CSE 150A-250A AI: Probabilistic Models

Lecture 7

Fall 2025

Trevor Bonjour
Department of Computer Science and Engineering
University of California, San Diego

Slides adapted from previous versions of the course (Prof. Lawrence, Prof. Alvarado, Prof Berg-Kirkpatrick)

Agenda

Gradescope
please assign pages.

Review

if not assigned
2% penalty
from HW 3

Markov chain Monte Carlo

Review

Approximate inference

Approximate inference

- Problem (for loopy BNs)

Approximate inference

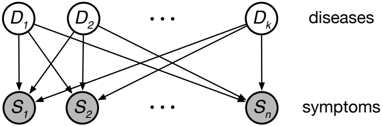
- Problem (for loopy BNs)

  Given a set E of evidence nodes, and a set Q of query nodes, how to estimate the posterior distribution $P(Q|E)$?

Approximate inference

- Problem (for loopy BNs)

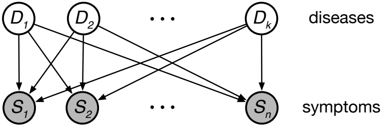
  Given a set E of evidence nodes, and a set Q of query nodes, how to estimate the posterior distribution $P(Q|E)$?



Approximate inference

- Problem (for loopy BNs)

  Given a set E of evidence nodes, and a set Q of query nodes, how to estimate the posterior distribution $P(Q|E)$?

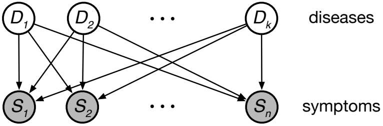


- Stochastic sampling methods

Approximate inference

- Problem (for loopy BNs)

  Given a set E of evidence nodes, and a set Q of query nodes, how to estimate the posterior distribution $P(Q|E)$?



- Stochastic sampling methods

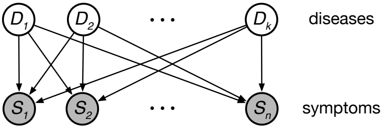
  LAST CLASS

  1. Rejection sampling — **slow**
  2. Likelihood weighting — **faster**

Approximate inference

- Problem (for loopy BNs)

  Given a set E of evidence nodes, and a set Q of query nodes, how to estimate the posterior distribution $P(Q|E)$?



- Stochastic sampling methods

  LAST CLASS

  1. Rejection sampling — **slow**
  2. Likelihood weighting — **faster**

  TODAY

  3. Markov chain Monte Carlo (MCMC) — **fastest**

Likelihood weighting

Likelihood weighting

- Make N forward passes through the BN:

Likelihood weighting

- Make N forward passes through the BN:

  Sample non-evidence nodes based on values of parents.
  Fix evidence nodes to desired values.

Likelihood weighting

- Make N forward passes through the BN:

  Sample non-evidence nodes based on values of parents.
  Fix evidence nodes to desired values.

- For single query and evidence nodes:

Likelihood weighting

- Make N forward passes through the BN:

  Sample non-evidence nodes based on values of parents.
  Fix evidence nodes to desired values.

- For single query and evidence nodes:

$$P(Q=q|E=e) \approx$$

Likelihood weighting

- Make N forward passes through the BN:

  Sample non-evidence nodes based on values of parents.
  Fix evidence nodes to desired values.

- For single query and evidence nodes:

$$P(Q=q|E=e) \approx \frac{\sum_{i=1}^{N} I(q,q_i) \overbrace{P(E=e|\mathrm{pa}_i(E))}^{\text{likelihood weight}}}{\sum_{i=1}^{N} P(E=e|\mathrm{pa}_i(E))}$$

Likelihood weighting

- Make N forward passes through the BN:

  Sample non-evidence nodes based on values of parents.
  Fix evidence nodes to desired values.

- For single query and evidence nodes:

$$P(Q=q|E=e) \approx \frac{\sum_{i=1}^{N} I(q, q_i) \overbrace{P(E=e|\mathrm{pa}_i(E))}^{\text{likelihood weight}}}{\sum_{i=1}^{N} P(E=e|\mathrm{pa}_i(E))}$$

- For multiple query and evidence nodes:

Likelihood weighting

- Make N forward passes through the BN:

  Sample non-evidence nodes based on values of parents.
  Fix evidence nodes to desired values.

- For single query and evidence nodes:

$$P(Q=q|E=e) \approx \frac{\sum_{i=1}^{N} I(q,q_i) \overbrace{P(E=e|\mathrm{pa}_i(E))}^{\text{likelihood weight}}}{\sum_{i=1}^{N} P(E=e|\mathrm{pa}_i(E))}$$

- For multiple query and evidence nodes:

$$P(Q=q, Q'=q'|E=e, E'=e')$$

$$\approx$$

Likelihood weighting

- **Make N forward passes through the BN:**

  Sample non-evidence nodes based on values of parents.
  Fix evidence nodes to desired values.

