Markov Chain Monte Carlo Methods

Presented by Chang Liu

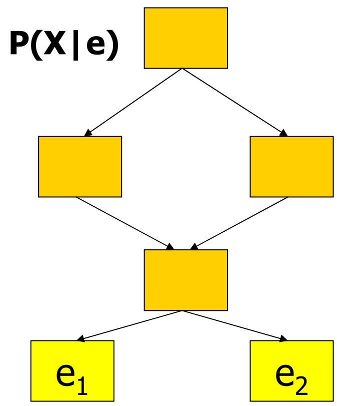
CS3710

2005.9.28

Outline

- Markov Chain
- Gibbs Sampling
- Build a Markov Chain
- Mixing Time in Using Markov Chain

Why Markov Chain?



- P(X|e)— the query we want to compute
- e_1 & e_2 are known evidence
- Sampling from the distribution P(X) is very different from the desired posterior

Markov Chain

- Markov Chain Monte Carlo
  - Objective: generates samples from the posterior distribution
  - Idea: a sampling process, that initially generates samples very different from the target posterior but gradually refines the samples so that they are closer and closer to the posterior.

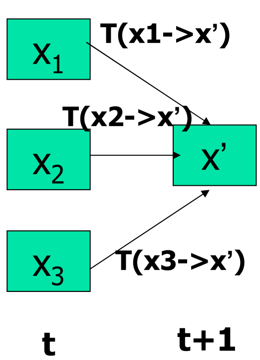
Markov Chain (Cont.)

- Based on
  - a state space $Val(X)$
  - a transition probability model T
- Defines a next-state distribution
  - for every state $x \in Val(X)$
- State space -- e.g. possible instantiations

Markov Chain (Cont.)

- Chain Dynamics

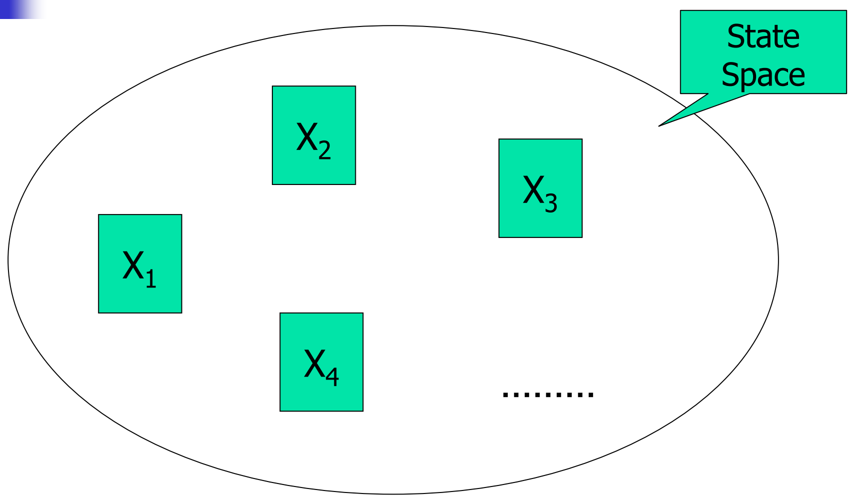
$$P^{(t+1)}(X^{(t+1)} = x') = \sum_{x \in Val(X)} P^{(t)}(X^{(t)} = x) T(x \to x')$$



$$P^{(t+1)}(X^{(t+1)} = x')$$

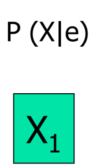
$$= \sum_{i=1}^{3} P^{(t)}(X^{(t)} = x_i) T(x_i \to x')$$

Markov Chain Monte Carlo (MCMC)



MCMC (Cont.)

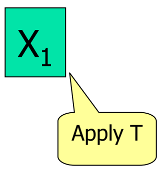
- given evidences e – state X_1 and P(X|e)



MCMC (Cont.)

