

Lecture 38: MRFs and Gibbs Distributions; the Ising Model

Charles A. Bouman

Purdue University

Monday, November 23, 2026

Outline

- 1 Loopy dependencies and the intractable partition function
- 2 Neighborhood systems and the definition of an MRF
- 3 Cliques and the Gibbs distribution
- 4 The Hammersley–Clifford theorem
- 5 What is and is not computable
- 6 The Ising model, and its conditional distribution
- 7 The critical temperature in 2-D
- 8 1-D and 2-D relationships; the M -level MRF

Reading: Chapter 14.1–14.4, *Foundations of Computational Imaging*.

Today's Big Ideas

- An MRF is a random field in which each pixel depends only on its neighbors, and its joint distribution is not the product of those conditional distributions.
- A Gibbs distribution gives an image a probability through an energy function, which is a sum of potential functions over cliques.
- The Hammersley–Clifford theorem says that an MRF and a Gibbs distribution are the same thing, so a model can be specified by writing down its energy function.
- The partition function cannot be computed, but it does not depend on the image, so the energy function alone is enough to find the most probable image and to compare two images.
- In 2-D the Ising model has a critical temperature, below which two pixels stay correlated no matter how far apart they are.

In 1-D an MRF is just a Markov chain, so nothing is gained. MRFs earn their place in two or more dimensions.

The problem: generalizing Markov chains to 2-D

- In 1-D, state sequence estimation of HMMs labels samples of a signal. In 2-D, the same approach corresponds to **image segmentation**.
- To formalize this, we must generalize the Markov chain to two or more dimensions. As with AR models, the generalization is not straightforward: it implies the need for **noncausal prediction**, which results in **loopy dependencies** in the model.
- These loopy dependencies generally make closed-form calculation of the PMF impossible, due to the intractable form of the **partition function**, or normalizing constant, of the distribution.

The workaround

Discrete MRF models do not require a 1-D causal ordering of the pixels, so they can produce more natural and isotropic image models. The disadvantage is that parameter estimation becomes much more difficult.

The key result required to work around this limitation is the Hammersley–Clifford theorem.

Neighborhood systems

Let S be a set of lattice points with elements $s \in S$. We write ∂s for the neighbors of s ; $\partial s \subset S$, so ∂ maps S to the power set 2^S . Not just any mapping qualifies — it must meet a symmetry constraint.

Definition: Neighborhood System

Let S be a set of lattice points; then any mapping $\partial : S \rightarrow 2^S$ is a **neighborhood system** if for all $s, r \in S$,

$$r \in \partial s \Rightarrow s \in \partial r \quad \text{and} \quad s \notin \partial s.$$

- If r is a neighbor of s , then s must be a neighbor of r , and s may not be a neighbor of itself.
- The definition is not restricted to a regular lattice. But on a regular lattice with a spatially invariant neighborhood, symmetry requires the neighbors of a point to be symmetrically distributed about each pixel.

$X_{(0,0)}$	$X_{(0,1)}$	$X_{(0,2)}$	$X_{(0,3)}$	$X_{(0,4)}$
$X_{(1,0)}$	$X_{(1,1)}$	$X_{(1,2)}$	$X_{(1,3)}$	$X_{(1,4)}$
$X_{(2,0)}$	$X_{(2,1)}$	$X_{(2,2)}$	$X_{(2,3)}$	$X_{(2,4)}$
$X_{(3,0)}$	$X_{(3,1)}$	$X_{(3,2)}$	$X_{(3,3)}$	$X_{(3,4)}$
$X_{(4,0)}$	$X_{(4,1)}$	$X_{(4,2)}$	$X_{(4,3)}$	$X_{(4,4)}$



Neighbors of $X_{(2,2)}$

The eight-point neighborhood of $X_{(2,2)}$, in gray.

Markov random fields

Definition: MRF

Let $X_s \in \Omega$ be a discrete- (continuous-) valued random field defined on the lattice S with neighborhood system ∂s . Further assume that the X has probability mass (density) function $p(x)$. Then we say that X is an **MRF** if for all $x \in \Omega$,

$$P\{X_s = x_s | X_r \text{ for } r \neq s\} = P\{X_s = x_s | X_{\partial s}\} ,$$

for all $x_s \in \Omega$.

So each pixel of the MRF depends only on its neighbors.

