Approximate inference (Ch. 14)
Bayesian Network: Efficiency

Last time we talked about how exact inference is fine if you have a polytree.

Otherwise, exact inference is exponential $O(2^n)$ and not really feasible.

Instead we use an approximate approach, specifically we will look at Monte Carlo approaches that utilize sampling (this let’s use balance runtime with accuracy).
Sampling can mean different things:

(1) Sample an unknown distribution
   - Much like running an experiment
     Tickle friend’s nose while asleep
     ... see how many times they react

(2) Sample from a known distribution
   - Might also call this “simulation”
     - Generate a random number to decide outcome of an event

we will use this way
Direct Sampling

The first method is called **direct sampling**, which is basically just running a simulation and tallying the results.

Today we will use this simple Bay-net (work):

- **P(a)** = 0.2
- **P(b|a)** = 0.4
- **P(b|¬a)** = 0.01
- **P(c|a,b)** = 1
- **P(c|a,¬b)** = 0.7
- **P(c|¬a,b)** = 0.3
- **P(c|¬a,¬b)** = 0
Direct Sampling

Direct Sampling algorithm:

- Loop this a lot (N times)
  - Repeat until all nodes have values:
    1. Find any node with all parents having been given a value already
    2. Generate a random number (0 to 1)
    3. Assign value to node based off of $P(\text{node} \mid \text{Parents}(\text{node}))$
- Calculate statistics
Direct Sampling

(1) Only node who has all parents with values is node A (as it has no parents)

(2) Pretend random value is: 0.183712

(3) Since 0.183712 \leq 0.2, set node A to a (true)
Direct Sampling

(1) Only node who has all parents with values is node B (as only A has a value)

(2) Pretend random value is: 0.910184

P(b|a), as A is a (i.e. a=true)

(3) Since 0.910184 > 0.4, set node B to be ¬b
Direct Sampling

(1) Only node left is C (has both parents)

(2) Pretend random value is: 0.634523

(3) Since 0.634523 < 0.7, set node C to $c$
Direct Sampling

After running the inner loop once, we have a sample of (in format [A,B,C]):

[a, \neg b, c]

... we would then repeat this process N times (outer loop) to get a bunch of these.

Pretend you got the results on the next slide.
Direct Sampling

1. [a, ¬b, c]
2. [a, b, c]
3. [¬a, b, c]
4. [¬a, ¬b, ¬c]
5. [¬a, ¬b, ¬c]
6. [¬a, ¬b, ¬c]
7. [¬a, ¬b, ¬c]
8. [¬a, ¬b, ¬c]
9. [¬a, ¬b, ¬c]
10. [¬a, ¬b, ¬c]

From here we can calculate statistics of anything...

For example:

\[ P(¬c) = \frac{\text{count of } "¬c"}{\text{total possible times}} \]

\[ = \frac{7}{10} \]
Direct Sampling

In fact, you can estimate $P(a, b, c)$ from this:

1. $[a, \neg b, c]$
2. $[a, b, c]$
3. $[\neg a, b, c]$
4. $[\neg a, \neg b, \neg c]$
5. $[\neg a, \neg b, \neg c]$
6. $[\neg a, \neg b, \neg c]$
7. $[\neg a, \neg b, \neg c]$
8. $[\neg a, \neg b, \neg c]$
9. $[\neg a, \neg b, \neg c]$
10. $[\neg a, \neg b, \neg c]$

$P(\neg a, \neg b, \neg c) = 0.7$
How would you compute: $P(a|b)$
Rejection Sampling

How would you compute:

\[ P(a|b) \]

You do the same counting, but only look at entries with “b” being true...

... thus \( P(a|b) = 0.5 \)
Rejection Sampling

This technique is called rejection sampling, as you reject/ignore any samples that do not have the given conditional information.

For direct sampling, with $N$ samples:

$$P(a, b, c) = P(a)P(b|a)P(c|b, a) = \lim_{N \to \infty} \frac{\text{count}(a,b,c)}{N}$$

... or more generally...

$$P(x_1, \ldots x_n) = \prod_{i=1}^{n} P(x_i|\text{Parents}(X_i)) = \lim_{N \to \infty} \frac{\text{count}(x_1, \ldots x_n)}{N}$$

Let us call the right hand side: $N_{PS}(a, b, c)$
Rejection Sampling

From here it is fairly easy to prove that rejection sampling is also finding the correct probability (assuming many samples):

\[ P(a|b) = \frac{N_{PS}(a, b)}{N_{PS}(b)} = \lim_{N \to \infty} \frac{\left( \frac{\text{count}(a,b)}{N} \right)}{\left( \frac{\text{count}(b)}{N} \right)} = \frac{\text{count}(a,b)}{\text{count}(b)} \]

... or let “x” be what we want to find and “e” be the given info (here “e” = \{b\}, but both “x” and “e” could be multiple variables, like

\[ P(x|e) = \lim_{N \to \infty} \frac{\left( \frac{\text{count}(x,e)}{N} \right)}{\left( \frac{\text{count}(e)}{N} \right)} = \frac{\text{count}(x,e)}{\text{count}(e)} \quad \text{“e”} = \{b,c\} \]
Rejection Sampling

As number of samples, \( N \), grows our accuracy of approximating probabilities gets better.

