Markov Chains Mixing Times **October 19, 2022**

Lecture 3: Total variation distance and distance from stationary

Lecturer: Yuval Peres Week: 2

In this lecture we will present some notions and tools that will be useful when studying the concept of mixing time. The first one is the general concept of time reversal.

Definition 3.32. Consider an irreducible transition matrix P with stationary distribution π. The ***time reversal matrix*** of P is denoted by $\widehat{P}$ and it is given by the formula

$$\pi(x)P(x,y) = \pi(y)\widehat{P}(y,x) \quad \forall\, x, y \in V.$$

Notice in the above definition that $\pi(y) > 0$ as P is irreducible. It is easy to check that $\widehat{P}$ is a transition matrix. Indeed,

$$\sum_{x\in V} \widehat{P}(y,x) = \frac{1}{\pi(y)} \sum_{x\in V} \pi(x)P(x,y) = \frac{1}{\pi(y)}\pi(y) = 1.$$

The transition matrix $\widehat{P}$ corresponds to the Markov chain that we obtain when we reverse time in P. Therefore, the stationary distribution π of P is also stationary for $\widehat{P}$:

$$\pi\widehat{P}(z) = \sum_{x\in V} \pi(x)\widehat{P}(x,z) = \sum_{x\in V} \pi(z)P(z,x) = \pi(z).$$

The following simple Markov chain will be used as a tool to study more complex ones in the future. A coupon collector desires to complete set of n different types of coupons. We suppose each coupon he acquires is equally likely to be each of the n types. In other words, if U_i is a random variable taking values in $\{1,\ldots,n\}$ that represents the i-th coupon, then (U_i) is a sequence of independent uniformly distributed random variables $\{1,\ldots,n\}$.

Let N_t denote the number of different types the collector obtained after the first t coupons. Let τ_k be the total number of coupons accumulated when the collection first contains k distinct coupons, that is, $\tau_k = \min\{t \colon N_t = k\}$. We are specially interested in the case $k = n$, that corresponds to the case when the collector has completed the collection. We might be interested in the expected number of coupons the collector needs to buy for this to happen, or in the variance of this random variable.

Clearly $N_0 = 0$. When the collector has coupons of k different types, there are $n-k$ types missing. Of the n possibilities for his next coupon, only $n-k$ will expand his collection. Hence

$$\mathbb{P}\{N_{t+1} = k+1 \mid N_t = k\} = \frac{n-k}{n}$$

and

$$\mathbb{P}\{N_{t+1} = k \mid N_t = k\} = \frac{k}{n}.$$

Every trajectory of this chain is non-decreasing. Once the chain arrives at state n (corresponding to a complete collection), it is absorbed there. We are interested in the number of steps required to reach the absorbing state.

Proposition 3.33. *Consider a collector attempting to collect a complete set of coupons. Assume that each new coupon is chosen uniformly and independently from the set of n possible types, and let τ be the (random) number of coupons collected when the set first contains every type. Then*

$$\mathbb{E}(\tau) = n \sum_{k=1}^{n} \frac{1}{k}.$$

Proof. The expectation $\mathbb{E}(\tau)$ can be computed by writing τ as a sum of geometric random variables. First, $\tau_1 = 1$. Next, observe that $\tau_{k+1} - \tau_k$ is a geometric random variable with parameter $p = \frac{n-k}{n}$ for $k = 1, \ldots, n-1$: after collecting τ_k coupons, there are $n-k$ types missing from the collection. Each subsequent coupon drawn has the same probability $(n-k)/n$ of being a type not already collected, until a new type is finally drawn. Then

$$\tau = \tau_n = \tau_1 + (\tau_2 - \tau_1) + \cdots + (\tau_n - \tau_{n-1}).$$

The expectation of these geometric variables is $\mathbb{E}(\tau_{k+1} - \tau_k) = n/(n-k)$ and

$$\mathbb{E}(\tau) = 1 + \sum_{k=1}^{n-1} \mathbb{E}(\tau_{k+1} - \tau_k) = 1 + n \sum_{k=1}^{n-1} \frac{1}{n-k} = n \sum_{k=1}^{n} \frac{1}{k}. \qquad \square$$

One can check that $|\sum_{k=1}^{n} 1/k - \log n| \leq 1$, whence $|\mathbb{E}(\tau) - n \log n| \leq n$. We will need to know more about the distribution of τ in future applications. Next, we are interested in estimating its variance.

