# 188 Section 8: Bayes Nets

Wednesday, October 8, 2017

# Bayes Nets

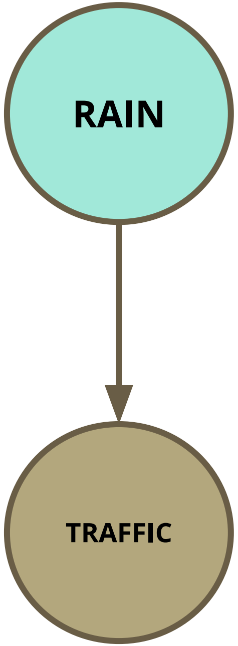
# Bayesian Networks

From last week's section on probability review, we know that asking questions from a probabilistic model requires that we know the **joint distribution**

But that's so big and messy!

**Bayes Nets:** Describes conditional independence relations.
Now we only need to worry about conditional distributions b/w variables of the model ⇒ **Implicitly gives us the joint distribution**

# Rain “influences” traffic




Bayes net + parent equation gives us the full joint of the model:

$$P(X_1, X_2, X_3 \ldots, X_n) = \prod_{i}^{n} P(X_i \mid Parents(X_i))$$

We get these tables for each of the nodes!

# Conditional Independence in Bayes Nets

# Conditional Independence

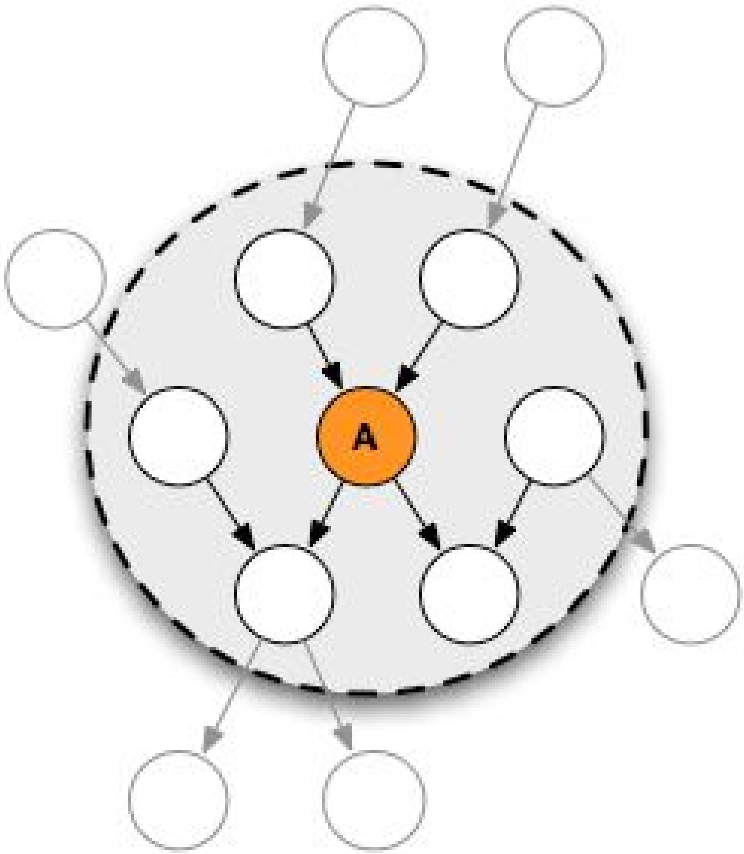
- **Regular Independence:** If X and Y are independent random variables, **$P(X \mid Y) = P(X)$,** and vice versa!

- **Conditional Independence:** If X is conditionally independent of Y **given** Z, **$P(X \mid Z, Y) = P(X \mid Z)$.**

  If we are given Z, X is not affected by event Y!

# Markov Blanket(Node) = ...

A node in a Bayes Net is conditionally independent from the rest of the net given **its parents, children, and all of its children's parents.**




Extension: A **leaf** node in a BN is conditionally independent from the others given all parents.

# D-Separation: “Are A and B independent, given D and F?”

For hairy situations, you might have a hard time looking at the graph and deciding independence.

Here’s **one way** to help you reason it out!...

(this is not the same technique used in lecture, but same principles. You should definitely understand the method from lecture. If they don’t specify you to consider active/inactive triplets, I find the following method much easier.)

