复旦大学大数据学院
School of Data Science, Fudan University

魏忠钰

Bayes' Net Sampling

Data Intelligence and Social Computing Lab (DISC)

December 14th, 2021

Outline

- Sampling

Approximate Inference in Bayes' Net

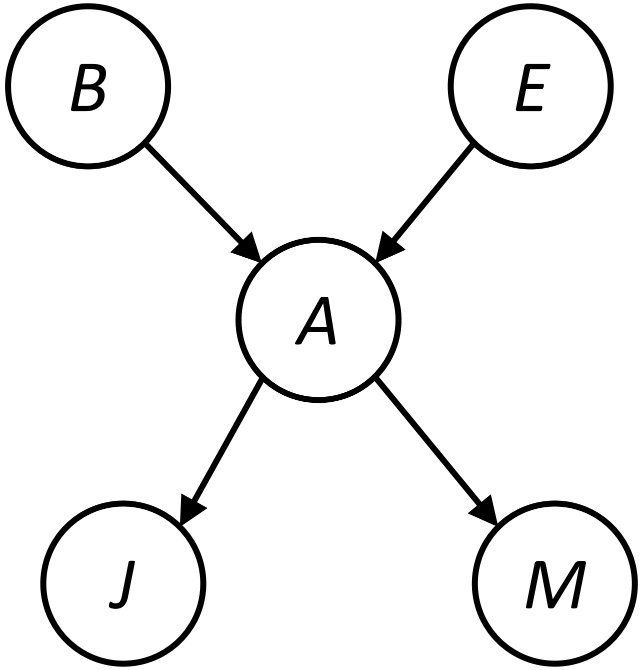
B: Burglary

E: Earthquake

A: Alarm

M: Mary calls

J: John calls



$$P(B \mid +j, +m) \propto_B P(B, +j, +m)$$
$$= \sum_{e,a} P(B, e, a, +j, +m)$$

Exact inference compute the results by enumeration.
Can be infeasible it the number of variables are large.

Can we get the results based on samples?

Using Maximum Likelihood Estimate to approximate the results.

{+b, +e, +a, +j, +m}
{+b, +e, +a, +j, +m}
{+b, -e, +a, +j, +m}
{-b, -e, +a, +j, +m}
{-b, -e, +a, +j, +m}
{-b, -e, +a, +j, -m}

B	-
+b	3/5 = 0.6
-b	2/5 = ~~.04~~ 0.4

$$P(B \mid +j, +m)$$

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 sampling?
  - Inference: getting a sample is faster than exact computation
  - Learning: the distribution S might not be available for exact computation (not covered in this course)

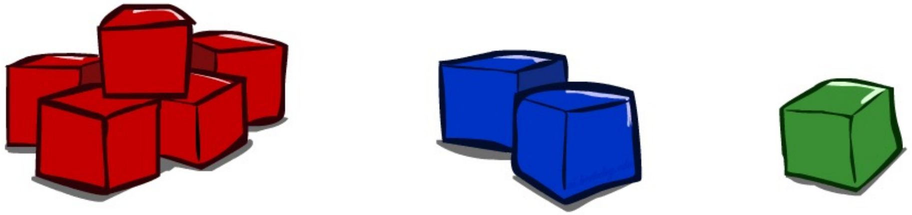
Sampling Procedure

- 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

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

- Example
  - If random() returns *u* = 0.83, then our sample is *C* = blue
  - E.g, after sampling 8 times:



Consistency of Sampling

- Suppose there are N total samples generated by approach ***S***
- $N_s(x_1, x_2, \dots x_n)$ is the number of times the event $(x_1, x_2, \dots x_n)$ occurs
- $P(x_1, x_2, \dots x_n)$ is the expected joint probability of the event

$$\hat{P}(x_1, x_2, \dots x_n) = \frac{N_s(x_1, x_2, \dots x_n)}{N}$$

$$\lim_{\mathrm{N}\to\infty} \hat{P}(x_1, x_2, \dots x_n) = P(x_1, x_2, \dots x_n)$$

- If the estimated probability $\hat{P}(x_1, x_2, \dots x_n)$ converges in the limit to its expected value, we call this sampling procedure is consistent.