Proposition 3.34. *Let τ be a coupon collector random variable, as in Proposition 3.33. Then*

$$Var(\tau) \leq 2n^2.$$

The above estimation tells us that τ is concentrated around its mean $\mathbb{E}(\tau)$. Indeed, one can apply Chebyshev's inequality to show that $\frac{\tau}{\mathbb{E}(\tau)}$ converges in probability to 1.

Proof. Notice that $(\tau_{k+1} - \tau_k)_{k=1}^{n-1}$ is a sequence of independent variables. Moreover, $\mathrm{Var}(\tau_1) = 0$. Therefore,

$$\mathrm{Var}(\tau) = \sum_{k=1}^{n-1} \mathrm{Var}(\tau_{k+1} - \tau_k).$$

Finally, the variance of a geometric variable can be bounded above by the square of its mean. Thus,

$$\mathrm{Var}(\tau) \leq n^2 \sum_{k=1}^{n-1} \frac{1}{(n-k)^2} \leq 2n^2. \qquad \square$$

The next result says that τ is unlikely to be much larger than its expected value.

Proposition 3.35. *Let τ be a coupon collector random variable, as in Proposition 3.33. For any $c > 0$,*

$$\mathbb{P}\{\tau > \lceil n \log n + cn \rceil\} \leq e^{-c}.$$

Proof. Let A_i be the event that the i-th type does not appear among the first $\lceil n \log n + cn \rceil$ coupons drawn. Observe first that

$$\mathbb{P}\{\tau > \lceil n \log n + cn \rceil\} = \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i).$$

Since each trial has probability $1 - n^{-1}$ of ***not*** drawing coupon i and the trials are independent, the right-hand side above is equal to

$$\sum_{i=1}^{n} \left(1 - \frac{1}{n}\right)^{\lceil n \log n + cn \rceil} \leq n \exp\left(-\frac{n \log n + cn}{n}\right) = e^{-c}. \qquad \square$$

As we defined in Lecture 1, the ***total variation distance*** between two probability distributions μ and ν on a finite state space V is defined as

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in V} |\mu(x) - \nu(x)|.$$

As it often happens with important concepts in mathematics, there are different equivalent definitions for the total variation distance, each of them useful depending on the situation. For instance, one can easily check that

$$\|\mu - \nu\|_{TV} = \sum_{\mu(x) > \nu(x)} \mu(x) - \nu(x) = \sum_{\nu(x) > \mu(x)} \nu(x) - \mu(x).$$

Similarly, we have the useful formula

$$\|\mu - \nu\|_{TV} = \max\{\mu(A) - \nu(A) \colon A \subseteq V\}.$$

This maximum is clearly attained for $A = \{x \in V \colon \mu(x) > \nu(x)\}$. A different formula appears when we consider functions from V to $[-1, 1]$.

Proposition 3.36. *Let μ and ν be two probability distributions on a finite state space V. Then*

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \max\left\{\int f \, d\mu - \int f \, dv \colon f \colon V \longrightarrow [-1, 1]\right\}.$$

Since V is a finite state space we have $\int f \, d\mu = \sum_{x \in V} f(x)\mu(x)$. We also use the notation $\mu f \colon = \int f \, d\mu$.

