# Chapter 4

# Coupling

In this chapter we move on to *coupling*, another probabilistic technique with a wide range of applications (far beyond discrete stochastic processes). The idea behind the coupling method is deceptively simple: to compare two probability measures $\mu$ and $\nu$, it is sometimes useful to construct a *joint* probability space with marginals $\mu$ and $\nu$. For instance, in the classical application of coupling to the convergence of Markov chains (Theorem 1.1.33), one simultaneously constructs *two* copies of a Markov chain—one of which is already at stationarity—and shows that they can be made to coincide after a random amount of time called the coupling time. We begin in Section 4.1 by defining coupling formally and deriving its connection to the total variation distance through the *coupling inequality*. We illustrate the basic idea on a classical Poisson approximation result, which we apply to the degree sequence of an Erdős-Rényi graph. In Section 4.2, we introduce the concept of *stochastic domination* and some related *correlation inequalities*. We develop a key application in percolation theory. Coupling of Markov chains is the subject of Section 4.3, where it serves as a powerful tool to derive mixing time bounds. Finally, we end in Section 4.4 with the *Chen-Stein method* for Poisson approximations, a technique that applies in particular in some natural settings with dependent variables.

## 4.1 Background

We begin with some background on coupling. After defining the concept formally and giving a few simple examples, we derive the coupling inequality, which provides a fundamental approach to bounding the distance between two distributions.

---

Version: December 20, 2023
Modern Discrete Probability: An Essential Toolkit

As an application, we analyze the degree distribution in the Erdős-Rényi graph model. Throughout this chapter, $(S, \mathcal{S})$ is a measurable space. Also we will denote by $\mu_Z$ the law of random variable $Z$.

### 4.1.1 Basic definitions

A formal definition of coupling follows. Recall (see Appendix B) that for measurable spaces $(S_1, \mathcal{S}_1)$ $(S_2, \mathcal{S}_2)$, we can consider the product space $(S_1 \times S_2, \mathcal{S}_1 \times \mathcal{S}_2)$ where

$$S_1 \times S_2 := \{(s_1, s_2) : s_1 \in S_1, s_2 \in S_2\}$$

is the Cartesian product of $S_1$ and $S_2$, and $\mathcal{S}_1 \times \mathcal{S}_2$ is the smallest $\sigma$-algebra on $S_1 \times S_2$ containing the rectangles $A_1 \times A_2$ for all $A_1 \in \mathcal{S}_1$ and $A_2 \in \mathcal{S}_2$.

**Definition 4.1.1** (Coupling)**.** *Let $\mu$ and $\nu$ be probability measures on the same measurable space $(S, \mathcal{S})$. A* coupling *of $\mu$ and $\nu$ is a probability measure $\gamma$ on the product space $(S \times S, \mathcal{S} \times \mathcal{S})$ such that the* marginals *of $\gamma$ coincide with $\mu$ and $\nu$, that is,* *coupling*

$$\gamma(A \times S) = \mu(A) \quad \textit{and} \quad \gamma(S \times A) = \nu(A), \qquad \forall A \in \mathcal{S}.$$

*For two random variables $X$ and $Y$ taking values in $(S, \mathcal{S})$, a* coupling *of $X$ and $Y$ is a joint variable $(X', Y')$ taking values in $(S \times S, \mathcal{S} \times \mathcal{S})$ whose law as a probability measure is a coupling of the laws of $X$ and $Y$. Note that, under this definition, $X$ and $Y$ need not be defined on the same probability space (but $X'$ and $Y'$ do need to). We also say that $(X', Y')$ is a coupling of $\mu$ and $\nu$ if the law of $(X', Y')$ is a coupling of $\mu$ and $\nu$.*

We give a few examples.

**Example 4.1.2** (Coupling of Bernoulli variables)**.** Let $X$ and $Y$ be Bernoulli random variables with respective parameters $0 \leq q < r \leq 1$. That is, $\mathbb{P}[X = 1] = q$ and $\mathbb{P}[Y = 1] = r$. Here $S = \{0, 1\}$ and $\mathcal{S} = 2^S$.

- *(Independent coupling)* One coupling of $X$ and $Y$ is $(X', Y')$ where $X' \stackrel{\mathrm{d}}{=} X$ and $Y' \stackrel{\mathrm{d}}{=} Y$ are *independent* of one another. Its law is

$$\Big(\mathbb{P}[(X', Y') = (i, j)]\Big)_{i,j \in \{0,1\}} = \begin{pmatrix} (1-q)(1-r) & (1-q)r \\ q(1-r) & qr \end{pmatrix}.$$

- *(Monotone coupling)* Another possibility is to pick $U$ uniformly at random in $[0, 1]$, and set $X'' = \mathbf{1}_{\{U \leq q\}}$ and $Y'' = \mathbf{1}_{\{U \leq r\}}$. Then $(X'', Y'')$ is a coupling of $X$ and $Y$ with law

$$\Big(\mathbb{P}[(X'', Y'') = (i, j)]\Big)_{i,j \in \{0,1\}} = \begin{pmatrix} 1-r & r-q \\ 0 & q \end{pmatrix}.$$

◀

**Example 4.1.3** (Bond percolation: monotonicity)**.** Let $G = (V, E)$ be a countable graph. Denote by $\mathbb{P}_p$ the law of bond percolation (Definition 1.2.1) on $G$ with density $p$. Let $x \in V$ and assume $0 \leq q < r \leq 1$. Using the monotone coupling in the previous example *on each edge independently* produces a coupling of $\mathbb{P}_q$ and $\mathbb{P}_r$. More precisely:

- Let $\{U_e\}_{e \in E}$ be independent uniforms on $[0, 1]$.
- For $p \in [0, 1]$, let $W_p$ be the set of edges $e$ such that $U_e \leq p$.

Thinking of $W_p$ as specifying the open edges in the percolation process on $G$ under $\mathbb{P}_p$, we see that $(W_q, W_r)$ is a coupling of $\mathbb{P}_q$ and $\mathbb{P}_r$ with the property that $\mathbb{P}[W_q \subseteq W_r] = 1$. Let $\mathcal{C}_x^{(q)}$ and $\mathcal{C}_x^{(r)}$ be the open clusters of $x$ under $W_q$ and $W_r$ respectively. Because $\mathcal{C}_x^{(q)} \subseteq \mathcal{C}_x^{(r)}$,

$$\begin{aligned}\theta(q) &:= \mathbb{P}_q[|\mathcal{C}_x| = +\infty] \\ &= \mathbb{P}[|\mathcal{C}_x^{(q)}| = +\infty] \\ &\leq \mathbb{P}[|\mathcal{C}_x^{(r)}| = +\infty] \\ &= \mathbb{P}_r[|\mathcal{C}_x| = +\infty] \\ &= \theta(r).\end{aligned}$$

(We made this claim in Section 2.2.4.) ◀

**Example 4.1.4** (Biased random walk on $\mathbb{Z}$)**.** For $p \in [0, 1]$, let $(S_t^{(p)})$ be nearest-neighbor random walk on $\mathbb{Z}$ started at 0 with probability $p$ of jumping to the right and probability $1 - p$ of jumping to the left. (See the gambler's ruin problem in Example 3.1.43.) Assume $0 \leq q < r \leq 1$. Using again the monotone coupling of Bernoulli variables above we produce a coupling of $(S_t^{(q)})$ and $(S_t^{(r)})$.

- Let $(X_i'', Y_i'')_i$ be an infinite sequence of i.i.d. monotone Bernoulli couplings with parameters $q$ and $r$ respectively.
- Define $(Z_i^{(q)}, Z_i^{(r)}) := (2X_i'' - 1, 2Y_i'' - 1)$. Note that $\mathbb{P}[2X_1'' - 1 = 1] = \mathbb{P}[X_1'' = 1] = q$ and $\mathbb{P}[2X_1'' - 1 = -1] = \mathbb{P}[X_1'' = 0] = 1 - q$, and similarly for $Y_i''$.
- Let $\hat{S}_t^{(q)} = \sum_{i \leq t} Z_i^{(q)}$ and $\hat{S}_t^{(r)} = \sum_{i \leq t} Z_i^{(r)}$.

Then $(\hat{S}_t^{(q)}, \hat{S}_t^{(r)})$ is a coupling of $(S_t^{(q)})$ and $(S_t^{(r)})$ such that $\hat{S}_t^{(q)} \leq \hat{S}_t^{(r)}$ for all $t$ almost surely. In particular, we deduce that for all $y$ and all $t$

$$\mathbb{P}[S_t^{(q)} \leq y] = \mathbb{P}[\hat{S}_t^{(q)} \leq y] \geq \mathbb{P}[\hat{S}_t^{(r)} \leq y] = \mathbb{P}[S_t^{(r)} \leq y].$$

◀

### 4.1.2 ▷ *Random walks: harmonic functions on lattices and infinite $d$-regular trees*

Let $(X_t)$ be a Markov chain on a finite or countably infinite state space $V$ with transition matrix $P$ and let $\mathbb{P}_x$ be the law of $(X_t)$ started at $x$. We say that a function $h : V \to \mathbb{R}$ is bounded if $\sup_{x\in V} |h(x)| < +\infty$. Recall from Section 3.3 that $h$ is harmonic (with respect to $P$) on $V$ if

$$h(x) = \sum_{y\in V} P(x,y)h(y), \qquad \forall x \in V.$$

We first give a coupling-based criterion for bounded harmonic functions to be constant. Recall that we treated the finite state-space case (where boundedness is automatic) in Corollary 3.3.3.

**Lemma 4.1.5** (Coupling and bounded harmonic functions)**.** *If, for all $y, z \in V$, there is a coupling $((Y_t)_t, (Z_t)_t)$ of $\mathbb{P}_y$ and $\mathbb{P}_z$ such that*

$$\lim_t \mathbb{P}[Y_t \neq Z_t] = 0,$$

*then all bounded harmonic functions on $V$ are constant.*

*Proof.* Let $h$ be bounded and harmonic on $V$ with $\sup_x |h(x)| = M < +\infty$. Let $y, z$ be any points in $V$. Then, arguing as in Section 3.3.1, $(h(Y_t))$ and $(h(Z_t))$ are martingales and, in particular,

$$\mathbb{E}[h(Y_t)] = \mathbb{E}[h(Y_0)] = h(y) \quad \text{and} \quad \mathbb{E}[h(Z_t)] = \mathbb{E}[h(Z_0)] = h(z).$$

So by Jensen's inequality (Theorem B.4.15) and the boundedness assumption

$$\begin{aligned}|h(y) - h(z)| &= |\mathbb{E}[h(Y_t)] - \mathbb{E}[h(Z_t)]| \\ &\leq \mathbb{E}\,|h(Y_t) - h(Z_t)| \\ &\leq 2M\,\mathbb{P}[Y_t \neq Z_t] \\ &\to 0.\end{aligned}$$

So $h(y) = h(z)$. ■

**Harmonic functions on $\mathbb{Z}^d$** Consider random walk on $\mathbb{L}^d$ for $d \geq 1$. In that case, we show that all bounded harmonic functions are constant.

**Theorem 4.1.6** (Bounded harmonic functions on $\mathbb{Z}^d$)**.** *All bounded harmonic functions on $\mathbb{L}^d$ are constant.*

*Proof.* From (3.3.2), $h$ is harmonic with respect to random walk on $\mathbb{L}^d$ if and only if it is harmonic with respect to *lazy* random walk (Definition 1.1.31), that is, the walk that stays put with probability $1/2$ at every step. Let $\mathbb{P}_y$ and $\mathbb{P}_z$ be the laws of lazy random walk on $\mathbb{L}^d$ started at $y$ and $z$ respectively. We construct a coupling $((Y_t),(Z_t)) = ((Y_t^{(i)})_{i\in[d]}, (Z_t^{(i)})_{i\in[d]})$ of $\mathbb{P}_y$ and $\mathbb{P}_z$ as follows: at time $t$, pick a coordinate $I \in [d]$ uniformly at random, then

- if $Y_t^{(I)} = Z_t^{(I)}$ then do nothing with probability $1/2$ and otherwise pick $W \in \{-1,+1\}$ uniformly at random, set $Y_{t+1}^{(I)} = Z_{t+1}^{(I)} := Z_t^{(I)} + W$ and leave the other coordinates unchanged;
- if instead $Y_t^{(I)} \neq Z_t^{(I)}$, pick $W \in \{-1,+1\}$ uniformly at random, and with probability $1/2$ set $Y_{t+1}^{(I)} := Y_t^{(I)} + W$ and leave $Z_t$ and the other coordinates of $Y_t$ unchanged, or otherwise set $Z_{t+1}^{(I)} := Z_t^{(I)} + W$ and leave $Y_t$ and the other coordinates of $Z_t$ unchanged.

It is straightforward to check that $((Y_t),(Z_t))$ is indeed a coupling of $\mathbb{P}_y$ and $\mathbb{P}_z$. To apply the previous lemma, it remains to bound $\mathbb{P}[Y_t \neq Z_t]$.

The key is to note that, for each coordinate $i$, the difference $(Y_t^{(i)} - Z_t^{(i)})$ is itself a nearest-neighbor random walk on $\mathbb{Z}$ started at $y^{(i)} - z^{(i)}$ with holding probability (i.e., probability of staying put) $1 - \frac{1}{d}$—until it hits $0$. Simple random walk on $\mathbb{Z}$ is irreducible and recurrent (Theorem 3.3.38). The holding probability does not affect the type of the walk. So $(Y_t^{(i)} - Z_t^{(i)})$ hits $0$ in finite time with probability $1$. Hence, letting $\tau^{(i)}$ be the first time $Y_t^{(i)} - Z_t^{(i)} = 0$, we have $\mathbb{P}[Y_t^{(i)} \neq Z_t^{(i)}] \leq \mathbb{P}[\tau^{(i)} > t] \to \mathbb{P}[\tau^{(i)} = +\infty] = 0$.

By a union bound,

$$\mathbb{P}[Y_t \neq Z_t] \leq \sum_{i\in[d]} \mathbb{P}[Y_t^{(i)} \neq Z_t^{(i)}] \to 0,$$

as desired. ■

Exercise 4.1 asks for an example of a non-constant (necessarily unbounded) harmonic function on $\mathbb{Z}^d$.

**Harmonic functions on $\mathbb{T}_d$** On trees, the situation is different. Let $\mathbb{T}_d$ be the infinite $d$-regular tree with root $\rho$. For $x \in \mathbb{T}_d$, we let $T_x$ be the subtree, rooted at $x$, of descendants of $x$.

**Theorem 4.1.7** (Bounded harmonic functions on $\mathbb{T}_d$)**.** *For $d \geq 3$, let $(X_t)$ be simple random walk on $\mathbb{T}_d$ and let $P$ be the corresponding transition matrix. Let $a$ be a neighbor of the root and consider the function*

$$h(x) := \mathbb{P}_x[X_t \in T_a \text{ for all but finitely many } t].$$

*Then $h$ is a non-constant, bounded harmonic function on $\mathbb{T}_d$.*

*Proof.* The function $h$ is bounded since it is defined as a probability, and by the usual first-step analysis

$$h(x) = \sum_{y: y\sim x} \frac{1}{d}\mathbb{P}_y[X_t \in T_a \text{ for all but finitely many } t] = \sum_y P(x,y)h(y),$$

so $h$ is harmonic on all of $\mathbb{T}_d$.

Let $b \neq a$ be a neighbor of the root. The key of the proof is the following lemma.

**Lemma 4.1.8.**

$$q := \mathbb{P}_a[\tau_\rho = +\infty] = \mathbb{P}_b[\tau_\rho = +\infty] > 0.$$

*Proof.* The equality of the two probabilities follows by symmetry. To see that $q > 0$, let $(Z_t)$ be simple random walk on $\mathbb{T}_d$ started at $a$ until the walk hits $\rho$ and let $L_t$ be the graph distance between $Z_t$ and the root. Then $(L_t)$ is a biased random walk on $\mathbb{Z}$ started at $1$ jumping to the right with probability $1 - \frac{1}{d}$ and jumping to the left with probability $\frac{1}{d}$. The probability that $(L_t)$ hits $0$ in finite time is $< 1$ because $1 - \frac{1}{d} > \frac{1}{2}$ when $d \geq 3$ by the gambler's ruin (Example 3.1.43). ■

Note that

$$h(\rho) \leq \left(1 - \frac{1}{d}\right)(1-q) < 1.$$

Indeed, if on the first step the random walk started at $\rho$ moves away from $a$, an event of probability $1 - \frac{1}{d}$, then it must come back to $\rho$ in finite time to reach $T_a$. Similarly, by the strong Markov property (Theorem 3.1.8),

$$h(a) = q + (1-q)h(\rho).$$

Since $h(\rho) \neq 1$ and $q > 0$, this shows that $h(a) > h(\rho)$. So $h$ is not constant. ■

### 4.1.3 Total variation distance and coupling inequality

In the examples of Section 4.1.1, we used coupling to prove monotonicity statements. Coupling is also useful to bound the distance between probability measures. For this, we need the coupling inequality.

**Total variation distance** Let $\mu$ and $\nu$ be probability measures on $(S, \mathcal{S})$. Recall the definition of the total variation distance

$$\|\mu - \nu\|_{\mathrm{TV}} := \sup_{A \in \mathcal{S}} |\mu(A) - \nu(A)|.$$

As the next lemma shows in the countable case, the total variation distance can be thought of as an $\ell^1$ distance on probability measures as vectors (up to a constant factor).

**Lemma 4.1.9** (Alternative definition of total variation distance)**.** *If $S$ is countable, then it holds that*

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

*Proof.* Let $E_* := \{x \,:\, \mu(x) \geq \nu(x)\}$. Then, for any $A \subseteq S$, by definition of $E_*$

$$\mu(A) - \nu(A) \leq \mu(A \cap E_*) - \nu(A \cap E_*) \leq \mu(E_*) - \nu(E_*).$$

Similarly, we have

$$\begin{aligned} \nu(A) - \mu(A) &\leq \nu(E_*^c) - \mu(E_*^c) \\ &= (1 - \nu(E_*)) - (1 - \mu(E_*)) \\ &= \mu(E_*) - \nu(E_*). \end{aligned}$$

The two bounds above are equal so $|\mu(A) - \nu(A)| \leq \mu(E_*) - \nu(E_*)$. Equality is achieved when $A = E_*$. Also

$$\begin{aligned} \mu(E_*) - \nu(E_*) &= \frac{1}{2}\left[\mu(E_*) - \nu(E_*) + \nu(E_*^c) - \mu(E_*^c)\right] \\ &= \frac{1}{2} \sum_{x \in S} |\mu(x) - \nu(x)|. \end{aligned}$$

That concludes the proof. ∎

Like the $\ell^1$ distance, the total variation distance is a metric. In particular, it satisfies the triangle inequality.

**Lemma 4.1.10** (Total variation distance: triangle inequality)**.** *Let $\mu, \nu, \eta$ be probability measures on $(S, \mathcal{S})$. Then*

$$\|\mu - \nu\|_{\mathrm{TV}} \leq \|\mu - \eta\|_{\mathrm{TV}} + \|\eta - \nu\|_{\mathrm{TV}}.$$

*Proof.* From the definition,

$$\begin{aligned}\sup_{A \in \mathcal{S}} |\mu(A) - \nu(A)| &\leq \sup_{A \in \mathcal{S}} \{|\mu(A) - \eta(A)| + |\eta(A) - \nu(A)|\} \\ &\leq \sup_{A \in \mathcal{S}} |\mu(A) - \eta(A)| + \sup_{A \in \mathcal{S}} |\eta(A) - \nu(A)|.\end{aligned}$$ ■

**Coupling inequality** We come to an elementary, yet fundamental inequality.

**Lemma 4.1.11** (Coupling inequality)**.** *Let $\mu$ and $\nu$ be probability measures on $(S, \mathcal{S})$. For any coupling $(X, Y)$ of $\mu$ and $\nu$,*

$$\|\mu - \nu\|_{\mathrm{TV}} \leq \mathbb{P}[X \neq Y].$$

*Proof.* For any $A \in \mathcal{S}$,

$$\begin{aligned}\mu(A) - \nu(A) &= \mathbb{P}[X \in A] - \mathbb{P}[Y \in A] \\ &= \mathbb{P}[X \in A,\ X = Y] + \mathbb{P}[X \in A,\ X \neq Y] \\ &\qquad - \mathbb{P}[Y \in A,\ X = Y] - \mathbb{P}[Y \in A,\ X \neq Y] \\ &= \mathbb{P}[X \in A,\ X \neq Y] - \mathbb{P}[Y \in A,\ X \neq Y] \\ &\leq \mathbb{P}[X \neq Y],\end{aligned}$$

and, similarly, $\nu(A) - \mu(A) \leq \mathbb{P}[X \neq Y]$. Hence

$$|\mu(A) - \nu(A)| \leq \mathbb{P}[X \neq Y].$$

Taking a supremum over $A$ gives the claim. ■

Here is a quick example.

**Example 4.1.12** (A coupling of Poisson random variables)**.** Let $X \sim \mathrm{Poi}(\lambda)$ and $Y \sim \mathrm{Poi}(\nu)$ with $\lambda > \nu$. Recall that a sum of independent Poisson is Poisson (see Exercise 6.7). This fact leads to a natural coupling: let $\hat{Y} \sim \mathrm{Poi}(\nu)$, $\hat{Z} \sim \mathrm{Poi}(\lambda - \nu)$ independently of $\hat{Y}$, and $\hat{X} = \hat{Y} + \hat{Z}$. Then $(\hat{X}, \hat{Y})$ is a coupling of $X$ and $Y$, and by the coupling inequality (Lemma 4.1.11)

$$\|\mu_X - \mu_Y\|_{\mathrm{TV}} \leq \mathbb{P}[\hat{X} \neq \hat{Y}] = \mathbb{P}[\hat{Z} > 0] = 1 - e^{-(\lambda - \nu)} \leq \lambda - \nu,$$

where we used $1 - e^{-x} \leq x$ for all $x$ (see Exercise 1.16). ◀

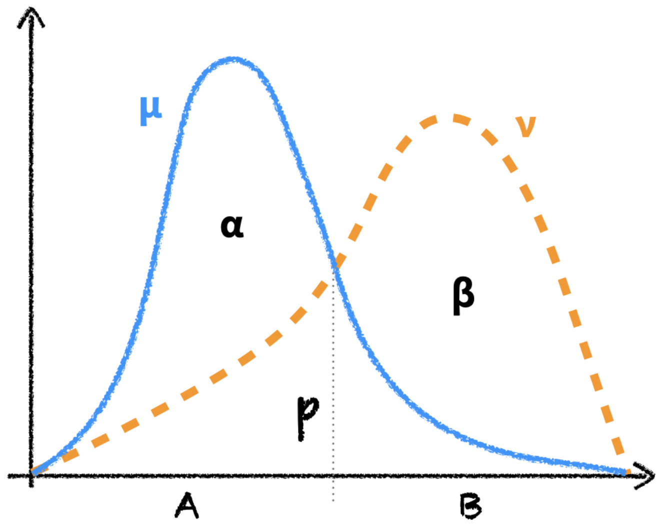


Figure 4.1: Proof by picture that: $1 - p = \alpha = \beta = \|\mu - \nu\|_{\mathrm{TV}}$.

Remarkably, the inequality in Lemma 4.1.11 is tight. For simplicity, we prove this in the finite case only.

**Lemma 4.1.13** (Maximal coupling)**.** *Assume $S$ is finite and let $\mathcal{S} = 2^S$. Let $\mu$ and $\nu$ be probability measures on $(S, \mathcal{S})$. Then,*

$$\|\mu - \nu\|_{\mathrm{TV}} = \inf\{\mathbb{P}[X \neq Y] : \textit{coupling } (X, Y) \textit{ of } \mu \textit{ and } \nu\}.$$

*Proof.* We construct a coupling which achieves equality in the coupling inequality. Such a coupling is called a *maximal coupling*.

*maximal coupling*

Let $A = \{x \in S : \mu(x) > \nu(x)\}$, $B = \{x \in S : \mu(x) \leq \nu(x)\}$ and

$$p := \sum_{x \in S} \mu(x) \wedge \nu(x), \quad \alpha := \sum_{x \in A} [\mu(x) - \nu(x)], \quad \beta := \sum_{x \in B} [\nu(x) - \mu(x)].$$

Assume $p > 0$ (otheriwse there is nothing to prove). First, two lemmas. See Figure 4.1 for a proof by picture.

**Lemma 4.1.14.**

$$\sum_{x \in S} \mu(x) \wedge \nu(x) = 1 - \|\mu - \nu\|_{\mathrm{TV}}.$$

*Proof.* We have

$$\begin{aligned}
2\|\mu-\nu\|_{\mathrm{TV}} &= \sum_{x\in S}|\mu(x)-\nu(x)| \\
&= \sum_{x\in A}[\mu(x)-\nu(x)] + \sum_{x\in B}[\nu(x)-\mu(x)] \\
&= \sum_{x\in A}\mu(x) + \sum_{x\in B}\nu(x) - \sum_{x\in S}\mu(x)\wedge\nu(x) \\
&= 2 - \sum_{x\in B}\mu(x) - \sum_{x\in A}\nu(x) - \sum_{x\in S}\mu(x)\wedge\nu(x) \\
&= 2 - 2\sum_{x\in S}\mu(x)\wedge\nu(x),
\end{aligned}$$

where we used that both $\mu$ and $\nu$ sum to 1. Rearranging gives the claim. ■

**Lemma 4.1.15.**

$$\sum_{x\in A}[\mu(x)-\nu(x)] = \sum_{x\in B}[\nu(x)-\mu(x)] = \|\mu-\nu\|_{\mathrm{TV}} = 1-p.$$

*Proof.* The first equality is immediate by the fact that $\mu$ and $\nu$ are probability measures. The second equality follows from the first one together with the second line in the proof of the previous lemma. The last equality is a restatement of the last lemma. ■

The maximal coupling is defined as follows:

- With probability $p$, pick $X = Y$ from $\gamma_{\min}$ where

$$\gamma_{\min}(x) := \frac{1}{p}\mu(x)\wedge\nu(x), \quad x \in S.$$

- Otherwise, pick $X$ from $\gamma_A$ where

$$\gamma_A(x) := \frac{\mu(x)-\nu(x)}{1-p}, \quad x \in A,$$

  and, independently, pick $Y$ from

$$\gamma_B(x) := \frac{\nu(x)-\mu(x)}{1-p}, \quad x \in B.$$

  Note that $X \neq Y$ in that case because $A$ and $B$ are disjoint.

The marginal law of $X$ is: for $x \in A$,

$$p\,\gamma_{\min}(x) + (1-p)\,\gamma_A(x) = \nu(x) + \mu(x) - \nu(x) = \mu(x),$$

and for $x \in B$,

$$p\,\gamma_{\min}(x) + (1-p)\,\gamma_A(x) = \mu(x) + 0 = \mu(x).$$

A similar calculation holds for $Y$. Finally $\mathbb{P}[X \neq Y] = 1 - p = \|\mu - \nu\|_{\mathrm{TV}}$. ■

**Remark 4.1.16.** *A proof of this result for general Polish spaces can be found in [dH, Section 2.5].*

We return to our coupling of Bernoulli variables.

**Example 4.1.17** (Coupling of Bernoulli variables (continued))**.** Recall the setting of Example 4.1.2. To construct the maximal coupling as above, we note that

$$A := \{0\}, \qquad B := \{1\},$$

$$p := \sum_x \mu(x) \wedge \nu(x) = (1-r) + q, \qquad 1 - p = \alpha = \beta := r - q,$$

$$(\gamma_{\min}(x))_{x=0,1} = \left(\frac{1-r}{(1-r)+q}, \frac{q}{(1-r)+q}\right),$$

$$\gamma_A(0) := 1, \qquad \gamma_B(1) := 1.$$

The law of the maximal coupling $(X''', Y''')$ is given by

$$\begin{aligned}\Big(\mathbb{P}[(X''', Y''') = (i,j)]\Big)_{i,j\in\{0,1\}} &= \begin{pmatrix} p\,\gamma_{\min}(0) & (1-p)\,\gamma_A(0)\gamma_B(1) \\ 0 & p\,\gamma_{\min}(1)\end{pmatrix} \\ &= \begin{pmatrix} 1-r & r-q \\ 0 & q\end{pmatrix}.\end{aligned}$$

Notice that it happens to coincide with the monotone coupling. ◀

**Poisson approximation** Here is a classical application of coupling: the approximation of a sum of independent Bernoulli variables with a Poisson. It gives a quantitative bound in total variation distance. Let $X_1, \ldots, X_n$ be independent Bernoulli random variables with parameters $p_1, \ldots, p_n$ respectively. We are interested in the case where the $p_i$s are "small." Let $S_n := \sum_{i \le n} X_i$. We approximate $S_n$ with a Poisson random variable $Z_n$ as follows: let $W_1, \ldots, W_n$ be independent Poisson random variables with means $\lambda_1, \ldots, \lambda_n$ respectively and define $Z_n := \sum_{i\le n} W_i$. We choose $\lambda_i = -\log(1 - p_i)$ for reasons that will become clear below. Note that $Z_n \sim \mathrm{Poi}(\lambda)$ where $\lambda = \sum_{i \le n} \lambda_i$.

**Theorem 4.1.18** (Poisson approximation)**.**

$$\|\mu_{S_n} - \mathrm{Poi}(\lambda)\|_{\mathrm{TV}} \leq \frac{1}{2}\sum_{i\leq n} \lambda_i^2.$$

*Proof.* We couple the pairs $X_i, W_i$ independently for $i \leq n$. Let

$$W_i' \sim \mathrm{Poi}(\lambda_i) \quad \text{and} \quad X_i' = W_i' \wedge 1.$$

Because of our choice $\lambda_i = -\log(1-p_i)$ which implies

$$1 - p_i = \mathbb{P}[X_i = 0] = \mathbb{P}[W_i = 0] = e^{-\lambda_i},$$

$(X_i', W_i')$ is indeed a coupling of $X_i, W_i$. Let $S_n' := \sum_{i\leq n} X_i'$ and $Z_n' := \sum_{i\leq n} W_i'$. Then $(S_n', Z_n')$ is a coupling of $S_n, Z_n$. By the coupling inequality

$$\begin{aligned}
\|\mu_{S_n} - \mu_{Z_n}\|_{\mathrm{TV}} &\leq \mathbb{P}[S_n' \neq Z_n'] \\
&\leq \sum_{i\leq n} \mathbb{P}[X_i' \neq W_i'] \\
&= \sum_{i\leq n} \mathbb{P}[W_i' \geq 2] \\
&= \sum_{i\leq n}\sum_{j\geq 2} e^{-\lambda_i}\frac{\lambda_i^j}{j!} \\
&\leq \sum_{i\leq n}\frac{\lambda_i^2}{2}\sum_{\ell\geq 0} e^{-\lambda_i}\frac{\lambda_i^\ell}{\ell!} \\
&= \sum_{i\leq n}\frac{\lambda_i^2}{2}.
\end{aligned}$$

■

**Mappings reduce the total variation distance** The following lemma will be useful.

**Lemma 4.1.19** (Mappings)**.** *Let $X$ and $Y$ be random variables taking values in $(S, \mathcal{S})$, let $h$ be a measurable map from $(S, \mathcal{S})$ to $(S', \mathcal{S}')$, and let $X' := h(X)$ and $Y' := h(Y)$. The following inequality holds*

$$\|\mu_{X'} - \mu_{Y'}\|_{\mathrm{TV}} \leq \|\mu_X - \mu_Y\|_{\mathrm{TV}}.$$

*Proof.* From the definition of the total variation distance, we seek to bound

$$\begin{aligned}&\sup_{A'\in\mathcal{S}'}\left|\mathbb{P}[X'\in A']-\mathbb{P}[Y'\in A']\right|\\&=\sup_{A'\in\mathcal{S}'}\left|\mathbb{P}[h(X)\in A']-\mathbb{P}[h(Y)\in A']\right|\\&=\sup_{A'\in\mathcal{S}'}\left|\mathbb{P}[X\in h^{-1}(A')]-\mathbb{P}[Y\in h^{-1}(A')]\right|.\end{aligned}$$

Since $h^{-1}(A')\in\mathcal{S}$ by the measurability of $h$, this last expression is less or equal than

$$\sup_{A\in\mathcal{S}}\left|\mathbb{P}[X\in A]-\mathbb{P}[Y\in A]\right|,$$

which proves the claim. ∎

**Coupling of Markov chains** In the context of Markov chains, a natural way to couple is to do so step by step. We will refer to such couplings as Markovian. An important special case is a Markovian coupling of a chain with itself.

