Uncertainty II

CS3243: Introduction to Artificial Intelligence - Lecture 10

28 March 2022

Contents

Reference: AIMA 4th Edition, Section 12; 13.1-13.2

Administrative Matters

Midterm Results & Lecture Schedule

- Midterm Examination
  - Midterm results finalised
  - Gradebook and accompanying summary announcement this week

- Today's lecture is the final content lecture
  - No lecture next week (Week 12) – focus on completing Project 3
  - 1-hour review lecture in Week 13 – tips for the Final?
    - This review lecture will not be recorded (no new examinable material)

Project 2 & Final Examination

- Project 2 results released
  - Appeals and clarifications by 30 March (Wednesday)

- Final Examination (30% assessment weight)
  - 25 April (Monday), 1700-1900hrs
  - Online like Midterm Examination
  - Adhere to proctoring protocols
  - ALL material assessable; more on CSP onwards

Upcoming…

- Deadlines (final set)
  - DQ10 - final diagnostic quiz (released today)
    - *Two attempts*
    - *Due this Sunday (3 April), 2359 hrs*
  - TA8 (released last Monday)
    - *Due this Sunday (3 April), 2359 hrs*
  - TA9 - final tutorial with assignment (released today)
    - *Due next Sunday (10 April), 2359 hrs*
  - Project 3
    - *Due next Sunday (10 April), 2359 hrs*

Bayes' Rule and Independence

Recap on Conditional Probabilities & Bayes' Rule

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

Note:
Pr[A | B] = Pr[A ∧ B] / Pr[B] --- (1)
Pr[B | A] = Pr[B ∧ A] / Pr[A] --- (2)

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

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

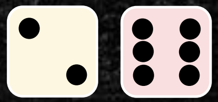
And thus from (4) and the definition above, we have Bayes' Rule:
P[A|B] = (P[B|A].P[A]) / P[B]

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

- Example: $\Pr[X_1 = 2 \mid X_1 + X_2 = 8] = ?$

$$= \frac{\overset{1/6}{\Pr[X_1+X_2=8|X_1=2]}\,.\,\overset{1/6}{\Pr[X_1=2]}}{\underset{5/36}{\Pr[X_1+X_2=8]}} = \frac{1}{5}$$

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



Chain Rule

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

- With more than two events, we have

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

- And by induction, we have

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

- Example:

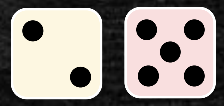
$$\Pr[A \wedge B \wedge C \wedge D] = \Pr[D|C \wedge B \wedge A].\Pr[C \wedge B \wedge A]$$
$$= \Pr[D|C \wedge B \wedge A].\Pr[C|B \wedge A].\Pr[B \wedge A]$$
$$= \Pr[D|C \wedge B \wedge A].\Pr[C|B \wedge A].\Pr[B|A].\Pr[A]$$

Independence

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

- A and B are independent if $\Pr[A \wedge B] = \Pr[A] \,.\, \Pr[B]$
- Equivalent to $\Pr[\, A \mid B \,] = \Pr[A]$
- Knowing B adds no information about A
  - B does not further categorise or classify A into sub-categories

- Example: rolling two dice



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

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

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 not 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 α true given 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: $\Pr[X] = \sum_{y \in X} \Pr[X = y]$

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

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: $\Pr[X] = \sum_{y \in X} \Pr[X = y]$
- $\Pr[\neg\text{cavity} \mid \text{toothache}] = \dfrac{\Pr[\neg\text{cavity} \wedge \text{toothache}]}{\Pr[\text{toothache}]}$

$$= \frac{0.016+0.064}{0.2} = 0.4$$

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

Power of Independence

- We have n random variables, $X_1, \ldots, X_n$, with domains of size d
  - How big is their joint distribution table?
    - $\underbrace{d \times d \times \ldots \times d}_{n \text{ times}} = d^n$

- Suppose that the n random variables, $X_1, \ldots, X_n$, are independent
  - How big is the joint distribution table now?
    - $\underbrace{d + d + \ldots + d}_{n \text{ times}} = dn$

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

Conditional Independence & Bayes' Rule

Conditional Independence

- Suppose that we test for pneumonia using two tests
  - Blood Test: $B \in \{\mathrm{T}, \mathrm{F}\}$
  - Throat Swab: $T \in \{\mathrm{T}, \mathrm{F}\}$

- Are they fully independent?
  - Tests were conducted independently
  - 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!

