Probabilistic Reasoning

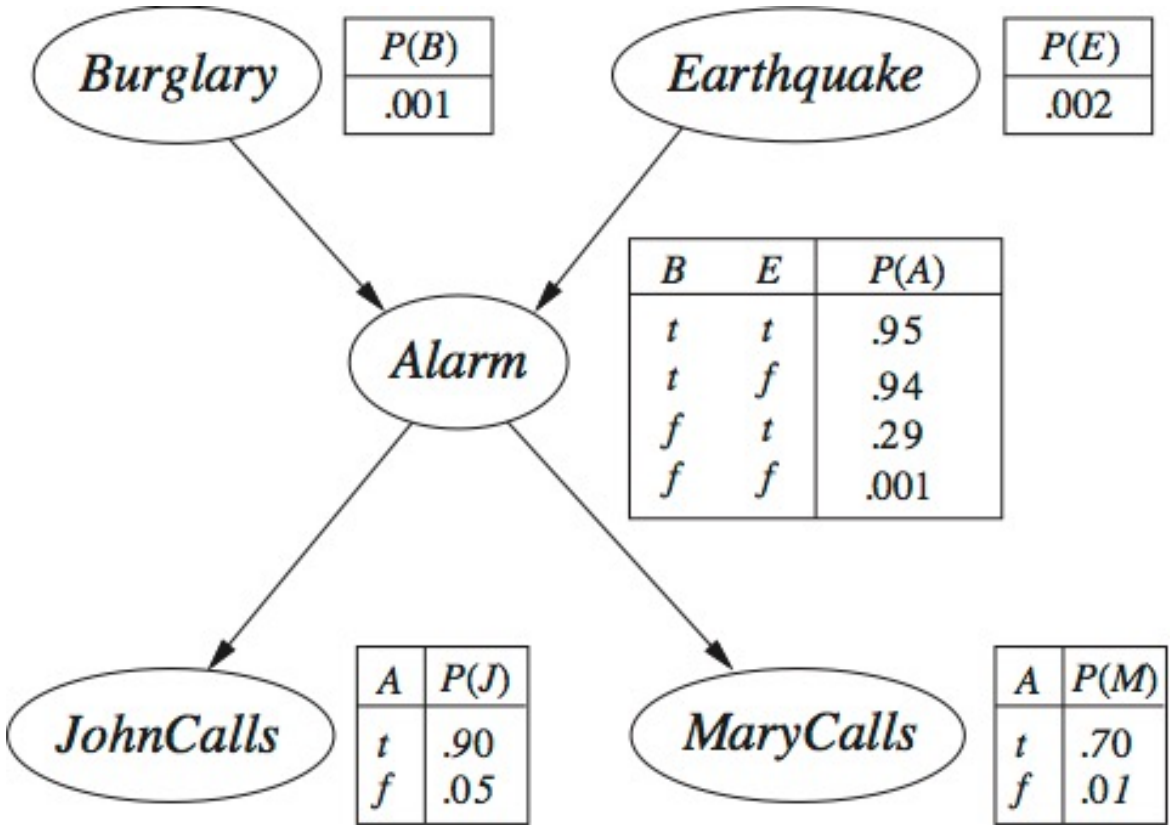
Larry Holder
School of EECS
Washington State University

Probabilistic Reasoning

- Full joint probability distribution
  - Can answer any query
  - But typically too large
- Conditional independence
  - Can reduce the number of probabilities needed
  - $P(X \mid Y,Z) = P(X \mid Z)$, if X independent of Y given Z
- Bayesian network
  - Concise representation of above

Bayesian Network

- Example



P(B)
.001

P(E)
.002

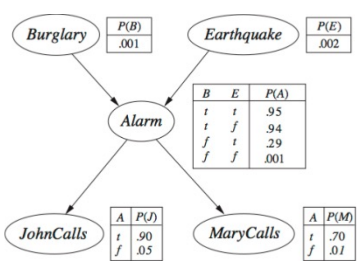
B	E	P(A)
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	P(J)
t	.90
f	.05

