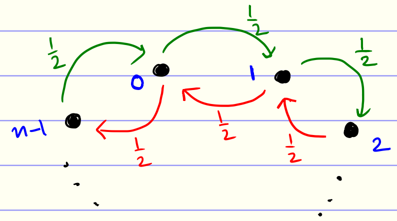
Section 0: Introduction.

In these notes, we discuss the long term behaviour of discrete time Markov chain on a finite space. A fundamental result about Markov chains is that a finite state irreducible aperiodic chain has a unique stationary distribution and regardless of the initial distribution the t-th time distribution of the Markov chain converges to its stationary distribution as t tends to infinity. Now the question is how large must t be until we can say that the Markov chain has reached stationarity (or reached equilibrium). One way is to define "total variation distance" and "mixing time". We will understand these concepts with the help of an example of a Markov chain, lazy $(p-q)$ biased random walk on the n-cycle. Then we will largely focus on finding out how the mixing time in the example changes as a function of the size of its state space. This will give us a general technique to make similar investigations into how the mixing time changes with the size and structure of other irreducible, aperiodic finite space Markov chains.

Section 1: setup and main result

We will consider a simple random walk on $\mathbb{Z}_n = \{0,1,2,\dots,n-1\}$ with the transition probability matrix P given by —



$$P(i,j) = \begin{cases} \frac{1}{2} & \text{if } j \equiv i+1 \bmod n \\ \frac{1}{2} & \text{if } j \equiv i-1 \bmod n \\ 0 & \text{otherwise} \end{cases}$$

Let $t \in \mathbb{N} \cup \{0\}$,

Let $\{Z_t\}_{t\geq 0}$ be a discrete Markov chain on the state space $\mathbb{Z}_n$ which satisfies

$\forall\, i,j \in \mathbb{Z}_n \qquad \mathbb{P}(Z_{t+1} = j \mid Z_t = i) = P(i,j)$, i.e, $\{Z_t\}_{t\geq 0}$ has transition matrix P.

Note that: i) $\{Z_t\}_{t\geq 0}$ is a Markov chain on a finite state space

ii) $\{Z_t\}_{t\geq 0}$ is irreducible

One can easily verify that $\pi(i) = \frac{1}{n}$ $\forall\, i \in \mathbb{Z}_n$ is the unique stationary distribution for $\{Z_t\}_{t\geq 0}$ i.e,

$$\forall j \in \mathbb{Z}_n \quad \pi(j) = \sum_{k \in \mathbb{Z}_n} \pi(k)\, P(k,j)$$

A key objective of these notes is to understand the limiting distribution of Z_t and the speed of convergence.

In order to quantify the speed of convergence we need an appropriate metric for measuring distance between distributions.
We will use the "Total Variation Distance".

Total Variation Distance: Let μ and ν be two probability distributions on $\mathcal{X}$. The total variation distance between μ and ν is denoted by $\|\mu-\nu\|_{TV}$ and is defined as

$$\|\mu-\nu\|_{TV} = \max_{A \subseteq \mathcal{X}} |\mu(A) - \nu(A)|$$

This definition is explicitly probabilistic; The distance between μ and ν is the maximum difference between the probabilities assigned to a single event by the two distributions.

It is well known that if the Markov chain is on a finite state space and is irreducible and aperiodic then it converges to stationary distribution.

Proposition : (Convergence) Suppose that a finite space $(\mathcal{X})$ Markov chain $\{X_t\}_{t\geq 0}$ with transition probability matrix P is irreducible and aperiodic with stationary distribution π. Then there exists constants $\alpha \in (0,1)$ and $C>0$ such that

(David A. Levin and Yuval Peres) page 52

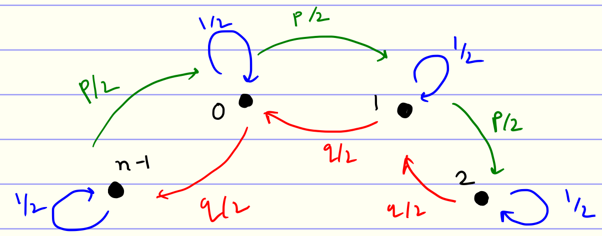
$$\max_{x\in\mathcal{X}} \| \mathbb{P}(X_t \in \cdot \mid X_0 = x) - \pi(\cdot) \|_{TV} \leq C\alpha^t$$

If we wish to apply the above proposition to $\{Z_t\}_{t\geq 0}$ then it would not satisfy the "aperiodic" assumption when the state space is $\mathbb{Z}_n$ with n even.

There is a standard method to make $\{Z_t\}_{t\geq 0}$ aperiodic i.e, we modify the transition probability matrix P of $\{Z_t\}_{t\geq 0}$ as follows:

Choose $0<p<1$, $q:=1-p$

$$\tilde{P}(i,j) = \begin{cases} \frac{1}{2} & j \equiv i \mod n \\ \frac{p}{2} & j \equiv i+1 \mod n \\ \frac{q}{2} & j \equiv i-1 \mod n \\ 0 & \text{otherwise} \end{cases}$$



$\{Z_t\}_{t\geq 0}$ with transition matrix $\tilde{P}$ will be referred to as the lazy $(p-q)$ biased random walk on n-cycle.

And we will denote $\tilde{P}$ by P.

So, by the above theorem, $\{Z_t\}_{t\geq 0}$ has a unique stationary distribution say π

and for each $i \in \mathbb{Z}_n$, $\| P(Z_t \in \cdot \mid Z_0 = i) - \pi(\cdot)\|_{TV} \longrightarrow 0$ as $t\to\infty$

Easy to verify that $\pi(i) = \frac{1}{n}$, $\forall i \in \mathbb{Z}_n$ is still the stationary distribution of $\{Z_t\}_{t\geq 0}$

While applying the "convergence" proposition to $\{Z_t\}_{t\geq 0}$, we see that the rate of convergence is atmost $C\alpha^t$ (i.e, exponential). Here the constant C depends on the size of state space $\mathbb{Z}_n$ and it may be very large depending on n $(\because |\mathbb{Z}_n| = n)$.

So, it may take a very long time to get the distance to stationarity to be small enough.

We want to express this "long time" in terms of n.

The idea is to fix a target distance to reach within stationary distribution and then try to express or bound the time / steps taken to reach that target distance in terms of n.

We denote $d(t) := \max_{i \in \mathbb{Z}_n} \| \mathbb{P}(Z_t \in \cdot \mid Z_0 = i) - \pi(\cdot) \|_{TV}$

We will later show that $d(t)$ is non-increasing in t.

Mixing Time: $t_{mix} := \min \{ t : d(t) \leq \frac{1}{4} \}$

t_{mix} denotes the first time the distance to stationarity is less than or equal to $\frac{1}{4}$.

