OSU CS 536
Probabilistic Graphical Models

Sampling-based Inference I

Scott Sanner

Reading:
Koller and Friedman
Ch 12

Using original slide material from UC Irvine course ICS 275 by Rina Dechter

Sampling Fundamentals

Sampling Fundamentals

Given a set of variables X = {X_1, X_2, … X_n} that represent joint probability distribution **π(X)** and some function g(X), we can compute expected value of **g(X)** :

$$E_{\pi}[g(x)] = \int g(x)\pi(X)dx$$

For **non-MCMC** methods, π(X) is the joint distribution P(X)

For **MCMC** methods, π(X) is the stationary distribution of a Markov Chain where π(X) = P(X)

Sampling From π(X)

A sample $\mathbf{S^t}$ is an instantiation:

$$S^t = \{x_1^t, x_2^t, \ldots, x_n^t\}$$

Given independent, identically distributed samples (iid) S^1, S^2, …S^T from π(X), it follows from **Strong Law of Large Numbers**:

$$E_\pi[g(x)] = \frac{1}{T}\sum_{t=1}^{T} g(S^t)$$

Forward Sampling

Forward Sampling

- Forward Sampling
  - Case with No evidence
  - Case with Evidence
  - N and Error Bounds

Forward Sampling No Evidence (Henrion 1988)

Input: Bayesian network

$X= \{X_1,\ldots,X_N\}$, N- #nodes, T - # samples

Output: T samples

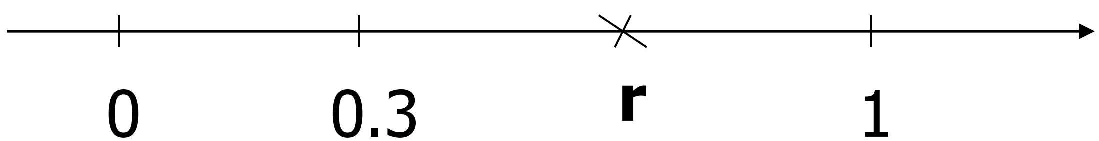
Process nodes in topological order – first process the ancestors of a node, then the node itself:

1. For t = 0 to T
2. For i = 0 to N
3. $X_i \leftarrow$ sample x_i^t from $P(x_i \mid pa_i)$

Sampling A Value

What does it mean to sample x_i^t from $P(X_i \mid pa_i)$?

- Assume $D(X_i)=\{0,1\}$
- Assume $P(X_i \mid pa_i) = (0.3, 0.7)$

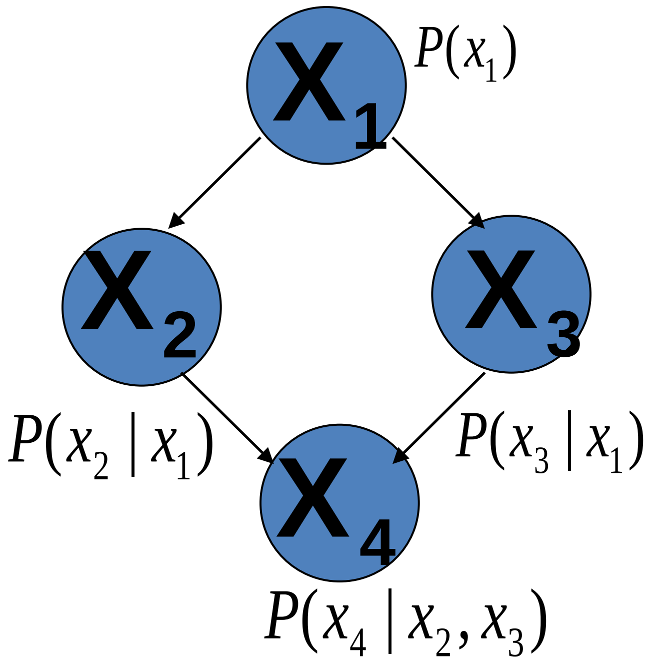


- Draw a uniform random number **r** from [0,1]