A	P(M)
t	.70
f	.01

Bayesian Network

- Bayesian network is a directed, acyclic graph
- Each node corresponds to a random variable
- A directed link from node X to node Y implies that X "influences" Y
  - X is the parent of Y
- Each node X has a conditional probability distribution P(X | Parents(X))
  - Quantifies the influence on X from its parent nodes
  - Conditional probability table (CPT)



Bayesian Networks

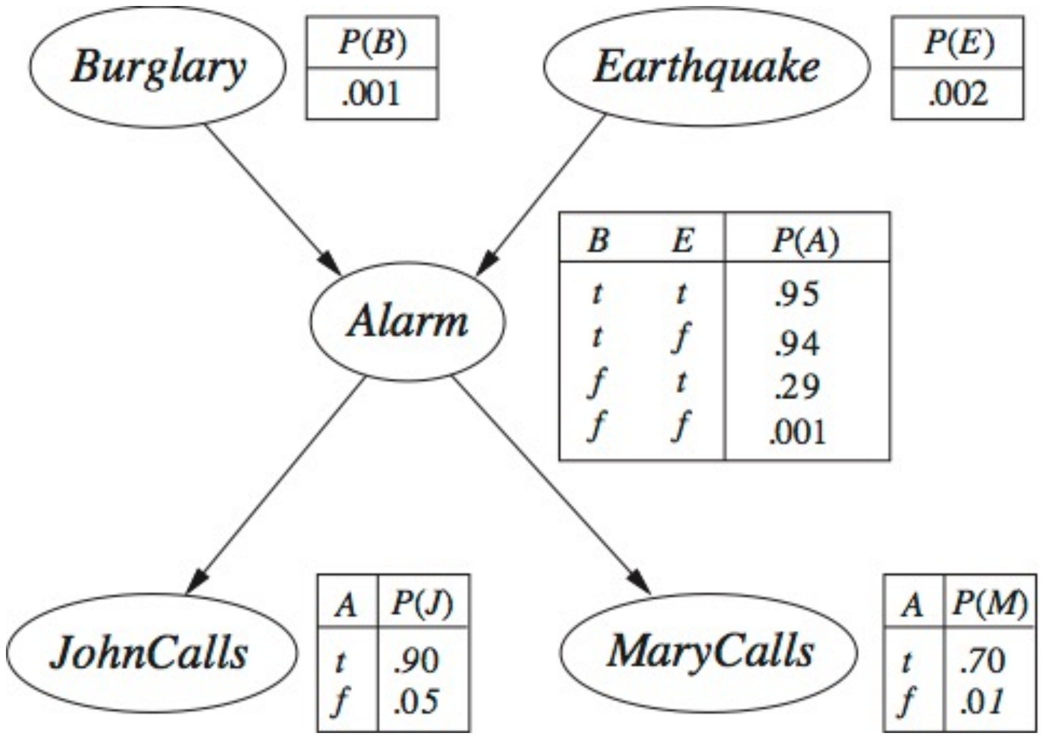
- Represents full joint distribution

$$P(X_1 = x_1 \wedge \ldots \wedge X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i \mid parents(X_i))$$

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$$

- Represents conditional independence
  - E.g., JohnCalls is independent of Burglary and Earthquake given Alarm