**Definition 4.1.20** (Markovian coupling)**.** *Let $P$ and $Q$ be transition matrices on the same state space $V$. A* Markovian coupling *of $P$ and $Q$ is a Markov chain $(X_t,Y_t)_t$ on $V\times V$ with transition matrix $R$ satisfying: for all $x,y,x',y'\in V$,*

*Markovian coupling*

$$\sum_{z'}R((x,y),(x',z'))=P(x,x'),$$

$$\sum_{z'}R((x,y),(z',y'))=Q(y,y').$$

We will give many examples throughout this chapter. See also Example 4.2.14 for an example of a coupling of Markov chains that is *not* Markovian.

### 4.1.4 ▷ *Random graphs: degree sequence in Erdős-Rényi model*

Let $G_n\sim\mathbb{G}_{n,p_n}$ be an Erdős-Rényi graph with $p_n:=\frac{\lambda}{n}$ and $\lambda>0$ (see Definition 1.2.2). For $i\in[n]$, let $D_i(n)$ be the degree of vertex $i$ and define

$$N_d(n):=\sum_{i=1}^{n}\mathbf{1}_{\{D_i(n)=d\}},$$

the number of vertices of degree $d$.

**Theorem 4.1.21** (Erdős-Rényi graph: degree sequence)**.**

$$\frac{1}{n}N_d(n) \to_{\mathrm{p}} f_d := e^{-\lambda}\frac{\lambda^d}{d!}, \qquad \forall d \geq 0.$$

*Proof.* We proceed in two steps:

1. we use the coupling inequality (Lemma 4.1.11) to show that the expectation of $\frac{1}{n}N_d(n)$ is close to $f_d$; and
2. we appeal to Chebyshev's inequality (Theorem 2.1.2) to show that $\frac{1}{n}N_d(n)$ is close to its expectation.

We justify each step as a lemma.

**Lemma 4.1.22** (Convergence of the mean)**.**

$$\lim_{n\to+\infty}\frac{1}{n}\mathbb{E}_{n,p_n}\left[N_d(n)\right] = f_d, \qquad \forall d \geq 1.$$

*Proof.* Note that the degrees $D_i(n)$, $i \in [n]$, are identically distributed (but not independent) so

$$\frac{1}{n}\mathbb{E}_{n,p_n}\left[N_d(n)\right] = \mathbb{P}_{n,p_n}[D_1(n) = d].$$

Moreover, by definition, $D_1(n) \sim \mathrm{Bin}(n-1, p_n)$. Let $S_n \sim \mathrm{Bin}(n, p_n)$ and $Z_n \sim \mathrm{Poi}(\lambda)$. Using the Poisson approximation (Theorem 4.1.18) and a Taylor expansion,

$$\begin{aligned}
\|\mu_{S_n} - \mu_{Z_n}\|_{\mathrm{TV}} &\leq \frac{1}{2}\sum_{i\leq n}(-\log(1-p_n))^2 \\
&= \frac{1}{2}\sum_{i\leq n}\left(\frac{\lambda}{n} + O(n^{-2})\right)^2 \\
&= \frac{\lambda^2}{2n} + O(n^{-2}).
\end{aligned}$$

We can further couple $D_1(n)$ and $S_n$ as

$$\left(\sum_{i\leq n-1} X_i, \sum_{i\leq n} X_i\right),$$

where the $X_i$s are i.i.d. $\mathrm{Ber}(p_n)$, that is, Bernoulli with parameter $p_n$. By the coupling inequality (Theorem 4.1.11),

$$\|\mu_{D_1(n)} - \mu_{S_n}\|_{\mathrm{TV}} \leq \mathbb{P}\left[\sum_{i\leq n-1} X_i \neq \sum_{i\leq n} X_i\right] = \mathbb{P}[X_n = 1] = p_n = \frac{\lambda}{n}.$$

By the triangle inequality for the total variation distance (Lemma 4.1.10) and the bounds above,

$$\begin{aligned}\frac{1}{2}\sum_{d\geq 0}|\mathbb{P}_{n,p_n}[D_1(n)=d]-f_d| &= \|\mu_{D_1(n)}-\mu_{Z_n}\|_{\mathrm{TV}}\\ &\leq \|\mu_{D_1(n)}-\mu_{S_n}\|_{\mathrm{TV}}+\|\mu_{S_n}-\mu_{Z_n}\|_{\mathrm{TV}}\\ &\leq \frac{\lambda+\lambda^2/2}{n}+O(n^{-2}).\end{aligned}$$

Therefore, for all $d$,

$$\left|\frac{1}{n}\mathbb{E}_{n,p_n}\left[N_d(n)\right]-f_d\right| \leq \frac{2\lambda+\lambda^2}{n}+O(n^{-2})\to 0,$$

as $n\to+\infty$. ∎

**Lemma 4.1.23** (Concentration around the mean)**.**

$$\mathbb{P}_{n,p_n}\left[\left|\frac{1}{n}N_d(n)-\frac{1}{n}\mathbb{E}_{n,p_n}\left[N_d(n)\right]\right|\geq\varepsilon\right]\leq\frac{2\lambda+1}{\varepsilon^2 n},\qquad \forall d\geq 1,\forall n.$$

*Proof.* By Chebyshev's inequality, for all $\varepsilon>0$

$$\mathbb{P}_{n,p_n}\left[\left|\frac{1}{n}N_d(n)-\frac{1}{n}\mathbb{E}_{n,p_n}\left[N_d(n)\right]\right|\geq\varepsilon\right]\leq\frac{\mathrm{Var}_{n,p_n}[\frac{1}{n}N_d(n)]}{\varepsilon^2}. \tag{4.1.1}$$

To compute the variance, we note that

$$\begin{aligned}&\mathrm{Var}_{n,p_n}\left[\frac{1}{n}N_d(n)\right]\\ &\quad=\frac{1}{n^2}\left\{\mathbb{E}_{n,p_n}\left[\left(\sum_{i\leq n}\mathbf{1}_{\{D_i(n)=d\}}\right)^2\right]-(n\,\mathbb{P}_{n,p_n}[D_1(n)=d])^2\right\}\\ &\quad=\frac{1}{n^2}\Big\{n(n-1)\mathbb{P}_{n,p_n}[D_1(n)=d,D_2(n)=d]\\ &\qquad\qquad+n\,\mathbb{P}_{n,p_n}[D_1(n)=d]-n^2\mathbb{P}_{n,p_n}[D_1(n)=d]^2\Big\}\\ &\quad\leq\frac{1}{n}+\Big\{\mathbb{P}_{n,p_n}[D_1(n)=d,D_2(n)=d]-\mathbb{P}_{n,p_n}[D_1(n)=d]^2\Big\},\end{aligned} \tag{4.1.2}$$

where we used the crude bound $\mathbb{P}_{n,p_n}[D_1(n)=d]\leq 1$. We bound the last line using a neat coupling argument. Let $Y_1$ and $Y_2$ be independent $\mathrm{Bin}(n-2,p_n)$, and let $X_1$ and $X_2$ be independent $\mathrm{Ber}(p_n)$. By separating the contribution of the edge between 1 and 2 from those of edges to other vertices, we see that the joint degrees $(D_1(n), D_2(n))$ have the same distribution as $(X_1 + Y_1, X_1 + Y_2)$. So the term in curly bracket above is equal to

$$\begin{aligned}
&\mathbb{P}[(X_1 + Y_1, X_1 + Y_2) = (d, d)] - \mathbb{P}[X_1 + Y_1 = d]^2 \\
&\quad = \mathbb{P}[(X_1 + Y_1, X_1 + Y_2) = (d, d)] - \mathbb{P}[(X_1 + Y_1, X_2 + Y_2) = (d, d)] \\
&\quad \leq \mathbb{P}[(X_1 + Y_1, X_1 + Y_2) = (d, d),\ (X_1 + Y_1, X_2 + Y_2) \neq (d, d)] \\
&\quad = \mathbb{P}[(X_1 + Y_1, X_1 + Y_2) = (d, d),\ X_2 + Y_2 \neq d] \\
&\quad = \mathbb{P}[X_1 = 0,\ Y_1 = Y_2 = d,\ X_2 = 1] \\
&\qquad + \mathbb{P}[X_1 = 1,\ Y_1 = Y_2 = d - 1,\ X_2 = 0] \\
&\quad \leq \mathbb{P}[X_2 = 1] + \mathbb{P}[X_1 = 1] \\
&\quad = \frac{2\lambda}{n}.
\end{aligned}$$

Plugging back into (4.1.2) we get $\mathrm{Var}_{n,p_n}\left[\frac{1}{n} N_d(n)\right] \leq \frac{2\lambda+1}{n}$, and (4.1.1) gives the claim. ■

Combining the lemmas concludes the proof of Theorem 4.1.21. ■

## 4.2 Stochastic domination

In comparing two probability measures, a natural relationship is that of "domination." For instance, let $(X_i)_{i=1}^n$ be independent $\mathbb{Z}_+$-valued random variables with

$$\mathbb{P}[X_i \geq 1] \geq p,$$

and let $S = \sum_{i=1}^n X_i$ be their sum. Now consider a separate random variable

$$S_* \sim \mathrm{Bin}(n, p).$$

It is intuitively clear that one should be able to bound $S$ from below by analyzing $S_*$ instead—which may be considerably easier. Indeed, in some sense, $S$ "dominates" $S_*$, that is, $S$ should have a tendency to be bigger than $S_*$. One expects more specifically that

$$\mathbb{P}[S > x] \geq \mathbb{P}[S_* > x].$$

Coupling provides a formal characterization of this notion, as we detail in this section.

In particular we study an important special case known as positive associations. Here a measure "dominates itself" in the following sense: conditioning on certain events makes other events more likely. That concept is formalized in Section 4.2.3.

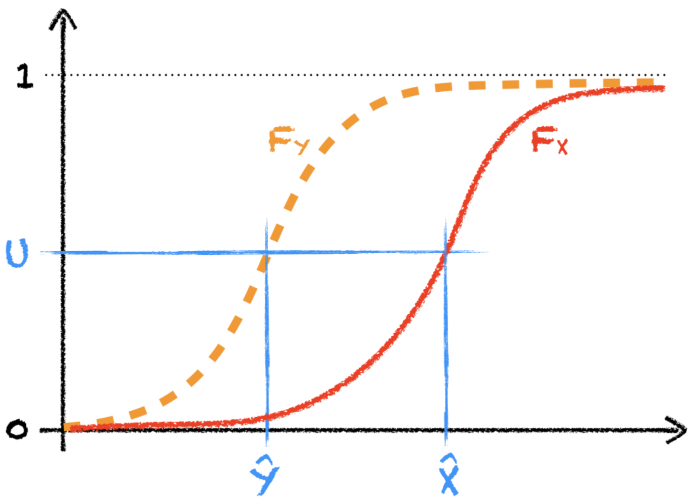


Figure 4.2: The law of $X$, represented here by its cumulative distribution function $F_X$ in solid, stochastically dominates the law of $Y$, in dashed. The construction of a monotone coupling, $(\hat{X}, \hat{Y}) := (F_X^{-1}(U), F_Y^{-1}(U))$ where $U$ is uniform in $[0,1]$, is also depicted.

### 4.2.1 Definitions

We start with the simpler case of real random variables then consider partially ordered sets, a natural setting for this concept.

**Ordering of real random variables** Recall that, intuitively, stochastic domination captures the idea that one variable "tends to take larger values" than the other. For real random variables, it is defined in terms of tail probabilities, or equivalently in terms of cumulative distribution functions. See Figure 4.2 for an illustration.

**Definition 4.2.1** (Stochastic domination)**.** *Let $\mu$ and $\nu$ be probability measures on $\mathbb{R}$. The measure $\mu$ is said to* stochastically dominate $\nu$*, denoted by $\mu \succeq \nu$, if for all $x \in \mathbb{R}$* *stochastic domination*

$$\mu\big[(x, +\infty)\big] \geq \nu\big[(x, +\infty)\big].$$

*A real random variable $X$* stochastically dominates $Y$*, denoted by $X \succeq Y$, if the law of $X$ dominates the law of $Y$.*

**Example 4.2.2** (Bernoulli vs. Poisson)**.** Let $X \sim \mathrm{Poi}(\lambda)$ be Poisson with mean $\lambda > 0$ and let $Y$ be a Bernoulli trial with success probability $p \in (0,1)$. In order for $X$ to stochastically dominate $Y$, we need to have

$$\mathbb{P}[X > \ell] \geq \mathbb{P}[Y > \ell], \qquad \forall \ell \geq 0.$$

This is always true for $\ell \geq 1$ since $\mathbb{P}[X > \ell] > 0$ but $\mathbb{P}[Y > \ell] = 0$. So it remains to consider the case $\ell = 0$. We have

$$1 - e^{-\lambda} = \mathbb{P}[X > 0] \geq \mathbb{P}[Y > 0] = p,$$

if and only if

$$\lambda \geq -\log(1-p).$$

◀

Note that stochastic domination does not require $X$ and $Y$ to be defined on the same probability space. However the connection to coupling arises from the following characterization.

**Theorem 4.2.3** (Coupling and stochastic domination)**.** *The real random variable $X$ stochastically dominates $Y$ if and only if there is a coupling $(\hat{X}, \hat{Y})$ of $X$ and $Y$ such that*

$$\mathbb{P}[\hat{X} \geq \hat{Y}] = 1. \tag{4.2.1}$$

*We refer to $(\hat{X}, \hat{Y})$ as a* monotone coupling *of $X$ and $Y$.*

*monotone coupling*

*Proof.* Suppose there is such a coupling. Then for all $x \in \mathbb{R}$

$$\mathbb{P}[Y > x] = \mathbb{P}[\hat{Y} > x] = \mathbb{P}[\hat{X} \geq \hat{Y} > x] \leq \mathbb{P}[\hat{X} > x] = \mathbb{P}[X > x].$$

For the other direction, define the cumulative distribution functions $F_X(x) = \mathbb{P}[X \leq x]$ and $F_Y(x) = \mathbb{P}[Y \leq x]$. Assume $X \succeq Y$. The idea of the proof is to use the following standard way of generating a real random variable (see Theorem B.2.7)

$$X \stackrel{\mathrm{d}}{=} F_X^{-1}(U), \tag{4.2.2}$$

where $U$ is a $[0,1]$-valued uniform random variable and

$$F_X^{-1}(u) := \inf\{x \in \mathbb{R} \,:\, F_X(x) \geq u\},$$

is a generalized inverse. It is natural to construct a coupling of $X$ and $Y$ by simply using the same uniform random variable $U$ in this representation, that is, we define $\hat{X} = F_X^{-1}(U)$ and $\hat{Y} = F_Y^{-1}(U)$. See Figure 4.2. By (4.2.2), this is a coupling of $X$ and $Y$. It remains to check (4.2.1). Because $F_X(x) \leq F_Y(x)$ for all $x$ by definition of stochastic domination, by the definition of the generalized inverse,

$$\mathbb{P}[\hat{X} \geq \hat{Y}] = \mathbb{P}[F_X^{-1}(U) \geq F_Y^{-1}(U)] = 1,$$

as required. ∎

**Example 4.2.4.** Returning to the example in the first paragraph of Section 4.2, let $(X_i)_{i=1}^n$ be independent $\mathbb{Z}_+$-valued random variables with $\mathbb{P}[X_i \geq 1] \geq p$ and consider their sum $S := \sum_{i=1}^n X_i$. Further let $S_* \sim \mathrm{Bin}(n, p)$. Write $S_*$ as the sum $\sum_{i=1}^n Y_i$ where $(Y_i)$ are independent Bernoullli variables with $\mathbb{P}[Y_i = 1] = p$. To couple $S$ and $S_*$, first set $(\hat{Y}_i) := (Y_i)$ and $\hat{S}_* := \sum_{i=1}^n \hat{Y}_i$. Let $\hat{X}_i$ be 0 whenever $\hat{Y}_i = 0$. Otherwise (i.e., if $\hat{Y}_i = 1$), generate $\hat{X}_i$ according to the distribution of $X_i$ conditioned on $\{X_i \geq 1\}$, independently of everything else. By construction $\hat{X}_i \geq \hat{Y}_i$ almost surely for all $i$ and as a result $\sum_{i=1}^n \hat{X}_i =: \hat{S} \geq \hat{S}_*$ almost surely, or $S \succeq S_*$ by Theorem 4.2.3. That implies for instance that $\mathbb{P}[S > x] \geq \mathbb{P}[S_* > x]$ as we claimed earlier. A slight modification of this argument gives the following useful fact about binomials

$$n \geq m,\ q \geq p \implies \mathrm{Bin}(n, q) \succeq \mathrm{Bin}(m, p).$$

Exercise 4.2 asks for a formal proof. ◀

**Example 4.2.5** (Poisson distribution)**.** Let $X \sim \mathrm{Poi}(\mu)$ and $Y \sim \mathrm{Poi}(\nu)$ with $\mu > \nu$. Recall that a sum of independent Poisson is Poisson (see Exercise 6.7). This fact leads to a natural coupling: let $\hat{Y} \sim \mathrm{Poi}(\nu)$, $\hat{Z} \sim \mathrm{Poi}(\mu - \nu)$ independently of $Y$, and $\hat{X} = \hat{Y} + \hat{Z}$. Then $(\hat{X}, \hat{Y})$ is a coupling and $\hat{X} \geq \hat{Y}$ a.s. because $\hat{Z} \geq 0$. Hence $X \succeq Y$. ◀

We record two useful consequences of Theorem 4.2.3.

**Corollary 4.2.6.** *Let $X$ and $Y$ be real random variables with $X \succeq Y$ and let $f : \mathbb{R} \to \mathbb{R}$ be a non-decreasing function. Then $f(X) \succeq f(Y)$ and furthermore, provided $\mathbb{E}|f(X)|, \mathbb{E}|f(Y)| < +\infty$, we have that*

$$\mathbb{E}[f(X)] \geq \mathbb{E}[f(Y)].$$

*Proof.* Let $(\hat{X}, \hat{Y})$ be the monotone coupling of $X$ and $Y$ whose existence is guaranteed by Theorem 4.2.3. Then $f(\hat{X}) \geq f(\hat{Y})$ almost surely so that, provided the expectations exist,

$$\mathbb{E}[f(X)] = \mathbb{E}[f(\hat{X})] \geq \mathbb{E}[f(\hat{Y})] = \mathbb{E}[f(Y)],$$

and furthermore $(f(\hat{X}), f(\hat{Y}))$ is a monotone coupling of $f(X)$ and $f(Y)$. Hence $f(X) \succeq f(Y)$. ∎

**Corollary 4.2.7.** *Let $X_1$, $X_2$ be independent random variables. Let $Y_1, Y_2$ be independent random variables such that $X_i \succeq Y_i$, $i = 1, 2$. Then*

$$X_1 + X_2 \succeq Y_1 + Y_2.$$

*Proof.* Let $(\hat{X}_1, \hat{Y}_1)$ and $(\hat{X}_2, \hat{Y}_2)$ be independent, monotone couplings of $X_1, Y_1$ and $X_2, Y_2$ on the same probability space. Then

$$X_1 + X_2 \stackrel{\mathrm{d}}{=} \hat{X}_1 + \hat{X}_2 \geq \hat{Y}_1 + \hat{Y}_2 \stackrel{\mathrm{d}}{=} Y_1 + Y_2.$$

■

**Example 4.2.8** (Binomial vs. Poisson)**.** A sum of $n$ independent Poisson variables with mean $\lambda$ is $\mathrm{Poi}(n\lambda)$. A sum of $n$ independent Bernoulli trials with success probability $p$ is $\mathrm{Bin}(n, p)$. Using Example 4.2.2 and Corollary 4.2.7, we get

$$\lambda \geq -\log(1-p) \quad \Longrightarrow \quad \mathrm{Poi}(n\lambda) \succeq \mathrm{Bin}(n, p). \tag{4.2.3}$$

The following special case will be useful later. Let $0 < \Lambda < 1$ and let $m$ be a positive integer. Then

$$\frac{\Lambda}{m-1} \geq \frac{\Lambda}{m-\Lambda} = \frac{m}{m-\Lambda} - 1 \geq \log\left(\frac{m}{m-\Lambda}\right) = -\log\left(1 - \frac{\Lambda}{m}\right),$$

where we used that $\log x \leq x - 1$ for all $x \in \mathbb{R}_+$ (see Exercise 1.16). So, setting $\lambda := \frac{\Lambda}{m-1}$, $p := \frac{\Lambda}{m}$ and $n := m - 1$ in (4.2.3), we get

$$\Lambda \in (0,1) \quad \Longrightarrow \quad \mathrm{Poi}(\Lambda) \succeq \mathrm{Bin}\left(m-1, \frac{\Lambda}{m}\right). \tag{4.2.4}$$

◀

**Ordering on partially ordered sets** The definition of stochastic domination hinges on the totally ordered nature of $\mathbb{R}$. It also extends naturally to posets. Let $(\mathcal{X}, \leq)$ be a *poset*, that is, for all $x, y, z \in \mathcal{X}$:

*poset*

- *(Reflexivity)* $x \leq x$;
- *(Antisymmetry)* if $x \leq y$ and $y \leq x$ then $x = y$; and
- *(Transitivity)* if $x \leq y$ and $y \leq z$ then $x \leq z$.

Throughout, we assume that $\mathcal{X}$ is a measurable space.

For instance the set $\{0,1\}^F$ is a poset when equipped with the relation $\mathbf{x} \leq \mathbf{y}$ if and only if $x_i \leq y_i$ for all $i \in F$, where $\mathbf{x} = (x_i)_{i \in F}$ and $\mathbf{y} = (y_i)_{i \in F}$. Equivalently the subsets of $F$, denoted by $2^F$, form a poset with the inclusion relation.

A *totally ordered* set satisfies in addition that, for any $x, y$, we have either $x \leq y$ or $y \leq x$. That is not satisfied in the previous example.

Let $\mathcal{F}$ be a $\sigma$-algebra over the poset $\mathcal{X}$. An event $A \in \mathcal{F}$ is *increasing* if $x \in A$ implies that any $y \geq x$ is also in $A$. A function $f : \mathcal{X} \to \mathbb{R}$ is *increasing* if $x \leq y$ implies $f(x) \leq f(y)$. Some properties of increasing events are derived in Exercise 4.4. *increasing*

**Definition 4.2.9** (Stochastic domination for posets)**.** *Let $(\mathcal{X}, \leq)$ be a poset and let $\mathcal{F}$ be a $\sigma$-algebra on $\mathcal{X}$. Let $\mu$ and $\nu$ be probability measures on $(\mathcal{X}, \mathcal{F})$. The measure $\mu$ is said to* stochastically dominate *$\nu$, denoted by $\mu \succeq \nu$, if for all increasing $A \in \mathcal{F}$*

$$\mu(A) \geq \nu(A).$$

*An $\mathcal{X}$-valued random variable $X$* stochastically dominates *$Y$, denoted by $X \succeq Y$, if the law of $X$ dominates the law of $Y$.*

As before, a *monotone coupling* $(\hat{X}, \hat{Y})$ of $X$ and $Y$ is one which satisfies $\hat{X} \geq \hat{Y}$ almost surely.

**Example 4.2.10** (Monotonicity of the percolation function)**.** We briefly revisit Example 4.1.3 to illustrate our definitions. Consider bond percolation on the $d$-dimensional lattice $\mathbb{L}^d$ (Definition 1.2.1). Here the poset is the collection of all subsets of edges, specifying the open edges, with the inclusion relation. Recall that the percolation function is given by

$$\theta(p) := \mathbb{P}_p[|\mathcal{C}_0| = +\infty],$$

where $\mathcal{C}_0$ is the open cluster of the origin. We argued in Example 4.1.3 (see also Section 2.2.4) that $\theta(p)$ is nondecreasing by considering the following alternative representation of the percolation process under $\mathbb{P}_p$: to each edge $e$, assign a uniform $[0, 1]$-valued random variable $U_e$ and declare the edge open if $U_e \leq p$. Using the same $U_e$s for two different values of $p$, say $p_1 < p_2$, gives a monotone coupling of the processes for $p_1$ and $p_2$. ◀

The existence of a monotone coupling is perhaps more surprising for posets. We prove the result in the finite case only, which will be enough for our purposes.

**Theorem 4.2.11** (Strassen's theorem)**.** *Let $X$ and $Y$ be random variables taking values in a* finite *poset $(\mathcal{X}, \leq)$ with the $\sigma$-algebra $\mathcal{F} = 2^{\mathcal{X}}$. Then $X \succeq Y$ if and only if there exists a monotone coupling $(\hat{X}, \hat{Y})$ of $X$ and $Y$.*

*Proof.* Suppose there is such a coupling. Then for all increasing $A$

$$\mathbb{P}[Y \in A] = \mathbb{P}[\hat{Y} \in A] = \mathbb{P}[\hat{X} \geq \hat{Y} \in A] \leq \mathbb{P}[\hat{X} \in A] = \mathbb{P}[X \in A].$$

The proof in the other direction relies on the max-flow min-cut theorem (Theorem 1.1.15). To see the connection with flows, let $\mu_X$ and $\mu_Y$ be the laws of $X$ and

$Y$ respectively, and denote by $\nu$ their joint distribution under the desired coupling. Noting that we want $\nu(x, y) > 0$ only if $x \geq y$, the marginal conditions on the coupling read

$$\sum_{y:x\geq y} \nu(x, y) = \mu_X(x), \qquad \forall x \in \mathcal{X}, \tag{4.2.5}$$

and

$$\sum_{x:x\geq y} \nu(x, y) = \mu_Y(y), \qquad \forall y \in \mathcal{X}. \tag{4.2.6}$$

These equations can be interpreted as flow-conservation constraints. Consider the following directed graph. There are two vertices, $(w, 1)$ and $(w, 2)$, for each element $w$ in $\mathcal{X}$ with edges connecting each $(x, 1)$ to those $(y, 2)$s with $x \geq y$. These edges have capacity $+\infty$. In addition there is a source $a$ and a sink $z$. The source has a directed edge of capacity $\mu_X(x)$ to $(x, 1)$ for each $x \in \mathcal{X}$ and, similarly, each $(y, 2)$ has a directed edge of capacity $\mu_Y(y)$ to the sink. The existence of a monotone coupling will follow once we show that there is a flow of strength $1$ between $a$ and $z$. Indeed, in that case, all edges from the source and all edges to the sink must be at capacity. If we let $\nu(x, y)$ be the flow on edge $\langle (x, 1), (y, 2)\rangle$, the systems in (4.2.5) and (4.2.6) encode conservation of flow on the vertices $(\mathcal{X} \times \{1\}) \cup (\mathcal{X} \times \{2\})$. Hence the flow between $\mathcal{X} \times \{1\}$ and $\mathcal{X} \times \{2\}$ yields the desired coupling. See Figure 4.3.

By the max-flow min-cut theorem (Theorem 1.1.15), it suffices to show that a minimum cut has capacity $1$. Such a cut is of course obtained by choosing all edges out of the source. So it remains to show that no cut has capacity less than $1$. This is where we use the fact that $\mu_X(A) \geq \mu_Y(A)$ for all increasing $A$. Because the edges between $\mathcal{X} \times \{1\}$ and $\mathcal{X} \times \{2\}$ have infinite capacity, they cannot be used in a minimum cut. So we can restrict our attention to those cuts containing edges from $a$ to $A_* \times \{1\}$ and from $Z_* \times \{2\}$ to $z$ for subsets $A_*, Z_* \subseteq \mathcal{X}$.

We must have

$$A_* \supseteq \{x \in \mathcal{X} \,:\, \exists y \in Z_*^c,\ x \geq y\},$$

to block all paths of the form $a \sim (x, 1) \sim (y, 2) \sim z$ with $x$ and $y$ as in the previous display; here $Z_*^c = \mathcal{X} \setminus Z_*$. In fact, for a minimum cut, we further have

$$A_* = \{x \in \mathcal{X} \,:\, \exists y \in Z_*^c,\ x \geq y\},$$

as adding an $x$ not satisfying this property is redundant. In particular $A_*$ is increasing: if $x_1 \in A_*$ and $x_2 \geq x_1$, then $\exists y \in Z_*^c$ such that $x_1 \geq y$ and, since $x_2 \geq x_1 \geq y$, we also have $x_2 \in A_*$.

