Artificial Intelligence

CS4365 --- Fall 2022

Bayesian Networks: Approximate Inference

Instructor: Yunhui Guo

Inference: The Bad News

- Computing the conditional probabilities by enumerating all relevant entries in the joint is expensive:

Exponential in the number of variables!

Possible Solutions

- **Exact methods**
  - Inferecen by enumeration and variable elimination
- **Approximate methods**
  - Approximate the joint distributions by drawing samples

Sampling

- Sampling is a lot like repeated simulation
  - tossing a coin, tosing a dice, …

- **Basic idea**
  - Draw N samples from a sampling distribution S
  - Compute an approximate posterior probability
  - Show this converges to the true probability P

- Why:
  - Learning: get samples from a distribution you don't know
  - Inference: getting a sample is faster than computing the right answer (e.g. with variable elimination)

Approximate Methods: Sampling

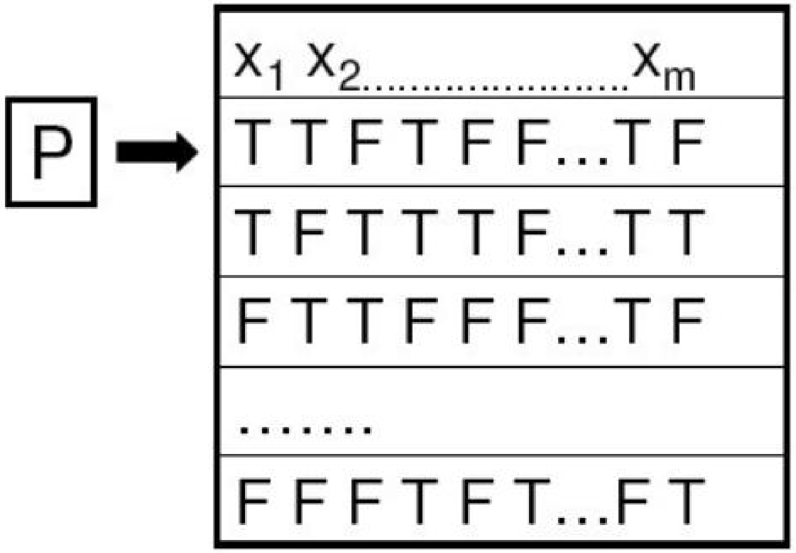
- **Sampling** = Powerful technique in many probabilistic problems
- General idea:
  - It is often difficult to compute and represent exactly the probability distribution of a set of variables
  - But, it is often easy to generate examples from the distribution

The number of rows too large for the table to be computed explicitly

x_1 $x_{2\ldots}x_m$	$P(X_1=x_1,X_2=x_2,\ldots,X_m=x_m)$
T T…T	0.95
T F…T	0.94
F T…T	0.29
F F…T	0.001
…….	……………….

Approximate Methods: Sampling

- **Sampling** = Powerful technique in many probabilistic problems
- General idea:
  - It is often difficult to compute and represent exactly the probability distribution of a set of variables
  - But, it is often easy to generate examples from the distribution



For a large number of samples,
$P(X_1=x_1, X_2=x_2, \ldots, X_m = x_m)$
is approximately equal to:

$$\frac{\text{\# of samples with } X_1=x_1 \text{ and } X_2=x_2 \ldots \text{and } X_m = x_m}{\text{Total \# of samples}}$$

Sampling from given distribution

- **Step 1**: Get sample u from uniform distribution over [0, 1)
  - E.g. random() in python
- **Step 2**: Convert this sample u into an outcome for the given distribution by having each outcome associated with a sub-interval of [0,1) with sub-interval size equal to probability of the outcome

- Example
  - If random() returns u = 0.83, then our sample is C = blue

$$0 \leq u < 0.6, \rightarrow C = red$$
$$0.6 \leq u < 0.7, \rightarrow C = green$$
$$0.7 \leq u < 1, \rightarrow C = blue$$

C	P(C)
red	0.6
green	0.1
blue	0.3

Sampling

- Prior Sampling
- Rejection Sampling
- Likelihood Weighting
- Gibbs Sampling

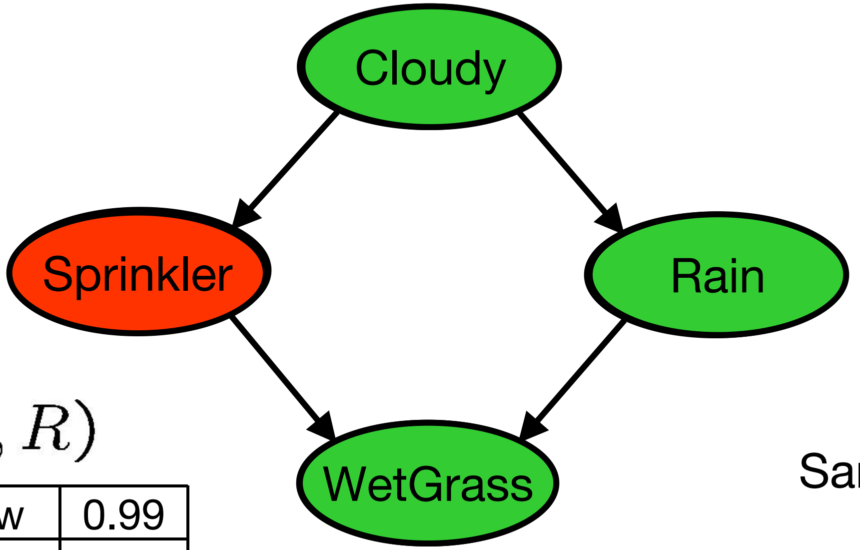
Prior Sampling

$P(C)$

+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

Samples:

+c, -s, +r, +w

-c, +s, -r, +w

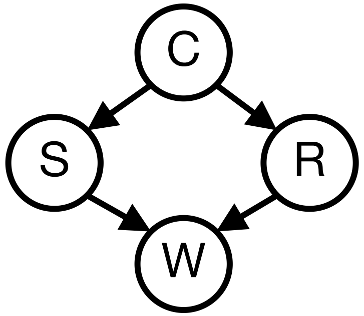
…

Prior Sampling

- For i=1, 2, …, n
  - Sample x_i from $P(X_i \mid Parents(X_i))$
- Return $(x_1, x_2, \ldots, x_n)$
- We'll get a bunch of samples from the BN:

  +c, -s, +r, +w
  +c, +s, +r, +w
  -c, +s, +r, -w
  +c, -s, +r, +w
  -c, -s, -r, +w

- Compute probability:
  - We have counts <+w:4, -w:1>
  - Normalize to get P(W) = <+w:0.8, -w:0.2>
  - What about P(C| +w)? P(C| +r, +w)? P(C| -r, -w)?

