Unit - II Probabilistic Reasoning

Acting under uncertainity - Bayesian Inference - Naive bayes models. Probabilistic Reasoning - Bayesian Networks - Exact inference in BN - approximate inference in BN - Causal Networks.

1 Acting under Uncertainity:-

$A \to B$ means if A is true then B is true, if we are not sure about whether A is true or not then we cannot express this statement, this situation is called uncertainity.

→ To represent uncertain knowledge, uncertain reasoning or Probabilistic reasoning is used.

Causes of uncertainity:-

Following are some leading causes of uncertainity to occur:

- → Information occured from unreliable Sources.
- → Experimental Errors.
- → Equipment fault.
- → Temparature Variation.
- → Climate change.

General theory of decision Theory:-

Decision theory = Probability theory + Utility theory.

Unit - II Probabilistic Reasoning

Acting under uncertainity - Bayesian Inference - Naive bayes models. Probabilistic Reasoning - Bayesian Networks - Exact inference in BN - approximate inference in BN - Causal Networks.

① Acting under Uncertainity:-

$A \rightarrow B$ means if A is true then B is true, if we are not sure about whether A is true or not then we cannot express this statement, this situation is called uncertainity.

→ To represent uncertain knowledge, uncertain reasoning or Probabilistic reasoning is used.

Causes of uncertainity:-

Following are some leading causes of uncertainity to occur:

- → Information occured from unrelIable sources.
- → Experimental Errors.
- → Equipment fault.
- → Temparature Variation.
- → Climate change.

General theory of decision Theory:-

Decision theory = Probability theory + Utility theory.

Bayes' Rule :-

The product rule can actually be written in

$P(a \cap b) = P(a|b)\ P(b)$ &

$P(a \cap b) = P(b|a)\ P(a)$

Equating the two right - hand sides & dividing,

$$\boxed{P(b|a) = P(a|b)\ P(b)\ /\ P(a)}$$

this equation is known as Bayes rule or law.

Baye's rule becomes.

$P(\text{cause} \mid \text{effect}) = P(\text{effect} \mid \text{cause})\ P(\text{cause})\ /\ P(\text{effect})$

2) Bayesian Inference :-

→ Bayesian inference is a method of statistical inference in which Baye's theorem is used to update the probability for a hypothesis.

→ Bayesian Inference derives the posterior probability as a consequence of two probability & a "likelihood function" derived from a statistical method.

According to Baye's theorem,

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}$$

where,

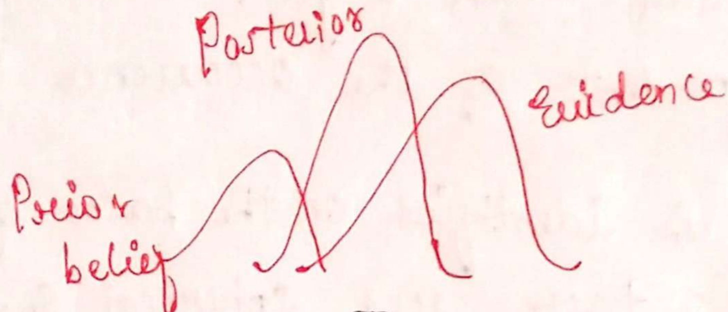
→ H stands for any hypothesis

→ P(H) is prior probability, estimate probability of each hypothesis.

→ E is the evidence.

→ P(H|E) is the posterior probability.

→ P(E|H) is called likelihood.



Types of Bayesian Inference :-

1) Exact inference in BN - Gaussing exact.

Example:-
→ Junction Tree algorithm (JTA)
- Belief Propagation method.
- Sum of Product method.

2) Approximate inference in BN - Guessing approximate.
- Stochastic &
- Deterministic.

Example
1) Markov chain Monte Carlo Algorithm.

Applications :-

Including Science, Engineering, Philosophy, medicine, sport & law.

3) Naive Baye's models:-

→ It is used supervised learning algorithm.

→ used for solving classification problems (Text)

→ It is a probabilistic classifier, which means it predicts on the basis of the probability of a object.

Why it is called Naive Bayes?

Naive : Independent of the occurrence of other features.

Fruit is identified on the bases of colour, shape & taste, red, spherical & sweet fruit.

Bayes: Depends on the principle of Baye's theorem.

Baye's theorem:-

→ known as baye's rule or Baye's Law.

→ Determine the probability of a hypothesis with prior knowledge.

→ Depends on the conditional probability.

formula:-

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

where,

$P(A|B)$ → Posterior Probability (target)

Hypothesis A on the observed event B.

$P(B)$ → Marginal Probability.

Probability of evidence.

(B|A) is likelihood Probability. ③

Evidence given that the probability of a hypothesis is true.

P(A) is prior probability.

Hypothesis before observing the evidence.

