Chapter 4

Coupling

Coupling is a quintessential probabilistic technique with a wide range of applications. The idea behind the coupling method is that, 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 1.22), one simultaneously constructs two copies of a Markov chain—one of which is at stationarity—and shows that they can be made to coincide after a random amount of time called the coupling time.

In this chapter, we discuss a number of applications of the coupling method in discrete probability, including applications of the important concept of stochastic domination and its connection to correlation inequalities.

4.1 Background

Throughout this section, we will denote by $\mu_Z$ the law of random variable $Z$.

4.1.1 Basic definitions

A formal definition of coupling follows. Recall 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 $\sigma$-field $S_1 \times S_2$ 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 (Coupling). Let $\mu$ and $\nu$ be probability measures on the same measurable space $(S, S)$. A coupling of $\mu$ and $\nu$ is a probability measure $\gamma$ on the product space $(S \times S, S \times S)$ such that the marginals of $\gamma$ coincide with $\mu$ and $\nu$, i.e.,

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

Similarly, for two random variables $X$ and $Y$ taking values in $(S, S)$, a coupling of $X$ and $Y$ is a joint variable $(X', Y')$ taking values in $(S \times S, S \times S)$ whose law 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 4.2 (Coupling of Bernoulli variables). Let $X$ and $Y$ be Bernoulli random variables with parameters $0 \leq q < r \leq 1$ respectively. That is, $P[X = 0] = 1 - q$ and $P[X = 1] = q$, and similarly for $Y$. Here $S = \{0, 1\}$ and $S = 2^S$.

- (Independent coupling) One coupling of $X$ and $Y$ is $(X', Y')$ where $X'$ and $Y'$ are independent. Its law is

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

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

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

Example 4.3 (Bond percolation: monotonicity). Let $G = (V, E)$ be a countable graph. Denote by $\mathbb{P}_p$ the law of bond percolation on $G$ with density $p$. Let $x \in V$ and assume $0 \leq q < r \leq 1$. Using the 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 $U_e \leq p$. 

199
Thinking of $W_p$ as specifying the open edges in the percolation process on $G$ under $P_p$, we see that $(W_q, W_r)$ is a coupling of $P_q$ and $P_r$ with the property that $P[W_q \subseteq W_r] = 1$. Let $C_x^{(q)}$ and $C_x^{(r)}$ be the open clusters of $x$ under $W_q$ and $W_r$ respectively. Because $C_x^{(q)} \subseteq C_x^{(r)}$,

$$\theta(q) := P_q[|C_x| = +\infty] = P[|C_x^{(q)}| = +\infty] \leq P[|C_x^{(r)}| = +\infty] = P_r[|C_x| = +\infty] = \theta(r),$$

as claimed in Section 2.2.4.

**Example 4.4** (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. Assume $0 \leq q < r \leq 1$. Using again the coupling of Bernoulli variables above we produce a coupling of $S_t^{(q)}$ and $S_t^{(r)}$.

- Let $(X_t^n, Y_t^n)_i$ be an infinite sequence of i.i.d. monotone Bernoulli couplings with parameters $q$ and $r$ respectively.
- Define $(Z_t^{(q)}, Z_t^{(r)}):= (2X_t^n - 1, 2Y_t^n - 1)$. Note that $P[2X_t^n - 1 = 1] = P[X_t^n = 1] = q$ and $P[2X_t^n - 1 = -1] = P[X_t^n = 0] = 1 - q$.
- 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 $(S_t^{(q)}, S_t^{(r)})$ such that $\hat{S}_t^{(q)} \leq \hat{S}_t^{(r)}$ for all $n$ almost surely. So for all $y$ and all $t$

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

**4.1.2 Harmonic functions on lattices and trees**

Let $(X_t)$ be a Markov chain on a (finite or) countable state space $V$ with transition matrix $P$ and let $P_x$ be the law of $(X_t)$ started at $x$.* Recall that a function $h : V \rightarrow \mathbb{R}$ is $P$-harmonic on $V$ (or harmonic for short) if

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

We first give a coupling-based criterion for harmonic functions to be constant.

---

*Requires: Section 3.3.1.
Lemma 4.5 (Coupling and bounded harmonic functions). If, for all \(y, z \in V\), there is a coupling \(((Y_t), (Z_t))\) of \(\mathbb{P}_y\) and \(\mathbb{P}_z\) such that
\[
\lim_{t \to \infty} \mathbb{P}[Y_t \neq Z_t] = 0,
\]
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 \(V\). By harmonicity, \((h(Y_t))\) 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 and the boundedness assumption
\[
|h(y) - h(z)| = |\mathbb{E}[h(Y_t)] - \mathbb{E}[h(Z_t)]| \leq \mathbb{E}|h(Y_t) - h(Z_t)| \leq 2M \mathbb{P}[Y_t \neq Z_t] \to 0.
\]
So \(h(y) = h(z)\).

Harmonic functions on \(\mathbb{Z}^d\) Consider simple random walk on \(\mathbb{Z}^d, d \geq 1\). In that case, we show that all bounded harmonic functions are constant.

**Theorem 4.6** (Bounded harmonic functions on \(\mathbb{Z}^d\)). All bounded harmonic functions on \(\mathbb{Z}^d\) are constant.

**Proof.** Clearly, \(h\) is harmonic with respect to simple random walk if and only if it is harmonic with respect to lazy simple random walk. Let \(\mathbb{P}_y\) and \(\mathbb{P}_z\) be the laws of lazy simple random walk on \(\mathbb{Z}^d\) started at \(y\) and \(z\). We construct a coupling \(((Y_t), (Z_t)) = ((Y^{(i)}_t)_{i \in [d]}, (Z^{(i)}_t)_{i \in [d]})\) of \(\mathbb{P}_y\) and \(\mathbb{P}_z\) as follows: at time \(t\), pick a coordinate \(I \in [d]\) uniformly at random, then
- if \(Y^{(I)}_t = Z^{(I)}_t\) then do nothing with probability \(1/2\) and otherwise pick \(W \in \{-1, +1\}\) uniformly at random, set \(Y_{t+1}^{(I)} = Z_{t+1}^{(I)} := Z^{(I)}_t + W\) and leave the other coordinates unchanged;
- if instead \(Y^{(I)}_t \neq Z^{(I)}_t\), pick \(W \in \{-1, +1\}\) uniformly at random, and with probability \(1/2\) set \(Y_{t+1}^{(I)} := Y^{(I)}_t + W\) and leave \(Z_t\) and the other coordinates of \(Y_t\) unchanged, or otherwise set \(Z_{t+1}^{(I)} := Z^{(I)}_t + W\) 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 random walk on \(\mathbb{Z}\) started at \(y^{(i)} - z^{(i)}\) with holding probability \(1 - \frac{1}{d}\) until it hits 0. Simple random walk on \(\mathbb{Z}\) is irreducible and recurrent. The holding probability does not affect the type of the walk, as can be seen for instance from the characterization in terms of effective resistance. 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.

### 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 4.7** (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 \(P\)-harmonic function on \(\mathbb{T}_d\).

**Proof.** The function \(h\) is clearly bounded and by the usual one-step trick
\[
h(x) = \sum_{y \sim x} \frac{1}{d} \mathbb{P}_y[X_t \in T_0 \text{ for all but finitely many } t] = \sum_y P(x, y)h(y),
\]
so \(h\) is \(P\)-harmonic.

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

**Lemma 4.8.**
\[
q := \mathbb{P}_a[\tau_\rho = +\infty] = \mathbb{P}_b[\tau_\rho = +\infty] > 0.
\]
Proof. The second equality 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 \( 0 \) 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 \( 1 - \frac{1}{d} > 2 \) when \( d \geq 3 \).

Note that
\[
\nu(q) \leq \left( 1 - \frac{1}{d} \right) (1 - q) < 1.
\]
Indeed if on the first step the random walk started at \( \nu \) moves away from \( a \), an event of probability \( 1 - \frac{1}{d} \), then it must come back to \( \nu \) in finite time to reach \( T_a \). Similarly, by the strong Markov property,
\[
\nu(a) = \nu + (1 - \nu) \nu(q).
\]
Since \( \nu(q) \neq 1 \) and \( q > 0 \), this shows that \( \nu(a) > \nu(q) \).

4.2 Coupling inequality

In the examples above, we used coupling to prove monotonicity statements. Coupling is also useful to bound the distance between probability measures.

4.2.1 Bounding the total variation distance via coupling

Let \( \mu \) and \( \nu \) be probability measures on \((S, \mathcal{S})\). Recall the definition of the total variation distance
\[
\|\mu - \nu\|_{TV} := \sup_{A \in \mathcal{S}} |\mu(A) - \nu(A)|.
\]

Lemma 4.9 (Coupling inequality). Let \( \mu \) and \( \nu \) be probability measures on \((S, \mathcal{S})\).

For any coupling \((X, Y)\) of \( \mu \) and \( \nu \),
\[
\|\mu - \nu\|_{TV} \leq \mathbb{P}[X \neq Y].
\]

Proof. For any \( A \in \mathcal{S} \),
\[
\mu(A) - \nu(A) = \mathbb{P}[X \in A] - \mathbb{P}[Y \in A]
= \mathbb{P}[X \in A, X = Y] + \mathbb{P}[X \in A, X \neq Y]
- \mathbb{P}[Y \in A, X = Y] - \mathbb{P}[Y \in A, X \neq Y]
= \mathbb{P}[X \in A, X \neq Y] - \mathbb{P}[Y \in A, X \neq Y]
\leq \mathbb{P}[X \neq Y],
\]

203
and, similarly, \( \nu(A) - \mu(A) \leq \mathbb{P}[X \neq Y] \). Hence
\[
|\mu(A) - \nu(A)| \leq \mathbb{P}[X \neq Y].
\]

Here is a simple example.

**Example 4.10** (A coupling of Poisson random variables). Let \( X \sim \text{Poi}(\lambda) \) and \( Y \sim \text{Poi}(\nu) \) with \( \lambda > \nu \). Recall that a sum of independent Poissons is Poisson. This fact leads to a natural coupling: let \( \hat{Y} \sim \text{Poi}(\nu) \), \( \hat{Z} \sim \text{Poi}(\lambda - \nu) \) independently of \( Y \), and \( \hat{X} = \hat{Y} + \hat{Z} \). Then \((\hat{X}, \hat{Y})\) is a coupling and
\[
\|\mu_X - \mu_Y\|_{TV} \leq \mathbb{P}[\hat{X} \neq \hat{Y}] = \mathbb{P}[\hat{Z} > 0] = 1 - e^{-(\lambda - \nu)} \leq \lambda - \nu.
\]

In fact, the inequality in Lemma 4.9 is tight. For simplicity, we prove this in the finite case only.

**Lemma 4.11** (Maximal coupling). Assume \( S \) is finite and let \( S = 2^S \). Let \( \mu \) and \( \nu \) be probability measures on \((S, S)\). Then,
\[
\|\mu - \nu\|_{TV} = \inf \{ \mathbb{P}[X \neq Y] : \text{coupling } (X, Y) \text{ of } \mu \text{ and } \nu \}.
\]

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

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

First, two lemmas. See Figure 4.1 for a proof by picture.

**Lemma 4.12.**
\[
\sum_{x \in S} \mu(x) \wedge \nu(x) = 1 - \|\mu - \nu\|_{TV}.
\]
Figure 4.1: Proof by picture that: $1 - p = \alpha = \beta = \|\mu - \nu\|_{\text{TV}}$.

Proof. We have

$$2\|\mu - \nu\|_{\text{TV}} = \sum_{x \in S} |\mu(x) - \nu(x)|$$

$$= \sum_{x \in A} [\mu(x) - \nu(x)] + \sum_{x \in B} [\nu(x) - \mu(x)]$$

$$= \sum_{x \in A} \mu(x) + \sum_{x \in B} \nu(x) - \sum_{x \in S} \mu(x) \wedge \nu(x)$$

$$= 2 - \sum_{x \in B} \mu(x) - \sum_{x \in A} \nu(x) - \sum_{x \in S} \mu(x) \wedge \nu(x)$$

$$= 2 - 2 \sum_{x \in S} \mu(x) \wedge \nu(x).$$

Lemma 4.13.

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

Proof. The first equality is immediate. The second equality follows from the second line in the proof of the previous 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 $X$ at $x \in S$ is
\[ p \gamma_{\min}(x) + (1 - p) \gamma_A(x) = \mu(x), \]
and similarly for $Y$. Finally $\mathbb{P}[X \neq Y] = 1 - p = \|\mu - \nu\|_{TV}$.

\[ \blacksquare \]

**Remark 4.14.** 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 4.15** (Coupling of Bernoulli variables (continued)). Let $X$ and $Y$ be Bernoulli random variables with parameters $0 \leq q < r \leq 1$ respectively. That is, $\mathbb{P}[X = 0] = 1 - q$ and $\mathbb{P}[X = 1] = q$, and similarly for $Y$. Here $S = \{0, 1\}$ and $\mathcal{S} = 2^S$. Let $\mu$ and $\nu$ be the laws of $X$ and $Y$ respectively. To construct the maximal coupling as above, we note that
\[ p := \sum_x \mu(x) \wedge \nu(x) = (1 - r) + q, \quad 1 - p = \alpha = \beta := r - q, \]
\[ A := \{0\}, \quad B := \{1\}, \]
\[ (\gamma_{\min}(x))_{x=0,1} = \left( \frac{1 - r}{(1 - r) + q}, \frac{q}{(1 - r) + q} \right), \quad \gamma_A(0) := 1, \quad \gamma_B(1) := 1. \]

The law of the maximal coupling $(X^m, Y^m)$ is given by
\[ \left( \mathbb{P}[(X^m, Y^m) = (i, j)] \right)_{i,j \in \{0,1\}} = \begin{pmatrix} p \gamma_{\min}(0) & (1 - p) \gamma_A(0) \gamma_B(1) \\ 0 & p \gamma_{\min}(1) \end{pmatrix} \]
\[ = \begin{pmatrix} 1 - r & r - q \\ 0 & q \end{pmatrix}, \]
which coincides with the monotone coupling. \[ \blacksquare \]
**Poisson approximation**  Here is a classical application of coupling. 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 \( S_n := \sum_{i \leq n} X_i \). 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 \( Z_n := \sum_{i \leq n} W_i \). We choose \( \lambda_i = - \log(1 - p_i) \) so as to ensure

\[
(1 - p_i) = P[X_i = 0] = P[W_i = 0] = e^{-\lambda_i}.
\]

Note that \( Z_n \sim \text{Poi}(\lambda) \) where \( \lambda = \sum_{i \leq n} \lambda_i \).

**Theorem 4.16** (Poisson approximation).

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

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

\[
W'_i \sim \text{Poi}(\lambda_i) \quad \text{and} \quad X'_i = W'_i \wedge 1.
\]

Because \( \lambda_i = -\log(1 - p_i) \), \( (X'_i, W'_i) \) is 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

\[
\|\mu_{S_n} - \mu_{Z_n}\|_{TV} \leq P[S'_n \neq Z'_n] \leq \sum_{i \leq n} P[X'_i \neq W'_i] = \sum_{i \leq n} P[W'_i \geq 2]
\]

\[
= \sum_{i \leq n} \sum_{j \geq 2} e^{-\lambda_i} \frac{\lambda_i^j}{j!} \leq \sum_{i \leq n} \frac{\lambda_i^2}{2} \sum_{\ell \geq 0} e^{-\lambda_i} \frac{\lambda_i^\ell}{\ell!} = \sum_{i \leq n} \frac{\lambda_i^2}{2}.
\]

Maps reduce total variation distance  The following lemma will be useful.