Using the Law of Large Numbers, we can find that the standard deviation for our estimates is:

\[
\frac{\sigma}{\sqrt{N}} \approx \frac{1}{\sqrt{N}}
\]

as we using probabilities, the mean & std dev in \([0,1]\) (small)

in rejection sampling, 
\( N = \) number non-rejected samples

So when we found \( P(a|b) = 0.5 \) (with 2 samples), we are 68.2% confident that \( P(a|b) \) is within:

\[
\left[ 0.5 - \frac{0.5}{\sqrt{2}}, 0.5 + \frac{0.5}{\sqrt{2}} \right] = [0.15, 0.85]\
\]
Good Sampling?

What is the general issue(s) with direct and/or rejection sampling? (When is it good?)
Good Sampling?

What is the general issue(s) with direct and/or rejection sampling? (When is it good?)

These sampling techniques are pretty good for finding non-conditional probability: $P(a,b,c)$

However, if the given information is restrictive, many samples will be rejected... leading to poor approximations of the probabilities
Good Sampling?

The given information (also called “evidence”) can be restrictive because:

1. the tables have low probabilities
2. many variables have to be satisfied

You will need exponentially more samples as you increase number of given variables.

If \( P(y) = P(z) = 0.5 \), this table shows number of samples for same accuracy:

| \( P(x) \) | 100 |
| \( P(x|y) \) | 200 |
| \( P(x|y,z) \) | 400 |
Likelihood Weighting

There a way to not waste time generating “rejected” samples called likelihood weighting.

As mentioned before, direct sampling is decent at finding non-conditional probabilities.

So for likelihood weighting we will assume we want to find a conditional probability.
Likelihood weighting, will weight samples:
For $P(a|b)$: $[a,b] \ w = 1$, $[\neg a,b] \ w=0.2$

If we did rejection sampling, we need about 5 $\neg a$ to actually get a ‘b’, so in 10 samples:
$[a,b]$, $[a,b]$, $[a,b]$, $[a,b]$, $[a,b]$, $[\neg a,b]$, $[\neg a,\neg b]$, $[\neg a,\neg b]$, $[\neg a,\neg b]$, $[\neg a,\neg b]$
Likelihood Weighting

Since we normalize, all we care about is the ratio between \([a,b]\) and \([\neg a,b]\)

In likelihood weighting, the weights create the correct ratio as “\([\neg a,b] : w=0.2\)” represents that you would actually need 5 of these to get a “true” sample
Likelihood Weighting

We will use a bit of notation here:

- $x =$ things we want the probability of
- $e =$ “evidence” or given info
- $y =$ anything else

So in our original sample network:

- $P(a|b) : x = \{a\}, e = \{b\}, y = \{c\}$
- $P(a|b,c) : x = \{a\}, e = \{b,c\}, y = \{\} $
- $P(a,b|c) : x = \{a,b\}, e = \{c\}, y = \{\}$

must be non-empty  
assume non-empty for this alg
Likelihood Weighting

Likelihood weighting algorithm:
- Assign all given variables into network
- \( w = 1 \) // our “weight”
- Do once for every node:
  (1) Find a node where all parents have values
  (2a) If node given info (in set “e”):
    \[ w = w \times P(\text{given} \mid \text{Parents(\text{given})}) \]
  (2b) Else (in sets “x” or “y”)
    Generate random number to determine T/F
- Repeat above a lot and calculate statistics
Likelihood Weighting

Since we are finding $P(a|b)$, we initially set $b=true$ in the network (and start $w=1$)

From here we need to loop through all three nodes, finding any node that has all of its parents with values

|   | $P(a)$ | 0.2 | $P(c|a,b)$ | 1   |
|---|-------|-----|-------------|-----|
|   | $P(b|a)$ | 0.4 | $P(c|a,\neg b)$ | 0.7 |
|   | $P(b|\neg a)$ | 0.01 | $P(c|\neg a,b)$ | 0.3 |
|   | $P(c|\neg a,\neg b)$ | 0  |               |     |
Likelihood Weighting

(1) A is only one with all parents having values, so pick A to look at

(2a) A is not given information, so we generate a random number: 0.746949

0.746949 > 0.2, so we set A to \( \neg a \)
Likelihood Weighting

(1) Here we could pick ‘b’ or ‘C’ as ‘b’ has its parent and C has values for ‘a’ and ‘b’ ... let’s pick B

(2b) B is given information, so we simply multiply “w” by the probability $P(b|\neg a)$

$$w = w \times P(b|\neg a) = 1 \times 0.01 = 0.01$$

as A sampled to be $\neg a$ this time

<table>
<thead>
<tr>
<th></th>
<th>$P(a)$</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(b</td>
<td>a)$</td>
<td>0.4</td>
</tr>
<tr>
<td>$P(b</td>
<td>\neg a)$</td>
<td>0.01</td>
</tr>
</tbody>
</table>

|      | $P(c|a,b)$ | 1   |
|------|------------|-----|
| $P(c|a,\neg b)$ | 0.7 |
| $P(c|\neg a,b)$ | 0.3 |
| $P(c|\neg a,\neg b)$ | 0  |
Likelihood Weighting