Baye's theorem can be rewritten as:

$$\underset{\text{Posterior Probability}}{P(Y/X)} = \frac{\overset{\text{likelihood}}{P(X/Y)}\ \overset{\text{class prior Probability}}{P(Y)}}{\underset{\text{predictor prior probability}}{P(X)}}$$

$$\text{posterior} = \frac{\text{Prior} \times \text{likelihood}}{\text{evidence}}$$

$$\underset{\text{posterior}}{P(H/E)} = \frac{\overset{\text{likelihood}}{P(E/H)} * \overset{\text{prior}}{P(H)}}{\underset{\text{prior probability evidence}}{P(E)}}$$

Steps in Naive Bayes model:-

S_1: Calculate the prior probability for given class labels.

By converting the given dataset into frequency labels.

S_2: find likelihood.

S_3: Use Baye's theorem to calculate posterior.

S_4: Higher Probability.

Example:-

If the weather is sunny, then the player shou[ld] play or [not]

Applying ... P(Yes) ...

Solution:-

First consider the below dataset.

	Outlook	Play
0	Rainy	Yes
1.	Sunny	Yes
2.	Overcast	Yes
3.	overcast	Yes
4.	Sunny	Yes
5.	Overcast	Yes
6.	Rainy	Yes

	Outlook	Play
7.	Sunny	Yes
8	Rainy	No
9	Sunny	No
10.	Sunny	Yes
11	Rainy	No
12.	Overcast	Yes
13.	Overcast	Yes.

frequency table for weather conditions.

Weather	Yes	No
Over cast	5	0
Rainy	2	2
Sunny	3	2
Total	10	4

likelihood table weather conditions:

Weather	No	Yes	
Overcast	0	5	5/14 = 0.35
Rainy	2	2	4/14 = 0.29
Sunny	2	3	5/14 = 0.35
All	4/14 = 0.29	10/14 = 0.71	

Applying Baye's theorem

$P(Yes| Sunny) = P(Sunny|Yes) * P(Yes) / P(Sunny)$

$P(Sunny|Yes) = 3/10 = 0.3$

$P(Sunny) = 0.35$

$P(Yes) = 0.71$

So $\boxed{P(Yes|Sunny) = 0.3 * 0.71 / 0.35 = 0.60}$

$P(No|Sunny) = P(Sunny|No) * P(No) / P(Sunny)$

$P(No|Sunny) = 2/4 = 0.5$

$P(No) = 0.29$

$P(Sunny) = 0.35$

So $\boxed{P(No|Sunny) = 0.5 * 0.29 / 0.35 = 0.41}$

Hence on a sunny day, player can play the game.

Advantages of Naive Bayes classifier :-

→ Fast & Easy ML algorithm.

→ Used for binary as well as multi-class classification.

→ Performs in multi-class prediction.

→ Text classification problems.

Disadvantages :-

All feature are independent or unrelated.

Applications:

-> used for Credit scoring

-> used in medical data classification

-> used in real - time prediction.

-> used in text classification (Spam filtering.
Sentiment analy

Types of Naive bayes model:-

3 types:

* Gaussian * multinomial * Bernouli.

Gaussian:

-> features follow a normal distribution

-> predictor take continuous Values instead of discreate.

multinomial:-

-> used when data is multinomial distributed

-> used for document classification problems.

-> Category such as sports, politics, education etc...

Bernouli:-

-> Similar to the multinomial classifier.

-> But the predictor Variables are independent boolean Variables.

-> If a particular word is present or not in a document.

Probabilistic Reasoning:-

* Probability can be defined as a chance that an uncertain event will occur. It is the numerical measure of the likelihood that an event will occur.
* Value of probability always remains between 0 & 1.

$0 \leq P(A) \leq 1$, where $P(A)$ is the probability of an event A.

$P(A) = 0$, indicates total uncertainty in an event A

$P(A) = 1$, indicates total certainty in an event A.

Axioms of probability:-

→ Given a set U (universe)

A probability function is a function defined over the subsets of U that maps each subset to the real numbers.

→ $P(U) = 1$, $P(A)$ $[0, 1]$

→ $P(A \in B) = P(A) + P(B) - P(A \cap B)$

→ if $A \cap B = \{\}$ then $P(A \cup B) = P(A) + P(B)$

Probability formula:-

$$\text{Probability of occurrence} = \frac{\text{No of desired outcomes}}{\text{Total no of outcomes}}$$

$P(\neg A)$ = probability of a not happening

$P(\neg A) + P(A) = 1$

Event : Each possible outcome of a variable

Sample space: collection of all possible events.

Random Variable : represents the event & objects

Prior probability : Computed before observing new information.

Posterior Probability : Combination of prior probability and new information.

Conditional Probability :-

* occurring an event when another event has already happened.

* Let's suppose, we want to calculate the event A when event B has already occurred, "the probability of A under the conditions of B", it can be written as,

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}}$$

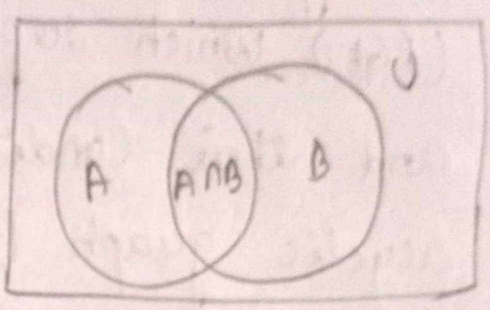
where,

$P(A \cap B)$ = Joint probability of A & B.

