MONOTONE COUPLING AND THE ISING MODEL

1. Perfect Matching in Bipartite Graphs

Definition 1. A *bipartite graph* is a graph $G = (V, \mathcal{E})$ whose vertex set V can be partitioned into two disjoint set V_I, V_O in such a way that every edge $e \in \mathcal{E}$ has one endpoint in V_I and one endpoint in V_O. The sets $V_I$and V_O in this partition will be referred to as the *input set* and the *output set*, respectively.

Definition 2. A *matching* (also called a *dimer covering*) in a bipartite graph G is an injective mapping $f : V_I \to V_O$ such that for every $x \in V_I$ there is an edge $e \in \mathcal{E}$ with endpoints x and $f(x)$.

Given a bipartite graph G, how can one tell if there is a perfect matching? There is an obvious *necessary* condition: it must be the case that for every subset $A \subset V_I$ there must be at least $|A|$ vertices $v \in V_O$ such that there is at least one edge connecting v to V_I. *Hall's matching theorem* asserts that this condition is also *sufficient*.

Theorem 1. *(Hall's Matching Theorem) Let G be a bipartite graph with input set V_I, output set V_O, and edge set $\mathcal{E}$. For any subset $A \subset V_I$, define ∂A to be the set of all vertices $y \in V_O$ that are endpoints of edges with one endpoint in A. Then there exists a matching $f : V_I \to V_O$ if and only if for every nonempty subset $A \subset V_I$,*

$$|\partial A| \geq |A|. \tag{1.1}$$

Proof. By induction on the cardinality of V_I. If $|V_I| = 1$ the result is trivially true. Suppose, then, that the result is true for all bipartite graphs such that $|V_I| \leq n$, and consider a bipartite graph G whose input set V_I has cardinality $n + 1$. There are two possibilities: either

Case 1: $|\partial A| \geq |A| + 1$ for every nonempty proper subset $A \subset V_I$; or
Case 2: $|\partial A| = |A|$ for some nonempty proper subset $A \subset V_I$.

Case 1: Choose any $x \in V_I$ and any $y \in \partial\{x\}$ (by hypothesis, $\partial\{x\}$ has at least two elements). Let G^* be the bipartite graph with input set $V_I^* = V_I - \{x\}$, output set $V_O^* = V_O - \{y\}$, and whose edges are the same as those of G, but with edges incident to either x or y deleted. The bipartite graph G^* satisfies the hypothesis (1.1), because in Case 1 every proper subset $A \subset V_I$ has $|\partial A| \geq |A| + 1$, so deleting the single vertex y from ∂A still leaves at least $|A|$ vertices. By the induction hypothesis, there is a perfect matching in G^*; this perfect matching extends to a perfect matching in the original graph G by setting $f(x) = y$.

Case 2: Let $A \subset V_I$ be a nonempty, proper subset of V_I such that $|\partial A| = |A|$. Clearly, for any subset $C \subset A$ we have $\partial C \subset \partial A$, and by the hypothesis of the theorem, $|\partial C| \geq |C|$. Since A has no more than n elements, it follows by the induction hypothesis that there is a perfect matching $f_A : A \to \partial A$.

Now let $B = V_I \setminus A$ be the set of all vertices in the input set V_I that are not elements of A. For every nonempty subset $D \subset B$, $\partial A \cup \partial D = \partial(A \cup D)$, by definition of the operator ∂. Hence, it must be the case that ∂D contains at least $|D|$ elements not in ∂A, because

otherwise $\partial(A \cup D)$ would have fewer than $|A \cup D|$ elements, contradicting the hypothesis of the theorem. Consequently, the hypotheses of the theorem are satisfied by the smaller bipartite graph G^* consisting of input set $V_I^* = B$, output set $V_O^* = \partial B \setminus \partial A$, and edges connecting all pairs of vertices in $V_I^* \times V_O^*$ that were connected in the original graph G. Consequently, by the induction hypotheses, there is a perfect matching $f_B : B \to \partial B$.

Finally, the perfect matchings f_A and f_B can be joined in the obvious manner to give a perfect matching in the original graph G.

□

2. The Birkhoff-von Neumann Theorem

A *doubly stochastic* matrix is a square matrix with nonnegative entries whose row sums and column sums are all 1. A *magic square* is a square matrix with nonnegative *integer* entries whose row sums and column sums are all equal; the common value of the row sums and column sums is called the *weight* of the square. Observe that if T is a magic square of weight $d \geq 1$, then one obtains a doubly stochastic matrix by dividing all entries of T by d. Conversely, if $\mathbf{P}$ is a doubly stochastic matrix with rational entries, then one may obtain a magic square by multiplying all entries by their least common denominator. The magic squares of weight 1 are called *permutation matrices*: for any $m \times m$ permutation matrix T, there exists a permutation σ of the set $[m]$ such that

$$\begin{aligned} T_{i,\sigma(i)} &= 1 \qquad \text{for all } i \in [m], \text{and} \\ T_{i,j} &= 0 \qquad \text{if } j \neq \sigma(i). \end{aligned} \tag{2.1}$$

Theorem 2. *Every doubly stochastic matrix is a convex combination (weighted average) of permutation matrices. Every magic square of weight d is the sum of d (not necessarily distinct) permutation matrices.*

Proof. We shall consider only the assertion about magic squares; the assertion about doubly stochastic matrices may be proved by similar arguments.) By definition, every magic square of weight 1 is a permutation matrix. Let T be an $m \times m$ magic square of weight $d > 1$. Consider the bipartite graph with $V_I = V_O = [m]$ such that, for any pair $(i, j) \in V_I \times V_O$, there is an edge from i to j if and only if $T_{i,j} > 0$.

Claim: The hypothesis (1.1) of the Matching Theorem is satisfied.

Proof. Let B be a subset of V_I with $r \leq m$ elements. Since T is a magic square of weight d, the sum of all entries $T_{i,j}$ such that $i \in B$ must be rd. The positive entries among these must all lie in the columns indexed by elements of ∂B; consequently, the sum of the entries $T_{i,j}$ such that $j \in \partial B$ must be at least rd. But this sum cannot exceed $d|\partial B|$, since the column sums of T are all d. □

The Matching Theorem now implies that there is a perfect matching in the bipartite graph. Since $V_I = V_O = [m]$, this perfect matching must be a permutation σ of the set $[m]$. By construction, the permutation matrix T^σ defined by equations (2.1) is dominated (entry by entry) by the magic square T, so the difference $T - T^\sigma$ is a magic square of weight $d - 1$. Thus, the assertion follows by induction on d. □