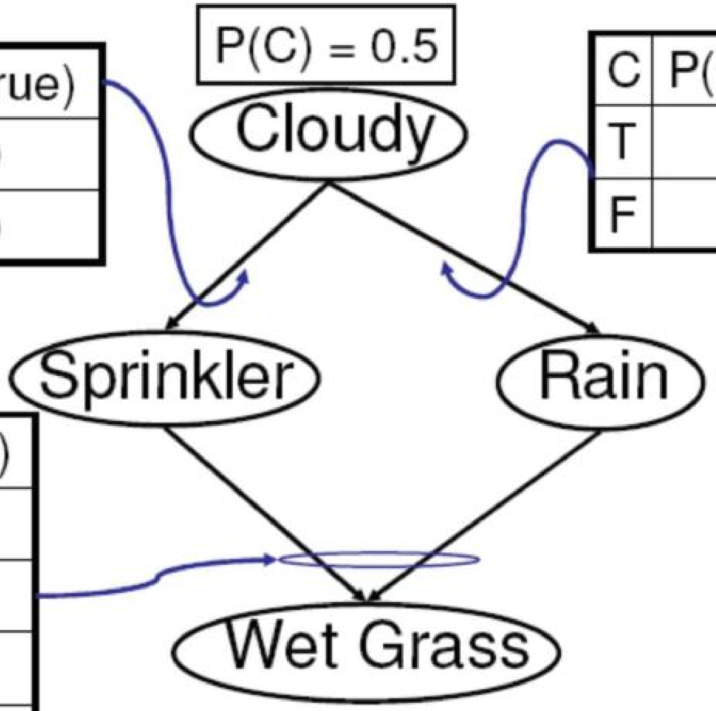


Rejection sampling

Sampling: An Example

- The lawn may be wet because the sprinkler was on or because it was raining (or both).

Compute P(C | +w)



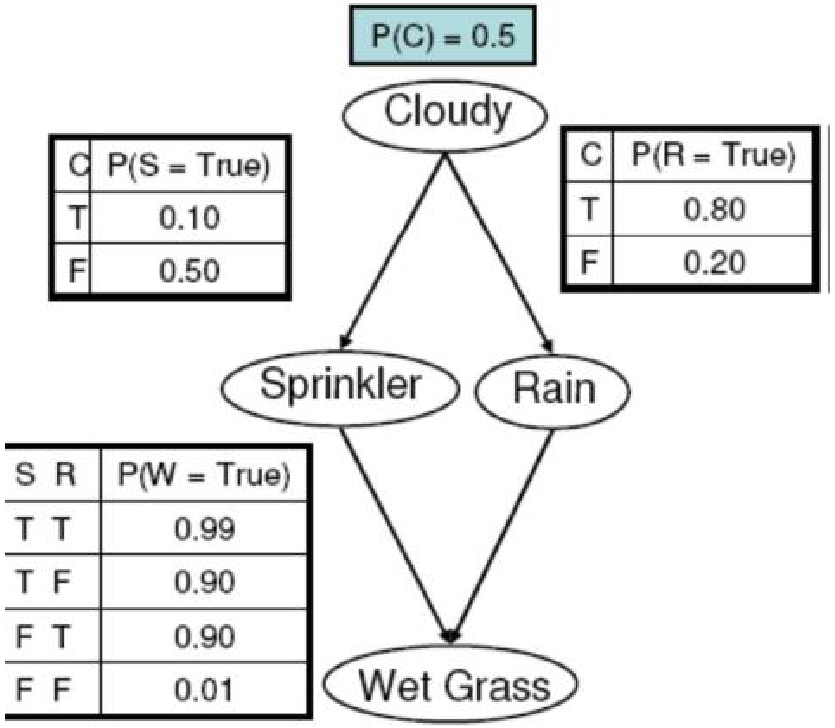
C	P(S = True)
T	0.10
F	0.50

C	P(R = True)
T	0.80
F	0.20

S	R	P(W = True)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

Sampling

C	S	R	W
T			



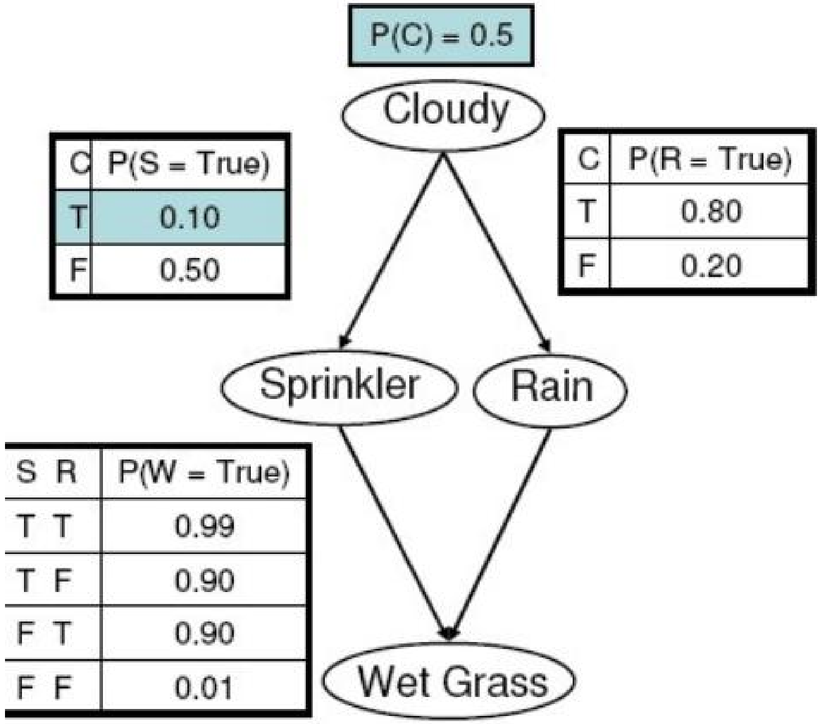
1. Randomly choose C.

    C = True with probability 0.5

    → C = True

Sampling

C	S	R	W
T	F		



1. Randomly choose C.

C = True with probability 0.5

→ C = True

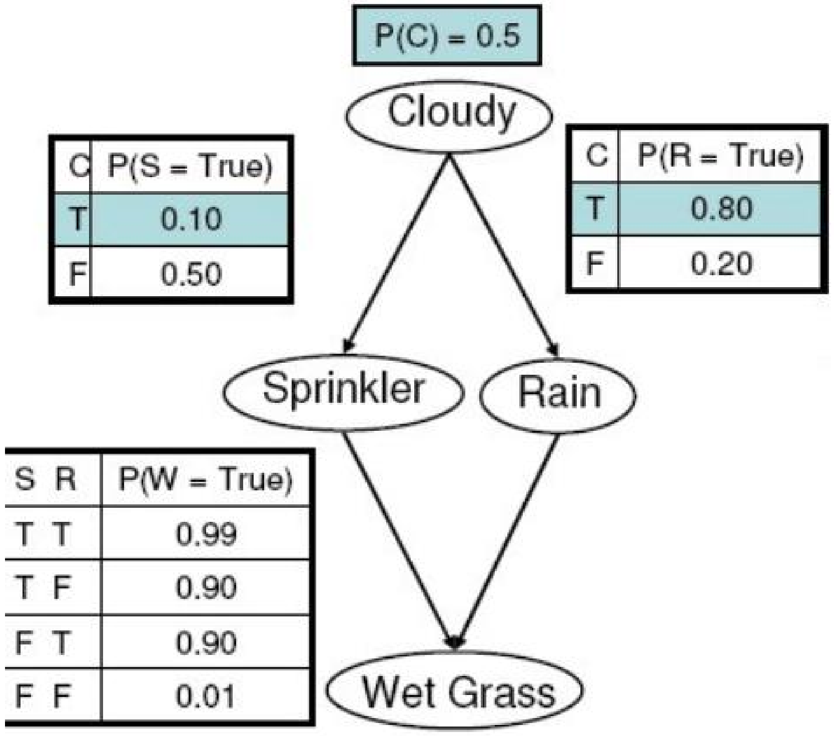
2. Randomly choose S.

