CS3243: Introduction to Artificial Intelligence

Semester 2, 2020

Uncertainty

AIMA Chapter 13

Outline

- Uncertainty
- Probability
- Syntax and Semantics
- Inference
- Independence and Bayes' Rule

Uncertainty: Motivating Example

Let taxi agent's action A_t = leave for airport t minutes before flight. **Will A_t get me there on time?**

- Sources of uncertainty:
  1. Partial observability (e.g., road state, other drivers' plans, …)
  2. Noisy sensors (e.g., traffic reports, fuel sensor, …)
  3. Uncertainty in action outcomes (e.g., flat tire, accident, …)
  4. Complexity in modeling and predicting traffic (e.g., congestion)
- Logical agent either
  1. risks falsehood: "A_{25} will get me there on time", or
  2. reaches weaker conclusion: "A_{25} will get me there on time **if** there's no accident on the bridge **and** it doesn't rain **and** my tires remain intact…"

Random Variables

Domains

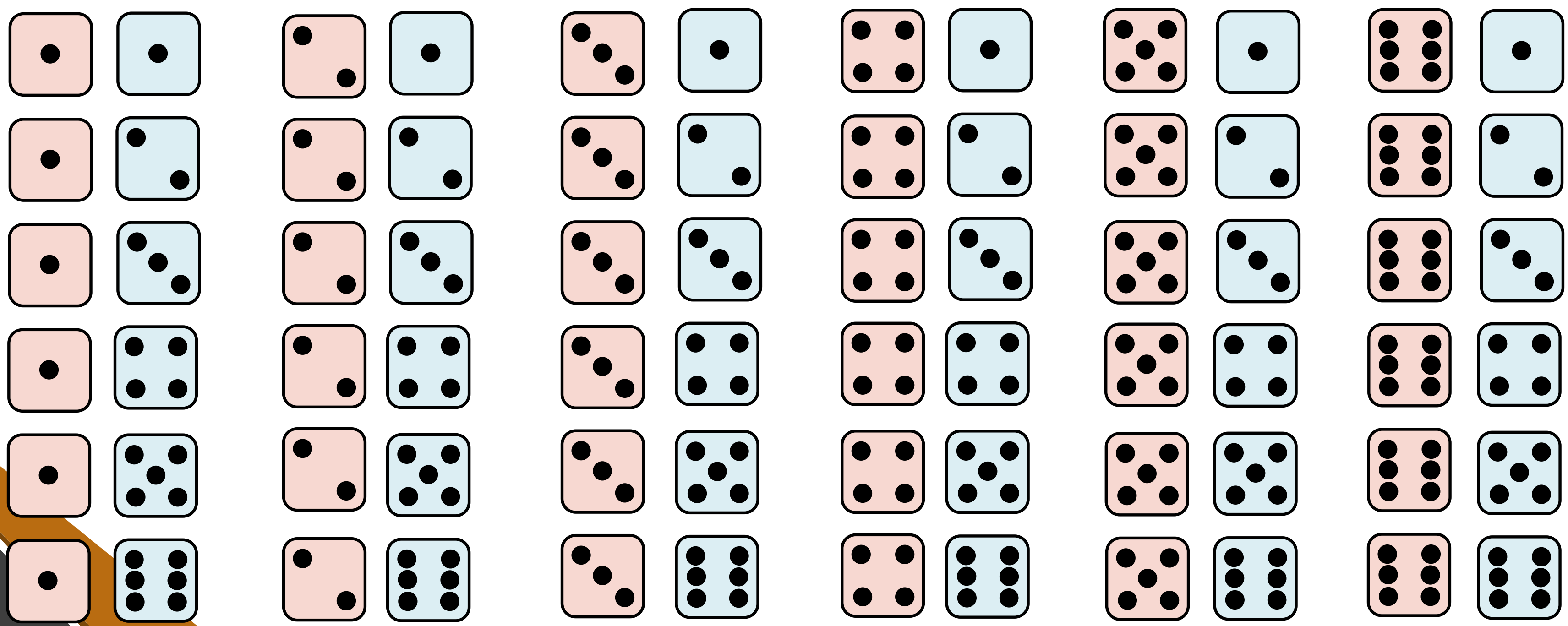
- Boolean: coin is either heads or tails (true or false)
- Discrete: a die can have values $\{1, \ldots, 6\}$

Events: subsets of domains

- $Heads(X)$ the coin flipped to heads
- $Even(X)$ the die has value $\in \{2,4,6\}$

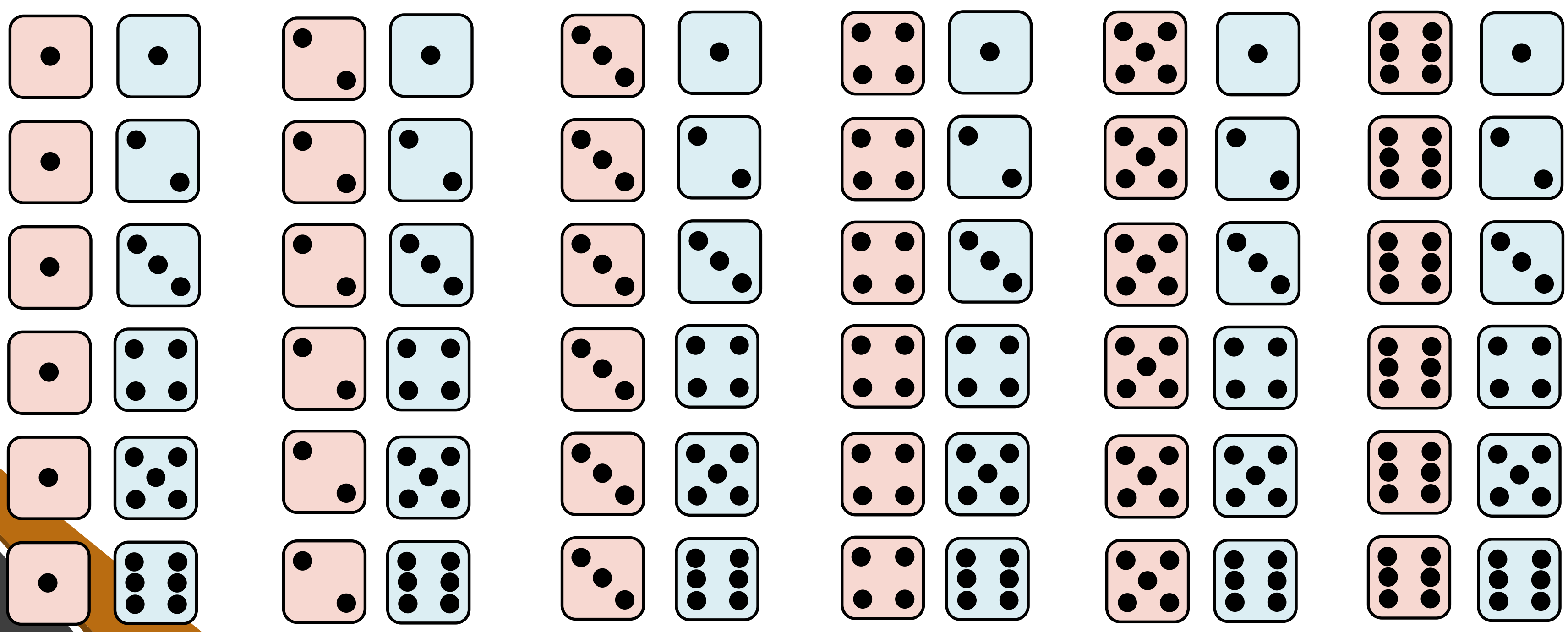
Events

- Given a random variable X, let D_X be its domain.
- **Atomic event (possible world):** an assignment of a value to each random variable; a singleton event
  - We roll two different dice



Events

- Red die = X_1, blue die = X_2
- Event: $X_1 + X_2 = 8$



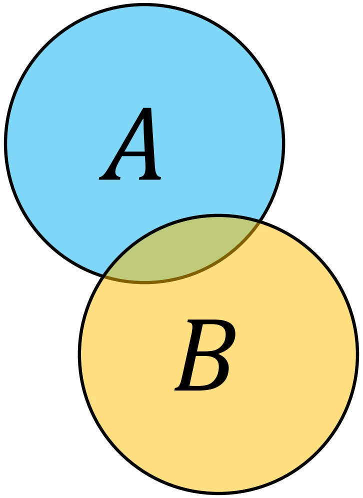
Axioms of Probability

- Let X be a random variable with finite domain D_X.
- A probability distribution over D_X assigns a value $p_X(x) \in [0,1]$ to every $x \in D_X$ s.t.

$$\sum_{x \in D_X} p_X(x) = 1$$

- For any event $A \subseteq D_X$ we have

$$\Pr[X \subseteq A] \equiv \Pr_X[A] = \sum_{x \in A} p_X(x)$$



- In particular

$$\Pr[A] + \Pr[B] = \Pr[A \cap B] + \Pr[A \cup B]$$

Joint Probability

- Given two random variables X and Y, the **joint probability** of an atomic event $(x, y) \in D_X \times D_Y$ is $p_{X,Y}(x, y) = \Pr[X = x \wedge Y = y]$
- In particular $p_X(x) = \sum_{y \in D_Y} p_{X,Y}(x, y)$

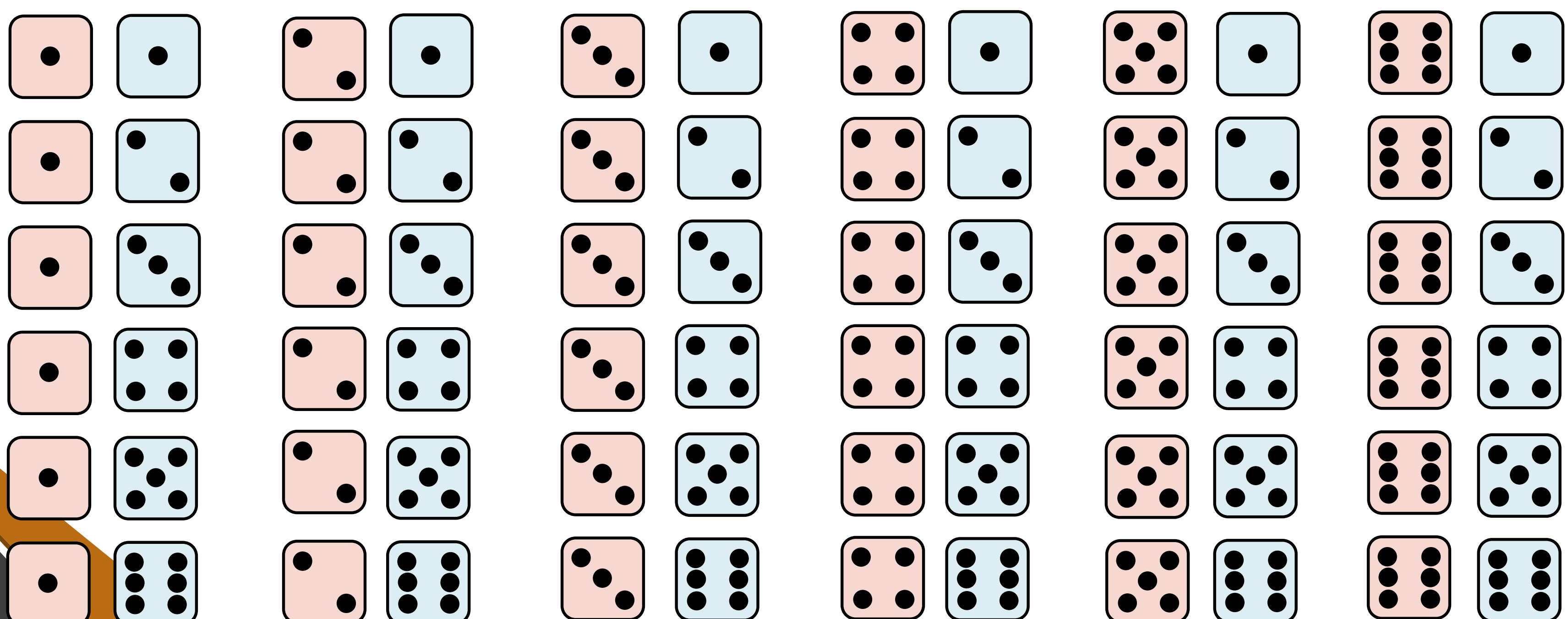
Income (in SGD)	15-24	25-34	35-44	45-54	55-64	65+
$< S\$2500$	0.062	0.051	0.037	0.019	0.015	0.039
$S\$2500 - S\$5000$	0.078	0.068	0.061	0.057	0.031	0.053
$> S\$5000$	0.015	0.051	0.094	0.119	0.111	0.039

$$\Pr[Age = (25 - 34)] = 0.051 + 0.068 + 0.051 = 0.17$$

Posterior/Conditional Probability

Probability that an event occurs, given that some other event occurs.

