xi_n), F\left(X_n', \xi_n\right)\right).$$

and we write $\mu_n = \mathrm{Law}(Z_n)$.

Lemma 7.4.12. *In the setting of Definition 7.4.11 suppose that for some $\gamma \in (0,1)$ we have*

$$\int_{\mathcal{X}} d\left(F(x,y), F\left(x',y\right)\right)\,\nu(\mathrm{d}y) \leq \gamma d\left(x,x'\right)$$

where d is the metric on $\mathcal{X}$. Then for the synchronized coupling we have

$$\lim_{n\to\infty} \mathbb{E}\left(1 \wedge d\left(X_n, X_n'\right)\right) = \lim_{n\to\infty} \int_{\mathcal{X}} 1 \wedge d\left(x,x'\right)\,\mu_n\left(\mathrm{d}x, \mathrm{d}x'\right) = 0.$$

Proof. Note that we can write

$$\mathbb{E}\left(1 \wedge d\left(X_n, X_n'\right)\right) = \mathbb{E}\left(\mathbb{E}\left(1 \wedge d\left(X_n, X_n'\right)|X_{n-1}, X_{n-1}'\right)\right).$$

Observe that

$$\begin{aligned}
\mathbb{E}\left(1 \wedge d\left(X_n, X_n'\right)|X_{n-1}, X_{n-1}'\right) &\leq 1 \wedge \mathbb{E}\left(d\left(X_n, X_n'\right)|X_{n-1}, X_{n-1}'\right) \\
&= 1 \wedge \mathbb{E}\left(d\left(F\left(X_{n-1}, \xi_{n-1}\right), F\left(X_{n-1}', \xi_{n-1}\right)\right)|X_{n-1}, X_{n-1}'\right) \\
&= 1 \wedge \int_{\mathcal{Y}} d\left(F\left(X_{n-1}, \xi_{n-1}\right), F\left(X_{n-1}', \xi_{n-1}\right)\right)\,\nu(\mathrm{d}y) \\
&\leq 1 \wedge \gamma d\left(X_{n-1}, X_{n-1}'\right).
\end{aligned}$$

Hence,

$$\mathbb{E}\left(1 \wedge d\left(X_n, X_n'\right)\right) \leq \mathbb{E}\left(1 \wedge \gamma d\left(X_n, X_n'\right)\right).$$

and so we can iterate the argument to get

$$\begin{aligned}
\mathbb{E}\left(1 \wedge \gamma d\left(X_{n-1}, X_{n-1}'\right)\right) &= \mathbb{E}\left(\mathbb{E}\left(1 \wedge \gamma d\left(X_{n-1}, X_{n-1}'\right)|X_{n-2}, X_{n-2}'\right)\right) \\
&\leq \mathbb{E}\left(1 \wedge \gamma^2 d\left(X_{n-2}, X_{n-2}'\right)\right) \\
&\;\vdots \\
&\leq \mathbb{E}\left(1 \wedge \gamma^n d\left(X_0, X_0'\right)\right).
\end{aligned}$$

Therefore,

$$\mathbb{E}\left(1 \wedge \gamma d\left(X_{n-1}, X_{n-1}'\right)\right) \leq \mathbb{E}\left(1 \wedge \gamma^n d\left(X_0, X_0'\right)\right) \overset{n\to\infty}{\longrightarrow} 0.$$

□

Theorem 7.4.13. *In the setting of Definition 7.4.11 suppose that for some $\gamma \in (0,1)$ we have*

$$\int_{\mathcal{X}} d\left(F(x,y), F\left(x',y\right)\right)\,\nu(\mathrm{d}y) \leq \gamma d\left(x,x'\right)$$

where d is the metric on $\mathcal{X}$. Then the random dynamical system has at most one invariant probability measure.

Proof. Let π_1 and π_2 be invariant distributions for a random dynamical system. Let $\mathrm{Law}\left(X_0, X_0'\right) = \pi_1 \otimes \pi_2$, such that $\mathrm{Law}(X_n) = \pi_1$ and $\mathrm{Law}\left(X_n'\right) = \pi_2$ by their invariant property. Then $\mu_n = \mathrm{Law}\left(X_n, X_n'\right)$ is a coupling of π_1 and π_2. Moreover, it is a synchronized coupling as X_0 and X_0' are independent. As we know the set of couplings is compact there exists a weakly convergent subsequence $(\mu_{n_k})_{k\in\mathbb{N}}$ to a coupling $\mu \in \mathcal{P}\left(\mathcal{X}^2\right)$. As each μ_n is a synchronized coupling using Lemma 7.4.12 we know that

$$\lim_{n\to\infty} \int_{\mathcal{X}} 1 \wedge d\left(x,x'\right)\,\mu_n\left(\mathrm{d}x, \mathrm{d}x'\right) = 0.$$

In particular, by the boundedness and continuity of $1 \wedge d(\cdot,\cdot)$ it follows that

$$\int_{\mathcal{X}} 1 \wedge d\left(x,x'\right)\,\mu\left(\mathrm{d}x, \mathrm{d}x'\right) = 0.$$

Hence, we can deduce that $\pi_1 = \pi_2$ using Lemma 7.4.6. □

7.5 Uniqueness of Invariant Probability Measures via Minorisation

Uniqueness via minorization is a probabilistic argument for the uniqueness of invariant probability measures. It is useful to also understand uniqueness arguments from this perspective as the random dynamical system formulation of a process may involve complicated functions and noises.

Definition 7.5.1. *Let μ, ν be two positive measures on a measurable space Ω. Suppose that $\mu, \nu \ll \eta$, then define*

$$\|\mu - \nu\|_{\mathrm{TV}} = \int_\Omega \left| \frac{\mathrm{d}\mu}{\mathrm{d}\eta} - \frac{\mathrm{d}\nu}{\mathrm{d}\eta} \right| \mathrm{d}\eta.$$

Remark 7.5.2.

1. *The existence of η in Definition 7.5.1 is not restrictive as $\eta = \mu + \nu$ is such that $\mu, \nu \ll \eta$. Furthermore, Definition 7.5.1 is independent of the choice η. One can see this, as for η such that $\mu, \nu \ll \eta$ then $(\mu + \nu) \ll \eta$, so*

$$\begin{aligned} \int_\Omega \left| \frac{\mathrm{d}\mu}{\mathrm{d}\eta} - \frac{d\nu}{\mathrm{d}\eta} \right| d\eta &= \int_\Omega \left| \frac{\mathrm{d}\mu}{\mathrm{d}(\mu+\nu)} - \frac{\mathrm{d}\nu}{\mathrm{d}(\mu+\nu)} \right| \frac{\mathrm{d}(\mu+\nu)}{\mathrm{d}\eta} \,\mathrm{d}\eta \\ &= \int_\Omega \left| \frac{\mathrm{d}\mu}{\mathrm{d}(\mu+\nu)} - \frac{\mathrm{d}\nu}{\mathrm{d}(\mu+\nu)} \right| \mathrm{d}(\mu+\nu). \end{aligned}$$

2. *Note how Definition 7.5.1 contains the definition of the total variation for the discrete case, given in Lemma 6.6.3, as one can just take η to be the measure that is 1 at each singleton point of the discrete space Ω.*

Exercise 7.5.3. *Show that if μ and ν are probability measures then Definition 7.5.1 gives*

$$\|\mu - \nu\|_{\mathrm{TV}} = 2 \sup \left(\{ |\mu(A) - \nu(A)| : A \subset \Omega, \text{ measurable} \} \right).$$

Lemma 7.5.4. *Let $a, b \geq 0$ then $|a - b| = a + b - 2a \wedge b$.*

With μ, ν and η be as in Definition 7.5.1 and using Lemma 7.5.4 it follows that

$$\left| \frac{\mathrm{d}\mu}{\mathrm{d}\eta} - \frac{\mathrm{d}\nu}{\mathrm{d}\eta} \right| = \frac{\mathrm{d}\mu}{\mathrm{d}\eta} + \frac{\mathrm{d}\nu}{\mathrm{d}\eta} - 2 \left(\frac{\mathrm{d}\mu}{\mathrm{d}\eta} \wedge \frac{\mathrm{d}\nu}{\mathrm{d}\eta} \right). \tag{7.5.1}$$

Definition 7.5.5. *For two positive measures μ and ν let $\mu \wedge \nu$ be the positive measure given by*

$$(\mu \wedge \nu)(A) = \int_A \left(\frac{\mathrm{d}\mu}{\mathrm{d}(\mu+\nu)} \wedge \frac{\mathrm{d}\nu}{\mathrm{d}(\mu+\nu)} \right) \mathrm{d}(\mu+\nu).$$

The measure $\mu \wedge \nu$ constructed in Definition 7.5.5 is the minimum of μ and ν and so can be thought of as their intersection.

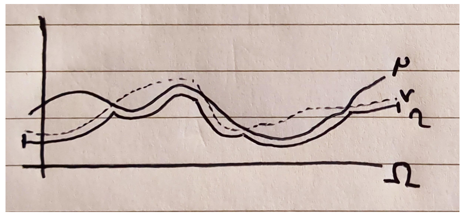


Figure 10: An illustration of the measure $\eta = \mu \wedge \nu$ constructed as in Definition 7.5.5 from measures μ and ν

Lemma 7.5.6. *For $\mu, \nu \in \mathcal{P}(\Omega)$ we have*

$$\|\mu - \nu\|_{\mathrm{TV}} = 2\left(1 - (\mu \wedge \nu)(\Omega)\right).$$

Proof. Using (7.5.1) it follows that

$$\begin{aligned}\|\mu - \nu\|_{\mathrm{TV}} &= \int_\Omega \left|\frac{\mathrm{d}\mu}{\mathrm{d}\eta} - \frac{\mathrm{d}\nu}{\mathrm{d}\eta}\right| \mathrm{d}\eta \\ &= \int_\Omega \frac{\mathrm{d}\mu}{\mathrm{d}\nu} + \frac{\mathrm{d}\nu}{\mathrm{d}\eta} - 2\left(\frac{\mathrm{d}\mu}{\mathrm{d}\eta} \wedge \frac{\mathrm{d}\nu}{\mathrm{d}\eta}\right) \mathrm{d}\eta \\ &= 1 + 1 - 2\left(\mu \wedge \nu\right)(\Omega) \\ &= 2\left(1 - (\mu \wedge \nu)(\Omega)\right).\end{aligned}$$

$\square$

Lemma 7.5.7. *Let $\mu, \nu \in \mathcal{P}(\Omega)$ be distinct and let*

$$\bar{\mu} = \frac{\mu - (\mu \wedge \nu)}{\frac{1}{2}\|\mu - \nu\|_{\mathrm{TV}}}$$

and

$$\bar{\nu} = \frac{\nu - (\mu \wedge \nu)}{\frac{1}{2}\|\mu - \nu\|_{\mathrm{TV}}}.$$

Then $\bar{\mu}, \bar{\nu} \in \mathcal{P}(\Omega)$ with

$$\mu - \nu = \frac{1}{2}\|\mu - \nu\|_{\mathrm{TV}}\left(\bar{\mu} - \bar{\nu}\right). \tag{7.5.2}$$

Proof. The equality (7.5.2) is immediate. Observe that

$$(\mu - (\mu \wedge \nu))(A) = \int_A \frac{\mathrm{d}\mu}{\mathrm{d}(\mu + \nu)} - \left(\frac{\mathrm{d}\mu}{\mathrm{d}(\mu + \nu)} \wedge \frac{\mathrm{d}\nu}{\mathrm{d}(\mu + \nu)}\right) \mathrm{d}(\mu + \nu) \geq 0$$

and so it is clear that $\bar{\mu}$ and $\bar{\nu}$ are positive measures. Moreover, by Lemma 7.5.6 we have

$$\begin{aligned}\frac{1}{2}\|\mu - \nu\|_{\mathrm{TV}} &= 1 - (\mu \wedge \nu)(\Omega) \\ &= \mu(\Omega) - (\mu \wedge \nu)(\Omega) \\ &= \nu(\Omega) - (\mu \wedge \nu)(\Omega)\end{aligned}$$

and so $\bar{\mu}(\Omega) = \bar{\nu}(\Omega) = 1$ which implies that $\bar{\mu}, \bar{\nu} \in \mathcal{P}(\Omega)$. $\square$

Lemma 7.5.8. *Let T be a linear operator on measures defined on Ω such that $T(\mathcal{P}(\Omega)) \subset \mathcal{P}(\Omega)$. Then for* *any $\mu, \nu \in \mathcal{P}(\Omega)$ we have*

$$\|T\mu - T\nu\|_{\mathrm{TV}} = \frac{1}{2}\|\mu - \nu\|_{\mathrm{TV}} \, \|T\bar{\mu} - T\bar{\nu}\|_{\mathrm{TV}} \leq \|\mu - \nu\|_{\mathrm{TV}}. \tag{7.5.3}$$

Proof. Note by the linearity of T that $\overline{T\mu} = T\bar{\mu}$. Thus applying T to both sides of (7.5.2) from Lemma 7.5.7 it follows that

$$\|T\mu - T\nu\|_{\mathrm{TV}} = \frac{1}{2}\|\mu - \nu\|_{\mathrm{TV}} \, \|T\bar{\mu} - T\bar{\nu}\|_{\mathrm{TV}} .$$

Applying the triangle inequality and the fact that $T(\mathcal{P}(\Omega)) \subset \mathcal{P}(\Omega)$ it is clear that $\|T\bar{\mu} - T\bar{\nu}\|_{\mathrm{TV}} \leq 2$. Therefore,

$$\|T\mu - T\nu\|_{\mathrm{TV}} = \frac{1}{2}\|\mu - \nu\|_{\mathrm{TV}} \, \|T\bar{\mu} - T\bar{\nu}\|_{\mathrm{TV}} \leq \|\mu - \nu\|_{\mathrm{TV}}.$$

□

Remark 7.5.9. *The transition operator, T, of a time homogeneous Markov chain, namely*

$$(T\mu)(A) = \int_{\mathcal{X}} P(x, A) \, \mu(\mathrm{d}x)$$

satisfies the requirements of Lemma 7.5.8.

Definition 7.5.10. *A transition probability P is minorized by $\eta \in \mathcal{P}(\mathcal{X})$ with constant $\alpha > 0$ if*

$$P(x, A) \geq \alpha \eta(A)$$

for all $x \in \mathcal{X}$ and $A \in \mathcal{B}(\mathcal{X})$.

