

# Uncertainty II

---

CS3243: Introduction to Artificial Intelligence – Lecture 10

27 March 2023



# Contents

---

1. Uncertainty Fundamentals
2. Bayesian Inference
3. Bayesian Networks
4. Decision-Theoretic Agents

Reference: AIMA 4<sup>th</sup> Edition, Section 12; 13.1-13.2



# Uncertainty Fundamentals

---



# Recap on Conditional Probabilities & Bayes' Rule

- $\Pr[A | B] = \frac{\Pr[A \wedge B]}{\Pr[B]}$  assuming that  $\Pr[B] > 0$

Note:

$$\Pr[A | B] = \Pr[A \wedge B] / \Pr[B] \text{ --- (1)}$$

$$\Pr[B | A] = \Pr[B \wedge A] / \Pr[A] \text{ --- (2)}$$

Substituting (3) into (2), we have:

$$\Pr[A \wedge B] = \Pr[B | A] \cdot \Pr[A] \text{ --- (4)}$$

Also, we know:

$$\Pr[B \wedge A] = \Pr[A \wedge B] \text{ --- (3)}$$

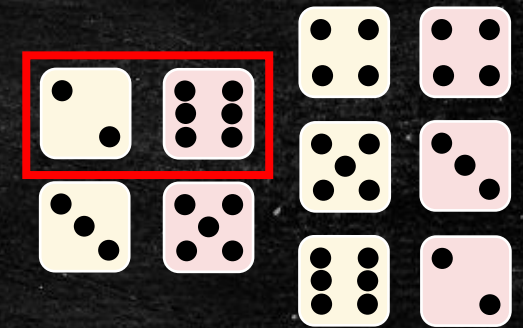
Substituting (4) into (1), we have Bayes' Rule:

$$\Pr[A|B] = (\Pr[B|A] \cdot \Pr[A]) / \Pr[B]$$

- Bayes' rule:  $\Pr[A | B] = \frac{\Pr[B|A] \Pr[A]}{\Pr[B]}$

- Example:  $\Pr[X_1 = 2 | X_1 + X_2 = 8] = \frac{\overset{1/6}{\Pr[X_1 + X_2 = 8 | X_1 = 2]} \cdot \overset{1/6}{\Pr[X_1 = 2]}}{\underset{5/36}{\Pr[X_1 + X_2 = 8]}} = \frac{1}{5}$

$X_1 = a$ : Die 1 rolls a  
 $X_2 = b$ : Die 2 rolls b  
 $a, b \in \{1, 2, 3, 4, 5, 6\}$





# Chain Rule

- With more than two events, we have

$$\begin{aligned}\Pr[R_1 \wedge \cdots \wedge R_k] &= \Pr[(R_1 \wedge \cdots \wedge R_{k-1}) \wedge R_k] \\ &= \Pr[R_k \mid R_{k-1} \wedge \cdots \wedge R_1] \cdot \Pr[R_{k-1} \wedge \cdots \wedge R_1]\end{aligned}$$

$$\Pr[A \mid B] = \frac{\Pr[A \wedge B]}{\Pr[B]}$$

- And by induction, we have

$$\Pr[R_1 \wedge R_2 \wedge \cdots \wedge R_k] = \prod_{j=1, \dots, k} \Pr[R_j \mid R_1 \wedge \cdots \wedge R_{j-1}]$$

- Example:

$$\begin{aligned}\Pr[A \wedge B \wedge C \wedge D] &= \Pr[D \mid C \wedge B \wedge A] \cdot \Pr[C \wedge B \wedge A] \\ &= \Pr[D \mid C \wedge B \wedge A] \cdot \Pr[C \mid B \wedge A] \cdot \Pr[B \wedge A] \\ &= \Pr[D \mid C \wedge B \wedge A] \cdot \Pr[C \mid B \wedge A] \cdot \Pr[B \mid A] \cdot \Pr[A]\end{aligned}$$