$P(B)$ = Marginal probability of B.

If the probability of A is given and we need to find the probability of B, then it will be given as,

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$



Example:-

In a class, there are 70% of the students who like English & 40% of the students who likes english & maths and then what is the present of students those who like english also like maths?

Solu:-

Let A is an event data that a students likes maths.

B is an event that a student likes english.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{0.4}{0.7}$$

$$= 57\%$$

Hence, 57% are the students who like english also like maths.

5) Bayesian Network:-

Bayesian n/w is a probabilistic graph (PGM) which represents a set of values and their conditional dependencies using a acyclic graph (DAG).

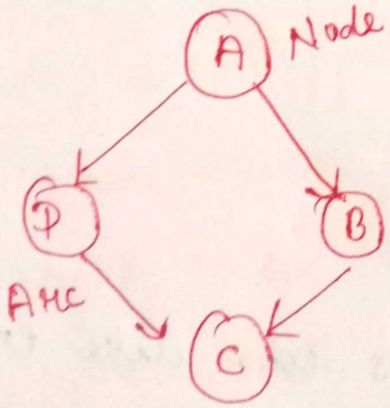
Burglary

Bayesian n/w : Directed acyclic graphs → Causal Structure.

Markov n/w : undirected graph → General dependencies.

Directed Acyclic Graph:-

1) Node represent random variable
2) Arcs represent direct Influence.
3) Nodes have conditional probability table.



→ A, B, C, D are random variables represented by the nodes of the n/w graph.

→ Node B, which is connected with node A by a directed arrow

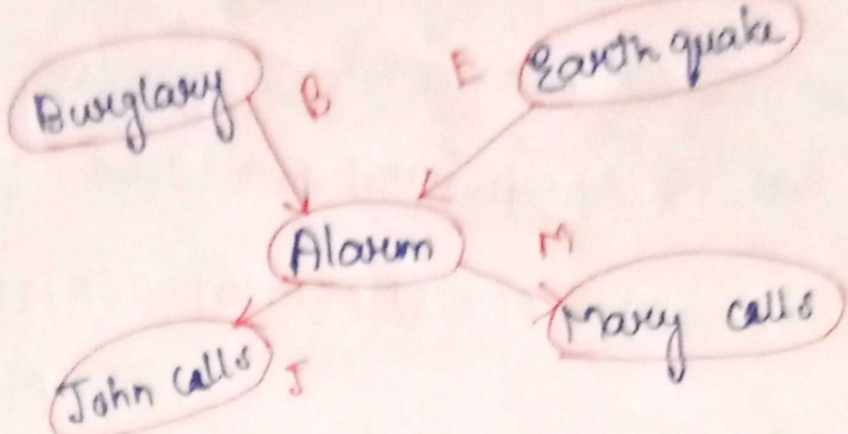
→ A is called parent of Node B.

→ Node C is independent of Node A.

Baysian network has mainly two components.

i) Causal Component

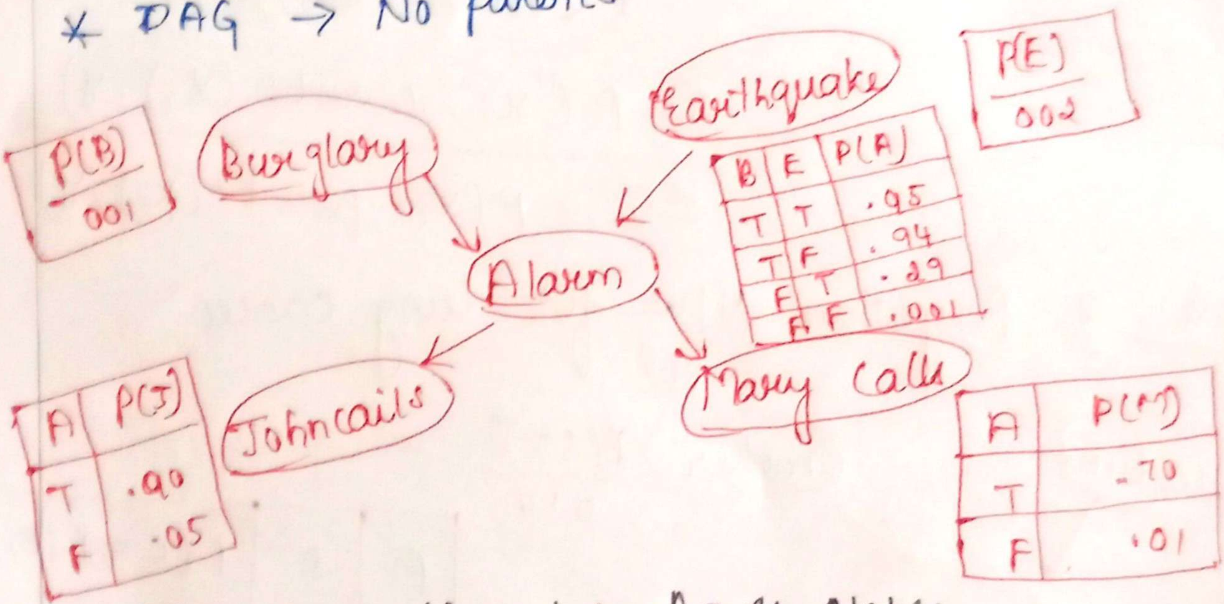
ii) Actual Numbers.