Observe further that, because $y \geq y$, the set $A_*$ also includes $Z_*^c$. If it were the case that $A_* \neq Z_*^c$, then we could construct a cut with lower or equal capacity

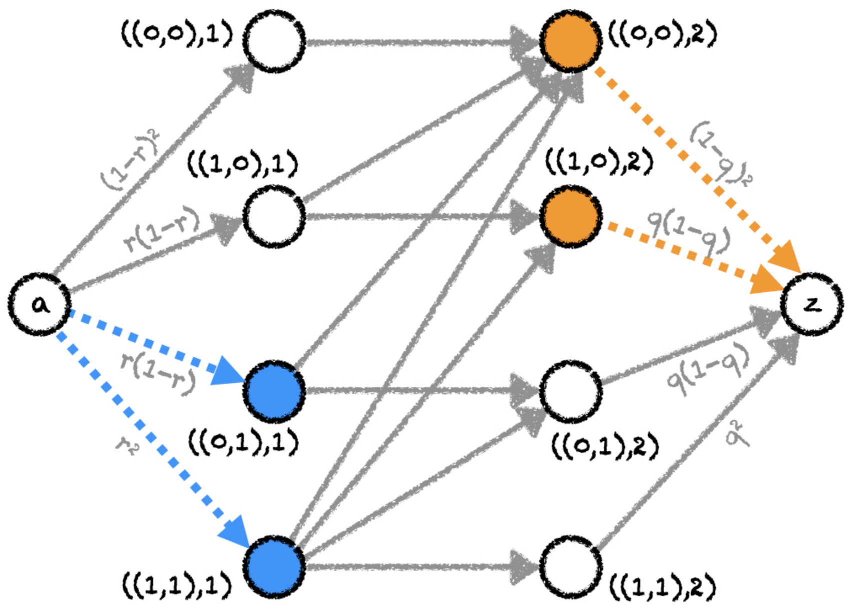


Figure 4.3: Construction of a monotone coupling through the max-flow representation for independent Bernoulli pairs with parameters $r$ (on the left) and $q < r$ (on the right). Edge labels indicate capacity. Edges without labels have infinite capacity. The dotted edges depict a suboptimal cut. The dark vertices correspond to the sets $A_*$ and $Z_*$ for this cut. The capacity of the cut is $r^2 + r(1-r) + (1-q)^2 + (1-q)q = r + (1-q) > r + (1-r) = 1$.

by fixing $A_*$ and setting $Z_* := A_*^c$: suppose $A_* \cap Z_*$ is nonempty; because $A_*$ is increasing, any $y \in A_* \cap Z_*$ is such that paths of the form $a \sim (x, 1) \sim (y, 2) \sim z$ with $x \geq y$ are cut by $x \in A_*$; so we do not need those $y$s in $Z_*$. Hence, for a minimum cut, we can assume that in fact $A_* = Z_*^c$. The capacity of the cut is

$$\mu_X(A_*) + \mu_Y(Z_*) = \mu_X(A_*) + 1 - \mu_Y(A_*) = 1 + (\mu_X(A_*) - \mu_Y(A_*)) \geq 1,$$

where the term in parenthesis is nonnegative by assumption and the fact that $A_*$ is increasing. That concludes the proof. ■

**Remark 4.2.12.** *Strassen's theorem (Theorem 4.2.11) holds more generally on Polish spaces with a closed partial order. See, e.g., [Lin02, Section IV.1.2] for the details.*

The proof of Corollary 4.2.6 immediately extends to:

**Corollary 4.2.13.** *Let $X$ and $Y$ be $\mathcal{X}$-valued random variables with $X \succeq Y$ and let $f : \mathcal{X} \to \mathbb{R}$ be an increasing function. Then $f(X) \succeq f(Y)$ and furthermore, provided $\mathbb{E}|f(X)|, \mathbb{E}|f(Y)| < +\infty$, we have that*

$$\mathbb{E}[f(X)] \geq \mathbb{E}[f(Y)].$$

**Ordering of Markov chains** Stochastic domination also arises in the context of Markov chains. We begin with an example. Recall the notion of a Markovian coupling from Definition 4.1.20. The following coupling of Markov chains is *not* Markovian.

**Example 4.2.14** (Lazier chain)**.** Consider a random walk $(X_t)$ on the network $\mathcal{N} = ((V, E), c)$ where $V = \{0, 1, \ldots, n\}$ and $i \sim j$ if and only if $|i - j| \leq 1$ (including self-loops). Let $\mathcal{N}' = ((V, E), c')$ be a modified version of $\mathcal{N}$ on the same graph where, for all $i$, $c(i, i) \leq c'(i, i)$. That is, if $(X'_t)$ is random walk on $\mathcal{N}'$, then $(X'_t)$ is "lazier" than $(X_t)$ in that it is more likely to stay put. To simplify the calculations, assume $c(i, i) = 0$ for all $i$.

Assume that both $(X_t)$ and $(X'_t)$ start at $i_0$ and define $M_s := \max_{t \leq s} X_t$ and $M'_s := \max_{t \leq s} X'_t$. Since $(X'_t)$ "travels less" than $(X_t)$ the following claim is intuitively obvious:

**Claim 4.2.15.**

$$M_s \succeq M'_s.$$

We prove this by producing a monotone coupling. First set $(\hat{X}_t)_{t \in \mathbb{Z}_+} := (X_t)_{t \in \mathbb{Z}_+}$. We then generate $(\hat{X}'_t)_{t \in \mathbb{Z}_+}$ as a "sticky" version of $(\hat{X}_t)_{t \in \mathbb{Z}_+}$. That is, $(\hat{X}'_t)$ follows exactly the same transitions as $(\hat{X}_t)$ (including the self-loops), but at each time it opts to stay where it currently is, say state $j$, for an extra time step with probability

$$\alpha_j := \frac{c'(j,j)}{\sum_{i:i\sim j} c'(i,j)},$$

which is in $[0,1]$ by assumption. Marginally, $(\hat{X}'_t)$ is a random walk on $\mathcal{N}'$. Indeed, we have by construction of the coupling that the probability of staying put when in state $j$ is

$$\alpha_j = \frac{c'(j,j)}{\sum_{i:i\sim j} c'(i,j)},$$

and, for $k \neq j$ with $k \sim j$, the probability of moving to state $k$ when in state $j$ is

$$\begin{aligned}
(1-\alpha_j)\frac{c(j,k)}{\sum_{i:i\sim j} c(i,j)} &= \left(\frac{[\sum_{i:i\sim j} c'(i,j)] - c'(j,j)}{\sum_{i:i\sim j} c'(i,j)}\right)\frac{c(j,k)}{\sum_{i:i\sim j} c(i,j)}\\
&= \left(\frac{\sum_{i:i\sim j} c(i,j)}{\sum_{i:i\sim j} c'(i,j)}\right)\frac{c'(j,k)}{\sum_{i:i\sim j} c(i,j)}\\
&= \frac{c'(j,k)}{\sum_{i:i\sim j} c'(i,j)},
\end{aligned}$$

where, on the second line, we used that $c'(i,j) = c(i,j)$ for $i \neq j$ and $i \sim j$. This coupling satisfies almost surely

$$\widehat{M}_s := \max_{t\leq s} \hat{X}_t \geq \max_{t \leq s} \hat{X}'_t =: \widehat{M}'_s$$

because $(\hat{X}'_t)_{t\leq s}$ visits a subset of the states visited by $(\hat{X}_t)_{t\leq s}$. In other words $(\widehat{M}_s, \widehat{M}'_s)$ is a monotone coupling of $(M_s, M'_s)$ and this proves the claim. ◀

As we indicated, the previous example involved an "asynchronous" coupling of the chains. Often a simpler step-by-step approach—that is, through the construction of a Markovian coupling—is possible. We specialize the notion of stochastic domination to that important case.

**Definition 4.2.16** (Stochastic domination of Markov chains)**.** *Let $P$ and $Q$ be transition matrices on a finite or countably infinite poset $(\mathcal{X}, \leq)$. The transition matrix $Q$ is said to* stochastically dominate *the transition matrix $P$ if*

$$x \leq y \implies P(x,\cdot) \preceq Q(y,\cdot). \tag{4.2.7}$$

*If the above condition is satisfied for $P = Q$, we say that $P$ is* stochastically monotone

The analogue of Strassen's theorem in this case is the following theorem, which we prove in the finite case only again.

**Theorem 4.2.17.** *Let $(X_t)_{t\in\mathbb{Z}_+}$ and $(Y_t)_{t\in\mathbb{Z}_+}$ be Markov chains on a finite poset $(\mathcal{X},\leq)$ with transition matrices $P$ and $Q$ respectively. Assume that $Q$ stochastically dominates $P$. Then for all $x_0 \leq y_0$ there is a coupling $(\hat{X}_t,\hat{Y}_t)$ of $(X_t)$ started at $x_0$ and $(Y_t)$ started at $y_0$ such that almost surely*

$$\hat{X}_t \leq \hat{Y}_t, \qquad \forall t.$$

*Furthermore, if the chains are irreducible and have stationary distributions $\pi$ and $\mu$ respectively, then $\pi \preceq \mu$.*

Observe that, for a Markovian, monotone coupling to exist, it is not generally enough for the weaker condition $P(x,\cdot) \preceq Q(x,\cdot)$ to hold for all $x$, as should be clear from the proof. See also Exercise 4.3.

*Proof of Theorem 4.2.17.* Let

$$\mathcal{W} := \{(x,y) \in \mathcal{X}\times\mathcal{X} \,:\, x \leq y\}.$$

For all $(x,y) \in \mathcal{W}$, let $R((x,y),\cdot)$ be the joint law of a monotone coupling of $P(x,\cdot)$ and $Q(y,\cdot)$. Such a coupling exists by Strassen's theorem and Condition (4.2.7). Let $(\hat{X}_t,\hat{Y}_t)$ be a Markov chain on $\mathcal{W}$ with transition matrix $R$ started at $(x_0,y_0)$. By construction, $\hat{X}_t \leq \hat{Y}_t$ for all $t$ almost surely. That proves the first half of the theorem.

For the second half, let $A$ be increasing on $\mathcal{X}$. Note that the first half implies that for all $s \geq 1$

$$P^s(x_0,A) = \mathbb{P}[\hat{X}_s \in A] \leq \mathbb{P}[\hat{Y}_s \in A] = Q^s(y_0,A),$$

because $\hat{X}_s \leq \hat{Y}_s$ and $A$ is increasing. Then, by a standard convergence result for irreducible Markov chains (i.e., (1.1.5)),

$$\pi(A) = \lim_{t\to+\infty} \frac{1}{t}\sum_{s\leq t} P^s(x_0,A) \leq \lim_{t\to+\infty} \frac{1}{t}\sum_{s\leq t} Q^s(y_0,A) = \mu(A).$$

This proves the claim by definition of stochastic domination. ■

An example of application of this theorem is given in the next subsection.

### 4.2.2 ▷ *Ising model: boundary conditions*

Consider the $d$-dimensional lattice $\mathbb{L}^d$. Let $\Lambda$ be a finite subset of vertices in $\mathbb{L}^d$ and define $\mathcal{X} := \{-1,+1\}^\Lambda$, which is a poset when equipped with the relation $\sigma \leq \sigma'$ if and only if $\sigma_i \leq \sigma'_i$ for all $i \in \Lambda$. Generalizing Example 1.2.5, for $\xi \in \{-1,+1\}^{\mathbb{L}^d}$, the (ferromagnetic) Ising model on $\Lambda$ with *boundary conditions* $\xi$ and inverse temperature $\beta$ is the probability distribution over spin configurations $\sigma \in \mathcal{X}$ given by

*boundary conditions*

$$\mu^\xi_{\beta,\Lambda}(\sigma) := \frac{1}{\mathcal{Z}_{\Lambda,\xi}(\beta)} e^{-\beta \mathcal{H}_{\Lambda,\xi}(\sigma)},$$

where

$$\mathcal{H}_{\Lambda,\xi}(\sigma) := -\sum_{\substack{i\sim j \\ i,j\in\Lambda}} \sigma_i\sigma_j - \sum_{\substack{i\sim j \\ i\in\Lambda, j\notin\Lambda}} \sigma_i \xi_j,$$

is the Hamiltonian and

$$\mathcal{Z}_{\Lambda,\xi}(\beta) := \sum_{\sigma\in\mathcal{X}} e^{-\beta\mathcal{H}_{\Lambda,\xi}(\sigma)},$$

is the partition function. For shorthand, we occasionally write $+$ and $-$ instead of $+1$ and $-1$.

For the all-$(+1)$ and all-$(-1)$ boundary conditions we denote the measure above by $\mu^+_{\beta,\Lambda}(\sigma)$ and $\mu^-_{\beta,\Lambda}(\sigma)$ respectively. In this section, we show that these two measures are "extreme" in the following sense.

**Claim 4.2.18.** *For all boundary conditions* $\xi \in \{-1,+1\}^{\mathbb{L}^d}$,

$$\mu^+_{\beta,\Lambda} \succeq \mu^\xi_{\beta,\Lambda} \succeq \mu^-_{\beta,\Lambda}.$$

Intuitively, because the ferromagnetic Ising model favors spin agreement, the all-$(+1)$ boundary condition tends to produce more $+1$s which in turn makes increasing events more likely. And vice versa.

The idea of the proof is to use Theorem 4.2.17 with a suitable choice of Markov chain.

**Stochastic domination** Recall that, in this context, vertices are often referred to as sites. Adapting Definition 1.2.8, we consider the single-site Glauber dynamics, which is the Markov chain on $\mathcal{X}$ which, at each time, selects a site $i \in \Lambda$ uniformly at random and updates the spin $\sigma_i$ according to $\mu^\xi_{\beta,\Lambda}(\sigma)$ conditioned on agreeing with $\sigma$ at all sites in $\Lambda\backslash\{i\}$. Specifically, for $\gamma \in \{-1,+1\}$, $i \in \Lambda$, and $\sigma \in \mathcal{X}$, let

$\sigma^{i,\gamma}$ be the configuration $\sigma$ with the state at $i$ being set to $\gamma$. Then, letting $n = |\Lambda|$, the transition matrix of the Glauber dynamics is

$$Q^{\xi}_{\beta,\Lambda}(\sigma, \sigma^{i,\gamma}) := \frac{1}{n} \cdot \frac{e^{\gamma\beta S^{\xi}_i(\sigma)}}{e^{-\beta S^{\xi}_i(\sigma)} + e^{\beta S^{\xi}_i(\sigma)}},$$

where

$$S^{\xi}_i(\sigma) := \sum_{\substack{j:j\sim i\\ j\in\Lambda}} \sigma_j + \sum_{\substack{j:j\sim i\\ j\notin\Lambda}} \xi_j.$$

All other transitions have probability $0$. It is straightforward to check that $Q^{\xi}_{\beta,\Lambda}$ is a stochastic matrix.

This chain is clearly irreducible. It is also reversible with respect to $\mu^{\xi}_{\beta,\Lambda}$. Indeed, for all $\sigma \in \mathcal{X}$ and $i \in \Lambda$, let

$$S^{\xi}_{\neq i}(\sigma) := \mathcal{H}_{\Lambda,\xi}(\sigma^{i,+}) + S^{\xi}_i(\sigma) = \mathcal{H}_{\Lambda,\xi}(\sigma^{i,-}) - S^{\xi}_i(\sigma),$$

Arguing as in Theorem 1.2.9, we have

$$\begin{aligned}
\mu^{\xi}_{\beta,\Lambda}(\sigma^{i,-})\, Q^{\xi}_{\beta,\Lambda}(\sigma^{i,-}, \sigma^{i,+}) &= \frac{e^{-\beta S^{\xi}_{\neq i}(\sigma)} e^{-\beta S^{\xi}_i(\sigma)}}{\mathcal{Z}_{\Lambda,\xi}(\beta)} \cdot \frac{e^{\beta S^{\xi}_i(\sigma)}}{n[e^{-\beta S^{\xi}_i(\sigma)} + e^{\beta S^{\xi}_i(\sigma)}]} \\
&= \frac{e^{-\beta S^{\xi}_{\neq i}(\sigma)}}{n\mathcal{Z}_{\Lambda,\xi}(\beta)[e^{-\beta S^{\xi}_i(\sigma)} + e^{\beta S^{\xi}_i(\sigma)}]} \\
&= \frac{e^{-\beta S^{\xi}_{\neq i}(\sigma)} e^{\beta S^{\xi}_i(\sigma)}}{\mathcal{Z}_{\Lambda,\xi}(\beta)} \cdot \frac{e^{-\beta S^{\xi}_i(\sigma)}}{n[e^{-\beta S^{\xi}_i(\sigma)} + e^{\beta S^{\xi}_i(\sigma)}]} \\
&= \mu^{\xi}_{\beta,\Lambda}(\sigma^{i,+})\, Q^{\xi}_{\beta,\Lambda}(\sigma^{i,+}, \sigma^{i,-}).
\end{aligned}$$

In particular $\mu^{\xi}_{\beta,\Lambda}$ is the stationary distribution of $Q^{\xi}_{\beta,\Lambda}$.

**Claim 4.2.19.**

$$\xi' \geq \xi \implies Q^{\xi'}_{\beta,\Lambda} \textit{ stochastically dominates } Q^{\xi}_{\beta,\Lambda}. \tag{4.2.8}$$

*Proof.* Because the Glauber dynamics updates a single site at a time, establishing stochastic domination reduces to checking simple one-site inequalities.

**Lemma 4.2.20.** *To establish* (4.2.8)*, it suffices to show that for all $i$ and all $\sigma \leq \tau$*

$$Q^{\xi}_{\beta,\Lambda}(\sigma, \sigma^{i,+}) \leq Q^{\xi'}_{\beta,\Lambda}(\tau, \tau^{i,+}). \tag{4.2.9}$$

*Proof.* Assume (4.2.9) holds. Let $A$ be increasing in $\mathcal{X}$ and let $\sigma \leq \tau$. Then, for the single-site Glauber dynamics, we have

$$Q^{\xi}_{\beta,\Lambda}(\sigma, A) = Q^{\xi}_{\beta,\Lambda}(\sigma, A \cap B_\sigma), \tag{4.2.10}$$

where

$$B_\sigma := \{\sigma^{i,\gamma} \,:\, i \in \Lambda,\ \gamma \in \{-1,+1\}\},$$

and similarly for $\tau$, $\xi'$. Moreover, because $A$ is increasing and $\tau \geq \sigma$,

$$\sigma^{i,\gamma} \in A \implies \tau^{i,\gamma} \in A, \tag{4.2.11}$$

and

$$\sigma^{i,-} \in A \implies \sigma^{i,+} \in A. \tag{4.2.12}$$

Letting

$$I^{\pm}_{\sigma,A} := \{i \in \Lambda \,:\, \sigma^{i,-} \in A\}, \qquad I^{+}_{\sigma,A} := \{i \in \Lambda \,:\, \sigma^{i,-} \notin A, \sigma^{i,+} \in A\},$$

and similarly for $\tau$, we have by (4.2.9), (4.2.10), (4.2.11), and (4.2.12),

$$\begin{aligned}
Q^{\xi}_{\beta,\Lambda}(\sigma, A) &= Q^{\xi}_{\beta,\Lambda}(\sigma, A \cap B_\sigma) \\
&= \sum_{i \in I^{+}_{\sigma,A}} Q^{\xi}_{\beta,\Lambda}(\sigma, \sigma^{i,+}) + \sum_{i \in I^{\pm}_{\sigma,A}} \left[Q^{\xi}_{\beta,\Lambda}(\sigma, \sigma^{i,-}) + Q^{\xi}_{\beta,\Lambda}(\sigma, \sigma^{i,+})\right] \\
&\leq \sum_{i \in I^{+}_{\sigma,A}} Q^{\xi'}_{\beta,\Lambda}(\tau, \tau^{i,+}) + \sum_{i \in I^{\pm}_{\sigma,A}} \left[Q^{\xi}_{\beta,\Lambda}(\sigma, \sigma^{i,-}) + Q^{\xi}_{\beta,\Lambda}(\sigma, \sigma^{i,+})\right] \\
&= \sum_{i \in I^{+}_{\sigma,A}} Q^{\xi'}_{\beta,\Lambda}(\tau, \tau^{i,+}) + \sum_{i \in I^{\pm}_{\sigma,A}} \frac{1}{n} \\
&\leq \sum_{i \in I^{+}_{\tau,A}} Q^{\xi'}_{\beta,\Lambda}(\tau, \tau^{i,+}) + \sum_{i \in I^{\pm}_{\tau,A}} \left[Q^{\xi'}_{\beta,\Lambda}(\tau, \tau^{i,-}) + Q^{\xi'}_{\beta,\Lambda}(\tau, \tau^{i,+})\right] \\
&= Q^{\xi'}_{\beta,\Lambda}(\tau, A),
\end{aligned}$$

as claimed, where on the fifth line we used that $I^{+}_{\sigma,A} \subseteq I^{+}_{\tau,A} \cup I^{\pm}_{\tau,A}$ by (4.2.11) and that $Q^{\xi'}_{\beta,\Lambda}(\tau, \tau^{i,+}) \leq 1/n$ for all $i$ (in particular for $i \in I^{+}_{\tau,A} \setminus I^{+}_{\sigma,A}$). ■

Returning to the proof of Claim 4.2.19, observe that

$$Q^{\xi}_{\beta,\Lambda}(\sigma, \sigma^{i,+}) = \frac{1}{n} \cdot \frac{e^{\beta S^{\xi}_i(\sigma)}}{e^{-\beta S^{\xi}_i(\sigma)} + e^{\beta S^{\xi}_i(\sigma)}} = \frac{1}{n} \cdot \frac{1}{e^{-2\beta S^{\xi}_i(\sigma)} + 1},$$

which is increasing in $S^{\xi}_i(\sigma)$. Now $\sigma \leq \tau$ and $\xi \leq \xi'$ imply that $S^{\xi}_i(\sigma) \leq S^{\xi'}_i(\tau)$. That proves the claim by Lemma 4.2.20. ■

Finally:

*Proof of Claim 4.2.18.* Combining Theorem 4.2.17 and Claim 4.2.19 gives the result. ■

**Remark 4.2.21.** *One can make sense of the limit of $\mu^+_{\beta,\Lambda}$ and $\mu^-_{\beta,\Lambda}$ when $|\Lambda| \to +\infty$, which is known as an infinite-volume Gibbs measure. For more, see for example [RAS15, Chapters 7-10].*

Observe that we have not used any special property of the $d$-dimensional lattice. Indeed Claim 4.2.18 in fact holds for any countable, locally finite graph with positive coupling constants. We give another proof in Example 4.2.33.

### 4.2.3 Correlation inequalities: FKG and Holley's inequalities

A special case of stochastic domination is positive associations. In this section, we restrict ourselves to posets of the form $\{0,1\}^F$ for $F$ finite. We begin with an example.

**Example 4.2.22** (Erdős-Rényi graph: positive associations)**.** Consider an Erdős-Rényi graph $G \sim \mathbb{G}_{n,p}$. Let $\mathcal{E} = \{\{x,y\} \,:\, x,y \in [n],\ x \neq y\}$. Think of $G$ as taking values in the poset $(\{0,1\}^{\mathcal{E}}, \leq)$ where a $1$ indicates that the corresponding edge is present. In fact observe that the law of $G$, which we denote as usual by $\mathbb{P}_{n,p}$, is a product measure on $\{0,1\}^{\mathcal{E}}$. The event $\mathcal{A}$ that $G$ is connected is increasing because adding edges cannot disconnect an already connected graph. So is the event $\mathcal{B}$ of having a chromatic number larger than $4$. Intuitively then, conditioning on $\mathcal{A}$ makes $\mathcal{B}$ more likely: the occurrence of $\mathcal{A}$ tends to be accompanied with a larger number of edges which in turn makes $\mathcal{B}$ more probable.

This is an example of a more general phenomenon. That is, for any non-empty increasing events $\mathcal{A}$ and $\mathcal{B}$, we have:

**Claim 4.2.23.**

$$\mathbb{P}_{n,p}[\mathcal{B} \,|\, \mathcal{A}] \geq \mathbb{P}_{n,p}[\mathcal{B}]. \tag{4.2.13}$$

Or, put differently, the conditional measure $\mathbb{P}_{n,p}[\,\cdot\,|\,\mathcal{A}]$ stochastically dominates the unconditional measure $\mathbb{P}_{n,p}[\,\cdot\,]$. This is a special case of what is known as Harris' inequality, proved below. Note that (4.2.13) is equivalent to $\mathbb{P}_{n,p}[\mathcal{A} \cap \mathcal{B}] \geq \mathbb{P}_{n,p}[\mathcal{A}]\,\mathbb{P}_{n,p}[\mathcal{B}]$, that is, to the fact that $\mathcal{A}$ and $\mathcal{B}$ are positively correlated. ◀

More generally:

**Definition 4.2.24** (Positive associations)**.** *Let $\mu$ be a probability measure on $\{0,1\}^F$ where $F$ is finite. Then $\mu$ is said to have* positive associations*, or is positively as-* positive associations *sociated, if for all increasing functions $f, g : \{0,1\}^F \to \mathbb{R}$*

$$\mu(fg) \geq \mu(f)\mu(g),$$

*where*

$$\mu(h) := \sum_{\omega \in \{0,1\}^F} \mu(\omega) h(\omega).$$

*In particular, for any increasing events $A$ and $B$ it holds that*

$$\mu(A \cap B) \geq \mu(A)\mu(B),$$

*that is, $A$ and $B$ are* positively correlated. *Denoting by $\mu(A \,|\, B)$ the conditional probability of $A$ given $B$, this is equivalent to* *positively correlated*

$$\mu(A \,|\, B) \geq \mu(A).$$

**Remark 4.2.25.** *Note that positive associations is concerned only with "monotone" events. See Remark 4.2.45.*

**Remark 4.2.26.** *A notion of negative associations, which is a somewhat more delicate concept, was defined in Remark 3.3.43. See also [Pem00].*

Let $\mu$ be positively associated. Note that if $A$ and $B$ are *decreasing*, that is, their complements are increasing (see Exercise 4.4), then *decreasing*

$$\begin{aligned} \mu(A \cap B) &= 1 - \mu(A^c \cup B^c) \\ &= 1 - \mu(A^c) - \mu(B^c) + \mu(A^c \cap B^c) \\ &\geq 1 - \mu(A^c) - \mu(B^c) + \mu(A^c)\mu(B^c) \\ &= \mu(A)\mu(B), \end{aligned}$$

or $\mu(A \,|\, B) \geq \mu(A)$. Similarly, if $A$ is increasing and $B$ is decreasing, we have $\mu(A \cap B) \leq \mu(A)\mu(B)$, or

$$\mu(A \,|\, B) \leq \mu(A). \tag{4.2.14}$$

Harris' inequality states that product measures on $\{0,1\}^F$ have positive associations. We prove a more general result known as the *FKG inequality*. For two configurations $\omega$, $\omega'$ in $\{0,1\}^F$, we let $\omega \wedge \omega'$ and $\omega \vee \omega'$ be the coordinatewise minimum and maximum of $\omega$ and $\omega'$.

**Definition 4.2.27** (FKG condition)**.** *Let $\mathcal{X} = \{0,1\}^F$ where $F$ is finite. A positive probability measure $\mu$ on $\mathcal{X}$ satisfies the* FKG condition *if* *FKG condition*

$$\mu(\omega \vee \omega')\,\mu(\omega \wedge \omega') \geq \mu(\omega)\,\mu(\omega'), \qquad \forall \omega, \omega' \in \mathcal{X}. \tag{4.2.15}$$

*This property is also known as log-supermodularity. We call such a measure an* FKG measure.

*FKG inequality*

**Theorem 4.2.28** (FKG inequality)**.** *Let $\mathcal{X} = \{0,1\}^F$ where $F$ is finite. Suppose $\mu$ is a positive probability measure on $\mathcal{X}$ satisfying the FKG condition. Then $\mu$ has positive associations.*

**Remark 4.2.29.** *Strict positivity is not in fact needed [FKG71]. The FKG condition is equivalent to a strong form of positive associations. See Exercise 4.8.*

Note that product measures satisfy the FKG condition with equality. Indeed if $\mu(\omega)$ is of the form $\prod_{f \in F} \mu_f(\omega_f)$ then

$$\begin{aligned}
\mu(\omega \vee \omega')\,\mu(\omega \wedge \omega') &= \prod_f \mu_f(\omega_f \vee \omega'_f)\ \mu_f(\omega_f \wedge \omega'_f)\\
&= \prod_{f:\omega_f = \omega'_f} \mu_f(\omega_f)^2 \prod_{f:\omega_f \neq \omega'_f} \mu_f(\omega_f)\mu_f(\omega'_f)\\
&= \prod_{f:\omega_f = \omega'_f} \mu_f(\omega_f)\mu_f(\omega'_f) \prod_{f:\omega_f \neq \omega'_f} \mu_f(\omega_f)\mu_f(\omega'_f)\\
&= \mu(\omega)\,\mu(\omega').
\end{aligned}$$

So the FKG inequality (Theorem 4.2.28) applies, for instance, to bond percolation and the Erdős-Rényi random graph model. The pointwise nature of the FKG condition also makes it relatively easy to check for measures which are defined explicitly up to a normalizing constant, such as the Ising model.

**Example 4.2.30** (Ising model with boundary conditions: checking FKG)**.** Consider again the setting of Section 4.2.2. We work on the space $\mathcal{X} := \{-1,+1\}^\Lambda$ rather than $\{0,1\}^F$. Fix a finite $\Lambda \subseteq \mathbb{L}^d$, $\xi \in \{-1,+1\}^{\mathbb{L}^d}$ and $\beta > 0$.

**Claim 4.2.31.** *The measure $\mu^\xi_{\beta,\Lambda}$ satisfies the FKG condition and therefore has positive associations.*

Intuitively, taking the minimum (or maximum) of two spin configurations tends to increase agreement and therefore leads to a higher likelihood. For $\sigma, \sigma' \in \mathcal{X}$, let $\overline{\tau} = \sigma \vee \sigma'$ and $\underline{\tau} = \sigma \wedge \sigma'$. By taking logarithms in the FKG condition and rearranging, we arrive at

$$\mathcal{H}_{\Lambda,\xi}(\overline{\tau}) + \mathcal{H}_{\Lambda,\xi}(\underline{\tau}) \leq \mathcal{H}_{\Lambda,\xi}(\sigma) + \mathcal{H}_{\Lambda,\xi}(\sigma'), \tag{4.2.16}$$

and we see that proving the claim boils down to checking an inequality for each term in the Hamiltonian (which, confusingly, has a negative sign in it).

When $i \in \Lambda$ and $j \notin \Lambda$ such that $i \sim j$, we have

$$\overline{\tau}_i \xi_j + \underline{\tau}_i \xi_j = (\overline{\tau}_i + \underline{\tau}_i)\xi_j = (\sigma_i + \sigma_i')\xi_j = \sigma_i \xi_j + \sigma_i' \xi_j. \tag{4.2.17}$$

For $i, j \in \Lambda$ with $i \sim j$, note first that the case $\sigma_j = \sigma_j'$ reduces to the previous calculation (with $\sigma_j = \sigma_j'$ playing the role of $\xi_j$), so we assume $\sigma_i \neq \sigma_i'$ and $\sigma_j \neq \sigma_j'$. Then

$$\overline{\tau}_i \overline{\tau}_j + \underline{\tau}_i \underline{\tau}_j = (+1)(+1) + (-1)(-1) = 2 \geq \sigma_i \sigma_j + \sigma_i' \sigma_j',$$

since 2 is the largest value the rightmost expression ever takes. We have established (4.2.16), which implies the claim.

Again, we have not used any special property of the lattice and the same result holds for countable, locally finite graphs with positive coupling constants. Note however that in the anti-ferromagnetic case, that is, if we multiply the Hamiltonian by $-1$, the above argument does not work. Indeed there is no reason to expect positive associations in that case. ◀

The FKG inequality in turn follows from a more general result known as *Holley's inequality*.

**Theorem 4.2.32** (Holley's inequality)**.** *Let $\mathcal{X} = \{0,1\}^F$ where $F$ is finite. Suppose $\mu_1$ and $\mu_2$ are positive probability measures on $\mathcal{X}$ satisfying* *Holley's inequality*

$$\mu_2(\omega \vee \omega')\, \mu_1(\omega \wedge \omega') \geq \mu_2(\omega)\, \mu_1(\omega'), \qquad \forall \omega, \omega' \in \mathcal{X}. \tag{4.2.18}$$

*Then $\mu_1 \preceq \mu_2$.*

Before proving Holley's inequality (Theorem 4.2.32), we check that it indeed implies the FKG inequality. See Exercise 4.5 for an elementary proof in the independent case, that is, of Harris' inequality.