  If **r** falls in [0,0.3], set $X_i = 0$

  If **r** falls in [0.3,1], set $X_i=1$

Forward sampling (example)



$\text{Evidence} : X_3 = 0$

// generate sample k

1. Sample x_1 from $P(x_1)$
2. Sample x_2 from $P(x_2 | x_1)$
3. Sample x_3 from $P(x_3 | x_1)$
4. If $x_3 \neq 0$, reject sample and start from 1, otherwise
5. sample x_4 from $P(x_4 | x_2, x_3)$

Forward Sampling-Answering Queries

Task: given n samples $\{S^1, S^2, \ldots, S^n\}$
estimate $P(X_i = x_i)$:

$$\overline{P}(X_i = x_i) = \frac{\#samples(X_i = x_i)}{T}$$

Basically, count the proportion of samples where $X_i = x_i$

Forward Sampling w/ Evidence

Input: Bayesian network

$X = \{X_1,\ldots,X_N\}$, N- #nodes

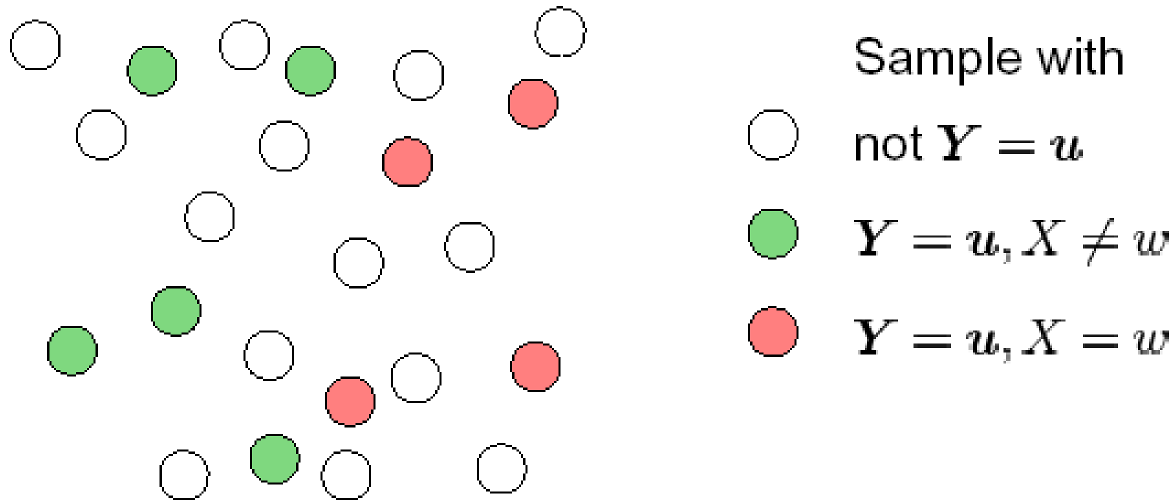
E – evidence, T - # samples

Output: T samples consistent with E

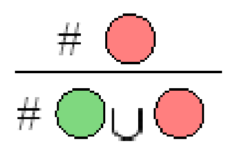
1. For t=1 to T
2. For i=1 to N
3. $X_i \leftarrow$ sample x_i^t from $P(x_i \mid pa_i)$
4. If X_i in E and $X_i \neq x_i$, **reject** sample:
5. i = 1 and go to step 2

Forward Sampling: Illustration

Let Y be a subset of evidence nodes s.t. Y=u



Approximation for $P^{\boldsymbol{X}}(X = w \mid \boldsymbol{Y} = \boldsymbol{u})$:



Forward Sampling –How many samples?

Theorem: Let $\pi_s(y)$ be the estimate of P(y) resulting from a randomly chosen sample set S with T samples. Then, to guarantee relative error at most **ε** with probability at least **1-δ** it is enough to have:

$$T \geq \frac{c}{P(y) \cdot \varepsilon^2} \bullet \frac{1}{\delta}$$

Derived from *Chebychev's Bound*.

$$P\left(\overline{P}(y) \notin [P(y) - \varepsilon, P(y) + \varepsilon]\right) \leq 2e^{-2N\varepsilon^2}$$

Forward Sampling - How many samples?

