AIFA: Reasoning Under Uncertainty

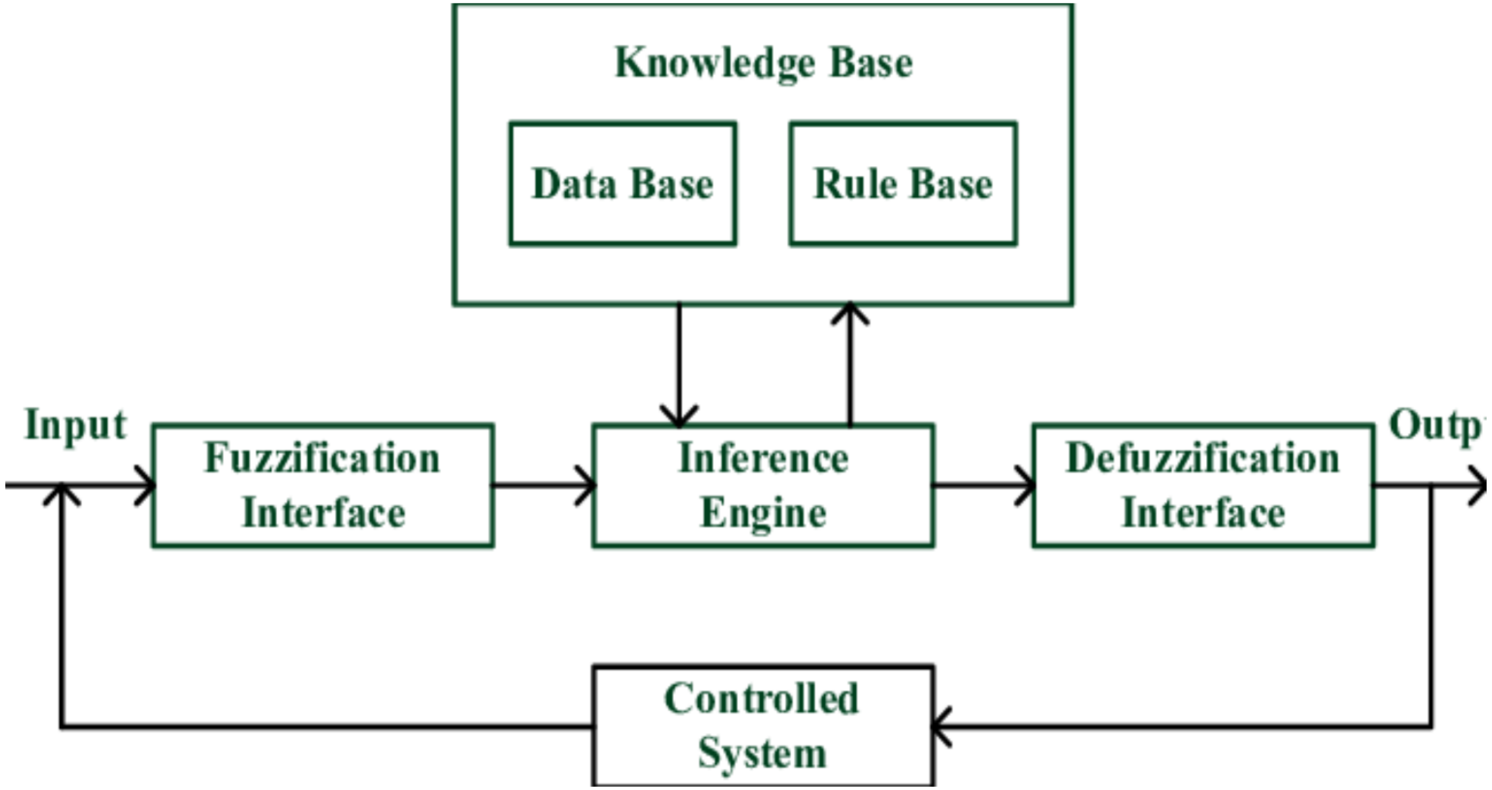
17/03/2025

Koustav Rudra

Fuzzy Inference System

- Rule Base: Contains IF-THEN Rules
- Fuzzification: Convert crisp inputs to Fuzzy set
- Inference engine: determines matching degree of current input
- Defuzzification: Convert fuzzy values to crisp values