Proof. First, for any $f\colon V \longrightarrow [-1,1]$ notice that

$$|\mu f - \nu f| = \left|\sum_{x\in V} f(x)(\mu(x)-\nu(x))\right| \leq \sum_{x\in V}|\mu(x)-\nu(x)| = 2\|\mu-\nu\|_{TV}.$$

Therefore, $\max\{\mu f - \nu f\colon f\colon V \longrightarrow [-1,1]\} \leq \|\mu-\nu\|_{TV}$. For the other direction, consider the function $g\colon V \longrightarrow [-1,1]$ given by

$$g(x) = \begin{cases} 1 & \text{if } \mu(x) > \nu(x); \\ 0 & \text{if } \mu(x) = \nu(x); \\ -1 & \text{if } \mu(x) < \nu(x). \end{cases}$$

Then, a simple computation shows that $\mu g - \nu g = 2\|\mu-\nu\|_{TV}$. □

Another characterization of the total variation distance uses the notion of coupling. Coupling is one of the most useful tools when studying Markov chains. Although it might provides visual and deep arguments, it often requires a careful treatment. Given two distribution probabilities μ and ν on a state space V, a ***coupling*** is a distribution γ on $V\times V$ that projects to μ and ν, that is, $\mu(A) = \gamma(A\times V)$ and $\nu(A) = \gamma(V\times A)$ for every $A\subseteq V$.

Example 3.37. Let μ be the uniform distribution on $\{1,2,3\}$ and let ν be the uniform distribution on $\{2,3,4\}$. One can see $V=\{1,2,3,4\}$ as their common state space. We want to find the coupling γ that maximizes the measure of the diagonal $\Delta = \{(x,x)\colon x\in V\}$. Let X be a random variable with distribution μ. If $X=2$ or X_3, let $Y=X$. If $X=1$, let $Y=4$. Let γ be the joint distribution of (X,Y). Then, γ is a coupling of μ and ν and $\gamma(\Delta)=\frac{2}{3}$. One can use the fact that $X=1$ with probability $\frac{1}{3}$ to show that this is the best we can do.

Proposition 3.38. *Let μ and ν be two probability distributions on a finite state space V. For $\Delta = \{(x,x)\colon x\in V\}$ we have*

$$\|\mu-\nu\|_{TV} = \min\{1-\gamma(\Delta)\colon \gamma \textit{ couples } \mu \textit{ and } \nu\}.$$

Proof. Given $A\subseteq V$ and a coupling γ of μ and ν,

$$\begin{aligned}\mu(A)-\nu(A) &= \gamma(A\times V) - \gamma(V\times A) \leq (A\times V) - \gamma(A\times A) = \gamma(A\times(V\setminus A)) \\ &\leq \gamma(\{(x,y)\colon x\neq y\}) = 1-\gamma(\Delta).\end{aligned}$$

Hence, taking maximum over $A\subset V$ gives $\|\mu-\nu\|_{TV} \leq \min\{1-\gamma(\Delta)\colon \gamma \text{ couples } \mu \text{ and } \nu\}$. For the other inequality, define

$$\gamma(x,y) = \begin{cases} \min\{\mu(x),\nu(x)\} & \text{if } \mu(x)=\nu(x); \\ \frac{(\mu(x)-\nu(x))_+\cdot(\nu(y)-\mu(y))_+}{\|\mu-\nu\|_{TV}} & \text{if } \mu(x)\neq\nu(x), \end{cases}$$

where $r_+ = \max\{0,r\}$ for any $r\in\mathbb{R}$. First, we need to check that γ is a coupling of μ and ν. Indeed, observe that

$$\sum_{y\in V}\gamma(x,y) = \min\{\mu(x),\nu(x)\} + \frac{(\mu(x)-\nu(x))_+}{\|\mu-\nu\|_{TV}}\|\mu-\nu\|_{TV} = \mu(x).$$

Similarly, $\sum_{x\in V}\gamma(x,y)=\nu(y)$. Finally,

$$\gamma(\{(x,y)\colon x\neq y\})=\sum_{x\in V}(\mu(x)-\nu(x))_+=\|\mu-\nu\|_{TV}.$$ □

Let P be the transition matrix of a Markov chain defined on a finite state space V. Given two probability measures μ and ν on V, we claim that $\|P\mu-P\nu\|_{TV}\leq\|\mu-\nu\|_{TV}$. In other words, the distance between μ and ν decreases after one step of the chain. In this case it is convenient to use the characterization of the total variation distance given in Proposition 3.36. Given $f\colon V\longrightarrow[-1,1]$, we can consider the function $Pf\colon V\longrightarrow[-1,1]$ given by $Pf(y)=\sum_{x\in V}f(x)P(x,y)$. Then,

$$2\|P\mu-P\nu\|_{TV}=\max\{\mu Pf-\nu Pf\colon f\colon V\longrightarrow[-1,1]\}\leq\max\{\mu f-\nu f\colon f\colon V\longrightarrow[-1,1]\}=2\|\mu-\nu\|_{TV}.$$

