Markov Process

Kexing Ying

April 20, 2022

Contents

1 Introduction and Review **2**

2 Markov Property **4**
2.1 Filtration and Simple Markov Property 4
2.2 Markov Property 6
2.3 Gaussian Measure and Gaussian Process 9
2.4 Kolmogorov's Extension Theorem 10

3 Strong Markov Property **12**
3.1 Transition Probability 12
3.2 Transition Operator 17
3.3 Stopping Times 18
3.4 Strong Markov Property 20

4 THMC With Discrete Time **24**
4.1 Recurrence and Transience 25
4.2 Existence and Uniqueness of Invariant Measures 27
4.3 Long Run Probabilities 31
4.4 Total Variation Distance 35
4.5 Law of Large Numbers for THMC 39
4.6 Reversible Markov Chain 41
4.7 Finite State Markov Chain 42

5 Invariant Measures in General State Space **48**
5.1 Weak Convergence and Feller 48
5.2 Invariant Measures and Lyapunov Function Test 49
5.3 Deterministic Contraction and Minorisation 52
5.4 Strong Feller Property 56
5.5 Invariant Sets 57
5.6 Random Dynamical System 59

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. □

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 $(\mathcal{F}_n^x)$ for (x_n) is

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

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

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

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 $s \le t$.

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. □

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

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

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. □

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. □

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. □

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 $x \in \mathcal{X}$, we have (recall that we denote the event $x_n^{-1}(\{x\})$ by $x_n = x$

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

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

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

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 $A \cap \{T \leq t\} \in \mathcal{F}_t, \forall t \geq 0$, $A \in \mathcal{F}_T$. Now, by considering

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

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

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

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 $n \geq 0$, we have $A \cap T = n$ is $\emptyset$ if $n \neq m$ 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 $T \wedge n \leq T$, 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 $n \geq 0$,

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

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 $l \leq k - 1$ 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),

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

Now, since $1 - \mathbf{1}_{\{T \leq k-1\}} = 1 - \sum_{l \leq k-1} \phi_l$, 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 $\mathcal{F}_n^x$,

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

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,

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

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,

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

□

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. □

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 $P^n_{ij} = \mathbb{P}(x_n = j \mid x_0 = i) > 0$. 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 $P^{n_1}_{ij}, P^{n_2}_{jk} > 0$. Then, by the C-K,

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

implying $i \longrightarrow k$ as required. □

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. □

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. □

4.2 Existence and Uniqueness of Invariant Measures

Lemma 4.7.

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

Proof. We may assume $i \sim j$ (since otherwise, both sides are 0). Denote $\eta_j = \sum_{n=1}^{\infty} \mathbf{1}_{x_n = j}$. Then (by the tail probability)

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

where the last equality holds for both cases where $\mathbb{P}_j(T_j < \infty) = 1$ or $\mathbb{P}_j(T_j < \infty) < 1$. $\square$

Proposition 4.1. If a state $j \in \mathcal{X}$ is transient, then $\sum_{n=1}^{\infty} P_{ij}^n < \infty$ and so, $\lim_{n \to \infty} P_{ij}^n = 0$.

Proof. From the above lemma, we have

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

where the right hand side is finite for $\mathbb{P}_j(T_j < \infty) < 1$. $\square$

Recalling that an invariant measure is invariant with respect to the transition operator, we have the following theorem.

Theorem 12. If π is an invariant probability measure (i.e. $\pi(A) = \int P(x, A)\pi(\mathrm{d}x)$) and $\pi(j) > 0$ for some $j \in \mathcal{X}$, then j is recurrent.

Proof. Since π is invariant, $\pi(j) = \int P_{ij}^n \pi(\mathrm{d}i) = \sum_{i \in \mathcal{X}} P_{ij}^n \pi(i)$ for any n. Then, we have

$$\infty = \sum_{n=1}^{\infty} \pi(j) = \sum_{n=1}^{\infty} \sum_{i \in \mathcal{X}} P_{ij}^n \pi(i) = \sum_{i \in \mathcal{X}} \pi(i) \sum_{n=1}^{\infty} P_{ij}^n$$
$$= \sum_{i \in \mathcal{X}} \pi(i) \frac{\mathbb{P}_i(T_j < \infty)}{\mathbb{P}_{\mathbb{j}}(T_j = \infty)} = \frac{\mathbb{P}_\pi(T_j < \infty)}{\mathbb{P}_j(T_j = \infty)}$$

where numerator of the last fraction is positive as $\pi(j) > 0$ and π is invariant. Thus, in order for the equation to hold, $\mathbb{P}_j(T_j = \infty) = 0$ and thus, j is recurrent. $\square$

Corollary 12.1. A transient Markov chain (every state is transient) does not have an invariant probability measure.

Proof. If π is invariant, then $\pi(j) = 0$ for all $j \in \mathcal{X}$ since the process is transient. Thus, $\pi(\mathcal{X}) = \sum_{i \in \mathcal{X}} \pi(i) = 0$ implying π is not a probability measure. $\square$

Definition 4.8 (Positive Recurrent)**.** A state $i \in \mathcal{X}$ is said to be positive recurrent if $\mathbb{E}_i T_i < \infty$. On the other hand, we say i is null-recurrent if $\mathbb{E}_i T_i = \infty$.

One may show positive recurrence and null-recurrence are class properties.

Let us now attempt to construct invariant measures from recurrent states. Let $i \in \mathcal{X}$ be recurrent. Then, for all $j \in \mathcal{X}$, let us define

$$\mu(j) = \mathbb{E}_i \left(\sum_{n=0}^{T_i - 1} \mathbf{1}_{x_n = j} \right).$$

This is the expected number of visits to j starting from i till the first time the chain returns to i. We note that $\mu(i) = 1$ since $x_n \neq i$ for all $n < T_i$ and $x_{T_i} = i$ be definition. Furthermore, as i is recurrent, $\mathbb{P}_i(T_i < \infty) = 1$,

$$\mu(j) = \mathbb{E}_i \left(\sum_{n=0}^{\infty} \mathbf{1}_{n < T_i} \mathbf{1}_{x_n = j} \right) = \sum_n \mathbb{P}_i(x_n = j, n < T_i).$$

Then, summing over all states, we have

$$\sum_{j \in \mathcal{X}} \mu(j) = \sum_{j \in \mathcal{X}} \sum_{n=0}^{\infty} \mathbb{P}_i(x_n = j, T_i > n) = \sum_{n=0}^{\infty} \mathbb{P}(T_j > n) = \mathbb{E}_i T_i.$$

Theorem 13. Let $i \in \mathcal{X}$ be recurrent. Then, the measure μ defined by

$$\mu(j) := \sum_{n=0}^{\infty} \mathbb{P}_i(x_n = j, T_i > n)$$

is invariant.

Proof. Suppose first $j \neq i$. Then, $\mu(j) = \sum_{n=1}^{\infty} \mathbb{P}_i(x_n = j, T_i > n)$ and so,

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

where T is the transition operator. Now considering the i case. Consider

$$\begin{aligned}
\mathbb{P}_i(T_i = n+1) &= \sum_{k \neq i} \mathbb{P}_i(T_i > n, x_{n+1} = i, x_n = k) \\
&= \sum_{k \neq i} \mathbb{P}_i(x_{n+1} \mid x_n = k, T_i > n) \mathbb{P}_i(T_i > n, x_n = k) \\
&= \sum_{k \neq i} P_{ki} \mathbb{P}_i(T_i > n, x_n = k) = \sum_{k \in \mathcal{X}} P_{ki} \mathbb{P}_i(T_i > n, x_n = k).
\end{aligned}$$

Thus,

$$T\mu(i) = \sum_{k\in\mathcal{X}}\sum_{n=0}^{\infty}\mathbb{P}_i(T_i > n, x_n = k)P_{ki} = \sum_{n=0}^{\infty})\mathbb{P}(T_i = n+1) = \mathbb{P}(T_i < \infty) = 1 = \mu(i).$$

Hence, $T\mu = \mu$ as required. □

We now ask whether or not the invariant measure is unique. Obviously, a scalar multiple of a invariant measure if invariant, so, we ask for uniqueness up to an scalar multiple.

Lemma 4.8. Let $i \in \mathcal{X}$ be recurrent and let μ be the invariant measure constructed from i. Then, if ν is an invariant measure, then

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

for all $k \in \mathcal{X}$. Note that since $\mu(i) = 1$, we may reformulate this as

$$\frac{\nu(k)}{\nu(i)} = \frac{\mu(k)}{\mu(i)}$$

Proof. Clearly, equality holds for $k = i$ and so assume $k \neq i$. Let L^n be the last visit the chain visits i before n. We note that

$$\Omega = A_0 \cup \bigcup_{m=0}^{n-1}\{L^n = m\}$$

where A_0 is the set of $\omega \in \Omega$ for which $(x_n(\omega))$ does not visit i before n. Then,

$$\begin{aligned}
P_{jk}^n = \mathbb{P}_j(x_n = k) &\geq \sum_{m=0}^{n-1}\mathbb{P}_j(x_n = k, L^n = m)\\
&= \sum_{m=0}^{n-1}\mathbb{P}(x_n = k, x_{n-1} \neq i, \cdots, x_{m+1} \neq i, x_m = i)\\
&= \sum_{m=0}^{n-1}\mathbb{P}(x_n = k, x_{n-1} \neq i, \cdots, x_{m+1} \neq i | x_m = i)P_{ji}^m\\
&= \sum_{m=0}^{n-1}\mathbb{P}(x_{n-m} = k, x_{n-m-1} \neq i, \cdots, x_1 \neq i | x_m = i)P_{ji}^m\\
&= \sum_{m=0}^{n-1}\mathbb{P}_i(x_{n-m} = k, T_i > n-m)P_{ji}^m
\end{aligned}$$

where the last equality holds as $x_{n-m} = k \neq i$. Thus,

$$\begin{aligned}
\nu(k) = T^n\nu(k) = \sum_{j\in\mathcal{X}}\nu(j)P_{jk}^m &\geq \sum_{j\in\mathcal{X}}\nu(j)\sum_{m=0}^{n-1}\mathbb{P}_j(x_{n-m} = k, T_i > n-m)P_{ji}^m\\
&= \sum_{m=0}^{n-1}\mathbb{P}_j(x_{n-m} = k, T_i > n-m)\sum_{j\in\mathcal{X}}\nu(j)P_{ji}^m = \sum_{m=0}^{n-1}\mathbb{P}_i(x_{n-m} = k, T_i > n+m)\nu(i).
\end{aligned}$$

Hence, change of index and taking the sum to ∞, we have the inequality as required. □

Theorem 14. A irreducible, recurrent Markov process has invariant measure unique up to a scalar multiple.

Proof. If ν be any invariant measure and let μ be the invariant measure constructed from i. Then, as $\mu(i) = 1$,

$$0 = \nu(i) - \nu(i)\mu(i) = T^n\nu(i) - \nu(i)T^n\mu(i) = \sum_{k \in \mathcal{X}} (\nu(k) - \nu(i)\mu(k))P^n_{ki}$$

where $\nu(k) - \nu(i)\mu(k) \geq 0$ for all k by the above lemma. Thus, $(\nu(k) - \nu(i)\mu(k))P^n_{ki} = 0$ for all n, k. Then, as the process is irreducible, for any k, i, $k \sim i$ and so, there exists some n such that $P^n_{ki} > 0$. Thus, $\nu(k) = \nu(i)\mu(k)$ for all $k \in \mathcal{X}$ implying $\nu = \nu(i)\mu$ as required. $\square$

Theorem 15. If a Markov process is irreducible and has an invariant probability measure π. Then, $\mathbb{E}_i T_i < \infty$ (positively recurrent) for all $i \in \mathcal{X}$ and

$$\pi(i) = \frac{1}{\mathbb{E}_i T_i} (> 0).$$

Proof. Since π is a probability measure, there exists some $i \in \mathcal{X}$ such that $\pi(i) > 0$. Thus, i is recurrent and thus, the chain is recurrent by irreducibility. Thus, taking i arbitrarily, by uniqueness, $\pi = \pi(i)\mu$ where μ is the invariant measure associated with i, namely,

$$\mu(j) = \sum_{n=0}^{\infty} \mathbb{P}_i(x_n = j, T_i > n).$$

Hence, since $\sum_{j \in \mathcal{X}} \mu(j) = \mathbb{E}_i T_i$, we have

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

as required. $\square$

Corollary 15.1. An irreducible Markov process has an invariant probability measure if and only if it is **positively recurrent**.

4.3 Long Run Probabilities

We have seen that if $\lim_{n \to \infty} P^n_{ij}$ exists for any j and we define $\nu(j) := \lim_{n \to \infty} P^n_{ij}$ then, ν is an invariant measure (Why?). We now ask whether or not the converse is true.