Fuzzy Inference System



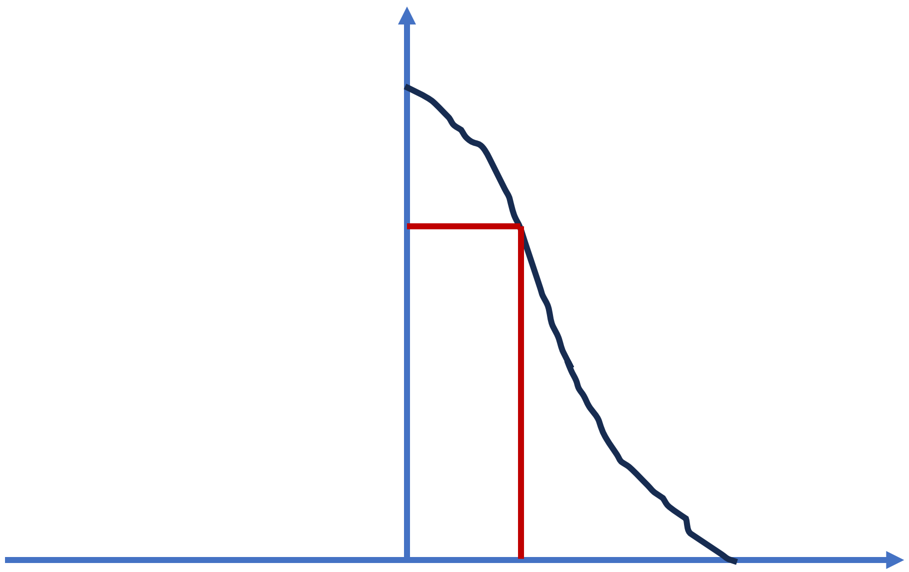
Fuzzification

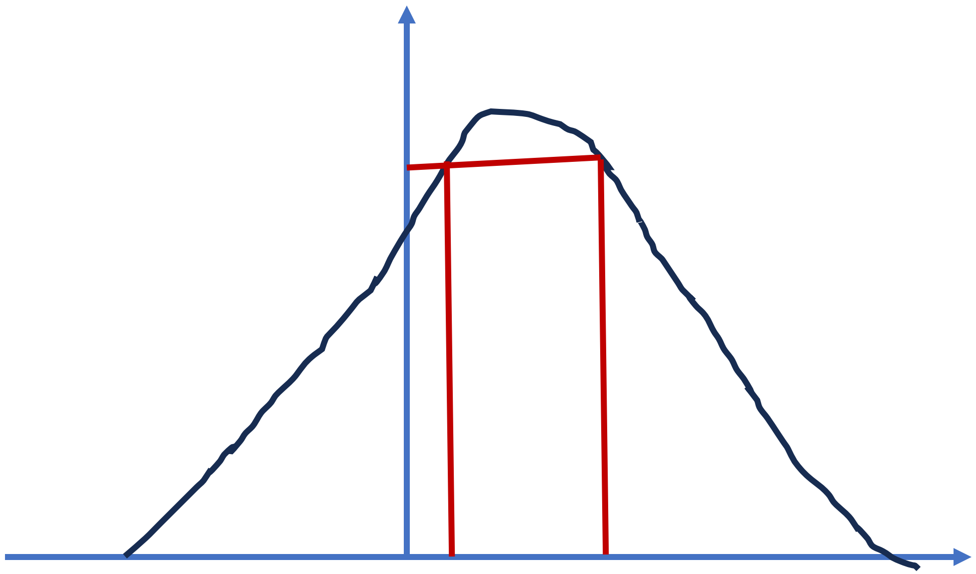
- Why do we need Fuzzification?
  - Rules are Fuzzy
- Converting a crisp value such as height = 5 ft 6 inches to a membership value of a fuzzy set, such as medium or tall
- Different ways of fuzzification – experimental/subjective
- Fuzzified value serves as input to the fuzzy rules

Defuzzification

- Converting a fuzzy term such as small shift
- To a crisp value such as 5 degrees
- Different methods --- such as COG (Centre of Gravity)

Defuzzification





AIFA: Reasoning Under Uncertainty

17/03/2025

Koustav Rudra

Handling uncertain knowledge

- $\forall_p symptom(p, Toothache) \rightarrow disease(p, Cavity)$
  - Not correct – toothache can be caused in many other cases

- $\forall_p symptom(p, Toothache) \rightarrow$
    - $disease(p, Cavity) \vee$
    - $disease(p, GumDisease) \vee$
    - …

Reasons for using probability

- Specification becomes too large
  - Difficult to get complete list of antecedents or consequents

- Theoretical ignorance
  - The complete set of antecedents not known

- Practical ignorance
  - The truth of antecedents not known

Reasons for using probability

- Probability that X is fat = 0.2

- If X is fat then X has coronary heart disease = 0.7

- P[X has CHD] = 0.2*0.7 + 0.8*Z

Probability Basics

- **Joint Probability**
  - P(A = a, C = c): joint probability that random variables A and C will take values a and c respectively