- given evidences e – state X_1 and P(X|e)

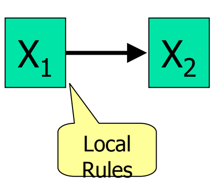
P (X|e)



MCMC (Cont.)

- given evidences e – state X_1 and P(X|e)

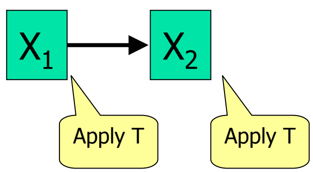
P (X|e)



MCMC (Cont.)

- given evidences e – state X_1 and $P(X|e)$

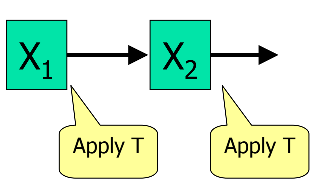
P (X|e)



MCMC (Cont.)

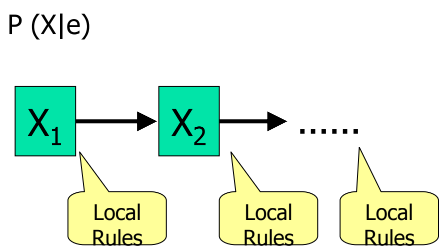
- given evidences e – state X_1 and $P(X|e)$

P (X|e)



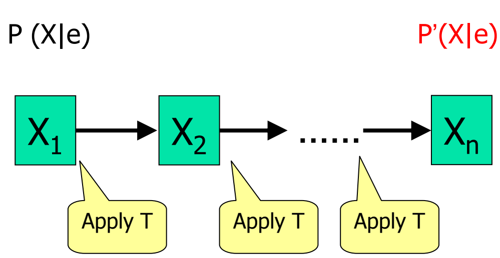
MCMC (Cont.)

- given evidences e – state X_1 and P(X|e)



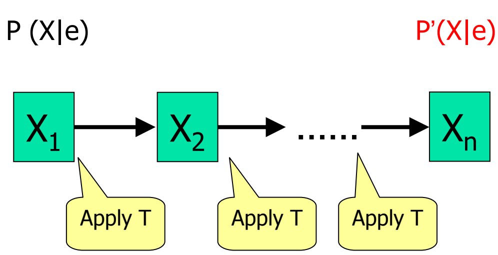
MCMC (Cont.)

- given evidences e – state X_1 and P(X|e)



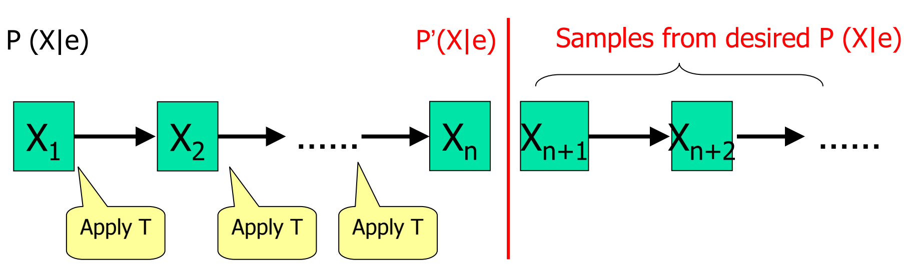
MCMC (Cont.)

- given evidences e – state X_1 and P(X|e)



MCMC (Cont.)

- given evidences e – state X_1 and P(X|e)



MCMC (Cont.)

- MCMC sampling process doesn't converge to a stationary distribution definitely
  - Stationary distribution

$$\pi(X = x') = \sum_{x \in Val(X)} \pi(X = x) T(x \to x')$$

- The stationary distribution is not unique, it depends on the initial states.

MCMC (Cont.)