*Proof of Theorem 4.2.28.* Assume that $\mu$ satisfies the FKG condition and let $f$, $g$ be increasing functions. Because of our restriction to positive measures in Holley's inequality, we will work with positive functions. This is done without loss of generality. Indeed, letting $\mathbf{0}$ be the all-0 vector, note that $f$ and $g$ are increasing if and only if $f' := f - f(\mathbf{0}) + 1 > 0$ and $g' := g - g(\mathbf{0}) + 1 > 0$ are increasing and that, moreover,

$$\begin{aligned} \mu(f'g') - \mu(f')\mu(g') &= \mu([f' - \mu(f')][g' - \mu(g')]) \\ &= \mu([f - \mu(f)][g - \mu(g)]) \\ &= \mu(fg) - \mu(f)\mu(g). \end{aligned}$$

In Holley's inequality, we let $\mu_1 := \mu$ and define the positive probability measure

$$\mu_2(\omega) := \frac{g(\omega)\mu(\omega)}{\mu(g)}.$$

We check that $\mu_1$ and $\mu_2$ satisfy the conditions of Theorem 4.2.32. Note that $\omega' \leq \omega \vee \omega'$ for any $\omega$ so that, because $g$ is increasing, we have $g(\omega') \leq g(\omega \vee \omega')$. Hence, for any $\omega, \omega'$,

$$\begin{aligned}
\mu_1(\omega)\mu_2(\omega') &= \mu(\omega)\frac{g(\omega')\mu(\omega')}{\mu(g)} \\
&= \mu(\omega)\mu(\omega')\frac{g(\omega')}{\mu(g)} \\
&\leq \mu(\omega \wedge \omega')\mu(\omega \vee \omega')\frac{g(\omega \vee \omega')}{\mu(g)} \\
&= \mu_1(\omega \wedge \omega')\mu_2(\omega \vee \omega'),
\end{aligned}$$

where on the third line we used the FKG condition satisfied by $\mu$.

So Holley's inequality implies that $\mu_2 \succeq \mu_1$. Hence, since $f$ is increasing, by Corollary 4.2.13

$$\mu(f) = \mu_1(f) \leq \mu_2(f) = \frac{\mu(fg)}{\mu(g)},$$

and the theorem is proved. ■

*Proof of Theorem 4.2.32*. The idea of the proof is to use Theorem 4.2.17. This is similar to what was done in Section 4.2.2. Again we use a single-site dynamic. For $x \in \mathcal{X}$ and $\gamma \in \{0, 1\}$, we let $x^{i,\gamma}$ be $x$ with coordinate $i$ set to $\gamma$. We write $x \sim y$ if $\|x - y\|_1 = 1$. Let $n = |F|$. We use a scheme analogous to the Metropolis algorithm (see Example 1.1.30). A natural symmetric chain on $\mathcal{X}$ is to pick a coordinate uniformly at random, and flip its value. We modify it to guarantee reversibility with respect to the desired stationary distributions, namely $\mu_1$ and $\mu_2$.

For $\alpha, \beta > 0$ small enough, the following transition matrix over $\mathcal{X}$ is irreducible and reversible with respect to its stationary distribution $\mu_2$: for all $i \in F$, $y \in \mathcal{X}$,

$$Q(y^{i,0}, y^{i,1}) = \frac{1}{n}\alpha \left\{\beta\right\},$$

$$Q(y^{i,1}, y^{i,0}) = \frac{1}{n}\alpha \left\{\beta \, \frac{\mu_2(y^{i,0})}{\mu_2(y^{i,1})}\right\},$$

$$Q(y, y) = 1 - \sum_{z:z\sim y} Q(y, z).$$

Let $P$ be similarly defined with respect to $\mu_1$ *with the same values of* $\alpha$ *and* $\beta$. For reasons that will be clear below, the value of $0 < \beta < 1$ is chosen small enough that the *sum* of the two expressions in brackets above is smaller than $1$ *for all* $y, i$ in both $P$ and $Q$. The value of $\alpha > 0$ is then chosen small enough that $P(y, y), Q(y, y) \geq 0$ *for all* $y$. Reversibility follows immediately from the first two equations. We call the first transition above an *upward transition* and the second one a *downward transition*.

By Theorem 4.2.17, it remains to show that $Q$ stochastically dominates $P$. That is, for any $x \leq y$, we want to show that $P(x, \cdot) \preceq Q(y, \cdot)$. We produce a monotone coupling $(\hat{X}, \hat{Y})$ of these two distributions. Because $x \leq y$, our goal is never to perform an upward transition in $x$ *simultaneously* with a downward transition in $y$. Observe that

$$\frac{\mu_1(x^{i,0})}{\mu_1(x^{i,1})} \geq \frac{\mu_2(y^{i,0})}{\mu_2(y^{i,1})} \tag{4.2.19}$$

by taking $\omega = y^{i,0}$ and $\omega' = x^{i,1}$ in Condition (4.2.18).

The coupling works as follows. Fix $x \leq y$. With probability $1 - \alpha$, set $(\hat{X}, \hat{Y}) := (x, y)$. Otherwise, pick a coordinate $i \in F$ uniformly at random. There are several cases to consider depending on the coordinates $x_i, y_i$ (with $x_i \leq y_i$ by assumption):

- $(x_i, y_i) = (0, 0)$: With probability $\beta$, perform an upward transition in both coordinates, that is, set $\hat{X} := x^{i,1}$ and $\hat{Y} := y^{i,1}$. With probability $1 - \beta$, set $(\hat{X}, \hat{Y}) := (x, y)$ instead.

- $(x_i, y_i) = (1, 1)$: With probability $\beta \frac{\mu_2(y^{i,0})}{\mu_2(y^{i,1})}$, perform a downward transition in both coordinates, that is, set $\hat{X} := x^{i,0}$ and $\hat{Y} := y^{i,0}$. With probability

$$\beta \left( \frac{\mu_1(x^{i,0})}{\mu_1(x^{i,1})} - \frac{\mu_2(y^{i,0})}{\mu_2(y^{i,1})} \right),$$

perform a downward transition in $x$ only, that is, set $\hat{X} := x^{i,0}$ and $\hat{Y} := y$. With the remaining probability, set $(\hat{X}, \hat{Y}) := (x, y)$ instead. Note that (4.2.19) and our choice of $\beta$ guarantees that this step is well-defined.

- $(x_i, y_i) = (0, 1)$: With probability $\beta$, perform an upward transition in $x$ only, that is, set $\hat{X} := x^{i,1}$ and $\hat{Y} := y$. With probability $\beta \frac{\mu_2(y^{i,0})}{\mu_2(y^{i,1})}$, perform a downward transition in $y$ only, that is, set $\hat{X} := x$ and $\hat{Y} := y^{i,0}$. With the remaining probability, set $(\hat{X}, \hat{Y}) := (x, y)$ instead. Again our choice of $\beta$ guarantees that this step is well-defined.

A little accounting shows that this is indeed a coupling of $P(x, \cdot)$ and $Q(y, \cdot)$. By construction, this coupling satisfies $\hat{X} \leq \hat{Y}$ almost surely. An application of Theorem 4.2.17 concludes the proof. ■

**Example 4.2.33** (Ising model revisited)**.** Holley's inequality implies Claim 4.2.18. To see this, just repeat the calculations of Example 4.2.30, where now (4.2.17) is replaced with an inequality. See Exercise 4.6. ◀

### 4.2.4 ▷ *Random graphs: Janson's inequality and application to the clique number in the Erdős-Rényi model*

Let $G = (V, E) \sim \mathbb{G}_{n,p_n}$ be an Erdős-Rényi graph (see Definition 1.2.2). By Claim 2.3.5, the property of being triangle-free has threshold $n^{-1}$. That is, the probability that $G$ contains a triangle goes to 0 or 1 as $n \to +\infty$ according to whether $p_n \ll n^{-1}$ or $p_n \gg n^{-1}$ respectively. In this section, we investigate what happens "at the threshold," by which we mean that we take $p_n = \lambda/n$ for some $\lambda > 0$ not depending on $n$.

For any subset $S$ of three distinct vertices of $G$, let $B_S$ be the event that $S$ forms a triangle in $G$. So

$$\varepsilon := \mathbb{P}_{n,p_n}[B_S] = p_n^3 \to 0. \tag{4.2.20}$$

Denoting the unordered triples of distinct vertices by $\binom{V}{3}$, let $X_n = \sum_{S \in \binom{V}{3}} \mathbf{1}_{B_S}$ be the number of triangles in $G$. By the linearity of expectation, the mean number of triangles is

$$\mathbb{E}_{n,p_n} X_n = \binom{n}{3} p_n^3 = \frac{n(n-1)(n-2)}{6} \left(\frac{\lambda}{n}\right)^3 \to \frac{\lambda^3}{6},$$

as $n \to +\infty$. *If the events $\{B_S\}_S$ were mutually independent,* $X_n$ would be binomially distributed and the event that $G$ is triangle-free would have probability

$$\prod_{S \in \binom{V}{3}} \mathbb{P}_{n,p_n}[B_S^c] = (1 - p_n^3)^{\binom{n}{3}} \to e^{-\lambda^3/6}. \tag{4.2.21}$$

In fact, by the Poisson approximation to the binomial (e.g., Theorem 4.1.18), we would have that the number of triangles converges weakly to $\mathrm{Poi}(\lambda^3/6)$.

In reality, of course, the events $\{B_S\}$ are not *mutually* independent. Observe however that, for *most* pairs $S, S'$, the events $B_S$ and $B_{S'}$ are in fact *pairwise* independent. That is the case whenever $|S \cap S'| \leq 1$, that is, whenever the edges connecting $S$ are disjoint from those connecting $S'$. Write $S \sim S'$ if $S \neq S'$ are not independent, that is, if $|S \cap S'| = 2$. The expected number of unordered pairs $S \sim S'$ both forming a triangle is

$$\Delta := \frac{1}{2} \sum_{\substack{S,S' \in \binom{V}{3} \\ S \sim S'}} \mathbb{P}_{n,p_n}[B_S \cap B_{S'}] = \frac{1}{2}\binom{n}{3}\binom{3}{2}(n-3)p_n^5 = \Theta(n^4 p_n^5) \to 0, \tag{4.2.22}$$

where the $\binom{n}{3}$ comes from the number of ways of choosing $S$, the $\binom{3}{2}$ comes from the number of ways of choosing the vertices in common between $S$ and $S'$, and the $n-3$ comes from the number of ways of choosing the third vertex of $S'$. Given that the events $\{B_S\}_{S \in \binom{V}{3}}$ are "mostly" independent, it is natural to expect that $X_n$ behaves asymptotically as it would in the independent case. Indeed we prove:

**Claim 4.2.34.**

$$\mathbb{P}_{n,p_n}[X_n = 0] \to e^{-\lambda^3/6}.$$

**Remark 4.2.35.** *In fact,* $X_n \xrightarrow{d} \mathrm{Poi}(\lambda^3/6)$. *See Exercises 2.18 and 4.9.*

The FKG inequality (Theorem 4.2.28) immediately gives one direction. Recall that $\mathbb{P}_{n,p_n}$, as a product measure over edge sets, satisfies the FKG condition and therefore has positive associations by the FKG inequality. Moreover the events $B_S^c$ are decreasing for all $S$. Hence, applying positive associations inductively,

$$\mathbb{P}_{n,p_n}\left[\bigcap_{S \in \binom{V}{3}} B_S^c\right] \geq \prod_{S \in \binom{V}{3}} \mathbb{P}_{n,p_n}[B_S^c] \to e^{-\lambda^3/6},$$

where the limit follows from (4.2.21). As it turns out, the FKG inequality also gives a bound in the other direction. This is known as *Janson's inequality*, which we state in a more general context.

**Janson's inequality** Let $\mathcal{X} := \{0,1\}^F$ where $F$ is finite. Let $B_i$, $i \in I$, be a finite collection of events of the form

$$B_i := \{\omega \in \mathcal{X} : \omega \geq \beta^{(i)}\},$$

for some $\beta^{(i)} \in \mathcal{X}$. Think of these as "bad events" corresponding to a certain subset of coordinates being set to 1. By definition, the $B_i$s are increasing. Assume $\mathbb{P}$ is a positive product measure on $\mathcal{X}$. Write $i \sim j$ if $\beta_r^{(i)} = \beta_r^{(j)} = 1$ for at least one $r$ and note that $B_i$ is independent of $B_j$ if $i \nsim j$. Set

$$\Delta := \sum_{\substack{\{i,j\} \\ i \sim j}} \mathbb{P}[B_i \cap B_j].$$

**Theorem 4.2.36** (Janson's inequality)**.** *Let $\mathcal{X} := \{0,1\}^F$ where $F$ is finite and $\mathbb{P}$ be a positive product measure on $\mathcal{X}$. Let $\{B_i\}_{i\in I}$ and $\Delta$ be as above. Assume further that there is $\varepsilon > 0$ such that $\mathbb{P}[B_i] \leq \varepsilon$ for all $i \in I$. Then*

$$\prod_{i\in I} \mathbb{P}[B_i^c] \leq \mathbb{P}[\cap_{i\in I} B_i^c] \leq e^{\frac{\Delta}{1-\varepsilon}} \prod_{i\in I} \mathbb{P}[B_i^c].$$

Before proving the theorem, we show that it implies Claim 4.2.34. We have already shown in (4.2.20) and (4.2.22) that $\varepsilon \to 0$ and $\Delta \to 0$. Janson's inequality (Theorem 4.2.36) immediately implies the claim by (4.2.21).

*Proof of Theorem 4.2.36.* The lower bound is the FKG inequality.

In the other direction, assume without loss of generality that $I = [m]$. The first step is to apply the chain rule to obtain

$$\mathbb{P}[\cap_{i\in I} B_i^c] \quad = \quad \prod_{i=1}^{m} \mathbb{P}[B_i^c \mid \cap_{j\in[i-1]} B_j^c].$$

The rest is clever manipulation. For $i \in [m]$, let $N(i) := \{\ell \in [m] \,:\, \ell \sim i\}$ and $N_<(i) := N(i) \cap [i-1]$. Note that $B_i$ is independent of $\{B_\ell \,:\, \ell \in [i-1]\backslash N_<(i)\}$. Hence,

$$\begin{aligned}
\mathbb{P}[B_i \mid \cap_{j\in[i-1]} B_j^c] &= \frac{\mathbb{P}\left[B_i \cap \left(\cap_{j\in[i-1]} B_j^c\right)\right]}{\mathbb{P}[\cap_{j\in[i-1]} B_j^c]} \\
&\geq \frac{\mathbb{P}\left[B_i \cap \left(\cap_{j\in N_<(i)} B_j^c\right) \cap \left(\cap_{j\in[i-1]\backslash N_<(i)} B_j^c\right)\right]}{\mathbb{P}[\cap_{j\in[i-1]\backslash N_<(i)} B_j^c]} \\
&= \mathbb{P}\left[B_i \cap \left(\cap_{j\in N_<(i)} B_j^c\right) \,\middle|\, \cap_{j\in[i-1]\backslash N_<(i)} B_j^c\right] \\
&= \mathbb{P}\left[B_i \,\middle|\, \cap_{j\in[i-1]\backslash N_<(i)} B_j^c\right] \\
&\qquad \times \mathbb{P}\left[\cap_{j\in N_<(i)} B_j^c \,\middle|\, B_i \cap \left(\cap_{j\in[i-1]\backslash N_<(i)} B_j^c\right)\right] \\
&= \mathbb{P}\left[B_i\right] \mathbb{P}\left[\cap_{j\in N_<(i)} B_j^c \,\middle|\, B_i \cap \left(\cap_{j\in[i-1]\backslash N_<(i)} B_j^c\right)\right],
\end{aligned}$$

where we used independence for the first term on the last line. By a union bound, the second term on the last line is

$$\begin{aligned}
&\mathbb{P}\left[\cap_{j\in N_<(i)} B_j^c \,\middle|\, B_i \cap \left(\cap_{j\in[i-1]\backslash N_<(i)} B_j^c\right)\right] \\
&\qquad \geq 1 - \sum_{j\in N_<(i)} \mathbb{P}\left[B_j \,\middle|\, B_i \cap \left(\cap_{j\in[i-1]\backslash N_<(i)} B_j^c\right)\right] \\
&\qquad \geq 1 - \sum_{j\in N_<(i)} \mathbb{P}\left[B_j \mid B_i\right],
\end{aligned}$$

where the last line follows from the FKG inequality. This requires some explanations:

- On the event $B_i$, all coordinates $\ell$ with $\beta_\ell^{(i)} = 1$ are fixed to 1, and the other ones are free. So we can think of $\mathbb{P}[\,\cdot \mid B_i]$ as a positive product measure on $\{0,1\}^{F'}$ with $F' := \{\ell \in [m] \,:\, \beta_\ell^{(i)} = 0\}$.
- The event $B_j$ is increasing, while the event $\cap_{j\in[i-1]\setminus N_<(i)} B_j^c$ is decreasing as the intersection of decreasing events (see Exercise 4.4).
- So we can apply the FKG inequality in the form (4.2.14) to $\mathbb{P}[\,\cdot \mid B_i]$.

Combining the last three displays and using $1 + x \le e^x$ for all $x$ (see Exercise 1.16), we get

$$\begin{aligned}
\mathbb{P}[\cap_{i\in I} B_i^c] &\le \prod_{i=1}^m \left[\mathbb{P}\left[B_i^c\right] + \sum_{j\in N_<(i)} \mathbb{P}\left[B_i \cap B_j\right]\right]\\
&\le \prod_{i=1}^m \mathbb{P}\left[B_i^c\right]\left[1 + \frac{1}{1-\varepsilon}\sum_{j\in N_<(i)} \mathbb{P}\left[B_i \cap B_j\right]\right]\\
&\le \prod_{i=1}^m \mathbb{P}\left[B_i^c\right]\,\exp\left(\frac{1}{1-\varepsilon}\sum_{j\in N_<(i)} \mathbb{P}\left[B_i \cap B_j\right]\right).
\end{aligned}$$

where we used the assumption $\mathbb{P}[B_i] \le \varepsilon$ on the second line. By the definition of $\Delta$, we are done. ■

### 4.2.5 ▷ *Percolation: RSW theory and a proof of Harris' theorem*

Consider bond percolation (Definition 1.2.1) on the two-dimensional lattice $\mathbb{L}^2$. Recall that the percolation function is given by

$$\theta(p) := \mathbb{P}_p[|\mathcal{C}_0| = +\infty],$$

where $\mathcal{C}_0$ is the open cluster of the origin. We know from Example 4.2.10 that $\theta(p)$ is non-decreasing. Let

$$p_c(\mathbb{L}^2) := \sup\{p \ge 0 \,:\, \theta(p) = 0\},$$

be the critical value. We proved in Section 2.2.4 that there is a non-trivial transition, that is, $p_c(\mathbb{L}^2) \in (0,1)$. See Exercise 2.3 for a proof that $p_c(\mathbb{L}^2) \in [1/3, 2/3]$.

Our goal in this section is to use the FKG inequality to improve this further to:

**Theorem 4.2.37** (Harris' theorem)**.**

$$\theta(1/2) = 0.$$

Or, put differently, $p_c(\mathbb{L}^2) \geq 1/2$.

**Remark 4.2.38.** *This bound is tight, that is, in fact $p_c(\mathbb{L}^2) = 1/2$. The other direction is known as* Kesten's theorem*. See, e.g., [BR06a].*

Here we present a proof of Harris' theorem that uses an important tool in percolation theory, the *Russo-Seymour-Welsh (RSW) lemma*, an application of the FKG inequality.

### Harris' theorem

To motivate the RSW lemma, we start with the proof of Harris' theorem.

*Proof of Theorem 4.2.37.* Fix $p = 1/2$. We use the dual lattice $\widetilde{\mathbb{L}}^2$ as we did in Section 2.2.4. Consider the annulus

$$\mathrm{Ann}(\ell) := [-3\ell, 3\ell]^2 \backslash [-\ell, \ell]^2.$$

The existence of a closed dual cycle inside $\mathrm{Ann}(\ell)$, an event we denote by $\mathrm{O_d}(\ell)$, prevents the possibility of an infinite open path from the origin in the primal lattice $\mathbb{L}^2$. See Figure 4.4. That is,

$$\mathbb{P}_{1/2}[|\mathcal{C}_0| = +\infty] \leq \prod_{k=0}^{K} \{1 - \mathbb{P}_{1/2}[\mathrm{O_d}(3^k)]\}, \tag{4.2.23}$$

for all $K$, where we took powers of 3 to make the annuli disjoint and therefore independent. To prove the theorem, it suffices to show that there is a constant $c^* > 0$ such that, for all $\ell$, $\mathbb{P}_{1/2}[\mathrm{O_d}(\ell)] \geq c^*$. Then the right-hand side of (4.2.23) tends to 0 as $K \to +\infty$.

To simplify further, thinking of $\mathrm{Ann}(\ell)$ as a union of four rectangles $[-3\ell, -\ell) \times [-3\ell, 3\ell]$, $[-3\ell, 3\ell] \times (\ell, 3\ell]$, etc., it suffices to consider the event $\mathrm{O}^{\#}_{\mathrm{d}}(\ell)$ that each one of these rectangles contains a closed dual path connecting its two shorter sides. To be more precise, for the first rectangle above for instance, the path connects $[-3\ell + 1/2, -\ell - 1/2] \times \{3\ell - 1/2\}$ to $[-3\ell + 1/2, -\ell - 1/2] \times \{-3\ell + 1/2\}$ and stays inside the rectangle. See Figure 4.4. By symmetry the probability that such a path exists is the same for all four rectangles. Denote that probability by $\rho_\ell$. Moreover the event that such a path exists is increasing so, although the four events

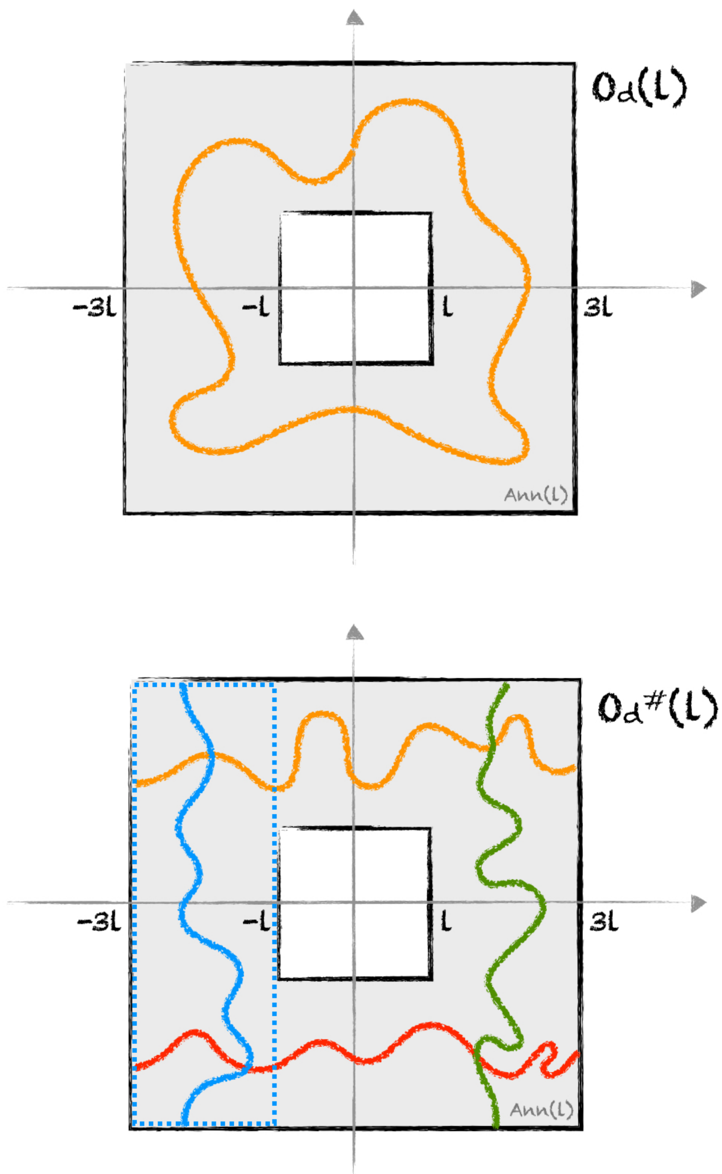


Figure 4.4: Top: the event $\mathrm{O_d}(\ell)$. Bottom: the event $\mathrm{O_d^{\#}}(\ell)$.

are not independent, we can apply the FKG inequality (Theorem 4.2.28). Hence, since $\mathrm{O}_{\mathrm{d}}^{\#}(\ell) \subseteq \mathrm{O}_{\mathrm{d}}(\ell)$, we finally get the bound

$$\mathbb{P}_{1/2}[\mathrm{O}_{\mathrm{d}}(\ell)] \geq \rho_\ell^4.$$

The RSW lemma and some symmetry arguments, both of which are detailed below, imply:

**Claim 4.2.39.** *There is some $c > 0$ such that, for all $\ell$,*

$$\rho_\ell \geq c.$$

That concludes the proof. ■

It remains to prove Claim 4.2.39. We first state the RSW lemma.

**RSW theory**

We have reduced the proof of Harris' theorem to bounding the probability that certain closed paths exist in the dual lattice. To be consistent with the standard RSW notation, we switch to the primal lattice and consider open paths. We also let $p$ take any value in $(0, 1)$.

Let $R_{n,\alpha}(p)$ be the probability that the rectangle

$$B(\alpha n, n) := [-n, (2\alpha - 1)n] \times [-n, n],$$

has an open path connecting its left and right sides with the path remaining inside the rectangle. Such a path is called an (open) *left-right crossing*. The event that a left-right crossing exists in a rectangle $B$ is denoted by $\mathrm{LR}(B)$. We similarly define the event, $\mathrm{TB}(B)$, that a *top-bottom crossing* exists in $B$. In essence, the RSW lemma says this: if there is a significant probability that a left-right crossing exists in the square $B(n, n)$, then there is a significant probability that a left-right crossing exists in the rectangle $B(3n, n)$. More precisely, here is a version of the theorem that will be enough for our purposes. (See Exercise 4.10 for a generalization.)

**Lemma 4.2.40** (RSW lemma)**.** *For all $n \geq 2$ (divisible by 4) and $p \in (0, 1)$,*

$$R_{n,3}(p) \geq \frac{1}{256} R_{n,1}(p)^{11} R_{n/2,1}(p)^{12}. \tag{4.2.24}$$

The right-hand side of (4.2.24) depends only on the probability of crossing a square from left to right. By a duality argument, at $p = 1/2$, it turns out that this probability is at least $1/2$ *independently of $n$*. Before presenting a proof of the RSW lemma, we detail this argument and finish the proof of Harris' theorem.

*Proof of Claim 4.2.39.* The point of (4.2.24) is that, if $R_{n,1}(1/2)$ is bounded away from $0$ uniformly in $n$, then so is the left-hand side. By the argument in the proof of Harris' theorem, this then implies that a closed cycle exists in $\mathrm{Ann}(n)$ with a probability bounded away from 0 as well. Hence to prove Claim 4.2.39 it suffices to give a lower bound on $R_{n,1}(1/2)$. It is crucial that this bound not depend on the "scale" $n$.

As it turns out, a simple duality-based symmetry argument does the trick. The following fact about $\mathbb{L}^2$ is a variant of the contour lemma (Lemma 2.2.14). Its proof is similar and Exercise 4.11 asks for the details (the "if" direction being the non-trivial implication).

**Lemma 4.2.41.** *There is an open left-right crossing in the primal rectangle* $[0, n+1] \times [0, n]$ *if and only if there is* no *closed top-bottom crossing in the dual rectangle* $[1/2, n + 1/2] \times [-1/2, n + 1/2]$.

By symmetry, when $p = 1/2$, the two events in Lemma 4.2.41 have equal probability. So they must have probability $1/2$ because they form a partition of the space of outcomes. By monotonicity, that implies $R_{n,1}(1/2) \geq 1/2$ for all $n$. The RSW lemma then implies the required bound. ∎

The proof of the RSW lemma involves a clever choice of event that relates the existence of crossings in squares and rectangles. (Combining crossings of squares into crossings of rectangles is not as trivial as it might look. Try it before reading the proof.)

*Proof of Lemma 4.2.40.* There are several steps in the proof.

**Step 1: it suffices to bound $R_{n,3/2}(p)$** We first reduce the proof to finding a bound on $R_{n,3/2}(p)$. Let $B_1' := B(2n, n)$ and $B_2' := [n, 5n] \times [-n, n]$. Note that $B_1' \cup B_2' = B(3n, n)$ and $B_1' \cap B_2' = [n, 3n] \times [-n, n]$. Then we have the implication

$$\mathrm{LR}(B_1') \cap \mathrm{TB}(B_1' \cap B_2') \cap \mathrm{LR}(B_2') \subseteq \mathrm{LR}(B(3n, n)).$$

See Figure 4.5. Each event on the left-hand side is increasing so the FKG inequality gives

$$R_{n,3}(p) \geq R_{n,2}(p)^2 R_{n,1}(p).$$

A similar argument over $B(2n, n)$ gives

$$R_{n,2}(p) \geq R_{n,3/2}(p)^2 R_{n,1}(p).$$

Combining the two, we have proved:

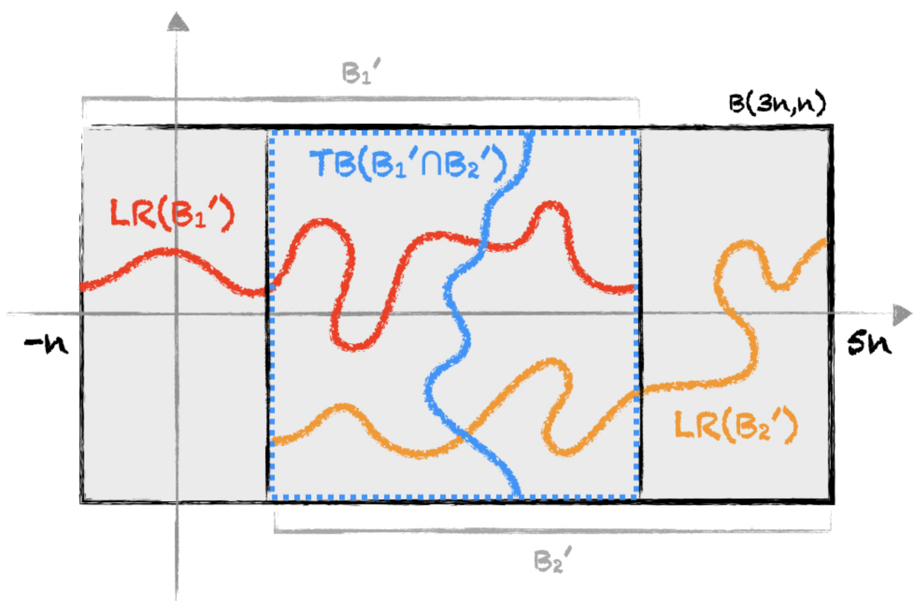


Figure 4.5: Illustration of the implication $\mathrm{LR}(B_1') \cap \mathrm{TB}(B_1' \cap B_2') \cap \mathrm{LR}(B_2') \subseteq \mathrm{LR}(B(3n, n))$.