- **Conditional Probability**
  - P(A = a | C = c): conditional probability that A will take the value a, given that C has taken value c
  - $P(A|C) = \frac{P(A,C)}{P(C)}$

Bayes Theorem

- Bayes theorem:
  - $P(C|A) = \frac{P(A|C)P(C)}{P(A)}$
  - P(C) known as the **prior probability** for class C
  - P(C|A) known as the **posterior probability**

Example of Bayes Theorem

- Given:
  - A doctor knows that meningitis (M) causes stiff neck (S) 50% of the time
    - $P(S|M) = 0.50$
  - Prior probability of any patient having meningitis is 1/50,000
    - $P(M) = \frac{1}{50000}$
  - Prior probability of any patient having stiff neck is 1/20
    - $P(S) = \frac{1}{20}$
- If a patient has stiff neck, what's the probability he/she has meningitis?
  - $P(M|S)$

Example of Bayes Theorem

- Given:
  - A doctor knows that meningitis (M) causes stiff neck (S) 50% of the time
    - $P(S|M) = 0.50$
  - Prior probability of any patient having meningitis is 1/50,000
    - $P(M) = \frac{1}{50000}$
  - Prior probability of any patient having stiff neck is 1/20
    - $P(S) = \frac{1}{20}$

- If a patient has stiff neck, what's the probability he/she has meningitis?
  - $P(M|S)$
  - $P(M|S) = \frac{P(S|M).P(M)}{P(S)} = \frac{0.50 \times \frac{1}{50000}}{\frac{1}{20}} = 0.0002$

Probability Distribution

- Describes joint probability distribution over a set of variables

- A set of random variables $Y_1, Y_2, \ldots, Y_n$
  - Each Y_i *can take on the set of possible values* $V(Y_i)$

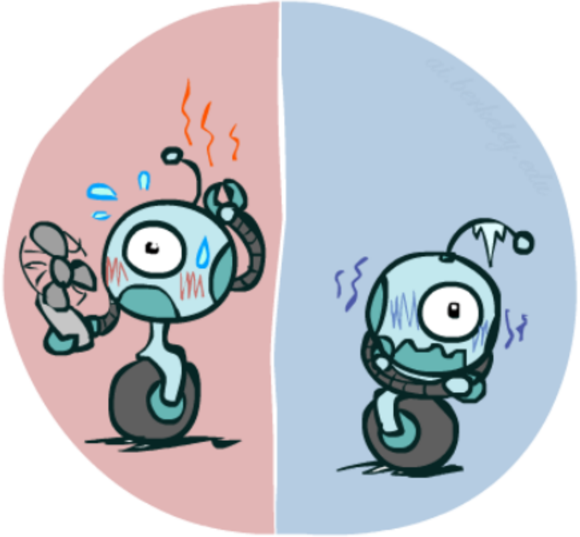
- Joint space of set of variables:
  - $V(Y_1) \times V(Y_2) \times V(Y_3) \ldots \times V(Y_n)$

- Each item in joint space corresponds to one of the possible assignments of values $< Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n >$

- Probability distribution over this joint space is called joint probability distribution

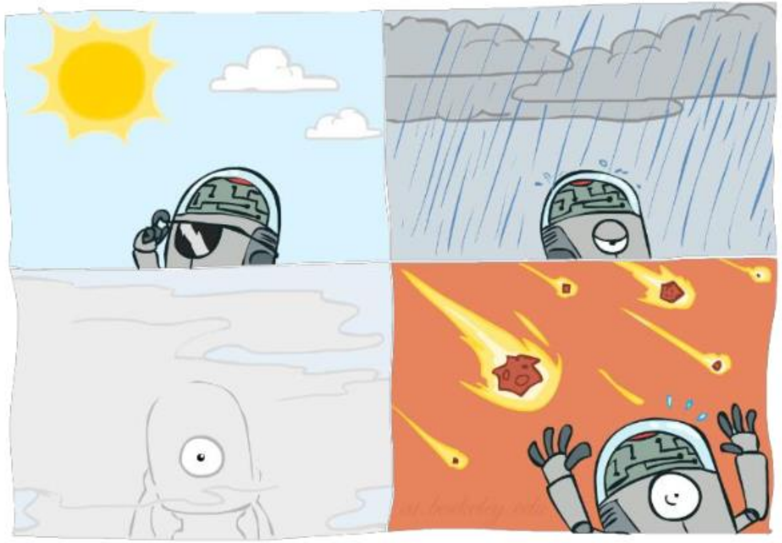
Probability Distributions

- 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 ∈ $\mathcal{S}$, $P(A) \geq 0$

- 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: $X1, X2, \ldots, Xn$

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

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

- 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:
- $P(X = x) = \sum_y P(X = x, Y = y)$

