Approximate Inference

Artificial Intelligence

School of Computer Science
The University of Adelaide

Inference on Bayesian Networks

Exact inference: computational expensive for a large BN.

- Number of multiplications approach to $O(n2^n)$

$$
\begin{aligned}
P(b|j,\neg m) &= \alpha \sum_e \sum_a P(b)P(e)P(a|b,e)P(j|a)P(\neg m|a) \\
&= \alpha\,[P(b)P(e)P(a|b,e)P(j|a)P(\neg m|a) \\
&\quad +P(b)P(e)P(\neg a|b,e)P(j|\neg a)P(\neg m|\neg a) \\
&\quad +P(b)P(\neg e)P(a|b,\neg e)P(j|a)P(\neg m|a) \\
&\quad +P(b)P(\neg e)P(\neg a|b,\neg e)P(j|\neg a)P(\neg m|\neg a)]
\end{aligned}
$$

Approximate inference:

- Approximately calculate the posterior probability.
- Use random sampling for inference.
- More samples leads to more accurate solutions.

Inference on Bayesian Networks

Exact inference: computational expensive for a large BN.

- Number of multiplications approach to $O(n2^n)$

Approximate inference with sampling:

- Direct Sampling (Prior Sampling).
- Rejection Sampling.
- Likelihood Weighting.
- Gibbs Sampling

Direct sampling

Markov Chain Monte Carlo (MCMC) sampling

Sampling

- What is sampling

Sampling is a statistical procedure to select the individual observations from the population.

- Why sampling

Statisticians attempt for the samples to represent the whole population in question.

- Example:

What is the probability of getting 3 when rolling a dice?

$$P(X = x_i) = \frac{number\ of\ times\ \{X = x_i\}}{total\ number\ of\ trials}$$

Sampling from a Distribution

How to sample a single discrete variable from a given distribution?

- Get a sample u from uniform distribution between [0,1).
    - In python : random()
- Map u to a specific instantiation of your random variable.

Weather (W)	P(W = w)
Sunny	0.3
Rain	0.3
Cold	0.3
Snow	0.1

$$0.0 \leq u < 0.3 \Rightarrow W = sunny$$
$$0.3 \leq u < 0.6 \Rightarrow W = rain$$
$$0.6 \leq u < 0.9 \Rightarrow W = cold$$
$$0.9 \leq u < 1.0 \Rightarrow W = snow$$

Sampling from a Distribution

Sample from a given distribution of a Variable.

- Given the distribution of discrete random variable *W.*

  values: $\{w_1, w_2, \ldots w_n\}$,

  corresponding probabilities: $p_1, p_2, \ldots, p_n$, $\sum_i p_i = 1$.

  - Get a sample u from uniform distribution in [0,1).
    In python : random()
  - Map u to a specific instantiation of W.

$$
\begin{array}{rl}
0 \leq u < p_1 & w_1 \\
p_1 \leq u < p_1 + p_2 & w_2 \\
p_1 + p_2 \leq u < p_1 + p_2 + p_3 & w_3 \\
p_1 + p_2 + \ldots + p_{n-1} \leq u < p_1 + p_2 + \ldots + p_{n-1} + p_n = 1 & w_n
\end{array}
$$

Direct Sampling

How to sample from a given distribution of Variables in BN?

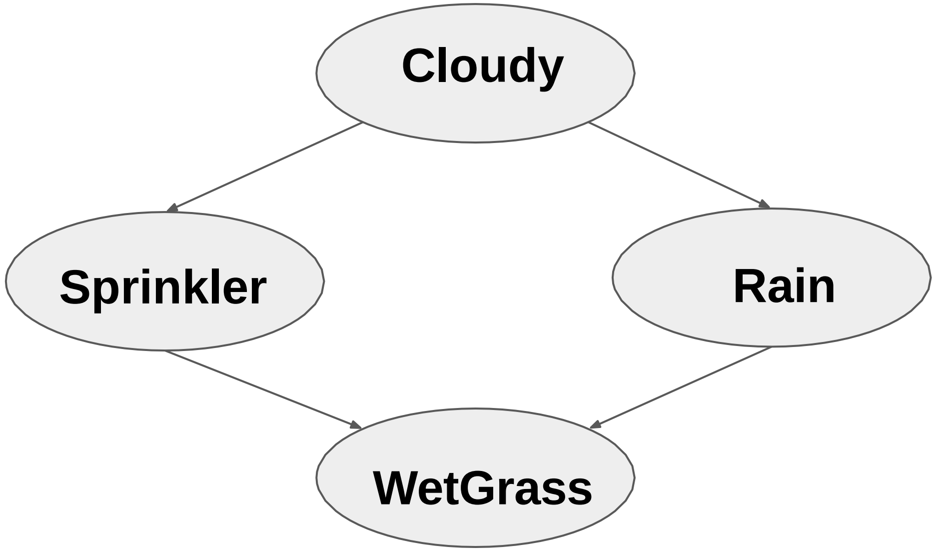
- Sample each variable in turn, in a topological order of a BN.
- Previously sampled variables value are used as condition for sampling the next node of BN.

Direct Sampling

How to sample from a given distribution of Variables in BN?

- Sample each variable in turn, in a topological order of a BN.
- Previously sampled variables value are used as condition for sampling the next node of BN.

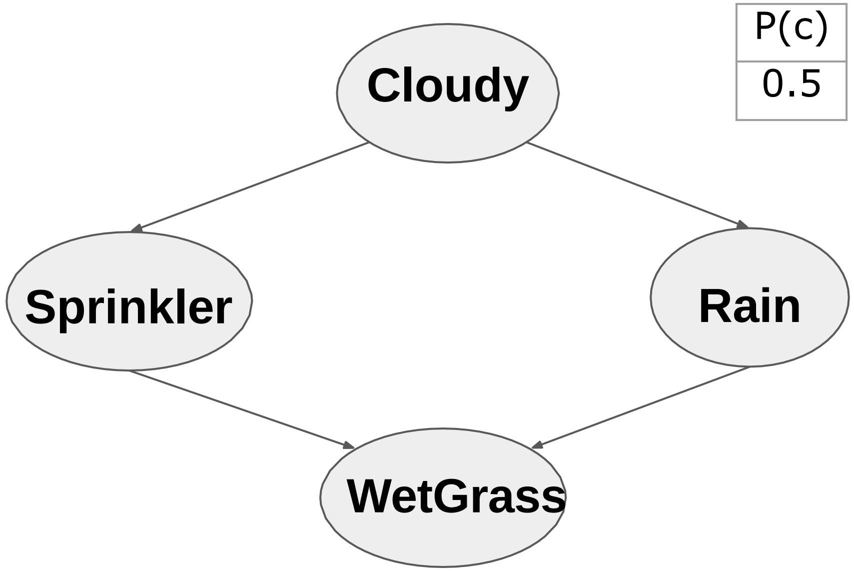
Example: Wet Grass Network.



