Sampling
(Monte Carlo Algorithms)

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

- 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

C	P(C)
red	0.6
green	0.1
blue	0.3

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

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

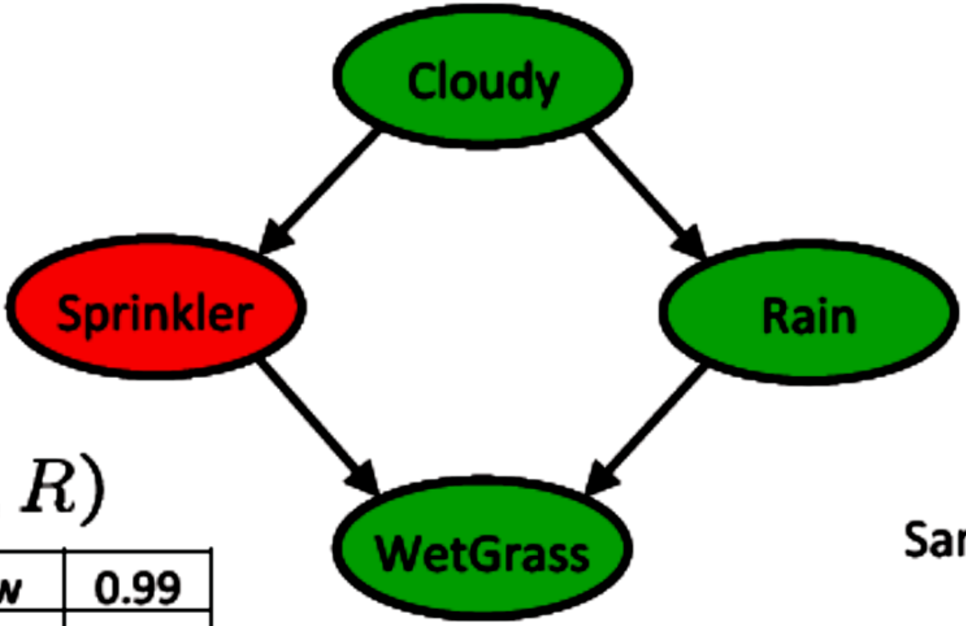
Prior Sampling

$P(C)$

+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

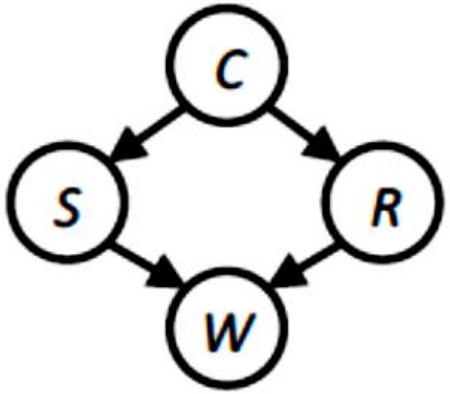
-c, +s, -r, +w

...

$$< Cloudy, Sprinkler, Rain, WetGrass > = < true, false, true, true >$$

$$P(< true, false, true, true >) = 0.5 \times 0.9 \times 0.8 \times 0.9 = 0.324$$

- 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



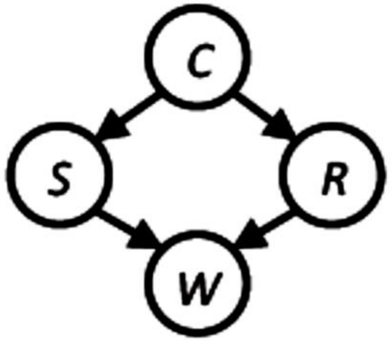
- If we want to know P(W)
  - We have counts <+w:4, -w:1>
  - Normalize to get P(W) = <+w:0.8, -w:0.2>
  - This will get closer to the true distribution with more samples

Suppose we generate M total samples and let $N(X_1, \ldots, X_n)$ be the number of samples that have produced the atomic event: $X_1, \ldots, X_n$. The important property of direct sampling is that:

$$\lim_{M \to \infty} \frac{N(X_1, \ldots, X_n)}{M} = P(X_1, \ldots, X_n)$$

Rejection Sampling

- What about P(C| +w)?   P(C| +r, +w)? P(C| -r, -w)?