We have now everything we need to measure the distance from stationary. For this purpose, define

$$d(t)=\max_{x\in V}\|P^t(x,\cdot)-\pi\|_{TV}. \tag{3.11}$$

Notice that $P^t(x,\cdot)=\delta_xP^t$. Since every distribution μ can be written as an average of distributions of the form δ_x, we also have

$$d(t)=\max\{\|\mu P^t-\pi\|_{TV}\colon \mu\text{ is a distribution on }V\}.$$

Instead of measuring the distance from stationary, we are also interested in measuring the distance between probability distributions of the form δ_x and δ_y as we run the chain. Define

$$\bar{d}(f)=\max\{x,y\in V\}\|P^t(x,\cdot)-P^t(y,\cdot)\|_{TV}. \tag{3.12}$$

These two distances are closely related.

Lemma 3.39. *If $d(t)$ and $\bar{d}(t)$ are defined as in (3.11) and (3.12), then*

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

Proof. From the triangle inequality for the total variation distance we obtain $\bar{d}(t)\leq 2d(t)$. To show that $d(t)\leq\bar(d)(t)$, notice that

$$\|P^t(x,\cdot)-\pi\|_{TV}=\left\|P^t(x,\cdot)-\sum_{y\in V}\pi(y)P^t(y,\cdot)\right\|_{TV}=\left\|\sum_{y\in V}\pi(y)(P^t(x,\cdot)-P^t(y,\cdot))\right\|_{TV}\leq\bar{d}(t).$$

□

The advantage of working with $\bar{d}$ is showed by the following result.

Lemma 3.40. *The function $\bar{d}$ is submultiplicative: $\bar{d}(s+t)\leq\bar{d}(s)\bar{d}(t)$.*

Proof. Fix x, $y \in V$ and let (X_t) and (Y_t) be Markov chains with transition matrix P that start at states x and y, respectively. We must bound $\|P^{t+s}(x,\cdot) - P^{t-s}(y,\cdot)\|_{TV}$. Using Proposition 3.38, we can consider a coupling γ_t of X_t and Y_t such that $\gamma_t(\{X_t \neq Y_t\}) = \|\delta_x P^t - \delta_y P^t\|_{TV} \leq \bar{d}(t)$. Moreover, notice that

$$P^{t+s}(x,w) = \sum_{z \in V} P^t(x,z) P^s(z,w) = \mathbb{E}(P^s(X_t, w)).$$

Therefore, for any $A \subseteq V$ we have

$$P^{t+s}(x,A) - P^{t+s}(y,A) = \mathbb{E}(P^s(X_t,A) - P^s(Y_t,A)) \leq \mathbb{E}(\bar{d}(s) 1_{X_t \neq X_y}) = \bar{d}(s)\mathbb{P}(X_t \neq Y_t) \leq \bar{d}(s)\bar{d}(t)$$

where the expectation is with respect to γ. □

As we will discuss in future lectures, the ***mixing time*** is defined as $t_{\text{mix}} = \min\{t \colon d(t) \leq \frac{1}{4}\}$. An important consequence of Lemma 3.40 is that $d(\ell t_{\text{mix}}) \leq 2^{-\ell}$ for any $\ell \in \mathbb{N}$.

Markov Chains Mixing Times October 21, 2022

Lecture 4: Coupling

Lecturer: Yuval Peres Week: 2

Definition 4.41. A ***Coupling*** of two probability distributions μ and ν is a pair of random variables (X, Y), defined on the same probability space, such that the marginal distribution of X is μ and the marginal distribution of Y is ν.

In this lecture, the couplings that we mainly discuss are ***Markovian Couplings***.