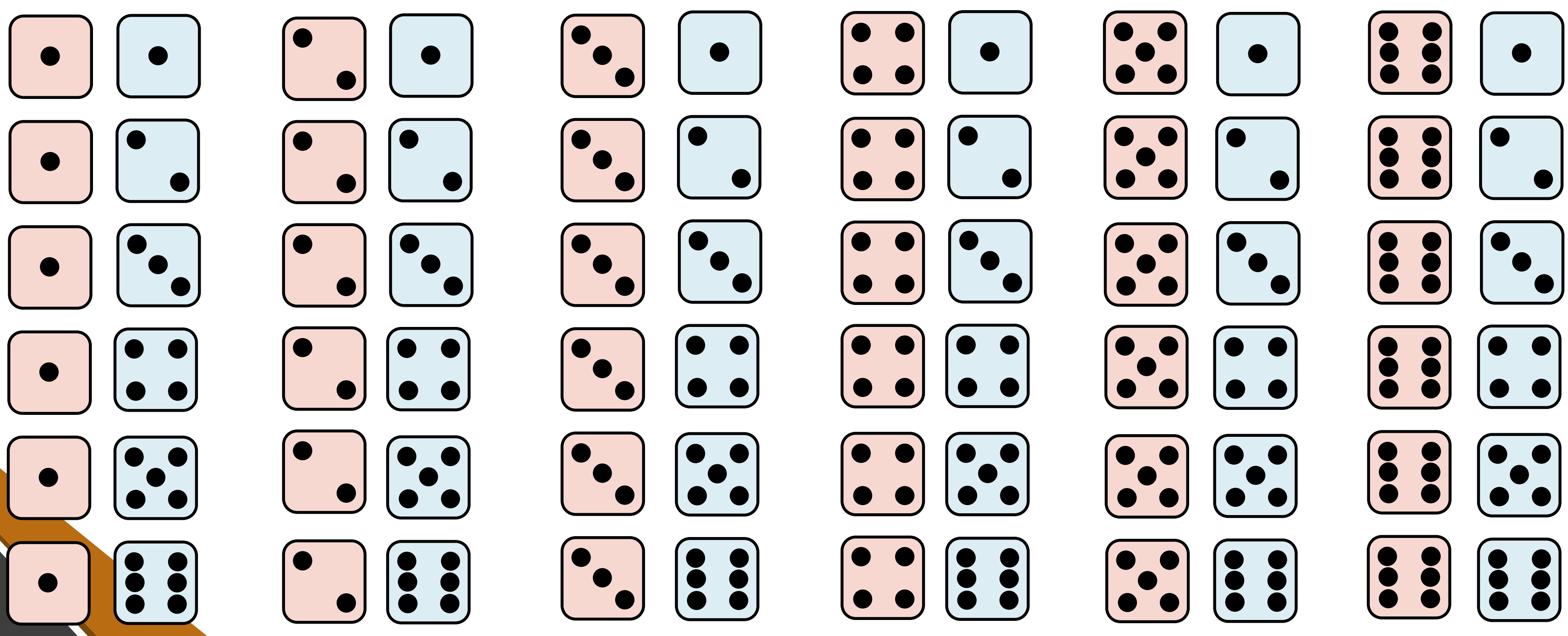
$$\Pr[X_1 = 2]$$



Posterior/Conditional Probability

Probability that an event occurs, given that some other event occurs.

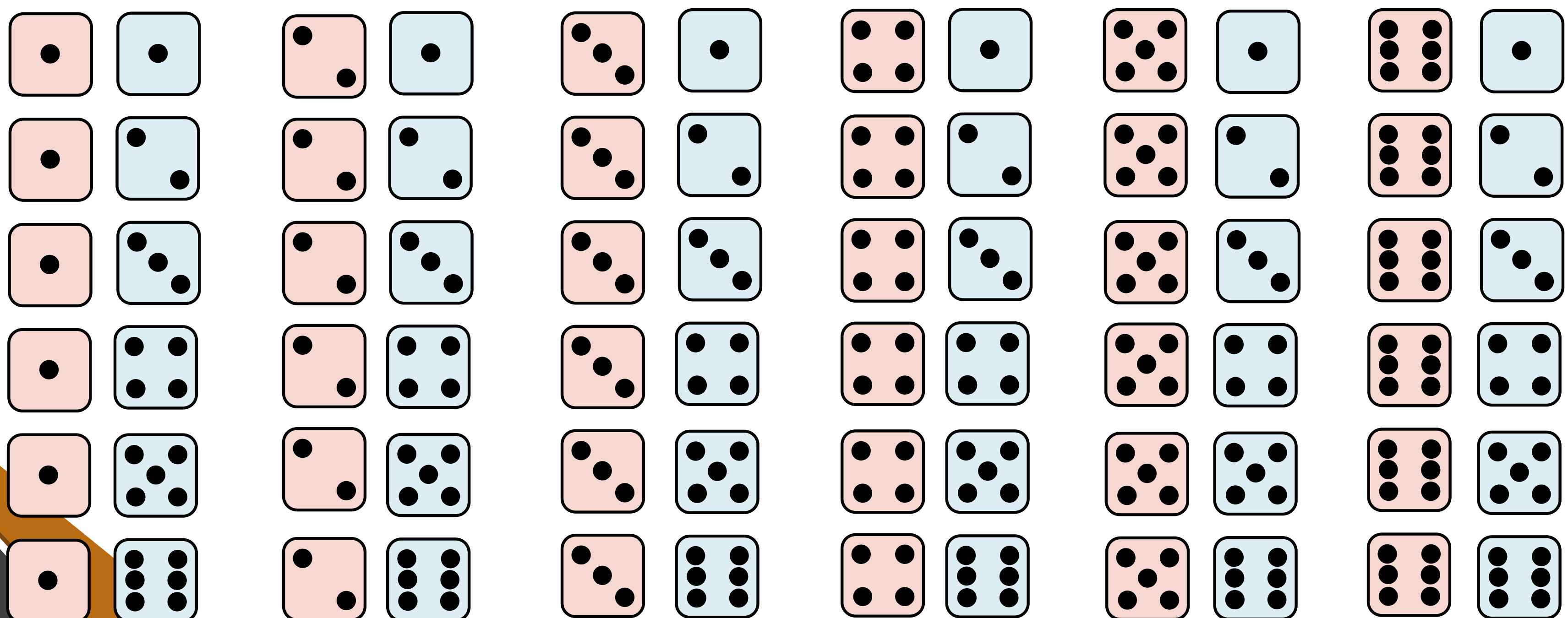
$$\Pr[X_1 = 2 \mid X_1 + X_2 = 8]$$



Posterior/Conditional Probability

Probability that an event occurs, given that some other event occurs.

$$\Pr[\, X_1 + X_2 = 8 \mid X_1 = 2 \,]$$



Posterior/Conditional Probability

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

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

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

Chain rule: derived by successive application of Bayes' rule:

$$\Pr[X_1 \wedge X_2 \wedge \cdots \wedge X_k] = \prod_{j=1,\ldots,k} \Pr\left[\, X_j \mid X_1 \wedge \cdots X_{j-1}\,\right]$$

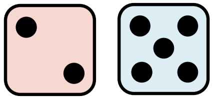
Independence

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"

Rolling two dice

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



Bayesian Inference

Instead of inferring statements of the form

'is X true given knowledge base?'

$$Y_1 \wedge \cdots \wedge Y_k \Rightarrow X?$$

we infer statements of the form

'What is the likelihood of an event X given the probabilities of other events?'

$$\Pr[\, X \mid Y_1, \ldots, Y_k \,] = ?$$

Inference by Enumeration

- Start with 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

- Start with 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: $\Pr[X] = \sum_{y \in X} \Pr[X = y]$
- $\Pr[\text{toothache} \vee \text{cavity}] =$

Inference by Enumeration

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

The Power of Independence

- We have n random variables $X_1, \ldots, X_n$; domains of size d. How big is their joint distribution table?
- Suppose that $X_1, \ldots, X_n$ are independent: maintain only dn values!
- Independence is good (if you can find it)

Conditional Independence

Suppose that we test for pneumonia using two tests

- Blood Test: B
- Throat Swab: T

- Are they fully independent?
- BUT: B, T independent given knowledge of **underlying cause** $S = \text{sick}$!

$$\Pr[\,B \wedge T \mid S\,] = \Pr[\,B \mid S\,]\Pr[\,T \mid S\,]$$

"Tests were conducted independently, only related by the underlying sickness"

Conditional Independence

- Write out full joint distribution using chain rule:
  $$\Pr[T_1 \wedge T_2 \wedge \cdots \wedge T_k \wedge S]$$
  $$= \Pr[\,T_1 \mid S\,]\Pr[\,T_2 \mid S\,] \cdots \Pr[\,T_k \mid S\,]\Pr[S]$$
- Joint distribution of n Boolean RVs: $2^n - 1$ entries.
- Conditional independence: linear!
- Conditional independence is more robust and common than absolute independence

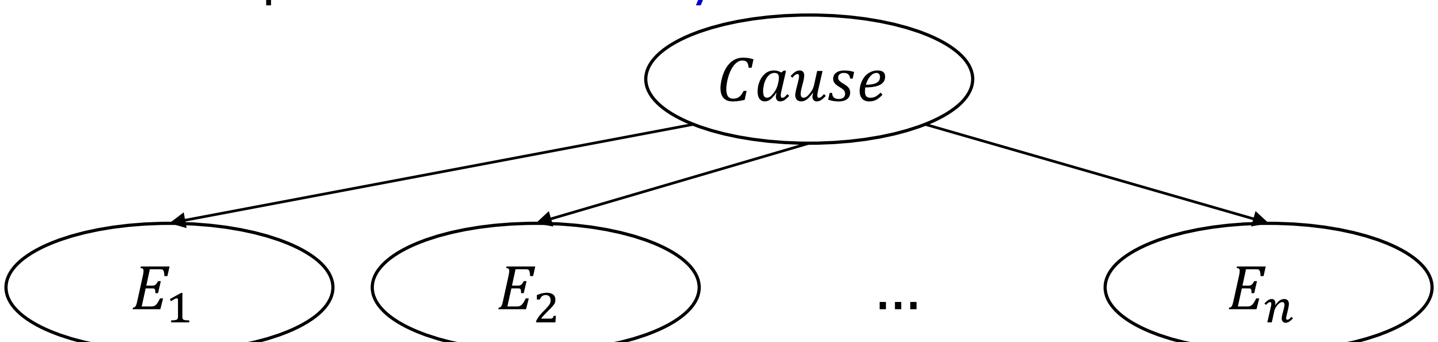
Bayes’ Rule and Conditional Independence

A cause (heavy rain) can have several conditionally independent effects (Alice takes umbrella, Bob takes umbrella, Claire takes umbrella…)

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

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

- This is an example of a naive Bayes model:



Normalization

- We are trying to diagnose the disease X. 70% of the population is healthy, 20% are carriers, and 10% are sick.
- A blood test will come back positive with the following probability:
  - $\Pr[\,T = 1 \mid X = healthy\,] = 0.1$
  - $\Pr[T = 1 \mid X = carrier] = 0.7$
  - $\Pr[\,T = 1 \mid X = sick\,] = 0.9$
- We run a test three times (independently) and obtain two positive (on tests 1 and 2) and one negative (on test 3). What is the likeliest value for X?