- B, T are independent *given* knowledge of underlying cause S = sick!
  - $\Pr[B \wedge T \mid S] = \Pr[B \mid S] . \Pr[T \mid S]$

Conditional Independence

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

$$\Pr[\, T_1 \wedge \cdots \wedge T_k \wedge S \,] = \Pr[\, T_1 \mid S \,].\Pr[\, T_2 \mid S \,].\ \dots\ .\Pr[\, T_k \mid S \,].\Pr[\, S \,]$$

- A joint distribution of 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: $n + 1$
  - linear!

- Conditional independence is more robust and common than absolute independence

Bayes' Rule & Conditional Independence

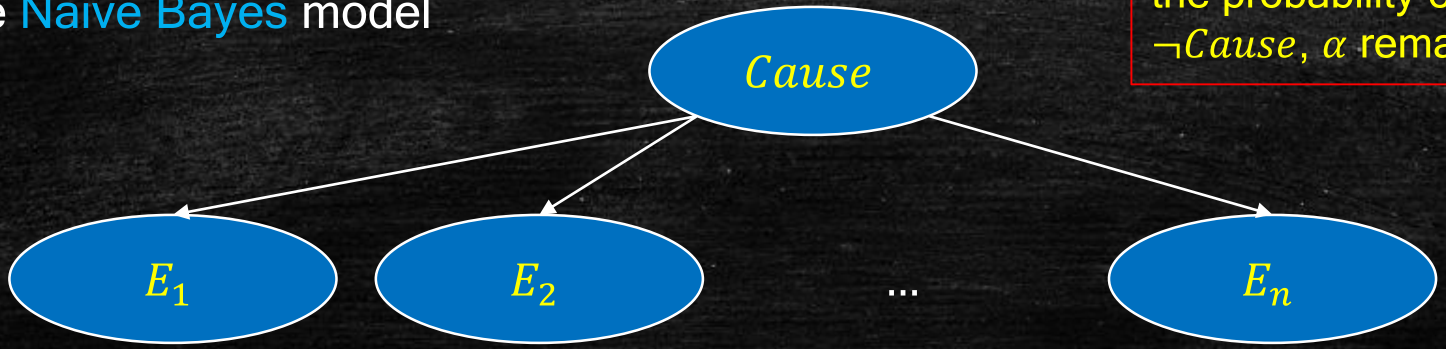
- A cause can have several conditionally independent effects
  - Cause: heavy rain
  - Conditionally independent effects: Alice brings umbrella, Bob brings umbrella, ...

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

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

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

  - This is the Naive Bayes model



Normalisation under Naive Bayes Algorithm

- Example
  - Suppose we are trying to diagnose a disease X
    - 70% of the population is $healthy$
    - 20% are $carrier$
    - 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?

Normalisation

- Example
  - What is the most likely value for X?
  - Need to determine $\Pr[\,X \mid T_1 = T_2 = 1, T_3 = 0\,] = \frac{\Pr[X]}{\Pr[T_1=T_2=1,T_3=0]}\,.\Pr[T_1 = T_2 = 1, T_3 = 0 \mid X]$
  - Notice 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 $\alpha = \frac{1}{\Pr[T_1=T_2=1,T_3=0]}$