The joint distribution is not the product of the conditionals

People have been tempted to write the joint distribution as the product of the associated conditional distributions; this may not generally be possible. It is typically the case that

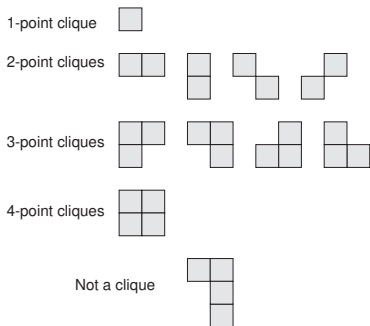
$$p(x) \neq \prod_{s \in S} p(x_s | x_{\partial s})$$

because the loopy dependencies of an MRF's noncausal structure render the Markov chain factorization invalid.

Cliques

Definition: Clique

A **clique** is any set of lattice points $c \subset S$ such that for every distinct pair of points $s, r \in c$, it is the case that $r \in \partial s$. Here S is a lattice and ∂s is a neighborhood system on S .



- Cliques are sets of points which are all neighbors of one another. At left: the cliques for an eight-point neighborhood system on a rectangular lattice.
- Any individual point $s \in S$ forms a clique with the single member $c = \{s\}$.
- Cliques and neighborhood systems are **alternative specifications of the same information**: each neighborhood system has a unique corresponding maximal set of cliques, and each set of cliques has a unique corresponding minimal neighborhood system.

Gibbs distributions

Definition: Gibbs Distribution

Let $p(x)$ be the probability mass (density) function of a discrete- (continuous-) valued random field $X_s \in \Omega$ defined on the lattice S with neighborhood system ∂s . Then we say that $p(x)$ is a **Gibbs distribution** if it can be written in the form

$$p(x) = \frac{1}{z} \exp \left\{ - \sum_{c \in \mathcal{C}} V_c(x_c) \right\} ,$$

where z is a normalizing constant known as the **partition function**, $\mathcal{C}$ is the set of all cliques, x_c is the vector containing values of x on the set c , and $V_c(x_c)$ is any function of x_c .

We sometimes refer to the function $V_c(x_c)$ as a **potential function** and the function

$$u(x) = \sum_{c \in \mathcal{C}} V_c(x_c)$$

as the associated **energy function**.

The Hammersley–Clifford theorem

Theorem: Hammersley–Clifford Theorem

Let S be an N -point lattice with neighborhood system ∂s , and let X be a discrete- (continuous-) valued random process on S with strictly positive probability mass (density) function $p(x) > 0$. Then X is an MRF if and only if $p(x)$ is a Gibbs distribution.

$$\left(\begin{array}{c} X \text{ is a Markov random field} \\ \& \\ \forall x, P\{X = x\} > 0 \end{array} \right) \begin{array}{c} \xLeftarrow{\text{Constructive}} \\ \xRightarrow{\text{Exists}} \end{array} \left(\begin{array}{c} P\{X = x\} \text{ has the form} \\ \text{of a Gibbs distribution} \end{array} \right)$$

Gibbs distributions are strictly positive, so the result carries the technical constraint that X has a strictly positive density. In practice this is a minor limitation, since the density can become exceedingly small even if it is nonzero.

Constructive direction: a Gibbs distribution is an MRF

Let $\mathcal{C}_s = \{c \in \mathcal{C} : s \in c\}$ be the cliques that contain x_s , and $\bar{\mathcal{C}}_s = \mathcal{C} - \mathcal{C}_s$ those that do not.

$$\begin{aligned} P\{X_s = x_s | X_r, r \neq s\} &= \frac{\frac{1}{Z} \exp \left\{ - \sum_{c \in \mathcal{C}} V_c(x_c) \right\}}{\sum_{x_s \in \Omega} \frac{1}{Z} \exp \left\{ - \sum_{c \in \mathcal{C}} V_c(x_c) \right\}} \\ &= \frac{\exp \left\{ - \sum_{c \in \mathcal{C}_s} V_c(x_c) \right\} \exp \left\{ - \sum_{c \in \bar{\mathcal{C}}_s} V_c(x_c) \right\}}{\sum_{x_s \in \Omega} \exp \left\{ - \sum_{c \in \mathcal{C}_s} V_c(x_c) \right\} \exp \left\{ - \sum_{c \in \bar{\mathcal{C}}_s} V_c(x_c) \right\}} \\ &= \frac{\exp \left\{ - \sum_{c \in \mathcal{C}_s} V_c(x_c) \right\}}{\sum_{x_s \in \Omega} \exp \left\{ - \sum_{c \in \mathcal{C}_s} V_c(x_c) \right\}} = \frac{\exp \{ -u_s(x_s | x_{\partial s}) \}}{\sum_{x_s \in \Omega} \exp \{ -u_s(x_s | x_{\partial s}) \}} \end{aligned}$$

