Random Walks: WEEK 7

1 The cut-off phenomenon

1.1 Summary of the two previous lectures

Recall that we are considering a Markov chain $(X_n,\ n \geq 0)$ with transition matrix P and a *finite* state space S, with $|S| = N$. We assume that the chain is *ergodic* (irreducible, aperiodic and positive-recurrent), thus there is a unique stationary and limiting distribution π, with $\pi = \pi P$, and $p_{ij}(n) \xrightarrow[n\to\infty]{} \pi_j,\ \forall i, j \in S$. Finally, we assume that the detailed balance equation is satisfied: $\pi_i\, p_{ij} = \pi_j\, p_{ji},\ \forall i, j \in S$.

Under these assumptions, the transition matrix P has N eigenvectors $\phi^{(0)}, \cdots, \phi^{(N-1)} \in \mathbb{R}^N$, and N corresponding eigenvalues $1 = \lambda_0 > \lambda_1 \geq \cdots \geq \lambda_{N-1} > -1$ such that $P\phi^{(k)} = \lambda_k \phi^{(k)}\ \forall 0 \leq k \leq N-1$. We define also $\lambda_* := \max_{1\leq k\leq N-1} |\lambda_k| = \max\{\lambda_1, -\lambda_{N-1}\}$.

In the previous lecture, we studied the rate of convergence of the distribution of this Markov chain towards π. If the initial state is $X_0 = i$, at time-step n the probability distribution is given by P_i^n. We measure the distance between this distribution and π in terms of the total variation (TV) distance: $\|P_i^n - \pi\|_{\text{TV}} := \frac{1}{2}\sum_{j\in S} |p_{ij}(n) - \pi_j|$. We proved the following bound:

Theorem 1.1 (Rate of Convergence)**.** Under the above assumptions,

$$\|P_i^n - \pi\|_{\text{TV}} \leq \frac{\lambda_*^n}{2\sqrt{\pi_i}},\ \forall i \in S, n \geq 1$$

Notice that this upper bound decays exponentially in terms of n. Today, we study a *reciprocal statement*, which gives a corresponding lower bound (under additional assumptions), that also decays exponentially in terms of n. We study examples where these two bounds are tight or loose. We also analyze an example of the *cut-off phenomenon*, where the actual TV distance does not have a smooth exponential decay in terms of n, but rather a sudden drop from 1 to 0 in a short interval.

1.2 Reciprocal statement

Before we present the statement, we need to better understand the concept of total variation distance.

Proposition 1.2. Let μ, ν be two probability distributions over the state space S. The following are equivalent definitions for the total variation distance $\|\mu - \nu\|_{\text{TV}}$:

$$\|\mu - \nu\|_{\text{TV}} = \frac{1}{2}\sum_{j\in S} |\mu_j - \nu_j| \tag{1}$$

$$= \max_{A\subset S} |\mu(A) - \nu(A)| \qquad \text{where } \mu(A) = \sum_{j\in A} \mu_j \tag{2}$$

$$= \frac{1}{2}\max_{\phi: S\to[-1,1]} |\mu(\phi) - \nu(\phi)| \qquad \text{where } \mu(\phi) = \sum_{j\in S} \mu_j \phi_j \tag{3}$$

Proving these equivalences is left as an exercise (suggestion: show that (1) $\leq$ (2) $\leq$ (3) $\leq$ (1)).

In the theorem below, we will use these facts about the eigenvectors of the transition matrix P, which have already been seen in class:

1. $\lambda_0 = 1$ and $\phi^{(0)} = (1, \cdots, 1)^T$;
2. $\phi_j^{(k)} = \dfrac{u_j^{(k)}}{\sqrt{\pi_j}}$ where $(u^{(k)})^T u^{(l)} = \delta_{kl}$.

