Artificial Intelligence: Foundations & Applications

Introduction to Probabilistic Reasoning

Prof. Partha P. Chakrabarti & Arijit Mondal
Indian Institute of Technology Kharagpur

Need for probabilistic model

- Till now, we have considered that intelligent agent has
  - Known environment
  - Full observability
  - Deterministic world
- Reasons for using probability
  - Example: $Toothache \implies Cavity$
    - Toothache can be caused by many other ways also, eg.
      $Toothache \implies GumDisease \lor Cavity \lor WisdomTeeth \lor \ldots$
  - Specifications become too large
  - Theoretical ignorance
  - Practical ignorance

Probabilistic reasoning

- Useful for prediction
  - Analyzing causal effect, predicting outcome
    - Given that I have cavity, what is the chance that I will have toothache?
- Useful for diagnosis
  - Analyzing causal effect, finding out reasons for a given effect
    - Given that I have toothache, what is the chance that it is caused by a cavity?
- Require a methodology to analyze both the scenarios

Axioms of probability

- **All probabilities lie between 0 and 1, ie.,** $0 \leq P(A) \leq 1$
- $P(True) = 1$ **and** $P(False) = 0$
- $P(A \vee B) = P(A) + P(B) - P(A \wedge B)$

- $P(A \wedge B) = P(A|B) \times P(B) = P(B|A) \times P(A)$
  - $P(A) = \dfrac{P(A|B) \times P(B)}{P(B|A)}$
  - **Bayes' rule**

Joint probability

Color \ Type	EV	SUV	Sedan	Truck
Red	0.05	0.20	0.00	0.10
Green	0.10	0.00	0.10	0.00
Black	0.00	0.10	0.05	0.10
White	0.10	0.00	0.10	0.00

- **What is the probability of a vehicle to be EV and Red?**

Joint probability

Color \ Type	EV	SUV	Sedan	Truck
Red	0.05	0.20	0.00	0.10
Green	0.10	0.00	0.10	0.00
Black	0.00	0.10	0.05	0.10
White	0.10	0.00	0.10	0.00

- **What is the probability of a vehicle to be EV and Red?**
- **What is the probability of a Red vehicle?**

Joint probability

Color \ Type	EV	SUV	Sedan	Truck
Red	0.05	0.20	0.00	0.10
Green	0.10	0.00	0.10	0.00
Black	0.00	0.10	0.05	0.10
White	0.10	0.00	0.10	0.00

- **What is the probability of a vehicle to be EV and Red?**
- **What is the probability of a Red vehicle?**
  - **Marginal probability:** $\sum_{T} P(C = Red \wedge T = *)$

Joint probability

Type / Color	EV	SUV	Sedan	Truck
Red	0.05	0.20	0.00	0.10
Green	0.10	0.00	0.10	0.00
Black	0.00	0.10	0.05	0.10
White	0.10	0.00	0.10	0.00

- **What is the probability of a vehicle to be EV and Red?**
- **What is the probability of a Red vehicle?**
  - **Marginal probability:** $\sum_{T} P(C = Red \wedge T = *)$
- **What is the probability of a vehicle to be EV given that it is Red?**

Joint probability

Color \ Type	EV	SUV	Sedan	Truck
Red	0.05	0.20	0.00	0.10
Green	0.10	0.00	0.10	0.00
Black	0.00	0.10	0.05	0.10
White	0.10	0.00	0.10	0.00

- **What is the probability of a vehicle to be EV and Red?**
- **What is the probability of a Red vehicle?**
  - **Marginal probability:** $\sum_{T} P(C = Red \wedge T = *)$
- **What is the probability of a vehicle to be EV given that it is Red?**
  - **Conditional probability:** $\frac{P(C = Red \wedge T = EV)}{P(C = Red)}$

Independence

- Two variables are independent if $P(X, Y) = P(X) \times P(Y)$ holds
  - It means that their joint distribution factors into a product two distributions
  - This can be expressed as $P(X|Y) = P(X)$
- Independence is a simplifying modeling assumption
  - Empirical joint distributions can be at best close to independent

