AIFA: APPROXIMATE INFERENCE

21/03/2024

Koustav Rudra

Approximate Inference: Sampling

- Sampling is a lot like repeated simulation
  - Predicting the weather, basketball games, …

- 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 we don't know
- Inference: getting a sample is faster than computing the right answer (e.g. with variable elimination)

Sampling

- 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

C	P(C)
red	0.6
green	0.1
blue	0.3

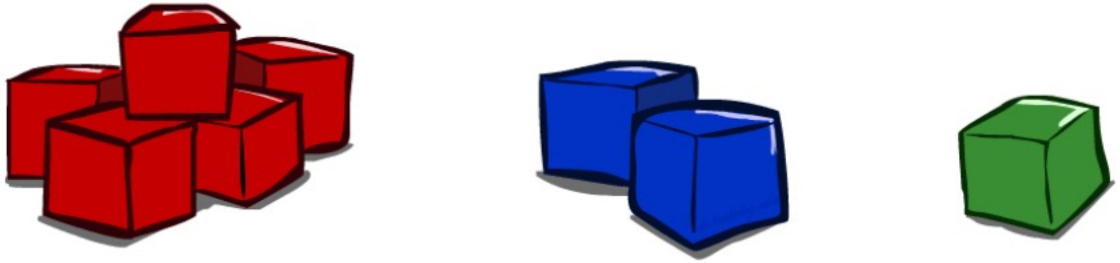
$0 \leq u < 0.6 \rightarrow C = red$

$0.6 \leq u < 0.7 \rightarrow C = blue$

$0.7 \leq u < 1 \rightarrow C = red$

If random() returns u=0.83
Our sample is C = blue

- E.g, after sampling 8 times:



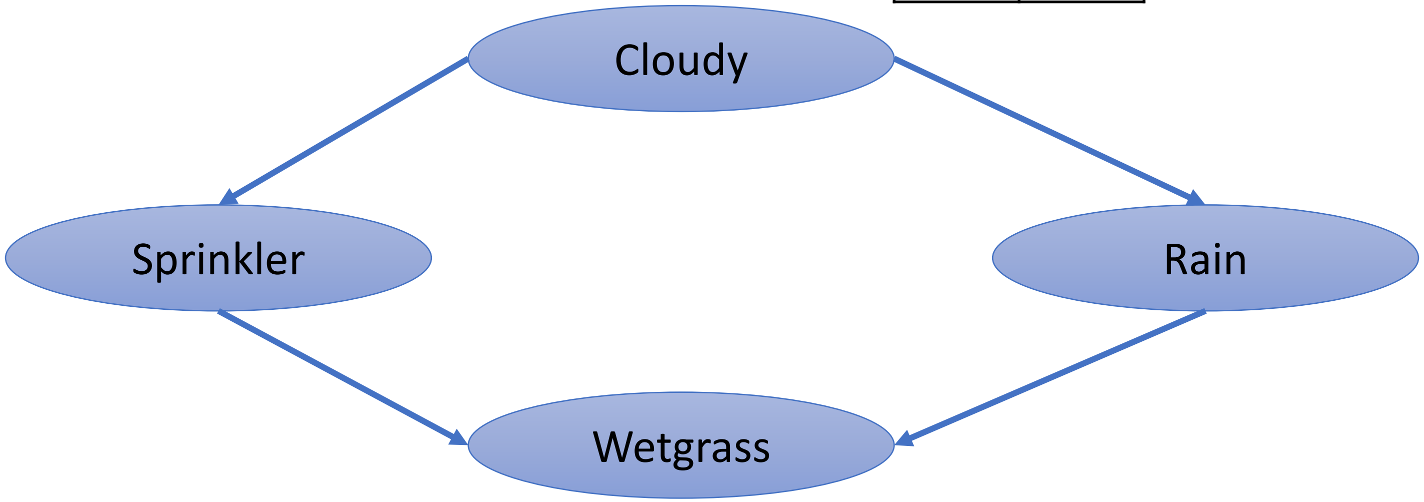
Sampling strategies

- Prior sampling
- Rejection sampling
- Likelihood weighting
- Gibbs sampling

Prior Sampling

P(C)	
+c	0.5

P(S|C)	
+c	0.1
-c	0.5



P(R|C)	
+c	0.8
-c	0.2

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

Samples:
+c, -s, +r, +w
-c, +s, -r, +w
…

Prior Sampling

- For i=1,2,…,n
  - Sample xi from P(Xi | Parents(Xi))
- Return (x1, x2, …, xn)

