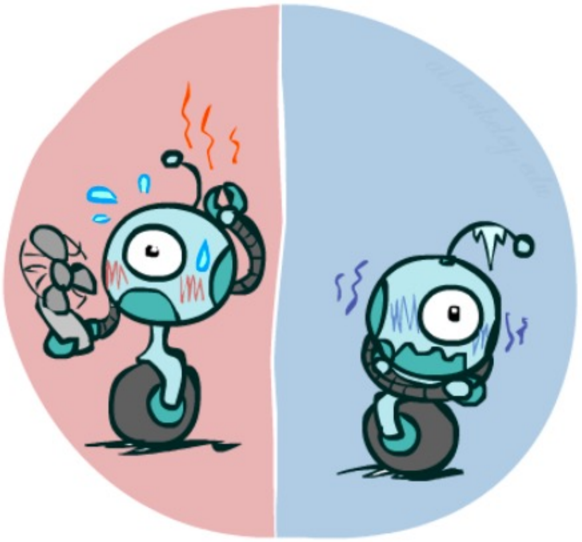
AIFA: Reasoning Under Uncertainty

12/03/2024

Koustav Rudra

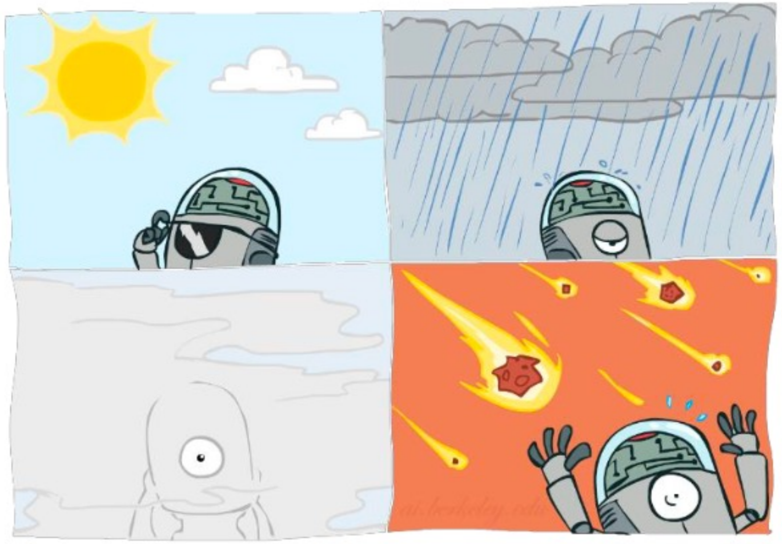
Probability Distributions

- A probability distribution is a description of how likely a random variable is to take on each of its possible states
- Notation: $P(X)$ is the probability distribution over the random variable X
- Associate a probability with each value



$P(T)$

T	P
hot	0.5
cold	0.5



$P(W)$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0

- Unobserved random variables have distributions
- A distribution is a TABLE of probabilities of values

Axioms of Probability

- The probability of an event A in the given sample space $\mathcal{S}$, denoted as $P(A)$, must satisfy the following properties:

- Non-negativity
  - For any event A $\in \mathcal{S}$, $P(A) \geq 0$

- All possible outcomes
  - Probability of the entire sample space is 1, $P(\mathcal{S}) = 1$

- Additivity of disjoint events
  - For all events A1, $A2 \in \mathcal{S}$ that are mutually exclusive ($A1 \cap A2 = \emptyset$), the probability that both events happen is equal to the sum of their individual probabilities, $P(A1 \vee A2) = P(A1) + P(A2)$

Joint Distributions

- A joint distribution over a set of random variables: $X1, X2, \dots, Xn$

- Specifies a real number for each assignment (or outcome):
  - $P(X1 = x1, X2 = x2, \dots, Xn = xn)$
  - $P(x1, x2, \dots, xn)$

- Must satisfy
  - $P(x1, x2, \dots, xn) \geq 0$
  - $\sum_{x1,x2,\dots,xn} \mathrm{P}(x1,x2,\dots,xn)=1$

- Size of distribution if n variables with domain sizes d?

