Markov Random Fields

- Noncausal model
- Advantages of MRF's
  - Isotropic behavior
  - Only local dependencies
- Disadvantages of MRF's
  - Computing probability is difficult
  - Parameter estimation is difficult
- Key theoretical result: Hammersley-Clifford theorem

Definition of Neighborhood System

- Define

  S - set of lattice points

  s - a lattice point, $s \in S$

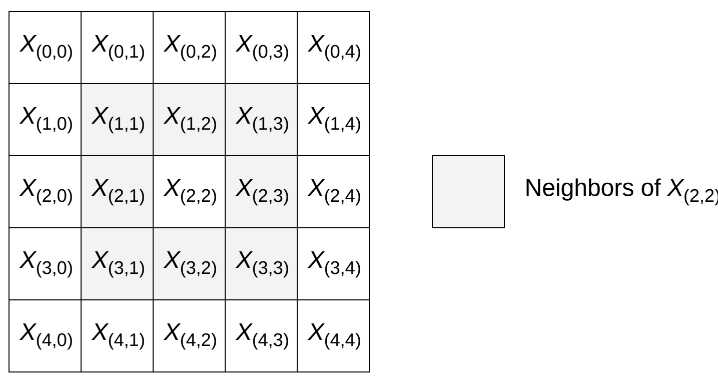
  X_s - the value of X at s

  ∂s - the neighboring points of s

- A neighborhood system ∂s must be symmetric

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

- Example of 8 point neighborhood



Markov Random Field

- Definition: A random object X on the lattice S with neighborhood system ∂s is said to be a Markov random field if for all $s \in S$

$$p(x_s | x_r \text{ for } r \neq s) = p(x_s | x_{\partial r})$$

- Problem: How do we write down the distribution for an MRF?

  Unfortunately

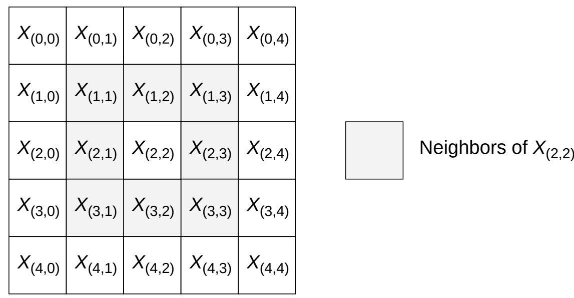
$$p(x) \neq \prod_{s \in S} p(x_s | x_r \text{ for } r \neq s)$$

Definition of Clique

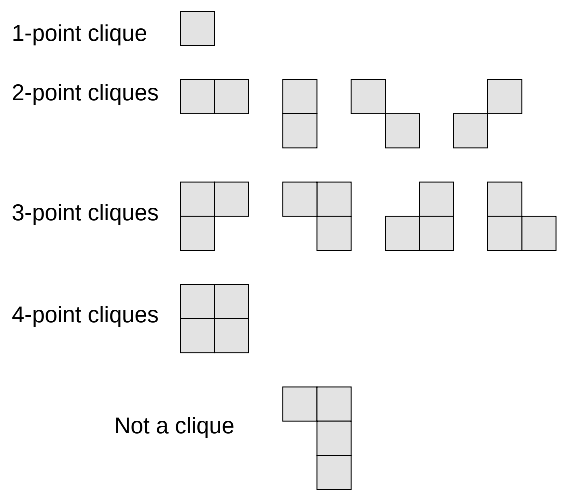
- A clique is a set of points, c, which are all neighbors of each other

$$\forall s, r \in c,\ r \in \partial s$$

- 8 point neighborhood system



- Example of cliques for 8 point neighborhood



Gibbs Distribution

x_c - The value of X at the points in clique c.

$V_c(x_c)$ - A potential function is any function of x_c.

- A (discrete) density is a Gibbs distribution if

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

$\mathcal{C}$ is the set of all cliques

Z is the normalizing constant for the density.

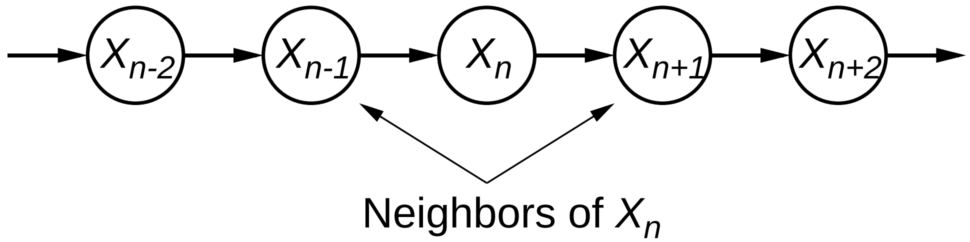
- Z is known as the **partition function**.
- $U(x) = \sum_{c \in \mathcal{C}} V_c(x_c)$ is known as the **energy function**.

Hammersley-Clifford Theorem[1]

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

- Gives you a method for writing the density for a MRF
- Does not give the value of Z, the partition function.
- Positivity, $P\{X = x\} > 0$, is a technical condition which we will generally assume.