3. STRASSEN'S MONOTONE COUPLING THEOREM

A *poset* is a partially ordered set $(\mathcal{X}, \leq)$. Recall that a partial order $\leq$ must satisfy the following properties: for all $x, y, z \in \mathcal{X}$,

$$x \leq x; \tag{3.1}$$

$$x \leq y \ \& \ y \leq x \Longrightarrow x = y; \tag{3.2}$$

$$x \leq y \ \& \ y \leq z \Longrightarrow x \leq z. \tag{3.3}$$

Posets occur frequently as state spaces in graphical models. An important special case is the *configuration space* $\Sigma = \{0,1\}^V$ of a spin system: here V is a set of *sites*, often the vertices of a lattice, and the elements of Σ are assignments of zeros and ones ("spins") to the sites ("configurations"). The partial order $\leq$ is defined as follows:

$$x \leq y \qquad \text{iff} \qquad x_s \leq y_s \ \forall \, s \in V.$$

An *ideal*[1] of a poset $(\mathcal{X}, \leq)$ is a subset $\mathcal{J} \subset \mathcal{X}$ with the property that if $x \in \mathcal{J}$ and $x \leq y$ then $y \in \mathcal{J}$. If μ and ν are two probability distributions on $\mathcal{X}$, say that ν *stochastically dominates* μ (and write $\mu \leq_D \nu$) if for every ideal $\mathcal{J}$,

$$\mu(\mathcal{J}) \leq \nu(\mathcal{J}). \tag{3.4}$$

Theorem 3. *(Strassen) Let $(\mathcal{X}, \leq)$ be a* finite *poset, and let μ, ν be probability distributions on $\mathcal{X}$. If $\mu \leq_D \nu$ then on some probability space (in fact, on any probability space supporting a random variable uniformly distributed on the unit interval) are defined $\mathcal{X}-$valued random variables M, N with distributions μ, ν, resepectively, such that*

$$M \leq N. \tag{3.5}$$

Equivalently, if $\mu \leq_D \nu$ then there is a probability distribution λ on $\mathcal{X} \times \mathcal{X}$ satisfying

(a) $\lambda(x,y) = 0$ *unless* $x \leq y$,
(b) $\sum_y \lambda(x,y) = \mu(x)$ *for all* $x \in \mathcal{X}$, *and*
(c) $\sum_x \lambda(x,y) = \nu(y)$ *for all* $y \in \mathcal{X}$.

Proof. First, observe that the two statements in the theorem are equivalent, because (i) if M, N exist, then the joint distribution

$$\lambda(x,y) := P\{M = x \ \text{ and } \ N = y\}$$

satisfies (a)- (c), and (ii) conversely, if λ exists, then the coordinate variables M, N on $\mathcal{X} \times \mathcal{X}$ will satisfy (3.5) and have marginals μ, ν, respectively. Thus, it suffices to prove the first statement. We shall only consider the case where the probability distributions μ, ν assign rational probabilities k/N (with a common denominator N) to the elements of the poset $\mathcal{X}$. The general case may be deduced from this by an approximation argument, which the reader will supply (EXERCISE!).

Case A: For every $x \in \mathcal{X}$, the probabilities $\mu(x)$ and $\nu(x)$ are either 0 or $1/N$.

Consider the bipartite graph with $V_I = \{x \in \mathcal{X} : \mu(x) = 1/N\}$ and $V_I = \{x \in \mathcal{X} : \nu(x) = 1/N\}$, where $x \in V_I$ and $y \in V_O$ are connected by an edge if and only if $x \leq y$. The hypothesis that μ is stochastically dominated by ν implies that the hypothesis (1.1) of the Matching Theorem is satisfied. Consequently, there is a perfect matching $f : V_I \to V_O$. Let M be an $\mathcal{X}-$valued random variable M with distribution μ (such a random variable

[1] sometimes called an *upper corner*

will exist on any probability space supporting a uniform-[0,1] random variable). Define $N = f(M)$. Then the pair (M, N) satisfies $M \leq N$, and the marginal distributions of M and N are μ and ν, as the reader will easily check.

Case B: For every $x \in \mathcal{X}$, the probabilities $\mu(x)$ and $\nu(x)$ are integer multiples of $1/N$.

For each $x \in \mathcal{X}$, if $\mu(x) = k/N$ then construct k "copies" $x_1, x_2, \ldots, x_k$ of x, and for each such copy set $\pi_I(x_i) = x$. Define V_I to be the set of all such copies, where x ranges over $\mathcal{X}$. Similarly, for each $y \in \mathcal{X}$, if $\nu(y) = m/N$ then construct m copies $y_1, y_2, \ldots, y_m$ of y, and for each such copy set $\pi_O(y_i) = y$. Define V_O to be the set of all such copies, where y ranges over $\mathcal{X}$. For each pair $x_i \in V_I$ and $y_j \in V_O$, put an edge from x_i to y_j if and only if $\pi_I(x_i) \leq \pi_O(y_j)$. Once again, it is easily verified that hypothesis (1.1) of the Matching Theorem is satisfied, since $\mu \leq \nu$. Consequently, there is a perfect matching $f : V_I \to V_O$.

Let U be a random variable that is uniformly distributed on V_I – such a random variable exists on any probability space supporting a Uniform-[0,1] random variable. Set $M = \pi_I(U)$ and $N = \pi_O(f(U))$; then the marginal distributions of M and N are μ and ν, respectively, and $M \leq N$, by construction. □

Problem 1. Prove that the general case of Strassen's theorem is implied by the rational case (the case where both probability distributions attach only rational probabilities to points of the poset).

HINTS: (A) Use the Bolzano-Weierstrass theorem to show that any sequence of probability measures on a finite set $\mathcal{X}$ has a convergent subsequence. (B) Show that if F, G are two probability measures on a finite poset $(\mathcal{X}, \leq)$ such that F is stochastically dominated by G then there are sequences of *rational* probability measures F_n, G_n on $\mathcal{X}$ such that

- $F_n \to F$ and $G_n \to G$; and
- $F_n \overset{D}{\leq} G_n$ for each n.

For this you might find it helpful to first consider the special case where the poset has a unique maximal element and a unique minimal element.

Corollary 4. *Let μ, ν be probability distributions on a finite poset $(\mathcal{X}, \leq)$ such that ν stochastically dominates μ. Then there is a Markov kernel (i.e., a transition probability matrix) $q(x, y)$ on $\mathcal{X}$ such that*

(A) $q(x, y) > 0$ only if $x \leq y$, and
(B) $\sum_{x \in \mathcal{X}} \mu(x) q(x, y) = \nu(x)$ for all $x \in \mathcal{X}$.

Proof. Assume for simplicity that $\mu(x) > 0$ for every x. Let λ be a probability distribution on $\mathcal{X} \times \mathcal{X}$ satisfying the conclusions (a), (b), (c) of Strassen's theorem. Then the transition probability matrix

$$q(x, y) = \lambda(x, y)/\mu(x)$$

satisfies the conclusions (A)– (B). □

4. CONSEQUENCES OF STRASSEN'S THEOREM

4.1. **Stochastic ordering of Markov kernels.**

Definition 3. Let $(\mathcal{X}, \leq)$ be a finite or countable poset, and let $p(x, y)$ and $q(x, y)$ be Markov kernels (i.e., stochastic matrices) on $\mathcal{X}$. The kernel $q(x, y)$ is said to stochastically dominate

the kernel $p(x,y)$ if for every pair of states $x, x' \in \mathcal{X}$ such that $x \leq x'$ the probability distribution $q(x',\cdot)$ stochastically dominates the distribution $p(x,\cdot)$.