Remark 7.5.11. *A transition probability being minorized can be thought of as the continuous analogue of a stochastic matrix being irreducible. Indeed the property of minorization plays an equivalent role in showing the uniqueness of invariant probability measures for continuous state space Markov chains.*

Exercise 7.5.12. *Let $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\Omega)$. Show that*

$$\eta = \sum_{n=1}^{\infty} 2^{-n} \mu_n \in \mathcal{P}(\mathcal{X}).$$

Lemma 7.5.13. *The metric space $(\mathcal{P}(\Omega), \|\cdot\|_{\mathrm{TV}})$ is a complete.*

Proof. Let $(\mu_n)_{n\in\mathbb{N}} \subset \mathcal{P}(\Omega)$ be a Cauchy in $\|\cdot\|_{\mathrm{TV}}$. Using Exercise 7.5.12 we know that $\eta = \sum_{n=1}^{\infty} 2^{-n}\mu_n \in \mathcal{P}(\mathcal{X})$. As $\mu_n \ll \eta$ for all $n \in \mathbb{N}$, it follows that

$$\|\mu_m - \mu_n\|_{\mathrm{TV}} = \left\| \frac{\mathrm{d}\mu_m}{\mathrm{d}\eta} - \frac{\mathrm{d}\mu_n}{\mathrm{d}\eta} \right\|_{L^1(\eta)} .$$

Therefore, $\left(\frac{\mathrm{d}\mu_n}{\mathrm{d}\eta}\right)_{n\in\mathbb{N}}$ is $L^1(\eta)$-Cauchy. Since $L^1(\eta)$ is complete, we know that $\left(\frac{\mathrm{d}\mu_n}{\mathrm{d}\eta}\right)_{n\in\mathbb{N}}$ has an $L^1(\eta)$-limit f with $f \geq 0$ and $\|f\|_{L^1(\eta)} = 1$. Setting $\mu \in \mathcal{P}(\mathcal{X})$ to be $\frac{\mathrm{d}\mu}{\mathrm{d}\eta} = f$ it follows that $\mu_n \to \mu$ in $\|\cdot\|_{\mathrm{TV}}$. □

Theorem 7.5.14. *Let P be a transition probability on $\mathcal{X}$ and suppose P is minorized by $\eta \in \mathcal{P}(\mathcal{X})$ with constant $\alpha \in (0,1)$, then the following hold.*

1. *P has a unique invariant probability measure π.*
2. *For any $\mu, \nu \in \mathcal{P}(\mathcal{X})$ it follows that*

$$\|T^n\mu - T^n\nu\|_{\mathrm{TV}} \leq (1-\alpha)^n \|\mu - \nu\|_{\mathrm{TV}}.$$

Proof. Recall that $(\mathcal{P}(\Omega), \|\cdot\|_{\mathrm{TV}})$ is a complete metric space, from Lemma 7.5.13. Hence, as statement 2. shows that T is a strict contraction we can show statement 1. by using Banach's fixed point theorem. To show statement 2. we note that for any measure $\lambda \in \mathcal{P}(\mathcal{X})$ and $A \in \mathcal{B}(\mathcal{X})$ we have

$$\begin{aligned}(T\lambda)(A) &= \int_{\mathcal{X}} P(x,A)\,\lambda(\mathrm{d}x) \\ &\geq \int_{\mathcal{X}} \alpha\eta(A)\,\lambda(\mathrm{d}x) \\ &= \alpha\eta(A).\end{aligned}$$

Hence, for any $\lambda \in \mathcal{P}(\mathcal{X})$ it follows that

$$\frac{1}{1-\alpha}(T\lambda - \alpha\eta) \in \mathcal{P}(\mathcal{X}).$$

Thus for any $\mu, \nu \in \mathcal{P}(\mathcal{X})$ we have

$$\begin{aligned}\|T\mu - T\nu\|_{\mathrm{TV}} &= (1-\alpha)\left\|\frac{T\mu - \alpha\eta}{1-\alpha} - \frac{T\nu - \alpha\eta}{1-\alpha}\right\|_{\mathrm{TV}} \\ &\leq (1-\alpha)(2).\end{aligned}$$

Using the equality in (7.5.3) of Lemma 7.5.8 it follows that

$$\begin{aligned}\|T\mu - T\nu\|_{\mathrm{TV}} &= \frac{1}{2}\|\mu - \nu\|_{\mathrm{TV}}\,\|T\bar{\mu} - T\bar{\nu}\|_{\mathrm{TV}} \\ &\leq \frac{1}{2}\|\mu - \nu\|_{\mathrm{TV}}\,(2(1-\alpha)) \\ &= (1-\alpha)\|\mu - \nu\|_{\mathrm{TV}}.\end{aligned}$$

Iterating this we arrive at statement 2. ☐

Corollary 7.5.15. *Under the assumptions of Theorem 7.5.14, let $\pi, \mu \in \mathcal{P}(\mathcal{X})$ with $T\pi = \pi$. Then*

$$\|T^n\mu - \pi\|_{\mathrm{TV}} \leq (1-\alpha)^n\|\mu - \pi\|_{\mathrm{TV}}.$$

7.6 P-Invariant Sets

Now we introduce a continuous analogue of closed communication classes.

Definition 7.6.1. *Let P be a transition function on $\mathcal{X}$, then $A \in \mathcal{B}(\mathcal{X})$ is P-invariant if $P(x,A) = 1$ for every $x \in A$.*

Exercise 7.6.2. *If $A \in \mathcal{B}(\mathcal{X})$ is P-invariant, and X is a Markov process with transition function P and initial distribution $\pi \in \mathcal{P}(\mathcal{X})$, then*

$$\mathbb{P}(X_0 \in A, \ldots, X_n \in A) = \pi(A).$$

If $A \subset \mathcal{X}$ is closed, and $\mathcal{X}$ is a complete separable metric space, then A is a complete separable metric space. Thus we can think of A as being a possible state space. Note that we can extend any $\tilde{\pi} \in \mathcal{P}(A)$ to an element $\pi \in \mathcal{P}(\mathcal{X})$ by letting

$$\pi(F) = \tilde{\pi}(F \cap A)$$

for all $F \in \mathcal{B}(\mathcal{X})$. Similarly, given $\pi \in \mathcal{P}(\mathcal{X})$ we can restrict π to a measure, $\tilde{\pi}$, on A. Namely, $\tilde{\pi} := \pi|_{\mathcal{B}(A)}$. However, it may not be the case that $\tilde{\pi}$ is a probability measure. If A is P-invariant then we can restrict a transition function P on $\mathcal{X}$ to a transition function $\tilde{P}$ on A. For any $F \in \mathcal{B}(\mathcal{X})$ one can let

$$\tilde{P}(x, B) = P(x, B \cap A) = P(x, B)$$

for any $x \in A$. For $\tilde{P}$ we have the corresponding operator $\tilde{T}$.

Lemma 7.6.3. *Let P be a transition function on $\mathcal{X}$ and let A be closed and P-invariant.*

1. *If $\tilde{\pi} \in \mathcal{P}(A)$ is $\tilde{P}$-invariant, then the extension of $\tilde{\pi}$ to $\mathcal{X}$, π, is P-invariant.*
2. *If $\pi \in \mathcal{P}(\mathcal{X})$ is P-invariant, then its restriction, $\tilde{\pi}$, to A is $\tilde{P}$-invariant.*

Proof.

1. As $\tilde{\pi}$ is $\tilde{P}$-invariant we have

$$\tilde{\pi}(C) = \int_A \tilde{P}(x, C)\, \tilde{\pi}(\mathrm{d}x)$$

for any $C \in \mathcal{B}(A) = A \cap \mathcal{B}(\mathcal{X})$. For $B \in \mathcal{B}(\mathcal{X})$ we have

$$\begin{aligned}
(T\pi)(B) &= \int_{\mathcal{X}} P(x, B)\, \pi(\mathrm{d}x) \\
&\overset{(1)}{=} \int_A P(x, B)\, \pi(\mathrm{d}x) \\
&= \int_A P(x, B \cap A)\, \pi(\mathrm{d}x) \\
&\overset{(2)}{=} \int_A \tilde{P}(x, B \cap A)\, \tilde{\pi}(\mathrm{d}x) \\
&= \tilde{\pi}(B \cap A) \\
&= \pi(B),
\end{aligned}$$

where in (1) we use that $\mathrm{supp}(\pi) \subset A$, and in (2) we use the fact that $\tilde{P}$ and $\tilde{\pi}$ agree on $A \cap \mathcal{B}(\mathcal{X})$.

2. Let $B \in \mathcal{B}(A) = A \cap \mathcal{B}(\mathcal{X})$, so that in particular we have $B \subset A$. Then

$$\begin{aligned}
\pi(B) &= \int_{\mathcal{X}} P(x, B)\, \pi(\mathrm{d}x) \\
&= \int_A P(x, B)\, \pi(\mathrm{d}x) + \int_{\mathcal{X} \setminus A} P(x, B)\, \pi(\mathrm{d}x).
\end{aligned}$$

By the invariance of A we have

$$\begin{aligned}
\pi(A) &= \int_A P(x, A)\, \pi(\mathrm{d}x) + \int_{\mathcal{X} \setminus A} P(x, A)\, \pi(\mathrm{d}x) \\
&= \pi(A) + \int_{\mathcal{X} \setminus A} P(x, A)\, \pi(\mathrm{d}x)
\end{aligned}$$

Therefore,

$$\int_{\mathcal{X} \setminus A} P(x, A)\, \pi(\mathrm{d}x) = 0$$

which implies that

$$\int_{\mathcal{X}\setminus A} P(x,B)\,\pi(\mathrm{d}x) = 0.$$

Hence,

$$\pi(B) = \int_A P(x,B)\,\pi(\mathrm{d}x).$$

As $\tilde{\pi}$ and $\tilde{P}$ coincide with π and P respectively on A, it follows that $\tilde{\pi}$ is $\tilde{P}$-invariant.

□

Theorem 7.6.4. *Let P be Feller on $\mathcal{X}$ and A be a compact P-invariant set, then the restriction of P to A is Feller. In particular, there exists a P-invariant $\pi \in \mathcal{P}(\mathcal{X})$.*

Proof. Let $\tilde{f} \in \mathcal{C}_b(A)$. Then by the Tietze Extension theorem, there exists $f \in \mathcal{C}_b(\mathcal{X})$ such that $f|_A = \tilde{f}$. Then for $x \in A$ we have

$$\begin{aligned}
\left(\tilde{T}\tilde{f}\right)(x) &= \int_A \tilde{f}(y)\tilde{P}(x,\mathrm{d}y) \\
&= \int_A f(y)P(x,\mathrm{d}y) \\
&\overset{(1)}{=} \int_{\mathcal{X}} f(y)P(x,\mathrm{d}y) \\
&= (Tf)(x),
\end{aligned}$$

where in (1) we have used the fact that A is P-invariant and $x \in A$. As P is Feller it follows that $Tf \in \mathcal{C}_b(\mathcal{X})$ as $f \in \mathcal{C}_b(\mathcal{X})$. Therefore, as $(Tf)|_A = \tilde{T}\tilde{f}$ we deduce that $\tilde{T}\tilde{f} \in \mathcal{C}_b(A)$ which shows that $\tilde{P}$ is Feller on A. Hence, by Corollary 7.3.8 there exists a $\tilde{P}$-invariant probability measure $\tilde{\pi} \in \mathcal{P}(A)$. Using Lemma statement 1. of 7.6.3 we can conclude that π, the extension of $\tilde{\pi}$ to $\mathcal{X}$, is P-invariant. □

We have seen with Theorem 7.6.4 that P-invariance can be used to deduce the existence of invariant measures. On the other hand, we would like to see how P-invariance can be used to make statements regarding all invariant measures. Thus working toward characterising the uniqueness of invariance measures. Consequently, for a P-invariant set A we introduce the sequence of sets $(A_n)_{n\in\mathbb{N}}$ where $A_0 = A$ and

$$A_n = \{x \in \mathcal{X} : P(x, A_{n-1}) > 0\} \tag{7.6.1}$$

for $n \geq 1$. It can be argued inductively that the sequence of sets $(A_n)_{n\in\mathbb{N}}$ is nested. More specifically, if we assume $A_{n-1} \subset A_n$ then for $x \in A_n$ we have that,

$$P(x, A_n) \overset{A_{n-1}\subset A_n}{\geq} P(x, A_{n-1}) \overset{x\in A_n}{>} 0,$$

which implies that $x \in A_{n+1}$.

Lemma 7.6.5. *Let $A \subset \mathcal{X}$ be P-invariant and let $(A_n)_{n\in\mathbb{N}}$ be as in* (7.6.1)*. Then for every $n \geq 1$ and $x \in A_n$ we have $P^n(x,A) > 0$.*

Proof. We proceed by induction on n. The case $n = 1$ follows by construction of the sequence A_1. Suppose the statement holds for $n-1$, then

$$\begin{aligned}
P^n(x,A) &\overset{(1)}{=} \int_{\mathcal{X}} P^{n-1}(y,A)P(x,\mathrm{d}y) \\
&\geq \int_{A_{n-1}} P^{n-1}(y,A)P(x,\mathrm{d}y) \\
&\overset{(2)}{>} 0,
\end{aligned}$$

where (1) is an application of the Chapman-Kolmogorov equation. Moreover, (2) follows by the inductive assumption as $y \in A_{n-1}$ implies that $P^{n-1}(y,A) > 0$, and we know that $P(x, A_{n-1}) > 0$ as $x \in A_n$. □

Remark 7.6.6. *Where the original construction of A_n was the set of points which with positive probability reach A_{n-1} in one step. What Lemma 7.6.5 provides is an alternative characterisation of the set A_n, which says that it is the set of points which with positive probability reach the set A.*

Proposition 7.6.7. *Let $A \subset \mathcal{X}$ be P-invariant and consider the sequence $(A_n)_{n\in\mathbb{N}}$ as given by (7.6.1). Suppose $\mathcal{X} = \bigcup_{n\in\mathbb{N}} A_n$, then for every P-invariant $\pi \in \mathcal{P}(\mathcal{X})$ we have $\pi(A) = 1$.*

Proof. Suppose that π is P-invariant with $\pi(A) < 1$. Then since $\pi(A_n) \nearrow 1$, there exists an $N \in \mathbb{N}$ such that $\pi(A_N \setminus A) > 0$. Therefore,

$$\begin{aligned}\pi(A) &= \int_{\mathcal{X}} P^N(x, A)\,\pi(\mathrm{d}x) \\ &\geq \int_{A_N} P^N(x, A)\,\pi(\mathrm{d}x) \\ &\geq \underbrace{\int_A P^N(x, A)\,\pi(\mathrm{d}x)}_{\pi(A)} + \underbrace{\int_{A_N\setminus A} P^N(x, A)\,\pi(\mathrm{d}x)}_{\overset{(1)}{>}0} \\ &> \pi(A),\end{aligned}$$

where (1) follows as $x \in A_N$ and so $P^N(x, A) > 0$ by Lemma 7.6.5, and we know that $\pi(A_N \setminus A) > 0$ by the choice of N. Therefore, we arrive at a contradiction and so $\pi(A) = 1$. ☐

Corollary 7.6.8. *Consider a random dynamical system*

$$X_{n+1} = F(X_n, \xi_n).$$

Suppose A is a P-invariant set, where P is a transition probability on $\mathcal{X}$ with the Feller property. Moreover, suppose that A is compact with $\bigcup_{n\in\mathbb{N}} A_n = \mathcal{X}$, and there exists a $\gamma \in (0, 1)$ such that

$$\mathbb{E}\left(d(F(x,\xi), F(y,\xi))\right) \leq \gamma d(x, y) \tag{7.6.2}$$

for all $x, y \in A$. Then there exists a unique P-invariant measure on $\mathcal{X}$.