**Lemma 4.17** (Mappings). Let \( X \) and \( Y \) be random variables taking values in \((S, S)\), let \( h \) be a measurable map from \((S, S)\) to \((S', S')\), and let \( X' := h(X) \) and \( Y' := h(Y) \). It holds that

\[
\|\mu_{X'} - \mu_{Y'}\|_{TV} \leq \|\mu_X - \mu_Y\|_{TV}.
\]

**Proof.** It holds that

\[
\sup_{A' \in S'} |P[X' \in A'] - P[Y' \in A']| = \sup_{A' \in S'} |P[h(X) \in A'] - P[h(Y) \in A']|
\]

\[
= \sup_{A' \in S'} |P[X \in h^{-1}(A')] - P[Y \in h^{-1}(A')]|
\]

\[
= \sup_{A \in S} |P[X \in A] - P[Y \in A]|.
\]
4.2.2 Erdős-Rényi graphs: degree sequence

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

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

**Theorem 4.18** (Erdős-Rényi graphs: degree sequence).

$$\frac{1}{n} N_d(n) \to_p f_d := e^{-\lambda} \frac{\lambda^d}{d!}, \quad \forall d \geq 1.$$

**Proof.** We proceed in two steps:

1. we use the coupling inequality (Lemma 4.9) to show that the expectation of $\frac{1}{n} N_d(n)$ is close to $f_d$;

2. we use Chebyshev’s inequality (Theorem 2.2) to show that $\frac{1}{n} N_d(n)$ is close to its expectation.

We prove each step as a lemma below.

**Lemma 4.19** (Convergence of the mean).

$$\frac{1}{n} \mathbb{E}_{n,p_n} [N_d(n)] \to f_d, \quad \forall d \geq 1.$$

**Proof.** Note that the $D_i(n)$s are identically distributed (but not independent) so

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

Moreover $D_1(n) \sim \text{Bin}(n-1, p_n)$. Let

$$S_n \sim \text{Bin}(n, p_n) \quad \text{and} \quad Z_n \sim \text{Poi}(\lambda).$$

By the Poisson approximation

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

We can couple $D_1(n)$ and $S_n$ as $(\sum_{i \leq n-1} X_i, \sum_{i \leq n} X_i)$ where the $X_i$s are i.i.d. Bernoulli with parameter $\frac{\lambda}{n}$. By the coupling inequality

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

By the triangle inequality for total variation distance,

$$\frac{1}{2} \sum_{d \geq 0} |\mathbb{P}_{n,p_n} [D_1(n) = d] - f_d| \leq \frac{\lambda + \lambda^2/2}{n} + O(n^{-2}).$$
Therefore,
\[
\left| \frac{1}{n} \mathbb{E}_{n,p_n} [N_d(n)] - f_d \right| \leq \frac{2\lambda + \lambda^2}{n} + O(n^{-2}) \to 0.
\]

Lemma 4.20 (Concentration around the mean).
\[
\mathbb{P}_{n,p_n} \left[ \left| \frac{1}{n} N_d(n) - \frac{1}{n} \mathbb{E}_{n,p_n} [N_d(n)] \right| \geq \varepsilon \right] \leq \frac{2\lambda + 1}{\varepsilon^2 n}, \quad \forall d \geq 1, \forall n.
\]

Proof. By Chebyshev’s inequality, for all $\varepsilon > 0$
\[
\mathbb{P}_{n,p_n} \left[ \left| \frac{1}{n} N_d(n) - \frac{1}{n} \mathbb{E}_{n,p_n} [N_d(n)] \right| \geq \varepsilon \right] \leq \frac{\text{Var}_{n,p_n} \left[ \frac{1}{n} N_d(n) \right]}{\varepsilon^2}.
\]
Note that
\[
\text{Var}_{n,p_n} \left[ \frac{1}{n} N_d(n) \right] = \frac{1}{n^2} \left\{ \mathbb{E}_{n,p_n} \left[ \left( \sum_{i \leq n} 1 \{ D_i(n) = d \} \right)^2 \right] - (n \mathbb{P}_{n,p_n} [D_1(n) = d])^2 \right\}
\]
\[
= \frac{1}{n^2} \left\{ n(n-1) \mathbb{P}_{n,p_n} [D_1(n) = d, D_2(n) = d] + n \mathbb{P}_{n,p_n} [D_1(n) = d] - n^2 \mathbb{P}_{n,p_n} [D_1(n) = d]^2 \right\}
\]
\[
\leq \frac{1}{n} + \left\{ \mathbb{P}_{n,p_n} [D_1(n) = d, D_2(n) = d] - \mathbb{P}_{n,p_n} [D_1(n) = d]^2 \right\}.
\]
We bound the second term using a neat coupling argument. Let $Y_1$ and $Y_2$ be independent Bin$(n - 2, p_n)$ and let $X_1$ and $X_2$ be independent Ber$(p_n)$. Then the term in curly bracket above is equal to
\[
\mathbb{P}[(X_1 + Y_1, X_1 + Y_2) = (d, d)] - \mathbb{P}[(X_1 + Y_1, X_2 + Y_2) = (d, d)]
\]
\[
\leq \mathbb{P}[(X_1 + Y_1, X_1 + Y_2) = (d, d), (X_1 + Y_1, X_2 + Y_2) \neq (d, d)]
\]
\[
= \mathbb{P}[(X_1 + Y_1, X_1 + Y_2) = (d, d), X_2 + Y_2 \neq d]
\]
\[
= \mathbb{P}[X_1 = 0, Y_1 = Y_2 = d, X_2 = 1] + \mathbb{P}[X_1 = 1, Y_1 = Y_2 = d - 1, X_2 = 0]
\]
\[
\leq \frac{2\lambda}{n}.
\]
So $\text{Var}_{n,p_n} \left[ \frac{1}{n} N_d(n) \right] \leq \frac{2\lambda + 1}{n}$.

Combining the lemmas concludes the proof of Theorem 4.18.
4.3 Stochastic domination

In comparing two probability measures, a natural relationship to consider is the notion of “domination.” To see why this might be useful, let \((X_i)_{i=1}^n\) be independent \(\mathbb{Z}_+\)-valued random variables with

\[ P[X_i \geq 1] \geq p, \]

and let \(S = \sum_{i=1}^n X_i\) be their sum. Consider also a random variable

\[ S^* \sim \text{Bin}(n, p). \]

Then it is intuitively clear that one should be able to obtain some bounds on \(S\) by studying \(S^*\) instead—which may be 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

\[ P[S > x] \geq P[S^* > x]. \]

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

In particular we study an important special case known as positive association. In that case a measure “dominates itself” in the following sense: conditioning on certain events makes other events more likely. This concept, which is formalized in Section 4.3.3, has numerous applications in discrete probability.

4.3.1 Definitions

We start with the case of real random variables.

**Ordering of real random variables** For real random variables, stochastic domination is defined as follows. See Figure 4.2 for an illustration.

**Definition 4.21** (Stochastic domination). Let \(\mu\) and \(\nu\) be probability measures on \(\mathbb{R}\). The measure \(\mu\) is said to stochastically dominate \(\nu\), denoted \(\mu \succeq \nu\), if for all \(x \in \mathbb{R}\)

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

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

**Example 4.22** (Bernoulli vs. Poisson). Let \(X \sim \text{Poi}(\lambda)\) be Poisson with mean \(\lambda > 0\) and let \(Y\) be a Bernoulli trial with success probability \(p \in (0, 1)\), i.e.,
Figure 4.2: The law of $X$, represented here by its cumulative distribution function $F_X$ in red, stochastically dominates the law of $Y$, in orange. 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.
\( P[Y = 1] = 1 - P[Y = 0] = p \). In order for \( X \) to stochastically dominate \( Y \), it suffices to have
\[
P[X > \ell] \geq P[Y > \ell], \quad \forall \ell \geq 0.
\]
This is always true for \( \ell \geq 1 \) since \( P[X > \ell] > 0 \) but \( P[Y > \ell] = 0 \). So it remains to consider the case \( \ell = 0 \). We have
\[
1 - e^{-\lambda} = P[X > 0] \geq P[Y > 0] = p,
\]
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. The connection to coupling arises from the following characterization.

**Theorem 4.23** (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
\[
P[\hat{X} \geq \hat{Y}] = 1. \tag{4.1}
\]
We refer to \((\hat{X}, \hat{Y})\) as a monotone coupling of \( X \) and \( Y \).

**Proof.** One direction is clear. Suppose there is such a coupling. Then for all \( x \in \mathbb{R} \)
\[
P[Y > x] = P[\hat{Y} > x] = P[\hat{X} \geq \hat{Y} > x] \leq P[\hat{X} > x] = P[X > x].
\]

For the other direction, define the cumulative distribution functions \( F_X(x) = P[X \leq x] \) and \( F_Y(x) = P[Y \leq x] \). Assume \( X \succeq X' \). The idea of the proof is to use the following standard way of generating a real random variable. Recall (e.g. [Dur10, Section 1.2]) that
\[
X \overset{d}{=} F_X^{-1}(U), \tag{4.2}
\]
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. Indeed \( F_X^{-1}(u) > x \) precisely when \( u > F_X(x) \). It is natural to construct a coupling of \( X \) and \( Y \) by simply using the same uniform random variable \( U \) in this representation, i.e., we define \( \hat{X} = F_X^{-1}(U) \) and \( \hat{Y} = F_Y^{-1}(U) \). See Figure 4.2. By (4.2), this is a coupling of \( X \) and \( Y \). It remains
to check (4.1). 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.

\textbf{Example 4.24.} Returning to the example in the first paragraph of Section 4.3, 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 \( S_* \sim \text{Bin}(n, p) \). Write \( S_* \) as the sum \( \sum_{i=1}^n Y_i \) where \( (Y_i) \) are independent \( \{0, 1\} \)-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 \( \hat{X}_i \geq \hat{Y}_i \) a.s. for all \( i \) and as a result \( \sum_{i=1}^n \hat{X}_i =: \hat{S} \geq \hat{S}_* \) a.s. or \( S \geq S_* \) by the previous theorem. That implies for instance that \( \mathbb{P}[S > x] \geq \mathbb{P}[S_* > x] \) as we claimed earlier. A special case of this argument gives the following useful fact about binomials

\[
n \geq m, q \geq p \implies \text{Bin}(n, q) \succeq \text{Bin}(m, p).
\]

\textbf{Example 4.25 (Poisson distribution).} Let \( X \sim \text{Poi}(\mu) \) and \( Y \sim \text{Poi}(\nu) \) with \( \mu > \nu \). Recall that a sum of independent Poisson is Poisson (use moment-generating functions or see e.g. [Dur10, Exercise 2.1.14]). This fact leads to a natural coupling: let \( \hat{Y} \sim \text{Poi}(\nu) \), \( \hat{Z} \sim \text{Poi}(\mu - \nu) \) 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 4.23.

\textbf{Corollary 4.26.} 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

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

\textbf{Proof.} Let \( (\hat{X}, \hat{Y}) \) be the monotone coupling of \( X \) and \( Y \) whose existence is guaranteed by Theorem 4.23. Then \( f(\hat{X}) \succeq f(\hat{Y}) \) a.s. 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 4.27. Let $X_1, X_2$ be independent random variables. Let $Y_1, Y_2$ be independent random variables such that $X_i \geq Y_i$, $i = 1, 2$. 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

$$X_1 + X_2 \sim \hat{X}_1 + \hat{X}_2 \succeq \hat{Y}_1 + \hat{Y}_2 \sim Y_1 + Y_2.$$  

Example 4.28 (Binomial vs. Poisson). A sum of $n$ Poisson variables with mean $\lambda$ is $\text{Poi}(n\lambda)$. A sum of $n$ Bernoulli trials with success probability $p$ is $\text{Bin}(n, p)$. Using Example 4.22 and Corollary 4.27, we get

$$\lambda \geq -\log(1 - p) \implies \text{Poi}(n\lambda) \succeq \text{Bin}(n, p).$$  

The following special case will be useful later. Let $0 < \Lambda < 1$ and let $m$ be an 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 $x \in \mathbb{R}$. So, setting $\lambda := \frac{\Lambda}{m-1}$, $p := \frac{\Lambda}{m}$ and $n := m - 1$ in (4.3), we get

$$\Lambda \in (0, 1) \implies \text{Poi}(\Lambda) \succeq \text{Bin} \left( m - 1, \frac{\Lambda}{m} \right).$$  

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 $(\mathcal{X}, \preceq)$ be a poset, i.e., for all $x, y, z \in \mathcal{X}$:

- [Reflexivity] $x \preceq x$,
- [Antisymmetry] if $x \preceq y$ and $y \preceq x$ then $x = y$,
- [Transitivity] if $x \preceq y$ and $y \preceq z$ then $x \preceq z$.  


For instance the set \(\{0, 1\}^F\) is a poset when equipped with the relation \(x \leq y\) if and only if \(x_i \leq y_i\) for all \(i \in F\). 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 \(y \leq x\).)

Let \(\mathcal{F}\) be a \(\sigma\)-field 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 \(f(x) \leq f(y)\).

**Definition 4.29** (Stochastic domination for posets). Let \((\mathcal{X}, \leq)\) be a poset and let \(\mathcal{F}\) be a \(\sigma\)-field 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 \(\hat{X} \geq \hat{Y}\) a.s.

**Example 4.30** (Monotonicity of the percolation function). We have already seen an example of stochastic domination in Section 2.2.4. Consider bond percolation on the \(d\)-dimensional lattice \(\mathbb{L}^d\). 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
\[
\theta(p) := \mathbb{P}_p[|C_0| = +\infty],
\]
where \(C_0\) is the open cluster of the origin. We argued in Section 2.2.4 that \(\theta(p)\) is non-decreasing 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 \(p\)-values, \(p_1 < p_2\), gives a monotone coupling of the processes for \(p_1\) and \(p_2\). It follows immediately that \(\theta(p_1) \leq \theta(p_2)\), where we used that the event \(\{|C_0| = +\infty\}\) is increasing.

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 4.31** (Strassen’s theorem). Let \(X\) and \(Y\) be random variables taking values in a finite poset \((\mathcal{X}, \leq)\) with the \(\sigma\)-field \(\mathcal{F} = 2^\mathcal{X}\). Then \(X \succeq Y\) if and only if there exists a monotone coupling \((\hat{X}, \hat{Y})\) of \(X\) and \(Y\).
Proof. One direction is clear. 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 1.9). 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 $\nu(x, y) = 0$ if $x \leq y$, the marginal conditions on the coupling read

$$
\sum_{y \leq x} \nu(x, y) = \mu_X(x), \quad \forall x \in \mathcal{X},
$$

and

$$
\sum_{x \geq y} \nu(x, y) = \mu_Y(y), \quad \forall y \in \mathcal{X}.
$$

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 are at capacity. If we let $\nu(x, y)$ be the flow on edge $(x, 1)$, the constraints above then impose the conservation of the 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 1.9), 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 above. In fact, for a minimum cut, we further have

$$
A_+ = \{ x \in \mathcal{X} : \exists y \in Z^c_+, x \geq y \},
$$

216
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 colored edges depict a suboptimal cut. The blue and orange vertices correspond respectively to the sets $A_*$ and $Z_*$ for this cut. The capacity of the cut is $r^2 + r(1-r) + (1-q)^2 + (1-q)q = r + (1-q) > r + (1-r) = 1$. 
as adding an $x$ not satisfying this property is redundant. See Figure 4.3. In particular $A_\ast$ is increasing: if $x_1 \in A_\ast$ and $x_2 \geq x_1$, then $\exists y \in Z_c^\ast$ such that $x_1 \geq y$ and, since $x_2 \geq x_1 \geq y$, the same $y$ works for $x_2$. Observe that, because $y \geq y$, the set $A_\ast$ also includes $Z_c^\ast$. If it were the case that $A_\ast \neq Z_c^\ast$, then we could construct a cut with lower or equal capacity by fixing $A_\ast$ and setting $Z_c^\ast := A_c^\ast$; because $A_\ast$ is increasing, any $y \in A_\ast \cap Z_c^\ast$ is such that paths of the form $a \sim (x, 1) \sim (y, 2) \sim z$ with $x \geq y$ are cut by $x \in A_\ast$. Hence, for a minimum cut, we can assume that in fact $A_\ast = Z_c^\ast$. The capacity of the cut is

$$\mu_X(A_\ast) + \mu_Y(Z_c^\ast) = \mu_X(A_\ast) + 1 - \mu_Y(A_\ast) = 1 + (\mu_X(A_\ast) - \mu_Y(A_\ast)) \geq 1,$$

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

Remark 4.32. Strassen’s theorem 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 4.26 immediately extends to:

Corollary 4.33. Let $X$ and $Y$ be $X$-valued random variables with $X \succ Y$ and let $f : X \rightarrow \mathbb{R}$ 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.

