BAYES' NETS

Lara J. Martin (she/they)
TA: Aydin Ayanzadeh (he)

11/14/2023

CMSC 671

By the end of class today, you will be able to:

- Recognize a simple solution for fixing MLE
- Represent the independence of random variables
- Calculate a probability using a Bayes' Net

Modified from slides by Dr. Chris Callison-Burch, Dr. Cynthia Matuszek, & Dr. Brian Hrolenok

COURSE SCHEDULE

- HW3 is due TONIGHT
  - You're allowed to use up to 3 of your 5 late days on it
- You can use up to 1 late day for each final project deadline (<u>except the last one</u>)
- First milestone due 11/16
  - NOTE: Meeting with me or Aydin is part of the grade (due when the proposal is due)
- Module 4 Presentations are 11/16
  - Summaries due tomorrow

REVIEW: MLE EXAMPLE

$$P(w_i|w_{i-1}) = \frac{count(w_{i-1}, w_i)}{count(w_{i-1})}$$

<s> I am Sam </s>
<s> Sam I am </s>
<s> I do not like green eggs and ham </s>

$P(\text{I}|\text{<s>}) = \frac{2}{3} = .67$ $\quad P(\text{Sam}|\text{<s>}) = \frac{1}{3} = .33$ $\quad P(\text{am}|\text{I}) = \frac{2}{3} = .67$

$P(\text{</s>}|\text{Sam}) = \frac{1}{2} = 0.5$ $\quad P(\text{Sam}|\text{am}) = \frac{1}{2} = .5$ $\quad P(\text{do}|\text{I}) = \frac{1}{3} = .33$

Example from "Speech and Language Processing", Chapter 3 by Jurafsky & Martin

REVIEW: ISSUE WITH MLE: ZEROS

I will not eat them in the rain.
I will not eat them on a train.
Not in the dark! Not in a tree!
Not in a car! You let me be!
I do not like them in a box.
I do not like them with a fox.
I will not eat them in a house.
I do not like them with a mouse.
I do not like them here or there.
I do not like them anywhere!
I do not like green eggs and ham!
I do not like them, Sam-I-am.

$$P(\text{"}them\text{"} \mid \text{"}I\ do\ not\ like\text{"}) = \frac{6}{7}$$

$$P(\text{"}pink\text{"} \mid \text{"}I\ do\ not\ like\text{"}) = 0$$

LAPLACE (ADD ONE) SMOOTHING

I will not eat them in the rain.
I will not eat them on a train.
Not in the dark! Not in a tree!
Not in a car! You let me be!
I do not like them in a box.
I do not like them with a fox.
I will not eat them in a house.
I do not like them with a mouse.
I do not like them here or there.
I do not like them anywhere!
I do not like green eggs and ham!
I do not like them, Sam-I-am.

$$P(\text{"}them\text{"}|\text{"}I\ do\ not\ like\text{"}) = \frac{6}{7}$$

$$P(\text{"}pink\text{"}|\text{"}I\ do\ not\ like\text{"}) = 0$$

Let V be the size of our vocabulary.

$$P(\text{"}them\text{"}|\text{"}I\ do\ not\ like\text{"}) = \frac{(6+1)}{(7+V)}$$

$$P(\text{"}pink\text{"}|\text{"}I\ do\ not\ like\text{"}) = \frac{(0+1)}{(7+V)}$$

Too much probability to unseen words

Explanation inspired by Dr. David A. Smith

GENERALIZED ADDITIVE SMOOTHING

I will not eat them in the rain.
I will not eat them on a train.
Not in the dark! Not in a tree!
Not in a car! You let me be!
I do not like them in a box.
I do not like them with a fox.
I will not eat them in a house.
I do not like them with a mouse.
I do not like them here or there.
I do not like them anywhere!
I do not like green eggs and ham!
I do not like them, Sam-I-am.

$$P(\text{"them"}|\text{"I do not like"}) = \frac{6}{7}$$

$$P(\text{"pink"}|\text{"I do not like"}) = 0$$

Let V be the size of our vocabulary.

$$P(\text{"them"}|\text{"I do not like"}) = \frac{(6+\lambda)}{(7+V)}$$

$$P(\text{"pink"}|\text{"I do not like"}) = \frac{(0+\lambda)}{(7+V)}$$

How do we pick a good λ?