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)

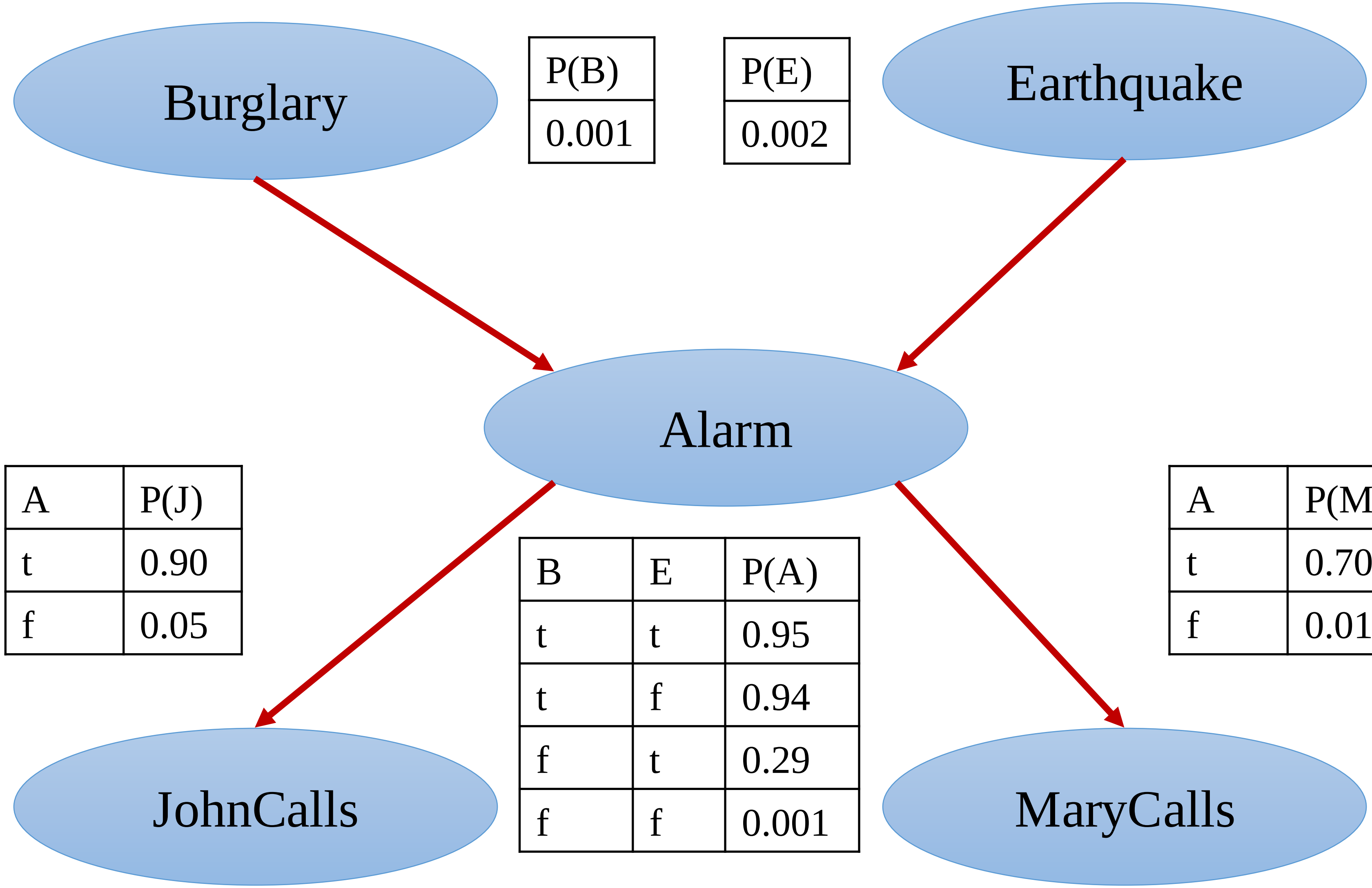
Example

- 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

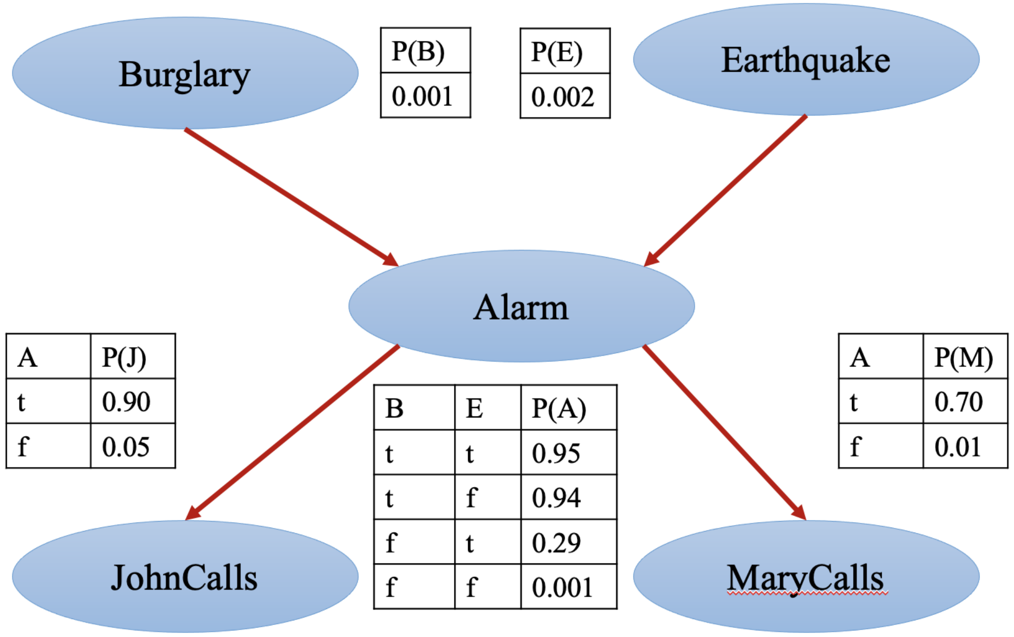
A	P(J)
t	0.90
f	0.05

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

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

