CS188 Section 9: Bayes Sampling

October 25th, 2017

Inference on Bayes Net

Inference given Joint Distribution

You'll be asked to calculate P(Query|evidence). How?

- **Enumeration (never really covered in discussion, but it's not hard)**
- **Variable Elimination (just earlier)**
- **Sampling (approximate)**

Inference By Enumeration

Goal: Find P(Query | Evidence)

Given:

Evidence:	$E_1 = e_1, E_2 = e_2, \ldots E_k = e_k$	
Query variable(s):	Q	
Hidden variables:	$H_1, H_2, \ldots H_R$	*(all the extraneous random vars in the Bayes Net)*

1. Compute full JPT via Bayes Net and parent equation
2. Select entries consistent with evidence to get P(Query, Evidence, Hidden)
3. Sum out (a.k.a. marginalize) all of H_i for P(Query, Evidence)
4. Normalize to get P(Query | Evidence)

Inference By Variable Elimination

Goal: Find P(Query | Evidence)

Given:

Evidence:	$E_1 = e_1, E_2 = e_2, \ldots E_k = e_k$	
Query variable(s):	Q	
Hidden variables:	$H_1, H_2, \ldots H_R$	*(all the extraneous random vars in the Bayes Net)*

Let factors = tables

1. Initially, each CPT is a factor.
   a. For each factor, select entries matching evidence.
2. Eliminate hidden variables:
   a. While hidden_variables not empty:
      i. New factor = Pick H, Join all factors with H, Sum out H
3. Normalize to get P(Query | Evidence)

Why eliminate?

You don't want to compute the full JPT like done in step one of enumeration. We eliminate as much as we can early so we never loop over a huge table.

Inference on Bayes Net: Sampling

Sampling

As used in Prior Sampling

(and Rejection Sampling)

Generate Samples from Bayes Net

We want to sample from the Joint Distribution, but we have a Bayes Net, not the actual JPT.

To sample from the JPT:

- Starting with roots of the Bayes Net, you'll sample from the marginal distributions.
- Then you'll sample from the selected CPT of children given parents
    - (on the sample you drew for the parents)

Example: Generate Samples from Bayes Net

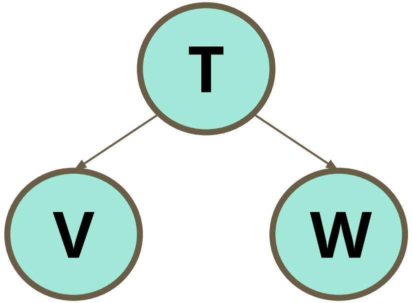
Sample Being Generated = (?, ?, ?)

Sample from:

P(T)

T =	**P**
+t	0.5
-t	0.5

→ got **+t**



Example: Generate Samples from Bayes Net

Sample Being Generated = (+t, ?, ?)

Fix parents, then sample from:

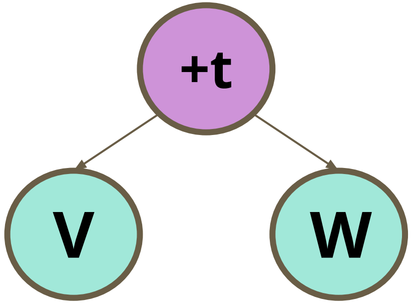
P(V|-t)

V =	P
+v	0.2
-v	0.8

P(V|+t)

V =	P
+v	0.9
-v	0.1

→ got -v



Example: Generate Samples from Bayes Net

Sample Being Generated = (+t, -v, ?)

Fix parents, then sample from:

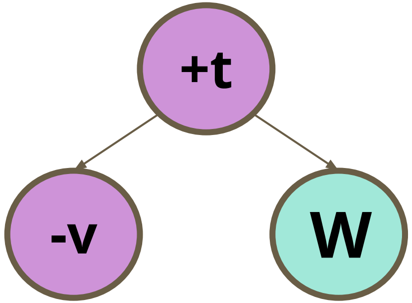
~~P(W|-t)~~

~~W =~~	~~P~~
~~+w~~	~~0.6~~
~~-w~~	~~0.4~~

P(W|+t)

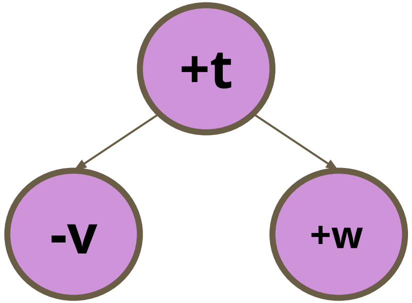
W =	P
+w	0.7
-w	0.3

→ got +w



Example: Generate Samples from Bayes Net

Generate one Sample = (+t, -v, +w)



In Prior Sampling, we always keep the sample.