Normalization

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

We don't care about $\frac{1}{\Pr[T_1 = T_2 = 1, T_3 = 0]}$! Set it to α.

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

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

BAYESIAN NETWORKS

AIMA Chapter 14.1 – 14.2

Bayesian Networks

- A graphical way of writing joint distributions
- Nodes are random variables
- Edge from X to Y: X directly influences 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): the distribution over X for each combination of parent values

Bayesian Networks

Given $X_1, \ldots, X_n$, write

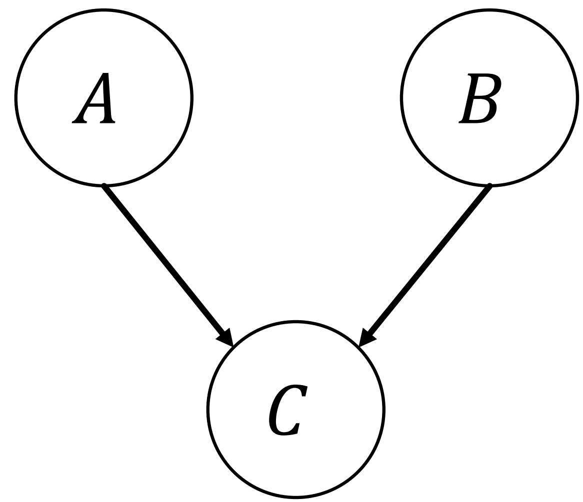
$$\Pr[X_1 \wedge \cdots \wedge X_n] = \prod_i \Pr[\, X_i \mid Parents(X_i) \,]$$

Examples

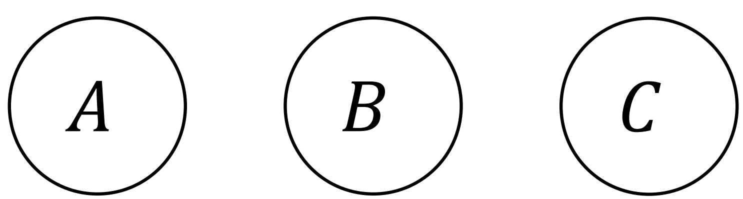
- $\Pr[A \wedge B \wedge C] = \Pr[\, C \mid A, B \,] \Pr[A] \Pr[B]$

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

(in logic: $A \vee B \rightarrow C$)



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

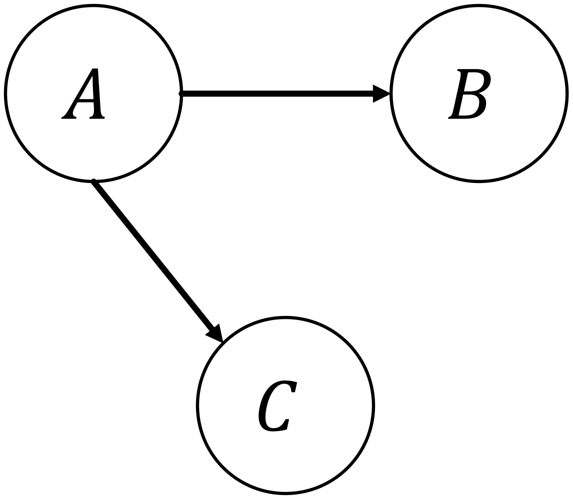


Examples

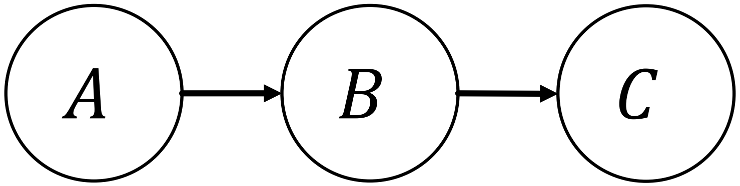
- $\Pr[A \wedge B \wedge C] = \Pr[\,C \mid A\,]\Pr[B \mid A]\Pr[A]$

Conditionally independent effects:
"A disease can cause two independent tests to be positive"

(in logic: $A \rightarrow B; A \rightarrow C$)

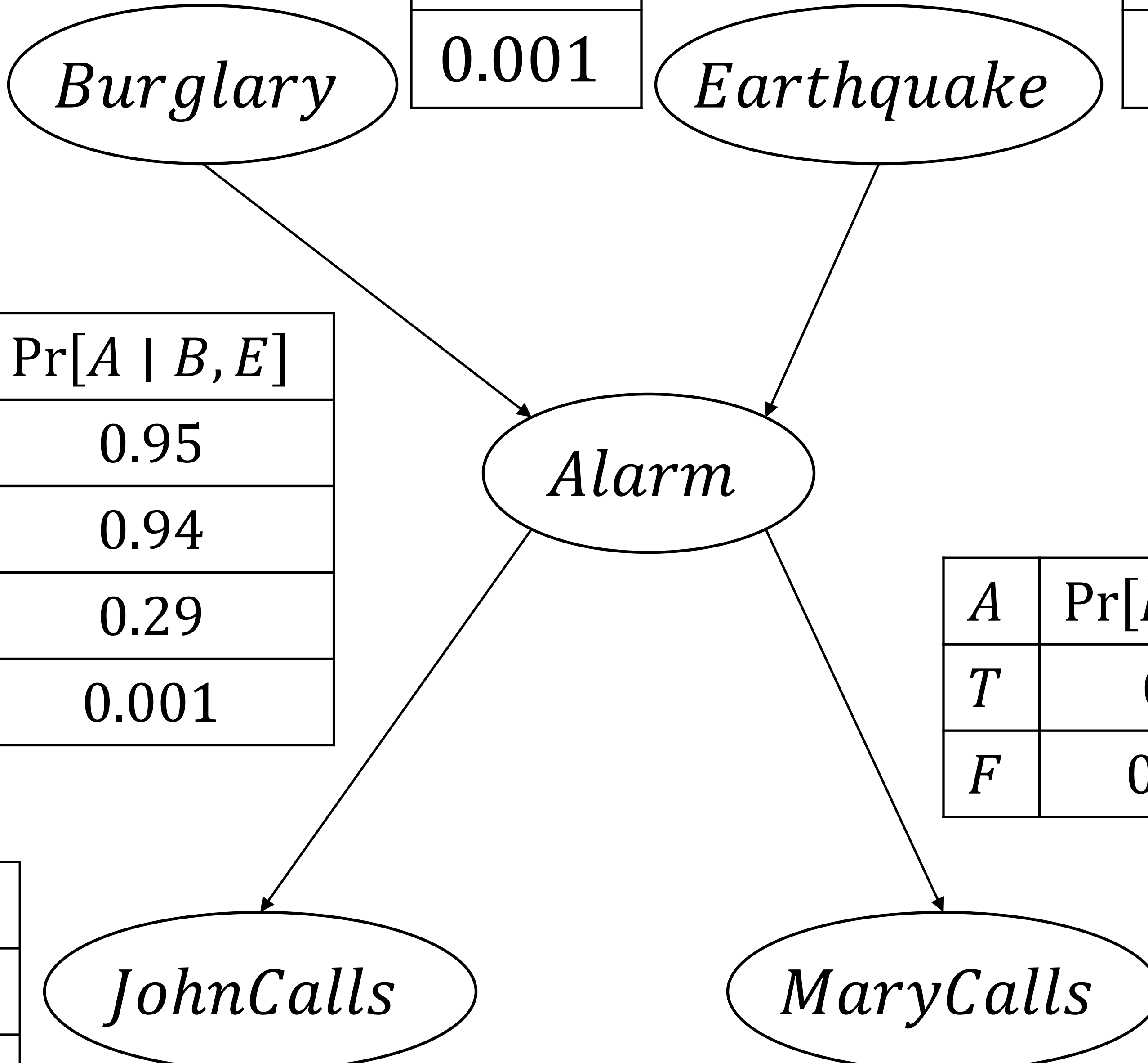


- $\Pr[A \wedge B \wedge C] = \Pr[C \mid B]\Pr[B \mid A]\Pr[A]$



Example With More Variables

- I'm at work
  - neighbor John calls to say my house alarm is ringing
  - neighbor Mary doesn't call
  - Alarm sometimes set off by minor earthquake.
  - Is there a burglar?
- Variables: B, E, A, J, M
- 5 binary variables: joint distribution table size $2^5 - 1$
- Exploit domain knowledge → smaller representation.



$\Pr[B]$
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

Bayesian Networks – Compactly Representing Joint Distributions

- Conditional probability table for Boolean 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, vs. $\mathcal{O}(2^n)$ for full joint distribution.
- For burglary network, $1 + 1 + 2 + 2 + 4 = 10$ numbers as compared to $2^5 - 1 = 31$ numbers for full joint distribution

Inference in Bayesian Networks

A Bayesian Network represents the full joint distribution; can infer any query.

$$\Pr[\, B = 1 \mid J = 1, M = 0 \,] = \frac{\Pr[B = 1, J = 1, M = 0]}{\Pr[J = 1, M = 0]} = ?$$

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

e.g.

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

Need to compute the cases $A = 0, E = 0; A = 1, E = 1; A = 1, E = 0$.

Constructing Bayesian Networks

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

Constructing Bayesian Networks

Consequence of chain rule, generally true!