Example: Independence

T	W	P(T,W)
Hot	Sun	0.4
Hot	Rain	0.1
Cold	Sun	0.2
Cold	Rain	0.3

Example: Independence

T	W	P(T,W)
Hot	Sun	0.4
Hot	Rain	0.1
Cold	Sun	0.2
Cold	Rain	0.3

- **Are *T* and *W* independent?**

Example: Independence

T	W	P(T,W)
Hot	Sun	0.4
Hot	Rain	0.1
Cold	Sun	0.2
Cold	Rain	0.3

- **Are *T* and *W* independent?**
- **Find marginal probabilities**

Example: Independence

T	W	P(T,W)
Hot	Sun	0.4
Hot	Rain	0.1
Cold	Sun	0.2
Cold	Rain	0.3

- **Are T and W independent?**
- **Find marginal probabilities**
  - $P(T = Hot) = ?$, $P(T = Cold) = ?$
  - $P(W = Sun) = ?$, $P(W = Rain) = ?$

Example: Independence

T	W	P(T,W)
Hot	Sun	0.4
Hot	Rain	0.1
Cold	Sun	0.2
Cold	Rain	0.3

- **Are T and W independent?**
- **Find marginal probabilities**
  - $P(T = Hot) = ?$, $P(T = Cold) = ?$
  - $P(W = Sun) = ?$, $P(W = Rain) = ?$
- **Now check for independence**

Example: Independence

- **Tossing of N fair coins**

$P(X_1)$	
H	**0.5**
T	**0.5**

$P(X_2)$	
H	**0.5**
T	**0.5**

$\ldots\ldots$

$P(X_n)$	
H	**0.5**
T	**0.5**

$P(X_1, X_2, \ldots, X_n)$

X_1	X_2	X_3	$\ldots$	$\ldots$	X_n	P
H					**H**	
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	
T					**H**	

$\}\, 2^n$

Chain rule

- **Product rule**

$$P(X, Y) = P(X) \times P(Y|X)$$

- **Chain rule**

$$\begin{aligned} P(X_1, X_2, \ldots, X_n) &= P(X_1) \times P(X_2|X_1) \times P(X_3|X_1, X_2) \times \ldots \times P(X_N|X_1, \ldots, X_{n-1}) \\ &= \prod_i P(X_i|X_1, \ldots, X_{i-1})) \end{aligned}$$

Conditional independence

- $P(Toothache, Cavity, Catch)$
- **If I have a cavity, the probability that the probe catches in it doesn't depend on whether I have a toothache:**

  $P(+catch| + toothache, +cavity) = P(+catch| + cavity)$

- **The same independence holds if I don't have a cavity:**

  $P(+catch| + toothache, -cavity) = P(+catch| - cavity)$

- **Catch is conditionally independent of Toothache given Cavity:**

  $P(Catch|Toothache, Cavity) = P(Catch|Cavity)$

Conditional independence

- **Absolute independence is very rare**
- **Conditional independence is the most basic and robust form of knowledge about uncertain environments**
- **Mathematically it is defined as -** X **is conditionally independent of** Y **given** Z

  $$P(X, Y|Z) = P(X|Z) \times P(Y|Z)$$

- **It can also be shown that**

  $$P(X|Z, Y) = P(X|Z)$$

Belief network

- Qualitative information
  - A belief network is a graph consist of the following:
  - Nodes – Set of random variables
  - Edges – Dependency of nodes. $X \longrightarrow Y$ means X has direct influence on Y
- Quantitative information
  - Each node has conditional probability table that quantifies the effects the parents have on the node

- The network is a directed acyclic graph (DAG), ie., without any cycle

Example

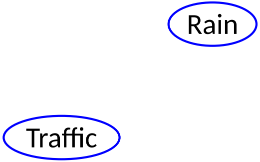
- **Consider the following knowledge base**
  - *Rain* **- Raining**
  - *Traffic* **- There may be traffic if it rains**
  - *Umbrella* **- People may carry umbrella if it rains**

Example

- **Consider the following knowledge base**
  - *Rain* **- Raining**
  - *Traffic* **- There may be traffic if it rains**
  - *Umbrella* **- People may carry umbrella if it rains**

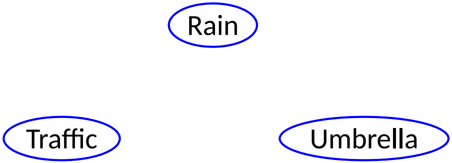
Example

- **Consider the following knowledge base**
  - *Rain* - **Raining**
  - *Traffic* - **There may be traffic if it rains**
  - *Umbrella* - **People may carry umbrella if it rains**



