Faculty of SET / School of Computer and Mathematical Sciences

COMP SCI 3007/7059/7659
Artificial Intelligence
Approximate Inference

Acknowledgement of Country

We acknowledge and pay our respects to the Kaurna people, the traditional custodians whose ancestral lands we gather on.

We acknowledge the deep feelings of attachment and relationship of the Kaurna people to the country and we respect and value their past, present and ongoing connection to the land and cultural beliefs.

Approximate Inference

AIMA C14.5

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 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?

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

e.g. 1000 trials and 30 times we get 3, then p(X=3) = 30/1000

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

$$
\begin{aligned}
0.0 \leq u < 0.3 &\Rightarrow W = sunny \\
0.3 \leq u < 0.6 &\Rightarrow W = rain \\
0.6 \leq u < 0.9 &\Rightarrow W = cold \\
0.9 \leq u < 1.0 &\Rightarrow W = snow
\end{aligned}
$$

e.g. we get 1000 samples, and number of Sunny is 200, so from the sample, the probability is 0.2, which is different with 0.3 in the table.

Why?

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*.

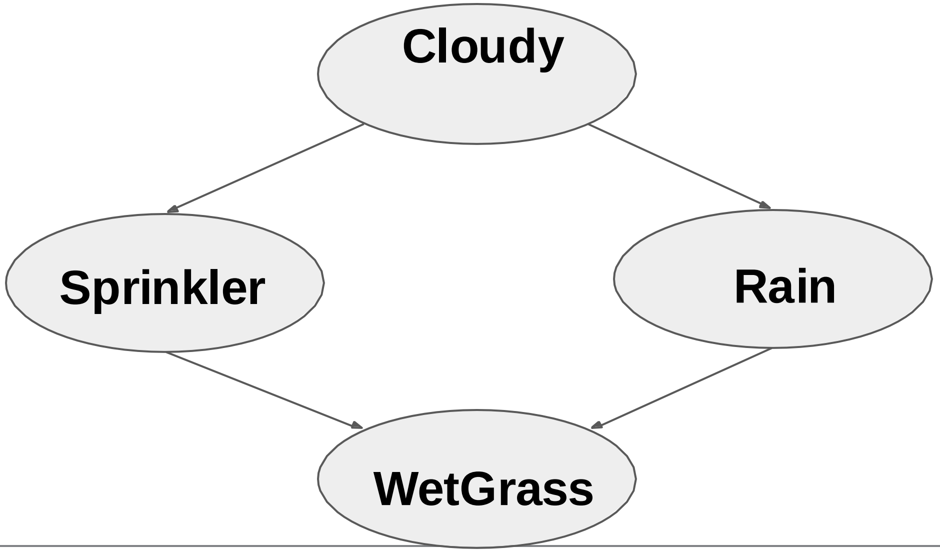
w_1	$0 \leq u < p_1$
w_2	$p_1 \leq u < p_1 + p_2$
w_3	$p_1 + p_2 \leq u < p_1 + p_2 + p_3$
$\vdots$	$\vdots$
w_n	$p_1 + p_2 + \ldots + p_{n-1} \leq u < p_1 + p_2 + \ldots + p_{n-1} + p_n = 1$

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.

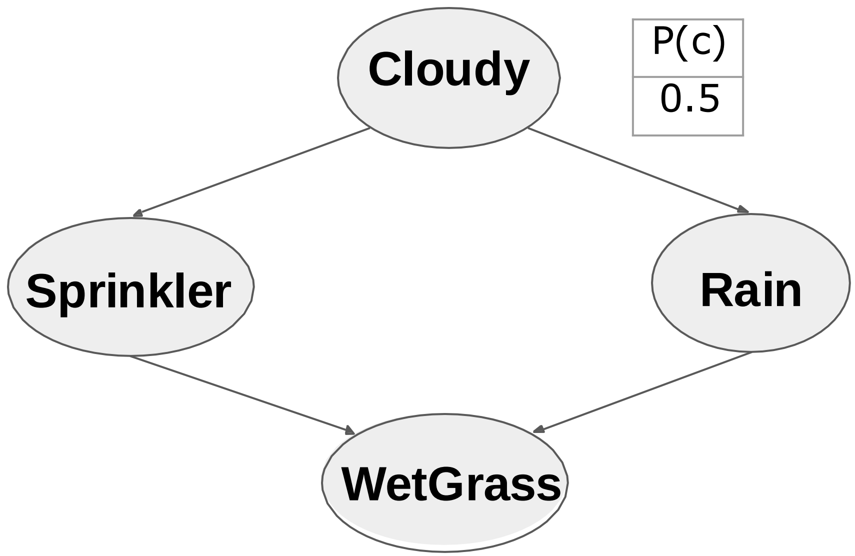
Example: Wet Grass Network.