- **For single query and evidence nodes:**

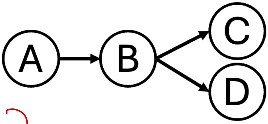
$$P(Q=q|E=e) \approx \frac{\sum_{i=1}^{N} I(q,q_i) \overbrace{P(E=e|\mathrm{pa}_i(E))}^{\text{likelihood weight}}}{\sum_{i=1}^{N} P(E=e|\mathrm{pa}_i(E))}$$

- **For multiple query and evidence nodes:**

$$P(Q=q, Q'=q'|E=e, E'=e') \approx \frac{\sum_{i=1}^{N} I(q,q_i)\, I(q',q_i')\, P(E=e|\mathrm{pa}_i(E))\, P(E'=e'|\mathrm{pa}_i(E'))}{\sum_{i=1}^{N} P(E=e|\mathrm{pa}_i(E))\, P(E'=e'|\mathrm{pa}_i(E'))}$$

$$\frac{\sum_{i=1}^{N} I(q,q_i)\, I(q',q'_i)\, P(E=e|\mathrm{pa}_i(E))\, P(E'=e'|\mathrm{pa}_i(E'))}{\sum_{i=1}^{N} P(E=e|\mathrm{pa}_i(E))\, P(E'=e'|\mathrm{pa}_i(E'))}$$

Problem: Estimate $P(a_0|c_1, d_1)$



Samples:

a_0, b_1, c_1, d_1 / (same) +

a_1, b_0, c_1, d_1 / $P(c_1|b_0)$ $P(d_1|b_0)$

a_0, b_1, c_1, d_1 / (same) +

Q. Estimate of $P(a_0|c_1, d_1)$ using likelihood weighting?

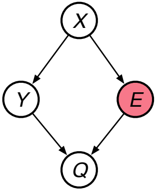
A	P(A)
a_0	1/5
a_1	4/5

A	B	P(B|A)
a_0	b_0	1/4
a_0	b_1	3/4
a_1	b_0	1/3
a_1	b_1	2/3

B	C	P(C|B)
b_0	c_0	1/5
b_0	c_1	4/5
b_1	c_0	3/5
b_1	c_1	2/5

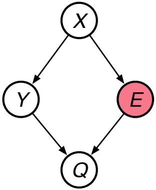
B	D	P(D|B)
b_0	d_0	3/4
b_0	d_1	1/4
b_1	d_0	1/3
b_1	d_1	2/3

Properties of likelihood weighting



Properties of likelihood weighting

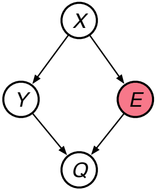
- It converges in the limit:



Properties of likelihood weighting

- It converges in the limit:

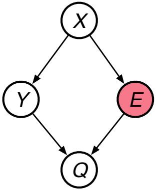
$$\lim_{N \to \infty} \frac{\sum_{i=1}^{N} I(q, q_i)\, P(E{=}e|X{=}x_i)}{\sum_{i=1}^{N} P(E{=}e|X{=}x_i)} =$$



Properties of likelihood weighting

- It converges in the limit:

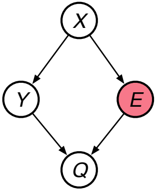
$$\lim_{N \to \infty} \frac{\sum_{i=1}^{N} I(q, q_i)\, P(E=e|X=x_i)}{\sum_{i=1}^{N} P(E=e|X=x_i)} = P(Q=q|E=e)$$



Properties of likelihood weighting

- It converges in the limit:

$$\lim_{N\to\infty} \frac{\sum_{i=1}^{N} I(q, q_i)\, P(E{=}e|X{=}x_i)}{\sum_{i=1}^{N} P(E{=}e|X{=}x_i)} = P(Q{=}q|E{=}e)$$

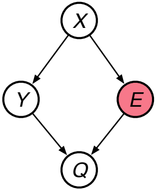


- It's more efficient than rejection sampling:

Properties of likelihood weighting

- It converges in the limit:

$$\lim_{N\to\infty} \frac{\sum_{i=1}^{N} I(q, q_i)\, P(E=e|X=x_i)}{\sum_{i=1}^{N} P(E=e|X=x_i)} = P(Q=q|E=e)$$



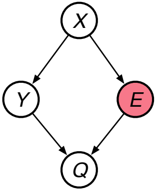
- It's more efficient than rejection sampling:

  No samples need to discarded.

Properties of likelihood weighting

- It converges in the limit:

$$\lim_{N\to\infty} \frac{\sum_{i=1}^{N} I(q, q_i)\, P(E=e|X=x_i)}{\sum_{i=1}^{N} P(E=e|X=x_i)} = P(Q=q|E=e)$$



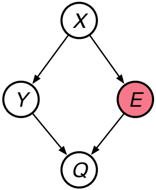
- It's more efficient than rejection sampling:

  No samples need to discarded.
  Descendants of evidence nodes are conditioned on evidence.

Properties of likelihood weighting

- It converges in the limit:

$$\lim_{N\to\infty} \frac{\sum_{i=1}^{N} I(q, q_i)\, P(E=e|X=x_i)}{\sum_{i=1}^{N} P(E=e|X=x_i)} = P(Q=q|E=e)$$



- It's more efficient than rejection sampling:

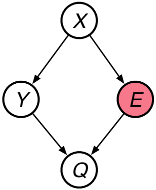
  No samples need to discarded.
  Descendants of evidence nodes are conditioned on evidence.

