STAT 403 5/19/25

Lecture 19

>> Don't miss
last lecture <<

Gibbs sampling
in practice

Lecture Notes VIII – Markov Chain Monte Carlo

Marina Meila
mmp@stat.washington.edu

Department of Statistics
University of Washington

April 2025

Notation ✓

Gibbs sampling ← mixing - in practice sampling subsets of variables

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

$x^t = \text{state at } t \in S$

$|S| = 2^n$

$x = (x_{1:n})$

1. Initialize $\boxed{x^0}$ 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

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.

M.C Ergodic $\Rightarrow$ $x^{t+T} \sim P_x^\infty$ stationary distribution

$x^{t+T} \perp x^t$ for T large

More detail

T_0 = # burn-in steps (time)

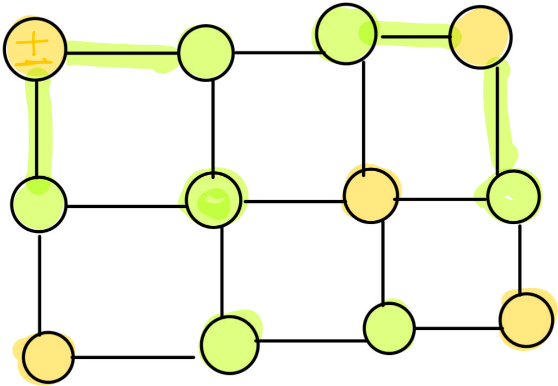
$x_0 \cdots x^{T_0}$ = first sample

$x_0 \longrightarrow P^\infty$

T = # sampling steps (time)

$x^t \sim P^\infty$

(assumption) $\Longrightarrow t > T_0$

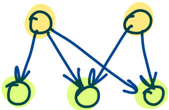


$T_0 > T$ in practice

T = mixing time

$\lambda_2(\underbrace{[P_{j|i}]}_{\text{transition matrix}}) \approx 0$

Bayes network

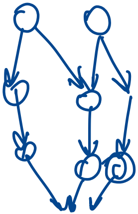


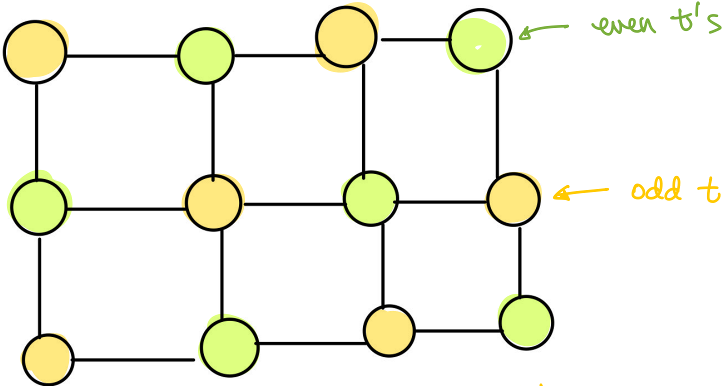
easy

hard computationally

Restricted Boltzmann Machine

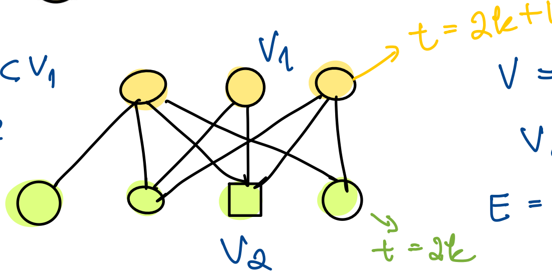
sequential





$i \in V_1 \Rightarrow$
$\text{neigh}_i \subset V_2$
$j \in V_2 \Rightarrow \text{neigh}_j \subset V_1$

Bipartite graph



$V = V_1 \cup V_2$

$V_1 \cap V_2 = \phi$

$E = \{(i,j) \,,\, i \in V_1, j \in V_2\}$