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

How to construct this network?

Conditional Independence

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

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

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

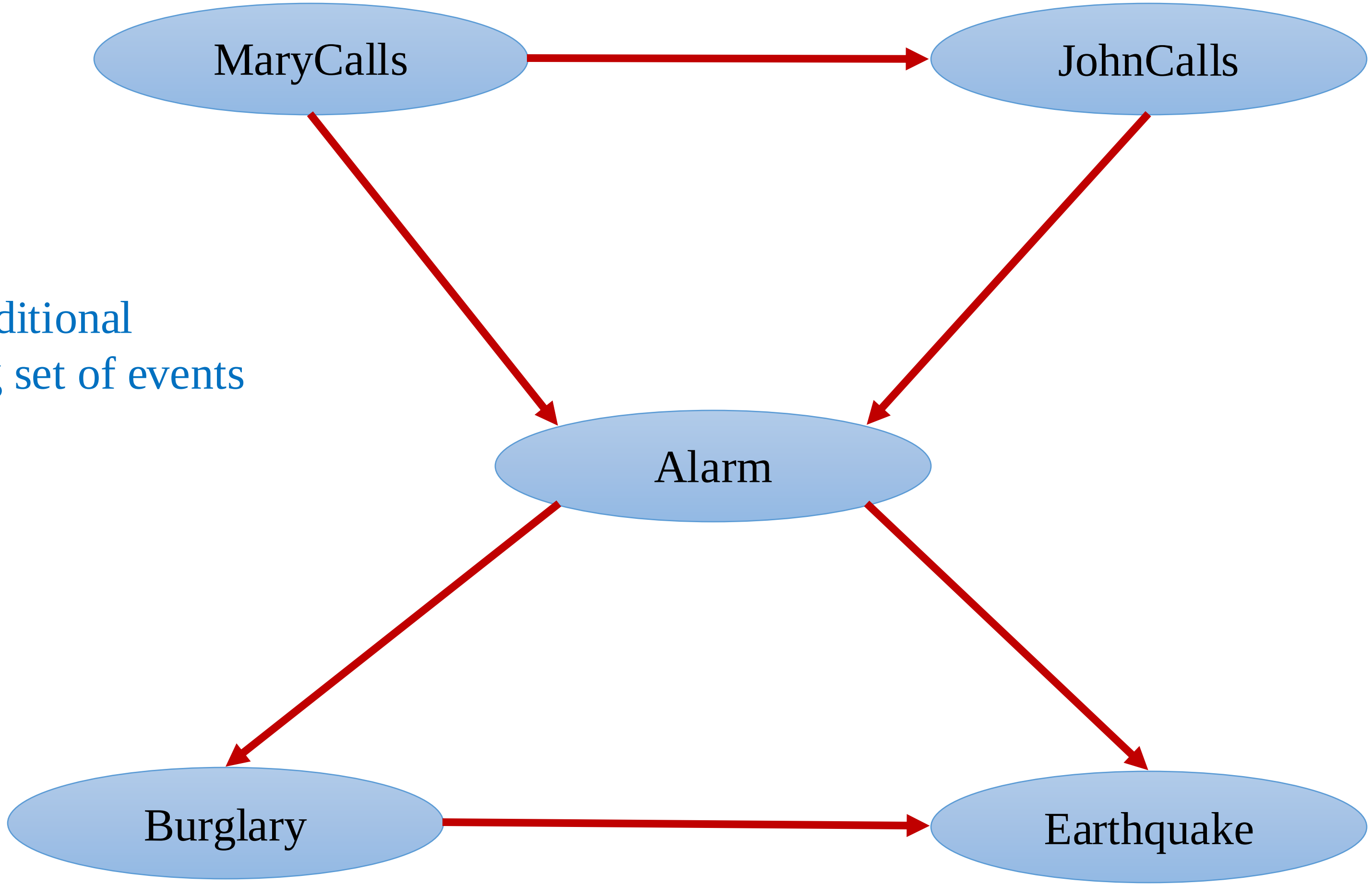
- $P(J, M, A, B, E) = P(J|A)P(M|A)P(A|B, E)P(B)P(E)$

