

Artificial Intelligence

CS 165A

Jan 17, 2019

T
o
d
a
y

→ Probabilistic Reasoning (Ch 14)

Announcements

- Course website:
<https://www.cs.ucsb.edu/~yuxiangw/classes/CS165A-2019winter/>
- Homework 1 is posted in the Assignment subdirectory,
 - **Due Jan 29.**
 - Homework submission box: HFH-2108 (look for a label with “Wang / 165A”)

Plan on “Reasoning with Uncertainty”

- Last time:
 - Probability theory
 - Joint probabilities, marginals, conditionals.
 - Independence
- Today
 - Conditional independences
 - Directed Graphical models --- Bayesian Networks (This lecture)
- Next Tuesday
 - D-separation / Markov Blankets
 - Undirected Graphical models --- Markov Random Field

Marginalization

- Given $P(X, Y)$, derive $P(X)$ (“marginalize away Y ”)

$$P(X) = \sum_Y P(X, Y) \quad \text{or} \quad P(X) = \int_Y P(X, Y)$$

- Equivalent: Given $P(X|Y)$ and $P(Y)$, derive $P(X)$

$$P(X) = \sum_Y P(X, Y) = \sum_Y P(X | Y)P(Y)$$

or

$$P(X) = \int_Y P(X, Y) = \int_Y P(X | Y)P(Y)$$

Marginalization (cont.)

- Marginalization is a common procedure
 - E.g., used to normalize Bayes' rule

$$P(H | D) = \frac{P(D | H) P(H)}{P(D)}$$

Known

?

- By *marginalization*,

$$\begin{aligned} P(D) &= \sum_H P(D, H) = \sum_H P(D | H) P(H) \\ &= P(D | H_1) P(H_1) + P(D | H_2) P(H_2) + \dots + P(D | H_N) P(H_N) \end{aligned}$$

Computing probabilities

- From the joint distribution $P(X_1, X_2, \dots, X_N)$ we can compute $P(\text{some variables})$ by summing over all the other variables
- For example
 - $P(X_2, \dots, X_N) = \sum_i P(X_{1i}, X_2, \dots, X_N)$
 - $P(X_1) = \sum_i \sum_j \dots \sum_p P(X_1, X_{2i}, X_{3j}, \dots, X_{Np})$
- For binary variables, from $P(X, Y)$
 - $P(Y) = \sum_i P(X_i, Y) = P(\neg X, Y) + P(X, Y)$
 - $P(X) = \sum_i P(X, Y_i) = P(X, \neg Y) + P(X, Y)$

Recap

- Goal: Probabilistic Inference
 - Calculate the conditional: **P(Event | Observed Evidence)**
- What do we need?
 - A model of the joint distribution
 - Read from the probability tables
 - Apply rules of probability theory:
 - ◆ Sum rule (Disjoint events)
 - ◆ Product rule (Independent Event)
 - ◆ Chain rule
 - ◆ Bayes rule

Independence

Absolute (or, marginal) Independence

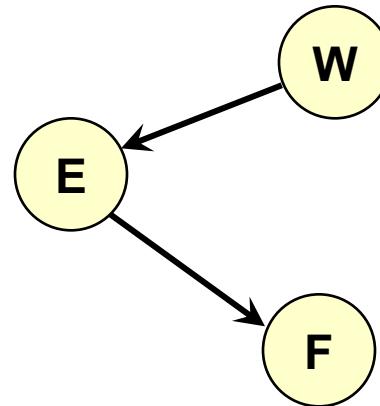
- X and Y are independent iff
 - $P(X, Y) = P(X) P(Y)$ [by definition]
 - $P(X | Y) = P(X)$ Since $P(X | Y) = P(X, Y)/P(Y) = P(X) P(Y)/P(Y)$

Conditional Independence

- If X and Y are (conditionally) independent given Z, then
 - $P(X | Y, Z) = P(X | Z)$
 - Example:
 - ◆ $P(\text{WetGrass} | \text{Season, Rain}) = P(\text{WetGrass} | \text{Rain})$

Conditional Independence

- In practice, conditional independence is more common than absolute independence.
 - $P(\text{Final exam grade} | \text{Weather}) \neq P(\text{Final exam grade})$
 - ◆ I.e., they are not independent
 - $P(\text{Final exam grade} | \text{Weather, Effort}) = P(\text{Final exam grade} | \text{Effort})$
 - ◆ But they are conditionally independent given Effort
- This leads to simplified rules for updating Bayes' Rule, and then onward to Bayesian networks