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

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

$\Pr[X]$ $\Pr[\,T_1 = 1 \mid X\,]$ $\Pr[\,T_2 = 1 \mid X\,]$ $\Pr[\,T_3 = 0 \mid X\,]$

$$\Pr[\,X = healthy \mid TestResults\,] = \alpha \times 0.7 \times 0.1 \times 0.1 \times 0.9 = 0.0063\alpha$$
$$\Pr[\,X = carrier \mid TestResults\,] = \alpha \times 0.2 \times 0.7 \times 0.7 \times 0.3 = 0.0294\alpha$$
$$\Pr[\,X = sick \mid TestResults\,] = \alpha \times 0.1 \times 0.9 \times 0.9 \times 0.1 = 0.0081\alpha$$

Conditional Probability Tables

- Given the chain rule and conditional independence assumption

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

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

- 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 on the Lecture so far?

- Was anything unclear?
- Do you need to clarify anything?

- Channels
  - Verbally on Zoom
  - On Archipelago
  - Via Zoom Chat

OR https://archipelago.rocks/app/resend-invite/56839002759

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

Note: the chain rule implies no cycles

Edges link dependent variables

- A conditional distribution for each node given its parents

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

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

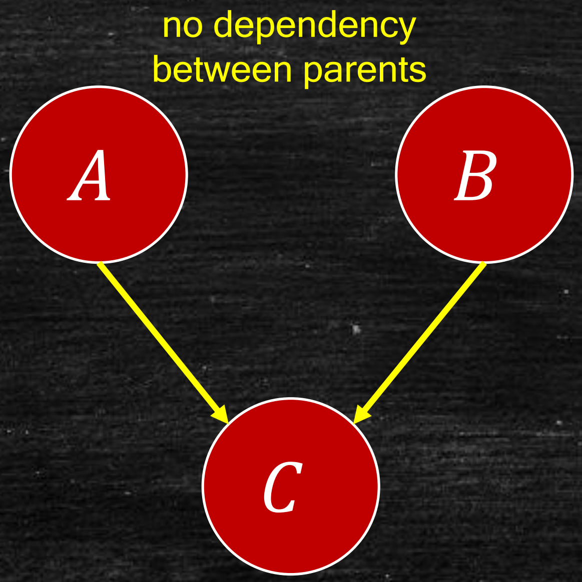
- In the simplest case, conditional distribution can be represented as a conditional probability table (CPT)
  - CPTs in the BN are the distribution over X for each combination of parent values

Relationships: Independent Events

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

Relationships: Independent Causes

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

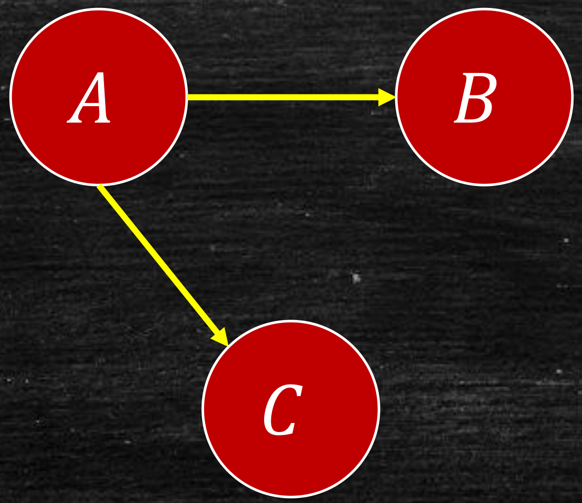


"I can be late either because of rain or because I was sick"

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

Relationships: Conditionally Independent Effects

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

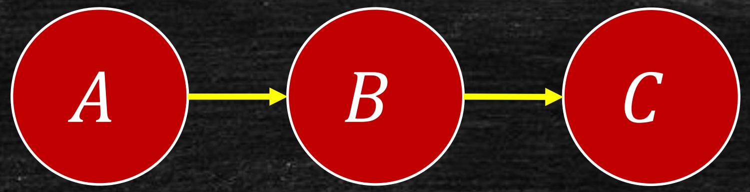


"A disease can cause two independent tests to be positive"

