Spectral gaps and comparisons of Markov chains

USTC, Hefei, China, Jan 16, 2025
Lecturer: Weiming Feng (ETH Zürich)

1 Glauber dynamics

In the last lecture, we have seen the Metropolis-Hastings chain for graph q-coloring. In this lecture, we will consider another well-known Markov chain, called *Glauber dynamics*. Again, we will start from the Glauber dynamics for graph q-coloring.

- Start from an arbitrary proper coloring $X \in [q]^V$.
- For each t from 1 to T:
  1. Sample a vertex $v \in V$ uniformly at random.
  2. Resample $X_v \sim \mu_v^{X_{V-v}}$, which is the marginal distribution on v conditional X_{V-v}. In other words, X_v is a uniform random color from $[q] \setminus X_{N(v)}$, where $N(v)$ is the set of neighbors of v in G.
- Return the coloring X.

Exercise 1.1. Show that the Glauber dynamics mixes in time $O(\frac{n}{\delta} n \log n)$ if $q \geq (2+\delta)\Delta$ using the coupling method.

The above chain can be easily generalized to a general high-dimensional distribution μ over $[q]^V$. In this lecture, we will study the mixing time of the Glauber dynamics. However, instead of using the coupling method, we will use the spectral analysis. Specifically, we will study the *eigenvalues* of the transition matrix of the Glauber dynamics to bound the mixing time.

2 The spectral analysis and reversible Markov chains

Let P be the transition matrix of a reversible Markov chain over the state space Ω with stationary distribution π. The $\langle \cdot, \cdot \rangle_\pi$ is the weighted inner product defined by $\langle f, g \rangle_\pi = \sum_{x \in \Omega} f(x) g(x) \pi(x)$.

Proposition 2.1 ([LPW17, Sec 12.1])**.** *The matrix P has N real eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ with corresponding eigenvectors $v_1, v_2, \ldots, v_N$ such that eigenvectors form an orthonormal basis of $\mathbb{R}^N$ with respect to the weighted inner product $\langle \cdot, \cdot \rangle_\pi$ (i.e. $\langle v_i, v_j \rangle_\pi = \mathbf{1}[i = j]$) and for $t \geq 1$,*

$$\frac{P^t(x,y)}{\pi(y)} = \sum_{i=1}^{N} \lambda_i^t v_i(x) v_i(y).$$

- *The largest eigenvalue $\lambda_1 = 1$, and the corresponding eigenvector is $v_1 = \mathbf{1}$. If P is irreducible, the eigenspace of λ_1 is one-dimensional, which means $\lambda_2 < \lambda_1$.*
- *If P is irreducible and aperiodic, then -1 is not an eigenvalue of P.*

By the detailed balance condition, we have

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

We can define a symmetric matrix Q by

$$Q(x,y) = \sqrt{\frac{\pi(x)}{\pi(y)}} P(x,y) = \sqrt{\frac{\pi(y)}{\pi(x)}} P(y,x) = Q(y,x).$$

Define the diagonal matrix D by $D_{x,x} = \pi(x)$. Then we have

$$Q = D^{1/2} P D^{-1/2}.$$

Let $|\Omega| = N$. Since Q is real symmetric, it has real eigenvalues. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ be the eigenvalues of Q with corresponding orthonormal eigenvectors $\varphi_1, \varphi_2, \ldots, \varphi_N$. Define vectors $v_1, v_2, \ldots, v_N$ by $v_i = D^{-1/2}\varphi_i$. Then we have

$$Pv_i = PD^{-1/2}\varphi_i = D^{-1/2}D^{1/2}PD^{-1/2}\varphi_i = D^{-1/2}Q\varphi_i = D^{-1/2}\lambda_i\varphi_i = \lambda_i v_i.$$

Hence, $v_1, v_2, \ldots, v_N$ are the eigenvectors of P with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$. Let $\langle f, g\rangle_\pi$ define the weighted inner product by $\langle f, g\rangle_\pi = \sum_{x\in\Omega} f(x)g(x)\pi(x)$. Then we have

$$\langle v_i, v_j\rangle_\pi = \langle D^{-1/2}\varphi_i, D^{-1/2}\varphi_j\rangle_\pi = \langle \varphi_i, \varphi_j\rangle = \mathbf{1}[i=j].$$