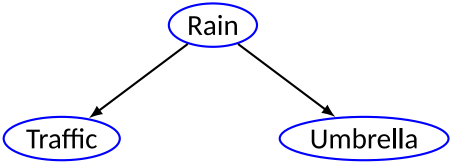
Example

- **Consider the following knowledge base**
  - *Rain* **- Raining**
  - *Traffic* **- There may be traffic if it rains**
  - *Umbrella* **- People may carry umbrella if it rains**



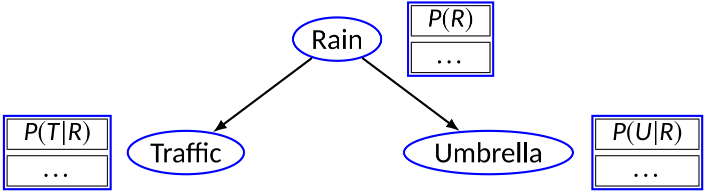
Example

- **Consider the following knowledge base**
  - *Rain* **- Raining**
  - *Traffic* **- There may be traffic if it rains**
  - *Umbrella* **- People may carry umbrella if it rains**

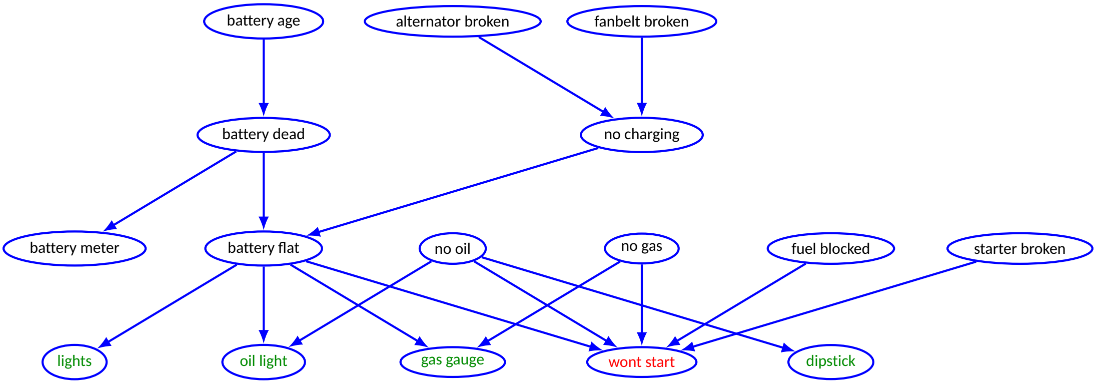


Example

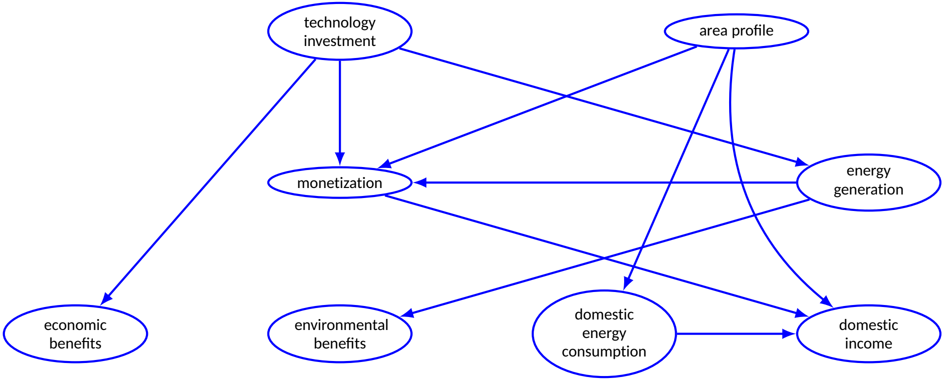
- **Consider the following knowledge base**
  - *Rain* - **Raining**
  - *Traffic* - **There may be traffic if it rains**
  - *Umbrella* - **People may carry umbrella if it rains**



Belief network example: car



Belief network example: renewable energy



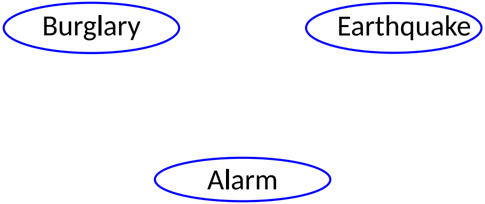
Classical example

- Burglar alarm at a home
  - It works almost perfectly
  - Sometime alarm goes off if there is a earthquake
- John calls police when he hears the alarm but sometimes he misinterpret telephone ringing as alarm and calls police too.
- Mary also calls police. But she loves loud music, therefore, she misses the alarm sometime.

