Lecture 24: 2 May, 2022

Madhavan Mukund

https://www.cmi.ac.in/~madhavan

Data Mining and Machine Learning
January–May 2022

Exact inference in Bayesian networks is hard

Hence - approximate inference - generate samples, count

- Topological order sample generation
  Many useless samples

- Rejection sampling

- Likelihood sampling
  Fix the evidence, use likelihood weighted

$$P(x \mid y, z) = \frac{P(x, y, z)}{P(y, z)}$$

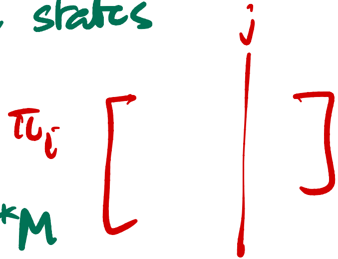
Markov Chain Monte Carlo (MCMC) methods

Markov chains

State transition matrix M - stochastic

Given some distribution π_i over states

$$\pi_{i+1} = \pi_i M$$



Stationary distribution $\pi^* = \pi^* M$

Ergodic ⇒ Stationary distribution

- Strongly connected – irreducible
- aperiodic

x^y

$x \; (x^2)^2$

Stationary distribution

- Unique soln of $x = xM$
- Reach π^* as $\lim_{n \to \infty} \pi_0 M^n$ – "mixing"
- $\lim_{n \to \infty} \frac{1}{n} (\pi_0 M^n)$

Markov Chain Monte Carlo (MCMC) methods

Interpretation of stationary distribution

$\pi^* = [p_1 \; p_2 \; \dots \; p_N]$ N states

In a "long" run of length M, s_i is visited $p_i M$ times

How do we use this for sampling?

Markov Chain Monte Carlo (MCMC) methods

Bayesian network, variables $v_1, v_2, \ldots, v_n$

Each configuration of v_is is a state

States $s_1, \ldots, s_N$

Associated $p_1, \ldots, p_N$ – probabilities we want to calculate

Set up a Markov chain on $s_1 \ldots s_N$, ergodic

s.t. $\pi^* = [p_1 \; p_2 \; \ldots \; p_N]$

Reversible Markov Chain (defn)

$$\pi_j P_{jk} = \pi_k P_{kj}$$

Moves $j \to k$ in limit

Moves $k \to j$ in limit

$M = [\ \ ]$

$M[j,k] = P_{jk}$

$\pi^* = [\pi_1 \ldots \pi_N]$

Why "reversible"?

$$\forall n \quad P[X_{n-1}=j \mid X_n=k] = P[X_n=j \mid X_{n-1}=k]$$

Markov Chain Monte Carlo (MCMC) methods

$$P[X_{n-1}=j \mid X_n=k] = P[X_n=k \mid X_{n-1}=j] \cdot \frac{P[X_{n-1}=j]}{P[X_n=k]}$$

$$P_{kj} = P_{jk} \cdot \frac{\pi_j}{\pi_k}$$

$$\pi_k P_{kj} = \pi_j P_{jk}$$

Markov Chain Monte Carlo (MCMC) methods

Suppose $\exists\ \underline{a_1, a_2, \ldots, a_N}$ s.t. (a)

$$a_j p_{jk} = a_k p_{kj}$$

$$\sum_k a_j p_{jk} = \sum_k a_k p_{kj}$$

$$a_j \cdot 1 =$$

$$a = a \cdot M$$

Given a non-reversible MC — make it reversible

Assume we know π

Construct a new Markov Chain

P_{jk} : Choose $k \neq j$ uniformly $\frac{1}{N-1}$

(a) $\pi_k \geq \pi_j$ — move to k with prob 1

(b) $\pi_k < \pi_j$ — move to k with $\frac{\pi_k}{\pi_j}$

— stay in j $1 - \frac{\pi_k}{\pi_j}$

Markov Chain Monte Carlo (MCMC) methods

$$P_{jk} = \begin{cases} \frac{1}{N-1} \min\left(\frac{\pi_k}{\pi_j}, 1\right) & , k \neq j \\ 1 - \sum_{l \neq j} P_{jl} & , k = j \end{cases}$$

$$P_{jk} = \frac{1}{N-1} \min\left(\frac{\pi_k}{\pi_j}, 1\right)$$

$k \neq j$

$$\pi_j \cdot P_{jk} = \frac{1}{N-1} \min(\pi_k, \pi_j)$$

Markov Chain Monte Carlo (MCMC) methods

$$\pi_j \cdot P_{jk} = \frac{1}{N-1} \min(\pi_k, \pi_j)$$

$$= \pi_k \frac{1}{N-1} \min\left(1, \frac{\pi_j}{\pi_k}\right)$$

$$= \pi_k P_{kj}$$

Metropolis Hastings Algorithm

Suppose $\pi_j P_{jk} \neq \pi_k P_{kj}$

Assume $\pi_j P_{jk} > \pi_k P_{kj}$

Add acceptance probabilities α_{jk}

$$\pi_j \cdot P_{jk} \cdot \alpha_{jk} = \pi_k \cdot P_{kj} \cdot \alpha_{kj}$$

Assume 1, wlog

Markov Chain Monte Carlo (MCMC) methods

$$\alpha_{jk} = \frac{\pi_k P_{kj}}{\pi_j P_{jk}}$$

In general $\alpha_{jk} = \min\left(\frac{\pi_k P_{kj}}{\pi_j P_{jk}}, 1\right)$

Gibbs Sampling

State is a valuation of $(v_1, v_2, \ldots, v_n)$

$s = (x_1, \ldots, x_n) \rightarrow s' = (y_1, \ldots, y_n)$

Move to a neighbor that differs in only one x_i

$(x_1, x_2, \ldots, x_i, \ldots, x_n)$

$\downarrow$

$(x_1, x_2, \ldots, x_i', \ldots x_n)$

Gibbs sampling

Current state is $s = (x_1, \ldots, x_n)$

Choose i uniformly

Resample x_i $\quad P(x_i \mid x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$

Subsumes Markov Blanket of x_i

$s = (\ldots, x_i, \ldots)$

$\downarrow$

$s' = (\ldots, x_i', \ldots)$

choose i

$\frac{1}{n} \cdot P(x_i' \mid \underbrace{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots x_n}_{\bar{x}_i})$

$= \frac{P(s_i')}{P(\bar{x})}$

$P_{k\ell} = \frac{1}{n} \frac{P(s_\ell)}{P(\bar{x})} = \pi_\ell$

$P_{\ell k} = \frac{1}{n} \frac{P(s_k)}{P(\bar{x})} = \pi_k$

Metropolis Hastings α_{jk}

$\min \left(\frac{\pi_k P_{kj}}{\pi_j P_{jk}} \frac{\pi_j}{\pi_k}, 1 \right)$

Markov Chain Monte Carlo (MCMC) methods

Gibbs Sampling = Metropolis Hastings with multiple !