UNIT → 3. DATE

Quantifying Uncertainty

Acting Under Uncertainty
Basic Probability Notation.
Inference Using full joint distribution
Independence
Bay's Rule & its ~~Rute~~ Use.

Probabilistic reasoning - Representing knowledge in an uncertain Domain
The Semantics of Baysian N/w or Baysian Belief Network
Efficient distrubution Representation of conditional distributon
Exact inference in Baysian N/w
Probabilistic Reasoning over time
- Time and uncertainty
Inference in temporal models
Hidden Markov Models
Kalman Filters.

Uncertainty

Acting Under Uncertainty

⇒ Uncertainty :

An agent may never know for certain what state it is in or where it will endup after a sequence of actions.

⇒ Uncertainty in systems (or agents) arises primarily because of problems in data, major causes being

- that data is missing or unavailable
- data is present but unreliable or ambiguous due to errors in measurement
- presence of multiple conflicting measurement. etc.

Or you can say :

⇒ Agents may need to handle uncertainty, whether due to partial observability, non-determinism or a combination of two.

⇒ Uncertainty may also be caused by the represented knowledge since it might represent the best guesses of the expert based on observations or statistical conclusions, which may not be appropriate in all the situation.

⇒ Agents designed to handle uncertainty by keeping track of belief state - a representation of all possible world states that it might be in - and generating a contingency plan that handles every possible eventuality that its sensors may report during execution.

DATE

⇒ This approach has significant drawbacks, like

- when interpreting partial sensor information, agent must consider every logically explanation, this leads to impossible large and complex belief set
- A correct contingent plan that handles every eventuality can grow arbitrarily large.
- Sometimes there is no plan that guaranteed to achieve the goal - yet the agent must act.

Uncertainty, Rational Decision, Utility Theory, Decision Theory

⇒ Rational decision: Depends on the relative importance of various goals and the likelihood that and degree to which they will be achieved.

⇒ Probability theory: probability provides a way of summarizing the uncertainty that comes from our laziness & Ignorance.

⇒ Utility theory: It says that every state has a degree of usefulness or utility, to an agent and agent will prefer states with higher utility.

⇒ Decision Theory: DT = probability theory + utility theory.

An agent is rational iff it chooses the action that yields the highest expected utility, averaged over all the possible outcomes of the action.

⇒ This is called the Maximum Expected Utility (MEU)

Probability Theory

⇒ The term probability is defined as a way of turning an opinion or an expectation in to number lying 0 and 1.

⇒ It basically reflects the likelihood of an event, or a chance that a particular event will occur.

⇒ probability of an event A is denoted by P(A) and defined as:

$$P(A) = \frac{(\text{No. of outcomes favourable to } A)}{(\text{Total no. of possible outcomes})}$$

⇒ The probability of all events $\{A_1, A_2 --, A_n\}$ must sum to certainty.

$$P(A_1) + P(A_2) + --- P(A_n) = 1$$

⇒ Tossing coin, getting two ~~conse~~ successive heads $\{HH, HT, TH, TT\}$

$P(HH) = 1/4 = 0.25$

⇒ $P(A') = 1 - P(A)$

$P(A \cup B) = P(A) + P(B)$ if A and B are mutually exclusive

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ if A and B are not mutually exclusive.

DATE

: Joint Probability :

⇒ JP is defined as the probability of occurrence of two independent events in conjunction.

⇒ P(A and B) = P(A) * P(B)
or P(A ∩ B)

⇒ Two events are said to be independent if the occurance of one event does not ~~affact~~ affect the probability of occurance of the other.

⇒ Tossing of two fair coin seprately.
The pro. of getting H on tossing a first coin
$P(A) = 0.5$
and tossing second coing, getting Head
$P(B) = 0.5$

The probability of getting H on both the coins is c/d joint probability.

$P(A \text{ and } B) = 0.5 * 0.5 = 0.25$

Similarily, the pro.

$P(A \text{ or } B) = 0.5 + 0.5 - (0.5 * 0.5)$
$= 0.75$

DATE

Joint Probability distribution of two Variables A and B :

Joint Probabilities	A	A'
B	0.20	0.12
B'	0.65	0.03

$P(A \text{ and } B) = 0.20$ $\quad$ $P(A' \text{ and } B') = 0.03$

$P(A \text{ and } B') = 0.65$ $\quad$ $P(A' \text{ and } B) = 0.65$

Joint probability of n variables require 2^n entries.