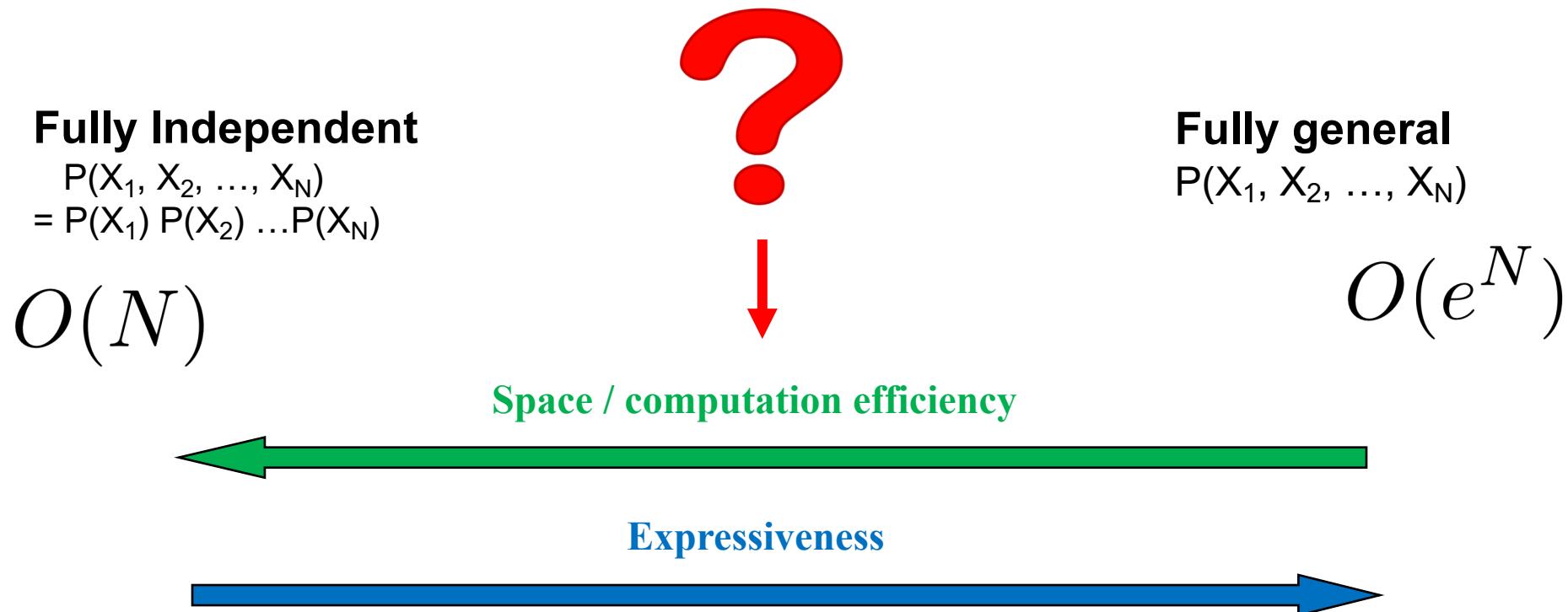


$$P(F | W, E) = P(F | E)$$

Quiz time: Representing a joint Probability

- Joint probability: $P(X_1, X_2, \dots, X_N)$
 - Defines the probability for any possible state of the world
 - Let the variables be binary. How many numbers does it take to define the joint distribution?
 - ◆ Defined by $2^N - 1$ independent numbers
- If the variables are independent, then
$$P(X_1, X_2, \dots, X_N) = P(X_1) P(X_2) \dots P(X_N)$$
 - How many numbers does it take to define the joint distribution?
 - ◆ Just N independent numbers!

Tradeoffs in our model choices



Idea:

1. Independent groups of variables?
2. Are all dependency created equal?

The Chain Rule again

$$\underline{P(X_1, X_2, \dots, X_N)} = \underline{P(X_1 | X_2, \dots, X_N)} P(X_2, \dots, X_N)$$

$$P(X_2, \dots, X_N) = \underline{P(X_2 | X_3, \dots, X_N)} P(X_3, \dots, X_N)$$

$$P(X_3, \dots, X_N) = \underline{P(X_3 | X_4, \dots, X_N)} P(X_4, \dots, X_N)$$

⋮

$$P(X_{N-1}, X_N) = \underline{P(X_{N-1} | X_N)} P(X_N)$$

$$\underline{P(X_1, X_2, \dots, X_N)} = \underline{P(X_1 | X_2, \dots, X_N)} \underline{P(X_2 | X_3, \dots, X_N)} \underline{P(X_3 | X_4, \dots, X_N)} \cdots \underline{P(X_{N-1} | X_N)} \underline{P(X_N)}$$

The Chain Rule again (cont.)

- Recursive definition:

$$P(X_1, X_2, \dots, X_N) = P(X_1 | X_2, \dots, X_N) P(X_2 | X_3, \dots, X_N) \dots P(X_{N-1} | X_N) P(X_N)$$

$$\uparrow \quad \text{or equivalently}$$

$$2^N - 1 = P(X_1) P(X_2)$$

How many values needed to represent this?
(assuming binary variables)

$$2^N - 1 = 1 + 2 + 4 + \dots + 2^{N-1}$$

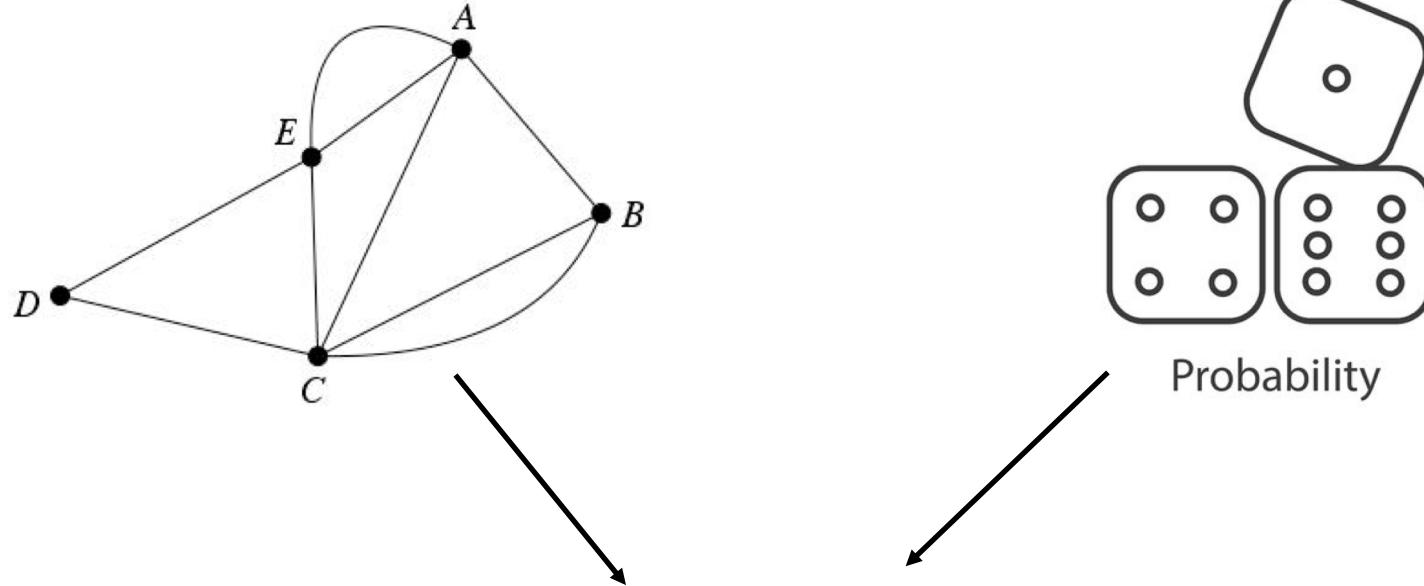
Note on number of independent values....