Direct Sampling

$P(C)$

P(c)
0.5



$P(S|C)$

C	P(s|C)
c	0.1
-c	0.5

$P(R|C)$

C	P(r|C)
c	0.8
-c	0.2

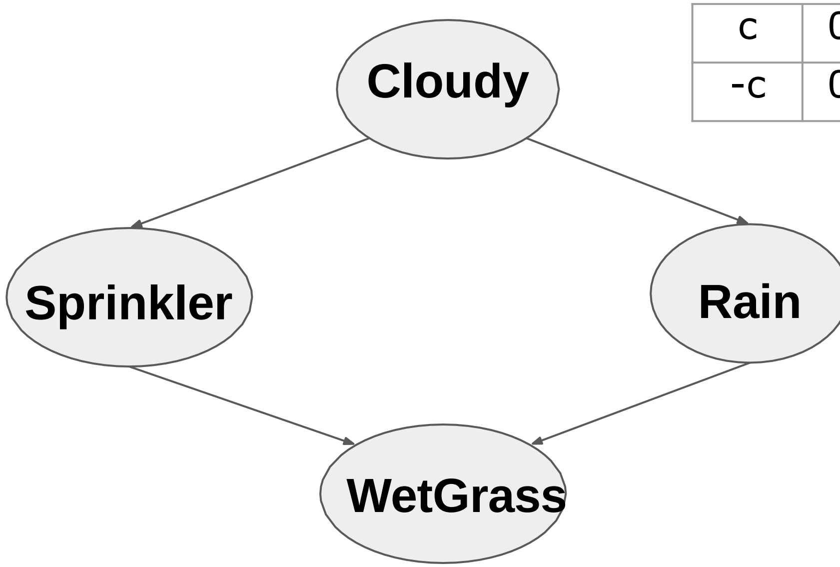
$P(W|S,R)$

S	R	P(w|S,R)
s	r	0.99
s	-r	0.90
-s	r	0.90
-s	-r	0.01

Direct Sampling

$P(C)$

c	0.5
-c	0.5



$P(S|C)$

c	s	0.1
	-s	0.9
-c	s	0.5
	-s	0.5

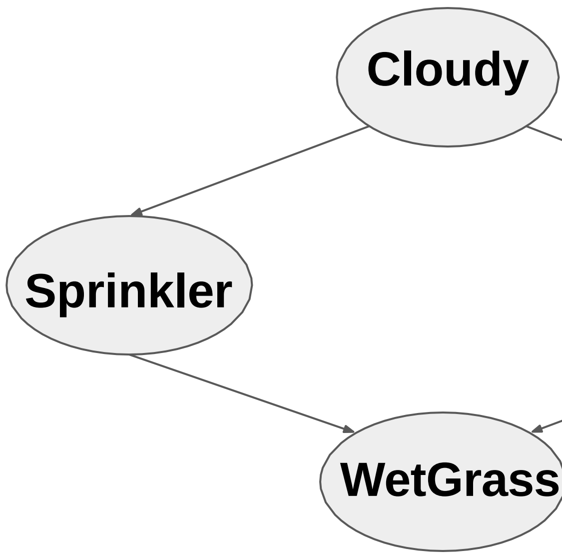
$P(R|C)$

c	r	0.8
	-r	0.2
-c	r	0.2
	-r	0.8

$P(W|S,R)$

s	r	w	0.99
		-w	0.01
	-r	w	0.90
		-w	0.10
-s	r	w	0.90
		-w	0.10
	-r	w	0.01
		-w	0.99

Direct Sampling



$P(C)$

c	0.5
-c	0.5

$P(S|C)$

c	s	0.1
	-s	0.9
-c	s	0.5
	-s	0.5

$P(R|C)$

c	r	0.8
	-r	0.2
-c	r	0.2
	-r	0.8

$P(W|S,R)$

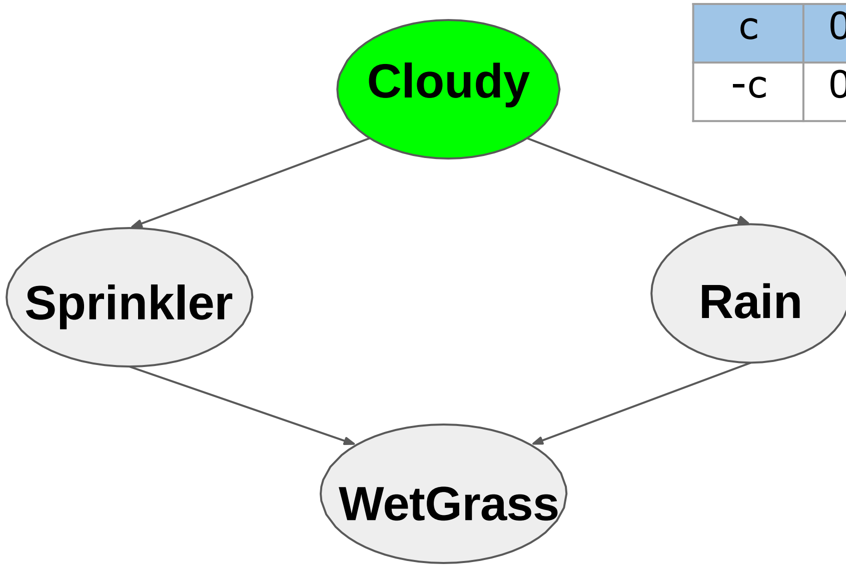
s	r	w	0.99
		-w	0.01
	-r	w	0.90
		-w	0.10
-s	r	w	0.90
		-w	0.10
	-r	w	0.01
		-w	0.99

- **Fix sampling order**
  - **{C, S, R,W}**
- **Sample examples given CPTs.**

Direct Sampling

$P(C)$

c	0.5
-c	0.5



$P(S|C)$

c	s	0.1
	-s	0.9
-c	s	0.5
	-s	0.5

$P(R|C)$

c	r	0.8
	-r	0.2
-c	r	0.2
	-r	0.8

$P(W|S,R)$

s	r	w	0.99
		-w	0.01
	-r	w	0.90
		-w	0.10
-s	r	w	0.90
		-w	0.10
	-r	w	0.01
		-w	0.99

$u = 0.22$

- **Fix sampling order**
  - **{C, S, R,W}**
- **Sample examples given CPTs.**
  - **{+c,**

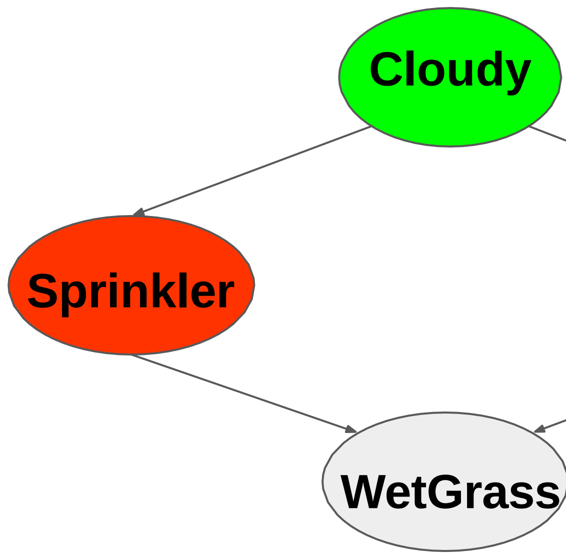
Direct Sampling

$P(C)$

c	0.5
-c	0.5

$P(S|C)$

c	s	0.1
	-s	0.9
-c	s	0.5
	-s	0.5



$P(R|C)$

c	r	0.8
	-r	0.2
-c	r	0.2
	-r	0.8

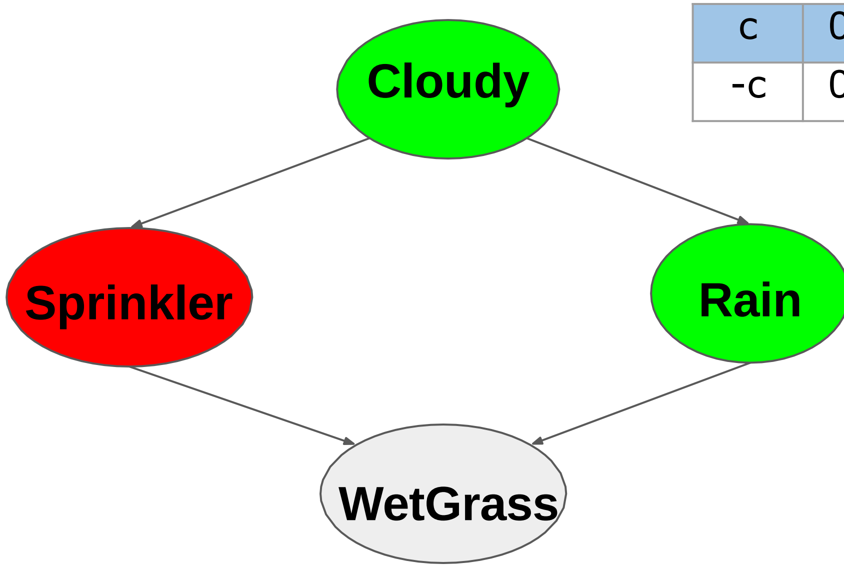
$P(W|S,R)$

s	r	w	0.99
		-w	0.01
	-r	w	0.90
		-w	0.10
-s	r	w	0.90
		-w	0.10
	-r	w	0.01
		-w	0.99

