Sampling Based Inference

Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

Weekly Objectives

- Learn basic sampling methods
  - Understand the concept of Markov chain Monte Carlo
  - Able to apply MCMC to the parameter inference of Bayesian networks
  - Know the mechanism of rejection sampling
  - Know the mechanism of importance sampling
- Learn sampling based inference
  - Understand the concept of Metropolis-Hastings Algorithm
  - Know the mechanism of Gibbs sampling
- Know a case study of sampling based inference
  - Understand the latent Dirichlet allocation model
  - Know the collapsed Gibbs sampling
  - Know how to derive Gibbs sampling formula for LDA

SAMPLING BASED INFERENCE

Detour: EM Algorithm

$$l(\theta) = \ln P(X|\theta) = \ln\left\{\sum_Z q(Z)\frac{P(X,Z|\theta)}{q(Z)}\right\} \geq \sum_Z q(Z)\ln\frac{P(X,Z|\theta)}{q(Z)} = Q(\theta,q)$$

$$Q(\theta,q) = E_{q(Z)}\ln P(X,Z|\theta) + H(q)$$

$$L(\theta,q) = \ln P(X|\theta) - \sum_Z \{q(Z)\ln\frac{q(Z)}{P(Z|X,\theta)}\}$$

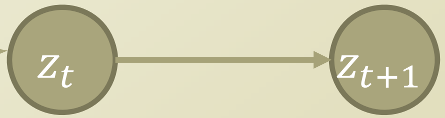
- EM algorithm
  - Finds the maximum likelihood solutions for models with latent variables
  - $P(X|\theta) = \sum_Z P(X,Z|\theta) \rightarrow \ln P(X|\theta) = \ln\{\sum_Z P(X,Z|\theta)\}$
- EM algorithm
  - Initialize θ^0 to an arbitrary point
  - Loop until the likelihood converges
    - Expectation step
      - $q^{t+1}(z) = argmax_q Q(\theta^t,q) = argmax_q L(\theta^t,q) = argmin_q KL(q||P(Z|X,\theta^t))$
      - → $q^t(z) = P(Z|X,\theta)$→**Assign Z by $\boldsymbol{P(Z|X,\theta)}$**
    - Maximization step
      - $\theta^{t+1} = argmax_\theta Q(\theta,q^{t+1}) = argmax_\theta L(\theta,q^{t+1})$
      - → fixed Z means that there is no unobserved variables
      - → Same optimization of ordinary MLE

Computing Expectation…. Sometimes, it can be hard

Markov Chain

- Markov chain
  - Each node has a probability distribution of states
    - i.e.) The probability that a state is the current state of a system
      - Concrete observation of a system: [1 0 0] → the system is at the first state
      - Stochastic observation of a system: [0.7 0.2 0.1] → the system is likely at the first state
    - The node has a vector of state probability distribution
  - Each link suggests a probabilistic state transition
    - If a system is at the first state, the probability distribution of the next state is [0.3 0.4 0.3]
    - The link has a matrix of state transition probability distribution.

$z_t = [0.5 \quad 0.2 \quad 0.3]$



$$P(z_{t+1}) = P(z_t)P(z_{t+1}|z_t) = z_t T_{i,j}$$

$$= [0.5 \quad 0.2 \quad 0.3]\begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.3 & 0.5 \\ 0.4 & 0.2 & 0.4 \end{bmatrix}$$

$$= [0.51 \quad 0.22 \quad 0.27]$$

- The system has three states, a, b, and c.
- Transition matrix is

$$\mathrm{P}(z_j|z_i) = T_{i,j} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.3 & 0.5 \\ 0.4 & 0.2 & 0.4 \end{bmatrix}$$

Properties of Markov Chain

- Accessible
  - ***i→j:*** State ***j*** is ***accessible*** from ***i*** if $T_{i,j}^{k} > 0\ and\ k \geq 0$
  - ***i*** $\leftrightarrow$ ***j***: State ***i*** and ***j*** ***communicate*** if ***i→j*** and ***j→i***
- Reducibility
  - A Markov chain is ***irreducible*** if $\boldsymbol{i} \leftrightarrow \boldsymbol{j}, \forall \boldsymbol{i} \in \boldsymbol{S}, \forall \boldsymbol{j} \in \boldsymbol{S}$
- Periodicity
  - State ***i*** has ***period d*** if $d = \gcd\{n: T_{i,j}^{n} > 0\}$
  - If ***d***=1, State ***i*** is ***aperiodic***.
- Transience
  - State ***j*** is ***recurrent*** if $\mathrm{P}(\inf(t \geq 1: X_t = j) < \infty | X_0 = j) = 1$
  - States which are not ***recurrent*** are ***transient***.
- Ergodicity
  - A state is ***ergodic*** if the state is (positive) ***recurrent*** and ***aperiodic***.
  - Markov chain is ergodic if all states are ergodic.

Stationary Distribution

- $RT_i = \min\{n > 0: X_n = i | X_0 = i\}$
  - Return time to state ***i*** after the departure from state ***i***
- Limit theorem of Markov chain
  - A friend in ISE dept. told me.....
  - If a Markov chain is irreducible and ergodic
    - $\pi_i = \lim_{n\to\infty} T_{i,j}^{(n)} = \frac{1}{E[RT_i]}$
    - π_i is uniquely determined by the set of equations
      - $\pi_i \geq 0, \sum_{i\in S} \pi_i = 1, \pi_j = \sum_{i\in S} \pi_i T_{i,j}$
  - How to compute $\boldsymbol{\pi}$ given $\boldsymbol{T}$
    - $\pi\left(I_{|S|,|S|} - T + 1_{|S|,|S|}\right) = 1_{1,|S|}$
      - $\pi_j = \sum_{i\in S} \pi_i T_{i,j} \rightarrow \pi_j - \sum_{i\in S} \pi_i T_{i,j} = 0 \rightarrow \pi\left(I_{|S|,|S|} - T\right) = 0$
        - To the above formula, apply $\sum_{i\in S} \pi_i = 1 \rightarrow \pi 1_{|S|,|S|} = 1_{1,|S|}$ to both sides
      - $\pi\left(I_{|S|,|S|} - T + 1_{|S|,|S|}\right) = 1_{1,|S|}$
  - Here, $\boldsymbol{\pi}$ is the stationary distribution!

```
>> T

T =

    0.7000    0.2000    0.1000
    0.2000    0.3000    0.5000
    0.4000    0.2000    0.4000

>> pi = ones(1,3) / (eye(3,3)-T+ones(3,3))

pi =

    0.5079    0.2222    0.2698

>> pi*T

ans =

    0.5079    0.2222    0.2698

>> pi(1)*T(1,2)

ans =

    0.1016

>> pi(2)*T(2,1)

ans =

    0.0444
```

Irreversible MC

```
>> T2 = [0 0.5 0.5 ; 0.25 0.5 0.25; 0.25 0.

T2 =

         0    0.5000    0.5000
    0.2500    0.5000    0.2500
    0.2500    0.2500    0.5000

>> pi2 = ones(1,3)/(eye(3,3)-T2+ones(3,3))

pi2 =

    0.2000    0.4000    0.4000

>> pi2*T2

ans =

    0.2000    0.4000    0.4000

>> pi2(1)*T2(1,2)

ans =

    0.1000

>> pi2(2)*T2(2,1)

ans =

    0.1000
```

Reversible MC

$\boldsymbol{\pi}$ is the stationary distribution

Reversible Markov chain

$\pi_i T_{i,j} = \pi_j T_{j,i}$

Detailed Balance, or **Balance Equation**