where $u_s(x_s | x_{\partial s}) = \sum_{c \in \mathcal{C}_s} V_c(x_c)$ is a function of x_s and its neighbors only. So

$$P\{X_s = x_s | X_r, r \neq s\} = P\{X_s = x_s | X_r, r \in \partial s\},$$

and X_s must be an MRF. The converse proof, that an MRF must have a Gibbs distribution, is not constructive. In practice one starts with a Gibbs distribution and computes the conditional dependencies from it.

The minimal neighborhood system

Most approaches start with a Gibbs distribution and then derive the associated MRF, so it matters how a set of cliques induces a neighborhood system. If X_S has a Gibbs distribution on S with cliques $\mathcal{C}$, what is the smallest neighborhood system ∂s guaranteed to fully describe that MRF?

Minimal neighborhood system

$$\partial s = \{r \in S : \exists c \in \mathcal{C} \text{ s.t. } r \neq s \text{ and } r, s \in c\} .$$

In words: ∂s must contain any pixel r that shares a clique with s , and it may not contain any more pixels.

- Hammersley–Clifford also justifies the pairwise Gibbs distributions of Chapter 6, since neighboring pixel pairs are cliques in any neighborhood system.
- But even for an eight-point neighborhood other cliques exist, so pairwise Gibbs distributions are not the most general class of MRFs for a given neighborhood system. They are nonetheless a rich and useful subclass.

What is and is not computable

Typically it is impossible to compute or even express in simple closed form the partition function. Theoretically,

$$z = \sum_{x \in \Omega^{|S|}} \exp \{-u(x)\} ,$$

but in practice this cannot be evaluated: it is summed over an **exponentially large set of states**. This blocks ML parameter estimation, which requires

$$\hat{\theta} = \arg \min_{\theta} \{u(x) + \log z(\theta)\} .$$

But the Gibbs distribution is still extremely valuable

The partition function is not a function of x , so optimization over x needs only the energy function:

$$\hat{x} = \arg \max_x \{\log p(x)\} = \arg \min_x \{u(x)\} .$$

Sampling generally requires only the likelihood ratio between alternative states, which is computed exactly:

$$\frac{p(x')}{p(x)} = \exp \{u(x) - u(x')\} .$$

The Ising model

The first serious attempts to generalize Markov chains to multiple dimensions came from the physics community, modeling magnetic dipoles in real materials.

- In the 1920s it was well known that magnetic materials exhibit a phase transition when the temperature falls below the **Curie temperature**. Below this **critical temperature** the material will spontaneously magnetize itself; above it, it will not.
- Wilhelm Lenz and his student Ernst Ising proposed the **Ising model** to mathematically model this behavior.
- A lattice of binary random variables models the dipole states. Each X_s takes the value $+1$ or -1 , a north or south orientation.
- The potential energy of a pair of neighboring dipoles is approximated as $-Jx_sx_r$, where $s \in \partial r$ and J is a positive constant. Misaligned ($x_s = -x_r$): energy J . Aligned ($x_s = x_r$): energy $-J$.

Total energy of the lattice

$$u(x) = - \sum_{\{s,r\} \in \mathcal{C}} Jx_sx_r ,$$

where $\mathcal{C}$ is the set of all neighboring pixel pairs (i.e., cliques).

The Boltzmann distribution

What is the distribution of the dipole states? For a physical system, the answer is the **Boltzmann distribution** (also known as the Gibbs distribution) of thermodynamics:

$$p(x) = \frac{1}{z} \exp \left\{ -\frac{1}{k_b \mathcal{T}} u(x) \right\} ,$$

where $u(x)$ is the energy associated with the state x . It is derived for a system in thermodynamic equilibrium by finding the distribution of states that maximizes the entropy of the system subject to a constraint that the expected energy be constant.

Boltzmann distribution of states for the Ising model

$$p(x) = \frac{1}{z} \exp \left\{ \frac{1}{k_b \mathcal{T}} \sum_{\{s,r\} \in \mathcal{C}} J_{s,r} x_s x_r \right\} .$$

Since this is a Gibbs distribution, we know $p(x)$ must be the distribution of an MRF with cliques given by $\mathcal{C}$.