$u = 0.81$

- **Fix sampling order**
  - **{C, S, R,W}**
- **Sample examples given CPTs.**
  - **{+c, -s,**

Direct Sampling



$P(C)$

c	0.5
-c	0.5

$P(S|C)$

c	s	0.1
	-s	0.9
-c	s	0.5
	-s	0.5

$P(R|C)$

c	r	0.8
	-r	0.2
-c	r	0.2
	-r	0.8

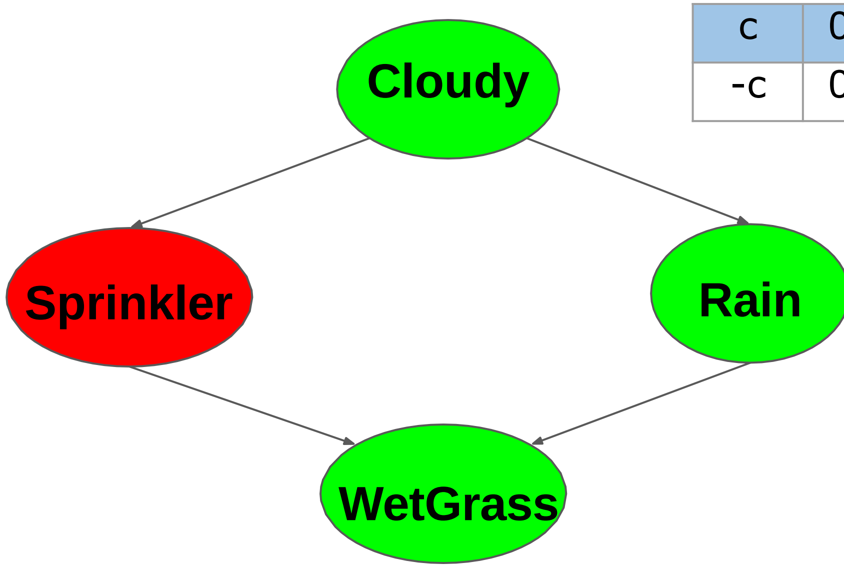
$P(W|S,R)$

s	r	w	0.99
		-w	0.01
	-r	w	0.90
		-w	0.10
-s	r	w	0.90
		-w	0.10
	-r	w	0.01
		-w	0.99

$u = 0.65$

- **Fix sampling order**
  - **{C, S, R,W}**
- **Sample examples given CPTs.**
  - **{+c, -s, +r,**

Direct Sampling



$P(C)$

c	0.5
-c	0.5

$P(S|C)$

c	s	0.1
	-s	0.9
-c	s	0.5
	-s	0.5

$P(R|C)$

c	r	0.8
	-r	0.2
-c	r	0.2
	-r	0.8

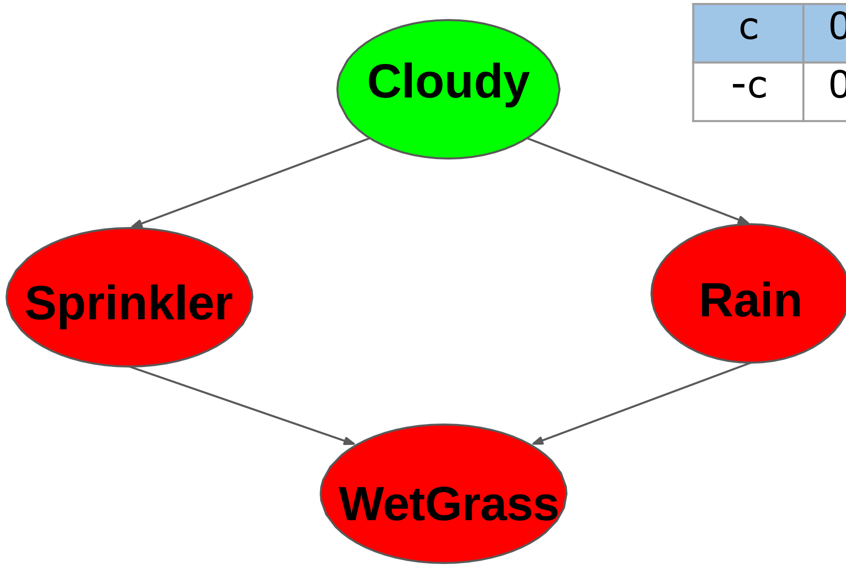
$P(W|S,R)$

s	r	w	0.99
		-w	0.01
	-r	w	0.90
		-w	0.10
-s	r	w	0.90
		-w	0.10
	-r	w	0.01
		-w	0.99

$u = 0.78$

- **Fix sampling order**
  - **{C, S, R,W}**
- **Sample examples given CPTs.**
  - **{+c, -s, +r, +w}**

Direct Sampling



$P(C)$

c	0.5
-c	0.5

$P(S|C)$

c	s	0.1
	-s	0.9
-c	s	0.5
	-s	0.5

$P(R|C)$

c	r	0.8
	-r	0.2
-c	r	0.2
	-r	0.8

$P(W|S,R)$

s	r	w	0.99
		-w	0.01
	-r	w	0.90
		-w	0.10
-s	r	w	0.90
		-w	0.10
	-r	w	0.01
		-w	0.99

- **Fix sampling order**
  - **{C, S, R,W}**
- **Sample examples given CPTs.**
  - **{+c, -s, +r, +w}**
  - **{-c, +s, -r, +w}**
  - **….**
  - **....**
  - **{+c, -s, -r, -w}**

Direct Sampling

- Given *N* samples, and the number of samples for a specific event is $N_{PS}(x_1, \ldots, x_n)$, then approximate inference with sampling gives the probability of this event:

$$S_{PS}(\text{-c,+s,+r,-w}) = \lim_{N\to\infty} \frac{N_{PS}(-c,s,r,-w)}{N}$$

$$S_{PS}(\text{-c,+s,+r,-w}) \approx \frac{N_{PS}(-c,s,r,-w)}{N}$$

- Why direct sampling works?

The sampling process generates samples with following probability as each sampling step depends only on the parent values:

$$S_{PS}(x_1 \ldots x_n) = \prod_{i=1}^{n} P(x_i \,|\, parents(X_i)) = P(x_1 \ldots x_n)$$

Recall global semantics of the BN.

Direct Sampling

We frequently want to estimate the probability of partially specified events $P(x_1, \ldots, x_m)$ with $m < n$.