Proof. The existence of an invariant measure follows from Theorem 7.6.4. By Proposition 7.6.7 any P-invariant probability measure $\pi \in \mathcal{P}(\mathcal{X})$ is an extension of $\tilde{\pi} \in \mathcal{P}(A)$ which is $\tilde{P}$-invariant by statement 2. of Lemma 7.6.3. Therefore, if two P-invariant measures on $\mathcal{X}$ existed, we would have two $\tilde{P}$-invariant measures on A. However, using (7.6.2) we know by Theorem 7.4.13 that $\tilde{P}$-invariant probability measure on A is unique. Therefore, P-invariant probability measures on $\mathcal{X}$ are unique. ☐

Example 7.6.9. *Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distribution $\mathcal{C}([0,1])$-valued random variables such that $\sup_{t\in[0,1]} |\xi_n(t)| \leq 1$ almost surely. Let $\varphi_t(x, f)$ be the solution to the differential equation*

$$\begin{cases}\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{1}{x(t)} - 2 + f(t) & t \in (0, \infty) \\ x(0) = x.\end{cases}$$

With $F(x, f) = \varphi_1(x, f)$ consider the Markov process $(X_n)_{n\in\mathbb{N}}$ where $X_0 \in (0, \infty)$ and

$$X_n = F(X_{n-1}, \xi_{n-1})$$

for $n \geq 1$. One can show that $\varphi_t(x, f)$ is indeed well-defined, that is, it exists and is unique. Recall that $\xi_n \in [-1, 1]$ almost surely, hence, the deterministic maps $F_-(x) = F(x, -1)$ and $F_+(x) = F(x, 1)$ are such *that*

$$X_{n+1} \in [F_-(X_n), F_+(X_n)] \tag{7.6.3}$$

almost surely. Observe the following.

1. *Note that $F_+(x)$ is the solution to $\frac{\mathrm{d}z}{\mathrm{d}t} = \frac{1}{z(t)} - 1$ with $z(0) = x$. For which $z(t) \equiv 1$ is an equilibrium solution, and*
$$\lim_{t\to\infty} \varphi_t(x,1) = \lim_{t\to\infty} z(t) = 1.$$
In particular, we note that $\varphi_n(x,1) = \underbrace{(F_+ \circ \cdots \circ F_+)}_{n}(x) =: F_+^n(x)$ and so $\lim_{n\to\infty} F_+^n(x) = 1$.

2. *Note that $F_-(x)$ is the solution to $\frac{\mathrm{d}z}{\mathrm{d}t} = \frac{1}{z(t)} - 3$ with $z(0) = x$. For which $z(t) \equiv \frac{1}{3}$ is an equilibrium solution, and*
$$\lim_{t\to\infty} \varphi_t(x,-1) = \lim_{t\to\infty} z(t) = \frac{1}{3}.$$
In particular, we note that $\varphi_n(x,-1) = \underbrace{(F_- \circ \cdots \circ F_-)}_{n}(x) =: F_-^n(x)$ and so $\lim_{n\to\infty} F_-^n(x) = 1$.

Now for a fixed $\epsilon > 0$ let $A = \left[\frac{1}{3} - \epsilon, 1 + \epsilon\right]$.

1. *On $(1,\infty)$ we note that $\frac{1}{z(t)} - 1 < 0$ and so it must be the case that $(F_+)^{-1}(1+\epsilon) > 1 + \epsilon$.*

2. *On $\left(0, \frac{1}{3}\right)$ we have that $\frac{1}{z(t)} - 3 > 0$ and so it must be the case that $(F_-)^{-1}(1-\epsilon) < 1 - \epsilon$.*

Consequently,

$$A \subset \left[(F_-)^{-1}\left(\frac{1}{3} - \epsilon\right), (F_+)^{-1}(1+\epsilon)\right].$$

In particular, using (7.6.3) this means that if $X_0 \in \left[(F_-)^{-1}\left(\frac{1}{3} - \epsilon\right), (F_+)^{-1}(1+\epsilon)\right]$ then $X_1 \in \left[\frac{1}{3} - \epsilon, 1 + \epsilon\right] = A$. Therefore, A is a P-invariant set. Recalling that

$$A_{n+1} := \{x \in \mathcal{X} : P(x, A_n) > 0\}$$

we can use similar arguments to show that

$$\left[(F_-^n)^{-1}\left(\frac{1}{3} - \epsilon\right), (F_+^n)^{-1}(1+\epsilon)\right] \subset A_n.$$

Moreover, using the observation regarding convergence to equilibrium solutions made previously, we note that

$$\bigcup_{n\in\mathbb{N}} A_n = (0,\infty).$$

Therefore, using Proposition 7.6.7 we deduce that for any invariant probability measure π we have that $\pi\left(\left[\frac{1}{3} - \epsilon, 1 + \epsilon\right]\right) = 1$. As $\epsilon > 0$ was arbitrary we conclude that $\pi\left(\left[\frac{1}{3}, 1\right]\right) = 1$. Note that since $F(x,\xi_n)$ is differentiable in x, it is continuous and so $(X_n)_{n\in\mathbb{N}}$ is Feller by Theorem 7.4.3. In particular, let $v(t) = \frac{\mathrm{d}\varphi_t(x,f)}{\mathrm{d}x}$ such that

$$\frac{\mathrm{d}v(t)}{\mathrm{d}t} = -\frac{1}{x^2(t)} v(t)$$

with $v(0) = 1$. Consequently, we have that

$$v(t) = \exp\left(-\int_0^t \frac{1}{x^2(s)}\,\mathrm{d}s\right).$$

Therefore,

$$F'(X_n, \xi_n) = v(t) = \exp\left(-\int_0^t \frac{1}{x^2(s)}\,\mathrm{d}s\right) \leq \exp\left(\frac{1}{1+\epsilon}\right) < 1.$$

So,

$$\mathbb{E}\left(|F(x,\xi) - F(y,\xi)|\right) < |x-y|.$$

Thus we have P-invariance on a compact set A such that $\bigcup_{n\in\mathbb{N}} A_n = \mathcal{X}$, allowing us to applying Corollary 7.6.8 to conclude that the Markov chain $(X_n)_{n\in\mathbb{N}}$ on $(0,\infty)$ has a unique invariant probability measure.

7.7 Solution to Exercises

Exercise 7.2.7

Solution. Let $A = \operatorname{supp}(\mu)$. Then by Proposition 7.2.6 we know that $\mu(A) = 1$. As $\operatorname{supp}(\nu) \subset \mathcal{X} \setminus A$ and $\nu(\operatorname{supp}(\nu)) = 1$ it follows that $\nu(\mathcal{X} \setminus A) = 1$ which implies $\nu(A) = 0$. However, let $\mathcal{X} = [0,1]$, $\mu = \delta_0$ and ν the Lebesgue measure. It is clear that $\mu \perp \nu$ as one can take $A = \{0\}$, however, $\operatorname{supp}(\mu) \cap \operatorname{supp}(\nu) = \{0\}$. ☐

Exercise 7.3.6

Solution. Suppose that $\mu \not\to \mu$ weakly. Then as $\mathcal{P}(\mathcal{X})$ with the weak topology is metrizable, say with the metric $d(\cdot,\cdot)$, there exists an $\epsilon > 0$ and a subsequence $(n_k)_{k\in\mathbb{N}}$ such that $d\left(\mu, \mu_{n_k}\right) \geq \epsilon$ for all $k \in \mathbb{N}$. However, as $(\mu_n)_{n=0}^{\infty}$ is tight it is relatively compact. Therefore, the sequence $(\mu_{n_k})_{k=0}^{\infty} \subseteq (\mu_n)_{n=0}^{\infty}$ must contain a convergent subsequence, which by assumption must have a limit μ. This contradicts the fact that $d\left(\mu, \mu_{n_k}\right) \geq \epsilon$ for all $k \in \mathbb{N}$. ☐

Exercise 7.3.10

Solution. For any $M > 0$ we have that

$$\begin{aligned}
\sup_{n\geq 0}\left(P^n\left(x, B(0,M)^c\right)\right) &= \sup_{n\geq 0}\left(\mathbb{P}_x\left(|X_n| \geq M\right)\right) \\
&\overset{\text{Chebyshev.}}{\leq} \frac{1}{G(M)} \sup_{n\geq 0}\left(\mathbb{E}_x\left(G(|X_n|)\right)\right) \\
&\overset{M\to\infty}{\longrightarrow} 0.
\end{aligned}$$

Hence, we deduce that $\{P^n(x,\bullet)\}_{n=1}^{\infty}$ is tight and so we can apply Theorem 7.3.7 to conclude. ☐

Exercise 7.4.2

Solution. Let $f \in \mathcal{B}_b(\mathcal{X})$, then

$$\begin{aligned}
(Tf)(x) &= \mathbb{E}(f(X_1)|X_0 = x) \\
&= \mathbb{E}(f(F(X_0,\xi_0))|X_0 = x) \\
&\overset{(1)}{=} \int_{\mathcal{Y}} f(F(x,y))\,\nu(\mathrm{d}y)
\end{aligned}$$

where in (1) we use the independence of the random variables $(\xi_i)_{i=0}^{\infty}$. ☐

Exercise 7.5.3

Solution. On the one hand, for any measurable set $A \subset \Omega$ we have

$$\begin{aligned}
\|\mu-\nu\|_{\mathrm{TV}} &= \int_\Omega \left|\frac{\mathrm{d}\mu}{\mathrm{d}\eta} - \frac{\mathrm{d}\nu}{\mathrm{d}\eta}\right| \mathrm{d}\eta \\
&= \int_A \left|\frac{\mathrm{d}\mu}{\mathrm{d}\eta} - \frac{\mathrm{d}\nu}{\mathrm{d}\eta}\right| \mathrm{d}\eta + \int_{\Omega\setminus A} \left|\frac{\mathrm{d}\mu}{\mathrm{d}\eta} - \frac{\mathrm{d}\nu}{\mathrm{d}\eta}\right| \mathrm{d}\eta \\
&\geq \left|\int_A \frac{\mathrm{d}\mu}{\mathrm{d}\eta} - \frac{\mathrm{d}\nu}{\mathrm{d}\eta} \mathrm{d}\eta - \int_{\Omega\setminus A} \frac{\mathrm{d}\mu}{\mathrm{d}\eta} - \frac{\mathrm{d}\nu}{\mathrm{d}\eta} \mathrm{d}\eta\right| \\
&= |\mu(A) - \nu(A) - \mu(\Omega\setminus A) + \nu(\Omega\setminus A)| \\
&= 2|\mu(A)-\nu(A)|.
\end{aligned}$$

Therefore, $\|\mu-\nu\|_{\mathrm{TV}} \geq 2\sup\left(\{|\mu(A)-\nu(A)| : A \subset \Omega \,\text{measurable}\}\right)$. On the other hand, let $A = \left\{\frac{\mathrm{d}\mu}{\mathrm{d}\eta} \geq \frac{\mathrm{d}\nu}{\mathrm{d}\eta}\right\}$. Then as

$$\int_A \frac{\mathrm{d}\mu}{\mathrm{d}\eta} \mathrm{d}\eta = \mu(A) = 1 - \mu(\Omega\setminus A) = 1 - \int_{\Omega\setminus A} \frac{\mathrm{d}\mu}{\mathrm{d}\eta} \mathrm{d}\eta,$$

it follows that

$$\int_A \frac{\mathrm{d}\mu}{\mathrm{d}\eta} - \frac{\mathrm{d}\nu}{\mathrm{d}\eta} \mathrm{d}\eta = \int_{\Omega\setminus A} \frac{\mathrm{d}\nu}{\mathrm{d}\eta} - \frac{\mathrm{d}\mu}{\mathrm{d}\eta} \mathrm{d}\eta = \frac{1}{2}\|\mu-\nu\|_{\mathrm{TV}}.$$

Therefore, as

$$0 \leq \int_A \frac{\mathrm{d}\mu}{\mathrm{d}\eta} - \frac{\mathrm{d}\nu}{\mathrm{d}\eta} \mathrm{d}\eta = \mu(A) - \nu(A) = |\mu(A)-\nu(A)|$$

it follows that $\|\mu-\nu\|_{\mathrm{TV}} \leq 2\sup\left(\{|\mu(A)-\nu(A)| : A \subset \Omega \,\text{measurable}\}\right)$, which completes the proof. □

Exercise 7.5.12

Solution. Clearly, $\eta \geq 0$, and

$$\eta(\Omega) = \sum_{n=1}^\infty 2^{-n}\mu_n(\Omega) = \sum_{n=1}^\infty 2^{-n} = 1.$$

Moreover, it is clear that $\eta(A) \leq \eta(B)$ for $A \subset B \subset \Omega$. If $A_1, A_2, \dots \subset \Omega$ are disjoint sets then

$$\eta\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{n=1}^\infty 2^{-n}\mu_n\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{n=1}^\infty\sum_{i=1}^\infty 2^{-n}\mu_n(A_i).$$

Noting that the sum is absolutely convergent, as the terms are non-negative and its bounded above by 1, we can exchange the order of summation to deduce that

$$\eta\left(\bigcup_{i=1}^\infty A_i\right) = \sum_{i=1}^\infty\sum_{n=1}^\infty 2^{-n}\mu_n(A_i) = \sum_{i=1}^\infty \eta(A_i).$$

Therefore, η is countably additive and defines a probability measure. □

Exercise 7.6.2

Solution. Proceed by induction on n. When $n = 1$ we have

$$\begin{aligned}
\pi(A) &= \mathbb{P}(X_0 \in A, X_1 \in \mathcal{X}) \\
&= \mathbb{P}(X_0 \in A, X_1 \in A) + \mathbb{P}(X_0 \in A, X_1 \in \mathcal{X}\setminus A) \\
&\overset{(1)}{=} \mathbb{P}(X_0 \in A, X_1 \in A) + 0 \\
&= \mathbb{P}(X_0 \in A, X_1 \in A),
\end{aligned}$$

where (1) follows by the invariance of A. Assuming that

$$\mathbb{P}(X_0 \in A, \ldots, X_n \in A) = \pi(A)$$

it follows that

$$\begin{aligned}
\pi(A) &= \mathbb{P}(X_0 \in A, \ldots, X_n \in A, X_{n+1} \in \mathcal{X}) \\
&= \mathbb{P}(X_0 \in A, \ldots, X_n \in A, X_{n+1} \in A) + \mathbb{P}(X_0 \in A, \ldots, X_n \in A, X_{n+1} \in \mathcal{X} \setminus A) \\
&\overset{(1)}{=} \mathbb{P}(X_0 \in A, \ldots, X_n \in A, X_{n+1} \in A) + 0 \\
&= \mathbb{P}(X_0 \in A, \ldots, X_n \in A, X_{n+1} \in A),
\end{aligned}$$

where (1) follows by the invariance of A. This completes the proof by induction. $\square$

8 Ergodic Theory

8.1 Birkhoff's Ergodic Theorem for Dynamical Systems

Definition 8.1.1. *A dynamical system consists of a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and a measure-preserving map* $\theta : \Omega \to \Omega$. *That is,* $\mathbb{P}\left(\theta^{-1}(A)\right) = \mathbb{P}(A)$ *for every* $A \in \mathcal{F}$.

Definition 8.1.2. *Given a measurable transformation* θ *on* $(\Omega, \mathcal{F}, \mathbb{P})$ *a set* A *such that* $\theta^{-1}(A) = A$ *is said to be* θ*-invariant. In particular, the* θ*-invariant* σ*-algebra* $\mathcal{I} \subset \mathcal{F}$ *is given by*

$$\mathcal{I} = \left\{A \in \mathcal{F} : \theta^{-1}(A) = A\right\}. \tag{8.1.1}$$

Exercise 8.1.3. *Show that* $\mathcal{I}$*, as given by* (8.1.1)*, is indeed a* σ*-algebra.*

Definition 8.1.4. *A measurable function* $f : \Omega \to \mathbb{R}$ *is* θ*-invariant if* $f \circ \theta = f$.