The β form of the Ising distribution

Make the standard modeling assumption of a 2-D rectangular lattice with a four-point neighborhood system and cliques formed by neighboring horizontal or vertical pixel pairs. With $\delta(x_r \neq x_s)$ the indicator function for $x_r \neq x_s$,

$$\begin{aligned} p(x) &= \frac{1}{z} \exp \left\{ \frac{J}{k_b \mathcal{T}} \sum_{\{r,s\} \in \mathcal{C}} x_r x_s \right\} \\ &= \frac{1}{z} \exp \left\{ \frac{J}{k_b \mathcal{T}} \sum_{\{r,s\} \in \mathcal{C}} (1 - 2\delta(x_r \neq x_s)) \right\} . \end{aligned}$$

The Ising distribution

$$p(x) = \frac{1}{z(\beta)} \exp \left\{ -\beta \sum_{\{r,s\} \in \mathcal{C}} \delta(x_r \neq x_s) \right\} , \quad \beta \triangleq \frac{2J}{k_b \mathcal{T}} .$$

β is a standard parameter of the Ising model, inversely proportional to the temperature. We write $z(\beta)$ to emphasize the dependence of the partition function on β . It is more convenient to work with the unitless β than with $\mathcal{T}$ directly.

The conditional distribution of a pixel

$p(x)$ is a Gibbs distribution with pairwise cliques; in 2-D the Ising model uses horizontal and vertical cliques, so by Hammersley–Clifford X must be an MRF with a four-point neighborhood. More directly:

$$p(x_s | x_{i \neq s}) = \frac{\exp \left\{ -\beta \sum_{r \in \partial s} \delta(x_s \neq x_r) \right\}}{\exp \left\{ -\beta \sum_{r \in \partial s} \delta(1 \neq x_r) \right\} + \exp \left\{ -\beta \sum_{r \in \partial s} \delta(-1 \neq x_r) \right\}} .$$

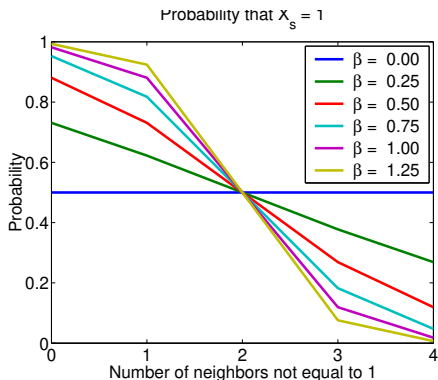
Define $v(x_s, x_{\partial s})$, the number of horizontal/vertical neighbors that differ from x_s :

$$v(x_s, x_{\partial s}) \triangleq \sum_{r \in \partial s} \delta(x_s \neq x_r) .$$

Conditional probability that $x_s = 1$

$$p(1 | x_{i \neq s}) = \frac{1}{1 + e^{2\beta[v(1, x_{\partial s}) - 2]}} .$$

Reading the conditional density



- The conditional probability is a function of $v(1, x_{\partial s})$ only — the number of neighbors *not* equal to one.
- When $\beta > 0$, as $v(1, x_{\partial s})$ increases, the probability that $X_s = 1$ decreases.
- Special case $v(1, x_{\partial s}) = 2$: half the neighbors are 1 and half are -1 , so $p(x_s | x_{\partial s}) = 1/2$.

- When β is very large and positive, a pixel is highly likely to have the same polarity as the majority of its neighbors.
- If $\beta \ll 0$, the opposite occurs: pixels are likely to have the opposite polarity to their neighbors.
- When $\beta = 0$, the pixels are all independent. Intuitively $\beta = 0$ corresponds to $\mathcal{T} = \infty$, so at infinite temperature the dipoles are highly disordered, with no correlation between them.

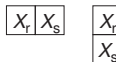
Energy as boundary length

With its four-point neighborhood, the energy function of the Ising model is equal to the total boundary length scaled by β :

$$\beta \sum_{\{r,s\} \in \mathcal{C}} \delta(X_r \neq X_s) = \beta(\text{Boundary length}) .$$

-	-	-	-	-	-	-	-
-	-	-	-	-	+	+	-
-	-	-	+	+	+	+	-
-	-	-	+	+	+	-	-
-	-	-	-	-	+	-	-
-	-	-	-	+	+	-	-
-	-	-	-	-	+	-	-
-	-	-	-	-	-	-	-

Cliques:



Boundary:



A caution

We might expect that longer boundaries become less probable as β increases. However, this is not necessarily the case, because as the length of a boundary increases the total number of possible boundaries of a given length can increase quite rapidly.