- But it can still be very slow:

Properties of likelihood weighting

- It converges in the limit:

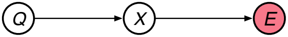
$$\lim_{N\to\infty} \frac{\sum_{i=1}^{N} I(q, q_i)\, P(E{=}e|X{=}x_i)}{\sum_{i=1}^{N} P(E{=}e|X{=}x_i)} = P(Q{=}q|E{=}e)$$



- It's more efficient than rejection sampling:

  No samples need to discarded.
  Descendants of evidence nodes are conditioned on evidence.

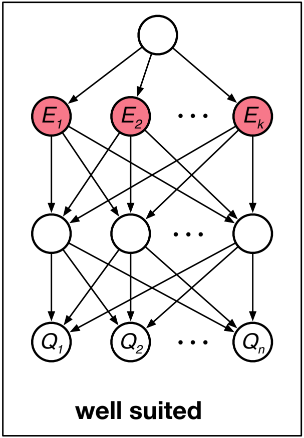
- But it can still be very slow:

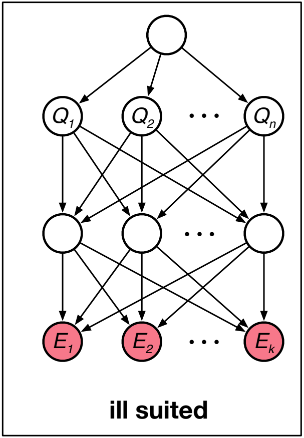


The worst case for likelihood weighting is when rare evidence is descended from query nodes.

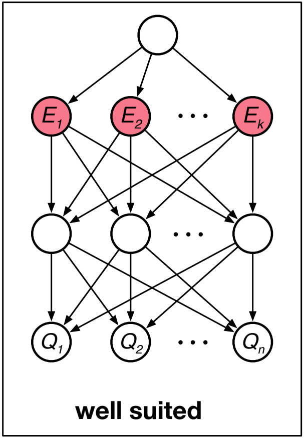
Best and worst cases for likelihood weighting

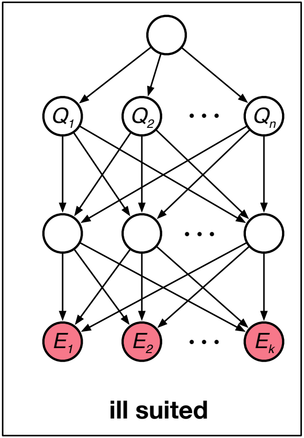
Best and worst cases for likelihood weighting





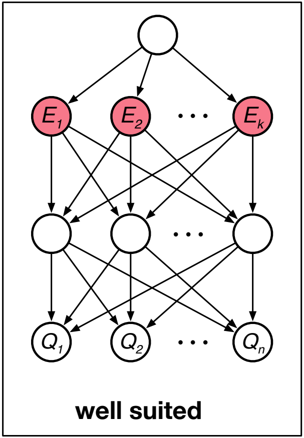
Best and worst cases for likelihood weighting

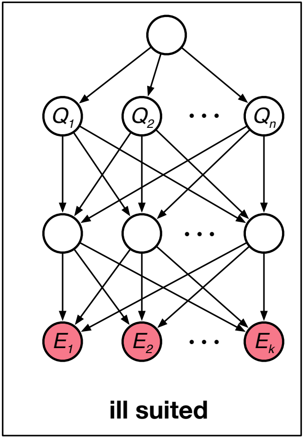




Left — rare evidence affects how query nodes are sampled.

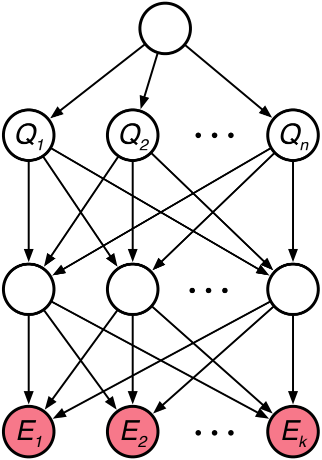
Best and worst cases for likelihood weighting





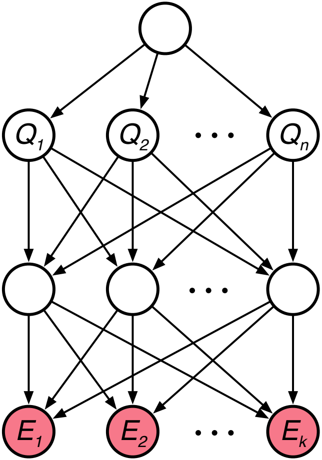
Left — rare evidence affects how query nodes are sampled.
Right — rare evidence is unlikely to occur with high probability.

What next?



What next?