Markov Chains are MRF's



- Neighbors of n are $\partial n = \{n-1, n+1\}$
- Cliques have the form $c = \{n-1, n\}$
- Density has the form

$$\begin{aligned} p(x) &= p(x_0) \prod_{n=1}^{N} p(x_n | x_{n-1}) \\ &= p(x_0) \exp\left\{ \sum_{n=1}^{N} \log p(x_n | x_{n-1}) \right\} \end{aligned}$$

- The potential functions have the form

$$V(x_n, x_{n-1}) = -\log p(x_n | x_{n-1})$$

1-D MRF's are Markov Chains

- Let X_n be a 1-D MRF with $\partial n = \{n-1, n+1\}$
- The discrete density has the form of a Gibbs distribution

$$p(x) = p(x_0) \exp\left\{ -\sum_{n=1}^{N} V(x_n, x_{n-1}) \right\}$$

- It may be shown that this is a Markov Chain.
- Transition probabilities may be difficult to compute.

The Ising Model

- First proposed to model 2-D magnetic structures.
- See the work of Peierls for an early treatment[13, 12].
- Kindermann and Snell have a very clear tutorial treatment in [9].
- Lattice geometry
  - S is a rectangular lattice of N pixels.
  - 4-point neighborhood system with cliques $c \in \mathcal{C}$.
  - Assume circular boundary conditions for now.
- Lattice energy
  - Each pixel $X_s \in \{-1, +1\}$ corresponding to north and south poles.
  - Potential of clique $\{r, s\} \in \mathcal{C}$ is $-\frac{J}{2} X_r X_s$.
  - Total energy is

$$u(x) = -\frac{J}{2} \sum_{\{r,s\} \in \mathcal{C}} X_r X_s \ .$$

Physical Basis of Gibbs Distribution

- What is the equilibrium distribution $p_e(x)$?
- Expected energy is

$$\mathcal{E}\{p_e\} = \sum_x p_e(x)\, u(x)$$

- Entropy is

$$\mathcal{H}\{p_e\} = \sum_x -p_e(x)\, \log p_e(x)$$

- First Law of Thermodynamics: Expected energy must be constant.
- Second Law of Thermodynamics: Entropy must be maximized.

$$p_e(x) = \arg \max_{p_e : \mathcal{E}\{p_e\} = \text{const}} \mathcal{H}\{p_e\}$$

- Solution is the Gibbs distribution!

$$p(x) = \frac{1}{z} \exp\left\{ -\frac{1}{kT} u(x) \right\}$$

  – T is tempurature
  – k is Boltzmann's constant

Distribution for Ising Model

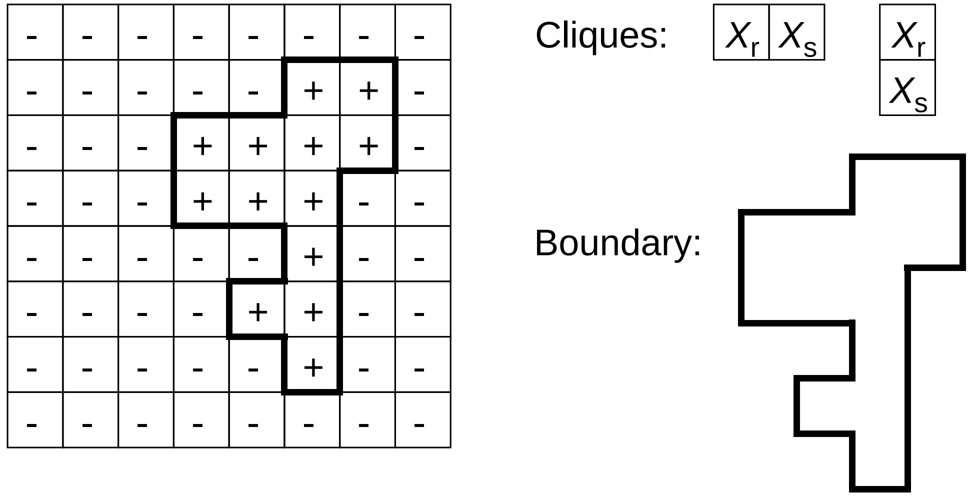
- Equalibrium distribution for Ising model is

$$\begin{aligned} p(x) &= \frac{1}{z}\exp\left\{\frac{J}{2kT}\sum_{\{r,s\}\in\mathcal{C}} X_r X_s\right\} \\ &= \frac{1}{z}\exp\left\{\frac{J}{kT}\sum_{\{r,s\}\in\mathcal{C}}\left(\frac{1}{2}-\delta(X_r\neq X_s)\right)\right\} \\ &= \frac{1}{z'}\exp\left\{-\beta\sum_{\{r,s\}\in\mathcal{C}}\delta(X_r\neq X_s)\right\} \end{aligned}$$

  where $\beta = \frac{J}{kT}$ is a model parameter and $\delta(X_r \neq X_s)$ is an indicator function for the event $X_r \neq X_s$.