**Lemma 4.2.42** (Proof of RSW: step 1)**.**

$$R_{n,3}(p) \geq R_{n,3/2}(p)^4 R_{n,1}(p)^3. \tag{4.2.25}$$

**Step 2: bounding $R_{n,3/2}(p)$** The heart of the proof is to bound $R_{n,3/2}(p)$ using an event involving crossings of squares. Let

$$\begin{aligned} B_1 &:= B(n,n) = [-n,n] \times [-n,n], \\ B_2 &:= [0,2n] \times [-n,n], \\ B_{12} &:= B_1 \cap B_2 = [0,n] \times [-n,n], \\ S &:= [0,n] \times [0,n]. \end{aligned}$$

Let $\Gamma_1$ be the event that there are paths $P_1, P_2$, where $P_1$ is a top-bottom crossing of $S$ and $P_2$ is an open path connecting the left side of $B_1$ to $P_1$ and stays inside $B_1$. Similarly let $\Gamma_2'$ be the event that there are paths $P_1', P_2'$, where $P_1'$ is a top-bottom crossing of $S$ and $P_2'$ is an open path connecting the right side of $B_2$ to $P_1'$ and stays inside $B_2$. Then we have the implication

$$\Gamma_1 \cap \mathrm{LR}(S) \cap \Gamma_2' \subseteq \mathrm{LR}(B(3n/2,n)).$$

See Figure 4.6. By symmetry $\mathbb{P}_p[\Gamma_1] = \mathbb{P}_p[\Gamma_2']$. Moreover, the events on the left-hand side are increasing so by the FKG inequality:

**Lemma 4.2.43** (Proof of RSW: step 2)**.**

$$R_{n,3/2}(p) \geq \mathbb{P}_p[\Gamma_1]^2 R_{n/2,1}(p). \tag{4.2.26}$$

**Step 3: bounding $\mathbb{P}_p[\Gamma_1]$** It remains to bound $\mathbb{P}_p[\Gamma_1]$. That requires several additional definitions. Let $P_1$ and $P_2$ be top-bottom crossings of $S$. There is a natural partial order over such crossings. The path $P_1$ divides $S$ into two subgraphs: $[P_1\}$ which includes the left side of $S$ (including edges on the left incident with $P_1$ but not those edges on $P_1$ itself) and $\{P_1]$ which includes the right side of $S$ (and $P_1$ itself). Then we write $P_1 \preceq P_2$ if $\{P_1] \subseteq \{P_2]$. Assuming $\mathrm{TB}(S)$ holds, one also gets the existence of a unique *rightmost crossing*. Roughly speaking, take the union of all top-bottom crossings of $S$ as sets of edges; then the "right boundary" of this set is a top-bottom crossing $P_S^*$ such that $P_S^* \preceq P$ for all top-bottom crossings $P$ of $S$. (We accept as a fact the existence of a unique rightmost crossing. See Exercise 4.11 for a related construction.)

*rightmost crossing*

Let $I_S$ be the set of (not necessarily open) paths connecting the top and bottom of $S$ and stay inside $S$. For $P \in I_S$, we let $P'$ be the reflection of $P$ in $B_{12} \backslash S$ through the $x$-axis and we let $\frac{P}{P'}$ be the union of $P$ and $P'$. Define $[\frac{P}{P'}\}$ to be the

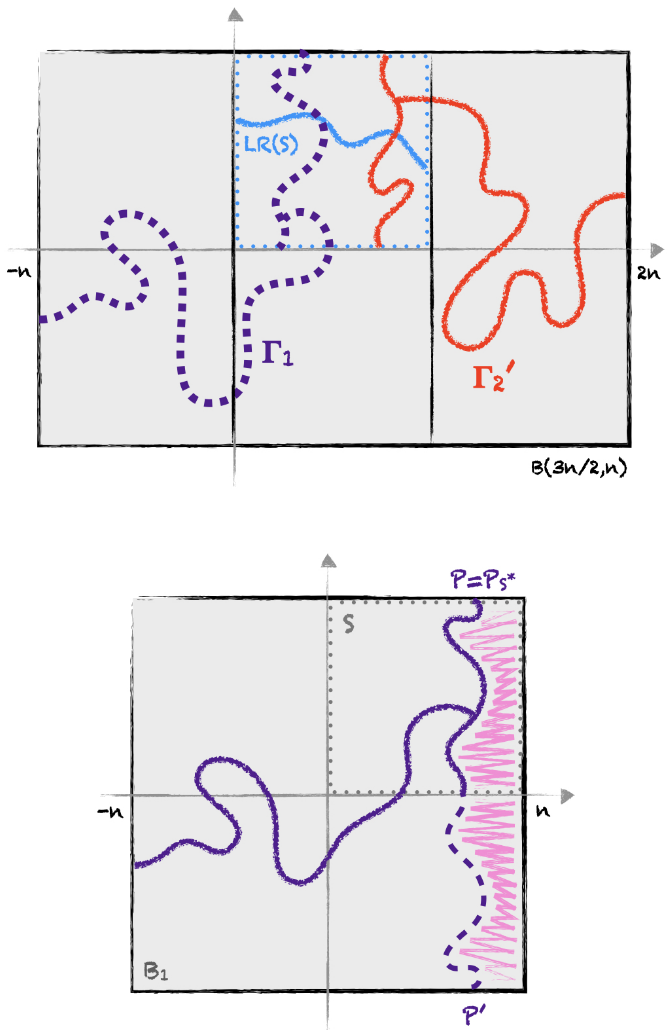


Figure 4.6: Top: illustration of the implication $\Gamma_1 \cap \mathrm{LR}(S) \cap \Gamma_2' \subseteq \mathrm{LR}(B(3n/2, n))$. Bottom: the event $\mathrm{LR}^+\big([\frac{P}{P'}\}\big) \cap \{P = P_S^*\}$; the dashed path is the mirror image of the rightmost top-bottom crossing in $S$; the shaded region on the right is the complement in $B_1$ of the set $[\frac{P}{P'}\}$. Note that, because in the bottom figure the left-right path must stay within $[\frac{P}{P'}\}$ by definition of $P_S^*$, the configuration shown in the top figure where a left-right path (dotted) "travels behind" the top-bottom crossing of $S$ cannot occur.

subgraph of $B_1$ to the left of $\frac{P}{P'}$ (including edges on the left incident with $\frac{P}{P'}$ but not those edges on $\frac{P}{P'}$ itself). Let $\mathrm{LR}^+\big([\frac{P}{P'}\}\big)$ be the event that there is a left-right crossing of $[\frac{P}{P'}\}$ *ending on* $P$, that is, that there is an open path connecting the left side of $B_1$ and $P$ that stays within $[\frac{P}{P'}\}$. See Figure 4.6. Note that the existence of a left-right crossing of $B_1$ implies the existence of an open path connecting the left side of $B_1$ to $\frac{P}{P'}$. By symmetry we then get

$$\mathbb{P}_p\left[\mathrm{LR}^+\big([\tfrac{P}{P'}\}\big)\right] \geq \frac{1}{2}\mathbb{P}_p[\mathrm{LR}(B_1)] = \frac{1}{2}R_{n,1}(p). \tag{4.2.27}$$

Now comes a subtle point. We turn to the rightmost crossing of $S$—for two reasons:

- First, by uniqueness of the rightmost crossing, $\{P^*_S = P\}_{P\in I_S}$ forms a *partition* of $\mathrm{TB}(S)$. Recall that we are looking to bound a probability *from below*, and therefore we have to be careful not to "double count."

- Second, the rightmost crossing has a Markov-like property. Observe that, for $P \in I_S$, the event that $\{P^*_S = P\}$ *depends only the bonds in* $\{P]$. In particular it is *independent of the bonds in* $[\frac{P}{P'}\}$, for example, of the event $\mathrm{LR}^+\big([\frac{P}{P'}\}\big)$. Hence

$$\mathbb{P}_p\left[\mathrm{LR}^+\big([\tfrac{P}{P'}\}\big) \mid P^*_S = P\right] = \mathbb{P}_p\left[\mathrm{LR}^+\big([\tfrac{P}{P'}\}\big)\right]. \tag{4.2.28}$$

  Note that the event $\{P^*_S = P\}$ is *not increasing*, as adding more open bonds can shift the rightmost crossing rightward. Therefore, we cannot use the FKG inequality here.

Combining (4.2.27) and (4.2.28), we get

$$\begin{aligned}
\mathbb{P}_p[\Gamma_1] &\geq \sum_{P\in I_S} \mathbb{P}_p[P^*_S = P]\,\mathbb{P}_p\left[\mathrm{LR}^+\big([\tfrac{P}{P'}\}\big) \mid P^*_S = P\right] \\
&\geq \frac{1}{2}R_{n,1}(p)\sum_{P\in I_S}\mathbb{P}_p[P^*_S = P] \\
&= \frac{1}{2}R_{n,1}(p)\,\mathbb{P}_p[\mathrm{TB}(S)] \\
&= \frac{1}{2}R_{n,1}(p)R_{n/2,1}(p).
\end{aligned}$$

We have proved:

**Lemma 4.2.44** (Proof of RSW: step 3)**.**

$$\mathbb{P}_p[\Gamma_1] \;\geq\; \frac{1}{2}R_{n,1}(p)R_{n/2,1}(p). \tag{4.2.29}$$

**Step 4: putting everything together** Combining (4.2.25), (4.2.26) and (4.2.29) gives that

$$\begin{aligned} R_{n,3}(p) &\geq R_{n,3/2}(p)^4 R_{n,1}(p)^3 \\ &\geq [\mathbb{P}_p[\Gamma_1]^2 R_{n/2,1}(p)]^4 R_{n,1}(p)^3 \\ &\geq \left[\left(\frac{1}{2} R_{n,1}(p) R_{n/2,1}(p)\right)^2 R_{n/2,1}(p)\right]^4 R_{n,1}(p)^3. \end{aligned}$$

Collecting the terms concludes the proof of the RSW lemma. ■

**Remark 4.2.45.** *This argument is quite subtle. It is instructive to read the remark after [Gri97, Theorem 9.3].*

## 4.3 Coupling of Markov chains and application to mixing

As we have seen, coupling is useful to bound total variation distance. In this section we apply the technique to bound the mixing time of Markov chains.

### 4.3.1 Bounding the mixing time via coupling

Let $P$ be an irreducible, aperiodic Markov transition matrix on the finite state space $V$ with stationary distribution $\pi$. Recall from Definition 1.1.35 that, for a fixed $0 < \varepsilon < 1/2$, the mixing time of $P$ is

$$\mathrm{t_{mix}}(\varepsilon) := \min\{t : d(t) \leq \varepsilon\},$$

where

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

It will be easier to work with

$$\bar{d}(t) := \max_{x,y \in V} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{\mathrm{TV}}.$$

The quantities $d(t)$ and $\bar{d}(t)$ are related in the following way.

**Lemma 4.3.1.**

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

*Proof.* The second inequality follows from an application of the triangle inequality.

For the first inequality, note that by definition of the total variation distance and the stationarity of $\pi$

$$\begin{aligned}
\|P^t(x,\cdot)-\pi\|_{\mathrm{TV}} &= \sup_{A\subseteq V}|P^t(x,A)-\pi(A)| \\
&= \sup_{A\subseteq V}\left|\sum_{y\in V}\pi(y)[P^t(x,A)-P^t(y,A)]\right| \\
&\leq \sup_{A\subseteq V}\sum_{y\in V}\pi(y)|P^t(x,A)-P^t(y,A)| \\
&\leq \sum_{y\in V}\pi(y)\left\{\sup_{A\subseteq V}|P^t(x,A)-P^t(y,A)|\right\} \\
&\leq \sum_{y\in V}\pi(y)\|P^t(x,\cdot)-P^t(y,\cdot)\|_{\mathrm{TV}} \\
&\leq \max_{x,y\in V}\|P^t(x,\cdot)-P^t(y,\cdot)\|_{\mathrm{TV}}.
\end{aligned}$$

■

**Coalescence** Recall that a Markovian coupling of $P$ with itself is a Markov chain $(X_t, Y_t)_t$ on $V\times V$ with transition matrix $Q$ satisfying: for all $x, y, x', y' \in V$,

$$\sum_{z'} Q((x,y),(x',z')) = P(x,x'),$$

$$\sum_{z'} Q((x,y),(z',y')) = P(y,y').$$

We say that a Markovian coupling is *coalescing* if further: for all $z\in V$, *coalescing*

$$x'\neq y' \implies Q((z,z),(x',y'))=0.$$

Let $(X_t, Y_t)$ be a coalescing Markovian coupling of $P$. By the coalescing condition, if $X_s = Y_s$ then $X_t = Y_t$ for all $t \geq s$. That is, once $(X_t)$ and $(Y_t)$ meet, they remain equal. Let $\tau_{\mathrm{coal}}$ be the *coalescence time* (also called *coupling time*), that is, *coalescence time*

$$\tau_{\mathrm{coal}} := \inf\{t\geq 0 : X_t = Y_t\}.$$

The key to the coupling approach to mixing times is the following immediate consequence of the coupling inequality (Lemma 4.1.11). For any starting point $(x, y)$,

$$\|P^t(x,\cdot)-P^t(y,\cdot)\|_{\mathrm{TV}} \leq \mathbb{P}_{(x,y)}[X_t\neq Y_t] = \mathbb{P}_{(x,y)}[\tau_{\mathrm{coal}} > t]. \tag{4.3.1}$$

Combining (4.3.1) and Lemma 4.3.1, we get the main tool of this section.

**Theorem 4.3.2** (Bounding the mixing time: coupling method)**.** *Let $(X_t, Y_t)$ be a coalescing Markovian coupling of an irreducible transition matrix $P$ on a finite state space $V$ with stationary distribution $\pi$. Then*

$$d(t) \leq \max_{x,y \in V} \mathbb{P}_{(x,y)}[\tau_{\mathrm{coal}} > t].$$

*In particular*

$$t_{\mathrm{mix}}(\varepsilon) \leq \inf\left\{t \geq 0 \,:\, \mathbb{P}_{(x,y)}[\tau_{\mathrm{coal}} > t] \leq \varepsilon,\ \forall x, y\right\}.$$

Note that a Markovian coupling can be made coalescing by modifying it as follows: when $X_t = Y_t$, perform one step of the chain to determine $X_{t+1}$ and set $Y_{t+1} := X_{t+1}$. That modification does not affect the coalescence time.

We give a few simple examples of the coupling method in the next subsection. First, we discuss a classical result.

**Example 4.3.3** (Doeblin's condition)**.** Let $P$ be a transition matrix on a countable space $V$. One form of *Doeblin's condition* (also called a *minorization condition*) is: there is $s \in \mathbb{Z}_+$ and $\delta > 0$ such that

*Doeblin's condition*

$$\sup_{z \in V} \inf_{w \in V} P^s(w, z) > \delta.$$

In words there is a state $z_0 \in V$ such that, starting from any state $w \in V$, the probability of reaching $z_0$ in exactly $s$ steps is at least $\delta$ (which does not depend on $w$). Assume such a $z_0$ exists.

We construct a coalescing Markovian coupling $(X_t, Y_t)$ of $P$. Assume first that $s = 1$ and let

$$\tilde{P}(w, z) = \frac{1}{1-\delta}\left[P(w, z) - \delta \mathbf{1}\{z = z_0\}\right].$$

It can be checked that $\tilde{P}$ is a stochastic matrix on $V$ provided $z_0$ satisfies the condition above (see Exercise 4.13). We use a technique known as *splitting*. While $X_t \neq Y_t$, at the next time step: (i) with probability $\delta$ we set $X_{t+1} = Y_{t+1} = z_0$, (ii) otherwise we pick $X_{t+1} \sim \tilde{P}(X_t, \cdot\,)$ and $Y_{t+1} \sim \tilde{P}(Y_t, \cdot\,)$ independently. On the other hand, if $X_t = Y_t$, we maintain the equality and pick the next state according to $P$. Put differently, the coupling $Q$ is defined as: if $x \neq y$,

*splitting*

$$Q((x, y), (x', y')) = \delta \mathbf{1}\{x' = y' = z_0\} + (1-\delta)\tilde{P}(x, x')\tilde{P}(y, y'),$$

while if $x = y$,

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

Observe that, in case (i) above, coalescence occurs at time $t + 1$. In case (ii), coalescence may or may not occur at time $t + 1$. In other words, while $X_t \neq$

$Y_t$, coalescence occurs at the next step with probability *at least* $\delta$. So $\tau_{\mathrm{coal}}$ is stochastically dominated by a geometric random variable with success probability $\delta$, or

$$\max_{x,y\in V} \mathbb{P}_{(x,y)}[\tau_{\mathrm{coal}} > t] \leq (1-\delta)^t.$$

By Theorem 4.3.2,

$$\max_{x\in V} \|P^t(x,\cdot) - \pi\|_{\mathrm{TV}} \leq (1-\delta)^t.$$

Exponential decay of the worst-case total variation distance to the stationary distribution is referred to as *uniform geometric ergodicity*.

*uniform geometric ergodicity*

Suppose now that $s > 1$. We apply the argument above to the chain $P^s$ this time. We get

$$\max_{x,y\in V} \mathbb{P}_{(x,y)}[\tau_{\mathrm{coal}} > ts] \leq (1-\delta)^t,$$

so that, after a change of variable,

$$\max_{x\in V} \|P^t(x,\cdot) - \pi\|_{\mathrm{TV}} \leq (1-\delta)^{\lfloor t/s \rfloor}.$$

So, we have shown that uniform geometric ergodicity is implied by Doeblin's condition.

We note however that the rate of decay derived from this technique can be very slow. For instance the condition always holds when $P$ is finite, irreducible and aperiodic (as follows from Lemma 1.1.32), but a straight application of the technique may lead to a bound depending badly on the size of the state space $V$ (see Exercise 4.14). ◀

### 4.3.2 ▷ *Random walks: mixing on cycles, hypercubes, and trees*

In this section, we consider lazy simple random walk on various graphs. By this we mean that the walk stays put with probability $1/2$ and otherwise picks an adjacent vertex uniformly at random. In each case, we construct a coupling to bound the mixing time. As a reference, we compare our upper bounds to the diameter-based lower bound we will derive in Section 5.2.3. Specifically, by Claim 5.2.25, for a finite, reversible Markov chain with stationary distribution $\pi$ and diameter $\Delta$ we have the lower bound

$$t_{\mathrm{mix}}(\varepsilon) = \Omega\left(\frac{\Delta^2}{\log(n \vee \pi_{\min}^{-1})}\right),$$

where $\pi_{\min}$ is the smallest value taken by $\pi$.

**Cycle**

Let $(Z_t)$ be lazy simple random walk on the cycle of size $n$, $\mathbb{Z}_n := \{0, 1 \ldots, n-1\}$, where $i \sim j$ if $|j - i| = 1 \pmod{n}$. For any starting points $x, y$, we construct a Markovian coupling $(X_t, Y_t)$ of this chain. Set $(X_0, Y_0) := (x, y)$. At each time, flip a fair coin. On heads, $Y_t$ stays put and $X_t$ moves one step, the direction of which is uniform at random. On tails, proceed similarly with the roles of $X_t$ and $Y_t$ reversed. Let $D_t$ be the clockwise distance between $X_t$ and $Y_t$. Observe that, by construction, $(D_t)$ is simple random walk on $\{0, \ldots, n\}$ and $\tau_{\mathrm{coal}} = \tau^D_{\{0,n\}}$, the first time $(D_t)$ hits $\{0, n\}$.

We use Markov's inequality (Theorem 2.1.1) to bound $\mathbb{P}_{(x,y)}[\tau^D_{\{0,n\}} > t]$. Denote by $D_0 = d_{x,y}$ the starting distance. By Wald's second identity (Theorem 3.1.40),

$$\mathbb{E}_{(x,y)}\left[\tau^D_{\{0,n\}}\right] = d_{x,y}(n - d_{x,y}).$$

Applying Theorem 4.3.2 and Markov's inequality we get

$$\begin{aligned} d(t) &\leq \max_{x,y \in V} \mathbb{P}_{(x,y)}[\tau_{\mathrm{coal}} > t] \\ &\leq \max_{x,y \in V} \frac{\mathbb{E}_{(x,y)}\left[\tau^D_{\{0,n\}}\right]}{t} \\ &= \max_{x,y \in V} \frac{d_{x,y}(n - d_{x,y})}{t} \\ &\leq \frac{n^2}{4t}, \end{aligned}$$

or:

**Claim 4.3.4.**

$$\mathrm{t_{mix}}(\varepsilon) \leq \frac{n^2}{4\varepsilon}.$$

By the diameter-based lower bound on mixing in Section 5.2.3, this bound gives the correct order of magnitude in $n$ up to logarithmic factors. Indeed, the diameter is $\Delta = n/2$ and $\pi_{\min} = 1/n$ so that Claim 5.2.25 gives

$$\mathrm{t_{mix}}(\varepsilon) \geq \frac{n^2}{64 \log n},$$

for $n$ large enough. Exercise 4.15 sketches a tighter lower bound.

### Hypercube

Let $(Z_t)_{t\in\mathbb{Z}_+}$ be lazy simple random walk on the $n$-dimensional hypercube $\mathbb{Z}_2^n := \{0,1\}^n$ where $i \sim j$ if $\|i-j\|_1 = 1$. We denote the coordinates of $Z_t$ by $(Z_t^{(1)},\ldots,Z_t^{(n)})$. This is equivalent to performing the Glauber dynamics chain on an empty graph (see Definition 1.2.8): at each step, we first pick a coordinate uniformly at random, then refresh its value. Because of the way the updating is done, the chain stays put with probability $1/2$ at each time as required.

Inspired by this observation, the coupling $(X_t, Y_t)$ started at $(x,y)$ is the following. At each time $t$, pick a coordinate $i$ uniformly at random in $[n]$, pick a bit value $b$ in $\{0,1\}$ uniformly at random independently of the coordinate choice. Set *both* $i$ coordinates to $b$, that is, $X_t^{(i)} = Y_t^{(i)} = b$. By design we reach coalescence when all coordinates have been updated at least once.

The following standard bound from the coupon collector's problem (see Example 2.1.4) is what is needed to conclude.

**Lemma 4.3.5.** *Let $\tau_{\mathrm{coll}}$ be the time it takes to update each coordinate at least once. Then, for any $c>0$,*

$$\mathbb{P}\left[\tau_{\mathrm{coll}} > \lceil n\log n + cn\rceil\right] \le e^{-c}.$$

*Proof.* Let $B_i$ be the event that the $i$-th coordinate has not been updated by time $\lceil n\log n + cn\rceil$. Then, using that $1-x \le e^{-x}$ for all $x$ (see Exercise 1.16),

$$\begin{aligned}
\mathbb{P}[\tau_{\mathrm{coll}} > \lceil n\log n + cn\rceil] &\le \sum_i \mathbb{P}[B_i]\\
&= \sum_i \left(1-\frac{1}{n}\right)^{\lceil n\log n + cn\rceil}\\
&\le n\exp\left(-\frac{n\log n + cn}{n}\right)\\
&= e^{-c}.
\end{aligned}$$

∎

Applying Theorem 4.3.2, we get

$$\begin{aligned}
d(\lceil n\log n + cn\rceil) &\le \max_{x,y\in V}\mathbb{P}_{(x,y)}[\tau_{\mathrm{coal}} > \lceil n\log n + cn\rceil]\\
&\le \mathbb{P}[\tau_{\mathrm{coll}} > \lceil n\log n + cn\rceil]\\
&\le e^{-c}.
\end{aligned}$$

Hence for $c_\varepsilon > 0$ large enough:

**Claim 4.3.6.**

$$\mathrm{t_{mix}}(\varepsilon) \leq \lceil n \log n + c_\varepsilon n \rceil.$$

Again we get a quick lower bound using the diameter-based result from Section 5.2.3. Here $\Delta = n$ and $\pi_{\min} = 1/2^n$ so that Claim 5.2.25 gives

$$\mathrm{t_{mix}}(\varepsilon) \geq \frac{n^2}{12 \log n + (4 \log 2) n} = \Omega(n),$$

for $n$ large enough. So the upper bound we derived above is off at most by a logarithmic factor in $n$. In fact:

**Claim 4.3.7.**

$$\mathrm{t_{mix}}(\varepsilon) \geq \frac{1}{2} n \log n - O(n).$$

*Proof.* For simplicity, we assume that $n$ is odd. Let $W_t$ be the number of 1s, or *Hamming weight*, at time $t$. Let $A$ be the event that the Hamming weight is $\leq n/2$. To bound the mixing time, we use the fact that for any $z_0$

*Hamming weight*

$$d(t) \geq \|P^t(z_0, \cdot) - \pi\|_{\mathrm{TV}} \geq |P^t(z_0, A) - \pi(A)|. \tag{4.3.2}$$

Under the stationary distribution, the Hamming weight is equal in distribution to a $\mathrm{Bin}(n, 1/2)$. In particular the probability that a majority of coordinates are $0$ is $1/2$. That is, $\pi(A) = 1/2$.

On the other hand, let $(Z_t)$ start at $z_0$, the all-$1$ vector. By the definition of $A$,

$$|P^t(z_0, A) - \pi(A)| = |\mathbb{P}[W_t \leq n/2] - 1/2|. \tag{4.3.3}$$

We use Chebyshev's inequality (Theorem 2.1.2) to bound the probability on the right-hand side. So we need to compute the expectation and variance of $W_t$.

Let $U_t$ be the number of (distinct) updated coordinates up to time $t$ in the Glauber dynamics representation of the chain discussed above. Observe that, conditioned on $U_t$, the Hamming weight $W_t$ is equal in distribution to $\mathrm{Bin}(U_t, 1/2) + (n - U_t)$ as the updated coordinates are uniform and the other ones are $1$. Thus we have

$$\begin{aligned}\mathbb{E}[W_t] &= \mathbb{E}[\mathbb{E}[W_t \mid U_t]] \\ &= \mathbb{E}\left[\frac{1}{2}U_t + (n - U_t)\right] \\ &= \mathbb{E}\left[n - \frac{1}{2}U_t\right] \\ &= n - \frac{1}{2}\, n\left[1 - \left(1 - \frac{1}{n}\right)^t\right] \\ &= \frac{n}{2}\left[1 + \left(1 - \frac{1}{n}\right)^t\right], \end{aligned} \tag{4.3.4}$$

where on the fourth line we used the fact that $\mathbb{E}[U_t] = n\left[1 - \left(1 - \frac{1}{n}\right)^t\right]$ by summing over the coordinates and using linearity of expectation.

As to the variance, using again the observation above about the distribution of $W_t$ given $U_t$,

$$\begin{aligned}\mathrm{Var}[W_t] &= \mathbb{E}[\mathrm{Var}[W_t \mid U_t]] + \mathrm{Var}[\mathbb{E}[W_t \mid U_t]] \\ &= \frac{1}{4}\mathbb{E}\left[U_t\right] + \frac{1}{4}\mathrm{Var}[U_t]. \end{aligned} \tag{4.3.5}$$

It remains to compute $\mathrm{Var}[U_t]$. Let $I_t^{(i)}$ be $1$ if coordinate $i$ has *not* been updated up to time $t$ and $0$ otherwise. Note that for $i \neq j$

$$\begin{aligned}\mathrm{Cov}[I_t^{(i)}, I_t^{(j)}] &= \mathbb{E}[I_t^{(i)} I_t^{(j)}] - \mathbb{E}[I_t^{(i)}]\mathbb{E}[I_t^{(j)}] \\ &= \left(1 - \frac{2}{n}\right)^t - \left(1 - \frac{1}{n}\right)^{2t} \\ &= \left(1 - \frac{2}{n}\right)^t - \left(1 - \frac{2}{n} + \frac{1}{n^2}\right)^t \\ &\leq 0, \end{aligned}$$

that is, $I_t^{(i)}$ and $I_t^{(j)}$ are negatively correlated, while

$$\mathrm{Var}[I_t^{(i)}] = \mathbb{E}[(I_t^{(i)})^2] - (\mathbb{E}[I_t^{(i)}])^2 \leq \mathbb{E}[I_t^{(i)}] = \left(1 - \frac{1}{n}\right)^t.$$

Then, writing $n - U_t$ as the sum of these indicators, we have

$$\begin{aligned}\mathrm{Var}[U_t] &= \mathrm{Var}[n - U_t] \\ &= \sum_{i=1}^{n} \mathrm{Var}[I_t^{(i)}] + 2\sum_{i<j} \mathrm{Cov}[I_t^{(i)}, I_t^{(j)}] \\ &\leq n\left(1 - \frac{1}{n}\right)^t.\end{aligned}$$

Plugging this back into (4.3.5), we get

$$\mathrm{Var}[W_t] \leq \frac{n}{4}\left[1 - \left(1 - \frac{1}{n}\right)^t\right] + \frac{n}{4}\left(1 - \frac{1}{n}\right)^t = \frac{n}{4}.$$

For $t_\alpha = \frac{1}{2} n \log n - n \log \alpha$ with $\alpha > 0$, by (4.3.4),

$$\mathbb{E}[W_{t_\alpha}] = \frac{n}{2} + \frac{n}{2} e^{t_\alpha(-1/n + \Theta(1/n^2))} = \frac{n}{2} + \frac{\alpha}{2}\sqrt{n} + o(1),$$

where we used that by a Taylor expansion, for $|z| \leq 1/2$, $\log(1 - z) = -z + \Theta(z^2)$. Fix $0 < \varepsilon < 1/2$. By Chebyshev's inequality, for $t_\alpha = \frac{1}{2} n \log n - n \log \alpha$ and $n$ large enough,

$$\mathbb{P}[W_{t_\alpha} \leq n/2] \leq \mathbb{P}[|W_{t_\alpha} - \mathbb{E}[W_{t_\alpha}]| \geq (\alpha/2)\sqrt{n}] \leq \frac{n/4}{(\alpha/2)^2 n} \leq \frac{1}{2} - \varepsilon,$$

for $\alpha$ large enough. By (4.3.2) and (4.3.3), that implies $d(t_\alpha) \geq \varepsilon$ and we are done. ■

The previous proof relies on a "distinguishing statistic." Recall from Lemma 4.1.19 that for any random variables $X, Y$ and mapping $h$ it holds that

$$\|\mu_{h(X)} - \mu_{h(Y)}\|_{\mathrm{TV}} \leq \|\mu_X - \mu_Y\|_{\mathrm{TV}},$$

where $\mu_Z$ is the law of $Z$. The mapping used in the proof of the claim is the Hamming weight. In essence, we gave a lower bound on the total variation distance between the laws of the Hamming weight at stationarity and under $P^t(z_0, \cdot)$. See Exercise 4.16 for a more general treatment of the distinguishing statistic approach.

