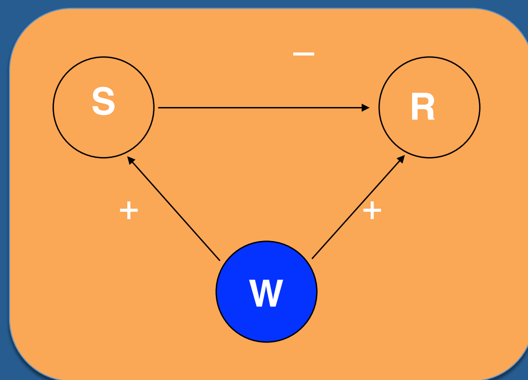


BAYESIAN NETWORKS

EECS 492
March 10, 2011

Today's Plan

- Last time, we saw how inference is "easy" when we have the full joint distribution
  - Can easily extract any marginal or conditional
- But the full joint distribution becomes very large
  - Do we really need to write the joint distribution for N fair coin flips?
- How do we exploit structure of problems in order to reduce memory complexity?

The problem with full joint distributions

- Joint distributions often contain *redundant* information
  - If x, y and z are independent, Pr(x,y,z)=Pr(x)Pr(y)Pr(z)
    - Pr(x,y,z) requires 8 quantities
      - (well, 7 since it must sum to 1)
    - Pr(x), Pr(y), Pr(z) requires 3 quantities
    - → *Independence yields space savings!*
  - If we noticed this factorization, we could compute any joint probabilities we wanted, on demand.
    - Don't need a complete table!
- Many problems have independent or conditionally independent random variables

How to avoid writing joint distributions: The Chain Rule

- The product rule gives us:
  - P(A , B) = P(A | B) P(B)
- We can apply this to more complicated situations:
  - P(A, B, C) = P(A | B, C) P(B, C)
  - And apply it again:
  - P(A, B, C) = P(A | B, C) P(B | C) P (C)
- Any joint probability can be expanded into products of this form.

What does this buy us?

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

- How many numbers to specify P(A, B, C, D) ?
  - 15
- How many numbers to specify P(A | B, C, D) ?
  - 8
- How many numbers to specify P(B | C, D) ?
  - 4
- How many numbers to specify P(C | D) ?
  - 2
- How many numbers to specify P(D) ?
  - 1
- What did this buy us? Nothing! 15 = 8 + 4 + 2 + 1

How to avoid writing joint distributions: Exploit Independence

$$\begin{array}{ccccccccc} P(A, B, C, D) & = & P(A \mid B, C, D) & & P(B \mid C, D) & & P(C \mid D) & & P(D) \\ 15 & = & 8 & + & 4 & + & 2 & + & 1 \end{array}$$

- A patient might have a **Cold** or a **Disease**. Both Colds and Diseases can cause **Bronchitis**, but only a Cold makes you **Achy.**
- Suppose C and D are independent
  - C: Patient has a *Cold*
  - D: Patient has a *Disease*
  - P(C | D) = P(C)

$$\begin{array}{cccccccc} P(A \mid B, C, D) & & P(B \mid C, D) & & P(C) & & P(D) & \\ 8 & + & 4 & + & 1 & + & 1 & = 14 \end{array}$$

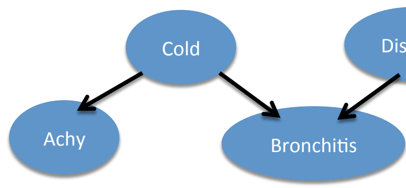
- Suppose A and B are *conditionally* independent given C
  - A: Patient symptom includes *Achy*
  - B: Patient symptom includes *Bronchitis*
  - P(A | B, C, D) = P(A | C, D)

$$\begin{array}{cccccccc} P(A \mid C, D) & & P(B \mid C, D) & & P(C) & & P(D) & \\ 4 & + & 4 & + & 1 & + & 1 & = 10 \end{array}$$