S = True with probability 0.10

→ S = False

Sampling

C	S	R	W
T	F	T	



1. Randomly choose C.

   C = True with probability 0.5

   → C = True

2. Randomly choose S.

   S = True with probability 0.10

   → S = False

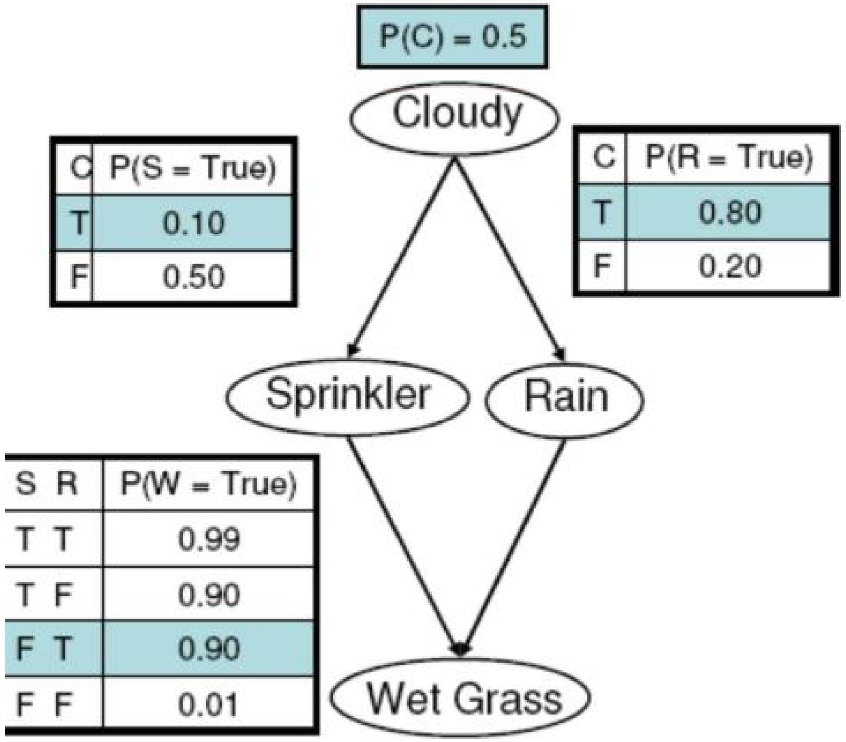
3. Randomly choose R.

   R = True with probability 0.80

   → R = True

Sampling

C	S	R	W
T	F	T	T



1. Randomly choose C.

    C = True with probability 0.5

    → C = True

2. Randomly choose S.

    S = True with probability 0.10

    → C = False

3. Randomly choose R.

    R = True with probability 0.80

    → R = True

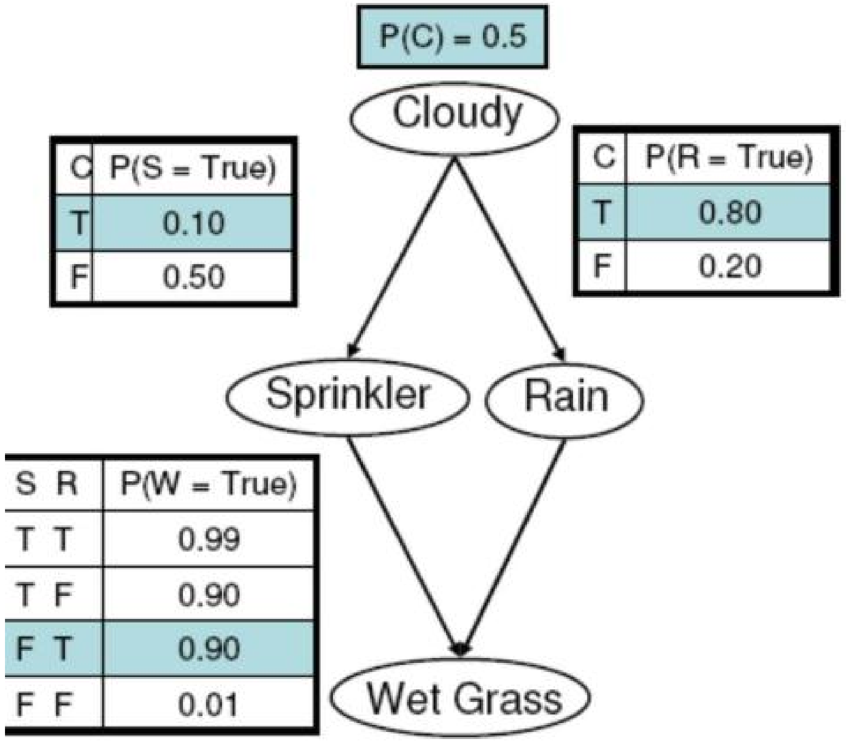
4. Random choose W.

    W = True with probability 0.90

    → W = True

Sampling

C	S	R	W
T	F	T	T



+c, -s, +r, +w
+c, +s, +r, +w
-c, +s, +r, -w
+c, -s, +r, +w
-c, -s, -r, +w

- Compute P(C | +w):
  - Gather all the samples with +w
  - Count +c and -c

Rejection Sampling: Example

- Suppose that we want to compute P(W = True | C = True) (In words: How likely is it that the grass will be wet given that the sky is cloudy)
- Compute lots of samples of (C,S,R,W)

  - N_c = Number of samples for which C = True
  - N_s = Number of samples for which W = True and C = True
  - N = Total number of samples

- N_c/N approximates P(C = True)
- N_s/N approximates P(W = True and C = True)

Therefore: N_s/N_c approximates:

P(W = True and C = True)/ P(C = True) = P(W = True | C = True)

Rejection Sampling: General Case

- Suppose that we want to compute $P(E_1 | E_2)$ (In words: How likely is it that the grass will be wet given that the sky is cloudy)
- Compute lots of samples of (C,S,R,W)
  - N_c = Number of samples for which C = True
  - N_s = Number of samples for which W = True and C = True
  - N = Total number of samples

