Markov Process

Kexing Ying

February 2, 2022

Contents

1 Introduction and Review

We will in this course assume the following notation:

- $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space;
- $\mathcal{X}$ is a Polish space, i.e. a separable, completely metrizable, topological space;
- $\mathcal{B}(\mathcal{X})$ is the Borel σ-algebra of $\mathcal{X}$.

Definition 1.1 (Stochastic Process)**.** A stochastic process $(x_n)_{n \in I}$ is a collection of random variables. In the case that $I = \mathbb{N}$ or $\mathbb{Z}$, we say that the stochastic process is discrete time. On the other hand if $I = \mathbb{R}_{\geq 0}$ or $[0,1] \subseteq \mathbb{R}$, then we say the process is continuous time.

We recall some definitions from elementary probability theory.

Definition 1.2 (Random Variable)**.** A random variable $x : \Omega \to \mathcal{X}$ is simply a measurable function.

Definition 1.3 (Probability Distribution)**.** Given a random variable $x : \Omega \to \mathcal{X}$, the probability distribution of x, denoted by $\mathcal{L}(x)$ is the push-forward measure of $\mathbb{P}$ along x, i.e.

$$\mathcal{L}(x) = x_* \mathbb{P} : A \in \mathcal{F} \mapsto \mathbb{P}(x^{-1}(A)).$$

Proposition 1.1. Let $x : \Omega \to \mathcal{X}$ be a random variable where $\mathcal{X}$ is countable, then

$$\mathcal{L}(x) = \sum_{i \in X} \mathbb{P}(x = i)\delta_i := \sum_{i \in X} x_* \mathbb{P}(\{i\})\delta_i$$

where δ_i is the Dirac measure concentrated at i.

Proof. Let $A \subseteq X$, then

$$\mathcal{L}(x)(A) = \sum_{i \in A} \mathcal{L}(x)(\{i\}) = \sum_{i \in X} \mathcal{L}(x)(\{i\})\delta_i(A) = \sum_{i \in X} x_* \mathbb{P}(\{i\})\delta_i(A),$$

as required. $\square$

Definition 1.4 (Independence)**.** Given random variables $x_1, \cdots, x_n$, we say $x_1, \cdots, x_n$ are independent if

$$\mathcal{L}((x_1, \cdots, x_n)) = \bigotimes_{i=1}^{n} \mathcal{L}(x_i),$$

where $\otimes$ denotes the product measure.

As the name suggests, we will in this course mostly focus on a class of stochastic processes known as Markov processes. These are processes in which given information about the process at the present time, its future is independent from its history. In particular, if (x_n) is a Markov process, given its value at x_k, the value of x_j is independent of the values of x_i for all $i < k < j$.

Definition 1.5 (Invariant Probability Measure)**.** A probability measure π is said to be an invariant probability measure or an invariant distribution of a Markov process $(x_n)_{n \in I}$ if for all $n \in I$, we have $\pi = \mathcal{L}(x_n)$.

A Markov chain started from an invariant distribution does is called a stationary Markov process as its distribution do not evolve and we say that the chain is in equilibrium.

In this course we will study the behaviour of the distribution of Markov processes. In particular, we ask

- does there exists an invariant measure? If so, is it unique?
- how does the distribution evolve over time?
- does $\mathcal{L}(x_n)$ converge as $n \to \infty$ (convergence in distribution)?

2 Markov Property

Let us now consider the Markov property in a more formal context.

2.1 Filtration and Simple Markov Property

In formation and filtration is an important notion not only for Markov processes but for stochastic processes in general.

In formally, the information of a random variable x is the collection of all possible events, i.e. the sigma algebra generated by x,

$$\sigma(x) = \sigma(\{x^{-1}(A) \mid A \in \mathcal{B}(\mathcal{X})\}).$$

In the case of a stochastic process (x_n), the information on the process up to time n is the σ-algebra generated by $x_0, \cdots, x_n$, i.e. $\sigma(x_0, \cdots, x_n)$.

With this in mind, we see that the notion of possible events evolving in time is naturally described by a sequence of increasing σ-algebras. We call such a sequence a filtration.

Definition 2.1 (Filtration)**.** A filtration is a sequence $(\mathcal{F}_n)$ of increasing sub-σ-algebras of $\mathcal{F}$.

Definition 2.2 (Adapted)**.** A stochastic process (x_n) is adapted to the filtration $(\mathcal{F}_n)$ if for all n, x_n is $\mathcal{F}_n$-measurable.

Definition 2.3 (Natural Filtration)**.** Given a stochastic process (x_n), the natural filtration $(\mathcal{F}^x_n)$ for (x_n) is

$$\mathcal{F}^x_n := \sigma(x_0, \cdots, x_n).$$

We note that by definition, a stochastic process is always adapted to its natural filtration.

Recalling the definition of conditional expectation, we introduce the following notations.

Definition 2.4 (Conditional Probability)**.** Given a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ and a random variable x, we define the conditional probability of x with respect to $\mathcal{G}$ to be

$$\mathbb{P}(x \in A \mid \mathcal{G}) := \mathbb{E}(\mathbf{1}_A(X) \mid \mathcal{G}),$$

for all $A \in \mathcal{B}(\mathcal{X})$ where $\mathbf{1}_A$ is the indicator function of A.

Furthermore, given random variables $x_0, \cdots, x_n$, we denote

$$\mathbb{P}(x \in A \mid x_0, \cdots, x_n) := \mathbb{P}(x \in A \mid \sigma(x_0, \cdots, x_n)).$$

Definition 2.5 (Simple Markov Property)**.** A stochastic process (x_n) with state space $\mathcal{X}$ is said to have the simple Markov property if for any $A \in \mathcal{B}(\mathcal{X})$ and $n \geq 0$, we have

$$\mathbb{P}(x_{n+1} \in A \mid x_0, \cdots, x_n) = \mathbb{P}(x_{n+1} \in A \mid x_n),$$

almost surely.

Unfolding the notation, the simple Markov property states that

$$\mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid \mathcal{F}^x_n) = \mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid \sigma(x_n)).$$

We call a stochastic process which has the simple Markov property a Markov process and we call $\mathcal{L}(x_0)$ the initial distribution. Furthermore, if the Markov process is discrete, we call it a Markov chain.

The definition of the simple Markov property can be generalized to continuous stochastic processes by taking the property to be $\mathbb{E}(\mathbf{1}_A(x_t) \mid \mathcal{F}_s^x) = \mathbb{E}(\mathbf{1}_A(x_t) \mid \sigma(x_s))$ for all $s \leq t$.

In the case that $\mathcal{X} = \mathbb{N}$, the simple Markov property is equivalent to the statement that

$$\mathbb{P}(x_{n+1} = j \mid x_0 = i_0, \cdots, x_n = i_n) = \mathbb{P}(x_{n+1} = j \mid x_n = i_n),$$

almost surely for every n where $i_0, \cdots, i_n \in \mathcal{X} = \mathbb{N}$

$$\mathbb{P}(x_0 = i_0, \cdots, x_n = i_n) > 0.$$

Lemma 2.1. Let $\mathcal{G} \subseteq \mathcal{F}$, $X : \Omega \to \mathcal{X}, Y : \Omega \to \mathcal{Y}$ be random variables such that X is $\mathcal{G}$-measurable, Y is independent of $\mathcal{G}$. Then, if $\phi : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is measurable such that $\phi(X, Y) \in L^1$, we have

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

almost surely.

Proof. Exercise. $\square$

Proposition 2.1. Let $\xi_1, \xi_2, \cdots$ be a sequence of independent random variables with state space $\mathcal{Y}$ and is independent with respect to $x_0 : \Omega \to \mathcal{X}$. Then, if $F : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ is a measurable function, we may define the stochastic process

$$x_{n+1} = F(x_n, \xi_{n+1}).$$

(x_n) is a Markov process.

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

$$\begin{aligned}\mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid x_0, \cdots, x_n) &= \mathbb{E}(\mathbf{1}_A(F(x_n, \xi_{n+1}) \mid x_0, \cdots, x_n) \\ &= \omega \mapsto \mathbb{E}(\mathbf{1}_A(F(x_n(\omega), \xi_{n+1})),\end{aligned}$$

where the second equality follows by the above lemma (setting $\phi = \mathbf{1}_A \circ F$ and observing that x_n is $\sigma(x_0, \cdots, x_n)$-measurable and ξ_{n+1} is independent of $\sigma(x_0, \cdots, x_n)$). Similarly,

$$\begin{aligned}\mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid x_n) &= \mathbb{E}(\mathbf{1}_A(F(x_n, \xi_{n+1}) \mid x_n) \\ &= \omega \mapsto \mathbb{E}(\mathbf{1}_A(F(x_n(\omega), \xi_{n+1})),\end{aligned}$$

we have $\mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid \mathcal{F}_n^x) = \mathbb{E}(\mathbf{1}_A(x_{n+1}) \mid \sigma(x_n))$ as required. $\square$

2.2 Markov Property

So far we have looked at the simple Markov property in which we have taken the filtration to be the natural filtration of the process. However, in the case that we are looking at multiple processes, we would like to consider a larger filtration such that each process is adapted. This motivates the general definition for the Markov property.

Definition 2.6. Let $(\mathcal{F}_t)_{t\in I}$ be a filtration indexed by the set I on the measurable space $(\Omega, \mathcal{F})$. A stochastic process $(x_t)_{t\in I}$ on $\mathcal{X}$ is a Markov process with respect to $\mathcal{F}_t$ if it is adapted to $\mathcal{F}_t$ and

$$\mathbb{P}(x_t \in A \mid \mathcal{F}_s) = \mathbb{P}(x_t \in A \mid x_s)$$

almost surely for all $s, t \in I$, $t > s$ and $A \in \mathcal{B}(\mathcal{X})$.