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

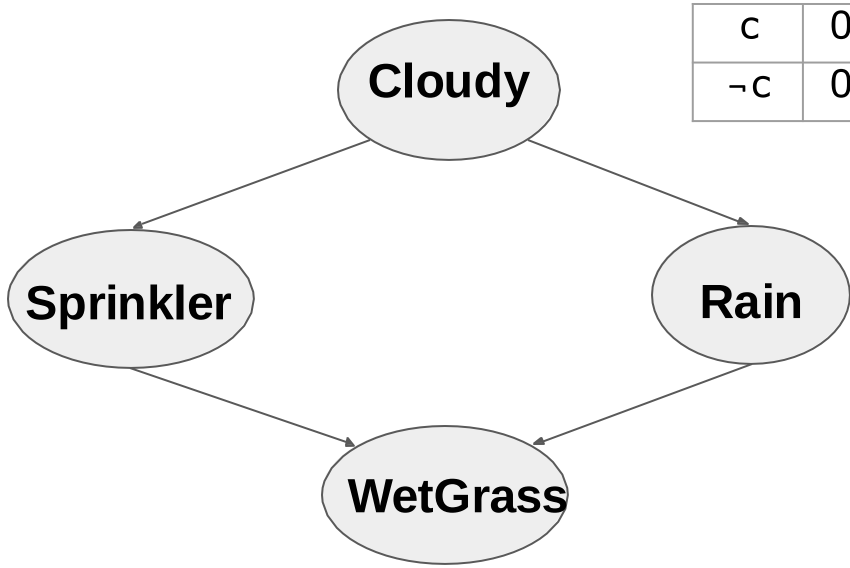
Direct Sampling

$P(C)$

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

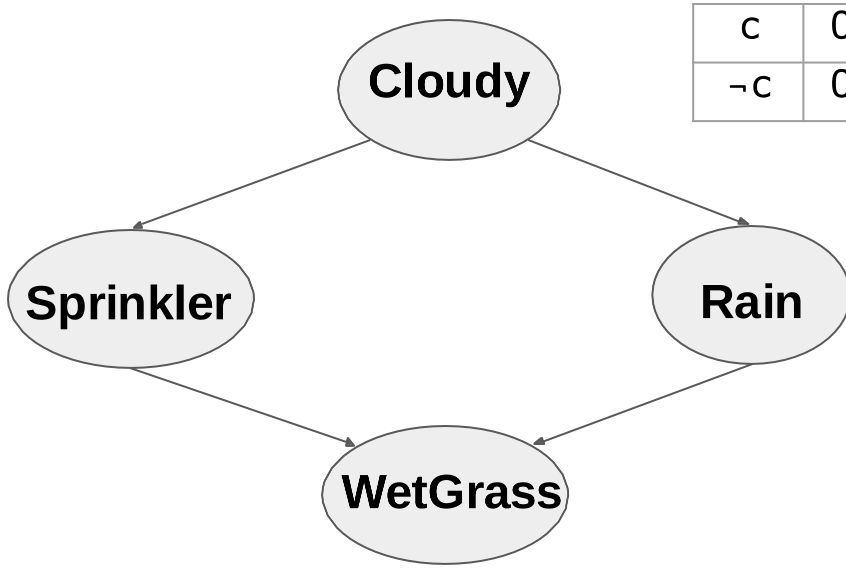
$P(W|S,R)$

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

$P(W|S, R)$

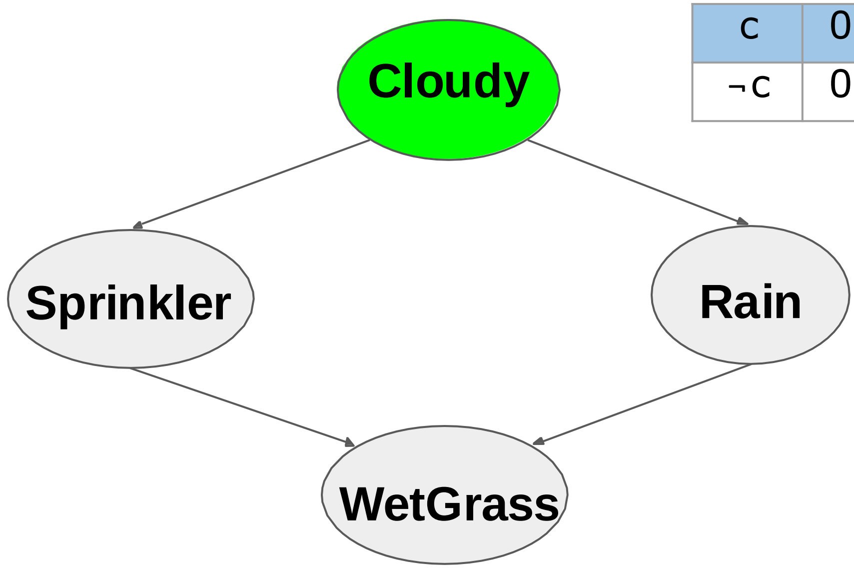
s	r	w	0.99
		¬w	0.01
	¬r	w	0.90
		¬w	0.10
¬s	r	w	0.90
		¬w	0.10
	¬r	w	0.01
		¬w	0.99