Exercise 8.1.5. *Let* $f : \Omega \to \mathbb{R}$ *be a* $\mathcal{F}$*-measurable function. Then* f *is invariant if and only if* f *is measurable with respect to* $\mathcal{I}$*, as given by* (8.1.1).

Similarly, a $\mathcal{F}$-measurable function $f : \Omega \to \mathbb{R}$ is invariant if and only if f is measurable with respect to $\mathcal{I}$ as defined above.

Definition 8.1.6. *Consider a dynamical system* $(\Omega, \mathcal{F}, \mathbb{P})$ *with* θ *a measure-preserving map. Then* θ *is ergodic if for any* $A \in \Omega$ *a* θ*-invariant set, we have* $\mathbb{P}(A) \in \{0, 1\}$*. Ergodicity is also a property of* $\mathbb{P}$*, and so* $\mathbb{P}$ *is said to be ergodic with respect to* θ.

Proposition 8.1.7. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *with* θ *be a dynamical system. Then the following are equivalent.*

1. $\mathbb{P}$ *is ergodic with respect to* θ.
2. *Every* θ*-invariant integrable function* $f : \Omega \to \mathbb{R}$ *is almost surely constant.*
3. *Every* θ*-invariant bounded function* $f : \Omega \to \mathbb{R}$ *is almost surely constant.*

Proof. $(2) \Rightarrow (3)$. This is clear as every bounded function is integrable with respect to a probability measure.
$(3) \Rightarrow (1)$. Let $f = \mathbf{1}_A$, where A is an invariant set. Then f is invariant and bounded and so is almost surely constant, that is $f \in \{0, 1\}$. Which implies that $\mathbb{P}(A) = f \in \{0, 1\}$ meaning $\mathbb{P}$ is ergodic.
$(1) \Rightarrow (2)$. Let f be an integrable and invariant function. Then by Exercise 8.1.5 it follows that f is $\mathcal{I}$-measurable. Consider the sets

- $A_+ := \{\omega \in \Omega : f(\omega) > \mathbb{E}(f)\}$,
- $A_- := \{\omega \in \Omega : f(\omega) < \mathbb{E}(f)\}$, and
- $A_0 := \{\omega \in \Omega : f(\omega) = \mathbb{E}(f)\}$.

These sets form a disjoint partition of Ω, and so by the ergodic property of $\mathbb{P}$ it follows that exactly one has full measure whilst the others have zero-measure. If $\mathbb{P}(A_+) = 1$ then

$$0 = \int_\Omega f - \mathbb{E}(f) \, \mathrm{d}\mathbb{P} = \int_{A_+} f - \mathbb{E}(f) \, \mathrm{d}\mathbb{P},$$

which implies that $f - \mathbb{E}(f) = 0$ almost surely on A_+ which is a contradiction. Similarly, we have that $\mathbb{P}(A_-) \neq 1$. Therefore, $\mathbb{P}(A_0) = 1$ which implies that f is almost surely constant. $\square$

Theorem 8.1.8 (Maximal Ergodic Theorem). *Let* $(\Omega, \mathcal{I}, \mathbb{P})$ *with* θ *be a dynamical system, where* $\mathcal{I}$ *is as given by* (8.1.1)*. Let* $f : \Omega \to \mathbb{R}$ *be such that* $\mathbb{E}(|f|) < \infty$*. Let*

$$S_N(\omega) = \sum_{n=0}^{N-1} f(\theta^n \omega)$$

and

$$M_N(\omega) = \max\{S_0(\omega), \dots, S_N(\omega)\}$$

where $S_0 = 0$*. Then*

$$\int_{\{M_N > 0\}} f(\omega)\mathbb{P}(d\omega) \geq 0$$

for every $N \geq 1$*.*

Proof. Observe that for $0 \leq k \leq N$ and every $\omega \in \Omega$ we have

$$f(\omega) + S_k(\theta\omega) - S_{k+1}(\omega)$$

and

$$S_k(\theta\omega) \leq M_N(\theta\omega).$$

This implies that,

$$f(\omega) + M_N(\theta\omega) \geq f(\omega) + S_k(\theta\omega) = S_{k+1}(\omega).$$

Therefore,

$$f(\omega) \geq \max(S_1(\omega), \dots, S_N(\omega)) - M_N(\theta\omega). \tag{8.1.2}$$

Furthermore, on the set $\{M_N > 0\}$ we have

$$M_N(\omega) = \max(S_1(\omega), \dots, S_N(\omega)).$$

Combined with (8.1.2) it follows that $f(\omega) \geq M_N(\omega) - M_N(\theta\omega)$ on $\{M_N > 0\}$. Note that as $M_N \geq 0$ we have

$$\begin{aligned}
\mathbb{E}(M_N) &= \int_\Omega M_N(\omega)\,\mathrm{d}\mathbb{P}(\omega) \\
&= \int_{\{M_N=0\}} M_N(\omega)\,\mathrm{d}\mathbb{P}(\omega) + \int_{\{M_N>0\}} M_N(\omega)\,\mathrm{d}\mathbb{P}(\omega) \\
&= 0\,(\mathbb{P}(\{M_N = 0\})) + \int_{\{M_N>0\}} M_N(\omega)\,\mathrm{d}\mathbb{P}(\omega) \\
&= \int_{\{M_N>0\}} M_N(\omega)\,\mathrm{d}\mathbb{P}(\omega).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\int_{\{M_N>0\}} f(\omega)\,\mathbb{P}(\mathrm{d}\omega) &\geq \int_{\{M_N>0\}} M_N(\omega) - M_N(\theta\omega)\,\mathbb{P}(\mathrm{d}\omega) \\
&\geq \mathbb{E}(M_N) - \int_{A_N} M_N(\omega)\,\mathbb{P}(\mathrm{d}\omega),
\end{aligned}$$

where $A_N := \{\theta\omega : M_N(\omega) > 0\} \subseteq \Omega$. Hence,

$$\int_{\{M_N>0\}} f(\omega)\,\mathbb{P}(\mathrm{d}\omega) \geq 0.$$

$\square$

Theorem 8.1.9 (Birkhoff's Ergodic Theorem)**.** *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *with* θ *be a dynamical system and let* $\mathcal{I}$ *be as in* (8.1.1). *For* $f : \Omega \to \mathbb{R}$ *with* $\mathbb{E}(|f|) < \infty$ *it follows that*

$$\lim_{N\to\infty} \left(\frac{1}{N} \sum_{n=0}^{N-1} f\left(\theta^n \omega\right) \right) = \mathbb{E}(f|\mathcal{I})$$

almost surely.

Proof. By replacing f with $f - \mathbb{E}(f|\mathcal{I})$, we can assume without loss of generality that $\mathbb{E}(f|\mathcal{I}) = 0$. Let

$$\overline{\eta} = \limsup_{n\to\infty} \frac{S_n}{n}$$

and

$$\underline{\eta} = \liminf_{n\to\infty} \frac{S_n}{n}.$$

Note that $\overline{\eta}(\theta\omega) = \overline{\eta}(\omega)$, so that for $\epsilon > 0$ it follows that

$$A^\epsilon = \{\overline{\eta}(\omega) > \epsilon\} \in \mathcal{I}.$$

Let

$$f^\epsilon(\omega) = (f(\omega) - \epsilon)\mathbf{1}_{A^\epsilon}(\omega)$$

to make analogous definition of S_N^ϵ and M_N^ϵ as those made in Theorem 8.1.8. Then using Theorem 8.1.8 it follows that

$$\int_{\{M_N^\epsilon > 0\}} f^\epsilon(\omega)\,\mathbb{P}(\mathrm{d}\omega) \geq 0$$

for $N \geq 1$. Next, observe that the sequence of sets $\{M_N^\epsilon > 0\}$ indexed by N is increasing to

$$B^\epsilon := \left\{ \sup_N S_N^\epsilon > 0 \right\} = \left\{ \sup_N \frac{S_N^\epsilon}{N} > 0 \right\}.$$

Noting that

$$\frac{S_N^\epsilon(\omega)}{N} = \begin{cases} 0 & \overline{\eta}(\omega) \leq \epsilon \\ \frac{S_N(\omega)}{N} - \epsilon & \overline{\eta}(\omega) > \epsilon, \end{cases}$$

it follows that

$$B^\epsilon = \{\overline{\eta} > \epsilon\} \cap \left\{ \sup_N \frac{S_N}{N} > \epsilon \right\} = \{\overline{\eta} > \epsilon\} = A^\epsilon.$$

As $\mathbb{E}\left(|f^\epsilon|\right) \leq \mathbb{E}\left(|f^\epsilon|\right) + \epsilon < \infty$, using the dominated convergence theorem we deduce that

$$0 \leq \lim_{N\to\infty} \int_{\{M_N^\epsilon > 0\}} f^\epsilon(\omega)\,\mathbb{P}(\mathrm{d}\omega) = \int_{A^\epsilon} f^\epsilon(\omega)\,\mathbb{P}(\mathrm{d}\omega).$$

Therefore,

$$\begin{aligned} 0 &\leq \int_{A^\epsilon} f^\epsilon(\omega)\,\mathbb{P}(\mathrm{d}\omega) \\ &= \int_{A^\epsilon} f(\omega) - \epsilon\,\mathbb{P}(\mathrm{d}\omega) \\ &= \int_{A^\epsilon} f(\omega)\,\mathbb{P}(\mathrm{d}\omega) - \epsilon\mathbb{P}\left(A^\epsilon\right) \\ &= \int_{A^\epsilon} \mathbb{E}(f|\mathcal{I})(\omega)\,\mathbb{P}(\mathrm{d}\omega) - \epsilon\mathbb{P}\left(A^\epsilon\right) \\ &= 0 - \epsilon\mathbb{P}\left(A^\epsilon\right). \end{aligned}$$

Therefore, $\mathbb{P}\left(A^\epsilon\right) = 0$ for every $\epsilon > 0$, implying that $\overline{\eta} \leq 0$ almost surely. Repeating a similar argument with $-f$ we see that $\underline{\eta} \geq 0$. Hence, $\overline{\eta} = \underline{\eta} = 0$ which completes the proof. $\square$

Corollary 8.1.10. *If a dynamical system $(\Omega, \mathcal{F}, \mathbb{P})$ with θ is ergodic, then*

$$\lim_{N\to\infty}\left(\frac{1}{N}\sum_{n=0}^{N-1} f\left(\theta^n\omega\right)\right) = \mathbb{E}(f)$$

almost surely.

Proof. As $\mathbb{E}(f|\mathcal{I})$ is $\mathcal{I}$-measurable, if θ is ergodic it follows by statement 2. of Proposition 8.1.7 that $\mathbb{E}(f|\mathcal{I})$ is almost surely constant. More specifically, $\mathbb{E}(f|\mathcal{I}) = \mathbb{E}(f)$. Therefore,

$$\lim_{N\to\infty}\left(\frac{1}{N}\sum_{n=0}^{N-1} f\left(\theta^n\omega\right)\right) \overset{\text{Thm } 8.1.9}{=} \mathbb{E}(f|\mathcal{I}) = \mathbb{E}(f).$$

□

8.2 Birkhoff's Ergodic Theorem for Markov Chains

Having seen how ergodic theorems can be derived for dynamical systems, it will be useful to investigate how to arrive at a dynamical system from a Markov chain. For this we investigate the space of sequences $\mathcal{X}^{\mathbb{N}}$, and more generally $\mathcal{X}^{\mathbb{Z}}$. More specifically, we consider the shift operator θ on these spaces, which is such that for $x := (x_0, x_1, \dots) \in \mathcal{X}^{\mathbb{N}}$ we have

$$\theta(x) = (x_1, x_2, \dots)$$

and similarly for $x := (\dots, x_{-1}, x_0, x_1, \dots)$. In particular, for $x \in \mathcal{X}^{\mathbb{N}}$ we let

$$(\theta_n x)(m) = x_{n+m},$$

and similarly for $x \in \mathcal{X}^{\mathbb{Z}}$. Note that on $\mathcal{X}^{\mathbb{Z}}$ we have $\theta = \theta_1$ and $\theta^{-1} = \theta_{-1}$. Therefore, as before we can consider

$$\mathcal{I} = \left\{C \in \mathcal{B}\left(\mathcal{X}^{\mathbb{Z}}\right) : \theta^{-1}C = C\right\}. \tag{8.2.1}$$

Thus to understand how the theory of dynamical systems can be used on Markov chains it will be important to understand how to construct two-side Markov processes on $\mathcal{X}^{\mathbb{Z}}$. To do so it will be necessary to work with a family of transition probabilities $P = (P(x,\cdot), x \in \mathcal{X})$ and a P-invariant probability measure π, that is $\pi = \int_{\mathcal{X}} P(x,\cdot)\pi(\mathrm{d}x)$.

8.2.1 Constructing Two-Sided Markov Processes

One approach to constructing a Markov chain $(X_n)_{n\in\mathbb{Z}}$ is referred to as the finite-dimensional approach. In this case, a probability measure $\mathbb{P}_\pi$ measure is constructed using finite-dimensional distributions and Kolmogorov's extension theorem. More specifically, $\mu_{n,m}$ denotes the distribution of $(X_{-n}, \dots, X_0, \dots, X_m)$ which is given by

$$P(z_{m-1}, \mathrm{d}z_m)\dots P(z_0, \mathrm{d}z_1)\dots P(z_{-n}, \mathrm{d}z_{-n+1})\pi(\mathrm{d}z_{-n}) = \prod_{k=-n}^{m-1} P(z_k, \mathrm{d}z_{k+1})\pi(\mathrm{d}z_{-n}).$$

By the invariance of π it follows that $(\mu_{n,m})_{n,m\in\mathbb{Z}}$ is a consistent family of probability measures. Therefore, using Theorem 3.2.7 it follows that $\mathbb{P}_\pi$ on $\mathcal{X}^{\mathbb{Z}}$ defines a stationary Markov chain with transition probabilities P and $\mathrm{Law}(X_n) = \pi$ for $n \in \mathbb{Z}$.

Another approach to constructing a Markov chain $(X_n)_{n\in\mathbb{Z}}$ is referred to as the time shift approach. Here we start with a Markov chain $(Y_n)_{n\in\mathbb{N}}$ which has π as an invariant initial distribution and let $\left(X_n^{(m)}\right)_{n\geq -m}$ be such that

$$\left(X_{-m}^{(m)}, X_{-(m-1)}^{(m)}, X_{-(m-2)}^{(m)}, \dots\right) = \left(X_{-(m-1)}^{(m-1)}, X_{-(m-2)}^{(m-1)}, X_{-(m-3)}^{(m-1)}, \dots\right)$$

with

$$\left(X_{-1}^{(1)}, X_1^{(1)}, X_0^{(1)}, \dots\right) = (Y_0, Y_1, Y_2, \dots).$$

In the limit the process $\left(X_n^{(m)}\right)$ gives the required Markov process on $\mathcal{X}^{\mathbb{Z}}$.