T	W	P
Hot	Sun	0.4
Hot	Run	0.1
Cold	Sun	0.2
Cold	Run	0.3

Probabilistic Models

- A probabilistic model is a joint distribution over a set of random variables

- Probabilistic models:
  - Random variables with domains
  - Assignments are called outcomes
  - Joint distributions: say whether
    - assignments (outcomes) are likely
  - Normalized: sum to 1.0
  - Ideally: only certain variables directly interact

T	W	P
Hot	Sun	0.4
Hot	Run	0.1
Cold	Sun	0.2
Cold	Run	0.3

Events

- An event is a set E of outcomes
- $P(E) = \sum_{x1,x2,\ldots,xn \in E} \mathrm{P}(x1,x2,\ldots,xn)$

- From a joint distribution, we can calculate the probability of any event
  - Probability that it's hot AND sunny?
  - Probability that it's hot?
  - Probability that it's hot OR sunny?

T	W	P
Hot	Sun	0.4
Hot	Run	0.1
Cold	Sun	0.2
Cold	Run	0.3

Marginal Probability Distribution

- Marginal probability distribution is the probability distribution of a single variable
- It is calculated based on the joint probability distribution $P(X,Y)$ using the sum rule:
- $P(X = x) = \sum_y P(X = x, Y = y)$

Marginal Distributions

- Marginal distributions are sub-tables which eliminate variables
- Marginalization (summing out): Combine collapsed rows by adding

T	W	P
Hot	Sun	0.4
Hot	Run	0.1
Cold	Sun	0.2
Cold	Run	0.3

$$P(t) = \sum_{s} P(t,s) = P(T)$$

T	P
Hot	0.5
Cold	0.5

$$P(s) = \sum_{t} P(t,s) = P(W)$$

W	P
Sun	0.6
Rain	0.4

$$P(X1 = x1) = \sum_{x2} P(X1 = x1, X2 = x2)$$

Conditional Probabilities

- $P(a|b) = \frac{P(a,b)}{P(b)}$
- $P(W = s|T = c) = \frac{P(W=s,T=c)}{P(T=c)}$
- $P(W = s|T = c) = \frac{0.2}{0.5}$

T	W	P
Hot	Sun	0.4
Hot	Run	0.1
Cold	Sun	0.2
Cold	Run	0.3

Conditional Distributions

- Conditional distributions are probability distributions over some variables given fixed values of others

Conditional Distributions

$P(W|T)$

$P(W|T = hot)$

W	P
sun	0.8
rain	0.2

$P(W|T = cold)$

W	P
sun	0.4
rain	0.6

Joint Distribution

$P(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Normalization Trick

T	W	P
Hot	Sun	0.4
Hot	Run	0.1
Cold	Sun	0.2
Cold	Run	0.3

- SELECT the joint probabilities matching the evidence

$P(c, W)$

T	W	P
Cold	Sun	0.2
Cold	Rain	0.3

- Normalize the selection

$P(W|T = c)$

W	P
Sun	0.2
Rain	0.3

Product Rule

- Sometimes have conditional distributions but want the joint
- $P(y)P(x|y) = P(x,y)$

Bayesian Network

Conditional Independence

- Two events A and B are conditionally independent given another event C with $P(C) > 0$:
  - $P(A \land B|C) = P(A|C)P(B|C)$

- $P(A|B \land C) = \frac{P(A \land B|C)}{P(B|C)}$
- $P(A|B \land C) = \frac{P(A|C) \land P(B|C)}{P(B|C)}$
- $P(A|B \land C) = P(A|C)$

Conditional Independence and Chain Rule

Trivial decomposition:
P(Traffic, Rain, Umbrella)=
P(Rain)P(Traffic|Rain)P(Umbrella|Rain,Traffic)

With assumption of conditional independence:
P(Traffic, Rain, Umbrella)=
P(Rain)P(Traffic|Rain)P(Umbrella|Rain)

Bayes nets / graphical models help us express conditional independence assumptions

Conditional Independence

- $P(x_1, \dots, x_n) = P(x_n | x_{n-1}, \dots, x_1) P(x_{n-1}, \dots, x_1)$
- $P(x_1, \dots, x_n) = P(x_n | x_{n-1}, \dots, x_1) P(x_{n-1} | x_{n-2}, \dots, x_1) \dots P(x_2 | x_1) P(x_1)$
- $P(x_1, \dots, x_n) = \prod_{i=1}^{n} P(x_i | x_{i-1}, \dots, x_1)$