- By the Hammersly-Clifford Theorem, X is a MRF with a 4-point neighborhood.

Interpretation of Ising Model



- Potential functions are given by

$$V(x_r, x_s) = \beta\delta(x_r \neq x_s)$$

- Energy function is given by

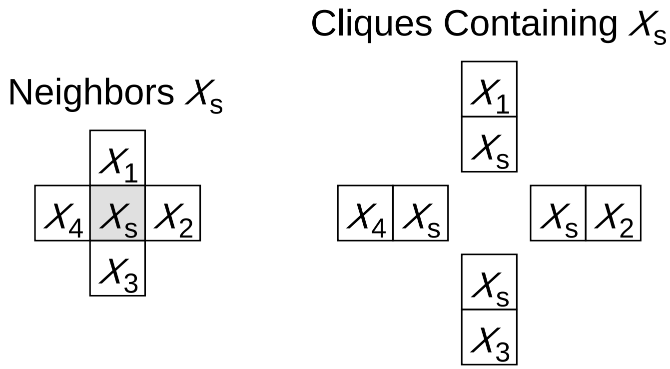
$$\sum_{c \in \mathcal{C}} V_c(x_c) = \beta(\text{Boundary length})$$

- Interpretation of probability density

$$p(x) = \frac{1}{z} \exp\{-\beta(\text{Boundary length})\}$$

- Longer boundaries $\Rightarrow$ less probable

Conditional Probability of a Pixel in Ising Model



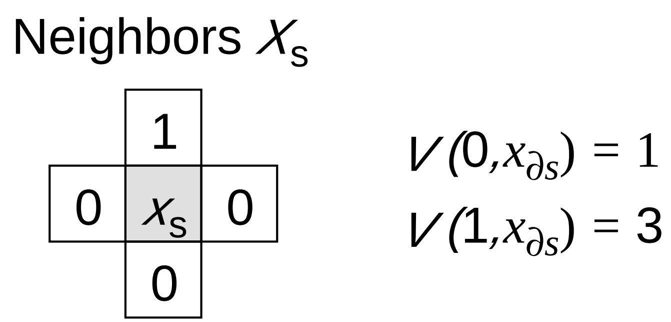
- The probability of a pixel given all other pixels is

$$p(x_s|x_{i\neq s}) = \frac{\frac{1}{Z}\exp\left\{-\sum_{c\in\mathcal{C}} V_c(x_c)\right\}}{\sum_{x_s=0}^{M-1}\frac{1}{Z}\exp\left\{-\sum_{c\in\mathcal{C}} V_c(x_c)\right\}}$$

- Notice: Any term $V_c(x_c)$ which does not include x_s cancels.

$$p(x_s|x_{i\neq s}) = \frac{\exp\left\{-\beta\sum_{i=1}^{4}\delta(x_s\neq x_i)\right\}}{\sum_{x_s=0}^{M-1}\exp\left\{-\beta\sum_{i=1}^{4}\delta(x_s\neq x_i)\right\}}$$

Conditional Probability of a Pixel in Ising Model (Continued)



- Define

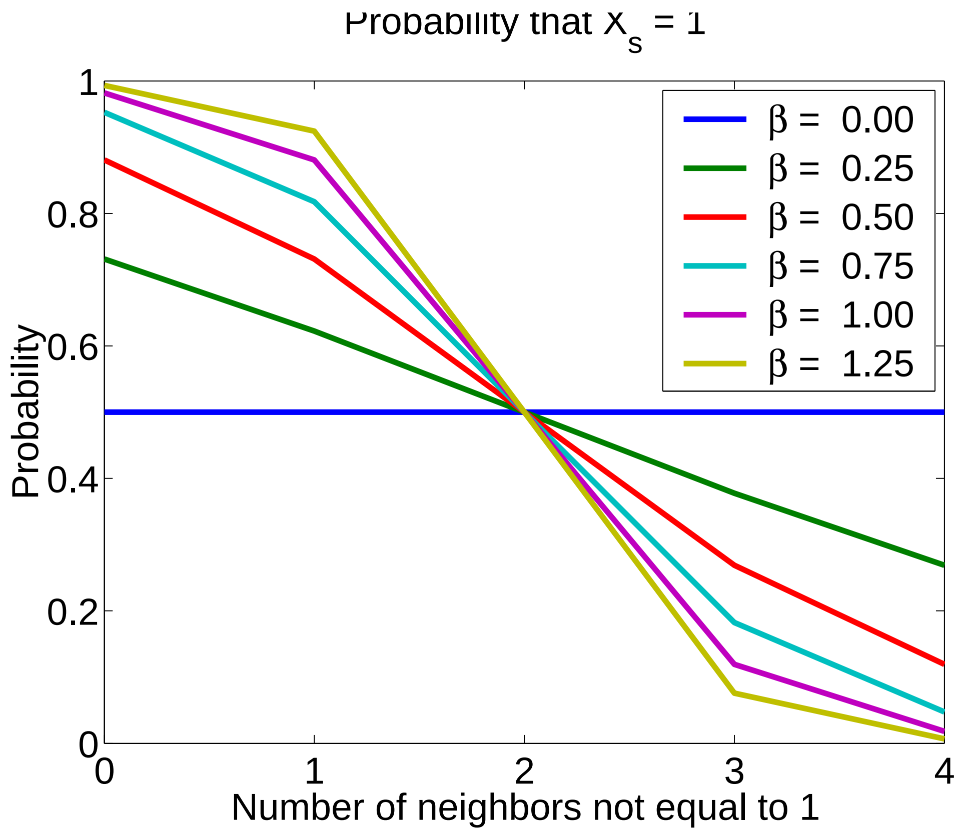
$$v(x_s, \partial x_s) \stackrel{\triangle}{=} \# \text{ of horzontal/vertical neighbors} \neq x_s$$