8.2.2 Birkhoff's Ergodic Theorem for Stationary Markov Chains as Dynamical Systems

Lemma 8.2.1. *The probability space $\left(\mathcal{X}^{\mathbb{Z}}, \mathcal{B}\left(\mathcal{X}^{\mathbb{Z}}\right), \mathbb{P}_\pi\right)$ with the shift operator θ is a dynamical system.*

Proof. Let $(X_n)_{n\in\mathbb{Z}}$ be a time-homogeneous Markov chain with law $\mathbb{P}_\pi$. By the finite-dimensional construction, it follows that

$$\mathbb{P}\left(X_{-n} \in A_{-n}, \dots, X_m \in A_m\right) = \int_{A_{-n}} \cdots \int_{A_m} P(x_{m-1}, \mathrm{d}x_m) \dots P(x_{-n}, \mathrm{d}x_{-(n-1)}\pi(\mathrm{d}x_{-n}), \tag{8.2.2}$$

similarly,

$$\mathbb{P}\left(X_{-(n-1)} \in A_{-n}, \dots, X_{m+1} \in A_m\right) = \int_{A_{-(n-1)}} \cdots \int_{A_{m+1}} P(x_m, \mathrm{d}x_{m+1}) \dots P(x_{-(n-1)}, \mathrm{d}x_{-(n-2)}\pi(\mathrm{d}x_{-(n-1)}). \tag{8.2.3}$$

Clearly, the right-hand sides of (8.2.2) and (8.2.3) are identical, and so

$$\mathbb{P}\left(X_{-n} \in A_{-n}, \dots, X_m \in A_m\right) = \mathbb{P}\left(X_{-(n-1)} \in A_{-n}, \dots, X_{m+1} \in A_m\right),$$

which is to say that θ preserving map and so defined a dynamical system. ☐

Remark 8.2.2. *Under the product topology, it is apparent that θ is continuous. Thus, the dynamical system of Lemma 8.2.1 is referred to as the continuous dynamical system.*

Definition 8.2.3.

1. *A measure $\mathbb{P}$ is ergodic for θ if for every $A \in \mathcal{I}$, where $\mathcal{I}$ is as given by (8.2.1), we have $\mathbb{P}(A) \in \{0, 1\}$.*
2. *An invariant measure π that induces an ergodic probability measure $\mathbb{P}_\pi$ for θ is said to be ergodic with respect θ.*

Applying Theorem 8.1.9 to $(\theta_n a)_{n\in\mathbb{Z}}$ gives Corollary 8.2.4.

Corollary 8.2.4. *Let $f : \mathcal{X} \to \mathbb{R}$ be integrable and let $\tilde{f} : \mathcal{X}^{\mathbb{Z}} \to \mathbb{R}$ be given by*

$$\tilde{f}(\dots, a_{-1}, a_0, a_1, \dots) = f(a_0).$$

Then $\tilde{f}\left(\theta_n a\right) = f(a_n)$, so that

$$\frac{1}{n}\sum_{k=1}^{n} f(a_k) \overset{n\to\infty}{\longrightarrow} \mathbb{E}_\pi\left(\tilde{f}|\mathcal{I}\right)$$

almost surely with respect to $\mathbb{P}_\pi$. Moreover, if π is ergodic then

$$\frac{1}{n}\sum_{k=1}^{n} f(a_k) \overset{n\to\infty}{\longrightarrow} \int_{\mathcal{X}} f \,\mathrm{d}\pi$$

almost surely with respect to $\mathbb{P}_\pi$.

Remark 8.2.5. *An analogous statement holds over $\mathcal{X}^{\mathbb{N}}$.*

8.2.3 Birkhoff's Ergodic Theorem for Markov Chains

Consider $(X_n)_{n\in\mathbb{Z}}$ a stationary time-homogeneous Markov chain on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with transition probability P and a P-invariant initial distribution π. Currently, our ergodic theory makes statements about

$(X_n)_{n\in\mathbb{Z}}$ as a dynamical system on $\left(\mathcal{X}^{\mathbb{Z}}, \mathcal{B}\left(\mathcal{X}^{\mathbb{Z}}\right), \mathbb{P}_\pi\right)$ with θ being the shift operator. Here we expand our ergodic theory to make statements about $(X_n)_{n\in\mathbb{Z}}$ in terms of $\mathbb{P}$ instead of $\mathbb{P}_\pi$. For this, let

$$I_P = \{\pi \in \mathcal{P}(\mathcal{X}) : T\pi = \pi\}.$$

Theorem 8.2.6. *Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary Markov process with $X_0 \sim \pi$, where π is an invariant probability measure. Then the following hold.*

1. *Let $f : \mathcal{X}^{\mathbb{Z}} \to \mathbb{R}$ be an integrable function. Let $\bar{f} = \mathbb{E}_\pi(f|\mathcal{I})$, then*

$$\frac{1}{n}\sum_{k=1}^{n} f\left(\theta^k X(\omega)\right) \overset{n\to\infty}{\longrightarrow} \bar{f}(X(\omega))$$

for almost every ω with respect to $\mathbb{P}$.

2. *Moreover, if π is ergodic then*

$$\frac{1}{n}\sum_{k=1}^{n} f\left(\theta^k X(\omega)\right) \overset{n\to\infty}{\longrightarrow} \int_{\mathcal{X}^{\mathbb{Z}}} f \,\mathrm{d}\mathbb{P}_\pi$$

for almost every ω with respect to $\mathbb{P}$.

Proof. For $f : \mathcal{X}^{\mathbb{Z}} \to \mathbb{R}$ in $L^1(\mathbb{P}_\pi)$ consider

$$E := \left\{ a \in \mathcal{X}^{\mathbb{Z}} : \frac{1}{n}\sum_{k=1}^{n} f\left(\theta^k a\right) \overset{n\to\infty}{\longrightarrow} \mathbb{E}_\pi\left(\tilde{f}|\mathcal{I}\right) \right\}.$$

By Corollary 8.2.4 it follows that $\mathbb{P}_\pi(E) = 1$. In particular, we have that

$$\mathbb{P}(\{\omega : X(\omega) \in E\}) = \mathbb{P}_\pi(E) = 1.$$

Therefore, for almost every ω with respect to $\mathbb{P}$ we have that $\frac{1}{n}\sum_{k=1}^{n} f\left(\theta^k X(\omega)\right)$ converges. ☐

Theorem 8.2.7. *Let $P = P(x,\cdot)$ be a transition probability with an invariant measure π. Let $(X_n)_{n\in\mathbb{Z}}$ be a time-homogeneous Markov process with transition probabilities P and initial position $X_0 = x$. Then for π-almost every $x \in \mathcal{X}$ the following hold.*

1. *For $f : \mathcal{X}^{\mathbb{Z}} \to \mathbb{R}$ an integrable function $\frac{1}{n}\sum_{k=1}^{n} f\left(\theta^k X(\omega)\right)$ converges for almost every ω with respect to $\mathbb{P}$.*

2. *Moreover, if π is ergodic, then*

$$\frac{1}{n}\sum_{k=1}^{n} f\left(\theta^k X(\omega)\right) \overset{n\to\infty}{\longrightarrow} \int_{\mathcal{X}} f \,\mathrm{d}\mathbb{P}_\pi$$

for almost every ω with respect to $\mathbb{P}$.

Proof. Suppose that $X_0 \sim \pi$. Then by Theorem 8.2.6 it follows that

$$\frac{1}{n}\sum_{k=1}^{n} f\left(\theta^k X(\omega)\right) \overset{n\to\infty}{\longrightarrow} \bar{f}(X(\omega)$$

for almost every ω with respect to $\mathbb{P}$. Therefore, by the dominated convergence theorem it follows that

$$\mathbb{E}\left(\frac{1}{n}\sum_{k=1}^{n} f\left(\theta^k X(\omega)\right)|\sigma(X_0)\right) \overset{n\to\infty}{\longrightarrow} \mathbb{E}\left(\bar{f}(X(\omega)|\sigma(X_0)\right)$$

for almost every ω with respect to $\mathbb{P}$. Therefore, for almost every $x \in \mathcal{X}$ with respect to π we have

$$\mathbb{E}\left(\frac{1}{n}\sum_{k=1}^{n} f\left(\theta^k X(\omega)\right)|X_0 = x\right) \overset{n\to\infty}{\longrightarrow} \mathbb{E}\left(\bar{f}(X(\omega)|X_0 = x\right)$$

for almost every ω with respect to $\mathbb{P}$. □

Corollary 8.2.8. *Let P be a transition probability with an ergodic invariant probability measure π. Then for $g : \mathcal{X} \to \mathbb{R}$ in $L^1(\pi)$ it follows that*

$$\frac{1}{n}\sum_{k=1}^{n} g(X_k(\omega)) \overset{n\to\infty}{\longrightarrow} \int_{\mathcal{X}} g \,\mathrm{d}\pi$$

for almost every ω with respect to $\mathbb{P}$.

Proof. Let $\tilde{g} : \mathcal{X}^{\mathbb{Z}} \to \mathbb{R}$ be given by $\tilde{g}(y) := g(y_0)$. Then using Theorem 8.2.7 with $\tilde{g}$ and noting that

$$\int_{\mathcal{X}^{\mathbb{Z}}} \tilde{g} \,\mathrm{d}\mathbb{P}_\pi = \int_{\mathcal{X}} g \,\mathrm{d}\pi$$

the result follows. □

Proposition 8.2.9. *Ergodic invariant probability measures for a time-homogeneous Markov chain are equal or mutually singular.*

Proof. Let π_1 and π_1 be distinct invariant probability measures. Let $f : \mathcal{X} \to \mathbb{R}$ be a bounded measurable function such that

$$\int_{\mathcal{X}} f \,\mathrm{d}\pi_1 = \int_{\mathcal{X}} f \,\mathrm{d}\pi_2, \tag{8.2.4}$$

which exists as the measures are distinct. Let $(X_n)_{n\in\mathbb{Z}}$ be a Markov chain with $X_0 = x$. Then for $i = 1, 2$ let

$$E_i := \left\{ x : X_0 = x,\ \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(X_k) = \int_{\mathcal{X}} f \,\mathrm{d}\pi_i \,\mathbb{P} \text{ almost everywhere} \right\}.$$

By Corollary 8.2.8 the set E_i is well-defined and is such that $\pi_i(E_i) = 1$. By (8.2.4) we have that $E_1 \cap E_2 = \emptyset$ and so $\pi_1(E_2) = 0$ which means that π_1 and π_2 are mutually singular if they are not equal. □

8.3 The Structure Theorem

The structure theorem relates to the invariant probability measures of a time-homogeneous Markov chain. More specifically, we consider a chain on $\mathcal{X}$ with transition probability P, and with T being the corresponding transition operator. Moreover, let

$$I_P := \{\pi \in \mathcal{P}(\mathcal{X}) : T\pi = \pi\} \tag{8.3.1}$$

consist of the probability measures that are invariant under T.

Exercise 8.3.1. *Let I_P be as given in* (8.3.1). *Show that the following statements hold.*

1. *If $\pi_1, \pi_2 \in I_P$ then $t\pi_1 + (1-t)\pi_2 \in I_P$ for $t \in [0, 1]$.*
2. *If T is Feller, then I_P is closed.*

Definition 8.3.2. *A probability measure $\pi \in I_P$ is an extremal of I_P if π cannot be decomposed as $\pi = t\pi_1 + (1-t)\pi_2$ for $t \in (0,1)$ and $\pi_1, \pi_2 \in I_P$ distinct.*

To work towards the structure theorem we need to understand the measurability of set-theoretic constructions.

Definition 8.3.3. *The set-theoretic difference of sets A and B is given by $A \triangle B := A \cup B \setminus (A \cap B)$.*

Proposition 8.3.4. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $A, B, (A_\alpha)_{\alpha\in\mathcal{A}}$ and $(B_\alpha)_{\alpha\in\mathcal{A}}$ be elements of $\mathcal{F}$. Then the following statements hold.*

1. $A^c \triangle B^c = A \triangle B$.
2. $\left(\bigcup_{\alpha\in\mathcal{A}} A_\alpha\right) \triangle \left(\bigcup_{\alpha\in\mathcal{A}} B_\alpha\right) \subset \bigcup_{\alpha\in\mathcal{A}} (A_\alpha \triangle B_\alpha)$.
3. *For a measurable function $f : \Omega \to \Omega$ we have $f^{-1}(A \triangle B) = f^{-1}(A) \triangle f^{-1}(B)$.*
4. $(A \triangle B) \triangle (B \triangle C) = A \triangle C$.
5. $\mathbb{P}(A \triangle B) = 0$ *implies that* $\mathbb{P}(A) = \mathbb{P}(B)$.

Proof.

1. Proceeding directly,
$$\begin{aligned} A^c \triangle B^c &= ((\Omega \setminus A) \setminus (\Omega \setminus B) \cup ((\Omega \setminus B) \setminus (\Omega \setminus A) \\ &= (A \setminus B) \cup (B \setminus A) \\ &= A \triangle B. \end{aligned}$$

2. Let $\omega \in \left(\bigcup_{\alpha\in\mathcal{A}} A_\alpha\right) \triangle \left(\bigcup_{\alpha\in\mathcal{A}} B_\alpha\right)$ then
$$\omega \in \left(\bigcup_{\alpha\in\mathcal{A}} A_\alpha\right) \setminus \left(\bigcup_{\alpha\in\mathcal{A}} B_\alpha\right) \cup \left(\bigcup_{\alpha\in\mathcal{A}} B_\alpha\right) \setminus \left(\bigcup_{\alpha\in\mathcal{A}} A_\alpha\right).$$
Suppose without loss of generality that $\omega \in \left(\bigcup_{\alpha\in\mathcal{A}} A_\alpha\right) \setminus \left(\bigcup_{\alpha\in\mathcal{A}} B_\alpha\right)$, with $\omega \in A_{\tilde{\alpha}}$. It follows that $\omega \in A_{\tilde{\alpha}} \setminus B_{\tilde{\alpha}}$ and so $\omega \in A_{\tilde{\alpha}} \triangle B_{\tilde{\alpha}}$. Therefore,
$$\left(\bigcup_{\alpha\in\mathcal{A}} A_\alpha\right) \triangle \left(\bigcup_{\alpha\in\mathcal{A}} B_\alpha\right) \subset \bigcup_{\alpha\in\mathcal{A}} (A_\alpha \triangle B_\alpha).$$

3. This follows from the fact that the pre-image distributed over unions and intersections.

4. Figure 11a is a visual representation of
$$((A \setminus B) \cup (B \setminus A)) \setminus ((B \setminus C) \cup (C \setminus B)) \tag{8.3.2}$$
and Figure 11b is a visual representation of
$$((B \setminus C) \cup (C \setminus B)) \setminus ((A \setminus B) \cup (B \setminus A)). \tag{8.3.3}$$
Noting that $(A \triangle B) \triangle (B \triangle C)$ is given by the union of (8.3.2) and (8.3.3) we use Figure 11c to conclude that
$$(A \triangle B) \triangle (B \triangle C) = A \triangle C.$$

5. As $A \setminus B$ and $B \setminus A$ are disjoint sets it follows that

$$\mathbb{P}(A \triangle B) = \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A).$$

Therefore, if $\mathbb{P}(A \triangle B) = 0$ it must be the case that $\mathbb{P}(A \setminus B) = \mathbb{P}(B \setminus A) = 0$. Noting that

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \setminus B)$$

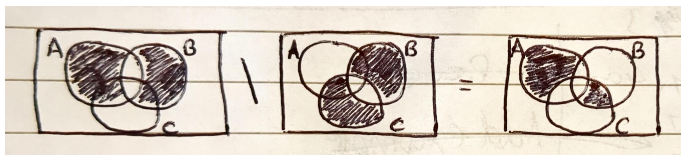
and

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A).$$

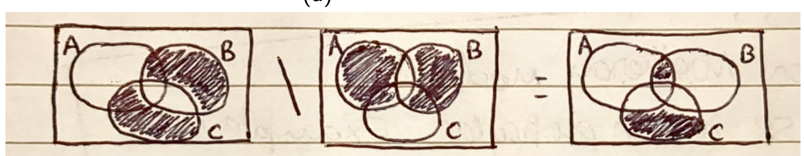
It follows that

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) = \mathbb{P}(B).$$

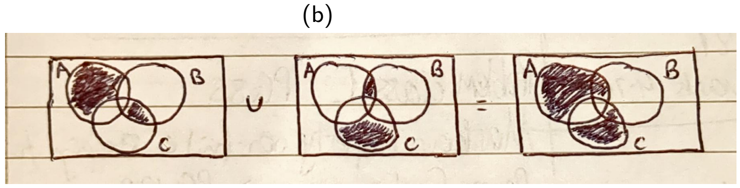
□



(a)



(b)



(c)

Figure 11: Visual proof for statement 3. of Proposition 8.3.4.