Example 4.34 (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. 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 4.35.

$$M_s \succeq M'_s.$$ We prove this by producing a monotone coupling. First set $(\hat{X}_t) := (X_t).$ We then generate $(\hat{X}'_t)$ as a “sticky” version of $(\hat{X}_t).$ 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 \(j\), for an extra time step with probability 

\[
\frac{c'(j, j) - c(j, j)}{\sum_{i \sim j} c'(i, j)},
\]

which is in \([0, 1]\) by assumption. Marginally, \((\hat{X}'_t)\) is a random walk on \(\mathcal{N}'\) because by construction 

\[
\frac{c'(j, j)}{\sum_{i \sim j} c'(i, j)} = \frac{c(j, j)}{\sum_{i \sim j} c(i, j)} + \left( \frac{\sum_{i \sim j} c(i, j)}{\sum_{i \sim j} c'(i, j)} \right) \frac{c'(j, j)}{\sum_{i \sim j} c(i, j)},
\]

and for \(i \neq j\) with \(i \sim j\)

\[
\frac{c'(i, j)}{\sum_{i' \sim j} c'(i', j)} = \left( \frac{\sum_{i' \sim j} c(i', j)}{\sum_{i' \sim j} c'(i', j)} \right) \frac{c(i, j)}{\sum_{i \sim j} c(i, j)},
\]

since \(c'(i, j) = c(i, j)\). This coupling satisfies 

\[
\hat{M}_s := \max_{t \leq s} \hat{X}_t \geq \max_{t \leq s} \hat{X}'_t =: \hat{M}'_s, \quad \text{a.s.}
\]

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

The previous example involved an asynchronous coupling of the chains. Often, a simpler step-by-step approach is possible.

**Definition 4.36** (Stochastic domination for Markov kernels). Let \(P\) and \(Q\) be transition matrices on a finite or countable poset \((\mathcal{X}, \leq)\). The transition matrix \(Q\) is said to stochastically dominate the transition matrix \(P\) if 

\[
x \leq y \implies P(x, \cdot) \preceq Q(y, \cdot). \tag{4.5}
\]

If the above condition is satisfied for \(P = Q\), we say that \(P\) is stochastically monotone.

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

**Theorem 4.37** (Strassen’s theorem for Markov kernels). Let \((X_t)\) and \((Y_t)\) 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 a.s. 

\[
\hat{X}_t \leq \hat{Y}_t, \quad \forall t.
\]
Furthermore, if the chains are irreducible and have stationary distributions $\pi$ and $\mu$ respectively, then $\pi \leq \mu$.

Observe that, for a step-by-step 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. Also you should convince yourself that the chains in Example 4.34 do not in general satisfy (4.5). (Which pairs $x, y$ cause problems?)

**Proof of Theorem 4.37.** Let

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

For all $(x, y) \in W$, 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 (4.5). Let $(\hat{X}_t, \hat{Y}_t)$ be a Markov chain on $W$ with transition matrix $R$ started at $(x_0, y_0)$. By construction, $\hat{X}_t \leq \hat{Y}_t$ for all $t$ a.s. That proves the first half of the theorem.

For the second half, let $A$ be increasing in $X$. Then, by the ergodic theorem for Markov chains (e.g. [Dur10, Exercise 6.6.4]),

$$\pi(A) \leftarrow \frac{1}{t} \sum_{s \leq t} 1_{\hat{X}_s \in A} \leq \frac{1}{t} \sum_{s \leq t} 1_{\hat{Y}_s \in A} \rightarrow \mu(A), \quad \text{a.s.}$$

where we used that $\hat{X}_s \in A$ implies $\hat{Y}_s \in A$ because $\hat{X}_s \leq \hat{Y}_s$ and $A$ is increasing. This proves the claim by definition of stochastic domination.

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

### 4.3.2 Ising model on $\mathbb{Z}^d$: extremal measures

Consider the $d$-dimensional lattice $\mathbb{L}^d$. Let $\Lambda$ be a finite subset of vertices in $\mathbb{L}^d$ and define $X := \{-1, +1\}^\Lambda$, which is a poset when equipped with the relation $\sigma \preceq \sigma'$ if and only if $\sigma_i \leq \sigma'_i$ for all $i \in \Lambda$. For shorthand, we occasionally write $+$ and $-$ instead of $+1$ and $-1$. For $\xi \in \{-1, +1\}^{2d}$, recall that the (ferromagnetic) Ising model on $\Lambda$ with boundary conditions $\xi$ and inverse temperature $\beta$ is the probability distribution over spin configurations $\sigma \in X$ given by

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

where

$$\mathcal{H}_{\Lambda, \xi}(\sigma) := -\sum_{\overset{i < j}{i, j \in \Lambda}} \sigma_i \sigma_j - \sum_{i \in \Lambda, j \notin \Lambda} \sigma_i \xi_j.$$
is the Hamiltonian and

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

is the partition function. (Warning: it is easy to get confused with the $-$ signs that cancel out in the exponent.) For the all-$(+1)$ and all-$(-1)$ boundary conditions we write respectively $\mu_{\beta, \Lambda}^+(\sigma)$ and $\mu_{\beta, \Lambda}^-(\sigma)$. In this section, we show that these two measures are extremal in the following sense. For all boundary conditions $\xi \in \{-1, +1\}^{1^d}$:

**Claim 4.38.**

$$\mu_{\beta, \Lambda}^+ \succeq \mu_{\beta, \Lambda}^\xi \succeq \mu_{\beta, \Lambda}^-.$$

In words, 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.

The idea of the proof is to use Theorem 4.37 with a suitable Markov chain.

**Stochastic domination** In this context, vertices are often referred to as sites.

Recall that the single-site Glauber dynamics of the Ising model 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 $\mu_{\beta, \Lambda}^\xi(\sigma)$ conditioned on agreeing with $\sigma$ at all sites in $\Lambda \setminus \{i\}$. 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 $n = |\Lambda|$, because the Ising measure factorizes, the transition matrix of the Glauber dynamics is simply

$$Q_{\beta, \Lambda}^\xi(\sigma, \sigma^{\gamma}) := \frac{1}{n} \cdot \frac{e^{\gamma \beta S_i^\xi(\sigma)}}{e^{-\beta S_i^\xi(\sigma)} + e^{\beta S_i^\xi(\sigma)}},$$

where

$$S_i^\xi(\sigma) := \sum_{j \sim i \setminus \Lambda} \sigma_j + \sum_{j \sim i \setminus \Lambda} \xi_j.$$

All other transitions have probability 0.

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

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

221
we have

\[
\mu_\beta^\xi (\sigma^{i,-}) Q_\beta^\xi (\sigma^{i,-}, \sigma^{i,+}) = \frac{e^{-\beta S^\xi_i (\sigma)} e^{-\beta S^\xi_i (\sigma)}}{Z_\Lambda^\xi (\beta)} \cdot \frac{e^{\beta S^\xi_i (\sigma)}}{n [e^{-\beta S^\xi_i (\sigma)} + e^{\beta S^\xi_i (\sigma)}]}
\]

\[
= \frac{e^{-\beta S^\xi_i (\sigma)}}{n Z_\Lambda^\xi (\beta)} \cdot \frac{e^{\beta S^\xi_i (\sigma)}}{n [e^{-\beta S^\xi_i (\sigma)} + e^{\beta S^\xi_i (\sigma)}]}
\]

\[
= \frac{e^{-\beta S^\xi_i (\sigma)} e^{\beta S^\xi_i (\sigma)}}{Z_\Lambda^\xi (\beta)} \cdot \frac{e^{\beta S^\xi_i (\sigma)}}{n [e^{-\beta S^\xi_i (\sigma)} + e^{\beta S^\xi_i (\sigma)}]}
\]

\[
= \mu_\beta^\xi (\sigma^{i,+}) Q_\beta^\xi (\sigma^{i,+}, \sigma^{i,-}).
\]

In particular \(\mu_\beta^\xi\) is the stationary distribution of \(Q_\beta^\xi\).

**Claim 4.39.**

\[
\xi' \geq \xi \implies Q_\beta^\xi \text{ stochastically dominates } Q_\beta^\xi'.
\] (4.6)

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

**Lemma 4.40.** To establish (4.6), it suffices to show that, for all \(\sigma \leq \tau\),

\[
Q_\beta^\xi (\sigma, \sigma^{i,+}) \leq Q_\beta^\xi' (\tau, \tau^{i,+}).
\] (4.7)

**Proof.** Assume (4.7) holds. Let \(A\) be increasing in \(X\) and let \(\sigma \leq \tau\). Then, for the single-site Glauber dynamics, we have

\[
Q_\beta^\xi (\sigma, A) = Q_\beta^\xi (\sigma, A \cap B_\sigma),
\] (4.8)

where

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

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

\[
\sigma^{i,\gamma} \in A \implies \tau^{i,\gamma} \in A,
\] (4.9)

and

\[
\sigma^{i,-} \in A \implies \sigma^{i,+} \in A.
\] (4.10)

Letting

\[
I_{\sigma, A}^\pm := \{ i \in \Lambda : \sigma^{i,-} \in A \}, \quad I_{\sigma, A}^+ := \{ i \in \Lambda : \sigma^{i,+} \in A \},
\]

222
and similarly for $\tau$, we have by (4.7), (4.8), (4.9), and (4.10),

$$
Q^\xi_{\beta,\Lambda}(\sigma, A) = Q^\xi_{\beta,\Lambda}(\sigma, A \cap B_\sigma)
= \sum_{i \in I^{\xi}_{\sigma,A}} Q^\xi_{\beta,\Lambda}(\sigma, \sigma^{i,+}) + \sum_{i \in I^{\xi}_{\sigma,A}} \left[ Q^\xi_{\beta,\Lambda}(\sigma, \sigma^{i,-}) + Q^\xi_{\beta,\Lambda}(\sigma, \sigma^{i,+}) \right]
\leq \sum_{i \in I^{\xi'}_{\sigma,A}} Q^{\xi'}_{\beta,\Lambda}(\tau, \tau^{i,+}) + \sum_{i \in I^{\xi}_{\sigma,A}} \frac{1}{n}
\leq \sum_{i \in I^{\xi'}_{\sigma,A}} Q^{\xi'}_{\beta,\Lambda}(\tau, \tau^{i,+}) + \sum_{i \in I^{\xi}_{\sigma,A}} \left[ Q^{\xi'}_{\beta,\Lambda}(\tau, \tau^{i,-}) + Q^{\xi'}_{\beta,\Lambda}(\tau, \tau^{i,+}) \right]
= Q^{\xi'}_{\beta,\Lambda}(\tau, A),
$$
as claimed.

Returning to the proof of Claim 4.39, 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.

Finally:


Observe that we have not used any special property of the $d$-dimensional lattice. Indeed Claim 4.38 in fact holds for any countable, locally finite graph with positive coupling constants.

4.3.3 Correlation inequalities: FKG and Holley’s inequalities

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

Example 4.41 (Erdős-Rényi graphs: positive associations). Consider an Erdős-Rényi graph $G \sim G_{n,p}$. Let $E = \{\{x, y\} : x, y \in [n], x \neq y\}$. Think of $G$ as taking values in the poset $[\{0, 1\}^E, \leq]$ 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 $\{0, 1\}^E$. The event $\mathcal{A}$ that $G$ is connected is increasing because adding edges cannot disconnect a 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. Indeed 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 a more general phenomenon. That is, for any non-empty increasing events $\mathcal{A}$ and $\mathcal{B}$, we have:

**Claim 4.42.**

$$\mathbb{P}_{n,p}[\mathcal{B} | \mathcal{A}] \geq \mathbb{P}_{n,p}[\mathcal{B}] . \tag{4.11}$$

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 (4.11) is equivalent to $\mathbb{P}_{n,p}[\mathcal{A} \cap \mathcal{B}] \geq \mathbb{P}_{n,p}[\mathcal{A}] \mathbb{P}_{n,p}[\mathcal{B}]$, i.e., to the fact that $\mathcal{A}$ and $\mathcal{B}$ are positively correlated. □

More generally:

**Definition 4.43** (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 associated, 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 $\mathcal{A}$ and $\mathcal{B}$ it holds that

$$\mu(\mathcal{A} \cap \mathcal{B}) \geq \mu(\mathcal{A})\mu(\mathcal{B}),$$

i.e., $\mathcal{A}$ and $\mathcal{B}$ are positively correlated.

**Remark 4.44.** Note that positive associations is concerned only with increasing events. See Remark 4.63.

**Remark 4.45.** A notion of negative associations, which is a somewhat more delicate concept, was defined in Remark 3.129. See also [Pem00].

Let $\mu$ be positively associated. Note that if $\mathcal{A}$ and $\mathcal{B}$ are decreasing, i.e. their
complements are increasing, then

\[\mu(A \cap B) = 1 - \mu(A^c \cup B^c)\]
\[= 1 - \mu(A^c) - \mu(B^c) + \mu(A^c \cap B^c)\]
\[\geq 1 - \mu(A^c) - \mu(B^c) + \mu(A^c)\mu(B^c)\]
\[= \mu(A)\mu(B).\]

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

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 \(\omega_0\) and \(\omega'_{0}\) be the coordinatewise minimum and maximum of \(\omega\) and \(\omega'\).

**Theorem 4.46 (FKG inequality).** Let \(X = \{0, 1\}^F\) where \(F\) is finite. Suppose \(\mu\) is a positive probability measure on \(X\) satisfying the FKG condition

\[\mu(\omega \lor \omega') \mu(\omega \land \omega') \geq \mu(\omega) \mu(\omega'), \quad \forall \omega, \omega' \in X.\]  

This property is also known as log-convexity or log-supermodularity. We call such a measure an FKG measure. Then \(\mu\) has positive associations.

**Remark 4.47.** Strict positivity is not in fact needed [FKG71]. The FKG condition is equivalent to a strong form of positive associations. See Exercise 4.4.

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

\[\mu(\omega \lor \omega') \mu(\omega \land \omega') = \prod_{f} \mu_f(\omega_f \lor \omega'_f) \mu_f(\omega_f \land \omega'_f)\]
\[= \prod_{f : \omega_f = \omega'_f} \mu_f(\omega_f)^2 \prod_{f : \omega_f \neq \omega'_f} \mu_f(\omega_f)\mu_f(\omega'_f)\]
\[= \prod_{f : \omega_f = \omega'_f} \mu_f(\omega_f)\mu_f(\omega'_f) \prod_{f : \omega_f \neq \omega'_f} \mu_f(\omega_f)\mu_f(\omega'_f)\]
\[= \mu(\omega) \mu(\omega').\]

So the FKG inequality applies, for instance, to bond percolation and Erdős-Rényi graphs. The pointwise nature of the FKG condition also makes it relatively easy to check it for measures which are defined explicitly up to a normalizing constant, such as the Ising model.
Example 4.48 (Ising model on $\mathbb{Z}^d$: checking FKG). Consider again the setting of Section 4.3.2. Of course we work on the space $\mathcal{X} := \{-1, +1\}^\Lambda$ rather than $\{0, 1\}^F$. Fix a finite $\Delta \subseteq \mathbb{Z}^d$, $\xi \in \{-1, +1\}^{\mathbb{Z}^d}$ and $\beta > 0$.

Claim 4.49. The measure $\mu^\xi_{\beta, \Lambda}$ satisfies the FKG condition and therefore has positive associations.

Intuitively, taking the maximum or minimum of two configurations tends to increase spin agreement and therefore leads to a higher likelihood. By taking logarithms in the FKG condition, one sees that proving the claim boils down to checking an inequality for each term in the Hamiltonian. For $\iota, j \in \Lambda$ and $i \not\in \Lambda$ such that $i \sim j$, we have

$$\tau_i \xi_j + \tau_i \xi_j = (\tau_i + \tau_i) \xi_j = \sigma_i \xi_j + \sigma'_i \xi_j.$$  \hspace{1cm} (4.13)

For $i, j \in \Lambda$ with $i \sim j$, note first that the case $\sigma_j = \sigma'_j$ reduces to the previous calculation, so we assume $\sigma_i \neq \sigma'_i$ and $\sigma_j \neq \sigma'_j$. Then

$$\tau_i \tau_j + \tau_i \tau_j = (+1)(+1) + (-1)(-1) = 2 \geq \sigma_i \sigma_j + \sigma'_i \sigma'_j,$$

since 2 is the largest value the rightmost expression ever takes. We have shown that

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

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, i.e., 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 4.50 (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

$$\mu_2(\omega \vee \omega') \mu_1(\omega \wedge \omega') \geq \mu_2(\omega) \mu_1(\omega'), \quad \forall \omega, \omega' \in \mathcal{X}. \hspace{1cm} (4.14)$$

Then $\mu_1 \preceq \mu_2$.