Prior Sampling

- This process generates samples with probability:

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

- …i.e. the BN's joint probability

- Let the number of samples of an event be $N_{PS}(x_1, x_2, \ldots, x_n)$

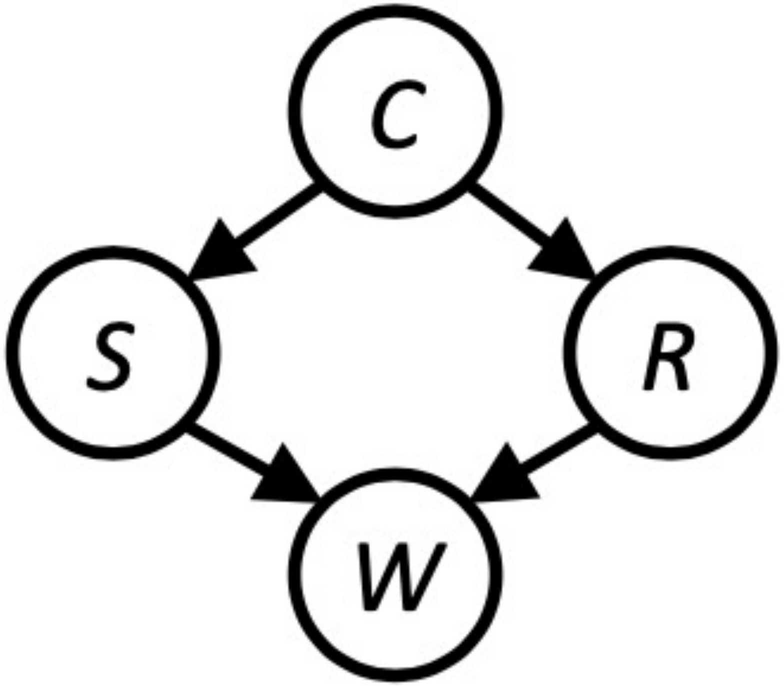
- **Then**

$$\begin{aligned} \lim_{N\to\infty} \widehat{P}(x_1, \ldots, x_n) &= \lim_{N\to\infty} N_{PS}(x_1, \ldots, x_n)/N \\ &= S_{PS}(x_1, \ldots, x_n) \\ &= P(x_1 \ldots x_n) \end{aligned}$$

- the sampling procedure is consistent

Prior Sampling

- 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



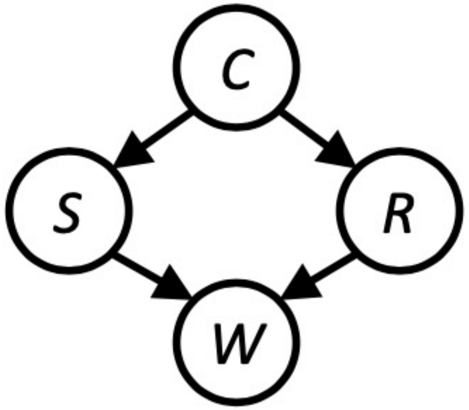
- If we want to know P(W)
- We have counts <+w:4, -w:1>
- Normalize to get P(W) = <+w:0.8, -w:0.2>
- This will get closer to the true distribution with more samples
- Can estimate anything else, too
- What about P(C| +w)? P(C| +r, +w)? P(C| -r, -w)?
- Fast: can use fewer samples if less time (what's the drawback?)

Rejection Sampling

- Let's say we want P(C)
  - No point keeping all samples around
  - Just tally counts of C as we go

- 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)



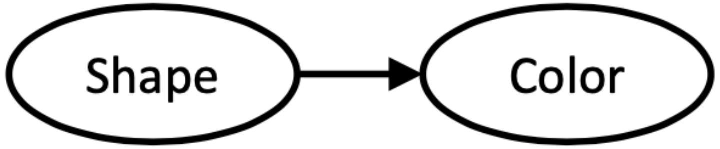
+c, -s, +r, +w
+c, +s, +r, +w
-c, +s, +r, -w
+c, -s, +r, +w
-c, -s, -r, +w

Rejection Sampling

- IN: evidence instantiation
- For i=1, 2, …, n
    - Sample x_i from $P(X_i \mid Parents(X_i))$
    - If x_i not consistent with evidence
        - Reject: Return, and no sample is generated in this cycle
- Return $(x_1, x_2, \ldots, x_n)$