- Random variables W, X, Y, and Z
 - $W = \{ w_1, w_2, w_3, w_4 \}$
 - $X = \{ x_1, x_2 \}$
 - $Y = \{ y_1, y_2, y_3 \}$
 - $Z = \{ z_1, z_2, z_3, z_4 \}$
- How many (independent) numbers are needed to describe:
 - $P(W)$ $4 - 1 = 3$
 - $P(X, Y)$ $2^*3 - 1 = 5$
 - $P(W, X, Y, Z)$ $4^*2^*3^*4 - 1 = 95$
 - $P(X | Y)$ $(2-1)^*3 = 3$
 - $P(W | X, Y, Z)$ $(4-1)^*(2^*3^*4) = 72$

Benefit of conditional independence

- If some variables are conditionally independent, the joint probability can be specified with many fewer than 2^N-1 numbers (or 3^N-1 , or 10^N-1 , or...)
- For example: (for binary variables W, X, Y, Z)
 - $P(W,X,Y,Z) = P(W) P(X|W) P(Y|W,X) P(Z|W,X,Y)$
 - ◆ $1 + 2 + 4 + 8 = 15$ numbers to specify
 - But if Y and W are independent given X, and Z is independent of W and X given Y, then
 - ◆ $P(W,X,Y,Z) = P(W) P(X|W) P(Y|X) P(Z|Y)$
 - $1 + 2 + 2 + 2 = 7$ numbers
- This is often the case in real problems.

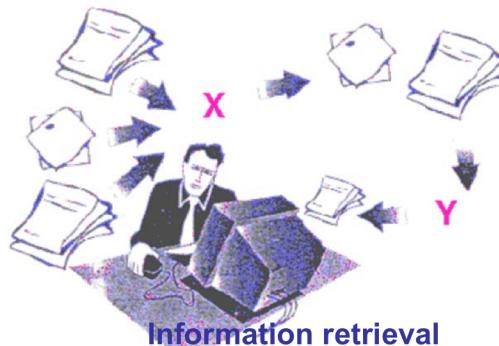
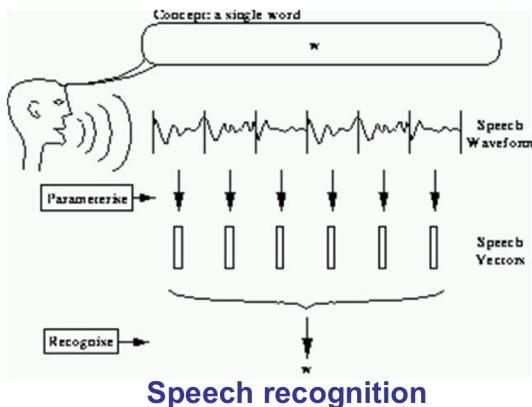
Graphical models



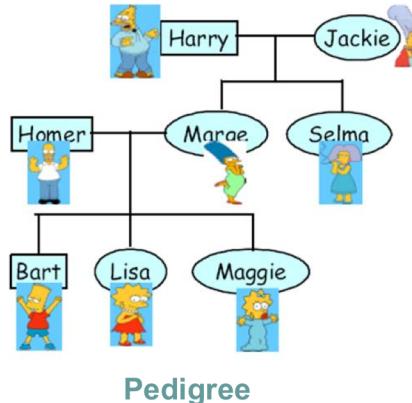
**Bayesian Networks / Belief Networks
Markov Random Fields**

$$\mathcal{D} \equiv \{X_1^{(i)}, X_2^{(i)}, \dots, X_m^{(i)}\}_{i=1}^N$$

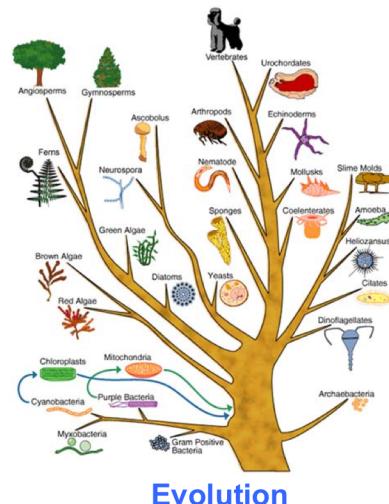
Many applications!