- The belief network represents conditional independence:
  - $P(x_i | x_i, \dots, x_1) = P(x_i | Parents(x_i))$

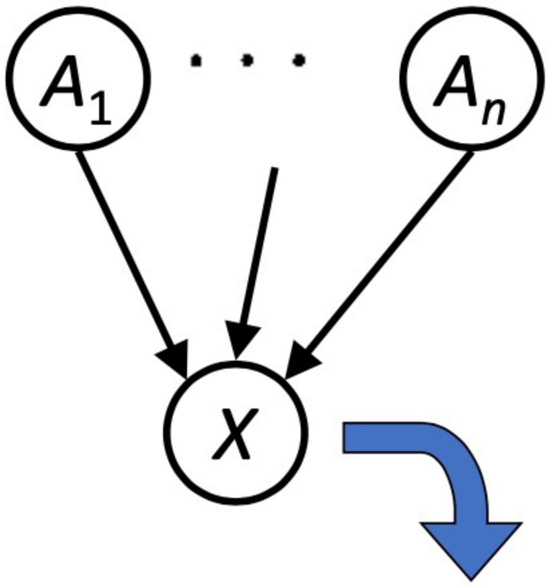
Bayes Nets

- A Bayes' net is an efficient encoding of a probabilistic model of a domain

- Questions we can ask:
  - Inference: given a fixed BN, what is $P(X \mid e)$?
  - Representation: given a BN graph, what kinds of distributions can it encode?
  - Modeling: what BN is most appropriate for a given domain?

Bayes Net Semantics

- A set of nodes, one per variable X
- A directed, acyclic graph
- A conditional distribution for each node
  - A collection of distributions over X, one for each combination of parents' values
  - $P(X|a_1, a_2, \ldots, a_n)$
- CPT: conditional probability table
- Description of a noisy "causal" process



$P(X|A_1 \ldots A_n)$

A Bayes net = Topology (graph) + Local Conditional Probabilities

Probabilities in BN

- Bayes' nets implicitly encode joint distributions
  - As a product of local conditional distributions
  - $P(x_i|x_{i-1}, x_{i-2}, \dots, x_1) = P(x_i|Parents(x_i))$

- $Parents(x_i)$: minimal set of predecessors of Xi in the total ordering such that other predecessors are conditionally independent of Xi given Parent(Xi)

Bayesian Network

- What is the issue with joint probability distribution?
  - Become intractably large as the number of variables grows
  - Specifying probabilities for atomic events is really difficult

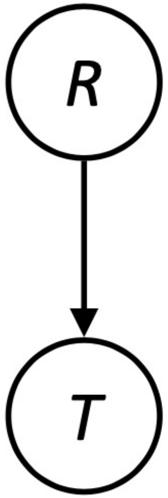
- How does Bayesian Network help?
  - Explore independence and conditional independence relationships among variables
  - To greatly reduce number of probabilities to be specified to define full joint distribution

Bayesian Network

- A set of random variables makes up the nodes of the network
  - Variables may be discrete or continuous
- A set of directed links or arrows connects pairs of nodes
  - Arrows represent probabilistic dependence among variables
- An arrow from X →Y indicates X is parent of Y
- Each node X_i $has\ a\ conditional\ probability\ distribution\ P(X_i|Parents(X_i))$
  - Quantifies the effect of the parents on the node
- The graph has no directed cycles (DAG)

Example1: Traffic

- Causal direction



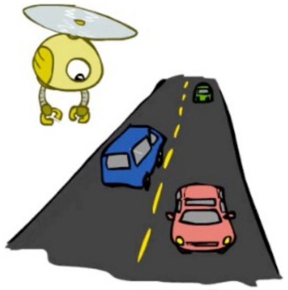
$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

+r	+t	3/4
	-t	1/4

-r	+t	1/2
	-t	1/2

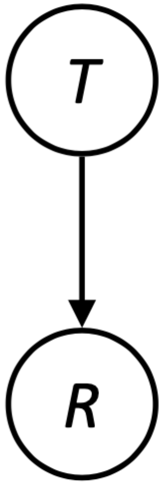


$P(T, R)$

+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

Example2: Traffic

- Reverse causality?