In Rejection Sampling, we might reject the sample.

Prior and Rejection Sampling

Approximate Inference: Prior Sampling

Main Idea

- Generate samples normally n times.
- Treat these as the sample space and compute probabilities.
  - **P(X = +x)** = **# Samples where X = +x** (divided by) **Total # of Samples**

For **inference**, you want to compute a conditional query **P(A| -b)**

- P(+a| -b) =
  - **# Samples where A=+a and B = -b** (divided by) **# of Sample where B = -b**
- P(-a| -b) =
  - **# Samples where A=-a and B = -b** (divided by) **# of Sample where B = -b**
- **a post-sampling rejection of samples that don't match B = -b**

Approximate Inference: Rejection Sampling

- For **inference**, want to compute some conditional query **P(A | -b)**
- Same as Prior Sampling, but throw away all samples that don't match the evidence +b **during the sample generation**--- as opposed to after you generated samples.

Implications:

- Can't stop sampling until you've gotten n samples that match your evidence
- Can only compute queries of the given evidence (you threw away everything else!)

Likelihood Weighting

Likelihood Weighting > Rejection Sampling

Want to find: P(A| -b)

- **Problem:** If evidence occurs with low likelihood, you'll waste a lot of time generating samples just to throw them away.
- **Solution:** Fix the evidence as "seen", sample, then account for the likelihood of having seen the evidence, given the sample you saw. Treat weighted samples as the sample space.

$$evidence = \{E_i = e_1, E_2 = e_2, \ldots\}$$
$$weight(sample) = \prod_i P(e_i \mid parents(e_i))$$

where $parents(e_i)$ is the sample values of $parents(E_i)$

Likelihood Weighting

Want to find: P(A| -b)

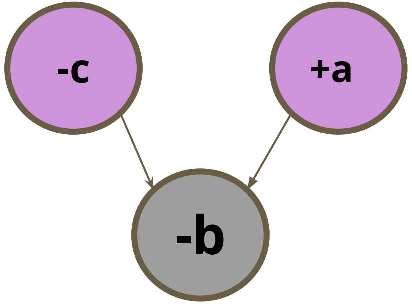
1. Generate samples, fixing evidence **(B = -b)** when we are required to sample it
2. Give weights to the samples (probability that had you not fixed the evidence, the evidence would have been realized)
3. Consider the weighted samples to be your sample space
    a. Analogy is that the samples were weighted 1 in rejection sampling
    b. Now, you sum the weights instead of counting the number of samples

Likelihood Weighting: Example for P(A|-b)

$$evidence = \{E_i = e_1, E_2 = e_2, ...\}$$
$$weight(sample) = \prod_i P(e_i \mid parents(e_i))$$

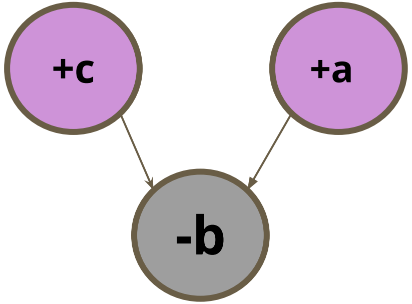
where $parents(e_i)$ is the sample values of $parents(E_i)$

Sample 1



Weight = P(-b|-c, +a)

Sample 2



Weight = P(-b|+c, +a)

Likelihood Weighting: Generate Samples

When we generate a sample, we still sample the Bayes Net from **top down**:

1. Sample regularly for everything upstream of the evidence.
   a. Note that because our evidence is ignored, things upstream are unaffected by it
2. When you encounter an evidence variable, **DON'T SAMPLE**
   a. **Fix the evidence** (i.e. $\mathbf{E_i = +e_i}$) instead of sampling.
   b. If we had truly sampled, the **likelihood of evidence** being the fixed value is dependent on the sampled upstream, specifically parent, variables
3. Anything downstream of $\mathbf{E_i}$ is sampled as regular.
   a. Note this means that because we fixed $E_i = +e_i$, things downstream are affected by it.
4. Once you have a sample, we **weight our sample** on the likelihood of actually having seen it (i.e. **weight = $\mathbf{P(E_i = +e_i \mid \text{sampled parents of } E_i)}$**) because we didn't truly sample.