Again, unfolding the notation, the above statement says

$$\mathbb{E}(\mathbf{1}_A(x_t) \mid \mathcal{F}_s) = \mathbb{E}(\mathbf{1}_A(x_t) \mid \sigma(x_s))$$

almost surely.

Proposition 2.2. If (x_t) is a Markov process with respect to the filtration $(\mathcal{F}_t)$, then it is also a Markov process with respect to its natural filtration $(\mathcal{F}^x_t)$.

Proof. Recalling that $\mathcal{F}^x_t \subseteq \mathcal{F}_t$ for all t, by the tower property of the conditional expectation, we have

$$\begin{aligned}
\mathbb{P}(x_{t+s} \in A \mid \mathcal{F}^x_s) &= \mathbb{E}(\mathbf{1}_A(x_{t+s}) \mid \mathcal{F}^x_s) \\
&= \mathbb{E}(\mathbb{E}(\mathbf{1}_A(x_{t+s}) \mid \mathcal{F}_s) \mid \mathcal{F}^x_s) \\
&= \mathbb{E}(\mathbb{E}(\mathbf{1}_A(x_{t+s}) \mid \sigma(x_s)) \mid \mathcal{F}^x_s),
\end{aligned}$$

where the equalities denotes equal a.e. Thus, as $\mathbb{E}(\mathbf{1}_A(x_{t+s}) \mid \sigma(x_s))$ is $\sigma(x_s)$-measurable, and thus $\mathcal{F}^x_s$-measurable (since $\sigma(x_s) \subseteq \sigma(x_r \mid r \leq s) = \mathcal{F}^x_s$)), we have

$$\mathbb{E}(\mathbb{E}(\mathbf{1}_A(x_{t+s}) \mid \sigma(x_s)) \mid \mathcal{F}^x_s) = \mathbb{E}(\mathbf{1}_A(x_{t+s}) \mid \sigma(x_s))$$

implying that the Markov property is satisfied. □

Theorem 1. If (x_t) is a Markov process with respect to the filtration $(\mathcal{F}_t)$, then

$$\mathbb{E}(f(x_t) \mid \mathcal{F}_s) = \mathbb{E}(f(x_t) \mid \sigma(x_s))$$

almost surely for any $f : \mathcal{X} \to \mathbb{R}$ bounded and measurable. In particular, this property is equivalent to the Markov property by choosing $f = \mathbf{1}_A$ for all $A \in \mathcal{B}(\mathcal{X})$.

Proof. By linearity, the property holds for simple functions. Furthermore, by the conditional monotone convergence theorem, the property holds for any non-negative bounded measurable functions. Finally, for arbitrary bounded measurable functions f, the result follows by taking $f = f^+ - f^-$ and applying the non-negative case. □

Proposition 2.3. Let $C \in \mathcal{F}_s$ and suppose $\mathcal{B}(\mathcal{X}) = \sigma(\mathcal{D})$ where $\mathcal{D}$ is a π-system (i.e. non-empty and closed under finite intersections), then, if

$$\mathbb{E}(\mathbf{1}_A(x_{t+s})\mathbf{1}_C) = \mathbb{E}(\mathbb{P}(x_{t+s} \in A \mid x_s)\mathbf{1}_C)$$

holds for any $A \in \mathcal{D}$, it holds for any $A \in \mathcal{B}(\mathcal{X})$.

Proof. Let $\mathcal{A}$ be the set of Borel sets which the equation holds. Then, by definition $\mathcal{D} \subseteq \mathcal{A}$ and so, it suffices to show $\mathcal{A}$ is a λ-system (i.e. $\mathcal{A}$ contains $\mathcal{X}$, closed under complements and closed under countable unions of increasing sets). Indeed, Dynkin's $\pi - \lambda$ theorem states that if $\mathcal{D}$ is a π-system, $\mathcal{A}$ is a λ-system and $\mathcal{D} \subseteq \mathcal{A}$, then $\sigma(\mathcal{D}) \subseteq \mathcal{A}$.

Clearly $\mathcal{X} \in \mathcal{A}$ since

$$\mathbb{E}(\mathbf{1}_{\mathcal{X}}(x_{t+s})\mathbf{1}_C) = \mathbb{E}(\mathbf{1}_C) = \mathbb{E}(\mathbb{P}(x_{t+s} \in \mathcal{X} \mid x_s)\mathbf{1}_C).$$

Suppose now $A \in \mathcal{A}$. Then, the property holds as $\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$ and so, the result follows by linearity. Finally, if $(A_n) \subseteq \mathcal{A}$ is increasing. Then by the monotone convergence theorem for conditional expectations, it follows that $\bigcup A_n \in \mathcal{A}$ and hence, $\mathcal{A}$ is a λ-system as required. $\square$

Proposition 2.4. Suppose $\mathbb{E}(f(x_{n+1}) \mid x_0, \cdots, x_n) = \mathbb{E}(f(x_{n+1}) \mid x_n)$ for any bounded measurable f. Then, if we have a sequence

$$0 \leq t_1 < t_2 < \cdots < t_{m-1} < t_m = n - 1,$$

where $n > 1, t_i \in \mathbb{N}$, for any bounded measurable functions f, h, we have

$$\mathbb{E}(f(x_{n+1})h(x_n) \mid x_{t_1}, \cdots, x_{t_m}) = \mathbb{E}(f(x_{n+1})h(x_n) \mid x_{n-1}).$$

Proof. Exercise. $\square$

As we will often use the bounded measurable functions, let us denote the set of bounded measurable functions $f : \mathcal{X} \to \mathbb{R}$ by $\mathcal{B}_b(\mathcal{X})$.

Lemma 2.2. Let $X, Y \in L_1$ and $\mathcal{G} \subseteq \mathcal{F}$. Then, if X is $\mathcal{G}$-measurable and $XY \in L_1$, we have

$$\mathbb{E}(XY \mid \mathcal{G}) = X\mathbb{E}(Y \mid \mathcal{G}).$$

We call this property "taking out what is known".

Proof. See problem sheet 1. $\square$

Theorem 2. Given a stochastic process (x_n) and indices $l < m < n$, TFAE.

- For any $f \in \mathcal{B}_b(\mathcal{X})$,
$$\mathbb{E}(f(x_n) \mid x_l, x_m) = \mathbb{E}(f(x_n) \mid x_m).$$
- For any $g \in \mathcal{B}_b(\mathcal{X})$,
$$\mathbb{E}(g(x_l) \mid x_m, x_n) = \mathbb{E}(g(x_l) \mid x_m).$$
- For any $f, g \in \mathcal{B}_b(\mathcal{X})$,
$$\mathbb{E}(f(x_n)g(x_l) \mid x_m) = \mathbb{E}(f(x_n) \mid x_m)\mathbb{E}(g(x_l) \mid x_m).$$

  That is to say, given now, the past is independent of the future.

Proof. Suppose the first statement holds, we will prove the third property. Let $f, g \in \mathcal{B}_b(\mathcal{X})$, then by the tower law and the above lemma, we have

$$\begin{aligned}\mathbb{E}(f(x_n)g(x_l) \mid x_m) &= \mathbb{E}(\mathbb{E}(f(x_n)g(x_l) \mid x_m, x_l) \mid x_m)\\ &= \mathbb{E}(g(x_l)\mathbb{E}(f(x_n) \mid x_m, x_l) \mid x_m)\\ &= \mathbb{E}(g(x_l)\mathbb{E}(f(x_n) \mid x_m) \mid x_m)\\ &= \mathbb{E}(f(x_n) \mid x_m)\mathbb{E}(g(x_l) \mid x_m)\end{aligned}$$

which is exactly the third property.

On the other hand, if the third property holds, for any $g, h \in \mathcal{B}_b(\mathcal{X})$, we have

$$\begin{aligned}\mathbb{E}(f(x_n)h(x_m)g(x_l)) &= \mathbb{E}(\mathbb{E}(f(x_n)g(x_l) \mid x_m))h(x_m))\\ &= \mathbb{E}(\mathbb{E}(f(x_n) \mid x_m))\mathbb{E}(g(x_l) \mid x_m)h(x_m))\\ &= \mathbb{E}(\mathbb{E}(\mathbb{E}(f(x_n) \mid x_m)g(x_l)h(x_m) \mid x_m))\\ &= \mathbb{E}(\mathbb{E}(f(x_n) \mid x_m)g(x_l)h(x_m))\end{aligned}$$

where the last equality is due to the law of total expectation. Now, by considering this equality implies that, for all $A = A_1 \cap A_2$ where $A_1 \in \sigma(x_l), A_2 \in \sigma(x_m)$, by choosing $g = \mathbf{1}_{x_l^{-1}(A_1)}$ and $h = \mathbf{1}_{x_m^{-1}(A_2)}$, we have

$$\int_A f(x_n)\mathrm{d}\mathbb{P} = \int_A \mathbb{E}(f(x_n) \mid x_m)\mathrm{d}\mathbb{P},$$

and so, $\mathbb{E}(f(x_n) \mid x_m) = \mathbb{E}(f(x_n) \mid x_l, x_m)$ almost surely. Hence, the first and third property are equivalent. Similarly, one can should that the second property is equivalent to the third property and hence the equivalence. □

Proposition 2.5. A stochastic process (x_n) is a Markov process if and only if one of the following conditions holds:

- for any $f_i \in \mathcal{B}_b(\mathcal{X})$,
$$\mathbb{E}\left(\prod_{i=1}^n f_i(x_i)\right) = \mathbb{E}\left(\prod_{i=1}^{n-1} f_i(x_i)\mathbb{E}(f_n(x_n) \mid x_{n-1})\right).$$
- for any $A_i \in \mathcal{B}(\mathcal{X})$,
$$\mathbb{P}(x_0 \in A_0, \cdots, x_n \in A_n) = \int_{\bigcap_{i=0}^{n-1}\{x_i \in A_i\}} \mathbb{P}(x_n \in A_n \mid x_{n-1})\mathrm{d}\mathbb{P}.$$

Proof. We note that by choosing $f_i = \mathbf{1}_{A_i}$, the first condition implies the second. On the other hand, the reverse implication follows by the standard routine of proving it for simple function and using monotone convergence. Thus, it suffices to establish an equivalence between the first condition and the Markov property. This is left as an exercise. □

2.3 Gaussian Measure and Gaussian Process

As one of the most important distributions in probability theory, let us in this short section introduce the Gaussian measure which we will again encounter later on with this course.