The constant $\frac{1}{4}$ is neither arbitrary nor it is unique in the definition of t_{mix}. It is an well accepted standard for a "close enough distance" to stationarity. Although any constant lesser than $\frac{1}{2}$ would work. [Will show this in Section 4, in proof of Lemma 2.viii]

So, for $t \geq t_{mix}$, we accept that the Markov chain is close enough to its stationary distribution.

David A. Levin and Yuval Peres Section 5.3.2.

Theorem 1: Let the Markov chain $\{Z_t\}_{t \geq 0}$, with transition probability matrix P be as above. Let π be the stationary distribution of $\{Z_t\}_{t \geq 0}$. Let $d(t)$ and t_{mix} be defined as above. Then

$$\frac{n^2}{32} \leq t_{mix} \leq n^2$$

Section 2: Overview of proof of Theorem 1.

Upper bound: We construct a Markovian Coupling of two particles performing lazy $(p-q)$ biased random walk on $\mathbb{Z}_n$, one particle starting from x and another starting from y.

Here in this proof, the two particle will never move simultaneously ensuring that they will not jump over one another when they come to within unit distance.

Until the two particles meet a fair coin is tossed, independent of all previous tosses, to determine which of the two particles will jump. The particle that is selected makes a clockwise increment with probability p and a counter-clockwise increment with probability q.

Once the two particle collide they make identical moves.

More precisely: Let $\{(X_t, Y_t)\}_{t\geq 0}$ be a Markov chain with the transition probabilities given below:

Choose $a, b \in \mathbb{Z}_n$.

① If $a \neq b$,

$$\mathbb{P}(X_{t+1} = a', Y_{t+1} = b' \mid X_t = a, Y_t = b) = \begin{cases} \frac{p}{2} & \text{if } b' = b,\ a' \equiv a+1 \bmod n \\ \frac{q}{2} & \text{if } b' = b,\ a' \equiv a-1 \bmod n \\ \frac{p}{2} & \text{if } a' = a,\ b' \equiv b+1 \bmod n \\ \frac{q}{2} & \text{if } a' = a,\ b' \equiv b-1 \bmod n \\ 0 & \text{otherwise} \end{cases}$$

② If $a = b$,

$$\mathbb{P}(X_{t+1} = a', Y_{t+1} = b' \mid X_t = a, Y_t = b) = \begin{cases} \frac{p}{2} & \text{if } a' = b',\ a' \equiv a+1 \bmod n \\ \frac{q}{2} & \text{if } a' = b',\ a' \equiv a-1 \bmod n \\ \frac{1}{2} & \text{if } a' = b',\ a' \equiv a \bmod n \\ 0 & \text{otherwise} \end{cases}$$

From the above definition of transition probabilities we get that

$$\mathbb{P}(X_{t+1} = a' \mid X_t = a, Y_t = b) = \begin{cases} \frac{p}{2} & \text{if } a' \equiv a+1 \mod n \\ \frac{q}{2} & \text{if } a' \equiv a-1 \mod n \\ \frac{1}{2} & \text{if } a' \equiv a \mod n \\ 0 & \text{otherwise} \end{cases} = P(a, a')$$

$$\mathbb{P}(Y_{t+1} = b' \mid X_t = a, Y_t = b) = \begin{cases} \frac{p}{2} & \text{if } b' \equiv b+1 \mod n \\ \frac{q}{2} & \text{if } b' \equiv b-1 \mod n \\ \frac{1}{2} & \text{if } b' \equiv b \mod n \\ 0 & \text{otherwise} \end{cases} = P(b, b')$$

We have that $\{(X_t, Y_t)\}_{t \geq 0}$ is a Markovian coupling of transition probability matrix P.

From the transition probabilities, it is also clear that –

if $X_s = Y_s$ then $X_t = Y_t \;\forall\; t \geq s$

We will consider the coupling time of the chain.

Define τ_{couple} to be the coupling time of the chains

$$\tau_{couple} = \min\{t : X_s = Y_s \text{ for all } s \geq t\}$$

We state a proposition which gives an inequality containing mixing time and τ_{couple}

Proposition 1 : $d(t) \leq \max_{x,y \in \mathbb{Z}_n} \mathbb{P}_{x,y}(\tau_{couple} > t)$

$$t_{mix} \leq 4 \max_{x,y \in \mathbb{Z}_n} \mathbb{E}_{x,y}(\tau_{couple})$$

So from proposition 1, we have $t_{mix} \leq 4 \max_{x,y \in \mathbb{Z}_n} \mathbb{E}_{x,y}(\tau_{couple})$ ---- (#)

We need to find the expectation of τ_{couple} given that $X_0 = x$ and $Y_0 = y$

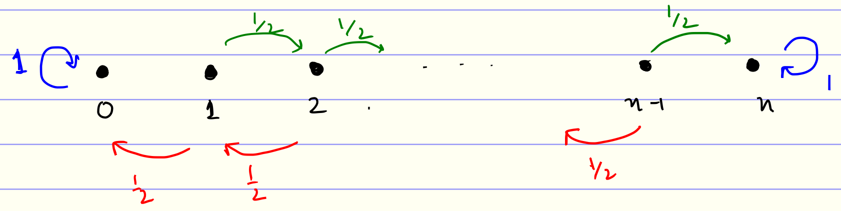
Define for $t = 0$, $D_0 = \min\{a \in \mathbb{N} \cup \{0\} \mid y \equiv x + a \mod n\}$

and for $t > 0$, D_t = clockwise distance from X_t to Y_t

Observe that Range $(D_t) = \{0, 1, 2, \ldots, n\}$ and $\{D_t\}_{t \geq 0}$ is a Markov chain with transition probabilities given by

$i \in \text{Range}(D_t)$ $\quad \mathbb{P}(D_{t+1} = j \mid D_t = i) = \mathbb{P}(\text{clockwise distance from } X_{t+1} \text{ to } Y_{t+1} = j \mid \text{clockwise distance from } X_t \text{ to } Y_t = i)$

$$= \begin{cases} \frac{1}{2} & |j-i| = 1, \; i \in \{1, 2, \ldots, n-1\} \\ 1 & j = 0, i = 0 \\ 1 & j = n, i = n \\ 0 & \text{otherwise} \end{cases}$$



Thus D_t performs simple random walk on $\{1, 2, \ldots, n-1\}$ and gets absorbed at 0 or n

So, $\{D_t\}_{t \geq 0}$ is equivalent to the Gambler's ruin chain

Define $\tau := \min\{t \geq 0 : D_t \in \{0, n\}\}$ = minimum time required for D_t to get absorbed in 0 or n