# Independence

- $A$  and  $B$  are independent if  $\Pr[A \wedge B] = \Pr[A] \cdot \Pr[B]$
- Equivalent to  $\Pr[A | B] = \Pr[A]$
- Knowing  $B$  adds no information about  $A$ 
  - $B$  does not further categorise or classify  $A$  into sub-categories
- Example: rolling two dice

Recall that:

$$\Pr[A|B] = \Pr[A \wedge B] / \Pr[B]$$

$$\Pr[B|A] = \Pr[B \wedge A] / \Pr[A]$$

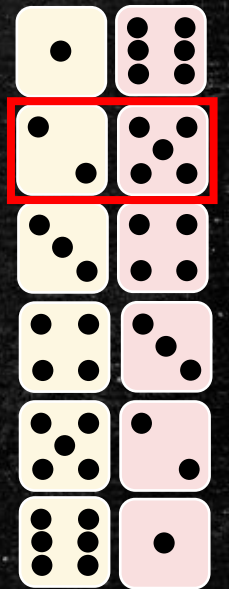
$$\Pr[X_1 = 2 | X_1 + X_2 = 7] = ?$$

$$= \frac{\Pr[X_1=2 \wedge X_1+X_2=7]}{\Pr[X_1+X_2=7]} = \frac{1/36}{6/36} = \frac{1}{6}$$

Since  $\Pr[X_1 = 2] = 1/6$

$[X_1 = 2]$  and  $[X_1 + X_2 = 7]$  are independent

Knowing  $[X_1 + X_2 = 7]$  does quantify the probability of  $[X_1 = 2]$  differently – no new information about  $[X_1 = 2]$  is added





# Bayesian Inference

---



# Performing Inference via Bayes' Rule

---

- Instead of inferring statements in the form

“is  $\alpha$  true given  $KB$ ?”

i.e.,  $R_1 \wedge \dots \wedge R_k \Rightarrow \alpha$ ?

- We infer statement of the form

“What is the likelihood of an event  $\alpha$  given the probabilities of other events?”

i.e.,  $\Pr[\alpha \mid R_1 \wedge \dots \wedge R_k] = ?$

Naturally occurs in everyday situations ...  
... e.g.,  $\Pr[\text{COVID-19} \mid \text{Fever} \wedge \text{Cough} \wedge \dots]$



# Inference by Enumeration

- Assuming we have the joint probability distribution

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- For any proposition (event)  $X$ , sum the atomic events  $y$  where  $X$  holds (i.e., calculate the **unconditional** or **marginal probability** of  $X$ ):  
$$\Pr[X] = \sum_{y \in X} \Pr[X = y]$$
- $\Pr[\text{toothache}] = 0.108 + 0.016 + 0.012 + 0.064 = 0.2$



# Inference by Enumeration

- Assuming we have the joint probability distribution

	toothache		¬toothache	
	catch	¬catch	catch	¬catch
cavity	0.108	0.012	0.072	0.008
¬cavity	0.016	0.064	0.144	0.576

- For any proposition (event)  $X$ , sum the atomic events  $y$  where  $X$  holds (i.e., calculate the **unconditional** or **marginal probability** of  $X$ ):  
 $\Pr[X] = \sum_{y \in X} \Pr[X = y]$

- $$\Pr[\neg \text{cavity} \mid \text{toothache}] = \frac{\Pr[\neg \text{cavity} \wedge \text{toothache}]}{\Pr[\text{toothache}]}$$
$$= \frac{0.016 + 0.064}{0.2} = 0.4$$

$$\begin{aligned}\Pr[\text{toothache}] &= 0.108 + 0.016 + 0.012 + 0.064 \\ &= 0.2\end{aligned}$$