Definition 2.7 (Gaussian Measure)**.** A measure μ on $\mathbb{R}^n$ is Gaussian if there exists a non-negative definite symmetric matrix K and $m \in \mathbb{R}^n$ such that the Fourier transform of μ is

$$\int_{\mathbb{R}^n} e^{i\langle \lambda, x\rangle} \mu(\mathrm{d}x) = e^{i\langle \lambda, m\rangle - \frac{1}{2}\langle K\lambda, \lambda\rangle}$$

for any $\lambda \in \mathbb{R}^n$. We call the matrix K the covariance of μ and m its mean.

We remark that if X is a random variable with distribution μ, then the Fourier transform of μ is simply the characteristic function of X, $\mathbb{E}(e^{i\langle \lambda, X\rangle})$.

Proposition 2.6. If μ is a Gaussian measure with covariance K and mean m is absolutely continuous with respect to the Lebesgue measure if and only if K is non-degenerate. In this case, for all $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\mu(A) = \int_A \frac{1}{\sqrt{(2\pi)^n \det K}} e^{-\frac{1}{2}\langle K^{-1}(x-m), x-m\rangle} \lambda(\mathrm{d}x).$$

We observe that if X is a random variable with Gaussian distribution μ, then as one might expect, $\mathbb{E}(X) = m$ and

$$\mathrm{Cov}(X_i, X_j) := \mathbb{E}(X_i - m_i)(X_j - m_j) = K_{ij}.$$

For this reason, we call K the covariance operator.

Theorem 3. If X is a Gaussian random variable (i.e. its distribution is Gaussian) on $\mathbb{R}^d$ with covariance K. Then, if $A : \mathbb{R}^d \to \mathbb{R}^m$ is a linear transformation, then AX is Gaussian with covariance AKA^T and mean Am.

Proof. Follows since,

$$\mathbb{E}e^{i\langle \lambda, AX\rangle} = \mathbb{E}e^{i\langle A^T\lambda, X\rangle} = e^{i\langle A^T\lambda, m\rangle - \frac{1}{2}\langle KA^T\lambda, A^T\lambda\rangle} = e^{i\langle \lambda, Am\rangle - \frac{1}{2}\langle AKA^T\lambda, \lambda\rangle}.$$

□

Proposition 2.7. Linear combinations of independent Gaussian random variables are also Gaussian.

Proof. Exercise. □

Definition 2.8 (Gaussian Process)**.** A stochastic process is Gaussian if its finite dimensional distributions are Gaussian.

2.4 Kolmogorov's Extension Theorem

Let (x_n) be a stochastic process with state space $\mathcal{X}$. Denote

$$\mathcal{X}^{\mathbb{N}_0} := \prod_{i=0}^{\infty} \mathcal{X} = \{(a_0, a_1, \cdots) \mid a_i \in \mathcal{X}\}.$$

We may consider (x_n) as a map from $\Omega \to \mathcal{X}^{\mathbb{N}_0}$ by defining

$$(x_n) : \Omega \to \mathcal{X}^{\mathbb{N}_0} : \omega \mapsto (x_n(\omega))_{n=0}^{\infty}.$$

We would like (x_n) to be measurable as a map from $\Omega \to \mathcal{X}^{\mathbb{N}_0}$ and so, let us first equip $\mathcal{X}^{\mathbb{N}_0}$ with a σ-algebra.

Definition 2.9. Given $\mathcal{X}_i$ complete separable metric spaces for $i \in \Lambda$, define the projection maps

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

Then, we define $\bigotimes_{i \in \Lambda} \mathcal{B}(\mathcal{X}_i) = \sigma(\pi_i \mid i \in \Lambda)$.

We note that this definition can be easily extended to arbitrary measurable spaces.

Proposition 2.8. If (x_n) is a stochastic process, then as a map from $\Omega \to \mathcal{X}^{\mathbb{N}_0}$, (x_n) is $\bigotimes_n \mathcal{B}(\mathcal{X})$-measurable.

With this in mind, we can push forward the probability measure along a stochastic process, inducing a measure on $\mathcal{X}^{\mathbb{N}_0}$. In particular, we have the measure space $(\mathcal{X}^{\mathbb{N}_0}, \bigoplus_n \mathcal{B}(\mathcal{X}), (x_n)_* \mathbb{P})$.

On the other hand, if we only consider the first n-components of the process (x_i), by the same argument, $(x_i)_{i=1}^n$ forms a measurable map from $\Omega \to \mathcal{X}^n$. Then, in this case, we call the push-forward measure $\mathcal{L}((x_i)_{i=1}^n) = (x_i)_{i=1}^n{}_* \mathbb{P}$ the joint distribution. These are known as the finite dimensional distributions of (x_n).

Definition 2.10. Let $(\mu_n)_{n=1}^{\infty}$ be a sequence of probability measures on $\mathcal{X}^n$ (i.e. μ_n is aprobability measures on $\mathcal{X}^n$). Then $(\mu_n)_{n=0}^{\infty}$ is said to satisfy Kolmogorov's consistency condition of

$$\mu_{n+1}(A_1 \times \cdots A_n \times \mathcal{X}) = \mu_n(A_1 \times \cdots \times A_n)$$

for all $n \geq 0$, $A_i \in \mathcal{B}(\mathcal{X})$.

Theorem 4 (Kolmogorov's Extension Theorem)**.** Let $(\mu_n)_{n=1}^{\infty}$ be a sequence of probability measures which are consistent. Then, there exists a unique probability measure μ on $\mathcal{X}^{\mathbb{N}_0}$ such that for any n, $A \in \bigotimes_{i=1}^n \mathcal{B}(\mathcal{X})$,

$$\mu(A \times \mathcal{X}^{\mathbb{N}_0}) = \mu_n(A).$$

In other words, if we denote pr_n the map

$$\mathrm{pr}_n : \mathcal{X}^{\mathbb{N}_0} \to \mathcal{X}^n : (a_i)_{i=1}^{\infty} \mapsto (a_i)_{i=1}^n,$$

μ is the unique measure which satisfies

$$(\mathrm{pr}_n)_* \mu = \mu_n.$$

Corollary 4.1. The finite dimensional distributions of a stochastic process (x_n) determines uniquely the probability distribution of the process on $\mathcal{X}^{\mathbb{N}_0}$.

Corollary 4.2. Given any consistent family of probability measures (μ_n), there exists a stochastic process with (μ_n) as its finite dimensional distributions.

Proof. Kolmogorov's extension theorem implies that there exists a compatible measure μ on $\mathcal{X}^{\mathbb{N}_0}$. Thus, it suffices to find a $\mathcal{X}^{\mathbb{N}_0}$-valued random variable with distribution μ. We will describe a trivial method for this purpose below. $\square$

Let μ be a probability measure on $\mathcal{X}$. Then, setting $\Omega = \mathcal{X}, \mathcal{F} = \mathcal{B}(\mathcal{X})$ and $\mathbb{P} = \mu$, it is clear that the push-forward of $\mathbb{P}$ along the identity map provides μ. Thus, the identity is a random variable with the distribution μ.

Thus, in the case of the corollary above, we set $\Omega = \mathcal{X}^{\mathbb{N}_0}$, $\mathcal{F} = \bigotimes \mathcal{B}(\mathcal{X})$, and $\mathbb{P} = \mu$. We call this probability space the canonical probability space and call the resulting process (π_n) the canonical process.

Definition 2.11 (Shift Operator)**.** For $n \in \mathbb{N}$, we define the n-th shift operator by

$$\theta_n : \mathcal{X}^{\mathbb{N}_0} \to \mathcal{X}^{\mathbb{N}_0} : (a_0, a_1, \cdots) \mapsto (a_n, a_{n+1}, \cdots).$$

For connivance, we will denote the above property by $\theta_n(a.) = (a_{n+}.)$.

Definition 2.12 (Stationary Process)**.** A stochastic process $(x.)$ is stationary if for any n,

$$\mathcal{L}(\theta_n x.) = \mathcal{L}(x.).$$

Straight away, by Kolmogorov's extension, we see that an equivalent definition for the stationary process is that the finite dimensional distributions of $(\theta_n x.)$ are the same as the finite dimensional distributions of $(x.)$.

In general, a process (x_n) is stationary if

$$\mathcal{L}(x_n, \cdots, x_{n+m}) = \mathcal{L}(x_0, \cdots, x_m)$$

for all $n, m \geq 0$. In the case that (x_n) is a Gaussian process, as $\mathcal{L}(x_{i_1} \cdots x_{i_n})$ is determined by $(\mathbb{E}x_{i_1}, \cdots, \mathbb{E}x_{i_n})$ and $\mathrm{Cov}(x_{i_k}, x_{i_l})$, it is stationary if

$$\mathbb{E}x_n = \mathbb{E}x_0$$

for all n and the covariances are shift invariant.

3 Strong Markov Property

We will in the section continue to let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $(\mathcal{F}_n)$ be a filtration on this probability space.

3.1 Transition Probability

Given a Markov process (x_n) on the state space $\mathcal{X}$, for $A \in \mathcal{B}(\mathcal{X})$, the function $\mathbb{P}(x_{n+1} \in A \mid x_n)$ is a Borel function of x_n. This function might depend on A, n and $n-1$. Suppose the case that this function does not depend on time (i.e. time homogeneous), that is there exists some function $\phi(x, A)$ such that

$$\mathbb{P}(x_{n+1} \in A \mid x_n) = \phi(x_n, A)$$

almost surely. Fixing x, under some regularities, its not difficult to show that $\phi(x, \cdot)$ form a probability measure. We shall assume this.