Theorem 1.3 (Reciprocal Statement)**.** Under the above assumptions, and with the additional condition $|\phi_0^{(k)}| = 1,\ |\phi_j^{(k)}| \leq 1,\ \forall k \in \{0, \cdots, N-1\}, j \in S$, we have

$$\|P_0^n - \pi\|_{\text{TV}} \geq \frac{\lambda_*^n}{2},\ \forall n \geq 1$$

Proof. Because of the additional assumption that all eigenvectors satisfy $|\phi_j^{(k)}| \leq 1, \forall j, k$, we can use the third definition of the TV distance:

$$\|P_0^n - \pi\|_{\text{TV}} = \frac{1}{2} \max_{\phi: S \to [-1,1]} |P_0^n(\phi) - \pi(\phi)| \geq \frac{1}{2} \max_{1 \leq k \leq N-1} |P_0^n(\phi^{(k)}) - \pi(\phi^{(k)})|$$

We do not include $k = 0$ in the above formula, because for $k = 0$, $|P_0^n(\phi^{(0)}) - \pi(\phi^{(0)})| = 1 - 1 = 0$. We compute the two terms on the right-hand side separately:

$$P_0^n(\phi^{(k)}) = \sum_{j \in S} p_{0j}(n) \phi_j^{(k)} = (P^n \phi^{(k)})_0 = (\lambda_k^n \phi^{(k)})_0 = \lambda_k^n \phi_0^{(k)}$$

and

$$\pi(\phi^{(k)}) = \sum_{j \in S} \pi_j \phi_j^{(k)} = \sum_{j \in S} \pi_j \phi_j^{(k)} \underbrace{(\phi_j^{(0)})}_{=1} = \sum_{j \in S} u_j^{(k)} u_j^{(0)} = (u^{(k)})^T u^{(0)} = 0 \text{ (for } k \neq 0)$$

Therefore,

$$\|P_0^n - \pi\|_{\text{TV}} \geq \frac{1}{2} \max_{1 \leq k \leq N-1} |P_0^n(\phi^{(k)}) - \pi(\phi^{(k)})| = \frac{1}{2} \max_{1 \leq k \leq N-1} |\lambda_k^n| \underbrace{|\phi_0^{(k)}|}_{=1} = \frac{1}{2} \max_{1 \leq k \leq N-1} |\lambda_k^n| = \frac{1}{2} \lambda_*^n$$

□

1.3 Examples

The upper and lower bounds for $\|P_0^n - \pi\|_{\text{TV}}$ seem to be very close, as they only differ by a multiplicative factor of $1/\sqrt{\pi_0}$. But if this factor is very large with respect to the inverse of the spectral gap $\gamma^{-1} = (1 - \lambda_*)^{-1}$, the bounds are no longer tight. In the next two examples, it is easy to check that all the assumptions are satisfied, including the condition in the reciprocal statement.

Example 1.4 (Random walk over odd cycle)**.** Consider the random walk over an odd cycle, where $S = \{0, \cdots, N-1\}$, N is odd, and $p_{ij} = \frac{1}{2}$ if $i - j \equiv \pm 1 \mod N$, 0 otherwise. As we proved in an exercise, in this case $\lambda_* = \cos(\pi/N) \approx 1 - \frac{\pi^2}{2N^2}$ (when N is large), so $\lambda_*^n \approx \exp(-\frac{\pi^2 n}{2N^2})$. On the other hand, $\pi_j = \frac{1}{N}\ \forall j \in S$. Therefore:

$$\frac{1}{2} \exp\left(-\frac{\pi^2 n}{2N^2}\right) \leq \|P_0^n - \pi\|_{\text{TV}} \leq \frac{\sqrt{N}}{2} \exp\left(-\frac{\pi^2 n}{2N^2}\right)$$

The upper and lower bounds become arbitrarily close to zero for $n \gg N^2$ and $n \gg N^2 \log N$, respectively. In this case, the bounds are relatively tight and describe well the true behaviour of the TV distance. Two parameters are at play here: the size of the state space (related to $\sqrt{\pi_0}$) and the spectral gap γ.

Example 1.5 (Lazy walk over the binary hypercube)**.** Fix a number d and consider the following variant of the Ehrenfest urns example: there are d balls numbered from 1 to d, partitioned over two urns which are labeled '0' and '1'. At each step, we pick a number t between 0 and d uniformly at random: if we pick $t = 0$, we do nothing; else we move ball t to the opposite urn. Each possible configuration of the balls is described by a vector $x \in \{0,1\}^d$, where x_t indicates which urn contains ball t. Hence the state space is $S = \{0,1\}^d$, with $N = |S| = 2^d$, and the transition probabilities are $p_{xy} = \frac{1}{d+1}$ if $x = y$ or x and y differ in exactly one coordinate, 0 otherwise.

Another way to look at this problem is as a lazy random walk over the d-dimensional binary cube. We make the walk lazy (i.e. with self-loops) because otherwise it would be periodic. We denote by 0 the all-zero vector in $\{0,1\}^d$, and index the 2^d eigenvalues and eigenvectors with elements $z \in \{0,1\}^d$.