Likelihood Weighting: Generate Samples

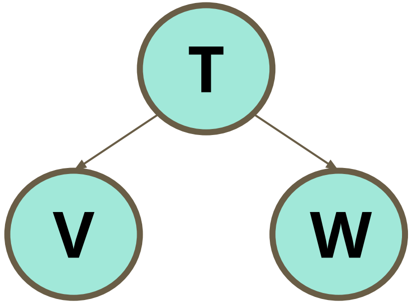
Say we want to know P(W|-v)
Sample Being Generated = (?, ?, ?)

First, sample from:

P(T)

T =	P
+t	0.5
-t	0.5

→ got **+t**



Likelihood Weighting: Generate Samples

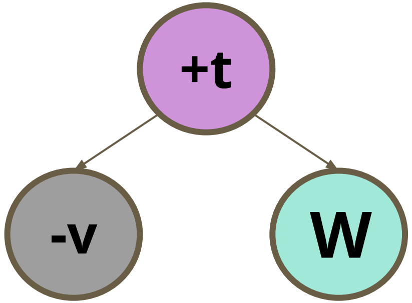
Say we want to know P(W|-v)
Sample Being Generated = (+t, ?, ?)

Instead of sampling from P(V|+t), **fix V = -v**:

P(V|+t)

V =	P
+v	0.9
-v	0.1

Because we didn't actually sample, the likelihood of the sample actually occurring is the **weight**



Likelihood Weighting: Generate Samples

Say we want to know P(W|-v)
Sample Being Generated = (**+t**, **-v** , ?)

Continue, so next sample from:

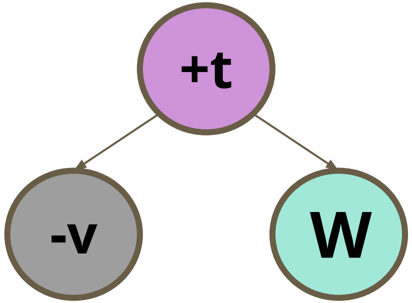
~~**P(W|-t)**~~

~~W =~~	~~P~~
~~+w~~	~~0.6~~
~~-w~~	~~0.4~~

P(W|+t)

W =	P
+w	0.7
-w	0.3

→ got **+w**

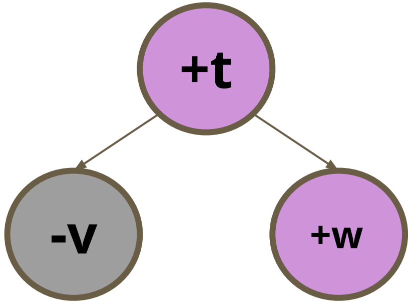


Likelihood Weighting: Generate Samples

Say we want to know P(W|-v)

Generated one Sample = (+t, **-v**, +w)

Likelihood weight = **P(V = -v | +t) = 0.1**



Likelihood Weighting: Important Stuff

- We're fixing evidence because we don't want to waste even a little bit of energy creating a sample that we would reject
- Because we ignore evidence until we encounter it during the sampling process, upstream variables are **NOT** affected by the evidence we fix
- In fact, we allow the sampling of upstream variables to determine the weights of our sample. **Weight = P(Evidence | Sampled Upstream)**
  - Could be having samples with low weight being overwhelmed by samples with high weight
  - In other words, super low (relative to other samples) weighted samples are basically "rejected"
    - INEFFICIENT!!!! If only we had account for evidence before sampling upstream variables.

Gibbs Sampling

Gibbs Sampling > Likelihood Weighting

Question: Does knowing the outcome of a child node affect the sampling of a parent node?

Yes, Gibbs Sampling tries to account for this so that we don't generate samples with upstream variables that would give the sample low weight.

Gibbs Sampling

Idea: Always sample **conditioned on everything (not just parents)**. That means we have to completely ditch the basic sampling method

How?

1. Start somewhere
2. Pick one part to change
3. Change it conditioned on everything else in the sample (parents, children, siblings, everything)
4. This is a new sample
5. Repeat