Definition 3.1. The set $P := \{P(x, A) \mid x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X})\}$ is said to be a family of transition probabilities if

- for all $x \in \mathcal{X}$, $P(x, \cdot)$ is a probability measure;
- for any $A \in \mathcal{B}(\mathcal{X})$, $P(\cdot, A)$ is Borel measurable.

As an example, consider the Markov chain (x_n) on $\mathcal{X}$ where $x_{n+1} := F(x_n, \xi_{n+1})$ for $x_0, \xi_1, \xi_2, \cdots$ independent with $(\xi_i : \Omega \to \mathcal{Y}) \sim \mu$ for all i. Then, for all $\omega \in \Omega$, we have (recall that we denote the event $x_n^{-1}(\{\omega\})$ by $x_n = \omega$

$$\begin{aligned}\mathbb{P}(x_{n+1} \in A \mid x_n = \omega) &= \mathbb{P}(F(x_n(\omega), \xi_{n+1}) \in A) \\ &= \int_{\mathcal{Y}} \mathbf{1}_A(F(x_n(\omega), \xi_{n+1}))\mathrm{d}\mathbb{P} \\ &= \int_{\mathcal{Y}} \mathbf{1}_A(F(x_n(\omega), y))\mu(\mathrm{d}y).\end{aligned}$$

Hence, defining

$$P(x, A) := \int_{\mathcal{Y}} \mathbf{1}_A(F(x, y))\mu(\mathrm{d}y),$$

$\{P(x, A)\}$ are the transition probabilities and

$$\mathbb{P}(x_{n+1} \in A \mid x_n = \omega) = P(x_n(\omega), A).$$

We remark that, in the case that the state space is countable, by σ-additivity, it is sufficient to work with transitional probabilities of singletons. In particular, the transitional probability is simply determined by

$$P(i, j) = P(i, \{j\}), i, j \in \mathcal{X}.$$

Proposition 3.1. Let (x_n) be a Markov chain such that there exists a transition probability P for which

$$\mathbb{P}(x_{n+1} \in A \mid x_n) = P(x_n, A)$$

for all $A \in \mathcal{B}(\mathcal{X})$. Then, for all $f \in \mathcal{B}_b(\mathcal{X})$,

$$\mathbb{E}(f(x_{n+1}) \mid x_n) = \int_{\mathcal{X}} f(y)P(x_n, \mathrm{d}y).$$

Proof. Choosing $f = \mathbf{1}_A$, we see that the property is true for simple functions and so, the results can be extended to all functions by the monotone convergence theorem. $\square$

Proposition 3.2. Let (x_n) as defined above, for all $f \in \mathcal{B}_b(\mathcal{X})$, we have

$$\mathbb{P}(x_{n+2} \in A \mid x_n) = \int_{\mathcal{X}} P(y, A)P(x_n, \mathrm{d}y).$$

Thus, we define the two-step transition probability by

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

such that $\mathbb{P}(x_{n+2} \in A \mid x_n) = P^2(x_n, A)$ almost surely.

Proof. By the tower law, we have

$$\begin{aligned}
\mathbb{P}(x_{n+2} \in A \mid x_n) &= \mathbb{E}(\mathbb{E}(\mathbf{1}_A(x_{n+2}) \mid x_{n+1}, x_n) \mid x_n) \\
&= \mathbb{E}(\mathbb{E}(\mathbf{1}_A(x_{n+2}) \mid x_{n+1}) \mid x_n) \\
&= \mathbb{E}(P(x_{n+1}, A) \mid x_n) \\
&= \int_{\mathcal{X}} P(y, A)P(x_n, \mathrm{d}y),
\end{aligned}$$

where the last equality follows by the above proposition. $\square$

The above process can be extended to k-steps by induction. In particular, for all $k \in \mathbb{N}$, we have

$$\mathbb{P}(x_{n+(k+1)} \in A) \mid x_n) = \int_{\mathcal{X}} P^i(y, A)P^j(x_n, \mathrm{d}y)$$

for any $i, j \in \mathbb{N}, i + j = k$. This is known as the Chapman-Kolmogorov equation.

Definition 3.2. A family

$$\{P^n(x, \cdot) \mid x \in \mathcal{X}, n \in \mathbb{N}_0\}$$

is said to be a transition function if

- $P^n(x, \cdot)$ is a transition probability for any x, n;
- $P^0(x, \cdot) = \delta_x$ for any x;
- (Chapman-Kolmogorov) for all $n, m \geq 0, x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{B})$,

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

Proposition 3.3. The Chapman-Kolmogorov equation is satisfied if and only if for all $f \in \mathcal{B}_b(\mathcal{X}), n, m \geq 0, x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{B})$,

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

Proof. Exercise. □

As alluded to above, the k-step transitional probabilities can be constructed from a 1-step transitional probabilities. In particular, given the 1-step transitional probability P,

1. set $P^0(x, \cdot) = \delta_x$;
2. set $P^1(x, \cdot) = P(x, \cdot)$;
3. for all $n > 1$, $x \in \mathcal{X}$, for all $A \in \mathcal{B}(\mathcal{X})$, set

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

It remains to show that this construction satisfies the Chapman-Kolmogorov equation. Indeed, by induction, suppose that the Chapman-Kolmogorov equation is satisfied for all $k \leq n + m$, then

$$\begin{aligned} P^{n+m+1}(x, A) &= \int_{\mathcal{X}} P(z, A) P^{n+m}(x, \mathrm{d}z) \\ &= \int_{\mathcal{X}} \left(\int_{\mathcal{X}} P(z, A) P^j(y, \mathrm{d}y) \right) P^{n+m-j}(x, \mathrm{d}y) \\ &= \int_{\mathcal{X}} P^{j+1}(y, A) P^{n+m-j}(x, \mathrm{d}y) \end{aligned}$$

for all $j = 0, 1, \cdots$, where the second and third equality follows by the inductive hypothesis and Fubini's theorem.

Definition 3.3. A transition probability P is said to be the transitional probability of the Markov chain (x_n) if

$$\mathbb{P}(x_{n+1} \in A \mid x_n) = P(x_n, A)$$

almost surely for all $A \in \mathcal{B}(\mathcal{X})$ and any $n \geq 0$.

If a Markov chain has a transitional probability P, then we say the Markov chain is time homogeneous.

From this point forward, unless otherwise stated, we will assume our Markov chains to be time homogeneous.

Theorem 5. Let (x_n) be a Markov process with transition probability P. Then

- $\mathbb{P}(x_{n+m} \in A \mid x_m) = P^n(x_m, A)$ almost surely for any $n, m \geq 0, A \in \mathcal{B}(\mathcal{X})$;
- if $\mathcal{L}(x_0) = \mu$, then

$$\mathbb{P}(x_n \in A) = \int_{\mathcal{X}} P^n(x, A) \mu(\mathrm{d}x).$$

Proof. The first property follows by induction. Indeed, if for some $k \in \mathbb{N}$, $\mathbb{P}(x_{k+m} \in A \mid x_m) = P^k(x_m, A)$, for any m, then

$$\begin{aligned}\mathbb{P}(x_{k+1+m} \in A \mid x_m) &= \mathbb{E}(\mathbb{P}(x_{k+1+m} \in A \mid \mathcal{F}_{k+m}) \mid x_m)\\ &= \mathbb{E}(P(x_{m+k}, A) \mid x_m)\\ &= \int_{\mathcal{X}} P(z, A) P^k(x_m \mathrm{d}z) = P^{k+1}(x_m, A)\end{aligned}$$

where the second to last equality follows as $\mathbb{E}(f(x_{k+m}) \mid x_m) = \int f(z) P^k(x_m, \mathrm{d}z)$.

The second property follows as

$$\mathbb{P}(x_n \in A) = \mathbb{E}(\mathbb{P}(x_n \in A \mid x_0)) = \mathbb{E}(P^n(x_0, A)) = \int_{\mathcal{X}} P^n(y, A)\mu(\mathrm{d}y)$$

as required. □

Theorem 6 (Einstein's Relation)**.** If (x_n) is a Markov chain with transition probability P and initial distribution μ, then, its finite dimensional distributions are

$$\mathbb{P}(x_0 \in A_0, \cdots, x_n \in A_n) = \int_{A_0} \cdots \int_{A_{n-1}} P(y_{n-1}, A_n) P(y_{n-2}, \mathrm{d}y_{n-1}) \cdots P(y_0, \mathrm{d}y_1)\mu(\mathrm{d}y_0).$$

We remark that the above definition provides a sequence of consistent measures. Namely, if we define

$$\mu(A_0 \times \cdots \times A_n) := \int_{A_0} \cdots \int_{A_{n-1}} P(y_{n-1}, A_n) P(y_{n-2}, \mathrm{d}y_{n-1}) \cdots P(y_0, \mathrm{d}y_1)\mu(\mathrm{d}y_0),$$

the sequence of measures (μ_n) is consistent.

Proof. For $A_0, A_1 \in \mathcal{B}(\mathcal{X})$, we observe

$$\begin{aligned}\mathbb{P}(x_0 \in A_0, x_1 \in A_1) &= \mathbb{E}(\mathbb{E}(\mathbf{1}_{A_0}(x_0)\mathbf{1}_{A_1}(x_1) \mid x_0))\\ &= \mathbb{E}(\mathbf{1}_{A_0}(x_0)\mathbb{E}(\mathbf{1}_{A_1}(x_1) \mid x_0))\\ &= \mathbb{E}(\mathbf{1}_{A_0}(x_0) P(x_0, A_1))\\ &= \int_{A_0} P(y_0, A_1)\mu(\mathrm{d}y_0).\end{aligned}$$

Hence, by induction, the relation follows. □

Theorem 7. If (x_n) is a process satisfying Einstein's relation, then it is a Markov chain with transition probability P.