Bayesian Networks



- $P(b, \neg e, a, j, m) = ?$

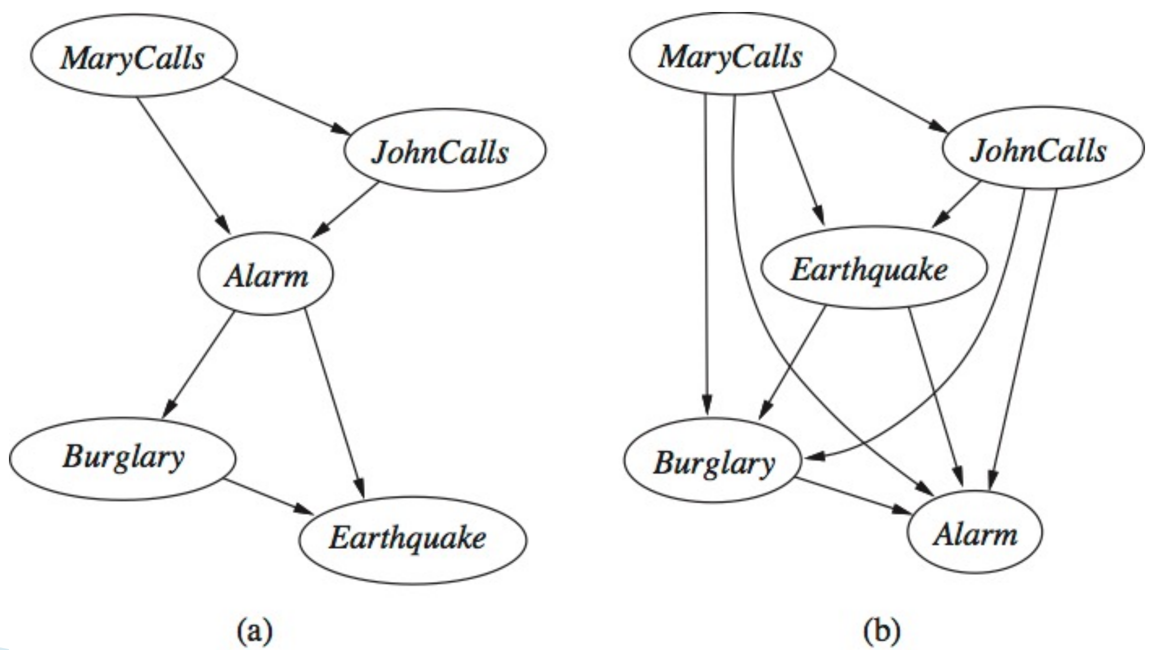
Constructing Bayesian Networks

- Determine set of random variables $\{X_1,\ldots,X_n\}$
- Order them so that causes precede effects
- For i = 1 to n do
  - Choose minimal set of parents for X_i such that $\mathbf{P}(X_i \mid X_{i-1},\ldots,X_1) = \mathbf{P}(X_i \mid Parents(X_i))$
  - For each parent X_k insert link from X_k to X_i
  - Write down the CPT, $\mathbf{P}(X_i \mid Parents(X_i))$

- E.g., Burglary, Earthquake, Alarm, JohnCalls, MaryCalls

Constructing Bayesian Networks

- Bad orderings lead to more complex networks with more CPT entries
  a) MaryCalls, JohnCalls, Alarm, Burglary, Earthquake
  b) MaryCalls, JohnCalls, Earthquake, Burglary, Alarm



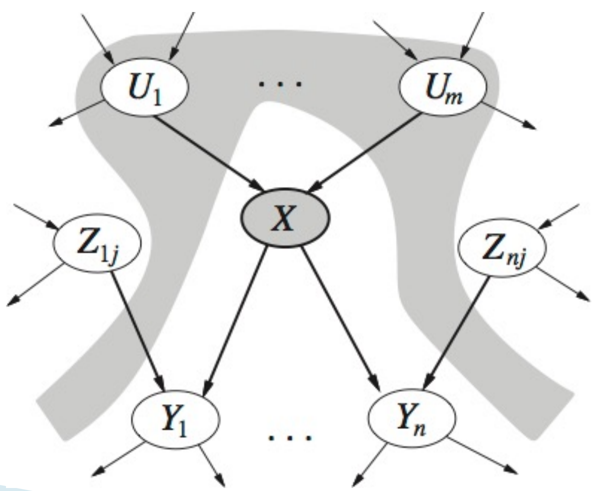
Constructing Bayesian Networks

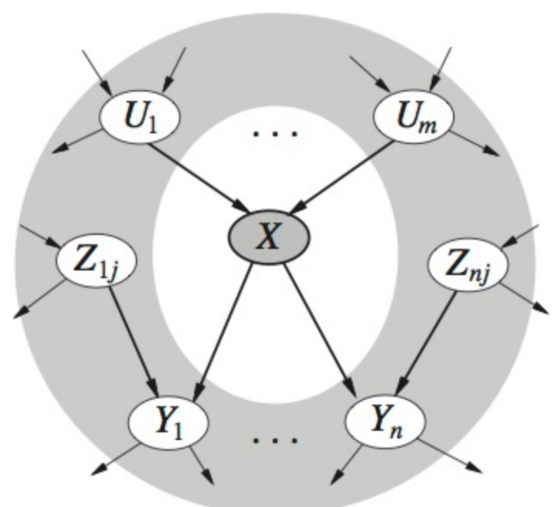
- Example: Tooth World

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	.108	.012	.072	.008
¬cavity	.016	.064	.144	.576