# Power of Independence

- We have  $n$  random variables,  $X_1, \dots, X_n$ , with domains of size  $d$ 
  - How big is their joint distribution table?
    - $\underbrace{d \times d \times \dots \times d}_{n \text{ times}} = d^n$
- Suppose that the  $n$  random variables,  $X_1, \dots, X_n$ , are independent
  - How big is the joint distribution table now?
    - $\underbrace{d + d + \dots + d}_{n \text{ times}} = dn$

If  $A$  and  $B$  are independent:  
 $\Pr[A \wedge B] = \Pr[A] \cdot \Pr[B]$   
 $\Pr[A | B] = \Pr[A]$

We no longer need to know the joint probabilities – e.g.,  $\Pr[A | B]$



# General Idea Behind Bayesian Networks

---

- Independence is good (if we can find it)
  - Less information (i.e., probabilities) to determine and store
  - Less to enumeration in order to determine probabilities
- Bayesian Networks try to work with some independence



# Conditional Independence & Bayes' Rule

---



# Conditional Independence

- Suppose that we test for COVID-19 using two tests

- Antigen Rapid Test (ART):  $A \in \{T, F\}$
- Breathalyser Test:  $B \in \{T, F\}$

- Are they fully independent?

- Tests were conducted independently
- Only related by the underlying sickness

As both tests are taken by the same patient, the outcomes of both tests are dependent on whether that patient is ill

But when we assume a patient is ill, the associated probabilities of both tests (on patients who are ill) are now independent!

- $A$ ,  $B$  are independent *given* knowledge of underlying cause  $S$  = sick!

- $\Pr[A \wedge B | S] = \Pr[A | S] \cdot \Pr[B | S]$



# Conditional Independence

- A joint distribution for  $n$  Boolean random variables results in  $2^n - 1$  entries

Note that we have  $2^n - 1$  since all the values sum to 1 and we can deduce the last value from the rest

- With conditional independence, writing out a full joint distribution using the chain rule becomes

$$\Pr[ T_1 \wedge \cdots \wedge T_{n-1} \wedge S ] = \Pr[ T_1 | S ] \cdot \Pr[ T_2 | S ] \cdot \dots \cdot \Pr[ T_{n-1} | S ] \cdot \Pr[ S ]$$

Notice that we need to store  $n - 1$  conditional probabilities for each value of  $S$  (assuming each  $T_i$  and  $S$  are Booleans), and thus  $2(n - 1)$  conditional probabilities

- With conditional independence: linear
  - From the example above:  $2(n - 1) + 1$
- Conditional independence is more robust and common than absolute independence

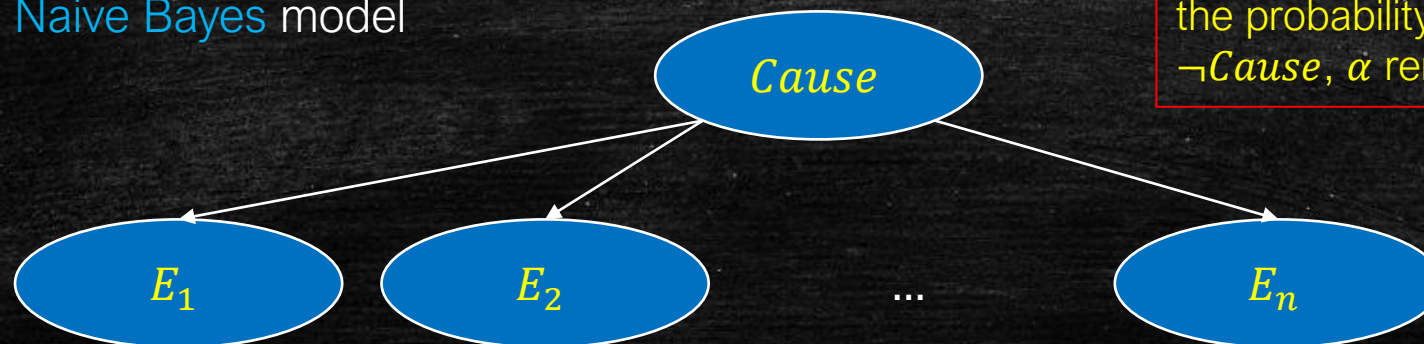


# Bayes' Rule & Conditional Independence

- A cause can have several conditionally independent effects
  - Cause: heavy rain
  - Conditionally independent effects: Alice brings umbrella, Bob brings umbrella, ...