This construction guarantees

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

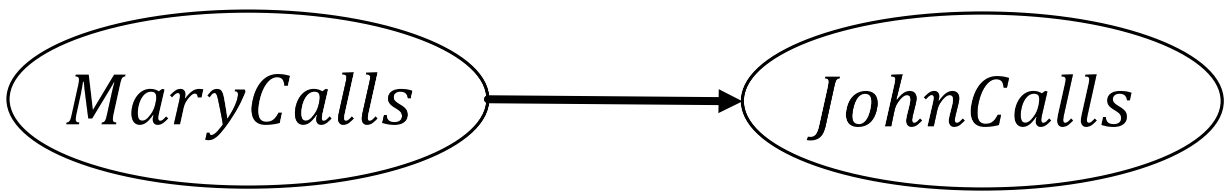
$$= \prod_i \Pr[\, X_i \mid Parents(X_i) \,]$$

By choice of parents

Network is acyclic (why??), and has no redundancies

Variable Order Matters

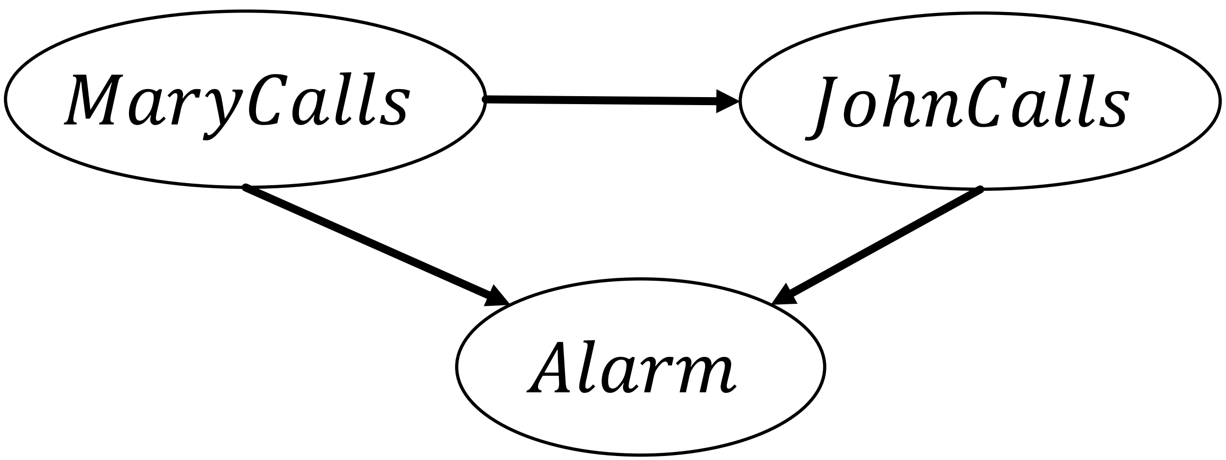
We choose the ordering M, J, A, B, E
(originally was B, E, A, M, J)



Is it true that $\Pr[J \mid M] = \Pr[J]$?

Variable Order Matters

We choose the ordering M, J, A, B, E

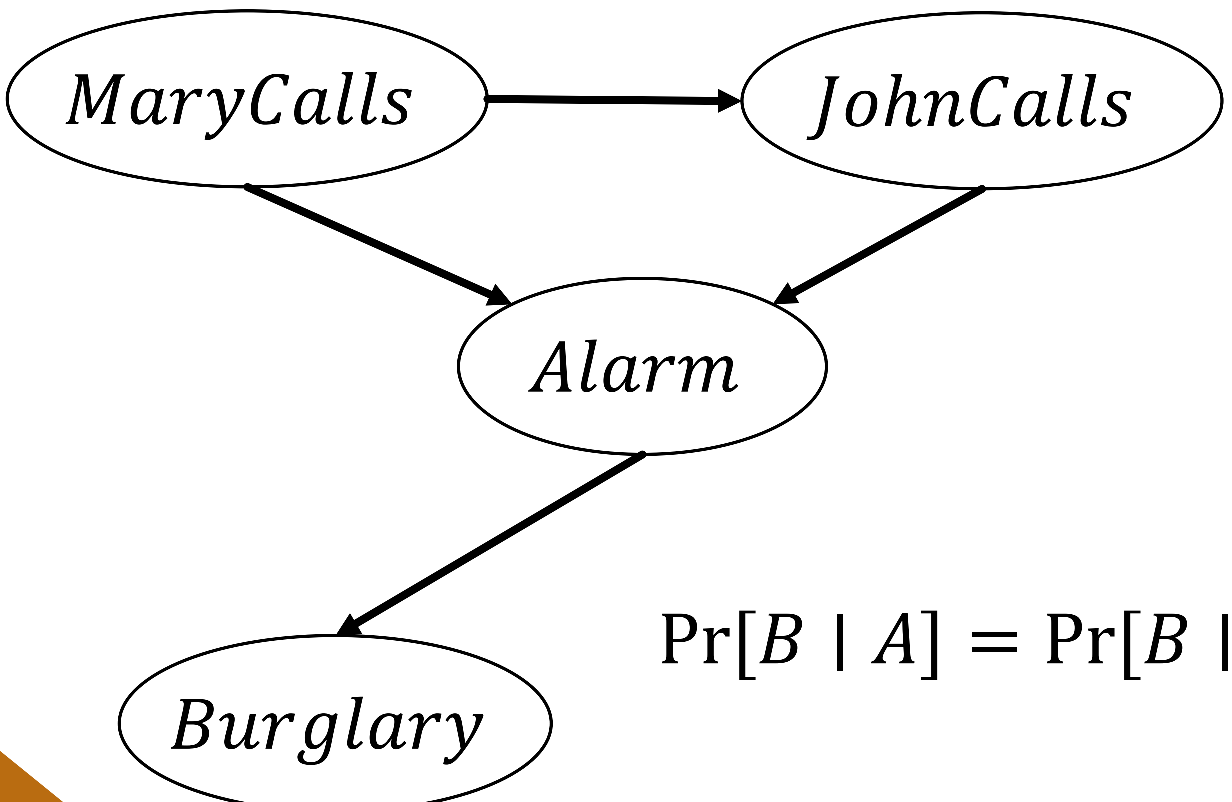


Is it true that

$$\Pr[\, A \mid M, J \,] = \Pr[A]$$
$$\Pr[\, A \mid M, J \,] = \Pr[A \mid J]$$
$$\Pr[\, A \mid M, J \,] = \Pr[A \mid M]$$

Variable Order Matters

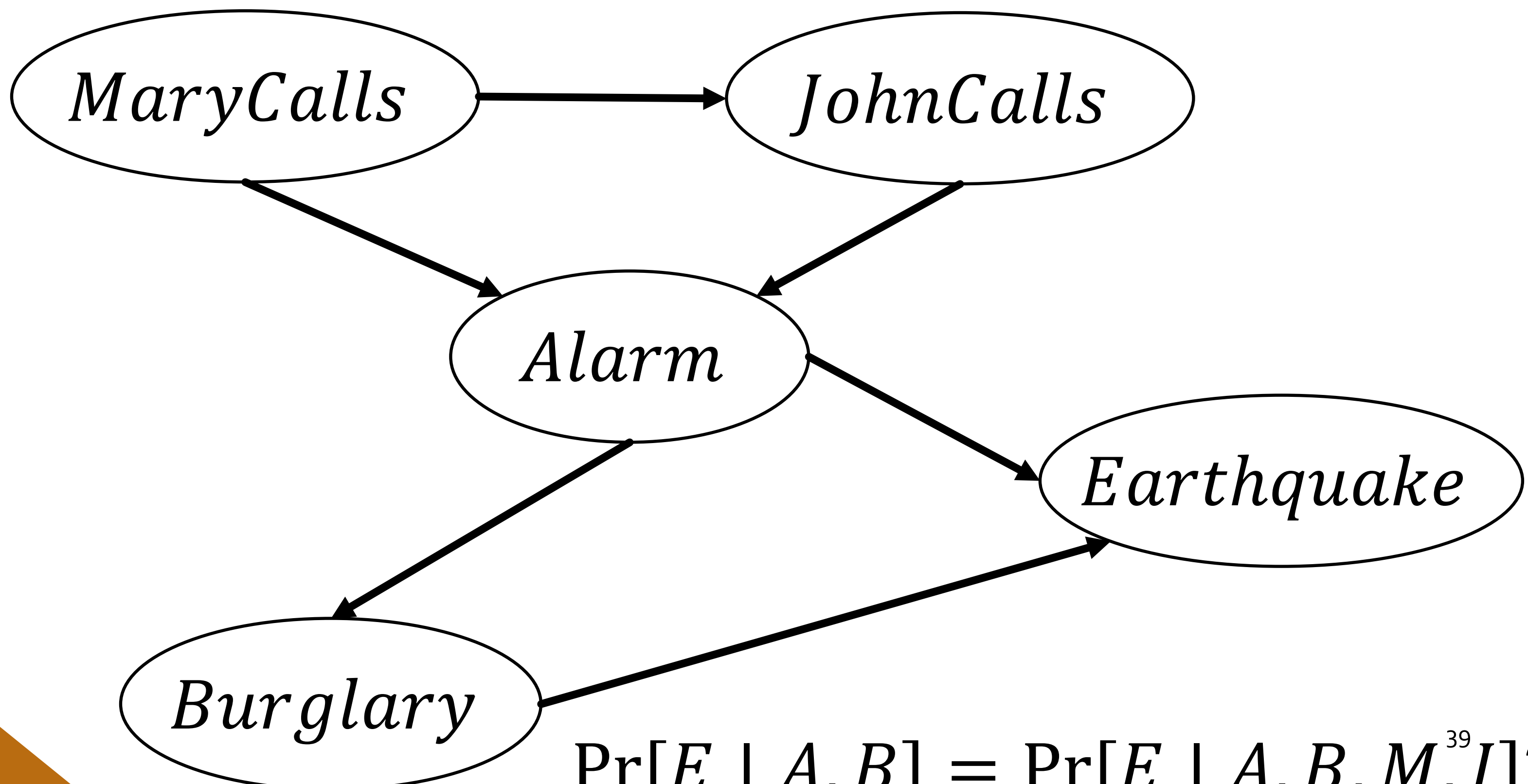
We choose the ordering M, J, A, B, E



$\Pr[B \mid A] = \Pr[B \mid A, M, J]$?

Variable Order Matters

We choose the ordering M, J, A, B, E

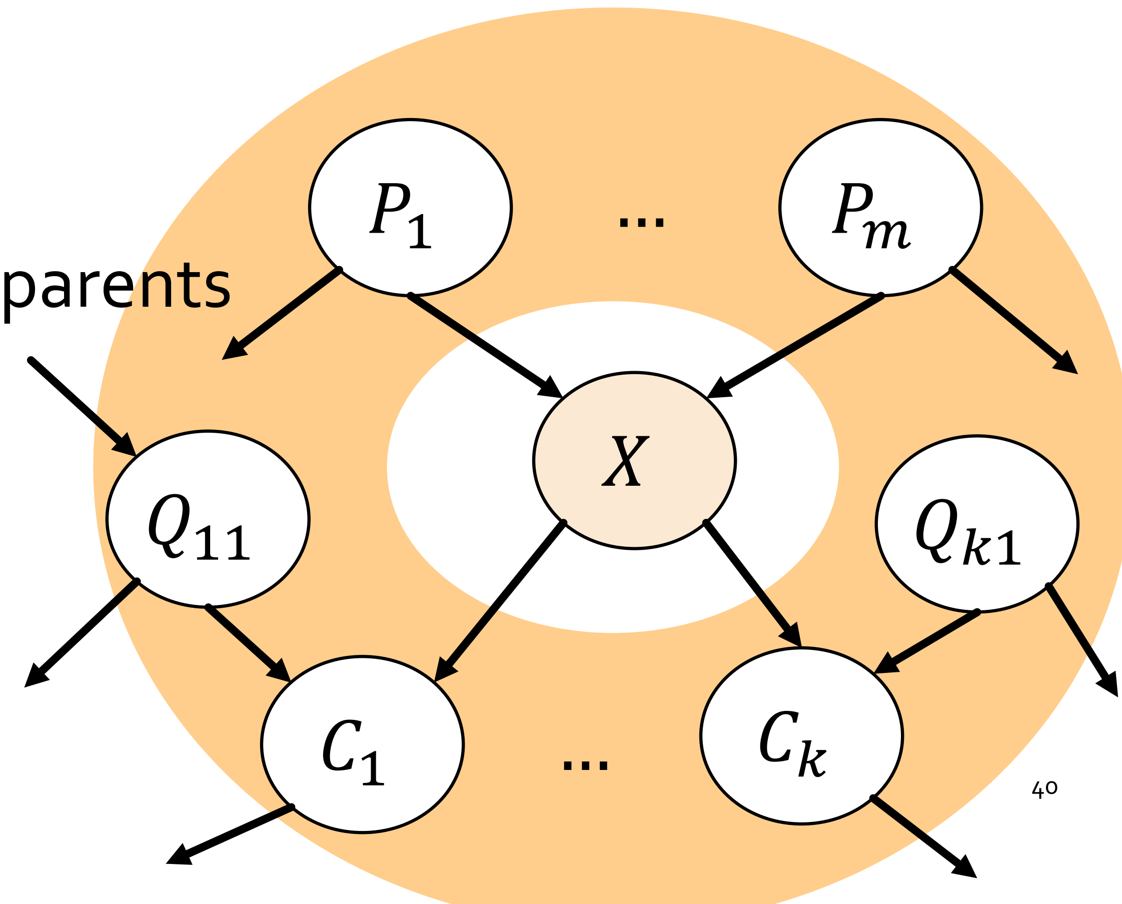


$\Pr[E \mid A, B] = \Pr[E \mid A, B, M, J]$?