Conditional Independence

- Node X is conditionally independent of its non-descendants (Z_{ij}'s) given its parents (U_i's)
- Markov blanket of node X is X's parents (U_i's), children (Y_i's) and children's parents (Z_{ij}'s)
- Node X is conditionally independent of all other nodes in the network given its Markov blanket





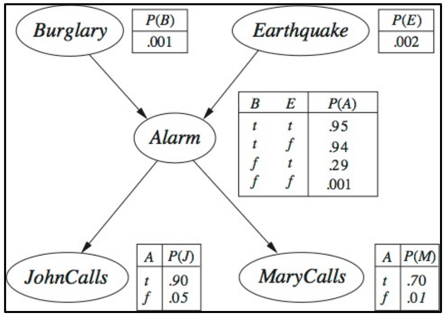
Inference in Bayesian Networks

- Want **P**(X | e)
- X is the query variable (can be more than one)
- e is an observed event, i.e., values for the evidence variables **E** = $\{E_1,\ldots,E_m\}$
- Any other variables **Y** are hidden variables

- Example
  - **P**(Burglary | JohnCalls=true, MaryCalls=true) = ?
  - X = Burglary
  - e = {JohnCalls=true, MaryCalls=true}
  - Y = {Earthquake, Alarm}

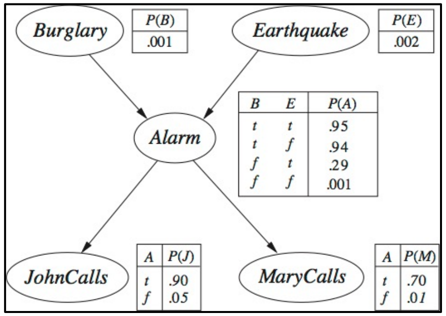
Inference by Enumeration

- Enumerate over all possible values for Y
  - $\mathbf{P}(X \mid \mathbf{e}) = \alpha \, \mathbf{P}(X,\mathbf{e}) = \alpha \, \Sigma\mathbf{y} \, \mathbf{P}(X,\mathbf{e},\mathbf{y})$
- Example
  - **P**(Burglary | JohnCalls=true, MaryCalls=true)
  - **P**(B | j, m) = ?