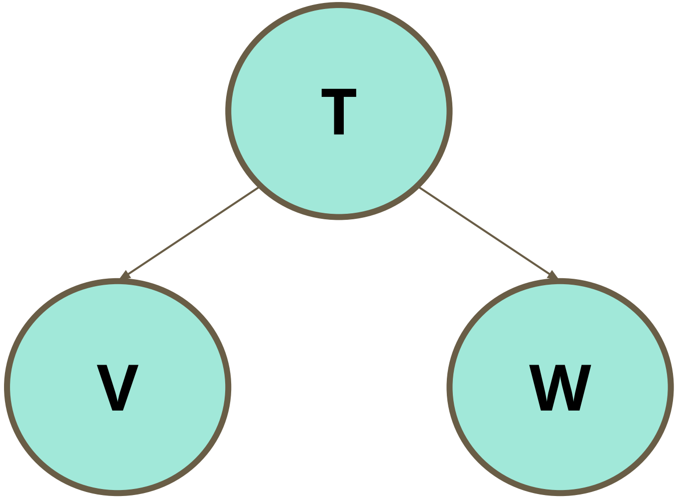
# D-Separation via Moralization: "Are A and B independent, given D and F?"

1. Draw ancestral graph (include ancestor nodes of all nodes mentioned in expression + node itself)
2. Moralize the graph
   a. For each pair of variables with a common child, draw an undirected edge (line) between them. (If a variable has more than two parents, draw lines between every pair of parents.)
3. "Disorient" the graph
   a. replacing the directed edges (arrows) with undirected edges (lines).
4. Delete the givens and their edges.
   a. **For the statement above, it would be D, F**
5. Read the graph
   a. If the variables are disconnected in this graph, they are **guaranteed to be independent**.
   b. If the variables are connected in this graph, they are **not guaranteed to be independent.**

You can prove independence via probability gymnastics too!

Using the information encoded in the Bayes Net and also all your probability rules and tricks

# Is V independent of W given T?




**From the Bayes Net:**

$$P(V, W, T) = P(T)P(W \mid T)(V \mid T)$$

$$P(W, T) = P(T)P(W \mid T)$$

**Conditional Probability Definition:**

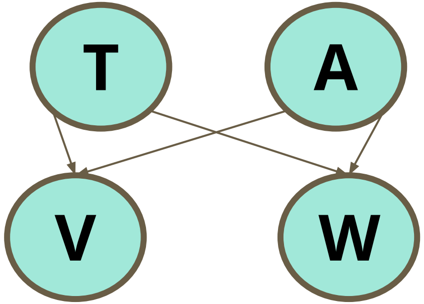
$$P(V \mid W, T) = \frac{P(V, W, T)}{P(W, T)}$$

**Combine and use definition of Conditional Independence:**

$$P(V \mid W, T) = \frac{P(V, W, T)}{P(W, T)} = \frac{P(T)P(W \mid T)P(V \mid T)}{P(T)P(W \mid T)} = P(V \mid T)$$

**YES**

# Is V independent of W given T?




**From the Bayes Net + Marginalization:**

$$P(V,W,T) = \sum_A P(A,V,W,T) = \sum_A P(A)P(T)P(W \mid T,A)P(V \mid T,A)$$

$$P(W,T) = \sum_A P(A,W,T) = \sum_A P(A)P(T)P(W \mid T,A)$$

**Conditional Probability Definition:** $P(V \mid W,T) = \dfrac{P(V,W,T)}{P(W,T)}$

**Combine:** $P(V \mid W,T) = \dfrac{P(V,W,T)}{P(W,T)} = \sum_A \dfrac{P(A)P(T)P(W \mid T,A)P(V \mid T,A)}{P(A)P(T)P(W \mid T,A)} = \sum_A P(V \mid T,A)$

**(Try to) use definition of Conditional Independence:**

$$P(V \mid W,T) = \sum_A P(V \mid T,A) \overset{?}{=} P(V \mid T)$$

**Not necessarily**

(see slides at the end if you don't see this)