- **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

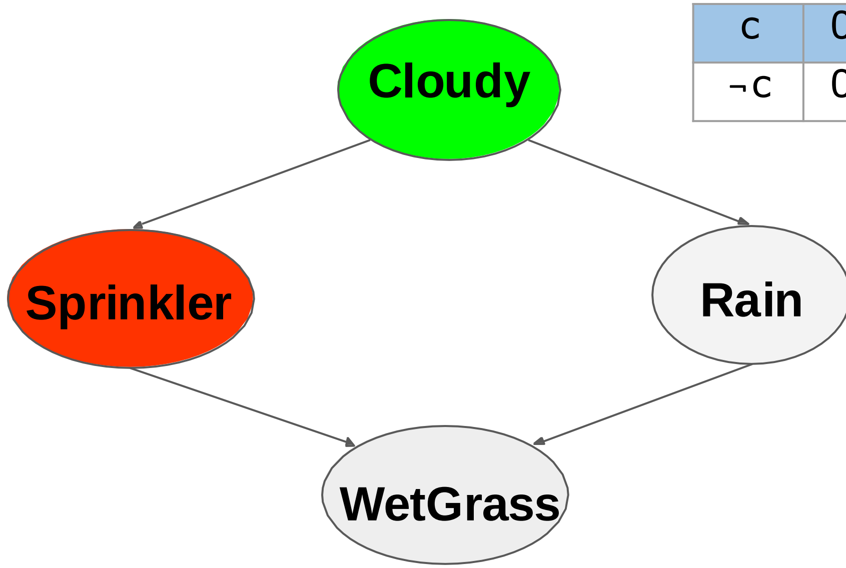
$u = 0.22$

- **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

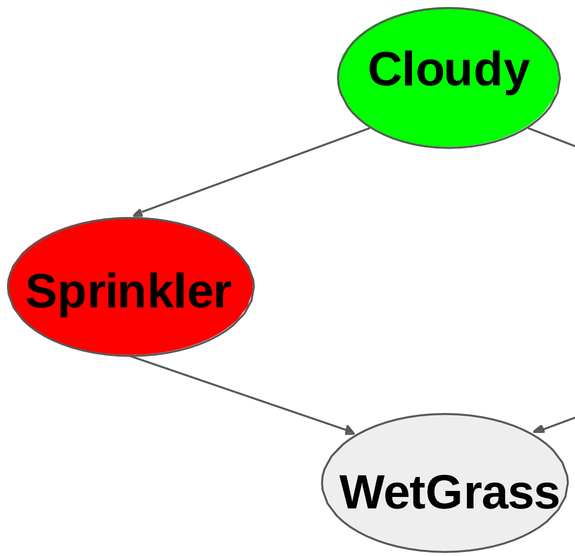
$u = 0.81$

- **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

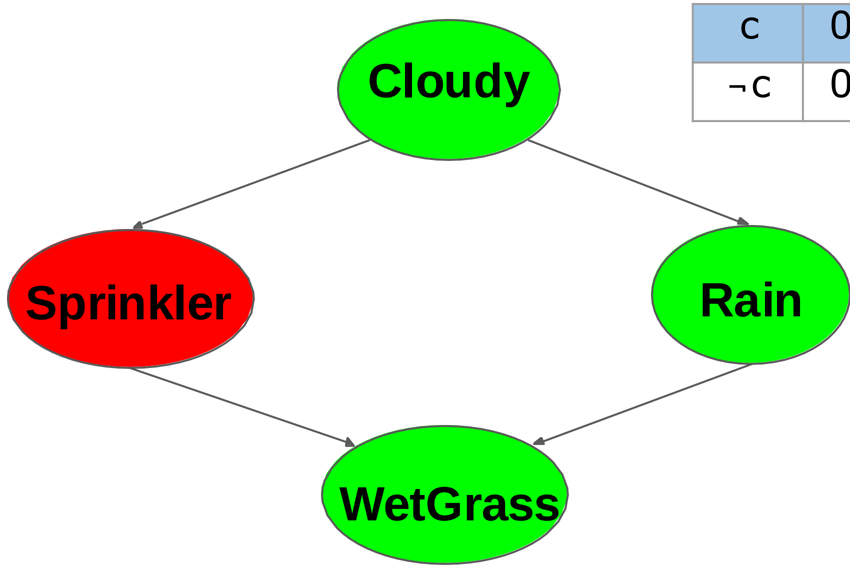
$P(W|S,R)$

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

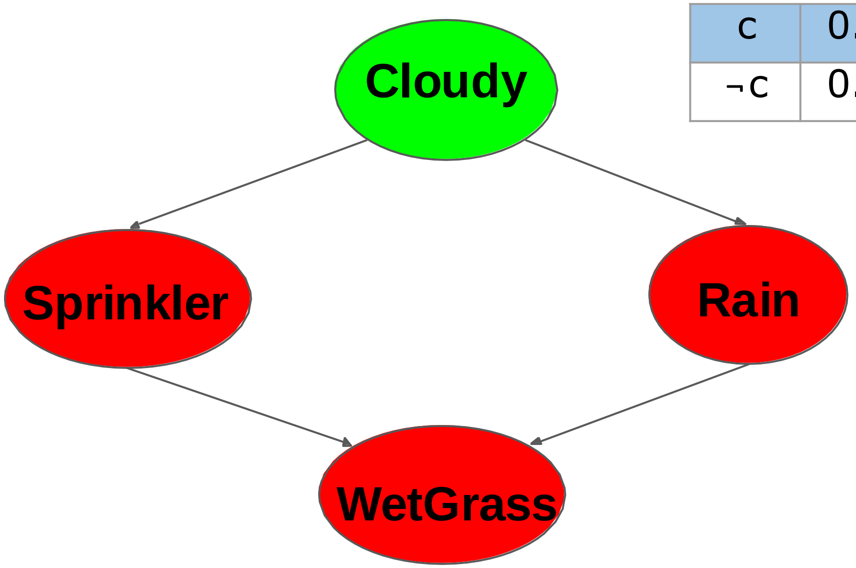
$u = 0.78$

- **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:

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

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

- Why does direct sampling work?

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.

Galton board

Attribution: Matemateca (IME USP)
https://en.wikipedia.org/wiki/File:Galton_box.webm

Direct Sampling

We frequently need to estimate the probability of partially specified events $P(x_1, \ldots, x_m)$ with $m<n$. *n here denotes the number of random variables.*

This can be approximated by

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

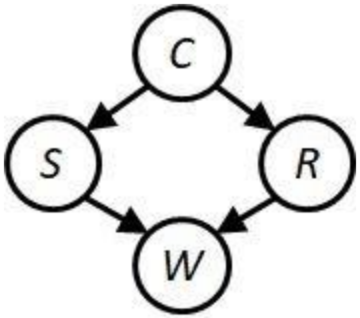
where $N_{PS}(x_1, \ldots xm)$ is now the number of samples among *N* whose values are consistent with $x_1, \ldots xm$

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} = ?

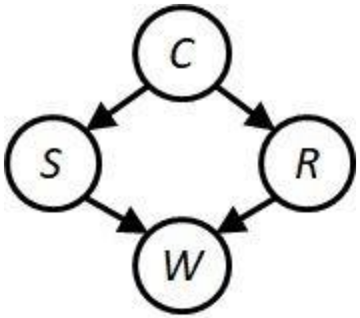


Direct Sampling

- 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
P{r} = 5/8
P(R|w} = ?
P(¬s|w} = ?



Rejection Sampling

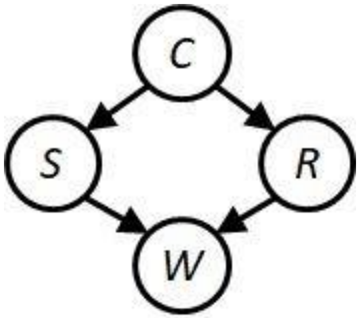
- Generate samples as follows.

{¬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}

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

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.

