Lecture 18

Gibbs Sampling

- Project + data POSTED
- Some lecture + Lab time
  - → about Project
- HW 6 posted
- HW 7 will be last HW
- Participation — don't miss the last chances

& Congrats to the many who have already participated A LOT !!

Lecture Notes VIII – Markov Chain Monte Carlo

Marina Meila
mmp@stat.washington.edu

Department of Statistics
University of Washington

April 2025

Notation

Gibbs sampling

The detailed balance

Metropolis-Hastings sampling

Notation

- $V = \{X_1, \ldots X_n\}$ set of random variables (nodes of a graphical model) also known as Markov network
- $X_{1:n} \in \{\pm 1\}$ (for simplicity)
- $E = \{(i, j),\, 1 \le i < j \le n\}$ edges of graph
- graph is not complete, some edges are missing
- We write $i \sim j$ for $(i, j) \in E$ or $(j, i) \in E$
- neigh_i = neighbors of X_i
- **Markov property** $\boxed{X_i \perp \text{all other variables} \,|\, \mathrm{neigh}_i}$
- $\mathrm{x} = (x_1, \ldots x_n) \in \{\pm 1\}^n = S$ an assignment to all variables in V

Notation

- $V = \{X_1, \ldots X_n\}$ set of random variables (nodes of a graphical model) also known as Markov network
- $X_{1:n} \in \{\pm 1\}$ (for simplicity)
- $E = \{(i,j),\ 1 \leq i < j \leq n\}$ edges of graph
- graph is not complete, some edges are missing
- We write $i \sim j$ for $(i,j) \in E$ or $(j,i) \in E$
- neigh_i = neighbors of X_i

- **Markov property** $\boxed{X_i \perp \text{all other variables} \,|\, \mathrm{neigh}_i}$

- $\mathrm{x} = (x_1, \ldots x_n) \in \{\pm 1\}^n = S$ an assignment to all variables in V
- Distribution over S

$$P(\mathrm{x}) = \frac{1}{Z} e^{-\phi(\mathrm{x})} \quad \text{with } \phi(\mathrm{x}) = \sum_{i=1}^{n} h_i x_i + \sum_{(i,j)\in E} h_{ij} x_i x_j \tag{1}$$

(and $h_{ij} > 0$ for all $(i,j) \in E$)
- $Z = \sum_{\mathrm{x}\in S} e^{-\phi(\mathrm{x})}$
- Usually, intractable to compute Z

Wanted samples $\mathrm{x}^1, \mathrm{x}^2, \ldots$ from P

Ising Model

our running example

n variables $\{X_{1:n}\} = V$

$\{(ij)\} = E$ edges

$S = \{\pm 1\}^n$ $\quad x_i \in \{\pm 1\}$

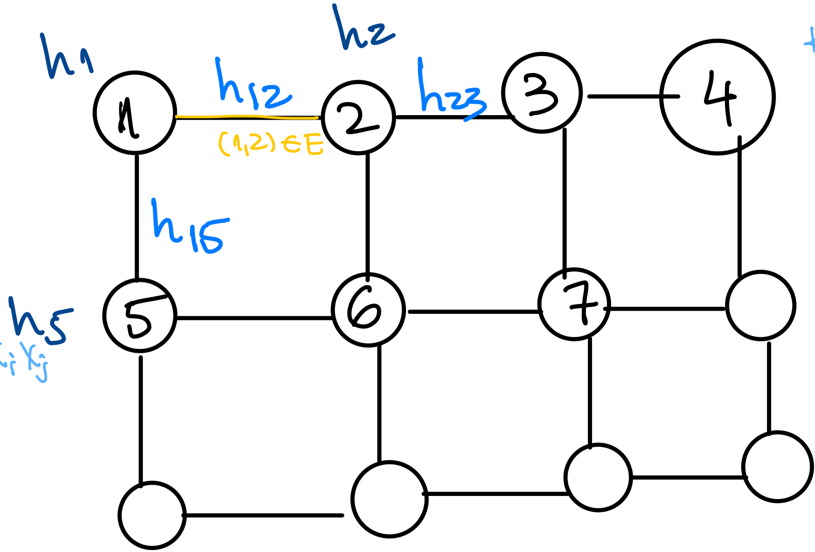
$|S| = 2^n$ $\Rightarrow$ can't calculate $Z = \sum_{x \in S} e^{-\varphi(x)}$ $\rightarrow P(x) = ?$