$P(T)$

+t	9/16
-t	7/16

$P(R|T)$

+t	+r	1/3
	-r	2/3
-t	+r	1/7
	-r	6/7

$P(T, R)$

+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

Causality?

- When Bayes' nets reflect the **true causal** patterns:
  - Often simpler (nodes have fewer parents)
  - Often easier to think about and to elicit from experts

- BNs need not actually be causal
  - Sometimes no causal net exists over the domain (especially if variables are missing)
  - End up with arrows that reflect correlation, not causation

- What do the arrows really mean?
  - Topology may happen to encode causal structure
  - Topology really encodes conditional independence

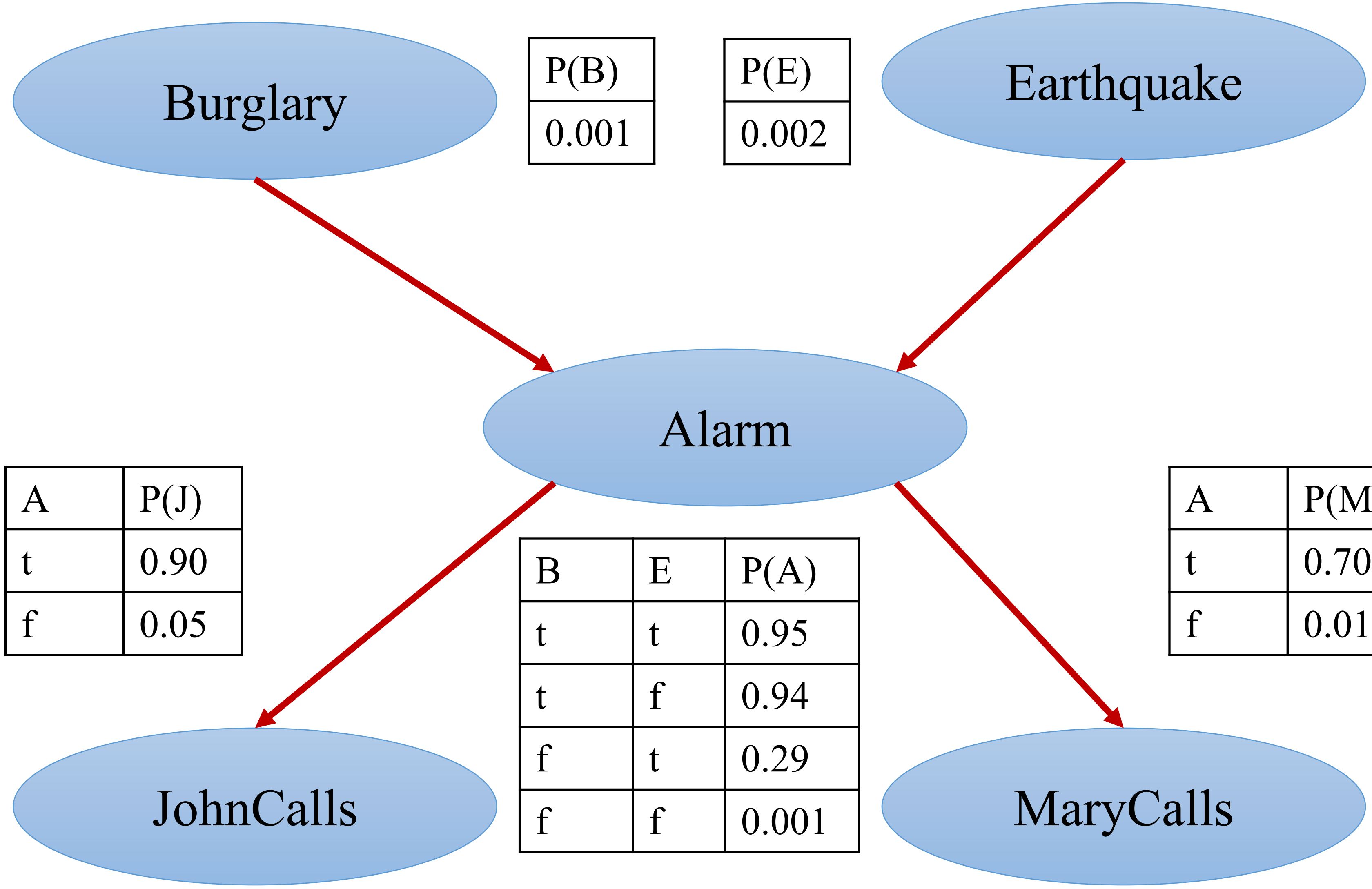
Example3: Home Alarm Network

- Burglar alarm at home
    - Fairly reliable at detecting a burglary
    - Responds at times to minor earthquakes