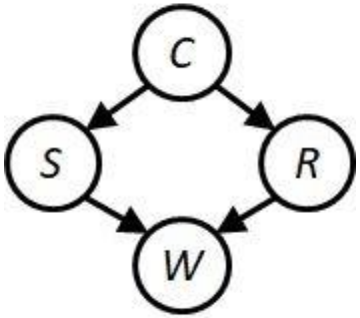
This can be approximated by

$$P(x_1, \ldots, x_m) \approx \frac{N_{PS}(x_1, \ldots, x_m)}{N}$$

where $N_{PS}(x_1, \ldots, x_m)$ is now the number of samples among N

Direct Sampling

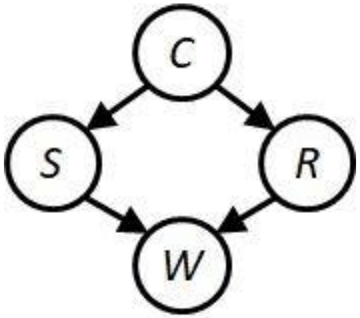
- Given following set of samples:
  - {-c, +s, -r, +w}
  - {-c, -s, +r, +w}
  - {-c, +s, +r, -w}
  - {-c, +s, -r, +w}
  - {+c, +s, +r, +w}
  - {-c, -s, +r, +w}
  - {+c, +s, -r, -w}
  - {-c, +s, +r, -w}



- P(-c, +s, +r, -w) = ?
- P(+r) = ?

Direct Sampling

- Given following set of samples:
  - {-c, +s, -r, +w}
  - {-c, -s, +r, +w}
  - {-c, +s, +r, -w}
  - {-c, +s, -r, +w}
  - {+c, +s, +r, +w}
  - {-c, -s, +r, +w}
  - {+c, +s, -r, -w}
  - {-c, +s, +r, -w}



- P(-c, +s, +r, -w) = 2/8 = 0.25
- P(+r) = 5/8 = 0.625
- P(R |+w) = ?, P(-s|+w) = ?

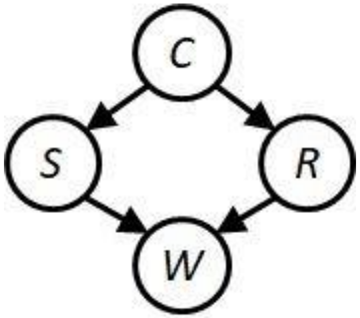
Rejection Sampling

$$P(X|e)$$

- Generate samples as follows.
  - {-c, +s, -r, +w}
  - {-c, -s, +r, +w}
  - {-c, +s, +r, -w}
  - {-c, +s, -r, +w}
  - {+c, +s, +r, +w}
  - {-c, -s, +r, +w}
  - {+c, +s, -r, -w}
  - {-c, +s, +r, -w}

P(R | +w) = ?

P(-s | +w) = ?



- Rejects the samples which does not match the evidence.

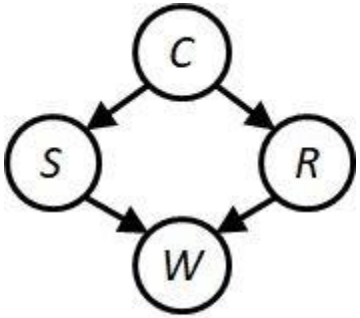
Rejection Sampling

$$P(X|e)$$

- Generate samples as follows.
    - {-c, +s, -r, +w}
    - {-c, -s, +r, +w}
    -
    - {-c, +s, -r, +w}
    - {+c, +s, +r, +w}
    - {-c, -s, +r, +w}
    -
    -

P(R | +w) = ?

P(-s | +w) = ?



- Rejects the samples which does not match the evidence.
- $\hat{P}(X|e)$ is estimated by counting how many times X = x occurs for samples which are consistent with observations.

Rejection Sampling

$$P(X|e)$$

- Generate samples as follows.
  - {-c, +s, -r, +w}
  - {-c, -s, +r, +w}
  -
  - {-c, +s, -r, +w}
  - {+c, +s, +r, +w}
  - {-c, -s, +r, +w}
  -
  -

$$P(X|e) \approx \frac{N_{PS}(X,e)}{N_{PS}(e)}$$

P(R | +w) = <3/5,2/5>

P(-s | +w) = 2/5

- Rejects the samples which does not match the evidence.
- $\hat{P}(X|e)$ is estimated by counting how many times X = x occurs for samples which are consistent with observations.

Likelihood Weighting

- **Rejection sampling:**

  Inefficient with P(e) being small: We sample too many examples that are inconsistent with evidence.

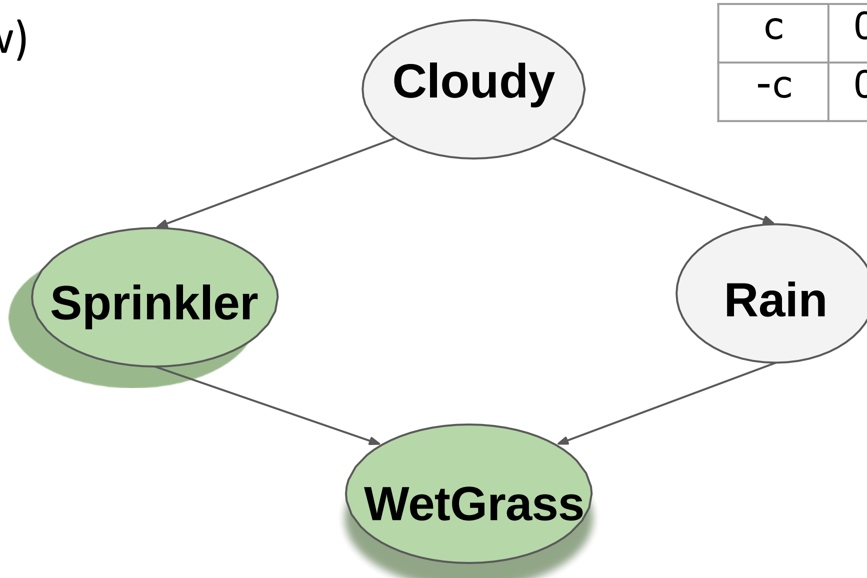
- **Likelihood weighting**
  - Samples examples which are consistent with evidence.
  - Each sample have a support value w (i.e., weight).
    - Initialize w of the generated sample as w = 1
    - Repeat
      - If variable is non-evidence : sample as usual.
      - If variable is evidence variable E: set E = e, and set $w = w * P(E = e \mid parents(E))$

Example : Likelihood Weighting

P(R|s,w)
Evidence: (+s,+w)

$P(C)$

c	0.5
-c	0.5



$P(S|C)$

c	s	0.1
	-s	0.9
-c	s	0.5
	-s	0.5

$P(R|C)$

c	r	0.8
	-r	0.2
-c	r	0.2
	-r	0.8

$P(W|S,R)$

s	r	w	0.99
		-w	0.01
	-r	w	0.90
		-w	0.10
-s	r	w	0.90
		-w	0.10
	-r	w	0.01
		-w	0.99

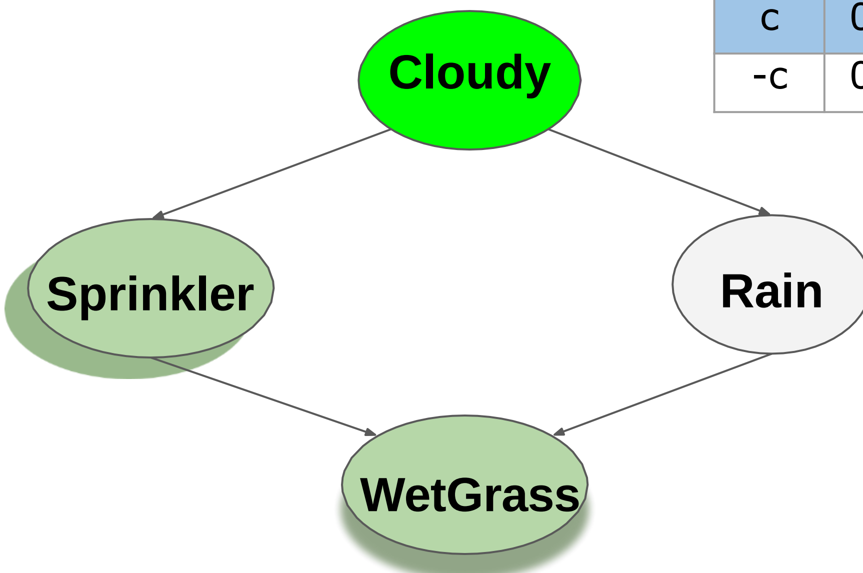
- **Fix sampling order {C, S, R, W}**
- **Likelihood weighted sampling with (+s, +w)**
  - **{_, _, _, _} , W = 1**