Before proving Holley’s inequality, we check that it indeed implies the FKG inequality. See Exercise 4.1 for an elementary proof in the independent case, i.e., of Harris’ inequality.
Proof of Theorem 4.46. 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 \( 0 \) be the all-0 vector, note that \( f \) and \( g \) are increasing if and only if \( f' := f - f(0) + 1 > 0 \) and \( g' := g - g(0) + 1 > 0 \) are increasing and that, moreover,

\[
\mu(f'g') - \mu(f')\mu(g') = \mu([f' - \mu(f')][g' - \mu(g')]) \\
= \mu([f - \mu(f)][g - \mu(g)]) \\
= \mu(fg) - \mu(f)\mu(g).
\]

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 Holley’s inequality. Note that \( \omega' \leq \omega \lor \omega' \) for any \( \omega \) so that, because \( g \) is increasing, we have \( g(\omega') \leq g(\omega \lor \omega') \). Hence, for any \( \omega, \omega' \),

\[
\mu_1(\omega)\mu_2(\omega') = \mu(\omega)\frac{g(\omega')\mu(\omega')}{\mu(g)} \\
= \mu(\omega)\mu(\omega')\frac{g(\omega')}{\mu(g)} \\
\leq \mu(\omega \land \omega')\mu(\omega \lor \omega')\frac{g(\omega \lor \omega')}{\mu(g)} \\
= \mu_1(\omega \land \omega')\mu_2(\omega \lor \omega'),
\]

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 4.33

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

and the theorem is proved.

Proof of Theorem 4.50. We use Theorem 4.37. This is similar to what was done in Section 4.3.2. This time we couple Metropolis-like chains. 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 \( n = |F| \).
For $\alpha, \beta > 0$ small enough, the following transition matrix over $X$ is irreducible and reversible w.r.t. its stationary distribution $\mu_2$: for all $i \in F$, $y \in X$,

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

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

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

Let $P$ be similarly defined for $\mu_1$ with the same values of $\alpha$ and $\beta$. For reasons that will be clear below, the value of $\beta$ is chosen so that the sum of the two expressions in brackets above is smaller than 1 for all $y, i$. The value of $\alpha$ is then chosen so that $P(x, x), Q(y, y) \geq 0$ for all $x, y$. 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 4.37, 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

$$\frac{\mu_1(x^{i,0})}{\mu_1(x^{i,1})} \geq \frac{\mu_2(y^{i,0})}{\mu_2(y^{i,1})} \tag{4.15}$$

by taking $\omega = y^{i,0}$ and $\omega' = x^{i,1}$ in Condition (4.14).

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 values of $x_i, y_i$ (with $x_i \leq y_i$ by assumption):

- $(x_i, y_i) = (0, 0)$: With probability $\beta$, perform an upward transition in both, i.e., 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, i.e., 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, i.e., set $\hat{X} := x^{i,0}$ and $\hat{Y} := y$. Note that (4.15) guarantees that the previous step is well-defined. With the remaining probability, set $(\hat{X}, \hat{Y}) := (x, y)$ instead.
- \((x_i, y_i) = (0, 1)\): With probability \(\beta\), perform an upward transition in \(x\) only, i.e., set \(\hat{X} := x^{i,1}\) and \(\hat{Y} := y\). With probability \(\beta \frac{p_2(y^{i,0})}{p_2(y^{i,1})}\), perform a downward transition in \(y\) only, i.e., set \(\hat{X} := x\) and \(\hat{Y} := y^{i,0}\). With the remaining probability, set \((\hat{X}, \hat{Y}) := (x, y)\) instead. (This is where we use the odd choice of \(\beta\).) By construction, this coupling satisfies \(\hat{X} \leq \hat{Y}\) a.s. An application of Theorem 4.37 concludes the proof.

Example 4.51 (Ising model: extremality revisited). Holley’s inequality gives another proof of Claim 4.38. To see this, just repeat the calculations of Example 4.48, where now (4.13) is replaced with an inequality. See Exercise 4.2.

4.3.4 Erdős-Rényi graph: Janson’s inequality and application to the containment problem

Let \(G = (V, E) \sim \mathbb{G}_{n,p}\) be an Erdős-Rényi graph. Repeating the computations of Section 2.3.2 (or see Claim 2.23), we see that 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 \(n \to +\infty\) depending on whether \(p \ll n^{-1}\) or \(p \gg n^{-1}\) respectively. In this section, we investigate what happens at the threshold. From now on, we assume that \(p = \lambda/n\) 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

\[
\varepsilon := \mathbb{P}_{n,p}[B_S] = p^3 \to 0.
\]

(4.16)

Let \(X_n = \sum_{S \in \binom{V}{3}} 1_{B_S}\) be the number of triangles in \(G\). By the linearity of expectation, the expected number of triangles is

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

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

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

(4.17)

In fact, by the Poisson approximation to the binomial (e.g. [Dur10, Theorem 3.6.1]), we would have that the number of triangles converges weakly to \(\text{Poi}(\lambda^3/6)\).

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 $|S \cap S'| \leq 1$, i.e., whenever the edges connecting $S$ are disjoint from those connecting $S'$. Write $S \sim S'$ if $S \neq S'$ are not independent, i.e. if $|S \cap S'| = 2$. The expected number of (unordered) pairs $S \sim S'$ both forming a triangle is

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

(4.18)

Given that the events $\{B_S\}$ are “mostly” independent, it is natural to expect that $X_n$ behaves asymptotically as it does in the independent case. Indeed we prove:

Claim 4.52.

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

Remark 4.53. In fact, $X_n \sim \text{Poi}(\lambda^3/6)$. See Exercises 2.15 and 4.5.

The FKG inequality immediately gives one direction. Recall that $\mathbb{P}_{n,p}$, 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

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

by (4.17). 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 $\beta^{(i)}_r = \beta^{(j)}_r = 1$ for some $r$ and note that $B_i$ is independent of $B_j$ if $i \sim j$. Set

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

**Theorem 4.54** (Janson’s inequality). Let $\mathcal{X}$, $\mathbb{P}$, $\{B_i\}_{i \in I}$ 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

$$
\prod_{i \in I} \mathbb{P}[B_i^c] \leq \mathbb{P}[\bigcap_{i \in I} B_i^c] \leq e^{\varepsilon \Delta} \prod_{i \in I} \mathbb{P}[B_i^c].
$$
Before proving the theorem, we show that it implies Claim 4.52. We have already shown in (4.16) and (4.18) that $\varepsilon \to 0$ and $\Delta \to 0$. Janson’s inequality immediately implies the claim by (4.17).

**Proof of Theorem 4.54.** The lower bound follows from the FKG inequality.

In the other direction, the first step is somewhat clear. We apply the chain rule to obtain

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

The rest is clever manipulation. W.l.o.g. assume $I = [m]$. 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] \setminus N_<(i)$\}. Then

$$
\mathbb{P}[B_i \cap \cap_{j \in [i-1]} B_j^c] = \frac{\mathbb{P}
[B_i \cap (\cap_{j \in [i-1]} B_j^c)]
}{\mathbb{P}[\cap_{j \in [i-1]} B_j^c]}
\geq \frac{\mathbb{P}
[B_i \cap (\cap_{j \in [i-1]} B_j^c)]
}{\mathbb{P}[(\cap_{j \in [i-1]} B_j^c) \setminus N_<(i)]}
= \mathbb{P}
[B_i \cap (\cap_{j \in [i-1]} B_j^c) \setminus N_<(i)]
\times \mathbb{P}
[\cap_{j \in N_<(i)} B_j^c | B_i \cap (\cap_{j \in [i-1]} B_j^c) \setminus N_<(i)]
= \mathbb{P}
[B_i]
\times \mathbb{P}
[\cap_{j \in N_<(i)} B_j^c | B_i \cap (\cap_{j \in [i-1]} B_j^c) \setminus N_<(i)].
$$

By a union bound the second term on the last line is

$$
\mathbb{P}
[\cap_{j \in N_<(i)} B_j^c | B_i \cap (\cap_{j \in [i-1]} B_j^c) \setminus N_<(i)]
\geq 1 - \sum_{j \in N_<(i)} \mathbb{P}
[B_j | B_i \cap (\cap_{j \in [i-1]} B_j^c) \setminus N_<(i)]
\geq 1 - \sum_{j \in N_<(i)} \mathbb{P}
[B_j | B_i],
$$

where the last line follows from the FKG inequality applied to the product measure $\mathbb{P}[\cdot | B_i]$ on \{0, 1\} with $F' := \{\ell \in [m] : \beta^{(i)}_\ell = 0\}$. Combining the last three
displays and using \(1 + z \leq e^z\), we get

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

By the definition of \(\Delta\), we are done.

### 4.3.5 Percolation on \(\mathbb{Z}^2\): RSW theory and a proof of Harris’ theorem

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

\[
\theta(p) := \mathbb{P}_p[|C_0| = +\infty],
\]

where \(C_0\) is the open cluster of the origin. We known from Example 4.30 that \(\theta(p)\) is non-decreasing. Let

\[
p_c(\mathbb{L}^2) := \sup\{p \geq 0 : \theta(p) = 0\},
\]

be the critical value. We proved in Section 2.2.4 that there is a non-trivial transition, i.e., \(p_c(\mathbb{L}^2) \in (0, 1)\). In fact we showed that \(p_c(\mathbb{L}^2) \in [1/3, 2/3]\) (see Exercise 2.2).

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

**Theorem 4.55** (Harris’ theorem).

\[
\theta(1/2) = 0.
\]

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

**Remark 4.56.** This bound is tight, i.e., in fact \(p_c(\mathbb{L}^2) = 1/2\). The other direction, known as Kesten’s theorem, is postponed to Section ?? where an additional ingredient is introduced, Russo’s formula.

Several proofs of Harris’ theorem are known. A particularly elegant one is sketched in Exercise 6.1. Here we present a proof that uses an important tool in percolation theory, the 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 4.55.* Fix $p = 1/2$. We use duality. Consider the $\mathbb{L}^2$ annulus

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

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

$$\mathbb{P}_{1/2}[|C_0| = +\infty] \leq \prod_{k=0}^{\ell} \{1 - \mathbb{P}_{1/2}[O_d(3^k)]\}, \quad (4.19)$$

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 $\ell$, $\mathbb{P}_{1/2}[O_d(\ell)] \geq c^*$. Then the r.h.s. of (4.19) tends to 0 as $K \to +\infty$. To simplify further, thinking of $\text{Ann}(\ell)$ as a union of four rectangles $[-3\ell, -\ell] \times [-3\ell, 3\ell], [-3\ell, 3\ell] \times (\ell, 3\ell], \text{etc.}$, it suffices to consider the event $O_d^\square(\ell)$ that each one of these rectangles contains a closed dual self-avoiding path connecting its two shorter sides. (More precisely, for the first rectangle above, 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, etc.) See Figure 4.4. By symmetry the probability that such a path exists is the same for all four rectangles. Denote it by $p_\ell$. Moreover the event that such a path exists is increasing so, although the four events are not independent, we can apply the FKG inequality. Hence, because $O_d^\square(\ell) \subseteq O_d(\ell)$, we finally get the bound

$$\mathbb{P}_{1/2}[O_d(\ell)] \geq p_\ell^4.$$

The RSW lemma and some symmetry arguments, both of which are detailed below, imply that there is some $c > 0$ such that, for all $\ell$:

**Lemma 4.57.**

$$p_\ell \geq c.$$

That concludes the proof.

It remains to prove Lemma 4.57. We first state the RSW lemma.
Figure 4.4: Top: the event $O_d(\ell)$. Bottom: the event $O_d^\#(\ell)$. 
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 self-avoiding 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 LR$(B)$. We similarly define the event, 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 4.6 for a generalization.)

**Lemma 4.58 (RSW lemma).** For all $n \geq 2$ (divisible by 4) and $p \in (0, 1)$,

$$R_{n,3}(p) \geq \frac{1}{4} R_{n,1}(p)^{11} R_{n/2,1}(p)^{12}. \quad (4.20)$$

The r.h.s. of (4.20) 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 Lemma 4.57.** The point of (4.20) is that, if $R_{n,1}(p)$ is bounded away from 0 uniformly in $n$, so is the l.h.s. By the argument in the proof of Harris’ theorem, this then implies that an open self-avoiding cycle exists in $\text{Ann}(n)$ with a probability bounded away from 0 as well. Hence to prove Lemma 4.57 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 $L^2$ is a variant of the contour lemma (Lemma 2.17). Its proof is similar and Exercise 4.7 asks for the details (the “if” direction being the non-trivial implication).

**Lemma 4.59.** 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 $[1/2, n + 1/2] \times [-1/2, n + 1/2]$. 

235
Figure 4.5: Illustration of the implication $\text{LR}(B_0^1) \cap \text{TB}(B_0^1 \cap B_0^2) \cap \text{LR}(B_0^2) \subseteq \text{LR}(B(3n, n))$.

By symmetry, when $p = 1/2$, the two events in Lemma 4.59 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 this proof.)

**Proof of Lemma 4.58.** 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

\[
\text{LR}(B_1') \cap \text{TB}(B_1' \cap B_2') \cap \text{LR}(B_2') \subseteq \text{LR}(B(3n, n)).
\]

See Figure 4.5. Each event on the l.h.s. 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:

**Lemma 4.60** (Proof of RSW: step 1).
\[ R_{n,3}(p) \geq R_{n,3/2}(p)^4 R_{n,1}(p)^3. \] (4.21)

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

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 self-avoiding 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 self-avoiding path connecting the right side of \( B_2 \) to \( P'_1 \) and stays inside \( B_2 \). Then we have the implication
\[ \Gamma_1 \cap \text{LR}(S) \cap \Gamma'_2 \subseteq \text{LR}(B(3n/2,n)). \]

See Figure 4.6. By symmetry \( \mathbb{P}_p[\Gamma_1] = \mathbb{P}_p[\Gamma'_2] \). Moreover, the events on the l.h.s. are increasing so by the FKG inequality:

**Lemma 4.61** (Proof of RSW: step 2).
\[ R_{n,3/2}(p) \geq \mathbb{P}_p[\Gamma_1]^2 R_{n/2,1}(p). \] (4.22)

**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 \( \{ P_2 \} \) 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 \( \text{TB}(S) \) holds, one also gets the existence of a unique rightmost crossing. Roughly speaking, take the rightmost crossing
Figure 4.6: Top: illustration of the implication $\Gamma_1 \cap \text{LR}(S) \cap \Gamma_2^\prime \subseteq \text{LR}(B(3n/2, n))$. Bottom: the event $\text{LR}^+([P^*]) \cap \{P = P^*_S\}$; the dashed path is the mirror image of the rightmost top-bottom crossing in $S$; the pink region is the complement in $B_1$ of the set $[P^*]$. Note that, because on the bottom figure the left-right path must stay within $[P^*]$, the configuration shown in the top figure where the purple left-right path “travels behind” the top-bottom crossing of $S$ cannot occur.
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 $P^*_S \leq P$ for all top-bottom crossings $P$ of $S$. (We accept as a fact the existence of a unique rightmost crossing. See Exercise 4.7 for a related construction.)