Lemma 1.6. The eigenvalues and eigenvectors of the transition probability matrix are

$$\lambda_z = 1 - \frac{2|z|}{d+1}, \text{ where } |z| = \text{number of non-zero components in } z$$

and

$$\phi_x^{(z)} = (-1)^{z \cdot x} \ \forall x \in \{0,1\}^d, \text{ where } z \cdot x = \textstyle\sum_{1 \le t \le d} z_t x_t$$

Proof.

$$(P\phi^{(z)})_x = \sum_{y \in S} p_{xy} \phi_y^{(z)} = \frac{1}{d+1}\,\phi_x^{(z)} + \frac{1}{d+1} \sum_{1 \le t \le d} \phi_{x+e_t}^{(z)}$$

where

$$\phi_{x+e_t}^{(z)} = (-1)^{z \cdot (x+e_t)} = (-1)^{z \cdot x}(-1)^{z \cdot e_t} = \phi_x^{(z)}(-1)^{z_t}$$

Thus

$$(P\phi^{(z)})_x = \frac{1}{d+1}\phi_x^{(z)} \left(1 + \sum_{1 \le t \le d} (-1)^{z_t}\right) = \frac{1}{d+1}\phi_x^{(z)}(1 + d - 2|z|) = \left(1 - \frac{2|z|}{d+1}\right) \phi_x^{(z)}$$

which proves that

$$P\phi^{(z)} = \left(1 - \frac{2|z|}{d+1}\right) \phi^{(z)}$$

□

Notice in particular that $\lambda_0 = 1$, $\phi^{(0)} = (1, \cdots, 1)^T$, and $|\phi_x^{(z)}| = 1$, $\forall x, z \in \{0,1\}^d$. The eigenvalues have high multiplicities: for $1 \le t \le d$, the eigenvalue $\lambda = 1 - \frac{2t}{d+1}$ corresponds to $\binom{d}{t}$ eigenvectors, namely all those $\phi^{(z)}$ with $|z| = t$. The eigenvectors are also orthogonal:

$$(\phi^{(z)})^T \phi^{(w)} = \sum_{x \in S} \phi_x^{(z)} \phi_x^{(w)} = \sum_{x \in S} (-1)^{z \cdot x}(-1)^{w \cdot x} = \sum_{x \in S} (-1)^{x \cdot (z+w)} = \begin{cases} 2^d & \text{if } z = w \\ 0 & \text{otherwise} \end{cases}$$

Finally, in this case, we obtain $\lambda_* = 1 - \frac{2}{d+1}$, so $\lambda_*^n \approx \exp(-\frac{2n}{d+1})$ for large d and n. On the other hand, the limiting distribution is uniform, so $\pi_x = 2^{-d}$ $\forall x \in S$. Therefore,

$$\frac{1}{2} \exp\left(-\frac{2n}{d+1}\right) \le \|P_0^n - \pi\|_{\text{TV}} \le 2^{\frac{d}{2}-1} \exp\left(-\frac{2n}{d+1}\right)$$

In this particular case, the bounds are loose, and do not really capture the true behaviour of the TV distance. Note that the state space is exponentially large. It turns out that in this example, the TV distance behaves in an unexpected way.

1.4 Cut-off phenomenon

When the cut-off phenomenon occurs, the value of the TV distance does not decay smoothly and exponentially, but rather it rapidly drops in a short interval. More concretely, there is a mixing time τ such that the distance remains close to 1 when $n < \tau$, and rapidly converges to 0 when $n > \tau$. We will prove that the last example observes the cut-off phenomenon, with mixing time $\tau = \frac{d+1}{4} \log d$.