Example : Likelihood Weighting

Evidence: (+s,+w)

$P(C)$

c	0.5
-c	0.5



$P(S|C)$

c	s	0.1
	-s	0.9
-c	s	0.5
	-s	0.5

$P(R|C)$

c	r	0.8
	-r	0.2
-c	r	0.2
	-r	0.8

$P(W|S,R)$

s	r	w	0.99
		-w	0.01
	-r	w	0.90
		-w	0.10
-s	r	w	0.90
		-w	0.10
	-r	w	0.01
		-w	0.99

- **Fix sampling order {C, S, R,W}**
- **Likelihood weighted sampling with (+s, +w)**
  - **{+c, _, _, _} , W = 1**

$u = 0.22$

Example : Likelihood Weighting

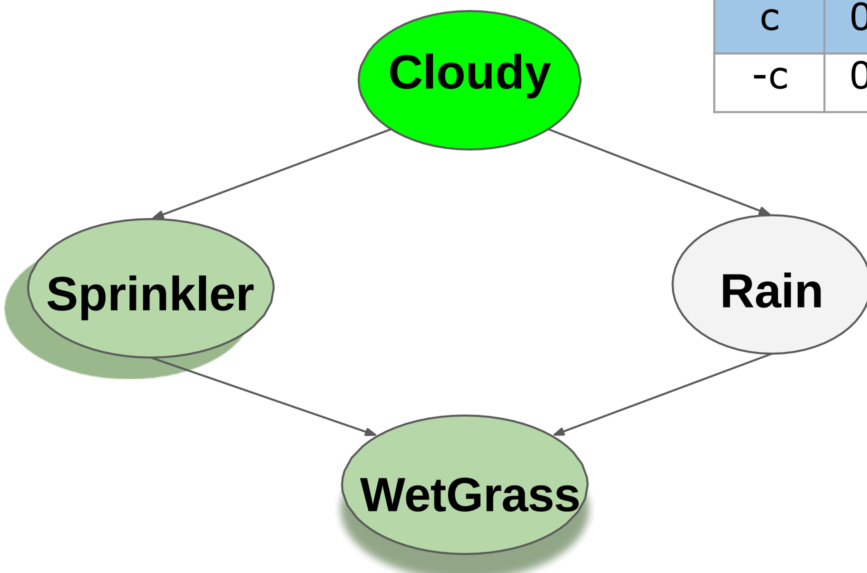
Evidence: (+s,+w)

$P(C)$

c	0.5
-c	0.5

$P(S|C)$

c	s	0.1
	-s	0.9
-c	s	0.5
	-s	0.5



$P(R|C)$

c	r	0.8
	-r	0.2
-c	r	0.2
	-r	0.8

$P(W|S,R)$

s	r	w	0.99
		-w	0.01
	-r	w	0.90
		-w	0.10
-s	r	w	0.90
		-w	0.10
	-r	w	0.01
		-w	0.99

- **Fix sampling order {C, S, R,W}**
- **Likelihood weighted sampling with (+s, +w)**
  - **{+c, +s, _, _} , W = 1*0.1**

$P(+s|+c)$

$u = 0.81$

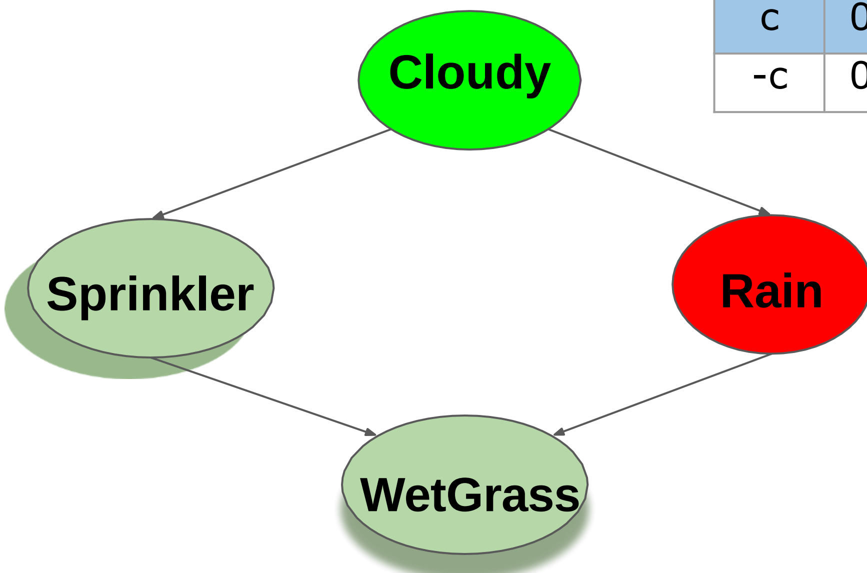
Not Used!

Example : Likelihood Weighting

Evidence: (+s,+w)

$P(C)$

c	0.5
-c	0.5



$P(S|C)$

c	s	0.1
	-s	0.9
-c	s	0.5
	-s	0.5

$P(R|C)$

c	r	0.8
	-r	0.2
-c	r	0.2
	-r	0.8

$P(W|S,R)$

s	r	w	0.99
		-w	0.01
	-r	w	0.90
		-w	0.10
-s	r	w	0.90
		-w	0.10
	-r	w	0.01
		-w	0.99

- **Fix sampling order {C, S, R,W}**
- **Likelihood weighted sampling with (+s, +w)**
  - **{+c, +s, -r, _} , W = 1*0.1**

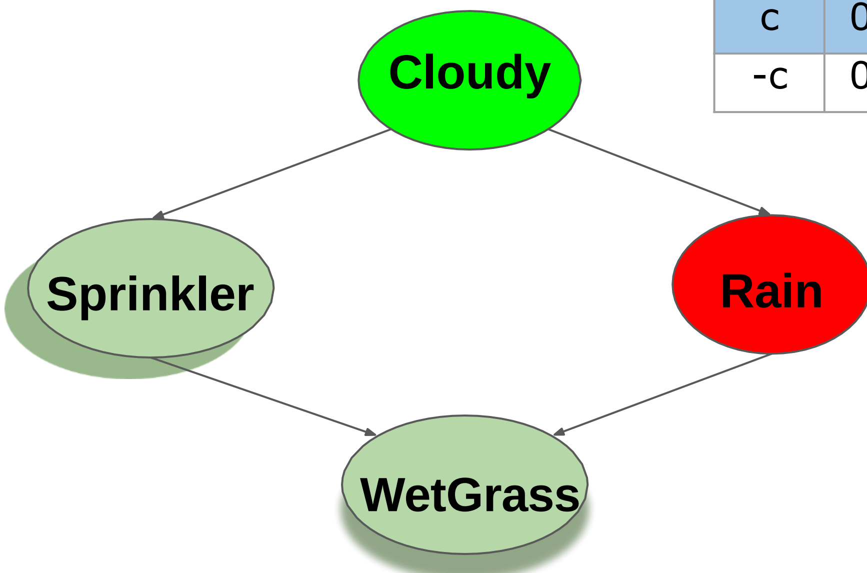
$u = 0.95$

Example : Likelihood Weighting

Evidence: (+s,+w)