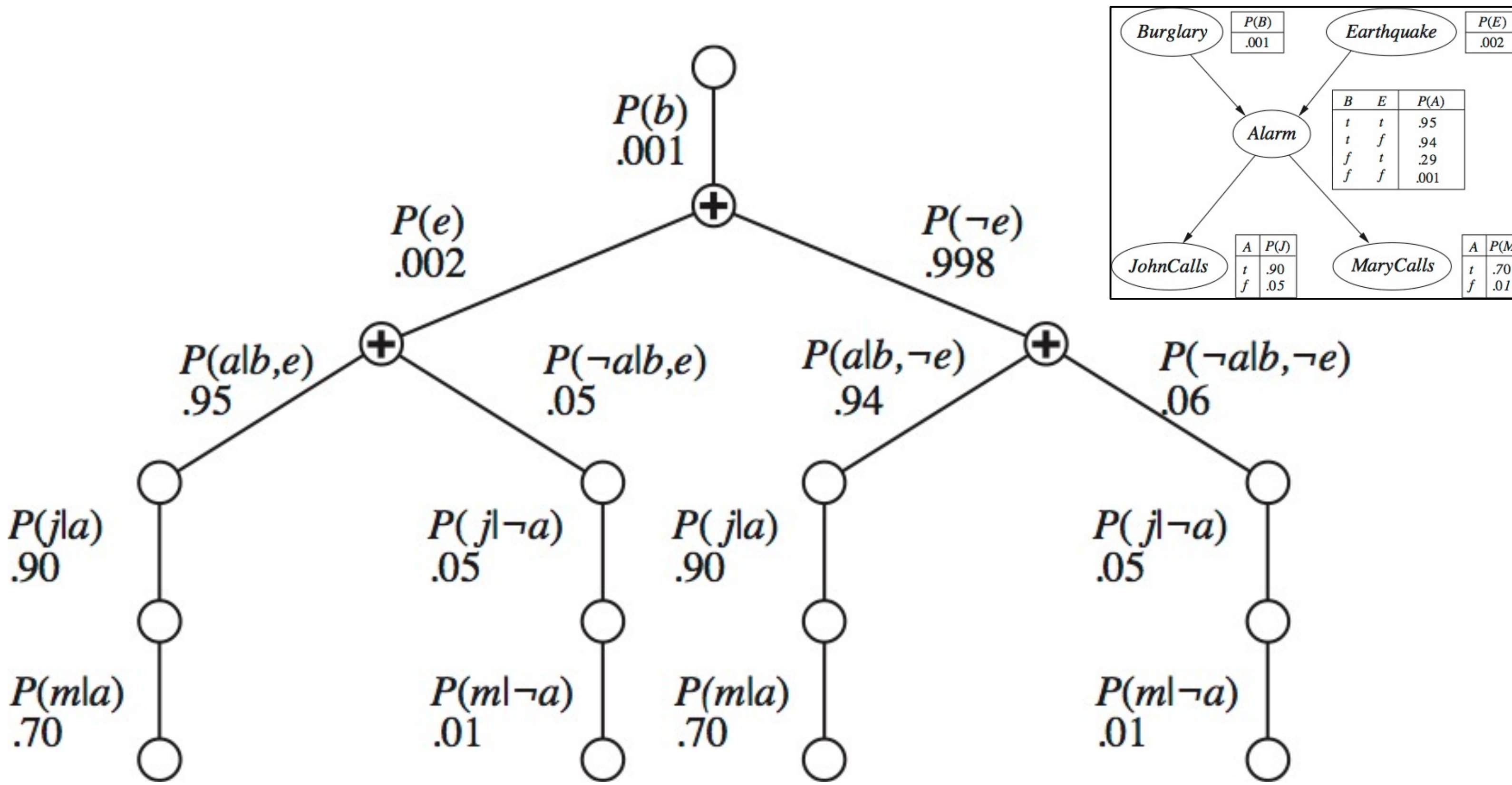
Inference by Enumeration

- $\mathbf{P}(B|j,m) = \alpha\, \mathbf{P}(B)\, \Sigma_E\; P(E)\, \Sigma_A\, \mathbf{P}(A|B,E)\, P(j|A)\, P(m|A)$
- $P(b|j,m) = \alpha\, P(b)\, \Sigma_E\, P(E)\, \Sigma_A\, P(A|b,E)\, P(j|A)\, P(m|A)$



Inference by Enumeration

- $P(b|j,m) = \alpha\, P(b) \sum_E P(E) \sum_A P(A|b,E)\, P(j|A)\, P(m|A)$



Inference by Enumeration

function ENUMERATION-ASK (X, **e**, bn) **returns** a distribution over X
inputs: X, the query variable
e, observed values of variables **E**
bn, a Bayes net with variables $\{X\} \cup \mathbf{E} \cup \mathbf{Y}$ // $\mathbf{Y}$ = *hidden variables*
$\mathbf{Q}(X) \leftarrow$ a distribution over X, initially empty
for each value x_i of X **do**
$\mathbf{Q}(x_i) \leftarrow$ ENUMERATE-ALL(bn.VARS, $\mathbf{e}_{x_i}$)
where $\mathbf{e}_{x_i}$ is **e** extended with $X = x_i$
return NORMALIZE($\mathbf{Q}(X)$)