How does ordering matter?

Conditional Independence

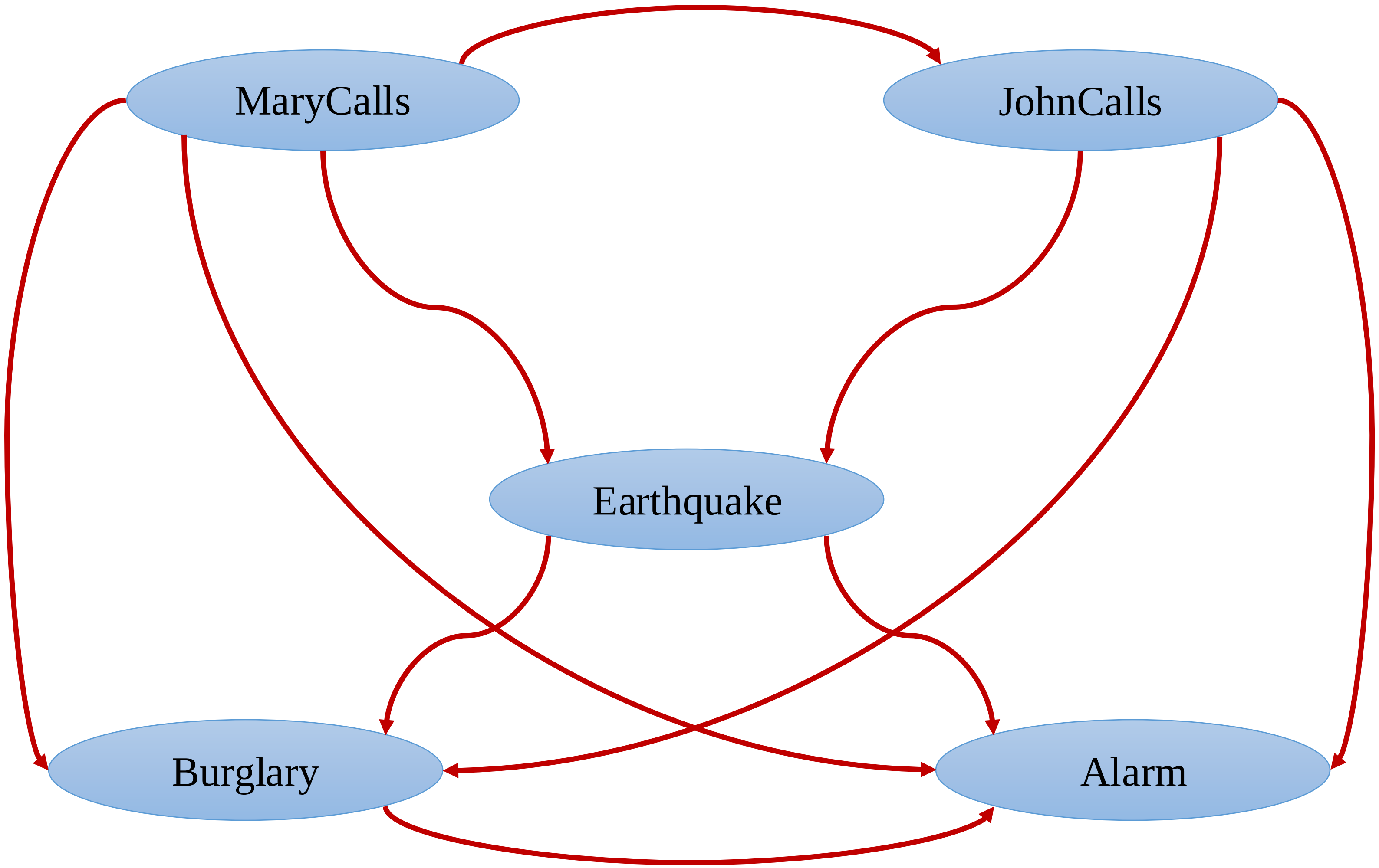
- Earthquake, Burglary, Alarm, JohnCalls, MaryCalls
- P(E|B,A,J,M)

