

CS 496 Lecture 11: Sampling via Markov Chains 101

Sidhanth Mohanty

October 22, 2025

1 Some sampling problems

Consider the following sampling problems.

1. **Random spanning tree.** Given a connected graph $G = (V, E)$, sample a uniformly random spanning tree T .
2. **Hardcore model on independent sets.** On a graph G of maximum degree Δ , sample an independent set $I \subseteq V$ with probability

$$\mu_\lambda(I) \propto \lambda^{|I|},$$

where $\lambda > 0$ is a parameter known as the *fugacity*.

3. **Ising model from a quadratic form.** Given a symmetric matrix $M \in \mathbb{R}^{n \times n}$, sample x from $\{\pm 1\}^n$ according to the distribution

$$\mu(x) \propto \exp\left(\frac{1}{2} x^\top M x\right).$$

The key motivating question for us is: *when is a sampling problem tractable or intractable?* The complexity of the above three sampling problems turns out to be well-understood.

1. **Random spanning trees.** It turns out that it is easy to sample a uniformly random spanning tree from any graph.
2. **Hardcore model.** This problem undergoes a *computational phase transition*—there is an efficient when $\lambda < \lambda_c(\Delta) := \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta} \approx \frac{e}{\Delta}$, and the problem is NP-hard above that threshold.
3. **Ising model.** This problem undergoes a similar computational phase transition—it is easy when $\lambda_{\max}(M) - \lambda_{\min}(M) < 1$, and NP-hard otherwise.

In each of these cases, the algorithm to sample from these distributions is to run some simple local Markov chain for $\text{poly}(n)$ time. The template that these Markov chains use is very versatile and generalizes far beyond the above examples.

2 The down-up walk and Glauber dynamics

Definition 2.1 (Down-up walk). Let $\mathcal{F} \subseteq \binom{[n]}{k}$ be a family of k -element subsets of $[n]$. A single step of the k -down-up walk from $S \in \mathcal{F}$:

- i. (*down*) pick $x \in S$ uniformly and set $S' \leftarrow S \setminus \{x\}$;
- ii. (*up*) pick $y \in [n] \setminus S'$ such that $S' \cup \{y\} \in \mathcal{F}$, and move to $S'' = S' \cup \{y\}$.

If each $T \in \mathcal{F}$ has a nonnegative weight $w(T)$, the up-step chooses y with probability proportional to $w(S' \cup \{y\})$ among feasible options.

Remark 2.2. For the problem of choosing a uniformly random spanning tree in a graph, $\mathcal{F}$ is the set of all spanning trees. A step deletes a random edge of the current tree, disconnecting the tree into two components, and then adds a random edge that reconnects these components.

A basic property one would want out of this Markov chain is that its stationary distribution is indeed the target distribution that one is interested in sampling from. For the down-up walk, this is easily verified by checking that the Markov chain satisfies the *detailed balance* condition.

Claim 2.3 (Detailed balance $\Rightarrow$ stationarity). If a distribution π and Markov kernel P satisfy

$$\pi(x) P(x \rightarrow y) = \pi(y) P(y \rightarrow x) \quad \text{for all } x, y,$$

then π is a stationary distribution of P .

Proof. One may verify that $\pi^\top P = \sum_x \pi(x) P(x \rightarrow y) = \sum_x \pi(y) P(y \rightarrow x) = \pi(y)$. □

A related Markov chain on product domains is *Glauber dynamics*

Definition 2.4 (Glauber dynamics). Let μ be a distribution on Σ^n for some alphabet Σ . A single step of the Glauber dynamics Markov chain is given as follows:

- i. pick $v \in [n]$ uniformly at random;
- ii. resample x_v from the conditional marginal $\mu(\cdot \mid x_{-v})$.

Remark 2.5. We can express Glauber dynamics as a down-up walk where $x \in \Sigma^n$ is identified with the set $\{(v, x_v) : v \in [n]\}$. Dropping (v, x_v) and adding (v, σ) according to the conditional marginal realizes Glauber dynamics as a down-up step.

A nice property of Markov chains whose stationary distribution satisfies detailed balance (known as *time-reversible Markov chains*)

Now that we have established canonical Markov chains whose stationary distribution is the target distribution we are interested in sampling from, it remains to understand why running the Markov chain for a short time horizon actually samples, i.e., the *mixing time* of the Markov chain.

3 Mixing times of random walks

In general, the distribution of the state of a Markov chain gets closer to the stationary distribution over time. How close should the distribution we are sampling from be to the stationary distribution for us to declare that our sampling algorithm was successful?

No distinguisher should be able to tell the distribution ν being sampled from apart from the target distribution μ .

