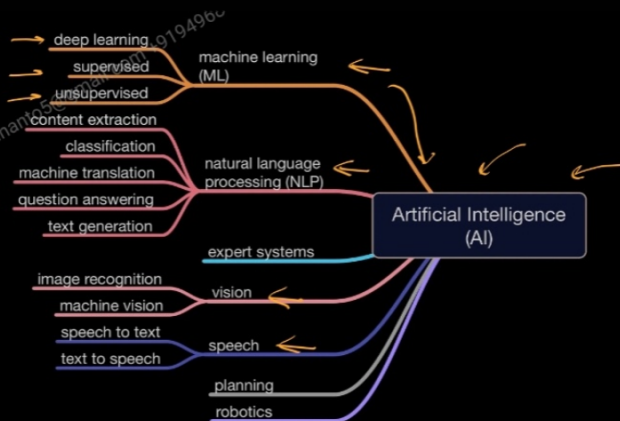


AI: Search: informed, uninformed, adversarial; logic, propositional, predicate; reasoning under uncertainty topics — conditional independence representation, exact inference through variable elimination, and approximate inference through sampling.



Bayesian NLP

Recap — Probability with example

eg) Two Teams — A & B
Each Team has 12 players

Team A → A players cheating using performance enhancing drug $P(\text{cheat}) = 0.6$

Team B → $P(\text{cheat}) = 0.1$

Q) A random player is chosen, both teams equally likely
What is the probability of him being cheater?

$C \rightarrow$ Indicator RV of cheater
 $A \rightarrow$ Indicator RV of Team A

Joint Distribution $P(C, A)$
Conditional Distribution $P(C|A)$, $P(A|C)$
Marginal Distribution $P(A)$, $P(C)$

$$\text{Marginal } P(A) = \sum_{k=0}^1 P(C=k, A) = \sum_{k=0}^1 P(C=k) \cdot P(A|C=k) = P(A) \cdot \left[\sum_{k=0}^1 P(C=k|A) \right] = 1$$

$$\text{Marginal } P(C) = \sum_{k=0}^1 P(C|A=k) = \sum_{k=0}^1 P(A=k) \cdot P(C|A=k) = P(C) \cdot \left[\sum_{k=0}^1 P(A=k|C) \right] = 1$$

$$\text{Conditional } P(A|C) = \frac{P(C, A)}{P(C)} = \frac{P(A) \cdot P(C|A)}{P(C)}$$

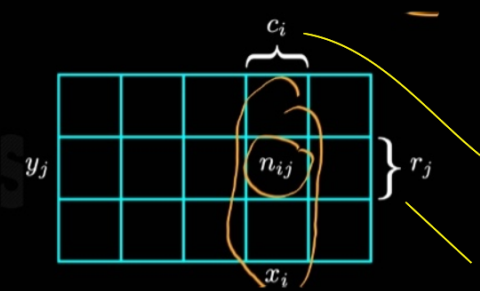
$$\text{Conditional } P(C|A) = \frac{P(C, A)}{P(A)} = \frac{P(A) \cdot P(C|A)}{P(A)}$$

$$\left. \begin{aligned} P(A, C) &= P(A|C) \cdot P(C) \\ &= P(C|A) \cdot P(A) \end{aligned} \right\} \text{chain rule}$$

Joint = Conditional · Marginal

	Cheater	Not a Cheater	
Team A	$0.5 \times 0.6 = 0.3$	$0.5 \times 0.4 = 0.2$	
Team B	$0.5 \times 0.1 = 0.05$	$0.5 \times 0.9 = 0.45$	
	$P(C)$	$P(\bar{C})$	

	Cheater	Not a Cheater	
Team A	0.3	0.2	$P(A)$
Team B	0.05	0.45	$P(B)$
	$P(C)$	$P(\bar{C})$	



$$n_{ij} = P(Y=y_j, X=x_i) \text{ Joint}$$

$$P(X=x_i) = c_i = \sum_y P(X=x_i, Y=y) \text{ Marginal}$$

$$P(Y=y_j) = r_j = \sum_x P(X=x, Y=y_j) \text{ Marginal}$$

$$P(X=x_i | Y=y_j) = \frac{n_{ij}}{r_j} \text{ Conditional}$$

$$P(Y=y_j | X=x_i) = \frac{n_{ij}}{c_i} \text{ Conditional}$$

chain rule

$$\text{Joint} = \text{Conditional} \times \text{Marginal}$$

Baye's rule

$$\text{Conditional} = \frac{\text{Joint}}{\text{Marginal}}$$

Let X and Y each have range $\{1, 2, 3, 4\}$. The following formula gives their joint pmf

$$P(X=i, Y=j) = \frac{i+j}{80}$$

Compute each of the following:

(a) $P(X=Y)$.

(b) $P(XY=6)$.

(c) $P(1 \leq X \leq 2, 2 < Y \leq 4)$.

$$\frac{2}{80} + \frac{4}{80} + \frac{6}{80} + \frac{8}{80} = \frac{20}{80} = \frac{1}{4}$$

$$\frac{5}{80} + \frac{5}{80} = \frac{1}{8}$$

$$= \frac{20}{80} + \frac{1}{8}$$

		X			
		1	2	3	4
Y	1	2	3	4	5
	2	3	4	5	6
	3	4	5	6	7
	4	5	6	7	8

Intro to Bayesian Networks

→ Why we care about joint distribution?

- We can calculate marginal by summation/integration
- We can calculate conditional by using chain/Bayes rule
- We can find conditional independence if it exists.

given $P(Y, X_1, X_2, \dots, X_n)$

$$① P(Y) = \sum_{X_1, X_2, \dots, X_n} P(Y, X_1, X_2, \dots, X_n)$$

$$② P(Y/X_1, X_2, \dots, X_n) = \frac{P(Y, X_1, X_2, \dots, X_n)}{\sum_Y P(Y, X_1, X_2, \dots, X_n)} = \frac{P(Y, X_1, X_2, \dots, X_n)}{P(X_1, X_2, \dots, X_n)}$$

$$③ P(Y, X_1) = P(Y) \cdot P(X_1)$$

To Represent Joint distribution of n -Indicator random variables $\Rightarrow 2^n - 1$ values

challenges

- store large table
- impossible data collection
- Need large data to learn the parameters

X_1	X_2	X_3	X_4	...	X_n	P
0	0	0	0	...	0	0.01
1	0	0	0	...	0	0.03
0	1	0	0	...	0	0.05
1	1	0	0	...	0	0.1
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
1	1	1	1	...	1	0.002

(Once the first $2^n - 1$ values are specified the last value is deterministic as the values need to sum to 1)

Conditional Parameterization — Another approach

d possible values for Y
 k possible values for each X component X_i
 m such components

$$P(Y, \bar{X}) = P(Y) \cdot P(\bar{X}|Y)$$

$$d \cdot k^m - 1 = d - 1 + d(k^m - 1)$$

$$d \cdot k^m - 1 = d \cdot k^m - 1$$

Same # parameters
 More { Natural, intuitive } representation
 • Easy interpretation / learning

Naïve Bayes Assumption

further if we assume conditional independence $P(X_1, X_2, \dots, X_n | Y) = P(X_1 | Y) \cdot \dots \cdot P(X_i | Y) \cdot \dots \cdot P(X_n | Y)$
 then #parameters are drastically reduced.

$$P(Y, \bar{X}) = P(Y) \cdot P(X_1 | Y) \cdot P(X_2 | Y) \cdot \dots \cdot P(X_n | Y)$$

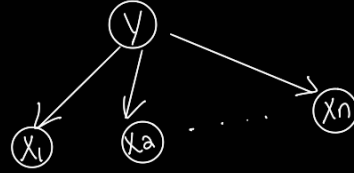
$$\xrightarrow{d-1 + d[k-1]m} d \cdot k^m - d + d - 1$$

Bayes's N/W

- expand on Naive Bayes's model
- Relax strong independence assumption in Naive Bayes's model
- Allow for more complex dependencies btw. variables

Graph → Joint Distribution
Nodes → Random Variable
Edges → Conditional dependence

eg. $Y, X_1, X_2, \dots, X_n \rightarrow$ Random variable
 $P(X_1, X_2, X_3, \dots, X_n) = P(Y) \cdot P(X_1|Y) \cdot P(X_2|Y) \dots P(X_n|Y) \rightarrow$ Naive

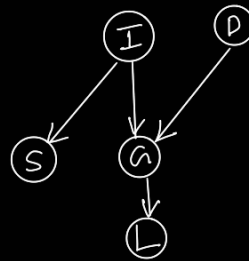


eg.)