We can easily compute $P(A)$ and $P(B)$ as

$$P(A) = P(A \text{ and } B) + P(A \text{ and } B')$$
$$= 0.20 + 0.65 = 0.85$$
$$P(B) = P(A \text{ and } B) + P(A' \text{ and } B)$$
$$= 0.20 + 0.12 = 0.32$$

In fact, we can compute probability of any logical combination of A and B.

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

$$P(A \text{ or } B) = 1 - (P(A \text{ or } B)')$$
$$= 1 - P(A' \text{ and } B')$$

⇒ So far, we have considered the probabilities of events and logical combinations of the events.

⇒ However, we can come across situations where the probability of a stmt is based on piece of evidence. This is c/d "Conditional Probability."

Conditional Probability

(Probability of an event depends on the probability of other event)

It is defined as the probability of the occurence of event H (hypothesis) provided an event E (evidence) is known to have occured.

$$P(H|E) = \frac{\text{No. of events favourable to H which are also favourable to E}}{\text{No. of events favourable to E}}$$

$$\boxed{P(H|E) = \frac{P(H \text{ and } E)}{P(E)}}$$

Here $P(E) \neq 0$

If events H and E defined on a sample space S of a random experiment are independent then

$$P(H|E) = P(H)$$

$$P(E|H) = P(E)$$

Q.1) If prob. of any person chosen at random being litrate as 0.40 and probability of any person chosen at random having age > 60 years as 0.005

find the probability of the fact that a person chosen at random of age > 60 years is litrate.

DATE

P(X is litrate and the age X > 60) =

= P(X is litrate) * P(age of X > 60)

= 0.90 * 0.005

= 0.002

P(X is litrate | Age x > 60)

$$= \frac{P(X \text{ is litrate and age } X > 60)}{P(\text{Age of } X > 60)}$$

= 0.002 / 0.005 = 0.4

Q.2 Sun is bright today P(sunny_today) = 0.6 (A)
Sun will be bright tomorrow P(sunny_tom) = 0.8, (B)
provided today the sun is bright.

Calculate

P(B|A),	P(~B|A),	P(B|~A),	P(~B|~A)
0.8 given	0.2 given	0.6 given	0.4 given

P(B, A), P(~B, A), P(B, ~A) P(~B, ~B)

We often have more no. of variables in that case drawing distribution table is not as simple.

Thomas Bayes proposed Bay's Theorem for two events using the concept of conditional probability.

DATE

Bay's Theorem (Thomas Bayes - 1763)

⇒ This theorem provides the model for reasoning where prior beliefs are combined with evidence to get estimates of uncertainty.

⇒ It relates the conditional probability and probabilities of events H and E.

⇒ The basic idea is to compute P(H|E) that represents the prob. assigned to H after taking into account the new piece of evidence E.

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

Derivation from the conditional probability

$$P(H|E) = P(H \text{ and } E) / P(E)$$

$$\Rightarrow P(H|E) * P(E) = P(H \text{ and } E) \quad \text{—— (1)}$$

similarly

$$P(E|H) * P(H) = P(E \text{ and } H) \quad \text{—— (2)}$$

from Eq^n (1) & (2)

$$P(H|E) * P(E) = P(E|H) * P(H)$$

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

This is Bay's Theorem.

We can also express the probability $P(H|E)$ in terms of $P(E|H)$, $P(E|\sim H)$, $P(H)$

$$\boxed{P(H|E) = \frac{P(E|H) * P(H)}{P(E|H) * P(H) + P(E|\sim H) * P(\sim H)}} \quad (A)$$

Proove

$$P(E) = P(E|H) * P(H) + P(E|\sim H) * P(\sim H)$$

We know

$$P(E) = P(E \text{ and } H) + P(E \text{ and } H')$$

using Conditional probability

$$P(E \text{ and } H) = P(E|H) * P(H)$$
$$P(E \text{ and } \sim H) = P(E|\sim H) * P(\sim H)$$

Therefore we can substitute above in eq^n (A)

⇒ $P(H)$ is known as prior probability of H. It is called prior because it does not consider any information regarding E.

⇒ $P(H|E)$ is known as posterior probability because it is derived from or depends on the specified value of E.

DATE

Q.3) pro. of mike has a cold 0.25, pro. of mike was obrerved sneezing when he had cold in past 0.9 and the prob Mike was observed sneezing when he did not have cold 0.20.
find the prob. of mike having a cold given that he sneezes.

