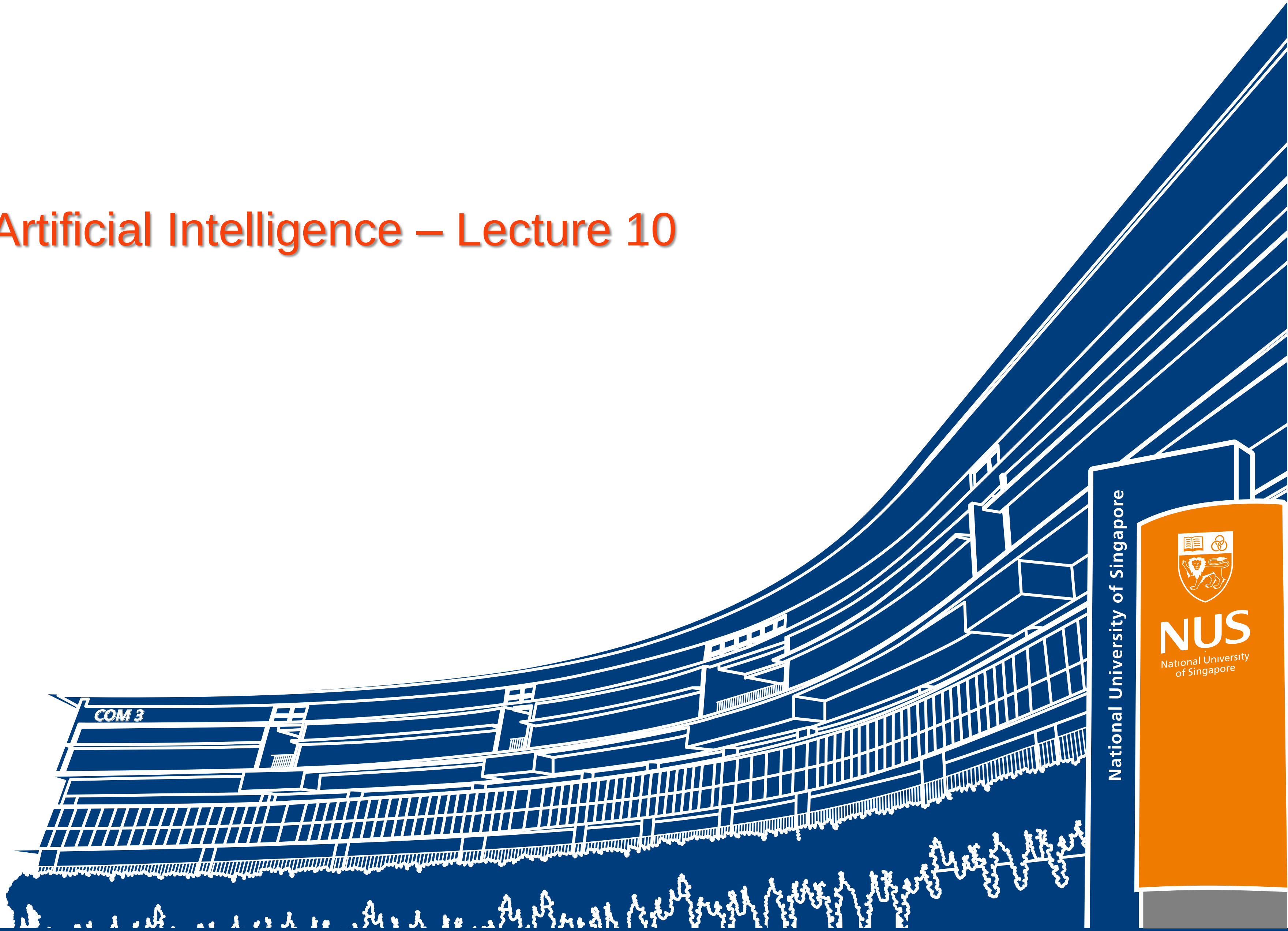
Uncertainty II

CS3243: Introduction to Artificial Intelligence – Lecture 10



NUS | Computing
National University of Singapore

1 Administrative Matters

Final Assessment

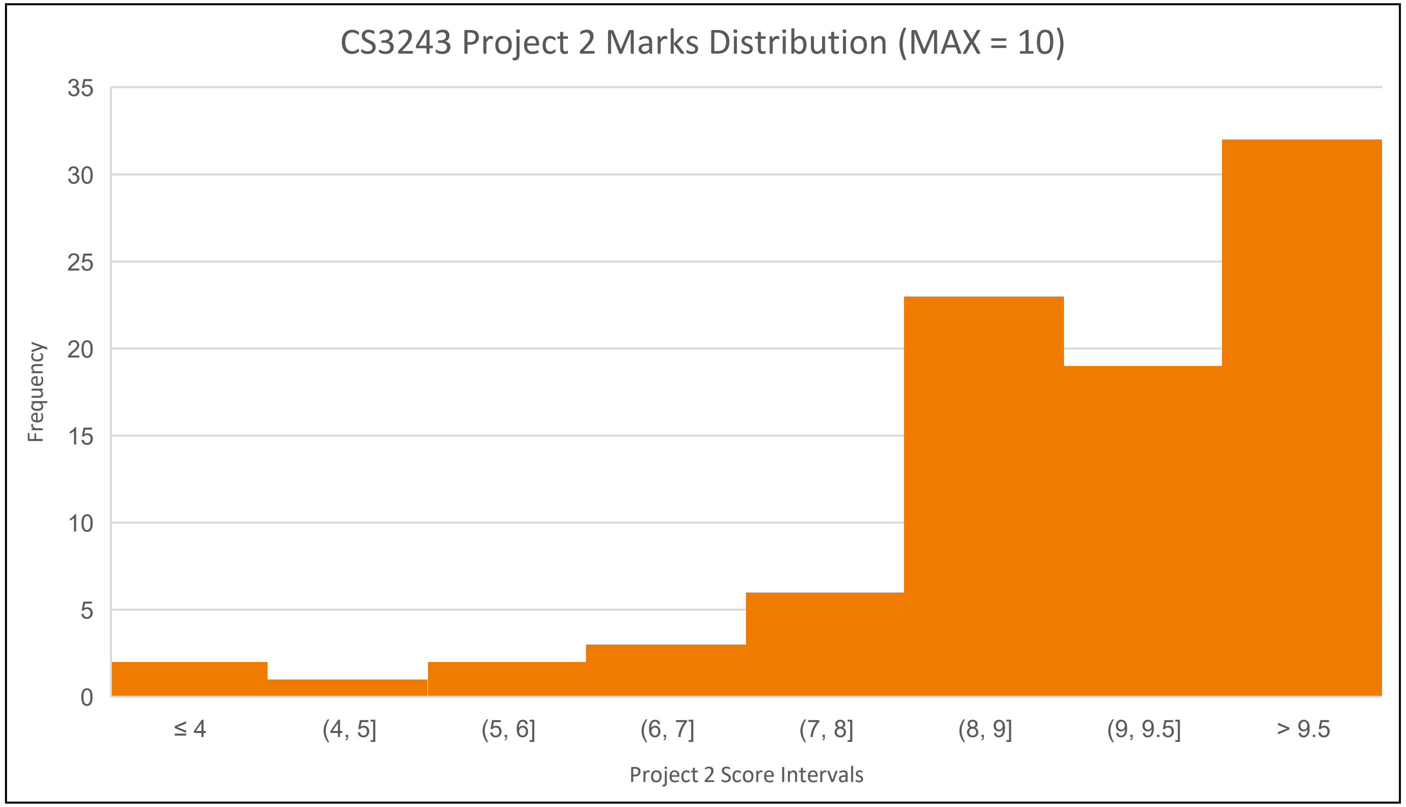
- Schedule & Venue
  - Tuesday (3 DEC), 0900-1100 hrs
  - MPSH 1A
- Format (similar to the Midterm)
  - Duration: 2 hours
  - Total: 60 marks
  - Closed-book + Cheat Sheet (1 × Double-sided A4 Sheet)
- Topics
  - All topics
  - About 75% of the paper will focus on topics not covered in the Midterm
- Practice Papers
  - Canvas > CS3243 > Flies > Past Papers > Final Papers