Definition 8.3.5. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then for $A, B \in \mathcal{F}$ we say $A \sim B$ if and only if $\mathbb{P}(A \triangle B) = 0$.*

Definition 8.3.6. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.*

1. *The σ-algebra $\mathcal{F}$ is complete with respect to $\mathbb{P}$ if whenever $B \in \mathcal{F}$ and $\mathbb{P}(B) = 0$, then for any $A \subset B$ it follows that $A \in \mathcal{F}$.*

2. *The completion $\bar{\mathcal{F}}$ of $\mathcal{F}$ with respect to $\mathbb{P}$ is the smallest σ-algebra containing $\mathcal{F}$ that is complete with respect to the measure $\mathbb{P}$.*

Recall, that we can view our time-homogeneous Markov process as the canonical stochastic process on the probability space $\left(\mathcal{X}^{\mathbb{Z}}, \mathcal{B}\left(\mathcal{X}^{\mathbb{Z}}\right), \mathbb{P}_\pi\right)$, where $\mathbb{P}_\pi$ is the law of the chain with transition probability P and initial distribution π. We can then consider θ and θ^{-1} measure preserving transformations on $\mathcal{X}^{\mathbb{Z}}$ and

$$\mathcal{F}_n^m := \bigvee_{k=-n}^{m} \sigma(X_k) \subset \mathcal{B}\left(\mathcal{X}^{\mathbb{Z}}\right).$$

Lemma 8.3.7. *Let $A \in \mathcal{B}\left(\mathcal{X}^{\mathbb{Z}}\right)$, then for any $\epsilon > 0$ there exists an $N > 0$ and $A_\epsilon \in \mathcal{F}_{-N}^{N}$ such that*

$$\mathbb{P}\left(A \triangle A_\epsilon\right) < \epsilon.$$

Proof. Consider

$$\mathcal{B}_0 := \left\{A \in \mathcal{B}\left(\mathcal{X}^{\mathbb{Z}}\right) : \text{for all } \epsilon > 0, \text{ there exists } N \in \mathbb{N} \text{ and } A_\epsilon \in \mathcal{F}_{-N}^{N} \text{ such that } \mathbb{P}(A \triangle A_\epsilon) < \epsilon\right\}.$$

Clearly, $\emptyset \in \mathcal{B}_0$. By statement 1. of Proposition 8.3.4 it follows that if $A \in \mathcal{B}_0$ then $A^c \in \mathcal{B}_0$. Now consider $(A_j)_{j\in\mathbb{N}} \subset \mathcal{B}_0$. By construction there exists a sequence $(N_j)_{j\in\mathbb{N}} \subset \mathbb{N}$ with corresponding events $A_j' \in \mathcal{F}_{-N_j}^{N_j}$ such that $\mathbb{P}\left(A_j \triangle A_j'\right) \leq \epsilon 2^{-j}$. Since $\mathbb{P}$ is a finite measure, there exists a J such that with $A := \bigcup_{j\in\mathbb{N}} A_j$ we have $\mathbb{P}\left(A \triangle \bigcup_{j\leq J} A_j\right) \leq \epsilon$. Therefore,

$$\begin{aligned}
\mathbb{P}\left(A \triangle \bigcup_{j\leq J} A_j'\right) &= \mathbb{P}\left(\left(A \triangle \bigcup_{j\leq J}\right) \triangle \left(\bigcup_{j\leq J} A_j \triangle \bigcup_{j\leq J} A_j'\right)\right) \\
&\leq \mathbb{P}\left(A \triangle \bigcup_{j\leq J} A_j\right) + \mathbb{P}\left(\bigcup_{j\leq J} A_j \triangle \bigcup_{j\leq J} A_j'\right) \\
&\leq \epsilon + \mathbb{P}\left(\bigcup_{j\leq J} \left(A_j \triangle A_j'\right)\right) \\
&\leq \sum_{j\leq J} 2\epsilon.
\end{aligned}$$

Hence, as $\bigcup_{j\leq J} A_j' \in \mathcal{F}_{-N}^{N}$ where $N = \max_{j\leq J}(N_j)$ we have that $\bigcup_{j\in\mathbb{N}} A_j \in \mathcal{B}_0$ which makes $\mathcal{B}_0$ a σ-algebra. It follows that $\mathcal{B}_0 = \mathcal{B}\left(\mathcal{X}^{\mathbb{Z}}\right)$ which completes the proof. ☐

Lemma 8.3.8. *For any $A \in \mathcal{I} := \left\{C \in \mathcal{B}\left(\mathcal{X}^{\mathbb{Z}}\right) : \theta^{-1}C = C\right\}$ and $l \in \mathbb{Z}$, there exists $\hat{A}_l \in \sigma(X_l)$ such that $A \sim \hat{A}_l$.*

Proof. Let $A \in \mathcal{I}$, then for any $\epsilon > 0$ we can use Lemma 8.3.7 to find a $N \in \mathbb{N}$ and $A_\epsilon \in \mathcal{F}_{-N}^{N}$ such that $\mathbb{P}\left(A \triangle A_\epsilon\right) < \epsilon$. Since,

$$\theta_{-1}\left(A \triangle A_\epsilon\right) = \theta_{-1}(A) \triangle \theta_{-1}\left(A_\epsilon\right) = A \triangle \theta_{-1}\left(A_\epsilon\right),$$

and $\mathbb{P}$ is θ-invariant, it follows that

$$\mathbb{P}\left(A \triangle \theta_{-k} A_\epsilon\right) < \epsilon \tag{8.3.4}$$

for all $k \geq 0$. In particular, observe that $\theta^{-(N+k)} A_\epsilon \in \mathcal{F}_k^{2N+k} \subset \mathcal{F}_k^{\infty}$ for fixed $k \geq 0$ and arbitrary $\epsilon > 0$. Therefore, with ϵ and N as previously given, fix k and let $\epsilon_m = \frac{1}{m}$. Furthermore, let

$$D_n^\epsilon := \theta^{-(N+k)} A_{\frac{\epsilon}{2^n}} \in \mathcal{F}_k^{\infty}$$

and

$$D := \bigcap_{m\geq 1} \bigcup_{n\geq 1} D_n^{\epsilon_m} \in \mathcal{F}_k^{\infty}.$$

Note that using (8.3.4) we have $\mathbb{P}\left(A \triangle D_n^\epsilon\right) < \frac{\epsilon}{2^n}$ and so

$$\mathbb{P}\left(\bigcap_{n\geq 1} \left(A \setminus D_n^{\epsilon_m}\right)\right) = \lim_{n\to\infty} \mathbb{P}\left(A \setminus D_n^{\epsilon_m}\right) \leq \lim_{n\to\infty} \frac{\epsilon_m}{2^n} = 0. \tag{8.3.5}$$

On the other hand,

$$\begin{aligned}\mathbb{P}(D \setminus A) &\leq \mathbb{P}\left(\bigcup_{n=1}^{\infty} D_n^{\epsilon_m} \setminus A\right)\\ &= \mathbb{P}\left(\bigcap_{n=1}^{\infty} (A \setminus D_n^{\epsilon_m})\right)\\ &\leq \frac{\epsilon}{2^n}\\ &\leq \frac{1}{m}\\ &\overset{m\to\infty}{\longrightarrow} 0.\end{aligned}$$

Therefore, $\mathbb{P}(D \setminus A) = 0$. Furthermore,

$$\begin{aligned}\mathbb{P}(A \setminus D) &= \mathbb{P}\left(A \setminus \bigcap_{m\geq 1}\bigcup_{n\geq 1} D_n^{\epsilon_m}\right)\\ &= \mathbb{P}\left(\bigcup_{m\geq 1}\bigcap_{n\geq 1} (A \setminus D_n^{\epsilon_m})\right)\\ &\overset{(8.3.5)}{=} 0\end{aligned}$$

and so $A \sim D$. Therefore, for any k with $D^{(k)} := D$ we have $D^{(k)} \in \mathcal{F}_k^{\infty}$ and $\mathbb{P}\left(A \triangle D^{(k)}\right) = 0$. Similarly using θ_{-1} for any k there is a $D^{(-k)} \in \mathcal{F}_{-\infty}^{(-k)}$ such that $\mathbb{P}\left(A \triangle D^{(-k)}\right) = 0$. Let $l \in (-k, k)$ then

$$\begin{aligned}\mathbb{E}\left(\mathbf{1}_A | \sigma(X_l)\right) &= \mathbb{E}\left(\mathbf{1}_A^2 | \sigma(X_l)\right)\\ &= \mathbb{E}\left(1_{D^{(-k)}}\mathbf{1}_{D^{(k)}} | \sigma(X_l)\right)\\ &= \mathbb{E}\left(1_{D^{(-k)}} | \sigma(X_l)\right)\mathbb{E}\left(1_{D^{(k)}} | \sigma(X_l)\right)\\ &= \left(\mathbb{E}\left(\mathbf{1}_A | \sigma(X_l)\right)\right)^2,\end{aligned}$$

which implies that $\mathbb{E}\left(\mathbf{1}_A | \sigma(X_l)\right)(\omega) \in \{0, 1\}$ almost surely. Let

$$\hat{A} := \left\{\omega \in \mathcal{X}^{\mathbb{Z}} : \mathbb{E}\left(\mathbf{1}_A | \sigma(X_l)\right)(\omega) = 1\right\} \in \sigma(X_l).$$

Then $\mathbb{E}\left(\mathbf{1}_A | \sigma(X_l)\right) = \mathbf{1}_{\hat{A}}$ and $\mathbb{E}\left(\mathbf{1}_{A^c} | \sigma(X_l)\right) = \mathbf{1}_{A^c}$. Then for any $E \in \sigma(X_l)$ we have

$$\mathbb{P}(A \cap E) = \mathbb{E}\left(\mathbb{E}\left(\mathbf{1}_A | \sigma(X_l)\right)\right) = \mathbb{P}\left(\hat{A} \cap E\right).$$

In particular, $\mathbb{P}\left(A \cap \hat{A}^c\right) = 0$ and $\mathbb{P}\left(\hat{A} \cap A^c\right) = 0$. Hence, $\mathbb{P}\left(A \triangle \hat{A}\right) = 0$. Thus letting $\hat{A}_l := \hat{A}$ we have $A \sim \hat{A}_l$. □

Proposition 8.3.9. *For any $A \in \mathcal{I} := \left\{C \in \mathcal{B}\left(\mathcal{X}^{\mathbb{Z}}\right) : \theta^{-1}C = C\right\}$, there exists $\bar{A} \in \mathcal{B}(\mathcal{X})$ such that*

$$A \sim \prod_{i\in\mathbb{Z}} \bar{A}.$$

Proof. By Lemma 8.3.8, there exists an $\hat{A} \in \sigma(X_0)$ with $A \sim \hat{A}$. Let $\bar{A} \in \mathcal{B}(\mathcal{X})$ be such that

$$\hat{A} := \left\{\omega \in \mathcal{X}^{\mathbb{Z}} : X_0(\omega) \in \bar{A}\right\}.$$

Then by the invariance of A with follows that

$$\mathbb{P}\left(A \triangle \theta_{-n}\hat{A}\right) = \mathbb{P}\left(\theta_{-n}\left(A \triangle \hat{A}\right)\right) = 0,$$

which implies that

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{Z}} A \triangle \theta^{-n}\hat{A}\right) = 0.$$

As $\theta_{-n}\hat{A} = \{\omega : X_n(\omega) \in \bar{A}\}$ we have that

$$\bigcap_{k=0}^{n} \theta_{-k}\hat{A} = \{\omega : X_0(\omega) \in \bar{A}, \ldots, X_n(\omega) \in \bar{A}\} \sim A.$$

Thus,

$$\{X_i \in \bar{A} : i \in \mathbb{Z}\} = \prod_{i\in\mathbb{Z}} \bar{A} \sim A.$$

□

Corollary 8.3.10. *Let π be an invariant probability measure for P. Then π is ergodic if and only if every π-invariant set $\bar{A}$ is such that $\pi\left(\bar{A}\right) \in \{0,1\}$.*

Proof. $(\Rightarrow)$. If $\bar{A}$ is π-invariant, then using Exercise 7.6.2 it follows that

$$\mathbb{P}_\pi\left(\prod_{i\in\mathbb{Z}} \bar{A}\right) = \pi\left(\bar{A}\right).$$

As we assume $\mathbb{P}_\pi$ is ergodic and $\prod_{i\in\mathbb{Z}} \bar{A}$ is a θ-invariant set, it follows that $\pi\left(\bar{A}\right) \in \{0,1\}$.
$(\Leftarrow)$. For $A \in \mathcal{I}$ we have from Proposition 8.3.9 that there exists a $\bar{A} \in \mathcal{B}(\mathcal{X})$ such that $A \sim \prod_{i\in\mathbb{Z}} \bar{A}$... □

Proposition 8.3.11. *Let π be an invariant probability measure for P. Then π is ergodic if and only if π is an extremal of I_P, where I_P is as given in* (8.3.1).