Definition 4.42. We define a coupling of Markov chains with transition matrix P to be a process $(X_t, Y_t)_{t=0}^{\infty}$ with the property that both (X_t) and (Y_t) are Markov chains with transition matrix P, although the two chains may possibly have different starting distributions.

Given a Markov chain on 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 $\chi \times \chi$ which satisfies, for all $x, y, x^{'}, y^{'}$,

$$\mathbf{P}\{X_{t+1} = x^{'} | X_t = x, Y_t = y\} = \mathbf{P}(x, x^{'}) \tag{4.13}$$

$$\mathbf{P}\{Y_{t+1} = y^{'} | X_t = x, Y_t = y\} = \mathbf{P}(y, y^{'}). \tag{4.14}$$

Remark 4.43. 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 more precisely, so that if $X_s = Y_s$, then $X_t = Y_t$ for $t \geq s$.

Proposition 4.44. *Let μ and ν be two probability distributions on X. Then*

$$\|\mu - \nu\|_{TV} = inf\{\boldsymbol{P}\{X \neq Y\} : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\}. \tag{4.15}$$

Proof. First, we note that for any coupling (X, Y) of μ and ν and any event $A \subseteq X$,

$$\begin{aligned}\mu(A) - \nu(A) &= \mathbf{P}\{X \subseteq A\} - \mathbf{P}\{Y \subseteq A\} \\ &\leq \mathbf{P}\{X \subseteq A, Y \nsubseteq A\} \\ &\leq \mathbf{P}\{X \neq Y\}.\end{aligned} \tag{4.16}$$

It immediately follows that

$$\|\mu - \nu\|_{TV} = inf\{\mathbf{P}\{X \neq Y\} : (X, Y) \text{ is a coupling of } \mu \text{ and } \nu\}. \tag{4.17}$$

□

Theorem 4.45. *Let $\{(X_t, Y_t)\}$ be coupling satisfying 4.41 for which $X_0 = x$ and $Y_0 = y$. Let τ_{coal} be the **coalescence time** of the chains:*

$$\tau_{coal} := min\{t : X_s = Y_s \text{ for all } s \geq t\}. \tag{4.18}$$

Then

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq \boldsymbol{P}_{x,y}\{\tau_{coal} > t\}. \tag{4.19}$$

Proof. Notice that $P^t(x,z) = \mathbf{P}_{x,y}\{X_t = z\}$ and $P^t(y,z) = \mathbf{P}_{x,y}\{Y_t = z\}$. Consequently, (X_t, Y_t) is a coupling of $P^t(x,\cdot)$ with $P^t(y,\cdot)$, whence Proposition 4.44 implies that

$$\|P^t(x,\cdot) - P^t(y,\cdot)\|_{TV} \leq \mathbf{P}_{x,y}\{X_t \neq Y_t\}. \tag{4.20}$$

□

Fact 4.46. *Applying Markov's inequality, we arrive at the inequality:*

$$\|P^t(x,\cdot) - P^t(y,\cdot)\|_{TV} \leq \boldsymbol{P}_{x,y}\{X_t \neq Y_t\} \leq \frac{\mathbb{E}(\tau_{coal})}{t}. \tag{4.21}$$

Now, we discuss some *Examples* which illustrates the importance of couplings.

Example 4.47. Random Walk on Hypercube

Choose $U_t \sim Unif\{1,\ldots,n\}$. Given initial states $x, y \in \{0,1\}^n = V$, we define X_t, Y_t via same U_t, b_t. Replace coordinate U_t by

$$b_t = \begin{cases} 1, & \text{with probability } \frac{1}{2} \\ 0, & \text{otherwise} \end{cases}$$

That is we first, pick among the n coordinates uniformly at random; suppose that coordinate i is selected. In both walks, replace the bit at coordinate i with the same random fair bit. From this time onwards, both walks will agree in the i-th coordinate.
If τ is the first time when all of the coordinates have been selected at least once, then the two walkers agree with each other from time τ onwards and $\tau_{coal} = \tau_n$ of the coupon collector problem. Therefore,

$$d(n\log n + cn) \leq \mathbf{P}\{\tau > n\log n + cn\} \leq e^{-c}. \tag{4.22}$$

It is immediate from the above that

$$t_{mix}(\epsilon) \leq n\log n + \log(1/\epsilon)n.$$

Example 4.48. Random walk on the cycle.