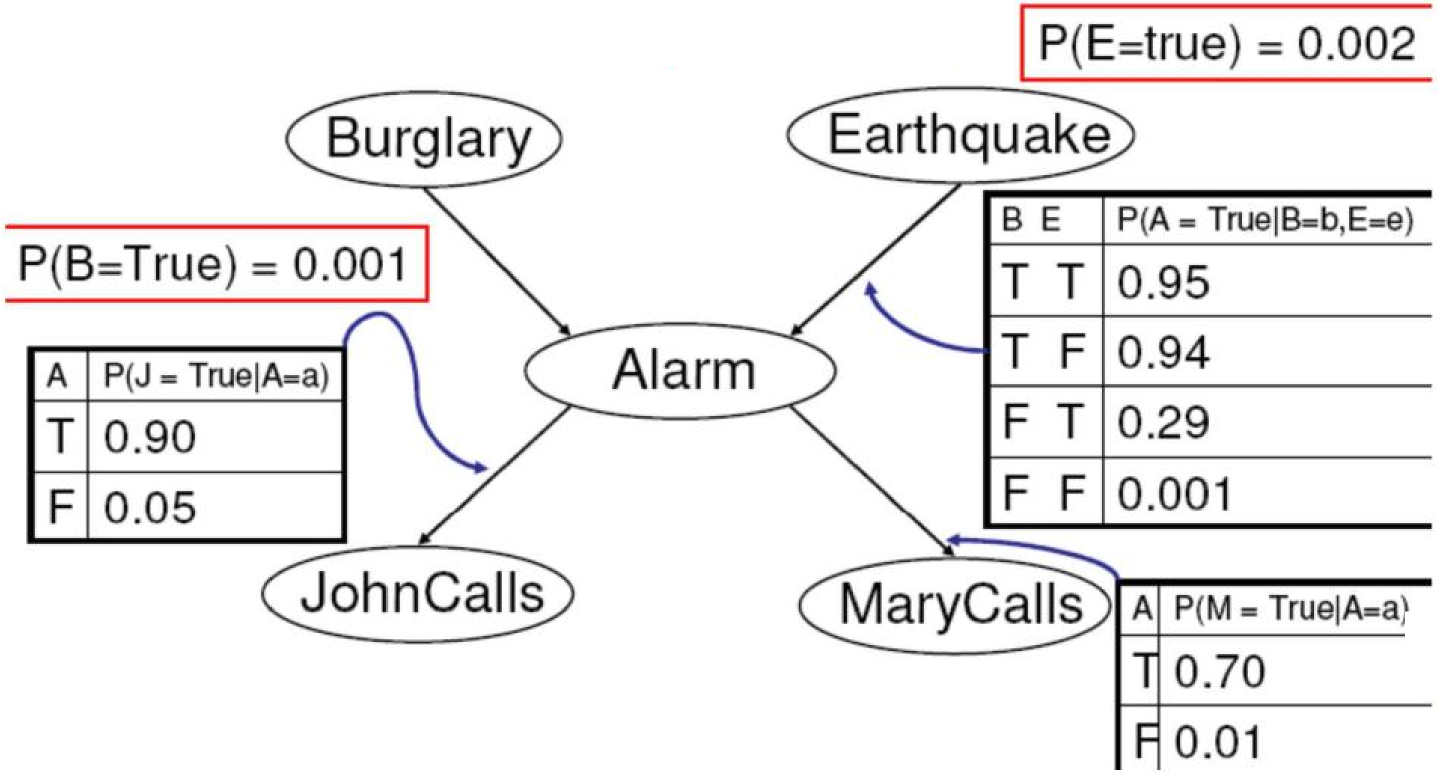
- N_c/N approximates $P(E_2)$
- N_s/N approximates $P(E_1 \text{ and } E_2)$

Therefore: N_s/N_c approximates:

$$P(E_1 \text{ and } E_2) / P(E_2) = P(E_1 | E_2)$$

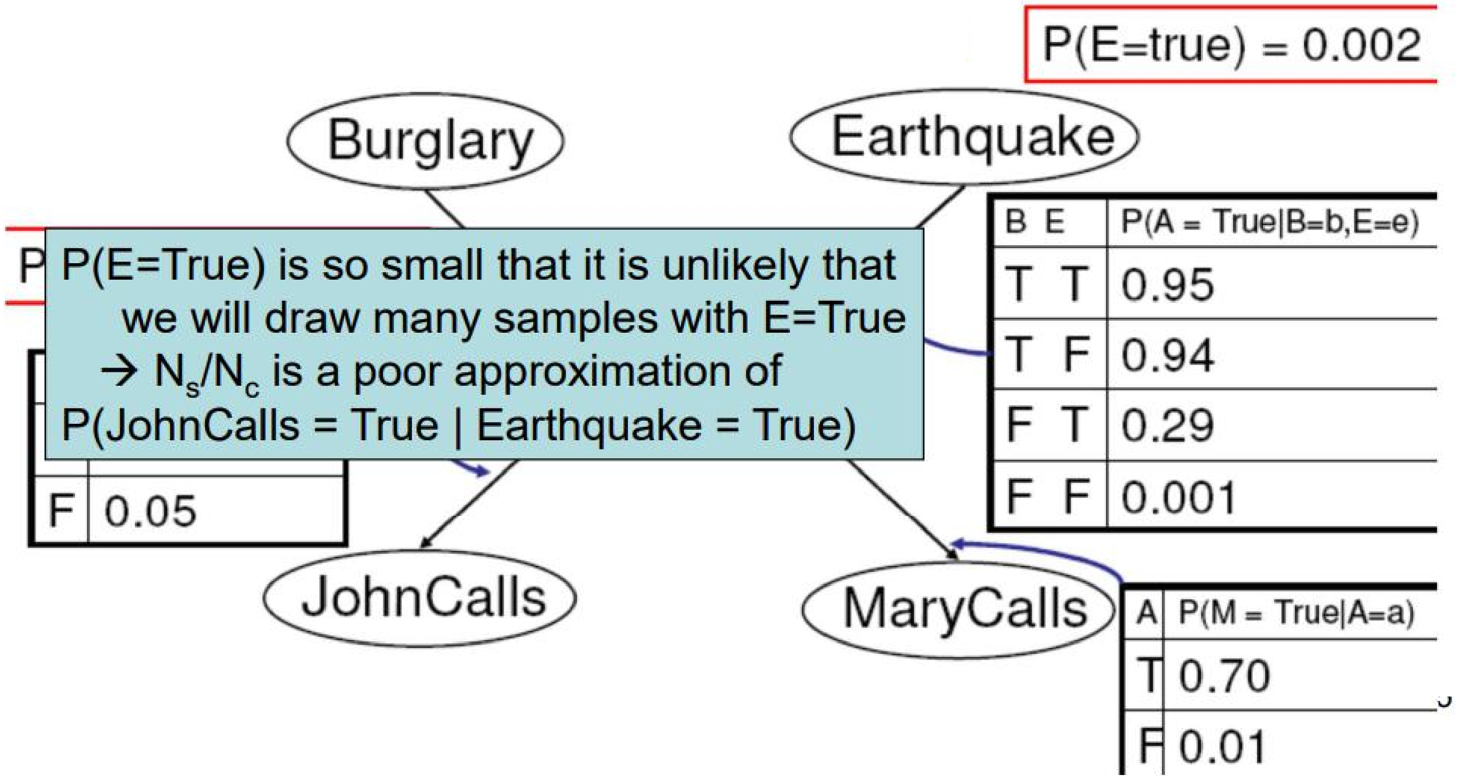
Problems with Rejection Sampling

- Probability is so low for some assignments of variables that will likely never be seen in the samples (unless a very large number of samples is drawn).