Proof. Einstein's relation can be extended to all bounded measurable functions through the usual process with the monotone convergence theorem and thus, we have for any $f_i \in \mathcal{B}_b(\mathcal{X})$,

$$\mathbb{E}\left(\prod_{i=0}^n f_i(x_i)\right) = \int \cdots \int \prod_{i=0}^n f_i(y_i) P(y_{n-1}, \mathrm{d}y_n) \cdots P(y_0, \mathrm{d}y_1)\mu(\mathrm{d}y_0).$$

By Fubini's theorem, the above becomes

$$
\begin{aligned}
\cdots &= \int f_n(y_n)P(y_{n-1}, \mathrm{d}y_n) \int \cdots \int \prod_{i=0}^{n-1} f_i(y_i) \prod_{i=1}^{n-1} P(y_i, \mathrm{d}y_{i+1})\mu(\mathrm{d}y_0) \\
&= \mathbb{E}(f_n(x_n) \mid x_{n-1} = y_{n-1}) \int \cdots \int \prod_{i=0}^{n-1} f_i(y_i) \prod_{i=1}^{n-1} P(y_i \mathrm{d}y_{i+1})\mu(\mathrm{d}y_0) \\
&= \mathbb{E}\left(\mathbb{E}(f_n(x_n) \mid x_{n-1}) \prod_{i=0}^{n-1} f_i(x_i) \right)
\end{aligned}
$$

which is equivalent to the Markov property. □

Theorem 8 (Existence of Markov Chain)**.** Given a family of transition probabilities on P on $\mathcal{X}$ and any probability measure μ_0 on $\mathcal{X}$, there exists a unique (in distribution) Markov process x with transition probability P and initial distribution μ_0.

Proof. Define μ_n on $\mathcal{X}^{n+1}$ such that

$$
\mu_n(A_0 \times A_n) := \int_{A_0} \cdots \int_{A_{n-1}} P(y_{n-1}, A_n)P(y_{n-2}, \mathrm{d}y_{n-1}) \cdots P(y_1, \mathrm{d}y_0)\mu_0(\mathrm{d}y_0).
$$

It is routine to check that this sequence of measures is well-defined and consistent, and thus, by the Kolmogorov extension theorem, there exists a unique $\mathbb{P}_\mu$ on $\mathcal{X}^\infty$ such that $\mathbb{P}_\mu$ projects to μ_n on $\mathcal{X}^{n+1}$. Thus, taking (π_n) to be the canonical process on the canonical space $(\mathcal{X}^\infty, \otimes\mathcal{B}(\mathcal{X}), \mathbb{P}_\mu)$, we have found a Markov process which satisfies the condition. □

Consider again the case where the state space is countable $\mathcal{X} = \mathbb{N}$. As mentioned previously, the transition probability on $\mathcal{X}$ is then determined by $p_{ij} = P(i, \{j\})$. As $P(i, \cdot)$ is a probability measure by definition,

$$
1 = P(i, \mathcal{X}) = \sum_{j \in \mathcal{X}} p_{ij}.
$$

In the case that $\mathcal{X}$ is finite, these p_{ij} can be represented as a matrix, motivating the definition of a stochastic matrix.

Definition 3.4 (Stochastic Matrix)**.** A matrix $p = (p_{ij})$ with $p_{ij} \geq 0$ is said to be a stochastic matrix if $\sum_{j \in \mathcal{X}} p_{ij} = 1$.

In the discrete case, our construction of the transition probability from the 1-step transition probability is straightforward. In particular, we obtain

$$
P^{n+1}(i, A) = \int_{\mathcal{X}} P(y, A)P^n(i, \mathrm{d}y) = \sum_{k \in \mathcal{X}} P(k, A)P^n(i, k).
$$

Thus, if we write $P(i, \{j\}) = p_{ij}$, then

$$
P^2(i, \{j\}) = \sum_{k \in \mathcal{X}} p_{ik}p_{kj} = ((p_{kl})^2_{k,l \in \mathcal{X}})_{ij},
$$

where the last term denotes matrix multiplication. Thus, by induction, we obtain that

$$
P^n(i, \{j\}) = \sum_{k_1 \in \mathcal{X}} \cdots \sum_{k_{n-1} \in \mathcal{X}} p_{ik_1}p_{k_1k_2} \cdots p_{k_{n-1}j} = ((p_{kl})^n_{k,l \in \mathcal{X}})_{ij}.
$$

3.2 Transition Operator

In the case of the transition probability P of a Markov chain (x_n), we have the relation

$$\mathbb{P}(x_{n+1} \in A) = \int_{\mathcal{X}} P(y, A)\mu_n(\mathrm{d}y)$$

where $\mu_n = \mathcal{L}(x_n)$. Thus, in some sense, the transitional probability the an operator on measures changing the distribution to the next time step. This motivates the following definition.

Definition 3.5 (Transition Operator)**.** The transition operator T^* given the transition probability P on the set of probability measures on $\mathcal{X}$ is defined to be

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

for all $A \in \mathcal{B}(\mathcal{X})$.

With this definition, we obtain that, given a Markov process (x_n) with the transition probability P, we have $(T^*)^n(\mathcal{L}(x_m)) = \mathcal{L}(x_{n+m})$.

Definition 3.6 (Dual Transition Operator)**.** The dual transition operator T_* given the transition probability P is defined to be an operator acting on $\mathcal{B}_b(\mathcal{X})$ such that

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

for all $f \in \mathcal{B}_b(\mathcal{X})$.

Equivalently, the dual transition operator acting on f is

$$T_*f(x) = \mathbb{E}(f(x_1) \mid x_0 = x)$$

where (x_n) is the Markov process associated with P.

Proposition 3.4. The above operators are dual in the sense that, for all $f \in \mathcal{B}_b(\mathcal{X})$,

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

Proof. Hint: first prove for $f = \mathbf{1}_A$. □

For simplicity, we will denote both T^* and T_* with T when there is no confusion.

We note that T^* extends to signed measures allowing us to show linearity (recall that the set of signed measures form a vector space over $\mathbb{R}$).

In the case that $\mathcal{X} = \{1, \cdots, N\}$ is finite, a probability measure ν is uniquely determined by its values on singletons $\{\nu(\{1\}), \cdots, \nu(\{N\})\}$. Then, if $P = (p_{ij})$ is a stochastic matrix, we have $T\nu = (\nu(\{i\}))_{i=1}^N P$.

3.3 Stopping Times

Definition 3.7 (Stopping Time)**.** A function $T : \Omega \to \overline{\mathbb{N}} := \{0, 1, \cdots\} \cup \{\infty\}$ is said to be a stopping time with respect to the filtration $(\mathcal{F}_n)$ if

$$\{\omega \mid T(\omega) = n\} \in \mathcal{F}_n$$

for all $n \geq 0$.

This definition can be easily generalized to continuous time with the codomain being $\overline{\mathbb{R}}_+$ and taking

$$\{\omega \mid T(\omega) \leq t\} \in \mathcal{F}_t$$

for all $t \geq 0$ instead. The generalized definition is consistent with the discrete version since $\{T \leq n\} = \bigcup_{k=1}^n T = k$ and $\{T = n\} = \{T \leq n+1\} \setminus \{T \leq n\}$ and thus, T is an $(\mathcal{F}_n)$ stopping time if and only if $\{T \leq n\} \in \mathcal{F}_n$ for all $n \geq 0$.

Definition 3.8 (Stopped Process)**.** Let T be a stopping time and let (x_n) be a stochastic process. We define the stopped process (x_n^T) by

$$x_n^T(\omega) := x_{\min\{n, T(\omega)\}}(\omega),$$

for all $\omega \in \Omega$, $n \in \overline{\mathbb{N}}$.

In some sense, the stopped process as the name suggests, stop the process once some condition has been achieved. Consider the random walk on the integers with the stopping time being the walk reaches 4. Then, for each $\omega \in \Omega$, the stopped process is the same as the process as long as $x_n(\omega) \leq 4$ while after $x_k(\omega) = 4$, the process stops in the sense that $x_n^T(\omega)$ is constant for all $n \geq k$.

The following three propositions are exercises.

Proposition 3.5. If S, T are stopping times, then so are

$$S \wedge T := \min\{S, T\} \text{ and } S \vee T := \max\{S, T\}.$$

Proposition 3.6. If (S_n) is a sequence of stopping times, then

$$\limsup_{n \to \infty} S_n \text{ and } \liminf_{n \to \infty} S_n$$

are stopping times.

Proposition 3.7. Constant functions are stopping times.

Definition 3.9 (Stopped σ-algebra)**.** Given a stopping time T, let $\mathcal{F}_\infty := \bigvee_{n=0}^\infty \mathcal{F}_n$ and define

$$\mathcal{F}_T := \{A \in \mathcal{F}_\infty \mid A \cap \{T = n\} \in \mathcal{F}_n, \forall n \geq 0\}.$$

For the continuous case, the set $\{T = n\}$ is replaced by $\{T \leq t\}$.

We note that in the case that T is a constant, $\mathcal{F}_T = \mathcal{F}_m$. Furthermore, if $S \leq T$ a.e. then $\mathcal{F}_S \subseteq \mathcal{F}_T$ (we assume the probability space is complete, i.e. null-sets are measurable).

Proposition 3.8. For a stopping time $T < \infty$, the stopped σ-algebra can be equivalently defined as

$$\mathcal{F}_T = \{A \in \mathcal{F} \mid A \cap \{T \leq t\} \in \mathcal{F}_t, \forall t \geq 0\}.$$

Proof. Clearly the right hand side is larger and so, it suffices to show that, for all $A \in \mathcal{F}$ such that $A \cap \{T \le t\} \in \mathcal{F}_t, \forall t \ge 0$, $A \in \mathcal{F}_T$. Now, by considering

$$A = \bigcup_{n \in \mathbb{N}} A \cap \{T \le n\},$$

where $A \cap \{T \le n\} \in \mathcal{F}_n$, we have

$$A = \bigcup_{n \in \mathbb{N}} A \cap \{T \le n\} \in \bigvee_{n=0}^{\infty} \mathcal{F}_n,$$

and hence, $A \in \mathcal{F}_T$ as required. ◻

Proposition 3.9. T is $\mathcal{F}_T$-measurable.