Held out data!

Explanation inspired by Dr. David A. Smith

BAYES' NETS

PROBABILISTIC MODELS

- Models describe how (a portion of) the world works
- What do we do with probabilistic models?
  - We (or our agents) need to reason about unknown variables, given evidence
  - Example: explanation (diagnostic reasoning)
  - Example: prediction (causal reasoning)
- Models are always simplifications
  - May not account for every variable
  - May not account for all interactions between variables

How can we know how variables interact?

INDEPENDENCE

- Two variables are independent if

$$\forall x, y : P(x, y) = P(x)P(y)$$

Meaning that their joint distribution *factors* into a product of two simpler distributions.

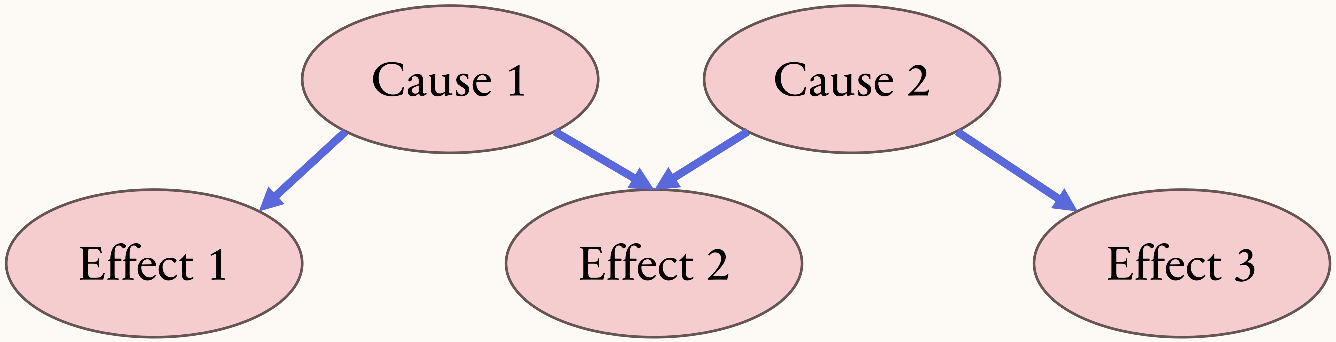
or

$$\forall x, y : P(x|y) = P(x)$$

- Independence is written as: $X \perp\!\!\!\perp Y$
- It is a *simplifying assumption* used for modeling.
- *Empirical* joint distributions can be used to approximate independence

REPRESENTING INDEPENDENCE

- Independent relationships can be thought of as **cause** & **effect**
- What can we assume about the relationship between
  {weather, traffic, cavity, toothache}?
- We can represent dependent relationships through directed acyclic graphs where the nodes are random variables and the edges represent the relationship from cause to effect

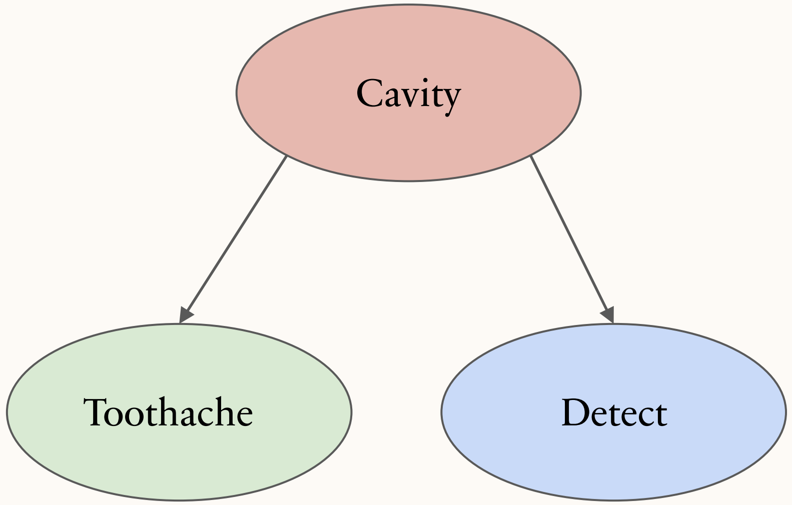


IF the graph accurately represents the independence structure, nodes are **independent** of their siblings given their immediate parents: $Toothache \perp Catch \mid Cavity$