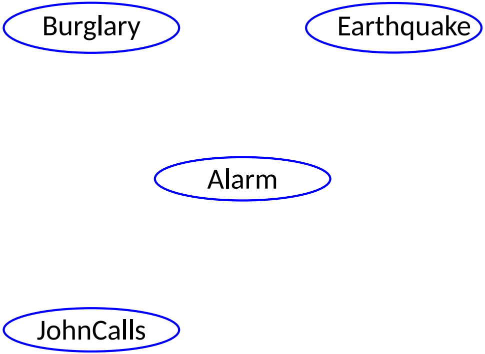
Belief network example

Belief network example

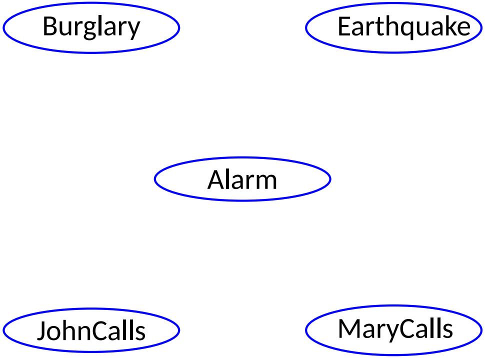
Belief network example



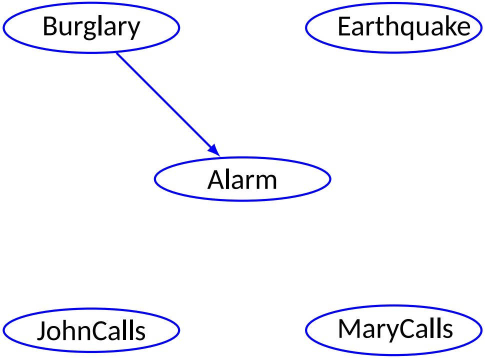
Belief network example



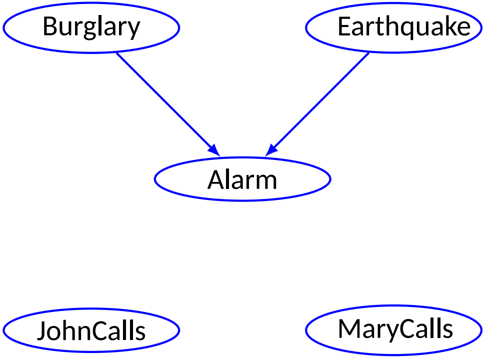
Belief network example



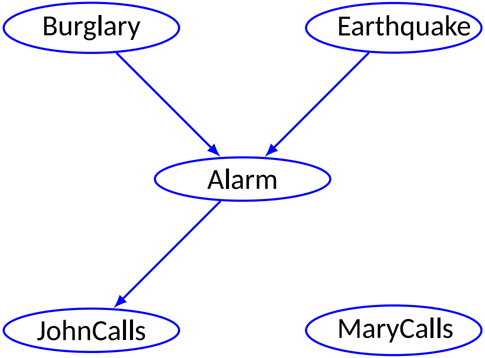
Belief network example



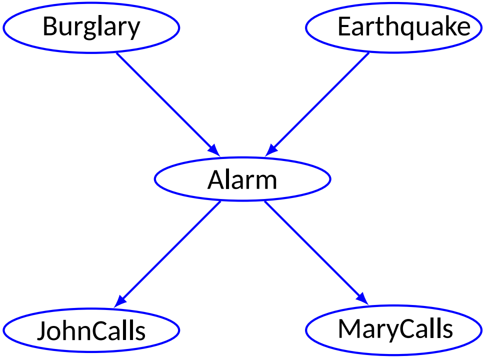
Belief network example



Belief network example

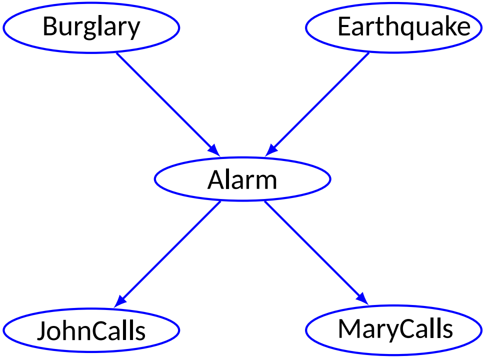


Belief network example

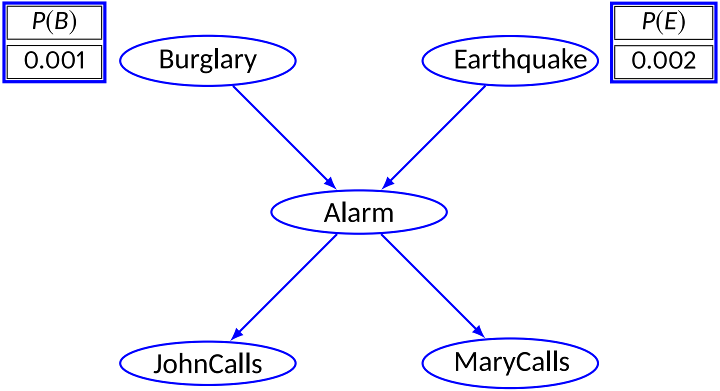


Belief network example

$P(B)$
0.001



Belief network example

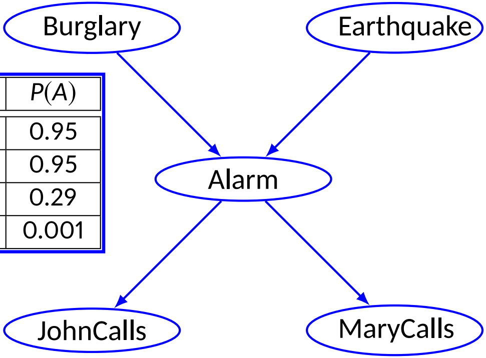


Belief network example

$P(B)$
0.001

$P(E)$
0.002

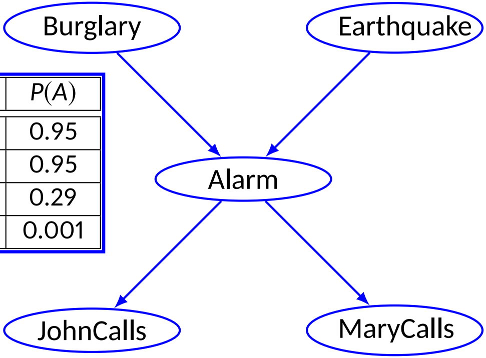
B	E	$P(A)$
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001