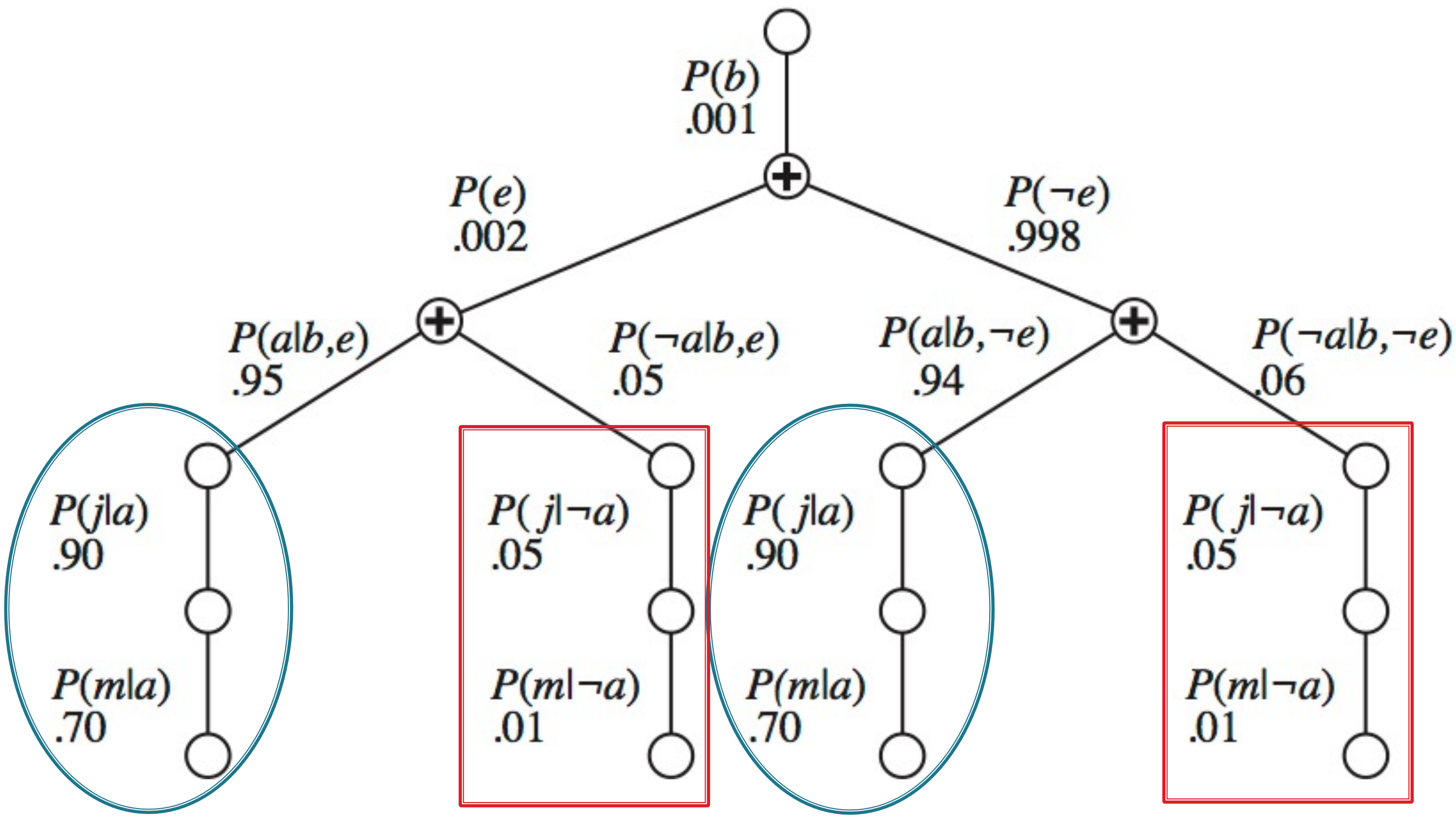
bn.VARS has variables in cause→effect order

function ENUMERATE-ALL ($vars$, **e**) **returns** a real number
if EMPTY? ($vars$) **then return** 1.0
$Y \leftarrow$ FIRST ($vars$)
if Y has value y in **e**
then return $P(y \mid parents(Y)) \times$ ENUMERATE-ALL(REST($vars$), **e**)
else return $\sum_y P(y \mid parents(Y)) \times$ ENUMERATE-ALL(REST($vars$), $\mathbf{e}_y$)
where $\mathbf{e}_y$ is **e** extended with $Y = y$