Problems with Sampling

- Probability is so low for some assignments of variables that will likely never be seen in the samples (unless a very large number of samples is drawn).
- P(JohnCalls = True | Earthquake = True)



Solution: Likelihood Weighting

- Suppose that E_2 contains a variable assignment of the form $X_i = v$

- Current approach:

  Generate samples until enough of them contain $X_i = v$

  Such samples are generated with probability

  $p = P(X_i = v \mid Parents(X_i))$

- Likelihood Weighting:

  Generate only samples with $X_i = v$

  Weight each sample by $\omega = p$

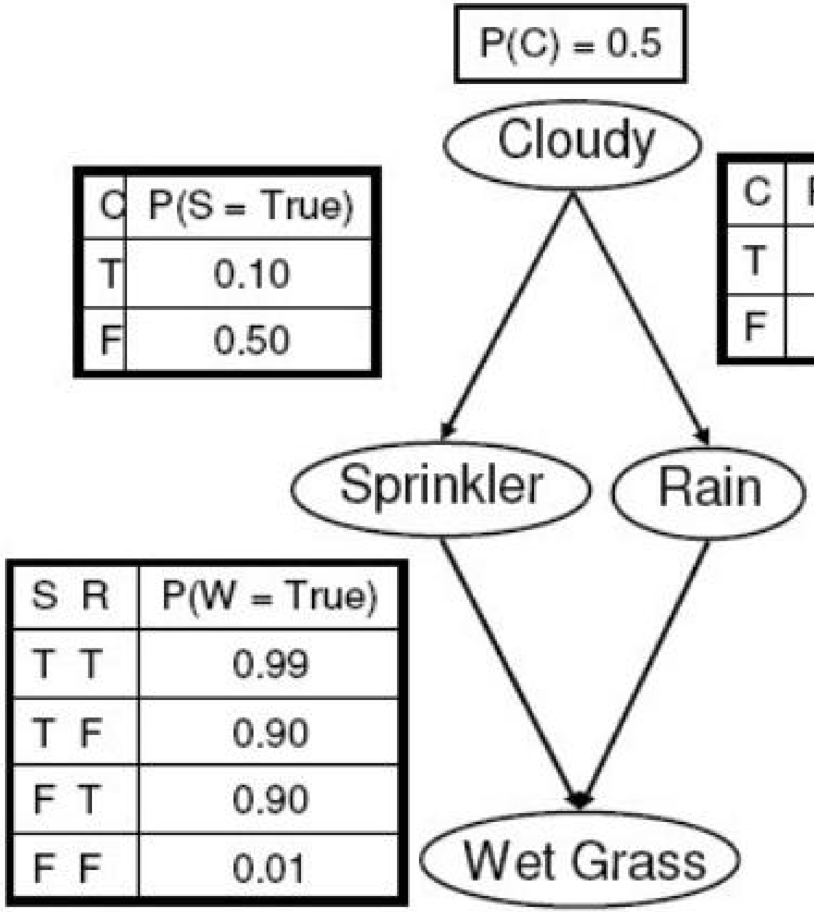
Solution: Likelihood Weighting

- Idea: fix evidence variables, sample only nonevidence variables,

- Weight each sample by the likelihood it accords the evidence

- The weights of samples derived from likelihood of evidence accumulated during sampling process

Solution: Likelihood Weighting

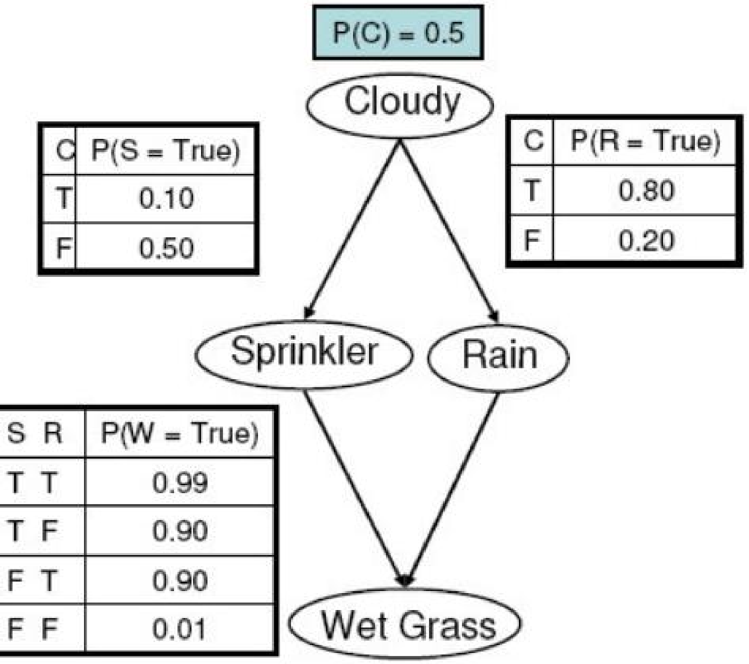


- Example: Suppose that we want to compute an inference with

E_2: (S= True, W = True)

Solution: Likelihood Weighting

$\omega = 1.0$



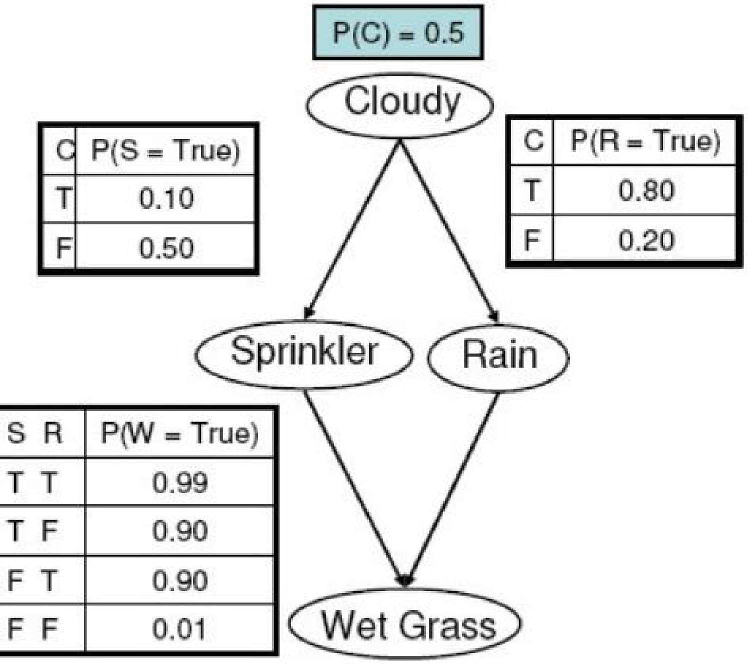
1. Randomly choose C.