List of all events occurring in this n/w.

Conditional Probability Tables (CPT):-

* Each node has CPT that gives the probability of each of its values given every possible combination of values for parents.
* DAG → No parents.



P(B)
.001

P(E)
.002

B	E	P(A)
T	T	.95
T	F	.94
F	T	.29
F	F	.001

A	P(J)
T	.90
F	.05

A	P(M)
T	.70
F	.01

Joint distribution for Bayes Nets:-

Joint distribution can be expressed as product of local conditional probabilities.

Bayesian N/w implicity defines a joint distribution.

$$\boxed{P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i \mid Parents(x_i))}$$

Example:-

Alarm has sounded but neither a burglary nor earthquake has occurred and both John & M.

$$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E) = P(J|A)\, P(M|A)\, P(A|\neg B \wedge \neg E)$$
$$P(\neg B)\, P($$

$$= 9.0 \times 7.0 \times 0.001 \times 0.999 \times 0.998$$

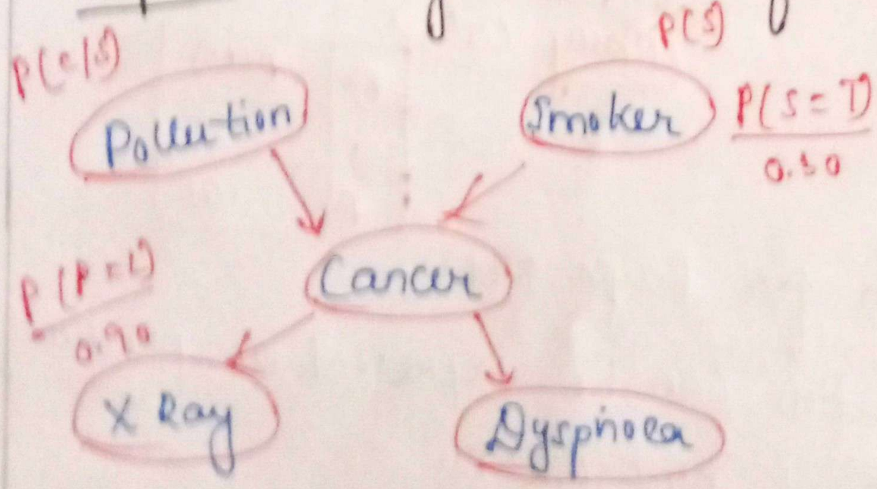
$$= 0.0062$$

Conditional probabilities can be computed from the joint distribution as follows

$$P(x_i \mid Parents(x_i)) = \frac{P(x_i, Parents(X_i))}{P(Parents(x_i))}$$

$$= \frac{\Sigma_y P(x_i, parents(x_i), y)}{\Sigma_{x_i, y} P(x_i, parents(x_i), y)}$$

Example: 2 Bayesian N/w for lung cancer.



P	S	P(C=T|P,S)
H	T	0.05
H	F	0.02
L	T	0.03
L	F	0.001

Chain rule:-

Reformulate the rule use of conditional probabilities.

$$P(x_1, \ldots x_n) = P(x_n / x_{n-1}, \ldots x_1)\, P(x_{n-1}, \ldots x_1)$$

Repeat process reduction of conjucative probabilities to a conditional dependency and a smaller conjunction

Final a big products.

$$P(x_1, \ldots x_n) = P(x_n / x_{n-1}, \ldots x_1)\, P(x_{n-1} / x_{n-2}, \ldots x_1) \ldots \ldots P(x_2/x_1)\, P(x_1) = \prod_{i=1}^{n} P(x_i / x_{i-1}, \ldots x_1)$$

This identity is called the chain rule,

Applications:

i) Prediction ii) Anomaly detection iii) Reasoning

iv) Time Series prediction v) Decision making under uncertainty.

6) Exact inference in BN:-

To compute the posterior probability distribution for a set of query variables, given some observed event usually, some assignment of values to a set of evidence variable.

X → Query variable

E → Set of evidence Variable

e → particular observed event

Y → denotes the hidden (non query) Variables.

Thus, the complete set of variables is $\{x$

posterior probability distribution $P(X|e)$.

Inference by enumeration:-

Any conditional probability can be computed

summing terms from full joint distribution.

$$P(X|e) = \alpha P(X, e) = \alpha \sum_{y} P(X, e, y).$$

Consider the query.