# Variable Elimination

# We know

$$P(X_1, X_2, X_3 \ldots, X_n) = \prod_i^n P(X_i \mid Parents(X_i))$$

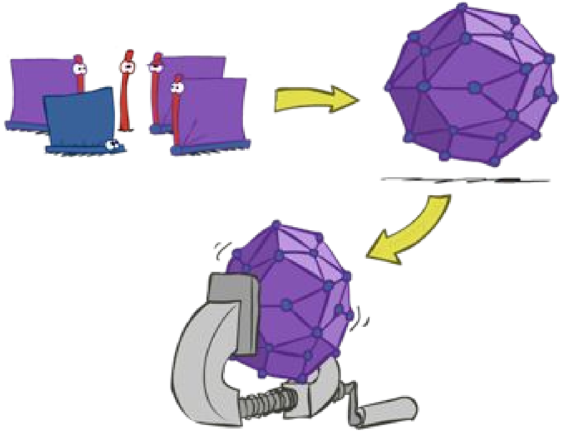
1. Bayes nets + the CPTs (conditional probability tables) + the Parent equation gives us the **full joint distribution**

2. If you have a full joint distribution, you can do magic with **marginalization and normalization** to get a specific CPT.

3. **Inference by enumeration:** We can compute out joint probability tables, then marginalize to get what we want, then apply formulas to get the answer to some query.

# But....

- **Inference by Enumeration is inefficient-** you build up the entire table then **marginalize** the probability that you care about. That's not very efficient.



# Remember: Normalization Trick

Say we have some CPT we are trying to find: **P(Y | +z)**

If we had the (marginal) JPT, **P(Y, Z)** we know we could use the **normalization trick** to get **P(Y | +z)**.

**P(Y, Z)**

| Y = | Z = | P |
|---|---|---|
| +y | +z | 0.2 |
| -y | +z | 0.3 |
| +y | -z | 0.1 |
| -y | -z | 0.4 |

Fix evidence →

**P(Y, +z)**

| Y = | P |
|---|---|
| +y | 0.2 |
| -y | 0.3 |

normalize →

**P(Y |+z)**

| Y = | P |
|---|---|
| +y | 0.4 |
| -y | 0.6 |

# Remember: Normalization Trick

So we know that

**P(Y |+z) is proportional to P(Y, +z)**

**LET'S FOCUS ON P(Y, +z)!!!! (later normalize)**

**P(Y, +z)**

| Y = | P |
|---|---|
| +y | 0.2 |
| -y | 0.3 |

normalize $\longrightarrow$

**P(Y |+z)**

| Y = | P |
|---|---|
| +y | 0.4 |
| -y | 0.6 |

# Find P(Y, +z)

Now we can use the Bayes net (+parent equation) and **marginalization:**
(note that Z is fixed to be +z)

$$P(Y,+z) = \sum_T \sum_V \sum_W \sum_U \sum_X P(T)P(U \mid T)P(V \mid T)P(W \mid T)P(Y \mid V,W)P(X \mid T)P(+z \mid X)$$

Oh god. **That's nasty.** It's the same problem as enumeration. Computing the JPT (and marginalizing after) will be inefficient.

But instead of computing a gigantic JPT, we'll notice that we can **push some of the summations inwards (marginalize early)**!!

# Eliminate X

alternate notation:

$f_1(+z, T)$

Let's push the summation on X as far in as we can (I wrote it nicely, but sometimes, usually, you rearrange terms).

$$P(Y,+z) = \sum_T \sum_V \sum_W \sum_U P(T)P(U \mid T)P(V \mid T)P(W \mid T)P(Y \mid V,W) \sum_X P(X \mid T)P(+z \mid X)$$

$f_1(+z|T)$

We can take this and just consider to be some **factor** that is determined by T and +z (X disappears as we sum over it).

This **factor** is small and easier to compute. Furthermore, it gets rid of X so we don't have to waste computation over X later.

# Eliminate the rest

You got rid of X

$$P(Y,+z) = \sum_T \sum_V \sum_W \sum_U P(T)P(U \mid T)P(V \mid T)P(W \mid T)P(Y \mid V,W)f_1(z+ \mid T)$$

Now we'll get rid of T:

(need to rearrange first)

$$P(Y,+z) = \sum_V \sum_W \sum_U P(Y \mid V,W) \sum_T P(T)P(U \mid T)P(V \mid T)P(W \mid T)f_1(z+ \mid T)$$

$$P(Y,+z) = \sum_V \sum_W \sum_U P(Y \mid V,W)f_2(U,V,W,+z)$$

And then U:

$$P(Y,+z) = \sum_V \sum_W P(Y \mid V,W) \sum_U f_2(U,V,W,+z)$$

$$P(Y,+z) = \sum_V \sum_W P(Y \mid V,W)f_3(V,W,+z)$$

And then V and W:

(because I'm lazy)

$$P(Y,+z) = f_5(Y,+z)$$

# Oh Schnitz. You found P(Y, +z)!!! Normalize!

Perfection! Now remember how we said this is proportional to our goal?

$$P(Y, +z) = f_5(Y, +z)$$

$$[P(Y, +z) = f_5(Y, +z)] \propto P(Y \mid +z)$$

... how did we get to our goal again?