Let $I_S$ be the set of self-avoiding (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 $P$ through the $x$-axis in $B_{12} \setminus S$ and we let $\frac{P'}{P}$ be the union of $P$ and $P'$. Define $\left[\frac{P}{P}\right]$ to be the 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 $LR^+\left(\left[\frac{P}{P}\right]\right)$ be the event that there is a left-right crossing of $\left[\frac{P}{P}\right]$ ending on $P$, i.e., that there is an open self-avoiding path connecting the left side of $B_1$ and $P$ that stays within $\left[\frac{P}{P}\right]$. See Figure 4.6. Note that the existence of a left-right crossing of $B_1$ implies the existence of an open self-avoiding path connecting the left side of $B_1$ to $\frac{P}{P}$. By symmetry we then get

$$\mathbb{P}_p \left[ LR^+\left(\left[\frac{P}{P}\right]\right) \right] \geq \frac{1}{2} \mathbb{P}_p[LR(B_1)] = \frac{1}{2} R_{n,1}(p). \quad (4.23)$$

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 $TB(S)$. 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 $\left[\frac{P}{P}\right]$, e.g. of the event $LR^+\left(\left[\frac{P}{P}\right]\right)$. Hence

$$\mathbb{P}_p \left[ LR^+\left(\left[\frac{P}{P}\right]\right) \mid P^*_S = P \right] = \mathbb{P}_p \left[ LR^+\left(\left[\frac{P}{P}\right]\right) \right]. \quad (4.24)$$

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 (4.23) and (4.24), we get

$$\mathbb{P}_p[\Gamma_1] \geq \sum_{P \in I_S} \mathbb{P}_p[P^*_S = P] \mathbb{P}_p \left[ LR^+\left(\left[\frac{P}{P}\right]\right) \mid P^*_S = P \right]$$

$$\geq \frac{1}{2} R_{n,1}(p) \sum_{P \in I_S} \mathbb{P}_p[P^*_S = P]$$

$$= \frac{1}{2} R_{n,1}(p) \mathbb{P}_p[TB(S)]$$

$$= \frac{1}{2} R_{n,1}(p) R_{n/2,1}(p).$$

239
We have proved:

**Lemma 4.62** (Proof of RSW: step 3).

\[ \mathbb{P}_p[\Gamma_1] \geq \frac{1}{2} R_{n,1}(p) R_{n/2,1}(p). \] (4.25)

**Step 4: putting everything together** Combining (4.21), (4.22) and (4.25) gives that

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

Collecting the terms concludes the proof of the RSW lemma.

**Remark 4.63.** This argument is quite subtle. It is instructive to read the remark after [Gri97, Theorem 9.3].

### 4.4 Couplings of Markov chains

As we have seen, coupling is useful to bound total variation distance. A natural application is mixing as we show in this section.

#### 4.4.1 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 that, for a fixed \( 0 < \varepsilon < 1/2 \), the mixing time of \( P \) is

\[ t_{\text{mix}}(\varepsilon) := \min\{ t : d(t) \leq \varepsilon \}, \]

where

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

It will be easier to work with

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

The quantities \( d(t) \) and \( \bar{d}(t) \) are related in the following way.
Lemma 4.64.

\[ d(t) \leq \bar{d}(t) \leq 2d(t), \quad \forall t. \]

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

\[ \| P^t(x, \cdot) - \pi \|_{TV} = \sup_{A \subseteq V} |P^t(x, A) - \pi(A)| \]

\[ = \sup_{A \subseteq V} \left| \sum_{y \in V} \pi(y) [P^t(x, A) - P^t(y, A)] \right| \]

\[ \leq \sup_{A \subseteq V} \sum_{y \in V} \pi(y) \| P^t(x, A) - P^t(y, A) \| \]

\[ \leq \sum_{y \in V} \pi(y) \left\{ \sup_{A \subseteq V} |P^t(x, A) - P^t(y, A)| \right\} \]

\[ \leq \sum_{y \in V} \pi(y) \| P^t(x, \cdot) - P^t(y, \cdot) \|_{TV} \]

\[ \leq \max_{x, y \in V} \| P^t(x, \cdot) - P^t(y, \cdot) \|_{TV}. \]

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

Markovian coupling A coupling of Markov chains with transition matrix \( P \) is a Markov chain \((X_t, Y_t)\) on \( V \times V \) such that both \((X_t)\) and \((Y_t)\) are Markov chains with transition matrix \( P \). For our purposes, the following special type of coupling will suffice.

Definition 4.65 (Markovian coupling). A Markovian coupling of \( P \) is a Markov chain \((X_t, Y_t)\) on \( V \times V \) with transition matrix \( Q \) satisfying:

- (Markovian coupling) 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:

- (Coalescing) For all \( z \in V \),

\[ x' \neq y' \implies Q((z, z), (x', y')) = 0. \]
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 \(\tau_{coal}\) be the coalescence time (also called coupling time), i.e.,

\[
\tau_{coal} := \inf \{ t \geq 0 : X_t = Y_t \}.
\]

The key to the coupling approach to mixing times is the following immediate consequence of the coupling inequality (Lemma 4.9). For any starting point \((x, y)\),

\[
\|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq \mathbb{P}_{(x,y)}[X_t \neq Y_t] = \mathbb{P}_{(x,y)}[\tau_{coal} > t]. \tag{4.26}
\]

Combining (4.26) and Lemma 4.64, we get the main result of this section.

**Theorem 4.66** (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

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

In particular

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

We give a few simple examples in the next subsection.

**Example 4.67** (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

\[
\sup_{z \in V} \inf_{w \in V} P^s(w, z) \geq \delta.
\]

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

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

It can be checked that \(\tilde{P}\) is a transition matrix on \(V\) provided \(z_0\) satisfies the condition above (see Exercise 4.9). We use a technique known as splitting. While \(X_t \neq Y_t\), at the next time step, (a) with probability \(\delta\) we set \(X_{t+1} = Y_{t+1} = z_0\), (b) otherwise we pick independently \(X_{t+1} \sim \tilde{P}(X_t, \cdot)\) and \(Y_{t+1} \sim \tilde{P}(Y_t, \cdot)\). 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 \( x \neq y \),

\[
Q((x, y), (x', y')) = \delta 1\{x' = y' = z_0\} + (1 - \delta) \tilde{P}(x, x') \tilde{P}(y, y'),
\]
while if \( x = y \),

\[
Q((x, x), (x', x')) = P(x, x'),
\]
and \( Q \) is 0 in the other cases.

Observe that, in case (a) above, coalescence occurs at time \( t + 1 \). In case (b), 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 \( \tau_{\text{coal}} \) is stochastically dominated by a geometric random variable with success probability \( \delta \), or

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

By Theorem 4.66,

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

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

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

\[
\max_{x, y \in V} \mathbb{P}(x, y)[\tau_{\text{coal}} > ts] \leq (1 - \delta)^t,
\]
so that, after a change of variable,

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

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, but a straightforward application of the technique may lead to a bound depending badly on the size of the state space \( V \) (see Exercise 4.10).

4.4.2 Markov chains: 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 derived in Section 2.4.7. Recall that, by Claim 2.76, for a finite,
reversible Markov chain with stationary distribution $\pi$ and diameter $\Delta$ we have the lower bound

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

for $\varepsilon > 0$, where $\pi_{\min}$ is the smallest value taken by $\pi$.

**Cycle**  Let $(Z_t)$ be lazy simple random walk on the cycle of size $n$, $Z_n := \{0, 1, \ldots, n - 1\}$, where $i \sim j$ if $|j - i| = 1 \pmod{n}$. For any starting points $x, y$, we construct a coalescing 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 $\tau_{\text{coal}} = \tau_{\{0,n\}}^D$, the first time $(D_t)$ hits $\{0, n\}$.

We use Markov’s inequality, Theorem 2.1, to bound $P_{(x,y)}[\tau_{\{0,n\}}^D > t]$. Denote by $D_0 = d_{x,y}$ the starting distance. By Wald’s second equation (e.g. [Dur10, Example 4.1.6]),

$$\mathbb{E}_{(x,y)}[\tau_{\{0,n\}}^D] = d_{x,y}(n - d_{x,y}).$$

Applying Theorem 4.66 and Markov’s inequality we get

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

$$\leq \max_{x,y \in V} \mathbb{E}_{(x,y)}[\tau_{\{0,n\}}^D] \frac{t}{t}$$

$$= \max_{x,y \in V} d_{x,y}(n - d_{x,y}) \frac{t}{t}$$

$$\leq \frac{n^2}{4t},$$

or:

**Claim 4.68.**

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

By our diameter-based lower bound on mixing in Section 2.4.7, 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 2.76 gives

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

for $n$ large enough. Exercise 4.11 sketches a tighter lower bound.
Hypercube Let \((Z_t)\) be lazy simple random walk on the \(n\)-dimensional hypercube \(Z_2^n := \{0, 1\}^n\) where \(i \sim j\) if \(\|i - j\|_1 = 1\). We denote the coordinates of \(Z_t\) by \((Z_t^{(1)}, \ldots, Z_t^{(n)})\). 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 \(b\), i.e., \(X_t^{(i)} = Y_t^{(i)} = b\). This is equivalent to performing the Glauber dynamics chain on an empty graph. Because of the way the updating is done, the chains stay put with probability \(1/2\) at each time as required. Clearly the chains coalesce when all coordinates have been updated at least once. The following standard bound on the coupon collector problem is what is needed to conclude.

Lemma 4.69 (Coupon collecting). Let \(\tau_{\text{coll}}\) be the time it takes to update each coordinate at least once. Then, for any \(c > 0\),

\[
\Pr[\tau_{\text{coll}} > [n \log n + cn]] \leq e^{-c}.
\]

Proof. Let \(B_i\) be the event that the \(i\)-th coordinate has not been updated by time \([n \log n + cn]\). Then

\[
\Pr[\tau_{\text{coll}} > [n \log n + cn]] \leq \sum_i \Pr[B_i] = \sum_i \left(1 - \frac{1}{n}\right)^{[n \log n + cn]} \leq n \exp\left(-\frac{n \log n + cn}{n}\right) = e^{-c}.
\]

Applying Theorem 4.66, we get

\[
d([n \log n + cn]) \leq \max_{x,y \in V} \Pr_{(x,y)}[\tau_{\text{coal}} > [n \log n + cn]] \\
\leq \Pr[\tau_{\text{coll}} > [n \log n + cn]] \\
\leq e^{-c}.
\]

Hence for \(c := c_\epsilon > 0\) large enough:

Claim 4.70.

\[
t_{\text{mix}}(\epsilon) \leq [n \log n + c_\epsilon n].
\]
Again we get a quick lower bound using our diameter-based result from Section 2.4.7. Here $\Delta = n$ and $\pi_{\text{min}} = 1/2^n$ so that Claim 2.76 gives

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

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

**Claim 4.71.**

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

**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 $z_0$

$$d(t) \geq \|P^t(z_0, \cdot) - \pi\|_{\text{TV}} \geq |P^t(z_0, A) - \pi(A)|. \quad (4.27)$$

Under the stationary distribution, the Hamming weight is equal in distribution to a $\text{Bin}(n, 1/2)$. 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 $z_0 = 1$, where 1 is the all-1 vector. Let $U_t$ be the number of updated coordinates up to time $t$ in the Glauber dynamics representation of the chain discussed above. We use Chebyshev’s inequality (Theorem 2.2) to bound the probability of event $A$ at time $t$. We first need to compute the expectation and variance of $W_t$. Observe that, conditioned on $U_t$, the Hamming weight $W_t$ is equal in distribution to $\text{Bin}(U_t, 1/2) + (n - U_t)$ as the updated coordinates are uniform and the other ones are 1. Thus we have

$$E[W_t] = E[E[W_t \mid U_t]]$$

$$= E\left[\frac{1}{2} U_t + (n - U_t)\right]$$

$$= n - \frac{1}{2} n \left[1 - \left(1 - \frac{1}{n}\right)^t\right]$$

$$= \frac{n}{2} \left[1 + \left(1 - \frac{1}{n}\right)^t\right], \quad (4.28)$$

where on the third line we used that $E[U_t] = n \left[1 - \left(1 - \frac{1}{n}\right)^t\right]$ by linearity of expectation, and

$$\text{Var}[W_t] = E[\text{Var}[W_t \mid U_t]] + \text{Var}[E[W_t \mid U_t]]$$

$$= \frac{1}{4} E[U_t] + \frac{1}{4} \text{Var}[U_t]. \quad (4.29)$$
It remains to compute $\text{Var}[U_t]$. Let $I_t^{(i)}$ be 1 if coordinate $i$ has not been updated up to time $t$ and 0 otherwise. Note that for $i \neq j$

$$\text{Cov}[I_t^{(i)}, I_t^{(j)}] = \mathbb{E}[I_t^{(i)}I_t^{(j)}] - \mathbb{E}[I_t^{(i)}]\mathbb{E}[I_t^{(j)}]$$

$$= \left(1 - \frac{2}{n}\right)^t - \left(1 - \frac{1}{n}\right)^{2t}$$

$$= \left(1 - \frac{2}{n}\right)^t - \left(1 - \frac{2}{n} + \frac{1}{n^2}\right)^t$$

$$\leq 0,$$

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

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

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

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

Plugging this back into (4.29), we get

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

For $t_\alpha = \frac{1}{2} n \log n - (\log 2\alpha)n$ with $\alpha > 0$, by (4.28),

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

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 (Theorem 2.2), for $t_\alpha = \frac{1}{2} n \log n - (\log 2\alpha)n$ 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 (4.27), that implies $d(t_\alpha) \geq \varepsilon$ and we are done. □

The proof of Claim 4.71 relies on a “distinguishing statistic.” Recall from Lemma 4.17 that for any random variables $X$, $Y$ and map $h$ it holds that $\|\mu_h(X) - \mu_h(Y)\|_{\text{TV}} \leq$
\[ \|\mu_X - \mu_Y\|_{TV}, \] where \( \mu_Z \) is the law of \( Z \). The map 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 4.12 for a more general instantiation of the distinguishing statistic approach.

**Remark 4.72.** The upper bound in Claim 4.70 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 that

\[
\lim_{n \to +\infty} \frac{t_{\text{mix}}^{(n)}(\varepsilon)}{t_{\text{mix}}^{(n)}(1 - \varepsilon)} = 1,
\]

where \( t_{\text{mix}}^{(n)}(\varepsilon) \) 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 6.2 for a necessary condition for cutoff.

**b-ary tree** Let \( (Z_t) \) be lazy simple random walk on the \( \ell \)-level rooted \( b \)-ary tree, \( \overline{T}_{b,\ell} \). 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. Hence 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 construct a coupling \( (X_t, Y_t) \) of this chain started at \( (x, y) \). Assume w.l.o.g. that \( x \) is no further from the root than \( y \), which we denote by \( x \preceq y \). 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 \( Y_t \) and proceed in the same manner. Do this until \( X_t \) and \( Y_t \) are on the same level.

- In the second stage, i.e., once the two chains are on the same level, at each time first let \( X_t \) move as a lazy random walk on \( \overline{T}_{b,\ell} \). Then let \( Y_t \) move in the same direction as \( X_t \), i.e., if \( X_t \) moves closer to the root, so does \( Y_t \) and so on.

By construction, \( X_t \preceq Y_t \) 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 \( X_t \preceq Y_t \), when \( X_t \) reaches the leaves, so does

248
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, so does $Y_t$. Hence $\tau_{\text{coal}} \leq \tau^*$. 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 $P(x,y)[\tau^* > t]$, we use Markov’s inequality (Theorem 2.1), 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 3.123),

$$\mathbb{E}_0[\tau_{0,\ell}] = c_{\mathcal{N}} \mathcal{R}(0 \leftrightarrow \ell),$$

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 $\mathbb{P}_{b,\ell}$ and the factor of 4 accounts for self-loops. Using network reduction techniques, we computed the effective resistance $\mathcal{R}(0 \leftrightarrow \ell)$ in Examples 3.104 and 3.105—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

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

which implies

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

Finally, applying Theorem 4.66 and Markov’s inequality and using (4.30), we get

$$d(t) \leq \max_{x,y \in V} P(x,y)[\tau^* > t] \leq \frac{\mathbb{E}_{(x,y)}[\tau^*]}{t} \leq \frac{\mathbb{E}_0[\tau_{0,\ell}]}{t} \leq \frac{4n}{t},$$

or:

249
Figure 4.7: Setup for the lower bound on the mixing time on a $b$-ary tree. (Here $b = 2$.)

**Claim 4.73.**

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

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 2.76 gives

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

for $n$ large enough. Here is a better lower bound. Intuitively the mixing time is significantly greater than the squared diameter because the chain tends to be pushed away from the root: going from the leaves on one side of the root to the leaves on the other typically takes time linear in $n$. Formally let $x_0$ be a leaf of $T_{b,\ell}$ and let $A$ be the set of vertices “on the other side of root (inclusively),” i.e., 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

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

to bound the mixing time. We claim that, started at $x_0$, the walk typically takes time linear in $n$ to reach $A$. Consider again the level $L_\ell$ of $X_\ell$. Using the expression for
the effective resistance above, we have

\[ P \ell \mathbb{P}(0 \leftrightarrow \ell) = \frac{1}{c(\ell)} \mathbb{P}(0 \leftrightarrow \ell) = \frac{1}{2b} \cdot \frac{b-1}{1-b^{-\ell}} = \frac{b-1}{2b - 2} = O \left( \frac{1}{n} \right). \]

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. 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,

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

for all \( t \leq \alpha \epsilon n \) with \( \alpha \epsilon > 0 \) small enough and

\[ \|P^{\alpha \epsilon n}(x_0, \cdot) - \pi\|_{TV} \geq |P^{\alpha \epsilon n}(x_0, A) - \pi(A)| > \epsilon. \]

We have proved that \( t_{mix}(\epsilon) \geq \alpha \epsilon n. \)

### 4.4.3 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 are given 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

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

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. Let

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

be the weighted diameter of \( H_0 \).