Note that: the event "$D_t \in \{0, n\}$" is equivalent to the event "$X_s = Y_s \; \forall s \geq t$"

Therefore $\tau_{\text{couple}} = \tau$

If k is the initial value of D_t i.e, at $t = 0$. the clockwise distance between $X_0 = x$ and $Y_0 = y$ is k

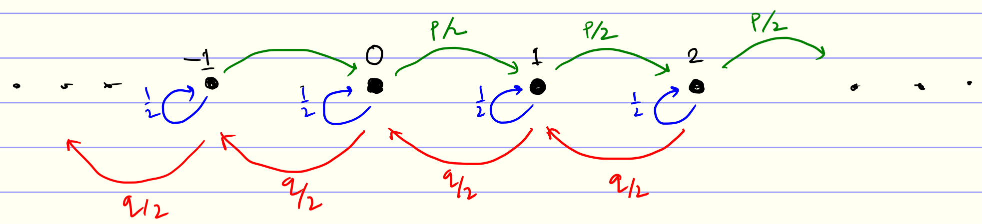
then $\quad E(\tau) = k(n-k)$ $\quad$ [HW 4, Book keeping exercises Q7]

Since $\tau_{\text{couple}} = \tau$, $\quad \mathbb{E}_{x,y}(\tau_{\text{couple}}) = k(n-k) \leq \frac{n}{2}\left(n - \frac{n}{2}\right) = \frac{n^2}{4}$

It follows from (#) that $t_{mix} \leq 4 \max\limits_{x,y \in \mathbb{Z}_n} \mathbb{E}_{x,y}(\tau_{\text{couple}}) \leq 4 \cdot \frac{n^2}{4} = n^2$

Lower bound: (Coupling is not used here)

Let S_t be a lazy $(p-q)$ biased random walk on $\mathbb{Z}$



Define $M_1, M_2, \ldots \overset{iid}{\sim} M \equiv \begin{cases} +1 & \text{w.p. } \frac{p}{2} \\ -1 & \text{w.p. } \frac{q}{2} \\ 0 & \text{w.p. } \frac{1}{2} \end{cases}$

Then $S_0 := 0$ and for $t \geq 1$ $S_t := S_0 + \sum_{i=1}^{t} M_i$

$$\mu_t := E(S_t) \underset{\substack{\downarrow \\ \text{linearity} \\ \text{of Expectation}}}{=} \sum_{i=1}^{t} E(M_i) \underset{\substack{\downarrow \\ \text{iid.}}}{=} tE(M) = t\left(\frac{p}{2}\cdot 1 + \frac{q}{2}\cdot(-1) + \frac{1}{2}\cdot 0\right) = t(p-q)/2$$

Define $X_0 = x_0$ and $X_t \equiv S_t \bmod n$.

Then X_t is a Markov chain on $\mathbb{Z}_n$ with transition probability matrix P.

Therefore $\{X_t\}_{t\geq 0}$ is same as the Markov chain $\{Z_t\}_{t\geq 0}$

So, $\{X_t\}_{t\geq 0}$ has the stationary distribution $\pi = \left(\frac{1}{n} : i \in \mathbb{Z}_n\right)$.

Define for $x, y \in \mathbb{Z}_n$, $\rho(x,y) := \min\{\alpha \in \mathbb{N}\cup\{0\} \mid y \equiv x + \alpha \bmod n \text{ or } y \equiv x - \alpha \bmod n\}$

$= \text{minimum distance between } x \text{ and } y \text{ on the } n\text{-cycle}$

Now for $t > 0$ set $A_t := \{k : \rho(k, \lfloor x_0 + \mu_t \rfloor \bmod n) \geq \frac{n}{4}\}$

See that $|A_t| \geq \frac{n}{2}$

Then $\pi(A_t) \geq \frac{1}{2}$ [Since π is uniform on $\mathbb{Z}_n$]

$$\mathrm{Var}(S_t) = \sum_{i=1}^{t} \mathrm{Var}(M_i) = t\,\mathrm{var}(M) = t\left(E(M^2) - E(M)\right) = t\left(\frac{p}{2}\cdot 1 + \frac{q}{2}\cdot 1 + \frac{1}{2}\cdot 0 - \left(\frac{p-q}{2}\right)^2\right)$$

$$= t\left(\frac{1}{2} - \frac{(p-q)^2}{4}\right)$$

$$= t\left(\frac{1}{2} - \frac{(p+q)^2}{4} + pq\right)$$

$$= t\left(\frac{1}{4} + pq\right)$$

$$\leq \frac{t}{2}$$

For each $t \geq 1$, S_t is a random variable with finite expectation μ_t and finite non-zero variance $\sigma_t^2 = t\left(\frac{1}{4} + pq\right)$.

Using Chebyshev's inequality, $\mathbb{P}\left(|S_t - \mu_t| \geq \frac{n}{4}\right) \leq \frac{16\sigma_t^2}{n^2} \leq \frac{8t}{n^2}$

Note that $\mathbb{P}(X_t \in A_t \mid X_0 = x_0) = \mathbb{P}\left(|S_t - \mu_t| \geq \frac{n}{4}\right) \leq \frac{8t}{n^2}$

So, for $t < \frac{n^2}{32}$, $\mathbb{P}(X_t \in A_t \mid X_0 = x) < \frac{1}{4}$

Then $d(t) \geq \pi(A_t) - \mathbb{P}(X_t \in A_t \mid X_0 = x_0) > \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$ for $t < \frac{n^2}{32}$

$\Rightarrow t_{mix} \geq \frac{n^2}{32}$

$\blacksquare$

Section 3: Here we will give the proof of proposition 1 but in a general setup.

$\{X_t\}_{t\geq 0}$ is an <u>irreducible</u>, <u>aperiodic</u> Markov chain on a <u>finite</u> state space $\mathcal{X}$.

Let the transition probability matrix of $\{X_t\}_{t\geq 0}$ be P.

For $x, x' \in \mathcal{X}$, $\mathbb{P}(X_t = x' \mid X_0 = x) = P^t(x, x')$

$P^t(x,\cdot)$ denotes the xth row of the t-th step transition probability matrix.

Let π be the unique stationary distribution of $\{X_t\}_{t\geq 0}$

We have $d(t) = \max_{x \in \mathcal{X}} \| P^t(x,\cdot) - \pi \|_{TV}$ and $t_{mix} = \min\{t : d(t) \leq \frac{1}{4}\}$

Proposition 1 : (In general setup) Suppose that for each pair of states x, y of $\mathcal{X}$ there is a Markovian coupling (X_t, Y_t) with $X_0 = x$ and $Y_0 = y$. For each such coupling let τ_{couple} be the coupling time of the chains.