Hence, $v_1, v_2, \ldots, v_N$ are orthonormal with respect to the weighted inner product $\langle\cdot,\cdot\rangle_\pi$. Any vector $f \in \mathbb{R}^N$ can be written as

$$f = \sum_{i=1}^{N} \langle f, v_i\rangle_\pi v_i.$$

This is because the coefficient vector $x_i = \langle f, v_i\rangle_\pi$ is the solution of linear system $Vx = f$, where V is the matrix with columns $v_1, v_2, \ldots, v_N$. The equation holds because the inverse of V is $V^T D$ and $(V^T D f)(x) = \langle f, x\rangle_\pi$. Again, it is not hard to verify that P can be decomposed as

$$P = \sum_{i=1}^{N} \lambda_i v_i v_i^T D \quad\iff\quad \frac{P(x,y)}{\pi(y)} = \sum_{i=1}^{N} \lambda_i v_i(x) v_i(y).$$

Consider

$$P^2 = \left(\sum_{i=1}^{N} \lambda_i v_i v_i^T D\right)\left(\sum_{i=1}^{N} \lambda_i v_i v_i^T D\right) = \sum_{ij} \lambda_i\lambda_j v_i \langle v_i, v_j\rangle_\pi v_j^T D = \sum_i \lambda_i^2 v_i v_i^T D.$$

Next, we take a look at the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$ of the transition matrix P. Since the row sum of P is 1, we have $\max_i |\lambda_i| \leq 1$. The largest eigenvalue $\lambda_1 = 1$, and the corresponding eigenvector is $v_1 = \mathbf{1}$. We show that, if P is irreducible, the eigenspace of λ_1 is one-dimensional. Let f be an eigenvector with eigenvalue $\lambda_1 = 1$. Then we have

$$(Pf)(x) = \sum_{y\in\Omega} P(x,y)f(y) = \mathbf{E}_{y\sim P(x,\cdot)}\left[f(y)\right] = f(x).$$

Let $x \in \Omega$ satisfying $f(x) = \max_{y\in\Omega} f(y)$. Then, for any $y \in \Omega$ such that $P(x,y) > 0$, since $f(x)$ is the average of $f(y)$'s, we have $f(y) = f(x)$. Otherwise, there exists $y \in \Omega$ such that $f(y) > f(x)$. We can repeat the above argument to show that if the Markov chain is irreducible, then f is a constant vector.

Exercise 2.2. Use the same argument as above to show that if P is irreducible and aperiodic, then -1 is not an eigenvalue of P. For any eigenvector f with eigenvalue -1, one can use the sign of $f(x)$ to partition the state space Ω into two parts. If $P(x,y) > 0$, then x and y are in different parts.

3 The spectral gap and relaxation time

Now, let us focus on the case that P is a positive semidefinite matrix. If P is not, we can replace P by $\frac{1}{2}(P+I)$, which means in every step, the chain stays in the same state with probability 1/2 and makes the transition defined by P with probability 1/2. Define the spectral gap

$$\gamma = 1 - \lambda_2.$$

For PSD P, define the relaxation time T_{rel} by

$$T_{\text{rel}} = \frac{1}{\gamma}.$$

In general case when P is not PSD, the relaxation time is defined by $\frac{1}{\gamma^*}$, where $\gamma^* = 1 - \lambda^*$ the called the absolute spectral gap, and $\lambda^* = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } P, \lambda \neq 1\}$. In this mini-course, we will mainly focus on the case that P is PSD.

Theorem 3.1 (mixing time upper bound)**.** *Let* $\pi_{\min} = \min_{x\in\Omega} \pi(x)$. *Then*

$$T_{mix}(\varepsilon) = O\left(T_{rel} \cdot \log \frac{1}{\varepsilon \pi_{\min}}\right).$$

Proof. By the spectral decomposition, we have

$$\frac{P^t(x,y)}{\pi(y)} = \sum_{i=1}^{N} \lambda_i^t v_i(x) v_i(y) = 1 + \underbrace{\sum_{i=2}^{N} \lambda_i^t v_i(x) v_i(y)}_{\text{error term}}.$$