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

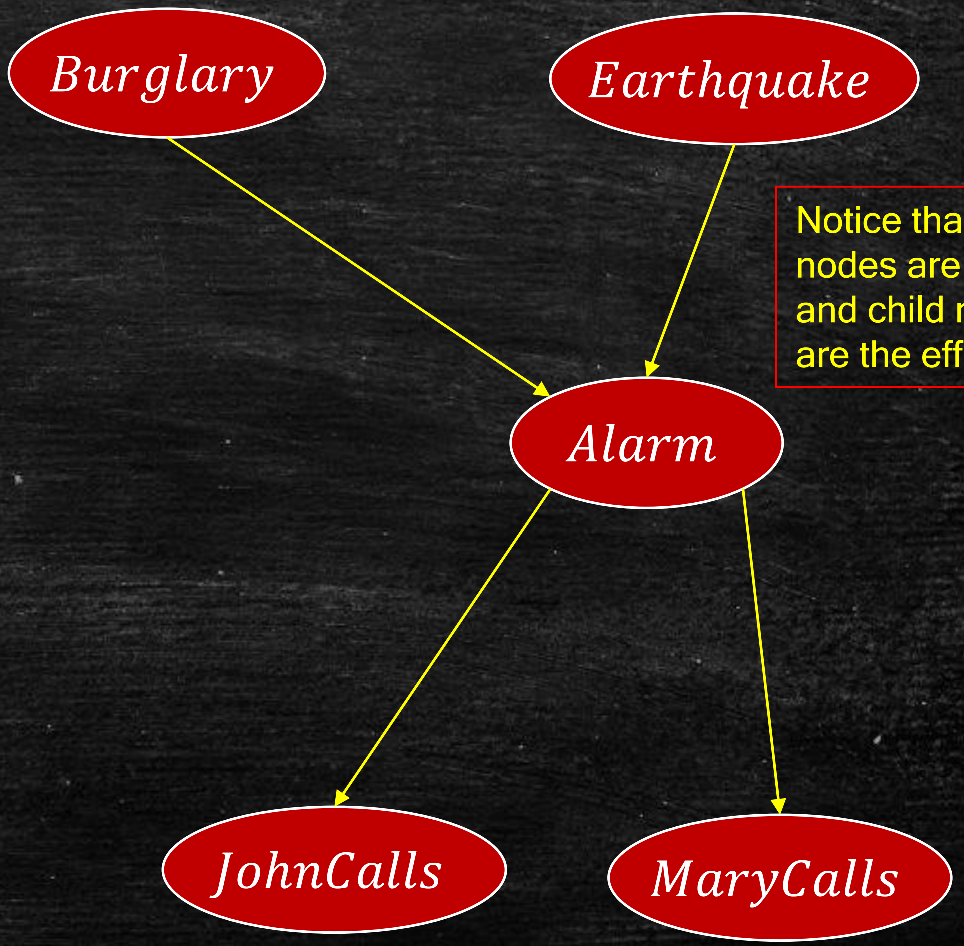
- You are out of the house...
  - J: your neighbour John calls to say your house alarm is ringing
  - M: another neighbour Mary does not call
  - A, E: Alarm sometimes set off by minor earthquake
  - B: Is there a burglar?

- 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 → 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



