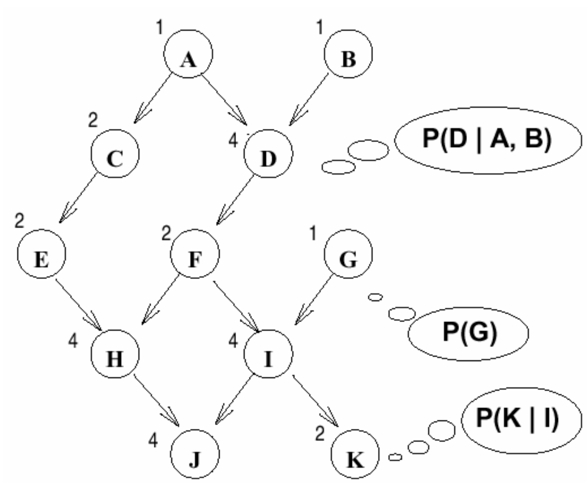
CSC384: Intro to Artificial Intelligence
Reasoning under Uncertainty-III

- Announcements:
  - Drop deadline is this Sunday Nov 5th.
  - All lecture notes needed for T3 posted (L13,…,L17).
  - T3 sample questions posted.
  - A3 posted.

Bayesian Networks

- A BN over variables $\{X_1, X_2, \ldots, X_n\}$ consists of:
  - a DAG (directed acyclic graph) whose nodes are the variables
  - a set of CPTs (conditional probability tables) $Pr(X_i \mid Par(X_i))$ for each X_i
- Key notions (see text for defn's, all are intuitive):
  - parents of a node: $Par(X_i)$
  - children of node
  - descendents of a node
  - ancestors of a node
  - family: set of nodes consisting of X_i and its parents
    - CPTs are defined over families in the BN

Example (Binary valued Variables)



- A couple CPTS are "shown"
- Explicit joint requires 2^{11} -1 =2047 parmtrs
- BN requires only 27 parmtrs (the number of entries for each CPT is listed)

Semantics of Bayes Nets.

- A Bayes net specifies that the joint distribution over the variable in the net can be written as the following product decomposition.
- $Pr(X_1, X_2, \ldots, X_n)$
  $= Pr(X_n \mid Par(X_n)) * Pr(X_{n-1} \mid Par(X_{n-1})) * \ldots * Pr(X_1 \mid Par(X_1))$
- This equation hold for any set of values $d_1, d_2, \ldots, d_n$ for the variables $X_1, X_2, \ldots, X_n$.

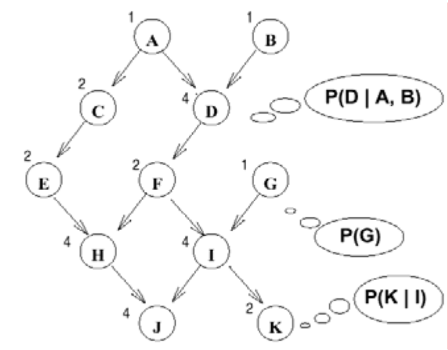
Semantics of Bayes Nets.

- E.g., say we have X_1, X_2, X_3 each with domain $Dom[X_i] = \{a, b, c\}$ and we have
  $Pr(X_1,X_2,X_3)$
  $= P(X_3|X_2)\ P(X_2)\ P(X_1)$

  Then
  $Pr(X_1=a,X_2=a,X_3=a)$
  $= P(X_3=a|X_2=a)\ P(X_2=a)\ P(X_1=a)$
  $Pr(X_1=a,X_2=a,X_3=b)$
  $= P(X_3=b|X_2=a)\ P(X_2=a)\ P(X_1=a)$
  $Pr(X_1=a,X_2=a,X_3=c)$
  $= P(X_3=c|X_2=a)\ P(X_2=a)\ P(X_1=a)$
  $Pr(X_1=a,X_2=b,X_3=a)$
  $= P(X_3=a|X_2=b)\ P(X_2=b)\ P(X_1=a)$
  …

Example (Binary valued Variables)

$Pr(a,b,c,d,e,f,g,h,i,j,k) =$
$Pr(a)$
$\times\ Pr(b)$
$\times\ Pr(c|a)$
$\times\ Pr(d|a,b)$
$\times\ Pr(e|c)$
$\times\ Pr(f|d)$
$\times\ Pr(g)$
$\times\ Pr(h|e,f)$
$\times\ Pr(i|f,g)$
$\times\ Pr(j|h,i)$
$\times\ Pr(k|i)$



Semantics of Bayes Nets.