We show the error term shrinks with time t. Using Cauchy-Schwarz inequality, we have

$$\begin{aligned}\left(\frac{P^t(x,y)}{\pi(y)} - 1\right)^2 = \left(\sum_{i=2}^{N} \lambda_i^t v_i(x) v_i(y)\right)^2 &\leq \left(\sum_{i=2}^{N} \lambda_i^t v_i^2(x)\right)\left(\sum_{i=2}^{N} \lambda_i^t v_i^2(y)\right) \\ &\leq \lambda_2^{2t} \left(\sum_{i=2}^{N} v_i^2(x)\right)\left(\sum_{i=2}^{N} v_i^2(y)\right).\end{aligned}$$

For general transition matrix P (not necessarily PSD), we can replace the λ_2 in the above inequality by $\lambda^* = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } P \text{ and } \lambda \neq 1\}$. Note that v_i's form an orthonormal basis with respect to the weighted inner product $\langle \cdot, \cdot \rangle_\pi$. Let δ_x denote the indicator function of x, i.e. $\delta_x(y) = 1[x = y]$. Then we have

$$1 = \delta_x(x) = \sum_{i=1}^{N} \langle \delta_x, v_i \rangle_\pi v_i(x) = \sum_{i=2}^{N} v_i^2(x) \pi(x).$$

Hence,

$$\left|\frac{P^t(x,y)}{\pi(y)} - 1\right| \leq \lambda_2^t \cdot \frac{1}{\sqrt{\pi(x)\pi(y)}} \leq \frac{\lambda_2^t}{\pi_{\min}}.$$

Since $2d_{\text{TV}}\left(P^t(x,\cdot), \pi\right)$ is all $\left|\frac{P^t(x,y)}{\pi(y)} - 1\right|$ for $y \in \Omega$ averaged by distribution π, if we set the time $t = O(T_{\text{rel}} \log \frac{1}{\varepsilon \pi_{\min}})$, then the total variation distance is at most ε. $\square$

Theorem 3.2 (mixing time lower bound)**.** *The mixing time can be lower bound by*

$$T_{mix}(\varepsilon) = \Omega\left(T_{rel} \cdot \log \frac{1}{\varepsilon}\right).$$

Proof. Let v_2 be the eigenvector corresponding to the second largest eigenvalue λ_2. Then we have $\mathbf{E}_\pi [Pv_2] = \sum_x \pi(x) \sum_y P(x,y)v_2(y) = \sum_{x,y} \pi(y)P(y,x)v_2(y) = \mathbf{E}_\pi [v_2]$. Note that $\lambda \neq 1$. If $\lambda_2 = 0$, then $\mathbf{E}_\pi [v_2] = \mathbf{E}_\pi [0] = 0$. If $\lambda_2 \neq 0$, then $\mathbf{E}_\pi [v_2] = \lambda_2 \mathbf{E}_\pi [v_2]$, which implies $\mathbf{E}_\pi [v_2] = 0$.

We can write the following inequality

$$\begin{aligned}
|\lambda_2^t v_2(x)| = \left|\sum_y P^t(x,y)v_2(y) - \sum_y \pi(y)v_2(y)\right| &\leq \sum_y |v_2(y)|\pi(y) \left|\frac{P^t(x,y)}{\pi(y)} - 1\right| \\
&\leq \sum_y \|v_2\|_\infty \pi(y) \left|\frac{P^t(x,y)}{\pi(y)} - 1\right|.
\end{aligned}$$

If we take x such that $|v_2(x)| = \|v_2\|_\infty$, then

$$\lambda_2^t \leq \sum_y \pi(y) \left|\frac{P^t(x,y)}{\pi(y)} - 1\right| = 2d_{\mathrm{TV}}\left(P^t(x,\cdot), \pi\right).$$

This gives the lower bound of the mixing time. □

4 Bound spectral gap via coupling

Theorem 4.1 ([Che98])**.** *Let Ω a metric space with metric ρ satisfying $\rho(x,y) \geq 1$ if $x \neq y$. Let $\theta \in (0,1)$. Suppose for any $x, y \in \Omega$, $X \sim P(x,\cdot)$ and $Y \sim P(y,\cdot)$ can be coupled such that*

$$\rho(X,Y) \leq \theta\rho(x,y).$$

For any eigenvalue $\lambda \neq 1$ of P, $|\lambda| \leq \theta$. In particular, the spectral gap of P is at least $1-\theta$.