Inference by Enumeration

- ENUMERATION–ASK evaluates trees using depth–first recursion
- Space complexity O(n)
- Time complexity $O(v^n)$, where each of n variables has v possible values

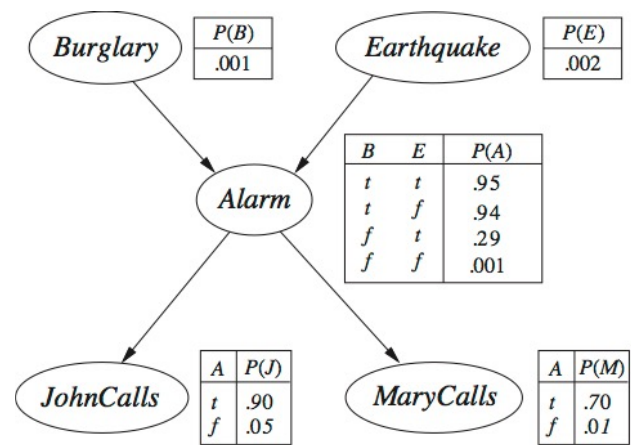
Inference by Enumeration



Note redundant computation

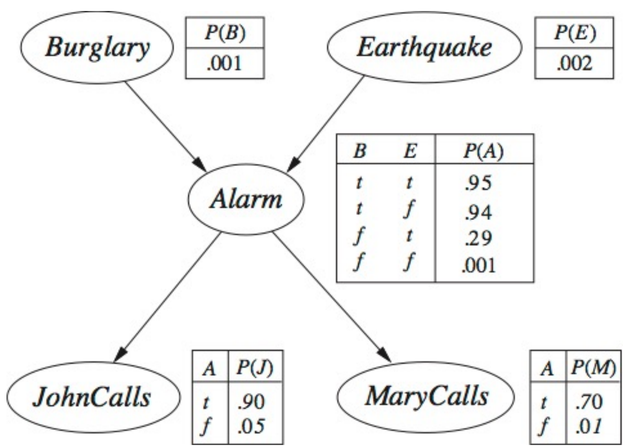
Efficient Inference

- Avoid redundant computation
  - Dynamic programming
  - Store intermediate computations and reuse
- Eliminate irrelevant variables
  - Variables that are not an ancestor of a query or evidence variable

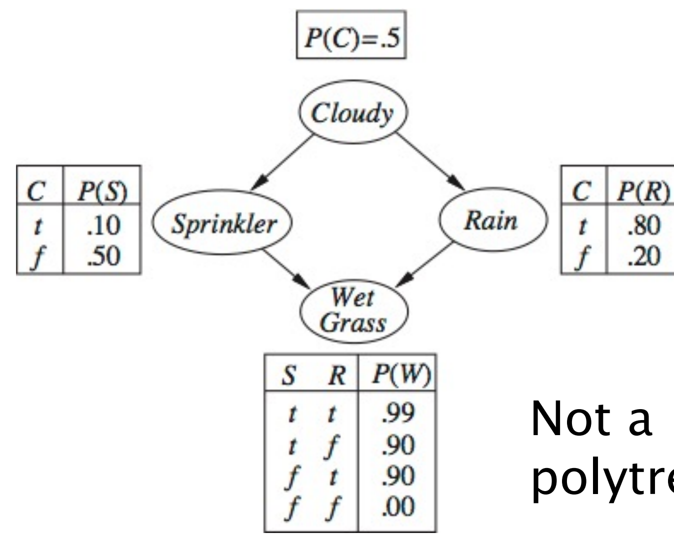


Complexity of Inference

- General case (any type of network)
  - Worst case space and time complexity is exponential
- <u>Polytree</u> is a network with at most one undirected path between any two nodes
  - Space and time complexity is linear in size of network