Outline

- Sampling
- Prior Sampling

Prior Sampling

- Sample each variable in topological order, say, C → S → R → W.
- Its value is sampled from the probability distribution conditioned on the values assigned to its parents.
- Note that parents' value will be assigned before their children.

+c	0.5
-c	0.5

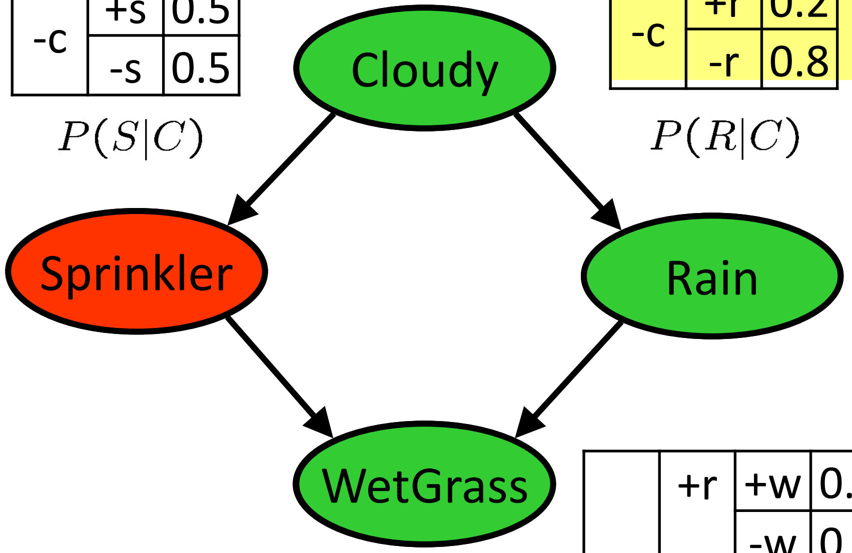
$P(C)$

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

$P(S|C)$

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

$P(R|C)$



+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

$P(W|S,R)$

For sample [+c, -s, +r, +w]

Highlighted probability distributions are used.

Prior Sampling

function PRIOR-SAMPLE(bn) **returns** an event sampled from the prior specified by bn
  inputs: bn, a Bayesian network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$

  $\mathbf{x} \leftarrow$ an event with n elements
  foreach variable X_i **in** $X_1, \ldots, X_n$ **do**
    $\mathbf{x}[i] \leftarrow$ a random sample from $\mathbf{P}(X_i \mid parents(X_i))$
  return x

Figure 14.13 A sampling algorithm that generates events from a Bayesian network. Each variable is sampled according to the conditional distribution given the values already sampled for the variable's parents.

Prior Sampling - consistency

- 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 $N_{PS}(x_1 \ldots x_n)$

- Then

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

- I.e., the sampling procedure is consistent

Example

- We' ll get a bunch of samples from the BN:

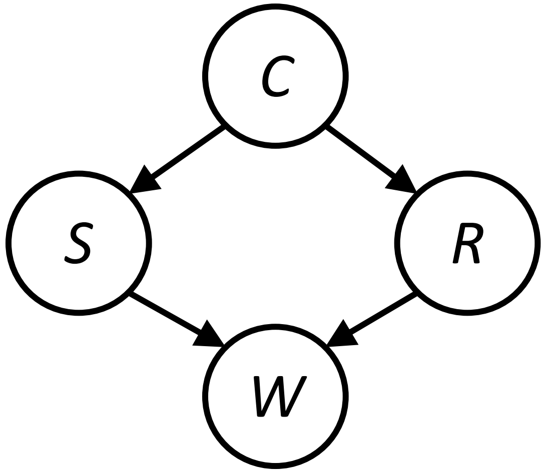
  +c, -s, +r, +w

  +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

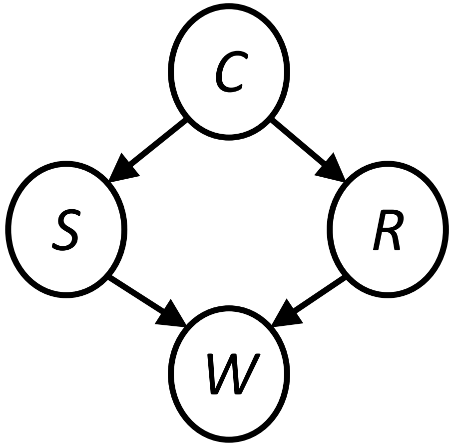
  - **What about P(C| +w)?   P(C| +r, -w)? P(C| -r, -w)?**

Outline

- Sampling
- Prior Sampling
- Reject Sampling

How about evidence?

- Let’s say we want P(C| +s)
- Rejection sampling
  - It generates samples from the prior distribution, like prior sampling.
  - It rejects all those samples do not match the evidence (S=+s in this case).
  - It counts the number of C = +c or -c in the remaining samples
  - 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

function REJECTION-SAMPLING(X, **e**, bn, N) **returns** an estimate of $\mathbf{P}(X|\mathbf{e})$
  inputs: X, the query variable
    e, observed values for variables **E**
    bn, a Bayesian network
    N, the total number of samples to be generated
  local variables: **N**, a vector of counts for each value of X, initially zero

  for j = 1 to N **do**
    x ← PRIOR-SAMPLE(bn)
    if x is consistent with **e then**
      N[x] ← **N**[x]+1 where x is the value of X in **x**
  return NORMALIZE(**N**)

Figure 14.14 The rejection-sampling algorithm for answering queries given evidence in a Bayesian network.

Reject Sampling - consistency

- The estimated probability with evidence is:

$$\hat{P}(X|\boldsymbol{e}) = \frac{N_{PS}(X,\boldsymbol{e})}{N_{PS}(\boldsymbol{e})} = \frac{N_{PS}(X,\boldsymbol{e})}{N} / \frac{N_{PS}(\boldsymbol{e})}{N}$$

- Then

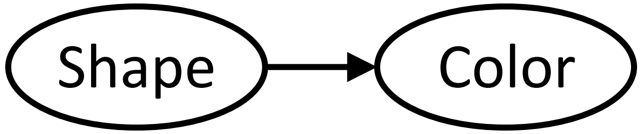
$$\lim_{N\to\infty} \hat{P}(X|\boldsymbol{e}) = \frac{P(X,\boldsymbol{e})}{P(\boldsymbol{e})} = P(X|\boldsymbol{e})$$

- The sampling procedure is consistent

Problem with rejection sampling:

- If evidence is unlikely, rejects lots of samples
- Probability of generating such evidence is not exploited in sampling
- Consider P(Shape|blue)

難收集樣本不適合



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

Outline

- Sampling
- Prior Sampling
- Reject Sampling
- Likelihood Weighting Sampling

Likelihood Weighting

- 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

- Weight each sample, initialized as 1.0
- Sample each variable in topological order, say, C → S → R → W.
- If a variable is not evidence, sample it as prior sampling. Keep the weight unchanged.
- If a variable is an evidence, use its assigned value. Multiply the weight by the probability of generating this value given its parents.