David A. Levin, Yuval Peres Corollary 5.5

Then $$d(t) \leq \max_{x,y \in \mathcal{X}} \mathbb{P}_{x,y}(\tau_{couple} > t)$$

and $$t_{mix} \leq 4 \max_{x,y \in \mathcal{X}} \mathbb{E}_{x,y}(\tau_{couple})$$

We will prove this proposition using two other propositions and two lemmas.

Lemma 1: If μ and ν are probability distributions on $\mathcal{X}$, then

$$\|\mu - \nu\|_{TV} = \frac{1}{2}\sum_{x\in\mathcal{X}} |\mu(x) - \nu(x)| = \sum_{\substack{x \in \mathcal{X} \\ \mu(x) \geq \nu(x)}} [\mu(x) - \nu(x)] = \sum_{\substack{x \in \mathcal{X} \\ \nu(x) \geq \mu(x)}} [\nu(x) - \mu(x)]$$

Proof: The proof of this lemma is done in Homework 2 Question 5

Before we give proposition 1, we give the definition of coupling of two probability distributions.

Coupling: A coupling of two probability distribution μ and ν (not necessarily on the same space) is a pair of random variables (X, Y) defined on a single probability space such that the marginal distribution of X is μ and marginal distribution of Y is ν.

That is a coupling (X, Y) satisfies $P_{X,Y}(X = x) = \mu(x)$

and $P_{X,Y}(Y = y) = \nu(y)$

Now we give the proposition 2.

Proposition 2: If μ and ν are two probability distributions on $\mathcal{X}$.

David A. Levin, Yuval Peres Proposition 4.7

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

Remark: We can also show that there exists a coupling (X,Y) of μ and ν such that $\|\mu-\nu\|_{TV} = \mathbb{P}(X\neq Y)$.

Such a coupling is called optimal coupling.

We define Markovian coupling precisely now.

Markovian Coupling: Given a Markov chain on $\mathcal{X}$ with transition matrix P, a Markovian coupling of two P-chains is a Markov chain $\{(X_t, Y_t)\}_{t\geq 0}$ with state space $\mathcal{X}\times\mathcal{X}$ which satisfies for all x, y, x', y',

$$\mathbb{P}(X_{t+1} = x' \mid X_t = x, Y_t = y) = P(x,x')$$

$$\mathbb{P}(Y_{t+1} = y' \mid X_t = x, Y_t = y) = P(y,y')$$

Remark: Any Markovian coupling of Markov chains with transition matrix P can be modified so that the two chains stay together at all times after their first simultaneous visit to a single state, i.e,

(proof of this remark is given in Section 4)

$$\text{if } X_s = Y_s \text{, then } X_t = Y_t \quad \forall t\geq s \quad \text{-----} \quad (*)$$

Proposition 3: Let $\{(X_t, Y_t)\}_{t\geq 0}$ be a Markovian coupling of P with $X_0 = x$ and $Y_0 = y$. Let the coupling satisfy $(*)$.

David A. Levin, Yuval Peres Theorem 5.4

Then we have $\tau_{couple} = \min\{t : X_s = Y_s \text{ for all } s\geq t\}$

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

(Proof in Section 4)

Now we introduce another quantity $\bar{d}(t) := \max_{x,y\in\mathcal{X}} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{TV}$.

This would help in proving certain results on $d(t)$ and understand its characteristic properties. (further given in Section 4)

Lemma 2: $d(t) \leq \bar{d}(t) \leq 2d(t)$ (Proof in Section 4)

David A. Levin, Yuval Peres Section 4.4

Assuming Propn 2, Propn. 3, lemma 1 and 2 we prove proposition 1.

Proof of proposition 1:

From Proposition 3, we have for x, y.

$$\|P^t(x,\cdot) - P^t(y,\cdot)\|_{TV} \leq \mathbb{P}_{x,y}(Z_{couple} > t)$$

Taking the maximum over $x,y \in \mathcal{X}$ on the right side,

$$\|P^t(x,\cdot) - P^t(y,\cdot)\|_{TV} \leq \max_{x,y} \mathbb{P}_{x,y}(Z_{couple} > t)$$

Now we take the maximum over $x,y \in \mathcal{X}$ on the left side, [since the right-hand side is an upper bound for all values of the left side]

$$\max_{x,y\in\mathcal{X}} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{TV} \leq \max_{x,y\in\mathcal{X}} \mathbb{P}_{x,y}(Z_{couple} > t)$$

$$\Rightarrow \quad \bar{d}(t) \leq \max_{x,y} \mathbb{P}_{x,y}(Z_{couple} > t)$$

Using Lemma 2, we have $d(t) \leq \bar{d}(t) \leq \max_{x,y} \mathbb{P}_{x,y}(Z_{couple} > t)$ ...... (##)

Now, observe that Z_{couple} is a non-negative random variable

For $t>0$, Using Markov's inequality we have,

$$\mathbb{P}_{x,y}(Z_{couple} > t) \leq \frac{\mathbb{E}_{x,y}(Z_{couple})}{t}$$

So, we can rewrite (##) as

$$d(t) \leq \max_{x,y} \frac{\mathbb{E}_{x,y}(Z_{couple})}{t}$$

$$\Rightarrow \quad d(t) \leq \frac{1}{t} \max_{x,y} \mathbb{E}_{x,y}(Z_{couple})$$

Now take $t = t_{mix}(\frac{1}{4}) = t_{mix}$, then $d(t_{mix}) \leq \frac{1}{4}$

So, $t_{mix}\, d(t_{mix}) \leq \max_{x,y} \mathbb{E}_{x,y}(Z_{couple})$

$$\Rightarrow \quad t_{mix} \leq \frac{1}{d(t_{mix})} \max_{x,y} \mathbb{E}_{x,y}(Z_{couple})$$

$$\Rightarrow \quad t_{mix} \leq 4 \max_{x,y} \mathbb{E}_{x,y}(Z_{couple})$$

Section 4: Here we give the proofs of lemma 2, Proposition 2 and 3 and proof of the remark (*).