Bayesian Networks

- We can graphically denote the independence of variables:

  A patient might have a ***Cold*** *or a* ***Disease****. Both Colds and Diseases can cause* ***Bronchitis****, but only a Cold makes you* ***Achy.***



- This *Graphical Model* directly encodes the relationship between joint and conditional distributions!
  - The graph is not merely a cartoon– it has specific, exploitable meanings.
  - Joint = product of conditional for each variable

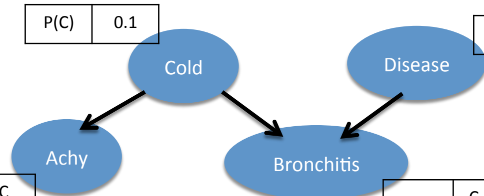
P(A, B, C, D) = P(C) P(D) P(A | C) P(B | C, D)

Bayesian Networks and CPTs

- Each node represents a *conditional probability* and has associated with it, a CPT:

P(C)	0.1

P(D)	0.01



	C	-C
P(A)	.8	.1

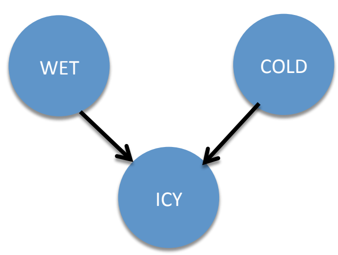
	C ^ D	C ^ -D	-C ^ D	-C ^ -D
P(B)	.9	.7	.8	.1

P(A, B, C, D) = P(C) P(D) P(A | C) P(B | C, D)

Reasoning about dependence

- Clearly LEAVE depends on OVERSLEEP. ARRIVE depends on LEAVE.
  - Intuitive causal relationships.
- Q: Does LEAVE depend on ARRIVE?
  - Yes: changing the order doesn't change independence
  - P(A,B) != P(A) P(B) ➔ P(B,A) != P(A)P(B)
- Q: Does ARRIVE depend on OVERSLEEP?
- Q: Are ARRIVE and OVERSLEEP conditionally independent given LEAVE?

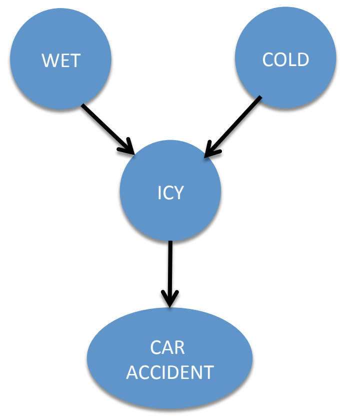
Reasoning about dependence



- Clearly, Icy is dependent on wet and cold.
- Q: Is wet dependent on icy?
- Q: Are wet and cold independent?
- Q: Are wet and cold *conditionally* independent given icy?

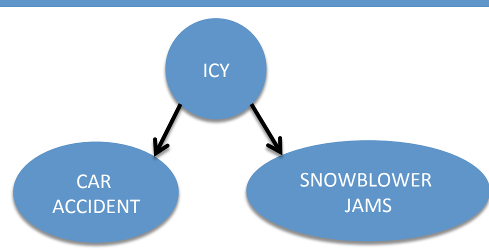
Reasoning about dependence

- From before:
  - WET and COLD are conditionally *dependent* given ICY.



- Q: Are WET and COLD conditionally dependent given ACCDIENT?

Reasoning about dependence



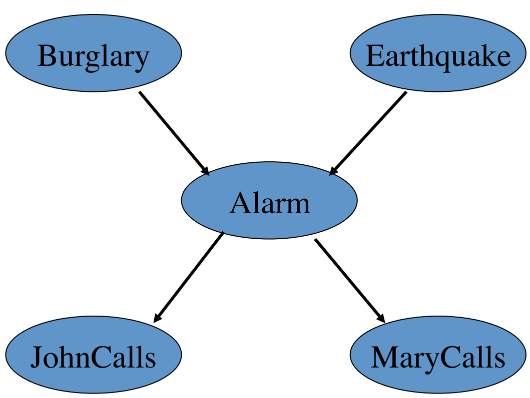
- Q: Are ACCIDENT and JAMS independent?
- Q: Are ACCIDENT and JAMS *conditionally* independent given ICY?

