Sampling Based Inference

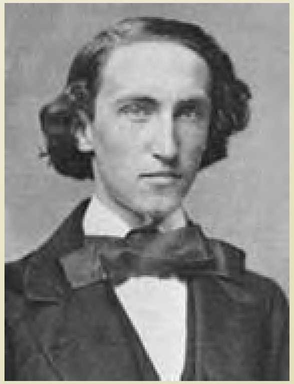
Il-Chul Moon
Dept. of Industrial and Systems Engineering
KAIST
icmoon@kaist.ac.kr

Weekly Objectives

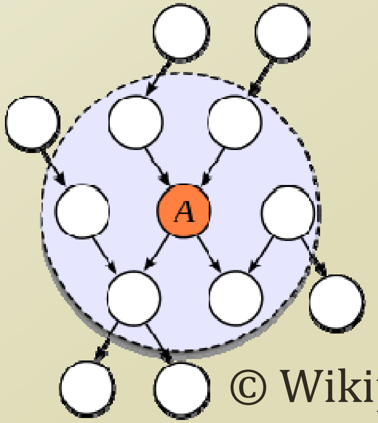
- Learn basic sampling methods
  - Understand the concept of Markov chain Monte Carlo
  - Able to apply MCMC to the parameter inference of Bayesian networks
  - Know the mechanism of rejection sampling
  - Know the mechanism of importance sampling
- Learn sampling based inference
  - Understand the concept 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

Gibbs Sampling

Josiah Willard Gibbs
(1839 - 1903)
- physicist
© Wikipedia



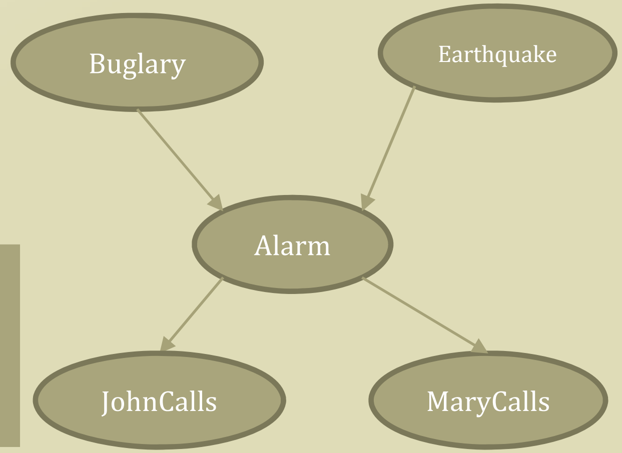
- Gibbs Sampling: A special case of M-H algorithm
  - Let's suppose $z^t = (z_k^t, z_{-k}^t) \rightarrow z^* = (z_k^*, z_{-k}^t)$
    - $T_{t,*}^{MH} = q(z^*|z^t)\alpha(z^*|z^t)$
    - $q(z^*|z^t) = P(z_k^*, z_{-k}^t|z_{-k}^t) = P(z_k^*|z_{-k}^t)$
  - Let's observe the balance equation
    - Should hold $\mathrm{P}(z^t)q(z^*|z^t) = \mathrm{P}(z^*)q(z^t|z^*)$
    - $\mathrm{P}(z^t)q(z^*|z^t) = \mathrm{P}(z_k^t, z_{-k}^t)P(z_k^*|z_{-k}^t) =$
      $\mathrm{P}(z_k^t|z_{-k}^t)P(z_{-k}^t)P(z_k^*|z_{-k}^t)$
      $= \mathrm{P}(z_k^t|z_{-k}^t)P(z_k^*, z_{-k}^t) = q(z^t|z^*)P(z^*)$
    - Always hold the balance equation!
  - Then, the acceptance probability becomes $\alpha(z^*|z^t) = 1$



© Wikipedia

- Example of Gibbs sampling
  - When the joint distribution is not known explicitly or is difficult to sample from directly, but the conditional distribution of each variable is known and is easy
    - P(E,JC,B|A=F,MC=T)=?
      - Hard to sample directly. Why?
    - Consider a conditional distribution $p(z_i|z_{-i}, e)$
      - P(E|B,A,JC,MC)=P(E|A,B)
      - P(JC|B,E,A,MC)=P(JC|A)
      - P(B|E,A,JC,MC)=P(B|A,E)
  - Update one random variable at a time

Can simplify using the Markov blanket



KAIST

Concept of Gibbs Sampling

- Each step involves **replacing** the value of one of the variables by a value drawn from the distribution of that variable conditioned on the values of the remaining variables
- Repeated either by cycling through the variables in some particular order or by choosing the variable to be updated at each step at random from some distribution
- Example
    1. Full joint probability : $p(z_1, z_2, z_3)$
    2. Sample $z_1 \sim p\left(z_1 \mid z_2^{(\tau)}, z_3^{(\tau)}\right) \rightarrow$ Obtain a value $z_1^{(\tau+1)}$
    3. Sample $z_2 \sim p\left(z_2 \mid z_1^{(\tau+1)}, z_3^{(\tau)}\right) \rightarrow$ Obtain a value $z_2^{(\tau+1)}$
    4. Sample $z_3 \sim p\left(z_3 \mid z_1^{(\tau+1)}, z_2^{(\tau+1)}\right) \rightarrow$ Obtain a value $z_3^{(\tau+1)}$

$\left\{z_1^{(\tau)}, z_2^{(\tau)}, z_3^{(\tau)}\right\}$ $\left\{z_1^{(\tau+1)}, z_2^{(\tau)}, z_3^{(\tau)}\right\}$ $\left\{z_1^{(\tau+1)}, z_2^{(\tau+1)}, z_3^{(\tau)}\right\}$ $\left\{z_1^{(\tau+1)}, z_2^{(\tau+1)}, z_3^{(\tau+1)}\right\}$

Gibbs Sampling Algorithm

- Full joint distribution, $p(\boldsymbol{z}) = p(z_1, \dots, z_M)$
- State = $\{z_i : i = 1, \dots, M\}$
- Algorithm
  1. Initialize $\{z_i : i = 1, \dots, M\}$
  2. For step $\tau = 1, \dots, T$:
     - Sample $z_1^{(\tau+1)} \sim p\left(z_1 \,|z_2^{(\tau)}, z_3^{(\tau)}, \dots, z_M^{(\tau)}\right)$
     - Sample $z_2^{(\tau+1)} \sim p\left(z_2 \,|z_1^{(\tau+1)}, z_3^{(\tau)}, \dots, z_M^{(\tau)}\right)$

$$\vdots$$

     - Sample $z_j^{(\tau+1)} \sim p\left(z_j \,|z_1^{(\tau+1)}, \dots, z_{j-1}^{(\tau+1)}, z_{j+1}^{(\tau)}, \dots, z_M^{(\tau)}\right)$

$$\vdots$$

     - Sample $z_M^{(\tau+1)} \sim p\left(z_M \,|z_1^{(\tau+1)}, z_2^{(\tau+1)}, \dots, z_{M-1}^{(\tau+1)}\right)$

Gibbs Sampling based GMM

- Hard to tell the performance with the simple GMM
  - Sampling based inference
    - Simulation based
  - EM based inference
    - Optimization based
- Real power of Gibbs sampler comes from collapsing! → Collapsed Gibbs Sampler
  - Let's look at more sophisticated model for collapsing technique