CONDITIONAL PROBABILITY TABLE (CPT)

If we have a cavity, this may be the reason why we have a toothache but also the reason why the dentist detects a cavity.

For **discrete** RV's we can represent the conditional probability as a table



p(Cav)	p(¬Cav)
0.2	0.8

	p(Det|Cav)	p(¬Det|Cav)
Cav=T	0.9	0.1
Cav=F	0.6	0.4

	p(Tth|Cav)	p(¬Tth|Cav)
Cav=T	0.6	0.4
Cav=F	0.1	0.9

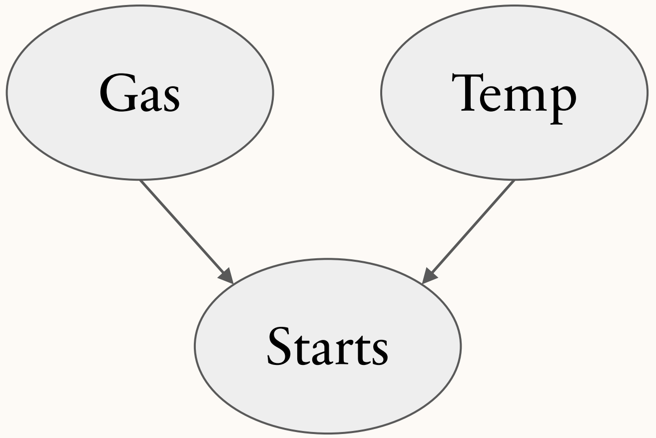
IF the graph accurately represents the independence structure, parents are **independent** if **not** conditioned on common children: $Gas \perp Temp$

BAYES' NETS: MULTIPLE PARENTS

Nodes can have more than one parent, or none

- CPT includes all parents
- Nodes without parents: marginals

“Car has fuel” “Weather is warm”



“Car starts”

	p(Starts|Gas,Temp)	p(¬Starts|Gas,Temp)
Gas=T, Temp=T	0.9	0.1
Gas=T, Temp=F	0.8	0.2
Gas=F, Temp=T	0.5	0.5
Gas=F,Temp=F	0.3	0.7

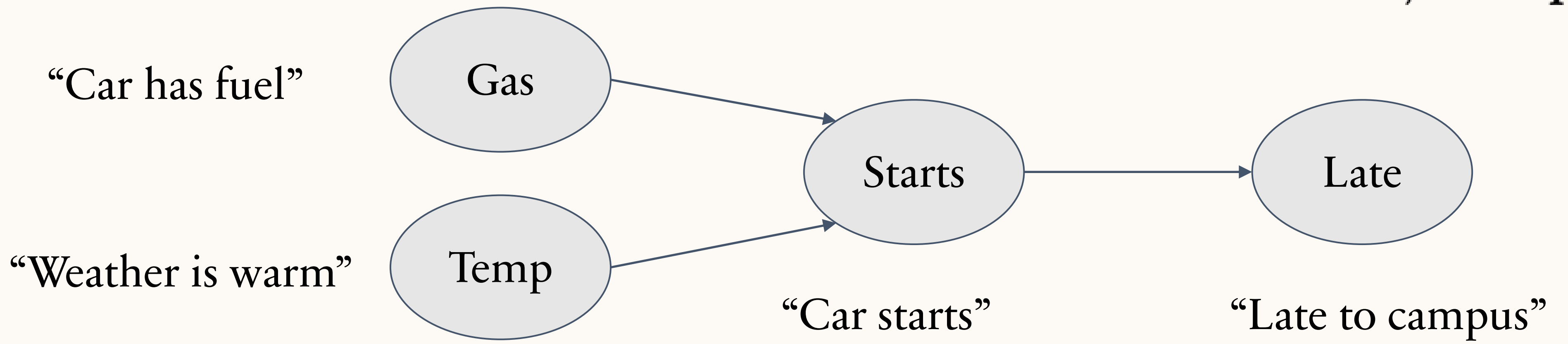
BUT if conditioned on a common child, parents are **no longer** independent (knowing effect influences the probability of both causes)