Proposition 1.7. Let c be a large positive constant. In the last example:

$$\text{If } n = \frac{d+1}{4}(\log d + c), \text{ then } \|P_0^n - \pi\|_{\text{TV}} \to 0 \text{ as } c \text{ increases.}$$
$$\text{If } n = \frac{d+1}{4}(\log d - c), \text{ then } \|P_0^n - \pi\|_{\text{TV}} \to 1 \text{ as } c \text{ increases.}$$

Proof of the first statement. Assume that $n = \frac{d+1}{4}(\log d + c)$. In what follows, the first inequality was proved in lecture notes 7, within the proof of the main theorem:

$$\begin{aligned}
\|P_0^n - \pi\|_{\text{TV}} &\leq \frac{1}{2}\left(\sum_{z \in S\setminus\{0\}} \lambda_z^{2n} \underbrace{\left(\phi_0^{(z)}\right)^2}_{=1}\right)^{1/2} = \frac{1}{2}\left(\sum_{t=1}^{d} \binom{d}{t}\left(1 - \frac{2t}{d+1}\right)^{2n}\right)^{1/2} \\
&\leq \frac{1}{2}\left(2\sum_{t=1}^{\lceil d/2 \rceil} \binom{d}{t}\left(1 - \frac{2t}{d+1}\right)^{2n}\right)^{1/2} \leq \frac{1}{\sqrt{2}}\left(\sum_{t=1}^{\lceil d/2 \rceil} \frac{d^t}{t!} \exp\left(-\frac{4tn}{d+1}\right)\right)^{1/2} \\
&\leq \frac{1}{\sqrt{2}}\left(\sum_{t=1}^{\infty} \frac{1}{t!} \exp\left(t \log d - \frac{4tn}{d+1}\right)\right)^{1/2} = \frac{1}{\sqrt{2}}\left(\sum_{t=1}^{\infty} \frac{1}{t!} e^{-tc}\right)^{1/2} \\
&= \frac{1}{\sqrt{2}}\left(\exp(e^{-c}) - 1\right)^{1/2} \approx \frac{1}{\sqrt{2}}\left(e^{-c}\right)^{1/2} = \frac{1}{\sqrt{2}} e^{-c/2}
\end{aligned}$$

Notice that the final expression approaches 0 exponentially as c increases, and does not depend on d. □

Proof of the second statement. Assume that $n = \frac{d+1}{4}(\log d - c)$. We use the second equivalent definition of the TV distance:

$$\|P_0^n - \pi\|_{\text{TV}} = \max_{A \subset S} |P_0^n(A) - \pi(A)| \geq |P_0^n(A) - \pi(A)|, \quad \forall A \in S$$

To obtain a good lower bound, we want to pick an appropriate set $A \subset S$ such that $P_0^n(A) \approx 0$ (for the chosen value of n) and $\pi(A) \approx 1$. As we will see, such a set is in the "center" of S: observe indeed that if the random variable X_∞ (with values in $\{0,1\}^d$) is distributed according to π, then $\mathbb{E}(|X_\infty|) = \frac{d}{2}$. The idea is therefore to include in A all states x with $|x| \approx \frac{d}{2}$. More concretely, we define $f : S \to \mathbb{Z}$ as $f(x) = d - 2|x| = \sum_{t=1}^{d} (-1)^{x_t}$, and $A = \{x \in S : |f(x)| \leq \beta\sqrt{d}\} = \{x \in S : \left||x| - \frac{d}{2}\right| \leq \frac{\beta}{2}\sqrt{d}\}$, where β is a parameter to be defined later.

Claim 1. $\pi(A) \geq 1 - \beta^{-2}$.

Proof.

$$\pi(A) = \mathbb{P}(X_\infty \in A) = \mathbb{P}\left(|f(X_\infty)| \leq \beta\sqrt{d}\right) = 1 - \mathbb{P}\left(|f(X_\infty)| > \beta\sqrt{d}\right) \geq 1 - \frac{\mathbb{E}\left(f^2(X_\infty)\right)}{\beta^2 d}$$

by Chebychev's inequality. The expected value of $f^2(X_\infty)$ is given by

$$\mathbb{E}(f^2(X_\infty)) = \sum_{x\in S} f^2(x)\,\pi_x = 2^{-d}\sum_{x\in S}\left(\sum_{t=1}^{d}(-1)^{x_t}\right)^2 = 2^{-d}\sum_{s,t=1}^{d}\sum_{x\in S}(-1)^{x_s}(-1)^{x_t}$$
$$= 2^{-d}\sum_{s,t=1}^{d}(\phi^{(e_s)})^T\phi^{(e_t)} = 2^{-d}\sum_{s,t=1}^{d}2^d\delta_{s,t} = 2^{-d}\,d\,2^d = d$$

So finally, we obtain $\pi(A) \geq 1 - \frac{d}{\beta^2 d} = 1 - \beta^{-2}$. □

Claim 2. $P_0^n(A) \leq (e^{c/2} - \beta)^{-2}$.