Let’s say we want P(C| +s)

- Same thing: tally C outcomes, but ignore (reject) samples which don’t have S=+s
- This is called rejection sampling
- It is also consistent for conditional probabilities (i.e., correct in the limit)

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

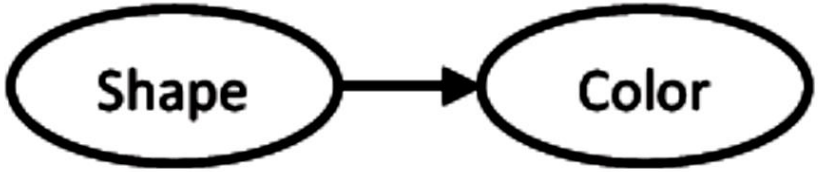
~~+c, -s, +r, +w~~
+c, +s, +r, +w
-c, +s, +r, -w
~~+c, -s, +r, +w~~
~~-c, -s, -r, +w~~

- Example
  - Estimate $\hat{P}(Rain|Sprinkler = true)$ using 100 samples
  - 27 samples have Sprinkler = true
  - Of these, 8 have Rain = true and 19 have Rain = false

$$\hat{P}(Rain = true \mid Sprinkler = true) = \frac{8}{27}$$

- Problem with rejection sampling:
  - If evidence is unlikely, rejects lots of samples
  - Evidence not exploited as you sample
  - Consider P(Shape|blue)

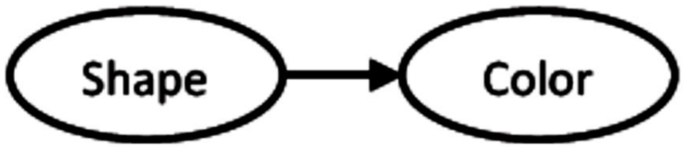
Suppose evidence is Color=blue



~~pyramid, green~~
~~pyramid, red~~
sphere, blue
~~cube, red~~
~~sphere, green~~

Likelihood Weighting

- Idea: fix evidence variables and sample the rest
  - Problem: sample distribution not consistent!
  - Solution: weight by probability of evidence given parents



~~pyramid, green~~
~~pyramid, red~~
sphere, blue
~~cube, red~~
~~sphere, green~~

pyramid, blue
pyramid, blue
sphere, blue
cube, blue
sphere, blue

- Main Idea
  - Fix evidence variables, sample only non-evidence variables
  - Weigh each sample by the likelihood of the evidence

- Set $w = 1$. For $i = 1$ to n

  If X_i is a non-evidence variable, sample $P(X_i|Parents(X_i))$

  If X_i is an evidence variable E_i, $w \leftarrow w \times P(E_i|Parents(E_i))$

- Then (X,w) forms a weighted sample

- IN: evidence instantiation
- w = 1.0
- for i=1, 2, …, n
  - if X_i is an evidence variable
    - X_i = observation x_i for X_i
    - Set w = w * $P(x_i \mid Parents(X_i))$
  - else
    - Sample x_i from $P(X_i \mid Parents(X_i))$
- return $(x_1, x_2, \ldots, x_n)$, w

$P(C)$

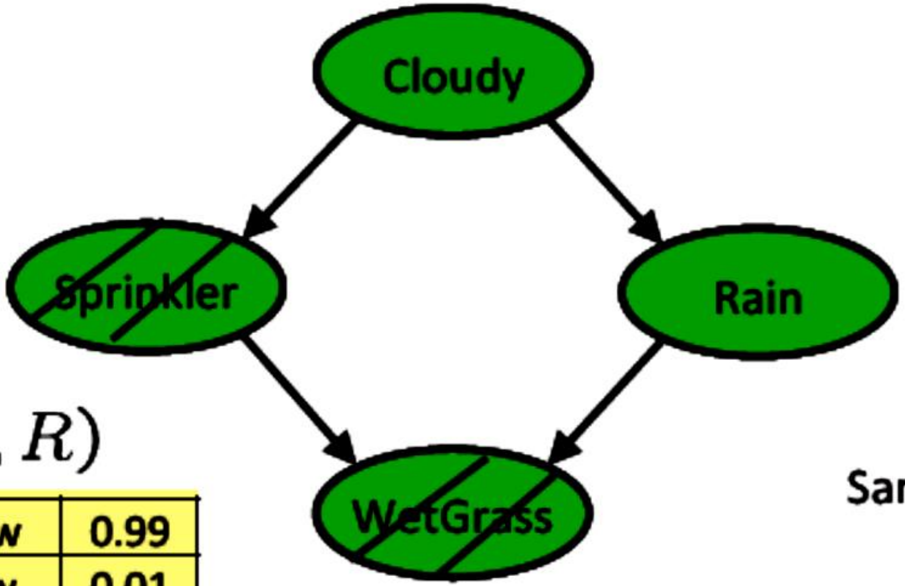
+c	0.5
-c	0.5

$P(S|C)$

+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

Evidence variables:
Sprinkler On is True
WetGrass is True

Samples:

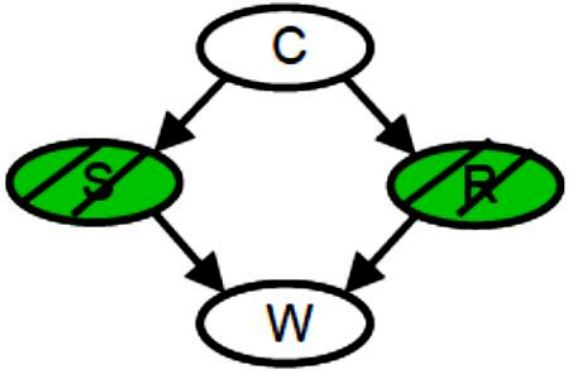
+c, +s, +r, +w

…

$$w = 1.0 \times 0.1 \times 0.99$$

weight of event <+c,+s,+r,+w> is .099

- 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
  - More of our samples will reflect the state of the world suggested by the evidence
- 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 *

- **An instance of MCMC (Markov Chain Monte Carlo)**

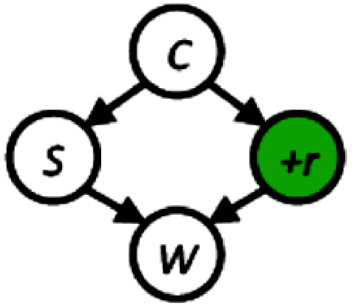
▪ *Procedure:* keep track of a full instantiation $x_1, x_2, \ldots, x_n$. Start with an arbitrary instantiation consistent with the evidence. Sample one variable at a time, conditioned on all the rest, but keep evidence fixed. Keep repeating this for a long time.

▪ *Property:* in the limit of repeating this infinitely many times the resulting sample is coming from the correct distribution

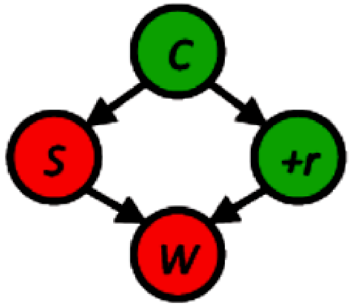
▪ *Rationale*: both upstream and downstream variables condition on evidence.

* - just FYI.

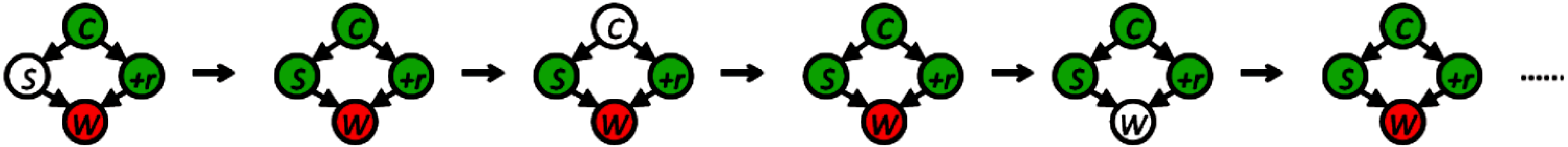
- Step 1: Fix evidence
  - R = +r



- Step 2: Initialize other variables
  - Randomly



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

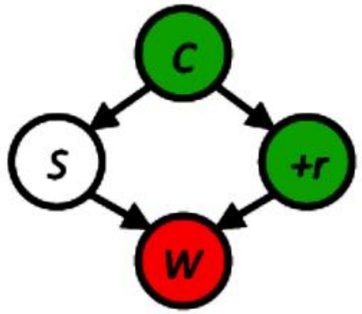


Sample from $P(S|+c, -w, +r)$ Sample from $P(C|+s, -w, +r)$ Sample from $P(W|+s, +c, +r)$

Efficient Resampling of One Variable *

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

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



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