Proposition 5. *Suppose that $p(x,y)$ and $q(x,y)$ are Markov kernels on a finite poset $(\mathcal{X},\leq)$ such that q stochastically dominates p. Then for any every pair of states $x, x' \in \mathcal{X}$ such that $x \leq x'$ there are Markov chains X_n and Y_n on $\mathcal{X}$ with transition probabilities $p(x,y)$ and $q(x,y)$, respectively, and initial states $X_0 = x$ and $Y_0 = x'$, such that*

$$X_n \leq Y_n \quad \textit{for all} \;\; n \geq 0$$

Furthermore, if $p(x,y)$ and $q(x,y)$ are irreducible and have stationary distributions π and ν, respectively, then ν stochastically dominates π.

Proof. Let $(\mathcal{X}\times\mathcal{X})_{\leq}$ be the subset of $\mathcal{X}\times\mathcal{X}$ consisting of all pairs (y,y') such that $y \leq y'$. By hypothesis, for each element $(y,y') \in (\mathcal{X}\times\mathcal{X})_{\leq}$ the probability distribution $p(y,\cdot)$ is stochastically dominated by $q(y',\cdot)$. Hence, by Strassen's theorem, there exists a probability distribution $R((y,y'),\cdot)$ on $(\mathcal{X}\times\mathcal{X})_{\leq}$ with marginals $p(y,\cdot)$ and $q(y',\cdot)$, that is,

$$\begin{aligned} \sum_{z':(z,z')\in(\mathcal{X}\times\mathcal{X})_{\leq}} R((y,y'),(z,z')) &= p(y,z) \quad \text{and} \\ \sum_{z:(z,z')\in(\mathcal{X}\times\mathcal{X})_{\leq}} R((y,y'),(z,z')) &= q(y',z') \end{aligned} \tag{4.1}$$

The matrix R on $(\mathcal{X}\times\mathcal{X})_{\leq}$ is stochastic, so for any initial point (x,x') we can build a Markov chain (X_n,Y_n) on $(\mathcal{X}\times\mathcal{X})_{\leq}$ with initial state (x,x') and transition probability matrix R (using an i.i.d. sequence of uniform-[0,1] random variables to make the transitions). By construction, the component sequences X_n, Y_n must satisfy $X_n \leq Y_n$, and by (4.1) the sequences X_n and Y_n will be Markov chains with transition probability matrices $p(x,y)$ and $q(x,y)$, respectively. This proves the first part of the proposition.

Suppose next that the transition probability matrices $p(x,y)$ and $q(x,y)$ are irreducible, so that the Markov chains X_n and Y_n have unique stationary distributions π and ν. Suppose also that both $p(x,y)$ and $q(x,y)$ are aperiodic. Then by Kolmogorov's convergence theorem for Markov chains, for any initial states $x \leq x'$,

$$\begin{aligned} X_n &\xrightarrow{\mathcal{D}} \pi \quad \text{and} \\ Y_n &\xrightarrow{\mathcal{D}} \nu. \end{aligned}$$

Since $X_n \leq Y_n$ for every n, it follows that π is stochastic dominated by ν (why?). □

EXERCISE: Show that the assumption that the Markov chains are aperiodic in the last argument is unnecessary.

4.2. **Stochastically monotone Markov kernels.**

Definition 4. Let $(\mathcal{X},\leq)$ be a finite or countable poset, and let $p(x,y)$ be a Markov kernel on $\mathcal{X}$. The kernel $p(x,y)$ is said to be *stochastically monotone* if for every pair of states $x, x' \in \mathcal{X}$ such that $x \leq x'$ the probability distribution $p(x',\cdot)$ stochastically dominates the probability distribution $p(x,\cdot)$.

Corollary 6. *If $p(x,y)$ is a stochastically monotone Markov kernel on a finite poset $(\mathcal{X},\leq)$ then for any two states x, x' such that $x \leq x'$ there exist, on some probability space, Markov chains X_n*

and X'_n, both with transition probability matrix $p(\cdot,\cdot)$, such that $X_0 = x$, $X'_0 = x'$, and $X_n \le X'_n$ for every n.

Proof. If $p(x,y)$ is stochastically monotone, then it stochastically dominates itself in the sense of Definition 3. Consequently, the corollary follows from Proposition 5. □

Corollary 7. *If $\mu_0, \mu_1, \ldots$ is a finite or infinite sequence of probability measures on a finite poset $\mathcal{X}$ such that μ_n is stochastically dominated by μ_{n+1} for each n, then on some probability space there are $\mathcal{X}-$valued random variables $X_0, X_1, \ldots$ such that $X_n \sim \mu_n$ and $X_n \le X_{n+1}$ for all $n \ge 0$.*

Proof. By the previous corollary, there are Markov kernels $q_n(x,y)$ such that $q_n(x,y) > 0$ only if $y \ge x$, and such that for each n and $x, y \in \mathcal{X}$,

$$\sum_x \mu_n(x) q_n(x,y) = \mu_{n+1}(y).$$

These Markov kernels can be used to build a time-inhomogeneous Markov chain X_n with $X_0 \sim \mu_0$ and steps $X_n \mapsto X_{n+1}$ made according to $q_n(X_n, \cdot)$: this can be done on any probability space with an infinite sequence U_n of i.i.d. uniform-[0,1] random variables. □

4.3. **Correlation inequalities.**

Definition 5. A probability distribution on a poset $\mathcal{X}$ is said to have *positive correlations* if for any two nondecreasing, nonnegative functions $f, g : \mathcal{X} \to \mathbf{R}_+$,

$$E_\mu fg \ge E_\mu f E_\mu g$$

Problem 2. (A) Show that if $\mathcal{X} = \mathbf{R}$ with the usual order $\le$ then every probability measure has positive correlations. NOTE: This is roughly equivalent to a famous theorem of mathematical statistics that is rarely covered in modern statistics classes (sadly).

(B) Show that if $\mathcal{X} = \mathbf{R}^d$ with the usual (coordinatewise) partial order $\le$ then every *product* probability measure has positive correlations. A *product probability measure* is defined to be the joint distribution of independent random variables $X_1, X_2, \ldots, X_d$. HINT: Induct on d, and begin the induction step by conditioning on (say) the last coordinate. If you aren't yet comfortable with conditional expectation in general, you may assume that the probability measure is discrete.

(C) Suppose that μ is a probability distribution on a poset $\mathcal{X}$ with positive correlations, and let $f : \mathcal{X} \to \mathbf{R}$ be a nondecreasing, nonnegative function such that $E_\mu f = 1$. Define ν to be the probability distribution on $\mathcal{X}$ such that ν has likelihood ratio f relative to μ. Prove that ν stochastically dominates μ.