- a Finite state Markov Chain has a unique stationary distribution

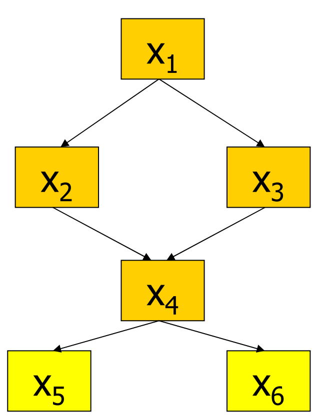
$$\Updownarrow$$

- this markov chain is regular
  - regular: exist some k, for each pair of states x and x', the probability of getting from x to x' in exactly k steps is greater than 0.

In general graphical model

- Target distribution ------ $P(\chi \mid E = e)$
- States ------ Instantiations ξ to χ
  - Some possible assignments to χ
- local transition models ------for each variable $x \in \chi - E$
- Combine all local transition models into a single chain (random select)

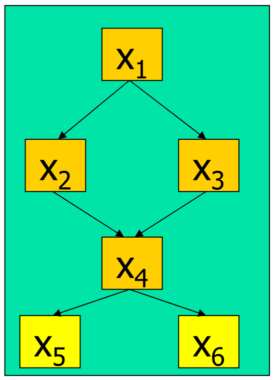
Gibbs Sampling



- Evidences:
  - x_5 =T
  - x_6 =T
- all variables have binary values T or F

Gibbs Sampling (Cont.)

$P(X_{rest} | e)$



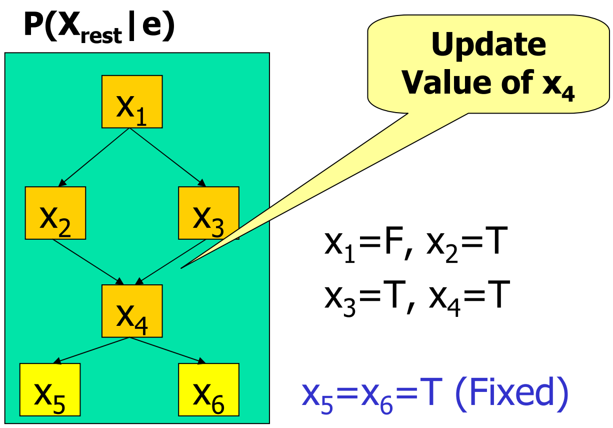
X_0

x_1=F, x_2=T
x_3=T, x_4=T

$x_5 = x_6$=T (Fixed)

Gibbs Sampling (Cont.)

$P(X_{rest} | e)$



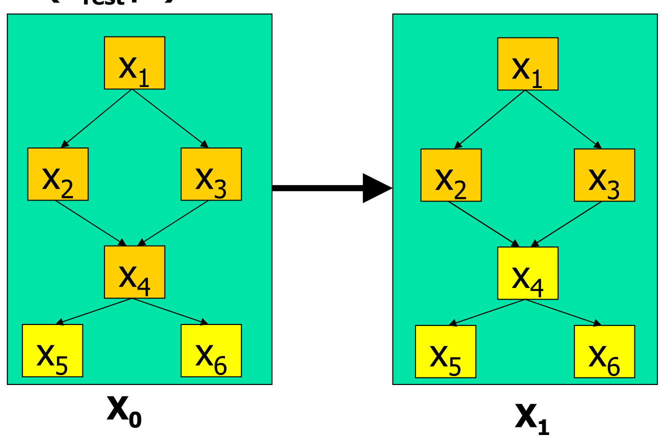
X_0

x_1=F, x_2=T
x_3=T, x_4=T

$x_5 = x_6$=T (Fixed)

Gibbs Sampling (Cont.)