$$Gas \not\perp Temp \mid Starts$$

BAYES' NETS: CHAINS

Nodes can have "grandparents" (chains)

$$Late \perp Gas, Temp \mid Starts$$



IF the graph accurately represents the independence structure, nodes are **independent** of their grandparents given their immediate parents

INDEPENDENCE IN BAYES' NETS

IF the graph accurately represents the independence structure, we can **factor** the joint probability into a convenient form

$$p(X_1, X_2, \ldots, X_D) = \prod_{i=1}^{D} p(X_i \mid \textsc{Parents}(X_i))$$

Nodes are **conditionally independent** of their **ancestors** and **siblings** (non-descendents) given their **parents.**

X is conditionally independent of Y given Z iff:

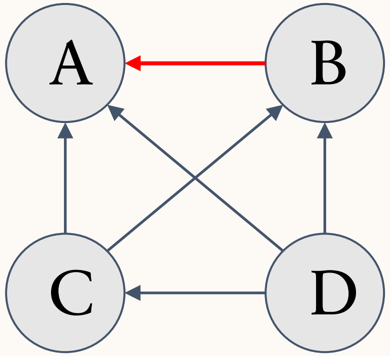
$\forall x, y, z:\ P(x, y|z) = P(x|z)P(y|z)$ or $\forall x, y, z:\ P(x|z, y) = P(x|z)$

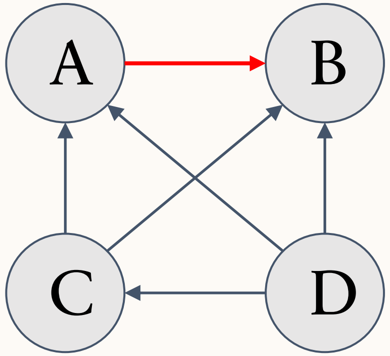
CAUSE AND EFFECT

If you know the cause and effect relationships for your problem, it's easy to build a Bayes Net.

Can you also **infer** cause and effect relationship **from** a graph? No!

$$p(A, B, C, D) = p(A \mid B, C, D)p(B \mid C, D)p(C \mid D)p(D)$$
$$= p(B \mid A, C, D)p(A \mid C, D)p(C \mid D)p(D)$$