Proof. Let $A = \{T = m\}$ and by construction, for all $n \ge 0$, we have $A \cap T = n$ is $\emptyset$ if $n \ne m$ and $\{T = n\}$ if $n = m$ both contained in $\mathcal{F}_n$. Thus, $\{T = m\} \in \mathcal{F}_T$ and hence T is $\mathcal{F}_T$-measurable. ◻

Proposition 3.10. For a stopping time $T < \infty$, and an adapted process (x_n),

- $\omega \mapsto x_{T(\omega)}(\omega)$ (denoted by x_T) is $\mathcal{F}_T$-measurable;
- for all n, x_n^T is $\mathcal{F}_T$-measurable.

Proof. We have $\{x_T \in A\} \cap \{T = n\} = \{x_n \in A\} \cap \{T = n\} \in \mathcal{F}_n$ as (x_n) is adapted. Thus, x_T is $\mathcal{F}_T$-measurable.

The second property follows as $x_n^T = x_{T \wedge n}$ where $T \wedge n$ is a stopping time. Thus, $x_{T \wedge n}$ is $\mathcal{F}_{T \wedge n}$-measurable. On the other hand, as $T \wedge n \le T$, we have $\mathcal{F}_{T \wedge n} \subseteq \mathcal{F}_T$ and thus, x_n^T is $\mathcal{F}_T$-measurable as required. ◻

Proposition 3.11. If $T < \infty$ is a $\mathcal{F}_T^x$-stopping time, then for all $n \ge 0$,

$$\{T = n\} \in \sigma(x_{T \wedge 0}, \cdots, x_{T \wedge n}).$$

That is to say T is a stopping time with respect to the natural filtration of stopped process x_n^T.

Proof. It suffices to show $\mathbf{1}_{T=n}$ is of the form $\phi(x_{T \wedge 0}, \cdots, x_{T \wedge n})$ for some $\phi \in \mathcal{B}_b(\mathcal{X}^{n+1})$. We will prove this by induction.

Suppose there exists some $\phi_l \in \mathcal{B}_b(\mathcal{X}^{l+1})$ for all $l \le k-1$ such that

$$\mathbf{1}_{\{T=l\}} = \phi(x_{T \wedge 0}, \cdots, x_{T \wedge l}).$$

We observe (factorisation lemma implies the existence of ψ),

$$\begin{aligned}
\mathbf{1}_{\{T=k\}} &= \mathbf{1}_{\{T=k\}} \mathbf{1}_{\{T \ge k\}} = \psi(x_0, \cdots, x_k) \mathbf{1}_{\{T \ge k\}} \\
&= \psi(x_{T \wedge 0}, \cdots, x_{T \wedge k}) \mathbf{1}_{\{T \ge k\}} \\
&= \psi(x_{T \wedge 0}, \cdots, x_{T \wedge k})(1 - \mathbf{1}_{\{T \le k-1\}}).
\end{aligned}$$

Now, since $1 - \mathbf{1}_{\{T \le k-1\}} = 1 - \sum_{l \le k-1} \phi_l$, the result follows. ◻

Proposition 3.12. Let us denote $\sigma(x^T) := \sigma(x_{T\wedge n} \mid n)$, then if $T < \infty$ is a stopping time with respect to $\mathcal{F}^x_n$,

$$\mathcal{F}_T = \sigma(x^T).$$

Proof. Clearly $\mathcal{F}_T \supseteq \sigma(x^T)$ so we will prove the reverse. Let $A \in \mathcal{F}_T$, then, for all $n \geq 0$, as $A \cap \{T = n\}$ is $\mathcal{F}_n$ measurable, by the factorisation lemma, there exists some $\psi \in \mathcal{B}_n(\mathcal{X}^{n+1})$ such that

$$\psi(x_0, \cdots, x_n) = \mathbf{1}_{A\cap\{T=n\}} = \mathbf{1}_A \mathbf{1}_{\{T=n\}}.$$

So,

$$\mathbf{1}_A \mathbf{1}_{\{T=n\}} = \mathbf{1}_A \mathbf{1}^2_{\{T=n\}} = \mathbf{1}_{\{T=n\}} \psi(x_0, \cdots, x_n) = \mathbf{1}_{\{T=n\}} \psi(x_{T\wedge 0}, \cdots, x_{T\wedge n}).$$

Thus, as $\{T = n\} \in \sigma(x_{T\wedge 0}, \cdots, x_{T\wedge n}) \subseteq \sigma(x^T)$ by the above lemma, $\mathbf{1}_A \mathbf{1}_{\{T=n\}}$ is $\sigma(x^T)$-measurable. Hence,

$$\mathbf{1}_A = \sum_{n=0}^{\infty} \mathbf{1}_A \mathbf{1}_{\{T=n\}} \in \sigma(x^T)$$

$\square$

3.4 Strong Markov Property

Recall the shift operator, we define the following operator.

Definition 3.10. Given a stopping time $T < \infty$ a.e. we define θ_T such that for all stochastic process $x. = (x_n)$

$$(\theta_T x.)_n(\omega) := x_{T(\omega)+n}(\omega).$$

Definition 3.11 (Strong Markov Property)**.** A stochastic process $x.$ is said to have the strong Markov property if for every stopping time $T < \infty$ a.e. and every $\Phi \in \mathcal{B}_b(\mathcal{X}^\infty)$, we have

$$\mathbb{E}(\Phi(\theta_T x.) \mid \mathcal{F}_T) = \mathbb{E}(\Phi(\theta_T x.) \mid x_T),$$

almost everywhere.

We may assume that Φ is independent in its components. Namely, the strong Markov property is equivalent to

$$\mathbb{E}\left(\prod_{i=1}^m f_i(\theta_T x.)_{n_i} \mid \mathcal{F}_T\right) = \mathbb{E}\left(\prod_{i=1}^m f_i(\theta_T x.)_{n_i} \mid x_T\right),$$

for some $f_i \in \mathcal{B}_b(\mathcal{X})$ and $n_1 < n_2 < \cdots < n_m$.

Our goal is to show that if (x_n) is a time homogeneous Markov process with transition probability P, then it has the strong Markov property.

Proposition 3.13. Let $T < \infty$ a.e. be a stopping time. Then

$$\mathbb{P}(x_{n+T} \in A \mid \mathcal{F}_T) = P^n(x_T, A)$$

almost everywhere. In particular, $(x_{n+T}) = \theta_T x.$ is a Markov process with transition probability P.

Proof. Let $f \in \mathcal{B}_b(\mathcal{X})$, then, as $\{T = \infty\}$ has measure 0, for all $B \in \mathcal{F}_T$,

$$\begin{aligned}
\int_B f(x_{n+T})\mathrm{d}\mathbb{P} &= \sum_{m=0}^{\infty} \int_{B\cap\{T=m\}} f(x_{n+m})\mathrm{d}\mathbb{P} \\
&= \sum_{m=0}^{\infty} \int_{B\cap\{T=m\}} \mathbb{E}(f(x_{n+m}) \mid \mathcal{F}_m)\mathrm{d}\mathbb{P} \\
&= \sum_{m=0}^{\infty} \int_{B\cap\{T=m\}} \int f(y)P^n(x_m, \mathrm{d}y)\mathrm{d}\mathbb{P} \\
&= \sum_{m=0}^{\infty} \int_{B\cap\{T=m\}} \int f(y)P^n(x_T, \mathrm{d}y)\mathrm{d}\mathbb{P} \\
&= \int_B \int f(y)P^n(x_T, \mathrm{d}y)\mathrm{d}\mathbb{P},
\end{aligned}$$

where the second equality follows as $B \cap \{T = m\}$ is $\mathcal{F}_m$-measurable by the definition of stopping time while the third equality follows the the property of the transition probability. Thus,

$$\mathbb{E}(f(x_{n+T}) \mid \mathcal{F}_T) = \int f(y)P^n(x_T, \mathrm{d}y).$$

Hence, choosing $f = \mathbf{1}_A$ completes the proof. $\square$

Recall that given any measure μ and transition probability P, there exists a unique probability measure $\mathbb{P}_\mu$ on $\mathcal{X}^\infty$ (distribution of the canonical process) which is the distribution of a Markov process (denoted by X^x with transition probabilities P and initial distribution $\mathbb{P}_\mu$. If $\mu = \delta_x$ the Dirac measure, then we denote $\mathbb{P}_{\delta_x}$ by $\mathbb{P}_x$.

We introduce the following notation. If $\Phi \in \mathcal{B}_b(\mathcal{X}^\infty)$, we denote

$$\mathbb{E}_x[\Phi] := \mathbb{E}(\Phi(X^x.)) = \int_{\mathcal{X}^\infty} \Phi \mathrm{d}\mathbb{P}_x,$$

where the equality follows by the change of variable formula.

Theorem 9 (Strong Markov Property for Finite Stopping Time)**.** Let (x_n) be a time homogeneous Markov process with transition probability P and let T be a finite stopping time. Then,

- $\theta_T x.$ is also a time homogeneous Markov process with transition probability P and initial value x_T;
- if $\Phi \in \mathcal{B}_b(\mathcal{X}^\infty)$, then
$$\mathbb{E}(\Phi(\theta_T x.) \mid \mathcal{F}_T) = \mathbb{E}_{x_T}[\Phi].$$
  That is to say, $\theta_T x.$ also has the strong Markov property.

We have already proved the first property. We shall provide a proof for a simpler case where $T = n$ though the proof also works for the strong case (see official notes).