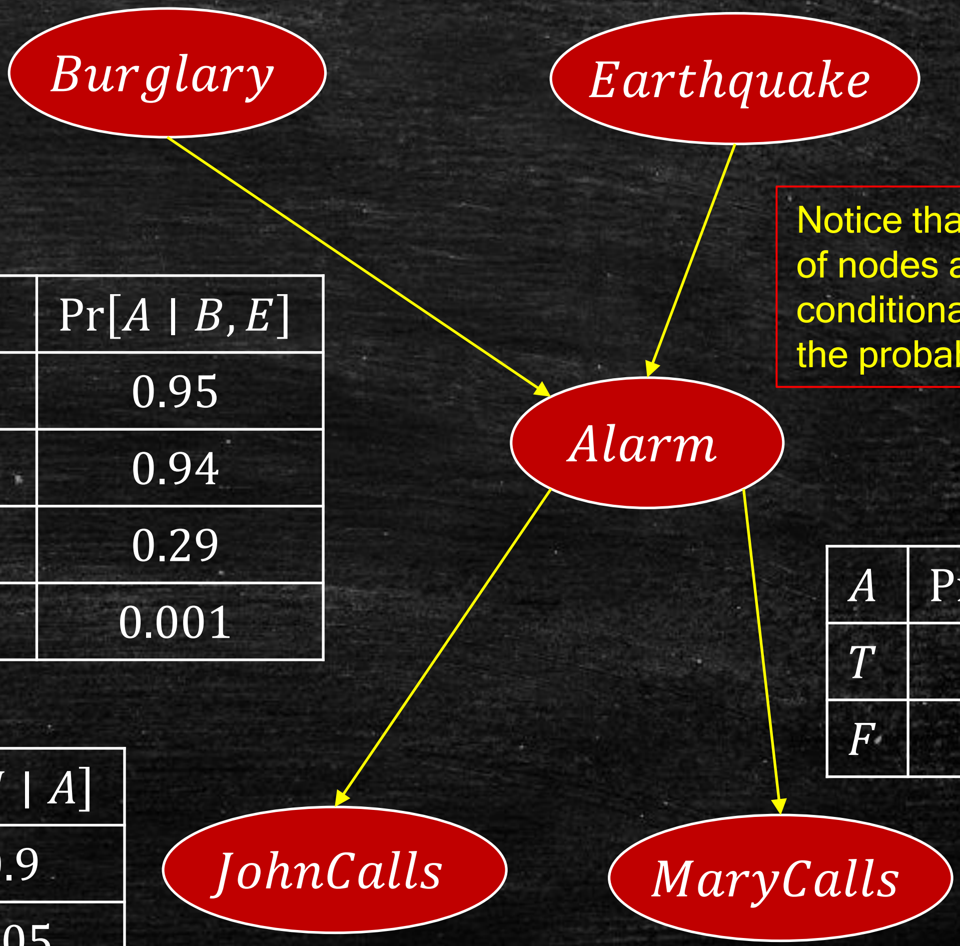
Notice that parent nodes are causes and child nodes are the effects