- Then

$$p(x_s | x_{i \neq s}) = \frac{\exp\{-\beta v(x_s, \partial x_s)\}}{\sum_{x'_s = \{-1,+1\}} \exp\{-\beta v(x'_s, \partial x_s)\}}$$

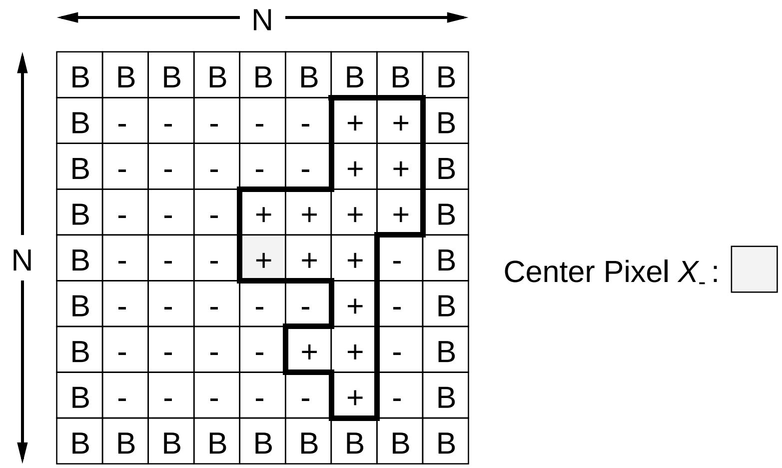
- When $\beta > 0$, X_s is most likely to be the majority neighboring class.

Conditional Distribution Plots



- $P\{X_s = 1 | X_r \text{ for } r \neq s\}$ for different values of β.

Critical Temperature Behavior[13, 12, 9]



- $\frac{1}{\beta}$ is analogous to temperature.
- Peierls showed that for $\beta > \beta_c$

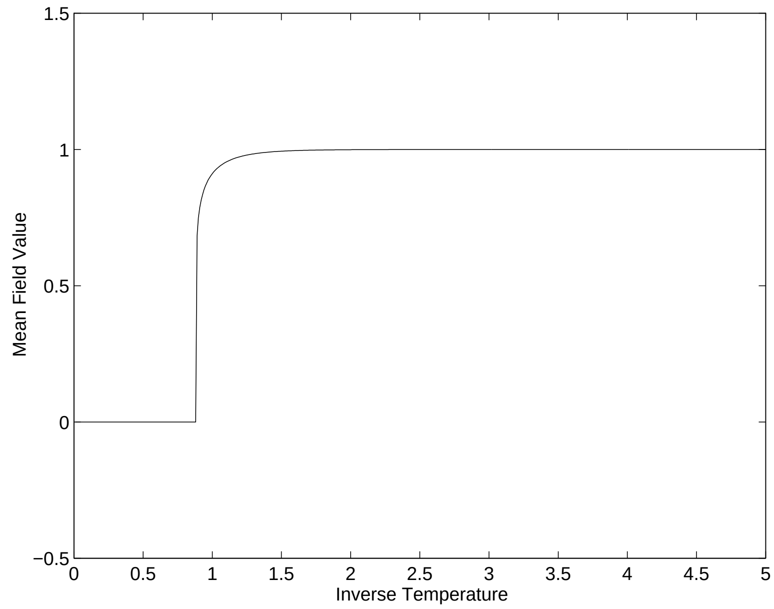
$$\lim_{N\to\infty} P(X_0 = 0|B = 0) \neq \lim_{N\to\infty} P(X_0 = 0|B = 1)$$

- The effect of the boundary does not diminish as $N \to \infty$!
- $\beta_c \approx .88$ is known as the critical temperature.
- Very nice proof of critical temperature in [9].