Dependence: Your turn!

- If X and Y are *unconditionally* dependent, which of the following are true?
  - A) X causes Y
  - B) Y causes X
  - C) Knowing something about X changes your belief about Y
  - D) X is Y's parent, or Y is X's parent.
  - E) X and Y share some common ancestor in the Bayes net

- If X and Y are *conditionally* dependent given Z, which of the following are true?
  - A) X and Y might also be *unconditionally* dependent.
  - B) Z causes X and Y
  - C) If X and Y are not unconditionally dependent, then Z is a common descendent of X and Y
  - D) I couldn't think of a fourth option.

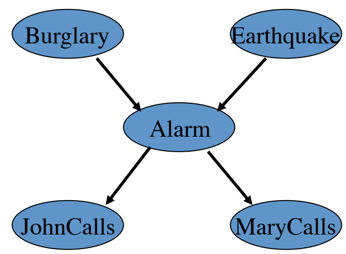
The Earthquake Network



- What is the joint distribution?
  - P(BEAJM) = P(B) P(E) P(A | BE) P(J | A) P (M | A)

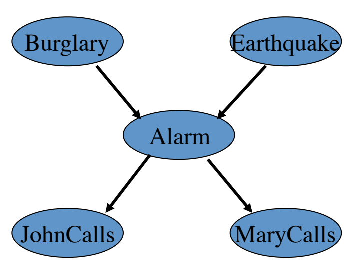
Elementary Inference

- How do I compute an arbitrary distribution?
  - For example, P(B, A | E) ?
- Method 1: Try to cleverly manipulate query to express in terms of known conditional probabilities.
  - Like we did with the diseases example
- Method 2: Methodically!

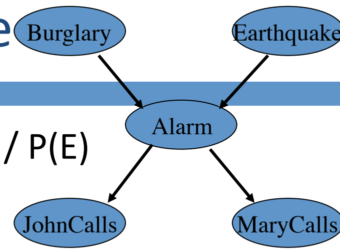


Elementary Inference

- Every distribution can be written in terms of the joint distribution and marginalizations of the joint distribution. *Our example: P(BA | E)*
- $P(BA \mid E) = P(BAE) / P(E)$
- Both P(BAE) and P(E) are marginalized versions of the joint distribution.



Elementary Inference



- Our query: P(BA | E) = P(BAE) / P(E)
- P(BA | E) = P(BAE) / P(E)

- Bayes net tells us:

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

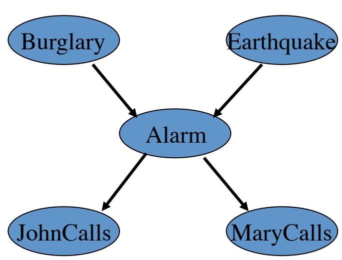
$$P(B,A,E) = \sum_M \sum_J P(B)P(E)P(A|B,E)P(J|A)P(M|A)$$

$$P(E) = \sum_B \sum_A \sum_M \sum_J P(B)P(E)P(A|B,E)P(J|A)P(M|A)$$

Tedious, but straight forward!

Restructuring a Graphical Model

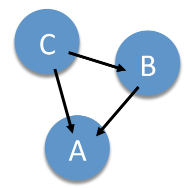
- Are there any other graphical models that are inferentially equivalent to this one?