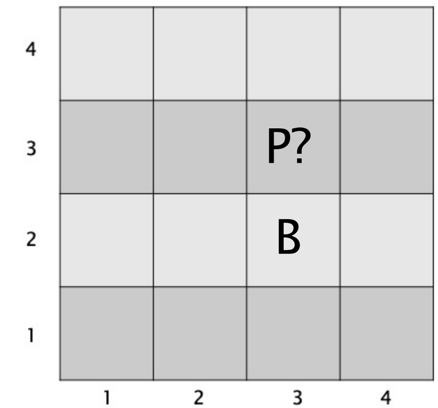
Polytree



Not a polytree

Wumpus World

- $\mathbf{P}(Pit_{3,3} \mid Breeze_{3,2}=true) = ?$

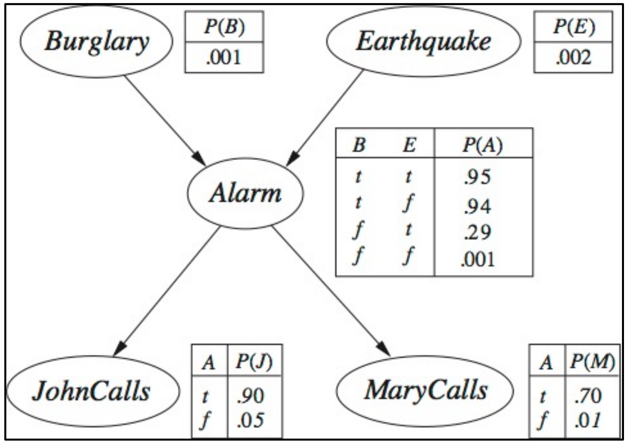


Approximate Inference

- Exact inference can be too expensive
- Approximate inference
  - Estimate probabilities from sample, rather than computing exactly
- Monte Carlo methods
  - Choose values for hidden variables
  - Compute query variables
  - Repeat and average
- Direct sampling
- Converges to exact inference

Direct Sampling

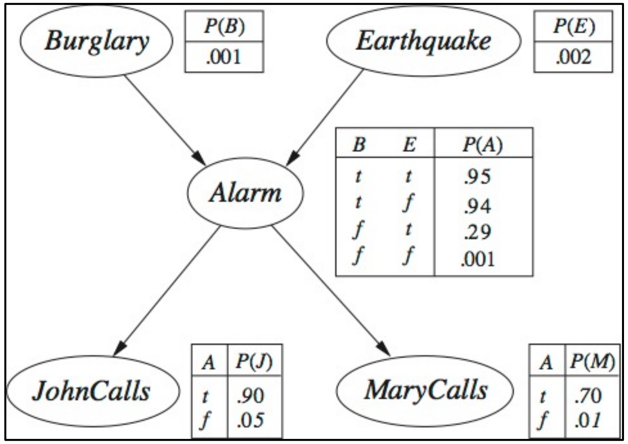
- Choose value for variables according to their CPT
  - Consider variables in topological order
- E.g.,
  - **P**(B) = ⟨0.001,0.999⟩, B=false
  - **P**(E) = ⟨0.002,0.998⟩, E=false
  - **P**(A|B=false,E=false) = ⟨0.001,0.999⟩, A=false
  - **P**(J|A=false) = ⟨0.05,0.95⟩, J=false
  - **P**(M|A=false) = ⟨0.01,0.99⟩, M=false
  - Sample is [false,false,false,false,false]



$$P(X = x_i) \approx \frac{|\,\text{samples where } X = x_i\,|}{|\,\text{samples}\,|}$$

Direct Sampling

- Another example



Bayes Net Software

- Commercial
  - Bayes Server (www.bayesserver.com)
  - BayesiaLab (www.bayesia.com)
  - HUGIN (www.hugin.com)
- Free
  - BayesPy (www.bayespy.org)
  - JavaBayes (www.cs.cmu.edu/~javabayes)
  - SMILE (www.bayesfusion.com)
- Sample networks
  - www.bnlearn.com/bnrepository

Summary: Probabilistic Reasoning

- Bayesian networks
  - Captures full joint probability distribution and conditional independence
- Exact inference
  - Intractable in worst case
- Approximate inference
  - Sampling
  - Converges to exact inference