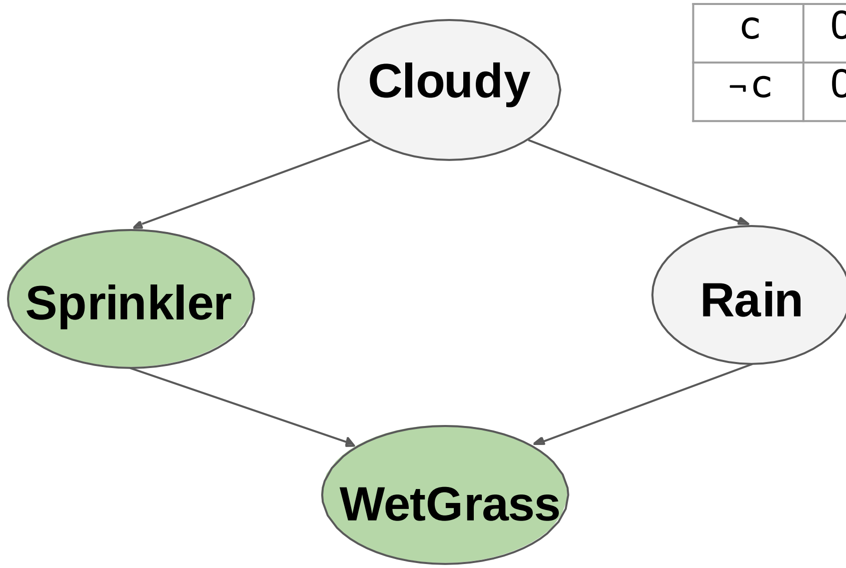
- **Likelihood weighting**
  - Samples examples which are consistent with evidence.
  - Each sample have a support value w (i.e., weight).
    - Initialize w of the generated sample as $w = 1$
    - Repeat
      - If variable is non-evidence : sample as usual.
      - If variable is evidence variable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

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

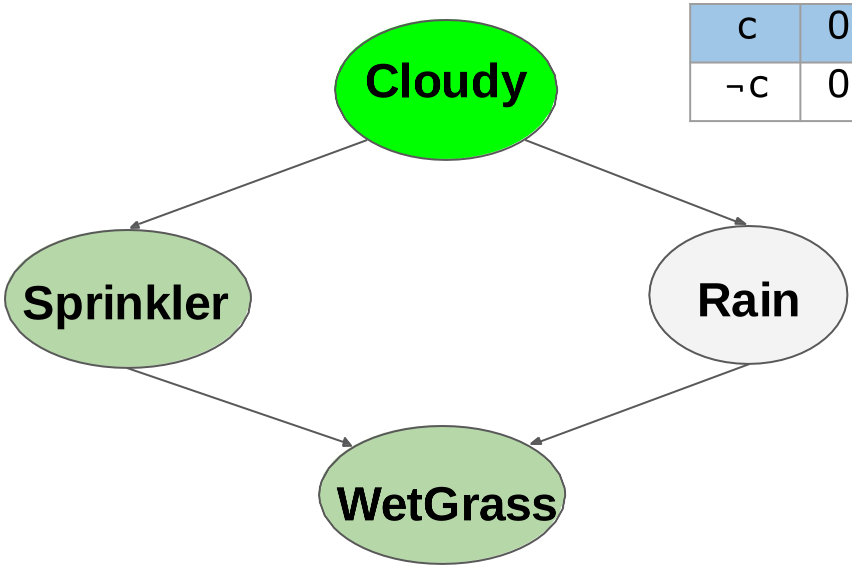
- **Fix sampling order {*C, S, R, W*}**
- **Likelihood weighted sampling with (*s, w*)**
  - **{ _, _, _, _ }** , w = 1

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

$P(W|S,R)$

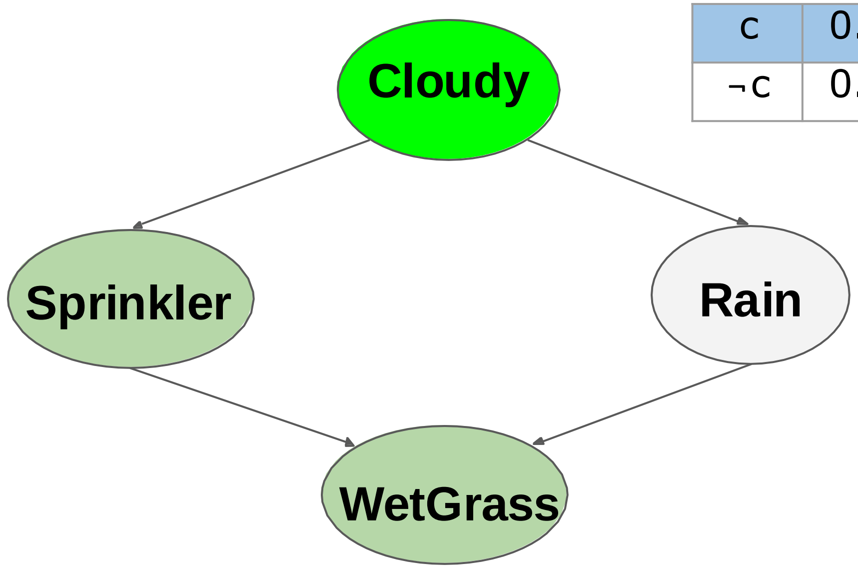
s	r	w	0.99
		¬w	0.01
	¬r	w	0.90
		¬w	0.10
¬s	r	w	0.90
		¬w	0.10
	¬r	w	0.01
		¬w	0.99

- **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

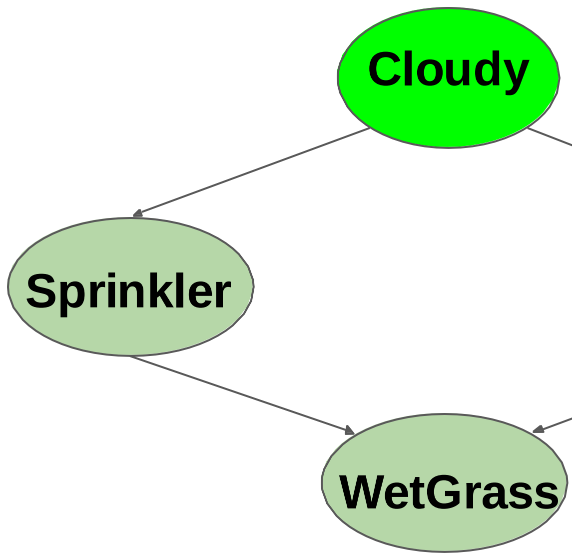
- **Fix sampling order {*C*, *S*, *R*,*W*}**
- **Likelihood weighted sampling with (*s*, *w*)**
  - **{*c*, *s*, _, _ }** , *w* = 1 * 0.1

$P(s|c)$

u is not needed!