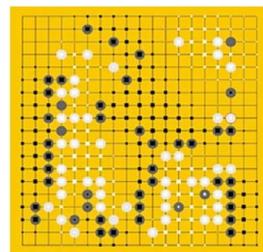
Information retrieval



Pedigree



Evolution



Games



Planning



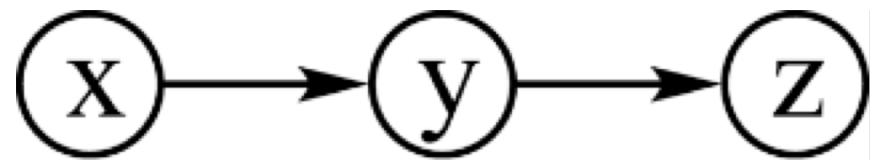
Computer vision



Robotic control

Two ways to think about Graphical Models

- A particular factorization of a joint distribution
 - $P(X,Y,Z) = P(X) P(Y|X) P(Z|Y)$
- A collection of conditional independences
 - $\{ X \perp Y | Z, \dots \}$



Represented using a graph!

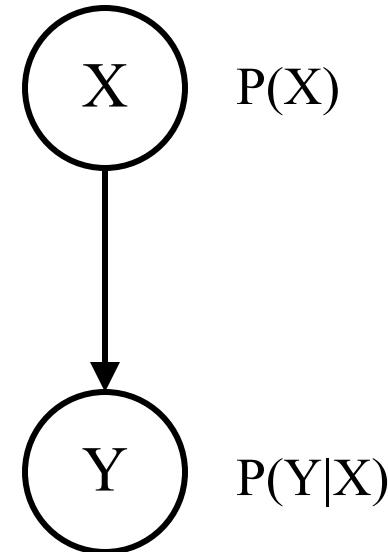
Belief Networks (a.k.a. Bayesian Networks)

a.k.a. Probabilistic networks, Belief nets, Bayes nets, etc.

- Belief network
 - A data structure (depicted as a graph) that represents the dependence among variables and allows us to concisely specify the joint probability distribution
 - The graph itself is known as an “influence diagram”
- A belief network is a **directed acyclic graph** where:
 - The nodes represent the set of random variables (one node per random variable)
 - Arcs between nodes represent *influence*, or *causality*
 - ◆ A link from node X to node Y means that X “directly influences” Y
 - Each node has a **conditional probability table** (CPT) that defines $P(\text{node} \mid \text{parents})$

Example

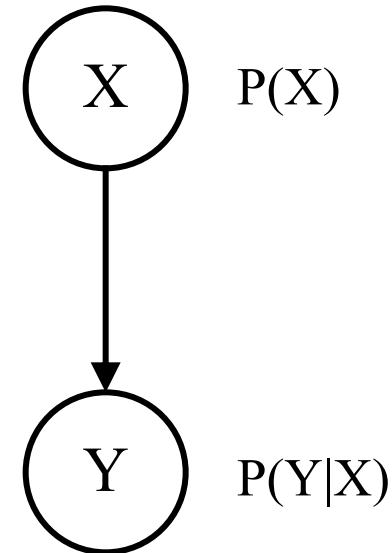
- Random variables X and Y
 - X – It is raining
 - Y – The grass is wet
- X has a *causal effect* on Y
Or, Y is a *symptom* of X
- Draw two nodes and link them
- Define the CPT for each node
 - $P(X)$ and $P(Y | X)$
- Typical use: we observe Y and we want to query $P(X | Y)$
 - Y is an *evidence variable*
 - X is a *query variable*



Try it

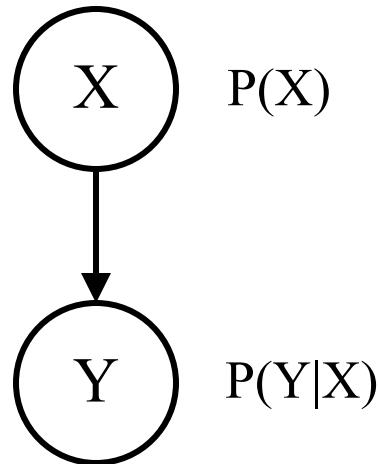
- What is $P(X | Y)$?
 - Given that we know the CPTs of each node in the graph

$$\begin{aligned} P(X | Y) &= \frac{P(Y | X)P(X)}{P(Y)} \\ &= \frac{P(Y | X)P(X)}{\sum_X P(X, Y)} \\ &= \frac{P(Y | X)P(X)}{\sum_X P(Y | X)P(X)} \end{aligned}$$

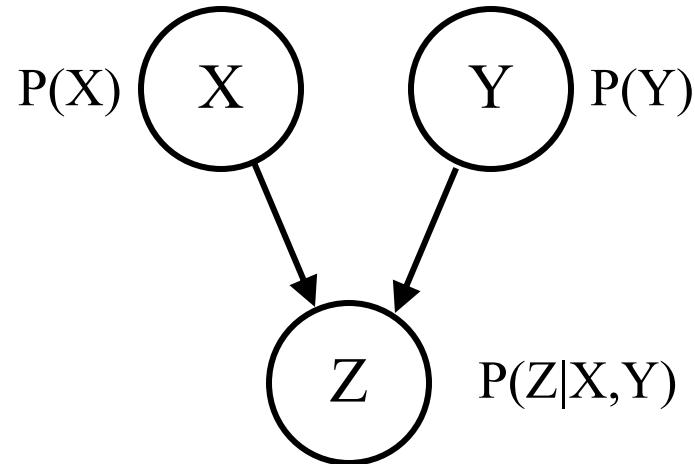


Belief nets represent the joint probability

- The joint probability function can be calculated directly from the network
 - It's the product of the CPTs of all the nodes
 - $P(\text{var}_1, \dots, \text{var}_N) = \prod_i P(\text{var}_i | \text{Parents}(\text{var}_i))$



$$\mathbf{P(X,Y)} = \mathbf{P(X) P(Y|X)}$$



$$\mathbf{P(X,Y,Z)} = \mathbf{P(X) P(Y) P(Z|X,Y)}$$

Example

I'm at work and my neighbor John called to say my home alarm is ringing, but my neighbor Mary didn't call. The alarm is sometimes triggered by minor earthquakes. Was there a burglar at my house?