Remark 4.2. For Metropolis-Hastings chain for graph q-coloring, ρ can be set to be the Hamming distance. The above theorem shows that if $q \geq (2+\delta)\Delta$, then the spectral gap of the Metropolis-Hastings chain is at least $\Omega(\frac{\delta}{n})$.

Proof. Let $f : \Omega \to \mathbb{R}$ be a function. Define the Lipschitz constant $L(f)$ of f by

$$L(f) = \sup_{x,y\in\Omega: x\neq y} \frac{|f(x) - f(y)|}{\rho(x,y)}.$$

Let f be an eigenvector of P with eigenvalue λ. Let $X \sim P(x,\cdot)$ and $Y \sim P(y,\cdot)$ be the coupled random variables. By linearity of expectation and triangle inequality, we have

$$|Pf(x) - Pf(y)| = |\mathbf{E}[f(X)] - \mathbf{E}[f(Y)]| \leq \mathbf{E}_{X,Y}[|f(X) - f(Y)|].$$

Using Lipschitz constant, we have

$$|Pf(x) - Pf(y)| \leq L(f) \cdot \mathbf{E}_{X,Y}[\rho(X,Y)] \leq L(f)\theta\rho(x,y).$$

On the other hand, the above inequality holds for any $x, y \in \Omega$. This implies

$$L(Pf) \leq \theta L(f).$$

Taking f as the eigenvector of λ implies

$$L(Pf) = L(\lambda f) = |\lambda| L(f) \leq \theta L(f).$$

Hence, $|\lambda| \leq \theta$ because f is not a constant function so that $L(f) > 0$. □

5 Markov chain comparison

Now, we know that the spectral gap of Metropolis-Hastings chain for graph q-coloring is at least $\Omega(\frac{\delta}{n})$. We show how to prove the spectral gap of Glauber dynamics is also $\Omega(\frac{\delta}{n})$ by comparing the spectral gap of two Markov chains, which implies $O(\frac{n^2}{\delta}\log q)$ mixing time for Glauber dynamics because $\pi_{\min} \geq 1/q^n$. One may notice that in Exercise 1.1, we have already used the coupling method to prove a faster $O(n \log n)$ mixing time of Glauber dynamics. Why do we need to prove this $O(n^2)$ mixing time? Here are the reasons.

- This example shows that compared to coupling stepwise contraction, the spectral gap often gives a slower mixing time.
- However, we use coloring as a running example in this mini-course. For other models (such as graph matching), it is hard to find a coupling proof of the mixing time but one can analyze the spectral gap. The upper and lower bounds for mixing time shows that up to a $\log \frac{1}{\pi_{\min}}$ factor (which is typically a polynomial in n), the mixing time is fully determined by the spectral gap.

We first give some abstract results on Markov chain comparison. Let P be a reversible Markov chain over Ω with stationary distribution π. For any functions $f, g : \Omega \to \mathbb{R}$, define the *Dirichlet form* of f by

$$\mathcal{E}(f, f) = \langle (I - P)f, f\rangle_\pi.$$

Using the detailed balance condition, we can verify that

$$\mathcal{E}(f, f) = \frac{1}{2} \sum_{x,y\in\Omega} \pi(x)P(x, y)(f(x) - f(y))^2.$$

Exercise 5.1. Verify the above identity.

Theorem 5.2. *The spectral gap of P can be characterized by the Dirichlet form as follows.*

$$\gamma = \min_{f\neq 0, f\perp_\pi 1} \frac{\mathcal{E}(f, f)}{\langle f, f\rangle_\pi},$$

where $f \perp_\pi 1$ denotes $\langle f, 1\rangle_\pi = 0$.