We defined random walk on the n-cycle. The underlying graph of this walk, $\mathbb{Z}_n$, has vertex set $\{1,2,\ldots,n\}$. and edges between j and k whenever $j \equiv k \pm 1 \mod n$.
Again, we have two markov chains X_t, Y_t. And, the coupling rules are as follows: If X chain jumps, Y doesn't jump, and vice-versa. And at the time of coalescence, X_t and Y_t becomes equal and continue together.
Let us denote $D_t =$ clockwise distance between X_t and Y_t.
If $X_t \neq Y_t$, then $0 < D_t < n$.

$$d\bar{(}t) \leq \max_{x,y} \mathbb{P}(\tau_{coal} > t) = \max_{k=D_0} \mathbb{P}(\tau^D_{\{0,1\}} > t) \leq \max_k \mathbb{E}_k(\tau^{SRW}_{\{0,n\}}) = \max_k \frac{k(n-k)}{t} \leq \frac{n^2}{4t}. \tag{4.23}$$

Therefore, $t_{mix} \leq n^2$. Now, $X_t = S_t \mod n$ where S_t is lazy walk on $\mathbb{Z}$.
$\mathbb{E}(S_t) = 0$ $Var(S_t) = \frac{t}{2}$. By Chebyshebs' inequality, we get that $\mathbb{P}(|S_t| \geq \sqrt{2}\sqrt{t}) < \frac{1}{4}$.

$$\tilde{A}_t = \{|S_t| \geq \sqrt{2t}\} \subseteq \{dist_{\mathbb{Z}_n}(X_t, 0) \geq \sqrt{2t}\} = A_t \tag{4.24}$$

Therefore, $\mathbb{P}_0(A_t) \leq \mathbb{P}_0(\tilde{A}_t) < \frac{1}{4}$. And, when we take $\sqrt{2t} = n/4$, $A_t = \{dist_{\mathbb{Z}_n}(X_t, 0) \geq n/4\}$, $\pi(A_t) \geq 1/2 \Rightarrow \pi(A_t) - \mathbb{P}_0(A_t) > 1/4 \Rightarrow t_{mix} > t = \frac{n^2}{32}$. Hence, it is proved that $\frac{n^2}{32} < t_{mix} \leq n^2$.

Example 4.49. Random walk on the torus.

The d-dimensional torus is the graph whose vertex set is the Cartesian product:

$$\mathbb{Z}_n^d = \mathbb{Z}_n \times \ldots \times \mathbb{Z}_n. \text{ (d times)}.$$

Vertices $x = (x_1, \ldots, x_d)$ and $y = (y_1, y_2, \ldots, y_d)$ are neighbours in $\mathbb{Z}_n^d$ if for some $j \in \{1, 2, \ldots, d\}$, we have $x^i = y^i$ for all $i \neq j$ and $x^j \equiv y^j \pm 1 \mod n$.
To connect a random walk (X_t) that begins at x with a random walk (Y_t) that begins at y, first choose one of the d locations at random. If the locations of the two walks coincide in the specified coordinate, we move both walks by $+1$, -1, or 0, with probability of 1/4, 1/4, and 1/2, respectively. If the locations of the two walks differ in the specified coordinate, we choose one of the chains at random to move while keeping the other unchanged. The selected walk is then moved by $+1$ or -1 in the specified coordinate, with the sign decided by a fair coin flip. Let $Xt = (X_t^1, \ldots, X_t^d)$ and $Y_t = (Y_t^1, \ldots, Y_t^d)$, and let

$$\tau_i := min\{t \geq 0 : X_t^i = Y_t^i\}$$

be the time required for the chains to agree in coordinate i. The clockwise difference between X_t^i and Y_t^i, viewed at the times when coordinate i is selected, behaves just as the coupling of the lazy walk on the cycle $\mathbb{Z}_n$ discussed above. Thus, the expected number of moves in coordinate i needed to make the two chains agree on that coordinate is not more than $n^2/4$. Since coordinate i is selected with probability $1/d$ at each move, there is a geometric waiting time between moves with expectation d,