Belief network example

$P(B)$
0.001

$P(E)$
0.002



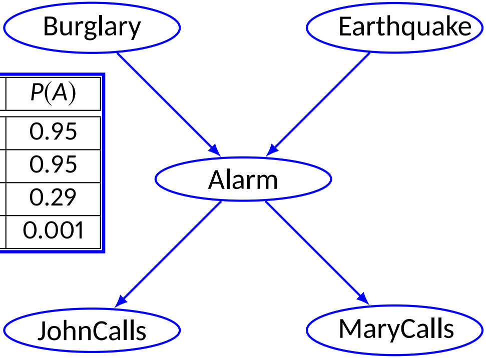
B	E	$P(A)$
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	$P(J)$
T	0.90
F	0.05

Belief network example

$P(B)$
0.001

$P(E)$
0.002



B	E	$P(A)$
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	$P(J)$
T	0.90
F	0.05

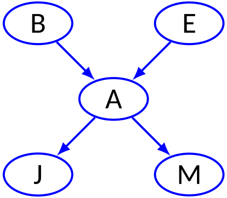
A	$P(M)$
T	0.70
F	0.01

Probabilities in Belief Nets

- **Bayes' net implicitly encode joint distributions**
  - **This can be determined from local conditional distributions**
  - **Need to multiply relevant conditional probabilities**
  - $P(X_1, X_2, \ldots, X_n) = \prod_{i}^{n} P(X_i | parents(X_i))$

Joint probability distribution

- **Probability of the event that the alarm has sounded but neither a burglary nor an earthquake has occurred, and both Mary and John call:**