- Two neighbors, on hearing alarm, calls police
    - John always calls when he hears the alarm
    - But sometimes confuses the telephone ringing with the alarm and calls then too

    - Mary likes loud music
    - But sometimes misses the alarm altogether

Belief Network



P(B)
0.001

P(E)
0.002

B	E	P(A)
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

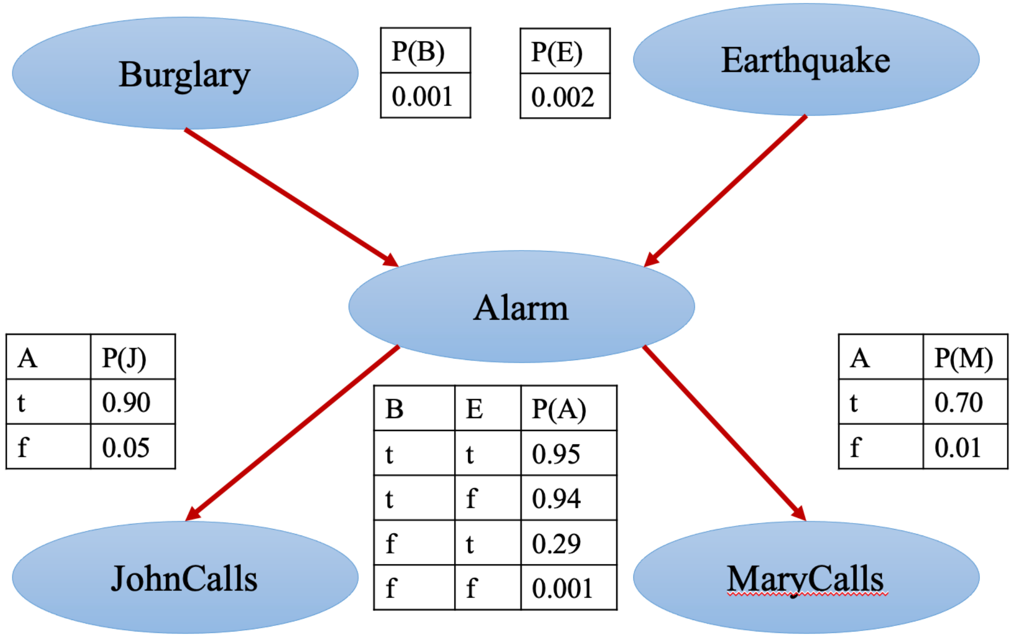
A	P(J)
t	0.90
f	0.05

A	P(M)
t	0.70
f	0.01

Joint probability distribution

- $P(x_1, \dots, x_n) = \prod_{i=1}^{n} P(x_i | Parents(x_i))$

- $P(J \wedge M \wedge A \wedge {\sim}B \wedge {\sim}E)$
  - $P(J|A)\ *$
  - $P(M|A)\ *$
  - $P(A|{\sim}B \wedge {\sim}E)\ *$
  - $P({\sim}B)\ *$
  - $\mathrm{P}({\sim}E)$



P(B)
0.001

P(E)
0.002

B	E	P(A)
t	t	0.95
t	f	0.94
f	t	0.29
f	f	0.001

A	P(J)
t	0.90
f	0.05

A	P(M)
t	0.70
f	0.01

- $P(J \wedge M \wedge A \wedge {\sim}B \wedge {\sim}E) = 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998 = 0.00062$