Proof. Let $v_1, v_2, \ldots, v_N$ be the eigenvectors of P with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_N$. Since $f \perp_\pi 1$, we have $f = \sum_{i=2}^{N} \langle f, v_i\rangle_\pi v_i$. We have

$$\mathcal{E}(f, f) = \langle f, f\rangle_\pi - \langle Pf, f\rangle_\pi = \sum_{i=2}^{N} (1 - \lambda_i)\langle f, v_i\rangle_\pi^2 \geq (1 - \lambda_2) \sum_{i=2}^{N} \langle f, v_i\rangle_\pi^2 = (1 - \lambda_2)\langle f, f\rangle_\pi.$$

On the other hand, we can take $f = v_2$ to make the equality hold. ☐

For any function f. The variance of f is defined by

$$\mathbf{Var}_\pi [f] = \mathbf{E}_\pi \left[f^2\right] - \mathbf{E}_\pi [f]^2 = \langle f, f\rangle_\pi - E[\pi]f^2.$$

Consider the function $\hat{f} = f - \mathbf{E}_\pi [f]$. It is easy to show that $\mathcal{E}(\hat{f}, \hat{f}) = \mathcal{E}(f, f)$ and $\mathbf{Var}_\pi [f] = \mathbf{Var}_\pi[\hat{f}] = \langle \hat{f}, \hat{f}\rangle_\pi$. The equation in Theorem 5.2 can be written as

$$\gamma = \min_{f\neq 0, f\perp_\pi 1} \frac{\mathcal{E}(f, f)}{\langle f, f\rangle_\pi} = \min_{f\neq 0, f\perp_\pi 1} \frac{\mathcal{E}(f - \mathbf{E}[f], f - \mathbf{E}[f])}{\langle f - \mathbf{E}[f], f - \mathbf{E}[f]\rangle_\pi} = \min_{\mathbf{Var}_\pi[f]\neq 0} \frac{\mathcal{E}(f, f)}{\mathbf{Var}_\pi [f]}.$$

Theorem 5.3. *Let γ_G and γ_M be the relaxation time of the Glauber dynamics and Metropolis-Hastings chain, respectively. Then*

$$\gamma_G \geq \gamma_M = \Omega\left(\frac{\delta}{n}\right).$$

Proof. Let P_G denote the transition matrix of the Glauber dynamics. Let P_M denote the transition matrix of the Metropolis-Hastings chain. Since two chains has the same stationary distribution π, we show that for any function f, the ratio of the Dirichlet form satisfies

$$\frac{\sum_{x,y} \pi(x) P_G(x,y)(f(x)-f(y))^2}{\sum_{x,y} \pi(x) P_M(x,y)(f(x)-f(y))^2} \geq 1,$$

where we use the convention that $\frac{0}{0} = 1$. To bound the ratio, it suffices to consider all $x \neq y$ such that $P_M(x,y) > 0$. Then, x and y differ only at one vertex. It holds that $P_M(x,y) = \frac{1}{nq}$ and $P_G(x,y) > \frac{1}{nq}$. This proves the ratio bound. The spectral gap of P_G can be written as

$$\gamma_G = \min_{\mathbf{Var}_\pi[f] \neq 0} \frac{\mathcal{E}_G(f,f)}{\mathbf{Var}_\pi[f]} \geq \frac{1}{2},$$

where $\mathcal{E}_G$ denotes the Dirichlet form of P_G. Let f the function that can achieve the minimum. We have

$$\gamma_G = \frac{\mathcal{E}_G(f,f)}{\mathbf{Var}_\pi[f]} \geq \frac{\mathcal{E}_M(f,f)}{\mathbf{Var}_\pi[f]} \geq \min_{g:\mathbf{Var}_\pi[g]\neq 0} \frac{\mathcal{E}_M(g,g)}{\mathbf{Var}_\pi[g]} = \gamma_M. \qquad \square$$

Theorem 5.2 also provides a way to upper bound the spectral gap by choosing an arbitrary function f. If the transition matrix P is PSD, an upper bound of spectral gap implies a lower bound of relaxation time, which implies a lower bound of mixing time.