If multiple given variables, \( w = \text{product of all (multiple times)} \)

(1) C is only node left... pick that

(2a) C is not given information, so generate random number to sample/simulate:
\[
0.987924 > 0.3, \text{ so set } C \text{ to } \neg c
\]

\[
\begin{array}{|c|c|}
\hline
P(a) & 0.2 \\
\hline
P(b|a) & 0.4 \\
\hline
P(b|\neg a) & 0.01 \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|}
\hline
P(c|a,b) & 1 \\
\hline
P(c|a,\neg b) & 0.7 \\
\hline
P(c|\neg a,b) & 0.3 \\
\hline
P(c|\neg a,\neg b) & 0 \\
\hline
\end{array}
\]
Likelihood Weighting

Now we have a single sample: $[\neg a, \neg c] : w=0.01$

We would then repeat this process, say $N$ times (make sure to reset $w=1$ every time)

Afterwards we would have a bunch of weighted samples where $b=true$ always
... pretend they turned out as the next slide
Likelihood Weighting

tells us $P(a, c|b)$

1. $[a, c] : w=0.4$
2. $[a, c] : w=0.4$
3. $[\neg a, c] : w=0.01$
4. $[\neg a, c] : w=0.01$
5. $[\neg a, \neg c] : w=0.01$
6. $[\neg a, \neg c] : w=0.01$
7. $[\neg a, \neg c] : w=0.01$
8. $[\neg a, \neg c] : w=0.01$
9. $[\neg a, \neg c] : w=0.01$
10. $[\neg a, \neg c] : w=0.01$

Rather than doing a direct tally, we sum the weights... so:

$$P(a|b) = \frac{0.4 + 0.4 + 0.01 + 0.01 + 0.01 + 0.01 + 0.01 + 0.01 + 0.01 + 0.01}{10} = 0.909$$

This is also just our normalization trick...

$P(a|b) = \alpha 0.8, P(\neg a|b) = \alpha 0.08$
You try it! Calculate $P(a|c)$ using this alg. and using these random numbers (20 of them):

<table>
<thead>
<tr>
<th>Random Numbers</th>
<th>0.784</th>
<th>0.859</th>
<th>0.934</th>
<th>0.760</th>
<th>0.543</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.532</td>
<td>0.967</td>
<td>0.229</td>
<td>0.781</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>0.168</td>
<td>0.439</td>
<td>0.873</td>
<td>0.415</td>
<td>0.471</td>
</tr>
<tr>
<td></td>
<td>0.053</td>
<td>0.646</td>
<td>0.694</td>
<td>0.325</td>
<td>0.368</td>
</tr>
</tbody>
</table>

use left to right, top to bottom
## Likelihood Weighting

1. \([\neg a, \neg b] : w=0\)  
   You should get these
2. \([\neg a, \neg b] : w=0\)  
   samples from the
3. \([\neg a, \neg b] : w=0\)  
   random simulation
4. \([\neg a, \neg b] : w=0\)  
5. \([\neg a, b] : w=0.3\)  
   Thus:
6. \([a, \neg b] : w=0.7\)  
   \(P(a|c) = \alpha 0.7\)
7. \([\neg a, \neg b] : w=0\)  
   \(P(\neg a|c) = \alpha 0.3\)
8. \([\neg a, \neg b] : w=0\)  
9. \([\neg a, \neg b] : w=0\)  
   So, \(P(a|c) = 0.7\)
10. \([\neg a, \neg b] : w=0\)
Likelihood Weighting

Any issues with this?
Likelihood Weighting

Any issues with this?

When \( w=0 \), this is basically like rejection sampling...

This happens because you do not consider the children when generating samples

In our example, \( A=\text{true} \) dominated the total weight (0.8 of 0.88), leading to accuracy issues
Likelihood Weighting

Why does this weight trick work? In our prob:

\[ P(a|c) = \alpha \text{TrueA}(w\text{For}(\text{inTable})) \]
\[ = \alpha P(a = \text{trueInTable}) \cdot w\text{For}(\text{inTable}) \]
\[ = \alpha \sum_b P(a, b) \cdot w\text{For}(\text{inTable}) \]
\[ = \alpha \sum_b P(a)P(b|a) \cdot w\text{For}(\text{inTable}) \]
\[ = \alpha \sum_b \underbrace{P(a)P(b|a)}_{\text{percent in table}} \cdot \underbrace{P(c|a, b)}_{\text{weight of sample}} \]
\[ = \alpha \sum_b P(a, b, c) \]
\[ = \alpha P(a, c) \]
I mentioned this in the algorithm, but did not do an example: weight’s product is cumulative.

So if we want to find \( P(a|b,c) \), say 3 samples: 

- \([a]\) : \( w = 0.4 \), \([a]\) \( w = 0.4 \times 1 = 0.4 \)
- \([\neg a]\) : \( w = 0.01 \times 0.3 = 0.003 \)