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

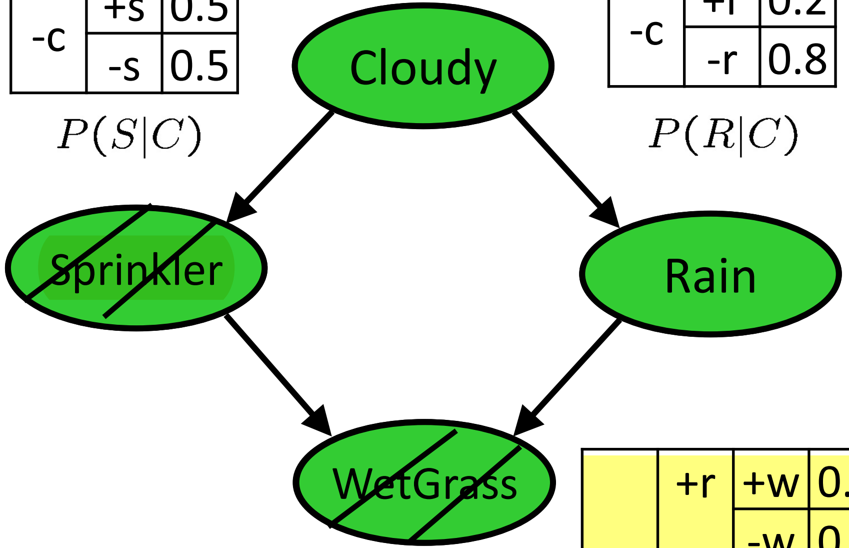
$P(S|C)$

+c	0.5
-c	0.5

$P(C)$

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

$P(R|C)$



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

$P(W|S,R)$

P(X | +s, +w)?

For sample [+c, +s, +r, +w]

$w = 1.0 \times 0.1 \times 0.99$

Likelihood Weighting

function LIKELIHOOD-WEIGHTING(X, **e**, bn, N) **returns** an estimate of $\mathbf{P}(X|\mathbf{e})$
inputs: X, the query variable
e, observed values for variables **E**
bn, a Bayesian network specifying joint distribution $\mathbf{P}(X_1, \ldots, X_n)$
N, the total number of samples to be generated
local variables: **W**, a vector of weighted counts for each value of X, initially zero

for j = 1 to N **do**
x, $w \leftarrow$ WEIGHTED-SAMPLE(bn, **e**)
$\mathbf{W}[x] \leftarrow \mathbf{W}[x] + w$ where x is the value of X in **x**
return NORMALIZE(**W**)

---

function WEIGHTED-SAMPLE(bn, **e**) **returns** an event and a weight

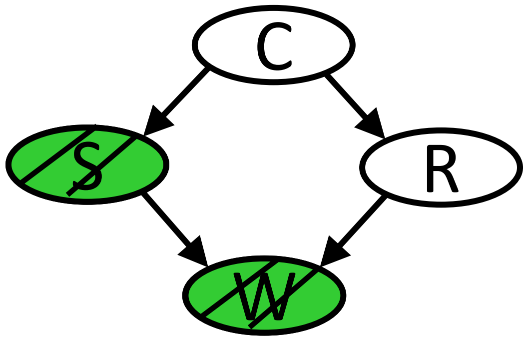
$w \leftarrow 1$; **x** $\leftarrow$ an event with n elements initialized from **e**
foreach variable X_i **in** $X_1, \ldots, X_n$ **do**
if X_i is an evidence variable with value x_i in **e**
then $w \leftarrow w \times P(X_i = x_i \mid parents(X_i))$
else $\mathbf{x}[i] \leftarrow$ a random sample from $\mathbf{P}(X_i \mid parents(X_i))$
return **x**, w

Figure 14.15 The likelihood-weighting algorithm for inference in Bayesian networks. In WEIGHTED-SAMPLE, each nonevidence variable is sampled according to the conditional distribution given the values already sampled for the variable's parents, while a weight is accumulated based on the likelihood for each evidence variable.

Likelihood Weighting

- Sampling distribution if z sampled and e fixed evidence

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



- Now, samples have weights

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

- Together, weighted sampling distribution is consistent

$$S_{\text{WS}}(z,e)\cdot w(z,e) = \prod_{i=1}^{l} P(z_i|\text{Parents}(z_i)) \prod_{i=1}^{m} P(e_i|\text{Parents}(e_i))$$

n 趨近於無限大

$$= P(\mathbf{z},\mathbf{e})$$