Critical Temperature Analysis[11]

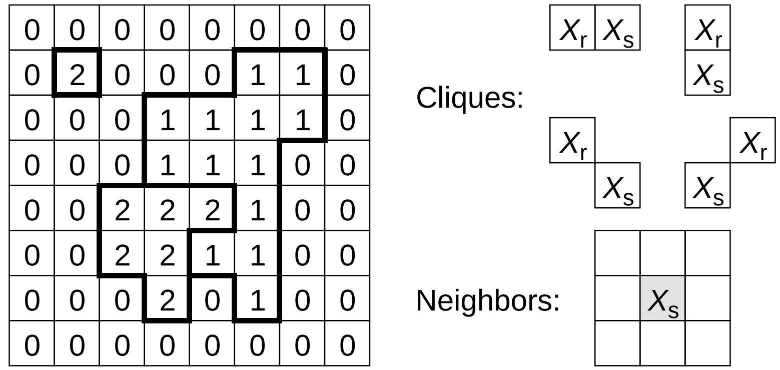
- Amazingly, Onsager was able to compute the following result as $N \rightarrow \infty$.

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



- Onsager also computed an analytic expression for $Z(T)$!

M-Level MRF[3]



- Define $\mathcal{C}_1 \triangleq$ (hor./vert. cliques) and $\mathcal{C}_2 \triangleq$ (diag. cliques)
- Then

$$V(x_r, x_s) = \begin{cases} \beta_1 \delta(x_r \neq x_s) & \text{for } \{x_r, x_s\} \in \mathcal{C}_1 \\ \beta_2 \delta(x_r \neq x_s) & \text{for } \{x_r, x_s\} \in \mathcal{C}_2 \end{cases}$$

- Define

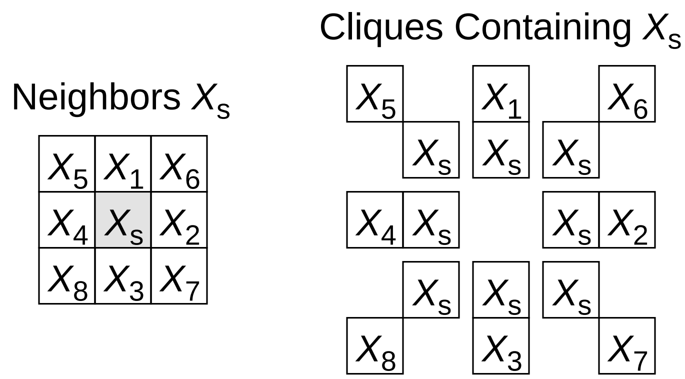
$$t_1(x) \triangleq \sum_{\{s,r\} \in \mathcal{C}_1} \delta(x_r \neq x_s)$$

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

- Then the probability is given by

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

Conditional Probability of a Pixel



- The probability of a pixel given all other pixels is

$$p(x_s|x_{i\neq s}) = \frac{\frac{1}{Z}\exp\left\{-\sum_{c\in\mathcal{C}} V_c(x_c)\right\}}{\sum_{x_s=0}^{M-1}\frac{1}{Z}\exp\left\{-\sum_{c\in\mathcal{C}} V_c(x_c)\right\}}$$

- Notice: Any term $V_c(x_c)$ which does not include x_s cancels.

$$p(x_s|x_{i\neq s}) = \frac{\exp\left\{-\beta_1\sum_{i=1}^{4}\delta(x_s\neq x_i) - \beta_2\sum_{i=5}^{8}\delta(x_s\neq x_i)\right\}}{\sum_{x_s=0}^{M-1}\exp\left\{-\beta_1\sum_{i=1}^{4}\delta(x_s\neq x_i) - \beta_2\sum_{i=5}^{8}\delta(x_s\neq x_i)\right\}}$$

Conditional Probability of a Pixel (Continued)

Neighbors X_s

1	1	0
1	x_s	0
0	0	0

$V_1(0, x_{\partial s}) = 2$ $V_2(0, x_{\partial s}) = 1$

$V_1(1, x_{\partial s}) = 2$ $V_2(1, x_{\partial s}) = 3$

- Define

$$v_1(x_s, \partial x_s) \triangleq \#\text{ of horz./vert. neighbors} \neq x_s$$
$$v_2(x_s, \partial x_s) \triangleq \#\text{ of diag. neighbors} \neq x_s$$

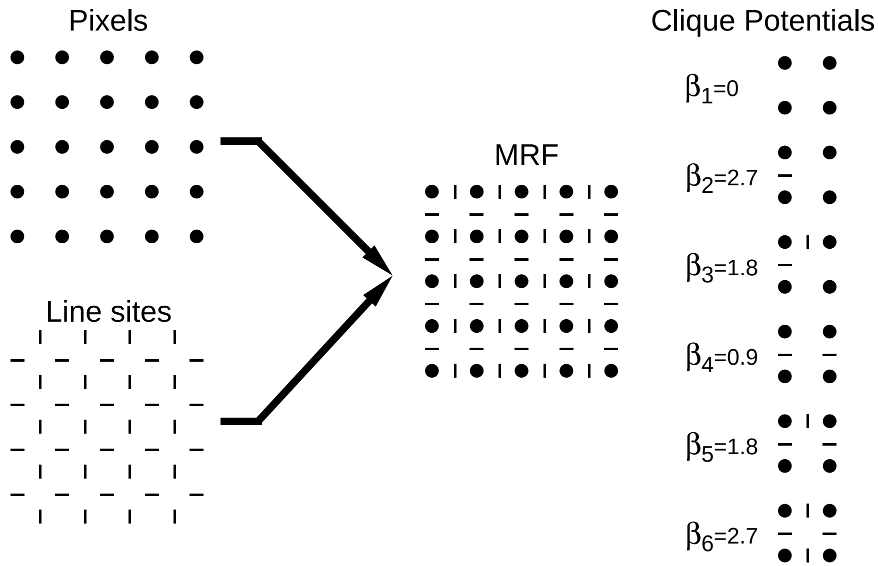
- Then

$$p(x_s|x_{i\neq s}) = \frac{1}{Z'} \exp\left\{-\beta_1 v_1(x_s, \partial x_s) - \beta_2 v_2(x_s, \partial x_s)\right\}$$