$P(C)$

c	0.5
-c	0.5



$P(S|C)$

c	s	0.1
	-s	0.9
-c	s	0.5
	-s	0.5

$P(R|C)$

c	r	0.8
	-r	0.2
-c	r	0.2
	-r	0.8

$P(W|S,R)$

s	r	w	0.99
		-w	0.01
	-r	w	0.90
		-w	0.10
-s	r	w	0.90
		-w	0.10
	-r	w	0.01
		-w	0.99

- **Fix sampling order {C, S, R,W}**
- **Likelihood weighted sampling with (+s, +w)**
  - **{+c, +s, -r, +w} , W = 1*0.1*0.9**

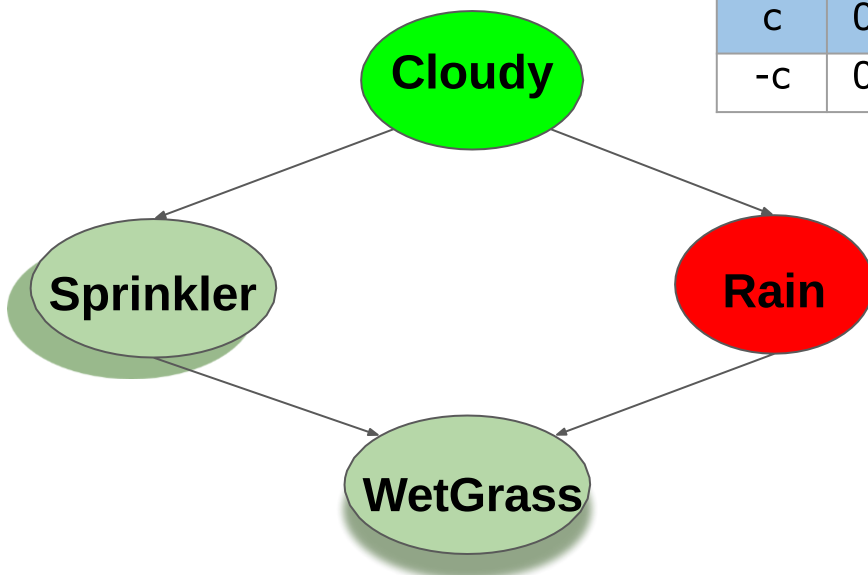
u = 0.78
Not Used

P(+s|+c) P(+w|+s,-r)

Evidence: (+s,+w)

Example : Likelihood Weighting

Evidence: (+s,+w)



$P(C)$

c	0.5
-c	0.5

$P(S|C)$

c	s	0.1
	-s	0.9
-c	s	0.5
	-s	0.5

$P(R|C)$

c	r	0.8
	-r	0.2
-c	r	0.2
	-r	0.8

$P(W|S,R)$

s	r	w	0.99
		-w	0.01
	-r	w	0.90
		-w	0.10
-s	r	w	0.90
		-w	0.10
	-r	w	0.01
		-w	0.99

- **Fix sampling order {C, S, R,W}**
- **Likelihood weighted sampling with (+s, +w)**
    - **{+c, +s, -r, +w} , W = 0.09**

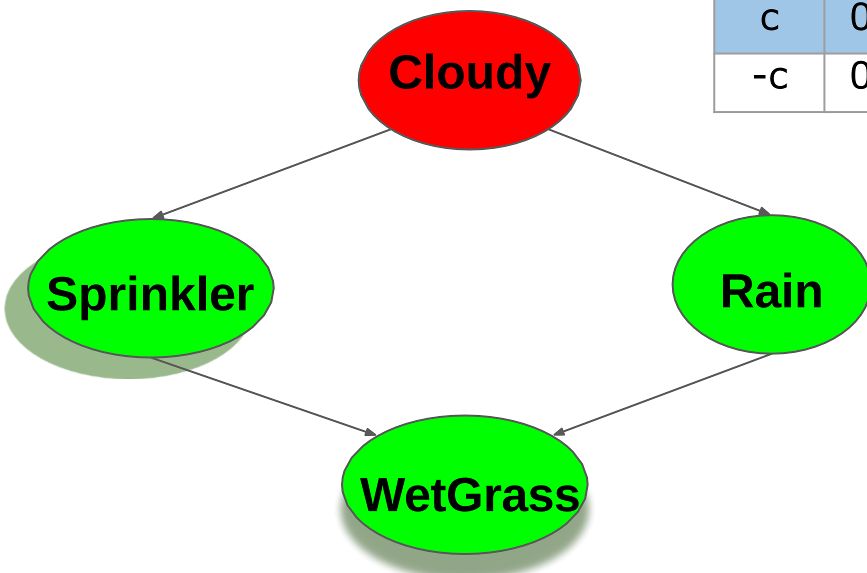
w is small,
meaning we have less confidence in our sample.

Example : Likelihood Weighting

Evidence: (+s,+w)

$P(C)$

c	0.5
-c	0.5



$P(S|C)$

c	s	0.1
	-s	0.9
-c	s	0.5
	-s	0.5

$P(R|C)$

c	r	0.8
	-r	0.2
-c	r	0.2
	-r	0.8

$P(W|S,R)$

s	r	w	0.99
		-w	0.01
	-r	w	0.90
		-w	0.10
-s	r	w	0.90
		-w	0.10
	-r	w	0.01
		-w	0.99

- **Fix sampling order {C, S, R,W}**
- **Likelihood weighted sampling with (+s, +w)**
  - **{+c, +s, -r, +w}, W = 0.09**
  - **{-c, +s, -r, +w}, W = 0.45**
  - **....**
  - **{+c, +s, +r, +w}, W = 0.099**

Example : Likelihood Weighting

Say the following $N = 100$ samples were generated from the wetgrass network with associated likelihood/weights:

c s r w

- 3 samples of $[true, true, true, true]$ with $w = 0.099$
- 2 samples of $[true, true, false, true]$ with $w = 0.09$
- 55 samples of $[false, true, true, true]$ with $w = 0.495$
- 40 samples of $[false, true, false, true]$ with $w = 0.45$

Notice that all samples are consistent with the evidence $Sprinkler = true$ and $WetGrass = true$.

The desired probability estimate is

P(R|s,w)=P(R,s,w)/P(s,w)= αP(R,s,w)= $\alpha\sum_C \boldsymbol{P(C,R,s,w,)}$

$$\hat{P}(Rain|Sprinkler = true, WetGrass = true)$$

(c,r,s,w) (-c,r,s,w)

$$= \alpha\langle 3 \times 0.099 + 55 \times 0.495, 2 \times 0.09 + 40 \times 0.45\rangle$$

(c,-r,s,w) (-c,-r,s,w)

$$= \alpha\langle 27.522, 18.18\rangle = \langle 0.60, 0.40\rangle$$

Likelihood Weighting

$$P(X|e)$$

For an arbitrary Bayesian Network, let X be the query variable, e be the values of the evidence variables and Y be the unobserved variables.

Let

- $N_{WS}(X, Y, e)$ be the number of samples generated for the event X, Y and e.
- $w(X, Y, e)$ be the weight of a sample corresponding to the event X, Y and e.

The the estimate for $P(X|e)$ is

$$\hat{P}(X|e) = \alpha \sum_{\forall Y} N_{WS}(X, Y, e) w(X, Y, e)$$

Likelihood Weighting : Why it works?