(more later)

5. THE ISING MODEL

The Ising model is a crude but extremely important mathematical model of a ferromagnetic metal introduced by Ising about 70 years ago. Its importance stems from the fact that it is the one of the simplest mathematical models to exhibit a *phase transition*: at high temperature, there is a unique equilibrium state for the system, but at temperatures below a certain critical temperature, there are several distinct equilibrium states. This corresponds to the physical phenomenon of spontaneous magnetization: If unmagnetized iron is cooled to a very low temperature, it will magnetize; and if a magnet is heated to a sufficiently high

temperature, it will demagnetize. The latter may be verified easily by experiment, using only a floppy disk and a household stove.

In the following sections the basic theory of the Ising model and an Markov chain, the *Gibbs sampler*, that can be used to simulate it.

5.1. **Gibbs States.** Let $\mathcal{X}$ be a finite set and $H : \mathcal{X} \to \mathbb{R}$ a function, called the *Hamiltonian* of the system. In physical applications $H(x)$ represents the *energy* of the system when it is in state x. The *Gibbs state* $\mu = \mu_\beta$ for inverse temperature β is the probability measure on $\mathcal{X}$ defined by

$$\mu_\beta(x) = e^{-\beta H(x)}/Z(\beta), \qquad \text{where} \tag{5.1}$$

$$Z(\beta) = \sum_{x\in\mathcal{X}} e^{-\beta H(x)}. \tag{5.2}$$

The normalizing constant $Z(\beta)$ is called the *partition function*. The family $\{\mu_\beta\}_{\beta>0}$ is a one-paramter *exponential family* of probability measures on $\mathcal{X}$, with $H(x)$ playing the role of the *natural parameter* and $\log Z(\beta)$ the role of the cumulant generating function. Observe that, since the sum in (5.2) is finite, the partition function is well-defined and (real-)analytic in the domain $\beta > 0$.

5.2. **The Ising Hamiltonian.** In condensed-matter physics, field theory, and various other parts of statistical physics, the state space $\mathcal{X}$ is often of the form

$$\mathcal{X} = A^\Lambda \tag{5.3}$$

where Λ is a set of *sites* (which we will also call *vertices*) and A is a finite set. Elements of Λ usually represent spatial locations, and elements of A may represent atomic elements (in models of alloys), presence (+1) or absence (0) of particles (in models of gases), *spins* (in models of magnetism and in quantum field theory), and so on. In the Ising model, A is the two-element set $A = \{\pm 1\}$, and Λ is the set of vertices of a graph G; the most interesting case (from the standpoint of the physicist) is where Λ is a subset of the $d-$dimensional integer lattice $\mathbb{Z}^d$. The *Ising Hamiltonian* is defined as follows: for any configuration $x \in \mathcal{X} := \{-1, +1\}^\Lambda$

$$H(x) = H_\Lambda(x) = J \sum_{\substack{i,j\in\Lambda:\\ i\sim j}} x_i x_j \tag{5.4}$$

Here $i \sim j$ means that vertices i, j are *nearest neighbors*, that is, there is an edge of the graph G connecting i and j; each edge counts only once in the sum. The constant J is called the *coupling constant*: if $J < 0$ the model is called *ferromagnetic*, and if $J > 0$ it is *anti-ferromagnetic*. Unless otherwise specified, it is henceforth assumed that $J = -1$. Observe that in this case, the system "prefers" configurations in which neighboring spins are aligned, as these have lower energy. The degree to which this is true depends, of course, on the inverse temperature β – for larger values of β the preference for low-energy states is stronger.

Let's focus on the case where Λ is the set of integer lattice points in a square or cube centered at the origin. The set of lattice points on the outer boundary of Λ (but not included in Λ) will be denoted $\partial\Lambda$. For physical reasons it is sometimes natural to fix the spins on the boundary $\partial\Lambda$ and to include the interactions between these boundary spins with the

neighboring interior spins in the Hamiltonian, so for any configuration ξ of spins ξ_i at boundary vertices $i \in \partial\Lambda$, let's write

$$H_{\Lambda,\xi}(x) = -\sum_{i,j\in\Lambda\,:i\sim j} x_i x_j + -\sum_{i\in\Lambda, j\in\partial\Lambda\,:i\sim j} x_i\xi_j.$$

When the boundary condition is $\xi_v = +1$ at all $v \in \partial\Lambda$, we will write H^+_Λ instead of $H_{\Lambda,\xi}$, and similarly for the boundary condition $\xi_v = -1$ at all $v \in$
$partial\Lambda$. The associated Gibbs states are the probability measures

$$\mu^\xi_{\beta,\Lambda}(x) = \exp\{-\beta H_{\Lambda,\xi}(x)\}/Z_{\Lambda,\xi}(\beta).$$

When ξ is the boundary configuration with all spins $+1$, we will write $\mu^+_{\beta,\Lambda}$ for $\mu^\xi_{\beta,\Lambda}$; similarly for the boundary configuration with all spins -1.

Important: Take note of the fact that the Ising Hamiltonian is invariant under spin reversal, that is, $H_{\Lambda,\xi}(x) = H_{\Lambda,-\xi}(-x)$. Consequently, the joint distribution of the spin configuration X under $\mu^+_{\beta,\Lambda}$ is the same as the joint distribution of the reversed spin configuration $-X$ under $\mu^-_{\beta,\Lambda}$.

Stochastic monotonicity plays an important role in the study of the Ising model. The relevant partial order is the coordinatewise order: in particular, for any configurations x, y we have $x \le y$ if and only if y can be obtained from x by flipping some of the $-$ spins in x to $+$, but leaving the $+$ spins in x alone.

Proposition 8. *For each $\beta > 0$, the probability measure $\mu^+_{\beta,\Lambda}$ stochastically dominates $\mu^-_{\beta,\Lambda}$.*

This will be proved in section 6 below, using the machinery developed in section 4.