You **MUST** **confirm your Finals Schedule on the Student Portal** (including time and venues)

More **specifics on topics and question types** will be covered in the **Week 13 Course Review Lecture**

Project 2 Results

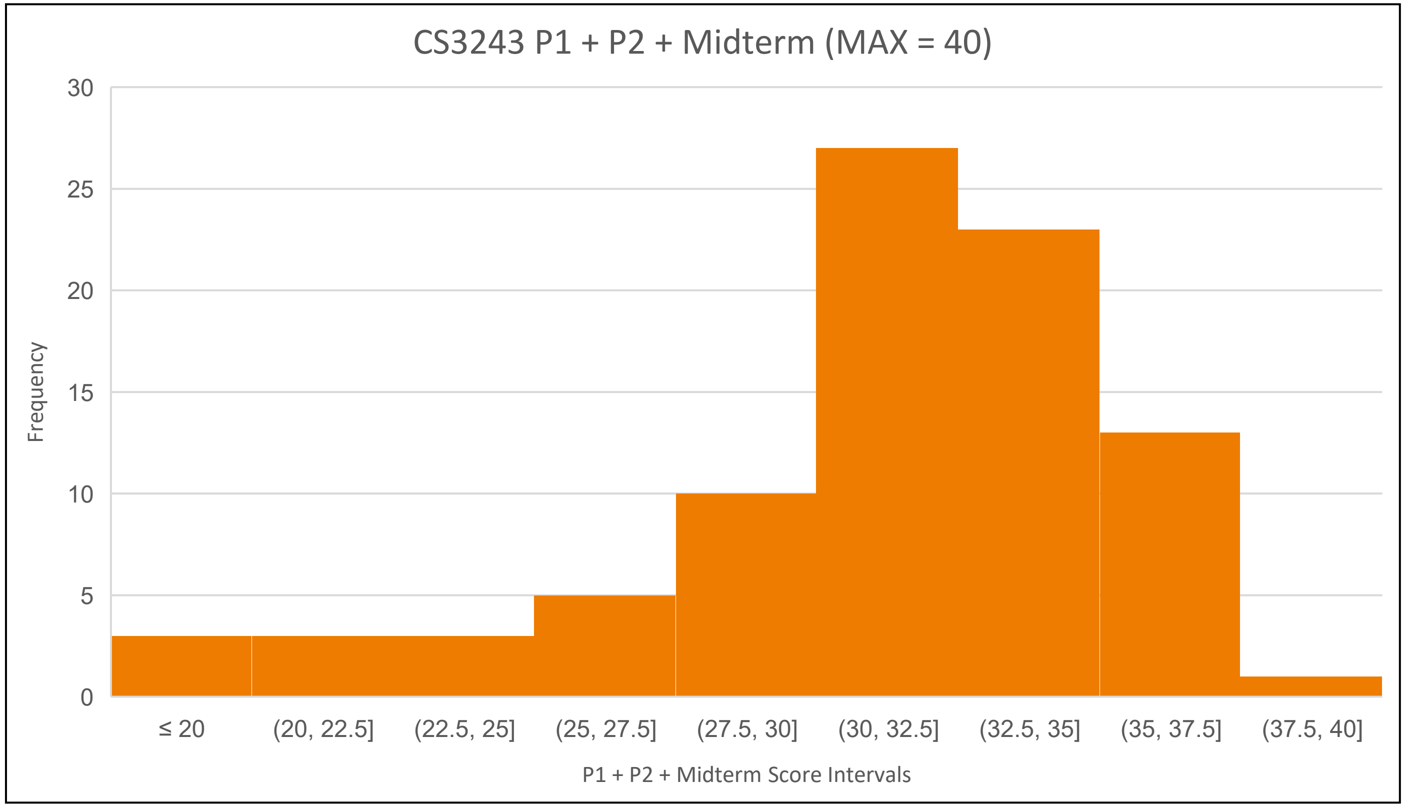
- Project 2.1 (MAX = 3)
  - Mean: 2.7
  - Median: 2.8
- Project 2.2 (MAX = 7)
  - Mean: 6.1
  - Median: 6.5
- Project 2 Bonus (MAX = 3)
  - Mean: 1.6
  - Median: 2



Current Overall Result

- Midterm (MAX = 20)
  - Mean: 12.8
  - Median: 12.8

- Project 1+2 + Bonus (MAX = 20)
  - Mean: 18.2
  - Median: 20.0



Student Feedback Exercise

NUS National University of Singapore

Your Voice Matters!

Be Constructive

Comments on your learning experience increase the value of your feedback.

Be Specific

Provide examples of how you think your teacher or the way the module is organised have helped (or not helped!) your learning.

Be Considerate

Improper language or personal comments are highly inappropriate, and undermine your feedback. Abusive comments are unacceptable.

Your feedback counts

Your constructive feedback helps professors to improve their modules and is one source of evidence for the university's appraisal decisions.

It's confidential

Your professors will never see your name. They will only get an aggregate report after the exam results have been released.

It's quick

Complete your module feedback on campus, at home, or on the go! It is easy to use and mobile compatible.

Student Feedback 2410

CS3243-INTRODUCTION TO ARTIFICIAL INTELLIGENCE (LECTURE)

88 Invited
1 Started
8 Responded
0 Opted Out

Evaluation ends on: 2024-11-22

9% Response Rate

Provide your feedback now >>

https://blue.nus.edu.sg/blue/

Upcoming...

- Deadlines
  - Tutorial Assignment 8 (released last week)
    - Post-tutorial submission: Due this Friday!
  - Tutorial Assignment 9 (released today)
    - Pre-tutorial Submission: Due this Sunday!
    - Post-tutorial Submission: Due next Friday!
  - Project 3 (released 14 October)
    - Due in Week 12

      10 November 2024

Final Content Lecture next week!

- No Lecture in Week 12
- Course Review Lecture in Week 13
  - Includes review of Final Exam Paper Structure

Project Consultations:
Thursdays (1900 hrs) via Zoom

Contents

- Uncertainty Fundamentals
- Bayesian Inference
- Bayesian Networks
- Decision-Theoretic Agents

2 Uncertainty Fundamentals

Recap: Conditional Probabilities & Bayes Rule

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

From above, we have :
Pr[A | B] = Pr[A ∧ B] / Pr[B] --- (1)
Pr[B | A] = Pr[B ∧ A] / Pr[A] --- (2)

Also, we have:
Pr[A ∧ B] = Pr[B ∧ A] --- (3)

From (2) and (3), we have:
Pr[A ∧ B] = Pr[B | A] · Pr[A] --- (4)

And thus from (4) and (1), we have Bayes Rule:
P[A|B] = (P[B|A] · P[A]) / P[B]

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

Driven by stochastic environment:
- What is Pr[$PIT_{2,1}$ | KB]?
- What is Pr[$PIT_{1,2}$ | KB]?
- How can we use the above to rationalise the action to take?
  ... revisited later in the lecture...