The Markov Blanket

A node is conditionally independent of everything else **given the values** of its:

- parents
- children
- its children's parents



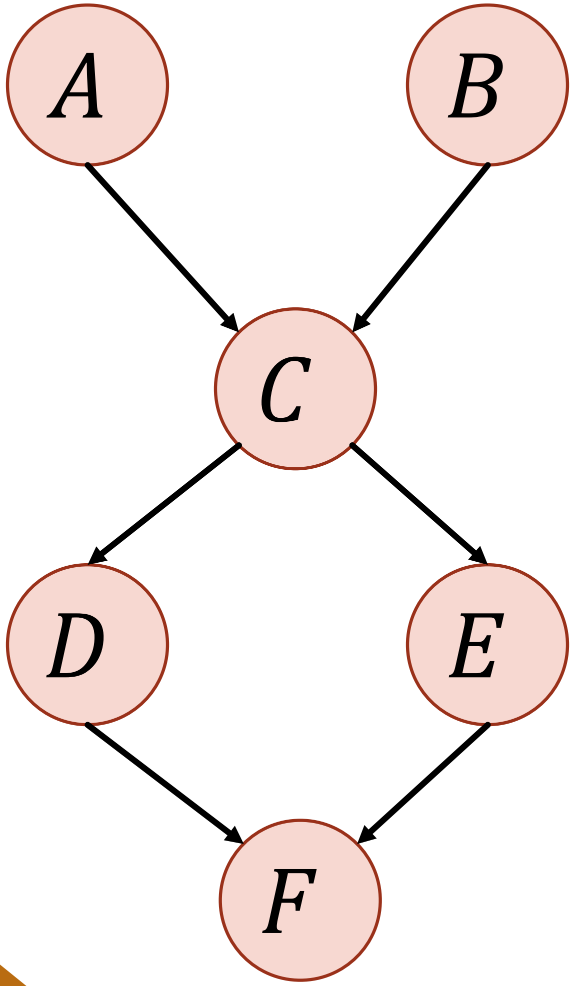
Putting it All Together

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:

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

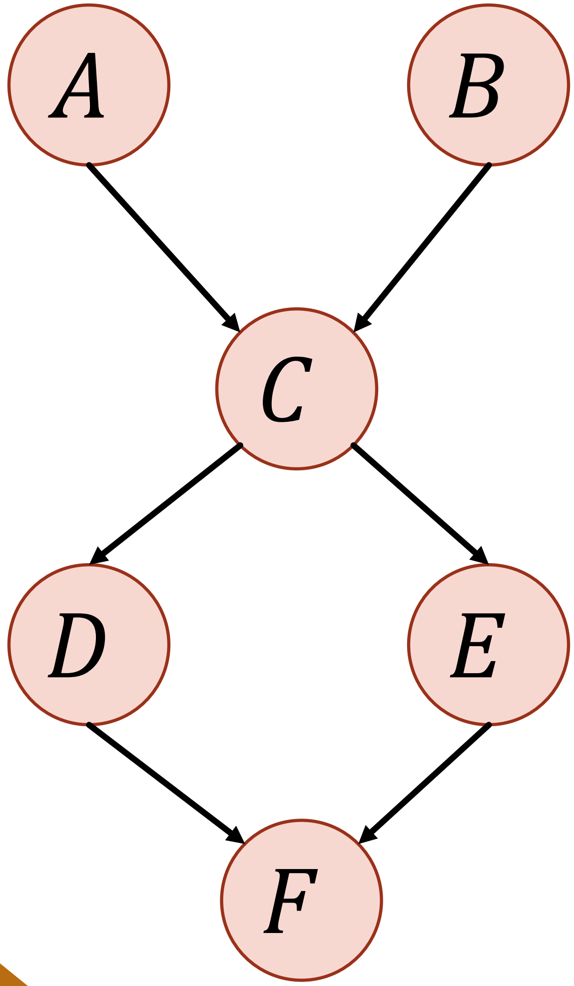
Bayesian Networks



A good way to factor joint distributions of random variables:

A variable is only conditionally dependent on its parents.

Conditional Independence in BN



Given variables X, Y and **known variables**

$$\mathcal{E} = \{E_1, \ldots, E_k\}$$

are X and Y independent given knowledge of $\mathcal{E}$?

Conditional Independence in BN

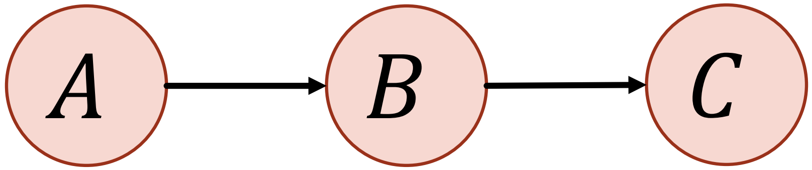
Given variables X, Y and **known variables** $\mathcal{E} = \{E_1, \dots, E_k\}$ are X and Y independent given knowledge of $\mathcal{E}$?

Can be shown

- using algebra (annoying and tedious):
$$\Pr[\, X \mid \mathcal{E} \,] = \cdots = \Pr[\, X \mid \mathcal{E}, Y \,]$$
- via counterexample (computing via the CPTs)

Can we show that two nodes are **necessarily independent?**

Causal Chains

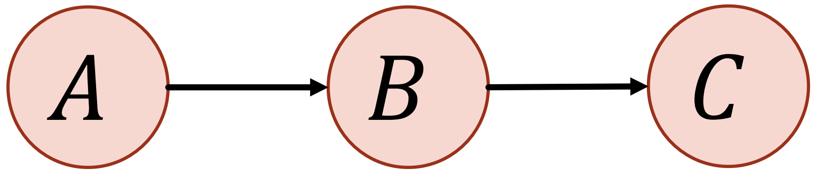


"Rain (A) causes traffic (B) which causes me to be late (C)"

Question: are A and C **necessarily independent**?

Question: are A and C **conditionally independent**, given B?

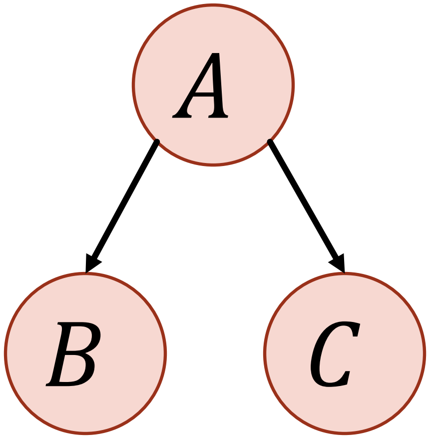
Causal Chains



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

$\Pr[\,C \mid A, B\,] = \Pr[\,C \mid B\,]$: given B, knowing A does not update my beliefs on C!

Common Cause

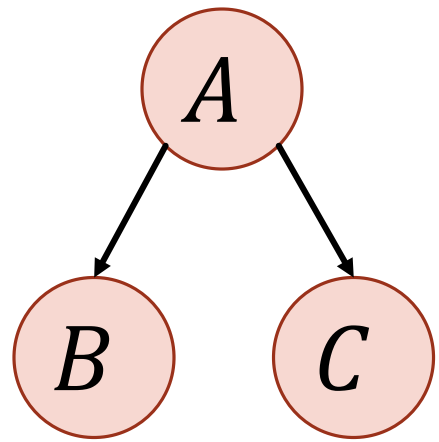


"If Batman is awake (A), he catches the Joker (B) and Bane (C)"

Question: are B and C **necessarily independent**?

Question: are B and C **conditionally independent**, given A?

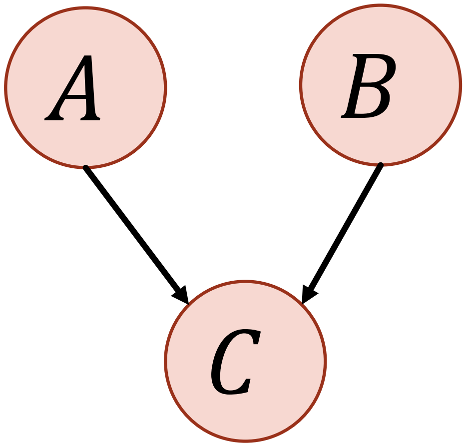
Common Cause



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

$\Pr[\,B \mid A, C\,] = \Pr[\,B \mid A\,]$: given A, knowing C does not update my beliefs on B!

Common Effect

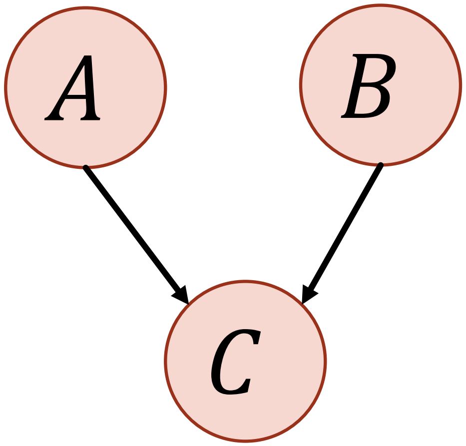


"The Joker (A) and Bane (B) could both rob the bank (C)"

Question: are A and B **necessarily independent**?

Question: are A and B **conditionally independent**, given C?

Common Effect



Observing an effect makes two causes **dependent**

- I know that the bank was robbed ($C = 1$)
- It could be either the Joker or Bane.
- If I know the Joker didn't do it –my belief about Bane doing it is higher!

$$\Pr[\,A \mid C, B\,] \neq \Pr[\,A \mid C\,]$$

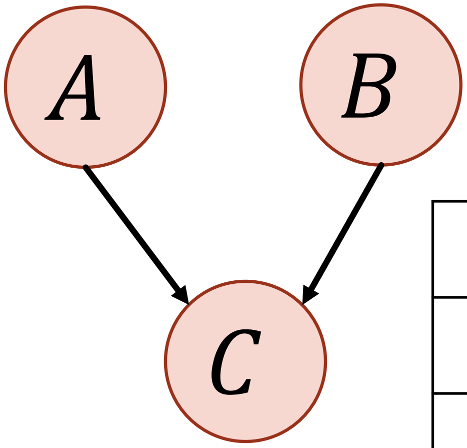
but

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

It's All About the CPTs

$\Pr[A] = 0.5$

$\Pr[B] = 0.5$



$A =$	$B =$	$\Pr[\, C \mid A, B \,] =$
1	1	1
1	0	1
0	1	1
0	0	0

$$\Pr[A = 1] = \Pr[\, A = 1 \mid B = 0 \,] = 0.5$$

but

$$\Pr[\, A = 1 \mid B = 0, C = 1 \,] = 1$$

General Case – d Separation

Given variables X, Y and **known variables**
$$\mathcal{E} = \{E_1, \ldots, E_k\}$$
are X and Y **surely** independent given $\mathcal{E}$?