   C = True with probability 0.5

   $\rightarrow$ C = True

Solution: Likelihood Weighting

$\omega = 1.0$



1. Randomly choose C.

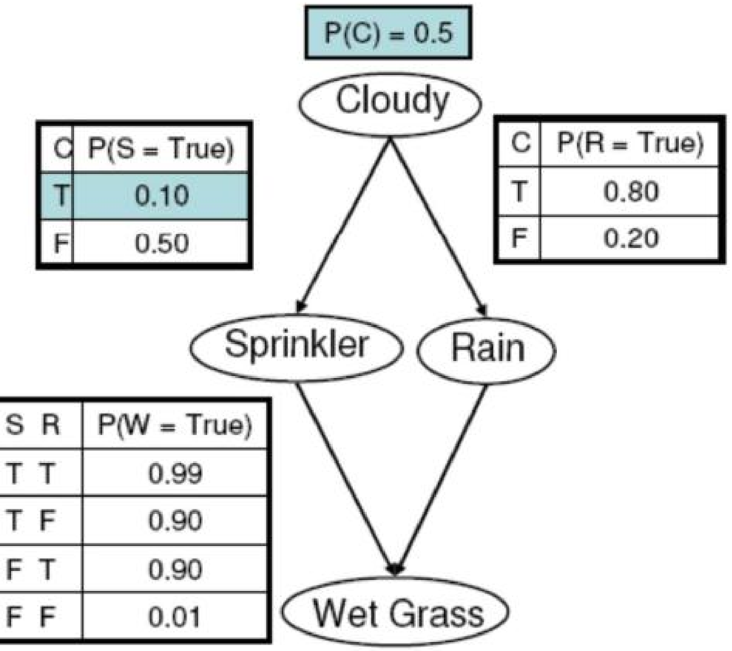
   C = True with probability 0.5

   → C = True

C is not one of the evidence variables, so we take a random sample as before

Solution: Likelihood Weighting

$\boxed{\omega = 1.0 \times 0.10}$



1. Randomly choose C.

   C = True with probability 0.5

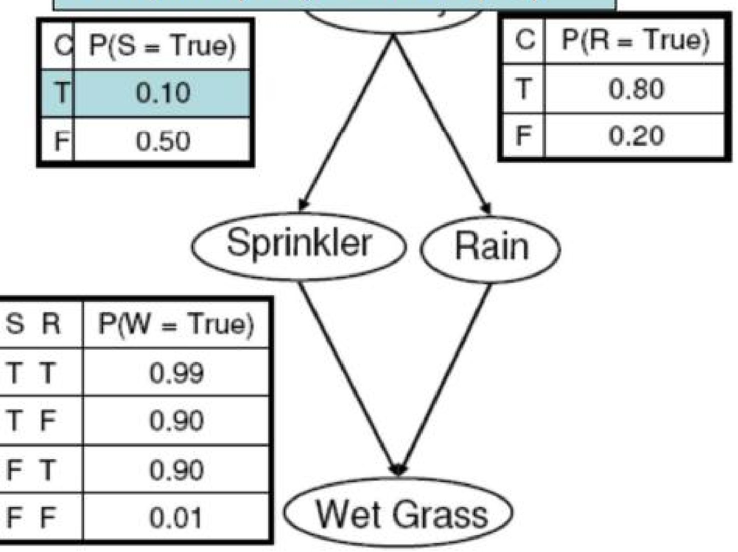
   → C = True

2. Set S = True

Solution: Likelihood Weighting

$\omega = 1.0 \times 0.10$

At the same time, we update the current weight of the sample by P(S = True | C)



C	P(S = True)
T	0.10
F	0.50

C	P(R = True)
T	0.80
F	0.20

S	R	P(W = True)
T	T	0.99
T	F	0.90
F	T	0.90
F	F	0.01

1. Randomly choose C.

   C = True with probability 0.5

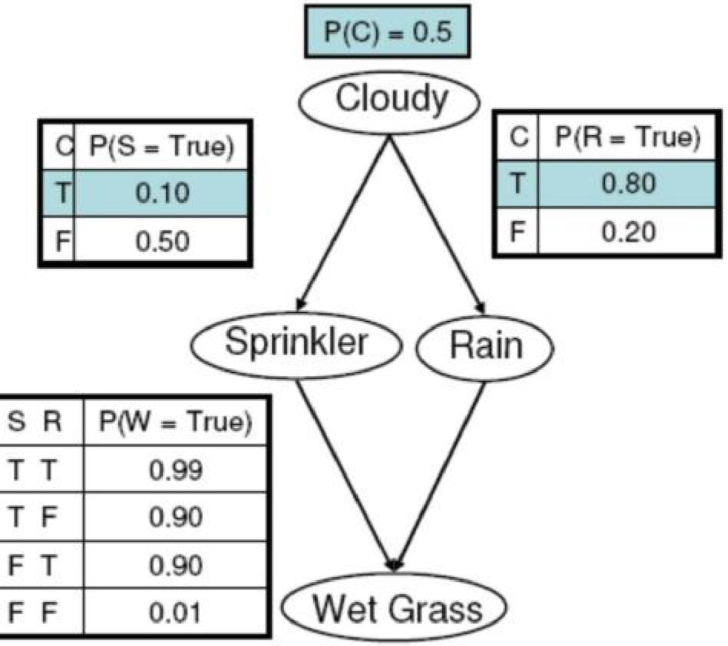
   → C = True

2. Set S = True

S is one of the evidence variables, so we fix its value without sampling

Solution: Likelihood Weighting

$\omega = 1.0 \times 0.10$



1. Randomly choose C.

   C = True with probability 0.5

   → C = True

2. Set S = True

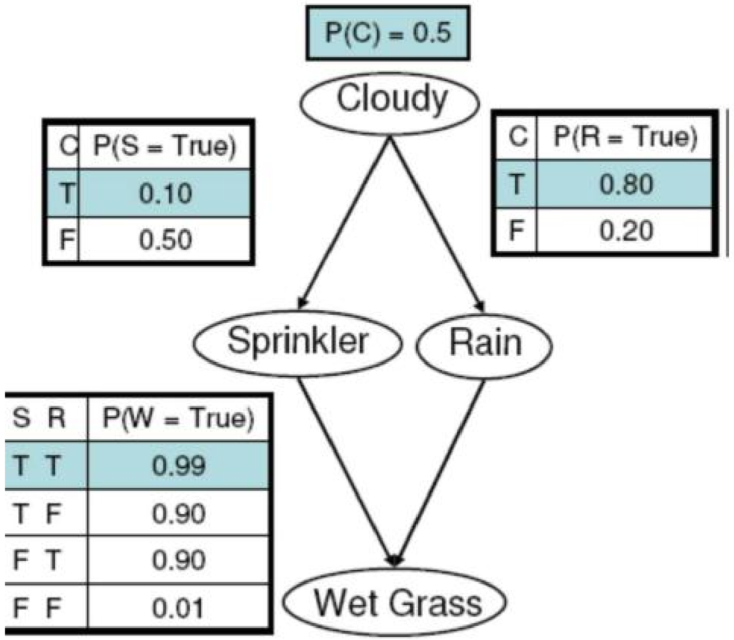
3. Randomly choose R.