Suppose the chain is irreducible and π is invariant, then, for all transient states $j \in \mathcal{X}$, $\lim_{n \to \infty} P^n_{ij} = 0$ and $\pi(j) = 0$ (as invariant measures do not charge transient states). Hence, $\lim_{n \to \infty} P^n_{ij} = \pi(j)$ for transient states. On the other hand, for situation is more difficult for recurrent states. Indeed, if we take the transition probability

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

then, $P^{2n+1} = P$ and $P^{2n} = \text{Id}$. Thus, $\lim_{n \to \infty} P^n_{ij}$ does not exist. On the other hand, this is an irreducible recurrent chain, and thus, processes an invariant measure. This motivates the study of periodic and non-periodic chains.

Denote $R(i) := \{n > 0 \mid P_{ii}^n > 0\}$ such that $n \in R(i)$ if there exists a path from i to i of length n. Clearly, if $n, m \in R(i)$, then $n+m \in R(i)$ (C-K) and if $R(i) \neq \emptyset$, then $|R(i)| = \infty$ (since if $n \in R(i)$ then $kn \in R(i)$ by the previous statement for all $k \in \mathbb{N}^*$.

Definition 4.9 (Period)**.** Given a state $i \in \mathcal{X}$, we define its period to be

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

Definition 4.10. A state $i \in \mathcal{X}$ is aperiodic if $d(i) = 1$ and periodic if $d(i) > 1$.

Proposition 4.2. Given $i, j \in \mathcal{X}$ such that $i \sim j$. Then $d(i) = d(j)$. Namely, the period is a class property.

Proof. Since $i \sim j$, there exists some n, m such that $P_{ij}^n, P_{ji}^m > 0$. Then, by C-K, $P_{ii}^{n+m} \geq P_{ij}^n + P_{ji}^m > 0$ and similarly for $P_{jj}^{n+m} > 0$, and hence, $n + m \in R(i) \cap R(j)$. Now, taking $k \in R(i)$, we have $k+n+m \in R(j)$ and so, $d(j) \mid k+n+m$ and $d(j) \mid n+m$ implying $d(j) \mid k$, and thus, $d(j) \leq d(i)$. By symmetry, $d(i) \leq d(j)$ and thus, $d(i) = d(j)$ as required. $\square$

Definition 4.11. A THMC (or P) is aperiodic if $d(i) = 1$ for all $i \in \mathcal{X}$ and is periodic with period d if $d(i) = d > 1$ for all $i \in \mathcal{X}$.

From the definition, we have the following proposition.

Corollary 15.2. An irreducible chain is either aperiodic or periodic.

Let $(x_n), (x_n')$ be independent time homogeneous Markov processes on $\mathcal{X} = \mathbb{N}$ with the same transition probability P. Furthermore, assume $(x_n), (x_n')$ has initial distribution μ and ν and write

$$P_\mu^n := \mathcal{L}(x_n), P_\nu^n := \mathcal{L}(x_n').$$

Lemma 4.9 (Doeblin Coupling)**.** $z_n := (x_n, x_n')$ is a THMC on $\mathcal{X}^2$ with the initial distribution $\mu \otimes \nu$ and transition probability Q, where

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

for all $i, i', j, j' \in \mathcal{X}$.

Denoting $T := \inf_{n \geq 0}\{x_n = x_n'\}$ (the coalessing time), we have the following lemmas.

Lemma 4.10 (Successful Coupling)**.** $\mathbb{P}(T < \infty) = 1$.

Lemma 4.11 (Coupling Inequality)**.** $\sum_{i \in \mathcal{X}} |\mathbb{P}(x_n = i) - \mathbb{P}(x_n' = i)| \leq 2\mathbb{P}(T > n)$. Thus, by the above lemma, the right hand side tends to 0 as $n \to \infty$.

Theorem 16. Let P be irreducible, aperiodic and positively recurrent (recall that this means $\mathbb{E}_i T_i < \infty$ for all $i \in \mathcal{X}$), and let π be its unique invariant probability measure. Then, for all $i \in \mathcal{X}$,

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

Proof. With the above constructions in mind, this theorem follows by taking $x_0 := i, x_0' := \pi$ and $\mathbb{P}(x_n = j) = P_{ij}^n$, $\mathbb{P}(x_n' = j) = \pi(j)$. $\square$

In order to prove the above lemmas, we need more machinery, namely coupling.

Definition 4.12 (Coupling)**.** Given two random variables $x, y : \Omega \to \mathcal{X}$, the coupling of the two is a random variable $z = (x', y')$ with state space $\mathcal{X}^2$ such that

$$\mathcal{L}(x) = \mathcal{L}(x') \text{ and } \mathcal{L}(y) = \mathcal{L}(y').$$

Definition 4.13 (Doeblin Coupling)**.** Let$(x_n), (x_n')$ be independent time homogeneous Markov processes on $\mathcal{X} = \mathbb{N}$ with the same transition probability P. Furthermore, assume $(x_n), (x_n')$ has initial distribution μ and ν. Then, the Doeblin coupling is the stochastic process defined by

$$z_n := (x_n, x_n').$$

Lemma 4.12 (Coupling lemma)**.** Let $(x_n), (x_n')$ be independent Markov processes with the same transition probability P. Then, defining

$$y_n := \begin{cases} x_n, & n < T, \\ x_n', & n \geq T, \end{cases}$$

where T is the coalessing time, (y_n) is a Markov process with the transition probability P and initial distribution $\mathcal{L}(x_0)$.

Proof. Let $f \in \mathcal{B}_b(\mathcal{X})$ and let

$$\mathcal{F}_n = \sigma(x_k \mid k \leq n) \vee \sigma(x_k' \mid k \leq n).$$

Then, $(x_n), (x_n')$ remains to be Markov processes with respect to the filtration $(\mathcal{F}_n)$ (by independence). Then,

$$\begin{aligned} \mathbb{E}(f(y_{n+1}) \mid \mathcal{F}_n) &= \mathbb{E}(f(y_{n+1})\mathbf{1}_{T \leq n} \mid \mathcal{F}_n) + \mathbb{E}(f(y_{n+1})\mathbf{1}_{T > n} \mid \mathcal{F}_n) \\ &= \mathbf{1}_{T \leq n}\mathbb{E}(f(x_{n+1}' \mid \mathcal{F}_n)) + \mathbf{1}_{T > n}\mathbb{E}(f(x_{n+1} \mid \mathcal{F}_n)) \\ &= \mathbf{1}_{T \leq n} T f(x_n') + \mathbf{1}_{T > n} T f(x_n) \\ &= \mathbf{1}_{T \leq n} T f(y_n) + \mathbf{1}_{T > n} T f(y_n) = T f(y) \end{aligned}$$

where T is the transition operator of the transition probability P. This implies (y_n) is a Markov process with respect to $(\mathcal{F}_n)$ (and hence to its natural filtration) with the same transition probability P. $\square$

Let us now prove the previously claimed lemmas.

Coupling Inequality. We are asked to prove

$$\sum_j |\mathbb{P}(x_n = j) - \mathbb{P}(x_n' = j)| \leq 2\mathbb{P}(T > n).$$

Defining (y_n) as above, as (x_n) and (y_n) have the same distribution,

$$\begin{aligned} |\mathbb{P}(x_n = j) - \mathbb{P}(x_n' = j)| &= |\mathbb{P}(y_n = j) - \mathbb{P}(x_n' = j)| \\ &= |\mathbb{P}(y_n = j) - \mathbb{P}(x_n' = j, n < T) - \mathbb{P}(y_n = j, n \geq T)| \\ &= |\mathbb{P}(y_n = j, n < T) - \mathbb{P}(x_n' = j, n < T)|. \end{aligned}$$

Hence, summing over j we have

$$\sum_j |\mathbb{P}(x_n = j) - \mathbb{P}(x'_n = j)| \leq \sum_j \mathbb{P}(y_n = j, n < T) + \sum_j \mathbb{P}(x'_n = j, n < T) \leq 2\mathbb{P}(n < T).$$

$\square$

Doeblin Coupling. The Doeblin Coupling asks us to prove the coupling $z_n := (x_n, x'_n)$ is a THMC on $\mathcal{X}^2$ with initial distribution $\mathcal{L}(x_0) \otimes \mathcal{L}(x'_0)$ and transition probability Q where

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

The initial distribution follows from independence. On the other hand,

$$\begin{aligned}\mathbb{P}(z_{n+1} = (j, j') \mid \mathcal{F}_n) &= \mathbb{P}(x_{n+1} = j, x'_{n+1} = j' \mid \mathcal{F}) \\ &= \mathbb{P}(x_{n+1} = j \mid x_n)\mathbb{P}(x'_{n+1} = j' \mid x'_n) \\ &= P(x_n, j)P(x'_n, j') = Q_{(x_n x'_n),(jj')}\end{aligned}$$

where the second equality holds by independence. $\square$

Before we can prove $T < \infty$ almost everywhere, let us construct a larger stopping time for which we know it is finite almost everywhere, hence proving the claim.

Lemma 4.13. If $i \in \mathcal{X}$ is aperiodic and recurrent, then there exists some N such that $P_{ii}^n > 0$ for all $n \geq N$.

Proof. Since i is recurrent, $\sum_{n=1}^{\infty} P_{ii}^n = \infty$ and so, $R(i) \neq \emptyset$. Since, by aperiodicity, $\gcd(R(i)) = 1$, there exists a finite subset $R_0 \subseteq R(i)$ such that $\gcd(R_0) = 1$. Then, by the Chinese remainder theorem, there exists some N such that for all $n > N$, $n \in R(i)$. $\square$

Corollary 16.1. If P is irreducible, aperiodic and recurrent, then, for all $i, j \in \mathcal{X}$, there exists some N such that for all $n \geq N$, $P_{ij}^n > 0$.

Proof. Since $i \to j$ as P is irreducible, there exists some m such that $P_{ij}^m > 0$. Then, by the above lemma, there exists some N such that for all $n \geq N$, $P_{ii}^n > 0$. So, for all $k \geq m + N$,

$$P_{ij}^k = P_{ij}^{(k-m)+m} \geq P^(k-m)_{ii}P_{ij}^m > 0$$

since $k - m \geq N$. $\square$

Lemma 4.14. Let P be irreducible, aperiodic and positively recurrent. Then, Q (as defined above from Doeblin coupling) is also irreducible and positively recurrent.

Proof. Firstly, $\pi \otimes \pi$ is an invariant probability measure for Q if π is invariant for P. Thus, it suffices to show Q is irreducible.

By induction, we can show that $Q_{(ii'),(jj')}^n = P_{ij}^n P_{i'j'}^n$. On the other hand, as P is irreducible, for any i, i', j, j' there exists some n, n' such that $P_{ij}^n, P_{i'j'}^{n'} > 0$. By the above corollary, we may choose sufficiently large n such that $P_{ij}^n, P_{i'j'}^n > 0$ and so, $Q_{(ii'),(jj')}^n = P_{ij}^n P_{i'j'}^n > 0$. $\square$

Finally, we can prove the successful coupling lemma.

Successful Coupling. We are required to prove $\mathbb{P}(T < \infty) = 1$. To achieve this, define

$$T_{(i,i)} := \inf\{z_n = (i,i)\} = \inf\{x_n = x_n' = i\}.$$

It is not difficult to see that $T \leq T_{(i,i)}$ and so, it suffices to show $\mathbb{P}(T_{(i,i)} < \infty) = 1$. Now, since Q is irreducible and recurrent, $\mathbb{P}_z(T_{(i,i)} < \infty) = 1$ for any $z \in \mathcal{X}^2$ and so,

$$\mathbb{P}(T < \infty) \geq \mathbb{P}(T_{(i,i)} < \infty) = 1.$$

□

Before moving on, let us first consider the following well-known lemma from elementary number theory.

Lemma 4.15. Let $S \subseteq \mathbb{N}$ be a non-empty additive set and let $d = \gcd(S)$. Then, there exists some $\kappa > 0$ such that $kd \in S$ for all $k > \kappa$.

Proof. We may assume $d = 1$ by dividing every element of S by d. In this case, by Bezout's lemma, there exists $a_1, \cdots, a_n \in \mathbb{Z}$ and $d_1, \cdots, d_n \in S$ such that $1 = \sum_{i=1}^n a_i d_i$. Then, defining $M = \sum_{i=1}^n d_i$, for all $k < M$, we observe that

$$NM + k = \sum_{i=1}^n N d_i + k \sum_{i=1}^n a_i d_i = \sum_{i=1}^n (N + k a_i) d_i.$$

Hence, for N sufficiently large, i.e. $N \geq N_0$ such that $N_0 + k a_i \geq 0$ for all i and $0 \leq k < M$, we have $N + k a_i \geq 0$ implying $NM + k \in S$ as S is additive. Thus, as all natural numbers larger than $N_0 M$ can be written in the form $NM + k$ where $N \geq N_0$ and $0 \leq k < M$, the claim follows. □

4.4 Total Variation Distance

Recall that in measure theory, we had considered the total variation of signed and complex measures. With regards to probability measures directly, the notion of total variation is less useful as the total variation of a probability measure is always 1. Instead we consider the total variation distance between two probability measures.