**NORMALIZATION!!!!!!**

# Put it all together: Variable Elimination

1. Fix your evidence (in the example, it would be Z = +z). Set up your goal as a **marginalization of the joint.**

2. **Break up the joint** using the Bayes Net (and the parent equation) to find your factors.

3. [1]**Eliminate each variable** ***V*** that is not in your goal (in the example, it's everything but Y and +z):
   a. Collect the factors that depend on ***V***
   b. Sum out ***V*** to create a factor that only depends on what remains in the factors you collected

4. The final factor is what you need to **normalize**!

# Summary: 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)... next time

# Inference By Enumeration

**Goal:** Find P(Query | Evidence)

**Given:**

*Evidence:* $E_1 = e_1, E_2 = e_2, \dots E_k = e_k$

*Query variable(s):* Q

*Hidden variables:* $H_1, H_2, \dots H_R$ *(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:* $E_1 = e_1, E_2 = e_2, \dots E_k = e_k$

*Query variable(s):* Q

*Hidden variables:* $H_1, H_2, \dots H_R$ *(all the extraneous random vars in the Bayes Net)*

**Let** factors = tables

1. Initially, each CPT is a factor. For each factor, select entries consistent with evidence.
2. Eliminate hidden variables:
   a. While hidden_variables not empty:
      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.

# [continued from slide 11]

# $\sum_A P( V \mid T, A) = P( V \mid T)$ ???

# Let's work with something concrete....

**Sanity Check**
Make sure you know why there are four tables...
T and A have 2 outcomes each.

[See speaker notes at bottom]

## $P(V \mid T, A) =$

**P(V|+t, +a)**

| V = | P |
|---|---|
| +v | 0.5 |
| -v | 0.5 |

**P(V|-t, +a)**

| V = | P |
|---|---|
| +v | 0.2 |
| -v | 0.8 |

**P(V|+t, -a)**

| V = | P |
|---|---|
| +v | 0.1 |
| -v | 0.9 |

**P(V|-t, -a)**

| V = | P |
|---|---|
| +v | 0.7 |
| -v | 0.3 |

## Each table sums up to 1

# Smash some tables together...

$$P(V \mid T, A) =$$

$P(V|T, +a)$

| V = | T= | P |
|---|---|---|
| +v | +t | 0.5 |
| -v | +t | 0.5 |
| +v | -t | 0.2 |
| -v | -t | 0.8 |

$P(V|T, -a)$

| V = | T= | P |
|---|---|---|
| +v | +t | 0.1 |
| -v | +t | 0.9 |
| +v | -t | 0.7 |
| -v | -t | 0.3 |

**Each table sums up to 2**

# One last smash

**P( V | T , A ) =**

| V = | T= | A= | P |
|---|---|---|---|
| +v | +t | +a | 0.5 |
| -v | +t | +a | 0.5 |
| +v | -t | +a | 0.2 |
| -v | -t | +a | 0.8 |
| +v | +t | -a | 0.1 |
| -v | +t | -a | 0.9 |
| +v | -t | -a | 0.7 |
| -v | -t | -a | 0.3 |

**Table sums up to 4**

# Sum over A

$$\sum_A P(V \mid T, A) =$$

**Table STILL sums up to 4**

| V = | T= | A= | P |
|---|---|---|---|
| +v | +t | +a | 0.5 |
| -v | +t | +a | 0.5 |
| +v | -t | +a | 0.2 |
| -v | -t | +a | 0.8 |
| +v | +t | -a | 0.1 |
| -v | +t | -a | 0.9 |
| +v | -t | -a | 0.7 |
| -v | -t | -a | 0.3 |

# Compare $\sum_A P( V \mid T, A)$ and $P( V \mid T )$

As a metric (if T can be either +t or -t) we know that **$P( V \mid T )$** sums up to 2

In the previous run through, **$\sum_A P( V \mid T, A)$** sums up to 4.

**Obviously, the two don't give us the same table…**

What happens if **A** is ALWAYS **+a** (never anything else)? What would that mean in terms of conditional independence?