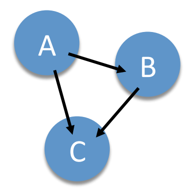
- Easy answer: we can always add edges
  - Resulting CPTs will be wasteful, but it's inferentially equivalent.
  - Adding edges == ignoring the structure in the problem

Reordering a Graphical Model

- Both of these models are inferentially equivalent:
  - P(A,B,C) = P(A | BC) P(B | C) P(C)
  - P(A,B,C) = P(C | BA) P (B | A) P(A)



- The hard part is determining which edges can be left out!

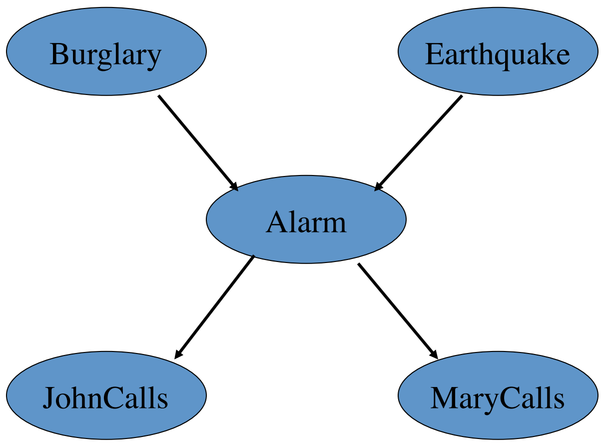


Reordering a Graphical Model

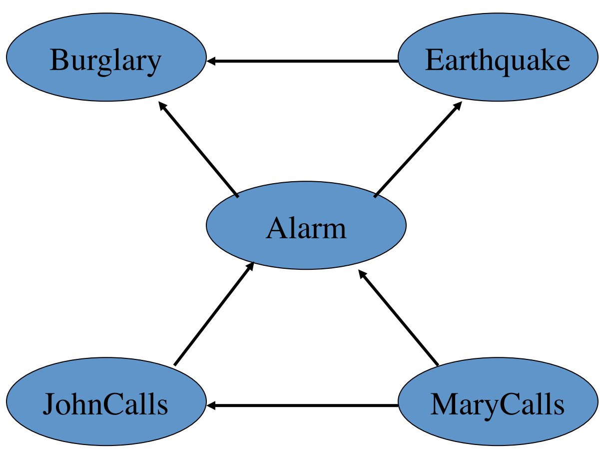
- Algorithm:
- Input: Graph g, with N variables
- Output: Graph g', with new variable order V

- for i = 1 : N
  - Add a node for variable V_i to the graph g'
    - Initially, no edges.
  - for j = 1 to i
    - Consider the original graph g. Suppose that we have observed all the variables in g', except V_i and V_j. If V_i and V_j are conditionally dependent, add an edge from V_i to V_j in g'

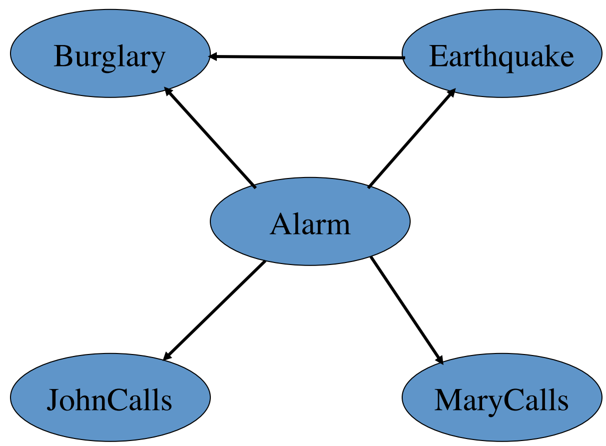
Node Order: BEAJM



Node Order: MJAEB



Your turn: AMEJB



Your turn: again!

- Which ordering will produce the most edges?
  - Try to reason about it without actually constructing the graphs by trial and error.
  - Easier: What node should be last?
  - Harder: Which nodes should be first?

Reflection

- Changing node order changes the number of edges
- In general, causal relationships lead to the smallest number of edges