P(Burglary | John calls) = true, Mary calls = true)

Burglary n/w.

$$P(B|j, m) = \alpha P(B, j, m) = \alpha \sum_{e} \sum_{a} P(B, j, m, e, a).$$

CPT.

$$P(b|j, m) = \alpha \sum_{a} \sum_{e} P(b) P(e) P(a|b, e) P(j|a) P(m|a)$$

product of probability will be $O(2n)$.

Complexity $O(n2n)$.

Enumeration - Ask algorithm:-

ENUMERATION - ASK algorithm, evaluates these expression trees using depth first, left-to-right recursion.

Space Complexity - No. of. variables.

Time Complexity - $O(2n)$ better than $O(n2n)$.

function ENUMERATION - ASK (x, e, bn) returns a distribution over X.

inputs: X, the query variable.

e, observed values for variables E.

bn, a Bayes net with variable Vars.

Q (X) ← a distribution over X, initially empty. for each value x_i of X do

$Q(x_i)$ ← ENUMERATE - ALL (Vars, e_{x_i})

where e_{x_i} is e extended with $X = x_i$

return NORMALIZER (Q (X))

function ENUMERATE - ALL (Vars, e) returns a real No

if EMPTY? (Vars) then return 1.0.

V ← FIRST (Vars)

if V is an evidence Variable with Value Vine.

then return P^ (v | Parents (V)) X ENUMERATE - ALL (REST (Vars),

else return $\sum_{v}$ P (v | parents (V)) X ENUMERATE - ALL (REST (1).

where,

e_v is extendended with V = v

Eg: $f_1(a,b) \rightarrow f_2(b,c) = f(a,b,c)$.

A	B	$f_1(A,B)$	B	C	$f_2(B,C)$	A	B	C	
T	T	.3	T	T	.2	T	T	T	.
T	F	.7	T	F	.8	T	T	F	.3 ×
F	T	.9	F	T	.6	T	F	T	.7 × .6
F	F	.1	F	F	.4	T	F	F	.7 × .4
						F	T	T	.9 × .2 =
						F	T	F	.9 × .8 = .72
						F	F	T	.1 × .6 = .06
						F	F	F	.1 × .4 = .04

Clustering Algorithm:-

→ Variable elimination algorithm is simple.

→ Computation of posterior probabilities $O(n_2)$

→ using clustering algorithms, this can be reduced to $O(n)$.

→ Multiply connected n/w can be converted into a polytree.

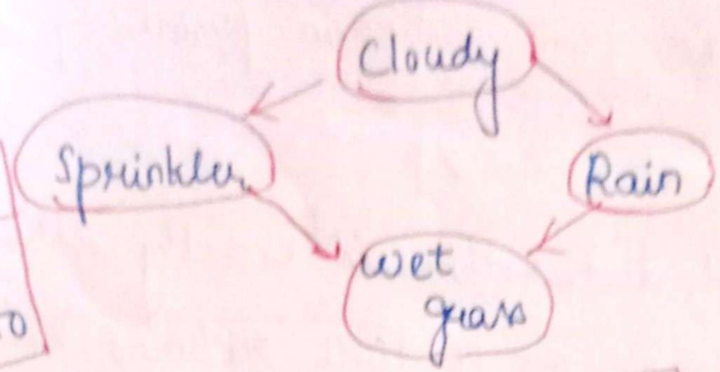
by combining sprinkler & rain node into cluster node called Sprinkler + Rain.

→ Two boolean nodes replaced by a mega-node.

4 possible values : TT, TF, FT, FF.

→ mega-nodes has only one parent.

$\boxed{P(C) = 5}$



	P(S)
T	.10
F	.50

C	P(R)
T	.80
F	.20

S	R	P(W)
T	T	.99
T	F	.90
F	T	.90
F	F	.00

$\boxed{P(C) = 5}$

	P(S+R = *)			
C	TT	TF	FT	FF
T	.03	.02	.72	.18
F	.10	.40	-10	.40

S+R	P(W)
TT	.99
TF	.90
FT	.90
FF	.00

I) Approximate Inference in Bayesian n/w:-

→ Approximate answers whose accuracy depends on the no samples generated.

→ Sampling applied to the Computation of posterior probabilities.

Two families of algorithm:-

i) Direct Sampling

ii) Markov Chain Sampling

Direct Sampling methods:-

→ Generation of samples from a known probability

$\boxed{\text{EX : } P(\text{coin}) = \langle 0.5, 0.5 \rangle}$

→ Sampling from this distribution is exactly like the coin: with probability 0.5 it will returns heads, and with probability 0.5 it will return tails.

function - PRIOR - SAMPLE (bn) return an event sampled from the prior specified by bn.

inputs: bn, a Bayesian n/w specifying joint distribution. $P(X_1, \ldots, X_n)$

$x \leftarrow$ an event with n elements.

for each variable X_i in $X_1, \ldots X_n$ do

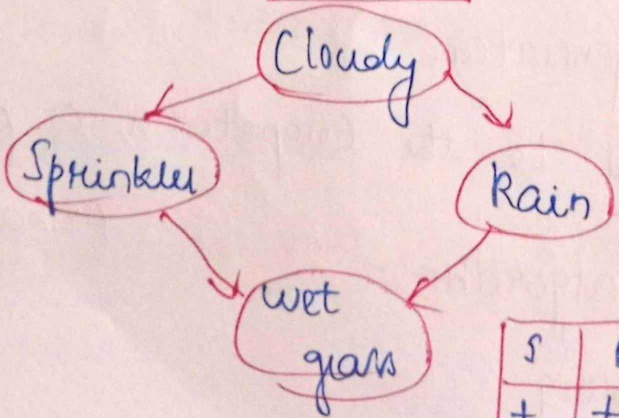
$x[i] \leftarrow$ a random sample $P(x_i \mid \text{parents}(x_i))$

return x.