$$\begin{aligned}\Pr[Cause \mid E_1, \dots, E_n] &= \frac{\Pr[Cause] \Pr[E_1, \dots, E_n \mid Cause]}{\Pr[E_1, \dots, E_n]} \\ &= \frac{\Pr[Cause]}{\Pr[E_1, \dots, E_n]} \cdot \prod_i \Pr[E_i \mid Cause] = \alpha \cdot \Pr[Cause] \cdot \prod_i \Pr[E_i \mid Cause]\end{aligned}$$

- This is the Naive Bayes model



When comparing, for example, the probability of *Cause* and  $\neg Cause$ ,  $\alpha$  remains constant



# Normalisation under Naive Bayes Algorithm

- Example
  - Suppose we are trying to diagnose a disease in a patient  $X$ 
    - 70% of the population is *healthy*
    - 20% are *carrier(s)*
    - 10% are *sick*
  - A test will come back *positive* with the following probability
    - $\Pr[\text{Test}(X) = \text{positive} \mid X = \text{healthy}] = 0.1$
    - $\Pr[\text{Test}(X) = \text{positive} \mid X = \text{carrier}] = 0.7$
    - $\Pr[\text{Test}(X) = \text{positive} \mid X = \text{sick}] = 0.9$
  - Three tests are run (independently) with the following results
    - Two *positive* (on tests 1 and 2)
    - One *negative* (on test 3)
  - What is the most likely value for  $X$ ?

$$\begin{aligned}\Pr[X = \text{healthy}] &= 0.7 \\ \Pr[X = \text{carrier}] &= 0.2 \\ \Pr[X = \text{sick}] &= 0.1\end{aligned}$$



# Normalisation

- Example

- What is the most likely value for  $X$ ?

- Need to determine  $\Pr[X | T_1 = T_2 = 1, T_3 = 0] = \frac{\Pr[X]}{\Pr[T_1=T_2=1, T_3=0]} \cdot \Pr[T_1 = T_2 = 1, T_3 = 0 | X]$

- Notice that  $\frac{1}{\Pr[T_1=T_2=1, T_3=0]}$  is constant over each  $X$  in  $\Pr[X | T_1 = T_2 = 1, T_3 = 0]$

- So only compute  $\Pr[X] \cdot \Pr[T_1 = T_2 = 1, T_3 = 0 | X]$  for all  $X$

- As defined earlier, we let  $\alpha = \frac{1}{\Pr[T_1=T_2=1, T_3=0]}$

$$\begin{aligned}\Pr[\text{Test}(X) = \text{positive} | X = \text{healthy}] &= 0.1 \\ \Pr[\text{Test}(X) = \text{positive} | X = \text{carrier}] &= 0.7 \\ \Pr[\text{Test}(X) = \text{positive} | X = \text{sick}] &= 0.9\end{aligned}$$

$$\begin{aligned}\Pr[X = \text{healthy}] &= 0.7 \\ \Pr[X = \text{carrier}] &= 0.2 \\ \Pr[X = \text{sick}] &= 0.1\end{aligned}$$

$$\begin{array}{cccc}\Pr[X] & \Pr[T_1 = 1 | X] & \Pr[T_2 = 1 | X] & \Pr[T_3 = 0 | X]\end{array}$$

$$\Pr[X = \text{healthy} | \text{TestResults}] = \alpha \times 0.7 \times 0.1 \times 0.1 \times 0.9 = 0.0063\alpha$$

$$\Pr[X = \text{carrier} | \text{TestResults}] = \alpha \times 0.2 \times 0.7 \times 0.7 \times 0.3 = 0.0294\alpha$$

$$\Pr[X = \text{sick} | \text{TestResults}] = \alpha \times 0.1 \times 0.9 \times 0.9 \times 0.1 = 0.0081\alpha$$