- Random (boolean) variables:
 - JohnCalls, MaryCalls, Earthquake, Burglar, Alarm
- The belief net shows the causal links
- This defines the joint probability
 - $P(\text{JohnCalls}, \text{MaryCalls}, \text{Earthquake}, \text{Burglar}, \text{Alarm})$
- What do we want to know? **P(B | J, $\neg M$)**

Why not **P(B | J, A, $\neg M$)** ?

Example

Burglary

Earthquake

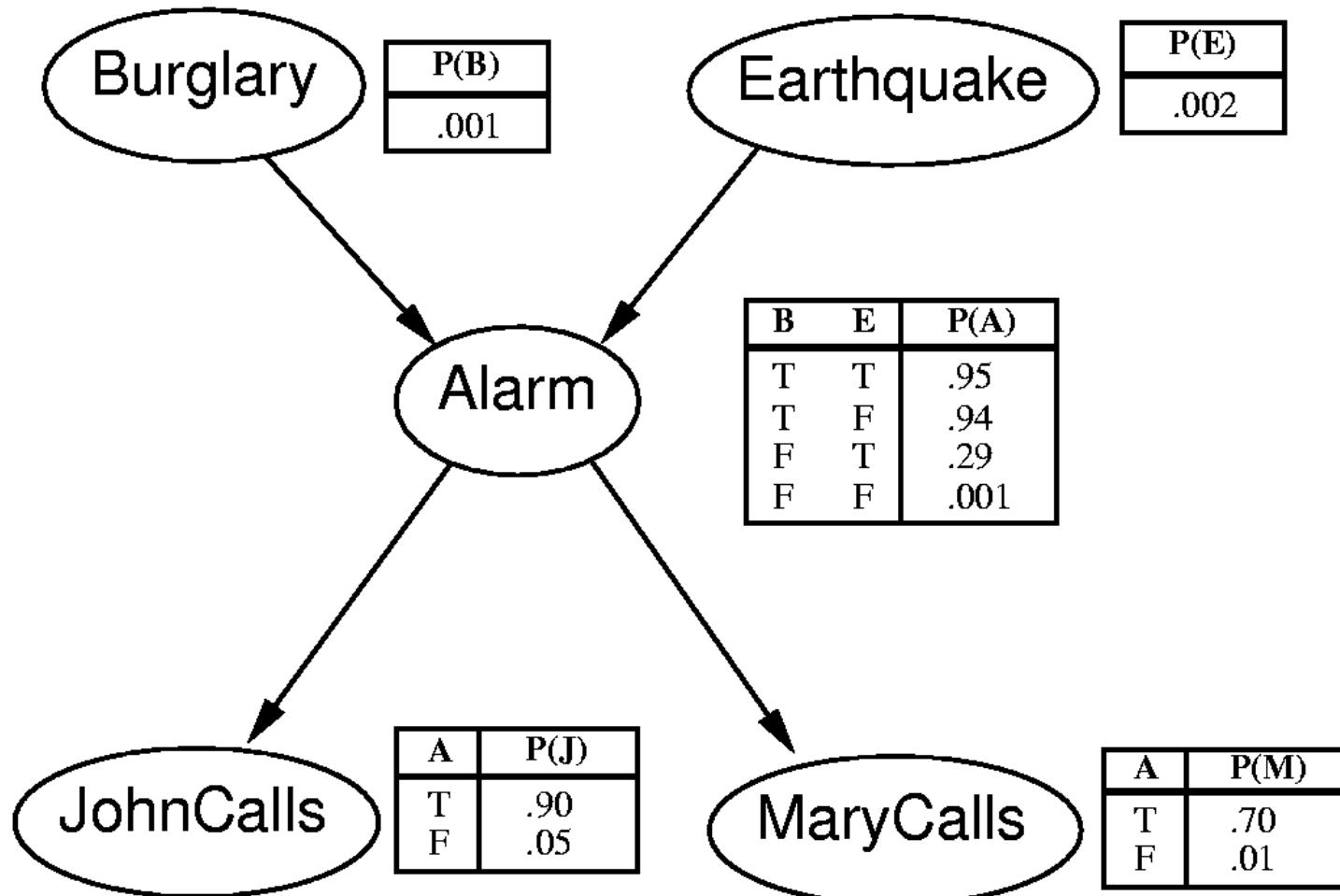
Alarm

JohnCalls

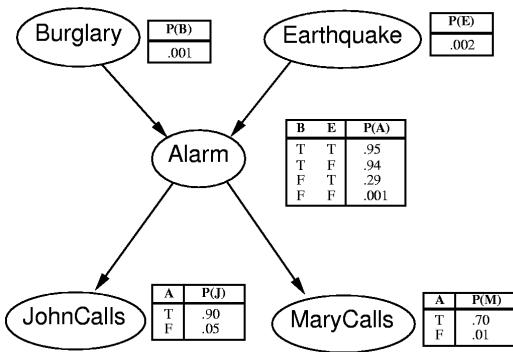
MaryCalls

Links and CPTs?