Chain Rule

- With more than two events, we have

$$\Pr[R_1 \wedge \cdots \wedge R_k] = \Pr[\,(R_1 \wedge \cdots \wedge R_{k-1}) \wedge R_k\,]$$

$$= \Pr[\,R_k \mid R_{k-1} \wedge \cdots \wedge R_1\,] \cdot \Pr[\,R_{k-1} \wedge \cdots \wedge R_1\,]$$

- And by induction, we have

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

- For example:

$$\Pr[A \wedge B \wedge C \wedge D] = \Pr[D|C \wedge B \wedge A] \cdot \Pr[C \wedge B \wedge A]$$

$$= \Pr[D|C \wedge B \wedge A] \cdot \Pr[C|B \wedge A] \cdot \Pr[B \wedge A]$$

$$= \Pr[D|C \wedge B \wedge A] \cdot \Pr[C|B \wedge A] \cdot \Pr[B|A] \cdot \Pr[A]$$

Independence

- A and B are independent if $\Pr[A \wedge B] = \Pr[A] \cdot \Pr[B]$
- Equivalent to $\Pr[\,A \mid B\,] = \Pr[A]$
- Knowing B adds no information about A
- Example:
  - Suppose we have a bag containing 8 Go stones
  - 3 white stones and 5 black stones
  - Event B: first stone drawn is white
  - Event A: second stone drawn is black
  - $\Pr[\,A \mid B\,] = \frac{5}{7}$; $\Pr[\,A \mid \neg B\,] = \frac{4}{7}$ (i.e., dependent)
  - The outcome of B changes the conditional probability $\Pr[\,A \mid B\,]$
  - If independent: $\Pr[\,A \mid B\,] = \Pr[\,A \mid \neg B\,]$

Recall that:
P[A|B] = P[A∧B] / P[B]
P[B|A] = P[B∧A] / P[A]

Image taken from Wikipedia

3 Bayesian Inference

Performing Inference via Bayes' Rule

- Instead of inferring statements in the form

"is α true given the KB ?"

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

- We infer statement of the form

"What is the likelihood of an event α given the probabilities of other events?"

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

Naturally occurs in everyday situations …
… e.g., Pr[COVID-19 | Fever ∧ Cough ∧ …]

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

Recall that $\Pr[\text{toothache}]$
$= 0.108 + 0.016 + 0.012 + 0.064$
$= 0.2$

- For any proposition (event) X, sum the atomic events y where X holds (i.e., calculate the unconditional or marginal probability of X):

$$\Pr[X] = \sum_{y \in X} \Pr[X = y]$$

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

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?

$$\underbrace{d \times d \times \dots \times d}_{n \text{ times}} = d^n$$

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

- Suppose that the n random variables, $X_1, \dots, X_n$, are independent
  - How big is their joint distribution table now?

$$\underbrace{d + d + \dots + d}_{n \text{ times}} = dn$$

If A and B are independent:
$\Pr[A \wedge B] = \Pr[A] \, . \Pr[B]$
$\Pr[\, A \mid B \,] = \Pr[A]$

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

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

4 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 !
  - $\Pr[A \wedge B \mid S] = \Pr[A \mid S] \cdot \Pr[B \mid S]$

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

$$\Pr[\, T_1 \wedge \cdots \wedge T_{n-1} \wedge S \,] = \Pr[\, T_1 \mid S \,] \cdot \Pr[\, T_2 \mid S \,] \cdot \ldots \cdot \Pr[\, T_{n-1} \mid S \,] \cdot \Pr[\, S \,]$$

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, ...

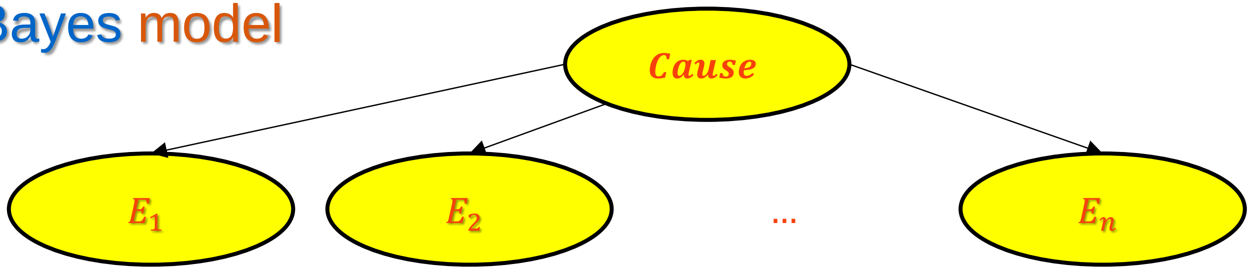
$$\Pr[Cause \mid E_1, \ldots, E_n] = \frac{\Pr[Cause]\Pr[E_1,\ldots,E_n \mid Cause]}{\Pr[E_1,\ldots,E_n]}$$

$$= \frac{\Pr[Cause]}{\Pr[E_1,\ldots,E_n]} \cdot \prod_i \Pr[\,E_i \mid Cause\,]$$

$$= \alpha.\Pr[Cause] \cdot \prod_i \Pr[\,E_i \mid Cause\,]$$

When comparing, for example, the probability of $Cause$ and $\neg Cause$, α remains constant, so the calculations may be normalised

- This is the Naive Bayes model



Application of the 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
    - $\Pr[Test(X) = positive \mid X = healthy] = 0.1$
    - $\Pr[Test(X) = positive \mid X = carrier] = 0.7$
    - $\Pr[Test(X) = positive \mid X = sick] = 0.9$
  - 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?

$$\Pr[X = healthy] = 0.7$$
$$\Pr[X = carrier] = 0.2$$
$$\Pr[X = sick] = 0.1$$

$$\Pr[Test(X) = positive \mid X = healthy] = 0.1$$
$$\Pr[Test(X) = positive \mid X = carrier] = 0.7$$
$$\Pr[Test(X) = positive \mid X = sick] = 0.9$$

Application of the Naive Bayes Algorithm

- Example
  - What is the most likely value for X?
    - Need to determine: $\Pr[X \mid T_1 = T_2 = 1, T_3 = 0] = \frac{\Pr[X]}{\Pr[T_1=T_2=1,T_3=0]} . \Pr[T_1 = T_2 = 1, T_3 = 0 \mid X]$
    - Recall that $\frac{1}{\Pr[T_1=T_2=1,T_3=0]}$ is constant over each X in $\Pr[X \mid T_1 = T_2 = 1, T_3 = 0]$
    - So only compute $\Pr[X] . \Pr[T_1 = T_2 = 1, T_3 = 0 \mid X]$ for all X
    - As defined earlier, we let $\alpha = \frac{1}{\Pr[T_1=T_2=1,T_3=0]}$

$\Pr[X = healthy] = 0.7$
$\Pr[X = carrier] = 0.2$
$\Pr[X = sick] = 0.1$

$\Pr[Test(X) = positive \mid X = healthy] = 0.1$
$\Pr[Test(X) = positive \mid X = carrier] = 0.7$
$\Pr[Test(X) = positive \mid X = sick] = 0.9$

$$\Pr[X = healthy \mid TestResults] = \alpha \times \underbrace{0.7}_{\Pr[X]} \times \underbrace{0.1}_{\Pr[T_1 = 1 \mid X]} \times \underbrace{0.1}_{\Pr[T_2 = 1 \mid X]} \times \underbrace{0.9}_{\Pr[T_3 = 0 \mid X]} = 0.0063\alpha$$

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

$$\Pr[X = sick \mid TestResults] = \alpha \times 0.1 \times 0.9 \times 0.9 \times 0.1 = 0.0081\alpha$$

Naive Bayes Lecture Exercise

5a. Consider the following table, which contains the performance data for 10 students taking the CS9999 course. The data for the i-th student is given in column i. Each student is represented by 3 Boolean variables, X_a, X_b, and X_c, as well as one class label, Y.