This requirement is formally articulated by the *total variation distance* (colloquially known as the “TV distance”) between a pair of distributions:

$$d_{\text{TV}}(\mu, \nu) = \max_{\mathcal{E} \text{ events}} |\Pr_{u \sim \mu}[\mathcal{E}] - \Pr_{u \sim \nu}[\mathcal{E}]|.$$

The event can be thought of as the space of all elements u such that $\mu(u) > \nu(u)$, and the distinguisher that outputs μ when this event occurs, and ν otherwise is optimal. In particular, no distinguisher can tell the two distributions apart with advantage better than the TV distance.

One nice fact about the TV distance is it can be characterized as the ℓ_1 distance between the two distributions.

$$d_{\text{TV}}(\mu, \nu) = \frac{1}{2} \|\mu - \nu\|_1.$$

For a Markov chain with transition matrix P and stationary distribution μ . Denote the eigenvalues of P by $1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_{|\Omega|} \geq -1$ and let the spectral gap be $\delta := 1 - \max\{|\lambda_2|, |\lambda_{|\Omega|}|\}$.

Theorem 3.1 (Spectral gap $\Rightarrow$ rapid mixing). *If the second-largest eigenvalue in magnitude is $1 - \delta$, then for any initial distribution ν and any $t \geq 0$,*

$$\|\nu P^t - \pi\|_{\text{TV}} \leq \frac{1}{2} \frac{1}{\mu_{\min}} (1 - \delta)^t.$$

Equivalently, to achieve total variation distance at most ε it suffices to take

$$t \geq \frac{1}{\delta} \log \left(\frac{1}{\varepsilon \mu_{\min}} \right).$$

Proof idea. Expand $\frac{\nu}{\mu} - 1$ in the eigenbasis and note that P multiplies the orthogonal component by at most $1 - \delta$ each step. (Full details omitted; it’s on the homework.) $\square$

4 Spectral independence

For a distribution μ , the essence of *spectral independence* is that if *all* pairwise correlations in μ (and also all partially pinned versions of μ) small, there is a large spectral gap for Markov chains such as the down-up walk and Glauber dynamics.

Definition 4.1 (Influence matrix). Let μ be a distribution on $\mathcal{F} \subseteq \binom{[n]}{k}$. Define the influence matrix $\Psi^\mu \in \mathbb{R}^{n \times n}$ by

$$\Psi_{i,j}^\mu = \Pr_{S \sim \mu}[i \in S \mid j \in S] - \Pr_{S \sim \mu}[i \in S].$$

For a *pinning* $T \subseteq [n]$, $|T| \leq k - 1$, let $\mu|T$ denote μ conditioned on $T \subseteq S$.

Definition 4.2 (Spectral independence). A family μ is κ -spectrally independent if for every pinning T in the support of μ ,

$$\|\Psi^{\mu|T}\|_{\text{op}} \leq \kappa.$$

Theorem 4.3 (Spectral independence $\Rightarrow$ rapid mixing). *Let μ be a distribution on $\mathcal{F} \subseteq \binom{[n]}{k}$. If for every pinning T the influence matrix satisfies $\|\Psi^{\mu|T}\|_{\text{op}} \leq O(1)$, then the down-up walk has spectral gap at least $n^{-O(\kappa)}$, hence mixes in $n^{O(\kappa)}$ time.*

The use of this conceptual framework for analyzing the mixing time of Markov chains began in the breakthrough work of Anari, Liu, Oveis Gharan, and Vinzant [ALGV19] where they proved that the down-up walk mixes rapidly on the bases of a matroid (see also the work of Anari, Liu, and Oveis Gharan [ALG21] for an application to the hardcore model). Ever since, this framework has been developed further, and the most evolved form is in the *localization schemes* framework of Chen and Eldan [CE22]. The roots of Theorem 4.3 can be traced back to the local-to-global method of Garland; a weak version of Theorem 4.3 was proved by Kaufman and Oppenheim [KO20], and the current version is due to Alev and Lau [AL20].

References

- [AL20] Vedat Levi Alev and Lap Chi Lau. Improved analysis of higher order random walks and applications. In *Proceedings of the 52nd annual ACM SIGACT symposium on theory of computing*, pages 1198–1211, 2020. [4](#)
- [ALG21] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. *SIAM Journal on Computing*, 53(6):FOCS20–1, 2021. [4](#)
- [ALGV19] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials II: high-dimensional walks and an fpras for counting bases of a matroid. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 1–12, 2019. [4](#)
- [CE22] Yuansi Chen and Ronen Eldan. Localization schemes: A framework for proving mixing bounds for markov chains. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 110–122. IEEE, 2022. [4](#)
- [KO20] Tali Kaufman and Izhar Oppenheim. High order random walks: Beyond spectral gap. *Combinatorica*, 40(2):245–281, 2020. [4](#)