Example



Joint probability? $P(J, \neg M, A, B, \neg E)$?



Calculate $P(J, \neg M, A, B, \neg E)$

Read the joint pf from the graph:

$$P(J, M, A, B, E) = P(B) P(E) P(A|B,E) P(J|A) P(M|A)$$

Plug in the desired values:

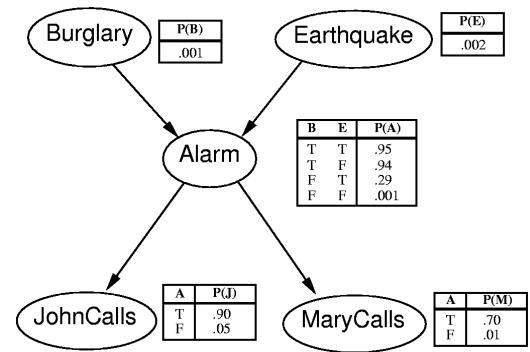
$$\begin{aligned}
 P(J, \neg M, A, B, \neg E) &= P(B) P(\neg E) P(A|B, \neg E) P(J|A) P(\neg M|A) \\
 &= 0.001 * 0.998 * 0.94 * 0.9 * 0.3 \\
 &= 0.0002532924
 \end{aligned}$$

How about $P(B | J, \neg M)$?

Remember, this means $P(B=\text{true} | J=\text{true}, M=\text{false})$

Calculate $P(B \mid J, \neg M)$

$$P(B \mid J, \neg M) = \frac{P(B, J, \neg M)}{P(J, \neg M)}$$



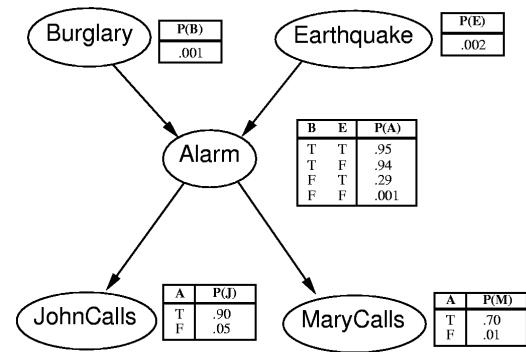
By marginalization:

$$= \frac{\sum_i \sum_j P(J, \neg M, A_i, B, E_j)}{\sum_i \sum_j \sum_k P(J, \neg M, A_i, B_j, E_k)}$$

$$= \frac{\sum_i \sum_j P(B)P(E_j)P(A_i \mid B, E_j)P(J \mid A_i)P(\neg M \mid A_i)}{\sum_i \sum_j \sum_k P(B_j)P(E_k)P(A_i \mid B_j, E_k)P(J \mid A_i)P(\neg M \mid A_i)}$$

Example

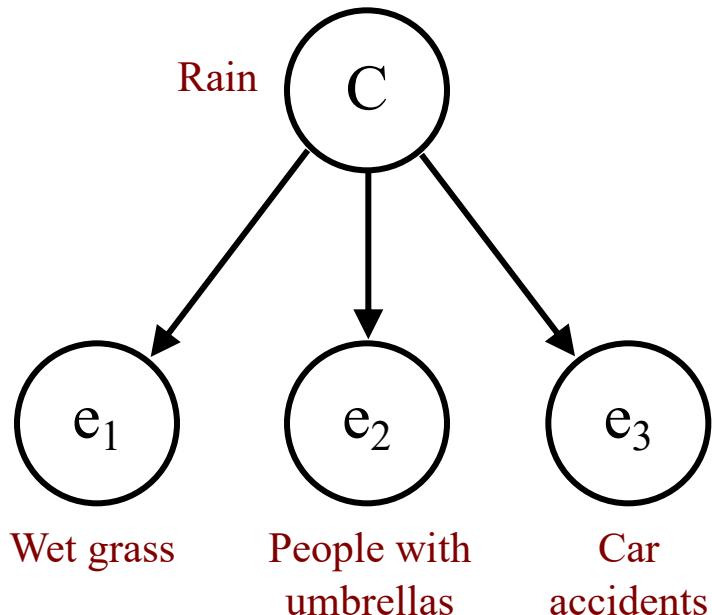
- Conditional independence is seen here
 - $P(\text{JohnCalls} | \text{MaryCalls}, \text{Alarm}, \text{Earthquake}, \text{Burglary}) = P(\text{JohnCalls} | \text{Alarm})$
 - So JohnCalls is independent of MaryCalls, Earthquake, and Burglary, given Alarm
- Does this mean that an earthquake or a burglary do not influence whether or not John calls?
 - No, but the influence is already accounted for in the Alarm variable
 - JohnCalls is conditionally independent of Earthquake, but not absolutely independent of it



Naive Bayes model

- A common situation is when a single cause directly influences several variables, which are all conditionally independent, given the cause.