Definition 4.14 (Total Variation Distance)**.** The total variation distance between two probability measure μ, ν is

$$\|\mu - \nu\|_{TV} := 2 \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|.$$

We observe that the total variation distance is simply the total variation of the signed measure obtained from $\mu - \nu$ (multiplied by 2 by convention).

It is easy to check that

- $\|\mu - \nu\|_{TV} = 0 \iff \mu = \nu$;
- $\|\mu - \nu\|_{TV} \leq 2$;

- $\|\mu - \nu\|_{TV} = 2$ if $\mu \perp \nu$. Indeed, if $\mu \perp \nu$, then there exists A, B such that $A \cup B = \Omega$ and $\mu(B) = \nu(A) = 0$. So, $\mu(A) - \nu(A) = \mu(A) + \mu(B) \geq \mu(A \cup B) = \mu(\Omega) = 1$.

Lemma 4.16. If $\mathcal{X}$ is countable and μ, ν are probability measures on $\mathcal{X}$, then

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

In particular, if you consider μ, ν as functions on $\mathcal{X}$, the right hand side is simply the L^1 norm of $\mu - \nu$.

Proof. Define $B = \{j \mid \mu(j) \geq \nu(j)\}$. Then, as μ, ν are probability measures, $\mu(B) - \nu(B) = \nu(B^c) - \mu(B^c)$. Thus,

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

On the other hand, for all measurable set $A \subseteq \mathcal{X}$,

$$\begin{aligned}
|\mu(A) - \nu(A)| &\leq |(\mu - \nu)(A \cap B) + (\mu - \nu)(A \cap B^c)| \\
&\leq (\mu - \nu)(A \cap B) \vee (\mu - \nu)(A \cap B^c) \\
&\leq (\mu - \nu)(B) \vee (\mu - \nu)(B^c) = \mu(B) - \nu(B) \\
&= \frac{1}{2} \sum_{j \in \mathcal{X}} |\mu(j) - \nu(j)|.
\end{aligned}$$

Hence, the required inequality follows by multiplying both sides by 2. □

Proposition 4.3. If $\mu = f\lambda$ and $\nu = g\lambda$. Then,

$$\|\mu - \nu\|_{TV} = \int |f - g| \mathrm{d}\lambda = \|f - g\|_{L^1}.$$

Proof. Exercise. By the same idea as above. □

Definition 4.15. A sequence of measures (μ_n) is said to converge in total variation to the measure μ if

$$\lim_{n \to \infty} \|\mu_n - \mu\|_{TV} = 0.$$

We note that convergence in total variation is different to convergence weakly. Indeed, consider the sequence of Dirac measures $(\delta_{1/n})$. For all bounded continuous functions f, we have

$$\int f \mathrm{d}\delta_{1/n} = f(1/n) \to f(0) = \int f \mathrm{d}\delta_0.$$

Thus $\delta_{1/n} \to \delta_0$ weakly. On the other hand, for all n, we have

$$\|\delta_{1/n} - \delta_0\|_{TV} = 2$$

and so, $\delta_{1/n}$ does not converge in total variation to δ_0.

Proposition 4.4. Given probability measures μ, ν,

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

Proof. See Sheffe's lemma. □

Corollary 16.2. Convergence in total variation implies convergence weakly.

Before moving further, let us remark that the space of all probability measures $\mathbb{P}(\Omega)$ on Ω forms a complete metric space with the distance $\|\cdot\|_{TV}$. We will return to this later.

With the notion the total variation distance in mind, we may reformulate our convergence theorem regarding the transition probability. Namely, we have

$$\sum_{j=1}^{\infty} |P_{ij}^n - \pi(j)| = \|P^n(i, \cdot) - \pi\|_{TV}.$$

Recall that, if $\mathcal{L}(x_0) = \mu$, then

$$\mu P^n(j) := \mathbb{P}(x_n = j) = \sum_{i=1}^{\infty} \mu(i) P_{ij}^n.$$

Lemma 4.17. If $\lim_{n \to \infty} \|P^n(i, \cdot) - \pi\|_{TV} = 0$ for all $i \in \mathcal{X}$, then $\lim_{n \to \infty} \|\mu P^n - \pi\|_{TV} = 0$.

Proof. We observe

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

As $\sum \mu(i) = 1$, for all $\epsilon > 0$, there exists some N such that $\sum_{i=N+1}^{\infty} \mu(i) < \epsilon/4$. Then,

$$\begin{aligned}
\sum_{j=1}^{\infty} \left| \sum_{i=N+1}^{\infty} \mu(i) P_{ij}^n - \sum_{i=1}^{\infty} \mu(i)\pi(j) \right| &\leq \sum_{i=N+1}^{\infty} \mu(i) \sum_{j=1}^{\infty} |P_{ij}^n - \pi(j)| \\
&\leq \sum_{i=N+1}^{\infty} \mu(i) \sum_{j=1}^{\infty} (|P_{ij}^n| + |\pi(j)|) \\
&< 2\frac{\epsilon}{4} = \frac{\epsilon}{2}.
\end{aligned}$$

On the other hand, as $\sum_{j=1}^{\infty} |P_{ij}^n - \pi(j)| \to 0$ as $n \to \infty$, there exists some M such that for all $n \geq M$, $\sum_{j=1}^{\infty} |P_{ij}^n - \pi(j)| < \epsilon/2$ and so, for $n \geq M$,

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

Hence, adding the two sums together, we have $\|\mu P^n - \pi\|_{TV} < \epsilon$ for $n \geq M$ implying $\lim_{n \to \infty} \|\mu P^n - \pi\|_{TV} = 0$ as required. □

Theorem 17. If (x_n) is an irreducible, aperiodic, positively recurrent THMC with $\mathcal{L}(x_0) = \mu$, then

$$\lim_{n\to\infty} \|\mu P^n - \pi\|_{TV} = 0.$$

Proof. Follows by the above lemma and the reformulation of the convergence theorem. □

We have so far considered aperiodic chains. Let us now consider what happens with periodic ones.

Lemma 4.18. Let (x_n) be irreducible and let $d > 1$. Then (x_n) has period d if and only if $\mathcal{X}$ is a disjoint union of sets $A_0, \cdots, A_{d-1}$ such that if $i \in A_n$ for some n and $j \in \mathcal{X}$ such that $P_{ij} > 0$, then $j \in A_{n+1}$ (we define $A_d := A_0$).

Proof. Suppose first that P has period d. Then, we define

$$A_n := \{j \in \mathcal{X} \mid P_{1j}^{kd+n} > 0 \text{ for some } k = 0, 1, \cdots\}$$

for $n = 0, \cdots, d-1$. Clearly, $\bigcup_{n=0}^{d-1} A_n = \mathcal{X}$ since P is irreducible and every natural number can be written as $kd + n$ for some $k \in \mathbb{N}$, $0, \leq n \leq d-1$.

Furthermore, $(A_n)_{n=0}^{d-1}$ is pair-wise disjoint since if $j \in A_{n_1} \cap A_{n_2}$, then $P_{1j}^{n_1+k_1 d}, P_{1j}^{n_2+k_2 d} > 0$ for some k_1, k_2. By irreducibility, there exists some m such that $P_{j1}^m > 0$ and so, $P_{11}^{n_1+k_1 d+m}, P_{11}^{n_2+k_2 d+m} > 0$, namely, $n_1 + k_1 d + m, n_2 + k_2 d + m \in R(i)$. Thus, $d \mid |n_1 - n_2|$ implying $n_1 = n_2$ as $0 \leq n_1, n_2 \leq d-1$.

On the other hand, if such a sequence of sets $(A_n)_{n=0}^{d-1}$ exists. By C-K, $P_{ii}^m = 0$ for all $d \nmid m$. Since the chain is irreducible, it follows $P_{ii}^{kd} > 0$ for some k, and hence, $d = d(i)$ as required. □

Proposition 4.5. Let T denote the transition operator. If for some n, $T^n \mu = \mu$, then

$$\hat{\mu} = \frac{1}{n} \sum_{k=1}^{n} T^k \mu$$

is an invariant measure, i.e. $T\hat{\mu} = \hat{\mu}$.

Proof. Follows straight away by computation,

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

□

With this in mind, if P has period d and μ is an invariant measure for the chain on A_0 (where $\mathcal{X} = \bigsqcup A_n$) as constructed above, then $\frac{1}{d}\sum_{k=1}^{d} P^k \mu$ is an invariant measure for P.

4.5 Law of Large Numbers for THMC

We recall the strong law of large numbers.

Theorem 18 (Kolmogorov's Strong Law of Large Numbers)**.** Let (ξ_n) be a sequence of real-valued, independent and identically distributed random variables. Then, if $\mathbb{E}|\xi_n| < \infty$ and $\mathbb{E}\xi_n = a$ for all n,

$$\frac{1}{n}\sum_{k=1}^{n} \xi_k \to a$$

almost everywhere and in L^1.

We may not apply this LLN directly to Markov chains as the sequence is not necessarily independent. Nonetheless, there is a version of the law of large numbers for THMCs.

Theorem 19 (Law of Large Numbers)**.** Let (x_n) be an irreducible positively recurrent THMC on $\mathcal{X} = \mathbb{N}$ and let π denote its invariant probability measure. Then, for all $f \in L^1$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(x_k) = \int_{\mathcal{X}} f \mathrm{d}\pi$$

almost everywhere.

We observe that Kolmogorov's SLLN is a special case of this by taking $f = \mathrm{id}$. Furthermore, by taking $f = \mathbf{1}_{\{i\}}$, $\mu \sim \mathcal{L}(x_0)$, we have

$$\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{i\}}(x_k) \to \pi(i)$$

almost everywhere. As we have already show, if (x_n) is aperiodic, $\lim_{n\to\infty} \mathbb{E}_\mu(\mathbf{1}_{\{i\}}(x_k)) = \lim_{n\to\infty} \mathbb{P}_\mu(x_n = i) = \pi(i)$. Thus, by dominated convergence,

$$\lim_{n\to\infty} \mathbb{E}_\mu \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{\{i\}}(x_k) = \pi(i).$$

Proof. Let $i \in \mathcal{X}$, $T = T_i$, T^k the successive return times to i, $x_0 = i$. Then it is clear that

$$\left\{ \sum_{l=T^k}^{T^{k+1}} f(x_l), k = 0, 1, \cdots \right\}$$

are independent, identically distributed for all $f \in L^1$. Suppose for now $f \geq 0$. Then,

$$\mathbb{E}\sum_{0}^{T} f(x_l) = \mathbb{E}\sum_{0}^{T}\sum_{j\in\mathcal{X}} f(j)\mathbf{1}_{x_l=j} = \sum_{j\in\mathcal{X}} f(j)\mathbb{E}\sum_{0}^{T}\mathbf{1}_{x_l=j}.$$

Then, by recalling that $\mathbb{E}_i \sum_{l=0}^{T} \mathbf{1}_{x_l=j} = \mu(j)$ which is a invariant measure, so proportional to π by the factor $\mathbb{E}_i T$, we have

$$\mathbb{E}\sum_{l=0}^{T} f(x_l) = \sum_{j\in\mathcal{X}} f(j)\mu(j) = \sum_{j} f(j)\pi(j)\mathbb{E}_i T = \mathbb{E}_i[T]\int f\mathrm{d}\pi < \infty.$$

Hence, our sequence of i.i.d. random variables is integrable, and we may applying SLLN to our constructed sequence, i.e.

$$\frac{1}{n}\sum_{l=0}^{T^n} f(x_l) \to \mathbb{E}_i[T]\int f \mathrm{d}\pi$$

almost everywhere.