Nodes: Represent random variables in the system

- I : Intelligence
- D : Difficulty
- G : Grade
- S : SAT Score
- L : Letter of Recommendation

Edges: Represent dependencies between variables.



Grade (G) depends on:

- Student's Intelligence (I)
- Exam's Difficulty (D)

SAT Score (S) depends on:

- Student's Intelligence (I)

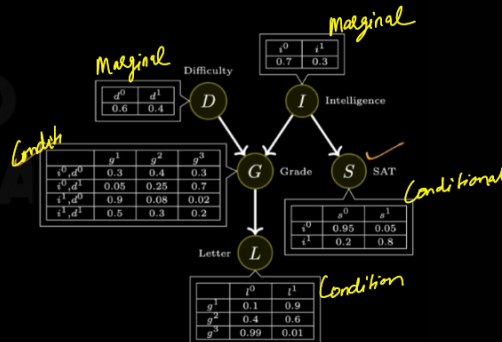
Recommendation Letter (L) depends on:

- Grade (G)

→ Each variable is influenced directly by parent node

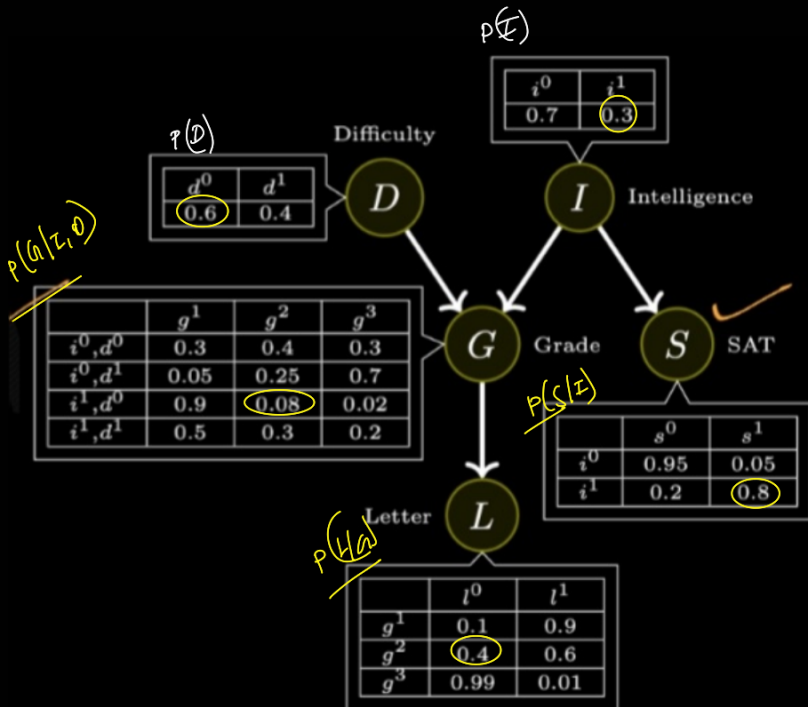
- A Bayesian network can be viewed as a **data structure**.
- It provides a **framework** for representing **joint distributions** compactly.
- Each node in the Bayesian network is associated with a **local probability model**.

- Each variable is associated with a **conditional probability distribution (CPD)** that depends on its parents in the network.



$$P(I, D, G, S, L) = P(I) \cdot P(D|I) \cdot P(G|I, D) \cdot P(S|I, D, G) \cdot P(L|I, D, G, S)$$

$$= P(I) \cdot P(D) \cdot P(G|I, D) \cdot P(S|I) \cdot P(L|G)$$



$$P(I=1, D=0, G=2, S=1, L=0) = P(I) \cdot P(D) \cdot P(G|I,D) \cdot P(S|I) \cdot P(L|G)$$

$$= (0.3) (0.6) \cdot 0.08 \cdot 0.8 \cdot 0.4$$

Reduction in parameters

- $P(I, D, G, S, L)$ without independence = $(2 \times 2 \times 3 \times 2 \times 2) - 1 = 47$
- with independence = $1 + 1 + 8 + 2 + 3 = 15$

The graph along with the tables of Marginal & Conditional distribution is a Bayesian NW.

Earthquakes and fires are independent events that will cause an alarm to go off. Suppose you hear an alarm. How does hearing on the radio that there's an earthquake change your beliefs?

- A) It increases the probability of a fire
- B) It decreases the probability of a fire
- C) It does not change the probability of a fire

Q) #parameters required to fully specify the Joint PDF of 3 boolean RVs. if no independence assumptions are made? = $2^3 - 1 = 7$

Q) Now assume X_1 is conditionally independent of X_3 given X_2 . Now how many parameters required?

$$P(X_1, X_2, X_3) = P(X_2) \cdot P(X_1, X_3 | X_2)$$

$$= P(X_2) \cdot P(X_1 | X_2) \cdot P(X_3 | X_2)$$

Parameters: $P(X_2)$ (1), $P(X_1 | X_2)$ (1+1), $P(X_3 | X_2)$ (1+1) = 5 parameters

Q) Suppose you are solving a binary classification problem, and your dataset has 4 attributes: A_1, A_2, A_3 , and A_4 . Each attribute can take 3 possible values, and the class label C is binary (taking values 0 or 1). If you modeled the full joint probability distribution of the attributes and the class label, how many parameters would you need to specify the distribution?

Without Naive Bayes' assumption

$$d \cdot k^m - 1 = (2 \cdot 3^4) - 1 = 161$$

with Naive Bayes

$$(d-1) + d[k-1]m = 1 + 2[2] \cdot 4 = 17$$

The attributes A_1, A_2, A_3 , and A_4 are conditionally independent given the class label C . Each attribute can take 3 possible values, and C is binary (0 or 1). How many parameters are required to specify the joint probability distribution under this conditional independence assumption?

What is the total number of parameters needed to model the joint probability distribution $P(Y, X_1, X_2, X_3, \dots, X_d)$ given that:

- k is the number of possible values Y can take.
- t is the number of possible values for each X_i , where $i \in \{1, 2, \dots, d\}$?

- A) $k \times t^d$
- B) $k \times t^d - 1$
- C) $k + t^d$
- D) $k + t^d - 1$

$$k \times (t \times t \times \dots \times t)$$

$$\rightarrow \text{Naive Bayes then} \\ (k-1) + k[t-1]d \\ k-1$$

Consider a two-class problem where the class label $y \in \{T, F\}$ and each training example X has 2 binary attributes X_1 and X_2 , both taking values in $\{T, F\}$. How many parameters will you need to estimate if we do not make the Naïve Bayes conditional independence assumption?

- A) 3
- B) 5
- C) 7
- D) 8

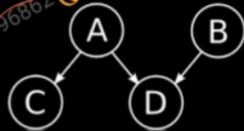
$$\text{Parameters} = 8 - 1 = 7$$

$$dk^{m-1} = (2 \cdot 2^2)^{-1} = 1/4$$

$$\text{with Naive Bayes} = (2-1) + 2[2-1] \cdot 2 = 1 + 4 = 5 \quad \text{option B}$$

Binary classification problem with n Binary attributes
without Naive Bayes assumption $\rightarrow 2^{n+1} - 1$
with Naive Bayes assumption $\rightarrow 2n + 1$

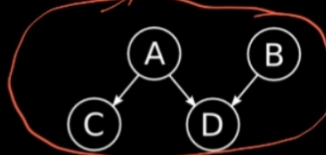
In the following graphical model, A, B, C , and D are binary random variables. How many parameters are needed to define the Conditional Probability Distributions (CPD's) for this Bayes Net?



$$P(A, B, C, D) = P(A) \cdot P(B|A) \cdot P(C|A, B) \cdot P(D|A, B, C) \\ = P(A) \cdot P(B) \cdot P(C|A) \cdot P(D|A, B) \\ \frac{1}{1} + \frac{1}{1} + (2 \cdot 1) + (4 \cdot 1) = 8 \text{ para}$$

Expression for the probability $P(A = 1, B = 1, C = 1, D = 1)$ in terms of the Bayes Net CPD parameters will be :

- A) $P(A = 1)P(B = 1)P(C = 1 | A = 0)P(D = 1 | A = 1, B = 0)$
- B) $P(A = 1)P(B = 1)P(C = 0 | A = 1)P(D = 1 | A = 0, B = 1)$
- C) $P(A = 1)P(B = 1)P(C = 1 | A = 1)P(D = 1 | B = 1)$
- D) $P(A = 1)P(B = 1)P(C = 1 | A = 1)P(D = 1 | A = 1, B = 1)$

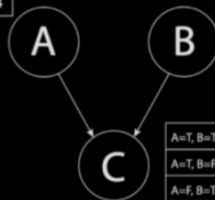


The formula for the joint probability distribution induced by the above Bayes Net:

- A) $P(A, B, C) = P(A)P(B | A)P(C | A)$
- B) $P(A, B, C) = P(A)P(B)P(C | A)$
- C) $P(A, B, C) = P(A)P(B)P(C | A, B)$
- D) $P(A, B, C) = P(A | B)P(B)P(C | B)$

$P(A=T)$	$P(A=F)$
1/4	3/4

$P(B=T)$	$P(B=F)$
3/4	1/4



	$P(C=T A,B)$	$P(C=F A,B)$
$A=T, B=T$	1	0
$A=T, B=F$	0	1
$A=F, B=T$	1/2	1/2
$A=F, B=F$	0	1

$$P(A, B, C) = P(A) \cdot P(B|A) \cdot P(C|A, B) \\ = \frac{P(A) \cdot P(B)}{1+1} + \frac{P(C|A, B)}{4} \\ = 6 \text{ parameter}$$

Answer: C

In above Q: What is the value of i) $P(C=T)$?