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

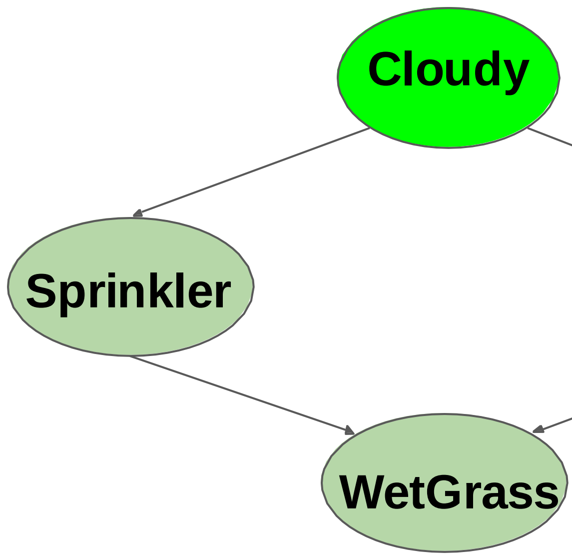
$u = 0.95$

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 is not needed!

$P(s|c)$ $P(w|s, \neg r)$

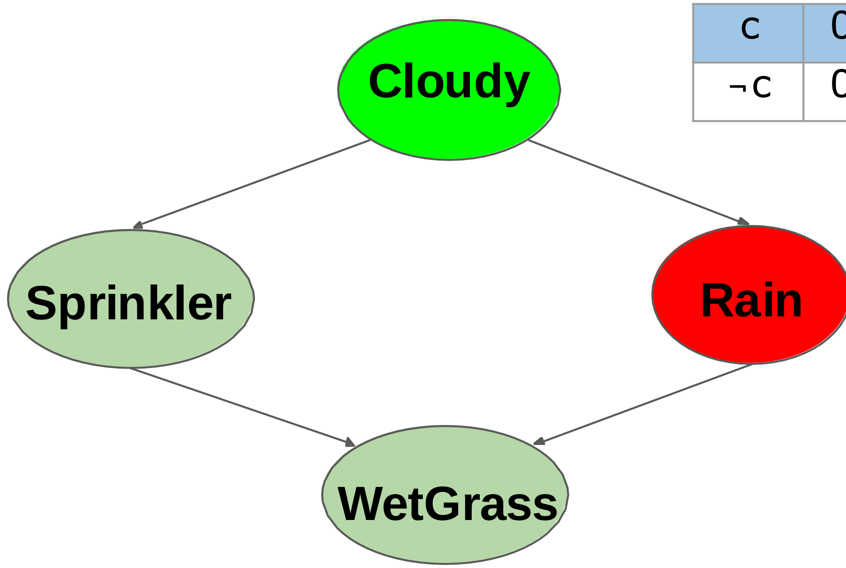
Evidence: (s,w)

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

$P(W|S,R)$

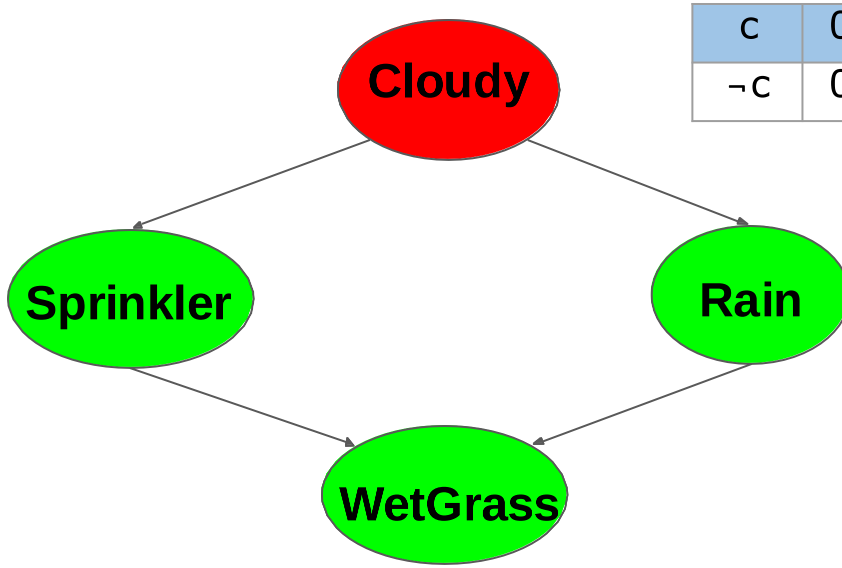
s	r	w	0.99
		¬w	0.01
	¬r	w	0.90
		¬w	0.10
¬s	r	w	0.90
		¬w	0.10
	¬r	w	0.01
		¬w	0.99

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

So the weight is very small,
meaning we have less likelihood in drawing this 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= 0.1*0.9 = 0.09
  - ***{¬c, s, ¬r, w},*** w= 0.5*0.9 = 0.45
  - **….**
  - ***{c, s, r, w},*** w= 0.1*0.99 = 0.099

Example : Likelihood Weighting

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

- 3 samples of $[\overset{C}{true}, \overset{S}{true}, \overset{R}{true}, \overset{W}{true}]$ 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)= $\alpha\sum_C P(C, R, s, w)$

$$\begin{aligned} & \hat{P}(Rain|Sprinkler = true, WetGrass = true) \\ = \quad & \alpha\langle \underset{(c,s,r,w)}{3 \times 0.099} + \underset{(\neg c,s,r,w)}{55 \times 0.495}, \underset{(c,s,\neg r,w)}{2 \times 0.09} + \underset{(\neg c,s,\neg r,w)}{40 \times 0.45}\rangle \\ = \quad & \alpha\langle 27.522, 18.18\rangle = \langle 0.60, 0.40\rangle \end{aligned}$$

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 (S_{WS}) is:

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

z = {c, ¬r}
e = {s, w}

where *l* is the number of nonevidence variables. For example:

$$S_{WS}(c,\ s,\ \neg r,\ w)=P(c)P(\neg r|c)$$

- Similarly the sample weights are:

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

Where *m* is the number of evidence variables. For example:

$$w(c, s, \neg r,\ w)=P(s|c)P(w|s,\ \neg r)$$

Likelihood Weighting : Why it works?