Now, by recalling $(T^{k+1} - T^k)_k$ is i.i.d., by applying SLLN, we have

$$\lim_{n\to\infty}\frac{T^n}{n} = \mathbb{E}_i T$$

almost everywhere. So, defining $\eta(n) := \sum_{k=1}^n \mathbf{1}_{x_k = i}$, we have $T^{\eta(n)} \leq n < T^{\eta(n)+1}$. Furthermore, as i is recurrent, $\eta(n) \to \infty$ and so, $T^{\eta(n)}/n \sim 1$. With this, we obtain

$$\frac{\eta(n)}{n}\frac{1}{\eta(n)}\sum_{l=0}^{T^{\eta(n)}} f(x_l) \leq \frac{1}{n}\sum_{l=0}^{n} f(x_l) \leq \frac{\eta(n)+1}{n}\frac{1}{\eta(n)+1}\sum_{l=0}^{T^{\eta(n)+1}} f(x_l).$$

Thus, as

$$\lim_{n\to\infty}\frac{\eta(n)}{n} = \lim_{n\to\infty}\frac{T^{\eta(n)}}{n}\frac{\eta(n)}{T^{\eta(n)}} = \frac{1}{\mathbb{E}_i T},$$

by sandwich, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{l=0}^{n} f(x_l) = \lim_{n\to\infty}\frac{\eta(n)}{n}\frac{1}{\eta(n)}\sum_{l=0}^{T^{\eta(n)}} f(x_l) = \frac{1}{\mathbb{E}_i T}\mathbb{E}_i T\int f\mathrm{d}\pi = \int f\mathrm{d}\pi,$$

almost everywhere as required. Now, applying the theorem to f^+, f^-, we obtain the theorem for the general case. □

We note that in the above proof, we had proved the limit

$$\frac{1}{n}\sum_{l=0}^{T^n} f(x_l) \to \mathbb{E}_i T\int f\mathrm{d}\pi.$$

Then, taking $f = \mathbf{1}_{\{j\}}$), we have

$$\frac{1}{n}\sum_{l=0}^{T^n}\mathbf{1}_{\{j\}}(x_l) \to \frac{\pi(j)}{\pi(i)}.$$

In words, this is saying the average time of a chain at j during a excursion to i is the proportion $\pi(j)/\pi(i)$.

Application: Empirical average, Monte Carlo Markov chain (see official notes).

4.6 Reversible Markov Chain

Let (x_n) be an irreducible THMC with transition probability P with invariant probability measure $\pi > 0$ and define $\hat{P}$ such that

$$\hat{P}_{ji} = \frac{\pi(i)}{\pi(j)} P_{ij}.$$

Then, as π is invariant,

$$\sum_{i \in \mathcal{X}} \hat{P}_{ji} = \sum_{i \in \mathcal{X}} \frac{\pi(i)}{\pi(j)} P_{ij} = 1,$$

so $\hat{P}$ is also a transition probability. Now, noting by elementary properties, we have

$$\mathbb{P}(x_n = j \mid x_{n+1} = i) = \frac{\mathbb{P}(x_{n+1} = i \mid x_n = j)\mathbb{P}(x_n = j)}{\mathbb{P}(x_{n+1} = i)} = \frac{\pi(j)}{\pi(i)} P_{ji} = \hat{P}_{ij}$$

if $\mathcal{L}(x_0) = \pi$. Thus, in some sense, if we take a Markov chain and run it backwards in time, we obtain another Markov chain with transition probability $\hat{P}$. This notion made formal with the following theorem.

Theorem 20. Let (x_n) be an irreducible, positively recurrent THMC with initial distribution π which is invariant. Then, for any $M \in \mathbb{N}$, defining $\hat{x}_n := x_{M-n}$, $(\hat{x}_n)$ is a THMC with transition probability $\hat{P}$ and initial distribution π where

$$\hat{P}_{ji} = \frac{\pi(i)}{\pi(j)} P_{ij}.$$

Proof. The initial distribution of $\hat{x}_n$ is π by construction and we observe

$$\begin{aligned}
\mathbb{P}(\hat{x}_0 = i_0, \cdots, \hat{x}_n = i_n) &= \mathbb{P}(x_M = i_0, \cdots, x_{M-n} = i_n) \\
&= \pi(i_n) P_{i_n i_{n-1}} \cdots P_{i_1 i_0} \\
&= \frac{\pi(i_n)}{\pi(i_{n-1})} P_{i_n i_{n-1}} \cdots \frac{\pi(i_1)}{\pi(i_0)} P_{i_1 i_0} \pi(i_0) \\
&= \hat{P}_{i_{n-1} i_n} \cdots \hat{P}_{i_1 i_0} \pi(i_0).
\end{aligned}$$

□

Corollary 20.1. If $\pi(i)P_{ij} = \pi(j)P_{ji}$, then $(\hat{x}_n)$ is a THMC with transition probability P and initial distribution π.

We note that if $\pi(i)P_{ij} = \pi(j)P_{ji}$, then

$$\sum_{i \in \mathcal{X}} \pi(i) P_{ij} = \pi(j) \sum_i P_{ji} = \pi(j)$$

implying π is invariant.

Definition 4.16 (Reversible)**.** A THMC (x_n) is said to be reversible with respect to π if $(\hat{x}_m)$ is a Markov chain with the same transition probability.

4.7 Finite State Markov Chain

In this section, we will consider specifically Markov chains on finite state spaces.

Proposition 4.6. Let $\mathcal{X} = \{1, \cdots, N\}$, the following are equivalent

- P is irreducible and aperiodic;
- P^n is irreducible for every $n \geq 1$;
- Let $\delta_n := \min_{i,j}(P^n)_{ij}$, then there exists some $n_0 \geq 1$ such that $\delta_{n_0} > 0$.

Proof. Clearly the second statement implies the first and in particular, if P has period $d > 1$, then P^d is reducible to each cycle.

Suppose the third statement is true, then P is irreducible since $P_ij^{n_0} > 0$ for all i, j. Furthermore, for any $m \geq 0$, we have

$$P_{jk}^{n_0+m} = \sum_l P_{jl}^m P_{lk}^{n_0} \geq \delta_0 \sum_l P_{jl}^m = \delta_0,$$

in particular, P^n has all positive entries for all $n \geq n_0$. Thus, $n_0, n_0 + 1, \cdots \in R(i)$ and so the chain is aperiodic. Furthermore, as for all $n \leq n_0$, we may find k sufficiently large such that $kn \geq n_0$ and so, $P^{kn} > 0$ implying P^n is irreducible.

Finally, assuming P is irreducible and aperiodic, for all i, there exists some N_i such that $P_{ii}^n > 0$ for all $n \geq N_i$. Thus, for any $i, j \in \mathcal{X}$, as P is irreducible, there exists some m_{ij} such that $P_{ij}^{m_{ij}} > 0$ and so, for all $n \geq N_i + m_{ij}$,

$$P_{ij}^n \geq P_{ii}^{n-m_{ij}} P_{ij}^{m_{ij}} > 0.$$

Thus, taking $n_0 = \max_{i,j} N_i + m_{ij}$, we have the required n_0. □

Lemma 4.19. Let (x_n) be a aperiodic, irreducible THMC on a finite state space with transition probability P. Then for any two states i, j,

$$\mathbb{E}_i(T_j^\alpha) < \infty$$

for any $\alpha \geq 1$.

Proof. As P is irreducible, we have $\mathbb{P}_j(T_i < \infty) = 1$. Furthermore, we note

$$\mathbb{E}_j(T_i^\alpha) = \sum_{n \geq 0} n^\alpha \mathbb{P}_j(T_i = n) \leq \sum_{n \geq 0} n^\alpha \mathbb{P}_j(T_i > n - 1).$$

Now, let n_0 such that $\delta_{n_0} > 0$, we have

$$\begin{aligned}
\mathbb{P}_j(T_i > n_0(k+1)) &\leq \mathbb{P}_j(x_{n_0(k+1)} \neq i, x_{n_0 k} \neq i, \cdots, x_{n_0} \neq i) \\
&= \mathbb{P}_j(x_{n_0(k+1)} \neq i \mid x_{n_0 k} \neq i, \cdots, x_{n_0} \neq i)\mathbb{P}_j((x_{n_0 k} \neq i, \cdots, x_{n_0} \neq i) \\
&= \mathbb{P}_j(x_{n_0(k+1)} \neq i \mid x_{n_0 k} \neq i)\mathbb{P}_j((x_{n_0 k} \neq i, \cdots, x_{n_0} \neq i) \\
&\leq (1 - \delta_{n_0})\mathbb{P}_j(T_j > n_0 k) \leq \cdots \leq (1 - \delta_{n_0})^{k-2},
\end{aligned}$$

where we used the inequality $\mathbb{P}_j(x_{n_0(k+1)} \neq i \mid x_{n_0 k} \neq i) \leq 1 - \delta_{n_0}$.

Now, by considering $\mathbb{P}_j(T_j > n-1) \leq \mathbb{P}_j(T_j > n_0 k - 1) \leq (1-\delta)^k$ for $n_0 k \leq n \leq n_0(k+1)$, we have

$$\mathbb{E}_j(T_i) \leq \sum_{n\geq 0}(1-\delta_{n_0})^{k(n)} \leq \sum_{k=0}^{\infty} n_0^\alpha (k+1)^\alpha (1+\delta)^k = n_0^\alpha \sum_{k=1}^{\infty} k^\alpha (1-\delta_{n_0})^{k-1} < \infty$$

□

Lemma 4.20. Let P be irreducible and aperiodic, then there exists some $n > 0$, $\delta > 0$ such that for every vector $\eta \in \mathbb{R}_+^N$,

$$\eta P^n \geq \delta \|\eta\|_1 \mathbf{1}$$

where $\mathbf{1}$ the the row vector of 1s and $\|\eta\|_1 = \sum_i |\eta_i|$.

Proof. We observe $\eta P^n(j) = \sum_{i\in\mathcal{X}} \eta(i) P_{ij}^n \geq \min_{i,j} P_{ij}^n \sum_{i\in\mathcal{X}} \eta(i)$. Thus, choosing $n_0, and \delta_{n_0}$ as above, the result follows. □

Lemma 4.21. Let P be a irreducible $N \times N$ stochastic matrix and define

$$T_n = \frac{1}{n}(P + P^2 + \cdots + P^n).$$

Then, there exists a number n_0 such that T_{n_0} has only positive entries. Furthermore, for all $\eta \in \mathbb{R}_+^N$,

$$\eta T_{n_0} \geq \delta \|\eta\|_1 \mathbf{1}$$

where $\delta > 0$ is chosen such that $\min_{i,j}(T_{n_0})_{ij} \geq \delta$.

Proof. The result follows from above in the case that P is periodic so suppose P has period $d \geq 1$. Then, $\mathcal{X} = A_0 \sqcup \cdots \sqcup A_{d-1}$ where P^d is aperiodic and irreducible on A_k. Thus, there exists some n_0 such that $P^{n_0 d} > 0$ in A_k. Then, if $j \in A_{k+1}$, as P is irreducible, there exists some $i_0 \in A_k$ such htat $P_{i_0 j} > 0$. Then, for all $i \in A_k$,

$$P_{ij}^{n_0 d + 1} \geq P_{ii_0}^{n_0 d} P_{i_0 j} > 0.$$

Thus, $P_{ij}^{n_0 d} + P_{ij}^{n_0 d+1} > 0$ for all $i, j \in A_k \cup A_{k+1}$. Hence, by induction,

$$P_{ij}^{n_0 d} + \cdots, P_{ij}^{n_0 d + d} > 0,$$

for all $i, j \in \mathcal{X}$ implying $T^n > 0$ for some n. □

Theorem 21 (Perron-Frobenius Theorem)**.** Let P be an irreducible $N \times N$ stochastic matrix. Then, the following holds,

1. 1 is a left eigenvalue of P and there exists exactly one left eigenvector for 1 up to multiplication by a constant.

   Furthermore, π can be chosen to have strictly positive entries and normalised such that $\sum_{i=1}^N \pi(i) = 1$. This vector is called the Perron-Frobenius vector of P.

2. Every eigenvalue λ of P must satisfy $|\lambda| \leq 1$. If P is furthermore aperiodic, all eigenvalues other than 1 satisfy $|\lambda| < 1$.

3. If P is periodic with period d it has eigenvalues

$$\lambda_j = e^{-\frac{2\pi i j}{d}}$$

for $j = 1, \cdots, d$ with associated eigenvectors μ_j such that

$$\mu_j(n) = (\lambda_j)^{-k} \pi(n)$$

if $n \in A_k$ where $\mathcal{X} = A_0 \sqcup \cdots \sqcup A_{d-1}$.

Proof. We will not by using the 3rd. claim for this course and the proof is left in the official notes.

If $\mu P = \lambda \mu$, then $\|\mu P\|_1 = |\lambda| \|\mu\|_1$. But $\|\mu P\|_1 = \sum_{i,j} |\mu(i) P_{ij}| \leq \sum_i |\mu(i)| = \|\mu\|_1$. Thus, $|\lambda| \leq 1$ as claimed.