- ii) $P(A=T, B=T)$
- iii) $P(A=T, B=T | C=T)$

$$\text{ii) } P(A=T, B=T) = \sum_C P(A=T, B=T, C) = P(A) \cdot P(B) = \frac{3}{16}$$

$$\text{iii) } P(A=T, B=T | C=T) = \frac{P(A=T, B=T) \cdot P(C=T | A=T, B=T)}{P(C=T)} = \frac{3/16 \cdot 1}{15/32} = \frac{6}{15} = \frac{2}{5}$$

$$\text{i) } P(C=A) = \sum_{A,B} [P(A, B, C)] \\ = \sum_{A,B} P(A) \cdot P(B) \cdot P(C|A, B) \\ = \left(\frac{3}{16} \cdot 1 \right) + \left(\frac{1}{16} \cdot 0 \right) + \left(\frac{9}{16} \cdot \frac{1}{2} \right) + \left(\frac{3}{16} \cdot 0 \right) = \frac{3}{16} + \frac{9}{32} \\ = \frac{15}{32}$$

Suppose $A \perp B$. Independence symbol ☆
 Determine the missing entries (x, y) of the joint distribution $P(A, B)$, where A and B take values in $\{0, 1\}$.

$$\begin{aligned} P(A=0, B=0) &= 0.1 \\ P(A=0, B=1) &= 0.3 \\ P(A=1, B=0) &= x \\ P(A=1, B=1) &= y \end{aligned}$$

$$P(A, B) = P(A) \cdot P(B) =$$

$$P(A=0) \cdot P(B=0) = 0.1 - a$$

$$P(A=0) \cdot P(B=1) = 0.3 - b$$

$$P(A=1) \cdot P(B=0) = \frac{6}{10} \times \frac{1}{4} = \frac{3}{20} = 0.15$$

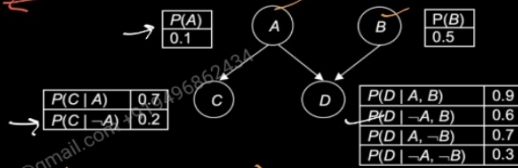
$$P(A=1) \cdot P(B=1) = \frac{6}{10} \times \frac{3}{4} = \frac{9}{20} = 0.45$$

$$\begin{aligned} a+b &\Rightarrow P(A=0) = 0.4 \\ P(A=1) &= 0.6 \\ a &\Rightarrow P(B=0) = \frac{0.1}{0.4} \\ &= 0.25 \\ b &\Rightarrow P(B=1) = 0.75 \end{aligned}$$

Consider the following Bayesian Network containing four Boolean random variables.

► Compute $P(\neg A, B, \neg C, D)$ ←

- ☒ A. 0.216
- ☐ B. 0.054
- ☐ C. 0.024
- ☐ D. 0.006
- ☐ E. None of the above



$$P(A, B, C, D) = P(A) \cdot P(B) \cdot P(C|A) \cdot P(D|A, B)$$

$$\begin{aligned} P(\neg A, B, \neg C, D) &= P(\neg A) \cdot P(B) \cdot P(\neg C|\neg A) \cdot P(D|\neg A, B) \\ &= 0.9 \times 0.5 \times 0.8 \times 0.6 \\ &= \frac{9 \times 5 \times 8 \times 6}{10000} = 0.216 \end{aligned}$$

ii) Compute $P(A|B, C, D) = \frac{P(A, B, C, D)}{\sum_A P(A, B, C, D)}$

$$= \frac{0.1 \times 0.5 \times 0.7 \times 0.9}{(0.1 \times 0.5 \times 0.7 \times 0.9) + (0.9 \times 0.5 \times 0.2 \times 0.6)}$$

$$= \frac{2160}{2160 + 540} = \frac{2160}{2700} = \frac{45 \times 12}{95 \times 540} = \frac{12}{130} = \frac{6}{65}$$

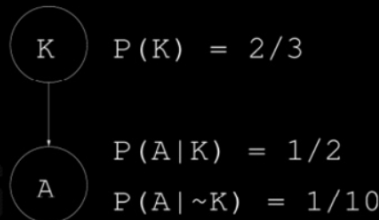
option E

Half of all kangaroos in the zoo are angry, and $2/3$ of the zoo is comprised of kangaroos. Only 1 in 10 of the other animals are angry.

What's the probability that a randomly chosen animal is an angry kangaroo?

$$P(A, K) = P(K) \cdot P(A|K) = \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3}$$

Answer : 1/3



ii) $P(A) = P(A|K) \cdot P(K) + P(A|\neg K) \cdot P(\neg K)$

$$P(A) = \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{10} \cdot \frac{1}{3}$$

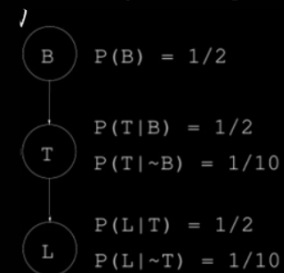
$$= \frac{2}{6} + \frac{1}{30} = \frac{1}{3} + \frac{1}{30} = \frac{11}{30}$$

ii) Probability of chosen Animal is an angry one $\rightarrow P(A)$?

Half of all potatoes are big. A big potato is more likely to be tall.

A tall potato is more likely to be lovable.

What's the probability that a big lovable potato is tall?



$$P(T|B, L) = \frac{P(T, B, L)}{P(B, L)} = \frac{\frac{1}{8} \times \frac{20}{3}}{\frac{20}{24}} = \frac{5}{6}$$

$$\begin{aligned} P(T, B, L) &= P(B) \cdot P(T|B) \cdot P(L|T) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8} \\ P(B, L) &= \sum_T P(B) \cdot P(T|B) \cdot P(L|T) = \frac{1}{2} \left[\frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \times \frac{1}{10} \right] \\ &= \frac{1}{2} \left[\frac{1}{4} + \frac{1}{20} \right] = \frac{3}{20} \end{aligned}$$

Suppose you are given the following set of data with three Boolean input variables a, b , and c , and a single Boolean output variable K .

Assume we are using a naive Bayes classifier to predict the value of K from the values of the other variables. According to the naive Bayes classifier, what is $P(K=1 | a=1 \wedge b=1 \wedge c=0)$?

a	b	c	K
1	0	1	1
1	1	1	1
0	1	1	0
1	1	0	0
1	0	1	0
0	0	0	1
0	0	0	1
0	0	1	0

$$P(K | A, B, C) = \frac{P(K) \cdot P(A|K) \cdot P(B|K) \cdot P(C|K)}{P(K) \cdot P(A|K) \cdot P(B|K) \cdot P(C|K) + P(\bar{K}) \cdot P(A|\bar{K}) \cdot P(B|\bar{K}) \cdot P(C|\bar{K})}$$

$$= \frac{\frac{1}{32}}{\frac{1}{32} + \frac{1}{32}} = \frac{1}{2}$$

$$Q) P(K=0 | A=1, B=1) = \frac{P(\bar{K}) \cdot P(A|\bar{K}) \cdot P(B|\bar{K})}{P(\bar{K}) \cdot P(A|\bar{K}) \cdot P(B|\bar{K}) + P(K) \cdot P(A|K) \cdot P(B|K)} = \frac{1/2 \cdot 1/2 \cdot 1/2}{1/2 \cdot 1/2 \cdot 1/2 + 1/2 \cdot 1/2 \cdot 1/2} = \frac{2}{3}$$

$$\begin{aligned} P(K) &= \frac{4}{8} = 1/2 \\ P(A|K) &= 2/4 = 1/2 \\ P(B|K) &= 1/4 \\ P(C|K) &= 1/2 \\ P(\bar{K}) &= 1/2 \\ P(A|\bar{K}) &= 1/2 \\ P(B|\bar{K}) &= 1/2 \\ P(C|\bar{K}) &= 1/4 \end{aligned}$$