# Conditional Probability Tables

---

- Given the chain rule and conditional independence assumption

$$\begin{aligned}\Pr[ Cause \mid Effect ] &= \frac{\Pr[Cause]}{\Pr[Effect]} \cdot \Pr[Effect \mid Cause] \\ &= \alpha \cdot \Pr[Cause] \cdot \prod_{i=1, \dots, k} \Pr[Effect_i \mid Cause]\end{aligned}$$

- We only need the Conditional Probability Table (CPT) with
  - Each  $\Pr[Effect_i \mid Cause]$
  - $\Pr[Cause]$
  - i.e.,  $k + 1$  entries (assuming  $k$  effects)



# Bayesian Networks

---



# Representing Bayesian Networks (BN)

- Represent joint distributions via a **graph**
  - **Vertices** are **random variables**
  - An **edge** from  **$X$**  to  **$Y$**   $\rightarrow X$  directly influences  **$Y$**   
some correlation – assume  $X$  causes  $Y$
- A **conditional distribution** for each node given its parents

$$\Pr[X \mid Parents(X)]$$

- In the simplest case, **conditional distribution** can be represented as a **conditional probability table (CPT)**
  - **CPTs** in the **BN** are the distribution over  **$X$**  for each combination of parent values

Note that the chain rule implies no cycles (BN is a DAG)

Edges link dependent variables

We want lower problem complexity – i.e., fewer parents

Note that CS3243 will only cover the specification and application of BN (not implementation)



## Relationships: Independent Events

---

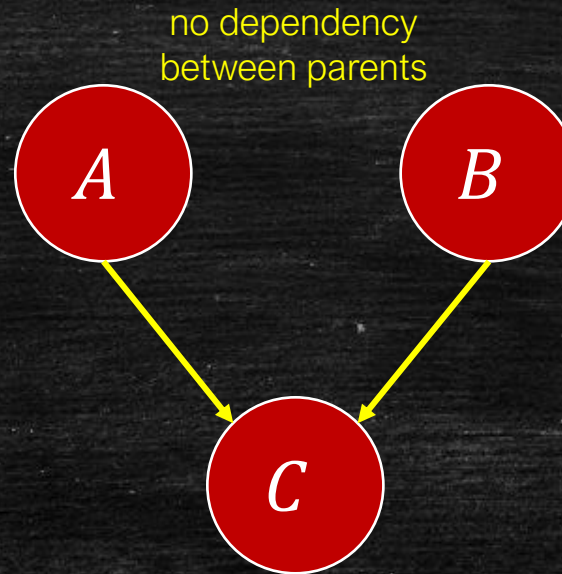
- $\Pr[A \wedge B \wedge C] = \Pr[C] \Pr[A] \Pr[B]$





# Relationships: Independent Causes

- $\Pr[A \wedge B \wedge C] = \Pr[C \mid A, B] \Pr[A] \Pr[B]$



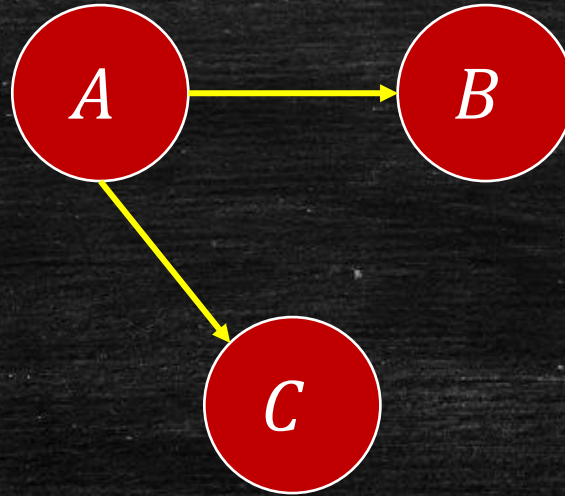
“I can be late either because of rain or because I was sick”