   R = True with probability 0.80

   → R = True

Solution: Likelihood Weighting

$\omega = 1.0 \times 0.10 \times 0.99$



1. Randomly choose C.

   C = True with probability 0.5

   → C = True

2. Set S = True

3. Randomly choose R.

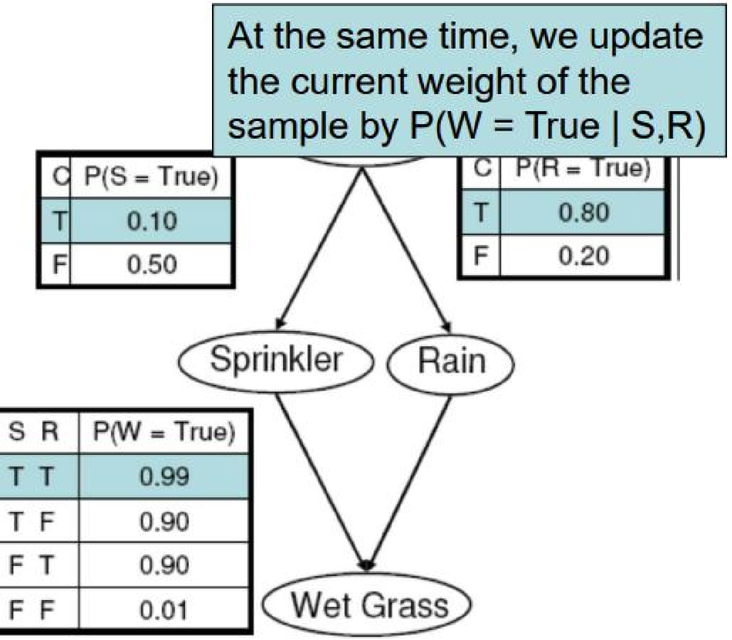
   R = True with probability 0.80

   → R = True

4. Set W = True

Solution: Likelihood Weighting

$\omega = 1.0 \times 0.10 \times 0.99$



1. Randomly choose C.

    C = True with probability 0.5

    → C = True

2. Set S = True

3. Randomly choose R.

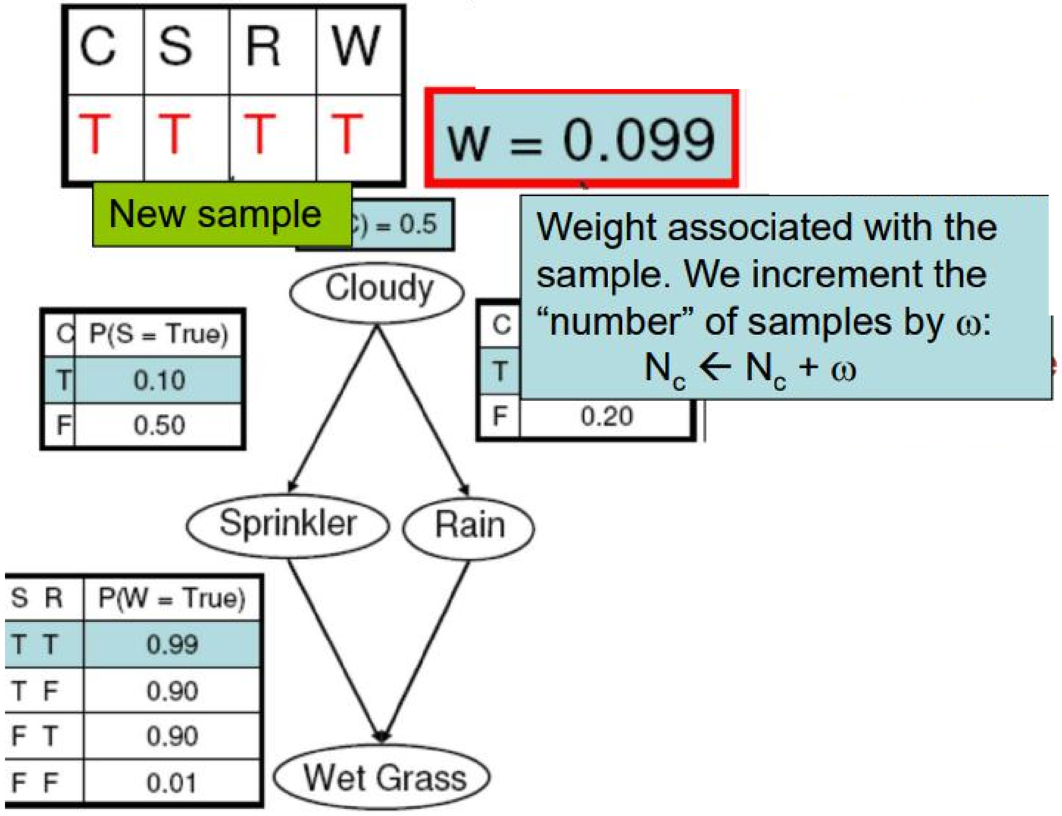
    R = True with probability 0.80

    → R = True

4. Set W = True

    W is one of the evidence variables, so we fix its value without sampling

Solution: Likelihood Weighting



1. Randomly choose C.

   C = True with probability 0.5

   → C = True

2. Set S = True

3. Randomly choose R.

   R = True with probability 0.80

   → R = True

4. Set W = True

   W is one of the evidence variables, so we fix its value without sampling

Likelihood Weighting

- $N_c = 0; N_s = 0;$

1. Generate a random assignment of the variables, fixing the variables assigned in E_2
2. Assign the sample a weight ω = probability that this sample would have been generated if we did not fix the value of the variables in E_2
3. $N_c \leftarrow N_c + \omega$
4. If the sample matches E_1 $\quad N_s \leftarrow N_s + \omega$
5. Repeat until we have "enough" samples

N_s/N_c is an estimate of $P(E_1|E_2)$

Likelihood Weighting

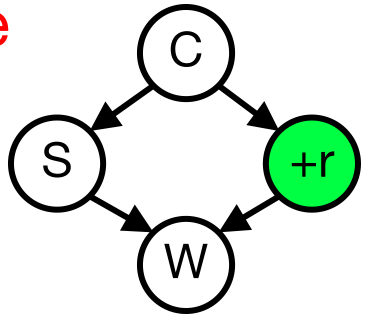
- Likelihood weighting is good
  - We have taken evidence into account as we generate the sample
  - E.g. here, W's value will get picked based on the evidence values of S, R
- Likelihood weighting doesn't solve all our problems
  - Evidence influences the choice of downstream variables, but not upstream ones (C isn't more likely to get a value matching the evidence)
- We would like to consider evidence when we sample every variable
  - Gibbs sampling

Gibbs Sampling

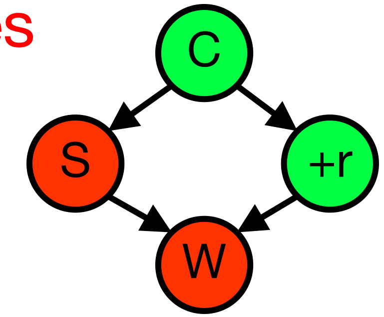
- Procedure: keep track of a full instantiation $x_1, x_2, \ldots, x_n$.
- Start arbitrary instantiation consistent with the evidence.
- Sample one variable at a time, conditioned on all the rest, but keep evidence fixed.
- Keep repeating this for a long time.