$P(B)$
0.001

$P(E)$
0.002

B	E	$P(A)$
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

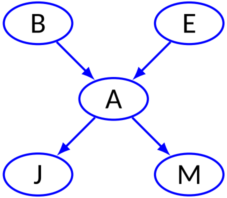
A	$P(J)$
T	0.90
F	0.05

A	$P(M)$
T	0.70
F	0.01

Joint probability distribution

- **Probability of the event that the alarm has sounded but neither a burglary nor an earthquake has occurred, and both Mary and John call:**

  $P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$



$P(B)$
0.001

$P(E)$
0.002

B	E	$P(A)$
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

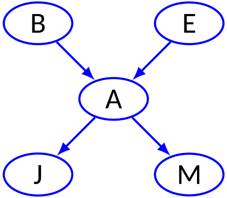
A	$P(J)$
T	0.90
F	0.05

A	$P(M)$
T	0.70
F	0.01

Joint probability distribution

- **Probability of the event that the alarm has sounded but neither a burglary nor an earthquake has occurred, and both Mary and John call:**

$$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E)$$
$$= P(J|A) \times P(M|A) \times P(A|\neg B \wedge \neg E) \times P(\neg B) \times P(\neg E)$$
$$= 0.9 \times 0.7 \times 0.001 \times 0.999 \times 0.998$$
$$= 0.00062$$



$P(B)$
0.001

$P(E)$
0.002

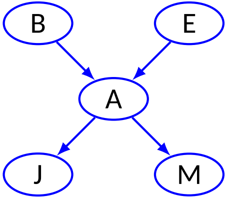
B	E	$P(A)$
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	$P(J)$
T	0.90
F	0.05

A	$P(M)$
T	0.70
F	0.01

Joint probability distribution: $P(A)$

$P(A)$



$P(B)$
0.001

$P(E)$
0.002

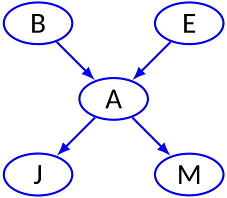
B	E	$P(A)$
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	$P(J)$
T	0.90
F	0.05

A	$P(M)$
T	0.70
F	0.01

Joint probability distribution: $P(A)$

$$P(A) = P(A\bar{B}\bar{E}) + P(A\bar{B}E) + P(AB\bar{E}) + P(ABE)$$



$P(B)$
0.001

$P(E)$
0.002

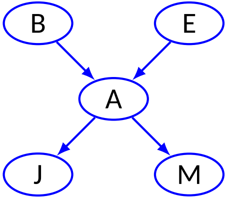
B	E	$P(A)$
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	$P(J)$
T	0.90
F	0.05

A	$P(M)$
T	0.70
F	0.01

Joint probability distribution: $P(A)$

$$
\begin{aligned}
P(A) \\
&= P(A\bar{B}\bar{E}) + P(A\bar{B}E) + P(AB\bar{E}) + P(ABE) \\
&= P(A|\bar{B}\bar{E}) \times P(\bar{B}\bar{E}) + P(A|\bar{B}E) \times P(\bar{B}E) + P(A|B\bar{E}) \times P(B\bar{E}) + P(A|BE) \times P(BE)
\end{aligned}
$$



$P(B)$
0.001

$P(E)$
0.002

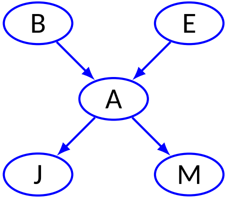
B	E	$P(A)$
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	$P(J)$
T	0.90
F	0.05

A	$P(M)$
T	0.70
F	0.01

Joint probability distribution: $P(A)$

$$
\begin{aligned}
P(A) \\
&= P(A\bar{B}\bar{E}) + P(A\bar{B}E) + P(AB\bar{E}) + P(ABE) \\
&= P(A|\bar{B}\bar{E}) \times P(\bar{B}\bar{E}) + P(A|\bar{B}E) \times P(\bar{B}E) + P(A|B\bar{E}) \times P(B\bar{E}) + P(A|BE) \times P(BE) \\
&= 0.001 \times 0.999 \times 0.998 + 0.29 \times 0.999 \times 0.002 + 0.95 \times 0.001 \times 0.998 \\
&\quad + 0.95 \times 0.001 \times 0.002 \\
&= 0.0025
\end{aligned}
$$