PYQ

GATE 2024 -- Data Science and Artificial Intelligence (DA) -- Sample Question Paper

Q.28	Given a K -class dataset containing N points, where sample points are described using D discrete features with each feature capable of taking V values, how many parameters need to be estimated for Naïve Bayes Classifier?
(A)	$V^D K$
(B)	$K V^D$
(C)	$V^D K + K$
(D)	$K(V + D)$

Q.20	Given a dataset with K binary-valued attributes (where $K > 2$) for a two-class classification task, the number of parameters to be estimated for learning a naïve Bayes classifier is
(A)	$2^K + 1$
(B)	$2K + 1$
(C)	$2^{K+1} + 1$
(D)	$K^2 + 1$

GATE DA 2024

Q.58	Consider two events T and S . Let $\bar{T}$ denote the complement of the event T . The probability associated with different events are given as follows: $P(T) = 0.6, \quad P(S T) = 0.3, \quad P(S \bar{T}) = 0.6$ Then, $P(T S)$ is _____ (rounded off to two decimal places).
------	---

$$P(T|S) = \frac{P(T) \cdot P(S|T)}{P(T) \cdot P(S|T) + P(\bar{T}) \cdot P(S|\bar{T})} = \frac{0.6 \times 0.3}{0.6 \times 0.3 + 0.4 \times 0.6} = \frac{4}{4+10} = \frac{2}{7}$$

Q.24	Consider five random variables U, V, W, X , and Y whose joint distribution satisfies: $P(U, V, W, X, Y) = P(U)P(V)P(W U, V)P(X W)P(Y W)$ Which ONE of the following statements is FALSE?
(A)	Y is conditionally independent of V given W
(B)	X is conditionally independent of U given W
(C)	U and V are conditionally independent given W ✓
(D)	Y and X are conditionally independent given W

- ⇒
- $V \perp\!\!\!\perp Y$
 - $P(X|W) \perp\!\!\!\perp U, V, Y$
 - $P(Y|W) \perp\!\!\!\perp U, V, X$

Sidenote $\Rightarrow A|B \perp\!\!\!\perp C$

$$P(A|B, C) = \frac{P(A, B, C)}{P(B, C)} = \frac{P(A, C|B) \cdot P(B)}{P(C|B) \cdot P(B)} = \frac{P(A|B) \cdot P(C|B) \cdot P(B)}{P(C|B) \cdot P(B)} = P(A|B)$$

WHY THIS METHOD ALWAYS WORKS (given joint density factorization) *next page*

Because any joint probability distribution that factorizes can be represented as a Bayesian network, and the factorization uniquely tells you:

- the nodes
- the parent sets
- the edges

This gives you the complete BN structure and all conditional independences.

Here is a clean, systematic, always-works method to extract the Bayesian network structure from a given factorization like:

$$P(U, V, W, X, Y) = P(U) P(V) P(W|U, V) P(X|W) P(Y|W)$$

✓ SYSTEMATIC METHOD TO FIND THE BAYESIAN NETWORK (BN) FROM A FACTORIZATION

Every factor of the form

$$P(\text{child} \mid \text{parents})$$

tells you exactly **one node and its parents in the Bayesian network**.

Follow this rule:

STEP 1: Identify each conditional factor.

Look at each term:

1. $P(U) \rightarrow U$ has **no parents**
2. $P(V) \rightarrow V$ has **no parents**
3. $P(W \mid U, V) \rightarrow W$ has parents **U and V**
4. $P(X \mid W) \rightarrow X$ has parent **W**
5. $P(Y \mid W) \rightarrow Y$ has parent **W**

STEP 2: Draw the arrows: parent $\rightarrow$ child

From the above:

- $U \rightarrow W$
- $V \rightarrow W$
- $W \rightarrow X$
- $W \rightarrow Y$

U and V also have no parents.

STEP 3: Combine into a directed graph

The Bayesian network is:

markdown

Copy code



This exactly corresponds to the factorization.

STEP 4: Use d-separation to reason about independence

Once the DAG is drawn, you can get all conditional independence statements by using three rules: $\rightarrow$ Directed Acyclic Graph

1. Chains: $A \rightarrow B \rightarrow C$

Condition on $B \Rightarrow A \perp C \mid B$

2. Forks: $A \leftarrow B \rightarrow C$

Condition on $B \Rightarrow A \perp C \mid B$

3. Colliders: $A \rightarrow B \leftarrow C$

- Without conditioning: $A \perp C$
- Conditioning on B (or its descendants): A becomes dependent on C



Q.12	Three fair coins are tossed independently. T is the event that two or more tosses result in heads. S is the event that two or more tosses result in tails. What is the probability of the event $T \cap S$?
(A)	0 ✓
(B)	0.5
(C)	0.25
(D)	1

Q.36	Consider the following joint distribution of random variables X and Y : $f(x, y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 \leq x \leq 2, 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$ Which one or more of the following statements is/are correct?
(A) ✓	X and Y are mutually uncorrelated.
(B) ✓	X and Y are mutually independent. ✓ True.
(C) ✗	The mean of X is 1.
(D) ✗	The mean of Y is 0.5. A2B True

Independence $\Rightarrow$ Joint PDF factors

Independence $\Rightarrow$ Uncorrelated $\Rightarrow A \perp B$ definitely true

$$f(x)_{\text{marginal}} = \int_0^1 x \frac{(1+3y^2)}{4} dy = \frac{x}{4} \left(y + \frac{3y^3}{3} \right) \Big|_0^1 = x/2$$

$$E(X) = \int_0^2 x f(x) dx = \int_0^2 x \cdot \frac{x}{2} dx = \frac{1}{2} \left[\frac{x^3}{3} \right]_0^2 = \frac{8}{6}$$

$$f(y)_{\text{marginal}} = \int_0^2 \frac{x(1+3y^2)}{4} dx = \frac{1+3y^2}{4} \left[\frac{x^2}{2} \right]_0^2 = \frac{1+3y^2}{2}$$

$$E(Y) = \int_0^1 y (1+3y^2) dy = \frac{1}{2} \left[\frac{y^2}{2} + \frac{3y^4}{4} \right]_0^1 = \frac{1}{2} \left[\frac{1}{2} + \frac{3}{4} \right] = \frac{5}{8}$$

Bayesian N/W (cont...)

→ Joint distribution → Way too big to represent
→ Hard to estimate

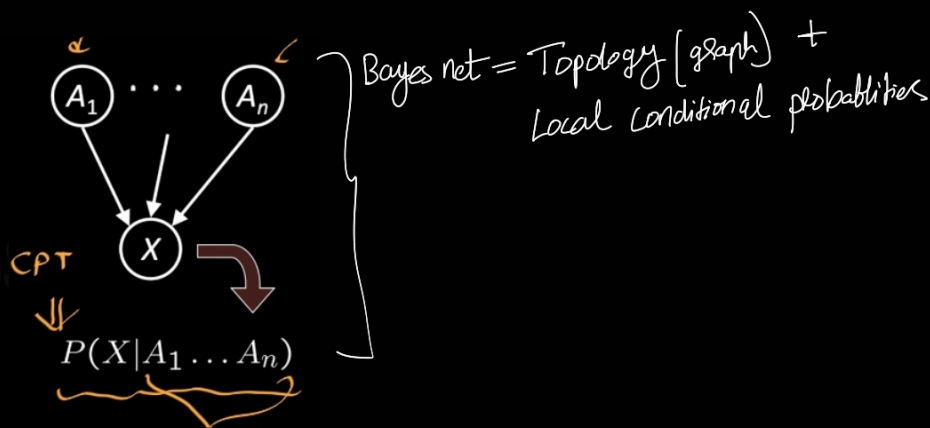
Baye's Net aka Belief Net → Local distributions (conditional probabilities)
→ Local since only parent child conditional probabilities are represented or modelled.

DAG → Directed acyclic graph
→ Nodes (RVs), edges (direct influence),

- A set of nodes, one per variable X
- A directed acyclic graph
- A conditional distribution for each node
 - A collection of distributions over X, one for each combination of parents' values

$$P(X | a_1 \dots a_n)$$

• CPT: conditional probability table



In a Bayesian Network, we must quantify the local relationships between a variable (node) and its parents (nodes). In particular, we specify a conditional probability table (CPT) for each variable (node) in the network, which is a conditional probability distribution for a variable given its parent variables (parent nodes).