**Remark 4.3.8.** *The upper bound in Claim 4.3.6 is indeed off by a factor of* 2*. See [LPW06, Theorem 18.3] for an improved upper bound and a discussion of the so-called* cutoff phenomenon*. The latter refers to the fact that for all* $0 < \varepsilon < 1/2$ *it can be shown in this case that* *cutoff*

$$\lim_{n \to +\infty} \frac{\mathrm{t}_{\mathrm{mix}}^{(n)}(\varepsilon)}{\mathrm{t}_{\mathrm{mix}}^{(n)}(1 - \varepsilon)} = 1,$$

*where* $\mathrm{t}_{\mathrm{mix}}^{(n)}(\varepsilon)$ *is the mixing time on the* $n$*-dimensional hypercube. In words, for large* $n$*, the total variation distance drops from* 1 *to* 0 *in a short time window. See Exercise 5.10 for a necessary condition for cutoff.*

**$b$-ary tree**

Let $(Z_t)_{t\in\mathbb{Z}_+}$ be lazy simple random walk on the $\ell$-level rooted $b$-ary tree, $\widehat{\mathbb{T}}_b^\ell$, with $\ell \geq 2$. The root, $0$, is on level $0$ and the leaves, $L$, are on level $\ell$. All vertices have degree $b+1$, except for the root which has degree $b$ and the leaves which have degree $1$. By Example 1.1.29 (noting that laziness makes no difference), the stationary distribution is

$$\pi(x) := \frac{\delta(x)}{2(n-1)},$$

where $n$ is the number of vertices and $\delta(x)$ is the degree of $x$. We used that a tree on $n$ vertices has $n-1$ edges (Corollary 1.1.7). We construct a coupling $(X_t, Y_t)$ of this chain started at $(x, y)$. Assume without loss of generality that $x$ is no further from the root than $y$, which we denote by $x \preccurlyeq y$ (which, here, does *not* mean that $y$ is a descendant of $x$). The coupling has two stages:

- In the first stage, at each time, flip a fair coin. On heads, $Y_t$ stays put and $X_t$ moves one step chosen uniformly at random among its neighbors. Similarly, on tails, reverse the roles of $X_t$ and $Y_t$. Do this until $X_t$ and $Y_t$ are on the same level.

- In the second stage, that is, once the two chains are on the same level, at each time first let $X_t$ move as a lazy simple random walk on $\widehat{\mathbb{T}}_b^\ell$. Then let $Y_t$ move in the same direction as $X_t$, that is, if $X_t$ moves closer to the root, so does $Y_t$ and so on.

By construction, $X_t \preccurlyeq Y_t$ for all $t$. The key observation is the following. Let $\tau^*$ be the first time $(X_t)$ visits the root *after visiting the leaves*. By time $\tau^*$, the two chains have necessarily met: because $X_t \preccurlyeq Y_t$, when $X_t$ reaches the leaves, so does $Y_t$; after that time, the coupling is in the second stage so $X_t$ and $Y_t$ remain on the same level; in particular, when $X_t$ reaches the root (after visiting the leaves), so does $Y_t$. Hence $\tau_{\mathrm{coal}} \leq \tau^*$. Intuitively, the mixing time is indeed dominated by the time it takes to reach the root from the worst starting point, a leaf. See Figure 4.7 and the corresponding lower bound argument.

To estimate $\mathbb{P}_{(x,y)}[\tau^* > t]$, we use Markov's inequality (Theorem 2.1.1), for which we need a bound on $\mathbb{E}_{(x,y)}[\tau^*]$. We note that $\mathbb{E}_{(x,y)}[\tau^*]$ is less than the mean time for the walk to go from the root to the leaves and back. Let $L_t$ be the level of $X_t$ and let $\mathcal{N}$ be the corresponding network (where the conductances are equal to the number of edges on each level of the tree). In terms of $L_t$, the quantity we seek to bound is the mean of $\tau_{0,\ell}$, the commute time of the chain $(L_t)$ between the states $0$ and $\ell$. By the commute time identity (Theorem 3.3.34),

$$\mathbb{E}[\tau_{0,\ell}] = c_{\mathcal{N}}\, \mathscr{R}(0 \leftrightarrow \ell), \tag{4.3.6}$$

where

$$c_{\mathcal{N}} = 2 \sum_{e=\{x,y\}\in\mathcal{N}} c(e) = 4(n-1),$$

where we simply counted the number of edges in $\widehat{\mathbb{T}}_b^\ell$ and the extra factor of 2 accounts for self-loops. Using network reduction techniques, we computed the effective resistance $\mathscr{R}(0 \leftrightarrow \ell)$ in Examples 3.3.21 and 3.3.22—without self-loops. Of course adding self-loops does not affect the effective resistance as we can use the same voltage and current. So, ignoring them, we get

$$\mathscr{R}(0 \leftrightarrow \ell) = \sum_{j=0}^{\ell-1} r(j, j+1) = \sum_{j=0}^{\ell-1} b^{-(j+1)} = \frac{1}{b} \cdot \frac{1-b^{-\ell}}{1-b^{-1}}, \tag{4.3.7}$$

which implies

$$\frac{1}{b} \leq \mathscr{R}(0 \leftrightarrow \ell) \leq \frac{1}{b-1} \leq 1.$$

Finally, applying Theorem 4.3.2 and Markov's inequality and using (4.3.6), we get

$$\begin{aligned} d(t) &\leq \max_{x,y\in V} \mathbb{P}_{(x,y)}[\tau^* > t] \\ &\leq \max_{x,y\in V} \frac{\mathbb{E}_{(x,y)}[\tau^*]}{t} \\ &\leq \frac{\mathbb{E}[\tau_{0,\ell}]}{t} \\ &\leq \frac{4n}{t}, \end{aligned}$$

or:

**Claim 4.3.9.**

$$\mathrm{t_{mix}}(\varepsilon) \leq \frac{4n}{\varepsilon}.$$

This time the diameter-based bound is far off. We have $\Delta = 2\ell = \Theta(\log n)$ and $\pi_{\min} = 1/2(n-1)$ so that Claim 5.2.25 gives

$$\mathrm{t_{mix}}(\varepsilon) \geq \frac{(2\ell)^2}{12\log n + 4\log(2(n-1))} = \Omega(\log n),$$

for $n$ large enough.

Here is a better lower bound. We take $b = 2$ to simplify. Intuitively the mixing time is significantly greater than the squared diameter because the chain tends to be pushed away from the root. Consider the time it takes to go from the leaves

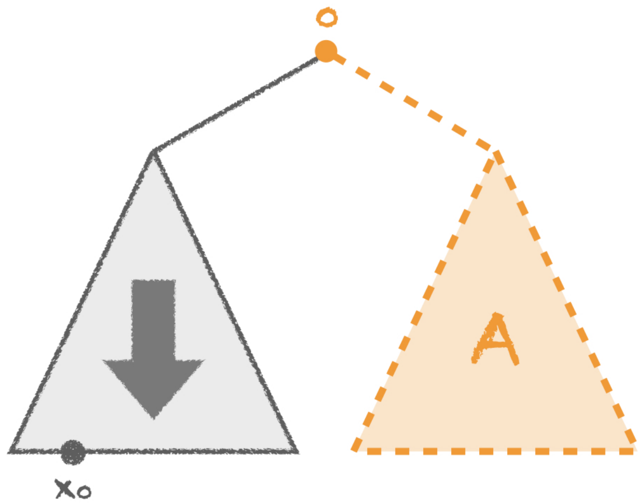


Figure 4.7: Setup for the lower bound on the mixing time on a $b$-ary tree. (Here $b = 2$.)

on one side of the root to the leaves on the other, both of which have substantial weight under the stationary distribution. That typically takes time exponential in the diameter—that is, linear in $n$. Indeed one first has to reach the root, which by the gambler's ruin problem (Example 3.1.43), takes an exponential in $\ell$ number of "excursions" (see Claim 3.1.44 (ii)).

Formally let $x_0$ be a leaf of $\widehat{\mathbb{T}}_b^\ell$ and let $A$ be the set of vertices "on the other side of root (inclusively)," that is, vertices whose graph distance from $x_0$ is at least $\ell$. See Figure 4.7. Then $\pi(A) \geq 1/2$ by symmetry. We use the fact that

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

to bound the mixing time from below. We claim that, started at $x_0$, the walk takes time linear in $n$ to reach $A$ with nontrivial probability.

Consider again the level $L_t$ of $X_t$. Using definition of the effective resistance (Definition 3.3.19) as well as the expression for it in (4.3.7), we have

$$\mathbb{P}_\ell[\tau_0 < \tau_\ell^+] = \frac{1}{c(\ell)\,\mathscr{R}(0 \leftrightarrow \ell)} = \frac{1}{b^\ell} \cdot \frac{b-1}{1-b^{-\ell}} = \frac{b-1}{b^\ell - 1} = O\left(\frac{1}{n}\right).$$

Hence, started from the leaves, the number of excursions back to the leaves needed to reach the root for the first time is geometric with success probability $O(n^{-1})$. Each such excursion takes time at least 2 (which corresponds to going right back to the leaves after the first step). So $P^t(x_0, A)$ is bounded above by the probability that at least one such excursion was successful among the first $t/2$ attempts. That is,

$$P^t(x_0, A) \leq 1 - \left(1 - O\left(n^{-1}\right)\right)^{t/2} < \frac{1}{2} - \varepsilon,$$

for all $t \leq \alpha_\varepsilon n$ with $\alpha_\varepsilon > 0$ small enough and

$$\|P^{\alpha_\varepsilon n}(x_0, \cdot) - \pi\|_{\mathrm{TV}} \geq |P^{\alpha_\varepsilon n}(x_0, A) - \pi(A)| > \varepsilon.$$

We have proved that $\mathrm{t_{mix}}(\varepsilon) \geq \alpha_\varepsilon n$.

### 4.3.3 Path coupling

*Path coupling* is a method for constructing Markovian couplings from "simpler" couplings. The building blocks are one-step couplings starting from pairs of initial states that are close in some "dissimilarity graph."

Let $(X_t)$ be an irreducible Markov chain on a finite state space $V$ with transition matrix $P$ and stationary distribution $\pi$. Assume that we have a *dissimilarity graph* $H_0 = (V_0, E_0)$ on $V_0 := V$ with edge weights $w_0 : E_0 \to \mathbb{R}_+$. This graph need not have the same edges as the transition graph of $(X_t)$. We extend $w_0$ to the *path metric*

*dissimilarity graph, path metric*

$$w_0(x, y) := \inf\left\{\sum_{i=0}^{m-1} w_0(x_i, x_{i+1}) \,:\, x = x_0, x_1, \ldots, x_m = y \text{ is a path in } H_0\right\},$$

where the infimum is over all paths connecting $x$ and $y$ in $H_0$. We call a path achieving the infimum a *minimum-weight path*. It is straightforward to check that $w_0$ satisfies the triangle inequality. Let

$$\Delta_0 := \max_{x,y} w_0(x, y),$$

be the *weighted diameter* of $H_0$.

**Theorem 4.3.10** (Path coupling method)**.** *Assume that*

$$w_0(u, v) \geq 1,$$

*for all $\{u, v\} \in E_0$. Assume further that there exists $\kappa \in (0, 1)$ such that:*

- (Local couplings) *For all $x, y$ with $\{x, y\} \in E_0$, there is a coupling $(X^*, Y^*)$ of $P(x, \cdot)$ and $P(y, \cdot)$ satisfying the following* contraction *property*

$$\mathbb{E}[w_0(X^*, Y^*)] \leq \kappa\, w_0(x, y). \tag{4.3.8}$$

*Then*

$$d(t) \leq \Delta_0 \, \kappa^t,$$

*or*

$$\mathrm{t}_{\mathrm{mix}}(\varepsilon) \leq \left\lceil \frac{\log \Delta_0 + \log \varepsilon^{-1}}{\log \kappa^{-1}} \right\rceil.$$

*Proof.* The crux of the proof is to extend (4.3.8) to arbitrary pairs of vertices.

**Claim 4.3.11** (Global coupling)**.** *For all $x, y \in V$, there is a coupling $(X^*, Y^*)$ of $P(x, \cdot)$ and $P(y, \cdot)$ such that* (4.3.8) *holds.*

Iterating the coupling in this last claim immediately implies the existence of a coalescing Markovian coupling $(X_t, Y_t)$ of $P$ such that

$$\begin{aligned}
\mathbb{E}_{(x,y)}[w_0(X_t, Y_t)] &= \mathbb{E}_{(x,y)}\left[\mathbb{E}[w_0(X_t, Y_t) \,|\, X_{t-1}, Y_{t-1}]\right] \\
&\leq \mathbb{E}_{(x,y)}\left[\kappa \, w_0(X_{t-1}, Y_{t-1})\right] \\
&\leq \cdots \\
&\leq \kappa^t \, \mathbb{E}_{(x,y)}[w_0(X_0, Y_0)] \\
&= \kappa^t \, w_0(x, y) \\
&\leq \kappa^t \, \Delta_0.
\end{aligned}$$

By assumption, $\mathbf{1}_{\{x \neq y\}} \leq w_0(x, y)$ so that by the coupling inequality and Lemma 4.3.1, we have

$$d(t) \leq \bar{d}(t) \leq \max_{x,y} \mathbb{P}_{(x,y)}[X_t \neq Y_t] \leq \max_{x,y} \mathbb{E}_{(x,y)}[w_0(X_t, Y_t)] \leq \kappa^t \, \Delta_0,$$

which implies the theorem.

**Remark 4.3.12.** *In essence, $w_0$ satisfies a form of Lyapounov condition (i.e.,* (3.3.15)*) with a "geometric drift." See, e.g., [MT09, Chapter 15].*

It remains to prove Claim 4.3.11.

*Proof of Claim 4.3.11.* Fix $x', y' \in V$ such that $\{x', y'\}$ is *not* an edge in the dissimilarity graph $H_0$. The idea is to combine the local couplings on a minimum-weight path between $x'$ and $y'$ in $H_0$. Let $x' = x_0 \sim \cdots \sim x_m = y'$ be such a path. For all $i = 0, \ldots, m-1$, let $(Z^*_{i,0}, Z^*_{i,1})$ be a coupling of $P(x_i, \cdot)$ and $P(x_{i+1}, \cdot)$ satisfying the contraction property (4.3.8).

Then we proceed as follows. Set $Z^{(0)} := Z^*_{0,0}$ and $Z^{(1)} := Z^*_{0,1}$. Then iteratively pick $Z^{(i+1)}$ according to the law $\mathbb{P}[Z^*_{i,1} \in \cdot \,|\, Z^*_{i,0} = Z^{(i)}]$. By induction on $i$, $(X^*, Y^*) := (Z^{(0)}, Z^{(m)})$ is then a coupling of $P(x', \cdot)$ and $P(y', \cdot)$. See Figure 4.8.

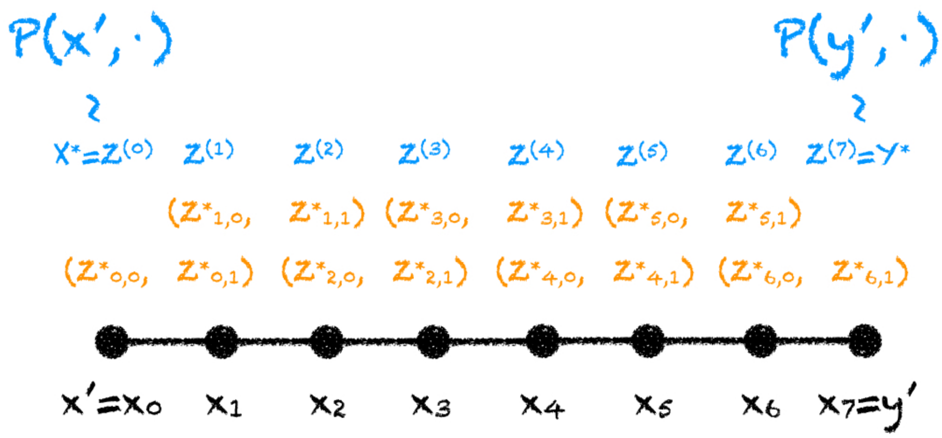


Figure 4.8: Coupling of $P(x', \cdot)$ and $P(y', \cdot)$ constructed from a sequence of local couplings $(Z^*_{0,0}, Z^*_{0,1}), \ldots, (Z^*_{0,m-1}, Z^*_{0,m-1})$.

To be more formal, define the transition matrix

$$R_i(z^{(i)}, z^{(i+1)}) := \mathbb{P}[Z^*_{i,1} = z^{(i+1)} \,|\, Z^*_{i,0} = z^{(i)}].$$

Observe that

$$\sum_{z^{(i+1)}} R_i(z^{(i)}, z^{(i+1)}) = 1, \tag{4.3.9}$$

and

$$\sum_{z^{(i)}} P(x_i, z^{(i)}) R_i(z^{(i)}, z^{(i+1)}) = P(x_{i+1}, z^{(i+1)}), \tag{4.3.10}$$

by construction of the coupling $(Z^*_{i,0}, Z^*_{i,1})$ and the definition of $R_i$. The law of the full coupling

$$(Z^{(0)}, \ldots, Z^{(m)})$$

is

$$\begin{aligned}&\mathbb{P}[(Z^{(0)}, \ldots, Z^{(m)}) = (z^{(0)}, \ldots, z^{(m)})]\\ &\qquad = P(x_0, z^{(0)}) R_0(z^{(0)}, z^{(1)}) \cdots R_{m-1}(z^{(m-1)}, z^{(m)}).\end{aligned}$$

Using (4.3.9) and (4.3.10) inductively gives respectively

$$\begin{aligned}\mathbb{P}[X^* = z^{(0)}] = \mathbb{P}[Z^{(0)} = z^{(0)}] = P(x_0, z^{(0)}),\\ \mathbb{P}[Y^* = z^{(m)}] = \mathbb{P}[Z^{(m)} = z^{(m)}] = P(x_m, z^{(m)}),\end{aligned}$$

as required.

By the triangle inequality for $w_0$, the coupling $(X^*, Y^*)$ satisfies

$$\begin{aligned}\mathbb{E}[w_0(X^*, Y^*)] &= \mathbb{E}\left[w_0(Z^{(0)}, Z^{(m)})\right]\\ &\leq \sum_{i=0}^{m-1} \mathbb{E}\left[w_0(Z^{(i)}, Z^{(i+1)})\right]\\ &\leq \sum_{i=0}^{m-1} \kappa\, w_0(x_i, x_{i+1})\\ &= \kappa\, w_0(x', y'),\end{aligned}$$

where, on the third line, we used (4.3.8) for adjacent pairs and the last line follows from the fact that we chose a minimum-weight path. ■

That concludes the proof of the theorem. ■

We illustrate the path coupling method in the next subsection. See Exercise 4.17 for an optimal transport perspective on the path coupling method.

### 4.3.4 ▷ *Ising model: Glauber dynamics at high temperature*

Let $G = (V, E)$ be a finite, connected graph with maximal degree $\bar{\delta}$. Define $\mathcal{X} := \{-1, +1\}^V$. Recall from Example 1.2.5 that the (ferromagnetic) Ising model on $V$ with inverse temperature $\beta$ is the probability distribution over spin configurations $\sigma \in \mathcal{X}$ given by

$$\mu_\beta(\sigma) := \frac{1}{\mathcal{Z}(\beta)} e^{-\beta \mathcal{H}(\sigma)},$$

where

$$\mathcal{H}(\sigma) := -\sum_{i \sim j} \sigma_i \sigma_j,$$

is the Hamiltonian and

$$\mathcal{Z}(\beta) := \sum_{\sigma \in \mathcal{X}} e^{-\beta \mathcal{H}(\sigma)},$$

is the partition function. In this context, recall that vertices are often referred to as sites. The single-site Glauber dynamics (Definition 1.2.8) of the Ising model is the Markov chain on $\mathcal{X}$ which, at each time, selects a site $i \in V$ uniformly at random and updates the spin $\sigma_i$ according to $\mu_\beta(\sigma)$ conditioned on agreeing with $\sigma$ at all sites in $V \backslash \{i\}$. Specifically, for $\gamma \in \{-1, +1\}$, $i \in V$, and $\sigma \in \mathcal{X}$, let $\sigma^{i,\gamma}$ be the configuration $\sigma$ with the state at $i$ being set to $\gamma$. Then, letting $n = |V|$, the transition matrix of the Glauber dynamics is

$$\begin{aligned} Q_\beta(\sigma, \sigma^{i,\gamma}) &:= \frac{1}{n} \cdot \frac{e^{\gamma \beta S_i(\sigma)}}{e^{-\beta S_i(\sigma)} + e^{\beta S_i(\sigma)}} \\ &= \frac{1}{n} \left\{ \frac{1}{2} + \frac{1}{2} \tanh(\gamma \beta S_i(\sigma)) \right\}, \end{aligned} \tag{4.3.11}$$

where

$$S_i(\sigma) := \sum_{j \sim i} \sigma_j.$$

All other transitions have probability $0$. Recall that this chain is irreducible and reversible with respect to $\mu_\beta$. In particular $\mu_\beta$ is the stationary distribution of $Q_\beta$.

In this section we give an upper bound on the mixing time, $t_{\mathrm{mix}}(\varepsilon)$, of $Q_\beta$ using path coupling. We say that the Glauber dynamics is *fast mixing* if $t_{\mathrm{mix}}(\varepsilon) = O(n \log n)$. We first make a simple observation: *fast mixing*

**Claim 4.3.13** (Glauber dynamics: lower bound on mixing)**.**

$$t_{\mathrm{mix}}(\varepsilon) = \Omega(n), \qquad \forall \beta > 0.$$

*Proof.* Similarly to what we did in Section 4.3.2 in the context of random walk on the hypercube (but for a lower bound this time), we use a coupon collecting argument (see Example 2.1.4). Let $\bar{\sigma}$ be the all-$(-1)$ configuration and let $A$ be the set of configurations where at least half of the sites are $+1$. Then, by symmetry, $\mu_\beta(A) = \mu_\beta(A^c) = 1/2$ where we assumed for simplicity that $n$ is odd. By definition of the total variation distance,

$$\begin{aligned} d(t) &\geq \|Q_\beta^t(\bar{\sigma}, \cdot) - \mu_\beta(\cdot)\|_{\mathrm{TV}} \\ &\geq |Q_\beta^t(\bar{\sigma}, A) - \mu_\beta(A)| \\ &= |Q_\beta^t(\bar{\sigma}, A) - 1/2|. \end{aligned} \tag{4.3.12}$$

So it remains to show that by time $c\,n$, for $c > 0$ small, the chain is unlikely to have reached $A$. That happens if, say, fewer than a third of the sites have been updated. Using the notation of Example 2.1.4, we are seeking a bound on $T_{n,n/3}$, that is, the time to collect $n/3$ coupons out of $n$.

We can write this random variable as a sum of $n/3$ independent geometric variables $T_{n,n/3} = \sum_{i=1}^{n/3} \tau_{n,i}$, where $\mathbb{E}[\tau_{n,i}] = \left(1 - \frac{i-1}{n}\right)^{-1}$ and $\mathrm{Var}[\tau_{n,i}] \leq \left(1 - \frac{i-1}{n}\right)^{-2}$. Hence, approximating the Riemann sums below by integrals,

$$\mathbb{E}[T_{n,n/3}] = \sum_{i=1}^{n/3} \left(1 - \frac{i-1}{n}\right)^{-1} = n \sum_{j=2n/3+1}^{n} j^{-1} = \Theta(n), \tag{4.3.13}$$

and

$$\mathrm{Var}[T_{n,n/3}] \leq \sum_{i=1}^{n/3} \left(1 - \frac{i-1}{n}\right)^{-2} = n^2 \sum_{j=2n/3+1}^{n} j^{-2} = \Theta(n). \tag{4.3.14}$$

So by Chebyshev's inequality (Theorem 2.1.2)

$$\mathbb{P}[|T_{n,n/3} - \mathbb{E}[T_{n,n/3}]| \geq \varepsilon\, n] \leq \frac{\mathrm{Var}[T_{n,n/3}]}{(\varepsilon\, n)^2} \to 0,$$

by (4.3.14). In view of (4.3.13), taking $\varepsilon > 0$ small enough and $n$ large enough, we have shown that for $t \leq c_\varepsilon n$ for some $c_\varepsilon > 0$

$$Q_\beta^t(\bar{\sigma}, A) \leq 1/3,$$

which proves the claim by (4.3.12) and the definition of the mixing time (Definition 1.1.35). ∎

**Remark 4.3.14.** *In fact, Ding and Peres proved that* $\mathrm{t_{mix}}(\varepsilon) = \Omega(n \log n)$ *for any graph on* $n$ *vertices [DP11]. In Claim 4.3.7, we treated the special case of the empty graph, which is equivalent to lazy random walk on the hypercube. See also Section 5.3.4 for a much stronger lower bound at low temperature for certain graphs with good "expansion properties."*

In our main result of this section, we show that the Glauber dynamics of the Ising model is fast mixing when the inverse temperature $\beta$ is small enough as a function of the maximum degree.

**Claim 4.3.15** (Glauber dynamics: fast mixing at high temperature)**.**

$$\beta < \bar{\delta}^{-1} \implies \mathrm{t_{mix}}(\varepsilon) = O(n \log n).$$

*Proof.* We use path coupling. Let $H_0 = (V_0, E_0)$ where $V_0 := \mathcal{X}$ and $\{\sigma, \omega\} \in E_0$ if $\frac{1}{2}\|\sigma - \omega\|_1 = 1$ (i.e., they differ in exactly one coordinate) with unit weight on all edges. To avoid confusion, we reserve the notation $\sim$ for adjacency in $G$.

Let $\{\sigma, \omega\} \in E_0$ differ at coordinate $i$. We construct a coupling $(X^*, Y^*)$ of $Q_\beta(\sigma, \cdot)$ and $Q_\beta(\omega, \cdot)$. We first pick the same coordinate $i_*$ to update. If $i_*$ is such that all its neighbors in $G$ have the same state in $\sigma$ and $\omega$, that is, if $\sigma_j = \omega_j$ for all $j \sim i_*$, we update $X^*$ from $\sigma$ according to the Glauber rule and set $Y^* := X^*$. Note that this includes the case $i_* = i$. Otherwise, that is, if $i_* \sim i$, we proceed as follows. From the state $\sigma$, the probability of updating site $i_*$ to state $\gamma \in \{-1, +1\}$ is given by the expression in brackets in (4.3.11), and similarly for $\omega$. Unlike the previous case, we cannot guarantee that the update is identical in both chains. In order to minimize the chance of increasing the distance between the two chains, we use a monotone coupling, which recall from Example 4.1.17 is maximal in the two-state case. Specifically, we pick a uniform random variable $U$ in $[-1, 1]$ and set

$$X^*_{i_*} := \begin{cases} +1 & \text{if } U \leq \tanh(\beta S_{i_*}(\sigma)), \\ -1 & \text{otherwise,} \end{cases}$$

and

$$Y^*_{i_*} := \begin{cases} +1 & \text{if } U \leq \tanh(\beta S_{i_*}(\omega)), \\ -1 & \text{otherwise.} \end{cases}$$

We set $X^*_j := \sigma_j$ and $Y^*_j := \omega_j$ for all $j \neq i^*$. The expected distance between $X^*$ and $Y^*$ is then

$$\begin{aligned} &\mathbb{E}[w_0(X^*, Y^*)] \\ &= 1 - \underbrace{\frac{1}{n}}_{\text{(a)}} + \underbrace{\frac{1}{n} \sum_{j : j \sim i} \frac{1}{2} \left|\tanh(\beta S_j(\sigma)) - \tanh(\beta S_j(\omega))\right|}_{\text{(b)}}, \end{aligned} \tag{4.3.15}$$

where: (a) corresponds to $i_* = i$ in which case $w_0(X^*, Y^*) = 0$; (b) corresponds to $i_* \sim i$ in which case $w_0(X^*, Y^*) = 2$ with probability

$$\frac{1}{2}|\tanh(\beta S_{i_*}(\sigma)) - \tanh(\beta S_{i_*}(\omega))|,$$

by our coupling; and otherwise $w_0(X^*, Y^*) = w_0(\sigma, \omega) = 1$. To bound (b), we note that for any $j \sim i$

$$|\tanh(\beta S_j(\sigma)) - \tanh(\beta S_j(\omega))| = \tanh(\beta(s+2)) - \tanh(\beta s), \qquad (4.3.16)$$

where

$$s := S_j(\sigma) \wedge S_j(\omega).$$

The derivative of $\tanh$ is maximized at $0$ where it is equal to $1$. So the right-hand side of (4.3.16) is $\leq \beta(s+2) - \beta s = 2\beta$. Plugging this back into (4.3.15) and using $1 - x \leq e^{-x}$ for all $x$ (see Exercise 1.16), we get

$$\mathbb{E}[w_0(X^*, Y^*)] \leq 1 - \frac{1 - \bar{\delta}\beta}{n} \leq \exp\left(-\frac{1 - \bar{\delta}\beta}{n}\right) = \kappa\, w_0(\sigma, \omega),$$

where

$$\kappa := \exp\left(-\frac{1 - \bar{\delta}\beta}{n}\right) < 1,$$

by our assumption on $\beta$. The diameter of $H_0$ is $\Delta_0 = n$. By Theorem 4.3.10,

$$\mathrm{t_{mix}}(\varepsilon) \leq \left\lceil \frac{\log \Delta_0 + \log \varepsilon^{-1}}{\log \kappa^{-1}} \right\rceil = \left\lceil \frac{n(\log n + \log \varepsilon^{-1})}{1 - \bar{\delta}\beta} \right\rceil,$$

which implies the claim. ■

**Remark 4.3.16.** *A slighlty more careful analysis shows that the condition* $\bar{\delta}\tanh(\beta) < 1$ *is enough for the claim to hold. See [LPW06, Theorem 15.1].*

## 4.4 Chen-Stein method

The Chen-Stein method serves to establish Poisson approximation results with quantitative bounds in certain settings with dependent variables that are common, for instance, in random graphs and string statistics.

**Setting** The basic setup is a sum of Bernoulli (i.e., $\{0,1\}$-valued) random variables $\{X_i\}_{i=1}^n$

$$W = \sum_{i=1}^n X_i, \tag{4.4.1}$$

where the $X_i$s are *not* assumed independent or identically distributed. Define

$$p_i = \mathbb{P}[X_i = 1], \tag{4.4.2}$$

and

$$\mathbb{E}[W] = \lambda := \sum_{i=1}^n p_i. \tag{4.4.3}$$

Letting $\mu$ denote the law of $W$ and $\pi$ be the Poisson distribution with mean $\lambda$, our goal is to bound $\|\mu - \pi\|_{\mathrm{TV}}$.

We first state the main bounds and give some examples of its use. We then motivate and prove the result, and return to further applications. Throughout the next two subsections, we use the notation in (4.4.1), (4.4.2) and (4.4.3).

### 4.4.1 Main bounds and examples

We begin with an elementary observation.

**Theorem 4.4.1** (Stein equation for the Poisson distribution)**.** *Let $\lambda > 0$. A non-negative integer-valued random variable $Z$ is $\mathrm{Poi}(\lambda)$ if and only if for all $g$ bounded*

$$\mathbb{E}[\lambda g(Z+1) - Zg(Z)] = 0. \tag{4.4.4}$$

The "only if" follows a direct calculation. The "if" follows from taking $g(z) := \mathbf{1}_{\{z=k\}}$ for all $k \geq 1$ and deriving a recursion. Exercise 4.18 asks for the details. One might expect that if the left-hand side of (4.4.4) is "small for many $g$s," then $Z$ is approximately Poisson.