where Z' is an easily computed normalizing constant

- When $\beta_1, \beta_2 > 0$, X_s is most likely to be the majority neighboring class.

Line Process MRF [5]



- Line sites fall between pixels
- The values $\beta_1, \cdots, \beta_2$ determine the potential of line sites
- The potential of pixel values is

$$V(x_s, x_r, l_{r,s}) = \begin{cases} (x_s - x_r)^2 & \text{if } l_{r,s} = 0 \\ 0 & \text{if } l_{r,s} = 1 \end{cases}$$

- The field is
  - Smooth between line sites
  - Discontinuous at line sites

Simulation

- Topics to be covered:
  - Gibbs sampler
  - Metropolis sampler
  - Hastings-Metropolis sampler

Generating Samples from a Gibbs Distribution

- How do we generate a random variable X with a Gibbs distribution?

$$p(x) = \frac{1}{Z} \exp\{-U(x)\}$$

- Generally, this problem is difficult.
- Markov Chains can be generated sequentially
- Non-causal structure of MRF's makes simulation difficult.

Gibbs Sampler[5]

- Replace each point with a sample from its conditional distribution

$$p(x_s|x_i^k\ i \neq s) = p(x_s|x_{\partial s})$$

- Scan through all the points in the image.
- Advantage
  - Eliminates need for rejections $\Rightarrow$ faster convergence
- Disadvantage
  - Generating samples from $p(x_s|x_{\partial s})$ can be difficult.

Gibbs Sampler Algorithm

Gibbs Sampler Algorithm:

1. Set $N = \#$ of pixels
2. Order the N pixels as $N = s(0), \cdots, s(N-1)$
3. Repeat for $k = 0$ to ∞
   (a) Form $X^{(k+1)}$ from $X^{(k)}$ via

$$X_r^{(k+1)} = \begin{cases} W & \text{if } r = s(k) \\ X_r^{(k)} & \text{if } r \neq s(k) \end{cases}$$

where $W \sim p\left(x_{s(k)} \middle| X_{\partial s(k)}^{(k)}\right)$

The Metropolis Sampler[10, 9]

- How do we generate a sample from a Gibbs distribution?

$$p(x) = \frac{1}{Z} \exp\{-U(x)\}$$

- Start with the sample x^k, and generate a new sample W with probability $q(w|x^k)$.

  Note: $q(w|x^k)$ must be symmetric.

$$q(w|x^k) = q(x^k|w)$$

- Compute $\Delta E(W) = U(W) - U(x^k)$, then do the following:

  If $\Delta E(W) < 0$

  – Accept: $X^{k+1} = W$

  If $\Delta E(W) \geq 0$

  – Accept: $X^{k+1} = W$ with probability $\exp\{-\Delta E(W)\}$
  – Reject: $X^{k+1} = x^k$ with probability $1 - \exp\{-\Delta E(W)\}$

Ergodic Behavior of Metropolis Sampler

- The sequence of random fields, X^k, form a Markov chain.
- Let $p(x^{k+1}|x^k)$ be the transition probabilities of the Markov chain.
- Then X^k is reversible

$$p(x^{k+1}|x^k)\exp\{-U(x^k)\} = \exp\{-U(x^{k+1})\}p(x^k|x^{k+1})$$

- Therefore, if the Markov chain is irreducible, then

$$\lim_{k\rightarrow\infty} P\{X^k = x\} = \frac{1}{Z}\exp\{-U(x)\}$$

- If every state can be reached, then as $k \rightarrow \infty$, X^k will be a sample from the Gibbs distribution.

Example Metropolis Sampler for Ising Model

	0	
1	x_s	0
	0	

- Assume $x_s^k = 0$.
- Generate a binary R.V., W, such that $P\{W = 0\} = 0.5$.

$$\begin{aligned} \Delta E(W) &= U(W) - U(x_s^k) \\ &= \begin{cases} 0 & \text{if } W = 0 \\ 2\beta & \text{if } W = 1 \end{cases} \end{aligned}$$