**Theorem 4.74** (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
\[
\mathbb{E}[w_0(X^*, Y^*)] \leq \kappa w_0(x, y).
\] (4.31)

Then
\[
d(t) \leq \Delta_0 \kappa^t,
\]
or
\[
t_{\text{mix}}(\varepsilon) \leq \left\lceil \frac{\log \Delta_0 + \log \varepsilon^{-1}}{\log \kappa} \right\rceil.
\]

Proof. The crux of the proof is to extend (4.31) to arbitrary pairs of vertices.

Claim 4.75 (Global coupling). For all \(x, y \in V\), there is a coupling \((X^*, Y^*)\) of \(P(x, \cdot)\) and \(P(y, \cdot)\) such that (4.31) holds.

Iterating the coupling in this claim immediately implies the existence of a coalescing Markovian coupling \((X_t, Y_t)\) of \(P\) such that
\[
\mathbb{E}_{(x, y)}[w_0(X_t, Y_t)] = \mathbb{E}_{(x, y)}[\mathbb{E}[w_0(X_t, Y_t) | X_{t-1}, Y_{t-1}]] \\
\leq \mathbb{E}_{(x, y)}[\kappa w_0(X_{t-1}, Y_{t-1})] \\
\leq \cdots \\
\leq \kappa^t \mathbb{E}_{(x, y)}[w_0(X_0, Y_0)] \\
= \kappa^t w_0(x, y) \\
\leq \kappa^t \Delta_0.
\]

By assumption, \(1_{\{x \neq y\}} \leq w_0(x, y)\) so that by the coupling inequality and Lemma 4.64, 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. It remains to prove Claim 4.75.

Proof of Claim 4.75. 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 (4.31). 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
Figure 4.8: Coupling of $P(x', \cdot)$ and $P(y', \cdot)$ constructed from a sequence of local couplings $(Z_{0,0}^*, Z_{0,1}^*), \ldots, (Z_{0,m-1}^*, Z_{0,m-1}^*)$.

$i$, $(X^*, Y^*) := (Z^{(0)}, Z^{(m)})$ is then a coupling of $P(x', \cdot)$ and $P(y', \cdot)$. Formally, define the transition matrix

$$R_i(z^{(i)}, z^{(i+1)}) := \mathbb{P}[Z_{i,1}^* = z^{(i+1)} \mid Z_{i,0}^* = z^{(i)}],$$

and observe that

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

(4.32)

and

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

(4.33)

by construction of the coupling $(Z_{i,0}^*, Z_{i,1}^*)$. See Figure 4.8. The law of the full coupling $(Z^{(0)}, \ldots, Z^{(m)})$ is

$$\mathbb{P}[(Z_0, \ldots, Z_m) = (z^{(0)}, \ldots, z^{(m)})] = P(x_0, z^{(0)}) R_0(z^{(0)}, z^{(1)}) \cdots R_{m-1}(z^{(m-1)}, z^{(m)}).$$

Using (4.32) and (4.33) inductively gives

$$\mathbb{P}[X^* = z^{(0)}] = \mathbb{P}[Z^{(0)} = z^{(0)}] = P(x_0, z^{(0)}),$$

$$\mathbb{P}[Y^* = z^{(m)}] = \mathbb{P}[Z^{(m)} = z^{(m)}] = P(x_m, z^{(m)}),$$

253
as required.

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

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

where, on the third line, we used (4.31) 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 two subsections. See Exercise 4.13 for an optimal transport perspective on the path coupling method.

### 4.4.4 Ising model: Glauber dynamics at high temperature

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

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

where

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

is the Hamiltonian and

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

is the partition function. In this context, recall that vertices are often referred to as sites. The single-site Glauber dynamics of the Ising model is the Markov chain on $X$ 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 $V \setminus \{i\}$. Specifically, for $\gamma \in \{-1, +1\}$, $i \in V$, and $\sigma \in X$, 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

\[
Q_\beta(\sigma, \sigma^{i\gamma}) := \frac{1}{n} \cdot \frac{e^{\gamma S_i(\sigma)}}{e^{-\beta S_i(\sigma)} + e^{\beta S_i(\sigma)}} = \frac{1}{n} \left\{ \frac{1}{2} + \frac{1}{2} \tanh(\gamma \beta S_i(\sigma)) \right\},
\]

(4.34)

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, \( t_{\text{mix}}(\varepsilon) \), of \( Q_\beta \) using path coupling. We say that the Glauber dynamics is fast mixing if \( t_{\text{mix}}(\varepsilon) = O(n \log n) \). We first make a simple observation:

**Claim 4.76** (Glauber dynamics: lower bound on mixing).

\[
t_{\text{mix}}(\varepsilon) = \Omega(n), \quad \forall \beta > 0.
\]

**Proof.** We use a coupon collecting argument. Let \( \hat{\sigma} \) 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, \( \mu_\beta(\hat{\sigma}) = \mu_\beta(A^c) = 1/2 \) where we assumed for simplicity that \( n \) is odd. By definition of the total variation distance

\[
d(t) \geq \|Q^t_\beta(\hat{\sigma}, \cdot) - \mu_\beta(\cdot)\|_{TV} \geq |Q^t_\beta(\hat{\sigma}, A) - \mu_\beta(A)| = |Q^t_\beta(\hat{\sigma}, A) - 1/2|.
\]

So it remains to show that by time \( cn \), 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 2.4, we are seeking a bound on \( T_{n,n/3} \), i.e., 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 \( \mathbb{E}[\tau_{n,i}] = (1 - \frac{i-1}{n})^{-1} \) and \( \text{Var}[\tau_{n,i}] = (1 - \frac{i-1}{n})^{-2} \). Hence

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

(4.35)

and

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

(4.36)
So by Chebyshev’s inequality (Theorem 2.2)

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

by (4.35) and (4.36). Taking \( \varepsilon > 0 \) small enough and \( n \) large enough, we have shown that, for \( t \leq c_\varepsilon n \) for some \( c_\varepsilon > 0 \),

\[ Q^t_\beta(\bar{\sigma}, A) \leq 1/3, \]

which proves the claim.

\[ \blacksquare \]

**Remark 4.77.** In fact, Ding and Peres proved that \( t_{\text{mix}}(\varepsilon) = \Omega(n \log n) \) for any graph on \( n \) vertices [DP11]. In Claim 4.71, we treated the special case of the empty graph, which is equivalent to lazy random walk on the hypercube.

In our main result, 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 4.78** (Glauber dynamics: fast mixing at high temperature).

\[ \beta < \tilde{\delta}^{-1} \implies t_{\text{mix}}(\varepsilon) = O(n \log n). \]

**Proof.** We use path coupling. Let \( H_0 = (V_0, E_0) \) where \( V_0 := \mathcal{X} \) and \( \{\sigma, \omega\} \in E_0 \) if \( \frac{1}{2}\|\sigma - \omega\|_1 = 1 \) with unit \( w_0 \)-weights 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 \( i_0 \) to update. If \( i_0 \) is such that all its neighbors in \( G \) have the same state in \( \sigma \) and \( \omega \), i.e., if \( \sigma_j = \omega_j \) for all \( j \sim i_0 \), we update \( X^* \) from \( \sigma \) according to the Glauber rule and set \( Y^* := X^* \). Note that this includes the case \( i_0 = i \). Otherwise, i.e. if \( i_0 \sim i \), we proceed as follows. From the state \( \sigma \), the probability of updating site \( i_0 \) to state \( \gamma \in \{-1, +1\} \) is given by the expression in brackets in (6.6), and similarly for \( \omega \). Unlike the previous case, we cannot guarantee that the update is identical in both chains. However, in order to minimize the chances of increasing the distance between the two chains, we pick a uniform-\([-1, 1]\) variable \( U \) and set

\[ X^*_{i_0} := \begin{cases} +1, & \text{if } U \leq \tanh(\beta S_{i_0}(\sigma)) \\ -1, & \text{o.w.} \end{cases} \]

and

\[ Y^*_{i_0} := \begin{cases} +1, & \text{if } U \leq \tanh(\beta S_{i_0}(\omega)) \\ -1, & \text{o.w.} \end{cases} \]
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
\[
\mathbb{E}[w_0(X^*, Y^*)] = 1 - \frac{1}{n} \sum_{j \sim i} \frac{1}{2} \left| \tanh(\beta S_j(\sigma)) - \tanh(\beta S_j(\omega)) \right|, \tag{4.37}
\]
where (a) corresponds to \( i_* = i \) in which case \( w_0(X^*, Y^*) = 0 \) and (b) corresponds to \( i_* \sim i \) in which case \( w_0(X^*, Y^*) = 2 \) with probability \( \frac{1}{2} \left| \tanh(\beta S_{i_*}(\sigma)) - \tanh(\beta S_{i_*}(\omega)) \right| \) by our coupling, and \( w_0(X^*, Y^*) = 1 \) otherwise. To bound (b), we note that for \( j \sim i \)
\[
\left| \tanh(\beta S_j(\sigma)) - \tanh(\beta S_j(\omega)) \right| = \tanh(\beta(s + 2)) - \tanh(\beta s), \tag{4.38}
\]
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 r.h.s. of (4.38) is \( \leq 2\beta \). Plugging this back into (4.37), 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 assumption. The diameter of \( H_0 \) is \( \Delta_0 = n \). By Theorem 4.74,
\[
\tau_{\text{mix}}(\varepsilon) \leq \left\lceil \frac{\log \Delta_0 + \log \varepsilon^{-1}}{\log \kappa^{-1}} \right\rceil = \left\lceil \frac{n(\log n + \log \varepsilon^{-1})}{1 - \delta \beta} \right\rceil,
\]
which implies the claim.

\[\Box\]

**Remark 4.79.** 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].

### 4.5 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. The basic setup is a sum of \(\{0, 1\}\)-valued random variables \(\{X_i\}_{i=1}^n\)

\[
W = \sum_{i=1}^n X_i
\]

(4.39)

where the \(X_i\)'s are not assumed independent or identically distributed. Define

\[
p_i = P[X_i = 1]
\]

(4.40)

and

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

(4.41)

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

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.

### 4.5.1 Main bounds and examples

We begin with a definition.

**Definition 4.80 (Stein coupling).** Using the notation in (4.39), (4.40) and (4.41), a Stein coupling is a pair \((U_i, V_i)\), for each \(i = 1, \ldots, n\), such that

\[
U_i \sim W, \quad V_i \sim W - 1|X_i = 1.
\]

Each pair \((U_i, V_i)\) is defined on a joint probability space, but different pairs are not.

The key bound is the following. It is proved in Section 4.5.2.

**Theorem 4.81 (Chen-Stein method).** Using the notation in (4.39), (4.40) and (4.41), let \((U_i, V_i), i = 1, \ldots, n\), be a Stein coupling. Then

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

(4.42)

where, as before, \(\mu\) denotes the law of \(W\) and \(\pi\) is the Poisson distribution with mean \(\lambda\).

As a first example, we derive a Poisson approximation result in the independent case. Compare to Theorem 4.16.
Example 4.82 (Independent $X_i$’s). Assume the $X_i$’s are independent. We prove the following:

Claim 4.83.

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

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

$$U_i = W$$

and

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

By independence,

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

as desired. Plugging into (4.42), we obtain the bound

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

$$\leq (1 \wedge \lambda^{-1}) \sum_{i=1}^{n} p_i \mathbb{E} |W - \sum_{j \neq i} X_j|$$

$$\leq (1 \wedge \lambda^{-1}) \sum_{i=1}^{n} p_i \mathbb{E} |X_i|$$

$$\leq (1 \wedge \lambda^{-1}) \sum_{i=1}^{n} p_i^2.$$  

Here is a less straightforward example.

Example 4.84 (Balls in boxes). Suppose we throw $k$ balls uniformly at random in $n$ boxes independently. Let

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

and let $W = \sum_{i=1}^{n} X_i$ be the number of empty boxes. Note that the $X_i$’s are not independent. (Think of what happens with one ball.) Note that

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

259
for all \(i\) and, hence,
\[
\lambda = n \left(1 - \frac{1}{n}\right)^k.
\]

For each \(i = 1, \ldots, n\), we generate the \((U_i, V_i)\) 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
\[
\mathbb{E}[U_i - V_i] = \mathbb{E}[U_i - V_i] = \lambda - \sum_{i=1}^{n} p_i \mathbb{E}[V_i].
\]

By the fact that \(V_i \sim U_i - 1 | X_i = 1\) and Bayes’ rule,
\[
\sum_{i=1}^{n} p_i \mathbb{E}[V_i] = \sum_{i=1}^{n} \mathbb{P}[X_i = 1] \sum_{k=1}^{n} (k - 1) \mathbb{P}[V_i = k - 1]
\]
\[
= \sum_{i=1}^{n} \sum_{k=1}^{n} (k - 1) \mathbb{P}[U_i = k | X_i = 1] \mathbb{P}[X_i = 1]
\]
\[
= \sum_{i=1}^{n} \sum_{k=1}^{n} (k - 1) \mathbb{P}[X_i = 1 | U_i = k] \mathbb{P}[U_i = k]
\]
\[
= \sum_{i=1}^{n} \sum_{k=1}^{n} (k - 1) \mathbb{E}[X_i | U_i = k] \mathbb{P}[U_i = k]
\]
\[
= \sum_{k=1}^{n} (k - 1) k \mathbb{P}[U_i = k]
\]
\[
= \mathbb{E}[U_i^2] - \mathbb{E}[U_i].
\]

It remains to compute \(\mathbb{E}[U_i^2] = \mathbb{E}[W^2]\). We have by symmetry
\[
\mathbb{E}[W^2] = n \mathbb{E}[X_1^2] + n(n - 1) \mathbb{E}[X_1 X_2]
\]
\[
= \lambda + n(n - 1) \left(1 - \frac{2}{n}\right)^k,
\]
so
\[
\|\mu - \pi\|_{TV} \leq (1 \land \lambda^{-1}) \left\{ \lambda^2 - n(n - 1) \left(1 - \frac{2}{n}\right)^k \right\}.
\]
This last example is generalized in Exercise 4.17.

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

**Theorem 4.85** (Chen-Stein: Dissociated Case). *Using the notation in (4.39), (4.40) and (4.41), suppose that for each \(i\) there is a neighborhood \(N_i \subseteq [n] \setminus \{i\}\) such that

\[
X_i \text{ is independent of } \{X_j : j \notin N_i \cup \{i\}\}.
\]

Then

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

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

\[U_i = W.\]

Then generate

\[(Y_j^{(i)})_{j \in N_i} \sim (X_j)_{j \in N_i} | \{X_k : k \notin N_i \cup \{i\}\}, X_i = 1,\]

and set

\[V_i = \sum_{k \notin N_i \cup \{i\}} X_k + \sum_{j \in N_i} Y_j^{(i)}.\]

Because the law of \(\{X_k : k \notin N_i \cup \{i\}\}\) is independent of the event \(\{X_i = 1\}\), the above scheme satisfies the conditions of the Stein coupling.
Therefore we can apply Theorem 4.81

\[ \|\mu - \pi\|_{TV} \leq (1 - \lambda^{-1}) \sum_{i=1}^{n} p_i |E[U_i - V_i]| \]

\[ = (1 - \lambda^{-1}) \sum_{i=1}^{n} p_i \left| \sum_{j=1}^{n} X_j - \sum_{k \in \mathcal{N}_i \cup \{i\}} X_k - \sum_{j \in \mathcal{N}_i} Y_j^{(i)} \right| \]

\[ = (1 - \lambda^{-1}) \sum_{i=1}^{n} p_i \left| X_i + \sum_{j \in \mathcal{N}_i} (X_j - Y_j^{(i)}) \right| \]

\[ \leq (1 - \lambda^{-1}) \sum_{i=1}^{n} p_i \left( E[X_i] + \sum_{j \in \mathcal{N}_i} [E[X_j] + E[Y_j^{(i)}]] \right) \]

\[ = (1 - \lambda^{-1}) \sum_{i=1}^{n} p_i \left( p_i + \sum_{j \in \mathcal{N}_i} [p_j + E[X_j | X_i = 1]] \right) \]

\[ = (1 - \lambda^{-1}) \sum_{i=1}^{n} \left\{ p_i^2 + \sum_{j \in \mathcal{N}_i} (p_i p_j + p_i E[X_j | X_i = 1]) \right\} \]

\[ = (1 - \lambda^{-1}) \sum_{i=1}^{n} \left\{ p_i^2 + \sum_{j \in \mathcal{N}_i} (p_i p_j + E[X_i X_j]) \right\} . \]

That concludes the proof. \[ \blacksquare \]

We give an example.