Consider that $(P\mathbf{1})_k = \sum_{j=1}^N P_{kj} = 1$ for all k, we have $P\mathbf{1} = \mathbf{1}$ implying 1 is an (right)eigenvalue of P with **right**-eigenvector $\mathbf{1}$. Now, since P^T has the same (right)eigenvalues as P, 1 is an eigenvalue of P^T implying 1 is an left-eigenvalue of P. Let $\pi \in \mathbb{R}^N$ such that $\pi P = \pi$ (where we can choose π to be a real vector as P is real). Then, we observe

$$\pi T_n \equiv \frac{1}{n} \pi (P + P^2 + \cdots + P^n) = \frac{1}{n} n \pi = \pi.$$

If $\pi > 0$ or $\pi < 0$ for all its entries, we may simply normalise as claimed, so suppose otherwise, $\alpha := \|\pi^+\|_1 \wedge \|\pi^-\|_1 > 0$. By previous lemmas, as P is irreducible, there exists some n such that $T_n > \delta > 0$ for all of its entries. Then,

$$\pi^+ T_n \geq \delta \alpha \mathbf{1} \text{ and, } \pi^- T_n \geq \delta \alpha \mathbf{1}.$$

So,

$$\begin{aligned}
\|\pi T_n\|_1 &= \|\pi^+ T_n - \delta\alpha\mathbf{1} + \delta\alpha\mathbf{1} - \pi^- T_n\|_1 \\
&\leq \|\pi^+ T_n - \delta\alpha\mathbf{1}\| + \|\pi^- T_n - \delta\alpha\mathbf{1}\| \\
&= \|\pi^+ T_n\| - \|\delta\alpha\mathbf{1}\|_1 + \|\pi^- T_n\| - \|\delta\alpha\mathbf{1}\|_1 \\
&= \|\pi T_n\|_1 - 2\delta\alpha n,
\end{aligned}$$

which is a contradiction for all $\alpha > 0$ and so, either $\pi^+ = 0$ or $\pi^- = 0$.

With this in mind, we may assume $\pi \geq 0$ and $\sum_{i=1}^N \pi(i) = 1$ and so,

$$\pi(i) = \pi T_n(i) = \sum_j \pi(j) T_{ji}^n \geq \delta \sum_j \pi(j) = \delta > 0,$$

and so, π has only positive entries.

Finally, we will prove the uniqueness of the eigenvector. Suppose $\pi, \tilde{\pi}$ are both normalised left-eigenvectors of P of eigenvalue 1. Then,

$$\pi - \tilde{\pi} = (\pi - \tilde{\pi}) P,$$

and we can choose $\pi - \tilde{\pi} \geq 0$. But,

$$0 = \sum_i \pi(i) - \sum_i \tilde{\pi}(i) = \sum_i (\pi - \tilde{\pi})(i),$$

we have $\pi(i) = \tilde{\pi}(i)$ for all i as required. $\square$

Corollary 21.1. An irreducible Markov chain on a finite state space is positively recurrent.

Proof. P-F tells us its unique invariant measure has positive charge at every state. Thus, recalling $\pi(i) = 1/\mathbb{E}_i T_i$, the chain is positively recurrent. $\square$

If $\mu_1, \cdots, \mu_k$ are invariant probability measures for P. Then, so are the convex combinations of these measures a invariant probability measure for P. That is to say, if $a_i \in [0, 1]$ and $\sum_{i=1}^k a_i = 1$, then $\sum_{i=1}^k a_i \mu_i$ is also a invariant probability measure for P (see exercise sheet).

Let us denote $I = \{\pi \in P(\mathcal{X}) \mid \pi P = \pi\}$ and we say $\pi \in I$ is the extremal of I if it cannot be written as a convex combination of other measures in I.

Theorem 22. Let P be a stochastic matrix. Then the set of its invariant probability measures are precisely all convex combinations of the Perron-Frobenius vectors of the restriction of P to its recurrent comminication classes.

Proof. See exercise sheet. $\square$

We observe that the Perron-Frobenius vectors of the restriction of P to its recurrent comminication classes are precisely the extremal of I.

Definition 4.17 (Sub-Stochastic Matrix)**.** An $N \times N$ matrix with non-negative entries P is a sub-stochastic matrix if

$$\sum_{j \in \mathcal{X}} P_{ij} \leq 1.$$

Sub-stochastic matrices are in general obtained by restricting to transient communication classes.

Lemma 4.22. Let P be an irreducible sub-stochastic matrix which is not a stochastic matrix. Then, for any $\mu \in \mathbb{N}$, $\mu P^n \to 0$. Furthermore there exists some $a \in (0, 1)$ such that $\|\mu P^n\|_1 \leq a^n$. In particular, the eigenvalues of P^n has modulus less than 1 and $\mathrm{id} - P$ is invertible.

Proof. Clearly, if $\mu = 0$ then we are done, so suppose otherwise. Then, we may decompose $\mu = \mu^+ - \mu^-$ and so, we may take $\mu > 0$ and hence, normalise it to be a probability measure. Define

$$T^n = \frac{1}{n}(P + \cdots + P^n).$$

Then, as $\|\mu P\|_1 \leq \|\mu\|_1$, we have

$$\|\mu P^{n+1}\|_1 \leq \frac{1}{n}(\|\mu P^{n+1}\|_1 + \cdots + \|\mu P^{n+1}\|_1) \leq \frac{1}{n}(\|\mu P^{n+1}\|_1 + \cdots + \|\mu P\|_1) = \frac{1}{n}\|\mu P T^n\|_1,$$

where the last equality holds as all entries are non-negative.

Then, using the same argument as lemma 4.21, one may show there exists some n_0 such that T^{n_0} has only positive entries. Namely, for all i,

$$\mu T^{n_0}(i) = \sum_{k \in \mathcal{X}} \mu(k) T^{n_0}_{ki} \geq \min_k T^{n_0}_{ki} =: \delta > 0.$$

Now, as P is not a stochastic matrix, there exists some row i_0 such that $\sum_{j\in\mathcal{X}} P_{i_0 j} = 1-\alpha$ for some $\alpha \in (0,1]$. Thus, denoting e_{i_0} the vector which is 1 at the i_0-th position and 0 otherwise, $\|e_{i_0}P\| = 1-\alpha$ and so,

$$\begin{aligned}\|\mu P^{n_0+1}\|_1 &\leq \|\mu P T^{n_0}\|_1 = \|\mu T^{n_0}P - \delta e_{i_0}P + \delta e_{i_0}P\|_1 \\ &\leq \|(\mu T^{n_0} - \delta e_{i_0})P\|_1 + \|\delta e_{i_0}P\|_1 \leq \|\mu T^{n_0} - \delta e_{i_0}\| + \delta(1-\alpha) \\ &= \|\mu T^{n_0}\|_1 - \delta\|e_i\|_1 + \delta(1-\alpha) = \|\mu T^{n_0}\|_1 - \delta\alpha \leq 1 - \delta\alpha.\end{aligned}$$

Thus,

$$\|\mu P^{k(n_0+1)}\|_1 \leq (1-\delta\alpha)^k,$$

and so, defining $a = (1-\delta\alpha)^{\frac{1}{n_0+1}}$, we have

$$\|\mu P^n\|_1 \leq a^n$$

as required. $\square$

Theorem 23 (Minorisation)**.** If there exist some $j_0 \in \mathcal{X}$, $\delta > 0$ such that $P_{ij_0} \geq \delta$ for all $i \in \mathcal{X}$. Then, (μP^n) is a Cauchy sequence for any $\mu \in P(\mathcal{X})$. Furthermore, as $(P(\mathcal{X}), \|\cdot\|_{TV})$ is complete, $\mu P^n \to \pi$ for some $\pi \in P(\mathcal{X})$, $\pi(j_0) \geq \delta$ and

$$\|\mu P^n - \pi\|_1 \leq 2(1-\delta)^n.$$

Proof. Similar proof to above (Hint: $P^n_{ij_0} > 0$). $\square$

As the final topic for finite chains, we will discuss the long run behaviour of a finite Markov chain. As we have seen, in the case a chain is irreducible, aperiodic, positively recurrent (which is automatic for finite chains), then $\lim_{n\to\infty} P^n_{ij} = \pi(j)$ where π is an invariant probability measure. We will now consider the case where the chain is reducible.

Definition 4.18 (Sink)**.** A state $i \in \mathcal{X}$ is said to be a sink if it is recurrent and $[i] = \{i\}$.

Let us denote

$$B_i := \{\omega \in \Omega \mid \exists n_0, x_n(\omega) = i,\ \forall n \geq n_0\}.$$

Proposition 4.7. If $i \in \mathcal{X}$ is a sink, then

$$\lim_{n\to\infty} P^n_{ji} = P_j(B_i).$$

Proof. Let $B^n_i := \{x_n = i\}$. We note B^n_i is increasing and $B_i = \bigcup_n B^n_i$. Thus, by the continuity of measures, we have

$$\lim_{n\to\infty} P^n_{ij} = \lim_{n\to\infty} \mathbb{P}_j(B^n_i) = \mathbb{P}_j\left(\bigcup_n B^n_i\right) = \mathbb{P}_j(B_i).$$

$\square$

Let us now denote $f(j) := \mathbb{P}_j(B_i)$. Then, we find

$$\begin{aligned} f(j) &= \mathbb{E}(\mathbf{1}_{B_i} \mid x_0 = j) = \mathbb{E}(\mathbb{E}(\mathbf{1}_{B_i} \mid \sigma(x_0) \vee \sigma(x_1)) \mid x_0 = j) \\ &= \mathbb{E}(\mathbb{E}(\mathbf{1}_{B_i} \mid x_1) \mid x_0 = j) = \mathbb{E}(f(x_1) \mid x_0 = j) \\ &= \sum_{k \in \mathcal{X}} f(k) P_{jk} = (Pf)(j). \end{aligned}$$

Thus, f is a right eigenvector of P with eigenvalue 1 (and is known as "harmonic"), namely $f = Pf$.

In the case that the minimal class is not a singleton, we may consider the minimal class as a single node and use the above method. From which, we may then work with the minimal class redistributing the probabilities to the ratio of the whole chain.

Proposition 4.8. Let P be a stochastic matrix. Decomposing $\mathcal{X} = T \cup \bigcup_{i=1}^k A_i$ where T are the transient states and A_i are the recurrent communicating classes, then by reordering we can write

$$P = \begin{pmatrix} T & S \\ 0 & \tilde{P} \end{pmatrix},$$

where $\tilde{P} = \bigoplus P_i$ for which P_i are stochastic matrices corresponding to A_i and T is a sub-stochastic matrix. Furthermore, denoting $A_{ij} = \mathbb{P}_i(x_n$ eventually ends in $j)$, we have

$$A := (A_{ij}) = (\mathrm{id} - T)^{-1} S.$$

Proof. Exercise (Hint: show $A_{ij} = (TA)_{ij} + S_{ij}$). □

5 Invariant Measures in General State Space

5.1 Weak Convergence and Feller

We recall the transition operator $T^* : \mu \mapsto (A \mapsto \int P(x, A)\mu(\mathrm{d}x))$ and the dual transition operator $T_* : f \mapsto (x \mapsto \int f(y)P(x, \mathrm{d}y))$, and the relation

$$\int f \mathrm{d}T^*\mu = \int T_* f \mathrm{d}\mu.$$

We note that one may deduce P, T^* and T_* from one another and in general, we will denote T for both T^* and T_*.

Definition 5.1 (Weak Convergence of Measures)**.** A sequence of measures (μ_n) is said to converge weakly to μ if for any bounded continuous real-valued function ϕ,

$$\lim_{n \to \infty} \int \phi \mathrm{d}\mu_n = \int \phi \mathrm{d}\mu.$$

The definition of weak convergence is inspired by the following lemma.

Lemma 5.1. Let μ, ν be measures on a separable complete metric space $\mathcal{X}$. Then, $\mu = \nu$ if for every bounded real-value uniformly continuous function f, we have

$$\int f \mathrm{d}\mu = \int f \mathrm{d}\nu.$$

Furthermore, the space of measures $P(\mathcal{X})$ can be equipped with a topology known as the weak topology which is metrizable in which $\mu_n \to \mu$ weakly if and only if $d(\mu_n, \mu) \to 0$.

Proposition 5.1. If $\mathcal{X}$ is discrete, then any function is continuous. So, $\mu_n \to \mu$ weakly if and only if $\mu_n(A) \to \mu(A$ for all measurable A (choosing $\phi = \mathbf{1}_A$).

Proposition 5.2. If $x_n \to x$ in $\mathcal{X}$, then $\delta_{x_n} \to \delta_x$ weakly.

Proposition 5.3. If $\mathcal{X} = \mathbb{R}$, defining $F_n(x) = \mu((-\infty, x])$ and $F(x) = \mu((-\infty, x])$, $\mu_n \to \mu$ weakly if and only if $F_n(x) \to F(x)$ at all points of continuity of F.

It is easy to check that the above holds, except perhaps the last proposition for which a more general proof is presented in the probability theory notes.

Definition 5.2 (Feller)**.** A time homogeneous Markov process with transition operator T is Feller if Tf is continuous whenever f is bounded continuous.