Proof. $(\Rightarrow)$. Suppose that π is not extremal, with $\pi = t\pi_1 + (1-t)\pi_2$ for $t \in (0,1)$ and $\pi_1, \pi_2 \in \mathcal{P}(\mathcal{X})$ distinct. Moreover, suppose that π is ergodic. Then for any θ-invariant set A we have

$$t\mathbb{P}_{\pi_1}(A) + (1-t)\mathbb{P}_{\pi_2}(A) \in \{0,1\},$$

thus $\mathbb{P}_{\pi_1}(A) = \mathbb{P}_{\pi_2} = 0$ or $\mathbb{P}_{\pi_1}(A) = \mathbb{P}_{\pi_2} = 1$ making π_1 and π_2 ergodic. Therefore, by Proposition 8.2.9 π_1 and π_2 are mutually singular. Let $E \in \mathcal{B}(\mathcal{X})$ be such that $\pi_1(E) = 1$, and $\pi_2(E) = 0$, so that $\mathbb{P}_{\pi_1}\left(\prod_{i\in\mathbb{Z}} E\right) = 1$ and $\mathbb{P}_{\pi_2}\left(\prod_{i\in\mathbb{Z}} E\right) = 0$. It follows that

$$\mathbb{P}_\pi\left(\prod_{i\in\mathbb{Z}} E\right) = t\mathbb{P}_{\pi_1}\left(\prod_{i\in\mathbb{Z}} E\right) + (1-t)\mathbb{P}_{\pi_2}\left(\prod_{i\in\mathbb{Z}} E\right) = t < 1,$$

which contradicts π being ergodic.
$(\Leftarrow)$. Suppose π is not ergodic. Then by Corollary 8.3.10 there exists a π-invariant set F with $\pi(F) := t \in (0,1)$. Let $\pi_1, \pi_2 \in \mathcal{P}(\S)$ be given by

$$\pi_1(B) = \frac{1}{t}\pi(B \cap F),$$

and

$$\pi_2(B) = \frac{1}{1-t}\pi\left(B \cap F^c\right).$$

By the π-invariance of F we have $P(x, F) = 1$ for almost every $x \in F$ with respect to π. Using similar arguments to those made in Lemma 7.6.3, we have that π restricted to F is invariant. That is, π_1 is invariant. On the other hand,

$$\pi\left(F^c\right) = \int_F P\left(x, F^c\right) \mathrm{d}\pi + \int_{F^c} P\left(x, F^c\right) \mathrm{d}\pi = \int_{F^c} P\left(x, F^c\right) \mathrm{d}\pi,$$

which implies that $P\left(x, F^c\right) = 1$ for almost every $x \in F^c$ with respect to π. Therefore, π_2 is invariant, and thus $\pi_1, \pi_2 \in I_P$ which contradicts π being extremal. $\square$

Theorem 8.3.12. *Given a time-homogeneous transition probability P, with transition operator T let I_P be as given by* (8.3.1)*. Then for*

$$\mathcal{E} = \{\pi \in \mathcal{P}(\mathcal{X}) : \pi \text{ ergodic}\} \cap \mathcal{I}_P$$

the following statements hold.

1. *I_P is convex with being the set $\mathcal{E}$ of its extremal points.*
2. *For $\pi_1, \pi_2 \in \mathcal{E}$, either π_1 and π_2 are equal or they are mutually singular.*
3. *Every $\pi \in I_P$ can be written as $\pi = t\pi_1 + (1-t)\pi_2$ for some $\pi_1, \pi_2 \in \mathcal{E}$ and $t \in [0, 1]$.*

Proof.

1. I_P being convex follows from statement 1. of Exercise 8.3.1. Proposition 8.3.11 shows that the set of extremal points of I_P coincides with $\mathcal{E}$.
2. This is shown in Proposition 8.2.9.
3. This follows from statement 1.

$\square$

Corollary 8.3.13. *If a time-homogeneous Markov process admits more than one invariant measure, it admits at least two ergodic invariant measures.*

Proof. Let $\pi_1, \pi_2 \in I_P$ be distinct. We are done if $\pi_1, \pi_2 \in \mathcal{E}$. In any other case, we must have that either π_1 or π_2 is not ergodic. Suppose without loss of generality that $\pi_1 \in I_P \setminus \mathcal{E}$. By statement 3. of Theorem 8.3.12 we can write $\pi_1 = t\mu_1 + (1-t)\mu_2$ where $\mu_1, \mu_2 \in \mathcal{E}$. It is clear that μ_1 is not equal to μ_2, otherwise $\pi_1 = \mu_1$ which contradicts π_1 not being ergodic. $\square$

Corollary 8.3.14. *If a time-homogeneous Markov process has a unique invariant measure π, then π is ergodic.*

Proof. In this case $I_P = \{\pi\}$ and so π is an extremal point. Therefore, by statement 1. of Theorem 8.3.12 we have that π is ergodic. $\square$

Proposition 8.3.15. *Let $A \subset \mathcal{X}$ be a P-invariant set. Let $A_0 = A$ and inductively let*

$$A_n = \{x \in \mathcal{X} : P\left(x, A_{n-1}\right) > 0\}.$$

Suppose $\mathcal{X} = \bigcup_{n=1}^{\infty} A_n$ and $A = \bigcup_{k=1}^{m} B_k$ for $(B_k)_{k=1,\ldots,m}$ disjoint closed P-invariant sets. If the time-homogeneous Markov chain restricted to B_k has a unique invariant measure π_k, then π_k is ergodic. Moreover, the $(\pi_k)_{k=1,\ldots,m}$ are the are the only ergodic invariant probability measures of the chain.

Proof. On B_k, the ergodicity follows from Corollary 8.3.14. If $\pi \in I_P$, then using Proposition 7.6.7 it follows that $\pi(A) = 1$. Using arguments made in the proof of Proposition 8.3.11 we have that the restriction on π onto the B_k is an invariant probability measure for P. By the uniqueness of invariant probability measures on B_k we can uniquely write

$$\pi = \sum_{k=1}^{m} \pi(B_k)\pi_k.$$

□

8.4 Solution to Exercises

Exercise 8.1.3

Solution. As $\emptyset \in \mathcal{F}$, and $\theta^{-1}(\emptyset) = \emptyset$ we have $\emptyset \in \mathcal{I}$. Suppose that $A \in \mathcal{I}$, then

$$\theta^{-1}\left(A^c\right) = \left(\theta^{-1}(A)\right)^c = A^c,$$

and so $A^c \in \mathcal{I}$. Now consider $(A_n)_{n\in\mathbb{N}} \subset \mathcal{I}$, then

$$\theta^{-1}\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \bigcup_{n\in\mathbb{N}} \theta^{-1}\left(A_n\right) = \bigcup_{n\in\mathbb{N}} A_n,$$

and so $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{I}$. Therefore, $\mathcal{I}$ forms a σ-algebra. □

Exercise 8.1.5

Solution. Let $f = \mathbf{1}_A$. Note that

$$f = \mathbf{1}_A = \mathbf{1}_{\omega:\omega\in A}$$

and

$$f \circ \theta = \mathbf{1}_{\omega:\theta\omega\in A} = \mathbf{1}_{\theta^{-1}(A)}.$$

Therefore, $f \circ \theta = f$ if and only if $A = \theta^{-1}(A)$ which is to say that $A \in \mathcal{I}$. Therefore, $f \circ \theta = f$ if and only if f is $\mathcal{I}$-measurable. Extending the above argument by linearity it follows that $f = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}$ is θ-invariant if and only if f is measurable with respect to θ. As measurability is preserved under limits, and any non-negative function is the limit of a sequence of simple functions it follows that $f : \Omega \to [0, \infty]$ is θ-invariant if and only if f is $\mathcal{I}$-measurable. Then for general $f : \Omega \to [0, \infty]$, using the decomposition $f = f^+ - f^-$, we conclude that $f : \Omega \to \mathbb{R}$ is θ-invariant if and only if f is $\mathcal{I}$-measurable. □

Exercise 8.3.1

Solution.

1. This follows directly from the linearity of the transition operator T.
2. T being Feller means that it is a continuous map from $\mathcal{P}(\mathcal{X})$ to $\mathcal{P}(\mathcal{X})$ under the weak topology. Therefore, for $(\pi_n)_{n\in\mathbb{N}} \subset I_P$ a sequence converging weakly to a limit π we have that

$$T\pi = T \lim_{n\to\infty} \pi_n = \lim_{n\to\infty} T\pi_n = \lim_{n\to\infty} \pi_n = \pi.$$

Therefore, $\pi \in I_P$ which means that I_P is closed.

□

9 Appendix

9.1 Gaussian Measures

Definition 9.1.1. *A measure μ on $\mathbb{R}^n$ is Gaussian if there exists a non-negative symmetric matrix K and vector $m \in \mathbb{R}^n$ such that*

$$\int_{\mathbb{R}^n} e^{i\langle \lambda, x\rangle} \mu(\mathrm{d}x) = \exp\left(i\langle \lambda, m\rangle - \frac{1}{2}\langle K\lambda, \lambda\rangle\right).$$

If K is non-degenerate then the density with respect to the Lebesgue measure is given by

$$\frac{1}{\sqrt{(2\pi)^n \det(K)}} \exp\left(-\frac{1}{2}\langle K^{-1}(x-m), x-m\rangle\right).$$

In this case, m is called the mean, and K the covariance operator.

Remark 9.1.2.

1. *We are using the notation*
$$\langle u, v\rangle = u^\top v.$$
2. *A Gaussian measure is specified entirely by its mean and covariance operator.*

Theorem 9.1.3. *If X is a Gaussian random variable on $\mathbb{R}^n$ with mean m, covariance operator K and $A : \mathbb{R}^d \to \mathbb{R}^d$ is a linear map. Then AX is a Gaussian random variable with covariance operator $AKA^\top$.*

Proof. For $\lambda \in \mathbb{R}^n$ we have

$$\begin{aligned}
\mathbb{E}\left(\exp\left(i\langle \lambda, AX\rangle\right)\right) &= \mathbb{E}\left(\exp\left(i\left\langle A^\top\lambda, X\right\rangle\right)\right) \\
&\overset{(1)}{=} \exp\left(i\left\langle \lambda, Am\right\rangle - \frac{1}{2}\left\langle KA^\top\lambda, A^\top\lambda\right\rangle\right) \\
&= \exp\left(i\langle \lambda, Am\rangle - \frac{1}{2}\left\langle AKA^\top\lambda, \lambda\right\rangle\right)
\end{aligned}$$

where (1) follows as X is a Gaussian random variable with covariance operator K and mean m. Therefore, as a Gaussian is determined by its mean and covariance, we deduce that AX is a Gaussian random variable with mean Am and covariance $AKA^\top$. ☐

Proposition 9.1.4. *If $\{X_1, \ldots, X_n\}$ are independent Gaussian random variables on $\mathbb{R}^d$ then for $a_i \in \mathbb{R}$ the random variable $\sum_{i=1}^n a_i X_i$ is also Gaussian.*

Proof. Suppose that each X_i has mean and covariance operator m_i and K_i respectively. Then for $\lambda \in \mathbb{R}^n$ observe that

$$\begin{aligned}
\mathbb{E}\left(\exp\left(i\left\langle\lambda,\sum_{i=1}^n a_iX_i\right\rangle\right)\right) &= \mathbb{E}\left(\exp\left(i\sum_{i=1}^n \langle a_i\lambda, X_i\rangle\right)\right)\\
&= \mathbb{E}\left(\prod_{i=1}^n \exp\left(i\left\langle a_i\lambda, X_i\right\rangle\right)\right)\\
&\overset{(1)}{=} \prod_{i=1}^n \mathbb{E}\left(\exp(i\langle a_i\lambda, X_i\rangle)\right)\\
&= \prod_{i=1}^n \exp\left(i\langle a_i\lambda, m_i\rangle - \frac{1}{2}\langle K_i(a_i\lambda), a_i\lambda\rangle\right)\\
&= \exp\left(\sum_{i=1}^n i\langle\lambda, a_im_i\rangle - \frac{1}{2}\left\langle\left(a_i^2K_i\right)\lambda,\lambda\right\rangle\right)\\
&= \exp\left(i\left\langle\lambda,\sum_{i=1}^n a_im_i\right\rangle - \frac{1}{2}\left\langle\sum_{i=1}^n a_i^2K_i\lambda,\lambda\right\rangle\right).
\end{aligned}$$

Where (1) is justified by the independence of the X_i. Hence, $\sum_{i=1}^n a_iX_i$ is a Gaussian measure with mean $\sum_{i=1}^n a_im_i$ and covariance operator $\sum_{i=1}^n a_i^2K_i$. □

9.2 The Doeblin Coupling

Definition 9.2.1. *For random variables X and Y with state space $\mathcal{X}$, a coupling is a random variable $Z = (X', Y')$ with state space $\mathcal{X}\times\mathcal{X}$ such that*

1. $\mathrm{Law}(X) = \mathrm{Law}\left(X'\right)$, *and*
2. $\mathrm{Law}(Y) = \mathrm{Law}\left(Y'\right)$.

Remark 9.2.2. *Note that for a given X and Y there can be lots of different couplings. This is because $\mathrm{Law}(X)$ and $\mathrm{Law}(Y)$ do not determine $\mathrm{Law}(Z)$, but only determine its marginals.*

Definition 9.2.3. *Consider independent Markov chains $(X_n)_{n\in\mathbb{N}}$ and $(X_n')_{n\in\mathbb{N}}$ on a discrete state space $\mathcal{X}$, with transition probabilities P, and initial distributions μ and ν respectively. The process $(Z_n)_{n\in\mathbb{N}}$ given by $Z_n = (X_n, X_n')$ is known as the Doeblin coupling. Moreover, $T := \inf\{n \geq : X_n = X_n'\}$ is known as the coalescing time of the chains $(X_n)_{n\in\mathbb{N}}$ and $(X_n')_{n\in\mathbb{N}}$.*

Exercise 9.2.4. *Let $(\mathcal{G}_n)_{n\in\mathbb{N}}$ and $(\mathcal{G}_n')_{n\in\mathbb{N}}$ be filtrations of independent σ-algebras. Let $(X_n)_{n\in\mathbb{N}}$ be a time-homogeneous Markov process with respect to the filtration $(\mathcal{G}_n)_{n\in\mathbb{N}}$. Then, $(X_n)_{n\in\mathbb{N}}$ is a time-homogeneous Markov process with respect to the filtration $(\mathcal{G}_n\vee\mathcal{G}_n')_{n\in\mathbb{N}}$, where $\mathcal{G}_n\vee\mathcal{G}_n'$ is the σ-algebra generated by $\mathcal{G}_n\cup\mathcal{G}_n'$.*

Lemma 9.2.5. *Assume the setup of Definition 9.2.3 and let*

$$Y_n = \begin{cases} X_n & n < T\\ X_n' & n \geq T.\end{cases}$$

Then $(Y_n)_{n\in\mathbb{N}}$ is a Markov process with initial distribution μ and transition probabilities P. In particular,

$\mathrm{Law}(Y) = \mathrm{Law}(X)$.