Again, since the product σ-algebra on $\mathcal{X}^\infty$ is determined by the π-system of cylindrical sets, it is sufficient to check the property for functions of the form $x \mapsto \prod_{i=1}^m f_i(x_{n_i})$ for

$f_i \in \mathcal{B}_b(\mathcal{X})$. Thus, the property is equivalent to

$$\mathbb{E}\left(\prod_{i=1}^{k} f_i(x_{n_i+T}) \mid \mathcal{F}_T\right) = \mathbb{E}_{x_T}\left[\prod_{i=1}^{k} f_i \circ \pi_{n_i}\right],$$

where π_{n_i} is the n_i-th projection map from $\mathcal{X}^\infty$.

Let us first consider the case where $T = \mathrm{id}$. Then,

$$\mathbb{E}_x\left[\prod_{i=1}^{k} f_i \circ \pi_{n_i}\right] = \int \cdots \int \prod_{i=1}^{m} f_i(y_i) \prod_{j=1}^{m} P^{n_j - n_{j-1}}(y_{j-1}, \mathrm{d}y_j).$$

In fact, $\prod_{j=1}^{m} P^{n_j-n_{j-1}}(y_{j-1}, \mathrm{d}y_j)$ is the distribution of $(\pi_{n_1}, \cdots, \pi_{n_m})$ on $(\mathcal{X}^\infty, \bigotimes \mathcal{B}(\mathcal{X}), \mathbb{P}_x)$.

Proposition 3.14. Let (x_n) be a time homogeneous Markov chain with transition probability P. Then, for any $\Phi \in \mathcal{B}_b(\mathcal{X}^\infty)$,

$$\mathbb{E}(\Phi(\theta_n x.) \mid \mathcal{F}_n)(\omega) = \mathbb{E}_{x_n(\omega)}[\Phi]$$

almost everywhere.

Proof. We may assume $\Phi = \prod_{i=1}^k f_i$, $f_i \in \mathcal{B}_b(\mathcal{X})$ and it suffices to show that, for $m_1 < \cdots < m_k$,

$$\mathbb{E}\left(\prod_{i=1}^{k} f_i(x_{n+m_i}) \mid \mathcal{F}_n\right) = \mathbb{E}_{x_n}\left[\prod_{i=1}^{k} f_i(x_{m_i})\right].$$

We will induct on k. For $k = 1$,

$$\mathbb{E}(f(x_{n+m}) \mid \mathcal{F}_n)(\omega) = \int f(y) P^m(x_n, \mathrm{d}y)$$

almost surely by the Markov property. Suppose now, the property holds for $k-1$. Then,

$$\begin{aligned}
&\mathbb{E}\left(\prod_{i=1}^{k} f_i(x_{n+m_i}) \mid \mathcal{F}_n\right) \\
&= \mathbb{E}\left(\prod_{i=1}^{k-1} f_i(x_{n+m_i}) \mathbb{E}(f_k(x_{n+m_k}) \mid \mathcal{F}_{n+m_{k-1}}) \mid \mathcal{F}_n\right) \\
&= \mathbb{E}\left(\prod_{i=1}^{k-1} f_i(x_{n+m_i}) \int f_k(y_k) P^{m_k - m_{k-1}}(x_{n+m_{k-1}}, \mathrm{d}y_k) \mid \mathcal{F}_n\right) \\
&= \mathbb{E}_{x_n}\left[\prod_{i=1}^{k-1} f_i(x_{n+m_i}) \int f_k(y_k) P^{m_k - m_{k-1}}(y_{k-1}, \mathrm{d}y_k)\right] \\
&= \int \cdots \int \prod_{i=1}^{k-1} f_i(y_i) \int f_k(y_k) P^{m_k - m_{k-1}}(y_{k-1}, \mathrm{d}y_k) \prod_{i=1}^{k-1} P^{m_i - m_{i-1}}(y_{i-1}, \mathrm{d}y_i) \\
&= \int \cdots \int \prod_{i=1}^{k} f_i(y_i) \prod_{j=1}^{k} P^{m_i - m_{i-1}}(y_{i-1}, \mathrm{d}y_i) = \mathbb{E}_{x_n}\left[\prod_{i=1}^{k} f_i(x_{m_i})\right]
\end{aligned}$$

as required. $\square$

Theorem 10 (Strong Markov Property for non-finite Stopping Times)**.** Let (x_n) be a time homogeneous Markov process with transition probability P and let T be a stopping time. Then for any $\Phi : \mathcal{B}_b(\mathcal{X}^\infty)$,

$$\mathbb{E}(\Phi(\theta_T x.)\mathbf{1}_{\{T<\infty\}} \mid \mathcal{F}_T)(\omega) = \mathbb{E}_{x_{T(\omega)}}[\Phi]$$

on $\{T < \infty\}$ almost everywhere.

More or less the same as the finite case since we are working on $\{T < \infty\}$.

Recall that the transition operator T^* is an operator acting on the space of measures such that, for μ a measure on $\mathcal{X}$,

$$T^*\mu(A) = \int P(y, A)\mu(\mathrm{d}y).$$

We say μ is invariant if $\mu = T^*\mu$. If $\mathcal{L}(x_0) = \pi$ is an invariant probability measure, then $\mathcal{L}(x_n) = \pi$ for all $n \geq 0$. Thus, the distribution of the process does not change with time. Indeed, by definition

$$\mathcal{L}(x_{n+1})(A) = \mathbb{P}(x_{n+1} \in A) = \int P(y, A)\mathcal{L}(x_n)(\mathrm{d}y),$$

and so follows by induction.

Definition 3.12 (Invariant)**.** A measure μ is said to be invariant if $\mu = T^*\mu$ where T^* is the transition operator with respect to some transition probability.

Definition 3.13 (Stationary)**.** A process (x_n) is said to be stationary such that $\theta_n x.$ has invariant distributions for all $n \geq 0$.

Proposition 3.15. A time homogeneous Markov process with invariant initial distribution is stationary.

Proof. Suppose (x_n) is a Markov process with initial distribution π such that π is invariant. Then, we denote $\mathbb{P}_\pi$ the distribution of $x.$ on $\mathcal{X}^\infty$, namely $\mathbb{P}_\pi = (x.)_*\mathbb{P}$. By the Markov property, $\theta_n x.$ is a Markov process with transition probability P and initial distribution $\mathcal{L}((\theta_n x.)_0) = \mathcal{L}(x_n) = \pi$. Hence, $\mathcal{L}(\theta_n x.) = \mathbb{P}_\pi$ and so, (x_n) is a stationary process. $\square$

4 THMC With Discrete Time

We will in this section consdier time homogeneous Markov chains (THMC) on discrete time.

We will in this section denote $\mathcal{X} = \{1, 2, \cdots, N\}$ if $|\mathcal{X}| < \infty$ and $\mathcal{X} = \{1, 2, \cdots\}$ otherwise. Given a THMC on $\mathcal{X}$ with transition probability P with initial distribution ν, then, we write $\mathcal{L}(x_n) =: \nu P^n$. As ν is discrete, it is represented by a vector $v \in [0,1]^{\mathcal{X}}$ such that for all $i \in \mathcal{X}$, $\nu(\{i\}) = v_i$. For short hand we write $\nu(i) := \nu(\{i\})$. Then,

$$\nu P(i) = \sum_{k \in \mathcal{X}} \nu(k) P_{ki}.$$

Definition 4.1 (Accessible)**.** Let $i, j \in \mathcal{X}$. Then we say j is accessible from i if there exists some n such that $P_{ij}^n = \mathbb{P}(x_n = j \mid x_0 = i) > 0$. We denote this by $i \longrightarrow j$.

Definition 4.2 (Communicating)**.** For $i, j \in \mathcal{X}$, i, j are communicating if $i \longrightarrow j$ and $j \longrightarrow i$. We denoted this by $i \sim j$.

We note the communicating is not necessarily an equivalence relation as a state might not communicate with itself.

Definition 4.3 (Communicating Class)**.** A communicating class for some $i \in \mathcal{X}$ is the set $[i] = \{j \in \mathcal{X} \mid i \sim j\}$.

Definition 4.4 (Irreducible)**.** A chain is irreducible if there exists only one communicating class, otherwise it is reducible.

Lemma 4.1. Communicating is transitive.

Proof. Suppose $i, j, k \in \mathcal{X}$ and $i \longrightarrow j, j \longrightarrow k$, then, there exists some n_1, n_2 such that $P_{ij}^{n_1}, P_{jk}^{n_2} > 0$. Then, by the C-K,

$$\begin{aligned} P_{ik}^{n_2+n_1} &= P^{n_2+n_1}(i, \{k\}) = \int P^{n_2}(y, \{k\}) P^{n_1}(i, \mathrm{d}y) = \sum_{l \in \mathcal{X}} P^{n_2}(l, \{k\}) P^{n_1}(i, \{l\}) \\ &\geq P^{n_2}(j, \{k\}) P^{n_1}(i, \{j\}) = P_{jk}^{n_2} P_{ij}^{n_1} > 0, \end{aligned}$$

implying $i \longrightarrow k$ as required. $\square$

Lemma 4.2. Let $i, j \in \mathcal{X}$ and suppose $i \longrightarrow j$. Then, any element in $[j]$ is accessible from any element in $[i]$.

Proof. Let $i' \in [i]$ and $j' \in [j]$. Then, by definition

$$i' \longrightarrow i \longrightarrow j \longrightarrow j'$$

implying $i' \longrightarrow j'$ by transitivity. $\square$

With this lemma in hand, we may define a partial order (antisymmetric) on the communicating classes. In particular, we say $[i] \leq [j]$ if every element of i can be assessed from any element of j. Equivalently, $j \longrightarrow i$.

Definition 4.5 (Minimal)**.** Am communicating class $[i]$ is said to be minimal (or closed) if there does **not** exist a communicating class $[j] \neq [i]$ such that $[j] \leq [i]$.