In logic:  $A \vee B \rightarrow C$



# Relationships: Conditionally Independent Effects

- $\Pr[A \wedge B \wedge C] = \Pr[C \mid A] \Pr[B \mid A] \Pr[A]$



“A disease can cause two independent tests to be positive”

In logic:  $A \rightarrow B; A \rightarrow C$



## Relationships: Causal Chain

---

- $\Pr[A \wedge B \wedge C] = \Pr[C \mid B] \Pr[B \mid A] \Pr[A]$





# Bayesian Networks Example

---



# Example Context

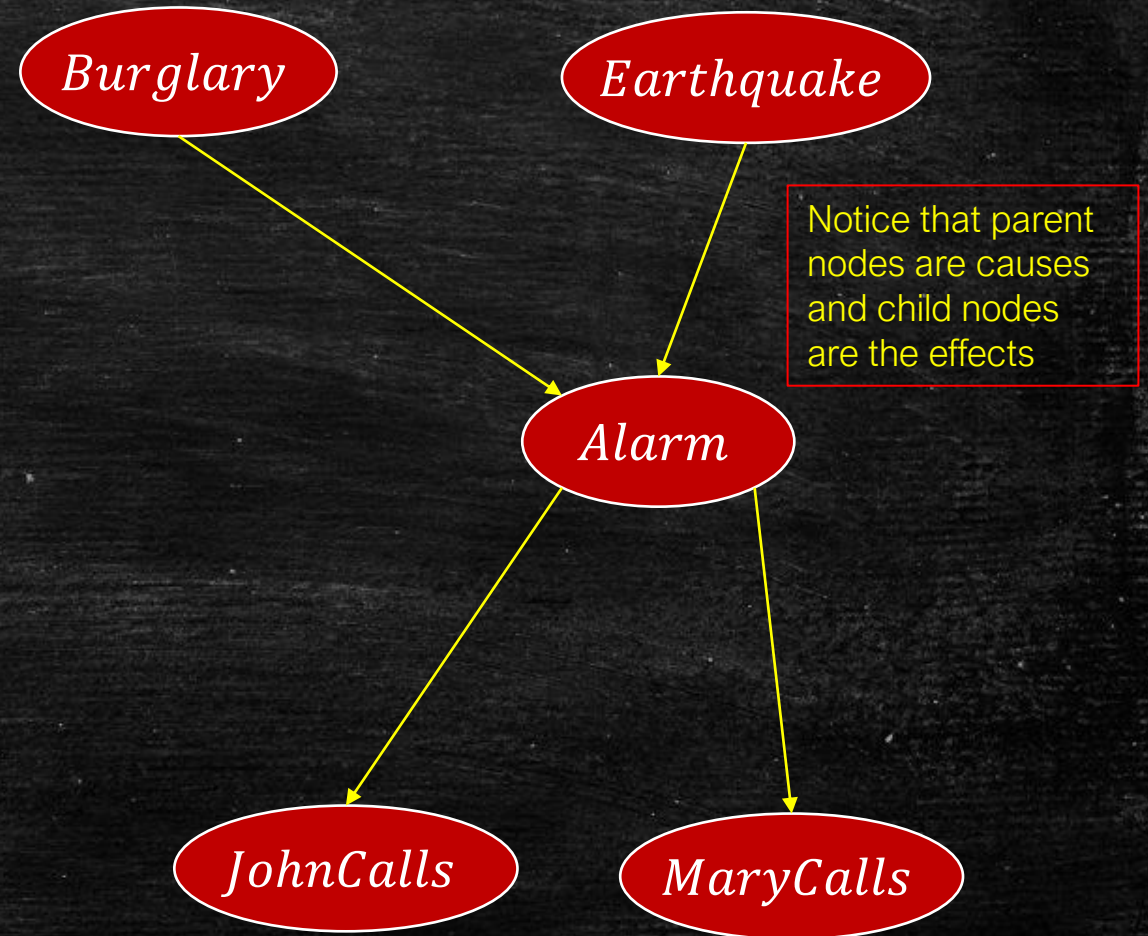
---

- You are out of the house...
  - $J$ : your neighbour John calls to say your house alarm is ringing
  - $M$ : another neighbour Mary does not call
  - $A, E$ : Alarm sometimes set off by minor earthquake
  - $B$ : Is there a burglar?
- Five binary variables
  - Joint distribution table size is  $2^5 - 1$
- Use domain knowledge to construct a Bayesian Network
  - Define the dependencies between variables
  - Fewer dependencies  $\rightarrow$  fewer probabilities (i.e., smaller CPTs)



# Example Bayesian Network

- Some domain knowledge...
  - The alarm is triggered by a burglary or earthquake
    - Independent Causes
  - John and Mary aren't friendly with each other; assume they do not check with each other before calling
  - They are mindful of privacy, so, would not directly observe a burglary at your house
  - They never notice earthquakes
    - Conditionally Independent Effects





# Example Bayesian Network

- Some domain knowledge...
  - We know the **crime rate** in the neighbourhood, which gives  $\Pr[B]$
  - We know the **likelihood of earthquakes** where you live, which gives  $\Pr[E]$