The critical temperature: 1-D versus 2-D

In 1-D there is no critical behavior

Ising first assumed the lattice to be only 1-D. Any 1-D MRF of order P is a 1-D Markov chain of order P , so this approximation allowed exact analysis. Let X_n be the resulting 1-D process with $X_0 = 1$. As $n \rightarrow \infty$ the Markov chain converges to a stationary distribution and X_n becomes independent of X_0 : in 1-D the initial state has no lasting effect.

In 2-D, Gibbs distributions can no longer be represented as simple Markov chains. The analysis is much more difficult but yields much richer behavior. In 1936 Peierls first provided a precise argument that in 2-D the Ising model exhibits the critical temperature behavior: below a specific critical temperature, dipoles in the random field **remain correlated regardless of how far apart they are**.

The statement

For $\beta > \beta_c$, it can be shown that

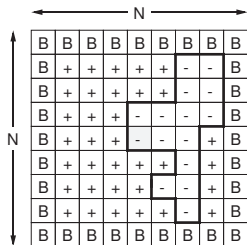
$$\lim_{N \rightarrow \infty} P\{X_0 = -1 | X_B = 1\} < 1/2 .$$

By choosing β sufficiently large, the limit can be made less than ϵ for any $\epsilon > 0$.

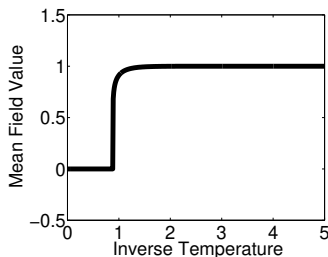
Onsager's closed-form result

In 1944, Onsager obtained closed-form expressions for properties of the Ising model as $N \rightarrow \infty$. One representative result is the conditional expectation of a pixel given that the boundary at infinity is $X_B = 1$:

$$\lim_{N \rightarrow \infty} E_N[X_0 | B = 1] = \begin{cases} \left(1 - \frac{1}{(\sinh(\beta))^4}\right)^{1/8} & \text{if } \beta > \beta_c, \\ 0 & \text{if } \beta < \beta_c \end{cases}.$$



Center Pixel X_0 : 



For $\beta < \beta_c \approx 0.88$ the expected value is zero — what you would expect if the boundary became negligible as it tended to infinity. For $\beta > \beta_c$ (i.e., $T < T_c$) this symmetry is broken and the expectation moves toward one.

Why the symmetry breaks

- For large values of β , the expectation approaches one, which means that $X_0 = 1$ with high probability.
- Intuitively, this happens because when β is above the critical value the random field's distribution becomes **bimodal**, tending to be either nearly all values of one or nearly all values of -1 .
- Without the boundary condition, the probability of X and $-X$ — the random field with the sign of each pixel reversed — must be the same, due to the symmetry in the density $p(x) = \frac{1}{Z} \exp \left\{ \frac{1}{k_b T} \sum_{\{s,r\} \in \mathcal{C}} J_{X_s X_r} \right\}$.
- The assumption of a boundary of $B = 1$ breaks this symmetry and favors the mode associated with a majority of ones in the lattice. In this case, values of -1 occur only in primarily isolated locations.

Contrast with 1-D

This behavior is distinctly different from the behavior of an ergodic 1-D Markov chain, in which the effect of the initial value decays with time.

1-D: MRFs and Markov chains are equivalent

Property: Equivalence of 1-D MRFs and Markov chains

Let $\{X_n\}_{n=0}^N$ be a discrete-time random process with strictly positive PMF taking values on a finite discrete set. Then X is a P^{th} -order MRF if and only if it is a P^{th} -order Markov chain.

- The sums in the proof can be replaced with integrals, so a similar result holds for discrete-time processes with strictly positive densities.
- In some sense this would seem to be a major disappointment. If MRFs and Markov chains are equivalent, why bother with MRFs at all? The answer is that **MRFs are rarely used in 1-D for this exact reason.**
- In two or more dimensions, MRFs are quite different from 2-D Markov processes and can exhibit interesting and useful characteristics.
- As a practical matter, a Markov chain model is almost always preferred in 1-D since, in many ways, it is more tractable.