4.1 Recurrence and Transience

Definition 4.6 (Recurrent & Transient)**.** Let $\mathcal{X}$ be countable. A state $i \in \mathcal{X}$ is recurrent if $\mathbb{P}(T_i < \infty \mid x_0 = i) = 1$ where $T_i := \inf\{n \geq 1 \mid x_n = i\}$ is the first hitting time of i by the process. If a process is not recurrent at i, then i is called a transient state.

If every element of $\mathcal{X}$ is recurrent, then the process is called recurrent. Similarly, the process is called transient if every state is transient.

Let us introduce the following notation:

- $\mathbb{P}_i(A) := \mathbb{P}(A \mid x_0 = i)$,
- $\mathbb{P}_i(T_i < \infty) := \mathbb{P}(T_i < \infty \mid x_0 = i)$,
- $\mathbb{E}_i(Y) := \mathbb{E}(Y \mid x_0 = i)$.

We shall see later that this existence of a recurrent state implies the existence of and invariant measure. Also, recurrent and transient are properties invariant on communicating classes.

Lemma 4.3. Given two states $i, j \in \mathcal{X}$, $i \longrightarrow j$ if and only if $\mathbb{P}_i(T_j < \infty) > 0$. Also,

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

Proof. We note that $\{T_j < \infty\} = \bigsqcup_{n=1}^{\infty}\{T_j = n\}$ and so, if $P_{ij}^n > 0$, $0 < P_{ij}^n = \mathbb{P}_i(x_j = n) \leq \mathbb{P}(\{T_j < \infty\})$. On the other hand,

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

and so, if $\mathbb{P}_i(T_j < \infty) > 0$, then, $P_{ij}^n > 0$ for some n. $\square$

Let us define the following sequence of stopping times. Let $T_j^0 = 0$, $T_j^1 = T_j$ and

$$T_j^{n+1} := \inf\{k \geq T_j^n \mid x_k = j.$$

That is the next returning time after T_j^n. In the case that there is no confusion, we simply denote $T^n = T_j^n$.

Lemma 4.4. If $j \in \mathcal{X}$ is recurrent, then $\{T_j^{n+1} - T_j^n, n \geq 0\}$ are independent. Furthermore, $T_j^{n+1} - T_j^n$ are identically distributed and

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

for all $n, m = 1, 2, \cdots$.

Proof. It is sufficient to show that for every n, $T^{n+1} - T^n$ is independent of $\mathcal{F}_{T^n} = \sigma(\theta_{T^n} x.)$ (since $T^{k+1} - T^k$ is $\mathcal{F}_{T^n}$-measurable for all $k \leq n$, so $\sigma(T^{k+1} - T^k) \subseteq \mathcal{F}_{T^n}$). Now, since j is recurrent, $\mathbb{P}_j(T_j < \infty) = 1$ and so, for any $n \geq 1$, by the strong Markov property,

$$\mathbb{P}(T^{n+1} - T^n = m \mid \mathcal{F}_{T^n})(\omega) = \mathbb{P}_{x_{T^n(\omega)}}(T = m) = \mathbb{P}_j(T = m).$$

Hence, taking expectation on both sides, we obtain

$$\mathbb{P}(T_j^{n+1} - T_j^n = m) = \mathbb{P}_j(T_j = m).$$

Now, taking $A \in \mathcal{F}_{T^n}$, we have

$$\mathbb{P}(A \cap \{T^{n+1} - T^n = m\}) = \mathbb{E}(\mathbf{1}_A \mathbb{P}_j(T = m)) = \mathbb{P}(A)\mathbb{P}(T^{n+1} - T^n = m)$$

and thus, is independent. □

Lemma 4.5. Let $i, j \in \mathcal{X}$ and $k \geq 1$. Then

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

Proof. Define

$$\Phi : \mathcal{X}^\infty \to \mathbb{R} : (a_n) \mapsto \begin{cases} 1, & a_n = j \text{ for some } n, \\ 0, & \text{otherwise.} \end{cases}$$

Then, $T_j^{k+1} < \infty$ if and only if $\Phi(\theta_{T^k} x.) = 1$. Again, by the strong Markov property, we have

$$\begin{aligned} \mathbb{E}(\mathbf{1}_{\{T_j^{k+1} < \infty\}} \mathbf{1}_{T_j < \infty} \mid \mathcal{F}_{T_j}) &= \mathbb{E}(\Phi(\theta_{T^k} x.) \mathbf{1}_{T_j < \infty} \mid \mathcal{F}_{T_j}) \\ &= \mathbb{E}_j(\Phi(x.)) \mathbf{1}_{T_j < \infty} = \mathbb{E}_j(\mathbf{1}_{T_j^k < \infty}) \mathbf{1}_{T_j < \infty}. \end{aligned}$$

Hence, the result follows by taking expectation on both sides. □

Definition 4.7. Let $\eta_j := \sum_{n=1}^\infty \mathbf{1}_{s_n = j}$. This is known as the occupation time of j and counts the number of visits to j.

We see that $\mathbb{E}_j \eta_j = \sum_{n=1}^\infty \mathbb{P}_j(x_n = j) = \sum_{n=1}^\infty P_{jj}^n$.

Theorem 11 (Recurrence Criterion)**.** A state $j \in \mathcal{X}$ is transient if and only if $\sum_{n=1}^\infty P_{jj}^n < \infty$ and recurrent if and only if $\sum_{n=1}^\infty P_{jj}^n = \infty$.

Proof. We have $\mathbb{E}_j \eta_j = \sum_{n=1}^\infty P_{jj}^n$. On the other hand, by the tail probability formula,

$$\mathbb{E}_j \eta_j = \sum_{n=1}^\infty \mathbb{P}(\eta_j \geq n) = \sum_{n=1}^\infty \mathbb{P}_j(T_j^n < \infty) = \sum_{n=1}^\infty (\mathbb{P}_j(T_j < \infty))^n$$

where $\mathbb{P}_j(T_j^n < \infty) = (\mathbb{P}_j(T_j < \infty))^n$ by induction using the above lemma. As the right hand side is a geometric series, the sum is convergent if and only if $\mathbb{P}_j(T_j < \infty) < 1$, i.e. j is transient. Hence, $\sum_{n=1}^\infty P_{jj}^n < \infty$ if and only if j is transient as required. □

With the above criterion, we can show whether or not a state is recurrent or transient by considering the transition probabilities. As an example, consider again the symmetric walk on $\mathbb{Z}$. For all $i \in \mathbb{Z}$, we have

$$P_{ii}^{2n} = \binom{2n}{n} \left(\frac{1}{2}\right)^{2n}$$

as a path starting from i and ending at i in $2n$ steps results in n steps upwards and n steps down. On the other hand $P_{ii}^{2n+1} = 0$. Hence, we have

$$\sum_{n=1}^{\infty} P_{ii}^{n} = \sum_{n=1}^{\infty} P_{ii}^{2n} = \sum_{n=1}^{\infty} \binom{2n}{n} \left(\frac{1}{2}\right)^{2n}.$$

By Stirling's formula, $n! \sim (n/e)^n \sqrt{2\pi n}$, and so, $\sum_{n=1}^{\infty} \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \sim c \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} = \infty$ for some constant c. Thus, every state in the symmetric walk is recurrent.

Corollary 11.1. If $j \in [i]$ where $i, j \in \mathcal{X}$. Then both i, j are recurrent or transient.

Proof. By symmetry, it suffices to show that i is recurrent implies j is. As $i \sim j$, there exists some m_1, m_2 such that $P_{ij}^{m_1}, P_{ji}^{m_2} > 0$. Then,

$$\sum_{n=1}^{\infty} P_{jj}^{n} \geq P_{jj} + \cdots + P_{jj}^{m_1+m_2} + \sum_{n=1}^{\infty} P_{ji}^{m_2} P_{ii}^{n} P_{ij}^{m_1} \geq P_{ij}^{m_1} P_{ji}^{m_2} \sum_{n=1}^{\infty} P_{ii}^{n} = \infty.$$

Hence j is recurrent as required. □

Corollary 11.2. Let $j \in \mathcal{X}$. Then j is recurrent if and only if $\mathbb{P}_j(\{x_n = j \text{ i.o.}\}) = 1$, and if transient if and only if $\mathbb{P}_j(\{x_n = j \text{ i.o.}\}) = 0$.

Proof. We see that $\{x_n = j \text{ i.o.}\} = \{\eta_j = \infty\}$ and $\{\eta_j > m\} = \{T_j^{m+1} < \infty\}$. Then,

$$\mathbb{P}_j(\eta_j > m) = \mathbb{P}_j(T_j^{m+1} < \infty) = (P_j(T_j < \infty))^{m+1}.$$

Then, by continuity from above,

$$\begin{aligned}\mathbb{P}(\{x_n = j \text{ i.o.}\}) = \mathbb{P}(\eta_j = \infty) = \mathbb{P}\left(\bigcup_{m=1}^{\infty} \{\eta_j > m\}\right) \\ = \lim_{m \to \infty} \mathbb{P}(\eta_j > m) = \lim_{m \to \infty} (P_j(T_j < \infty))^{m+1}.\end{aligned}$$

So, j is recurrent $(P_j(T_j < \infty) = 1)$ if and only if $\mathbb{P}(\{x_n = j \text{ i.o.}\}) = 1$ and transient if and only if $\mathbb{P}(\{x_n = j \text{ i.o.}\}) = 0$ as required. □

Lemma 4.6. Let $\mathcal{X}$ be finite. A state is recurrent if and only if it is in a closed class.

Proof. The reverse direction is left as an exercise. Suppose $[i]$ is not closed, then there exists some $j \in [i]$ and $k \notin [i]$ such that $P_{jk} > 0$. Then, by definition, a path starting from j arriving at k cannot return to $[i]$ and thus, $\mathbb{P}_j(T_j < \infty) < 1$ implying j is transient. □