The Gibbs states for the Ising model have an important property known in the study of random fields as the *Markov property*. Let $\Lambda \subset \Lambda_*$ be two finite sets of sites. For any spin configuration $X = (X_v)_{v\in\Lambda_*}$ and any subset $F \subset L_*$ denote by X_F the restriction of the spin configuration to F, that is,

$$X_F = (X_v)_{v\in F}.$$

Proposition 9. *For any value of $\beta > 0$, the conditional distribution under μ^+_{β,Λ_*} of X_Λ given the event that all of the spins in $\Lambda_* \setminus \Lambda$ are $+1$ is the same as the unconditional distribution of X_Λ under $\mu^+_{\beta,\Lambda}$. Explicitly, for any fixed configuration $\sigma_\Lambda = (\sigma_v)_{v\in\Lambda}$ of spins in Λ,*

$$\mu^+_{\beta,\Lambda_*}(X_\Lambda = \sigma_\Lambda \,|\, X_v = +1\ \forall\ v \in \Lambda_* \setminus \Lambda) = \mu^+_{\beta,\Lambda}(X_\Lambda = \sigma_\Lambda).$$

Proof. This is a routine calculation, using the definition (5.1) of a Gibbs state and the formula (5.4) for the Ising Hamiltonian. Denote by $(+1)_{\Lambda_*\setminus\Lambda}$ the spin configuration on $\Lambda_* \setminus \Lambda$ with all spins $+1$, and by $\sigma_\Lambda \vee 1$ the configuration on Λ_* with spins σ_v for $v \in \Lambda$ and spins $+1$ outside Λ. For notational ease, let's drop the subscript β on the Gibbs measures, since β is fixed throughout the argument. Then

$$\begin{aligned}
\mu^+_{\Lambda_*}(X_\Lambda = \sigma_\Lambda \,|\, X_{\Lambda_*\setminus\Lambda} = (+1)_{\Lambda_*\setminus\Lambda}) &= \frac{\exp\{-\beta H^+_{\Lambda_*}(\sigma_\Lambda \vee 1)\}}{\sum_{\sigma'\in\{+1,-1\}^\Lambda}\exp\{-\beta H^+_{\Lambda_*}(\sigma'_\Lambda \vee 1)\}} \\
&= \frac{\exp\{-\beta H^+_\Lambda(\sigma_\Lambda)\}}{\sum_{\sigma'\in\{+1,-1\}^\Lambda}\exp\{-\beta H^+_\Lambda(\sigma'_\Lambda)\}} \\
&= \mu^+_\Lambda(X_\Lambda = \sigma_\Lambda).
\end{aligned}$$

The key step is the second equality: this holds because all of the factors corresponding to edges of the lattices not in Λ cancel in the numerator and denominator. □

5.3. **Phase Transition in Dimension 2.** The observable physical phenomenon of spontaneous magnetization (and demagnetization) has a mathematical analogue in the Ising model in dimensions two and higher, a fact discovered by R. PEIERLS in the 1930s, some years after Ising introduced his model. [2] Let $G = (\mathbb{Z}^2, \mathcal{E})$ be the standard two-dimensional lattice, and let Λ_n be the square of side $2n+1$ centered at the origin o. Denote by $\mu_n^+ = \mu_{\beta,\Lambda_n}^+$ and $\mu_n^- = \mu_{\beta,\Lambda_n}^-$ the Gibbs states with external boundary conditions z^+ and z^- on the square Λ_n.

Theorem 10. *There exists β_c satisfying $0 < \beta_c < \infty$ such that the following is true: For each vertex $i \in \mathbb{Z}^2$*

$$\lim_{n\to\infty} \mu_n^+\{X_i = +1\} > 1/2 \qquad \text{if } \beta > \beta_c \tag{5.5}$$

$$\lim_{n\to\infty} \mu_n^+\{X_i = +1\} = 1/2 \qquad \text{if } \beta \le \beta_c. \tag{5.6}$$

The fact that $\beta_c < \infty$ is, in essence, Peierls discovery. I do not know who first proved that $\beta_c > 0$. Peierls' argument will be presented in section 7 below.

6. GLAUBER DYNAMICS – THE GIBBS SAMPLER

6.1. **Gibbs Sampler.** In the early 1960s the physicist Glauber proposed a Markov chain that in principle could be used for Markov chain Monte Carlo studies of the Gibbs distributions $\mu_{\beta,\Lambda}^{\pm}$. Subsequently, various people (notably Swendsen and Wang) have discovered much more effective algorithms for simulation, so Glauber dynamics isn't used for MCMC purposes any more. Nevertheless, Glauber's Markov chain provides some useful theoretical information about the Gibbs distributions. Furthermore, the rate of convergence of this Markov chain remains an important area of study.

Glauber dynamics is an instance of what is now called the *Gibbs sampler*, a general strategy for designing a Markov chain for MCMC studies of a probability distribution on a product space. It is similar in certain respects to the Metropolis algorithm, and in many situations works with similar computational efficiency. The strategy is as follows. Suppose you want to simulate the distribution μ of a random vector $Y = (Y_1, Y_2, \ldots, Y_m)$, and for one reason or another can't do this directly. Start with an initial configuration $Y^0 = (y_1^0, y_2^0, \ldots, y_m^0)$. At each step $n = 0, 1, 2, \ldots$, update the current state Y^n by first choosing an index $i \in [m]$ at random and then replacing the ith coordinate by re-sampling from the conditional distribution of Y_i given $Y_j = y_j^n$ for all $j \neq i$, that is, replace Y_i^n by Y_i^{n+1} where

$$P(Y_i^{n+1} = u \,|\, Y_j^n = y_j \ \ \forall\, j \neq i) = \mu(Y_i = u \,|\, Y_j = y_j \ \ \forall\, j \neq i)).$$

Since the updating rule uses only the current state, the resulting sequence of states Y^n is a Markov chain.

Proposition 11. *Assume that the distribution μ is discrete. Then μ is a stationary distribution for the Markov chain Y^n, and the Markov chain is reversible with respect to μ.*

[2] Ising's Ph. D. thesis supervisor LENZ had suggested to Ising that a phase transition might exist in the Ising model; Ising was able to prove that there is *no* phase transition in one dimension, but proved nothing about the behavior in higher dimensions.

Proof. Exercise: check the detailed balance conditions (it shouldn't take more than about three lines). □

6.2. **Monotonicity properties of the Glauber chain.** Fix a domain Λ and a boundary configuration ξ. Let $q(x,y) = q_{\beta,\xi}(x,y)$ be the transition probability matrix of the Glauber chain (Gibbs sampler) for the Ising model on Λ at inverse temperature β. Because only one site can be changed at any step, the only transitions that can occur are between configurations x, x' that differ only at a single site, say site i. Hence, all of the terms in the expression (5.4) except those that involve the spins at nearest neighbors of site i are the same in $H(x)$ and $H(x')$, so the transition probabilities $q(x,x')$ involve only the spins at the neighbors $j \sim i$. Thus, the transition probabilities have a simple closed form: for any spin configurations y and any site $i \in \Lambda$,

$$q(y, y^{+,i}) = \frac{e^{-\beta S_i(y)}}{e^{-\beta S_i(y)} + e^{+\beta S_i(y)}} \quad \text{and} \tag{6.1}$$

$$q(y, y^{-,i}) = \frac{e^{+\beta S_i(y)}}{e^{-\beta S_i(y)} + e^{+\beta S_i(y)}}$$

where

$$S_i(y) = \sum_{j \in \Lambda\,:\, j\sim i} y_j + \sum_{j \in \partial\Lambda\,:\, j\sim i} \xi_j$$

and where $y^{+,i}$ and $y^{-,i}$ are the spin configurations obtained from y by re-setting the ith spin to $+1$ and -1, respectively.