- Note that this means we can compute the probability of any setting of the variables using only the information contained in the CPTs of the network.

Constructing a Bayes Net

- It is always possible to construct a Bayes net to represent any distribution over the variables $X_1, X_2,\ldots, X_n$, using any ordering of the variables.
  - Take any ordering of the variables (say, the order given). From the chain rule we obtain.

    $$Pr(X_1,\ldots,X_n) = Pr(X_n|X_1,\ldots,X_{n-1})Pr(X_{n-1}|X_1,\ldots,X_{n-2})\ldots Pr(X_1)$$

  - Now for each Xi go through its conditioning set $X_1,\ldots,X_{i-1}$, and iteratively remove all variables X_j such that X_i is conditionally independent of X_j given the remaining variables. Do this until no more variables can be removed.
  - The final product will specify a Bayes net.

Constructing a Bayes Net

- The end result will be a product decomposition/Bayes net
  $Pr(X_n \mid Par(X_n))\ Pr(X_{n-1} \mid Par(X_{n-1}))\ldots\ Pr(X_1)$

- Now we specify the numeric values associated with each term $Pr(X_i \mid Par(X_i))$ in a CPT.
- Typically we represent the CPT as a table mapping each setting of $\{X_i, Par(X_i)\}$ to the probability of X_i taking that particular value given that the variables in $Par(X_i)$ have their specified values.
- If each variable has d different values.
  - We will need a table of size $d^{|\{X_i, Par(X_i)\}|}$.
  - That is, exponential in the size of the parent set.

- Note that the original chain rule $Pr(X_1,\ldots,X_n) = Pr(X_n|X_1,\ldots,X_{n-1})Pr(X_{n-1}|X_1,\ldots,X_{n-2})\ldots Pr(X_1)$ requires as much space to represent as specifying the probability of each individual event.

Causal Intuitions

- The BN can be constructed using an arbitrary ordering of the variables.
- However, some orderings will yield BN's with very large parent sets. This requires exponential space, and (as we will see later) exponential time to perform inference.
- Empirically, and conceptually, a good way to construct a BN is to use an ordering based on causality. This often yields a more natural and compact BN.

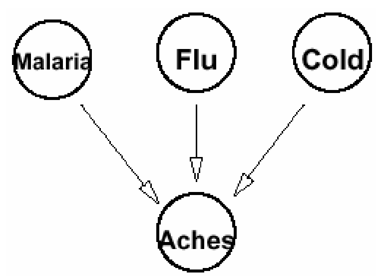
Causal Intuitions

- Malaria, the flu and a cold all "cause" aches. So use the ordering that causes come before effects
  Malaria, Flu, Cold, Aches

$Pr(M,F,C,A) = Pr(A|M,F,C)\ Pr(C|M,F)\ Pr(F|M)\ Pr(M)$

- Each of these disease affects the probability of aches, so the first conditional probability does not change.
- It is reasonable to assume that these diseases are independent of each other: having or not having one does not change the probability of having the others. So
  $Pr(C|M,F) = Pr(C)$
  $Pr(F|M) = Pr(F)$

Causal Intuitions



- This yields a fairly simple Bayes net.
- Only need one big CPT, involving the family of "Aches".

Causal Intuitions

- Suppose we build the BN for distribution P using the opposite ordering
  - i.e., we use ordering Aches, Cold, Flu, Malaria

  Pr(A,C,F,M) = Pr(M|A,C,F) Pr(F|A,C) Pr(C|A) Pr(A)

  - We can't reduce Pr(M|A,C,F). Probability of Malaria is clearly affected by knowing aches. What about knowing aches and Cold, or aches and Cold and Flu?
    - Probability of Malaria is affected by both of these additional pieces of knowledge

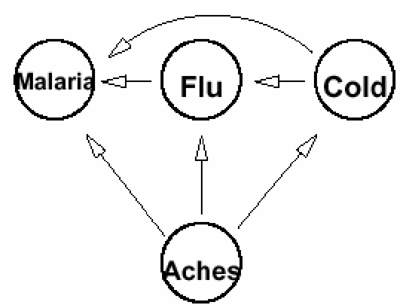
      Knowing Cold and of Flu lowers the probability of Aches indicating Malaria since they "explain away" Aches!