We note that $T\phi(x) = \int \phi(y)P(x, \mathrm{d}y)$ and so, T is Feller if and only if $x \mapsto P(x, \cdot)$ is continuous in the weak topology of $P(\mathcal{X})$.

Definition 5.3 (Strong-Feller)**.** A time homogeneous Markov process with transition operator T is Strong-Feller if Tf is continuous whenever f is bounded measurable.

Lemma 5.2. Let μ be a probability measure on a complete separable metric space. Then for every $\epsilon > 0$, there exists a compact set K such that $\mu(K) \geq 1 - \epsilon$.

Proof. Recall that totally bounded + complete implies compact. So, as $\mathcal{X}$ is complete, it suffices to find a totally bounded K satisfying $\mu(K) \geq 1 - \epsilon$.

As $\mathcal{X}$ is separable, there exists some $\{x_i\}_{i=1}^\infty \subseteq \mathcal{X}$ dense. So, for all $n \in \mathcal{N}$, $\mathcal{X} = \bigcup_{i=1}^\infty B_{1/n}(x_i)$. Then, by the continuity of measures, there exists some N_n such that

$$\mu\left(\bigcup_{i=1}^{N_n} B_{1/n}(x_i)\right) \geq 1 - \frac{\epsilon}{2^n}.$$

Thus, defining $K := \bigcap_{n=1}^\infty \bigcup_{i=1}^{N_n} B_{1/n}(x_i)$,

$$\mu(K^c) = \mu\left(\bigcup_{n=1}^\infty \left(\bigcup_{i=1}^{N_n} B_{1/n}(x_i)\right)^c\right) \leq \sum_{n=1}^\infty \mu\left(\bigcup_{i=1}^{N_n} B_{1/n}(x_i)\right)^c \leq \sum_{n=1}^\infty \frac{\epsilon}{2^n} = \epsilon,$$

implying $\mu(K) \geq 1 - \epsilon$ as required. Finally, K is totally bounded as for all $\delta > 0$, there exists some n such that $1/n < \delta$, and so, $\{B_{1/n}(x_i) \mid i = 1, \cdots, N_n\}$ is a finite cover of K with each element having radius $1/n < \delta$. □

This lemma motivates the definition of tightness (note the analogy with uniform integrability).

Definition 5.4 (Tight)**.** Let $M \subseteq P(\mathcal{X})$. Then, M is said to be tight if for all $\epsilon > 0$, there exists some compact $K \subseteq \mathcal{X}$ such that

$$\mu(K) \geq 1 - \epsilon$$

for all $\mu \in M$.

Theorem 24 (Prokhorov)**.** Let $\mathcal{X}$ be a separable complete metric space. Then a family $M \subseteq P(X)$ is tight if and only if M is relatively compact (i.e. for all $(\mu_n) \subseteq M$, there exists some $\mu \in P(\mathcal{X})$ such that $\mu_n \to \mu$ weakly).

Proof. See probability theory notes. □

5.2 Invariant Measures and Lyapunov Function Test

Theorem 25 (Krylov-Bogoliubov)**.** Let P be Feller on the complete separable metric space $\mathcal{X}$. If there exists some $x_0 \in \mathcal{X}$ such that the family of measures $\{P^n(x_0, \cdot) \mid n \in \mathbb{N}\} \subseteq P(\mathcal{X})$ is tight, then P has an invariant probability measure.

Proof. Define

$$\mu_N(A) := \frac{1}{N}\sum_{n=1}^N P^n(x_0, A)$$

for all $A \in \mathcal{B}(\mathcal{X})$. Then, $\{\mu_N\}$ is tight (by choosing the same K as $\{P^n(x_0, \cdot)\}$) and by Prokhorov's theorem, there exists some $\pi \in P(\mathcal{X})$ such that $\mu_{N_k} \to \pi$ weakly.

As mentioned previously, $T\pi = \pi$ if $\int f \, \mathrm{d}(T\pi) = \int f \mathrm{d}\pi$ for all bounded continuous functions f, and so, it suffices to show the latter. Indeed, by noting $P(\cdot, A)$ is continuous as T is

Feller,

$$
\begin{aligned}
T\pi(A) = \int P(y,A)\pi(\mathrm{d}y) &= \lim_{k\to\infty}\int P(y,A)\mu_{N_k}(\mathrm{d}y) \\
&= \lim_{k\to\infty}\int P(y,A)\frac{1}{N_k}\sum_{n=1}^{N_k}P^n(x_0,\mathrm{d}y) \\
&= \lim_{k\to\infty}\frac{1}{N_k}\sum_{n=1}^{N_k}\int P(y,A)P^n(x_0,\mathrm{d}y) \\
&= \lim_{k\to\infty}\frac{1}{N_k}\sum_{n=1}^{N_k}P^{n+1}(x_0,A) \\
&= \lim_{k\to\infty}\mu_{N_k}(A)+\frac{1}{N_k}(P^{N_k+1}(x_0,A)-P(x_0,A)).
\end{aligned}
$$

Thus, for all bounded continuous f, as $\int f(y)P^{N_k+1}(x_0,\mathrm{d}y)\leq\|f\|_\infty$,

$$
\begin{aligned}
\int f\,\mathrm{d}(T\pi) &= \lim_{k\to\infty}\int f(y)\mu_{N_k}(\mathrm{d}y)+\frac{1}{N_k}\int f(y)P^{N_k+1}(x_0,\mathrm{d}y)-\frac{1}{N_k}\int f(y)P(x_0,\mathrm{d}y) \\
&= \lim_{k\to\infty}\int f(y)\mu_{N_k}(\mathrm{d}y)=\int f\mathrm{d}\pi
\end{aligned}
$$

as required. □

Corollary 25.1. If $\mathcal{X}$ is compact, any Feller transition probability operator has an invariant probability measure.

Corollary 25.2. If (x_n) is a Markov chain with $\mathcal{L}(x_0)=\delta_{x_0}$ on $\mathbb{R}^n$ with Feller transition probability P. Then, there exist an invariant probability measure if any of the following holds:

- $\sup_n\mathbb{E}|x_n|^p<\infty$ for some $p>0$;
- $\sup_n\mathbb{E}\log(|x_n|+1)<\infty$.

Proof. By definition, $P^n(x_0,\cdot)=\mathcal{L}(x_n)$, and so, for all M,

$$
P^n(x_0,\overline{B_M(0)}^c)=\mathbb{P}(|x_n|>M)\leq\frac{\sup_n\mathbb{E}\log(|x_n|+1)}{\log(M+1)}\to 0,
$$

by Markov's inequality. Thus, $\{P^n(x_0,\cdot)\}$ is tight implying the existence of an invariant measure with Krylov-Bogoliubov.

Similar proof for the first case. □

Proposition 5.4. Let P be a transition function on $\mathcal{X}$ and let $V:\mathcal{X}\to\mathbb{R}_+$ be Borel measurable. Then, if there exists some $\gamma\in(0,1)$ and $c>0$ such that

$$
TV(x)\leq\gamma V(x)+c,
$$

then, $T^nV(x)\leq\gamma^nV(x)+\frac{c}{1-\gamma}$.

Proof.

$$\begin{aligned}T^nV(x) &= \int_{\mathcal{X}} V(y)P^n(x,\mathrm{d}y) = \int_{\mathcal{X}}\int_{\mathcal{X}} V(y)P(z,\mathrm{d}yy)P^{n-1}(x,\mathrm{d}z)\\ &= \int_{\mathcal{X}} TV(z)P^{n-1}(x,\mathrm{d}z) \le \gamma\int_{\mathcal{X}} V(z)P^{n-1}(x,\mathrm{d}y) + c\\ &\le \cdots \le \gamma^n V(x) + \frac{c}{1-\gamma}.\end{aligned}$$

□

Definition 5.5 (Lyapunov Function)**.** Let $\mathcal{X}$ be a complete separable metric space and P a transition probability on $\mathcal{X}$. Then, a Borel measurable function $V : \mathcal{X} \to \overline{\mathbb{R}_+}$ is a Lyapunov function for P if

- $V^{-1}(\mathbb{R}_+) \neq \emptyset$;
- $V^{-1}([0,a])$ is compact for all $a \in \mathbb{R}$;
- there exists some $\gamma < 1$ and c such that $TV(x) \le \gamma V(x) + c$ for all x which $V(x) \neq \infty$.

Theorem 26 (Lyapunov Function Test)**.** If a transition function P is Feller and admits a Lyapunov function V, then, it has an invariant probability measure π.

Proof. Let $x_0 \in \mathcal{X}$ with $V(x_0) < \infty$ and let $a > 0$ and define $K_a := V^{-1}[0,a]$ which is compact. Then,

$$\begin{aligned}P^n(x_0, K_a^c) &= \int_{V(y)>a} P^n(x_0,\mathrm{d}y) \le \int \frac{V(y)}{a}P^n(x_0,\mathrm{d}y)\\ &= \frac{1}{a}T^nV(x_0) \le \frac{1}{a}\left(\frac{c}{1-\gamma} + \gamma^n V(x_0)\right).\end{aligned}$$

Thus, for all $\epsilon > 0$, choosing $a > \frac{1}{\epsilon}\left(\frac{c}{1-\gamma} + V(x_0)\right)$, we have $P^n(x_0, K_a) > 1-\epsilon$ for all n implying $\{P^n(x_0,\cdot)\}$ is tight which implies the existence of an invariant probability measure by Krylov-Bogoliubov. □

Proposition 5.5. Let P be a trnasition function on $\mathcal{X}$ and let $V : \mathcal{X} \to \mathbb{R}_+$ be a Borel measurable function. Then, if there exists some $\gamma \in (0,1)$, $c > 0$ such that $TV(x) \le \gamma V(x) + c$, every invariant probability measure π for P satisfies

$$\int_{\mathcal{X}} V\mathrm{d}\pi \le \frac{c}{1-\gamma}.$$

Proof. Let $M > 0$, then

$$\int V \wedge M\mathrm{d}\pi = \int T^n(V\wedge M)\mathrm{d}\pi \le \int \gamma^n V\wedge M + \frac{c}{1-\gamma}\mathrm{d}\pi.$$

By dominated convergence, by taking $n \to \infty$,

$$\in V \wedge M\mathrm{d}\pi \le \frac{c}{1-\gamma}$$

for all M. Thus, taking $M \uparrow \infty$, allows us to conclude the inequality. □

Corollary 26.1. Let $F : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ be Borel measurable and let (ξ_n) be i.i.d. on $\mathcal{Y}$ all of which are independent of x_0 on $\mathcal{X}$. Then, defining $x_{n+1} := F(x_n, \xi_{n+1})$, we have $TV(x) = \mathbb{E}V(F(x, \xi_n))$.

Now, if $F(\cdot, \xi_n(\omega))$ is continuous for all $\omega \in A$ where A is some set of probability 1, and there exists a Borel measurable function $V : \mathcal{X} \to \mathbb{R}_+$ with compact level sets such that there exists some $\gamma \in (0,1), c \geq 0$,

$$\mathbb{E}V(F(x, \xi_n)) \leq \gamma V(x) + c,$$

then (x_n) is Feller and $(x_\cdot)$ has at least one invariant probability measure.

Proof. The first claim follows by sequential continuity while the second follows straight away by the Lyapunov function test. □

5.3 Deterministic Contraction and Minorisation

So far, with the Lyapunov function test, we have provided a sufficient condition for the existence of an invariant probability measure. We will now consider their uniqueness.

Suppose π_1, π_2 are two probability measures on a complete separable space $\mathcal{X}$. Let μ be the coupling of π_1 and π_2, namely, $\mu \in P(\mathcal{X}^2)$ and $(\mathrm{pr}_1)_*\mu = \pi_1$ and $(\mathrm{pr}_2)_*\mu = \pi_2$ where $\mathrm{pr}_1, \mathrm{pr}_2$ are the two projection maps.

Lemma 5.3. If there exists a coupling μ of π_1 and π_2 such that $\mu(\Delta) = 1$ where $\Delta = \{(x,x) \mid x \in \mathcal{X}\}$, then $\pi_1 = \pi_2$. In particular, $\pi_1 = \pi_2$ if

$$\int_{\mathcal{X}^2} 1 \wedge d(x,y)\mu(\mathrm{d}x, \mathrm{d}y) = 0.$$

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

$$\begin{aligned} \pi_1(A) &= \mu(A \times \mathcal{X}) = \mu((A \times \mathcal{X}) \cap \Delta) \\ &= \mu((\mathcal{X} \times A) \cap \Delta) = \mu(\mathcal{X} \times A) = \pi_2(A) \end{aligned}$$

where the second equality follows as $\mu(\Delta) = 1$. Thus, $\pi_1 = \pi_2$ as required.