→ These local conditional distributions induce a global probability distribution over all network variables (the joint probability distribution).

P(variable | parents of variable)

implies the factorization:

$$P(X) = \prod_{i=1}^{|X|} P(X_i | \text{parents}(X_i))$$

Types of Reasoning

- 1) Causal Reasoning → Predict downstream
- 2) Evidential Reasoning → Reason about cause by looking at effects
- 3) Explaining Away → Reasoning about different causes interaction by looking at effect.

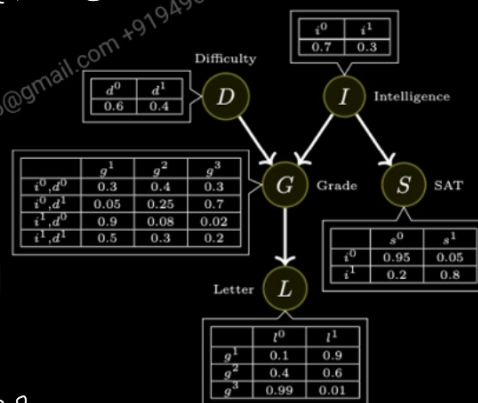
Causal Reasoning

eg a) Probability that student will get good recommendation letter?
b) Probability that student will get good recommendation letter given not intelligent?

Evidential Reasoning

eg c) What is the probability of student being intelligent?
a) What is the probability of student being intelligent given he scored grade = 3
e) What is the probability of exam Difficulty hard if student got poor recommendation letter.

$$P(D, I, G, S, L) = P(D) \cdot P(I) \cdot P(G | D, I) \cdot P(S | I) \cdot P(L | G)$$



$$a) P(L) = \sum_{I,D,G,S} P(I,D,G,S,L=1) = \sum_{I=0,1} P(I) \cdot \sum_{D=0,1} P(D) \cdot \sum_{S=0,1} P(S|I) \cdot \sum_{G=1,2,3} P(G|D,I) \cdot P(L=1|G)$$

$$b) P(L|I=0) = \frac{P(L=1, I=0)}{P(I=0)} = \frac{P(I=0) \cdot \sum_{D=0,1} P(D) \cdot \sum_{S=0,1} P(S|I) \cdot \sum_{G=1,2,3} P(G|D,I) \cdot P(L=1|G)}{P(I=0) \cdot \sum_{D=0,1} P(D) \cdot \sum_{S=0,1} P(S|I) \cdot \sum_{G=1,2,3} P(G|D,I) \cdot [P(L=1|G) + P(L=0|G)]}$$

$$c) P(I=1) = \text{from data}$$

$$d) P(I=1|G=3) = \frac{P(I=1, G=3)}{P(G=3)}$$

$$e) P(D=1|L=0) = \frac{P(D=1, L=0)}{P(L=0)}$$

Side Quest
What if we
additionally?

got to know about poor recommendation letter
ie $P(I=1|G=3, L=0)$?

$$= \frac{P(I=1, G=3, L=0)}{P(G=3, L=0)} = \frac{P(I=0, L=0|G=3) \cdot P(G=3)}{P(L=0|G=3) \cdot P(G=3)}$$

$$= \frac{P(I=0|G=3) \cdot P(L=0|G=3) \cdot P(G=3)}{P(L=0|G=3) \cdot P(G=3)}$$

Since
 $P(L|G=3)$ is
independent of
 I

Explaining Away

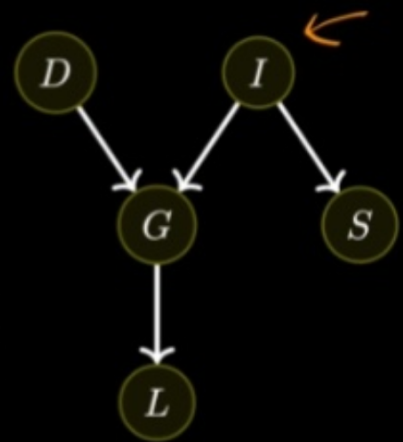
→ Probability of student being intelligent given grades & difficult course $P(I=1|G=3, D=1)$
→ Probability of difficult course given grades and good sat score $P(D=1|G=3, S=1)$

Independence Rules

→ Direct influence → X & Y are correlated
no matter any evidence
about any other variable

→ A node is not independent of its parents ever.
No matter the evidence (conditioning)

eg. $G \perp L | I?$ → NO
 $G \perp L | D?$ → NO
 $G \perp L | S?$ → NO



Chains

If Pascal woke up too early E, Pascal probably needs coffee C; if Pascal needs coffee, he's likely grumpy G. If he is grumpy then it's possible that the lecture won't go smoothly L. If the lecture does not go smoothly then the students will likely be sad S

$$E \rightarrow C \rightarrow G \rightarrow L \rightarrow S$$

• S is dependent on E, C, G, or L

• $S \perp E, C, G | L$ → knowing the value of L
other variables don't affect probability of S.

$$\Rightarrow P(S|L, G, C, E) = P(S|L)$$

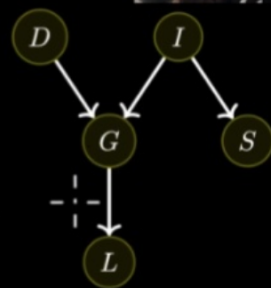
$$P(L|G, C, E) = P(L|G)$$

$$P(G|C, E) = P(G|C)$$

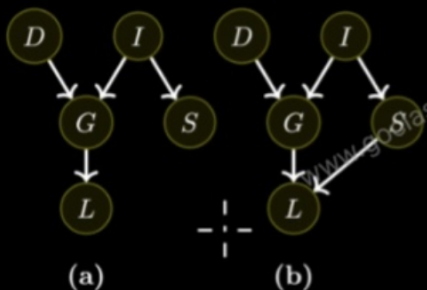
→ Influence on chains are mediated by intermediate variables knowing which the upstream factors becomes independent of downstream effects.

→ Forks → $P(L \perp S | G)$ — Yes

→ given the parent node as evidence/condition the child is independent of all other non parent nodes in earlier forks of the DAG. (A fork is a branch that splitted away. Earlier fork refers to splits branches in higher levels)



If we believe Graph (a) is how the world works then :



$$(L \perp S) | G$$

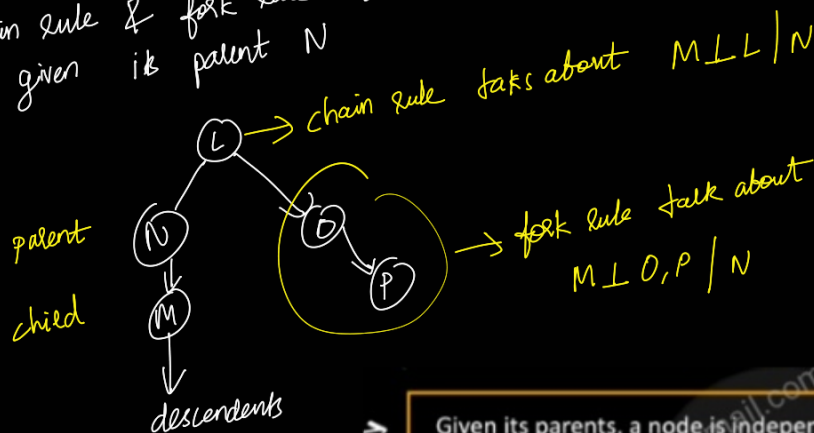
If we believe Graph(b) is how the world works

$$(L \not\perp S) | G$$

Chain & Fork Do not kill direct Influence → Not a new rule, just a classification

is $G \perp L | D, I$? → NO, this is a violation of direct influence
→ if X & Y are in parent child relation, then they are dependent on each other no matter evidence.

Chain rule & fork rule doesn't say descendants are independent for a node M given its parent N



Given its parents, a node is independent of all variables except its descendants

direct influence

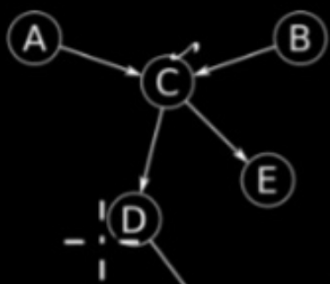
A node is not independent of its parents

A node is not independent of its parents even when we are given the values of other variables.

Given its parents, a node is independent of all variables except its descendants

fork & chain →

Collider rule → Parent Nodes may be unconditionally independent but is not conditionally independent given common descendant / child.



Q) Make all independence inferences?