2-D: Markov chains are MRFs, but not conversely

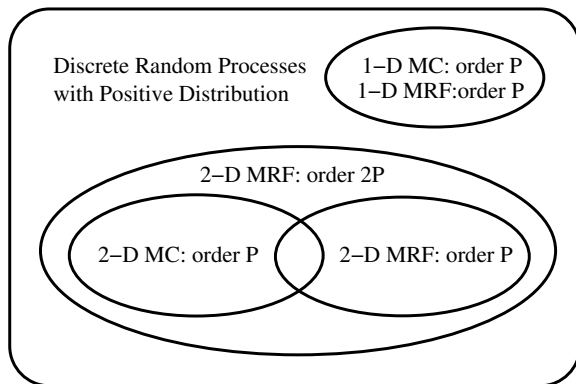
- A 2-D Markov chain of order P has the joint distribution $p(x) = \prod_{s \in S} P\{X_s = x_s | X_{s-W_P} = x_{s-W_P}\}$, which (assuming $p(x) > 0$) is a Gibbs distribution with cliques $c_s = \{s, s - W_P\}$. The induced neighborhood is the 2-D autocorrelation of the cliques with themselves.
- A P^{th} -order 2-D MRF has each pixel depending on a $(2P + 1) \times (2P + 1)$ noncausal neighborhood. So a **2-D Markov chain of order P must be a 2-D MRF of order $2P$** .
- The converse fails. For the Ising model, computing $P\{X_s = x_s | X_r = x_r, r < s\}$ requires summing the joint over all x_k for $k > s$. Some manipulation gives

$$P\{X_s = x_s | X_r = x_r, r < s\} = P\{X_s = x_s | X_r = x_r, r \in B_s\},$$

where B_s is a set of boundary pixels that precede s and lie along two rows of the image.

- B_s **extends indefinitely to the left and right**, so X_s is not a 2-D Markov chain for any fixed finite P . This makes sense: if all MRFs were Markov chains, two pixels would decorrelate when far apart — which does not happen when β is above the critical value.

The Venn diagram



Relationship between Markov chains (MC) and MRFs in 1-D and 2-D among the set of all discrete-state random processes with strictly positive distribution. In 1-D, Markov chain and MRF models are equivalent; in 2-D they are not. 2-D Markov chains of order P are always 2-D MRFs, but of order $2P$.

The M -level discrete MRF

The **M -level discrete MRF** is a simple and very useful generalization of the Ising model, typically with an eight-point neighborhood. Let $\mathcal{C}_1$ be the horizontal and vertical cliques and $\mathcal{C}_2$ the diagonal cliques; pixels x_s take values in $\{0, \dots, M-1\}$.

$$t_1(x) \triangleq \sum_{\{s,r\} \in \mathcal{C}_1} \delta(x_r \neq x_s) ,$$
$$t_2(x) \triangleq \sum_{\{s,r\} \in \mathcal{C}_2} \delta(x_r \neq x_s) ,$$

counting vertically/horizontally adjacent and diagonally adjacent pixels with different values.

Distribution of an M -level MRF

$$p(x) = \frac{1}{Z} \exp \{ -(\beta_1 t_1(x) + \beta_2 t_2(x)) \} .$$

The M -level MRF: conditional distribution

Define v_1 and v_2 , counting the horizontal/vertical and diagonal neighbors with a different value from X_s :

$$v_1(x_s, x_{\partial s}) \triangleq \sum_{r \in [\partial s]_1} \delta(x_s \neq x_r) ,$$
$$v_2(x_s, x_{\partial s}) \triangleq \sum_{r \in [\partial s]_2} \delta(x_s \neq x_r) ,$$

where $[\partial s]_1$ is the set of horizontally and vertically adjacent neighbors to s and $[\partial s]_2$ is the set of diagonally adjacent neighbors to s .

$$p(x_s | x_{r \neq s}) = \frac{\exp \{ -\beta_1 v_1(x_s, x_{\partial s}) - \beta_2 v_2(x_s, x_{\partial s}) \}}{\sum_{m=0}^{M-1} \exp \{ -\beta_1 v_1(m, x_{\partial s}) - \beta_2 v_2(m, x_{\partial s}) \}} .$$

In this form, it is clear that when $\beta_1, \beta_2 > 0$, the pixel X_s is more likely to take on a value that is shared by a number of its neighbors.