Now, by observing that $\{(x,y) \mid 1 \wedge d(x,y) = 0\} = \Delta$, if

$$\int_{\mathcal{X}^2} 1 \wedge d(x,y)\mu(\mathrm{d}x, \mathrm{d}y) = 0$$

then $1 \wedge d(x,y)\mu(\mathrm{d}x, \mathrm{d}y) = 0$ almost everywhere, implying $1 = \mu(\{1 \wedge d(x,y)\mu(\mathrm{d}x, \mathrm{d}y) = 0\}) = \mu(\Delta)$. □

Lemma 5.4. Let $\{\mu_n\}$ be a family of couplings of π_1 and π_2. Then $\{\mu_n\}$ is tight.

Proof. As π_1, π_2 are probability measures, they are themselves tight. Thus, for all $\epsilon > 0$, there exists some compact K_1, K_2 such that $\pi_i(K_i^c) < \epsilon/2$. Then, as $(K_1 \times K_2)^c \subseteq K_1^c \times \mathcal{X} \cup \mathcal{X} \times K_2^c$, we have

$$\mu_i((K_1 \times K_2)^c) \leq \mu(K_1^c \times \mathcal{X}) + \mu(\mathcal{X} \times K_2^c) = \pi_1(K_1^c) + \pi_2(K_2^c) < \epsilon.$$

Hence, as $K_1 \times K_2$ is compact, we have $\{\mu_n\}$ is tight as required. □

Lemma 5.5. If $\{\mu_n\}$ are couplings of π_1 and π_2, then so is any of its accumulation points (also known as limit/cluster points).

Proof. Suppose $\mu_{n_k} \to \mu$ weakly. Then, as the projection map is continuous,

$$\int f \mathrm{d}\pi_i = \lim_{n\to\infty} \int f \circ \mathrm{pr}_i \mathrm{d}\mu_n = \int f \circ \mathrm{pr}_i \mathrm{d}\mu,$$

for all bounded continuous f. Thus, $\int f \mathrm{d}\pi_i = \int f \, \mathrm{d}(\mathrm{pr}_i)_* \mu$ implying $\pi_i = (\mathrm{pr}_i)_* \mu$ as required. $\square$

Lemma 5.6. Let $x_{n+1} = F(x_n, \xi_{n+1}), y_{n+1} = F(y_n, \xi_{n+1})$ be Markov chains where ξ_i are i.i.d. where x_0, y_0 are independent and independent from ξ_i and let $\mu_n = \mathcal{L}((x_n, y_n))$. Then, if there exists some constant $\gamma \in (0, 1)$ such that

$$\mathbb{E}d(F(x, \xi_1), (y, \xi_1)) \leq \gamma d(x, y),$$

we have

$$\lim_{n\to\infty} \mathbb{E}(1 \wedge d(x_n, y_n)) = \lim_{n\to\infty} \int_{\mathcal{X}} 1 \wedge d \mathrm{d}\mu_n = 0$$

Proof. Define $\phi(t) = 1 \wedge t$. By noting that ϕ is convex, we may apply the conditional Jensen's inequality, namely

$$\begin{aligned}
\mathbb{E}(1 \wedge d(x_n, y_n)) &= \mathbb{E}(\mathbb{E}\phi(d(x_n, y_n)) \mid x_{n-1}, y_{n-1}) \\
&\leq \mathbb{E}\phi(\mathbb{E}(d(x_n, y_n) \mid x_{n-1}, y_{n-1})) \\
&= \mathbb{E}\phi(\mathbb{E}d(F(x_{n-1}, \xi_n), F(y_{n-1}, \xi_n))) \\
&\leq \mathbb{E}\phi(\gamma d(x_{n-1}, y_{n-1})) = \mathbb{E}(1 \wedge \gamma d(x_{n-1}, y_{n-1})).
\end{aligned}$$

By iterating this inequality, we obtain $\mathbb{E}(1 \wedge d(x_n, y_n)) \leq \mathbb{E}(1 \wedge \gamma^n d(x_0, y_0))$. Thus, as $1 \wedge \gamma^n d(x_0, y_0) \to 0$ as $n \ to \infty$ almost everywhere, by dominated convergence

$$\lim_{n\to\infty} \mathbb{E}(1 \wedge d(x_n, y_n)) = 0$$

as required. $\square$

Theorem 27 (Deterministic Contraction)**.** Let $x_{n+1} = F(x_n, \xi_{n+1})$ be a Markov chain where ξ_i are i.i.d. Then, if there exists some constant $\gamma \in (0, 1)$ such that

$$\mathbb{E}d(F(x, \xi_1), (y, \xi_1)) \leq \gamma d(x, y)$$

for all $x, y \in \mathcal{X}$, (x_n) has at most one invariant probability measure.

Proof. Let π_1, π_2 be invariant probability measures and let x_0, y_0 be independent random variables both independent from ξ_i such that $\mathcal{L}(x_0) = \pi_1$ and $\mathcal{L}(y_0) = \pi_2$. Then, as π_i are invariant, x_n, y_n has distribution π_1, π_2 respectively for all n.

Now, defining $\mu_i = \mathcal{L}((x_n, y_n))$, $\{\mu_n\}$ is a coupling of π_1 and π_2. By the above lemma, $\{\mu_n\}$ is tight and so, by Prokhorov's theorem, there exists a weakly convergent subsequence μ_{n_k} with limit μ which is also a coupling of π_1 and π_2. Thus, as by the above lemma,

$$\int 1 \wedge d \mathrm{d}\mu = \lim_{k\to\infty} \int 1 \wedge d \mathrm{d}\mu_{n_k} = 0,$$

we have $\pi_1 = \pi_2$ as required. $\square$

Definition 5.6 (Minorisation)**.** Let $\eta \in P(\mathcal{X})$. We say a family of transition probabilities $P = (P(x, \cdot))$ is minorised by η if there exists some $a > 0$ such that for all $x \in \mathcal{X}$,

$$P(x, \cdot) \geq a\eta.$$

In the finite state case, minorisation is saying that $P(i, j) \geq a\eta(j)$ for all $i, j \in \mathcal{X}$. Thus, if we take η to be the vector with 1 in the j_0-th position and 0 everywhere else, P is minorised by η if and only if $P(i, j_0) \geq a$ for all i.

Before moving on, let us introduce another alternative definition for the total variation of measures which will be helpful.

Proposition 5.6. Let μ, ν be positive measures on Ω. Let η be a positive measure such that $\mu \ll \eta$ and $\nu \ll \eta$. Then,

$$\|\mu - \nu\|_{TV} = \int \left| \frac{\mathrm{d}\mu}{\mathrm{d}\eta} - \frac{\mathrm{d}\nu}{\mathrm{d}\eta} \right| \mathrm{d}\eta.$$

We note that such an η always exists by simply taking $\eta = \mu + \nu$.

We note that this formulation is independent of the choice of η. Indeed,

$$\begin{aligned}\int \left| \frac{\mathrm{d}\mu}{\mathrm{d}\eta} - \frac{\mathrm{d}\nu}{\mathrm{d}\eta} \right| \mathrm{d}\eta &= \int \frac{\mathrm{d}(\mu+\nu)}{\mathrm{d}\eta} \left| \frac{\mathrm{d}\mu}{\mathrm{d}(\mu+\nu)} - \frac{\mathrm{d}\nu}{\mathrm{d}(\mu+\nu)} \right| \mathrm{d}\eta \\ &= \int \left| \frac{\mathrm{d}\mu}{\mathrm{d}(\mu+\nu)} - \frac{\mathrm{d}\nu}{\mathrm{d}(\mu+\nu)} \right| \mathrm{d}(\mu+\nu)\,.\end{aligned}$$

Definition 5.7. Given measures μ, ν, we define

$$\mu \wedge \nu := \left(\frac{\mathrm{d}\mu}{\mathrm{d}\eta} \wedge \frac{\mathrm{d}\nu}{\mathrm{d}\eta} \right) \eta$$

where $\mu, \nu \ll \eta$. This definition is independent of the choice of η.

Lemma 5.7. Given measures μ, ν,

$$\|\mu - \nu\|_{TV} = \mu(\Omega) + \nu(\Omega) - 2\mu \wedge \nu(\Omega)$$

which equals $2(1 - \mu \wedge \nu(\Omega))$ if $\mu, \nu \in P(\Omega)$.

Lemma 5.8. The space $P(\mathcal{X})$ is complete under $\|\cdot\|_{TV}$.

Proof. Let (μ_n) be a Cauchy sequence of probability measures and let

$$\eta := \sum_{n=1}^{\infty} \frac{1}{2^n} \mu_n,$$

so that $\mu_n \ll \eta$ for all n.Thus,

$$\|\mu_n - \mu_m\|_{TV} = \int \left| \frac{\mathrm{d}\mu_n}{\mathrm{d}\eta} - \frac{\mathrm{d}\mu_m}{\mathrm{d}\eta} \right| \mathrm{d}\eta.$$

So, (μ_n) is Cauchy if and only if $(\mathrm{d}\mu_n/\mathrm{d}\eta)$ is Cauchy in L^1. As L^1 is complete, there exists some $f \in L^1$ such that $\mathrm{d}\mu_n/\mathrm{d}\eta \to f$ in L^1. So, $\mu_n \to \mu$ in total variation where $\mu = f\eta \in P(\mathcal{X})$. $\square$

Lemma 5.9. Let μ, ν be probability measures on $\mathcal{X}$. Then, denoting

$$\bar{\mu} := \frac{\mu - \mu \wedge \nu}{\frac{1}{2}\|\mu - \nu\|_{TV}},$$

and

$$\bar{\nu} := \frac{\nu - \mu \wedge \nu}{\frac{1}{2}\|\mu - \nu\|_{TV}},$$

$\bar{\mu}, \bar{\nu}$ are probability measures and

$$\mu - \nu = \frac{1}{2}\|\mu - \nu\|_{TV}(\bar{\mu} - \bar{\nu}).$$

Proof. Clear. $\square$

Corollary 27.1. Let μ, ν be probability measures on $\mathcal{X}$ and T a transition operator. Then

$$\|T\mu - T\nu\|_{TV} = \frac{1}{2}\|\mu - \nu\|_{TV}\|T\bar{\mu} - T\bar{\nu}\| \leq \|\mu - \nu\|_{TV}.$$

Theorem 28 (Geometric Convergence Theorem)**.** Suppose P is a transition probability on $\mathcal{X}$ minorised by a probability measure η (i.e. $P(x, \cdot) \geq a\eta$ for some $a \in (0,1)$). Then, P has a unique invariant probability measure π.

Furthermore, if $\mu, \nu \in P(\mathcal{X})$, we have

$$\|T^{n+1}\mu - T^{n+1}\nu\|_{TV} \leq (1-a)^{n+1}\|\mu - \nu\|_{TV}.$$

Proof. If m is a probability measure on $\mathcal{X}$, then

$$Tm = \int_{\mathcal{X}} P(x, \cdot)m(\mathrm{d}x) \geq a\eta.$$

Furthermore, $(Tm - a\eta)(\mathcal{X}) = 1 - a$. So,

$$\begin{aligned}\|Tm - T\tilde{m}\|_{TV} &= \|(Tm - a\eta) - (T\tilde{m} - a\eta)\|_{TV} \\ &\leq (1-a)\left\|\frac{Tm - a\eta}{1-a} - \frac{T\tilde{m} - a\eta}{1-a}\right\|_{TV} \leq 2(1-a).\end{aligned}$$

Hence, for $\mu, \nu \in P(\mathcal{X})$, using the above lemma

$$\|T\mu - T\nu\|_{TV} = \frac{1}{2}\|\mu - \nu\|_{TV}\|T\bar{\mu} - T\bar{\nu}\| \leq \frac{1}{2}\|\mu - \nu\|_{TV}2(1-a) = (1-a)\|\mu - \nu\|_{TV}.$$

Thus, by the Banach fixed point theorem, T has a unique fixed point, namely P has a unique invariant probability measure. $\square$

Corollary 28.1. If π is the invariant probability measure for T,

$$\|T^n\mu - \pi\|_{TV} \leq (1-a)^n\|\mu - \pi\|_{TV}.$$

We note that we may generalise the convergence theorem such that P has a unique invariant probability measure if there exists some n_0, $a \in (0,1)$ $\eta \in P(\mathcal{X})$ such that $P^{n_0}(x,\cdot) \geq a\eta$ by considering the the more general Banach fixed point theorem which only require T^n to be a strict contraction for some n.

5.4 Strong Feller Property

Definition 5.8 (Support)**.** Let μ be a measure on the separable metric space $\mathcal{X}$. Then, the support of μ is the closed set A such that A is the smallest closed set of full-measure, i.e.

$$\text{supp}(\mu) := \bigcap_{\substack{\mu(A^c)=0 \\ A \text{ closed}}} A.$$

Alternatively, the support is the set A such that any open set containing it has positive measure.

Theorem 29. If μ, ν are mutually singular probability measures, and is invariant for a transition operator T. Then, if T has the strong Feller property,

$$\text{supp}(\mu) \cap \text{supp}(\nu) = \emptyset.$$

Proof. As $\mu \perp \nu$, there exists some measurable $F \subseteq \mathcal{X}$ such that $\mu(F) = 1$ and $\nu(F) = 0$. Then, as T is strong Feller, $T\mathbf{1}_F(x) = P(x,F) \in [0,1]$ is continuous. Now, as ν is invariant,

$$0 = \nu(F) = \int \mathbf{1}_F \mathrm{d}\nu = \int \mathbf{1}_F \mathrm{d}T\nu = \int T\mathbf{1}_F \mathrm{d}\nu.$$

Since, $T\mathbf{1}_F(x) = P(x,F) \geq 0$, $\nu(T\mathbf{1}_F^{-1}(\{0\})) = \nu(\{T\mathbf{1}_F = 0\}) = 1$. Similarly, we have $\mu(T\mathbf{1}_F^{-1}(\{1\})) = 1$. Thus, as $T\mathbf{1}_F^{-1}(\{0\}), T\mathbf{1}_F^{-1}(\{1\})$ are closed as $T\mathbf{1}_F$ is continuous, we have

$$\text{supp}(\mu) \cap \text{supp}(\nu) \subseteq T\mathbf{1}_F^{-1}(\{1\}) \cap T\mathbf{1}_F^{-1}(\{0\}) = \emptyset.$$

□

Proposition 5.7. Let $g : \mathbb{R}^n \to \mathbb{R}_+$ be measurable such that $\int g \mathrm{d}\lambda = 1$. If $Tf(x) = \int f(y)g(x-y)\lambda(\mathrm{d}y) = f * g(x)$, then T has strong Feller property.

Proof. This follows from the fact $f : \mathbb{R}^n \to \mathbb{R}$ is bounded measurable and $g : \mathbb{R}^n \to \mathbb{R}$ is in L^1, then $f * g$ is a bounded continuous function. □

Proposition 5.8. Let $P : \mathcal{X}^2 \to \mathbb{R}$ be measurable such that $P(x, \mathrm{d}y) = P(x,y)\mu$ for some measure μ on $\mathcal{X}$. Then, if either (1) and (2) or (1) and (3) holds, P has the strong Feller property, where

1. $P(\cdot, y)$ is continuous for all y.

2. for all x, there exists some $a > 0$ such that

$$\sup_{z \in B_a(x)} P(z, \cdot) \in L^1(\mu).$$

3. for all x, there exists some $a > 0$ such that $\{P(z, y) \mid z \in B_a(x)\}$ is uniformly integrable.

We note that (2) implies (3).

5.5 Invariant Sets

Definition 5.9 (Invariant Sets)**.** Let P be a family of transition probabilities. A Borel set A is P-invariant if $P(x, A) = 1$ for all $x \in A$.

An easy example of an invariant set is a communication class.

It is easy to see that if A is an invariant set of the Markov chain (x_n), then

$$\mathbb{P}(x_0 \in A, \cdots, x_n \in A) = \pi(A)$$

where π is the initial distribution. Furthermore, if $\mathbb{P}_\pi$ is the stationary distribution on $\mathcal{X}^N$ where π is the initial distribution. Then, $\mathbb{P}_\pi(A^n) = \pi(A)$.

Since for an invariant set A, $P(x, A) = 1$ for all $x \in A$, $P|_A$ provides a family of transition probabilities on A. As we in general work with **complete** separable metric space, in order for the restriction to also be complete, we prefer to consider closed invariant sets such that Krylov-Bogoliubov can be applied.

Lemma 5.10. Let A be P-invariant and let π^0 be a probability measure on A, we can define a probability measure π on X with

$$\pi(B) := \pi^0(B \cap A).$$

Then, π^0 is invariant for $P|_A$ if and only if π is invariant for P.

Proof. Denoting T the transition operator, for all B,

$$T\pi(B) = \int_{\mathcal{X}} P(x, B)\pi(\mathrm{d}x) = \int_A P(x, B)\pi(\mathrm{d}x) = \int_A P(x, B \cap A)\pi(\mathrm{d}x) = \pi^0(B \cap A) = \pi(B),$$

where the second equality is due to $\pi(A) = 1$ and the third equality follows as for all $x \in A$, $P(x, B) = P(x, B \cap A)$.

Reverse direction is clear. ☐

Theorem 30. Let P be Feller and suppose there exists a compact P-invariant set A. Then, there exists an invariant probability measure for P.

Proof. Let P^0 be the restriction of P to A. Then, as A is compact, P^0 is tight. Then, for all $f : A \to \mathbb{R}$ bounded continuous, by Tietze's lemma, it extends to a bounded continuous function $\bar{f} : \mathcal{X} \to \mathbb{R}$. Thus, P^0 is Feller. Hence, as P^0 has an invariant probability measure by Krylov-Bogoliubov, the above lemma allows us to conclude P has an invariant probability measure. ☐

We can also use invariant sets to show the uniqueness of the invariant measure provided the invariant set is sufficiently absorbing. Consider the following sequence. Let A be invariant, $A_0 = A$ and $A_{n+1} = \{x \mid P(x, A_n) > 0\}$. We see that (A_n) is a sequence such that elements of A_{n+1} can reach inside A_n in 1 time step with positive probability.

Lemma 5.11. (A_n) is increasing.

Proof. We will show by induction $A_n \subseteq A_{n+1}$. Clearly $A_0 \subseteq A_1$ as $A_0 = A$ and so, for all $x \in A_0$, $P(x, A_0) = 1 > 0$ implying $x \in A_1$. Now, for all n, $x \in A_n$, by the inductive hypothesis $A_{n-1} \subseteq A_n$ and so, $P(x, A_n) \geq P(x, A_{n-1}) > 0$ implying $x \in A_{n+1}$ as required. $\square$

Lemma 5.12. Let A be P-invariant, then for any $n \geq 1$, for any $x \in A_n$, $P^n(x, A) > 0$.

Proof. Clear by Chapman-Kolmogorov. $\square$

Proposition 5.9. Let A be P-invariant. Then, if $\bigcup_{n=0}^\infty A_n = \mathcal{X}$, every invariant probability measure π of P is an invariant probability measure of P on A.

Proof. If $\pi(A) < 1$, then there exists some A_{n_0} with $\pi(A_{n_0} \setminus A) > 0$ (as $\lim_{n \to \infty} \pi(A_n) = 1$). Thus,

$$\begin{aligned}\pi(A) = T^{n_0}\pi(A) = \int P^{n_0}(x, A)\pi(\mathrm{d}x) &\geq \int_{A_{n_0}} P^{n_0}(x, A)\pi(\mathrm{d}x) \\ &= \int_A P^{n_0}(x, A)\pi(\mathrm{d}x) + \int_{A_{n_0} \setminus A} P^{n_0}(x, A)\pi(\mathrm{d}x) > \pi(A)\end{aligned}$$

which is a contradiction. Hence, π is a probability measure on A. Now as π is invariant on A as it is invariant on $\mathcal{X}$, we conclude the claim. $\square$

Corollary 30.1. If A is compact, P-invariant where P is Feller and $\bigcup_{n=0}^\infty A_n = \mathcal{X}$. Then, if there exists some $\gamma < 1$ such that

$$\mathbb{E}d(F(x, \xi_1), F(y, \xi_1)) \leq \gamma d(x, y)$$

for all $x, y \in A$, there exists a unique invariant probability measure for P.

Proof. Applying the deterministic contraction theorem on P restricted to A, we obtain that P has a unique invariant measure on A. Now, by the above lemmas, this invariant measure can be extended to $\mathcal{X}$ implying the existence of an invariant probability measure. On the other hand, if we can another invariant measure, it restricted on A is a invariant measure on A and hence, by uniqueness, they are equal. Thus, we obtain the uniqueness of the invariant measure. $\square$

5.6 Random Dynamical System

We recall the definition of initial value problem. Given $g : \mathbb{R}^n \to \mathbb{R}^n$, $f : \mathbb{R} \to \mathbb{R}^n$ measurable, a initial value problem

$$\begin{cases} \dot{x}(t) &= g(x(t)) + f(t), \\ x(t_0) &= x \end{cases}$$

is said to have solution $x \in C(a,b)$ if $t_0 \in (a,b)$ and for all $t \in (a,b)$,

$$x(t) = x + \int_{t_0}^{t} g(x(s))\mathrm{d}s + \int_{t_0}^{t} f(s)\mathrm{d}s.$$

If g is locally Lipschitz and f is continuous, $\dot{x}$ exists.

We recall the existence and uniqueness of solutions from second year.

Proposition 5.10 (Picard-Lindelöf)**.** If g is locally Lipschitz and f is locally bounded, then for every initial value, there exists a unique maximal solution. Furthermore, if there exists some c such that

$$\langle x, g(x) \rangle \leq c(1 + |x|^2)$$

for all x, the the solution exists on $\mathbb{R}$. This condition is called the one sided linear growth condition.

Proposition 5.11 (Flow Condition)**.** Suppose $\phi_{t_0,t}(x)$ is the global solution to the IVT with initial values $x(t_0) = x$, then,

$$\phi_{u,t}(x) = \phi_{s,t}(\phi_{u,s}(x))$$

for all $u < s < t$. We denote $\phi_t(x) = \phi_{0,t}(x)$.

Theorem 31. Suppose g is globally Lipschitz and f is bounded measurable. Then there exists a unique global solution $\phi_{t_0,t}(x)$. Furthermore, the map $(t,x) \mapsto \phi_t(x)$ is continuous (with repsect to both t, x) and the map $x \mapsto \phi_t(x)$ is differentiable. Denoting $V_t = (D\phi_t)_{x_0}(v_0)$ for all $x_0, v_0 \in \mathbb{R}^n$, V_t is the unique solution to

$$\begin{cases} \dot{V}(t) &= (Dg)_{\phi_t(x_0)}(v(t)), \\ V(0) &= v_0. \end{cases}$$

Finally, $(x,v) \mapsto (Dg)_x(v)$ is continuous and Lipschitz.

Corollary 31.1. If $\langle Dg(x)(v), v \rangle \leq -c(x)|v|^2$ for some c. Then,

$$|v_t| \leq e^{-\int_0^t c(\phi_s(x))\mathrm{d}s}$$

where $v_t = (D\phi_t)_{\phi_t(x)}(v)$.

Proof.

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}|v_t|^2 &= 2\langle v_t, \frac{\mathrm{d}}{\mathrm{d}t}v_t \rangle = 2\langle v_t, (Dg)_{\phi_t(x)}(v_t) \rangle \\ &\leq -2c(\phi_t(x))|v_t|^2, \end{aligned}$$

which implies the result by integrating both sides. $\square$

Thus, if $c(x) \geq c > 0$, then

$$|\phi_t(x) - \phi_t(y)| \leq \|D\phi_t\|_\infty |x-y| \leq e^{-ct}|x-y|.$$

We also introduce a more general system where we allow g to evolve with time. Namely, $g : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ and consider

$$\begin{cases} \dot{x}(t) &= g(t, x(t)), \\ x(t_0) &= x. \end{cases}$$

Proposition 5.12. If $|g(t,x) - g(t,y)| \leq K|x-y|$ i.e. g is Lipschitz with respect to the second argument, then, for any initial value, there exists a unique global solution which is differentiable in time. In particular,

$$x(t) = x_0 + \int_{t_0}^{t} g(\gamma, x(\gamma)) \mathrm{d}\gamma.$$

Also, if $\phi_{t_0,t}(x)$ denotes this solution, the map $x \mapsto \phi_{t_0,t}(x)$ is differentiable (and hence also continuous).

With the above in mind, we come back to Markov processes by constructing a Markov process with a dynamical system. Namely, we will vary f. Denote $\phi_t(x,f)$ the solution to

$$\begin{cases} \dot{x}(t) &= g(x(t)) + f(t), \\ x(0) &= x, \end{cases}$$

and assume that there exists a unique global solution for any initial value. Denoting the solution at time 1 as Φ, i.e. $\Phi(x,f) = \phi_1(x,f)$, if $(\xi_n : \Omega \to C_b(\mathbb{R}))$ is a sequence of continuous iid. random variables (where $C_b(\mathbb{R})$ is the space of bounded continuous functions on $\mathbb{R}$)

$$\begin{cases} \dot{x}(t) &= g(x(t)) + \xi_n(t, \omega), \\ x(0) &= x, \end{cases}$$

we define $x_0 := x, x_1 := \Phi(x, \xi_1), \cdots, x_n := \Phi(x_{n-1}, \xi_n), \cdots$. Indeed, (x_n) is a Markov chain as (ξ_n) are independent.

See official notes for a detailed example of such a random process and a proof that such an process has an invariant measure using P-invariant sets.