Proof of Proposition 2:

For any coupling (X,Y) of μ and ν and any event $A \subset \mathcal{X}$.

$$\mu(A)-\nu(A) = \mathbb{P}(X\in A) - \mathbb{P}(Y\in A) = \mathbb{P}(X\in A, Y\notin A) + \underbrace{\mathbb{P}(X\in A, Y\in A) - \mathbb{P}(Y\in A)}_{\leq 0}$$

$$\leq \mathbb{P}(X\in A, Y\notin A)$$

$$\leq \mathbb{P}(X\neq Y)$$

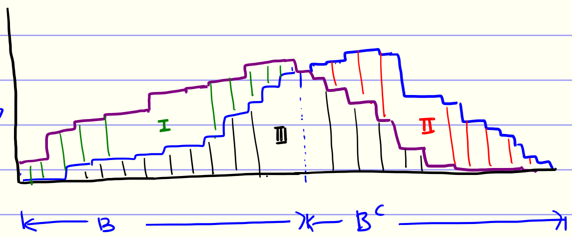
$$\Rightarrow \|\mu-\nu\|_{TV} \leq \inf\{\mathbb{P}(X\neq Y) : (X,Y) \text{ is a coupling of } \mu \text{ and } \nu\}$$

Other direction:

Intuition

Area of Region I $= \sum_{x\in B} [\mu(x)-\nu(x)] = \|\mu-\nu\|_{TV}$

Area of Region II $= \sum_{x\in B^c} [\nu(x)-\mu(x)] = \|\mu-\nu\|_{TV}$



Area of Region III $= 1-\|\mu-\nu\|_{TV}$

We will construct a coupling such that $\mathbb{P}(X\neq Y) = \|\mu-\nu\|_{TV}$

we do so by forcing X and Y to be equal as often as possible

Region III is bounded by $\mu(x)\wedge\nu(x) = \min\{\mu(x),\nu(x)\}$

Our coupling proceeds by choosing a point in the union of the Region I and Region III and setting X to be the x-coordinate of the point.

If the point is in III, then we set $Y=X$.

If the point is in I, then we sample a point from II independently and set Y to be the xth coordinate of that point.

Formal Setup:

Define $p(x) = \mu(x)\wedge\nu(x) = \min\{\mu(x),\nu(x)\}$ for $x\in\mathcal{X}$

$$\sum_{x\in\mathcal{X}} \mu(x)\wedge\nu(x) = \sum_{\substack{x\in\mathcal{X}\\ \mu(x)\leq\nu(x)}} \mu(x) + \sum_{\substack{x\in\mathcal{X}\\ \mu(x)>\nu(x)}} \nu(x)$$

$$= \underbrace{\sum_{\substack{x\in\mathcal{X}\\ \mu(x)\leq\nu(x)}} \mu(x) + \sum_{\substack{x\in\mathcal{X}\\ \mu(x)>\nu(x)}} \mu(x)}_{=1} + \sum_{\substack{x\in\mathcal{X}\\ \mu(x)>\nu(x)}} \nu(x) - \sum_{\substack{x\in\mathcal{X}\\ \mu(x)>\nu(x)}} \mu(x).$$

$$= 1 - \sum_{\substack{x \in X \\ u(x) > v(x)}} [u(x) - v(x)] = 1 - \sum_{\substack{x \in X \\ u(x) \geq v(x)}} [u(x) - v(x)]$$

$$= 1 - \| u - v \|_{TV}$$

Now we flip a coin with probability of heads equal to p

1. If the coin comes up heads, then sample Z according to the probability distribution $\gamma_{III}(x) = \frac{u(x) \wedge v(x)}{p}$.

   and set $X = Y = Z$

2. If the coin comes up tails, then sample X according to the probability distribution –

$$\gamma_I(x) = \begin{cases} \frac{u(x) - v(x)}{\| u - v \|_{TV}} & \text{if } u(x) \geq v(x) \\ 0 & \text{otherwise} \end{cases}$$

   and sample Y according to the probability distribution

$$\gamma_{II}(x) = \begin{cases} \frac{v(x) - u(x)}{\| u - v \|_{TV}} & \text{if } v(x) \geq u(x) \\ 0 & \text{otherwise} \end{cases}$$

Lemma 1 ensures that $\gamma_I(x)$ and $\gamma_{II}(x)$ are probability distributions

Observe that,

$$p\gamma_{III} + (1-p)\gamma_I = u$$

and

$$p\gamma_{III} + (1-p)\gamma_{II} = v$$

so that the distribution of X is u and of Y is v.

In case the coin lands tails up, then $X \neq Y$ since γ_I and γ_{II} are positive on disjoint subsets of X.

$X = Y$ if and only if the coin lands heads up

So, $P(X \neq Y) = \mathbb{P}(\text{coin lands tails up}) = 1 - p = \| u - v \|_{TV}$

Proof of Proposition 3 :

Observe that $p^t(x,z) = \mathbb{P}(X_t = z \mid X_0 = x)$

$$= \mathbb{P}(X_t = z \mid X_0 = x, Y_0 = y)$$

$$= \mathbb{P}_{x,y}(X_t = z)$$

Similarly, $p^t(y,z) = \mathbb{P}_{x,y}(Y_t = z)$

We see that (X_t, Y_t) has marginal distribution $p^t(x,\cdot)$ and $p^t(y,\cdot)$ and (X_t, Y_t) is defined on a single probability space $\mathcal{X} \times \mathcal{X}$.
So, (X_t, Y_t) is a coupling of $p^t(x,\cdot)$ and $p^t(y,\cdot)$.

From Proposition 1, we get $\| p^t(x,\cdot) - p^t(y,\cdot) \|_{TV} \leq \mathbb{P}_{x,y}(X_t \neq Y_t)$

If $X_t \neq Y_t$, then $\tau_{couple} > t$ [Follows from definition of τ_{couple}]

Since $\{(X_t, Y_t)\}$ satisfies $(*)$,

$$\tau_{couple} = \min\{t : X_s = Y_s \text{ for all } s \geq t\}$$

$$= \min\{t : X_t = Y_t\}$$ [$X_t = Y_t$ implies that $X_s = Y_s \ \forall s \geq t$]

Now if $\tau_{couple} > t$, then $X_t \neq Y_t$

So, $\{X_t \neq Y_t\} \equiv \{\tau_{couple} > t\}$

$$\Rightarrow \mathbb{P}_{x,y}(X_t \neq Y_t) = \mathbb{P}_{x,y}(\tau_{couple} > t)$$

Therefore $\| p^t(x,\cdot) - p^t(y,\cdot) \|_{TV} \leq \mathbb{P}_{x,y}(\tau_{couple} > t)$

Proof of lemma 2 : $d(t) \underset{(2)}{\leq} \bar{d}(t) \underset{(1)}{\leq} 2d(t)$

1) $\| p^t(x,\cdot) - p^t(y,\cdot) \|_{TV} \overset{\text{Triangle Inequality}}{\leq} \| p^t(x,\cdot) - \pi \|_{TV} + \| \pi - p^t(y,\cdot) \|_{TV}$

