STAT 403

5/21/25

Lecture 20

Gibbs – proof

M.H

Lecture Notes VIII – Markov Chain Monte Carlo

Marina Meila
mmp@stat.washington.edu

Department of Statistics
University of Washington

April 2025

Notation

Gibbs sampling

Detailed Balance

The detailed balance

Metropolis-Hastings sampling

Notation

- $V = \{X_1, \dots X_n\}$ 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** $\boxed{X_i \perp \text{all other variables} \,|\, \mathrm{neigh}_i}$

- $\mathrm{x} = (x_1, \dots x_n) \in \{\pm 1\}^n = S$ 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 $\mathrm{x}^1, \mathrm{x}^2, \dots$ from P

Gibbs sampling algorithm

$x^{t+1} \perp x^{t'} \mid x^t$ for all $t' < t$

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 $X_i^t \mid \mathrm{neigh}_i \sim \mathrm{Bernoulli}(\pi_i^+)$
   - 2.3 Every T steps (where T is a LARGE number), output x^t

$\ldots x^t, x^{t+1}, \ldots$

Q Why does this work?

A1 We are sampling from a Markov chain on S (transition probability matrix P on ~~next~~ page 7)

A2 If we take enough steps T, the distribution converges to the stationary distribution of this chain, let's call it ~~P~~ π'. 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 $\underset{\tilde{\pi}}{P^\infty} = P$ the distribution we wanted to sample from.

NEXT →

π' on $S = \{\pm 1\}^n$ is stationary distribution of Gibbs sampling

$\tilde{\pi} = P$

wanted

→ p6 Detailed Balance

→ p7 Transition Prob Gibbs

→ p8 $\dfrac{\pi^+}{\pi^-}$

p5

The detailed balance

Theorem

Let π be a distribution over S, and P a transition matrix of a Markov chain. Then if the following **detailed balance** *holds, π is the stationary distribution of* P*.*

Gibbs

desired distribution

$$\pi(x)P(x,x') = \pi(x')P(x',x) \quad (9)$$

$$\overset{\pi(x)}{x} \underset{P_{x|x'}}{\overset{P_{x'|x}}{\rightleftarrows}} \overset{\pi(x')}{x'}$$

Is it true for Gibbs?

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

1) $x_{-i} \neq x'_{-i}$ for any i

differ in ≥ 2 variables

$$\pi(x) \cdot 0 = \pi(x') \cdot 0 \checkmark$$

2) $x = x'$ $\checkmark$

3) $x_{-i} = x'_{-i}$ for some i

$x_i \neq x'_i$

$x_i = 1$

$x'_i = -1$

$$\cancel{\frac{1}{Z}} e^{-\varphi(x)} \cdot \pi^{-}(x_{-i}) = \cancel{\frac{1}{Z}} e^{-\varphi(x')} \pi^{+}(x_{-i})$$

$$\frac{\pi^{+}(x_{-i})}{\pi^{-}(x_{-i})} = \frac{\varphi(x)}{\varphi(x')}$$

?!

The transition probability P of Gibbs sampling

- The transitions are between states x, x' that only differ in one variable i. All the other transition probabilities are 0.
- If x^t and x' differ only in variable i, then

$\Downarrow$

$x^t_{-i} = x'_{-i}$

$$P(x^{t+1} = x'|x) \; = \; \begin{cases} \pi_i^+(x_{-i}) & x_i' = +1 \\ \pi_i^-(x_{-i}) & x_i' = -1 \end{cases} \tag{8}$$

Gibbs sampling idea

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

- 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

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

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

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

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

$\Rightarrow$ Detailed Balance Holds!

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

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

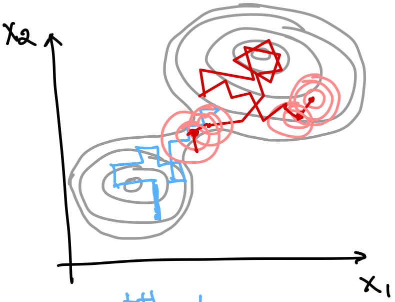
- $X_i \mid \text{neigh}_i \sim \text{Bernoulli}(\pi_i^+)$

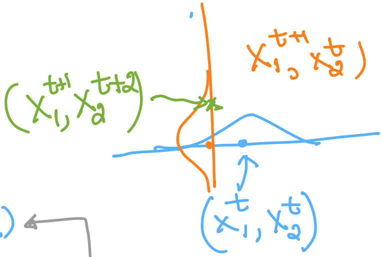
plug into #6

Example

MH $\quad S(x, x')$ proposal distribution

$x = (x_1, x_2)$





Gibbs

$$x_1^{t+1} \mid x_2^{t} \sim P(x_1 \mid x_2)$$

$$x_2^{t+2} \mid x_1^{t+1} \sim P(x_2 \mid x_1)$$

$$\vdots$$

$$P(x_{1,2}) = \frac{1}{Z} e^{-\varphi(x_1, x_2)}$$

Metropolis-Hastings (MH) idea

- MH is a **rejection sampling** algorithm
- We sample $x' \mid x^{t-1} \sim S$ a **proposal distribution**
- Then we **accept** $x^t = x'$ with some **acceptance probability** $a(x, x')$ that ensures the detail balance
- (if we don't accept, $x^t = x^{t-1}$)

- With MH, we have more flexibility in exploring the sample space around x^{t-1} than with Gibbs

Metropolis-Hastings algorithm

In Proposal distribution $\boxed{S(x, x') \propto \text{transition probability } x \to x'}$
no need to be normalized, no need to be symmetric

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 Sample $x' \sim S(x^{t-1}, x')$ ← propose
   - 2.2 Compute acceptance probability

$$a(x^{t-1}, x') = \min\left(1, \frac{P(x')S(x', x^{t-1})}{P(x^{t-1})S(x^{t-1}, x')}\right). \tag{10}$$

   - 2.3 $x^t = \begin{cases} x' & \text{w.p. } a \\ x^{t-1} & \text{w.p. } 1-a \end{cases}$ accept & move ← reject & stay
   - 2.4 Every T steps (where T is a LARGE number), output x^t

P = desired stationary distribution

S = proposed transition

Does it satisfy the detailed balance?

$(x^t \rightarrow x^{t+1})$
$x \rightarrow x'$
transitioning

- If proposal x' rejected ✓ $\Rightarrow x^t = x^{t-1}$ ✓
- If x' accepted

$\Pr[x \rightarrow x']$ =

x' proposed | x

accepted

$$P(x', x) = S(x', x)a(x', x) \tag{11}$$

$$P(x)P(x, x') = P(x)S(x, x')\min\left(1, \frac{P(x')S(x', x)}{P(x)S(x, x')}\right) \tag{12}$$

$$= \boxed{\min\left(P(x')S(x', x), P(x)S(x, x')\right)} \tag{13}$$

$$= P(x')P(x', x) \qquad \text{by symmetry} \tag{14}$$

detailed balance left hand

detailed bal. right hand ⇒ same !! expression