STAT 403

5/14/25

Lecture 17

Using Cond. Prob. in M.C., graphical models intro

L VIII posted
M.Chain refresher

Markov Chain Monte Carlo

$X_{1,2,\dots} \in \{\pm 1\}$ or other S discrete

goal: answer statistical question by MC (= estimate θ)

when: large systems, multivariate p, or P

Markov Chain

$X_1 - X_2 - X_3 - X_4$

$X_1 - X_2 - \cdots - X_{t-1} - X_t - X_{t+1} - \cdots X_T$

Independence properties — Markov properties

$X_t \perp X_{1:t-2} \mid \underline{\underline{X_{t-1}}}$

$X_t \perp X_{t+2:T} \mid \underline{\underline{X_{t+1}}}$

$X_t \perp \underset{X_{t'}}{\text{all other}} \mid \underbrace{X_{t-1}, X_{t+1}}_{\text{Markov blanket of } X_t}$ $\quad t' \neq t, t\pm 1$

Transition Matrix $\underline{P}$

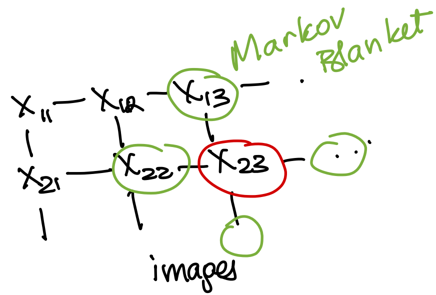
$X_t \in S$ discrete finite for $t = 1:T$

- $P[X_1]$ initial state probability
- $P_{ij} = [X_{t+1} = j \mid X_t = i]$

$P = [P_{ij}]_{i,j=1:m}$

$|S| = m$

$\sum_{j=1}^{m} P_{ij} = 1 \Leftrightarrow P$ is stochastic matrix



$\underline{P_{j|i}} \equiv \boxed{P_{ij}}$ transition matrix $= P[X_j | X_i] = P_{X_j|X_i}$

$P_{X_i} \equiv P_i$ marginal prob of X_i

$P_{X_i^t} =$ —— " —— X_i^t

$P_{X_i X_j} =$ joint marginal of (X_i, X_j)

Probabilities in M.C.

Joint distribution

$$P_{X_1 \cdots X_T}(x_1, \cdots x_T) = P_{X_1} \cdot \underbrace{P_{X_2 \cdots X_T | X_1}}_{P_{X_2|X_1} \cdot P_{X_3 \cdots X_T | X_1 X_2}} = P_{X_1} P_{X_2|X_1} P_{X_3 \cdots X_T | X_1 X_2} = P_{X_1} P_{X_2|X_1} P_{X_3|X_1 X_2} P_{X_4 \cdots X_T | X_1 X_2 X_3} =$$

$$= P_{X_1} \cdot \prod_{t=1}^{T-1} \underbrace{P_{X_{t+1} | X_t X_{t-1} \cdots X_1}}_{P_{X_{t+1}|X_t}}$$

Markov

$(x_1 \cdots x_T) = (+1, +1, -1, -1, +1)$

$+ \; + \; - \; - \; +$

Chain Rule

Conditional Prob.

$P_{X_T | X_1 \cdots X_{T-1}} = P_{X_T | X_{T-1}}$

$P_{X_t | X_1 \cdots X_{t-1}} = P_{X_t | X_{t-1}}$

$$P_{X_t | X_1 \cdots X_{t-1}, X_{t+1}, \cdots X_T} = \underline{P_{X_t | X_{t-1}, X_{t+1}}} = \frac{P_{X_t, X_{t+1} | X_{t-1}}}{P_{X_{t+1} | X_{t-1}}}$$

chaine

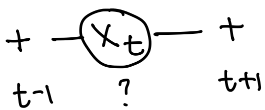
$$P_{X_t, X_{t+1} | X_{t-1}} \to \underbrace{P_{X_t | X_{t-1}}}_{} \underbrace{P_{X_{t+1} | X_t}}_{}{}_{, \cancel{X_{t-1}}}$$

$$P_{X_{t+1} | X_{t-1}} \to P^{(\cdot, +)}_{X_{t+1} X_t | X_{t-1}} + P^{(\cdot, -)}_{X_{t+1} X_t | X_{t+1}} =$$