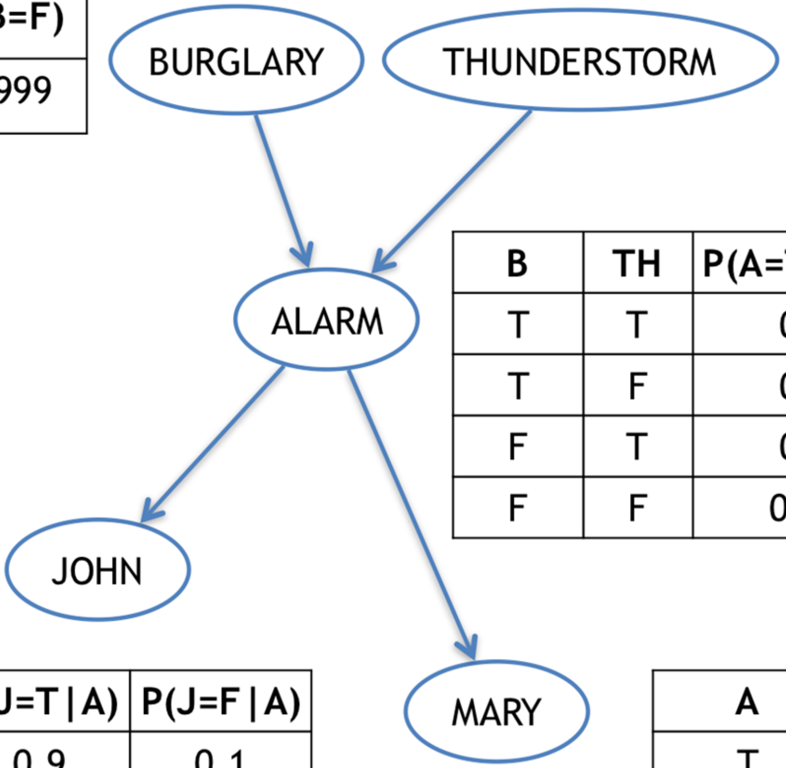


A graph for a given set of RVs is **not necessarily unique**.

USING BAYES' NETS - EXAMPLE

You are on vacation, and you've asked your neighbors (John & Mary) to keep an eye on your house while you are away. They'll call you if your house alarm goes off.

Your alarm system could be triggered because of an actual burglar or because a thunderstorm sets it off.



P(B=T)	P(B=F)
0.001	0.999

P(TH=T)	P(TH=F)
0.002	0.998

B	TH	P(A=T|B,TH)	P(A=F|B,TH)
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

A	P(J=T|A)	P(J=F|A)
T	0.9	0.1
F	0.05	0.95

A	P(M=T|A)	P(M=F|A)
T	0.7	0.3
F	0.01	0.99

USING BAYES' NETS - EXAMPLE

What's the probability that

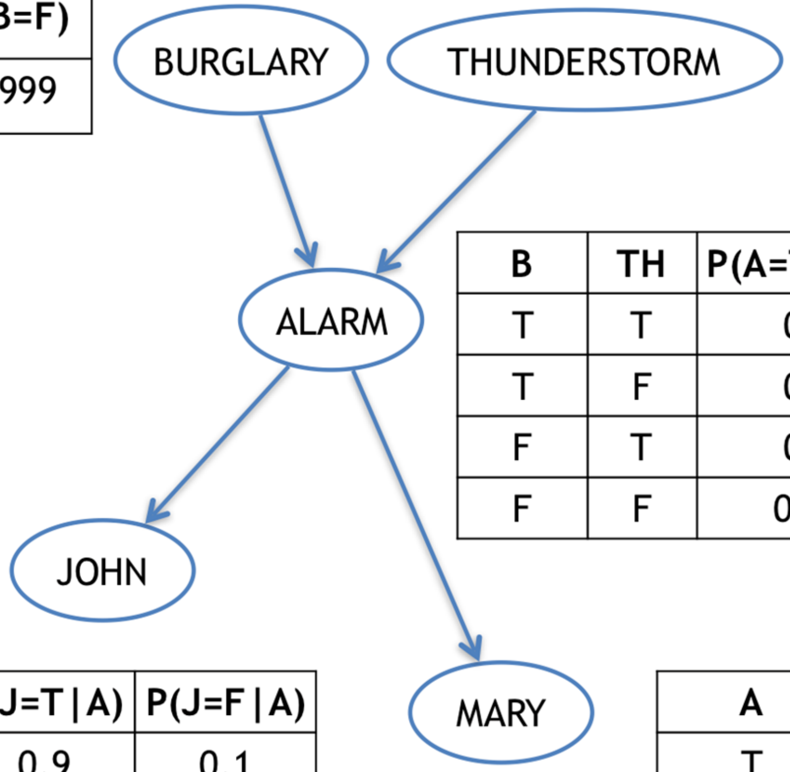
- Both neighbors call
- The alarm goes off
- There is no burglar
- There is no storm

p(j,m,a,¬b,¬t) =
p(j|a) p(m|a) p(a|¬b,¬t) p(¬b) p(¬t) =
(.9) (.7) (.001) (.999) (.998) = 0.00062

Joint probability table: 2^5=32 cells

CPT factorization: 20 cells

$$p(X_1, X_2, \ldots, X_D) = \prod_{i=1}^{D} p(X_i \mid \text{P}\textsc{arents}(X_i))$$



