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?

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

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

Gibbs Sampler[4]

- 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[9, 8]

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

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

- 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

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

- If every state can be reached, then as $k \to \infty$, 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 $X_s^{k+1} = W$

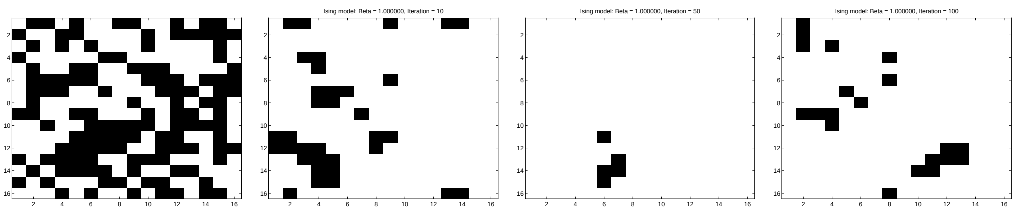
If $\Delta E(W) \geq 0$

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

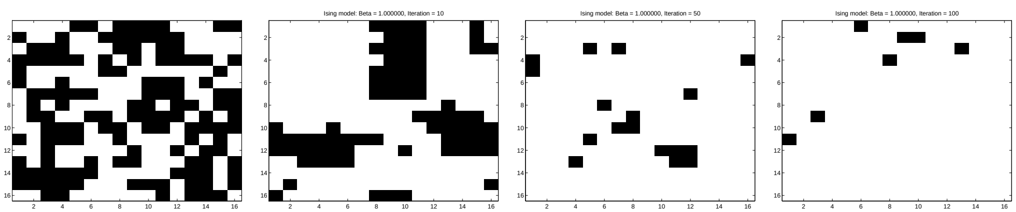
- Repeat this procedure for each pixel.
- **Warning:** for $\beta > \beta_c$ convergence can be extremely slow!

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

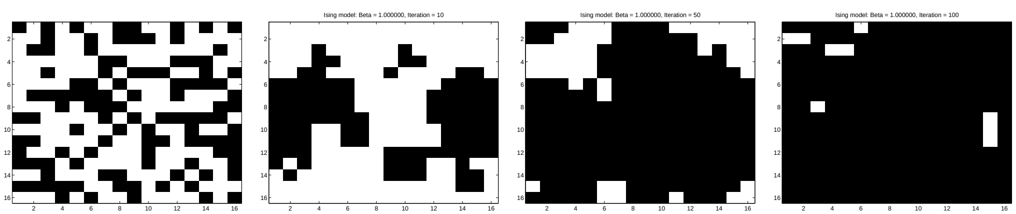
- Test 1



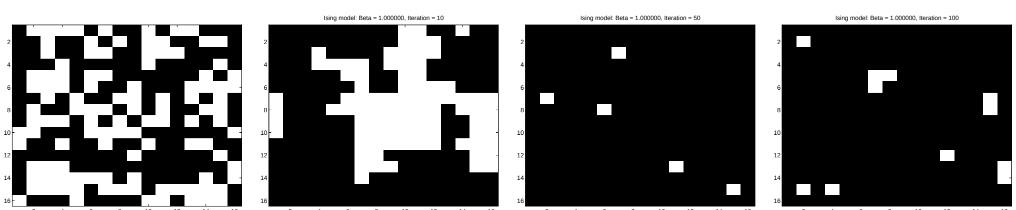
- Test 2



- 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[7, 10]

- 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[6]

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

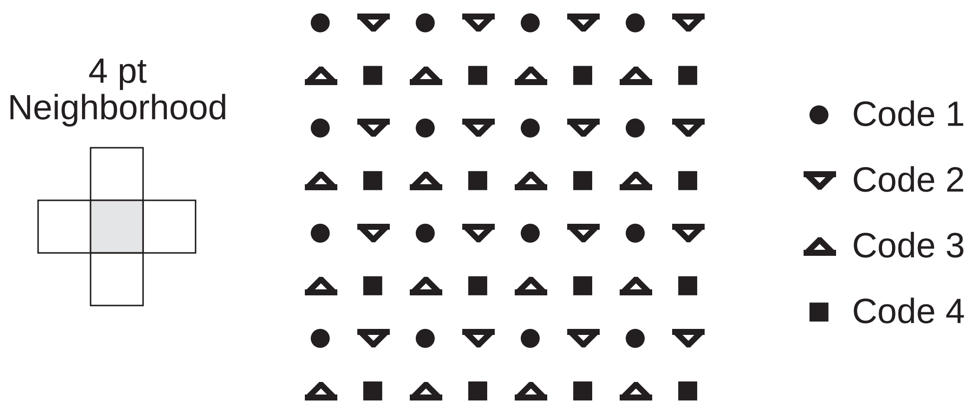
$$\begin{aligned}\hat{\beta} &= \arg\max_{\beta} \left\{ \frac{1}{Z(\beta)} \exp\left\{-\beta t_1(x)\right\} \right\} \\ &= \arg\max_{\beta} \left\{-\beta t_1(x) - \log Z(\beta)\right\}\end{aligned}$$

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

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

- Partition function can not be computed.

Coding Method/Maximum Pseudolikelihood[1, 2]



- 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[3]

- 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[11, 12]
  - 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[5]
  - Requires an identifiable parameterization.

References

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

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

[3] 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.

[4] 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.

[5] 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, CA, 1986.

[6] 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.

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

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

[9] 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.

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

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

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