- $P(J) = ?$

Conditional Independence

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

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

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

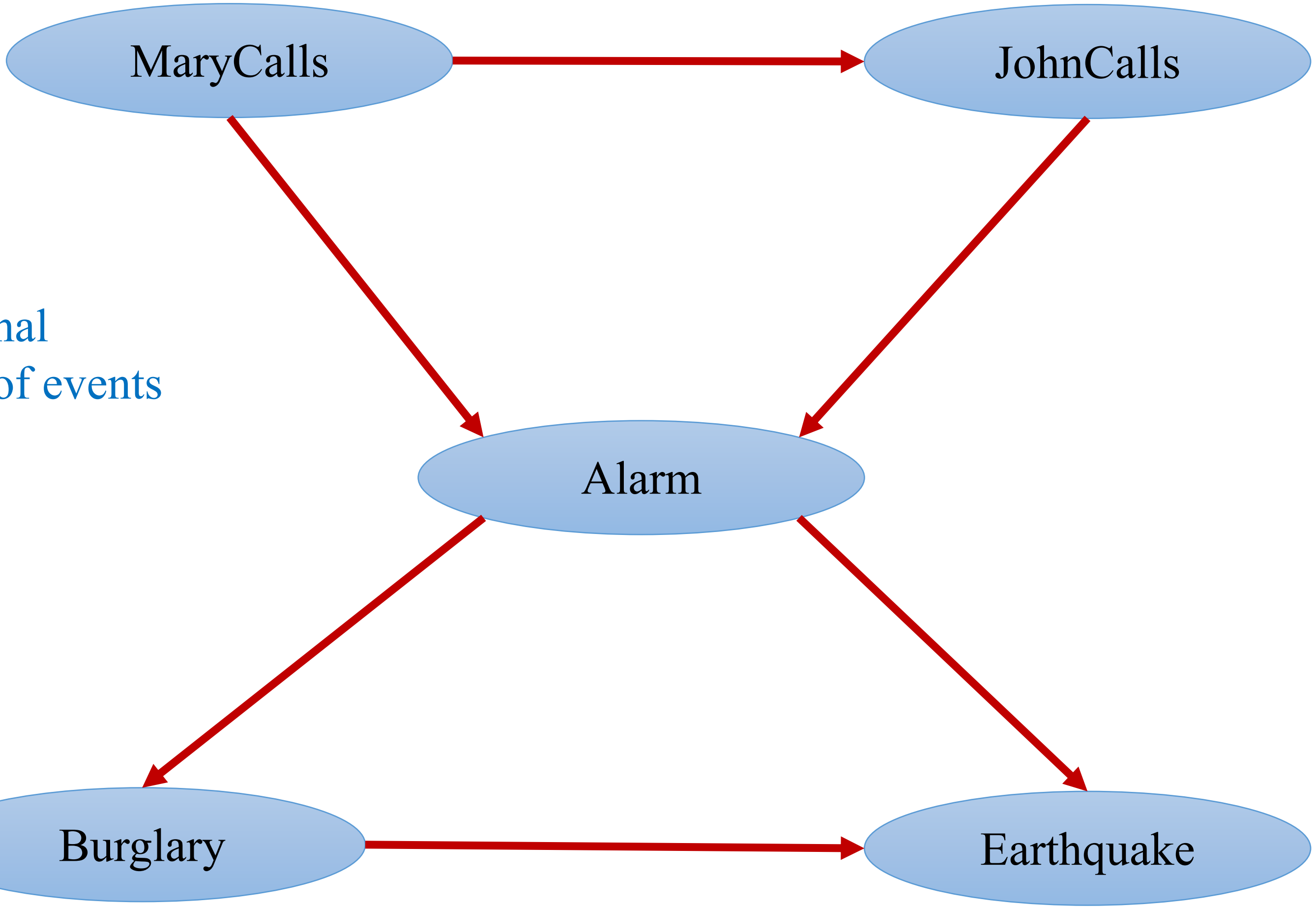
- $P(J, M, A, B, E) = P(J|A)P(M|A)P(A|B, E)P(B)P(E)$

How does ordering matter?

Conditional Independence

- Earthquake, Burglary, Alarm, JohnCalls, MaryCalls
- P(E|B,A,J,M)