- Together, weighted sampling distribution is consistent.

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

-

$$
\begin{aligned}
\hat{P}(x \mid \mathbf{e}) &= \alpha \sum_{\mathbf{y}} N_{WS}(x,\mathbf{y},\mathbf{e}) w(x,\mathbf{y},\mathbf{e}) \\
&\approx \alpha' \sum_{\mathbf{y}} S_{WS}(x,\mathbf{y},\mathbf{e}) w(x,\mathbf{y},\mathbf{e}) \\
&= \alpha' \sum_{\mathbf{y}} P(x,\mathbf{y},\mathbf{e}) \\
&= \alpha' P(x,\mathbf{e}) = P(x \mid \mathbf{e}) \ .
\end{aligned}
$$

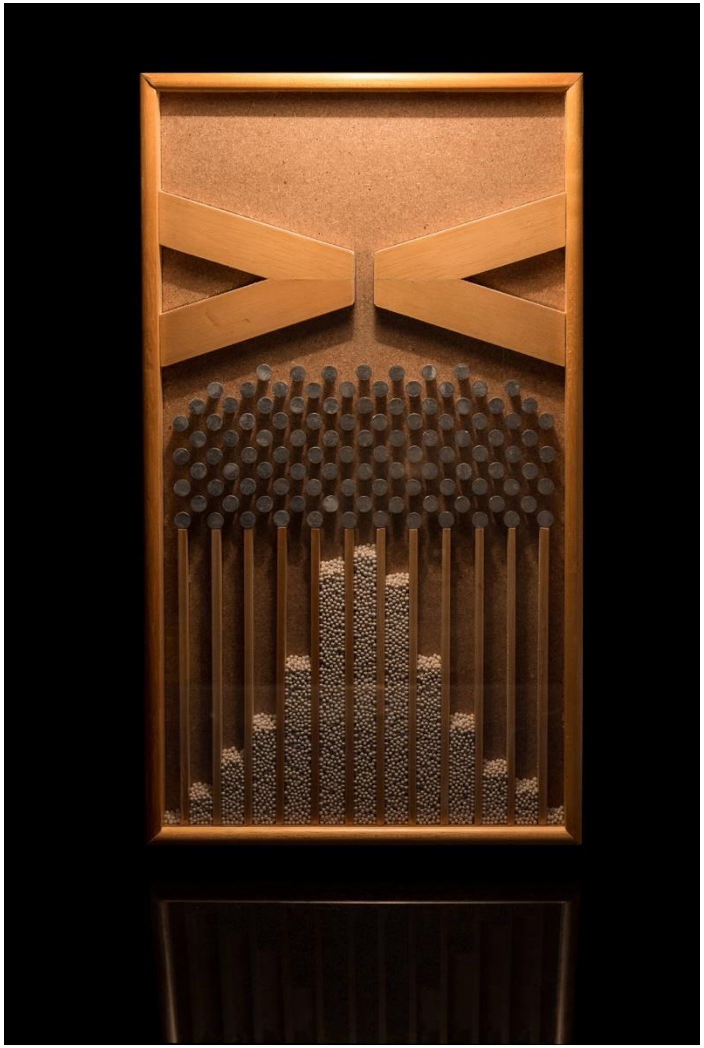
x = {¬r}
y = {c}
e = {s, w}

Likelihood Weighting: the Problem

- **Likelihood weighting**
  - More efficient than rejection sampling.
  - Performance (estimation accuracy/efficiency) decreases if the number of evidence variables increases -- samples could have very low weights.

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

  - The problem is exacerbated if the evidence is late in the variable ordering when we do the sampling.

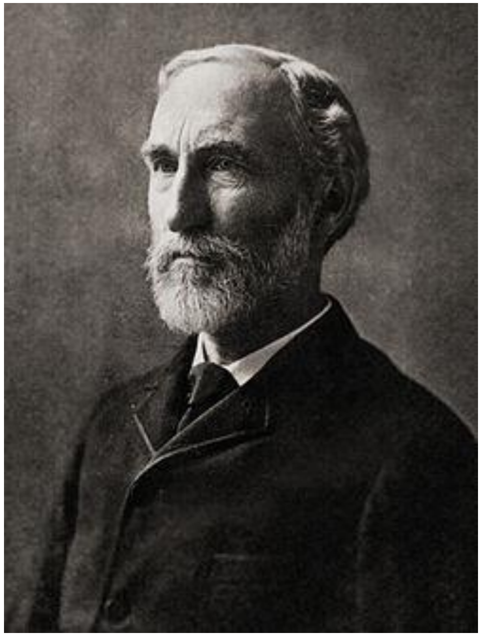


Attribution: Matemateca (IME USP)
https://upload.wikimedia.org/wikipedia/commons/c/c1/Galton_box.jpg

MCMC Methods

- **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.