Total prob

$$= P_{X_{t+1} | X_t = +} P^{(+)}_{X_t | X_{t-1}} + P_{X_{t+1} | X_t = -} \cdot P_{X_t = - | X_{t-1}}$$

$t+1$	$+$	$-$
$t=+$	0.8	0.2
$-$	0.2	0.8



$+ \; - \; ? \; - \; +$

$+ \rightarrow + \rightarrow +$

0.8

$+ \rightarrow - \rightarrow +$

0.2 0.2

$$P_{X_t=+|X_{t-1}=+} \, P_{X_{t+1}=+|X_t=+} = 0.8 \times 0.8 = 0.64$$

$$P_{X_t=-|X_{t-1}=+} \, P_{X_{t+1}=+|X_t=-} = 0.2 \times 0.2 = 0.04$$

$$P_{X_t=+|X_{t-1}=X_{t+1}=+} = \frac{0.64}{0.64+0.04} = \frac{16}{17}$$

$$P_{X_t=-|\ldots\ldots} = \frac{1}{17}$$

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

- $V = \{X_1, \dots X_n\}$ 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

$V = \{\text{nodes}\}$

$E = \{\text{edges}\}$

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

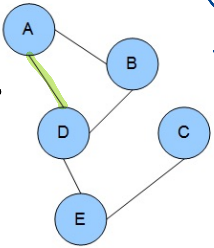
- $\mathrm{x} = (x_1, \dots x_n) \in \boxed{\{\pm 1\}^n = S}$ an assignment to all variables in V

$|S| = 2^n$

- P_{ABCDE} = joint ?
- P_A = ? marginal
- $P_{A|BC}$ = ? conditional

MCMC: answers by sampling

$A \sim B$



$V = \{A, B, \dots D\}$

$E = \{AB, AD, BD, CE, DE\}$

$E \sim C$
$C \sim E$

$A \not\sim E$

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

$$\varphi = h_A + h_D - h_C - h_B - h_E$$
$$+ h_{AD} - h_{BD} - h_{AB}$$
$$+ h_{CE} - h_{DE}$$

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

$$\boxed{P \propto e^{-\varphi}}$$

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

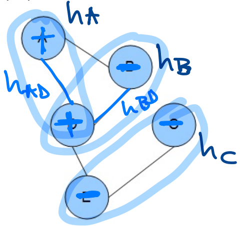
"Energy"

Ising Model

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

(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



S too large

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

WITHOUT Z

~~An example~~

What can we do without Z?

$$P(x) \propto e^{-\varphi(x)}$$

can't know $P(x)$

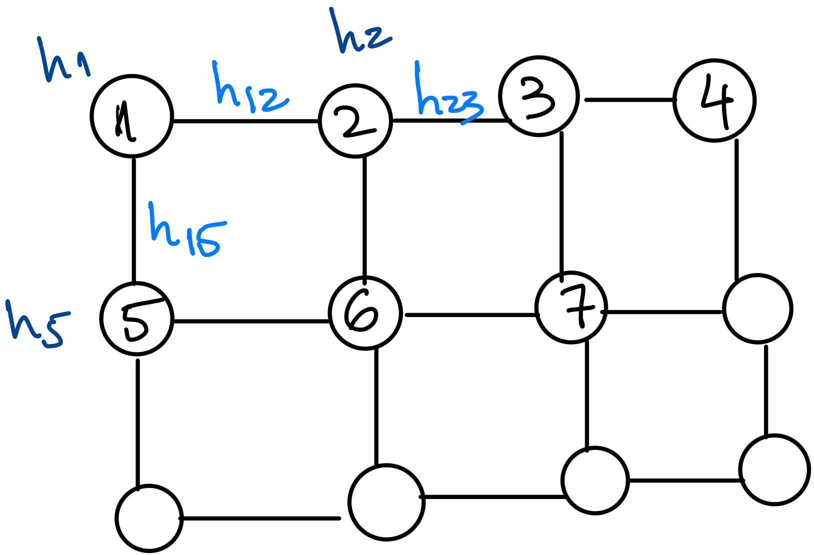
$$\frac{P(x)}{P(x')} = e^{-\varphi(x)+\varphi(x')}$$

can compute ratios

(Unnormalized P)

Ising Model

our running example



Simplify :

$h_1 = h_2 = \dots = 0$

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

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

HW 6 graph