Proposition 12. *For any fixed $\beta > 0$ and any two boundary configurations satisfying $\xi \le \xi_*$ the Markov kernel $q_{\beta,\xi}(\cdot,\cdot)$ is stochastically dominated by $q_{\beta,\xi_*}(\cdot,\cdot)$. Consequently, for each fixed $\beta > 0$ and each boundary configuration ξ, the Markov kernel $q_{\beta,\xi}(\cdot,\cdot)$ is stochastically monotone.*

Proof. Let $x \le x_*$ be two spin configurations such that all spins of x are $\le$ the corresponding spins of x_*. We must show that the distribution $q_{\beta,\xi}(x,\cdot)$ is stochastically dominated by $q_{\beta_*,\xi_*}(x_*,\cdot)$. Since Glauber dynamics changes at most one coordinate at a time, it suffices to show that for any two configurations $y \le y_*$ and any coordinate (site) i,

$$q_{\beta,\xi}(y, y^{+,i}) \le q_{\beta,\xi_*}(y_*, y_*^{+,i}) \tag{6.2}$$

where $y^{+,i}$ is the configuration obtained from y by re-setting the ith coordinate spin to $+1$. (Mental exercise: Convince yourself that this implies that upper corner probabilities for $q_{\beta,\xi}(y,\cdot)$ are dominated by the corresponding upper corner probabilities for $q_{\beta_*,\xi_*}(y_*,\cdot)$.)

Look closely at the equations (6.1) that give the transition probabilities for the Glauber chain. If the configuration y is replaced by a configuration $y_* \ge y$, then for any *interior* site i (that is, one that doesn't have a neighbor $j \in \partial\Lambda$)

$$\sum_{j\,:\,j\sim i} y_j \le \sum_{j\,:\,j\sim i} (y_*)_j.$$

Consequently, the effect of replacing y by y_* while keeping β fixed is to *increase* (or at least not to decrease) the probability of flipping the spin at site i to a $+$. (To see this, check that the exponential in the numerator of $q(y, y^{+,i})$ goes up, and the inverse exponential in the denominator goes down.) Note that changing ξ to ξ_* has no effect on this transition probability, because for *interior* sites i the transition probabilities (6.1) do not involve boundary spins.

Now consider the case where $i \in \Lambda$ has at least one neighbor $j \in \partial\Lambda$. If ξ is replaced by ξ_* and y by y_*, then

$$\sum_{j\in\Lambda\,:\,j\sim i} y_j + \sum_{j\in\partial\Lambda\,:\,j\sim i} \xi_j \le \sum_{j\in\Lambda\,:\,j\sim i} (y_*)_j + \sum_{j\in\partial\Lambda\,:\,j\sim i} (\xi_*)_j.$$

Consequently, by (6.1), the probability of flipping the spin at site i to a $+$ is higher when ξ is replaced by ξ_* and y by y_*. This proves (6.2). □

Corollary 13. *For any fixed $\beta > 0$ and any two boundary configurations satisfying $\xi \le \xi_*$, the corresponding Gibbs distributions satisfy*

$$\mu^{\xi}_{\beta,\Lambda} \overset{D}{\le} \mu^{\xi_*}_{\beta,\Lambda}. \tag{6.3}$$

This follows immediately from the preceding proposition and Proposition 5.

Stochastic monotonicity considerations provide an approach to making sense of what statistical physicists call the *thermodynamic limit* for the Ising model. Let Λ_m be the square (or cube) of side length $2m+1$ centered at the origin, and fix $\beta > 0$. Denote by

$$\mu^+_m = \mu^+_{\beta,\Lambda_m} \quad \text{and}$$
$$\mu^-_m = \mu^-_{\beta,\Lambda_m}$$

the Gibbs states for the Ising model at inverse temperature β with boundary spins all set at $+$ (for μ^+_m) or $-$ (for μ^-_m). Fix a finite set F of sites, and denote by

$$X_F = (X_u)_{u\in F}$$

the random vector of spins at the sites $u \in F$. Since F is finite, it will be entirely contained in Λ_m for all sufficiently large m.

Corollary 14. *For any finite set F of sites denote by $G_F = \{X_v = +1 \ \forall\ v \in F\}$ the event that all of the spins in F are $+1$. For any value of $\beta > 0$, if m is large enough that $F \subset \Lambda_m$ then*

$$\mu^-_m(G_F) \le \mu^-_{m+1}(G_F) \le \mu^+_{m+1}(G_F) \le \mu^+_m(G_F). \tag{6.4}$$

Consequently, the limits

$$\mu^+_\infty(G_F) := \lim_{m\to\infty} \mu^+_m(G_F) \quad \textit{and} \quad \mu^-_\infty(G_F) := \lim_{m\to\infty} \mu^-_m(G_F) \tag{6.5}$$

exist, and satisfy

$$\mu^-_\infty(G_F) \le \mu^+_\infty(G_F). \tag{6.6}$$

Proof. The existence of the limits and the final inequality follow immediately from the inequalities (6.4). These in turn follow from the stochastic monotonicity relation (6.3) and the Markov property of the Gibbs measures (Proposition 9). The Markov property implies that the $\mu^+_{m+1}-$distribution of the spin vector X_{Λ_m} given the event

$$X_{\Lambda_{m+1}\setminus\Lambda_m} = \sigma_{\Lambda_{m+1}\setminus\Lambda_m} := \sigma$$

is the unconditional distribution of X_{Λ_m} under the Gibbs measure $\mu^\sigma_{\Lambda_m}$. But relation (6.3) implies that the distribution of X_{Λ_m} under μ^+_m stochastically dominates the distribution of X_{Λ_m} under $\mu^\sigma_{\Lambda_m}$, for any configuration σ of spins on $\Lambda_{m+1} \setminus \Lambda_m$. Since the (indicator function of the) event G_F is nondecreasing in the partial order $\le$, it follows that for every choice of σ,

$$\mu^+_{m+1}(G_F \,|\, X_{\Lambda_{m+1}\setminus\Lambda_m} = \sigma) \le \mu^+_m(G_F).$$

The inequality $\mu^+_{m+1}(G_F) \le \mu^+_m(G_F)$ therefore follows from elementary properties of conditional expectation. Similar arguments prove the other inequalities. □

7. Peierls' Contour Argument

Peierls' argument is based on the observation that the Ising Hamiltonian H_Λ defined by (**??**) depends only on the number of $+/-$ nearest neighbor pairs in the configuration:

$$H_\Lambda(x) = -2JL_\Lambda(x) + C_\Lambda \qquad \text{where} \tag{7.1}$$

$$L_\Lambda(x) = \sum_{\substack{i\in\Lambda, j\in V\\ i\sim j}} \delta(x_i, -x_j), \tag{7.2}$$

with $\delta(\cdot,\cdot)$ being the Kronecker delta function. Evaluation of $L_\Lambda(x)$ can be accomplished by partitioning the vertices of $\Lambda \cup \partial\Lambda$ into (maximal) connected clusters of $+$ spins and $-$ spins in x, as in Figure ; $L_\Lambda(x)$ is the number of edges in $\Lambda \cup \partial\Lambda$ connecting $+$ clusters to $-$ clusters. For two-dimensional graphs, $L_\Lambda(x)$ may be evaluated by drawing *boundary contours* around the connected clusters, as shown in the following lemma. For the remainder of this section, assume that G is the standard two-dimensional integer lattice $\mathbb{Z}^2$.