know $e^{-\varphi(x)}$ unnormalized

energy

$P(x) = \frac{1}{Z} e^{-\varphi(x)}$

$x = (x_1 \dots x_n)$



$$\frac{P(x)}{P(x')} = \frac{e^{-\varphi(x)}}{e^{-\varphi(x')}} = \dots$$

$$\varphi = \sum_{i \in V} h_i x_i + \sum_{(ij) \in E} h_{ij} x_i x_j$$

Want: samples from $P(x)$ iid

Simplify:

$h_1 = h_2 = \dots = 0$

$h_{12} = h_{15} = \dots = 1$

$$\Rightarrow \varphi(x) = \sum_{ij \in E} \underbrace{x_i x_j}_{\pm 1}$$

Gibbs sampling idea

$$P(\mathrm{x}) = \frac{1}{Z} e^{-\phi(\mathrm{x})} \quad \text{with } \phi(\mathrm{x}) = \sum_{i=1}^{n} h_i x_i + \sum_{(i,j)\in E} h_{ij} x_i x_j$$

- We cannot sample directly from P
- But we can sample each $X_i \sim P_{X_i | X_{-i}} = P_{X_i | \mathrm{neigh}_i}$ for any $i = 1 : n$

Pr Xi | all other variables

$P_{x_i | \mathrm{neigh}(x_i)}$

$$\varphi(x) = \sum h_i x_i + \sum h_{ij} x_i x_j$$

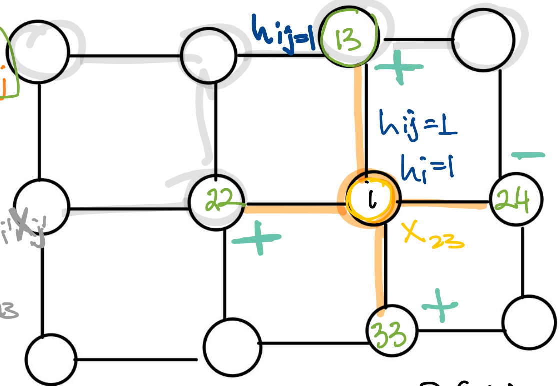
$$= \frac{e^{-\varphi(x)}}{e^{-\varphi(x_{23}=1,\, x_{-23})} + e^{-\varphi(x_{23}=-1,\, x_{-23})}} = \frac{e^{-\left(h_{23} x_{23} + \sum_{j\sim i} h_{ij} x_{23} x_j\right)}}{e^{-(+1)\left[h_{23} + \sum_{j\sim i} h_{ij} x_j\right]} + e^{-(-1)[\quad]}}$$

$$= \frac{e^{-x_{23}\cdot A}}{e^{-A} + e^{+A}}$$

$$h_{23} x_{23} + \sum_{j\sim i} h_{ij} x_{23} x_j$$

$i = x_{23}$

$$+ \sum_{j \neq x_{23}} h_j x_j + \sum_{\substack{(i'j') \in E \\ i',j' \neq x_{23}}} h_{i'j'} x_{i'} x_{j'}$$



$x_{23} = \pm 1$

$$A = \underset{h_i}{1} + (1+1+1-1)$$

$$= 3$$

$$P_{x_{23} \mid x_{22}, x_{24}, x_{33}, x_{13}, x_{11}, x_{12} \,\cdots} = \frac{P(x)}{P(x_{-23})} =$$

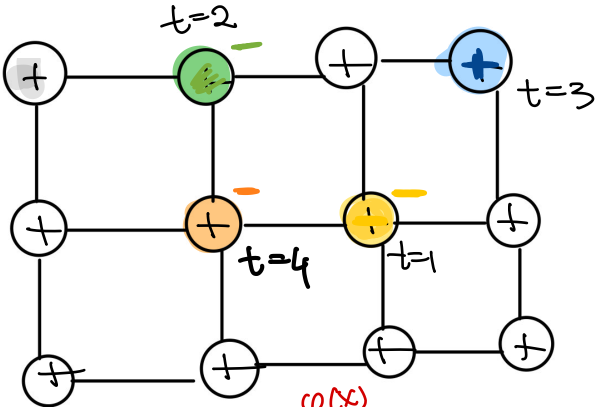
$$\underbrace{P_{V \setminus \{x_{23}\}}}_{-x_{23}} = P(x_{23}=+1, x_{-23}) + P(x_{23}=-1, x_{-23})$$