$P(X_{rest} | e)$

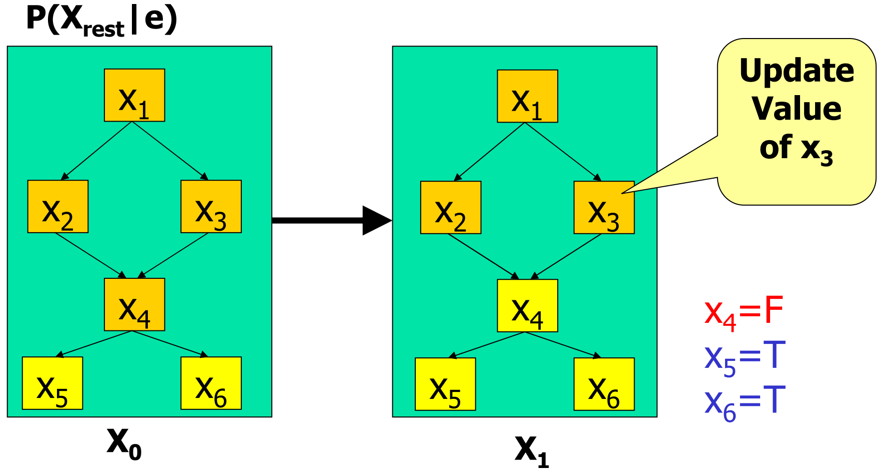


x_1=F, x_2=T,
x_3=T,

x_4=F
x_5=T
x_6=T

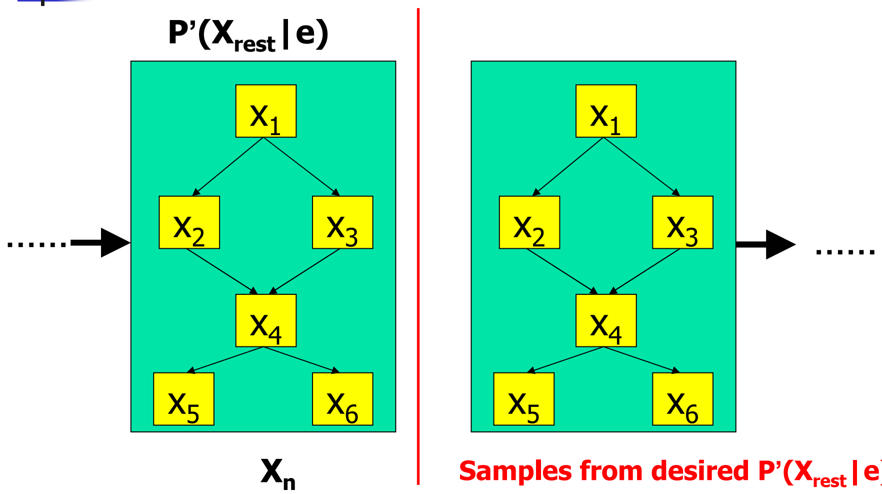
Gibbs Sampling (Cont.)

$P(X_{rest} | e)$



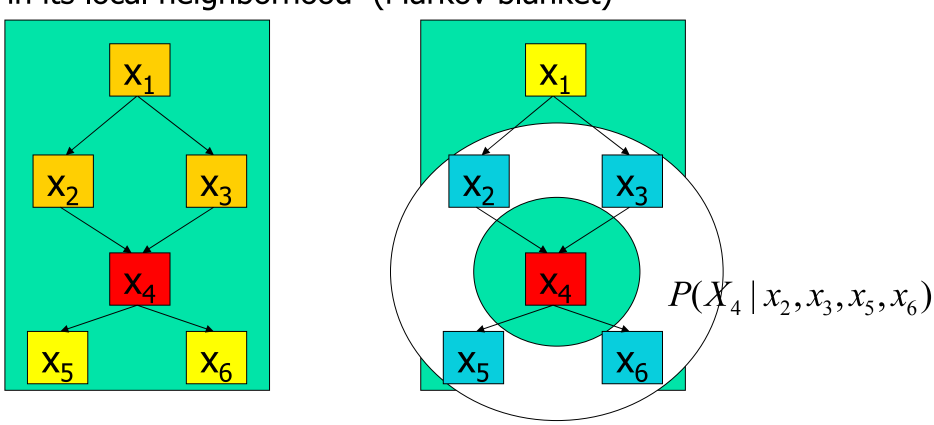
x_4=F
x_5=T
x_6=T

Gibbs Sampling (Cont.)



