Markov Process

Kexing Ying

March 9, 2022

Contents

1 Invariant Measures in General State Space

1.1 Weak Convergence and Feller

We recall the transition operator $T^* : \mu \mapsto (A \mapsto \int P(x,A)\mu(\mathrm{d}x))$ 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 1.1 (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 1.1. 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 1.1. 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 1.2. If $x_n \to x$ in $\mathcal{X}$, then $\delta_{x_n} \to \delta_x$ weakly.

Proposition 1.3. If $\mathcal{X} = \mathbb{R}$, defining $F_n(x) = \mu((-\infty, x])$ and $F(x) = \mu((-\infty, x])$, $\mu_n \to \mu$ 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 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 $T\phi(x) = \int \phi(y)P(x,\mathrm{d}y)$ 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 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 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 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 $M \subseteq P(X)$ 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. □

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 $\mu_{N_k} \to \pi$ 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 $\int f(y)P^{N_k+1}(x_0,\mathrm{d}y) \le \|f\|_\infty$,

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

as required. □

Corollary 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,

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

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 1.4. 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

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

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

Proof.

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

□

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 $V : \mathcal{X} \to \overline{\mathbb{R}_+}$ 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 $TV(x) \leq \gamma V(x) + c$ 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 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,

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

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

Proposition 1.5. 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 $TV(x) \leq \gamma V(x) + c$, every invariant probability measure π for P satisfies

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

Proof. Let $M > 0$, then

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

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

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

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

Corollary 3.1. Let $F : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ 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. □

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 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 1.3. 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 1.4. 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 1.5. 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 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 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 4 (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 1.6 (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 1.6. 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 1.7. 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 1.7. 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 1.8. 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 1.9. 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 4.1. 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 5 (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 5.1. 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.

1.4 Strong Feller Property

Definition 1.8 (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 6. 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,

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

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

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

$\square$

Proposition 1.7. Let $g : \mathbb{R}^n \to \mathbb{R}_+$ be measurable such that $\int g \text{d}\lambda = 1$. If $Tf(x) = \int f(y)g(x-y)\lambda(\text{d}y) = f * g(x)$, 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. $\square$

Proposition 1.8. Let $P : \mathcal{X}^2 \to \mathbb{R}$ be measurable such that $P(x, \text{d}y) = P(x, y)\mu$ 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).

1.5 Invariant Sets

Definition 1.9 (Invariant Sets)**.** Let P be a family of transition probabilities. A Borel set B 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 1.10. 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 7. 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 1.11. (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. □

Lemma 1.12. 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. □

Proposition 1.9. 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. □

Corollary 7.1. 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. □

1.6 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 1.10 (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 1.11 (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 8. 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 8.1. 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. □

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 1.12. 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.