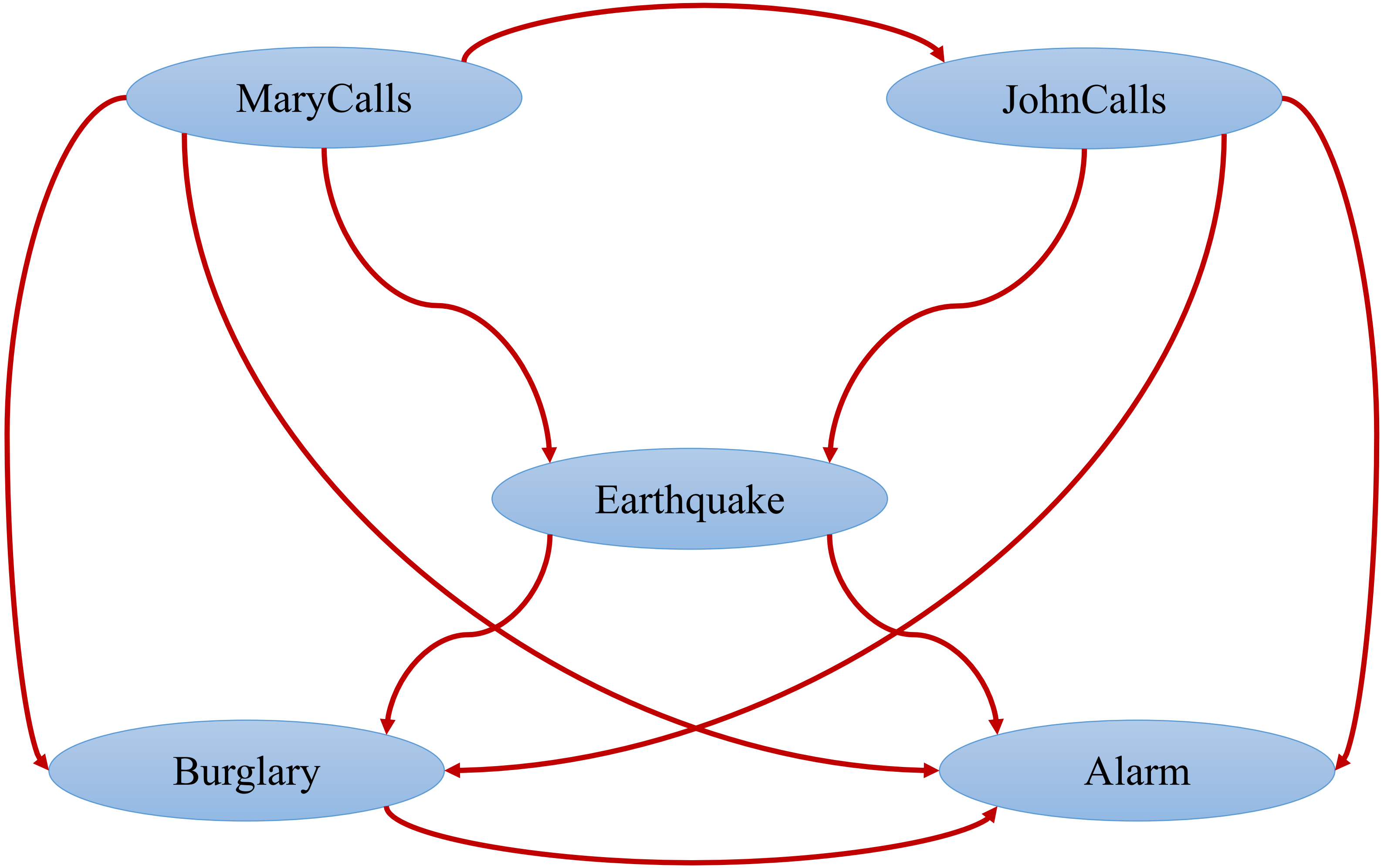
May have to define conditional probability of confusing set of events



Conditional Independence

- Alarm, Burglary, Earthquake, JohnCalls, MaryCalls

May have to construct
large probability tables



Incremental Network Construction

- Choose the set of relevant variables X_i, that describe the domain

- Choose an ordering for the variables [important step]

- While there are variables left:
  - Pick a variable X and add a node for it
  - Set Parents(X) to some minimal set of existing nodes such that the conditional independence property is satisfied
  - Define conditional probability table for X

Why do we construct Bayes Network?

To answer queries related to joint probability distribution

Size of a Bayes Net

- How big is a joint distribution over N Boolean variables?
  - 2^N
- How big is an N-node net if nodes have up to k parents?
  - $O(N * 2^k)$
- Both give you the power to calculate
- BNs: Huge space savings!
- Also easier to elicit local CPTs
- Also faster to answer queries (coming)

Thank You