Theorem: Let $\pi_s(y)$ be the estimate of P(y) resulting from a randomly chosen sample set S with T samples. Then, to guarantee relative error at most **ε** with probability at least **1-δ** it is enough to have:

$$T \geq \frac{4}{P(y)\cdot\varepsilon^2}\ln\frac{2}{\delta}$$

Derived from *Hoeffding's Bound* (full proof is given in Koller).

$$P\left(\overline{P}(y) \notin [P(y)-\varepsilon, P(y)+\varepsilon]\right) \leq 2e^{-2N\varepsilon^2}$$

Forward Sampling: Performance

Advantages:

- $P(x_i \mid pa(x_i))$ is readily available
- Samples are independent!

Drawbacks:

- If evidence E is rare (P(e) is low), then we will reject most of the samples!
- Since P(y) in estimate of N is unknown, must estimate P(y) from samples themselves!
- If P(e) is small, T will become very big!

Gibbs MCMC Sampling

Gibbs Sampling

- A sample **t**$\in$[1,2,…],is an instantiation of all variables in the network:

$$x^t = \{X_1 = x_1^t, X_2 = x_2^t, \ldots, X_N = x_N^t\}$$

- Sampling process
  - Fix values of observed variables e
  - Instantiate node values in sample x^0 at random
  - Generate samples $x^1, x^2, \ldots x^T$ from P(x|e)
  - Compute posteriors from samples

Ordered Gibbs Sampler

Generate sample x^{t+1} from x^t :

Process All Variables In Some Order

$$X_1 = x_1^{t+1} \leftarrow P(x_1 \mid x_2^t, x_3^t, \ldots, x_N^t, e)$$

$$X_2 = x_2^{t+1} \leftarrow P(x_2 \mid x_1^{t+1}, x_3^t, \ldots, x_N^t, e)$$

...

$$X_N = x_N^{t+1} \leftarrow P(x_N \mid x_1^{t+1}, x_2^{t+1}, \ldots, x_{N-1}^{t+1}, e)$$

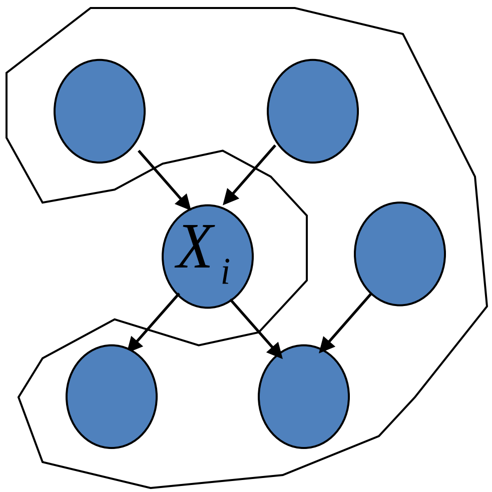
In short, for i=1 to N:

$$X_i = x_i^{t+1} \leftarrow \textbf{sampled from } P(x_i \mid x^t \setminus x_i, e)$$

Gibbs Sampling: Markov Blanket for a Variable

$$\text{Important}: P(x_i \mid x^t \setminus x_i) = P(x_i \mid markov^t \setminus x_i):$$

$$P(x_i \mid x^t \setminus x_i) \propto P(x_i \mid pa_i) \prod_{X_j \in ch_i} P(x_j \mid pa_j)$$



Markov blanket:

$$M(X_i) = pa_i \cup ch_i \cup \left(\bigcup_{X_j \in ch_j} pa_j \right)$$

Given *Markov blanket*

(parents, children, and their parents),

X_i is independent of all other nodes

Ordered Gibbs Sampling Algorithm

Input: X, E

Output: T samples $\{x^t\}$

- Fix evidence E
- Generate samples from $P(X \mid E)$

1. For t = 1 to T (compute samples)
2. For i = 1 to N (loop through variables)
3. $X_i \leftarrow$ sample x_i^t from $P(X_i \mid \mathit{markov}^t \setminus X_i)$