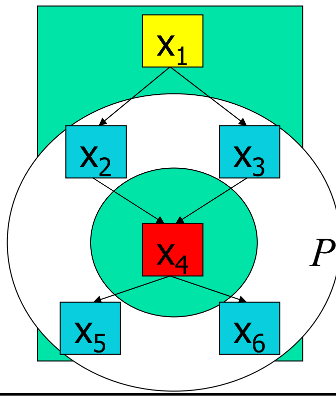
Gibbs Sampling (Cont.)

Keep resampling each variable using the value of variables in its local neighborhood (Markov blanket)



Gibbs Sampling (Cont.)

- Gibbs sampling takes advantage of the structure
- Markov blanket makes the variable independent from the rest of the network



$$P(X_4 \mid x_2, x_3, x_5, x_6)$$

Build a Markov Chain

- Key idea of MH algorithm
  - builds a reversible Markov Chain
  - For the move proposed by the proposal distribution
    - Either accept it and take a transition to state x'
    - Or reject it and stay at current state x

Build a Markov Chain (Cont.)

- For each pair of states x and x'

$x \neq x'$

$$T(x \to x') = T^Q(x \to x') A(x \to x')$$

$x = x'$

$$T(x \to x') = T^Q(x \to x') + \sum_{x' \neq x} T^Q(x \to x')(1 - A(x \to x'))$$

Build a Markov Chain (Cont.)

- From detailed balance equation

$$\pi(x) T(x \to x') = \pi(x') T(x' \to x)$$

- We can get that

$$A(x \to x') = \min[1, \frac{\pi(x') T^Q(x' \to x)}{\pi(x) T^Q(x \to x')}]$$

Build a Markov Chain (Cont.)

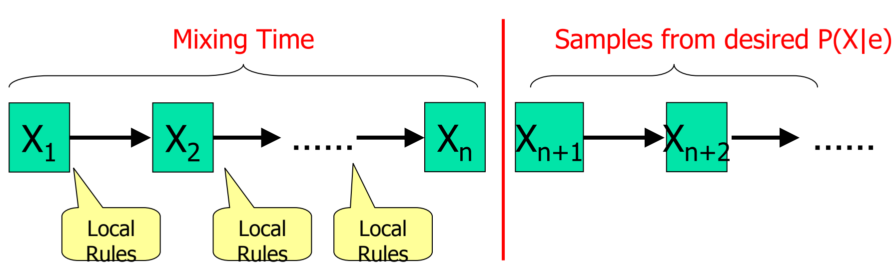
- Compared MH with Gibbs
  - For Gibbs

$$
\begin{aligned}
& A(u_i, x_i \rightarrow u_i, x'_i) \\
&= \min[\ 1, \frac{P(x'_i \mid u_i) T^Q(u_i, x'_i \rightarrow u_i, x_i)}{P(x_i \mid u_i) T^Q(u_i, x_i \rightarrow u_i, x'_i)}] \\
&= \min[\ 1, \frac{P(x'_i \mid u_i) P(x_i \mid u_i)}{P(x_i \mid u_i) P(x'_i \mid u_i)}] \\
&= \min[\ 1,1] = 1
\end{aligned}
$$

  - Special MH, which acceptance probability is 1.

Mixing Time in Using Markov Chain

- Mixing Time
  - The number of steps we take until we collect a sample from the target distribution. (# = n)



Summary

- Markov Chain Monte Carlo method attempts to generate samples from posterior distribution
- MH algorithm is a general scheme for specifying a Markov chain.
- Gibbs sampling is a special case that takes advantage of the network structure (Markov Blanket)

Thanks

Any question?