$$P(C, e_1, e_2, e_3) = P(C) P(e_1 | C) P(e_2 | C) P(e_3 | C)$$

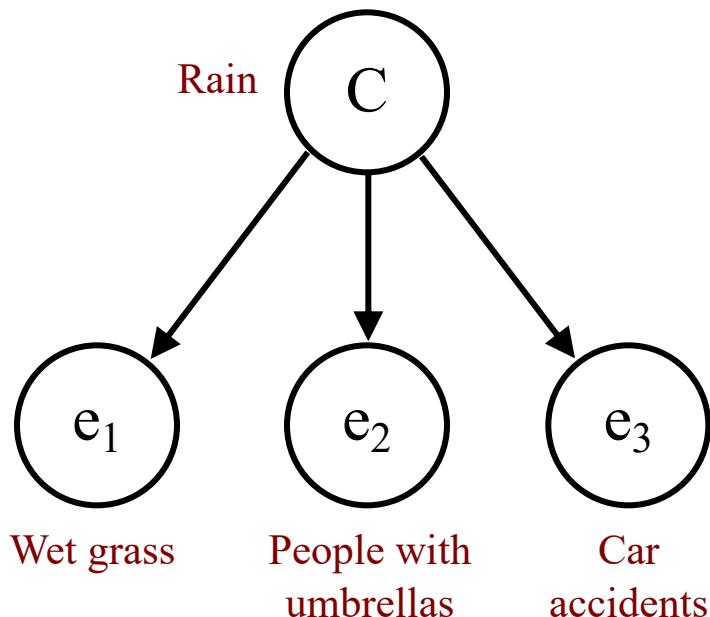


In general,

$$P(C, e_1, \dots, e_n) = P(C) \prod_i P(e_i | C)$$

Naive Bayes model

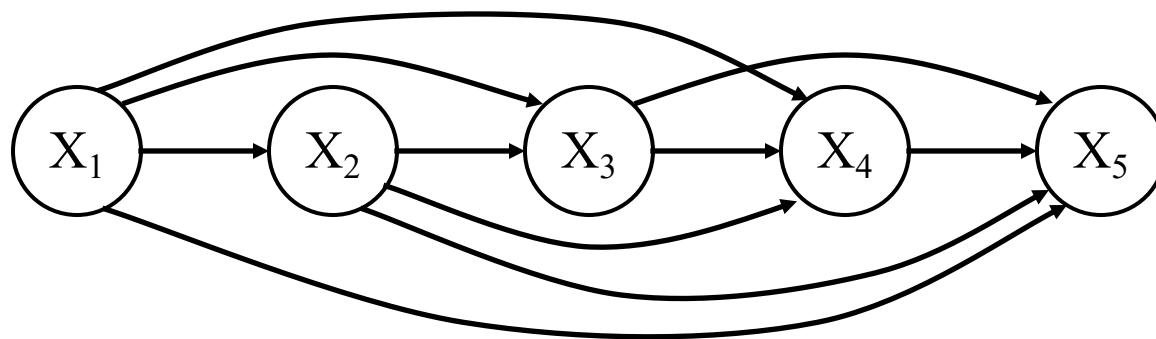
- Typical query for naive Bayes:
 - Given some evidence, what's the probability of the cause?
 - $P(C | e_1) = ?$
 - $P(C | e_1, e_3) = ?$



$$P(C | e_1) = \frac{P(e_1 | C)P(C)}{P(e_1)}$$
$$= \frac{P(e_1 | C)P(C)}{\sum_C P(e_1 | C)P(C)}$$

Drawing belief nets

- What would a belief net look like if all the variables were fully dependent?



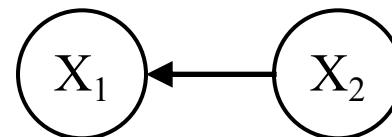
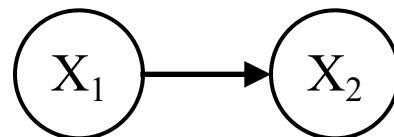
$$P(X_1, X_2, X_3, X_4, X_5) = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)P(X_4|X_1, X_2, X_3)P(X_5|X_1, X_2, X_3, X_4)$$

- But this isn't the only way to draw the belief net when all the variables are fully dependent

Fully connected belief net

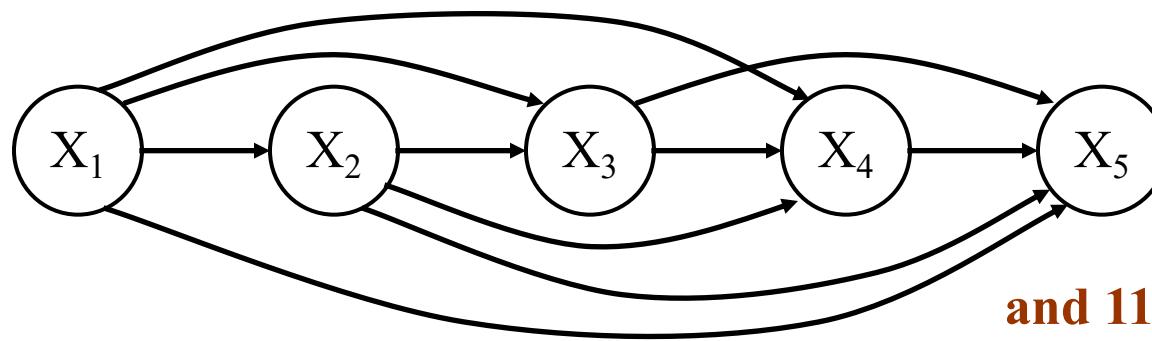
- In fact, there are $N!$ ways of connecting up a fully-connected belief net
 - That is, there are $N!$ ways of ordering the nodes

For $N=2$



$$P(X_1, X_2) = ?$$

For $N=5$



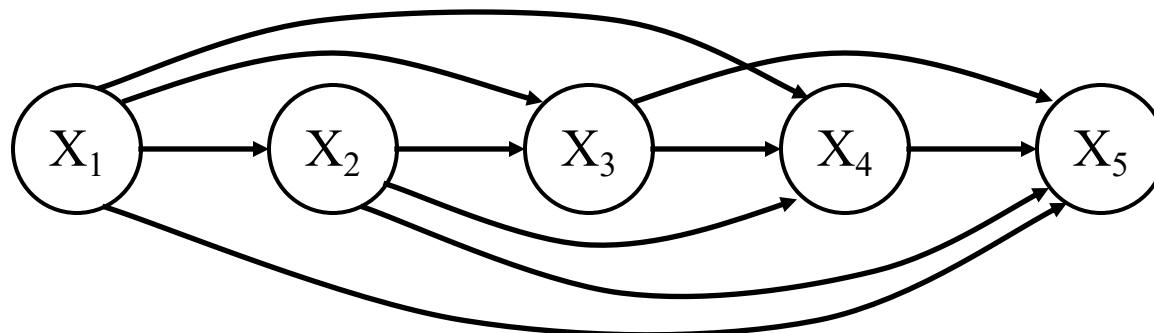
$$P(X_1, X_2, X_3, X_4, X_5) = ?$$

and 119 others...