Lemma 15. *For each vertex $i \in \Lambda \cup \partial\Lambda$, let $K_i = K_i(x)$ be the maximal connected set of vertices j such that sites i and j have the same spin in configuration x. Then for any two vertices i, j such that $K_i \neq K_j$ there is a simple closed curve $\gamma = \gamma_{i,j}$, called a* boundary contour *(possibly empty) separating K_i from K_j. The curve γ is a finite union of horizontal and vertical segments in the dual lattice. Each such segment bisects an edge connecting a vertex in K_i to a vertex in K_j.*

Proof. The curve γ may be constructed using a "maze-walking" algorithm. Begin by choosing an edge e connecting K_i to K_j (if there is one), and let the first segment γ_1 of γ be a perpendicular bisector of e. Define (oriented) segments γ_n, for $n = 2, 3, \ldots$, inductively, in such a way that if one traverses the segment γ_n then a vertex of K_i is on the right and a vertex of K_j is on the left. Eventually the sequence γ_n will enter a cycle. This cycle must include all of the segments γ_n because otherwise the right/left rule would be violated somewhere. Therefore, the cycle determines a closed curve. This closed curve must completely separate the regions K_i and K_j, because otherwise one of them could not be connected. Consult your local topologist for further details. □

Corollary 16. $L_\Lambda(x) = \sum_{i,j} |\gamma_{i,j}|$. □

Assume now that the region Λ is a square. Fix a vertex $i \in \Lambda$, and let $x \in \mathcal{X}$ be a configuration such that $x_{\Lambda^c} = z^+_{\Lambda^c}$. If $x_i = -1$, then it must be that the vertex i is completely surrounded by a contour that separates it from $\partial\Lambda$, as the vertices outside Λ all have $+$ spins. In particular, the boundary of the connected cluster $K_i = K_i(x)$ of $-$ spins to which vertex i belongs contains a unique contour $\gamma := \gamma_{i,\infty}$ that separates K_i from the exterior Λ^c of the square Λ. (Note that this contour may in general surround other connected components K_j.) Define C_γ ot be the set of all vertices j that are surrounded by γ; define configuration $\tilde{x}$ to be the configuration obtained from x by flipping all spins inside $\gamma_{i,\infty}$

$$(\tilde{x})_j = \begin{cases} -x_j \text{ if } j \in C_\gamma \\ +x_j \text{ if } j \notin C_\gamma. \end{cases} \tag{7.3}$$

Lemma 17. *Let $x \in \mathcal{X}$ be any configuration such that $x_i = -1$, and let $\gamma = \gamma_{i,\infty}$ be the contour that separates K_i from Λ^c. If $\tilde{x}$ is the configuration defined by (7.3), then*

$$\frac{\mu^+_\Lambda(x)}{\mu^+_\Lambda(\tilde{x})} = \exp\{-2\beta J|\gamma|\}. \tag{7.4}$$

Proof. For all nearest neighbor pairs j, k, the spin products $x_j x_k$ and $\tilde{x}_j \tilde{x}_k$ are related as follows:

$$\begin{aligned} x_j x_k &= -\tilde{x}_j \tilde{x_k} && \text{if } j,k \text{ are on opposite sides of } \gamma; \\ &= +\tilde{x}_j \tilde{x_k} && \text{otherwise.} \end{aligned}$$

Consequently,

$$H_\Lambda(\tilde{x}) - H_\Lambda(x) = 2J|\gamma|.$$

□

Lemma 18. *The mapping $x \mapsto (\tilde{x}, \gamma)$ is one-to-one.*

Proof. Given $(\tilde{x}, \gamma)$, one can recover x by negating in the region C_γ surrounded by γ. □

Proposition 19. *For each $\beta > 0$ and each square Λ containing vertex i,*

$$\mu^+_\Lambda(X_i = -1) \le \sum_{n=4}^{\infty} n 3^n e^{-2\beta J n}. \tag{7.5}$$

Proof. On the event $X_i = -1$ the connected cluster K_i of $-$ spins containing the vertex i must be separated from the connected cluster K_∞ of $+$ spins containing the vertices on the boundary $\partial\Lambda$. Let Γ be the boundary contour of $K_i = K_i(X)$ that separates K_i from K_∞. By Lemmas 18 and 17, the $\mu^+_\Lambda-$probability that $X_i = -1$ and $\Gamma = \gamma$ satisfies

$$\mu^+_\Lambda(X_i = -1 \text{ and } \Gamma = \gamma) \le \exp\{-2\beta J|\gamma|\}.$$

Consequently,

$$\mu^+_\Lambda(X_i = -1) \le \sum_\gamma \exp\{-2\beta J|\gamma|\},$$

where the sum is over all contours in the (dual) integer lattice surrounding i. To estimate the number of such surrounding contours of length k, observe that any such contour must intersect the vertical upward ray emanating from vertex i at some point within distance k of i. Starting from this intersection point, the contour is formed by attaching successive line segments, one at a time; at each stage, there are at most 3 such segments to choose from. Hence, the number of surrounding contours of length k is at most $k3^k$. The estimate (7.5) now follows easily. □

Since the sum on the right side of inequality (7.5) is less than $1/2$ for all sufficiently large values of β Proposition 19, together with Theorem **??**, implies that (5.5) holds at low temperature.

8. The High Temperature Limit

In this section we shall prove the following proposition, which implies that (5.6), and hence also (**??**), hold at high temperature.

Proposition 20.

$$\tanh(-4\beta J) < 1/4 \Longrightarrow \lim_{n\to\infty} \mu_n^+\{X_i = -1\} = 1/2. \tag{8.1}$$

The proof, unlike Peierls' argument, does not really depend on planarity of the underlying graph, and may be extended not only to the higher-dimensional integer lattices but to arbitrary vertex-regular graphs (graphs with the property that all vertices have the same number of incident edges). We shall only discuss the case $G = \mathbb{Z}^2$.

8.1. **Bernoulli-p Site Percolation.** The upper bound of $1/4$ in (8.1) for $\tanh(-4\beta J)$ emerges from the world of *site percolation*. In its simplest incarnation, site percolation has to do wtih the connectivity properties of the random graph obtained from the two-dimensional integer lattice by tossing a $p-$coin at every vertex, then erasing the vertex, and all edges incident to it, if the coin toss results in a T. *Percolation* is the event that the resulting subgraph of $\mathbb{Z}^2$ has an infinite connected cluster of vertices, equivalently, that $\mathbb{Z}^2$ has an infinite connected cluster of $H-$vertices.

