1 Background

Definition 1.1. A sequence of random variables $(X_n)_{n\geq 0}$ taking values in a space S is called a Markov chain if for all $x_0, \ldots, x_n \in S$ such that $\mathbb{P}(X_0 = x_0, \ldots, X_{n-1} = x_{n-1}) > 0$ we have

$$\mathbb{P}(X_n = x_n \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}) = \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}) .$$

In other words, the future of the process is independent of the past given the present.

For an event A we write $\mathbb{P}_i(A)$ to denote $\mathbb{P}(A \mid X_0 = i)$.

A Markov chain is defined by its transition matrix P given by

$$P(i,j) = \mathbb{P}(X_1 = j \mid X_0 = i) \quad \forall\, i,j \in S.$$

A probability distribution π is called stationary/invariant if $\pi P = \pi$, i.e. if $X_0 \sim \pi$, then $X_n \sim \pi$ for all $n \geq 0$. A Markov chain is called **reversible** if

$$\pi(x)P(x,y) = \pi(y)P(y,x)\ \forall x,y.$$

Definition 1.2. A Markov chain is called **irreducible** if for all $x,y \in E$ there exists $n \geq 0$ such that $P^n(x,y) > 0$. It is called **aperiodic**, if g.c.d.$\{n \geq 1 : P^n(x,x) > 0\} = 1$ for all x.

Theorem 1.3. *Suppose that X is an irreducible and aperiodic Markov chain on a finite state space with invariant distribution π. Then for all x,y we have*

$$P^t(x,y) \to \pi(y) \text{ as } t \to \infty.$$

Definition 1.4. Let μ, ν be two probability measures on S. The total variation distance between μ and ν is defined to be

$$\|\mu - \nu\|_{\mathrm{TV}} = \max_{A \subseteq S} (\mu(A) - \nu(A)).$$

Exercise 1.5. *Show that*

$$\|\mu - \nu\|_{\mathrm{TV}} = \tfrac{1}{2}\sum_x |\mu(x) - \nu(x)| = \sum_{x:\mu(x)>\nu(x)} \max(\mu(x) - \nu(x), 0).$$

Proposition 1.6. *Let μ and ν be two probability distributions on S. Then*

$$\|\mu - \nu\|_{\mathrm{TV}} = \inf\{\mathbb{P}(X \neq Y) :\ (X,Y) \text{ is a coupling of } \mu \text{ and } \nu\}$$

and there is a coupling achieving the infimum above. We will call this coupling the optimal coupling of μ and ν.

Proof. It is easy to see using the definition of total variation that $\|\mu - \nu\|_{\mathrm{TV}} \leq \mathbb{P}(X \neq Y)$ for any coupling (X,Y). We now construct the optimal coupling. Let $p = \sum_x \min(\mu(x), \nu(x))$. Toss a coin with probability of Heads equal to p. If Heads comes up, then sample Z according to $f(x) = \min(\mu(x), \nu(x))/p$ and set $X = Y = Z$. If Tails comes up, then sample X according to h and Y according to g, where

$$h(x) = \frac{\mu(x) - \nu(x)}{1-p}\mathbf{1}(\mu(x) > \nu(x)) \text{ and } g(x) = \frac{\nu(x) - \mu(x)}{1-p}\mathbf{1}(\nu(x) > \mu(x)).$$

It is easy to check that this is indeed an optimal coupling of μ and ν. $\square$

We now define

$$d(t) = \max_x \left\| P^t(x,\cdot) - \pi \right\|_{\mathrm{TV}} \quad \text{and} \quad \overline{d}(t) = \max_{x,y} \left\| P^t(x,\cdot) - P^t(y,\cdot) \right\|_{\mathrm{TV}}$$

Proposition 1.7. *For all $t > 0$*

$$d(t) \leq \overline{d}(t) \leq 2d(t).$$

Moreover, the function $\overline{\mathrm{d}}$ is submultiplicative, i.e. for all $s, t > 0$

$$\overline{d}(t+s) \leq \overline{d}(t)\overline{d}(s).$$

Proof. We leave the first assertion as an exercise.

For the second one, fix x and y and let (X, Y) be the optimal coupling of $P^s(x,\cdot)$ and $P^s(y,\cdot)$. By the Markov property $P^{s+t}(x,z) = \mathbb{E}\big[P^t(X,z)\big]$ and $P^{s+t}(y,z) = \mathbb{E}\big[P^t(Y,z)\big]$. So

$$\begin{aligned} \left\| P^{s+t}(x,\cdot) - P^{s+t}(y,\cdot) \right\|_{\mathrm{TV}} &\leq \mathbb{E}\left[\mathbf{1}(X \neq Y) \cdot \frac{1}{2} \sum_z \left| P^t(X,z) - P^t(Y,z) \right| \right] \\ &\leq \mathbb{E}\big[\mathbf{1}(X \neq Y) \cdot \overline{d}(t) \big] = \mathbb{P}(X \neq Y)\, \overline{d}(t). \end{aligned}$$

Maximising over x and y and using that (X, Y) is an optimal coupling completes the proof. □

Definition 1.8. A coupling of Markov chains with transition matrix P is a process $(X_t, Y_t)_t$ so that both X and Y are Markov chains with transition matrix P and with possibly different starting distributions. A Markovian coupling of P is a coupling of Markov chains which is itself a Markov chain which also satisfies that for all x, x', y, y'

$$\mathbb{P}\big(X_1 = x' \mid X_0 = x, Y_0 = y\big) = P(x,x') \quad \text{and} \quad \mathbb{P}\big(Y_1 = y' \mid X_0 = x, Y_0 = y\big) = P(y,y').$$

A coupling is called coalescent, if whenever there exists s such that $X_s = Y_s$, then $X_t = Y_t$ for all $t \geq s$.

Theorem 1.9. *Let (X, Y) be a Markovian coalescent coupling with $X_0 = x$ and $Y_0 = y$. Let*

$$\tau_{\mathrm{couple}} = \inf\{t \geq 0 : X_t = Y_t\}.$$

Then

$$\left\| P^t(x,\cdot) - P^t(y,\cdot) \right\|_{\mathrm{TV}} \leq \mathbb{P}_{x,y}(\tau_{\mathrm{couple}} > t)\,.$$

Proof. Since the coupling is Markovian we have

$$P^t(x,x') = \mathbb{P}_{x,y}\big(X_t = x'\big) \quad \text{and} \quad P^t(y,y') = \mathbb{P}_{x,y}\big(Y_t = y'\big)\,.$$

So (X_t, Y_t) is a coupling of $P^t(x,\cdot)$ and $P^t(y,\cdot)$, and hence we get

$$\left\| P^t(x,\cdot) - P^t(y,\cdot) \right\|_{\mathrm{TV}} \leq \mathbb{P}_{x,y}(X_t \neq Y_t) \leq \mathbb{P}_{x,y}(\tau_{\mathrm{couple}} > t)\,.$$

This finishes the proof. □