- If $X_a = 1$, then the student did well in Assessment 1, else, $X_a = 0$.
- If $X_b = 1$, then the student did well in Assessment 2, else, $X_b = 0$.
- If $X_c = 1$, then the student did well in Assessment 3, else, $X_c = 0$.
- The class label $Y = 1$ if the student did well in the CS9999 course, else, $Y = 0$.

i	1	2	3	4	5	6	7	8	9	10
X_a	0	0	0	0	0	1	0	1	0	1
X_b	0	0	0	0	0	0	0	0	1	1
X_c	1	1	0	0	1	1	1	1	0	0
Y	0	0	0	0	1	1	1	1	1	1

Assume that a Naïve Bayes model is to be utilised. More specifically, assume that we are to adopt the following Bayesian Network.

(i) [2 marks] Complete the following probability tables – i.e., fill in the last column in each of the tables given below.

Solution:

Y	$P(Y)$
0	
1	

Y	X_a	$P(X_a\|Y)$
0	0	
0	1	
1	0	
1	1	

Y	X_b	$P(X_b\|Y)$
0	0	
0	1	
1	0	
1	1	

Y	X_c	$P(X_c\|Y)$
0	0	
0	1	
1	0	
1	1	

(ii) [2 mark] Assume that we observe the assessments for a new student, where $X_a = 0$, $X_b = 0$, and $X_c = 1$. Is this student more likely to do well in the CS9999 course (i.e., have $Y = 1$)?

Option **A**: The new student is more likely to do well in CS9999 (i.e., have $Y = 1$)

Option **B**: The new student is more likely to not do well in CS9999 (i.e., have $Y = 0$)

Option **C**: Both $Y = 1$ and $Y = 0$ are equally likely

Option **D**: There is not enough information to determine which outcome is more likely

Naive Bayes Lecture Exercise

5a. Consider the following table, which contains the performance data for 10 students taking the CS9999 course. The data for the i-th student is given in column i. Each student is represented by 3 Boolean variables, X_a, X_b, and X_c, as well as one class label, Y.

- If $X_a = 1$, then the student did well in Assessment 1, else, $X_a = 0$.
- If $X_b = 1$, then the student did well in Assessment 2, else, $X_b = 0$.
- If $X_c = 1$, then the student did well in Assessment 3, else, $X_c = 0$.
- The class label $Y = 1$ if the student did well in the CS9999 course, else, $Y = 0$.

i	1	2	3	4	5	6	7	8	9	10
X_a	0	0	0	0	0	1	0	1	0	1
X_b	0	0	0	0	0	0	0	0	1	1
X_c	1	1	0	0	1	1	1	1	0	0
Y	0	0	0	0	1	1	1	1	1	1

Assume that a Naïve Bayes model is to be utilised. More specifically, assume that we are to adopt the following Bayesian Network.

(i) [2 marks] Complete the following probability tables – i.e., fill in the last column in each of the tables given below.

Solution:

Y	P(Y)
0	
1	

Y	X_a	P(X_a|Y)
0	0	
0	1	
1	0	
1	1	

Y	X_b	P(X_b|Y)
0	0	
0	1	
1	0	
1	1	

Y	X_c	P(X_c|Y)
0	0	
0	1	
1	0	
1	1	

(ii) [2 mark] Assume that we observe the assessments for a new student, where $X_a = 0$, $X_b = 0$, and $X_c = 1$. Is this student more likely to do well in the CS9999 course (i.e., have $Y = 1$)?

Option **A**: The new student is more likely to do well in CS9999 (i.e., have $Y = 1$)

Option **B**: The new student is more likely to not do well in CS9999 (i.e., have $Y = 0$)

Option **C**: Both $Y = 1$ and $Y = 0$ are equally likely

Option **D**: There is not enough information to determine which outcome is more likely

The CS3243 Final Examination **WILL** include a problem like this!

Conditional Probability Tables

- Given the chain rule and conditional independence assumption

$$\Pr[\,Cause \mid Effect\,] = \frac{\Pr[Cause]}{\Pr[Effect]} . \Pr[Effect \mid Cause]$$

$$= \alpha . \Pr[Cause] . \prod_{i=1,\ldots,k} \Pr[\,Effect_i \mid Cause\,]$$

- 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)

Questions about the Lecture?

- Was anything unclear?
- Do you need to clarify anything?
- Ask on Archipelago
  - Specify a question
  - Upvote someone else's question

Invitation Link (Use NUS Email --- starts with E)
https://archipelago.rocks/app/resend-invite/19950643218

5 Bayesian Networks

Representing Bayesian Networks (BN)

- Represent joint distributions via a graph
  - Vertices are random variables
  - An edge from X to Y → X directly influences Y

    some correlation – assume X causes Y

- Each vertex has a conditional distribution given its parents

$$\Pr[\, Z \mid Parents(Z) \,]$$

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

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

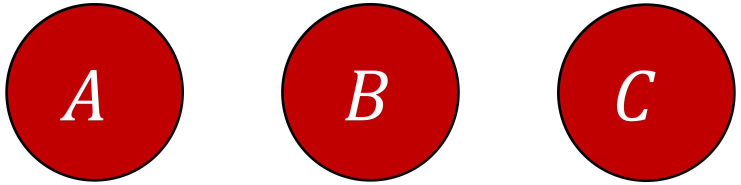
Edges link dependent variables

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

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

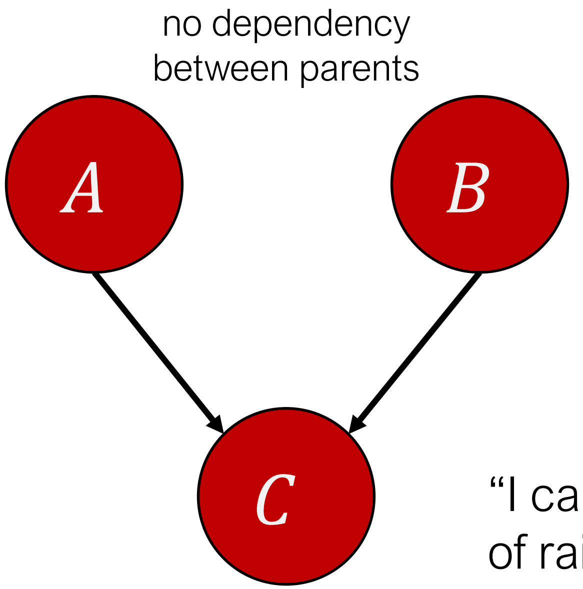
Relationships: Independent Events

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



Relationships: Independent Causes

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

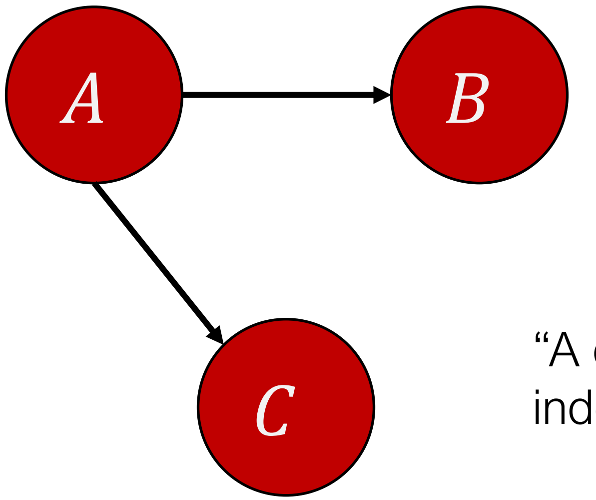


“I can be late either because of rain or because I was sick”