Solution

H - Mike has cold
E - Mike sneezes

$P(H|E)$

$P(H) = 0.25$

$P(E|H) = 0.90$

$P(E|\sim H) = 0.20$

$$P(H|E) = \frac{P(E|H) * P(H)}{P(E|H) * P(H) + P(E|\sim H) * P(\sim H)}$$

$$= \frac{0.9 * 0.25}{(0.9)(0.25) + (0.20)(0.75)}$$

$$= \frac{0.9 * 0.25}{0.375}$$

$$= 0.6$$

Therefore we conclude that Mike's probability of having a cold given that he sneezes is equal to 0.6.

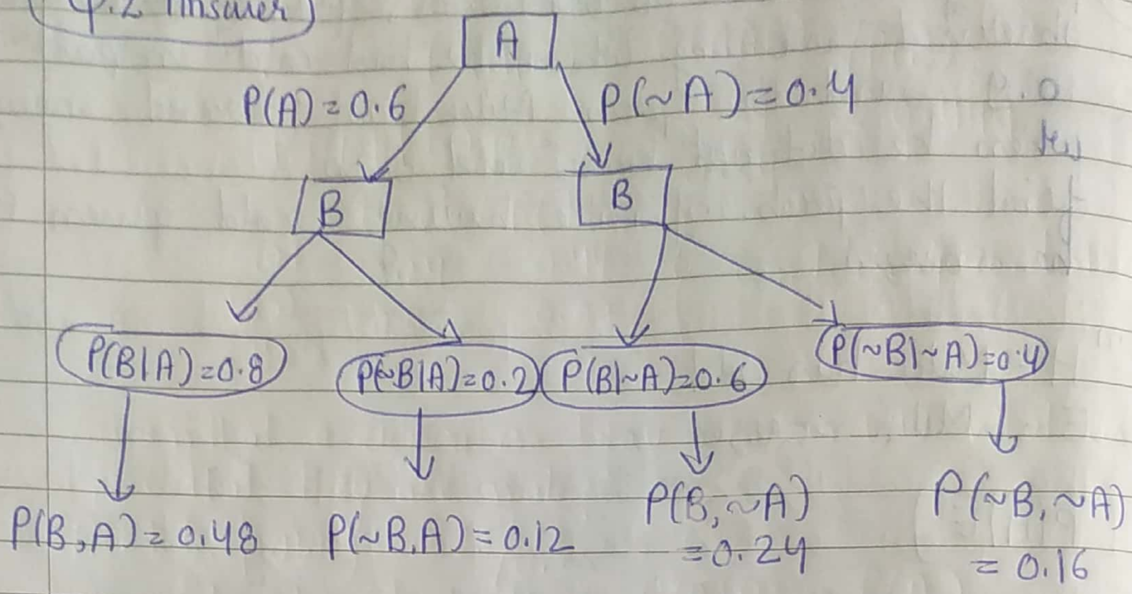
Similarly we can determine his probability of having a cold if he was not sneezing

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

$$= \frac{[1-0.9] * 0.25}{1-0.375}$$

$$= \frac{0.025}{0.625} = 0.04$$

Q.2 Answer



Independence: If two events A & B are independent

$$P(A|B) = P(A)$$
$$P(B|A) = P(B)$$

or

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{P(A)\,\cancel{P(B)}}{\cancel{P(B)}}$$

Independence assertions are usually based on knowledge of the domain.

The HBO Cable n/w took a survey of 500 subscribers to determine people's favourite show

	Male	Female	Total
Game of thrones	80	120	200
West World	100	25	125
others	50	125	175
Total	230	270	500

Now divide this table by 500 and we will get probability distribution

	M	F	Total
GoT	0.16	0.24	0.4
WW	0.2	0.05	0.25
others	0.1	0.25	0.35
Total	0.46	0.54	1

+ = 1

Marginal probability Distribution

$P(F \cap GoT) = 0.24$ Joint Prob.

→ Joint prob. distribution sums to 1

Marginal prob. = $P(GoT) = 0.4$

$P(Male) = 0.46$ what is the probability of an HBO subscribers being male?

→ What is the pro. of an HBO subscriber preferring WW?

$P(WW) = 0.25$

→ What is the prob of an HBO subscriber being male AND prefering WW? $P(Male \cap WW) = 0.2$

→ What is the prob of an HBO subscribers being Male Or prefering WW?

$P(M \cup WW)$

$= 0.16 + 0.2 + 0.1 + 0.05$

$= 0.51$

or $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
$= 0.46 + 0.25 - 0.2 = 0.51$

Conditional Probability ! a girl first got [illegible] HBO subscription. What is the chance that her favorite show will be GoT?

$$P(A|B) = P(A \cap B) / P(B)$$

$$P(GoT \mid Female) = \frac{0.24}{0.54} = 0.444$$

	Female	P(show|F)
GoT	0.24	0.44
WW	0.05	0.09
others	0.25	0.46
		1

} conditional pro. distribution

Given that a subscriber's fav show is WW. What is the pro. that they are male?