marginal $\sum_{\text{all} \setminus x_{23}}$

$$P(x_{23}=+1 \mid x_{-23}) = \frac{e^{-3}}{e^{3}+e^{-3}} = \pi_i^+$$

Theorem Ergodic Markov chain $\quad P_X^{(t)} \to P_X^{\infty}$ for any x_0

x^0 initial state



To show: $\quad X^t \sim P^* \propto e^{-\varphi(x)} \leftarrow$ to show $P_X^{\infty} = P^*$

$X^{t'}$ independent

Gibbs sampling idea

$$P(x) = \frac{1}{Z} e^{-\phi(x)} \quad \text{with } \phi(x) = \sum_{i=1}^{n} h_i x_i + \sum_{(i,j)\in E} h_{ij} x_i x_j \quad \leftarrow \text{Ising}$$

- We cannot sample directly from P
- But we can sample each $X_i \sim P_{X_i|X_{-i}} = P_{X_i|\text{neigh}_i}$ for any $i = 1:n$
- Why? For any x_{-i} let

$$\pi_i^+ + \pi_i^- = 1$$

$$\pi_i^+ = Pr[X_i = +|x_{-i}] \quad \pi_i^- = Pr[X_i = -|x_{-i}] \tag{2}$$

$$\pi_i^+ = \frac{P(X_i = +1, x_{-i})}{P(x_{-i})} = \frac{P(X_i = +1, x_{-i})}{P(X_i = +1, x_{-i}) + P(X_i = -1, x_{-i})} \tag{3}$$

$$\pi_i^- = \frac{P(X_i = -1, x_{-i})}{P(X_i = +1, x_{-i}) + P(X_i = -1, x_{-i})} \quad \text{hence} \tag{4}$$

$$\frac{\pi_i^+}{\pi_i^-} = \frac{P(X_i = +1, x_{-i})}{P(X_i = -1, x_{-i})} = \frac{e^{-\phi(X_i=+1, x_{-i})}}{e^{-\phi(X_i=-1, x_{-i})}} = e^{-\phi(X_i=+1, x_{-i}) + \phi(X_i=-1, x_{-i})} \tag{5}$$

$$= e^{-2h_i - 2\sum_{j\sim i} h_{ij} x_j} \tag{6}$$

$$1 = \pi_i^+ + \pi_i^- \quad \text{hence } \pi_i^+ = \frac{e^{-2h_i - 2\sum_{j\sim i} h_{ij}}}{1 + e^{-2h_i - 2\sum_{j\sim i} h_{ij}}} = \frac{e^{-h_i - \sum_{j\sim i} h_{ij}}}{e^{h_i + \sum_{j\sim i} h_{ij}} + e^{-h_i - \sum_{j\sim i} h_{ij}}} \tag{7}$$

- $X_i \mid \text{neigh}_i \sim \text{Bernoulli}(\pi_i^+)$

Gibbs sampling algorithm

1. Initialize x^0 with some arbitrary values
2. For $t = 1, 2, \ldots$ we will sample sequentially $x^t | x^{t-1}$ as follows
   2.1 Pick $i \in 1 : n$ uniformly at random
   2.2 Sample $X_i^t \,|\, \mathrm{neigh}_i \sim \mathrm{Bernoulli}(\pi_i^+)$
   2.3 Every T steps (where T is a LARGE number), output x^t

Q Why does this work?

A1 We are sampling from a Markov chain on S (transition probability matrix P on next page)

A2 If we take enough steps T, the distribution converges to the stationary distribution of this chain, let's call it P. We take a sample from it x^T

A3 If we continue for another T steps, the chain has "forgotten" about x^T; the new sample x^{2T} is independent of x^T. Etc, samples $x^{T, 2T, \ldots NT}$ are i.i.d. from P^∞

TODO To show that $P^\infty = P$ the distribution we wanted to sample from.