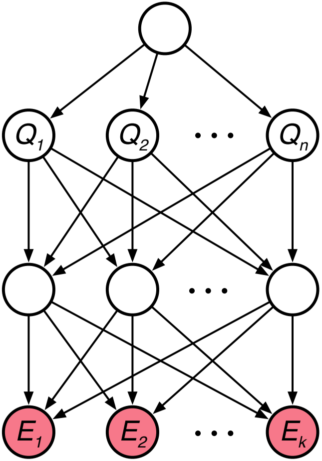
To handle this case, especially with rare evidence, we need the evidence nodes to affect how other nodes are sampled.



What next?

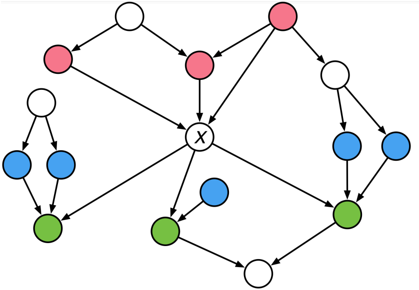
To handle this case, especially with rare evidence, we need the evidence nodes to affect how other nodes are sampled.

We need a way to sample nodes **in any order**—not only in a forward pass when they are conditioned on their parents.

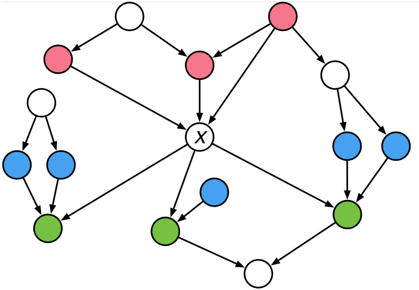


Markov blanket

Markov blanket

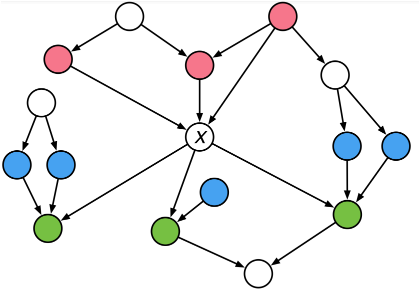


Markov blanket



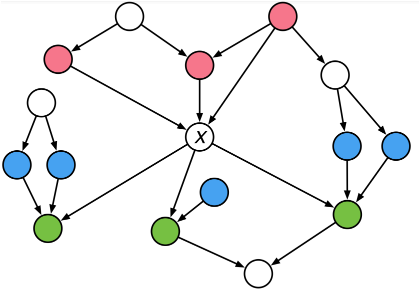
- Definition

Markov blanket



- **Definition**

  The Markov blanket B_X of a node X consists of its parents, children, and spouses (i.e., parents of children).

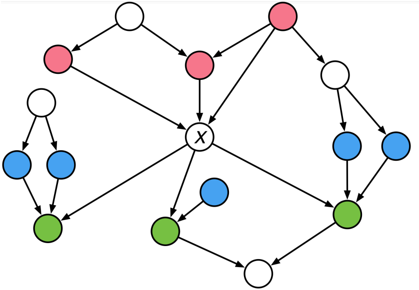


- **Definition**

  The Markov blanket B_X of a node X consists of its **parents**, **children**, and **spouses** (i.e., parents of children).

- **Theorem**

Markov blanket



- **Definition**

  The Markov blanket B_X of a node X consists of its parents, children, and spouses (i.e., parents of children).

- **Theorem**

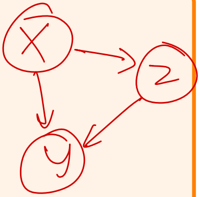
  The node X is conditionally independent of **the nodes outside** its Markov blanket given the nodes inside its Markov blanket.

Test your understanding

Let X be a node in a belief network.
Let B_X denote its Markov blanket (i.e., parents, children, spouses). Let Y be any node such that $Y \notin X \cup B_X$.

Test your understanding

Let X be a node in a belief network.
Let B_X denote its Markov blanket (i.e., parents, children, spouses). Let Y be any node such that $Y \notin X \cup B_X$.



Q. Which of these is TRUE?

A. The parents, children, and spouses of X are non-overlapping sets of nodes.

B. The parents, children, and spouses of X are non-overlapping in a polytree.

C. $P(X|B_X, Y) = P(X|B_X)$ is **only** guaranteed to be true in a polytree.

D. All are true.

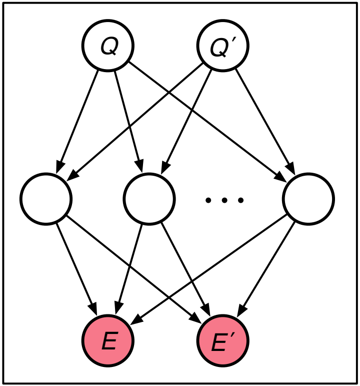
E. None are true.

Markov chain Monte Carlo

Approximate inference

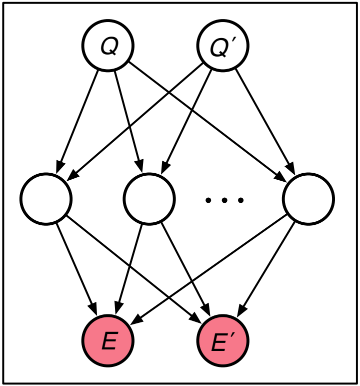
Query nodes Q, Q'

Evidence nodes E, E'



Query nodes Q, Q'

Evidence nodes E, E'



How to estimate $P(Q=q, Q'=q' | E=e, E'=e')$?