Drawing belief nets (cont.)

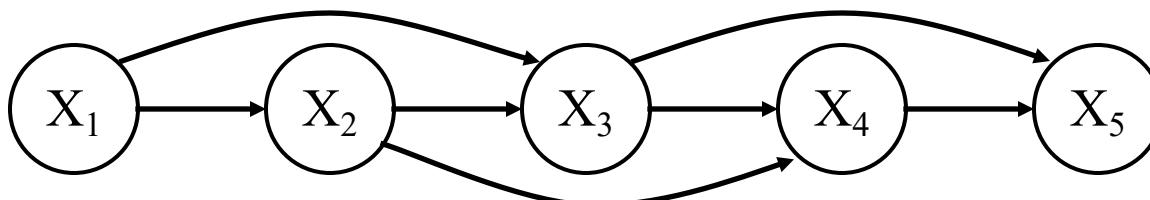
Fully-connected net displays the joint distribution

$$P(X_1, X_2, X_3, X_4, X_5) = P(X_1) P(X_2|X_1) P(X_3|X_1, X_2) P(X_4|X_1, X_2, X_3) P(X_5|X_1, X_2, X_3, X_4)$$



But what if there are conditionally independent variables?

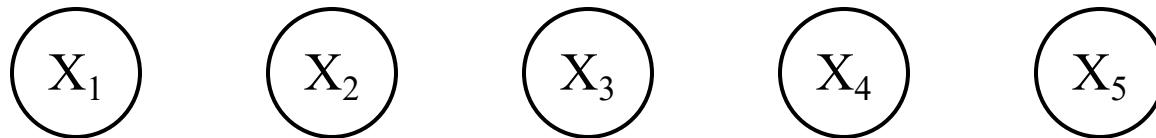
$$P(X_1, X_2, X_3, X_4, X_5) = P(X_1) P(X_2|X_1) P(X_3|X_1, X_2) P(X_4|X_2, X_3) P(X_5|X_3, X_4)$$



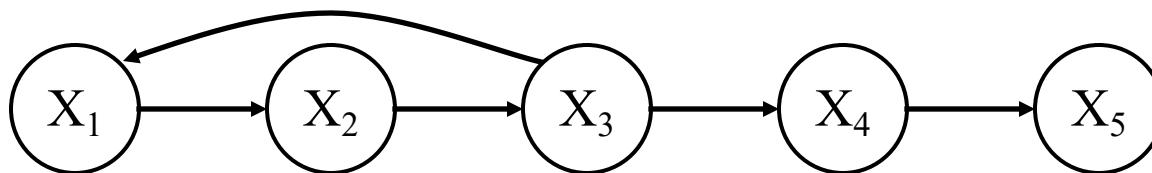
Drawing belief nets (cont.)

What if the variables are all independent?

$$P(X_1, X_2, X_3, X_4, X_5) = P(X_1) P(X_2) P(X_3) P(X_4) P(X_5)$$



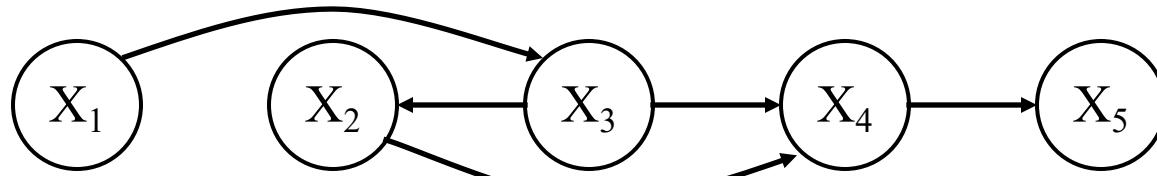
What if the links are drawn like this:



Not allowed – not a DAG

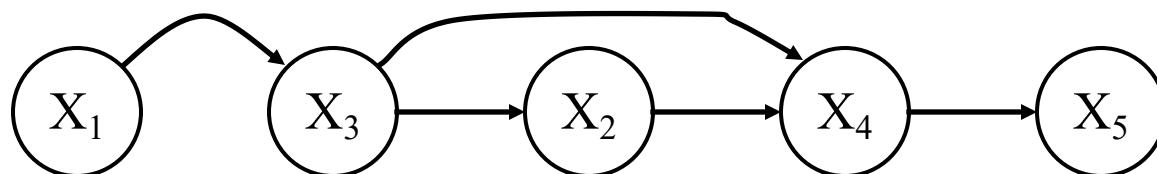
Drawing belief nets (cont.)

What if the links are drawn like this:



$$P(X_1, X_2, X_3, X_4, X_5) = P(X_1) P(X_2 | X_3) P(X_3 | X_1) P(X_4 | X_2, X_3) P(X_5 | X_4)$$

It can be redrawn like this:



All arrows going left-to-right