

# Markov Process

Kexing Ying

February 28, 2022

## Contents

<b>1</b>	<b>Invariant Measures in General State Space</b>	<b>2</b>
1.1	Weak Convergence and Feller . . . . .	2
1.2	Invariant Measures and Lyapunov Function Test . . . . .	3
1.3	Deterministic Contraction and Minorisation . . . . .	6
1.4	Strong Feller Property . . . . .	10

# 1 Invariant Measures in General State Space

## 1.1 Weak Convergence and Feller

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

$$\int f dT^*\mu = \int T_*f d\mu.$$

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 1.1** (Weak Convergence of Measures). A sequence of measures  $(\mu_n)$  is said to converge weakly to  $\mu$  if for any bounded continuous real-valued function  $\phi$ ,

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

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

**Lemma 1.1.** Let  $\mu, \nu$  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

$$\int f d\mu = \int f d\nu.$$

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

**Proposition 1.1.** If  $\mathcal{X}$  is discrete, then any function is continuous. So,  $\mu_n \rightarrow \mu$  weakly if and only if  $\mu_n(A) \rightarrow \mu(A)$  for all measurable  $A$  (choosing  $\phi = \mathbf{1}_A$ ).

**Proposition 1.2.** If  $x_n \rightarrow x$  in  $\mathcal{X}$ , then  $\delta_{x_n} \rightarrow \delta_x$  weakly.

**Proposition 1.3.** If  $\mathcal{X} = \mathbb{R}$ , defining  $F_n(x) = \mu((-\infty, x])$  and  $F(x) = \mu((-\infty, x])$ ,  $\mu_n \rightarrow \mu$  weakly if and only if  $F_n(x) \rightarrow F(x)$  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 1.2** (Feller). A time homogeneous Markov process with transition operator  $T$  is Feller if  $Tf$  is continuous whenever  $f$  is bounded continuous.

We note that  $T\phi(x) = \int \phi(y)P(x, dy)$  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 1.3** (Strong-Feller). A time homogeneous Markov process with transition operator  $T$  is Strong-Feller if  $Tf$  is continuous whenever  $f$  is bounded measurable.

**Lemma 1.2.** Let  $\mu$  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,  $\{B_{1/n}(x_i) \mid i = 1, \dots, N_n\}$  is a finite cover of  $K$  with each element having radius  $1/n < \delta$ .  $\square$

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

**Definition 1.4** (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 1** (Prokhorov). Let  $\mathcal{X}$  be a separable complete metric space. Then a family  $M \subseteq P(\mathcal{X})$  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  $\mu_n \rightarrow \mu$  weakly).

*Proof.* See probability theory notes.  $\square$

## 1.2 Invariant Measures and Lyapunov Function Test

**Theorem 2** (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  $\mu_{N_k} \rightarrow \pi$  weakly.

As mentioned previously,  $T\pi = \pi$  if  $\int f d(T\pi) = \int f d\pi$  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,

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

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

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

as required.  $\square$

**Corollary 2.1.** If  $\mathcal{X}$  is compact, any Feller transition probability operator has an invariant probability measure.

**Corollary 2.2.** 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  $M$ ,

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

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.  $\square$

**Proposition 1.4.** Let  $P$  be a transition function on  $\mathcal{X}$  and let  $V : \mathcal{X} \rightarrow \mathbb{R}_+$  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,  $T^n V(x) \leq \gamma^n V(x) + \frac{c}{1-\gamma}$ .

*Proof.*

$$\begin{aligned}
T^n V(x) &= \int_{\mathcal{X}} V(y) P^n(x, dy) = \int_{\mathcal{X}} \int_{\mathcal{X}} V(y) P(z, dy) P^{n-1}(x, dz) \\
&= \int_{\mathcal{X}} TV(z) P^{n-1}(x, dz) \leq \gamma \int_{\mathcal{X}} V(z) P^{n-1}(x, dz) + c \\
&\leq \dots \leq \gamma^n V(x) + \frac{c}{1-\gamma}.
\end{aligned}$$

□

**Definition 1.5** (Lyapunov Function). Let  $\mathcal{X}$  be a complete separable metric space and  $P$  a transition probability on  $\mathcal{X}$ . Then, a Borel measurable function  $V : \mathcal{X} \rightarrow \overline{\mathbb{R}}_+$  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  $TV(x) \leq \gamma V(x) + c$  for all  $x$  which  $V(x) \neq \infty$ .

**Theorem 3** (Lyapunov Function Test). If a transition function  $P$  is Feller and admits a Lyapunov function  $V$ , then, it has an invariant probability measure  $\pi$ .

*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,

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

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

**Proposition 1.5.** Let  $P$  be a transition function on  $\mathcal{X}$  and let  $V : \mathcal{X} \rightarrow \mathbb{R}_+$  be a Borel measurable function. Then, if there exists some  $\gamma \in (0, 1)$ ,  $c > 0$  such that  $TV(x) \leq \gamma V(x) + c$ , every invariant probability measure  $\pi$  for  $P$  satisfies

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

*Proof.* Let  $M > 0$ , then

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

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

$$\int V \wedge M d\pi \leq \frac{c}{1-\gamma}$$

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

**Corollary 3.1.** Let  $F : \mathcal{X} \times \mathcal{Y} \rightarrow \mathcal{X}$  be Borel measurable and let  $(\xi_n)$  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  $V : \mathcal{X} \rightarrow \mathbb{R}_+$  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  $(x)$  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.  $\square$

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

**Lemma 1.3.** If there exists a coupling  $\mu$  of  $\pi_1$  and  $\pi_2$  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

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

*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

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

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

**Lemma 1.4.** Let  $\{\mu_n\}$  be a family of couplings of  $\pi_1$  and  $\pi_2$ . Then  $\{\mu_n\}$  is tight.

*Proof.* As  $\pi_1, \pi_2$  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.  $\square$

**Lemma 1.5.** If  $\{\mu_n\}$  are couplings of  $\pi_1$  and  $\pi_2$ , then so is any of its accumulation points (also known as limit/cluster points).

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

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

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

**Lemma 1.6.** Let  $x_{n+1} = F(x_n, \xi_{n+1}), y_{n+1} = F(y_n, \xi_{n+1})$  be Markov chains where  $\xi_i$  are i.i.d. where  $x_0, y_0$  are independent and independent from  $\xi_i$  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

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

*Proof.* Define  $\phi(t) = 1 \wedge t$ . By noting that  $\phi$  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  $1 \wedge \gamma^n d(x_0, y_0) \rightarrow 0$  as  $n \rightarrow \infty$  almost everywhere, by dominated convergence

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

as required.  $\square$

**Theorem 4** (Deterministic Contraction). Let  $x_{n+1} = F(x_n, \xi_{n+1})$  be a Markov chain where  $\xi_i$  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  $\pi_1, \pi_2$  be invariant probability measures and let  $x_0, y_0$  be independent random variables both independent from  $\xi_i$  such that  $\mathcal{L}(x_0) = \pi_1$  and  $\mathcal{L}(y_0) = \pi_2$ . Then, as  $\pi_i$  are invariant,  $x_n, y_n$  has distribution  $\pi_1, \pi_2$  respectively for all  $n$ .

Now, defining  $\mu_i = \mathcal{L}((x_n, y_n))$ ,  $\{\mu_n\}$  is a coupling of  $\pi_1$  and  $\pi_2$ . By the above lemma,  $\{\mu_n\}$  is tight and so, by Prokhorov's theorem, there exists a weakly convergent subsequence  $\mu_{n_k}$  with limit  $\mu$  which is also a coupling of  $\pi_1$  and  $\pi_2$ . Thus, as by the above lemma,

$$\int 1 \wedge dd\mu = \lim_{k \rightarrow \infty} \int 1 \wedge dd\mu_{n_k} = 0,$$

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

**Definition 1.6** (Minorisation). Let  $\eta \in P(\mathcal{X})$ . We say a family of transition probabilities  $P = (P(x, \cdot))$  is minorised by  $\eta$  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  $\eta$  to be the vector with 1 in the  $j_0$ -th position and 0 everywhere else,  $P$  is minorised by  $\eta$  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 1.6.** Let  $\mu, \nu$  be positive measures on  $\Omega$ . Let  $\eta$  be a positive measure such that  $\mu \ll \eta$  and  $\nu \ll \eta$ . Then,

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

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

We note that this formulation is independent of the choice of  $\eta$ . Indeed,

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

**Definition 1.7.** Given measures  $\mu, \nu$ , we define

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

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

**Lemma 1.7.** Given measures  $\mu, \nu$ ,

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

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

**Lemma 1.8.** The space  $P(\mathcal{X})$  is complete under  $\|\cdot\|_{TV}$ .

*Proof.* Let  $(\mu_n)$  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,

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



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

**Lemma 1.9.** Let  $\mu, \nu$  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 4.1.** Let  $\mu, \nu$  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 5** (Geometric Convergence Theorem). Suppose  $P$  is a transition probability on  $\mathcal{X}$  minorised by a probability measure  $\eta$  (i.e.  $P(x, \cdot) \geq a\eta$  for some  $a \in (0, 1)$ ). Then,  $P$  has a unique invariant probability measure  $\pi$ .

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

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

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 5.1.** If  $\pi$  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 more general Banach fixed point theorem which only require  $T^n$  to be a strict contraction for some  $n$ .

## 1.4 Strong Feller Property

**Definition 1.8** (Support). Let  $\mu$  be a measure on the separable metric space  $\mathcal{X}$ . Then, the support of  $\mu$  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 6.** If  $\mu, \nu$  are mutually singular probability measures, and  $\mu$  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  $\nu$  is invariant,

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

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

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

**Proposition 1.8.** Let  $P : \mathcal{X}^2 \rightarrow \mathbb{R}$  be measurable such that  $P(x, dy) = P(x, y)\mu$  for some measure  $\mu$  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  $y$ .