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

Relationships: Conditionally Independent Effects

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

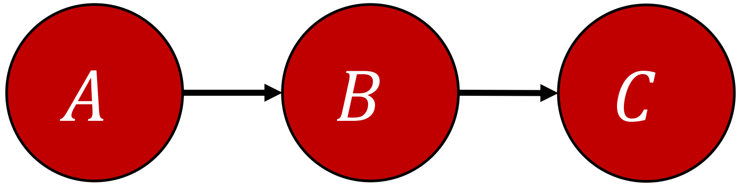


"A disease can cause two independent tests to be positive"

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

Relationships: Causal Chain

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



Markov Blanket

- A variable is conditionally independent of all other nodes in the network, given
  - its parents
  - its children
  - its children's parents
- This refers to a variable's Markov blanket

AIMA Textbook pp. 437

- For each variable within a Bayesian Network, all variables are conditionally independent apart from the variables listed above

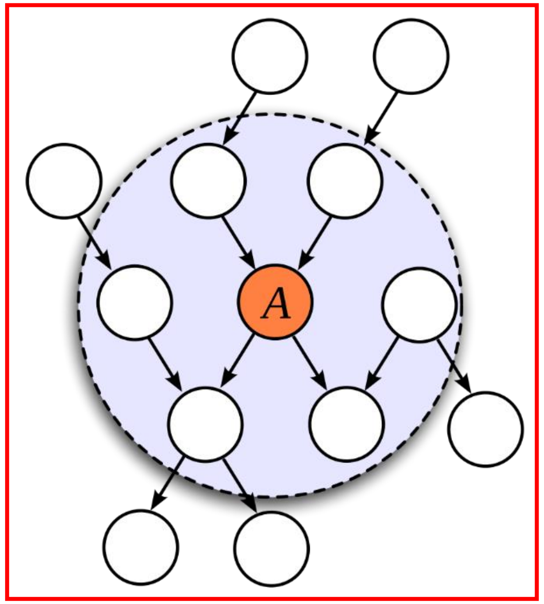


Image taken from Wikipedia

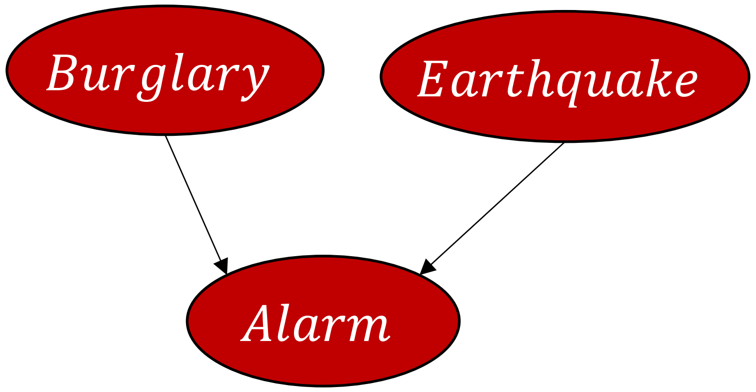
6 Bayesian Networks Example

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 → fewer probabilities (i.e., smaller CPTs)

Example Bayesian Network

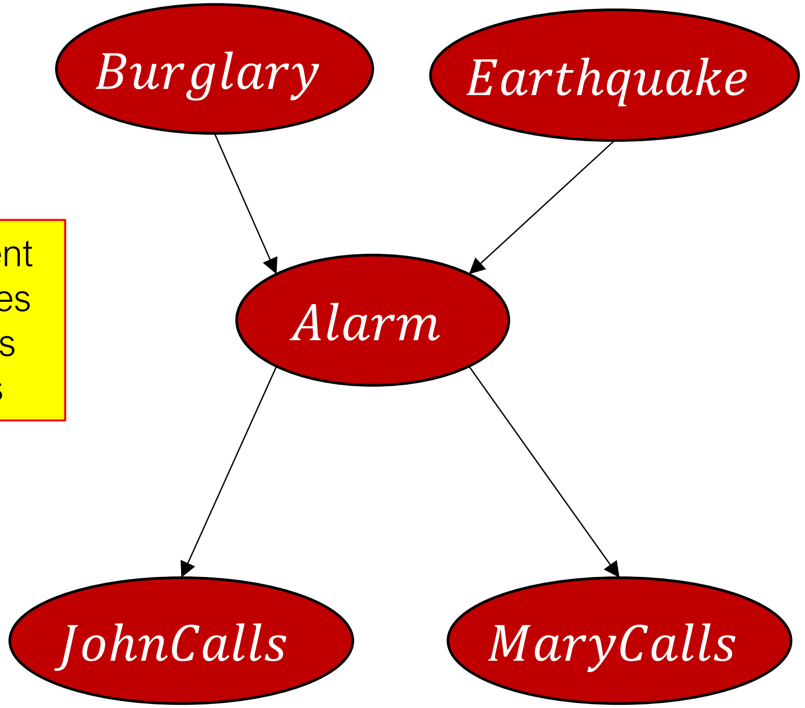
- Some domain knowledge...
  - The alarm is triggered by a burglary or earthquake
    - Independent Causes



Example Bayesian Network

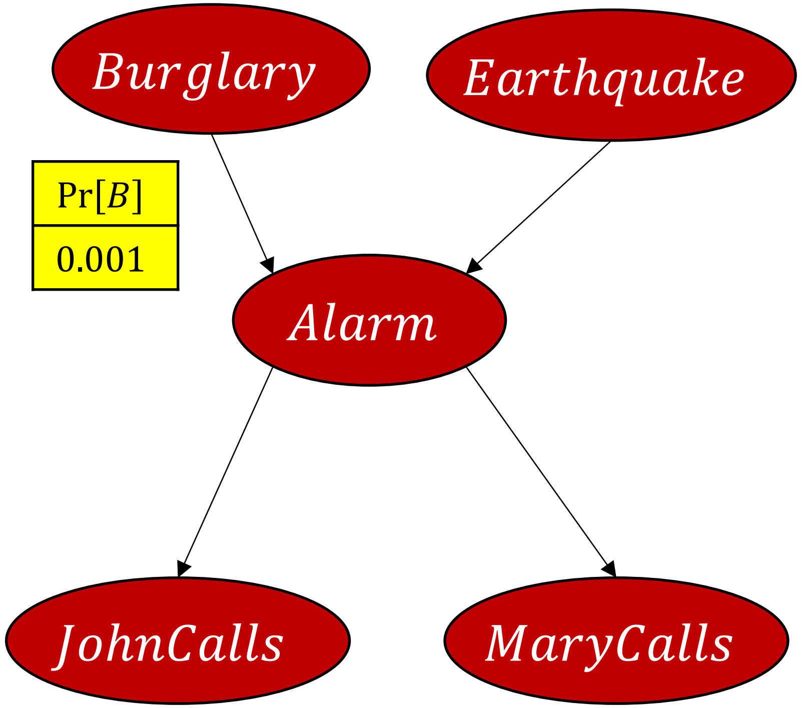
- Some domain knowledge...
  - The alarm is triggered by a burglary or earthquake
    - Independent Causes
  - John and Mary aren't friendly with each other; assume they do not check with each other before calling
  - They are mindful of privacy, so, would not directly observe a burglary at your house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

