

Approximate Inference 1

1

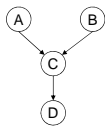
Forward Sampling

- This section on approximate inference relies on samples / particles
- Full particles: complete assignment to all network variables eg. $(X_1 = x_1, X_2 = x_2, \dots, X_N = x_N)$

2

Forward Sampling

- **Topological sort or order:** An ordering of the nodes in the DAG where X comes before Y in the ordering if there is a directed path from X to Y in the graph.
- A topological order is equivalent to a partial order on the nodes of the graph
- There may be several topological orderings



Examples of Topological orders:
A,B,C,D
B,A,C,D

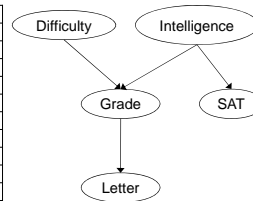
3

Forward Sampling

Student Example

D	P(D)
low	0.6
high	0.4

D	I	G	P(G D,I)
low	low	C	0.3
low	low	B	0.4
low	low	A	0.3
low	high	C	0.02
low	high	B	0.08
low	high	A	0.9
high	low	C	0.7
high	low	B	0.25
high	low	A	0.05
high	high	C	0.2
high	high	B	0.3
high	high	A	0.5



I	P(I)
low	0.7
high	0.3

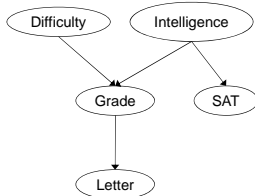
I	S	P(S I)
low	low	0.95
low	high	0.05
high	low	0.2
high	high	0.8

G	L	P(L G)
C	weak	0.99
C	strong	0.01
B	weak	0.4
B	strong	0.6
A	weak	0.1
A	strong	0.9

4

Forward Sampling

Topological ordering: D, I, G, S, L



1. Sample D from P(D) (Say you get D=high)
2. Sample I from P(I) (Say you get I=low)
3. Sample G from P(G|I=low,D=high) (Say you get G=C)
4. Sample S from P(S|I=low) (Say you get S=low)
5. Sample L from P(L|G=C) (Say you get L=weak)

You now have a sample (D=high, I=low, G=C, S=low, L=weak)

5

Forward Sampling

Suppose you want to calculate $P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ using forward sampling on a Bayesian network. The algorithm:

1. Do a topological sort of the nodes in the Bayesian network.
2. For $j = 1$ to $NUM_SAMPLES$
For each node i in the ordering (starting from the top of the Bayesian network down)
Sample the value \hat{x}_i from the distribution $P(X_i | Parents(X_i))$
Add $\{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n\}$ to your collection of samples
3. Let
$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \approx \frac{\# \text{ of samples with } X_1 = x_1, X_2 = x_2, \dots, X_n = x_n}{NUM_SAMPLES}$$

6

Forward Sampling

- How do you sample from $P(X_i | Parents(X_i))$?
- Note: $P(X_i | Parents(X_i))$ is a multinomial distribution $P(x_i^1, \dots, x_i^k | \theta_1, \dots, \theta_k)$?

7

Forward Sampling

How do you sample from a multinomial distribution $P(x_i^1, \dots, x_i^k | \theta_1, \dots, \theta_k)$?

- Generate a sample s uniformly from $[0, 1]$
- Partition interval into k subintervals: $[0, \theta_1), [\theta_1, \theta_1 + \theta_2), \dots$
- More generally, the i th interval is

$$\left[\sum_{j=1}^{i-1} \theta_j, \sum_{j=1}^i \theta_j \right)$$

- If s is in the i th interval, the sample value is x_i
- Use binary search to find the interval for s in time $O(\log k)$

8

Forward Sampling

Suppose your list of samples looks like the following table:

D	I	G	S	L
low	low	B	low	weak
low	high	A	high	strong
low	high	A	high	weak
high	high	A	high	strong
high	low	C	low	weak

$$P(I=high) = 3/5 = 0.6$$

Note that this value becomes a lot more accurate as the number of samples heads to infinity.

9

Forward Sampling

- From a set of samples $D = \{\xi[1], \dots, \xi[M]\}$, we can estimate the expectation of any function f as:

$$\hat{E}_D(f) = \frac{1}{M} \sum_{m=1}^M f(\xi[m])$$

- To estimate $P(y)$

$$\hat{P}_D(y) = \frac{1}{M} \sum_{m=1}^M I\{y[m] = y\}$$

This will return 1 if the sample $\xi[m]$ matches the values in y (eg. y is I=high) and 0 otherwise.

10

Rejection Sampling

11

Rejection Sampling

What if we want to estimate $P(y|E=e)$?

- **Rejection sampling**: do forward sampling but throw out samples where $E \neq e$

Example:

$$P(I=high|L=weak) = 1/3$$

D	I	G	S	L
low	low	B	low	weak
low	high	A	high	strong
low	high	A	high	weak
high	high	A	high	strong
high	low	C	low	weak

12

Rejection Sampling

What if the evidence $E=e$ is very very rare?

- For example, if $P(e) = 0.001$, then for 10,000 samples, we get 10 unrejected samples
- To obtain at least M^* unrejected samples, we need to generate on average $M = M^*/P(e)$ samples
- If evidence is rare, we end up generating a lot of samples which wastes time

13

Rejection Sampling

Bad news:

- Rare evidence is the norm!
- As # of evidence variables $k = |E|$ grows, the probability of the evidence decreases exponentially with k

Need something better than rejection sampling!

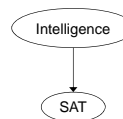
14

Likelihood Weighting

15

Likelihood Weighting

Intuition: Weight samples according to probability of the evidence



I	P(I)
low	0.7
high	0.3

I	S	P(S I)
low	low	0.95
low	high	0.05
high	low	0.2
high	high	0.8

Drawing I = high and S = high should be 80% of a sample

Drawing I = low and S = high should be 5% of a sample

16

Likelihood Weighting

Weighted particles:

$$D = \langle \xi[1], w[1] \rangle, \dots, \langle \xi[M], w[M] \rangle$$

Estimate:

$$\hat{P}_D(y | e) = \frac{\sum_{m=1}^M w[m] \mathbf{I}\{y[m] = y\}}{\sum_{m=1}^M w[m]}$$

17

Likelihood Weighting

```

Procedure LW-Sample(
     $\beta$ , // Bayesian network over  $\mathcal{X}$ 
     $\mathbf{Z}=\mathbf{z}$  // Event in the network
)

```

1. Let X_1, \dots, X_n be a topological ordering of \mathcal{X}
2. $w \leftarrow 1$
3. **for** $i = 1, \dots, n$
4. $\mathbf{u}_i \leftarrow \mathbf{x} \langle \text{Pa}_{X_i} \rangle$ // Assignment to Pa_{X_i} in X_1, \dots, X_{i-1}
5. **if** $X_i \notin \mathbf{Z}$ **then**
6. Sample x_i from $P(X_i | \mathbf{u}_i)$
7. **else**
8. $x_i \leftarrow \mathbf{z} \langle X_i \rangle$ // Assignment to X_i in \mathbf{z}
9. $w \leftarrow w \cdot P(x_i | \mathbf{u}_i)$ // Multiply weight by probability of desired value
10. **return** $(x_1, \dots, x_n), w$

18