The following key result in some sense helps to formalize this intuition. We prove it by constructing a Markov chain that "interpolates" between $\mu$ and $\pi$, where (4.4.4) will arise naturally (see Section 4.4.2).

**Theorem 4.4.2** (Chen-Stein method)**.** *Let $W \sim \mu$ and $\pi \sim \mathrm{Poi}(\lambda)$. Then there exists a function $h : \{0,1,\ldots,n+1\} \to \mathbb{R}$ such*

$$\|\mu - \pi\|_{\mathrm{TV}} = \mathbb{E}\left[-\lambda h(W+1) + Wh(W)\right]. \tag{4.4.5}$$

*Moreover $h$ satisfies the following Lipschitz condition: for all $y, y' \in \{0,1,\ldots,n+1\}$,*

$$|h(y') - h(y)| \leq (1 \wedge \lambda^{-1})|y' - y|. \tag{4.4.6}$$

By bounding the right-hand side of (4.4.5) for any function satisfying (4.4.6), we get a Poisson approximation result for $\mu$.

One way to do this is to construct a certain type of coupling. We begin with a definition, which will be justified in the corollary below. We write $X \sim Y|A$ to mean that $X$ is distributed as $Y$ conditioned on the event $A$.

**Definition 4.4.3** (Stein coupling)**.** *A* Stein coupling *is a pair* $(U_i, V_i)$*, for each* $i = 1, \ldots, n$*, such that*

Stein coupling

$$U_i \sim W, \qquad V_i \sim W - 1 | X_i = 1.$$

*Each pair* $(U_i, V_i)$ *is defined on a joint probability space, but different pairs do not need to.*

How such a coupling is constructed will become clearer in the examples below.

**Corollary 4.4.4.** *Let* $(U_i, V_i)$*,* $i = 1, \ldots, n$*, be a Stein coupling. Then*

$$\|\mu - \pi\|_{\mathrm{TV}} \leq (1 \wedge \lambda^{-1}) \sum_{i=1}^{n} p_i \, \mathbb{E}|U_i - V_i|. \tag{4.4.7}$$

*Proof.* By (4.4.5), using the facts that $\lambda = \sum_{i=1}^n p_i$ and $W = \sum_{i=1}^n X_i$, we get

$$\begin{aligned}
&\|\mu - \pi\|_{\mathrm{TV}} \\
&= \mathbb{E}\left[-\lambda h(W+1) + W h(W)\right] \\
&= \mathbb{E}\left[-\left(\sum_{i=1}^n p_i\right) h(W+1) + \left(\sum_{i=1}^n X_i\right) h(W)\right] \\
&= \sum_{i=1}^n \left(-p_i \mathbb{E}\left[h(W+1)\right] + \mathbb{E}\left[X_i h(W)\right]\right) \\
&= \sum_{i=1}^n \left(-p_i \mathbb{E}\left[h(W+1)\right] + \mathbb{E}\left[h(W) \,|\, X_i = 1\right] \mathbb{P}[X_i = 1]\right) \\
&= \sum_{i=1}^n p_i \left(-\mathbb{E}\left[h(W+1)\right] + \mathbb{E}\left[h(W) \,|\, X_i = 1\right]\right).
\end{aligned}$$

Let $(U_i, V_i)$, $i = 1, \ldots, n$, be a Stein coupling (Definition 4.4.3). Then, we can rewrite this last expression is

$$\begin{aligned}
&= \sum_{i=1}^n p_i \left(-\mathbb{E}\left[h(U_i + 1)\right] + \mathbb{E}\left[h(V_i + 1)\right]\right) \\
&\leq \sum_{i=1}^n p_i \mathbb{E}\left[\left|h(U_i + 1) - h(V_i + 1)\right|\right].
\end{aligned}$$

By (4.4.6), we finally get

$$\|\mu - \pi\|_{\mathrm{TV}} \leq (1 \wedge \lambda^{-1}) \sum_{i=1}^{n} p_i \, \mathbb{E}|U_i - V_i|,$$

which concludes the proof. ■

As a first example, we derive a Poisson approximation result in the independent case. Compare to Theorem 4.1.18.

**Example 4.4.5** (Independent $X_i$s)**.** Assume the $X_i$s are independent. We prove the following:

**Claim 4.4.6.**

$$\|\mu - \pi\|_{\mathrm{TV}} \leq (1 \wedge \lambda^{-1}) \sum_{i=1}^{n} p_i^2.$$

We use the following Stein coupling. For each $i = 1, \ldots, n$, we let

$$U_i = W$$

and

$$V_i = \sum_{j:j \neq i} X_j.$$

By independence,

$$V_i = W - X_i \sim W - 1 | X_i = 1,$$

as desired. Plugging into (4.4.7), we obtain the bound

$$\begin{aligned}
\|\mu - \pi\|_{\mathrm{TV}} &\leq (1 \wedge \lambda^{-1}) \sum_{i=1}^{n} p_i \, \mathbb{E}|U_i - V_i| \\
&\leq (1 \wedge \lambda^{-1}) \sum_{i=1}^{n} p_i \, \mathbb{E} \left| W - \sum_{j \neq i} X_j \right| \\
&\leq (1 \wedge \lambda^{-1}) \sum_{i=1}^{n} p_i \, \mathbb{E}\,|X_i| \\
&\leq (1 \wedge \lambda^{-1}) \sum_{i=1}^{n} p_i^2.
\end{aligned}$$

◀

Here is a less straightforward example.

**Example 4.4.7** (Balls in boxes)**.** Suppose we throw $k$ balls uniformly at random in $n$ boxes independently. Let

$$X_i = \mathbf{1}\{\text{box } i \text{ is empty}\},$$

and let $W = \sum_{i=1}^n X_i$ be the number of empty boxes. Note that the $X_i$s are *not* independent. In particular, we cannot use Theorem 4.1.18. Note that

$$p_i = \left(1 - \frac{1}{n}\right)^k,$$

for all $i$ and, hence,

$$\lambda = n\left(1 - \frac{1}{n}\right)^k.$$

For each $i = 1, \ldots, n$, we generate the coupling $(U_i, V_i)$ in the following way. We let $U_i = W$. If box $i$ is empty then $V_i = W - 1$. Otherwise, we re-distribute all balls in box $i$ among the remaining boxes and let $V_i$ count the number of empty boxes $\neq i$. By construction, both conditions of the Stein coupling are satisfied. Moreover we have almost surely $V_i \leq U_i$ so that

$$\sum_{i=1}^n p_i \mathbb{E}|U_i - V_i| = \sum_{i=1}^n p_i \mathbb{E}[U_i - V_i] = \lambda^2 - \sum_{i=1}^n p_i \mathbb{E}[V_i].$$

By the fact that $V_i \sim W - 1 | X_i = 1$ and Bayes' rule,

$$\begin{aligned}\sum_{i=1}^n p_i \mathbb{E}[V_i] &= \sum_{i=1}^n \mathbb{P}[X_i = 1] \sum_{k=1}^n (k-1)\mathbb{P}[V_i = k-1] \\ &= \sum_{i=1}^n \sum_{k=1}^n (k-1)\mathbb{P}[W = k \,|\, X_i = 1]\, \mathbb{P}[X_i = 1] \\ &= \sum_{i=1}^n \sum_{k=1}^n (k-1)\mathbb{P}[X_i = 1 \,|\, W = k]\, \mathbb{P}[W = k].\end{aligned}$$

Now we use the fact that $\mathbb{P}[X_i = 1 \,|\, W = k] = \mathbb{E}[X_i \,|\, W = k]$ because $X_i$ is an indicator variable. So the last line above is

$$
\begin{aligned}
&= \sum_{i=1}^{n} \sum_{k=1}^{n} (k-1)\mathbb{E}[X_i \,|\, W = k]\, \mathbb{P}[W = k] \\
&= \sum_{k=1}^{n} (k-1)\mathbb{E}\left[ \sum_{i=1}^{n} X_i \,\middle|\, W = k \right] \mathbb{P}[W = k] \\
&= \sum_{k=1}^{n} (k-1)k\, \mathbb{P}[W = k] \\
&= \mathbb{E}[W^2] - \mathbb{E}[W].
\end{aligned}
$$

It remains to compute $\mathbb{E}[W^2]$. We have by symmetry

$$
\begin{aligned}
\mathbb{E}[W^2] &= n\,\mathbb{E}[X_1^2] + n(n-1)\mathbb{E}[X_1 X_2] \\
&= \lambda + n(n-1)\left(1 - \frac{2}{n}\right)^k,
\end{aligned}
$$

so by Corollary 4.4.4

$$
\|\mu - \pi\|_{\mathrm{TV}} \leq (1 \wedge \lambda^{-1}) \left\{ n^2 \left(1 - \frac{1}{n}\right)^{2k} - n(n-1)\left(1 - \frac{2}{n}\right)^k \right\}.
$$

When $k = n \log n + Cn$ for instance, it can be checked that $\|\mu - \pi\|_{\mathrm{TV}} = O(\log n / n)$. ◀

This last example is generalized in Exercise 4.22.

In special settings, one can give useful general bounds by constructing an appropriate Stein coupling. We give an important example next. Recall that $[n] = \{1, \ldots, n\}$.

**Theorem 4.4.8** (Chen-Stein method: dissociated case)**.** *Suppose that for each $i$ there is a neighborhood $\mathcal{N}_i \subseteq [n] \setminus \{i\}$ such that*

$$
X_i \text{ is independent of } \{X_j : j \notin \mathcal{N}_i \cup \{i\}\}.
$$

*Then*

$$
\|\mu - \pi\|_{\mathrm{TV}} \leq (1 \wedge \lambda^{-1}) \sum_{i=1}^{n} \left\{ p_i^2 + \sum_{j \in \mathcal{N}_i} (p_i p_j + \mathbb{E}[X_i X_j]) \right\}.
$$

*Proof.* We use the following Stein coupling. Let

$$U_i = W.$$

Then generate

$$(Y_j^{(i)})_{j\in\mathcal{N}_i} \sim (X_j)_{j\in\mathcal{N}_i} | \{X_k : k \notin \mathcal{N}_i \cup \{i\}\}, X_i = 1,$$

and set

$$V_i = \sum_{k\notin\mathcal{N}_i\cup\{i\}} X_k + \sum_{j\in\mathcal{N}_i} Y_j^{(i)}.$$

Because the law of $\{X_k : k \notin \mathcal{N}_i \cup \{i\}\}$ (and therefore of the first term in $V_i$) is independent of the event $\{X_i = 1\}$, the above scheme satisfies the conditions of the Stein coupling.

Hence we can apply Corollary 4.4.4. The construction of $(U_i, V_i)$ guarantees that $U_i - V_i$ depends *only* on "$i$ and its neighborhood." Specifically, we get

$$\begin{aligned}
\|\mu - \pi\|_{\mathrm{TV}} &\leq (1 \wedge \lambda^{-1}) \sum_{i=1}^n p_i \, \mathbb{E}|U_i - V_i| \\
&= (1 \wedge \lambda^{-1}) \sum_{i=1}^n p_i \, \mathbb{E}\left| \sum_{j=1}^n X_j - \sum_{k\notin\mathcal{N}_i\cup\{i\}} X_k - \sum_{j\in\mathcal{N}_i} Y_j^{(i)} \right| \\
&= (1 \wedge \lambda^{-1}) \sum_{i=1}^n p_i \, \mathbb{E}\left| X_i + \sum_{j\in\mathcal{N}_i} (X_j - Y_j^{(i)}) \right| \\
&\leq (1 \wedge \lambda^{-1}) \sum_{i=1}^n p_i \left( \mathbb{E}|X_i| + \sum_{j\in\mathcal{N}_i} (\mathbb{E}|X_j| + \mathbb{E}|Y_j^{(i)}|) \right),
\end{aligned}$$

where we used the triangle inequality. Recalling that $p_i = \mathbb{P}[X_i = 1] = \mathbb{E}[X_i] = \mathbb{E}|X_i|$ and the definition of $Y_j^{(i)}$, the last expression above is

$$\begin{aligned}
&= (1 \wedge \lambda^{-1}) \sum_{i=1}^n p_i \left( p_i + \sum_{j\in\mathcal{N}_i} [p_j + \mathbb{E}[|X_j| | X_i = 1] \right) \\
&= (1 \wedge \lambda^{-1}) \sum_{i=1}^n \left\{ p_i^2 + \sum_{j\in\mathcal{N}_i} (p_i p_j + p_i \mathbb{E}[X_j | X_i = 1]) \right\} \\
&= (1 \wedge \lambda^{-1}) \sum_{i=1}^n \left\{ p_i^2 + \sum_{j\in\mathcal{N}_i} (p_i p_j + \mathbb{E}[X_i X_j]) \right\}.
\end{aligned}$$

That concludes the proof. ■

Next we give an example of the previous theorem.

**Example 4.4.9** (Longest head run)**.** Let $0 < q < 1$ and let $Z_1, Z_2, \ldots$ be i.i.d. Bernoulli random variables with success probability $q = \mathbb{P}[Z_i = 1]$. We are interested in the distribution of $R$, the length of the longest run of 1s starting in the first $n$ tosses. For any positive integer $t$, let $X_1^{(t)} := Z_1 \cdots Z_t$ and

$$X_i^{(t)} := (1 - Z_{i-1}) Z_i \cdots Z_{i+t-1}, \qquad i \geq 2.$$

The event $\{X_i^{(t)} = 1\}$ indicates that a head run of length at least $t$ *starts* at the $i$-th toss. Now define

$$W^{(t)} := \sum_{i=1}^{n} X_i^{(t)}.$$

The key observation is that

$$\{R < t\} = \{W^{(t)} = 0\}. \tag{4.4.8}$$

Notice that, for fixed $t$, the $X_i^{(t)}$s are neither independent nor identically distributed. However, they exhibit a natural neighborhood structure as in Theorem 4.4.8. Indeed let

$$\mathcal{N}_i^{(t)} := \{\alpha \in [n] : |\alpha - i| \leq t\} \setminus \{i\}.$$

Then, $X_i^{(t)}$ is independent of $\{X_j^{(t)} : j \notin \mathcal{N}_i \cup \{i\}\}$. For example,

$$X_i^{(t)} = (1 - Z_{i-1}) Z_i \cdots Z_{i+t-1},$$

and

$$X_{i+t+1}^{(t)} = (1 - Z_{i+t}) Z_{i+t+1} \cdots Z_{i+2t},$$

do not depend on any common $Z_j$, while $X_i^{(t)}$ and

$$X_{i+t}^{(t)} = (1 - Z_{i+t-1}) Z_{i+t} \cdots Z_{i+2t-1},$$

both depend on $Z_{i+t-1}$.

We compute the quantities needed to apply Theorem 4.4.8. We have

$$p_1^{(t)} := \mathbb{E}[Z_1 \cdots Z_t] = \prod_{j=1}^{t} \mathbb{E}[Z_j] = q^t,$$

and, for $i \geq 2$,

$$\begin{aligned} p_i^{(t)} &:= \mathbb{E}[(1 - Z_{i-1}) Z_i \cdots Z_{i+t-1}] \\ &= \mathbb{E}[1 - Z_{i-1}] \prod_{j=i}^{i+t-1} \mathbb{E}[Z_j] \\ &= (1-q) q^t \\ &\leq q^t. \end{aligned}$$

For $i \geq 1$ and $j \in \mathcal{N}_i^{(t)}$, observe that a head run of length at least $t$ cannot start simultaneously at $i$ and $j$. So $\mathbb{E}[X_i^{(t)} X_j^{(t)}] = 0$ in that case. We also have

$$\lambda^{(t)} := \mathbb{E}[W^{(t)}] = q^t + (n-1)(1-q)q^t \in [n(1-q)q^t, nq^t],$$

and

$$\left| \mathcal{N}_i^{(t)} \right| \leq 2t.$$

We are ready to apply Theorem 4.4.8. We get, letting $\mu^{(t)}$ denote the law of $W^{(t)}$ and $\pi^{(t)}$ be the Poisson distribution with mean $\lambda^{(t)}$,

$$\begin{aligned} \|\mu^{(t)} - \pi^{(t)}\|_{\mathrm{TV}} &\leq (1 \wedge (\lambda^{(t)})^{-1}) \sum_{i=1}^n \left\{ (p_i^{(t)})^2 + \sum_{j \in \mathcal{N}_i^{(t)}} \left( p_i^{(t)} p_j^{(t)} + \mathbb{E}[X_i^{(t)} X_j^{(t)}] \right) \right\} \\ &\leq (1 \wedge (n(1-q)q^t)^{-1}) \left[ nq^{2t} + 2tnq^{2t} \right] \\ &\leq \frac{1}{(1-q)n} (1 \wedge (nq^t)^{-1}) [2t+1] (nq^t)^2. \end{aligned}$$

This bound is non-asymptotic—it holds for any $q, n, t$. One special regime of note is $t = \log_{1/q} n + C$ with large $n$. In that case, we have $nq^t \to C'$ as $n \to +\infty$ for some $0 < C' < +\infty$ and the total variation above is of the order to $O(\log n/n)$.

Going back to (4.4.8), we finally obtain when $t = \log_{1/q} n + C$ that

$$\left| \mathbb{P}[R < t] - e^{-\lambda^{(t)}} \right| = O\left( \frac{\log n}{n} \right),$$

where recall that $R$ and $\lambda^{(t)}$ implicitly depend on $n$. ◀

### 4.4.2 Some motivation and proof

The idea behind the Chen-Stein method is to interpolate between $\mu$ and $\pi$ in Theorem 4.4.2 by constructing a Markov chain with initial distribution $\mu$ and stationary distribution $\pi$. Here we use a discrete-time, finite Markov chain.

*Proof of Theorem 4.4.2.* We seek a bound on

$$\begin{aligned}\|\mu - \pi\|_{\mathrm{TV}} &= \sup_{A \subseteq \mathbb{Z}_+} |\mu(A) - \pi(A)| \\ &= \mu(A^*) - \pi(A^*) \\ &= \sum_{z \in A^*} (\mu(z) - \pi(z)),\end{aligned} \tag{4.4.9}$$

where $A^* = \{z \in \mathbb{Z}_+ : \mu(z) > \pi(z)\}$, by Lemma 4.1.15. Since $W \leq n$ almost surely, $\mu(z) = 0$ for all $z > n$ which implies that $A^* \subseteq \{0, 1, \ldots, n\}$. In particular, it will suffice to bound $\mu(z) - \pi(z)$ for $0 \leq z \leq n$. We also assume $\lambda < n$ (the case $\lambda = n$ being uninteresting).

**Constructing the Markov chain** It will be convenient to truncate $\pi$ at $n$, that is, we define

$$\bar{\pi}(z) = \begin{cases} \pi(z) & 0 \leq z \leq n, \\ 1 - \Pi(n) & z = n + 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $\Pi(z) = \sum_{w \leq z} \pi(w)$ is the cumulative distribution function of the Poisson distribution with mean $\lambda$. We construct a Markov chain with stationary distribution $\bar{\pi}$. We will also need the chain to be aperiodic and irreducible over $\{0, 1, \ldots, n + 1\}$.

We choose the transition matrix $(P(x, y))_{0 \leq x, y \leq n+1}$ to be that of a birth-death chain reversible with respect to $\bar{\pi}$, that is, we require $P(x, y) = 0$ unless $|x - y| \leq 1$ and

$$\frac{P(x, x+1)}{P(x+1, x)} = \frac{\bar{\pi}(x+1)}{\bar{\pi}(x)}, \qquad \forall x \in [n]. \tag{4.4.10}$$

For $x < n$,

$$\frac{\bar{\pi}(x+1)}{\bar{\pi}(x)} = \frac{\pi(x+1)}{\pi(x)} = \frac{e^{-\lambda} \lambda^{x+1}/(x+1)!}{e^{-\lambda} \lambda^x / x!} = \frac{\lambda}{x+1}.$$

In view of this, we want $P(x, x+1) \propto \lambda$ and $P(x, x-1) \propto x$. We choose the proportionality constant to ensure that all transition probabilities are in $[0, 1]$. Specifically, for $x \neq y$, the nonzero transition probabilities take values

$$P(x, y) = \begin{cases} \frac{1}{2n} \lambda, & \text{if } 0 \leq x \leq n,\ y = x + 1, \\ \frac{1}{2n} x, & \text{if } 1 \leq x \leq n,\ y = x - 1, \\ \frac{1}{2n} \lambda \frac{\pi(n)}{1 - \Pi(n)}, & \text{if } x = n + 1,\ y = n. \end{cases} \tag{4.4.11}$$

The probability of staying put is: $1 - \frac{1}{2n}\lambda$ if $x = 0$, $1 - \frac{1}{2n}x - \frac{1}{2n}\lambda$ if $1 \leq x \leq n$, and $1 - \frac{1}{2n}\lambda\frac{\pi(n)}{1-\Pi(n)}$ if $x = n+1$. Those are all strictly positive when $\lambda < n$. Hence by construction $P$ is aperiodic and irreducible, and it satisfies the detailed balance conditions (4.4.10).

Recalling (3.3.6), the Laplacian is

$$\begin{aligned}\Delta f(x) &= \sum_y P(x,y)[f(y) - f(x)] \\ &= P(x,x+1)[f(x+1) - f(x)] - P(x,x-1)[f(x) - f(x-1)] \\ &= \lambda g(x+1) - x g(x),\end{aligned}$$

for $0 \leq x \leq n$, where we defined

$$g(x) := \frac{f(x) - f(x-1)}{2n}, \qquad x \in \{1, \ldots, n+1\}, \tag{4.4.12}$$

and $g(0)$ is arbitrary. At $x = n+1$,

$$\Delta f(n+1) = -\lambda n \frac{\pi(n)}{1 - \Pi(n)} g(n+1).$$

It is a standard fact (see Exercise 4.19) that the expectation of the Laplacian under the stationary distribution is 0. Inverting the relationship (4.4.12), for any $g : \{0, \ldots, n+1\} \to \mathbb{R}$, there is a corresponding $f$ (unique up to an additive constant). So we have shown that if $Z \sim \bar{\pi}$ then

$$\mathbb{E}\left[(\lambda g(Z+1) - Z g(Z))\mathbf{1}_{\{Z \leq n\}} - \lambda n \frac{\pi(n)}{1 - \Pi(n)} g(Z)\mathbf{1}_{\{Z = n+1\}}\right] = 0,$$

that is,

$$\mathbb{E}\left[(\lambda g(Z+1) - Z g(Z))\mathbf{1}_{\{Z \leq n\}}\right] = \lambda n \pi(n) g(n+1).$$

Notice that, if $g$ is extended to a bounded function on $\mathbb{Z}_+$, $\lambda$ is fixed and $Z \sim \mathrm{Poi}(\lambda)$, then taking $n \to +\infty$ recovers Theorem 4.4.1 by dominated convergence (Proposition B.4.14).*

**Markov chains calculations** By the convergence theorem for Markov chains (Theorem 1.1.33),

$$P^t(y,z) \to \bar{\pi}(z)$$

*The above argument is more natural in the setting of continuous-time Markov chains, but we will not introduced them here.

for all $0 \leq y \leq n+1$ and $0 \leq z \leq n+1$ as $t \to +\infty$. Hence, letting $\delta_z(x) = \mathbf{1}_{\{x=z\}}$, by telescoping

$$\begin{aligned}\delta_z(y) - \bar{\pi}(z) &= \lim_{t\to+\infty} \mathbb{E}_y[\delta_z(X_0) - \delta_z(X_t)] \\ &= \lim_{t\to+\infty} \sum_{s=0}^{t-1} \mathbb{E}_y[\delta_z(X_s) - \delta_z(X_{s+1})], \end{aligned} \tag{4.4.13}$$

where the subscript of $\mathbb{E}$ indicates the initial state. We will later take expectations over $\mu$ and use the fact that $\pi(z) = \bar{\pi}(z)$ on $\{0, 1 \ldots, n\}$ to interpolate between $\mu$ and $\pi$.

First, we use standard Markov chains facts to compute (4.4.13). Define for $y \in \{1, \ldots, n+1\}$

$$g_z^t(y) := \frac{1}{2n} \sum_{s=0}^{t-1} (\mathbb{E}_y[\delta_z(X_s)] - \mathbb{E}_{y-1}[\delta_z(X_s)]), \tag{4.4.14}$$

and $g_z^t(0) := 0$. The function $g_z^t(y)$ is, up to a factor (whose purpose will be clearer below), the difference between the expected number of visits to $z$ up to time $t-1$ when started at $y$ and $y-1$ respectively. It depends on $\mu$ only through $\lambda$ and $n$. By Chapman-Kolmogorov (Theorem 1.1.20) applied to the first step of the chain,

$$\begin{aligned}\mathbb{E}_y[\delta_z(X_{s+1})] = P(y, y+1)\, \mathbb{E}_{y+1}[\delta_z(X_s)] \\ + P(y, y)\, \mathbb{E}_y[\delta_z(X_s)] + P(y, y-1)\, \mathbb{E}_{y-1}[\delta_z(X_s)].\end{aligned}$$

Using that $P(y, y+1) + P(y, y) + P(y, y-1) = 1$ and rearranging we get for $0 \leq y \leq n$ and $0 \leq z \leq n+1$

$$\begin{aligned}&\sum_{s=0}^{t-1} \mathbb{E}_y[\delta_z(X_s) - \delta_z(X_{s+1})] \\ &= \sum_{s=0}^{t-1} \Big\{ - P(y, y+1)(\mathbb{E}_{y+1}[\delta_z(X_s)] - \mathbb{E}_y[\delta_z(X_s)]) \\ &\qquad\qquad + P(y, y-1)(\mathbb{E}_y[\delta_z(X_s)] - \mathbb{E}_{y-1}[\delta_z(X_s)]) \Big\} \\ &= -2nP(y, y+1) g_z^t(y+1) + 2nP(y, y-1) g_z^t(y) \\ &= -\lambda g_z^t(y+1) + y g_z^t(y), \end{aligned} \tag{4.4.15}$$

where we used (4.4.11) on the last line.

We establish after the proof of the theorem that $g_z^t(y)$ has a well-defined limit. That fact is not immediately obvious as the limit is the "difference of two infinities." But a simple coupling argument does the trick.

**Lemma 4.4.10.** *Let $g_z^t : \{0, 1, \ldots, n+1\} \to \mathbb{R}$ be defined in (4.4.14). Then there exists a bounded function $g_z^\infty : \{0, 1, \ldots, n+1\} \to \mathbb{R}$ such that for all $0 \le z \le n+1$ and $0 \le y \le n+1$,*

$$g_z^\infty(y) = \lim_{t \to +\infty} g_z^t(y).$$

In fact, an explicit expression for $g_z^\infty$ can be derived via the following recursion. That expression will be helpful to establish the Lipschitz condition in Theorem 4.4.2.

**Lemma 4.4.11.** *For all $0 \le y \le n$ and $0 \le z \le n+1$,*

$$\delta_z(y) - \bar{\pi}(z) = -\lambda g_z^\infty(y+1) + y g_z^\infty(y).$$

*Proof.* Combine (4.4.13), (4.4.15), and Lemma 4.4.10. ■

Lemma 4.4.11 leads to the following formula for $g_z^\infty$, which we establish after the proof of the theorem.

**Lemma 4.4.12.** *For $1 \le y \le n+1$ and $0 \le z \le n+1$,*

$$g_z^\infty(y) = \begin{cases} \frac{\Pi(y-1)}{y\pi(y)} \bar{\pi}(z), & \text{if } z \ge y, \\ -\frac{1-\Pi(y-1)}{y\pi(y)} \bar{\pi}(z), & \text{if } z < y. \end{cases} \tag{4.4.16}$$

*and $g_z^\infty(0) = 0$.*

**Interpolating between $\mu$ and $\pi$** For $A \subseteq \{0, 1, \ldots, n\}$, define

$$g_A^\infty(y) := \sum_{z \in A} g_z^\infty(y).$$

We obtain the following key bound.

**Lemma 4.4.13** (Chen's equation)**.** *Let $W \sim \mu$ and $\pi \stackrel{\mathrm{d}}{=} \mathrm{Poi}(\lambda)$. Then,*

$$\|\mu - \pi\|_{\mathrm{TV}} = \mathbb{E}\left[-\lambda g_{A^*}^\infty(W+1) + W g_{A^*}^\infty(W)\right] \tag{4.4.17}$$

*where $A^* = \{z \in \mathbb{Z}_+ : \mu(z) > \pi(z)\}$.*

*Proof.* Fix $z \in \{0, 1, \ldots, n\}$. Multiplying both sides in Lemma 4.4.11 by $\mu(y)$ and summing over $y$ in $\{0, 1, \ldots, n\}$ gives

$$\mu(z) - \pi(z) = \mathbb{E}\left[-\lambda g_z^\infty(W+1) + W g_z^\infty(W)\right].$$

Now summing over $z$ in $A^* \subseteq \{0, 1, \ldots, n\}$ and using (4.4.9) gives the claim. ■

Lemma 4.4.12 can be used to derive a Lipschitz constant for $g_A^\infty$. That lemma is also established after the proof of the theorem.

**Lemma 4.4.14.** *For* $A \subseteq \{0, 1, \ldots, n\}$ *and* $y, y' \in \{0, 1, \ldots, n+1\}$,

$$|g_A^\infty(y') - g_A^\infty(y)| \leq (1 \wedge \lambda^{-1})|y' - y|.$$

Lemmas 4.4.13 and 4.4.14 imply the theorem with $h := g_{A^*}^\infty$. ∎

**Proofs of technical lemmas** It remains to prove Lemmas 4.4.10, 4.4.12 and 4.4.14.

*Proof of Lemma 4.4.10.* We use a coupling argument. Let $(Y_s, \tilde{Y}_s)_{s=0}^{+\infty}$ be an independent Markovian coupling of $(Y_s)$, the chain started at $y-1$, and $(\tilde{Y}_s)$, the chain started at $y$. Let $\tau$ be the first time $s$ that $Y_s = \tilde{Y}_s$. Because $Y_s$ and $\tilde{Y}_s$ are independent and $P$ is a birth-death chain with strictly positive nearest-neighbor and staying-put transition probabilities, the coupled chain $(Y_s, \tilde{Y}_s)_{s=0}^{+\infty}$ is aperiodic and irreducible over $\{0, 1, \ldots, n+1\}^2$. By the exponential tail of hitting times, Lemma 3.1.25, it holds that $\mathbb{E}[\tau] < +\infty$.