Fun Fact!

Monte Carlo methods are usually traced to physicists at Los Alamos in 1940s!

Fun Fact!

Monte Carlo methods are usually traced to physicists at Los Alamos in 1940s!

- Stanisław Ulam (inspired by solitaire!) and Von Neumann (rejection sampling).
- Interested in modeling the probabilistic behavior of collections of atomic particles.
- The term 'Monte-Carlo' was coined at Los Alamos.

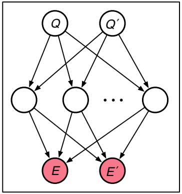
MCMC - Gibbs Sampling

- Initialization

MCMC - Gibbs Sampling

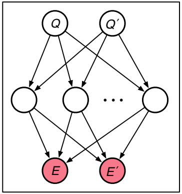
- Initialization
  Fix evidence nodes to observed values e, e'.

- **Initialization**
  Fix evidence nodes to observed values e, e'.
  Initialize non-evidence nodes to random values.

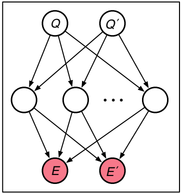


MCMC - Gibbs Sampling

- **Initialization**
  Fix evidence nodes to observed values e, e'.
  Initialize non-evidence nodes to random values.
- **Repeat N times**



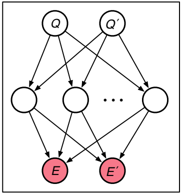
- **Initialization**
  Fix evidence nodes to observed values e, e'.
  Initialize non-evidence nodes to random values.
- **Repeat N times**
  Pick a non-evidence node X at random.



MCMC - Gibbs Sampling

- **Initialization**
  Fix evidence nodes to observed values e, e'.
  Initialize non-evidence nodes to random values.

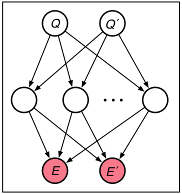
- **Repeat N times**
  Pick a non-evidence node X at random.
  Use **Bayes rule** to compute $P(X|B_X)$.



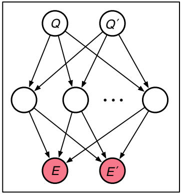
MCMC - Gibbs Sampling

- **Initialization**
  Fix evidence nodes to observed values e, e'.
  Initialize non-evidence nodes to random values.

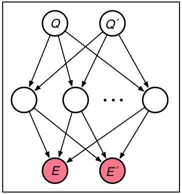
- **Repeat N times**
  Pick a non-evidence node X at random.
  Use **Bayes rule** to compute $P(X|B_X)$.
  Resample $x \sim P(X|B_X)$.



MCMC - Gibbs Sampling

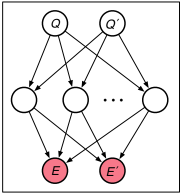


- **Initialization**
  Fix evidence nodes to observed values e, e'.
  Initialize non-evidence nodes to random values.
- **Repeat** N **times**
  Pick a non-evidence node X at random.
  Use **Bayes rule** to compute $P(X|B_X)$.
  Resample $x \sim P(X|B_X)$.
  Take a snapshot of all the nodes in the BN.



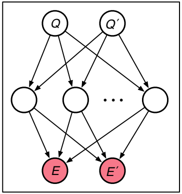
- **Initialization**
  Fix evidence nodes to observed values e, e'.
  Initialize non-evidence nodes to random values.
- **Repeat N times**
  Pick a non-evidence node X at random.
  Use **Bayes rule** to compute $P(X|B_X)$.
  Resample $x \sim P(X|B_X)$.
  Take a snapshot of all the nodes in the BN.
- **Estimate**

MCMC - Gibbs Sampling



- **Initialization**
  Fix evidence nodes to observed values e, e'.
  Initialize non-evidence nodes to random values.
- **Repeat** N **times**
  Pick a non-evidence node X at random.
  Use **Bayes rule** to compute $P(X|B_X)$.
  Resample $x \sim P(X|B_X)$.
  Take a snapshot of all the nodes in the BN.
- **Estimate**
  Count the snapshots $N(q, q') \leq N$ with $Q = q$ and $Q' = q'$.

MCMC - Gibbs Sampling



- **Initialization**
  Fix evidence nodes to observed values e, e'.
  Initialize non-evidence nodes to random values.
- **Repeat** N **times**
  Pick a non-evidence node X at random.
  Use **Bayes rule** to compute $P(X|B_X)$.
  Resample $x \sim P(X|B_X)$.
  Take a snapshot of all the nodes in the BN.
- **Estimate**
  Count the snapshots $N(q, q') \leq N$ with $Q = q$ and $Q' = q'$.