P(B=T)	P(B=F)
0.001	0.999

P(TH=T)	P(TH=F)
0.002	0.998

B	TH	P(A=T|B,TH)	P(A=F|B,TH)
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

A	P(J=T|A)	P(J=F|A)
T	0.9	0.1
F	0.05	0.95

A	P(M=T|A)	P(M=F|A)
T	0.7	0.3
F	0.01	0.99

YOUR TURN: USING BAYES' NETS

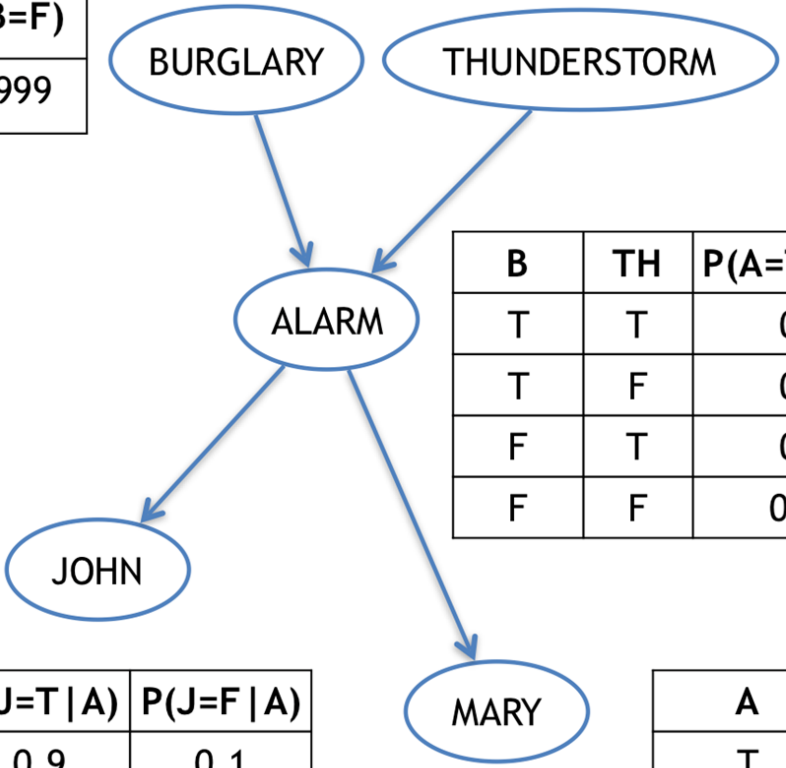
In general, there's a 4 step process to solve **any** query about a Bayes' Net:

1. Write the query as a statement about probabilities
2. Rewrite statement in terms of the joint probability distribution
3. Factor the joint probability using Bayes' Net independencies
4. Simplify, and plug in numbers from CPTs

$$p(X_1, X_2, \ldots, X_D) = \prod_{i=1}^{D} p(X_i \mid \text{PARENTS}(X_i))$$

Calculate the probability that there is a burglar if both John and Mary call.

P(B=T)	P(B=F)
0.001	0.999



P(TH=T)	P(TH=F)
0.002	0.998

B	TH	P(A=T|B,TH)	P(A=F|B,TH)
T	T	0.95	0.05
T	F	0.94	0.06
F	T	0.29	0.71
F	F	0.001	0.999

A	P(J=T|A)	P(J=F|A)
T	0.9	0.1
F	0.05	0.95

A	P(M=T|A)	P(M=F|A)
T	0.7	0.3
F	0.01	0.99

Steps 1 & 2: $p(b \mid j, m) = \frac{p(b, j, m)}{p(j, m)} = \alpha \cdot p(b, j, m)$

$$= \alpha \sum_{h_1} \sum_{h_2} p(b, Th = h_1, Al = h_2, j, m)$$

Do steps 3 & 4.

BAYES' NETS BIG PICTURE

- Two problems with using **full joint distribution tables** as our probabilistic models:
  - Unless there are only a few variables, the joint is *way* too big to represent explicitly
  - Hard to learn (estimate) anything empirically about more than a few variables at a time
- **Bayes' nets:** a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
  - More properly called graphical models
  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions