① Approximate Inference w/ Sampling.

(1.1) Motivation:

Exact inference is intractable for many types of BNs.

(1.2) Idea:

> A BN is a generative model.
> You can generate samples from it.
> Use the samples to estimate the probabilities you want.

(1.3) Simple Example:

$$\sum_R P(R \mid S) = 1$$

S	P(S)
☀	0.8
☁	0.2

S	R	P(R | S)
☀	☂	0.85
☀	☂	0.15
☁	☂	0.20
☁	☂	0.80

> w/ these coins you can generate samples.

		P(☂)
1. ☀	☂	$\frac{\#(☂)}{7}$
2. ☀	☂	
3. ☁	☂	P(☂)
4. ☀	☂	
5. ☀	☂	$= 1 - P(☂)$
6. ☁	☂	
7. ☀	☂	

> Query: ~~P(R | ☀)~~

$P(R) = ?$ ~~P(R = ☂)~~ or ~~P(R = ☂)~~

>  $P(☀) = 0.8$

 $P(☁) = 0.2$

Flip this coin: if ☀ then use sunny coin. else use cloudy coin

>  Sunny &  Cloudy.

$P(☂) = 0.85$ $P(☂) = 0.20$

$$S \rightarrow R$$

S	$P(S)$
s	0.8
$-s$	0.2

	S	R	$P(R \mid S)$
1	s	r	0.85
	s	$-r$	0.15
2	$-s$	r	0.2
	$-s$	$-r$	0.8

> Flip a coin for s

s

> Flip the coin 1 to obtain value for R

$-r$

$(s, -r)$

use coin 2 to get R value.

$(-s, -r)$

(s, r)

$\vdots$

$(s, -r)$

$Pr(s, r)$

$Pr(S = s)$

$Pr(R = r)$

2.4 Are these probability estimates "correct"?

Consistency: A sampling estimate is consistent if in the limit it yields the exact probability.

i.e) If $\hat{\mathbb{P}}(X \mid E=e) = \#_s(X,e) / \#_s(e)$ is the sampling estimate

Then $\hat{P}$ is consistent if

$$\lim_{N \to \infty} \hat{P}_r(X \mid E=e) = P_r(X \mid E=e)$$

We did not cover this in class but useful to know

i.e)
$$= \lim_{N \to \infty} \frac{\#_s(x,e)}{N} \cdot \frac{1}{\frac{\#_s(e)}{N}}$$

Straightforward

$$\lim_{N \to \infty} \frac{\#_s(x,e)}{N} = P(x,e)$$

$$\lim_{N \to \infty} \frac{\#_s(e)}{N} = P(e)$$

Because of how we defined the sampling process

$$\Rightarrow \quad = \frac{P(x,e)}{P(e)}$$

$$= P(x \mid e)$$

~~Intuition~~

(3.1) **Algorithm** : Input : $Q, E = e, H$ $\{X = Q, H\}$

1. Fix $E = e$

2. Topologically sort Q, E, H

3. In sorted order, sample values for each var : $\Pr(X_i \mid \text{Parents}(X_i))$

   ↳ evidence vars are fixed to i/ps

4. For each sample $(q, e, h) = S_j$ weight it by ~~$\prod_{q,h} \Pr(e \mid q,h)$~~

$$wt_j \leftarrow \prod_{e_i \in e} \Pr(e_i \mid \text{Parents}(E_i))$$

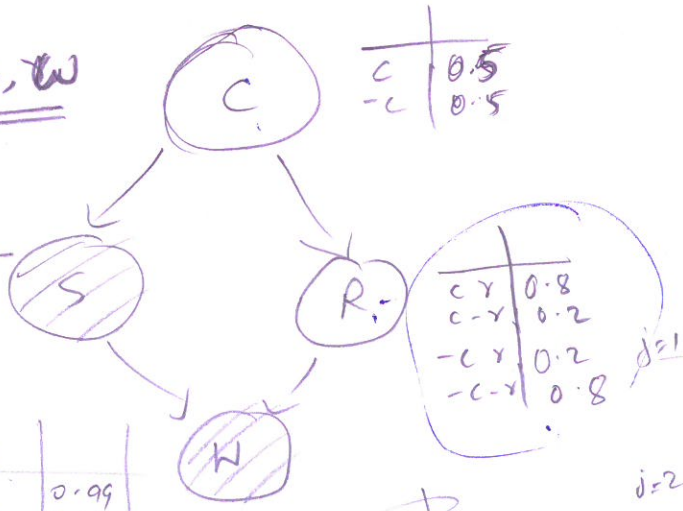
w_1	S_1
w_2	S_2
$\vdots$	$\vdots$
w_n	S_n

q a a a

e e $\vdots$ e

5. $$\Pr(q \mid e) = \frac{\sum_{S_j \in S_q} wt_j}{\sum_{S_j \in S} wt_j}$$

e.g. $e = S, W$



C	
C	0.5
-C	0.5

C	S	S|C
C	S	0.1
C	-S	0.9
-C	S	0.5
-C	-S	0.5

C	R	R|C
C	R	0.8
C	-R	0.2
-C	R	0.2
-C	-R	0.8

S	R	W	
S	R	W	0.99
S	R	-W	0.01
S	-R	W	0.9
S	-R	-W	0.1
-S	R	W	0.9
-S	R	-W	0.1
-S	-R	W	0.01

Samples	Weight
	$\Pr(S\|C) \times \Pr(W\|R)$
$j=1$ C S R W	0.1×0.99 ✓
C S -R W	0.5×0.9 ✓
$j=2$ C S -R W	
S R W	

$\Pr(C \mid S, W) =$ Sum of ✓ wts / Total wts.

(4) Monte Carlo Markov Chain

Motivation ~~Idea~~: Let's ensure that evidence influences both upstream and downstream.

Idea: > Starting from top scratch is the problem

~~> Why don't we keep~~

> Sample ~~every~~ variables conditioned on rest

> No re-start from scratch / top.

Gibbs Sampling

> start w/ an arbitrary instantiation but consistent w/ evidence. $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$

> Select one variable (keep evidence fixed) X_i

> Sample a new value conditioned on the rest $\Pr(X_i \mid X_{/i})$

all but X_i

> Repeat!

BN summary

> BN compactly specifies joint distributions in terms of CPTs.

> Any prob. query can be evaluated over a BN!

↳ Remarkable since we only store 'n' CPTs.

> Inference in BNs is the key challenge

>> Exact inference is hard

>>> For some n/ws is tractable.

>> Sampling, approx inference, is useful in practice.

> How to obtain the structure of n/w?

↳ domain experts
↳ some structure learning

> How to obtain CPTs?

↳ hand-code → not tenable
↳ learn (see example in following lect).

Oct 24th Lecture

Plan

Inference in BN

(1) Recap of Sampling based inference

(2) MCMC : Gibbs sampling

(3) Bayesian Network summary

Learning

(1) Learning from examples

(2) Basic supervised learning recipe

(3) Naive Bayes

(4) Logistic Regression

$Q = Q_1, Q_2 \; Q_3$

$\underline{q} = q_1, -q_2,$

$\underline{q_1 \; q_2 \; -q_3}$

$\underline{q_1 \; -q_2 \; -q_3}$

$\underline{q_1 \; -q_2 \; q_3}$

$\underline{q_1 \; -q_2 \; q_3}$

1 Sampling Based Inference - Recap

> Exact inference is intractable for general BNs

$O \to O \to O \ldots$ > VE is tractable - polynomial $O(d^2 n)$ (size of the dom; the # of vars) for chains. More generally for poly-tree

Motivation

> When $\exists$ more than one path to any node, inference is intractable - exponential in BN size.

> VO helps but isn't enough.

2

> obtain samples from BN.

> Use samples to estimate probabilities

Idea

3

> Suppose there are samples:

B	A	C	E
b,	a,	-c,	e
-b,	a,	c,	-e
b,	-a,	c,	-e

What is $Pr(c)$?

What is $Pr(a|-c)$?

Example

$$Pr(c) = \frac{\#(c)}{N} = \frac{2}{3} \qquad Pr(-c) = \frac{1}{3}$$

$$Pr(a|-c) = \frac{\#(a,-c)}{\#(-c)} = \frac{1}{1} = 1$$

4

Question

How do we obtain samples?

5 - cont'd

> Likelihood weighted sampling

↳ Don't sample evidence vars

↳ Keep their values fixed

↳ Sample the rest

Beware: Probs are inconsistent

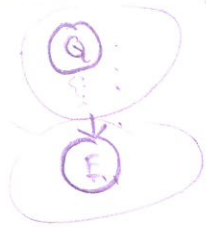
> If E is downstream from Q

Count	E	Q
$w(s_1) = Pr(e \mid q_1)$	s_1	
$w(s_2) = Pr(e \mid q_2)$	s_2	
	s_3	
$w(s_n) = Pr(e \mid q_n)$		

> Q samples don't account for evidence E

↳ ok if E is the top of the n/w

> But not neccessarily true in general



> Idea ?

Formally

$$> Pr(q \mid e) = \frac{\sum_{s_i \in S_q} w(s_i)}{\sum_{s_i \in S_q} w(s_i) + \sum_{s_i \in S_{\neg q}} w(s_i)}$$

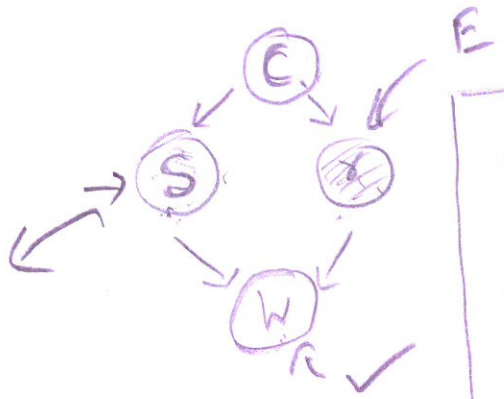
$> w(s_i) = Pr(e \mid q_i)$

⑥ Markov Chain Monte Carlo (MCMC)

Motivation

Evidence should influence both up & down stream vars!