Notice that parent nodes are causes and child nodes are the effects



Example Bayesian Network

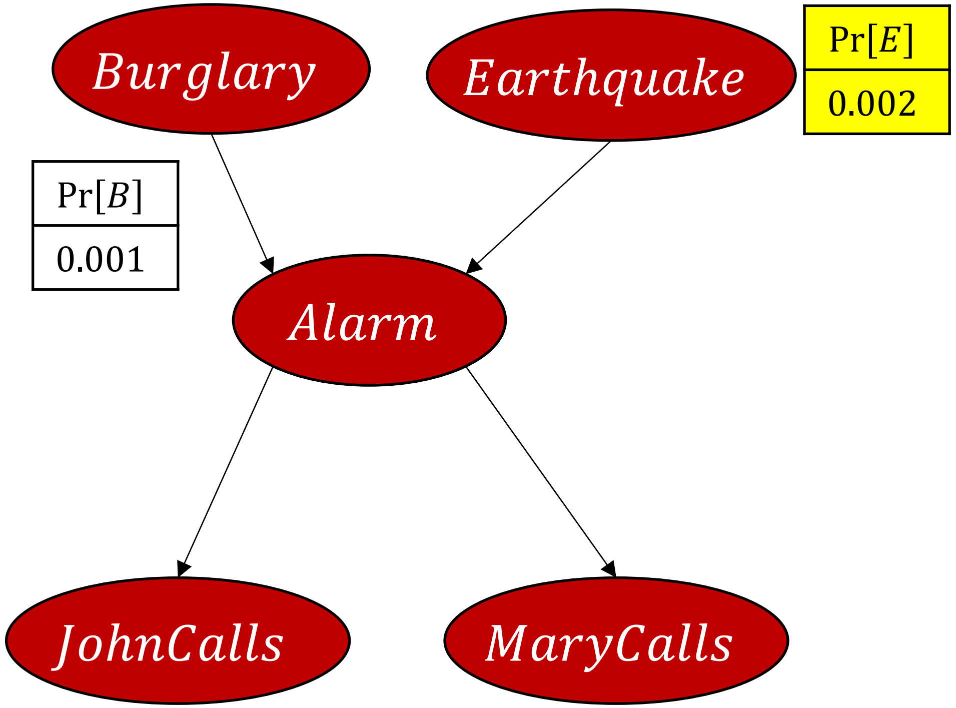
- Some domain knowledge...
  - We know the crime rate in the neighbourhood, which gives $\Pr[B]$



$\Pr[B]$
0.001

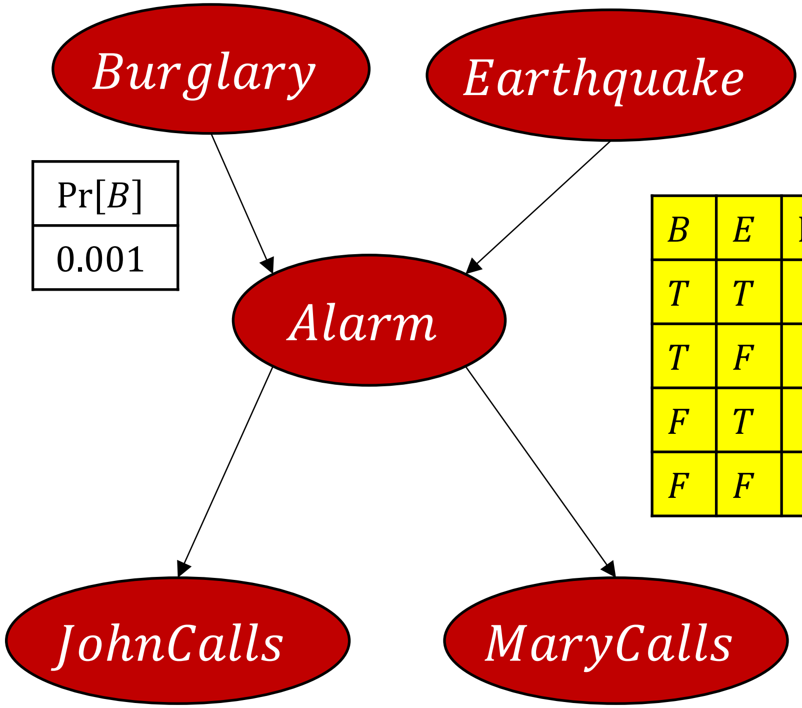
Example Bayesian Network

- 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 $\Pr[E]$



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 $\Pr[A \mid B, E]$



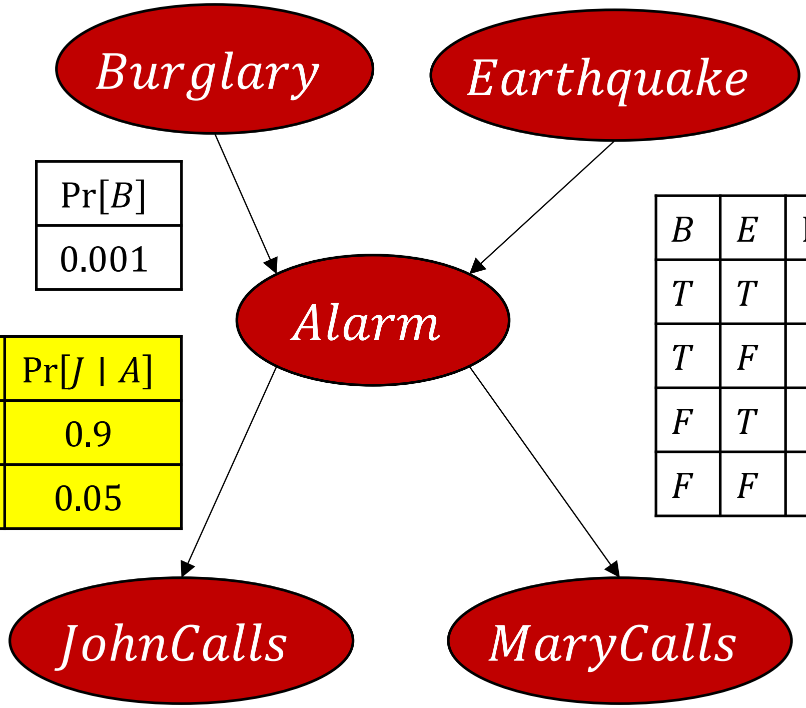
$\Pr[B]$
0.001

$\Pr[E]$
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

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 $\Pr[A \mid B, E]$
  - Based on experience, we also know how likely John and Mary are to call when the alarm sounds, which gives $\Pr[J \mid A]$ and $\Pr[M \mid A]$ respectively