- In a BN, let **E** represents all evidence variables, **Z** represents all nonevidence variables including the query variable X. The sampling probability distribution is:

l: the number of nonevidence variables.

$$S_{WS}(\mathbf{z}, \mathbf{e}) = \prod_{i=1}^{l} P(z_i \,|\, parents(Z_i))$$

The sampling process: $S_{WS}(C, s, R, w) = P(C)P(R|C)$

- Now, samples have weights.

m: the number of nonevidence variables.

$$w(\mathbf{z}, \mathbf{e}) = \prod_{i=1}^{m} P(e_i \,|\, parents(E_i))$$

e.g., w=P(s|c)P(w|s,-r) for {+c, +s, -r, +w}

Likelihood Weighting : Why it works?

- Together, weighted sampling distribution is consistent.

$$S_{WS}(\mathbf{z},\mathbf{e})w(\mathbf{z},\mathbf{e}) = \prod_{i=1}^{l} P(z_i \,|\, parents(Z_i)) \prod_{i=1}^{m} P(e_i \,|\, parents(E_i))$$
$$= P(\mathbf{z},\mathbf{e})$$

- $$P(X|e) = \alpha \sum_{\forall Y} P(X,Y,e)$$
$$= \alpha \sum_{\forall Y} S_{WS}(X,Y,e)w(X,Y,e)$$
$$= \lim_{N\to\infty} \alpha \sum_{\forall Y} N_{WS}(X,Y,e)w(X,Y,e)$$
$$\approx \alpha \sum_{\forall Y} N_{WS}(X,Y,e)w(X,Y,e)$$

X: query variable
Y: nonevidence variable and not X
Z: includes X and Y.
{X,Y, e}: all random variables

MCMC Methods

- **Likelihood weighting**
  - More efficient than rejection sampling.
  - Efficiency decreases if the samples have low weights.

    i.e. w approaches 0, when $P(\mathbf{e})$ is small.
  - You sample downstream in BN without reasoning about evidence probability being low.

$P(C \mid s, r, w)$

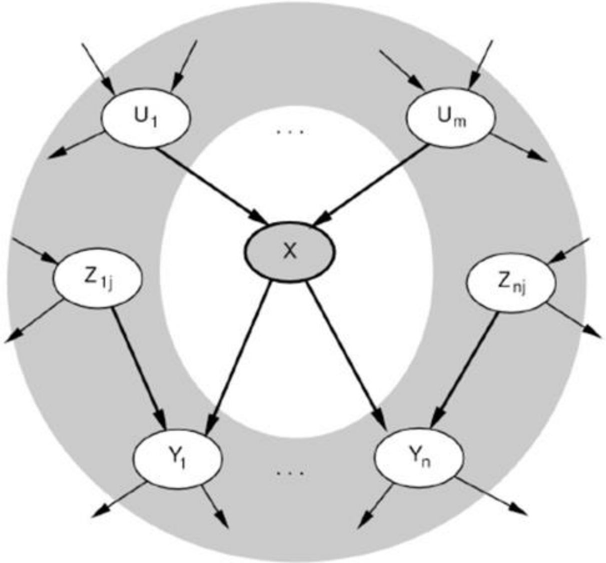
- **Markov Chain Monte Carlo (MCMC) Methods**
  - Generate samples with high probability accounting for evidence being low probability.
  - Gibbs sampling is a special instance of MCMC methods which we will study.

Gibbs Sampling

- Generates an event by making a random change to preceding event
  - Think that network is in a **current state** which species an event.
  - **Next state** is reached by sampling a value for one non-evidence variable X to be conditioned on the current values of X's **Markov blanket variables.**
  - Gibbs sampler thus wonders randomly in the state space by flipping one variable at a time while keeping evidence variables fixed.
- As sampling settles into a dynamic equilibrium, the fraction of time spend on each state is proportional its posterior probability.

Recall Markov Blanket

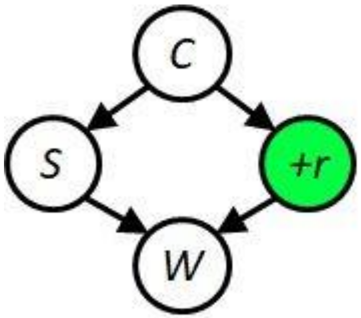
Recall that the Markov Blanket of a variable comprises of the parents, children, and children's parents of the variable.



In a BN, a node is conditionally independent of all others given the Markov Blanket of the node.

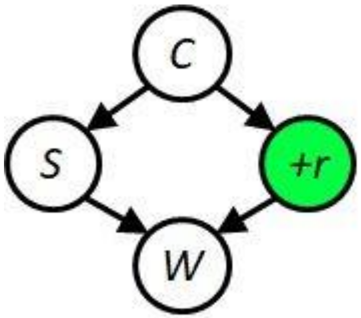
Gibbs Sampling Example (P(S|+r))

Step 1: initialize evidence

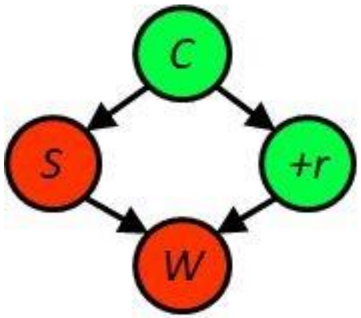


Gibbs Sampling Example (P(S|+r))

Step 1: initialize evidence



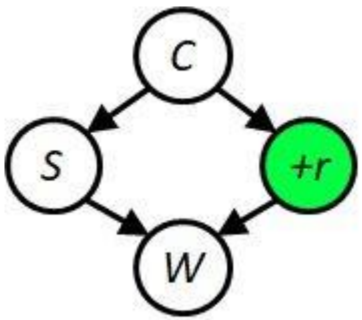
Step 2: initialize other variables (random)



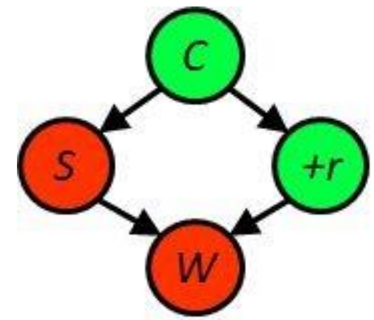
Initial state, e.g,: {+c, -s, +r, –w}

Gibbs Sampling Example (P(S|+r))

Step 1: initialize evidence



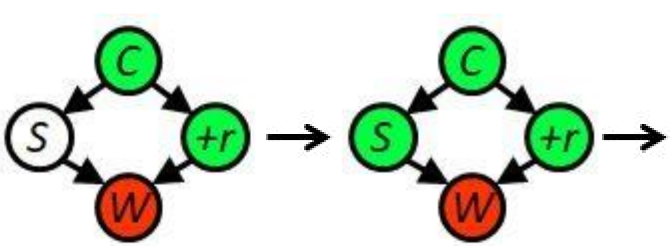
Step 2: initialize other variables (random)



Initial state, e.g,: {+c, -s, +r, –w}

Step 3: Repeat following

- Choose a nonevidence variable X (at random). Here X is S, C, W.
- Sample X given the current values of X's Markov blanket variables.