Proposition 21. *If $p < 1/4$ then percolation occurs with probability 0.*

Proof. It is enough to show that for any vertex i, the probability that i is part of an infinite connected cluster of Hs is zero. Denote by K the (maximal) connected cluster of vertices containing i at which the coin toss is H. Define sets $F_0, F_1, F_2, \ldots$ inductively as follows: Let $F_0 = \{i\}$, and for each $n \geq 0$ define F_{n+1} to be the set of all vertices at which the coin toss is H that are nearest neighbors of vertices in F_n and that have not been listed in $\cup_{j=0}^n F_j$. I claim that

$$E|F_{n+1}| \leq 4pE|F_n|. \tag{8.2}$$

To see this, observe that, for each vertex $j \in F_n$ there are at most 4 vertices adjacent to j that can be included in F_{n+1}. For each of these, the conditional probability that it is included in F_{n+1}, given the coin tosses that have resulted in constructing $F_0, F_1, \ldots, F_n$, is at most p; consequently, the expected number that are included is no more than $4p$.

The cluster K is the union of the sets $F_0, F_1, \ldots$, and so its expected cardinality is bounded by $\sum_n E|F_n|$. By inequality (8.2), if $4p < 1$ then $E|K| < \infty$, in which case K is finite with probability 1. □

Fix a site $i \in V = \mathbb{Z}^2$, and denote by Λ_n the square of side $2n+1$ centered at the origin. Say that i percolates to $\partial\Lambda_n$ if the connected cluster of Hs containing site i extends to the boundary of Λ_n, equivalently, if there is a path of $H-$vertices from i ot the boundary of Λ_n. Denote this event by $A(i,n)$.

Corollary 22. *If $p < 1/4$ then $\lim_{n\to\infty} P_p(A(i,n)) = 0$ for each site i.* □

8.2. **Monotone Coupling of Gibbs States.**

Proposition 23. *Fix $\beta > 0$, and set $p = \tanh(-4\beta J)$. On some probability space $(\Omega, \mathcal{F}, P)$, there exist $\mathcal{X}-$valued random variables $Z^{(n)} \leq Y^{(n)}$ with marginal distributions μ_n^- and μ_n^+, respectively, such that*

$$P(Z_i^{(n)} \neq Y_i^{(n)}) \leq P_p(A(i,n)), \tag{8.3}$$

where $P_p(A(i,n))$ is the probability that site i percolates to $\partial\Lambda_n$ in Bernoulli-p site percolation.

Observe that this proposition and Corollary 22 imply Proposition 20, because Corollary 22 implies that the probability that site i percolates to $\partial\Lambda_n$ converges to zero if $p = \tanh(-4\beta J) < 1/4$. The proof of Proposition 23 will use the following lemma, which explains the occurrence of the quantity $\tanh(-4\beta J)$.

Lemma 24. *For any two configurations z, y such that $z \le y$, and for any finite regions $\Sigma \subset \Lambda$ and any site $i \in \Lambda - \Sigma$,*

$$\mu^+_\Lambda(X_i = +1 \,|\, X_\Sigma = y_\Sigma) - \mu^-_\Lambda(X_i = +1 \,|\, X_\Sigma = z_\Sigma) \le \tanh(-4\beta J). \tag{8.4}$$

Proof. In view of the Markov property (Proposition **??**), it suffices to show that for any two configurations x, y,

$$|\mu^x_{\Lambda - i}(X_i = +1) - \mu^y_{\Lambda - i}(X_i = +1)| \le \tanh(-4\beta J). \tag{8.5}$$

The two probabilities in (8.5) are easily calculated, as they depend only on the spins x_j, y_j at the four nearest neighbors of i. The maximum discrepancy occurs when the $x-$spins are all $+1$ and the $y-$spins are all -1: it is $\tanh(-4\beta J)$. □

Proof of Proposition 23. Fix n, and abbreviate $\Lambda = \Lambda_n$, $Z = Z^{(n)}$, and $Y = Y^{(n)}$. There are $N := (2n+1)^2$ sites in the square Λ: label these $1, 2, \ldots, N$ in order, starting from the sites at distance 1 from $\partial\Lambda$, then proceeding through the sites at distance 2 from $\partial\Lambda$, and so on, but omitting site i until the very end, so that it is listed as site N. We will construct Z, Y one site at a time, proceeding through the sites $1, 2, \ldots, N$ in order, using independent uniform-$(0,1)$ random variables $U_1, U_2, \ldots, U_N$. (The values $Z_i = -1$ and $Y_i = +1$ are determined by the requirement that the marginal distributions of Z and Y are μ^-_n and μ^+_n, respectively.)

To construct Z_1, Y_1, use the uniform U_1 to choose $\pm$ spins from the Gibbs distributions $\mu^-_n(X_1 \in dx)$ and $\mu^+_n(X_1 \in dx)$. By Proposition **??**, these distributions are stochastically ordered, so the assignment of spins may be done in such a way that $Z_1 \le Y_1$. Moreover,by Lemma 24,

$$|\mu^-_n(X_1 = +1) - \mu^+_n(X_1 = +1)| \le p,$$

so the probability that $Z_1 < Y_1$ is no larger than p. Hence, the assignment of spins may be dome in such a way that the event $Z_1 < Y_1$ is contained in the event $U_1 < p$.

Now suppose that Z_j, Y_j, for $1 \le j < m$, are defined. Conditional on the event $Z_j = z_j$, $Y_j = y_j$, with $z_j \le y_j$, use the uniform random variable U_m to assign the spins $\pm$ at Z_m and Y_m using the conditional distributions

$$Z_m \sim \mu^-_n(X_m \in dx \,|\, X_j = z_j \;\forall\; 1 \le j < m) \text{ and}$$
$$Y_m \sim \mu^+_n(X_m \in dx \,|\, X_j = y_j \;\forall\; 1 \le j < m).$$

Since $z_j \le y_j$, these conditional distributions are again stochastically ordered, by Proposition **??** and Proposition **??**; consequently, the assignment of spins may be done in such a way that $Z_m \le Y_m$. Moreover, by Lemma 24, the conditional probability that $Z_m < Y_m$, given the assignments $Y_j = y_j$ and $Z_j = z_j$ for $1 \le j < m$, is, once again, no larger than p; consequently, the assignments may be done in such a way that the event $Z_m < Y_m$ is contained in the event $U_m < p$.

It remains to show that the inequality (8.3) holds. By construction, $Z_j < Y_j$ can only occur if $U_j < p$. Moreover, by Corollary **??**, if in the course of the construction it develops

that $Z_j = Y_j$ for all sites j in a contour that surrounds site i, then it must be the case that $Z_i = Y_i$, as the conditional distributions of the spins Z_k and Y_k will coincide for all sites k after completion of the contour. Thus, $Z_i \neq Y_i$ can only occur if there is a connected path of sites j leading from site i to $\partial\Lambda$ along which $U_j < p$. But this is precisely the event that site i percolates to $\partial\Lambda$ in Bernoulli-p percolation. $\square$