Andrey Markov



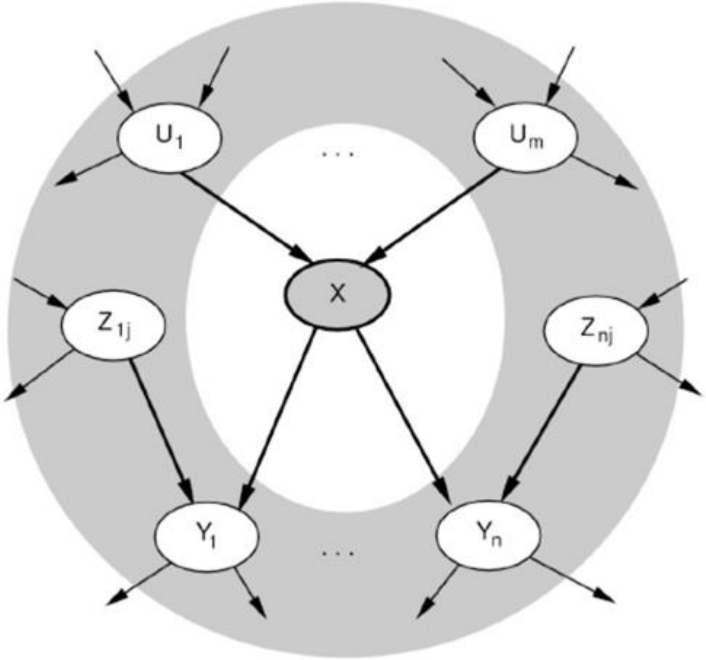
Josiah Willard Gibbs

Gibbs Sampling

- Generates an event by making a random change to preceding event
  - Think that network is in a **current state** which specifies 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

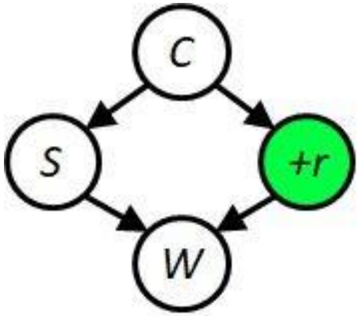
Recall that the Markov Blanket of a variable comprises of the parents, children, and children's parents of the variable.



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

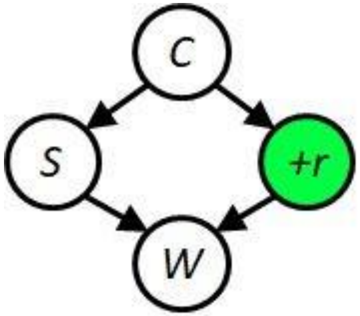
Gibbs Sampling Example (P(S|r))

Step 1: initialize evidence

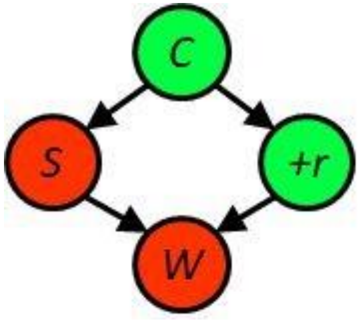


Gibbs Sampling Example (P(S|r))

Step 1: initialize evidence



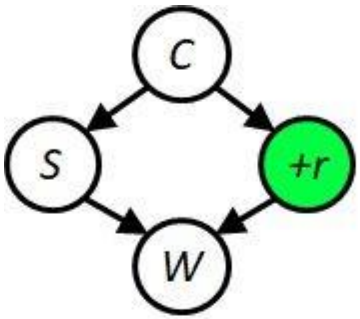
Step 2: initialize other variables (random)



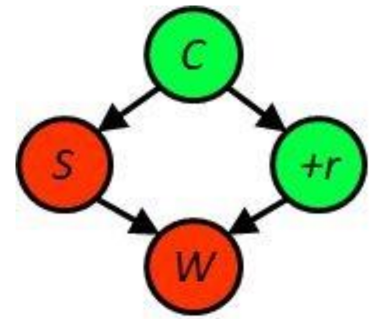
Initial state, e.g,: {c, ¬s, r, ¬w}

Gibbs Sampling Example (P(S|r))

Step 1: initialize evidence



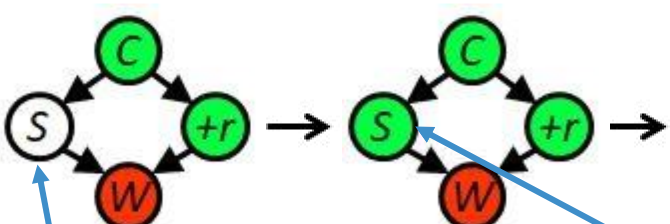
Step 2: initialize other variables (random)



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.

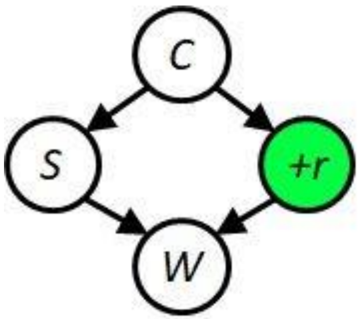


Sample from S with P(S|c, ¬w, r)

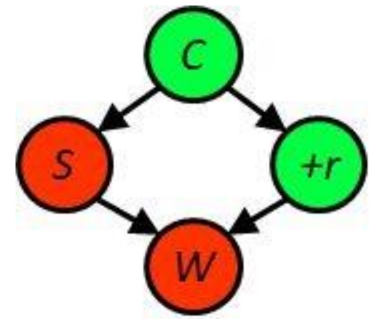
Suppose the result is true (S=s), the we get a new sample {c, s, r, ¬w}

Gibbs Sampling Example (P(S|r))

Step 1: initialize evidence



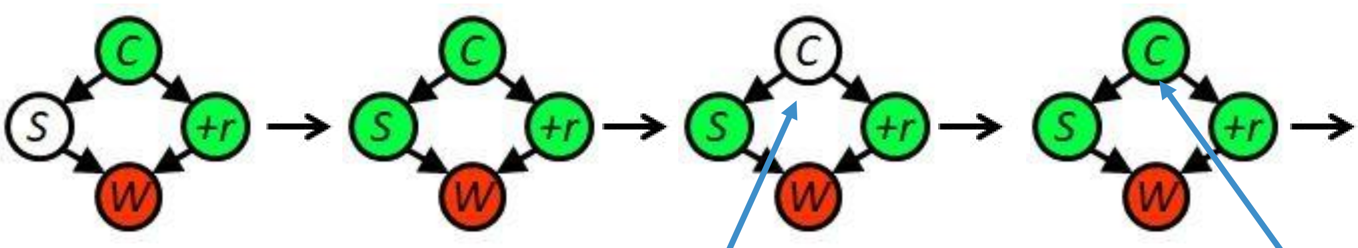
Step 2: initialize other variables (random)



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.