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]

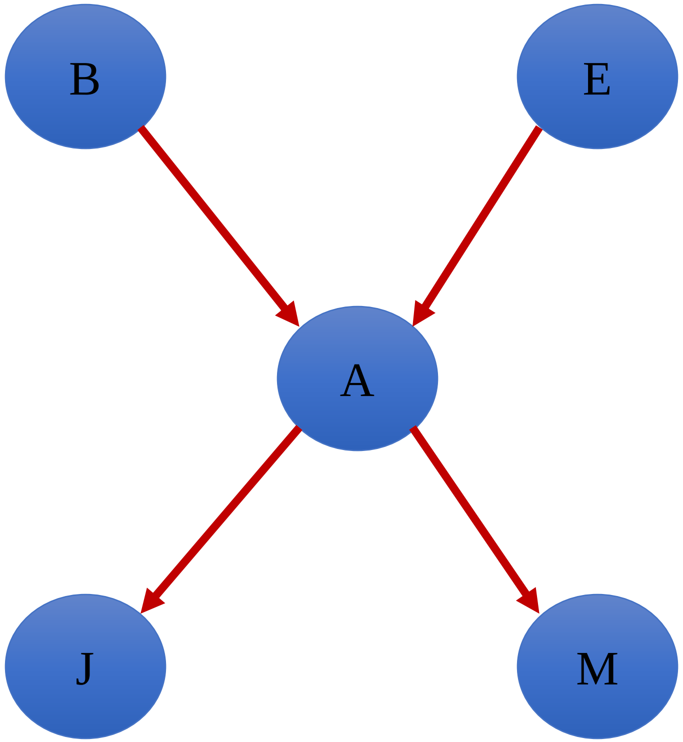
- While there are variables left:
  - Pick a variable X and add a node for it
  - Set Parents(X) to some minimal set of existing nodes such that the conditional independence property is satisfied
  - Define conditional probability table for X

Why do we construct Bayes Network?

To answer queries related to joint probability distribution

Bayesian Network: Topological Semantics

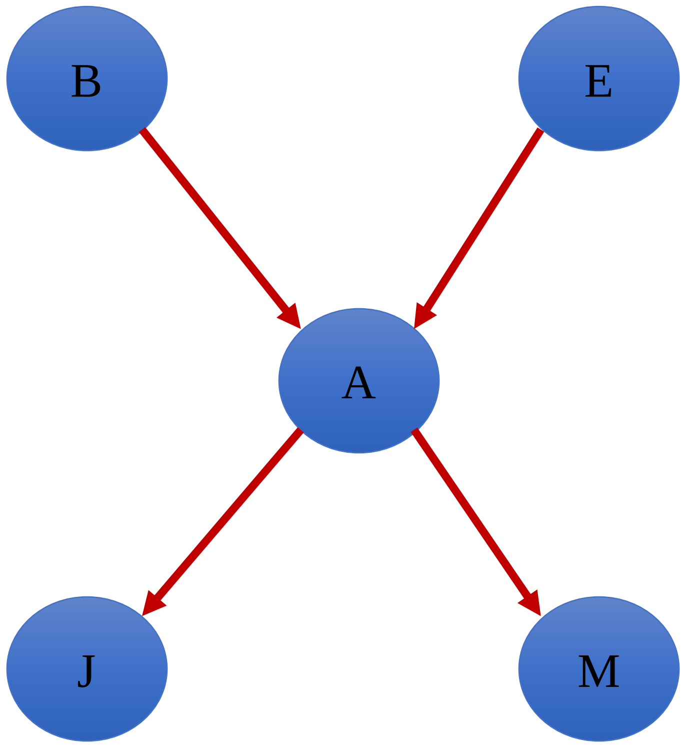
- A node is conditionally independent of its non-descendants, given its parents



JohnCalls is independent of Burglary and Earthquake given the value of Alarm

Bayesian Network: Topological Semantics

- A node is conditionally independent of all other nodes in the network, given its parents, children, and children's parents
  - Markov Blanket



Burglary is independent of JohnCalls and MaryCalls given the value of Alarm and Earthquake