If time permits: the Kindermann–Snell proof (1 of 2)

Goal: for β sufficiently large, $P_N\{X_0 = -1 | B = 1\} < 1/3$, hence $\lim_{N \rightarrow \infty} P_N\{X_0 = B\} \geq 2/3$.

- Let $\Omega = \{+1, -1\}^{N^2}$ and let $L(x)$ be the total boundary length between neighboring pixels of opposite polarity. Then $P\{X = x | X_B = 1\} = \frac{1}{Z} \exp\{-\beta L(x)\}$.
- Let $c(x)$ be the closed boundary around X_0 that includes all pixels of value -1 , of length $L(c(x))$. Let $\Omega_{c_o} = \{x \in \Omega : c(x) = c_o \text{ and } x_0 = -1\}$.
- Flip the pixels inside c_o : $x'_s = f_{c_o}(x) = 1$ if s lies inside the boundary c_o , and x_s if s lies outside. The flipped set Ω'_{c_o} is in one-to-one correspondence with Ω_{c_o} , so $|\Omega_{c_o}| = |\Omega'_{c_o}|$, and

$$L(f(x)) = L(x) - L(c_o) .$$

The bound

$$P_N\{c(X) = c_o, X_0 = -1 | B = 1\} \leq e^{-\beta L(c_o)} .$$

If time permits: the Kindermann–Snell proof (2 of 2)

Let C_L contain all boundaries c_o of length L whose interior includes X_0 . Summing the bound over all boundaries,

$$\begin{aligned} P_N\{X_0 = -1 | B = 1\} &= \sum_{L=4}^{\infty} \sum_{c_o \in C_L} P\{c(x) = c_o, X_0 = -1 | B = 1\} \\ &\leq \sum_{L=4}^{\infty} \sum_{c_o \in C_L} e^{-\beta L} = \sum_{L=4}^{\infty} |C_L| e^{-\beta L}. \end{aligned}$$

How many boundaries of length L are there? Boundaries have even length, so $|C_L| = 0$ for L odd. A boundary must be within $L/2$ of X_0 , so there are L^2 ways to choose its starting point, four possibilities for the initial direction, and three choices at each subsequent step: $|C_L| \leq 4L^2 3^{L/2}$ for L even. Setting $L = 2n$,

$$P_N\{X_0 = -1 | B = 1\} \leq \sum_{n=2}^{\infty} 4(2n)^2 3^{2n} e^{-\beta 2n} = \sum_{n=2}^{\infty} 16n^2 (9e^{-2\beta})^n < 1/3$$

for β sufficiently large, which proves the desired result.

Summary

- A neighborhood system requires $r \in \partial s \Rightarrow s \in \partial r$ and $s \notin \partial s$. X is an MRF if $P\{X_s = x_s | X_r, r \neq s\} = P\{X_s = x_s | X_{\partial s}\}$. The joint distribution is *not* the product of the conditionals.
- A clique is a set of points that are all neighbors of one another. A Gibbs distribution is $p(x) = \frac{1}{z} \exp\{-\sum_{c \in \mathcal{C}} V_c(x_c)\}$ with energy function $u(x) = \sum_{c \in \mathcal{C}} V_c(x_c)$.
- **Hammersley–Clifford:** with $p(x) > 0$, X is an MRF if and only if $p(x)$ is a Gibbs distribution.
- z is a sum over exponentially many states, so ML parameter estimation is blocked; but $\hat{x} = \arg \min_x \{u(x)\}$ and $p(x')/p(x) = \exp\{u(x) - u(x')\}$ need only u .
- Ising model: $p(x) = \frac{1}{z(\beta)} \exp\{-\beta \sum_{\{r,s\} \in \mathcal{C}} \delta(x_r \neq x_s)\}$ with $\beta = 2J/(k_b T)$. In 2-D, for $\beta > \beta_c \approx 0.88$, correlations extend to infinite distances.
- In 1-D, P^{th} -order MRFs and Markov chains are equivalent. In 2-D, a Markov chain of order P is an MRF of order $2P$, but the Ising model is not a 2-D Markov chain of any finite order.

Reading: Chapter 14.1–14.4. Next lecture: Chapter 15, stochastic simulation. Lab 7 due Dec. 4 with HW 12.