  - The alarm company provided us with the **statistics of the alarm system**, which gives  $\Pr[A | B, E]$
  - Based on experience, we also know **how likely John and Mary are to call when the alarm sounds**, which gives  $\Pr[J | A]$  and  $\Pr[M | A]$  respectively

$\Pr[B]$
0.001

*Burglary*

$\Pr[E]$
0.002

*Earthquake*

$B$	$E$	$\Pr[A   B, E]$
$T$	$T$	0.95
$T$	$F$	0.94
$F$	$T$	0.29
$F$	$F$	0.001

Notice that parents of nodes are in the conditional part of the probabilities

*Alarm*

$A$	$\Pr[J   A]$
$T$	0.9
$F$	0.05

*JohnCalls*

$A$	$\Pr[M   A]$
$T$	0.7
$F$	0.01

*MaryCalls*



# Example Bayesian Network

- From the context, we know

- $J = \text{True}$
- $M = \text{False}$

- We want to know if

- $B = \text{True}$

- $\Pr[B = 1 \mid J = 1 \wedge M = 0]$

Recall that:

$$\Pr[X \mid Y] = \frac{\Pr[X \wedge Y]}{\Pr[Y]}$$

$$= \frac{\Pr[B = 1 \wedge J = 1 \wedge M = 0]}{\Pr[J = 1 \wedge M = 0]}$$

Need to also consider  $A \wedge E$ ; so, 4 versions

- $\Pr[J, M, A, B, E] =$

Given chain rule and conditional dependencies

$$\Pr[J \mid A] \cdot \Pr[M \mid A] \cdot \Pr[A \mid B, E] \cdot \Pr[B] \Pr[E]$$

Remove  $\alpha$  from consideration

$\Pr[B]$
0.001

*Burglary*

$\Pr[E]$
0.002

*Earthquake*

$B$	$E$	$\Pr[A \mid B, E]$
$T$	$T$	0.95
$T$	$F$	0.94
$F$	$T$	0.29
$F$	$F$	0.001

*Alarm*

$A$	$\Pr[M \mid A]$
$T$	0.7
$F$	0.01

$A$	$\Pr[J \mid A]$
$T$	0.9
$F$	0.05

*JohnCalls*

*MaryCalls*



# Example Bayesian Network

- Calculate  $\Pr[B = 1 \mid J = 1 \wedge M = 0]$  for

- $A = 0, E = 0$
- $A = 1, E = 0$
- $A = 0, E = 1$
- $A = 1, E = 1$

- Use  $\Pr[J, M, A, B, E]$   
 $= \Pr[J \mid A] \Pr[M \mid A] \Pr[A \mid B, E] \Pr[B] \Pr[E]$