Idea: any general graph can be broken down into the three cases described above, to determine conditional independence of X, Y given knowledge of $\mathcal{E}$.

General Case – d Separation

Given variables X, Y and **known variables**

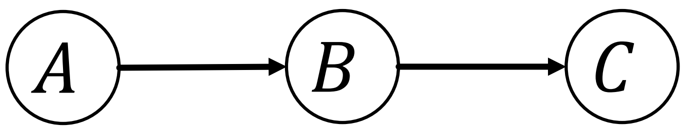
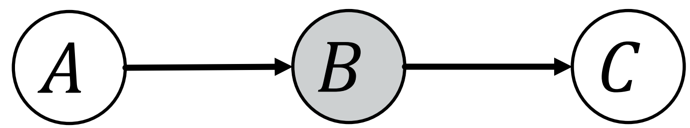
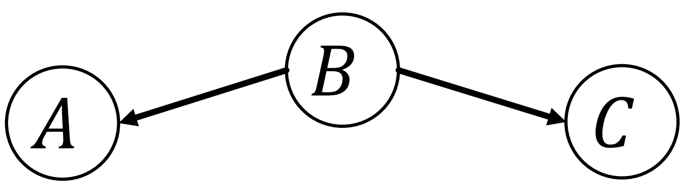
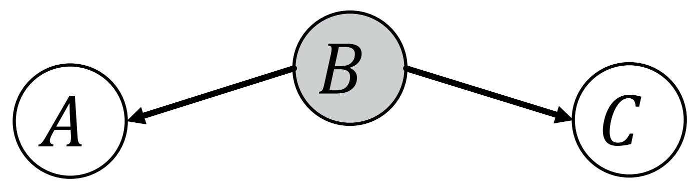
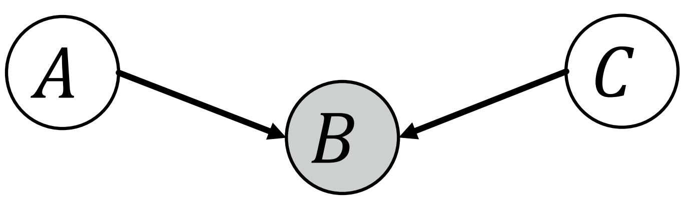
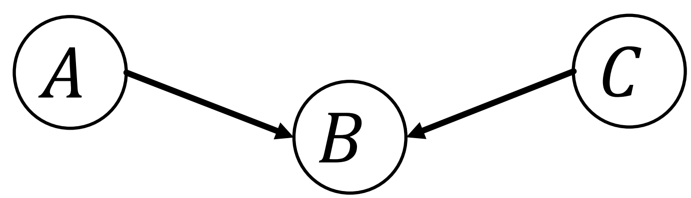
$$\mathcal{E} = \{E_1, \dots, E_k\}$$

are X and Y **surely** independent given $\mathcal{E}$?

- Check every **undirected** path between X and Y (ignore direction of arcs).
- If all paths are not **active** then X and Y are independent given $\mathcal{E}$.

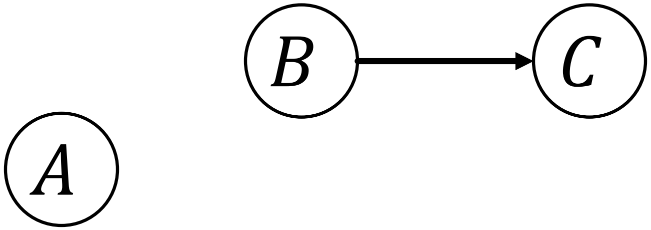
General Case – d Separation

- A path is **active** iff every triple on path is active

Active	Inactive
	
	
	

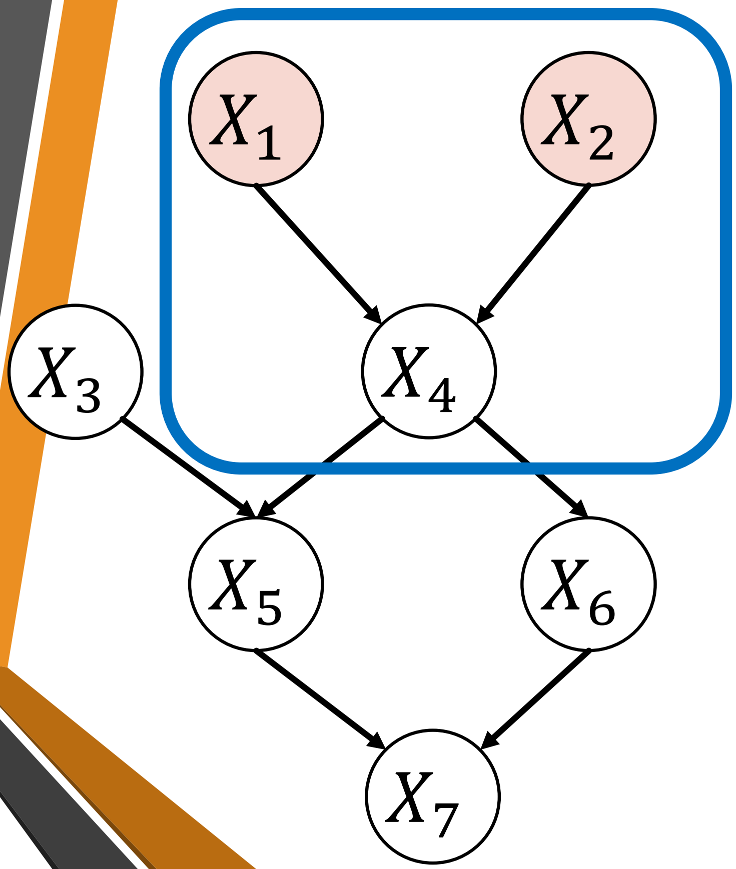
General Case – **Analyze the Graph**

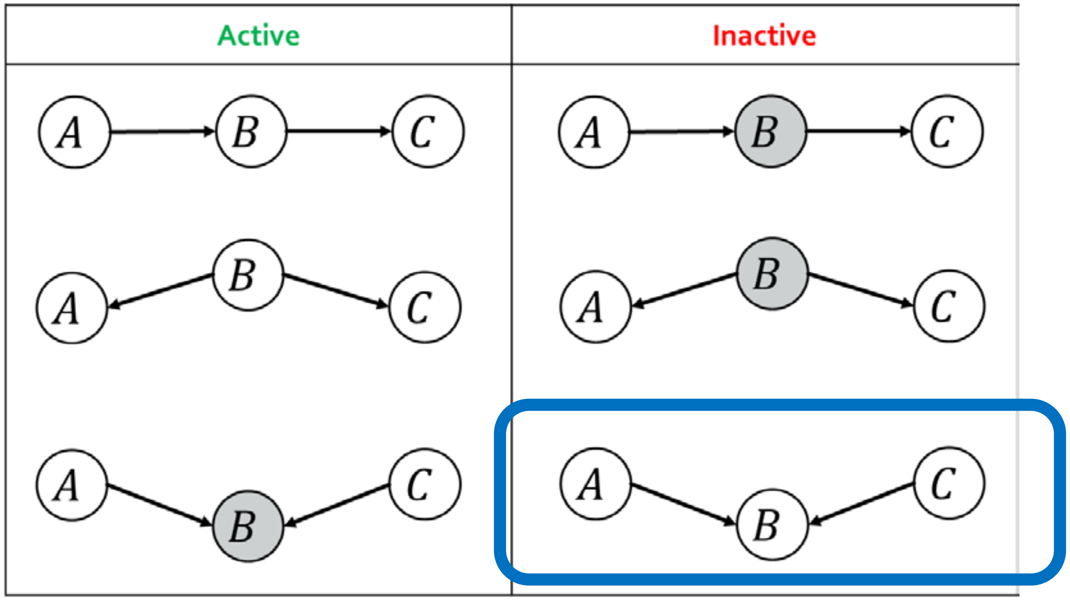
- Degenerate cases:
  - Disconnected variables: always independent.
  - Directly connected variables: never **surely** independent.



General Case – Analyze the Graph

- All paths must be inactive.
- A path is **active** iff every triple on path is active
- **One** inactive triple ⇒ path is **inactive**!



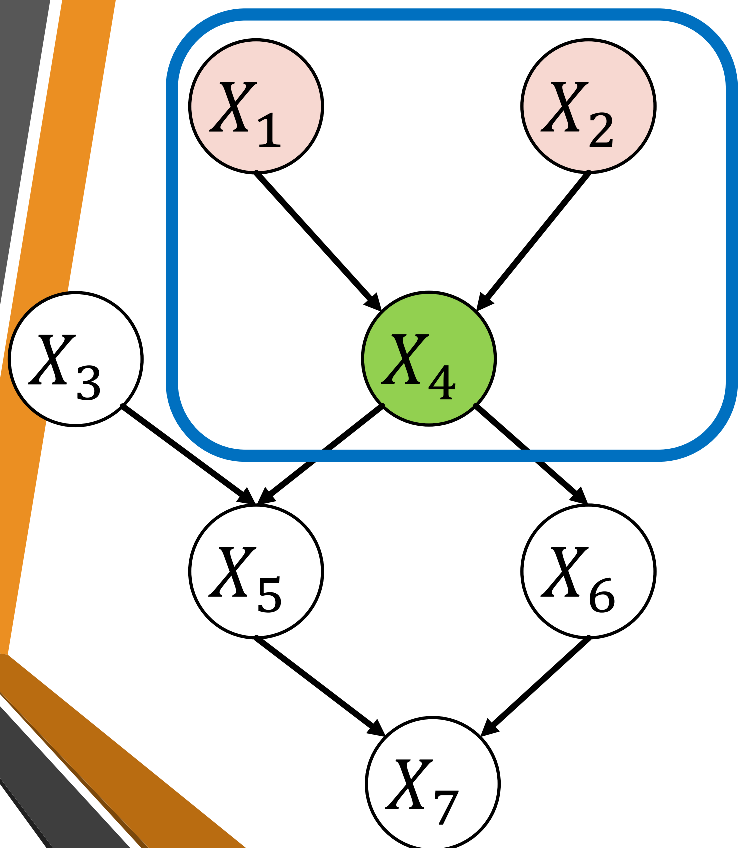


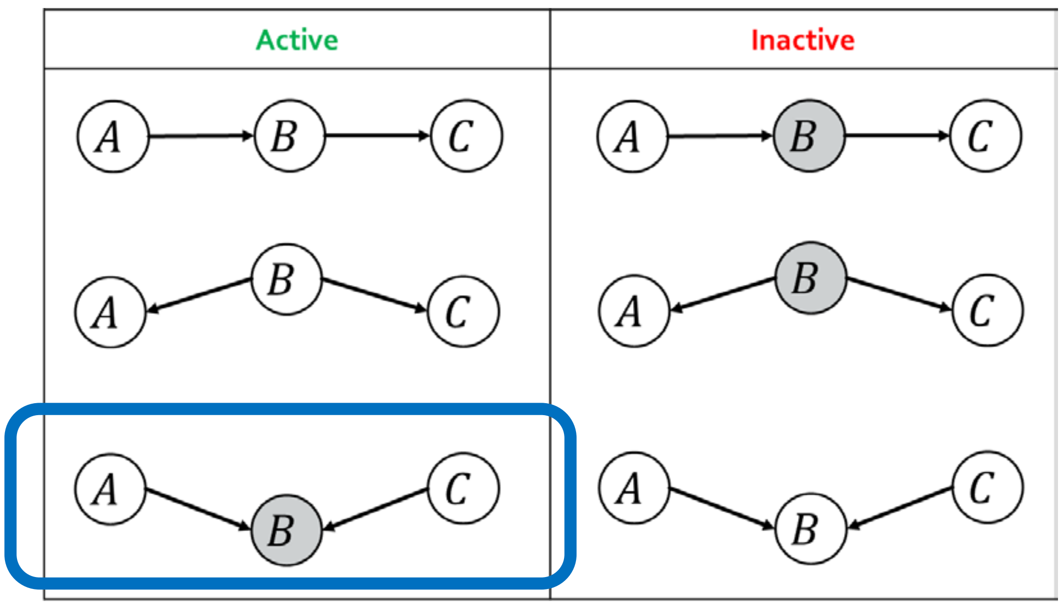
Is X_1 independent of X_2?