**Example 4.86** (Longest head run). Let \( 0 < q < 1 \) and let \( Z_1, Z_2, \ldots \) be i.i.d. Bernoulli random variables with \( q = P[Z_1 = 1] \). We are interested in the distribution of \( R \), the length of the longest run of 1’s starting in the first \( n \) tosses. For a positive integer \( t \), define

\[ X_1^{(t)} = Z_1 \cdots Z_t, \]

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

and

\[ W^{(t)} = \sum_{i=1}^{n} X_i^{(t)}. \]
The event $\{X_i^{(t)} = 1\}$ indicates that a head run of length at least $t$ starts at the $i$-th toss. The key observation is that

$$\{R < t\} = \{W^{(t)} = 0\}. \quad (4.43)$$

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 4.85. 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 4.85. We have

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

and, for $i \geq 2$,

$$p_i^{(t)} = \mathbb{E}[(1 - Z_{i-1})Z_i \cdots Z_{i+t-1}]$$

$$= \mathbb{E}[1 - Z_{i-1}] \prod_{j=i}^{i+t-1} \mathbb{E}[Z_j]$$

$$= (1 - q)q^{t}$$

$$\leq q^{t}.$$

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 $j$, where by definition of the neighborhoods $j$ is within $t$ of $i$. Hence, $\mathbb{E}[X_i^{(t)}X_j^{(t)}] = 0$ in that case. We also have

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

263
and 

\[ |\mathcal{N}_i^{(t)}| \leq 2t. \]

We are ready to apply Theorem 4.85. We get

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

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 \( n^{-1} \log n \).

Going back to (4.43), we finally obtain when \( t = \log_{1/q} n + C \) that

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

\[ \blacktriangle \]

4.5.2 Motivation and proof

The idea behind the Chen-Stein method is to interpolate between \( \mu \) and \( \pi \) in Theorem 4.81 by constructing a Markov chain with initial distribution \( \mu \) and stationary distribution \( \pi \). There are many ways to define such a chain. Here we use a discrete-time, finite Markov chain for simplicity.

Proof of Theorem 4.81. We seek a bound on

\[
\|\mu - \pi\|_{TV} = \sup_{A \subseteq \mathbb{Z}_+} |\mu(A) - \pi(A)| \\
= \mu(A^*) - \pi(A^*) \\
= \sum_{z \in A^*} (\mu(z) - \pi(z)),
\]

where \( A^* = \{ z \in \mathbb{Z}_+ : \mu(z) > \pi(z) \} \), by Lemma 4.13. 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 \( 0 \leq z \leq n \).
Constructing the Markov chain

For this purpose, it will be convenient to truncate \( \pi \) at \( n \), that is, we define

\[
\tilde{\pi}(z) = \begin{cases} 
\pi(z), & 0 \leq z \leq n, \\
1 - \Pi(n), & z = n + 1, \\
0, & \text{o.w.}
\end{cases}
\]

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 \( \tilde{\pi} \). We will also need the chain to be aperiodic and irreducible over \( \{0, 1, \ldots, n + 1\} \). We choose the transition matrix \( P(x, y) \) to be that of a birth-death chain reversible with respect to \( \tilde{\pi} \), that is, we require

\[
P(x, y) = P(y, x)_{x \neq y}
\]

and

\[
P(x, x) = \frac{\pi(x + 1)}{\pi(x)}, \quad \forall x \in [n].
\]

(4.45)

For \( x < n \),

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

In view of this, we let the nonzero transition probabilities take values

\[
P(x, y) = \begin{cases} 
C \lambda, & \text{if } y = x + 1, \ 0 \leq x \leq n, \\
C x, & \text{if } y = x - 1, \ 1 \leq x \leq n, \\
1 - C x - C \lambda, & \text{if } y = x, \ 1 \leq x \leq n, \\
1 - C \lambda, & \text{if } y = x = 0, \\
C \lambda \frac{\pi(n)}{1 - \Pi(n)}, & \text{if } y = x - 1, \ x = n + 1, \\
1 - C \lambda \frac{\pi(n)}{1 - \Pi(n)}, & \text{if } y = x = n + 1,
\end{cases}
\]

(4.46)

for a constant \( C \in (0, +\infty) \) small enough that all probabilities above are strictly positive. Then, \( P \) is aperiodic, irreducible and satisfies the detailed balance conditions (4.45).

Interpolating between \( \mu \) and \( \pi \)

By the convergence theorem for Markov chains, Theorem 1.22,

\[
P^t(y, z) \to \pi(z)
\]

for all \( 0 \leq y \leq n + 1 \) and \( 0 \leq z \leq n \) as \( t \to +\infty \). Hence, letting \( \delta_z(x) = 1_{\{x=z\}} \),

\[
\delta_z(y) - \pi(z) = \lim_{t \to +\infty} \mathbb{E}_y[\delta_z(X_0) - \delta_z(X_t)]
\]

\[
= \lim_{t \to +\infty} \sum_{s=0}^{t-1} \mathbb{E}_y[\delta_z(X_s) - \delta_z(X_{s+1})],
\]

(4.47)
and, taking expectation over $\mu$,

$$
\mu(z) - \pi(z) = \lim_{t \to +\infty} \mathbb{E}_{\mu}[\delta_z(X_0) - \delta_z(X_t)]
= \lim_{t \to +\infty} \sum_{s=0}^{t-1} \mathbb{E}_{\mu}[\delta_z(X_s) - \delta_z(X_{s+1})],
$$

(4.48)

where the subscript of $\mathbb{E}$ indicates the initial distribution or state, and we used the fact that $\pi(z) = \tilde{\pi}(z)$ on $\{0, 1 \ldots, n\}$.

**Markov chain calculations**  We use standard Markov chains facts to re-write the expression above in a more useful form. By Chapman-Kolmogorov, Theorem 1.13, applied to the first step of the chain,

$$
\mathbb{E}_y[\delta_z(X_{s+1})] = P(y, y+1)\mathbb{E}_{y+1}[\delta_z(X_s)] + P(y, y)\mathbb{E}_y[\delta_z(X_s)] + P(y, y-1)\mathbb{E}_{y-1}[\delta_z(X_s)].
$$

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 $0 \leq z \leq n$

$$
\sum_{s=0}^{t-1} \mathbb{E}_y[\delta_z(X_s) - \delta_z(X_{s+1})]
= \sum_{s=0}^{t-1} \mathbb{E}_y[\delta_z(X_s) - \delta_z(X_{s+1})]
= \sum_{s=0}^{t-1} \left\{ -P(y, y+1)(\mathbb{E}_{y+1}[\delta_z(X_s)] - \mathbb{E}_y[\delta_z(X_s)])
+ P(y, y-1)(\mathbb{E}_y[\delta_z(X_s)] - \mathbb{E}_{y-1}[\delta_z(X_s)]) \right\}
= -P(y, y+1)C^{-1}g_z^t(y+1) + P(y, y-1)C^{-1}g_z^t(y),
$$

(4.49)

where we defined

$$
g_z^t(y) = C \sum_{s=0}^{t-1} (\mathbb{E}_y[\delta_z(X_s)] - \mathbb{E}_{y-1}[\delta_z(X_s)]).
$$

(4.50)

The function $g_z^t(y)$ is the difference between the expected number of visits to $z$ when started at $y$ and $y-1$ respectively. We establish after the proof of the theorem that it has a well-defined limit. That fact is not immediately obvious as the limit of $g_z^t(y)$ is difference of two infinities.
**Lemma 4.87.** Let \( g_z^t : \{0, 1, \ldots, n + 1\} \to \mathbb{R} \) be defined in (4.50). Then there exists a bounded function \( g_z^\infty : \{0, 1, \ldots, n + 1\} \to \infty \) such that for all \( 0 \leq z \leq n \) and \( 0 \leq y \leq n + 1 \),
\[
g_z^\infty(y) = \lim_{t \to +\infty} g_z^t(y).
\]

**Chen’s equation and a recursion** For \( A \subseteq \{0, 1, \ldots, n\} \), let
\[
g_A^\infty(y) = \sum_{z \in A} g_z^\infty(y).
\]

We obtain the following key bound.

**Lemma 4.88** (Chen’s equation). Let \( W \sim \mu \) and \( \pi \sim \text{Poi}(\lambda) \). Then,
\[
\|\mu - \pi\|_{TV} = \mathbb{E} \left[ -\lambda g_\mu^\infty(W + 1) + W g_\pi^\infty(W) \right] \tag{4.51}
\]
where \( A^* = \{ z \in \mathbb{Z}_+ : \mu(z) > \pi(z) \} \).

**Proof.** Combine (4.44), (4.46), (4.48), (4.49) and Lemma 4.87. \( \blacksquare \)

The expectation on the right-hand side of (4.51) can actually be used to characterize the Poisson distribution and is the starting point of a simpler proof of the Chen-Stein method. See Exercise 4.14.

In fact, an explicit expression for \( g_z^\infty \) can be derived via the following recursion. That expression in turn will be helpful to bound the right-hand side of (4.51). For notational convenience, define
\[
g_z^\infty(0) = g_z^\infty(n + 2) = 0.
\]
for all \( 0 \leq z \leq n \).

**Lemma 4.89.** For all \( 0 \leq y \leq n \) and \( 0 \leq z \leq n \),
\[
\delta_z(y) - \pi(z) = -\lambda g_z^\infty(y + 1) + yg_z^\infty(y).
\]

**Proof.** Combine (4.47), (4.46), (4.49) and Lemma 4.87. \( \blacksquare \)

Lemma 4.89 leads to the following formula for \( g_z^\infty \), which we establish after the proof of the theorem.

**Lemma 4.90.** For \( 0 \leq y \leq n + 1 \) and \( 0 \leq z \leq n \),
\[
g_z^\infty(y) = \begin{cases} 
\frac{\Pi(y-1)}{y\pi(y)} \pi(z), & \text{if } z \geq y, \\
\frac{1-\Pi(y-1)}{y\pi(y)} \pi(z), & \text{if } z < y.
\end{cases} \tag{4.52}
\]
Lemma 4.90 in turn can be used to derive a Lipschitz constant for $g_A^\infty$. That lemma is also established after the proof of the theorem.

**Lemma 4.91.** For $A \subseteq \{0, 1, \ldots, n\}$ and $y, y' \in \{0, 1, \ldots, n + 1\}$,

$$|g_A^\infty(y') - g_A^\infty(y)| \leq (1 \wedge \lambda^{-1})|y' - y|.$$

**Using the Stein coupling** We now use Lemma 4.88 and Definition 4.80 to derive the final bound in the claim. By (4.51), using the facts that $\lambda = \sum_{i=1}^n p_i$ and $W = \sum_{i=1}^n X_i$, we get

$$\|\mu - \pi\|_{TV} = \mathbb{E} \left[ -\lambda g_A^\infty (W + 1) + W g_A^\infty (W) \right] \tag{4.53}$$

$$= \mathbb{E} \left[ - \left( \sum_{i=1}^n p_i \right) g_A^\infty (W + 1) + \left( \sum_{i=1}^n X_i \right) g_A^\infty (W) \right]$$

$$= \sum_{i=1}^n (-p_i \mathbb{E} [g_A^\infty (W + 1)] + \mathbb{E} [g_A^\infty (W) | X_i = 1] \mathbb{P}[X_i = 1])$$

$$= \sum_{i=1}^n p_i (-\mathbb{E} [g_A^\infty (W + 1)] + \mathbb{E} [g_A^\infty (W) | X_i = 1]). \tag{4.54}$$

Let $(U_i, V_i), i = 1, \ldots, n$, be a Stein coupling. Then, we can re-write this last expression as

$$\sum_{i=1}^n p_i (-\mathbb{E} [g_A^\infty (W + 1)] + \mathbb{E} [g_A^\infty (W) | X_i = 1])$$

$$= \sum_{i=1}^n p_i (-\mathbb{E} [g_A^\infty (U_i + 1)] + \mathbb{E} [g_A^\infty (V_i + 1)]). \tag{4.55}$$

By Lemma 4.91, combining (4.54) and (4.55) gives

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

which concludes the proof. \hfill \blacksquare
Additional proofs  It remains to prove Lemmas 4.87, 4.90 and 4.91.

Proof of Lemma 4.87. We use a coupling argument. Let \((Y_s, \tilde{Y}_s)_{s=0}^{+\infty}\) be an independent 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 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 3.26, 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 \((Y_s)\)). By the Strong Markov property, Theorem 3.11, the resulting chain \((Y^*_s, \tilde{Y}^*_s)\) is also a coupling of the chain started at \(y - 1\) and \(y\) respectively. Using this coupling, we re-write

\[
g_z^t(y) = C \sum_{s=0}^{t-1} (\mathbb{E}_{y}[\delta_z(X_s)] - \mathbb{E}_{y-1}[\delta_z(X_s)])
\]

\[
= C \sum_{s=0}^{t-1} \mathbb{E}[\delta_z(\tilde{Y}^*_s) - \delta_z(Y^*_s)]
\]

\[
= C \mathbb{E} \left[ \sum_{s=0}^{t-1} (\delta_z(\tilde{Y}^*_s) - \delta_z(Y^*_s)) \right].
\]

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 allows to take the limit, leading to

\[
g_z^\infty(y) = \lim_{t \to +\infty} C \mathbb{E} \left[ \sum_{s=0}^{t-1} (\delta_z(\tilde{Y}^*_s) - \delta_z(Y^*_s)) \right]
\]

\[
= C \mathbb{E} \left[ \sum_{s=0}^{+\infty} (\delta_z(\tilde{Y}^*_s) - \delta_z(Y^*_s)) \right]
\]

\[
< +\infty.
\]

That concludes the proof.
Proof of Lemma 4.90. Our starting point is Lemma 4.89, from which we deduce the recursive formula
\[
g^\infty_z(y + 1) = \frac{1}{\lambda} \left( yg^\infty_z(y) + \pi(z) - \delta_z(y) \right),
\]
(4.56)
for \(0 \leq y \leq n\) and \(0 \leq z \leq n\). Hence,
\[
g^\infty_z(1) = \frac{1}{\lambda} \{\pi(z) - \delta_z(0)\},
\]
(4.57)
\[
g^\infty_z(2) = \frac{1}{\lambda} \left\{ g^\infty_z(1) + \pi(z) - \delta_z(1) \right\} \\
= \frac{1}{\lambda} \left\{ \frac{1}{\lambda} \{\pi(z) - \delta_z(0)\} + \pi(z) - \delta_z(1) \right\} \\
= \frac{1}{\lambda^2} \{\pi(z) - \delta_z(0)\} + \frac{1}{\lambda} \{\pi(z) - \delta_z(1)\},
\]
\[
g^\infty_z(3) = \frac{1}{\lambda} \left\{ 2g^\infty_z(2) + \pi(z) - \delta_z(2) \right\} \\
= \frac{1}{\lambda} \left\{ 2 \frac{1}{\lambda^2} \{\pi(z) - \delta_z(0)\} + 2 \frac{1}{\lambda} \{\pi(z) - \delta_z(1)\} + \pi(z) - \delta_z(2) \right\} \\
= \frac{2}{\lambda^3} \{\pi(z) - \delta_z(0)\} + \frac{2}{\lambda^2} \{\pi(z) - \delta_z(1)\} + \frac{1}{\lambda} \{\pi(z) - \delta_z(2)\},
\]
and so forth. We posit the general formula
\[
g^\infty_z(y) = \frac{(y - 1)!}{\lambda^y} \sum_{k=0}^{y-1} \lambda^k \frac{1}{k!} \{\pi(z) - \delta_z(k)\},
\]
(4.58)
for \(1 \leq y \leq n + 1\) and \(0 \leq z \leq n\). The formula is straightforward to confirm by induction. Indeed it holds for \(y = 1\) as can be seen in (4.57) (and recalling that \(0! = 1\) by convention) and, assuming it holds for \(y\), we have by (4.56)
\[
g^\infty_z(y + 1) = \frac{1}{\lambda} \left\{ yg^\infty_z(y) + \pi(z) - \delta_z(y) \right\} \\
= \frac{1}{\lambda} \left\{ y \frac{(y - 1)!}{\lambda^y} \sum_{k=0}^{y-1} \lambda^k \frac{1}{k!} \{\pi(z) - \delta_z(k)\} + \pi(z) - \delta_z(y) \right\} \\
= \frac{y!}{\lambda^{y+1}} \sum_{k=0}^{y-1} \lambda^k \frac{1}{k!} \{\pi(z) - \delta_z(k)\} + \frac{1}{\lambda} \{\pi(z) - \delta_z(y)\} \\
= \frac{y!}{\lambda^{y+1}} \sum_{k=0}^{y} \lambda^k \frac{1}{k!} \{\pi(z) - \delta_z(k)\},
\]
270
as desired.