Example Bayesian Network

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



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

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

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

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

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

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

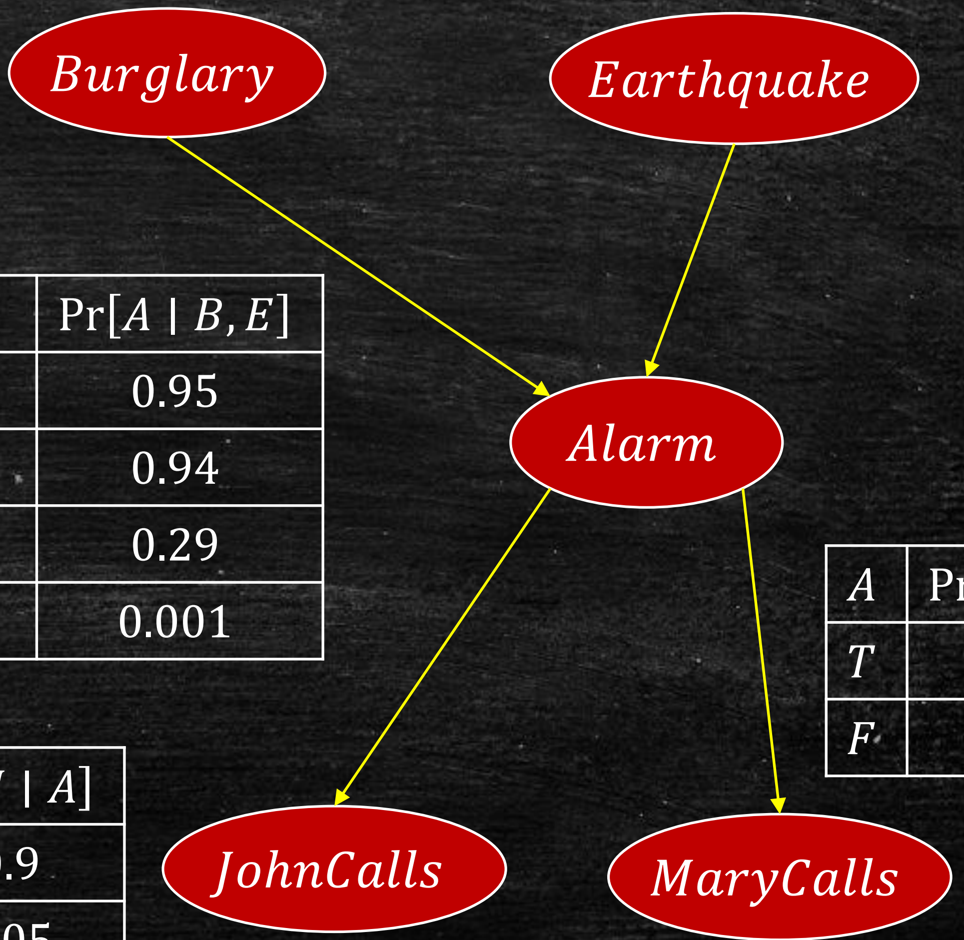
Need to also consider $A \wedge E$; so, 4 versions

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

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

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

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

- Compare the sum against similar calculations 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 $X = True$

- If each variable has $\leq k$ parents, network representation requires $\mathcal{O}(n2^k)$ 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

Constructing the Bayesian Network

Algorithm to Construct BN

1. Choose an ordering of variables $X_1, \dots, X_n$
2. For $i = 1, \dots, n$:
   - Add node X_i to the network
   - Select minimal set of parents from $X_1, \dots, X_{i-1}$ such that

     $$\Pr[\, X_i \mid Parents(X_i)\,] = \Pr[X_i \mid X_1, \dots, X_{i-1}]$$

   - Link every parent to X_i
   - Write down CPT for $\Pr[\, X_i \mid Parents(X_i)\,]$

Ordering is important, but not covered in CS3243

Determine which nodes on BN will influence X_i

Need to check all subsets - exponential!

Algorithm to Construct BN

- This construction guarantees

$$\Pr[X_1, \dots, X_n] = \prod_i \Pr[X_i \mid X_1, \dots, X_{i-1}]$$

Consequence of chain rule

$$= \prod_i \Pr[\, X_i \mid Parents(X_i)\,]$$

By choice of parents

- Network is acyclic, and has no redundancies

Algorithm assures no cycles since we only define parents for new nodes added; we never let a new node become a parent of a node that was added earlier

Given that minimum parents chosen

Summary

- We want to compute $\Pr[\, X_1 = a \mid X_2 = b \,]$

1. Bayes' rule:

$$\Pr[\, a \mid b \,] = \frac{\Pr[a, b]}{\Pr[b]} = \alpha \Pr[a, b]$$

2. Total Probability:

$$\Pr[a, b] = \sum_{x_3 \in X_3} \dots \sum_{x_n \in X_n} \Pr[a, b, x_3, \dots, x_n]$$

3. Bayesian Network Factoring:

$$\sum_{x_3 \in X_3} \dots \sum_{x_n \in X_n} \prod_{j} \Pr[\, x_j \mid Parents(X_j) \,]$$

Questions on the Lecture?

- Was anything unclear?
- Do you need to clarify anything?

- Channels
  - Verbally on Zoom
  - On Archipelago
  - Via Zoom Chat

OR https://archipelago.rocks/app/resend-invite/56839002759