Yes!

General Case – Analyze the Graph

- All paths must be inactive.
- A path is **active** iff every triple on path is active
- **One** inactive triple ⇒ path is **inactive**!



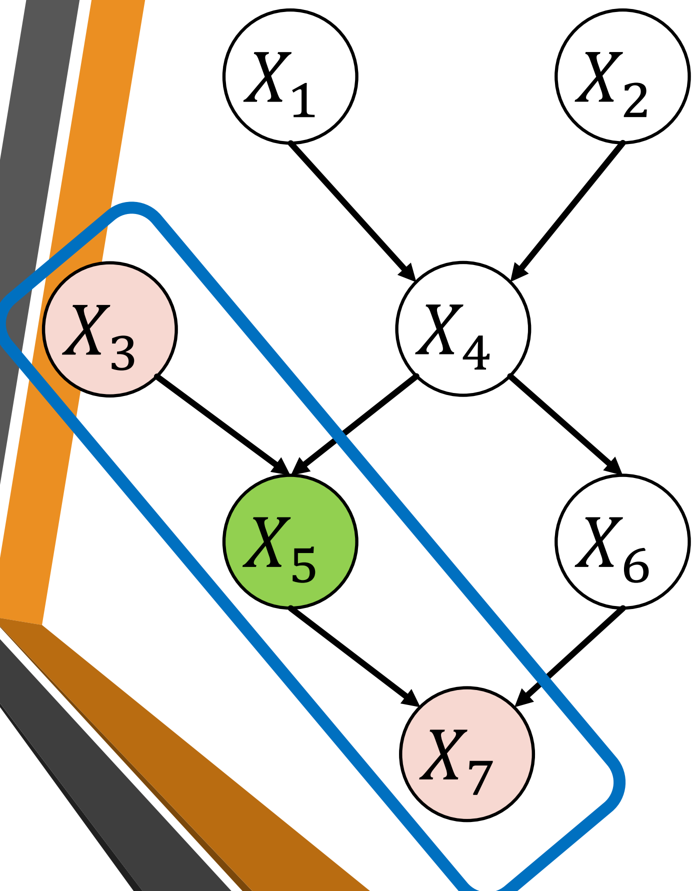


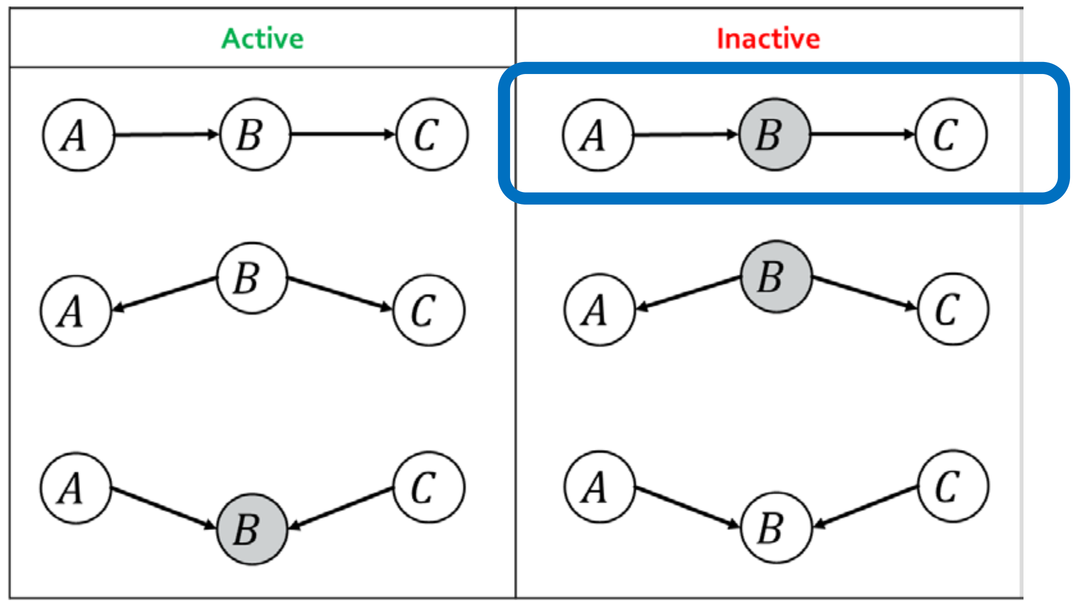
Is X_1 independent of X_2 given X_4?

No!

General Case – Analyze the Graph

- All paths must be inactive.
- A path is **active** iff every triple on path is active
- **One** inactive triple ⇒ path is **inactive**!

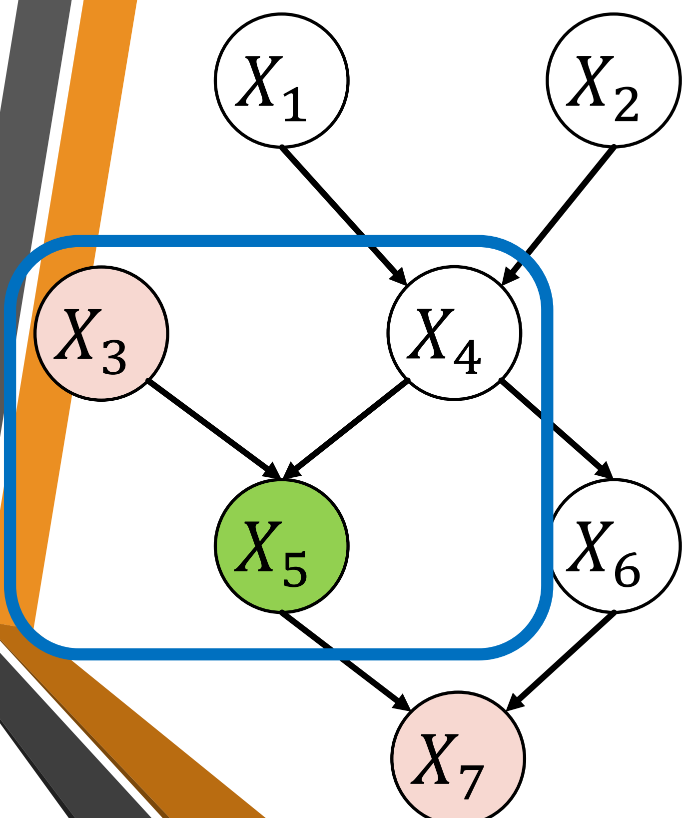


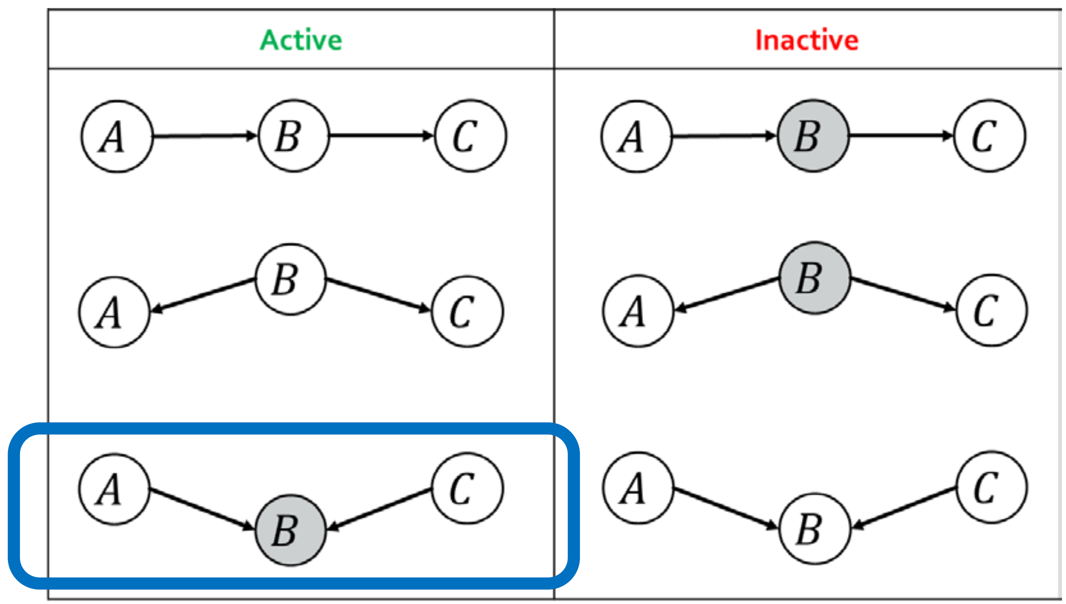


Is X_3 independent of X_7 given X_5?

General Case – Analyze the Graph

- All paths must be inactive.
- A path is **active** iff every triple on path is active
- **One** inactive triple ⇒ path is **inactive**!