Taking the maximum over $x \in X$ on the right side,

$$\| p^t(x,\cdot) - p^t(y,\cdot) \|_{TV} \leq \max_{x \in X} \| p^t(x,\cdot) - \pi \|_{TV} + \max_{x \in X} \| \pi - p^t(y,\cdot) \|_{TV}$$

$$\Rightarrow \| p^t(x,\cdot) - p^t(y,\cdot) \|_{TV} \leq 2d(t)$$

Taking the maximum over $x,y \in X$ on the left side,

$$\Rightarrow \max_{x,y \in X} \| p^t(x,\cdot) - p^t(y,\cdot) \| \leq 2d(t)$$

$$\Rightarrow \bar{d}(t) \leq 2d(t)$$

2) Since π is a stationary distribution,

$$\pi(A) = \sum_{y\in\mathcal{X}} \pi(y)\, P^t(y,A) \quad \text{and} \quad \sum_{y\in\mathcal{X}} \pi(y) = 1$$

$$|P^t(x,A) - \pi(A)| = \left| \sum_{y\in\mathcal{X}} \pi(y) \left[P^t(x,A) - P^t(y,A) \right] \right|$$

$$\leq \sum_{y\in\mathcal{X}} \pi(y)\, \| P^t(x,\cdot) - P^t(y,\cdot) \|_{TV}$$

$$\leq \| P^t(x,\cdot) - P^t(y,\cdot) \|_{TV}$$

$$\leq \max_{x,y\in\mathcal{X}} \| P^t(x,\cdot) - P^t(y,\cdot) \|_{TV} = \bar{d}(t)$$

Now, we take the maximum over $A \subseteq \mathcal{X}$ on the left side,

$$\max_{A\subseteq\mathcal{X}} |P^t(x,A) - \pi(A)| \leq \bar{d}(t)$$

$$\Rightarrow \quad \| P^t(x,\cdot) - \pi \|_{TV} \leq \bar{d}(t)$$

Now, maximize over $x \in \mathcal{X}$,

$$d(t) = \max_{x\in\mathcal{X}} \| P^t(x,\cdot) - \pi \|_{TV} \leq \bar{d}(t)$$

We state other results that are useful in understanding properties of $d(t)$.
We write these as a continuation of Lemma 2.

Lemma 2 : i) $d(t) \leq \bar{d}(t) \leq 2d(t)$ (Proved above

Let $\mathcal{P}$ be the set containing all probability distributions on $\mathcal{X}$.

ii) $d(t) = \sup_{\mu\in\mathcal{P}} \| \mu P^t - \pi \|_{TV}$

iii) $\bar{d}(t) = \sup_{\mu,\nu\in\mathcal{P}} \| \mu P^t - \nu P^t \|_{TV}$

iv) $\bar{d}(s+t) \leq \bar{d}(s)\, \bar{d}(t)$

v) $\mu,\nu \in \mathcal{P}$, then $\| \mu P - \nu P \|_{TV} \leq \| \mu - \nu \|_{TV}$
Using this show $\| \mu P^{t+1} - \pi \| \leq \| \mu P^t - \pi \|$
ie, advancing in time t, the chain can only move closer to π.

vi) $d(t)$, $\bar{d}(t)$ are non-increasing in t

vii) c and t are positive integers, then $d(ct) \leq \bar{d}(ct) \leq (\bar{d}(t))^c$

viii) Define $t_{mix}(\varepsilon) := \min\{ t : d(t) \leq \varepsilon \}$.
Using vii) show that $t_{mix}(\varepsilon) \leq \lceil \log_2 \varepsilon^{-1} \rceil\, t_{mix}$

Proof of ii) $d(t) = \sup_{\mu} \| \mu P^t - \pi \|_{TV}$

$$\| \mu P^t - \pi \|_{TV} = \frac{1}{2} \sum_{y \in X} | \mu P^t(y) - \pi(y) |$$

since $\sum_{x \in X} \mu(x) = 1$.

and $\sum_{x \in X} \mu(x)\pi(y) = \pi(y) \sum_{x \in X} \mu(x)$

$$= \frac{1}{2} \sum_{y \in X} \left| \sum_{x \in X} \mu(x) P^t(x,y) - \sum_{x \in X} \mu(x) \pi(y) \right|$$

$$\leq \frac{1}{2} \sum_{y \in X} \sum_{x \in X} \mu(x) | P^t(x,y) - \pi(y) |$$

$$\leq \sum_{x \in X} \mu(x) \frac{1}{2} \sum_{y \in X} | P^t(x,y) - \pi(y) |$$

$$\leq \sum_{x \in X} \mu(x) \| P^t(x,\cdot) - \pi \|_{TV}$$

$$\leq \max_{x \in X} \| P^t(x,\cdot) - \pi \|_{TV} = d(t)$$

Taking the maximum over μ on the left side,

$$\max_{\mu} \| \mu P^t - \pi \|_{TV} \leq d(t)$$

For the other direction, we have $d(t) = \max_{x \in X} \| P^t(x,\cdot) - \pi \|_{TV}$.

since X is a finite set, let the maximum occur for say x_0,

ie, $d(t) = \max_{x \in X} \| P^t(x,\cdot) - \pi \|_{TV} = \| P^t(x_0,\cdot) - \pi \|_{TV}$

Now define $\mu(x)$ a probability distribution on X such that

$$\mu(x_0) = 1 \quad \text{and} \quad \mu(x) = 0 \text{ for } x \neq x_0$$

then $\| \mu P^t - \pi \|_{TV} = \frac{1}{2} \sum_{y \in X} | \mu P^t(y) - \pi(y) |$

$$= \frac{1}{2} \sum_{y \in X} \left| \sum_{x \in X} \mu(x) P^t(x,y) - \pi(y) \right|$$

$$= \frac{1}{2} \sum_{y \in X} | \mu(x_0) P^t(x_0,y) - \pi(y) |$$

$$= \frac{1}{2} \sum_{y \in X} | P^t(x_0,y) - \pi(y) |$$

$$= \| P^t(x_0,\cdot) - \pi \|_{TV}$$

$$= d(t).$$

So, there exists an u such that $d(t) = \|u P^t - \pi\|_{TV}$

$$\Rightarrow \quad d(t) = \sup_{u \in \mathcal{P}} \|u P^t - \pi\|_{TV}$$

Proof of iii) $\quad \bar{d}(t) = \sup_{u,v \in \mathcal{P}} \|u P^t - v P^t\|_{TV}$.