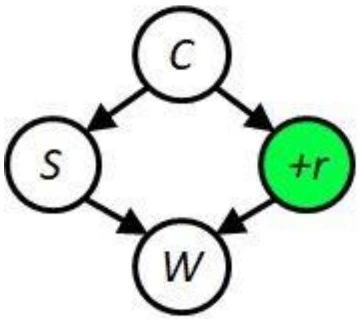
{c, s, r, ¬w}

C = c, hence {c, s, r, ¬w}

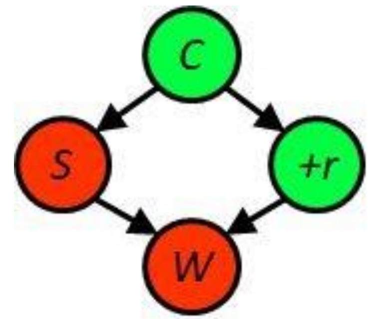
Sample from C with P(C|s, r)

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

Step 1: initialize evidence



Step 2: initialize other variables (random)

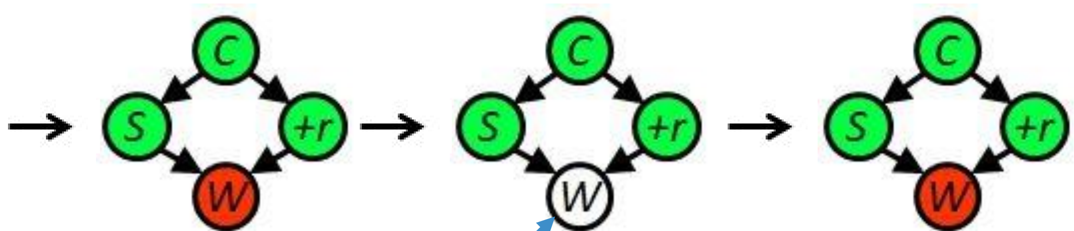


Initial state, e.g,: {c, ¬s, r, ¬w}

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}

{c, s, r, ¬w}

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

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}

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

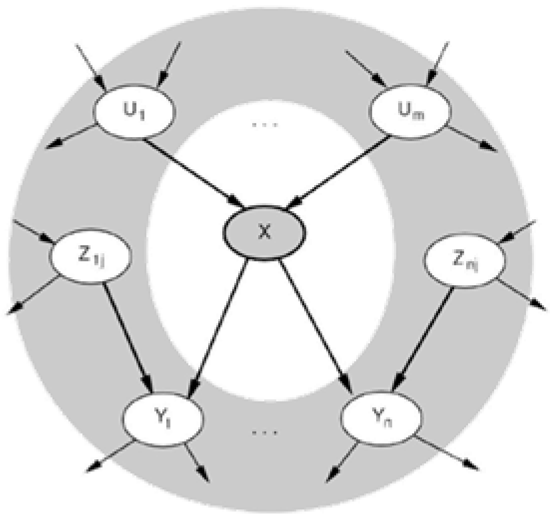
$$= < 0.37, 0.63 >$$

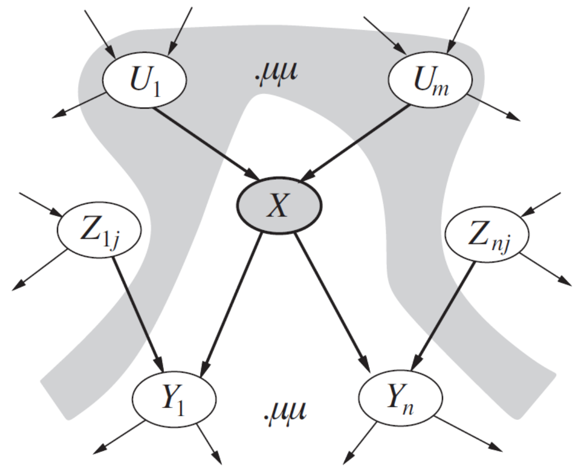
Probability given Markov Blanket

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

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

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





Gibbs Sampling Example (P(S|r))

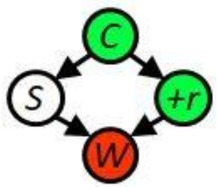
Sample from *P(S|c, ¬w, r)*
Sample from *P(C|s, r)*

Given:

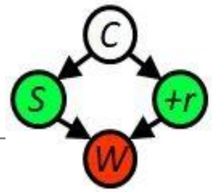
$$
\begin{aligned}
&\mathbf{P}(X_i|MB(X_i)) \\
&= \mathbf{P}(X_i|Parents(X_i), \mathbf{Y}, \mathbf{Z}_1, \ldots, \mathbf{Z}_\ell) \\
&= \alpha\mathbf{P}(X_i|Parents(X_i), \mathbf{Z}_1, \ldots, \mathbf{Z}_\ell)\mathbf{P}(\mathbf{Y}|Parents(X_i), X_i, \mathbf{Z}_1, \ldots, \mathbf{Z}_\ell) \\
&= \alpha\mathbf{P}(X_i|Parents(X_i))\mathbf{P}(\mathbf{Y}|X_i,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

Then,

P(S|c, ¬w, r) = αP(S|c)P(¬w|S,r)



P(C|s, r) = α'P(C)P(s|C)P(r|C)



Gibbs Sampling

- As sampling settles into a dynamic equilibrium, the fraction of time spend on each state is proportional to its posterior probability.

$$\{c,\ s,\ r,\ w\}$$
$$\{c,\ \neg s,\ r,\ w\}$$
$$\{\neg c,\ s,\ r,\ w\}$$
$$\{\neg c,\ \neg s,\ r,\ w\}$$