PRIOR - SAMPLE:

$$S_{PS}(x_1, \ldots x_n) = \prod_{i=1}^{n} P(x_i \mid \text{Parents}(x_i))$$

$\boxed{P(C) = 5}$



C	P(S)
t	.10
f	.50

C	P(R)
t	.80
f	.20

S	R	P(W)
t	t	.99
t	f	.90
f	t	.90

h sampling step depends only on the parent values S_{ps} (11)

$(x_1, \ldots x_n) = P(x_1, \ldots x_n)$

$$\lim_{N \to \infty} \frac{N_{PS}(x_1 \ldots x_n)}{N} = S_{ps}(x_1, \ldots x_n) = P(x_1 \ldots x_n)$$

S_{ps} (true, false, true, true) $= 0.5 \times 0.9 \times 0.8 \times 0.9$

$= 0.324.$

Rejection Sampling:-

function REJECTION - SAMPLING (X, e, bn, N) returns an estimate of P(X / e).

inputs : X, the query variable.

e, observed value for variable E.

bn, a Bayesian n/w

N, the total no of samples to be generated local variable N a vector of counts for each value of X, initially zero.

for j = 1 to N do

X ← PRIOR - SAMPLE (bn)

if X is consistent with e then,

N(x) ← N[x] + 1 where x is the value of X in x.

return NORMALIZE (N).

Rejection Sampling Algorithm:-

Let $P(x|e)$ be the estimated distribution

$$\hat{P}(x|e) = \alpha\, N_{PS}(x,e) = \frac{N_{PS}(x,e)}{N_{PS}(e)}$$

$$\hat{P}(x|e) \simeq \frac{P(x,e)}{P(e)} = P(x|e).$$

Rejection sampling produce a consistent estimate of true probability

Estimate P(Rain / Sprinkler = true), using 100 samples of the 100 that we generate, suppose that 73 have Sprinkler = false and are rejected, while 27 have sprinkler = true of the 27, 8 have rain = true & 19 have rain = false.

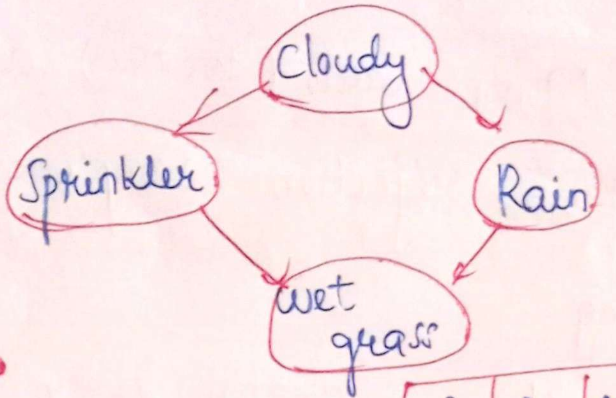
P(Rain | sprinkler = true) NORMALIZE (<8, 19>) = <0.296, 0.704> 26.

Likelihood Weighting:-

→ Likelihood weighting avoids the inefficiency of rejection sampling.

→ Generates only events that are consistent with the evidence.

$\boxed{P(C) = 5}$



C	P(S)
t	.10
f	.50

C	P(R)
t	.80
f	.20

S	R	P(W)
t	t	.99
t	f	.90
f	t	.90
f	f	.00

Cloudy is an evidence variable $\leftarrow$ Wx $P[\text{Cloudy} = \text{true}) = 0.5$

Sprinkler is not an evidence variable $\leftarrow$ $P[\text{Sprinkler} \mid \text{Cloudy} = \text{true})$

weight for a given sample X is the product.

$$w(z, e) = \prod_{i=1}^{m} P(e_i \mid \text{parents}(E_i))$$

$$S_{ws}(z, e)\, w(z, e) = \prod_{i=1}^{l} P(z_i \mid \text{parents})(z_i)) \prod_{i=1}^{m} P(e_i / \text{parents}(E_i))$$

Calculated as follows,

$$\hat{P}(x/e) = \alpha \sum_{y}^{l} N_{ws}(x, y, e)\, w(x, y, e)$$

$$\simeq \alpha' \sum_{y} S_{ws}(x, y, e)\, w(x, y, e)$$

$$\simeq \alpha' \sum_{y} P(x, y, e)$$

$$= \alpha' p(x, e) = P(x/e)$$

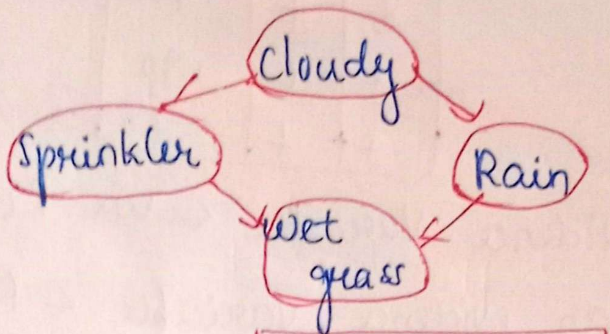
Hence, likelihood weighting returns consistent estimates.