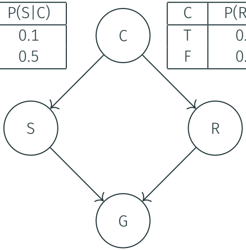
$$P(Q = q, Q' = q' | E = e, E' = e') \approx \frac{N(q, q')}{N}$$

Estimate $P(R = 1 \mid S = 1, G = 1)$

P(C)
0.5

C	P(S|C)
0	0.1
1	0.5

C	P(R|C)
T	0.8
F	0.2



S	R	P(G|S,R)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Gibbs Sampling Example

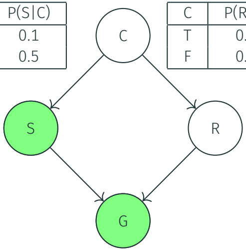
Estimate $P(R = 1 \mid S = 1, G = 1)$

- **Initialization**
  - Set evidence: s_1, g_1

P(C)
0.5

C	P(S|C)
0	0.1
1	0.5

C	P(R|C)
T	0.8
F	0.2



S	R	P(G|S,R)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Gibbs Sampling Example

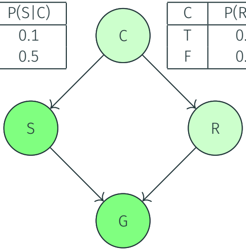
Estimate $P(R = 1 \mid S = 1, G = 1)$

- **Initialization**
  - Set evidence: s_1, g_1
  - Randomly set non-evidence variables: c_1, r_1

P(C)
0.5

C	P(S|C)
0	0.1
1	0.5

C	P(R|C)
T	0.8
F	0.2



S	R	P(G|S,R)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Gibbs Sampling Example

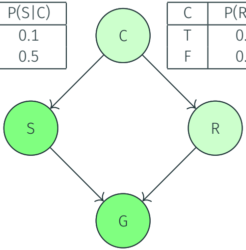
Estimate $P(R = 1 \mid S = 1, G = 1)$

- **Initialization**
  - Set evidence: s_1, g_1
  - Randomly set non-evidence variables: c_1, r_1
- Repeat N times:

P(C)
0.5

C	P(S|C)
0	0.1
1	0.5

C	P(R|C)
T	0.8
F	0.2



S	R	P(G|S,R)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Gibbs Sampling Example

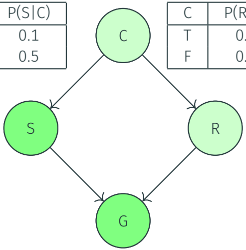
Estimate $P(R = 1 \mid S = 1, G = 1)$

- **Initialization**
  - Set evidence: s_1, g_1
  - Randomly set non-evidence variables: c_1, r_1
- Repeat N times:
  - Pick variable to update from $\{R, C\}$ uniformly at random: R

P(C)
0.5

C	P(S|C)
0	0.1
1	0.5

C	P(R|C)
T	0.8
F	0.2

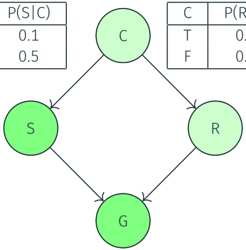


S	R	P(G|S,R)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Gibbs Sampling Example

Estimate $P(R = 1 \mid S = 1, G = 1)$

- **Initialization**
    - Set evidence: s_1, g_1
    - Randomly set non-evidence variables: c_1, r_1
- Repeat N times:
    - Compute $P(R|c_1, s_1, g_1)$ using **Bayes rule**



P(C)
0.5

C	P(S|C)
0	0.1
1	0.5

C	P(R|C)
T	0.8
F	0.2

S	R	P(G|S,R)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Gibbs Sampling Example

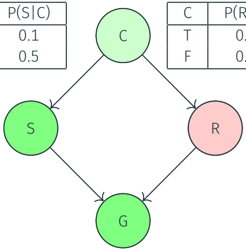
Estimate $P(R = 1 \mid S = 1, G = 1)$

- **Initialization**
  - Set evidence: s_1, g_1
  - Randomly set non-evidence variables: c_1, r_1
- Repeat N times:
  - Sample R from $P(R|c_1, s_1, g_1)$: r_0
    Take a snapshot

P(C)
0.5

C	P(S|C)
0	0.1
1	0.5

C	P(R|C)
T	0.8
F	0.2



S	R	P(G|S,R)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Gibbs Sampling Example

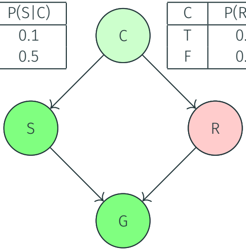
Estimate $P(R = 1 \mid S = 1, G = 1)$

- **Initialization**
  - Set evidence: s_1, g_1
  - Randomly set non-evidence variables: c_1, r_1
- Repeat N times:
  - Pick variable to update from $\{R, C\}$ uniformly at random: **C**

P(C)
0.5

C	P(S|C)
0	0.1
1	0.5

C	P(R|C)
T	0.8
F	0.2



S	R	P(G|S,R)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Gibbs Sampling Example

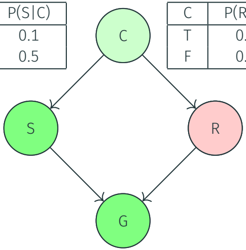
Estimate $P(R = 1 \mid S = 1, G = 1)$

- **Initialization**
  - Set evidence: s_1, g_1
  - Randomly set non-evidence variables: c_1, r_1
- Repeat N times:
  - Compute $P(C|r_0, s_1)$ using **Bayes rule**