Theorem 5.4. *Suppose $q > (1+\delta)\Delta$, where $\delta > 0$ is a constant. The spectral gaps of Glauber dynamics and Metropolis-Hastings chain for graph coloring can be upper bounded by*

$$\gamma_G, \gamma_M \leq O\left(\frac{1}{n}\right).$$

Proof. Fix an arbitrary vertex v. We partition all colorings Ω into two parts Ω_R and Ω_{R^c}, where Ω_R contains all colorings such that v takes the color 0, say the color red. Define a function f such that for all $x \in \Omega_R$, $f(x) = q-1$ and for all $x \in \Omega_{R^c}$, $f(x) = -1$. It is easy to verify that $f \perp_\pi 1$ and $f \neq 0$. We can compute

$$\langle f, f\rangle_\pi = \sum_{x\in\Omega} \pi(x) = \frac{(q-1)^2}{q} + \frac{q-1}{q} = q-1.$$

For the Dirichlet form of both Glauber dynamics and Metropolis-Hastings chain, we can bound

$$\begin{aligned}
\mathcal{E}(f,f) &= \frac{1}{2}\sum_{x\in\Omega_R, y\in\Omega_{R^c}} \pi(x)P(x,y)(f(x)-f(y))^2 + \frac{1}{2}\sum_{x\in\Omega_{R^c}, y\in\Omega_R} \pi(x)P(x,y)(f(x)-f(y))^2 \\
&\leq \frac{1}{2}\sum_{x\in\Omega_R, y\in\Omega_{R^c}} \pi(x)P(x,y)q^2 + \frac{1}{2}\sum_{x\in\Omega_{R^c}, y\in\Omega_R} \pi(x)P(x,y)q^2 \\
&= q^2 \sum_{x\in\Omega_R, y\in\Omega_{R^c}} \pi(x)P(x,y). && \text{(by reversibility)}
\end{aligned}$$

The last term can be written as $\sum_{x\in\Omega_R}\pi(x)\sum_{y\in\Omega_{R^c}}P(x,y)$, where the sum of transition probabilities is at most $1/n$. This is because to move from x to Ω_{R^c}, the chain needs to pick vertex v, which happens with probability $1/n$. Hence, the last term is at most $\frac{q^2}{qn}=\frac{q}{n}$.

$$\mathcal{E}(f,f)=\frac{q}{n}.$$

This implies $\gamma_G,\gamma_M\le O(\frac{1}{n})$. □

Finally, given the bounds on spectral gap, to bound the relaxation time and mixing time of Glauber dynamics, we need to verify that the transition matrix P_G is PSD. The simplest trick is to consider lazy Glauber dynamics, where in every step, the chain is lazy with probability 1/2. However, the transition matrix of lazy Glauber dynamics is indeed PSD [DGU14]. We will prove this fact in next lecture. We can directly use Theorem 3.1 and Theorem 3.2 to get the mixing time of Glauber dynamics:

$$\begin{aligned}T_{\text{mix}}(\varepsilon)&=O\left(T_{\text{rel}}\cdot\log\frac{1}{\varepsilon\pi_{\min}}\right)=O\left(\frac{n}{\delta}\left(n\log q+\log\frac{1}{\varepsilon}\right)\right),\\ T_{\text{mix}}(\varepsilon)&=\Omega\left(T_{\text{rel}}\cdot\log\frac{1}{\varepsilon}\right)=\Omega\left(n\log\frac{1}{\varepsilon}\right).\end{aligned}$$

There are many advanced techniques to compare the spectral gap of Markov chains. For example, one can compare a Markov chain P (say Glauber dynamics or Metropolis-Hastings) to the one step mixing chain Q, where $Q(x,y)=\pi(y)$ for all $x,y\in\Omega$. The spectral gap of Q is 1. One can use a path of transitions in P to *mimic* one transition in Q. The spectral gap of P can be captured by the length of the paths and the *congestion* of the paths. This kind of techniques is called the *canonical path* [Jer03] and the *path method* [LPW17, Sec 13.4].

References

[Che98] Mu-Fa Chen. Trilogy of couplings and general formulas for lower bound of spectral gap. In *Probability towards 2000 (New York, 1995)*, volume 128 of *Lect. Notes Stat.*, pages 123–136. Springer, New York, 1998.

[DGU14] Martin Dyer, Catherine Greenhill, and Mario Ullrich. Structure and eigenvalues of heat-bath Markov chains. *Linear Algebra Appl.*, 454:57–71, 2014.

[Jer03] Mark Jerrum. *Counting, sampling and integrating: algorithms and complexity.* Springer Science & Business Media, 2003.

[LPW17] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times.* American Mathematical Society, Providence, RI, 2017.