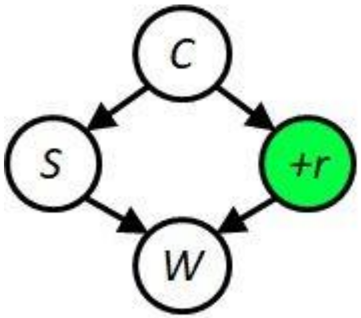


Sample from $P(S|+c,-w,+r)$

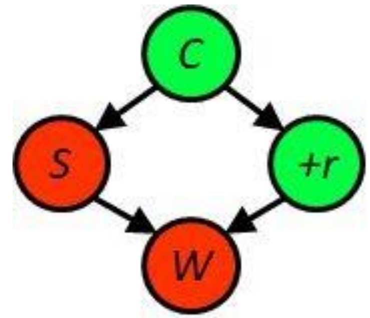
Suppose the result is true, the we get a new sample {+c, +s, +r, –w}

Gibbs Sampling Example (P(S|+r))

Step 1: initialize evidence



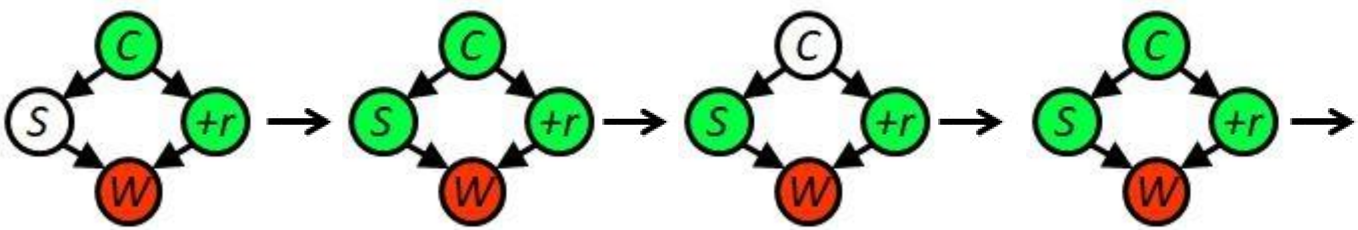
Step 2: initialize other variables (random)



Initial state, e.g,: {+c, -s, +r, –w}

Step 3: Repeat following

- Choose a nonevidence variable X (at random). Here X is S, C, W.
- Sample X given the current values of X's Markov blanket variables.



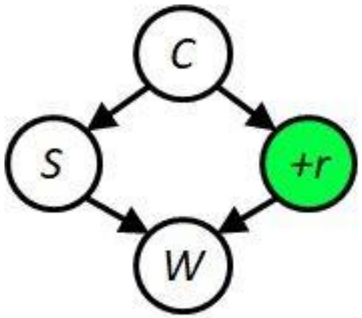
{+c, +s, +r, –w}

Sample from $P(C| + s, +r)$

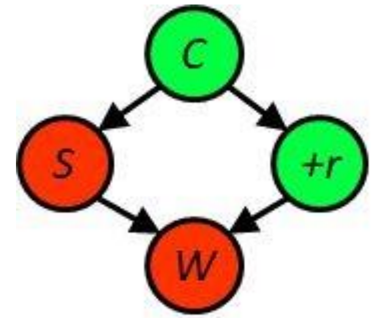
Suppose the result is true, then we get another sample {+c, +s, +r, –w}

Gibbs Sampling Example (P(S|+r))

Step 1: initialize evidence



Step 2: initialize other variables (random)

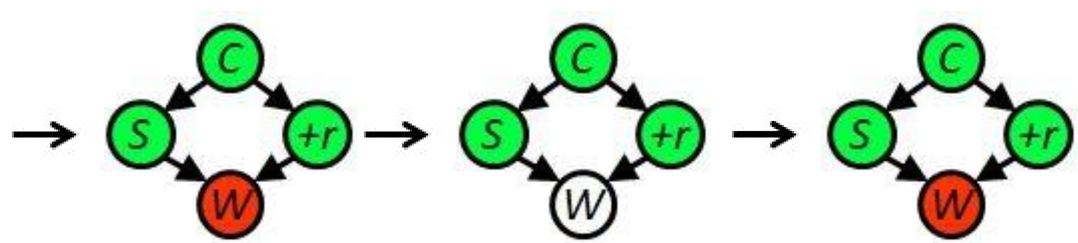


Initial state, e.g,: {+c, -s, +r, –w}

Step 3: Repeat following

- Choose a nonevidence variable X (at random). Here X is S, C, W.
- Sample X given the current values of X's Markov blanket variables.

……



{+c, +s, +r, –w}

Sample from $P(W|+s,+r)$

{+c, +s, +r, -w}

Gibbs Sampling Example (P(S|+r))

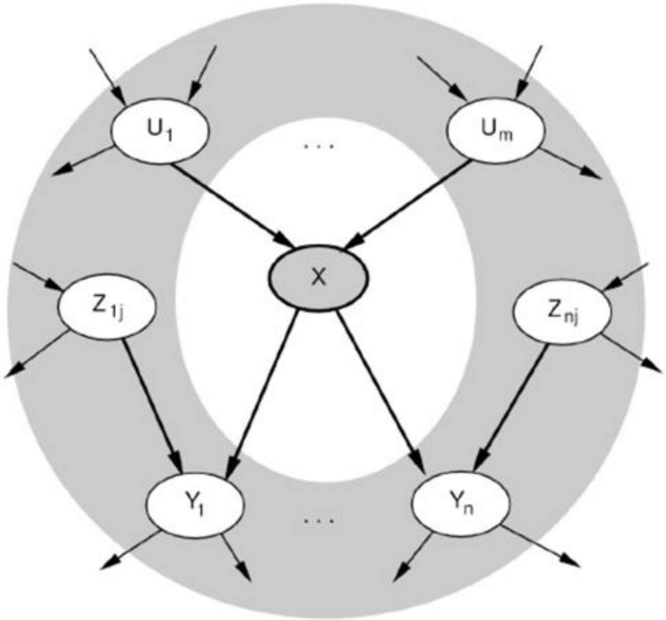
Now suppose we get 100 samples with Gibbs Sampling.

- All samples were satisfying observation, i.e. {**Rain = true**}
- 37 of them had {**Sprinkler = true**}
- Which means, 63 of them had {**Sprinkler = false**}

$$P(S|Rain = true) = \alpha < 37, 63 >$$

$$=< 0.37, 0.63 >$$

Probability given Markov Blanket



Let $\mathbf{Y}$ be the children of X_i

$\mathbf{Z}_j$ be the parents of Y_j other than X_i

$$\begin{aligned}
&\mathbf{P}(X_i|MB(X_i)) \\
&= \mathbf{P}(X_i|Parents(X_i), \mathbf{Y}, \mathbf{Z}_1, \ldots, \mathbf{Z}_\ell) \\
&= \alpha \mathbf{P}(X_i|Parents(X_i), \mathbf{Z}_1, \ldots, \mathbf{Z}_\ell)\mathbf{P}(\mathbf{Y}|Parents(X_i), X_i, \mathbf{Z}_1, \ldots, \mathbf{Z}_\ell) \\
&= \alpha \mathbf{P}(X_i|Parents(X_i))\mathbf{P}(\mathbf{Y}|X_i, \mathbf{Z}_1, \ldots, \mathbf{Z}_\ell) \\
&= \alpha \mathbf{P}(X_i|Parents(X_i)) \prod_{Y_j \in Children(X_i)} P(Y_j|Parents(Y_j))
\end{aligned}$$

Gibbs Sampling Example (P(S|+r))

$$\text{Sample from } P(S|+c, -w, +r)$$

$$\text{Sample from } P(C|+s, +r)$$

$$P\left(X|mb(X)\right)$$

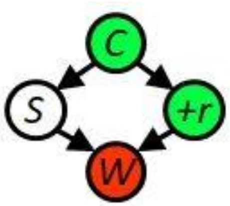
- As in BN,

P(X|variables of Markov Blanket of X)

$$= \alpha P\left(X|parents\left(X\right)\right) \times \prod_{Y_j \in Children(X)} P\left(Y_j|parents\left(Y_j\right)\right)$$

That is,

P(S|+c, -w, +r) = αP(S|+c)P(-w|+r)



P(C|+s, +r) = α'P(C)P(+s|C)P(+r|C)