P(C)
0.5

C	P(S|C)
0	0.1
1	0.5

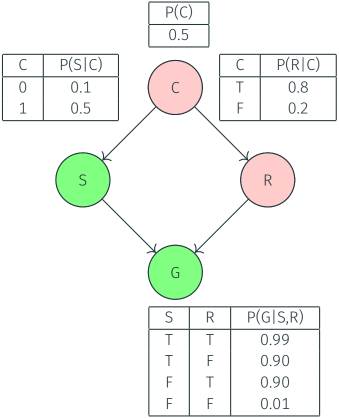
C	P(R|C)
T	0.8
F	0.2



S	R	P(G|S,R)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Estimate $P(R = 1 \mid S = 1, G = 1)$

- **Initialization**
  - Set evidence: s_1, g_1
  - Randomly set non-evidence variables: c_1, r_1
- Repeat N times:
  - Sample C from $P(C|r_0, s_1)$: c_0
    Take a snapshot



P(C)
0.5

C	P(S|C)
0	0.1
1	0.5

C	P(R|C)
T	0.8
F	0.2

S	R	P(G|S,R)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Gibbs Sampling Example

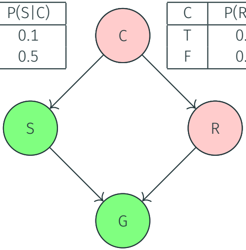
Estimate $P(R = 1 \mid S = 1, G = 1)$

- **Initialization**
  - Set evidence: s_1, g_1
  - Randomly set non-evidence variables: c_1, r_1
- Repeat N times:
  - Pick variable to update from $\{R, C\}$ uniformly at random: **C**

P(C)
0.5

C	P(S|C)
0	0.1
1	0.5

C	P(R|C)
T	0.8
F	0.2



S	R	P(G|S,R)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Gibbs Sampling Example

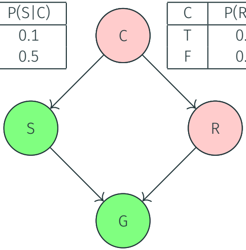
Estimate $P(R = 1 \mid S = 1, G = 1)$

- **Initialization**
  - Set evidence: s_1, g_1
  - Randomly set non-evidence variables: c_1, r_1
- Repeat N times:
  - Compute $P(C|r_0, s_1)$ using **Bayes rule**

P(C)
0.5

C	P(S|C)
0	0.1
1	0.5

C	P(R|C)
T	0.8
F	0.2



S	R	P(G|S,R)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

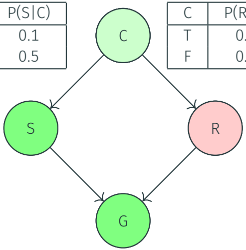
Estimate $P(R = 1 \mid S = 1, G = 1)$

- **Initialization**
  - Set evidence: s_1, g_1
  - Randomly set non-evidence variables: c_1, r_1
- Repeat N times:
  - Sample C from $P(C|r_0, s_1)$: c_1
    Take a snapshot

P(C)
0.5

C	P(S|C)
0	0.1
1	0.5

C	P(R|C)
T	0.8
F	0.2



S	R	P(G|S,R)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Gibbs Sampling Example

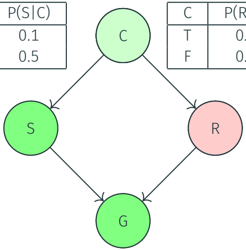
Estimate $P(R = 1 \mid S = 1, G = 1)$

- **Initialization**
  - Set evidence: s_1, g_1
  - Randomly set non-evidence variables: c_1, r_1
- Repeat N times:
  -
- Count the snapshots with r_1: N_{r_1}

P(C)
0.5

C	P(S|C)
0	0.1
1	0.5

C	P(R|C)
T	0.8
F	0.2



S	R	P(G|S,R)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Gibbs Sampling Example

Estimate $P(R = 1 \mid S = 1, G = 1)$

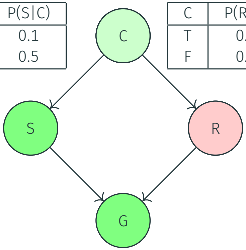
- **Initialization**
  - Set evidence: s_1, g_1
  - Randomly set non-evidence variables: c_1, r_1
- Repeat N times:
  -
- Count the snapshots with r_1: N_{r_1}

$$P(R = 1|S = 1, G = 1) \approx \frac{N_{r_1}}{N}$$

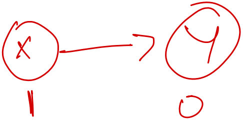
P(C)
0.5

C	P(S|C)
0	0.1
1	0.5

C	P(R|C)
T	0.8
F	0.2



S	R	P(G|S,R)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

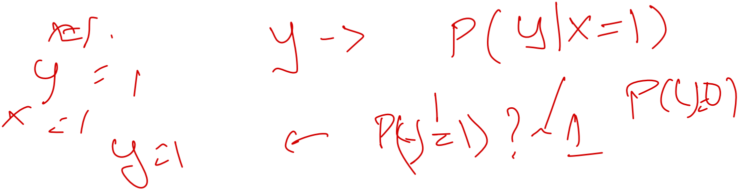


$P(X=1) = 0.5$

$P(Y=1 \mid X=1) = 1$

Q. (A) True or (B) False

Gibbs MCMC could get stuck if the relationship between two random variables is *deterministic*.