Let u, v be two probabilities on space $\mathcal{X}$.

$$\|u P^t - v P^t\|_{TV} = \frac{1}{2} \sum_{x \in X} \left| u P^t(x) - \sum_{y \in X} v(y) P^t(y, z) \right|$$

$$\leq \frac{1}{2} \sum_{x \in X} \sum_{y \in X} v(y) \, |u P^t(x) - P^t(y, x)|$$

$$\leq \sum_{y \in X} v(y) \, \frac{1}{2} \sum_{x \in X} |u P^t(x) - P^t(y, x)|$$

$$\leq \sum_{y \in X} v(y) \, \|u P^t - P^t(y, \cdot)\|_{TV}$$

$$\leq \max_{y \in X} \|u P^t - P^t(y, \cdot)\|_{TV}$$

So, we have $\|u P^t - v P^t\|_{TV} \leq \max_{y \in X} \|u P^t - P^t(y, \cdot)\|_{TV}$

Now take δ_y where $\delta_y(z) = \begin{cases} 1 & \text{if } z = y \\ 0 & \text{otherwise} \end{cases}$

and u to be the two distributions

$$\Rightarrow \|\delta_y P^t - u P^t\|_{TV} \leq \max_{x \in X} \|\delta_y P^t - P^t(x, \cdot)\|_{TV}$$

$$\Rightarrow \|P^t(y, \cdot) - u P\|_{TV} \leq \max_{x \in X} \|P^t(y, \cdot) - P^t(x, \cdot)\|_{TV}$$

Taking maximum over y on the both sides

$$\max_{y \in X} \|P^t(y, \cdot) - u P^t\|_{TV} \leq \max_{x, y \in X} \|P^t(y, \cdot) - P^t(x, \cdot)\|_{TV}$$

$$\Rightarrow \|u P^t - v P^t\|_{TV} \leq \max_{y \in X} \|u P^t - P^t(y, \cdot)\|_{TV} \leq \max_{x, y \in X} \|P^t(y, \cdot) - P^t(x, \cdot)\|_{TV} = \bar{d}(t)$$

$$\Rightarrow \sup_{u, v} \|u P^t - v P^t\|_{TV} \leq \max_{x, y \in X} \|P^t(y, \cdot) - P^t(x, \cdot)\|_{TV} = \bar{d}(t)$$

Since X is finite, then suppose that $\max_{x, y \in X} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV}$ occurs for x_0, y_0.

then take $u = \delta_{x_0}$ and $v = \delta_{y_0}$,

then $\|u P^t - v P^t\| = \|P^t(x_0, \cdot) - P^t(y_0, \cdot)\|_{TV}$

Therefore, $\bar{d}(t) = \sup_{u, v \in \mathcal{P}} \|u P^t - v P^t\|_{TV}$

Proof of iv) $\bar{d}(s+t) \leq \bar{d}(s)\,\bar{d}(t)$

Fix $x, y \in \mathcal{X}$. Let (X_s, Y_s) be the optimal coupling of

$P^s(x,\cdot)$ and $P^s(y,\cdot)$ [Existence is guaranteed by Propn 1]

$\mathbb{P}$ denotes the joint distribution of (X_s, Y_s).

Then $\| P^s(x,\cdot) - P^s(y,\cdot) \| = \mathbb{P}(X_s \neq Y_s)$

$$P^{s+t}(x,w) = \sum_z \mathbb{P}(X_s = z)\, P^t(z,w) = \mathbb{E}\left(P^t(X_s, w)\right)$$

For a set A, summing over $w \in A$ shows that

$$\begin{aligned} P^{s+t}(x,A) - P^{s+t}(y,A) &= \mathbb{E}\left(P^t(X_s,A) - P^t(Y_s,A)\right) \\ &\leq \mathbb{E}\left(\bar{d}(t)\, \mathbb{1}_{\{X_s \neq Y_s\}} \right) \\ &\leq \mathbb{P}(X_s \neq Y_s)\, \bar{d}(t) \end{aligned}$$

[$\because$ X_s, Y_s has marginal distribution $P^s(x,\cdot)$ and $P^s(y,\cdot)$ respectively]

$$\mathbb{P}(X_s \neq Y_s) = \| P^s(x,\cdot) - P^s(y,\cdot)\|_{TV} \leq \max_{x,y} \| P^s(x,\cdot) - P^s(y,\cdot)\|_{TV} = \bar{d}(s)$$

$$\Rightarrow \bar{d}(s+t) \leq \mathbb{P}(X_s \neq Y_s)\, \bar{d}(t) \leq \bar{d}(s)\, \bar{d}(t)$$

Proof of v) $\| \mu P - \nu P \|_{TV} \leq \| \mu - \nu \|_{TV}$.

$$\begin{aligned} \| \mu P - \nu P \|_{TV} &= \frac{1}{2} \sum_{x \in \mathcal{X}} | \mu P(x) - \nu P(x) | \\ &= \frac{1}{2} \sum_{x \in \mathcal{X}} \Big| \sum_{y \in \mathcal{X}} \mu(y) P(y,x) - \sum_{y \in \mathcal{X}} \nu(y) P(y,x) \Big| \\ &= \frac{1}{2} \sum_{x \in \mathcal{X}} \Big| \sum_{y \in \mathcal{X}} P(y,x) \left[\mu(y) - \nu(y) \right] \Big| \\ &\leq \frac{1}{2} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{X}} P(y,x)\, | \mu(y) - \nu(y) | \\ &\leq \frac{1}{2} \sum_{y \in \mathcal{X}} | \mu(y) - \nu(y) | \sum_{x \in \mathcal{X}} P(y,x) \\ &\leq \frac{1}{2} \sum_{y \in \mathcal{X}} | \mu(y) - \nu(y) | \\ &\leq \| \mu - \nu \|_{TV} \end{aligned}$$

So, it follows that $\| \mu P^t - \nu P^t \|_{TV} \leq \| \mu P - \nu P \|_{TV} \leq \| \mu - \nu \|_{TV}$

If $\nu = \pi$ is a stationary distribution, then.

$$\| \mu P^{t+1} - \pi \|_{TV} = \| \mu P^{t+1} - \pi P^{t+1} \|_{TV} \le \| \mu P^t - \pi P^t \|_{TV} = \| \mu P^t - \pi \|_{TV}$$

Apply the above result for $\mu = \mu P^t$ and $\nu = \pi P^t$.

i.e, advancing the chain can only move closer to the stationary distribution

Proof of vi) : From v) we have for any $\mu \in \mathcal{P}$ $\| \mu P^{t+1} - \pi \| \le \| \mu P^t - \pi \|_{TV}$

Taking the supremum on the right hand side wrt $\mu \in \mathcal{P}$.

$$\Rightarrow \| \mu P^{t+1} - \pi \|_{TV} \le \sup_{\mu \in \mathcal{P}} \| \mu P^t - \pi \|_{TV} = d(t)$$

Taking the supremum on the left hand side wrt $\mu \in \mathcal{P}$.

$$\Rightarrow \quad d(t+1) = \sup_{\mu \in \mathcal{P}} \| \mu P^{t+1} - \pi \|_{TV} \le d(t).$$

$\Rightarrow$ $d(t)$ is non-increasing in t

From v) we also have $\| \mu P^{t+1} - \nu P^{t+1} \|_{TV} \le \| \mu P^t - \nu P^t \|_{TV}$ for any $\mu, \nu \in \mathcal{P}$.