If $\Delta E(W) < 0$

- Accept $X_s^{k+1} = W$

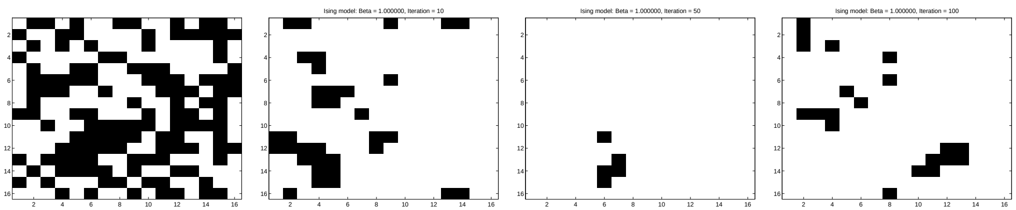
If $\Delta E(W) \geq 0$

- Accept: $X_s^{k+1} = W$ with probability $\exp\{-\Delta E(W)\}$
- Reject: $X_s^{k+1} = x_s^k$ with probability $1 - \exp\{-\Delta E(W)\}$

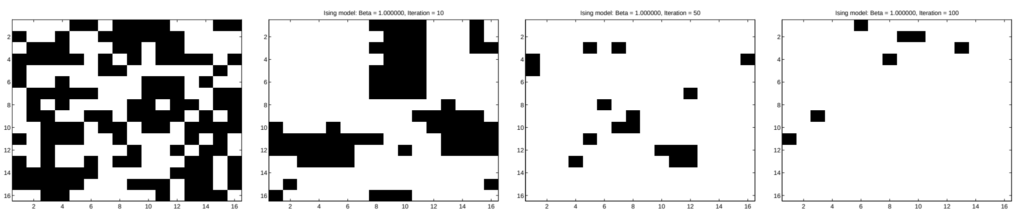
- Repeat this procedure for each pixel.
- **Warning:** for $\beta > \beta_c$ convergence can be extremely slow!

Example Simulation for Ising Model($\beta = 1.0$)

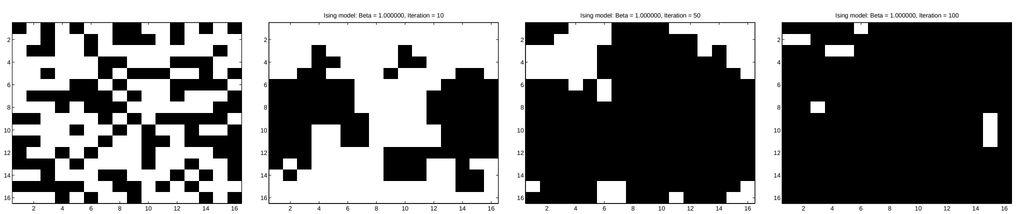
- Test 1



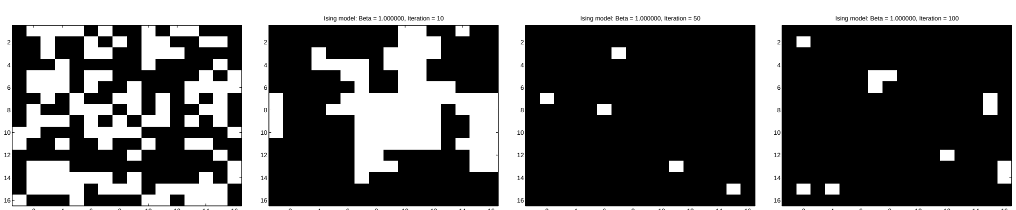
- Test 2



- Test 3



- Test 4



Uniform Random   10 Iterations   50 Iterations   100 Iterations

Advantages and Disadvantages of Metropolis Sampler

- Advantages
  - Can be implemented whenever ΔE is easy to compute.
  - Has guaranteed geometric convergence.
- Disadvantages
  - Can be slow if there are many rejections.
  - Is constrained to use a symmetric transition function $q(x^{k+1}|x^k)$.

Hastings-Metropolis Sampler[8, 14]

- Hastings and Peskun generalized the Metropolis sampler for transition functions $q(w|x^k)$ which are not symmetric.
- The acceptance probability is then

$$\alpha(x_s^k, w) = \min\left\{1, \frac{q(x^k|w)}{q(w|x^k)} \exp\{-\Delta E(w)\}\right\}$$

- Special cases

$$q(w|x^k) = q(x^k|z) \Rightarrow \text{conventional Metropolis}$$
$$q(w_s|x^k) = p(x_s^k|x_{\partial s}^k)\big|_{x_s^k = w_s} \Rightarrow \text{Gibbs sampler}$$

- Advantage
  - Transition function may be chosen to minimize rejections[7]

Parameter Estimation for Discrete State MRF's

- Topics to be covered:
  - Why is it difficult?
  - Coding/maximum pseudolikehood
  - Least squares

Why is Parameter Estimation Difficult?