Properties of MCMC

Properties of MCMC

Under reasonable conditions...

Properties of MCMC

Under reasonable conditions...

1. This sampling procedure defines an ergodic (**irreducibile** and **aperiodic**) Markov chain over the non-evidence nodes of the BN.

Properties of MCMC

Under reasonable conditions...

1. This sampling procedure defines an ergodic (**irreducibile** and **aperiodic**) Markov chain over the non-evidence nodes of the BN.
2. The stationary distribution of this Markov chain is equal to the BN's posterior distribution over its non-evidence nodes.

Properties of MCMC

Under reasonable conditions...

1. This sampling procedure defines an ergodic (**irreducibile** and **aperiodic**) Markov chain over the non-evidence nodes of the BN.
2. The stationary distribution of this Markov chain is equal to the BN's posterior distribution over its non-evidence nodes.
3. Theoretical guarantees for mixing time, in practice we use burn in time.
4. The estimates from MCMC converge in the limit:

$$\lim_{N \to \infty} \frac{N(q, q')}{N} \rightarrow P(Q = q, Q' = q' | E = e, E' = e')$$

MCMC versus likelihood weighting (LW)

MCMC versus likelihood weighting (LW)

- How they sample

MCMC versus likelihood weighting (LW)

- How they sample

$$\left.\begin{array}{r}\text{LW} \\ \text{MCMC}\end{array}\right\} \text{ samples non-evidence nodes from } \left\{\begin{array}{l}P(X|\mathrm{pa}(X)) \\ P(X|B_X)\end{array}\right.$$

MCMC versus likelihood weighting (LW)

- How they sample

$$\left.\begin{array}{r}\text{LW}\\ \text{MCMC}\end{array}\right\}\ \text{samples non-evidence nodes from}\ \left\{\begin{array}{l}P(X|\mathrm{pa}(X))\\ P(X|B_X)\end{array}\right.$$

- Cost per sample

MCMC versus likelihood weighting (LW)

- How they sample

$$\left.\begin{array}{r}\text{LW}\\ \text{MCMC}\end{array}\right\}\text{ samples non-evidence nodes from }\left\{\begin{array}{l}P(X|\mathrm{pa}(X))\\ P(X|B_X)\end{array}\right.$$

- Cost per sample

LW can read off $P(X|\mathrm{pa}(X))$ from each CPT.

MCMC versus likelihood weighting (LW)

- How they sample

$$\left.\begin{array}{r}\text{LW}\\ \text{MCMC}\end{array}\right\}\text{ samples non-evidence nodes from }\left\{\begin{array}{l}P(X|\mathrm{pa}(X))\\ P(X|B_X)\end{array}\right.$$

- Cost per sample

LW can read off $P(X|\mathrm{pa}(X))$ from each CPT.
MCMC must compute $P(X|B_X)$ before each sample.

MCMC versus likelihood weighting (LW)

- How they sample

$$\left.\begin{array}{r}\text{LW}\\\text{MCMC}\end{array}\right\}\text{ samples non-evidence nodes from }\left\{\begin{array}{l}P(X|\mathrm{pa}(X))\\P(X|B_X)\end{array}\right.$$

- Cost per sample

  LW can read off $P(X|\mathrm{pa}(X))$ from each CPT.
  MCMC must compute $P(X|B_X)$ before each sample.

- Convergence

MCMC versus likelihood weighting (LW)

- How they sample

$$\left.\begin{array}{r}\text{LW}\\ \text{MCMC}\end{array}\right\}\text{ samples non-evidence nodes from }\left\{\begin{array}{l}P(X|\mathrm{pa}(X))\\ P(X|B_X)\end{array}\right.$$

- Cost per sample

LW can read off $P(X|\mathrm{pa}(X))$ from each CPT.
MCMC must compute $P(X|B_X)$ before each sample.

- Convergence

LW is slow for rare evidence in leaf nodes.

MCMC versus likelihood weighting (LW)

- How they sample

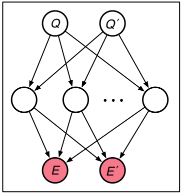
$$\left.\begin{array}{r}\text{LW}\\ \text{MCMC}\end{array}\right\}\text{ samples non-evidence nodes from }\left\{\begin{array}{l}P(X|\mathrm{pa}(X))\\ P(X|B_X)\end{array}\right.$$

- Cost per sample

  LW can read off $P(X|\mathrm{pa}(X))$ from each CPT.
  MCMC must compute $P(X|B_X)$ before each sample.

- Convergence

  LW is slow for rare evidence in leaf nodes.
  MCMC can be much faster in this situation.



That's all folks!