Proof. Let $\left(\mathcal{F}_n^0\right)_{n\in\mathbb{N}}$ and $\left(\mathcal{G}_n^0\right)_{n\in\mathbb{N}}$ be the natural filtrations of $(X_n)_{n\in\mathbb{N}}$ and $(X_n')_{n\in\mathbb{N}}$ respectively. Then let $\mathcal{F}_n = \mathcal{F}_n^0 \vee \mathcal{G}_n^0$. For $f \in \mathcal{B}_b(\mathcal{X})$ it follows that

$$\begin{aligned}\mathbb{E}(f(Y_{n+1})|\mathcal{F}_n) &= \mathbb{E}\left(f(Y_{n+1})\mathbf{1}_{\{T\le n\}}|\mathcal{F}_n\right) + \mathbb{E}\left(f(Y_{n+1})\mathbf{1}_{\{T>n\}}|\mathcal{F}_n\right)\\ &= \mathbf{1}_{\{T\le n\}}\mathbb{E}\left(f\left(X_{n+1}'\right)|\mathcal{F}_n\right) + \mathbf{1}_{\{T>n\}}\mathbb{E}(f(X_{n+1})|\mathcal{F}_n)\\ &= \mathbf{1}_{\{T\le n\}}Pf\left(X_n'\right) + \mathbf{1}_{\{T>n\}}Pf(X_n)\\ &= \mathbf{1}_{\{T\le n\}}Pf(Y_n) + \mathbf{1}_{\{T>n\}}Pf(Y_n)\\ &= Pf(Y_n).\end{aligned}$$

Where we have implicitly used the result of Exercise 9.2.4. Furthermore, we have used the fact that T is a $(\mathcal{F}_n)_{n\in\mathbb{N}}$-stopping time, and hence $\{T \le n\}$ and $\{T > n\}$ are $\mathcal{F}_n$-measurable. Hence, $(Y_n)_{n\in\mathbb{N}}$ is a Markov process with transition probability P. Moreover, it is clear that $Y_0 = X_0$, as if $T = 0$ then $Y_0 = X_0' = X_0$ by the definition of T. So that $\mathrm{Law}(Y_0) = \mathrm{Law}(X_0) = \mu$, which implies that $\mathrm{Law}(Y) = \mathrm{Law}(X)$ as $(Y_n)_{n\in\mathbb{N}}$ and $(X_n)_{n\in\mathbb{N}}$ have the same transition probability. □

Lemma 9.2.6. *Assume the setup of Definition 9.2.3, then*

$$\sum_{j\in\mathcal{X}} \left|\mathbb{P}(X_n = j) - \mathbb{P}\left(X_n' = j\right)\right| \le 2\mathbb{P}(T > n).$$

Proof. Let $j \in \mathcal{X}$ then

$$\begin{aligned}\left|\mathbb{P}(X_n = j) - \mathbb{P}\left(X_n' = j\right)\right| &\overset{(1)}{=} \left|\mathbb{P}(Y_n = j) - \mathbb{P}\left(X_n' = j\right)\right|\\ &= \left|\mathbb{P}(Y_n = j) - \mathbb{P}\left(X_n' = j, T > n\right) - \mathbb{P}\left(y_n = j, T \le n\right)\right|\\ &\overset{(2)}{=} \mathbb{P}\left(Y_n = j, T > n\right) - \mathbb{P}\left(Y_n' = j, T > n\right).\end{aligned}$$

Where (1) follows by the result of Lemma 9.2.5 and (2) is an application of the law of total probability. Therefore,

$$\begin{aligned}\sum_{j\in\mathcal{X}} \left|\mathbb{P}(X_n = j) - \mathbb{P}\left(X_n' = j\right)\right| &\le \sum_{j\in\mathcal{X}} \mathbb{P}(Y_n = j, T > n) + \mathbb{P}\left(X_n' = j, T > n\right)\\ &\le 2\mathbb{P}(T > n).\end{aligned}$$

□

Exercise 9.2.7. *Let $(X_n)_{n\in\mathbb{N}}$ and $(X_n')_{n\in\mathbb{N}}$ be independent Markov chains on a discrete state space $\mathcal{X}$ and transition probabilities P. Let $\mathcal{G}_n = \sigma(X_k : k \le n)$, $\mathcal{G}_n' = \sigma\left(X_k' : k \le n\right)$ and $\mathcal{F}_n = \mathcal{G}_n \vee \mathcal{G}_n'$. Show that*

$$\mathbb{P}\left(X_{n+1} = j, X_{n+1}' = j'|\mathcal{F}_n\right) = \mathbb{P}(X_{n+1} = j|X_n) \cdot \mathbb{P}\left(X_{n+1}' = j'|X_n'\right). \tag{9.2.1}$$

Lemma 9.2.8. *Assume the setup of Definition 9.2.3, then the Doeblin coupling Z_n is a time-homogeneous Markov chain on $\mathcal{X} \times \mathcal{X}$ with transition probabilities Q and initial distribution $\mu \otimes \nu$, where*

$$Q_{(i,i'),(j,j')} = P_{ij}P_{i'j'}$$

for all $i, i', j, j' \in \mathcal{X}$.

Proof. By the independence of X_0 and X_0' it follows immediately that $\mathrm{Law}(X_0, X_0') = \mu \otimes \nu$. With $(\mathcal{F}_n)_{n\in\mathbb{N}}$ as defined in the proof of Lemma 9.2.5, for $j, j' \in \mathcal{X}$ it follows that

$$\begin{aligned}
\mathbb{P}(Z_{n+1} = (j, j') | \mathcal{F}_n) &= \mathbb{P}\left(X_{n+1} = j, X_{n+1}' = j' | \mathcal{F}_n\right) \\
&\overset{\text{Ex 9.2.7}}{=} \mathbb{P}(X_{n+1} = j | X_n) \cdot \mathbb{P}\left(X_{n+1}' = j | X_n'\right) \\
&= P_{X_n, j} P_{X_n', j'} \\
&=: Q_{Z_n, (j,j')}.
\end{aligned}$$

Therefore, $\mathbb{P}(Z_{n+1}(j, j') | \mathcal{F}_n) = \mathbb{P}(Z_{n+1}(j, j') | Z_n)$ and so $(Z_n)_{n\in\mathbb{N}}$ is a time-homogeneous Markov process with transition probability Q. ☐

Lemma 9.2.9. *Suppose that $S \subset \mathbb{N}$ is a non-empty set with the property that for $s, s' \in S$ it follows that $s + s' \in S$. Then letting $d = \gcd(S)$, there must exists a $K > 0$ such that for any $k \geq K$ one has $kd \in S$.*

Proof. By considering $S' = \{\frac{s}{d} : s \in S\}$ we can assume without loss of generality that $d = 1$. Let $\{d_1, \ldots, d_n\} \subseteq S$ be such that $\gcd(\{d_1, \ldots, d_n\}) = 1$. By Bezout's identity we know that there exists integers $a_1, \ldots, a_n$ such that

$$\sum_{i=1}^{n} a_i d_i = 1.$$

Let $M = \sum_{i=1}^{n} d_i$. Then for $l = 0, \ldots, M-1$ it follows that

$$NM + l = \sum_{i=1}^{n} (N + l a_i) d_i.$$

As $k \leq M$ we can choose N_0 such that $N_0 + l a_i \geq 0$. Therefore, the sum on the right-hand side can be thought of as summing $N + l a_i$ copies of d_i for each i. Hence, by the additive property of S it follows that $NM + l \in S$ for all $l = 0, \ldots, M-1$ and $N \geq N_0$. Letting $l = 0$ we deduce that $kd = k \in S$ for every $k \geq N_0 M$. ☐

Lemma 9.2.10. *Suppose i is aperiodic and recurrent, then there exists an N such that $P_{ii}^n > 0$ for every $n > N$.*

Proof. As i is recurrent we know that $R(i) \neq \emptyset$. Moreover, if $n_1 \in R(i)$ and $n_2 \in R(i)$ then $P_{ii}^{n_1} > 0$ and $P_{ii}^{n_2} > 0$ and so $P_{ii}^{n_1+n_2} > 0$ by the Chapman-Kolomogorov equation. This implies that $n_1 + n_2 \in R(i)$ and so $R(i)$ has the additive property. Consequently, as i is aperiodic we have that $\gcd(R(i)) = 1$ and so by Lemma 9.2.9 that there exists an $N \in \mathbb{N}$ such that $n \in R(i)$ for all $n \geq N$. In other words, $P_{ii}^n > 0$ for $n \geq N$. ☐

Lemma 9.2.11. *If P is irreducible, aperiodic, and positive recurrent, then Q, as defined in Lemma 9.2.8, is irreducible and positive recurrent.*

Proof. For any $i \in \mathcal{X}$ we know by Lemma 9.2.10 that there exists an N such that $P_{ii}^n > 0$ for any $n \geq N$. Thus by the irreducibility of P we know that for any $j \in \mathcal{X}$ there exists an m such that $P_{ij}^m > 0$. Consequently,

$$P_{ij}^{n+m} \geq P_{ii}^n P_{ij}^m > 0$$

for all $n \geq N$. More succinctly, we can say that $P_{ij}^n > 0$ for sufficiently large n. Therefore, for $(i, j), (i', j') \in \mathcal{X} \times \mathcal{X}$ there exists some $N \in \mathbb{N}$ such that $P_{ij}^n > 0$ and $P_{i'j'}^n > 0$ which implies that

$$Q_{(i,i'),(j,j')}^n = P_{ij}^n P_{i'j'}^n > 0$$

for all $n \geq N$. Which proves that Q is irreducible. Consequently, with π being the invariant measure of P we know that $\pi \otimes \pi$ is an invariant measure of Q. Having established that Q is irreducible this implies that Q is positive recurrent. ☐

Lemma 9.2.12. *Assume the setup of Definition9.2.3 and in particular that P is irreducible, aperiodic and positive recurrent. Then,*

$$\mathbb{P}(T < \infty) = 1.$$

Proof. Let

$$T_{(i,i')} := \inf\left\{n \geq 1 : Z_n\left(X_n, X_n'\right) = (i, i')\right\}.$$

As Q is irreducible we know that

$$\mathbb{P}\left(T_{(i,i')} < \infty\right) = 1$$

for all $(i, i') \in \mathcal{X} \times \mathcal{X}$. In particular, letting $i = i'$ it is clear that $T \leq T_{(i,i)}$, hence,

$$\mathbb{P}(T < \infty) \geq \mathbb{P}\left(T_{(i,i)} < \infty\right) = 1.$$

□

9.3 Solution to Exercises

Exercise 9.2.4

Solution. Consider the set $\mathcal{D} = \{A \cap B : A \in \mathcal{G}_n, B \in \mathcal{G}_n'\}$. It is clear that $\mathcal{D}$ defines a π-system. Moreover, it is clear that $\mathcal{G} := \mathcal{G}_n \vee \mathcal{G}_n'$ forms a λ-system. So by the Dynkin π-λ theorem we know that $\sigma(\mathcal{D}) \subset \mathcal{G}$. However, as $\mathcal{G}_n \subseteq \mathcal{D}$ and $\mathcal{G}_n' \subseteq \mathcal{D}$, by taking $A = \mathcal{X}$ and $B = \mathcal{X}$ respectively, we know that $\mathcal{G}_n \cup \mathcal{G}_n' \subseteq \sigma(\mathcal{D})$ which implies that $\mathcal{G} = \sigma\left(\mathcal{G}_n \cup \mathcal{G}_n'\right) \subseteq \sigma(\mathcal{D})$, hence $\sigma(\mathcal{D}) = \mathcal{G}$. Now for $f \in \mathcal{B}_b(\mathcal{X})$ and $n \geq 0$ we know that $\mathbb{E}\left(f(X_{n+1})|\mathcal{G}_n\right)$ is $\mathcal{G}$ measurable. Moreover, for $A \cap B \in \mathcal{D}$ it follows that

$$\begin{aligned}\int_{A\cap B} \mathbb{E}\left(f(X_{n+1})|\mathcal{G}_n\right)\,\mathrm{d}\mathbb{P} &= \int \mathbb{E}\left(f(X_{n+1})\mathbf{1}_{A\cap B}|\mathcal{G}_n\right)\,\mathrm{d}\mathbb{P}\\ &\overset{(1)}{=} \int \mathbb{E}(f(X_{n+1})\mathbf{1}_{A\cap B})\,\mathrm{d}\mathbb{P}\\ &= \int_{A\cap B} f(X_{n+1})\,\mathrm{d}\mathbb{P},\end{aligned}$$

where in (1) we have used the fact that $f(X_{n+1})\mathbf{1}_{A\cap B}$ is independent of $\mathcal{G}_n$. Therefore, using our observation that $\sigma(\mathcal{D}) = \mathcal{G}$ we deduce that for all $C \in \mathcal{G}$ that

$$\int_C \mathbb{E}(f(X_{n+1})|\mathcal{G}_n)\,d\mathbb{P} = \int_C f(X_{n+1})\,\mathrm{d}\mathbb{P}.$$

Hence, $\mathbb{E}\left(f(X_{n+1})|\mathcal{G}_n \vee \mathcal{G}_n'\right) = \mathbb{E}\left(f(X_{n+1})|\mathcal{G}_n\right)$. Therefore, as (X_n) is Markov with respect to $\mathcal{G}_n$ we conclude that

$$\begin{aligned}\mathbb{E}\left(f(X_{n+1})|\mathcal{G}_n \vee \mathcal{G}_n'\right) &= \mathbb{E}\left(f(X_{n+1})|\mathcal{G}_n\right)\\ &= \mathbb{E}(f(X_{n+1})|X_n)\end{aligned}$$

and so (X_n) is Markov with respect to $\mathcal{G}_n \vee \mathcal{G}_n'$. □

Exercise 9.2.7

Solution. Let $\mathcal{G}_n = \sigma(X_k : k \leq n)$, $\mathcal{G}_n' = \sigma\left(X_k' : k \leq n\right)$ so that $\mathcal{F}_n = \mathcal{G}_n \vee \mathcal{G}_n'$. The right-hand side of (9.2.1) is equal to $\mathbb{P}(X_n = j|\mathcal{G}_n) \cdot \mathbb{P}\left(X_{n+1}' = j'|\mathcal{G}_n'\right)$ by the Markov property and so is clearly $\mathcal{F}_n$-measurable. Recall, that $\mathcal{F}_n$ is generated by sets of the form $G_1 \cap G_2$ where $G_1 \in \mathcal{G}_n$ and $G_2 \in \mathcal{G}'_n$. Using the independence of $\mathcal{G}_n$ and $\mathcal{G}'_n$ it follows that,

$$\begin{aligned}
\mathbb{E}\left(\mathbb{P}(X_{n+1} = j|\mathcal{G}_n) \cdot \mathbb{P}\left(X'_{n+1} = j'|\mathcal{G}'_n\right) \mathbf{1}_{G_1 \cap G_2}\right) &= \mathbb{E}\left(\mathbb{E}\left(\mathbf{1}_{\{X_{n+1}=j\}} \mathbf{1}_{G_1} | \mathcal{G}_n\right) \mathbb{E}\left(\mathbf{1}_{\{X'_{n+1}=j'\}} \mathbf{1}_{G_2} | \mathcal{G}'_n\right)\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(\mathbf{1}_{\{X_{n+1}=j\}} \mathbf{1}_{G_1} | \mathcal{G}_n\right)\right) \mathbb{E}\left(\mathbb{E}\left(\mathbf{1}_{\{X'_{n+1}=j'\}} \mathbf{1}_{G_2} | \mathcal{G}'_n\right)\right) \\
&= \mathbb{E}\left(\mathbf{1}_{\{X_{n+1}=j\}} \mathbf{1}_{G_1}\right) \mathbb{E}\left(\mathbf{1}_{\{X'_{n+1}=j'\}} \mathbf{1}_{G_2}\right) \\
&= \mathbb{E}\left(\mathbf{1}_{\{X_{n+1}=j\} \cap \{X'_{n+1}=j'\}} \mathbf{1}_{G_1 \cap G_2}\right) \\
&= \mathbb{E}\left(\mathbb{P}\left(X_{n+1} = j, X'_{n+1} = j'\right) \mathbf{1}_{G_1 \cap G_2}\right).
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathbb{P}\left(X_{n+1} = j, X'_{n+1} = j'\right) &= \mathbb{P}(X_{n+1} = j|\mathcal{G}_n) \cdot \mathbb{P}\left(X'_{n+1} = j'|\mathcal{G}'_n\right) \\
&= \mathbb{P}(X_{n+1} = j|X_n) \cdot \mathbb{P}\left(X'_{n+1} = j'|X'_n\right).
\end{aligned}$$

$\square$