The issue



S sampled from $Pr(S|C)$

> We want S to be influenced by E

> Likelihood weighted doesn't directly do this

Why? Starting to sample from top every time decouples S from r

The fix

Not from the top »

When sampling value for a var condition on the current values for all vars in the n/w

Why the fix works?

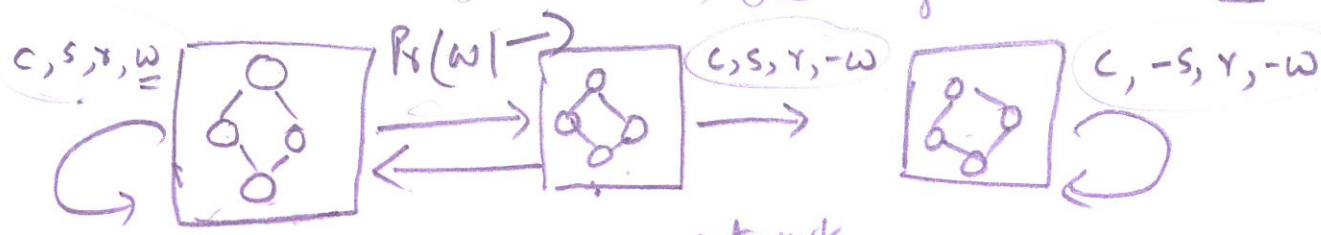
Suppose you have sampled $C = c$, $W = w$ & $R = r$ → Evidence

If you sample S from $Pr(S | C=c, W=w, R=r)$ you will see that S influenced by $\boxed{r}$

(7) -contd

Why 'does' Gibb's work?

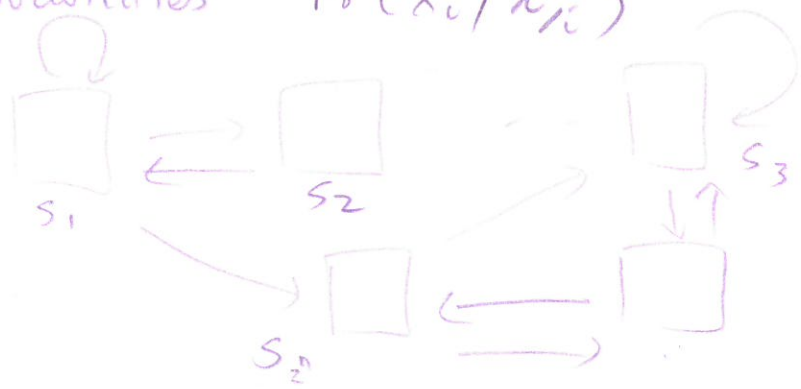
> Gibbs cycles through a Markov chain over diff configs of the vars



Markov ~~chain~~ network of config.

> Each configuration = Sample = state in MC

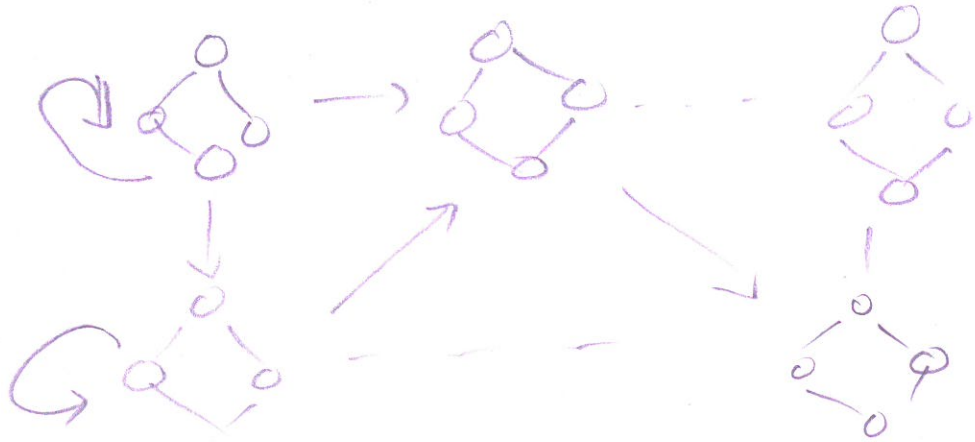
> A random walk on this chain cond. on the probabilities $Pr(x_i / x_{/i})$



> If you walk forever then stationary distribution $\pi \rightarrow Pr(S_i)$, the prob of sample according to the BN

Why does Gibbs work?

> The set of configuration that Gibbs cycles through forms a Markov Chain.



$2^4 = 16$ states

> Each state = sample

> walk through this chain = seq of samples.

> If you walk forever, the walk is transition guided by the $Pr(x_i \mid X_{/i})$ distribution

> One can show that the stationary distribution of this walk (ie) the prob of landing in each state

= Pr(sample) given the BN!

> So we get consistent estimates.