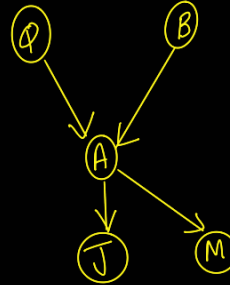
- 1) $F \perp A, C, E, B / D$
- 2) $D \perp A, B, E / C$
- 3) $E \perp A, B / C$
- 4) $A \perp B$
- 5) $A \not\perp B / C \rightarrow$ Collider Rule $\rightarrow$

Burglars and Earthquakes

Pearl lives in Los Angeles. It is a high-crime area.
 Pearl installed a burglar alarm.
 He asked his neighbors John & Mary to call him if they hear the alarm.
 This way he can come home if there is a burglary.
 Los Angeles is also earth-quake prone.
Alarm goes off when there is an earth-quake.

Variables: Burglary, Earthquake, Alarm, JohnCalls, MaryCalls

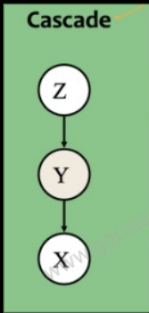
$B \quad Q \quad A \quad J \quad M$



$J \perp Q / A$
 $J \perp B / A$

$M \perp Q / A$
 $M \perp B / A$

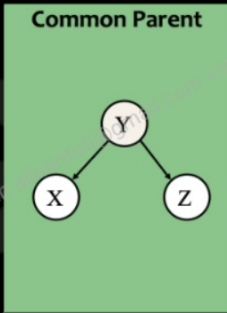
Once you know that the Alarm has gone off, John Calls depends solely on whether the alarm was triggered, not on what caused it (burglary or earthquake). Since John Calls only responds to the alarm, it becomes independent of the Burglary once the status of the Alarm is known.



$X \not\perp Z$

$X \perp Z / Y$

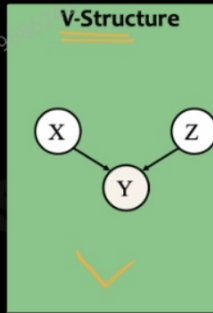
chain



$X \not\perp Z$

$X \perp Z / Y$

Fork



$X \perp Z$

$X \not\perp Z / Y$

Collider

aka

knowing Y couples X & Z

chain & Fork

$$P(X, Z / Y) = P(X / Y) \cdot P(Z / Y)$$

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

Collider (Explaining away)

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

$$P(X, Z / Y) \neq P(X / Y) \cdot P(Z / Y)$$

Explaining away

knowing Y decouples X & Z

Markov Blanket

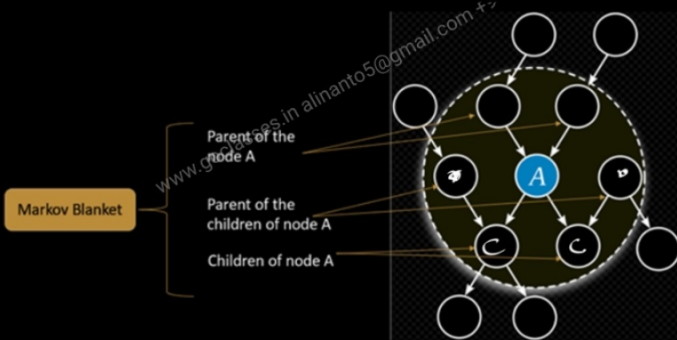
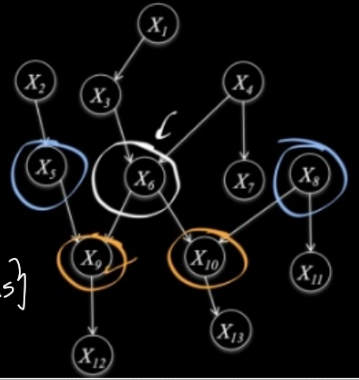
Co-parents: of a node x are all parent nodes of the children of x

Markov Blanket: of a node is the set containing its parents, children & co-parents.



Example: The Markov Blanket of X_6 is $\{X_3, X_4, X_5, X_7, X_8, X_9, X_{10}, X_{11}\}$

eg.
 → Coparents of $X_6 = X_7 \& X_8$
 → Markov Blanket of $X_6 = \{X_3, X_4, X_5, X_7, X_8, X_9, X_{10}, X_{11}\}$

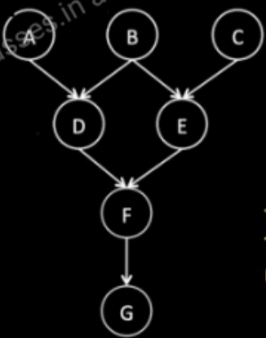


Blanking Theorem

A node is conditionally independent of every other node in the graph given its Markov Blanket as evidence.

Intuition

- chain law makes all upstream factors independent given parents
- Fork law makes all upstream forks and cousins independent given parents.
- chain law & fork law makes any upstream nodes of co-parents independent
- chain law also makes all downstream nodes independent



Consider the graph drawn below. Assume that each variable can only take on values true and false.

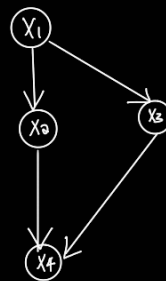
The minimum number of Bayes net parameters required to fully specify the distribution $P(G|A, B, C, D, E, F)$?

$$P(G|A, B, C, D, E, F) = P(G|F) = \text{only 2 parameters}$$

In a Bayesian network, consider the following structure:

- $X_1 \rightarrow X_2$
- $X_1 \rightarrow X_3$
- $X_2 \rightarrow X_4$
- $X_3 \rightarrow X_4$

Is node X_1 independent of X_4 given its Markov blanket? Yes



$$X_1 \perp\!\!\!\perp X_4 / X_2, X_3$$

D-separation (Direct Separated)

Algorithm to calculate whether $X \perp\!\!\!\perp Z \mid Y$ by looking at graph separation.
 Looks to see if there is active path between X & Z when variable Y is evidence.

Algorithm

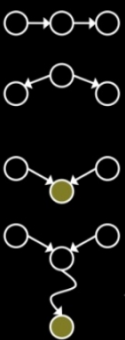
- Active path:
- Chain from one node to another where belief about A propagates to B since no intermediary node is observed as evidence.
 - Fork: Common cause node $A \leftarrow B \rightarrow C$ where B is unobserved $\Rightarrow$ Belief about A propagates to C
 - Collider: Common effect node $A \rightarrow B \leftarrow C$ where B or one of its descendants is evidence. Belief about A can be explained away to belief about B given cause.

Chain active path (when belief propagates)

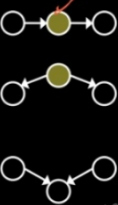
- A path is active if each triple is active
- Causal chain $A \rightarrow B \rightarrow C$ where B is unobserved $\rightarrow$ Both direction active path.
- if a intermediate node in the path is observed or evidence, then belief doesn't propagate

Evidence stops belief propagation in both direction

Active Triples



Inactive Triples



- Causal chain $A \rightarrow B \rightarrow C$ where B is unobserved (either direction)
- Common cause $A \leftarrow B \rightarrow C$ where B is unobserved
- Common effect (aka v-structure) $A \rightarrow B \leftarrow C$ where B or one of its descendants is observed

NOTE: All it takes to block a path is a single inactive segment



(x)

(y)

(z)

- If there is "some" active path between X & Z given Y , then we can say that the information can flow from X to $Z \Rightarrow X$ & Z are dependent
- If there is "No" active path between X & Z given Y , then we can say that there no way that the information can flow from X to $Z \Rightarrow X$ & Z are independent $\Rightarrow X$ & Z are d-separated

D-separation ("direct separated") in Bayesian networks

Algorithm to calculate whether $X \perp\!\!\!\perp Z \mid Y$ by looking at graph separation
 Look to see if there is active path between X and Z when variables Y are observed:

If no active path between X and Y , then X and Z are d-separated with respect to Y

If X and Y are d-separated by Z , then X and Y are conditionally independent given Z

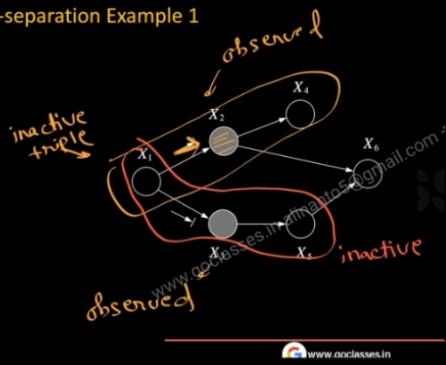
A is d-separated from B given C if every undirected path between them is blocked.