Likelihood Weighting

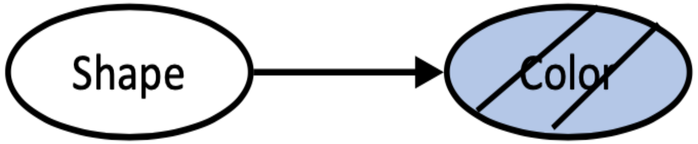
- Problem with rejection sampling:
  - If evidence is unlikely, rejects lots of samples
  - Evidence not exploited as you sample
  - Consider P(Shape|blue)



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

Idea: fix evidence variables and sample the rest

- Problem: sample distribution not consistent!
- Solution: weight by probability of evidence given parents



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

Likelihood Weighting

$P(C)$

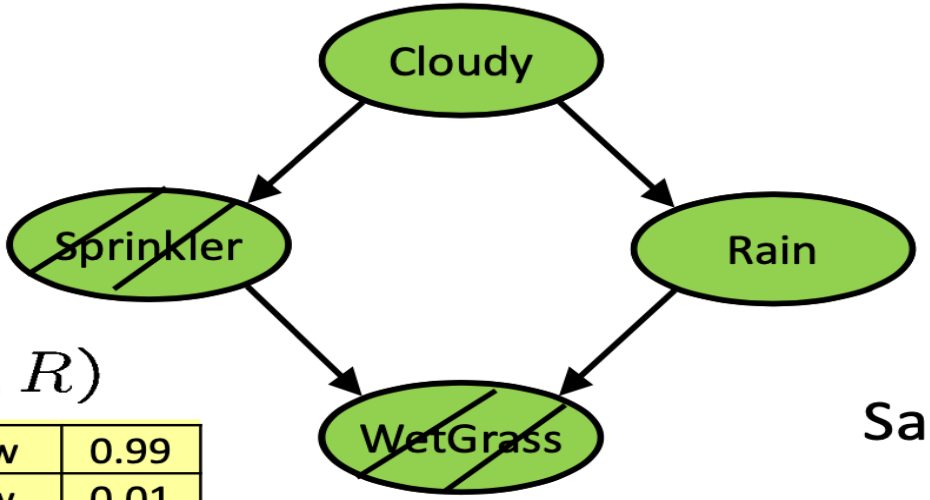
+c	0.5
-c	0.5

$P(S|C)$

+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

…

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

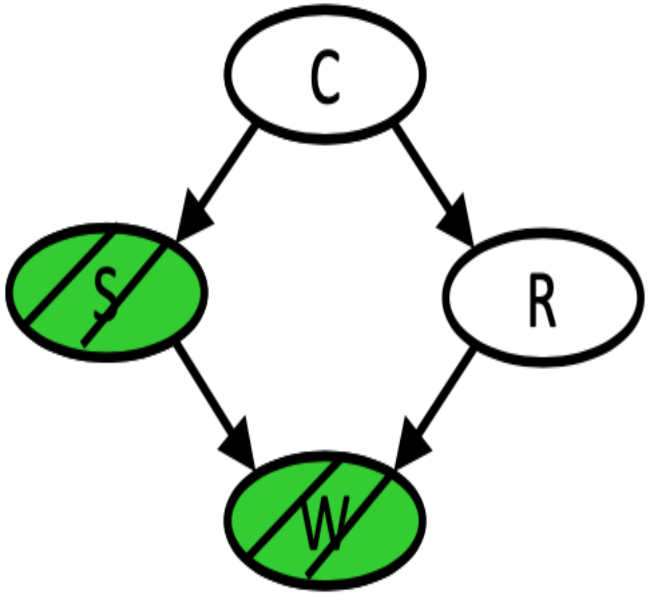
P(Rain|Sprinkler=True, WetGrass=True)

Likelihood Weighting

- IN: evidence instantiation
- $w = 1.0$
- for $i=1, 2, \ldots, n$
  - 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

Likelihood Weighting

- Sampling distribution if z sampled and e fixed evidence
  - $S_{WS}(z,e) = \prod_{i=1}^{l} P(z_i|Parents(z_i))$
- Now, samples have weights
  - $w(z,e) = \prod_{i=1}^{m} P(e_i|Parents(e_i))$
- Together, weighted sampling distribution is consistent
  - $S_{WS}(z,e)w(z,e) = \prod_{i=1}^{l} P(z_i|Parents(z_i)) \prod_{i=1}^{m} P(e_i|Parents(e_i))$
  - $S_{WS}(z,e)w(z,e) = P(z,e)$



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
  - 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

Thank You