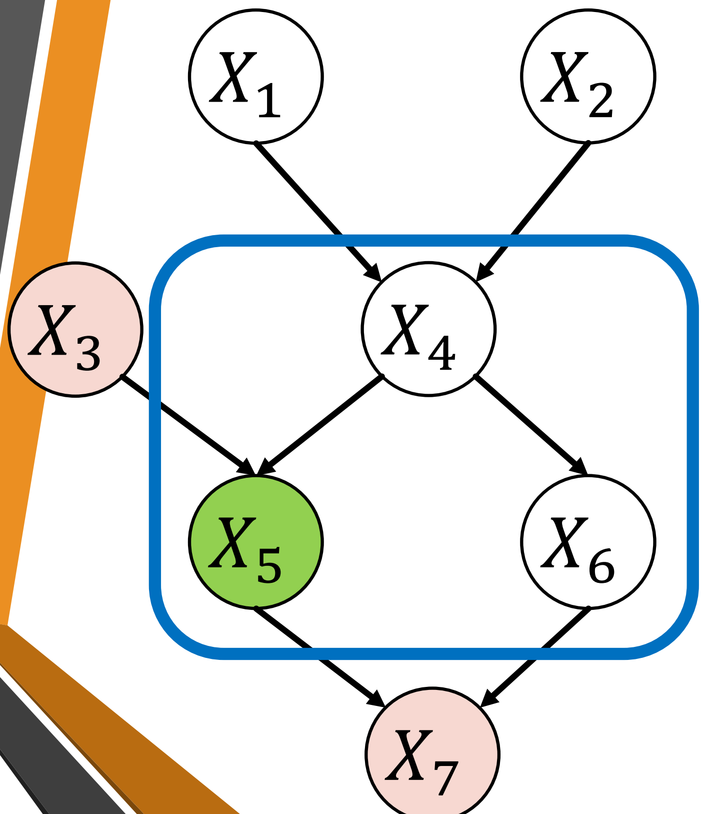


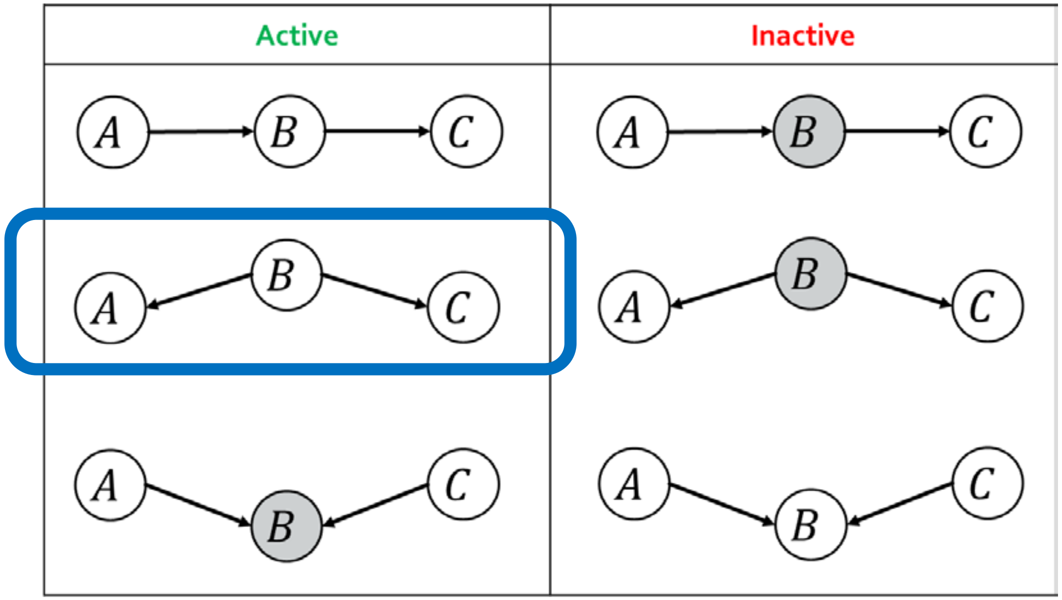


Is X_3 independent of X_7 given X_5?

General Case – Analyze the Graph

- All paths must be inactive.
- A path is **active** iff every triple on path is active
- **One** inactive triple ⇒ path is **inactive**!

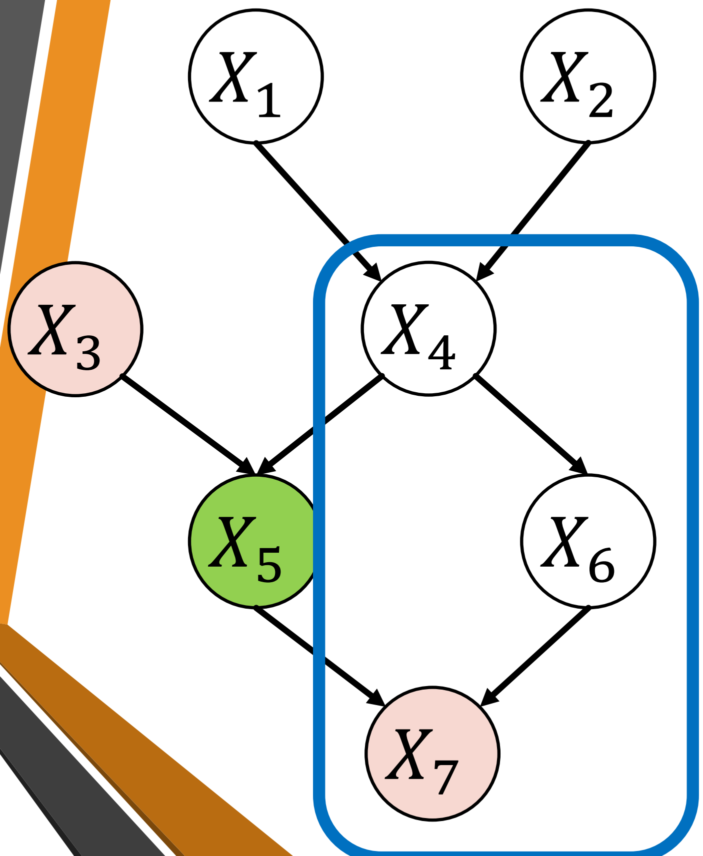


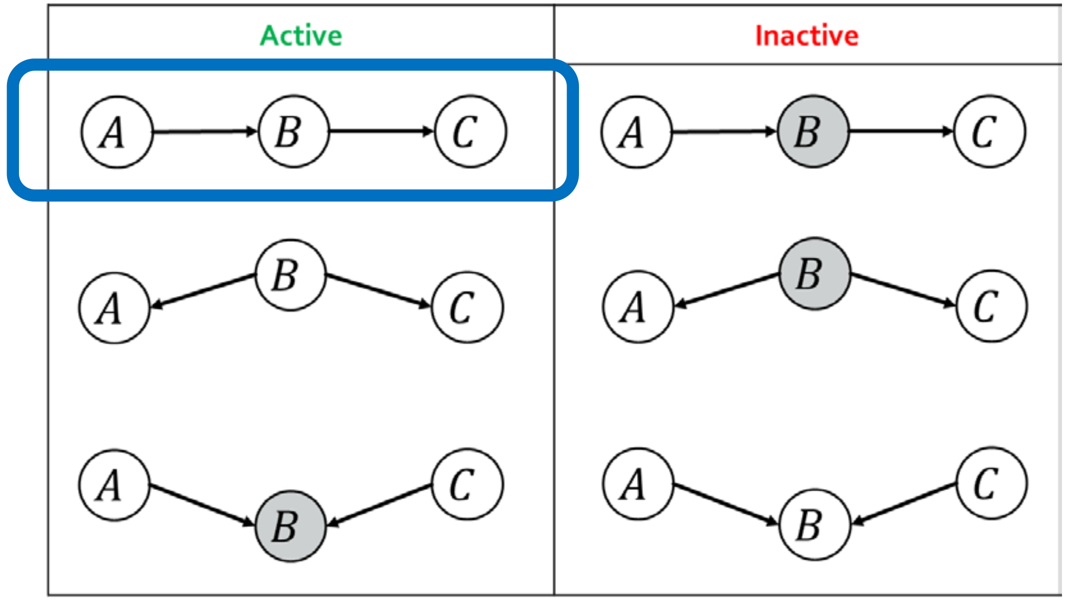


Is X_3 independent of X_7 given X_5?

General Case – Analyze the Graph

- All paths must be inactive.
- A path is **active** iff every triple on path is active
- **One** inactive triple ⇒ path is **inactive**!



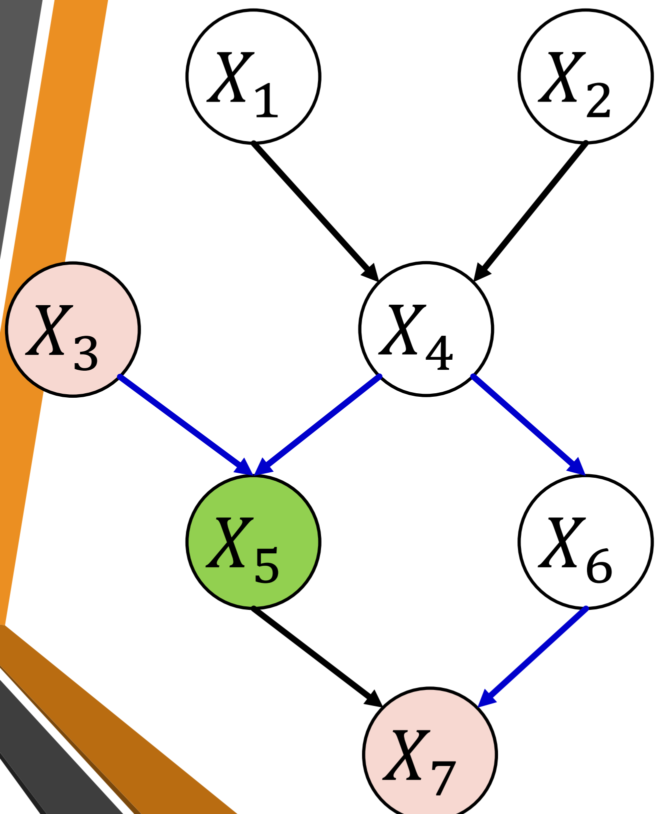


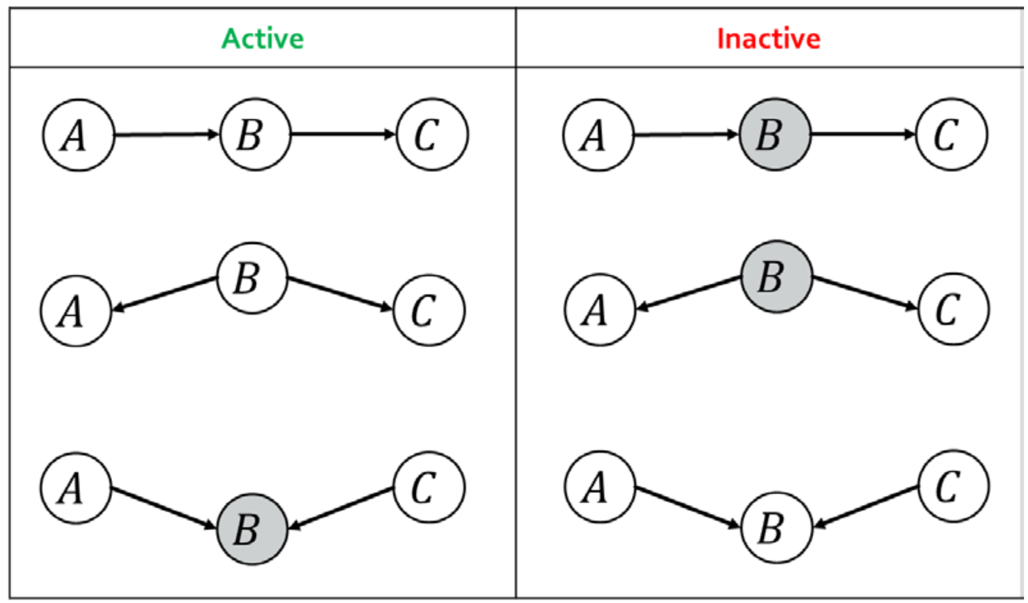
Is X_3 independent of X_7 given X_5?

No!

General Case – Analyze the Graph

- All paths must be inactive.
- A path is **active** iff every triple on path is active
- **One** inactive triple ⇒ path is **inactive**!





Is X_3 independent of X_7 given X_5?

No!

X_3, X_5, X_4, X_6, X_7 form an **active path**