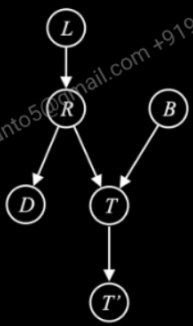
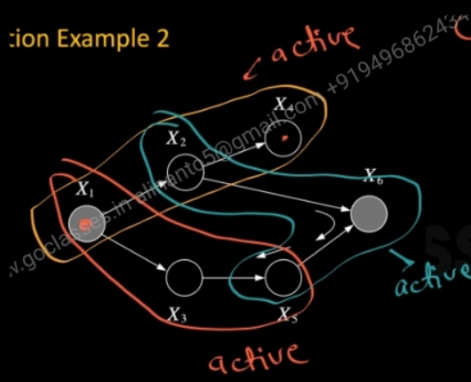
Algorithm

- Identify all inactive triple (given Y)
- Identify all paths from X to Y
- if $\exists$ path from X to Y without inactive triple $\Rightarrow X \not\perp\!\!\!\perp Y \mid Z$
- if $\nexists$ active path from X to $Y \Rightarrow X \perp\!\!\!\perp Y \mid Z$
 X & Y are D-separated by Z

D-separation Example 1



Example 2



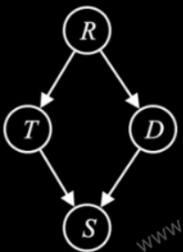
$L \perp\!\!\!\perp T' | T$ True — chain inactive
 $L \perp\!\!\!\perp B$ True — collider inactive
 $L \perp\!\!\!\perp B | T$ False — collider active
 $L \perp\!\!\!\perp B | T'$ False — collider active
 $L \perp\!\!\!\perp B | T, R$ True — collider active, chain inactive

D-separation theorem $\xrightarrow{\text{Analogy}}$ Bayes Ball Boume theorem (BBB)

	Active	Inactive
Chain :	Ball passes	Ball bounces
Fork :	Ball passes	Ball bounces
Collider :	Ball passes	Ball bounces

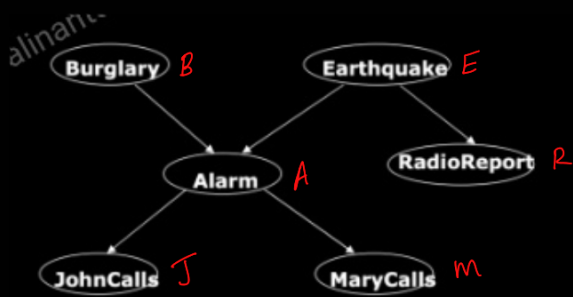
} Ball bounces from evidence/observation

} Ball passes only if it collides with observation



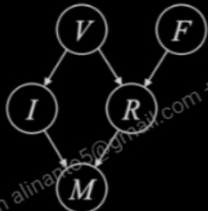
$T \perp\!\!\!\perp D$ False — fork active, collider inactive
 $T \perp\!\!\!\perp D | R$ True — fork & collider inactive
 $T \perp\!\!\!\perp D | R, S$ False — fork inactive, collider active

Practice Questions



- 1) $E \perp\!\!\!\perp B | M$ — False
- 2) $B \perp\!\!\!\perp M$ — False
- 3) $B \perp\!\!\!\perp R | E$ — True } ☆
- 4) $B \perp\!\!\!\perp R | M$ — False

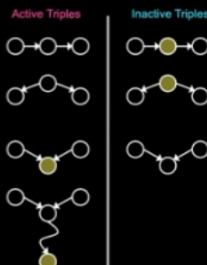
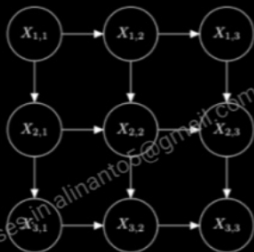
11.3.



- 1) $V \perp\!\!\!\perp M | R$ → False
- 2) $M \perp\!\!\!\perp F | R$ → False —☆☆☆
- 3) $V \perp\!\!\!\perp F$ → True —☆☆☆☆ — 2 paths
- 4) $V \perp\!\!\!\perp F | M$ → False
- 5) $V \perp\!\!\!\perp F | I$ → True

6) $V \perp\!\!\!\perp M | I, R$ True ● guaranteed to be true
○ not guaranteed to be true

Question 11: For this question consider the Bayes' Net below with 9 variables.



Which random variables are independent of $X_{3,1}$? (Leave blank if the answer is none.)

- ☐ $X_{1,1}$ ☐ $X_{1,2}$ ☐ $X_{1,3}$ ☐ $X_{2,1}$ ☐ $X_{2,2}$ ☐ $X_{2,3}$ ☐ $X_{3,2}$ ☐ $X_{3,3}$

There is at least one active path between $X_{3,1}$ and every other node.

- 13) Which RV are independent of $X_{3,1}$ given $X_{1,1} \rightarrow X_{1,2}, X_{1,3}$ —☆☆ → Quick Answer ⇒ Mistake
- 14) " " " " " " " " $X_{1,1} \neq X_{3,3} \rightarrow$ No nodes

15) Consider a Bayes Net with the following graph:



$P(A | C, E) = P(A | C)$ True

- 16) $P(A, E | G) = P(A | G) \cdot P(E | G)$ False
- 17) $P(A | B, G) = P(A)$ True
- 18) $P(A | B, G) = P(A | G)$ False
- 19) $P(E, G | D) = P(E | D) \cdot P(G | D)$ False
- 20) $P(A, B | F) = P(A | F) \cdot P(B | F)$ True

15) Statement is true if $A \perp\!\!\!\perp E | C$

- 1) True since ① $A \rightarrow C \rightarrow E$ → Inactive
- ② $D \rightarrow G \leftarrow F$ → Inactive

16) True if $A \perp\!\!\!\perp E | G$ — ① $A \rightarrow C \rightarrow E$ — active
 ② $A \rightarrow D \rightarrow G \leftarrow F \leftarrow E$ (active)

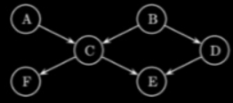
17) True if $A \perp\!\!\!\perp B$ — ① $A \rightarrow D \leftarrow B$ → Inactive
 ② $A \rightarrow C \rightarrow E \rightarrow F \rightarrow G \leftarrow D \leftarrow B$ Inactive

18) True if $A \perp\!\!\!\perp B | G$ — ① $A \rightarrow D \leftarrow B$ → active
 ② $A \rightarrow C \rightarrow E \rightarrow F \rightarrow G \leftarrow D \leftarrow B$ active

19) True if $E \perp\!\!\!\perp G | D$ — ① $E \rightarrow F \rightarrow G$ — active
 ② $E \leftarrow C \leftarrow A \rightarrow D \rightarrow G$ — Inactive

20) True if $A \perp\!\!\!\perp B | F$ — ① $A \rightarrow D \rightarrow G \leftarrow F$ — Inactive $A \rightarrow D \leftarrow B$
 ② $A \rightarrow C \rightarrow E \rightarrow F \rightarrow G \leftarrow D \leftarrow B$ — Inactive

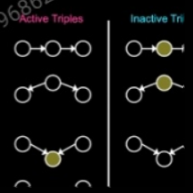
21) Consider the following Bayesian network. State whether the given conditional independences are implied by the net structure.



T/F $I \langle A, \{ \}, B \rangle$

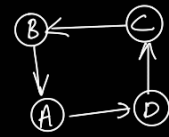
Side Note

$I \langle X, Z, Y \rangle$
 $\Rightarrow X \perp\!\!\!\perp Y | Z$



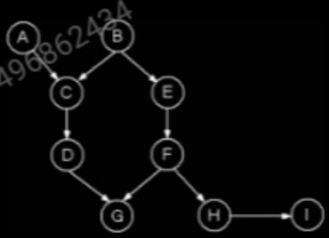
- 21) True
 22) $I \langle A, \{E\}, D \rangle = A \perp\!\!\!\perp D / E$ — False
 23) $I \langle A, \{F\}, D \rangle = A \perp\!\!\!\perp D / F$ — False — ☆

24) The following product of factors corresponds to a valid Bayesian Network over the variables A, B, C and D : $P(A|B)P(B|C)P(C|D)P(D|A)$. — False



cyclic $\Rightarrow$ DAG $\Rightarrow$ Not a valid Bayesian N/w

25) The figure below shows a Bayesian network with 9 variables, all of which are binary.