$P(B)$
0.001

$P(E)$
0.002

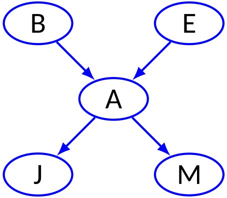
B	E	$P(A)$
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	$P(J)$
T	0.90
F	0.05

A	$P(M)$
T	0.70
F	0.01

Joint probability distribution: $P(J)$

$P(J)$



$P(B)$
0.001

$P(E)$
0.002

B	E	$P(A)$
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

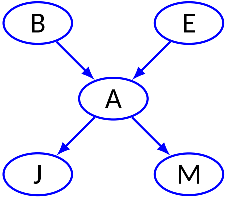
A	$P(J)$
T	0.90
F	0.05

A	$P(M)$
T	0.70
F	0.01

Joint probability distribution: $P(J)$

$P(J)$

$= P(JA) + P(J\bar{A})$



$P(B)$
0.001

$P(E)$
0.002

B	E	$P(A)$
T	T	0.95
T	F	0.95
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: $P(J)$

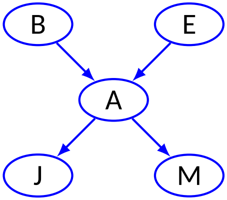
$P(J)$

$= P(JA) + P(J\bar{A})$

$= P(J|A) \times P(A) + P(J|\bar{A}) \times P(\bar{A})$

$= 0.9 \times 0.0025 + 0.05 \times (1 - 0.0025)$

$= 0.052125$



$P(B)$
0.001

$P(E)$
0.002

B	E	$P(A)$
T	T	0.95
T	F	0.95
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: $P(J)$

$P(J)$

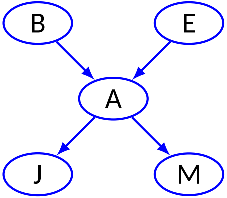
$= P(JA) + P(J\bar{A})$

$= P(J|A)\times P(A) + P(J|\bar{A})\times P(\bar{A})$

$= 0.9\times 0.0025 + 0.05\times(1 - 0.0025)$

$= 0.052125$

$P(AB)$



$P(B)$
0.001

$P(E)$
0.002

B	E	$P(A)$
T	T	0.95
T	F	0.95
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: $P(J)$

$P(J)$

$= P(JA) + P(J\bar{A})$

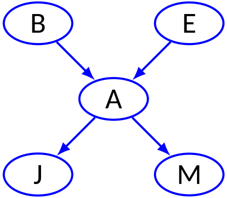
$= P(J|A) \times P(A) + P(J|\bar{A}) \times P(\bar{A})$

$= 0.9 \times 0.0025 + 0.05 \times (1 - 0.0025)$

$= 0.052125$

$P(AB)$

$= P(ABE) + P(AB\bar{E})$



$P(B)$
0.001

$P(E)$
0.002

B	E	$P(A)$
T	T	0.95
T	F	0.95
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: $P(J)$

$P(J)$

$= P(JA) + P(J\bar{A})$

$= P(J|A) \times P(A) + P(J|\bar{A}) \times P(\bar{A})$

$= 0.9 \times 0.0025 + 0.05 \times (1 - 0.0025)$

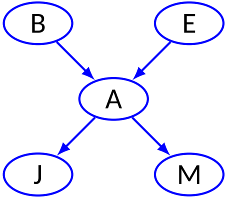
$= 0.052125$

$P(AB)$

$= P(ABE) + P(AB\bar{E})$

$= 0.95 \times 0.001 \times 0.002 + 0.95 \times 0.001 \times 0.998$

$= 0.00095$



$P(B)$
0.001

$P(E)$
0.002

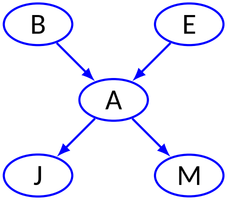
B	E	$P(A)$
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	$P(J)$
T	0.90
F	0.05