$\Pr[B]$
0.001

$\Pr[E]$
0.002

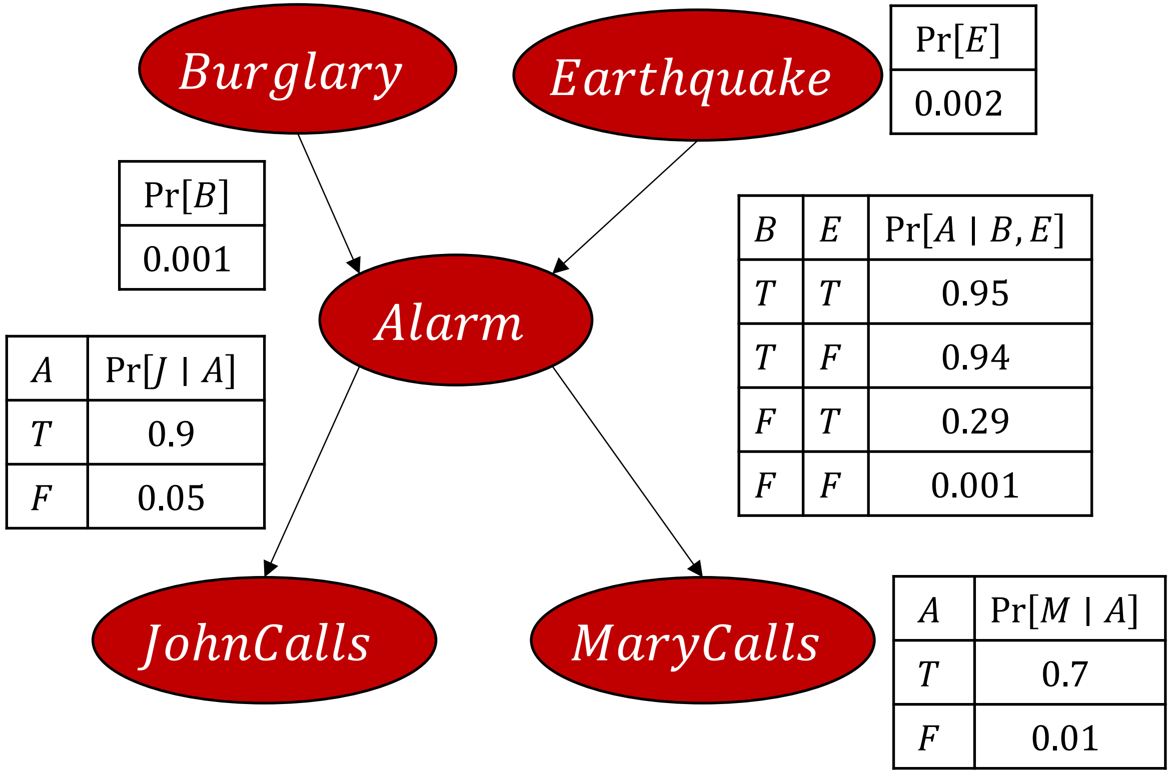
B	E	$\Pr[A \mid B, E]$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

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

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

Example Bayesian Network

- From the context, we know
  - $J = True$
  - $M = False$
- We want to know if
  - $B = True$



$\Pr[B]$
0.001

$\Pr[E]$
0.002

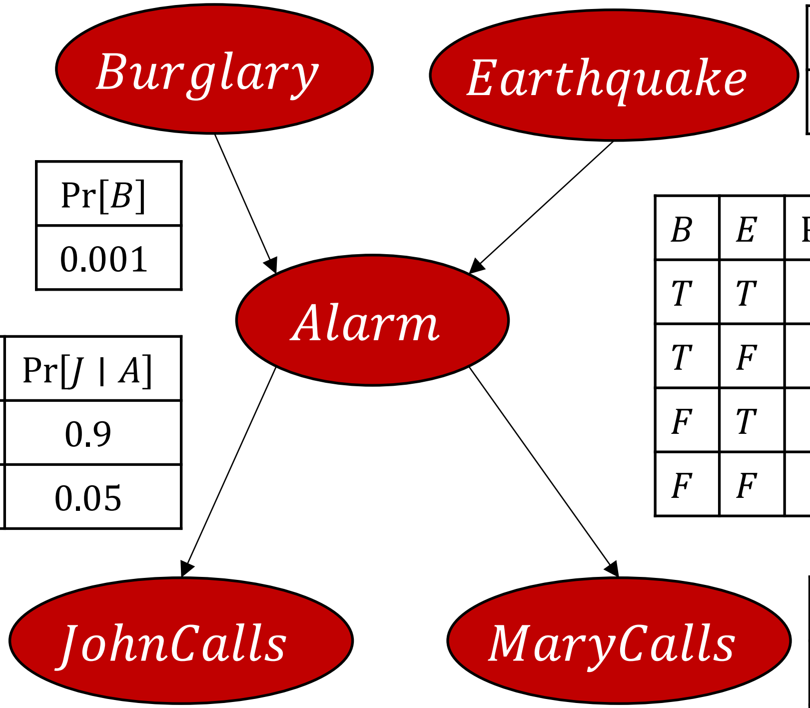
B	E	$\Pr[A \mid B, E]$
T	T	0.95
T	F	0.94
F	T	0.29
F	F	0.001

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

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

Example Bayesian Network

- From the context, we know
  - $J = True$
  - $M = False$
- We want to know if
  - $B = True$
- $\Pr[B = 1 \mid J = 1 \land M = 0]$



$\Pr[B]$
0.001

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

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

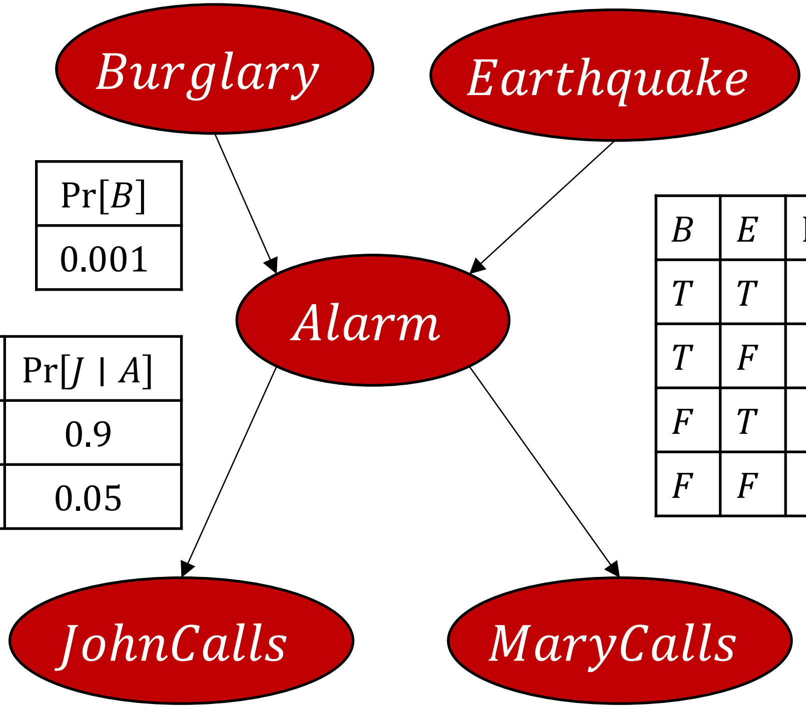
Example Bayesian Network

- From the context, we know
  - $J = True$
  - $M = False$
- We want to know if
  - $B = True$
- $\Pr[B = 1 \mid J = 1 \land M = 0]$