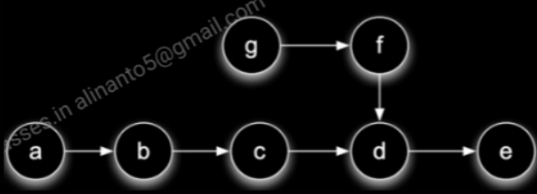
1. (3pts) Which of the following statements are always true for this Bayes net?

- (a) $P(A, B|G) = P(A|G)P(B|G)$;
- (b) $P(A, I) = P(A)P(I)$;
- (c) $P(B, H|E, G) = P(B|E, G)P(H|E, G)$;
- (d) $P(C|B, F) = P(C|F)$.

B

- a) $\Rightarrow A \perp\!\!\!\perp B / G$ — F
- b) $\Rightarrow A \perp\!\!\!\perp I$ — T
- c) $\Rightarrow B \perp\!\!\!\perp H | E, G$ — F
- d) $\Rightarrow C \perp\!\!\!\perp B / F$ — F

26) Let $X = \{c\}, Y = \{b, d\}, Z = \{a, e, f, g\}$. Is $X \perp\!\!\!\perp Z | Y$? If yes, explain why. If no, show a path from X to Z that is not blocked. (2 points)



$C \perp\!\!\!\perp A, E / B, D$

$C \not\perp\!\!\!\perp F, H / B, D$

There is a active path from $C \in X$ to $F \in Z$ given Y

$\Rightarrow X \not\perp\!\!\!\perp Z | Y$

27) Let $X = \{c\}, Y = \{b, d\}, Z = \{a, e, f, g\}$. Suppose you are allowed to choose a set W such that $W \subset Z$. Then define $Z^* = Z/W$ and $Y^* = Y \cup W$. What is the smallest set W such that $X \perp\!\!\!\perp Z^* | Y^*$ is true? (2 points)



- 27) $W = \{f\}$
 $Z^* = \{a, e, g\}$
 $Y^* = \{b, d, f\}$ — Markov blanket of $X = \{c\}$

Set difference

28) Suppose that a patient can have a symptom, S , that can be caused by two different diseases (A and B). It is known that the variation of gene G plays a big role in the manifestation of disease A . The Bayes' Net and corresponding conditional probability tables for this situation are shown below. For each part, you may leave your answer as an arithmetic expression.

$P(G)$	
+g	0.1
-g	0.9

$G \perp\!\!\!\perp A$
 $A \perp\!\!\!\perp B$

$P(B)$	
+b	0.4
-b	0.6

$P(S A, B)$			
+a	+b	+s	1.0
+a	+b	-s	0.0
+a	-b	+s	0.9
+a	-b	-s	0.1
-a	+b	+s	0.8
-a	+b	-s	0.2
-a	-b	+s	0.1
-a	-b	-s	0.9

(a) [2 pts] Compute the following entry from the joint distribution:

$P(+g, +a, +b, +s) = \underline{P(+g)P(+a|+g)P(+b|+g)P(+s|+b, +a)} = (0.1)(1.0)(0.4)(1.0) = 0.04$

$P(A, B, S, G) = P(G) \cdot P(A|G) \cdot P(B) \cdot P(S|A, B)$

$P(A=1, B=1)_{S=1, G=1} = 0.1 \times 1.0 \times 0.4 \times 1.0 = 0.04$

29) $P(A=1|B=1) = ? = \frac{P(A=1, B=1)}{P(B=1)} = \frac{P(A) \cdot P(B)}{P(B)}$ — AIB
 $\stackrel{\text{Alternatively}}{\approx} \frac{\sum_{G \sim A} P(G) \cdot P(A|G) \cdot P(B) \cdot P(S|A, B)}{P(B)}$
 $= \sum_{G \sim A} P(G) \cdot P(A|G) \cdot \sum_{S \sim B} P(S|A, B) = 1$
 $= P(A)$

$\approx P(A|G) \cdot P(G) = 0.1 \times 1.0 + 0.9 \times 0.1 = 0.1 + 0.09 = 0.19$

How to show X is d-separated from Y given Z (where X, Y, Z are sets of nodes)

Short answer (definition):

X is d-separated from Y by Z (written $X \perp\!\!\!\perp Y \mid Z$ in the graph) iff **every** undirected path between any $x \in X$ and any $y \in Y$ is *blocked* by Z according to the d-separation rules below.

What it means for a single path to be *blocked* by Z

Take any (simple) path p between two nodes. For every intermediate node v on that path classify its local configuration on the path:

- **Collider** at v : the path passes through v like $\rightarrow v \leftarrow$.
 - This *blocks* the path **unless** $v \in Z$ or **some descendant of v is in Z** .
 - (So a collider is *activated* only if it or one of its descendants is conditioned on.)
- **Non-collider** at v : the path passes like $\rightarrow v \rightarrow$ or $\leftarrow v \rightarrow$ (chain or fork).
 - This *blocks* the path **if and only if** $v \in Z$.
 - (If the non-collider is **not** in Z the path remains *open* at v .)

A path is **blocked** by Z if at least one node on the path blocks it by the rules above. If **every** path between every $x \in X$ and $y \in Y$ is blocked, then X and Y are d-separated by Z .

Practical ways to show d-separation

1. By inspection (small graphs):

- Enumerate all simple paths from each $x \in X$ to each $y \in Y$.
- For each path check every intermediate node whether it's a collider or non-collider and apply the blocking rules above.
- If every path is blocked, conclude $X \perp\!\!\!\perp Y \mid Z$.

2. Bayes-ball algorithm (efficient):

- Run the Bayes-ball/graph reachability algorithm beginning from all nodes in X with Z marked as observed.
- If no ball can reach any node in Y , X is d-separated from Y given Z . (If some $y \in Y$ is reached, they are d-connected.)

3. Ancestors + moralization (graphical test):

- Take subgraph induced by ancestors of $X \cup Y \cup Z$.
- Moralize (connect parents of common child, drop directions) to get an undirected graph.
- Check whether Z separates X and Y in that undirected graph (graph-theoretic separation). If yes $\rightarrow$ d-separated.

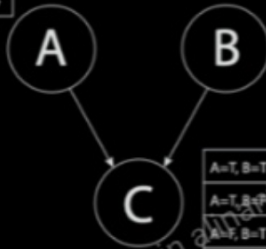
A few important notes / edge cases

- If $X \cap Y \neq \emptyset$ then trivially they are **not** independent (they share variables).
- If a variable in X or Y is in Z , that affects activation: conditioning on a variable in X or Y is allowed but changes reachability—use the rules above.
- d-separation is a purely graph-structural property; if the graph says X and Y are d-separated by Z , then in any distribution that factorizes according to the network they are conditionally independent: $P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$.

10) $P(A=T, B=T) = ?$

$P(A=T)$	$P(A=F)$
$1/4$	$3/4$

$P(B=T)$	$P(B=F)$
$3/4$	$1/4$



	P(C=T A,B)	P(C=F A,B)
A=T, B=T	1	0
A=T, B=F	0	1
A=F, B=T	1/2	1/2
A=F, B=F	0	1

$$A \perp B \Rightarrow P(A=T, B=T) = P(A) P(B) \\ = 1/4 \cdot 3/4 = 3/16$$

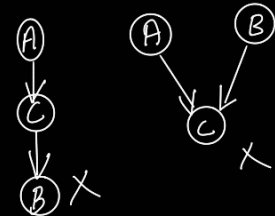
Alternatively,

$$P(A,B,C) = P(A) \cdot P(B) \cdot P(C|A,B)$$
$$P(A,B) = \sum_C P(A) \cdot P(B) \cdot P(C|A,B)$$
$$= P(A) \cdot P(B) \cdot \underbrace{\sum_C P(C|A,B)}_{=1}$$

For the following set of independence relations, draw a Bayesian network graph that encodes the relations, or show that no such graph exists.

- A is always independent of B
- A is independent of C given B
- A is not independent of C

not possible



Consider five random variables A, B, C, D , and E whose joint distribution satisfies the following factorization:

$$P(A, B, C, D, E) = P(A)P(B | A)P(C | A, B)P(D | B, C)P(E | C, D)$$

Which ONE of the following statements is FALSE?

- Which ONE of the following statements is FALSE?
- A. E is conditionally independent of A given C and D $E \perp\!\!\!\perp A \mid C, D$ — True
 - ✓ B. C is conditionally independent of E given D $C \perp\!\!\!\perp E \mid D$ — False
 - ✓ C. B and E are conditionally independent given A and C $B \perp\!\!\!\perp E \mid A, C$ — False
 - ✓ D. D and A are conditionally independent given B $D \perp\!\!\!\perp A \mid B$ False