Proof. In order to show that $P_0^n(A)$ is small for the chosen value of n (i.e. in order to show that in n steps, the chain does not have the time, starting from position 0, to reach the "center" of the state space S), we need to analyze the distribution of the random variable X_n conditioned on the starting point $X_0 = 0$. For this, it is convenient to use $\mathbb{P}_0(\cdot)$, $\mathbb{E}_0(\cdot)$ and $\mathrm{Var}_0(\cdot)$ as shorthand notations for $\mathbb{P}(\cdot|X_0 = 0)$, $\mathbb{E}(\cdot|X_0 = 0)$ and $\mathrm{Var}(\cdot|X_0 = 0)$, respectively. We then obtain (using the triangle inequality $|a + b| \geq |b| - |a|$ with $a = f(X_n) - \mathbb{E}_0(f(X_n))$ and $b = \mathbb{E}_0(f(X_n))$):

$$P_0^n(A) = \mathbb{P}_0(X_n \in A) = \mathbb{P}_0\left(|f(X_n)| \leq \beta\sqrt{d}\right) = \mathbb{P}_0\left(|f(X_n) - \mathbb{E}_0(f(X_n)) + \mathbb{E}_0(f(X_n))| \leq \beta\sqrt{d}\right)$$
$$\leq \mathbb{P}_0\left(|f(X_n) - \mathbb{E}_0(f(X_n))| \geq \mathbb{E}_0(f(X_n)) - \beta\sqrt{d}\right) \leq \frac{\mathrm{Var}_0(f(X_n))}{\left(\mathbb{E}_0(f(X_n)) - \beta\sqrt{d}\right)^2}$$

where we have again used Chebychev's inequality. The expectation can be computed as follows:

$$\mathbb{E}_0(f(X_n)) = \sum_{x\in S} p_{0x}(n) f(x) = \sum_{x\in S} p_{0x}(n)\sum_{t=1}^{d}(-1)^{x_t} = \sum_{t=1}^{d}\sum_{x\in S} p_{0x}(n)\phi_x^{(e_t)} = \sum_{t=1}^{d}\left(P^n\phi^{(e_t)}\right)_0$$
$$= \sum_{t=1}^{d}\lambda_{e_t}^n\phi_0^{(e_t)} = \sum_{t=1}^{d}\left(1 - \frac{2}{d+1}\right)^n = d\left(1 - \frac{2}{d+1}\right)^n \approx d\,\exp\left(-\frac{2n}{d+1}\right)$$
$$= d\,\exp\left(\frac{c - \log d}{2}\right) = \sqrt{d}\,e^{c/2}$$

In a similar manner, the variance is given by

$$\mathrm{Var}_0(f(X_n)) = \mathbb{E}_0(f(X_n)^2) - \mathbb{E}_0(f(X_n))^2 \approx \sum_{x\in S} p_{0x}(n) f(x)^2 - d\,e^c = \sum_{x\in S} p_{0x}(n)\sum_{s,t=1}^{d}(-1)^{x_s+x_t} - d\,e^c$$
$$= \sum_{s,t=1}^{d}\sum_{x\in S} p_{0x}(n)\phi_x^{(e_s+e_t)} - d\,e^c = \sum_{s,t=1}^{d}\left(P^n\phi^{(e_s+e_t)}\right)_0 - d\,e^c$$
$$= \sum_{t=1}^{d}\lambda_0^n\phi_0^{(0)} + \sum_{s\neq t}\lambda_{e_s+e_t}^n\phi_0^{(e_s+e_t)} - d\,e^c = d + d(d-1)\left(1 - \frac{4}{d+1}\right)^n - d\,e^c$$
$$\approx d + d(d-1)\exp(c - \log d) - d\,e^c \approx d$$

for large d. Gathering these last three computations together, we obtain

$$P_0^n(A) \leq \frac{d}{\left(\sqrt{d}\,e^{c/2} - \beta\sqrt{d}\right)^2} = (e^{c/2} - \beta)^{-2}$$

Joining the two claims together finally leads to the inequality

$$\|P_0^n - \pi\|_{\mathrm{TV}} \geq |P_0^n(A) - \pi(A)| = \pi(A) - P_0^n(A) \geq 1 - \beta^{-2} - (e^{c/2} - \beta)^{-2}$$

which can be made arbitrarily close to 1 by first choosing β large and then c such that $e^{c/2} \gg \beta$. □