Inference by Markov chain simulation

→ Markov Chain Monte Carlo (MCMC) alg. ... quite differently from rejection sampling and likelihood weighting.

→ MCMC state change similar to SA.

$\boxed{P(C) = 5}$



C	P(S)
t	.10
f	.50

C	P(R)
t	.80
f	.20

S	R	P(W)
t	t	.99
t	f	.90
f	t	.90
f	f	.00

P(Rain/sprinkler) = true, wet grass = true

Initial state is [true, true, false, true].

Cloudy:-

P(Cloudy | sprinkler = true, Rain = false)

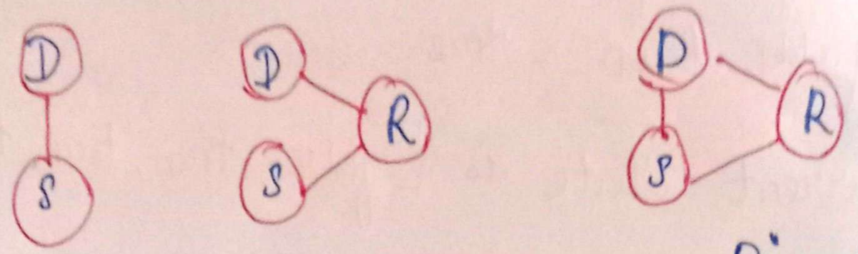
Suppose the result is Cloudy = false.

Then the new current state is [false, true, false, true].

Rain:-

Given the current values of its Markov blanket.

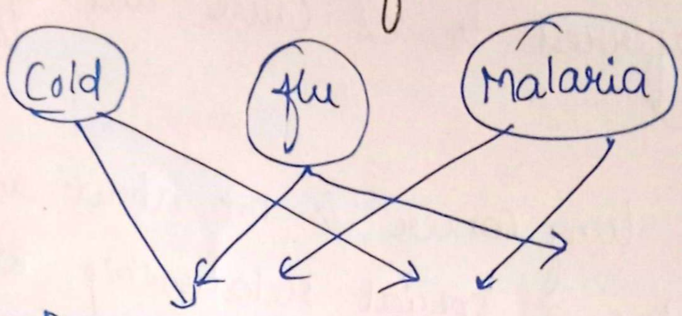
Types of cause :-



a) Direct cause   b) Common cause   c) Directed + Common

Example :-

Causal N/w for disease & symptoms :-

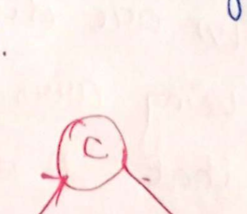


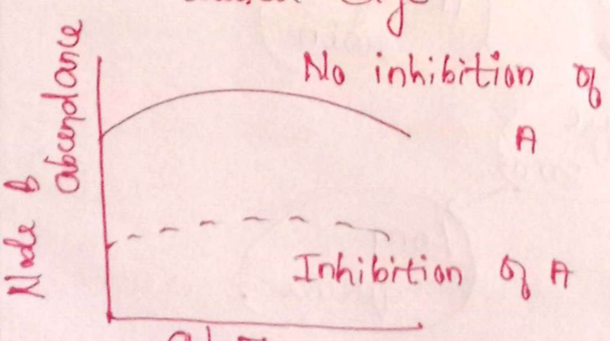
Causal Bayesian n/w with causes (diseases) Cold, flu and malaria and effects (Symptoms) Nausea & Headache.

Types of causal N/w :-

a) Directed causal edge may represent direct effect. Inhibitation of parent node A can change the abundance of the child node B.

Causal edge





a) Time

un measured node

Indirect effect (b)

indirect causal edge may represent indirect effect:- (4)

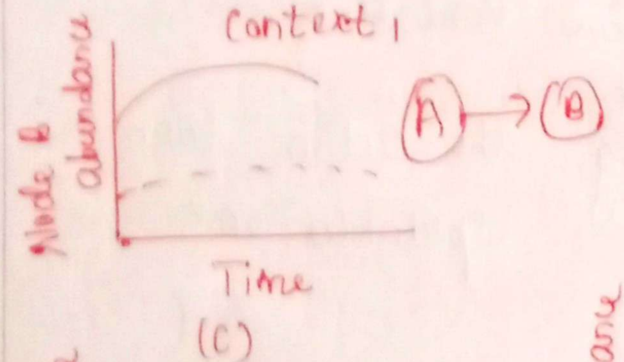
, That occur via unmeasured intermidiate node

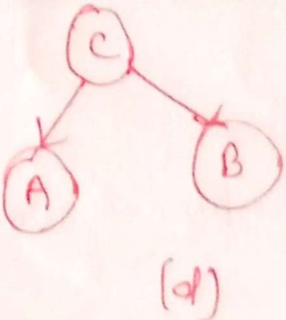
→ If node A causally influences node B via measured node C, the causal n/w should contain edges from A to C from C to B.