Gibbs Sampling Example: P(S|+r)

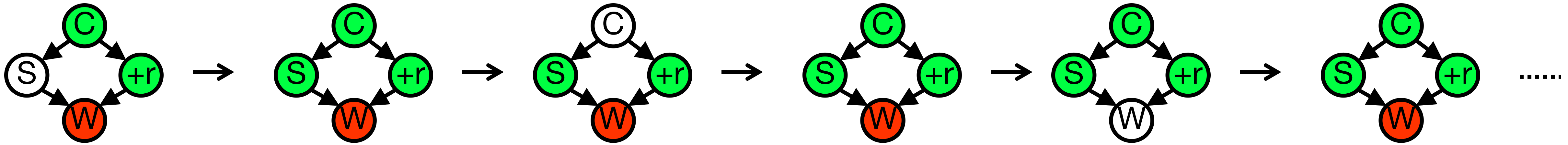
- Step 1: Fix evidence
  - R = +r



- Step 2: Initialize other variables
  - Randomly

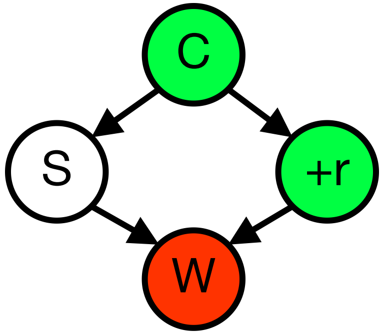


- Steps 3: Repeat
  - Choose a non-evidence variable X
  - Resample X from P(X | all other variables)



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

Efficient Resampling of One Variable

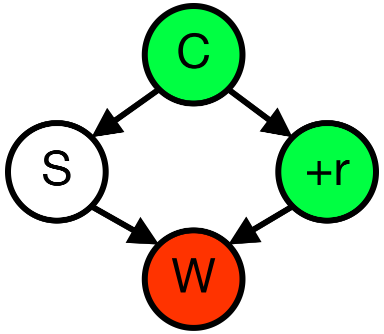


- Sample from P(S | +c, +r, -w)

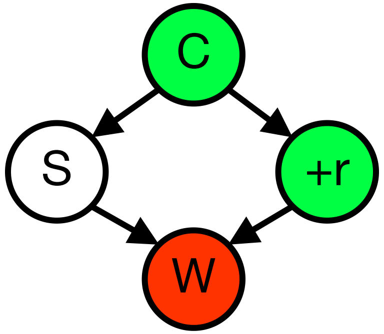
Efficient Resampling of One Variable

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

$$P(S| + c, +r, -w) = \frac{P(S, +c, +r, -w)}{P(+c, +r, -w)}$$



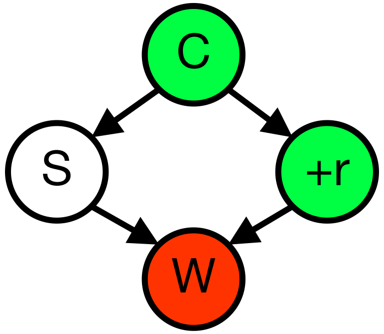
Efficient Resampling of One Variable



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

$$P(S|+c,+r,-w) = \frac{P(S,+c,+r,-w)}{P(+c,+r,-w)}$$
$$= \frac{P(S,+c,+r,-w)}{\sum_s P(s,+c,+r,-w)}$$

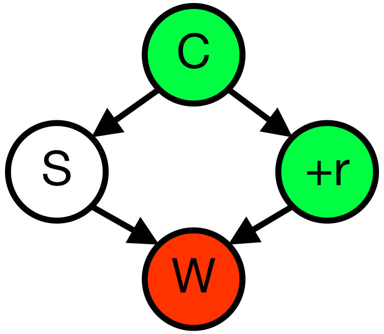
Efficient Resampling of One Variable



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

$$
\begin{aligned}
P(S|+c,+r,-w) &= \frac{P(S,+c,+r,-w)}{P(+c,+r,-w)} \\
&= \frac{P(S,+c,+r,-w)}{\sum_s P(s,+c,+r,-w)} \\
&= \frac{P(+c)P(S|+c)P(+r|+c)P(-w|S,+r)}{\sum_s P(+c)P(s|+c)P(+r|+c)P(-w|s,+r)}
\end{aligned}
$$

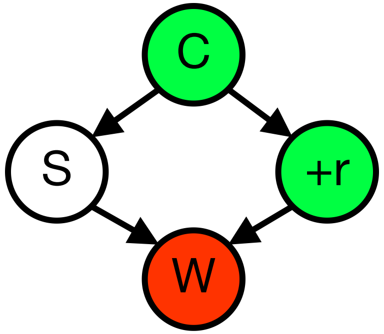
Efficient Resampling of One Variable



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

$$
\begin{aligned}
P(S|+c,+r,-w) &= \frac{P(S,+c,+r,-w)}{P(+c,+r,-w)} \\
&= \frac{P(S,+c,+r,-w)}{\sum_s P(s,+c,+r,-w)} \\
&= \frac{P(+c)P(S|+c)P(+r|+c)P(-w|S,+r)}{\sum_s P(+c)P(s|+c)P(+r|+c)P(-w|s,+r)} \\
&= \frac{P(+c)P(S|+c)P(+r|+c)P(-w|S,+r)}{P(+c)P(+r|+c)\sum_s P(s|+c)P(-w|s,+r)}
\end{aligned}
$$

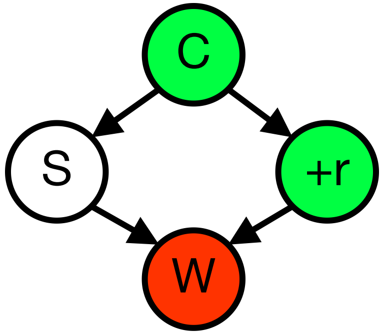
Efficient Resampling of One Variable



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

$$
\begin{aligned}
P(S|+c,+r,-w) &= \frac{P(S,+c,+r,-w)}{P(+c,+r,-w)} \\
&= \frac{P(S,+c,+r,-w)}{\sum_s P(s,+c,+r,-w)} \\
&= \frac{P(+c)P(S|+c)P(+r|+c)P(-w|S,+r)}{\sum_s P(+c)P(s|+c)P(+r|+c)P(-w|s,+r)} \\
&= \frac{P(+c)P(S|+c)P(+r|+c)P(-w|S,+r)}{P(+c)P(+r|+c)\sum_s P(s|+c)P(-w|s,+r)} \\
&= \frac{P(S|+c)P(-w|S,+r)}{\sum_s P(s|+c)P(-w|s,+r)}
\end{aligned}
$$

Efficient Resampling of One Variable



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

$$
\begin{aligned}
P(S|+c,+r,-w) &= \frac{P(S,+c,+r,-w)}{P(+c,+r,-w)} \\
&= \frac{P(S,+c,+r,-w)}{\sum_s P(s,+c,+r,-w)} \\
&= \frac{P(+c)P(S|+c)P(+r|+c)P(-w|S,+r)}{\sum_s P(+c)P(s|+c)P(+r|+c)P(-w|s,+r)} \\
&= \frac{P(+c)P(S|+c)P(+r|+c)P(-w|S,+r)}{P(+c)P(+r|+c)\sum_s P(s|+c)P(-w|s,+r)} \\
&= \frac{P(S|+c)P(-w|S,+r)}{\sum_s P(s|+c)P(-w|s,+r)}
\end{aligned}
$$

- Many things cancel out – only CPTs with S remain!
- More generally: only CPTs that have resampled variable need to be considered, and joined together

Further Reading on Gibbs Sampling*

- Gibbs sampling produces sample from the query distribution P(Q | e) in limit of re-sampling infinitely often

- Gibbs sampling is a special case of more general methods called **Markov chain Monte Carlo (MCMC)** methods
  - Metropolis-Hastings is one of the more famous MCMC methods (in fact, Gibbs sampling is a special case of Metropolis-Hastings)

- You may read about Monte Carlo methods – they're just sampling

Summary

- **Prior Sampling**: sampling from P
- **Rejection Sampling:** sampling form P(Q|e)
- **Likelihood Weighting:** sampling form P(Q|e)
- **Gibbs Sampling**: sampling form P(Q|e)