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

Recall that: $\Pr[X \mid Y] = \frac{\Pr[X \land Y]}{\Pr[Y]}$

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



$\Pr[E]$
0.002

$\Pr[B]$
0.001

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

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

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

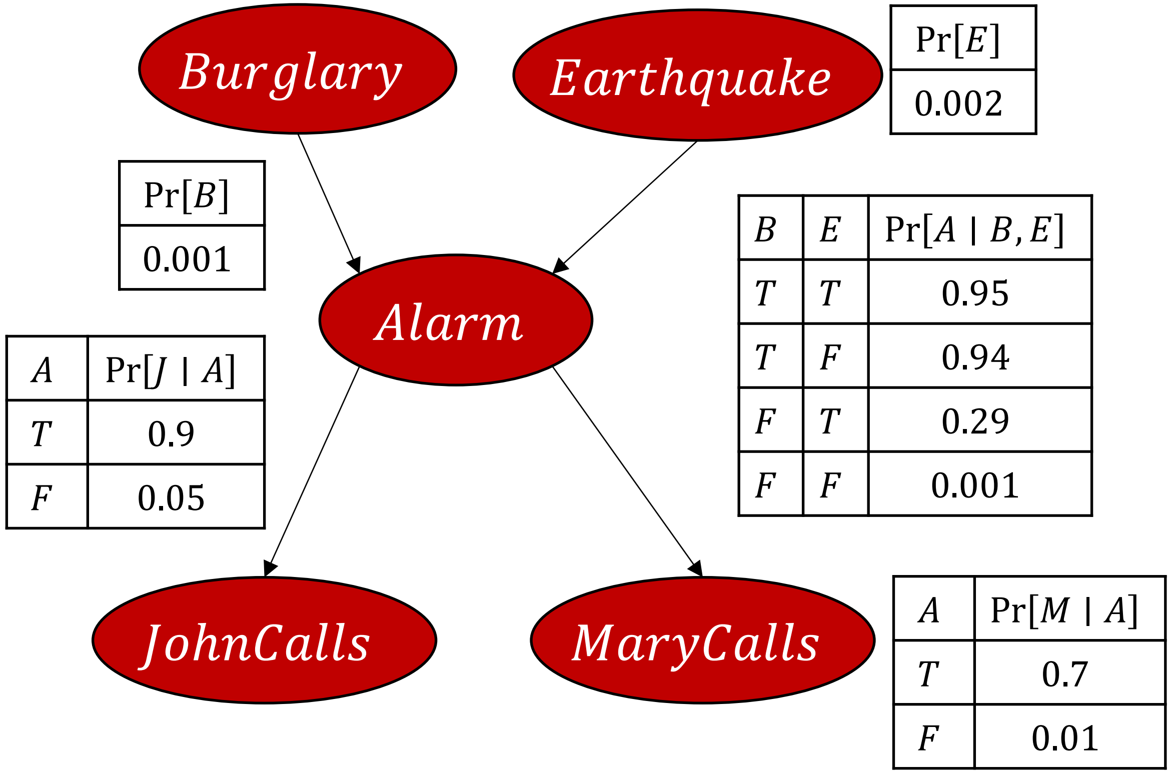
Example Bayesian Network

- From the context, we know
  - $J = True$
  - $M = False$
- We want to know if
  - $B = True$
- $\Pr[B = 1 \mid J = 1 \wedge M = 0]$
  $$= \frac{\Pr[B = 1 \wedge J = 1 \wedge M = 0]}{\Pr[J = 1 \wedge M = 0]}$$

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

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

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



$\Pr[E]$
0.002

$\Pr[B]$
0.001

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

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

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

Example Bayesian Network

- From the context, we know
  - $J = True$
  - $M = False$
- We want to know if
  - $B = True$
- $\Pr[B = 1 \mid J = 1 \wedge M = 0]$

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

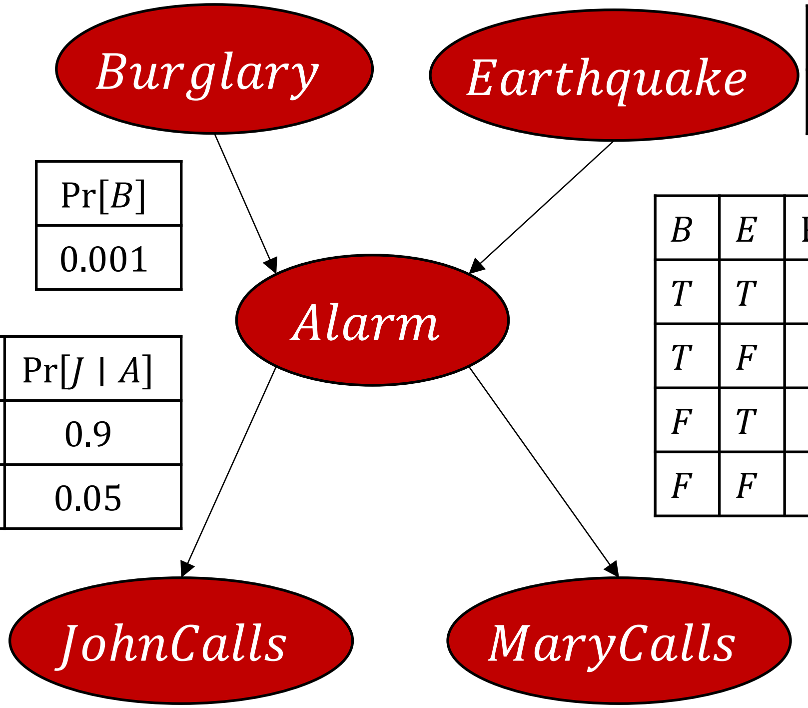
Recall that: $\Pr[X \mid Y] = \dfrac{\Pr[X \wedge Y]}{\Pr[Y]}$

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

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

Given chain rule and conditional dependencies

Remove α from consideration



$\Pr[B]$
0.001

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

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

Example Bayesian Network

- Calculate $\Pr[\, B = 1 \mid J = 1 \land M = 0 \,]$ for the following 4
  - $A = 0, E = 0$
  - $A = 1, E = 0$
  - $A = 0, E = 1$
  - $A = 1, E = 1$
- 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

$\Pr[B = 1, J = 1, M = 0, A = 1, E = 0]$

$= \Pr[\, j \mid a \,] \Pr[\neg m \mid a] \Pr[\, a \mid b, \neg e \,] \Pr[b] \Pr[\neg e]$

$= 0.9 \times \quad 0.3 \times \quad 0.94 \times \quad 0.001 \times 0.998 \quad \simeq 0.000253$

- Compare the sum of 4 probabilities above against sum for $B = 0$

$\Pr[B]$
0.001

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

A	$\Pr[M \mid A]$
T	0.7
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 $X = True$
    - Joint distribution table size is 2^5 - 1
- If each variable has $\leq k$ parents, network representation requires $\mathcal{O}\left(n2^k\right)$ values
    - Full joint distribution has $\mathcal{O}(2^n)$ values

Recall that we wanted fewer parents; this is why

- For burglary network, $1 + 1 + 2 + 2 + 4 = 10$ numbers
  as compared to $2^5 - 1 = 31$ numbers for full joint distribution

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

7 | 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 does not always lead to the same intermediate state!
  - 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

Questions about the Lecture?

- Was anything unclear?
- Do you need to clarify anything?
- Ask on Archipelago
  - Specify a question
  - Upvote someone else's question

Invitation Link (Use NUS Email --- starts with E)
https://archipelago.rocks/app/resend-invite/19950643218