Causal Intuitions

Pr(A,C,F,M) = Pr(M|A,C,F) Pr(F|A,C) Pr(C|A) Pr(A)

- Similarly, we can't reduce Pr(F|A,C).
- Pr(C|A) ≠ Pr(C)

Causal Intuitions



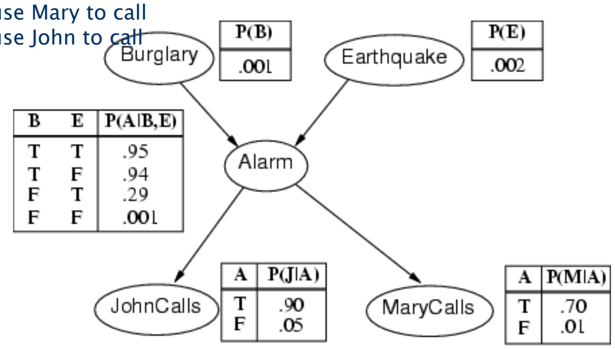
- Obtain a much more complex Bayes net. In fact, we obtain no savings over explicitly representing the full joint distribution (i.e., representing the probability of every atomic event).

Bayes Net Examples

- I'm at work, neighbor John calls to say my alarm is ringing, but neighbor Mary doesn't call. Sometimes it's set off by minor earthquakes. Is there a burglar?
- Variables: *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, *MaryCalls*
- Network topology reflects "causal" knowledge:
  - A burglar can set the alarm off
  - An earthquake can set the alarm off
  - The alarm can cause Mary to call
  - The alarm can cause John to call

Burglary Example

- A burglary can set the alarm off
- An earthquake can set the alarm off
- The alarm can cause Mary to call
- The alarm can cause John to call



P(B)
.001

P(E)
.002

B	E	P(A|B,E)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

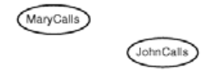
A	P(J|A)
T	.90
F	.05

A	P(M|A)
T	.70
F	.01

- # of Params:1 + 1 + 4 + 2 + 2 = 10 (vs. 2^5–1 = 31)

Example of Constructing Bayes Network

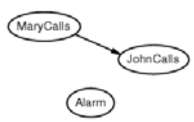
- Suppose we choose the ordering *M, J, A, B, E*



P(J | M) = P(J)?

Example continue…

- Suppose we choose the ordering *M, J, A, B, E*



P(J | M) = P(J)? No
P(A | J, M) = P(A | J)? P(A | J, M) = P(A)?

Example continue…

- Suppose we choose the ordering *M, J, A, B, E*

P(J | M) = P(J)? No
P(A | J, M) = P(A | J)? P(A | J, M) = P(A)?
P(B | A, J, M) = P(B | A)?
P(B | A, J, M) = P(B)?

Example continue…

- Suppose we choose the ordering M, J, A, B, E

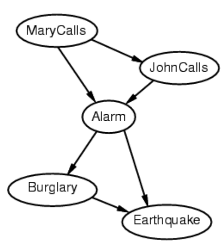
$P(J \mid M) = P(J)$? No
$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$?
$P(B \mid A, J, M) = P(B \mid A)$? Yes
$P(B \mid A, J, M) = P(B)$? No
$P(E \mid B, A, J, M) = P(E \mid A)$?
$P(E \mid B, A, J, M) = P(E \mid A, B)$?

Example continue…

- Suppose we choose the ordering M, J, A, B, E

$P(J \mid M) = P(J)$? No
$P(A \mid J, M) = P(A \mid J)$? $P(A \mid J, M) = P(A)$?
$P(B \mid A, J, M) = P(B \mid A)$? Yes
$P(B \mid A, J, M) = P(B)$? No
$P(E \mid B, A, J, M) = P(E \mid A)$? No
$P(E \mid B, A, J, M) = P(E \mid A, B)$? Yes

Example continue…



- Deciding conditional independence **is hard** in non-causal directions!
- (Causal models and conditional independence seem hardwired for humans!)
- Network is **less compact**: 1 + 2 + 4 + 2 + 4 = 13 numbers needed!