- Consider the ML estimate of β for an Ising model.
- Remember that

$$t_1(x) = (\#\ \text{horz. and vert. neighbors of different value.})$$

- Then the ML estimate of β is

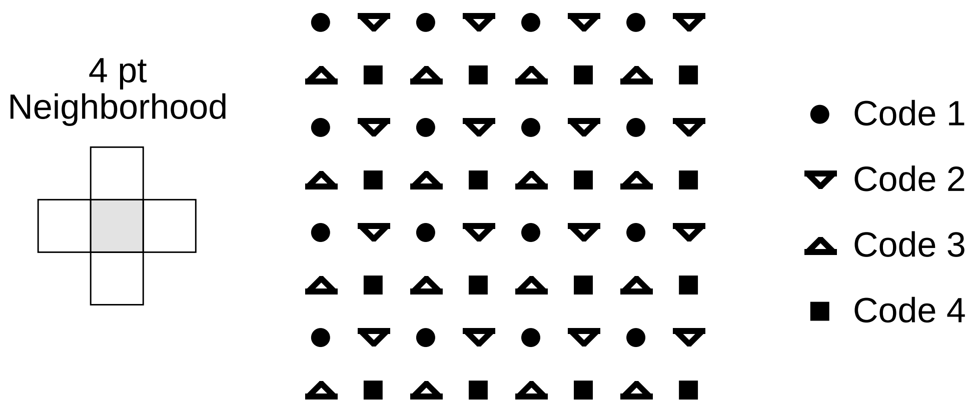
$$\begin{aligned}\hat{\beta} &= \arg\max_{\beta} \left\{ \frac{1}{Z(\beta)} \exp\{-\beta t_1(x)\} \right\} \\ &= \arg\max_{\beta} \{-\beta t_1(x) - \log Z(\beta)\}\end{aligned}$$

- However, $\log Z(\beta)$ has an intractable form

$$\log Z(\beta) = \log \sum_x \exp\{-\beta t_1(x)\}$$

- Partition function can not be computed.

Coding Method/Maximum Pseudolikelihood[2, 3]



- Assume a 4 point neighborhood
- Separate points into four groups or codes.
- Group (code) contains points which are conditionally independent given the other groups (codes).

$$\hat{\beta} = \arg\max_{\beta} \prod_{s \in \text{Code}_k} p(x_s | x_{\partial s})$$

- This is tractable (but not necessarily easy) to compute

Least Squares Parameter Estimation[4]

- It can be shown that for an Ising model

$$\log \frac{P\{X_s = 1 | x_{\partial s}\}}{P\{X_s = 0 | x_{\partial s}\}} = -\beta \left(V_1(1|x_{\partial s}) - V_1(0|x_{\partial s}) \right)$$

- For each unique set of neighboring pixel values, $x_{\partial s}$, we may compute
  - The observed rate of $\log \frac{P\{X_s=1|x_{\partial s}\}}{P\{X_s=0|x_{\partial s}\}}$
  - The value of $(V_1(1|x_{\partial s}) - V_1(0|x_{\partial s}))$
  - This produces a set of over-determined linear equations which can be solved for β.
- This least squares method is easily implemented.

Theoretical Results in Parameter Estimation for MRF's

- Inconsistency of ML estimate for Ising model[15, 16]
  - Caused by critical temperature behavior.
  - Single sample of Ising model cannot distinguish between high β with mean $1/2$, and low β with large mean.
  - Not identifiable
- Consistency of maximum pseudolikelihood estimate[6]
  - Requires an identifiable parameterization.

References

[1] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B*, 36(2):192–236, 1974.

[2] J. Besag. Efficiency of pseudolikelihood estimation for simple Gaussian fields. *Biometrica*, 64(3):616–618, 1977.

[3] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society B*, 48(3):259–302, 1986.

[4] H. Derin and H. Elliott. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-9(1):39–55, January 1987.

[5] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-6:721–741, November 1984.

[6] S. Geman and C. Graffigne. Markov random field image models and their applications to computer vision. In *Proc. of the Intl Congress of Mathematicians*, pages 1496–1517, Berkeley, California, 1986.

[7] P. J. Green and X. liang Han. Metropolis methods, Gaussian proposals and antithetic variables. In P. Barone, A. Frigessi, and M. Piccioni, editors, *Stochastic Models, Statistical methods, and Algorithms in Image Analysis*, pages 142–164. Springer-Verlag, Berlin, 1992.

[8] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[9] R. Kindermann and J. Snell. *Markov Random Fields and their Applications*. American Mathematical Society, Providence, 1980.

[10] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1091, 1953.

[11] L. Onsager. Crystal statistics i. a two-dimensional model. *Physical Review*, 65:117–149, 1944.

[12] R. E. Peierls. On Ising's model of ferromagnetism. *Proc. Camb. Phil. Soc.*, 32:477–481, 1936.

[13] R. E. Peierls. Statistical theory of adsorption with interaction between the adsorbed atoms. *Proc. Camb. Phil. Soc.*, 32:471–476, 1936.

[14] P. H. Peskun. Optimum Monte-Carlo sampling using Markov chains. *Biometrika*, 60(3):607–612, 1973.

[15] D. Pickard. Asymptotic inference for an Ising lattice iii. non-zero field and ferromagnetic states. *J. Appl. Prob.*, 16:12–24, 1979.

[16] D. Pickard. Inference for discrete Markov fields: The simplest nontrivial case. *Journal of the American Statistical Association*, 82:90–96, March 1987.