- For example, for  $A = 1, E = 0$ , we have

$$\begin{aligned} & \Pr[B = 1, J = 1, M = 0, A = 1, E = 0] \\ &= \Pr[j \mid a] \Pr[\neg m \mid a] \Pr[a \mid b, \neg e] \Pr[b] \Pr[\neg e] \\ &= 0.9 \times 0.3 \times 0.94 \times 0.001 \times 0.998 \simeq 0.000253 \end{aligned}$$

- Compare the sum against similar calculations for  $B = 0$

$\Pr[B]$	$\Pr[E]$
0.001	0.002

$B$	$E$	$\Pr[A \mid B, E]$
$T$	$T$	0.95
$T$	$F$	0.94
$F$	$T$	0.29
$F$	$F$	0.001

$A$	$\Pr[J \mid A]$	$A$	$\Pr[M \mid A]$
$T$	0.9	$T$	0.7
$F$	0.05	$F$	0.01



# Compactly Represent Joint Distributions

- Conditional probability for Boolean variable  $X$  with  $k$  Boolean parents has  $2^k$  rows
  - All possible parent values
- Each row requires one number  $p$  for  $X = \text{True}$
- If each variable has  $\leq k$  parents, network representation requires  $\mathcal{O}(n2^k)$  values
  - Full joint distribution has  $\mathcal{O}(2^n)$  values
- For burglary network,  $1 + 1 + 2 + 2 + 4 = 10$  numbers as compared to  $2^5 - 1 = 31$  numbers for full joint distribution

Recall that we wanted fewer parents; this is why

We will not delve further into the implementational details of Bayesian Networks for CS3243  
– please review Chapter 13 as further reading for more details



# Coping with Uncertainty: Decision-Theoretic Agents

---



# Recap on the Issues with Uncertainty

---

- **Deterministic environment**
  - No uncertainty
  - Planning is possible
- **Stochastic environment**
  - Uncertainty in Transition Model
    - Taking an action can lead to different intermediate states
  - Planning is difficult
    - Must now account for all possible intermediate states
    - Model-based agents operating in real-time?
      - Even if not planning, must still model transitions

We now know how to account for uncertainty – i.e., define the likelihood of an intermediate state, but how should we use this?



# The Decision-Theoretic Agent

---

- Rationality in the face of uncertainty
  - Probability Theory – accounting for uncertainty
  - Utility Theory – accounting for value (subjective and dependent on specific agent)
  - Decision Theory = Probability Theory + Utility Theory
- General idea
  - Agent is rational only *iff* it chooses actions that maximise utility
  - Maximum Expected Utility (MEU) Principle
    - Pick action with highest utility weighted over probable outcomes



# The Decision-Theoretic Agent

**function** DT-AGENT(*percept*) **returns** an *action*

**persistent:** *belief\_state*, probabilistic beliefs about the current state of the world  
*action*, the agent's action

update *belief\_state* based on *action* and *percept*

calculate outcome probabilities for actions,  
given action descriptions and current *belief\_state*

select *action* with highest expected utility  
given probabilities of outcomes and utility information

**return** *action*

Agent determines an action with each percept (not planning)

Maintains BELIEF STATE – i.e., what it currently believes about the problem environment

Accounting for the uncertainty – e.g., via a Bayesian Network – to determine the likelihood of intermediate states given actions

For each action,  $a_i$ , that can be taken at the current state  $s'$ , determine

$$\sum_{s \in T} utility(s) \times probability(s)$$

where  $T$  denotes the set of states that one may transition to from state  $s'$  when action  $a_i$  is taken

Choose the action that yields the highest value