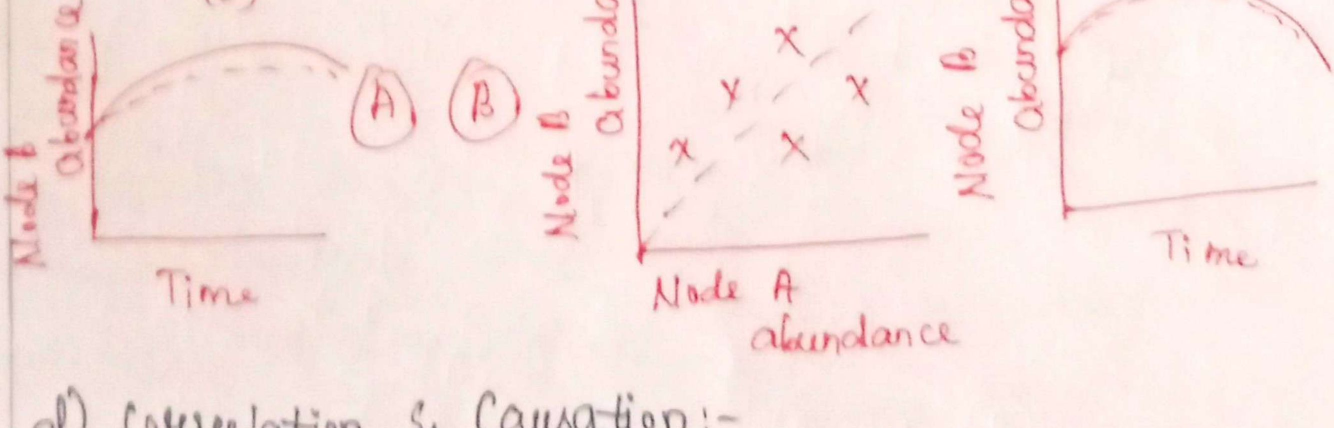
→ If node C is not measured.

c) Causal edges depend on biological Context:-

Causal edge from A to B appears in Context 1 not in Context 2.





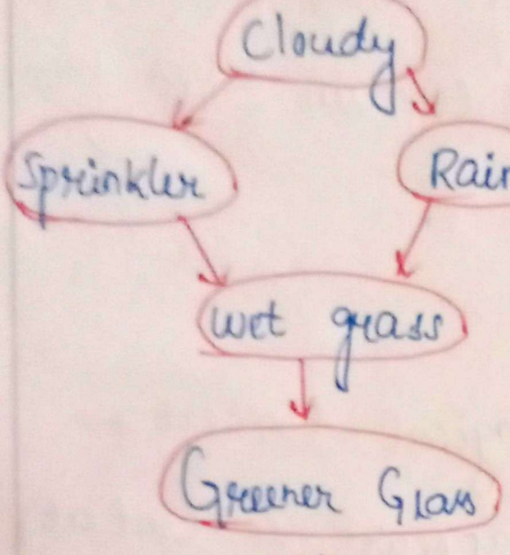


d) Correlation & Causation:-

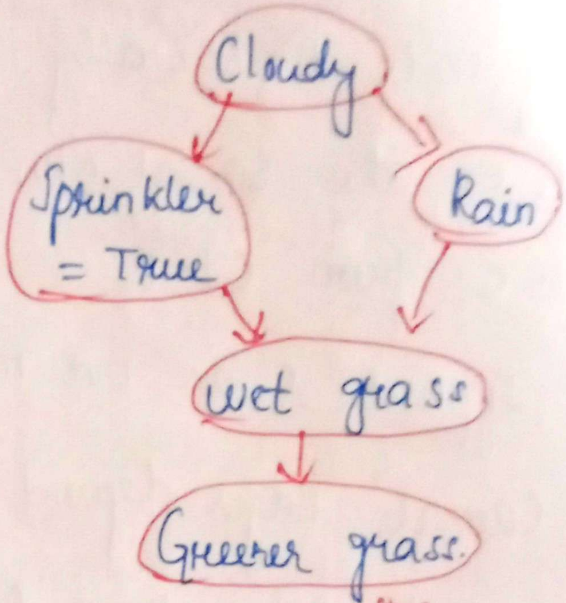
Node A & B are correlated owing to regulation by the same node (C) but in this example no sequence of mechanistic events links A to B and thus inhibition of A does not change the abundance of B.

There is no causal edge from A to B

Another Example:-



a)



(b)

a) A causal Bayesian n/w representing cause-effect among five Variables.

b) The n/w after performing the action "turn Sprinkler on".