A	$P(M)$
T	0.70
F	0.01

Joint probability distribution: $P(JB)$

$P(JB)$



$P(B)$
0.001

$P(E)$
0.002

B	E	$P(A)$
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

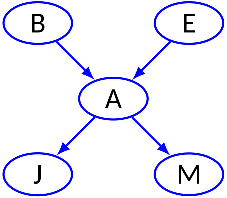
A	$P(J)$
T	0.90
F	0.05

A	$P(M)$
T	0.70
F	0.01

Joint probability distribution: $P(JB)$

$P(JB)$

$= P(JBA) + P(JB\bar{A})$



$P(B)$
0.001

$P(E)$
0.002

B	E	$P(A)$
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	$P(J)$
T	0.90
F	0.05

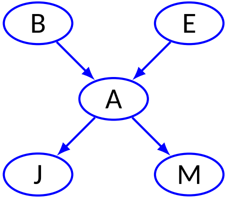
A	$P(M)$
T	0.70
F	0.01

Joint probability distribution: $P(JB)$

$P(JB)$

$= P(JBA) + P(JB\bar{A})$

$= P(J|AB) \times P(AB) + P(J|\bar{A}B) \times P(\bar{A}B)$



$P(B)$
0.001

$P(E)$
0.002

B	E	$P(A)$
T	T	0.95
T	F	0.95
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: $P(JB)$

$P(JB)$

$= P(JBA) + P(JB\bar{A})$

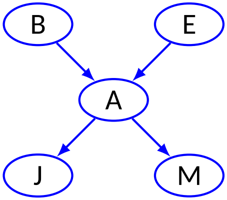
$= P(J|AB) \times P(AB) + P(J|\bar{A}B) \times P(\bar{A}B)$

$= P(J|A) \times P(AB) + P(J|\bar{A}) \times P(\bar{A}B)$

$= 0.9 \times 0.00095 + 0.05 \times 0.00005$

$= 0.00086$

$P(J|B)$



$P(B)$
0.001

$P(E)$
0.002

B	E	$P(A)$
T	T	0.95
T	F	0.95
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: $P(JB)$

$P(JB)$

$$= P(JBA) + P(JB\bar{A})$$

$$= P(J|AB) \times P(AB) + P(J|\bar{A}B) \times P(\bar{A}B)$$

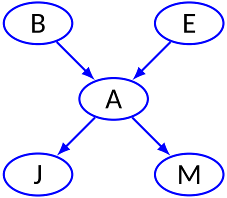
$$= P(J|A) \times P(AB) + P(J|\bar{A}) \times P(\bar{A}B)$$

$$= 0.9 \times 0.00095 + 0.05 \times 0.00005$$

$$= 0.00086$$

$P(J|B)$

$$= \frac{P(JB)}{P(B)} = \frac{0.00086}{0.001} = 0.86$$



$P(B)$
0.001

$P(E)$
0.002

B	E	$P(A)$
T	T	0.95
T	F	0.95
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

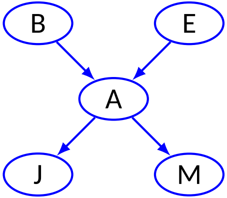
Inferences using belief networks

- **Diagnostic inferences (effects to causes)**
- **Given that JohnCalls, infer that** $P(Burglary|JohnCalls) = 0.016$

- **Causal inferences (causes to effects)**
- **Given Burglary, infer that**

  $P(JohnCalls|Burglary) = 0.86$

  $P(MaryCalls|Burglary) = 0.67$



$P(B)$
0.001

$P(E)$
0.002

B	E	$P(A)$
T	T	0.95
T	F	0.95
F	T	0.29
F	F	0.001

A	$P(J)$
T	0.90
F	0.05

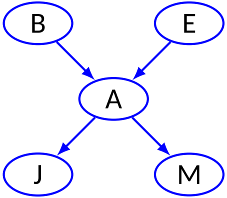
A	$P(M)$
T	0.70
F	0.01

Inferences using belief networks

- **Inter-causal inferences (between causes to a common effect)**
- **Given Alarm, we have** $P(Burglary|Alarm) = 0.376$
- **Also, if it is given that Earthquake is true, then**
  $P(Burglary|Alarm \wedge Earthquake) = 0.003$

- **Mixed inferences**

  $P(Alarm|JohnCalls \wedge \neg Earthquake) = 0.003$



$P(B)$
0.001

$P(E)$
0.002

B	E	$P(A)$
T	T	0.95
T	F	0.95
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

Thank you!