$$\mathbb{E}x, y(\tau_i) \leq \frac{dn^2}{4}.$$

The coupling time we are interested in is $\tau_{couple} = \max_{1 \leq i \leq d} \tau_i$, and we can bound the maximum by a sum to get

$$\mathbb{E}_{x,y}(\tau_{couple}) \leq \frac{d^2 n^2}{4}.$$

Therefore,

$$\mathbb{P}_{x,y}\{\tau_{couple} > t\} \leq \frac{\mathbb{E}_{x,y}(\tau_{couple})}{t} \leq \frac{1}{t}\frac{d^2 n^2}{4}.$$

Hence, we can say that $t_{mix} \leq d^2 n^2$.

Definition 4.50. A family of σ-algebras $\{\mathcal{F}_t\}_{t \in \mathbb{N}}$ is said to be a ***filtration*** if $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$

A filtration is Markovian if the σ-algebra generated by $X_0, X_1, \cdots, X_t$ is contained in $\mathcal{F}_t$ i.e., $\sigma(X_0, X_1, \cdots, X_t) \subset \mathcal{F}_t$ for any t it satisfies the following equation,

$$\mathbf{P}(X_{t+1} = y | \mathcal{F}_t) = P(X_t, y). \tag{4.25}$$

Consider, the lazy random walk on hypercube. Let $U_t \sim \{1, 2, \ldots, n\}$ be i.i.d. and $b_t \sim \text{Bernoulli}(\frac{1}{2})$ be i.i.d. Whence, the lazy random walk can be modelled as,

$$X_{t+1}(j) = \begin{cases} X_t(j) & \text{if } j \neq U_t \\ b_t & \text{if } j = U_t \end{cases} \tag{4.26}$$

Fix $X_0 = x$, and define $\mathcal{F}_t = \sigma(U_0, U_1, \dots, U_t, b_0, b_1, \dots, b_t)$. Then it can be observed that $\sigma(X_0, X_1, \cdots, X_t) \subset \mathcal{F}_t$. It can be easier to think of filtrations as refining sets of partitions of the sample space.

Definition 4.51. A non-negative integer valued random variable τ is called a ***stopping time*** if for any t $\{\tau \leq t\} \in \mathcal{F}_t$.

Hitting time of a set $A \subseteq \mathcal{X}$ is defined as

$$\tau_A = \min\{t : X_t \in A\}$$

Definition 4.52. ***Stationary time*** is a stopping time with respect to a Markovian filtration $\{\mathcal{F}_t\}$ such that,

$$\mathbf{P}_x\{X_\tau = y\} = \pi(y)$$

In other words, stationary time is the random time after which the distribution of the Markov chain achieves stationarity.

Definition 4.53. A stationary time is said to be ***strong stationary*** if X_τ is independent of τ, i.e.

$$\mathbf{P}_x\{X_\tau = y | \tau = t\} = \pi(y)\ \forall t \geq 0$$

Example 4.54. Let (X_t) be an irreducible Markov chain with state space $\mathcal{X}$ and stationary distribution π. Let U be a random object distributed as π and assume $\tau = \min\{t \geq 0 : X_t = U\}$. Let $\mathcal{F}_t = \sigma(U, X_0, X_1, \dots, X_t)$. The time τ is an $\mathcal{F}_t$-stopping time, and because $X_t = U$, it follows that τ is a stationary time. If the Markov chain starts at x then $\mathbf{P}_x\{X_\tau = y | \tau = 0\} = \delta_x(y)$. This shows that the time is not strong stationary.

Consider lazy random walk on hypercube, with $U_t \sim \text{Uniform}\{1, 2, \dots, n\}$ where U_t denotes the co-ordinate chosen at time t. The time when all the co-ordinates are chosen, $\tau = \min\{t : \{U_1, \dots, U_t\} = [n]\}$, is a strong stationary time.