Modify the coupling $(Y_s, \tilde{Y}_s)$ to enforce $\tilde{Y}_s = Y_s$ for all $s \geq \tau$ (while not changing $(Y_s)$), that is, to make it coalescing. By the Strong Markov property (Theorem 3.1.8), the resulting chain $(Y_s^*, \tilde{Y}_s^*)$ is also a Markovian coupling of the chain started at $y-1$ and $y$ respectively. Using this coupling, we rewrite

$$\begin{aligned} g_z^t(y) &= \frac{1}{2n} \sum_{s=0}^{t-1} (\mathbb{E}_y[\delta_z(X_s)] - \mathbb{E}_{y-1}[\delta_z(X_s)]) \\ &= \frac{1}{2n} \sum_{s=0}^{t-1} \mathbb{E}[\delta_z(\tilde{Y}_s^*) - \delta_z(Y_s^*)] \\ &= \frac{1}{2n} \mathbb{E}\left[\sum_{s=0}^{t-1} (\delta_z(\tilde{Y}_s^*) - \delta_z(Y_s^*))\right]. \end{aligned}$$

The random variable inside the expectation is bounded in absolute value by

$$\left|\sum_{s=0}^{t-1} (\delta_z(\tilde{Y}_s^*) - \delta_z(Y_s^*))\right| \leq \tau,$$

uniformly in $t$. Indeed, after $s = \tau$, the terms in the sum are 0, while before $s = \tau$ the terms are bounded by 1 in absolute value. By the integrability of $\tau$, the dominated convergence theorem (Proposition B.4.14) allows to take the limit, leading to

$$\begin{aligned} g_z^\infty(y) &= \lim_{t\to+\infty} \frac{1}{2n}\,\mathbb{E}\left[\sum_{s=0}^{t-1}(\delta_z(\tilde{Y}_s^*) - \delta_z(Y_s^*))\right] \\ &= \frac{1}{2n}\,\mathbb{E}\left[\sum_{s=0}^{+\infty}(\delta_z(\tilde{Y}_s^*) - \delta_z(Y_s^*))\right] \\ &< +\infty. \end{aligned}$$

That concludes the proof. ■

*Proof of Lemma 4.4.12.* Our starting point is Lemma 4.4.11, from which we deduce the recursive formula

$$g_z^\infty(y+1) = \frac{1}{\lambda}\left\{y g_z^\infty(y) + \pi(z) - \delta_z(y)\right\}, \tag{4.4.18}$$

for $0 \le y \le n$ and $0 \le z \le n$.

We guess a general formula and then check it. By (4.4.18),

$$g_z^\infty(1) = \frac{1}{\lambda}\left\{\pi(z) - \delta_z(0)\right\}, \tag{4.4.19}$$

$$\begin{aligned} g_z^\infty(2) &= \frac{1}{\lambda}\left\{g_z^\infty(1) + \pi(z) - \delta_z(1)\right\} \\ &= \frac{1}{\lambda}\left\{\frac{1}{\lambda}\left\{\pi(z) - \delta_z(0)\right\} + \pi(z) - \delta_z(1)\right\} \\ &= \frac{1}{\lambda^2}\left\{\pi(z) - \delta_z(0)\right\} + \frac{1}{\lambda}\{\pi(z) - \delta_z(1)\}, \end{aligned}$$

$$\begin{aligned} g_z^\infty(3) &= \frac{1}{\lambda}\left\{2g_z^\infty(2) + \pi(z) - \delta_z(2)\right\} \\ &= \frac{1}{\lambda}\left\{2\frac{1}{\lambda^2}\left\{\pi(z) - \delta_z(0)\right\} + 2\frac{1}{\lambda}\{\pi(z) - \delta_z(1)\} + \pi(z) - \delta_z(2)\right\} \\ &= \frac{2}{\lambda^3}\left\{\pi(z) - \delta_z(0)\right\} + \frac{2}{\lambda^2}\{\pi(z) - \delta_z(1)\} + \frac{1}{\lambda}\{\pi(z) - \delta_z(2)\}, \end{aligned}$$

and so forth. We posit the general formula

$$g_z^\infty(y) = \frac{(y-1)!}{\lambda^y}\sum_{k=0}^{y-1}\frac{\lambda^k}{k!}\{\pi(z) - \delta_z(k)\}, \tag{4.4.20}$$

for $1 \le y \le n+1$ and $0 \le z \le n$.

The formula is straightforward to confirm by induction. Indeed it holds for $y = 1$ as can be seen in (4.4.19) (and recalling that $0! = 1$ by convention) and, assuming it holds for $y$, we have by (4.4.18)

$$\begin{aligned}
g_z^\infty(y+1) &= \frac{1}{\lambda}\{y g_z^\infty(y) + \pi(z) - \delta_z(y)\} \\
&= \frac{1}{\lambda}\left\{ y \frac{(y-1)!}{\lambda^y} \sum_{k=0}^{y-1} \frac{\lambda^k}{k!}\{\pi(z) - \delta_z(k)\} + \pi(z) - \delta_z(y) \right\} \\
&= \frac{y!}{\lambda^{y+1}} \sum_{k=0}^{y-1} \frac{\lambda^k}{k!}\{\pi(z) - \delta_z(k)\} + \frac{1}{\lambda}\{\pi(z) - \delta_z(y)\} \\
&= \frac{y!}{\lambda^{y+1}} \sum_{k=0}^{y} \frac{\lambda^k}{k!}\{\pi(z) - \delta_z(k)\},
\end{aligned}$$

as desired.

We rewrite (4.4.20) according to whether the term $\delta_z(y) = \mathbf{1}\{z = y\}$ plays a role in the equation. For $z \ge y > 0$, the equation simplifies to

$$\begin{aligned}
g_z^\infty(y) &= \frac{(y-1)!}{\lambda^y} \sum_{k=0}^{y-1} \frac{\lambda^k}{k!}\pi(z) \\
&= \frac{1}{y} \frac{y!}{e^{-\lambda}\lambda^y} \sum_{k=0}^{y-1} \frac{e^{-\lambda}\lambda^k}{k!}\pi(z) \\
&= \frac{\Pi(y-1)}{y\pi(y)}\pi(z).
\end{aligned}$$

For $0 \le z < y$, we get instead

$$\begin{aligned}
g_z^\infty(y) &= \frac{(y-1)!}{\lambda^y} \left\{ \left( \sum_{k=0}^{y-1} \frac{\lambda^k}{k!}\pi(z) \right) - \frac{\lambda^z}{z!} \right\} \\
&= \frac{1}{y} \frac{y!}{e^{-\lambda}\lambda^y} \left\{ \left( \sum_{k=0}^{y-1} \frac{e^{-\lambda}\lambda^k}{k!}\pi(z) \right) - \pi(z) \right\} \\
&= \frac{\Pi(y-1) - 1}{y\pi(y)}\pi(z).
\end{aligned}$$

The cases $z = n+1$ are analogous. That concludes the proof. ∎

*Proof of Lemma 4.4.14.* It suffices to prove that, for $A \subseteq \{0, 1, \ldots, n\}$ and $y \in \{0, 1, \ldots, n\}$,

$$|g_A^\infty(y+1) - g_A^\infty(y)| \leq (1 \wedge \lambda^{-1}), \tag{4.4.21}$$

and then use the triangle inequality.

We start with the case $y \geq 1$. We use the expression derived in Lemma 4.4.12. For $1 \leq y < z \leq n+1$,

$$\begin{aligned} g_z^\infty(y+1) - g_z^\infty(y) &= \frac{\Pi(y)}{(y+1)\pi(y+1)}\bar{\pi}(z) - \frac{\Pi(y-1)}{y\pi(y)}\bar{\pi}(z) \\ &= \bar{\pi}(z)\frac{1}{y\pi(y)}\left\{\frac{y}{\lambda}\Pi(y) - \Pi(y-1)\right\}, \end{aligned}$$

where we used that $\pi(y+1)/\pi(y) = \lambda/(y+1)$. We show that the expression in curly brackets is non-negative. Indeed, taking out the term $k' = 0$ in the first sum below and changing variables, we get

$$\begin{aligned} &\frac{y}{\lambda}\sum_{k'=0}^{y}\frac{e^{-\lambda}\lambda^{k'}}{(k')!} - \sum_{k=0}^{y-1}\frac{e^{-\lambda}\lambda^k}{k!} \\ &= \frac{y}{\lambda}e^{-\lambda} + \sum_{k=0}^{y-1}\frac{e^{-\lambda}\lambda^{(k+1)-1}}{(k+1)!/y} - \sum_{k=0}^{y-1}\frac{e^{-\lambda}\lambda^k}{k!} \\ &\geq \frac{y}{\lambda}e^{-\lambda} + \sum_{k=0}^{y-1}\frac{e^{-\lambda}\lambda^k}{k!} - \sum_{k=0}^{y-1}\frac{e^{-\lambda}\lambda^k}{k!} \\ &\geq 0. \end{aligned}$$

So $g_z^\infty(y+1) - g_z^\infty(y) \geq 0$ for $1 \leq y < z$. A similar calculation, which we omit, shows that the same inequality holds for $z < y \leq n$. The cases $y = 0$, which are analogous, are detailed below.

For notational convenience, it will be helpful to define $g_z^\infty(n+2) = 0$ for all $z$. Then, for $y = n+1$ and $z \leq n$, we get

$$g_z^\infty(n+2) - g_z^\infty(n+1) = 0 + \frac{1 - \Pi(n)}{(n+1)\pi(n+1)}\pi(z) \geq 0.$$

Moreover, by telescoping,

$$0 = g_z^\infty(n+2) - g_z^\infty(0) = \sum_{y=0}^{n+1}\{g_z^\infty(y+1) - g_z^\infty(y)\}.$$

We have argued that all the terms in this last sum are non-negative—with the sole exception of the term $y = z$. Hence, for a fixed $0 \leq z \leq n$, it must be that the maximum of $|g_z^\infty(y+1) - g_z^\infty(y)|$ is achieved at $y = z$. The case $z = n+1$ is analogous. By definition of $g_z^\infty$, for a fixed $0 \leq y \leq n$, it holds that $\sum_z \{g_z^\infty(y+1) - g_z^\infty(y)\} = 0$ and the maximum of $|g_A^\infty(y+1) - g_A^\infty(y)|$ over $A \subseteq \{0, 1, \ldots, n\}$ is achieved at $A = \{y\}$. It remains to bound that last case.

We have, using $\pi(y+1)/\pi(y) = \lambda/(y+1)$ again, that for $1 \leq y \leq n$

$$\begin{aligned}
&|g_y^\infty(y+1) - g_y^\infty(y)| \\
&= \left| -\frac{1 - \Pi(y)}{(y+1)\pi(y+1)}\pi(y) - \frac{\Pi(y-1)}{y\pi(y)}\pi(y) \right| \\
&= \frac{1}{\lambda} \sum_{k \geq y+1} e^{-\lambda} \frac{\lambda^k}{k!} + \frac{1}{y} \sum_{k=0}^{y-1} e^{-\lambda} \frac{\lambda^k}{k!} \\
&= \frac{e^{-\lambda}}{\lambda} \left\{ \sum_{k'=1}^{y} \frac{\lambda^{k'}}{(k')!} \frac{k'}{y} + \sum_{k \geq y+1} \frac{\lambda^k}{k!} \right\} \\
&\leq \frac{e^{-\lambda}}{\lambda} \left\{ \sum_{k \geq 1} \frac{\lambda^k}{k!} \right\} \\
&= \frac{e^{-\lambda}}{\lambda} \left\{ e^{\lambda} - 1 \right\} \\
&= \frac{1 - e^{-\lambda}}{\lambda}.
\end{aligned}$$

For $\lambda \geq 1$, we have $\frac{1-e^{-\lambda}}{\lambda} \leq \frac{1}{\lambda} = (1 \wedge \lambda^{-1})$, while for $0 < \lambda < 1$ we have $\frac{1-e^{-\lambda}}{\lambda} \leq \frac{\lambda}{\lambda} = 1 = (1 \wedge \lambda^{-1})$ by Exercise 1.16.

It remains to consider the cases $y = 0$. Recall that $g_z^\infty(0) = 0$. By Lemma 4.4.12, for $z \geq 1$,

$$\begin{aligned}
g_z^\infty(1) - g_z^\infty(0) &= g_z^\infty(1) \\
&= \frac{\Pi(0)}{\pi(1)} \bar{\pi}(z) \\
&\geq 0.
\end{aligned}$$

And

$$g_0^\infty(1) - g_0^\infty(0) = g_0^\infty(1) = -\frac{1 - \Pi(0)}{\pi(1)}\pi(0) = -\frac{1 - e^{-\lambda}}{\lambda}.$$

So we have established (4.4.21) and that concludes the proof. ■

### 4.4.3 ▷ *Random graphs: clique number at the threshold in the Erdős-Rényi model*

We revisit the subgraph containment problem of Section 2.3.2 (and Section 4.2.4). Let $G_n \sim \mathbb{G}_{n,p_n}$ be an Erdős-Rényi graph with $n$ vertices and density $p_n$. Let $\omega(G)$ be the *clique number* of a graph $G$, that is, the size of its largest clique. We showed previously that the property $\omega(G) \geq 4$ has threshold function $n^{-2/3}$. Here we consider what happens when

$$p_n = Cn^{-2/3},$$

for some constant $C > 0$. We use the Chen-Stein method in the form of Theorem 4.4.8.

For an enumeration $S_1, \ldots, S_m$ of the 4-tuples of vertices in $G_n$, let $A_1, \ldots, A_m$ be the events that the corresponding 4-cliques are present and define $Z_i = \mathbf{1}_{A_i}$. Then $W = \sum_{i=1}^m Z_i$ is the number of 4-cliques in $G_n$. We argued previously (see Claim 2.3.4) that

$$q_i := \mathbb{E}[Z_i] = p_n^6,$$

and

$$\lambda := \mathbb{E}[W] = \binom{n}{4} p_n^6.$$

In our regime of interest, $\lambda$ is of constant order.

Observe that the $Z_i$s are not independent because the 4-tuples may share potential edges. However they admit a neighborhood structure as in Theorem 4.4.8. Specifically, for $i = 1, \ldots, m$, define

$$\mathcal{N}_i = \{j : S_i \text{ and } S_j \text{ share at least two vertices}\} \setminus \{i\}.$$

Then the conditions of Theorem 4.4.8 are satisfied, that is, $X_i$ is independent of $\{Z_j : j \notin \mathcal{N}_i \cup \{i\}\}$. We argued previously (again see Claim 2.3.4) that

$$|\mathcal{N}_i| = \binom{4}{3}(n-4) + \binom{4}{2}\binom{n-4}{2} = \Theta(n^2),$$

where the first term counts the number of $S_j$s sharing exactly three vertices with $S_i$, in which case $\mathbb{E}[Z_i Z_j] = p_n^9$, and the second term counts those sharing two, in which case $\mathbb{E}[Z_i Z_j] = p_n^{11}$.

We are ready to apply the bound in Theorem 4.4.8. Let $\pi$ be the Poisson distribution with mean $\lambda$. Using the formulas above, we get when $p_n = Cn^{-2/3}$

$$\begin{aligned}
&\|\mu - \pi\|_{\mathrm{TV}} \\
&\leq (1 \wedge \lambda^{-1}) \sum_{i=1}^{n} \left\{ q_i^2 + \sum_{j \in \mathcal{N}_i} (q_i q_j + \mathbb{E}[Z_i Z_j]) \right\} \\
&\leq (1 \wedge \lambda^{-1}) \binom{n}{4} \\
&\quad \times \left[ p_n^{12} + \left\{ \binom{4}{3}(n-4)(p_n^{12} + p_n^{9}) + \binom{4}{2}\binom{n-4}{2}(p_n^{12} + p_n^{11}) \right\} \right] \\
&= (1 \wedge \lambda^{-1})\, \Theta(n^4 p_n^{12} + n^5 p_n^9 + n^6 p_n^{11}) \\
&= (1 \wedge \lambda^{-1})\, \Theta(n^4 n^{-8} + n^5 n^{-6} + n^6 n^{-22/3}) \\
&= (1 \wedge \lambda^{-1})\, \Theta(n^{-1}),
\end{aligned}$$

which goes to 0 as $n \to +\infty$.

See Exercise 4.21 for an improved bound.

## Exercises

**Exercise 4.1** (Harmonic function on $\mathbb{Z}^d$: unbounded)**.** Give an example of an unbounded harmonic function on $\mathbb{Z}$. Give one on $\mathbb{Z}^d$ for general $d$. [Hint: What is the simplest function after the constant one?]

**Exercise 4.2** (Binomial vs. Binomial)**.** Use coupling to show that

$$n \geq m,\ q \geq p \implies \mathrm{Bin}(n, q) \succeq \mathrm{Bin}(m, p).$$

**Exercise 4.3** (A chain that is not stochastically monotone)**.** Consider random walk on a network $\mathcal{N} = ((V, E), c)$ where $V = \{0, 1, \ldots, n\}$ and $i \sim j$ if and only if $|i - j| = 1$ (in particular, not including self-loops). Show that the transition matrix is, in general, *not* stochastically monotone (see Definition 4.2.16).

**Exercise 4.4** (Increasing events: properties)**.** Let $\mathcal{F}$ be a $\sigma$-algebra over the poset $\mathcal{X}$. Recall that an event $A \in \mathcal{F}$ is increasing if $x \in A$ implies that any $y \geq x$ is also in $A$ and that a function $f : \mathcal{X} \to \mathbb{R}$ is increasing if $x \leq y$ implies $f(x) \leq f(y)$.

(i) Show that an event $A \in \mathcal{F}$ is increasing if and only if the indicator function $\mathbf{1}_A$ is increasing.

(ii) Let $A, B \in \mathcal{F}$ be increasing. Show that $A \cap B$ and $A \cup B$ are increasing.

(iii) An event $A$ is decreasing if $x \in A$ implies that any $y \leq x$ is also in $A$. Show that $A$ is decreasing if and only if $A^c$ is increasing.

(iv) Let $A, B \in \mathcal{F}$ be decreasing. Show that $A \cap B$ and $A \cup B$ are decreasing.

**Exercise 4.5** (Harris' inequality: alternative proof)**.** We say that $f : \mathbb{R}^n \to \mathbb{R}$ is *coordinatewise nondecreasing* if it is nondecreasing in each variable while keeping the other variables fixed.

(i) *(Chebyshev's association inequality)* Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be coordinatewise nondecreasing and let $X$ be a real random variable. Show that

$$\mathbb{E}[f(X)g(X)] \geq \mathbb{E}[f(X)]\mathbb{E}[g(X)].$$

[Hint: Consider the quantity $(f(X) - f(X'))(g(X) - g(X'))$ where $X'$ is an independent copy of $X$.]

(ii) *(Harris' inequality)* Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ be coordinatewise nondecreasing and let $X = (X_1, \ldots, X_n)$ be independent real random variables. Show by induction on $n$ that

$$\mathbb{E}[f(X)g(X)] \geq \mathbb{E}[f(X)]\mathbb{E}[g(X)].$$

[Hint: You may need Lemma B.6.15.]

**Exercise 4.6.** Provide the details for Example 4.2.33.

**Exercise 4.7** (FKG: sufficient conditions)**.** Let $\mathcal{X} := \{0,1\}^F$ where $F$ is finite and let $\mu$ be a positive probability measure on $\mathcal{X}$. We use the notation introduced in the proof of Holley's inequality (Theorem 4.2.32).

(i) To check the FKG condition, show that it suffices to check that, for all $x \leq y \in \mathcal{X}$ and $i \in F$,
$$\frac{\mu(y^{i,1})}{\mu(y^{i,0})} \geq \frac{\mu(x^{i,1})}{\mu(x^{i,0})}.$$
[Hint: Write $\mu(\omega \vee \omega')/\mu(\omega)$ as a telescoping product.]

(ii) To check the FKG condition, show that it suffices to check (4.2.15) only for those $\omega, \omega' \in \mathcal{X}$ such that $\|\omega - \omega'\|_1 = 2$ and neither $\omega \leq \omega'$ nor $\omega' \leq \omega$. [Hint: Use (i).]

**Exercise 4.8** (FKG and strong positive associations)**.** Let $\mathcal{X} := \{0,1\}^F$ where $F$ is finite and let $\mu$ be a positive probability measure on $\mathcal{X}$. For $\Lambda \subseteq F$ and $\xi \in \mathcal{X}$, let
$$\mathcal{X}^\xi_\Lambda := \{\omega_\Lambda \times \xi_{\Lambda^c} \,:\, \omega_\Lambda \in \{0,1\}^\Lambda\},$$
where $\omega_\Lambda \times \xi_{\Lambda^c}$ agrees with $\omega$ on coordinates in $\Lambda$ and with $\xi$ on coordinates in $F\backslash\Lambda$. Define the measure $\mu^\xi_\Lambda$ over $\{0,1\}^\Lambda$ as
$$\mu^\xi_\Lambda(\omega_\Lambda) := \frac{\mu(\omega_\Lambda \times \xi_{\Lambda^c})}{\mu(\mathcal{X}^\xi_\Lambda)}.$$

That is, $\mu^\xi_\Lambda$ is $\mu$ conditioned on agreeing with $\xi$ on $F\backslash\Lambda$. The measure $\mu$ is said to be *strongly positively associated* if $\mu^\xi_\Lambda(\omega_\Lambda)$ is positively associated for all $\Lambda$ and $\xi$. Prove that the FKG condition is equivalent to strong positive associations. [Hint: Use Exercise 4.7 as well as the FKG inequality.]

**Exercise 4.9** (Triangle-freeness: a second proof)**.** Consider again the setting of Section 4.2.4.

(i) Let $e_t$ be the minimum number of edges in a $t$-vertex union of $k$ not mutually vertex-disjoint triangles. Show that, for any $k \geq 2$ and $k \leq t < 3k$, it holds that $e_t > t$.

(ii) Use Exercise 2.18 to give a second proof of the fact that $\mathbb{P}[X_n = 0] \to e^{-\lambda^3/6}$.

**Exercise 4.10** (RSW lemma: general $\alpha$)**.** Let $R_{n,\alpha}(p)$ be as defined in Section 4.2.5. Show that for all $n \geq 2$ (divisible by 4) and $p \in (0,1)$

$$R_{n,\alpha}(p) \geq \left(\frac{1}{2}\right)^{2\alpha-2} R_{n,1}(p)^{6\alpha-7} R_{n/2,1}(p)^{6\alpha-6}.$$

**Exercise 4.11** (Primal and dual crossings)**.** Modify the proof of Lemma 2.2.14 to prove Lemma 4.2.41.

**Exercise 4.12** (Square-root trick)**.** Let $\mu$ be an FKG measure on $\{0,1\}^F$ where $F$ is finite. Let $A_1$ and $A_2$ be increasing events with $\mu(A_1) = \mu(A_2)$. Show that

$$\mu(A_1) \geq 1 - \sqrt{1 - \mu(A_1 \cup A_2)}.$$

**Exercise 4.13** (Splitting: details)**.** Show that $\tilde{P}$, as defined in Example 4.3.3, is a transition matrix on $V$ provided $z_0$ satisfies the condition there.

**Exercise 4.14** (Doeblin's condition in finite case)**.** Let $P$ be a transition matrix on a finite state space.

(i) Show that Doeblin's condition (see Example 4.3.3) holds when $P$ is finite, irreducible and aperiodic.

(ii) Show that Doeblin's condition holds for lazy random walk on the hypercube with $s = n$. Use it to derive a bound on the mixing time.

**Exercise 4.15** (Mixing on cycles: lower bound)**.** Let $(Z_t)$ be lazy, simple random walk on the cycle of size $n$, $\mathbb{Z}_n := \{0, 1 \ldots, n-1\}$, where $i \sim j$ if $|j - i| = 1 \pmod{n}$. Assume $n$ is divisible by 4 and fix $0 < \varepsilon < 1/2$.

(i) Let $A = \{n/2, \ldots, n-1\}$. By coupling $(Z_t)$ with lazy, simple random walk on $\mathbb{Z}$, show that

$$P^{\alpha n^2}(n/4, A) < \frac{1}{2} - \varepsilon,$$

for $\alpha \leq \alpha_\varepsilon$ for some $\alpha_\varepsilon > 0$. [Hint: You may want to use Chebyshev's inequality (Theorem 2.1.2) or Kolmogorov's maximal inequality (Corollary 3.1.46).]

(ii) Deduce that

$$t_{\mathrm{mix}}(\varepsilon) \geq \alpha_\varepsilon n^2.$$

**Exercise 4.16** (Lower bound on mixing: distinguishing statistic)**.** Let $X$ and $Y$ be random variables on a finite state space $S$. Let $h : S \to \mathbb{R}$ be a measurable real-valued map. Assume that

$$\mathbb{E}[h(Y)] - \mathbb{E}[h(X)] \geq r\sigma,$$

where $r > 0$ and $\sigma^2 := \max\{\mathrm{Var}[h(X)], \mathrm{Var}[h(Y)]\}$. Show that

$$\|\mu_X - \mu_Y\|_{\mathrm{TV}} \geq 1 - \frac{8}{r^2}.$$

[Hint: Consider the interval on one side of the midpoint between $\mathbb{E}[h(X)]$ and $\mathbb{E}[h(Y)]$.]

**Exercise 4.17** (Path coupling and optimal transport)**.** Let $V$ be a finite state space and let $P$ be an irreducible transition matrix on $V$ with stationary distribution $\pi$. Let $w_0$ be a metric on $V$. For probability measures $\mu, \nu$ on $V$, let

$$W_0(\mu, \nu) := \inf\left\{\mathbb{E}[w_0(X,Y)] \,:\, (X,Y) \text{ is a coupling of } \mu \text{ and } \nu\right\},$$

be the so-called *Wasserstein distance* (or *transportation metric*) between $\mu$ and $\nu$.

*Wasserstein distance*

(i) Show that $W_0$ is a metric. [Hint: See the proof of Claim 4.3.11.]

(ii) Assume that the conditions of Theorem 4.3.10 hold. Show that for any probability measures $\mu, \nu$

$$W_0(\mu P, \nu P) \leq \kappa\, W_0(\mu, \nu).$$

(iii) Use (i) and (ii) to prove Theorem 4.3.10.

**Exercise 4.18** (Stein equation for the Poisson distribution)**.** Let $\lambda > 0$. Show that a non-negative integer-valued random variable $Z$ is $\mathrm{Poi}(\lambda)$ if and only if for all $g$ bounded

$$\mathbb{E}[\lambda g(Z+1) - Zg(Z)] = 0.$$

**Exercise 4.19** (Laplacian and stationarity)**.** Let $P$ be an irreducible transition matrix on a finite or countably infinite spate space $V$. Recall the Laplacian operator is

$$\Delta f(x) = \left[\sum_y P(x,y) f(y)\right] - f(x),$$

provided the sum if finite. Show that a probability distribution $\mu$ over $V$ is stationary for $P$ if and only if for all bounded measurable functions

$$\sum_{x \in V} \mu(x) \Delta f(x) = 0.$$

**Exercise 4.20** (Chen-Stein method for positively related variables)**.** Using the notation in (4.4.1), (4.4.2) and (4.4.3), suppose that for each $i$ we can construct a coupling $\{(X_j^{(i)} : j = 1, \ldots, n), (Y_j^{(i)} : j \neq i)\}$ with $(X_j^{(i)})_j \sim (X_j)_j$ such that

$$(Y_j^{(i)}, j \neq i) \sim (X_j^{(i)}, j \neq i) | X_i^{(i)} = 1 \qquad \text{and} \qquad Y_j^{(i)} \geq X_j^{(i)}, \ \forall j \neq i.$$

Show that

$$\|\mu - \pi\|_{\mathrm{TV}} \leq (1 \wedge \lambda^{-1}) \left\{ \mathrm{Var}(W) - \lambda + 2 \sum_{i=1}^{n} p_i^2 \right\}.$$

**Exercise 4.21** (Chen-Stein and 4-cliques)**.** Use Exercise 4.20 to give an improved asymptotic bound in the setting of Section 4.4.3.

**Exercise 4.22** (Chen-Stein for negatively related variables)**.** Using the notation in (4.4.1), (4.4.2) and (4.4.3), suppose that for each $i$ we can construct a coupling $\{(X_j^{(i)} : j = 1, \ldots, n), (Y_j^{(i)} : j \neq i)\}$ with $(X_j^{(i)})_j \sim (X_j)_j$ such that

$$(Y_j^{(i)}, j \neq i) \sim (X_j^{(i)}, j \neq i) | X_i^{(i)} = 1 \qquad \text{and} \qquad Y_j^{(i)} \leq X_j^{(i)}, \forall j \neq i.$$

Show that

$$\|\mu - \pi\|_{\mathrm{TV}} \leq (1 \wedge \lambda^{-1}) \{\lambda - \mathrm{Var}(W)\}.$$

## Bibliographic remarks

**Section 4.1** The coupling method is generally attributed to Doeblin [Doe38]. The standard reference on coupling is [Lin02]. See that reference for a history of coupling and a facsimile of Doeblin's paper. See also [dH]. Section 4.1.2 is based on [Per, Section 6] and Section 4.1.4 is based on [vdH17, Section 5.3].

**Section 4.2** Strassen's theorem is due to Strassen [Str65]. Harris' inequality is due to Harris [Har60]. The FKG inequality is due to Fortuin, Kasteleyn, and Ginibre [FKG71]. A "four-function" version of Holley's inequality, which also extends to distributive lattices, was proved by Ahlswede and Daykin [AD78]. See for example [AS11, Section 6.1]. An exposition of submodularity and its connections to convexity can be found in [Lov83]. For more on Markov random fields, see for example [RAS15]. Section 4.2.4 follows [AS11, Sections 8.1, 8.2, 10.1]. Janson's inequality is due to Janson [Jan90]. Boppana and Spencer [BS89] gave the proof presented here. For more on Janson's inequality, see [JLR11, Section 2.2]. The presentation in Section 4.2.5 follows closely [BR06b, Sections 3 and 4]. See also [BR06a, Chapter 3]. Broadbent and Hammersley [BH57, Ham57] initiated the study of the critical value of percolation. Harris' theorem was proved by Harris [Har60] and Kesten's theorem was proved two decades later by Kesten [Kes80], confirming non-rigorous work of Sykes and Essam [SE64]. The RSW lemma was obtained independently by Russo [Rus78] and Seymour and Welsh [SW78]. The proof we gave here is due to Bollobás and Riordan [BR06b]. Another short proof of a version of the RSW lemma for critical site percolation on a triangular lattice was given by Smirnov; see for example [Ste]. The type of "scale invariance" seen in the RSW lemma plays a key role in the contemporary theory of critical two-dimensional percolation and of two-dimensional lattice models more generally. See for example [Law05, Gri10a].

**Section 4.3** The material in Section 4.3 borrows heavily from [LPW06, Chapters 5, 14, 15] and [AF, Chapter 12]. Aldous [Ald83] was the first author to make explicit use of coupling to bound total variation distance to stationarity of finite Markov chains. The link between couplings of Markov chains and total variation distance was also used by Griffeath [Gri75] and Pitman [Pit76]. Example 4.3.3 is based on [Str14] and [JH01]. For a more general treatment, see [MT09, Chapter 16]. The proof of Claim 4.3.7 is partly based on [LPW06, Proposition 7.13]. See also [DGG$^+$00] and [HS07] for alternative proofs. Path coupling is due to Bubley and Dyer [BD97]. The optimal transport perspective on the path coupling method in Exercise 4.17 is from [LPW06, Chapter 14]. For more on optimal transport, see for example [Vil09]. The main result in Section 4.3.4 is taken from [LPW06, Theorem 15.1]. For more background on the so-called "critical slowdown" of the Glauber dynamics of Ising and Potts models on various graphs, see [CDL$^+$12, LS12].

**Section 4.4** The Chen-Stein method was introduced by Chen in [Che75] as an adaptation of the Stein method [Ste72] to the Poisson distribution. The presentation in Section 4.4 is inspired heavily by [Dey] and [vH16]. Example 4.4.9 is taken from [AGG89]. Further applications of the Chen-Stein and Stein methods to random graphs can be found in [JLR11, Chapter 6].