Taking the supremum over $\mu, \nu \in \mathcal{P}$ on the right hand side,

$$\| \mu P^{t+1} - \nu P^{t+1} \|_{TV} \le \sup_{\mu, \nu \in \mathcal{P}} \| \mu P^t - \nu P^t \| = \bar{d}(t)$$

Taking the supremum over $\mu, \nu \in \mathcal{P}$ on the left hand side,

$$\Rightarrow \bar{d}(t+1) = \sup_{\mu, \nu \in \mathcal{P}} \| \mu P^{t+1} - \nu P^{t+1} \| \le \bar{d}(t)$$

$\Rightarrow$ $\bar{d}(t)$ is non-increasing in t.

Proof of vii) c and t are positive integers.

From i) it follows that $d(ct) \le \bar{d}(ct)$.

See that $\bar{d}(ct) = \bar{d}(t + (c-1)t) \le \bar{d}(t)\, \bar{d}((c-1)t)$ [From iv (submultiplicative property)]

Continuing in the same way we get $\bar{d}(ct) \le (\bar{d}(t))^c$

Proof of viii) $t_{mix}(\varepsilon) = \min\{t : d(t) \le \varepsilon\}$.

Then $t_{mix} = t_{mix}(1/4)$ and $d(t_{mix}) \le \frac{1}{4}$

Let ℓ be a positive integer, then from vii),

$$d(\ell t_{mix}(\varepsilon)) \leq \bar{d}(\ell t_{mix}(\varepsilon)) \leq \left(\bar{d}(t_{mix}(\varepsilon))\right)^{\ell} \leq \left(2\, d(t_{mix}(\varepsilon))\right)^{\ell}$$
$$\leq (2\varepsilon)^{\ell}$$

Take $\varepsilon = 1/4$, then $d(\ell t_{mix}) \leq 2^{-\ell}$

For any $\varepsilon > 0$, choose ℓ such that $\ell \geq \lceil \log_2 \varepsilon^{-1} \rceil$

then $\ell \geq \log_2 \varepsilon^{-1}$

$$\Rightarrow 2^{\ell} \geq \frac{1}{\varepsilon}$$

$$\Rightarrow 2^{-\ell} \leq \varepsilon$$

So, for such a ℓ, $d(\ell t_{mix}) \leq 2^{-\ell} \leq \varepsilon$

The $\ell t_{mix} \in \{t : d(t) \leq \varepsilon\}$

This implies that $t_{mix}(\varepsilon) \leq \ell t_{mix}$, for $\ell \geq \lceil \log_2 \varepsilon^{-1} \rceil$

$$\therefore \; t_{mix}(\varepsilon) \leq \lceil \log_2 \varepsilon^{-1} \rceil \, t_{mix}$$

See that $d(\ell t_{mix}(\varepsilon)) \leq (2\varepsilon)^{\ell}$

We need an $\varepsilon < \frac{1}{2}$ to make this inequality meaningful. [For $\varepsilon < \frac{1}{2}$, $(2\varepsilon)^{\ell} < 1$]

$\varepsilon = \frac{1}{4}$ is just a well accepted standard.

Proof of Remark (*): Suppose $\{(X_t, Y_t)\}_{t \geq 0}$ is a Markovian coupling of a Markov chain on a state space $\mathcal{X}$ and transition probability matrix P.

Then from definition of Markovian coupling, we have for all x, y, x', y' in $\mathcal{X}$,

$$\mathbb{P}(X_{t+1} = x' \mid X_t = x, Y_t = y) = P(x, x') \quad \text{----- (i)}$$

and $$\mathbb{P}(Y_{t+1} = y' \mid X_t = x, Y_t = y) = P(y, y') \quad \text{----- (ii)}$$

Now we will modify the transition probabilities of (X_t, Y_t) so that the modified chain is a Markovian coupling and it satisfies (*).

If $x \neq y$, we keep $\mathbb{P}(X_{t+1} = x', Y_{t+1} = y' \mid X_t = x, Y_t = y)$ as it is.

If $x=y$, we define $\mathbb{P}(X_{t+1}=x', Y_{t+1}=y' \mid X_t=x, Y_t=y) = \begin{cases} P(x,x') & \text{if } x'=y' \\ 0 & \text{otherwise} \end{cases}$

From this definition, we see that if $X_s=Y_s$ then $X_t=Y_t$ $\forall\, t \geq s$.

i.e, it satisfies $(*)$

Now we need to check that the new transition probability is still a Markovian coupling of P.

$$\mathbb{P}(X_{t+1}=x' \mid X_t=x, Y_t=y) = \sum_{y \in \mathcal{X}} \mathbb{P}(X_{t+1}=x', Y_{t+1}=y' \mid X_t=x, Y_t=y)$$

$$= \begin{cases} \sum_{y \in \mathcal{X}} \mathbb{P}(X_{t+1}=x', Y_{t+1}=y' \mid X_t=x, Y_t=y) & \text{if } x \neq y \\ \sum_{y \in \mathcal{X}} \mathbb{P}(X_{t+1}=x', Y_{t+1}=y' \mid X_t=x, Y_t=y) & \text{if } x=y \end{cases}$$

$$= \begin{cases} \mathbb{P}(X_{t+1}=x' \mid X_t=x, Y_t=y) & \text{if } x \neq y \\ \mathbb{P}(X_{t+1}=x', Y_{t+1}=y' \mid X_t=x, Y_t=y) & \text{if } x=y \end{cases}$$

$$= \begin{cases} P(x,x') & \text{if } x \neq y & [\text{from equ (i)}] \\ P(x,x') & \text{if } x=y & [\text{Defined as so}] \end{cases}$$

$$= P(x,x')$$

Using the same argument, we show that $\mathbb{P}(Y_{t+1}=y' \mid X_t=x, Y_t=y) = P(y,y')$.

With these modified transition probabilities, $\{(X_t,Y_t)\}_{t\geq 0}$ is a Markovian coupling of P and it satisfies $(*)$.

Section 5 : Application and References.

The coupling technique used in obtaining the bounds for mixing time, can be applied to other Markov chains as —

i) Random Walk on the Hypercube

ii) Random Walk on the torus

iii) Random Walk on a finite binary tree

The above examples are discussed in chapter 5 of Markov chains and Mixing times by David A. Levin and Yuval Peres

References :

Markov chains and Mixing Times
- David A. Levin and Yuval Peres.