We re-write (4.58) depending on whether the term \( \delta_z(y) = 1\{ z = y \} \) plays a role in the equation. For \( z \geq y \), the equation simplifies to

\[
g_z^\infty(y) = \frac{(y - 1)! \lambda^k}{\lambda^y} \sum_{k=0}^{y-1} \frac{\lambda^k}{k!} \pi(z)
\]

\[
= \frac{1}{y} \frac{y!}{e^{-\lambda} \lambda^y} \sum_{k=0}^{y-1} \frac{e^{-\lambda} \lambda^k}{k!} \pi(z)
\]

\[
= \Pi(y - 1) \frac{1}{y \pi(y)} \pi(z).
\]

For \( z < y \), we get instead

\[
g_z^\infty(y) = \frac{(y - 1)! \lambda^k}{\lambda^y} \left\{ \sum_{k=0}^{y-1} \frac{\lambda^k}{k!} \pi(z) - \frac{\lambda^z}{z!} \right\}
\]

\[
= \frac{1}{y} \frac{y!}{e^{-\lambda} \lambda^y} \left\{ \sum_{k=0}^{y-1} \frac{e^{-\lambda} \lambda^k}{k!} \pi(z) - \pi(z) \right\}
\]

\[
= \Pi(y - 1) - \frac{1}{y \pi(y)} \pi(z).
\]

That concludes the proof. \( \blacksquare \)

**Proof of Lemma 4.91.** It suffices to prove that, for \( A \subseteq \{0,1,\ldots,n\} \) and \( y \in \{0,1,\ldots,n\} \),

\[ |g_A^\infty(y + 1) - g_A^\infty(y)| \leq (1 - \lambda^{-1}), \]

and then use the triangle inequality.

We use the expression derived in Lemma 4.90. For \( y < z \),

\[
g_z^\infty(y + 1) - g_z^\infty(y) = \frac{\Pi(y)}{(y + 1) \pi(y + 1)} \pi(z) - \frac{\Pi(y - 1)}{y \pi(y)} \pi(z)
\]

\[
= \pi(z) \frac{1}{y \pi(y)} \left\{ \frac{y}{\lambda} \Pi(y) - \Pi(y - 1) \right\},
\]

where we used that \( \pi(y + 1)/\pi(y) = \lambda/(y + 1) \). We show that the expression in
curly brackets is non-negative. Indeed,

\[
y \sum_{k'=0}^{y} \frac{e^{-\lambda \lambda' k'}}{(k')!} - \sum_{k=0}^{y-1} \frac{e^{-\lambda \lambda k}}{k!}
\]

\[
= \frac{y}{\lambda} e^{-\lambda} + \sum_{k=0}^{y-1} \frac{e^{-\lambda \lambda (k+1)-1}}{(k+1)!/y} - \sum_{k=0}^{y-1} \frac{e^{-\lambda \lambda k}}{k!}
\]

\[
\geq \frac{y}{\lambda} e^{-\lambda} + \sum_{k=0}^{y-1} \frac{e^{-\lambda \lambda k}}{k!} - \sum_{k=0}^{y-1} \frac{e^{-\lambda \lambda k}}{k!}
\]

\[
\geq 0.
\]

So \(g_z^\infty(y + 1) - g_z^\infty(y) \geq 0\) for \(0 \leq y < z\). A similar calculation, which we omit, shows that the same inequality holds for \(z < y \leq n\). For \(y = n + 1\), 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,

\[
0 = g_z^\infty(n + 2) - g_z^\infty(0) = \sum_{y=1}^{n+1} \{g_z^\infty(y + 1) - g_z^\infty(y)\}.
\]

We have shown that all terms in the last sum are non-negative, with the exception of the term \(y = z\). Hence, for \(0 \leq y \leq n\), it must be that the maximum of \(|g_z^\infty(y + 1) - g_z^\infty(y)|\) is achieved at \(z = y\). In fact, for \(0 \leq y \leq n\), it must be that 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

\[
|g_y^\infty(y + 1) - g_y^\infty(y)|
= \left| - \frac{1 - \Pi(y)}{(y + 1)\pi(y + 1)} \pi(y) - \frac{\Pi(y - 1)}{y\pi(y)} \pi(y) \right|
= \frac{1}{\lambda} \sum_{k \geq y+1} e^{-\lambda} \frac{\lambda^k}{k!} + \frac{1}{y} \sum_{k=0}^{y-1} e^{-\lambda} \frac{\lambda^k}{k!}
= \frac{e^{-\lambda}}{\lambda} \left\{ \sum_{k'=1}^{y} \frac{\lambda^{k'}}{(k')!} y + \sum_{k \geq y+1} \frac{\lambda^k}{k!} \right\}
\leq \frac{e^{-\lambda}}{\lambda} \left\{ \sum_{k \geq 1} \frac{\lambda^k}{k!} \right\}
= \frac{e^{-\lambda}}{\lambda} \{ e^\lambda - 1 \}
= \frac{1 - e^{-\lambda}}{\lambda}.
\]

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 \( \frac{1 - e^{-\lambda}}{\lambda} \leq \frac{\lambda}{\lambda} = 1 = (1 \wedge \lambda^{-1}) \).

**4.5.3 Erdős-Rényi: subgraph containment at the threshold**

We revisit the subgraph containment problem of Section 2.3.2 (and Section 4.3.4). 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 \( G \), i.e., 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 \( p_n = Cn^{-2/3} \). We use the Chen-Stein method.

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 \( Z_i = 1_{A_i} \). Then \( W = \sum_{i=1}^{m} Z_i \) is the number of 4-cliques in \( G_n \). We argued previously 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 4.85. 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 4.85 are satisfied, that is, $X_i$ is independent of $\{Z_j : j \notin \mathcal{N}_i \cup \{i\}\}$. We argued previously 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 $\mathbb{E}[Z_i, Z_j] = p_n^0$, 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 4.85. Using the formulas above, we get

$$\|\mu - \pi\|_{TV} \leq (1 \wedge \lambda^{-1}) \sum_{i=1}^{n} \left\{ \frac{n^2}{4} + \sum_{j \in \mathcal{N}_i} (q_i q_j + \mathbb{E}[Z_i, Z_j]) \right\}$$

$$\leq (1 \wedge \lambda^{-1}) \binom{n}{4} \left[ p_n^{12} + \left\{ \binom{4}{3}(n - 4)(p_n^{12} + p_n^9) + \binom{4}{2} \binom{n - 4}{2} (p_n^{12} + p_n^{11}) \right\} \right]$$

$$= (1 \wedge \lambda^{-1}) \Theta(n^4 p_n^{12} + n^5 p_n^9 + n^6 p_n^{11})$$

$$= (1 \wedge \lambda^{-1}) \Theta(n^4 n^{-8} + n^5 n^{-6} + n^6 n^{-22/3})$$

$$= (1 \wedge \lambda^{-1}) \Theta(1),$$

which goes to 0 as $n \to +\infty$ when $p_n = Cn^{-2/3}$.

See Exercise 4.16 for an improved bound.

**Exercises**

**Exercise 4.1** (Harris’ inequality: alternative proof). We say that $f : \mathbb{R}^n \to \mathbb{R}$ is nondecreasing if it is nondecreasing in each variable while keeping the other variables fixed.
• (Chebyshev’s association inequality) Let \( f : \mathbb{R} \to \mathbb{R} \) and \( g : \mathbb{R} \to \mathbb{R} \) be nondecreasing and let \( X \) be a real random variable. Show that
\[
E[f(X)g(X)] \geq E[f(X)]E[g(X)].
\]
[Hint: Consider the quantity \((f(X) - f(X'))(g(X) - g(X'))\) where \( X' \) is an independent copy of \( X \).

• (Harris’ inequality) Let \( f : \mathbb{R}^n \to \mathbb{R} \) and \( g : \mathbb{R}^n \to \mathbb{R} \) be nondecreasing and let \( X = (X_1, \ldots, X_n) \) be independent real random variables. Show by induction on \( n \) that
\[
E[f(X)g(X)] \geq E[f(X)]E[g(X)].
\]

Exercise 4.2. Provide the details for Example 4.51.

Exercise 4.3 (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 4.50).

a) 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.]

b) To check the FKG condition, show that it suffices to check (4.12) 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 a].

Exercise 4.4 (FKG and strong positive association). 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
\[
\mathcal{X}_\Lambda^\xi := \{\omega_\Lambda \times \xi_{\Lambda^c} : \omega_\Lambda \in \{0, 1\}^\Lambda\},
\]
where \( \omega_\Lambda \times \xi_{\Lambda^c} \) agrees with \( \omega \) on coordinates in \( \Lambda \) and with \( \xi \) on coordinates in \( F \setminus \Lambda \). Define the measure \( \mu_\Lambda^\xi \) over \( \{0, 1\}^\Lambda \) as
\[
\mu_\Lambda^\xi(\omega_\Lambda) := \frac{\mu(\omega_\Lambda \times \xi_{\Lambda^c})}{\mu(\mathcal{X}_\Lambda^\xi)}.
\]
That is, \( \mu_\Lambda^\xi \) is \( \mu \) conditioned on agreeing with \( \xi \) on \( F \setminus \Lambda \). The measure \( \mu \) is said to be strongly positively associated if \( \mu_\Lambda^\xi(\omega_\Lambda) \) is positively associated for all \( \Lambda \) and \( \xi \). Prove that the FKG condition is equivalent to strong positive association. [Hint: Use Exercise 4.3 as well as the FKG inequality.]
Exercise 4.5 (Triangle-freeness: a second proof). Consider again the setting of Section 4.3.4.

a) 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 $e_t > t$.

b) Use Exercise 2.15 to give a second proof of the fact that $P[X_n = 0] \to e^{-\lambda^3/6}$.

Exercise 4.6 (RSW lemma: general $\alpha$). Let $R_n,\alpha(p)$ be as defined in Section 4.3.5. 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 4.7 (Primal and dual crossings). Modify the proof of Lemma 2.17 to prove Lemma 4.59.

Exercise 4.8 (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 4.9 (Splitting: details). Show that $\tilde{P}$, as defined in Example 4.67, is a transition matrix on $V$ provided $z_0$ satisfies the condition there.

Exercise 4.10 (Doeblin’s condition in finite case). 1. Show that Doeblin’s condition holds when $P$ is finite, irreducible and aperiodic.

2. 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 4.11 (Mixing on cycles: lower bound). Let $(Z_t)$ be lazy, simple random walk on the cycle of size $n$, $Z_n := \{0, 1, \ldots, n - 1\}$, where $i \sim j$ if $|j - i| = 1$ (mod $n$).

a) Let $A = \{n/2, \ldots, n - 1\}$. By coupling $(Z_t)$ with lazy, simple random walk on $Z$, show that

$$P^{\alpha n^2}(n/4, A) < \frac{1}{2} - \varepsilon,$$

for $\alpha \leq \alpha_\varepsilon$ for some $\alpha_\varepsilon > 0$. [Hint: Use Kolmogorov’s maximal inequality (e.g. [Dur10, Theorem 2.5.2]).]
b) Deduce that
\[ \tau_{\text{mix}}(\varepsilon) \geq \alpha \varepsilon n^2. \]

**Exercise 4.12** (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 \( \sigma^2 := \max\{\text{Var}[h(X)], \text{Var}[h(Y)]\} \). Show that

\[ \|\mu_X - \mu_Y\|_{TV} \geq 1 - \frac{8}{r^2}. \]

[Hint: Consider the interval on one side of the midpoint between \( \mathbb{E}[h(X)] \) and \( \mathbb{E}[h(Y)] \).]

**Exercise 4.13** (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 transportation metric or Wasserstein distance between \( \mu \) and \( \nu \).

a) Show that \( W_0 \) is a metric. [Hint: Proof of Claim 4.75.]

b) Assume that the conditions of Theorem 4.74 hold. Show that for any probability measures \( \mu, \nu \)

\[ W_0(\mu P, \nu P) \leq \kappa W_0(\mu, \nu). \]

c) Use a) and b) to prove Theorem 4.74.

**Exercise 4.14** (Stein equation for the Poisson distribution). Let \( \lambda > 0 \). Show that a non-negative integer-valued random variable \( Z \) is \( \text{Poi}(\lambda) \) if and only if for all \( g \) bounded

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

**Exercise 4.15** (Chen-Stein for positively related variables). Using the notation in (4.39), (4.40) and (4.41), suppose that for each \( i \) we can construct a coupling \( \{(X_i : i = 1, \ldots, n), (Y^{(i)}_j : j \neq i)\} \) such that

\( (Y^{(i)}_j, j \neq i) \sim (X_j, j \neq i) | X_i = 1 \) and \( Y^{(i)}_j \geq X_j, \forall j \neq i \).

Show that

\[ \|\mu - \pi\|_{TV} \leq (1 \wedge \lambda^{-1}) \left\{ \text{Var}(W) - \lambda + 2 \sum_{i=1}^{n} \tilde{p}_i^2 \right\}. \]
Exercise 4.16 (Chen-Stein and 4-cliques). Use Exercise 4.15 to give an improved asymptotic bound in the setting of Section 4.5.3.

Exercise 4.17 (Chen-Stein for negatively related variables). Using the notation in (4.39), (4.40) and (4.41), suppose that for each $i$ we can construct a coupling $\{(X_i : i = 1, \ldots, n), (Y_j^{(i)} : j \neq i)\}$ such that

\[
(Y_j^{(i)}, j \neq i) \sim (X_j, j \neq i) | X_i = 1 \quad \text{and} \quad Y_j^{(i)} \leq X_j, \forall j \neq i.
\]

Show that

\[
\|\mu - \pi\|_{TV} \leq (1 + \lambda^{-1}) \{\lambda - \text{Var}(W)\}.
\]

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 4.2.2 is based on [vdH14, Section 5.3] and Section 4.1.2 is based on [Per, Section 6].

Section 4.3 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 e.g. [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 e.g. [RAS]. Section 4.3.4 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 4.3.5 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-rigourous work of Sykes and Essam [SE64]. The RSW theorem was obtained independently to 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 theorem for critical site percolation on the triangular lattice was given by Smirnov. See e.g. [Ste]. The type of “scale invariance” seen in the RSW theorem plays a key role in the contemporary theory of critical two-dimensional percolation and of two-dimensional lattice models more generally. See e.g. [Law05, Gri10a].
Section 4.4  The material in Section 4.4 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 4.67 is based on [Str14] and [JH01]. For a more general treatment, see [MT09, Chapter 16]. The proof of Claim 4.71 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 4.13 is from [LPW06, Chapter 14]. For more on optimal transport, see e.g. [Vil09]. The main result in Section 4.4.4 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]. The connection between sampling and counting was first considered by Jerrum, L. Valiant and V. Vazirani [JVV86]. For more on this topic, see e.g. [Sin93].

Section 4.5  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 4.5 is inspired by [Dey] and [vH16]. Example 4.86 is from [AGG89]. Further applications of the Chen-Stein and Stein methods to random graphs can be found in [JLR11, Chapter 6].