AIFA: Conditional Independence and d-separation

17/03/2025

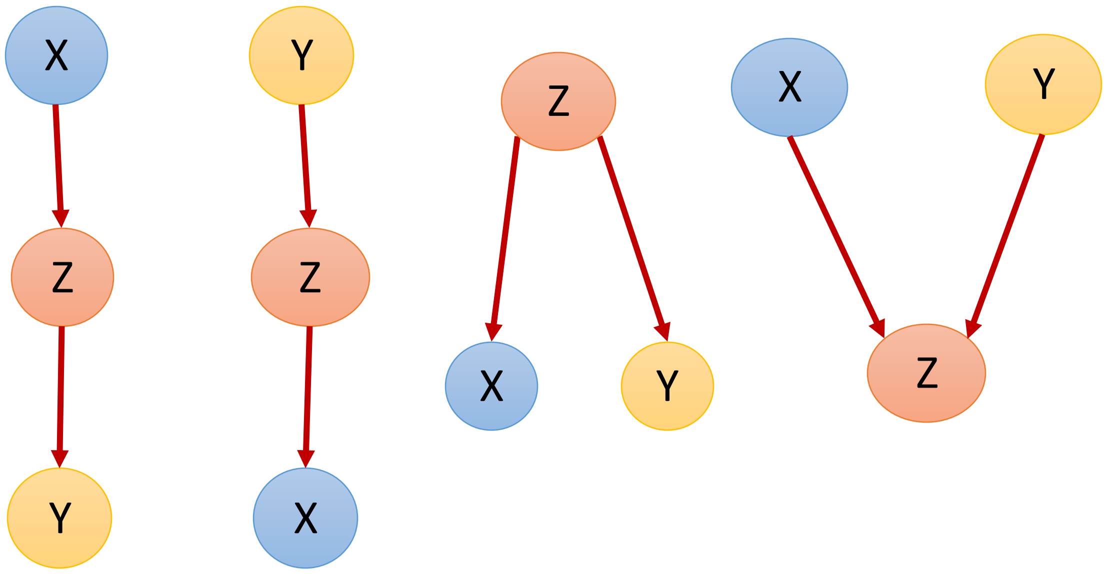
Koustav Rudra

D-separation

- A path is blocked given a set of nodes E if there is a node Z on the path for which one of three conditions hold:
  - Z is in E and Z has one arrow on the path leading in and one arrow out
  - Z is in E and Z has both path arrows leading out
  - Neither Z nor any descendent of Z is in E and both path arrows lead into Z

- If every undirected path from a node in X to a node in Y is d-separated by a given set of evidence nodes E
  - X and Y are conditionally independent given E

- A set of nodes E d-separates two set of nodes X and Y if every undirected path from a node in X to a node in Y is blocked given E



Diagnostic Causal InterCausal Mixed

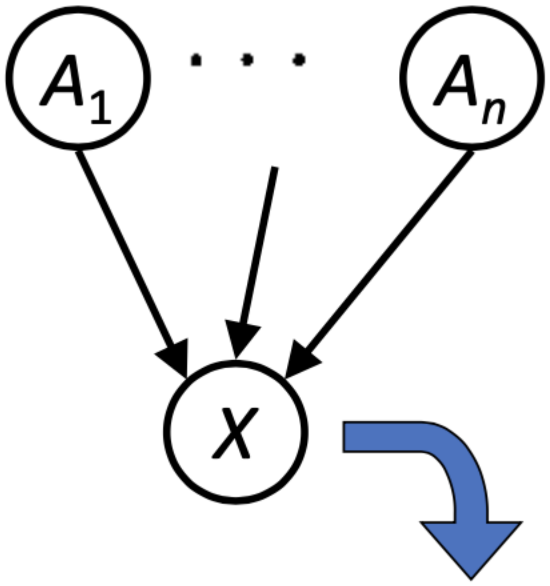
Bayes Nets

- 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 |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

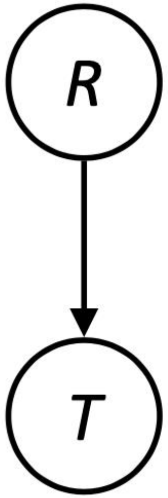
Probabilities in BN

- 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)

Example: Traffic

- Causal direction



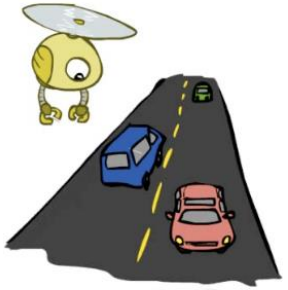
$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

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

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

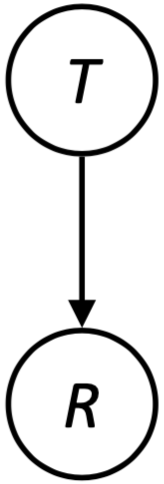


$P(T, R)$

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

Example: 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

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?
  - $O(N * 2^{k+1})$
- 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