$P(Male \mid WW)$ $0.2 / 0.25 = 0.80$

If independent then $P(A|B) = P(A)$

$P(WW \mid Female) = 0.05 / 0.54 = 0.093$

$P(WW) = 0.25$

Therefore not indepent $0.093 \neq 0.25$

Q. Given the prob. of the stmt "John has a viral" is 0.20, prob. of John being observed sneezing when he had viral is 0.8, John being observed sneezing when he did not have viral is 0.2. find pro:

1) John having viral if he is seen sneezing.

2)

Q. Consider the problem of tossing of four fair coins separately. Let the prob. of getting head H on each coin is 025. find prob.

1) H on all four coins
2) H and tail on 2 coins each
3) H on 3 coins and tail on one coin

Q. Find the prob of a person chosen at random having age >80 years is female.

1) prob of choosing any person at random female is about 0.48
2) prob of choosing any person at random having age >80 years is 0.002.

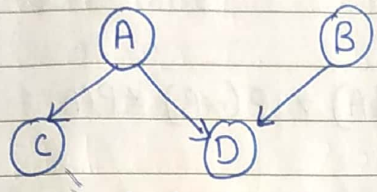
$P(WW \cap F) = 0.05$
$P(WW) * P(F) = 0.14$) $\neq$ not equal
$0.54 * 0.25$

Not independent

DATE

Inference Using Baysian N/w OR Exact Inference in Baysian N/w

How we can compute the influence of a node on any other node ? using BBN/w



"what is the likelihood or probability that the hypothesis is A, given C" ?

$P(A|C) = ?$

In this n/w we have four variables.

$$\boxed{P(A|C) = \frac{P(A,C)}{P(C)}} \quad \text{---} \; (1)$$

where $P(A,C) = \sum_{B,D \in \{T,F\}} P(A,B,C,D)$ ——— (2)

$$P(C) = \sum_{A,B,D \in \{T,F\}} P(A,B,C,D) \quad \text{---} \; (3)$$

$$\sum_{B,D \in \{T,F\}} P(A,B,C,D) = P(A,B,C,D) + P(A,\sim B,C,D) + P(A,B,C,\sim D) + P(A,\sim B,C,\sim D) \quad \text{---} \; (4)$$

$$\sum_{A,B,D \in \{T,F\}} P(A,B,C,D) = P(A,B,C,D) + P(A,B,C,\sim D) + P(A,\sim B,C,D) + P(A,\sim B,C,\sim D) + P(\sim A,B,C,D) + P(\sim A,\sim B,C,D) + P(\sim A,B,C,\sim D) + P(\sim A,\sim B,\sim C,\sim D) \quad \text{---} \; (5)$$

put eqn no (4) & (5) in eqn no (1) you can calculate $P(A|C)$.

DATE

The Values of each component of the $P(A,C)$ may be computed :

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

$= 0.7 * \overset{*0.4}{0.6} * 0.3$

$= 0.0504$

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

$= 0.4 * 0.4 \times 0.4 * 0.3$

$= 0.019$

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

$= 0.3 \times 0.4 \times 0.6 \times 0.3$

$= 0.0216$

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

$= 0.6 \times 0.4 * 0.4 \times 0.3$

$= 0.0288$

So the value of numerator $P(A,C)$

$= 0.0504 + 0.019 + 0.0216 + 0.0288$

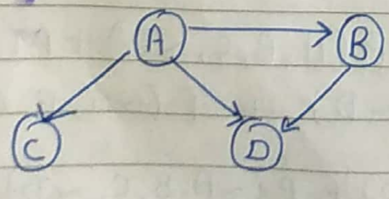
$P(A,C) = 0.12$

Similarly, the values of components of the denominator can be computed.

$P(C) = 0.3608$

$P(A|C) = \frac{0.12}{0.3608} = 0.33$ (appro.)

Q.



$P(A) = 0.4$

$P(B|A) = 0.5$

$P(B|\sim A) = 0.1$

$P(C|A) = 0.6$

$P(C|\sim A) = 0.3$

$P(D|A,B) = 0.8$

$P(D|A,\sim B) = 0.3$

$P(D|\sim A,B) = 0.3$

$P(D,|\sim A,\sim B) = 0.05$

Generate Conditional pro. table and compute:

① Joint prob $P(A,B,C,D)$

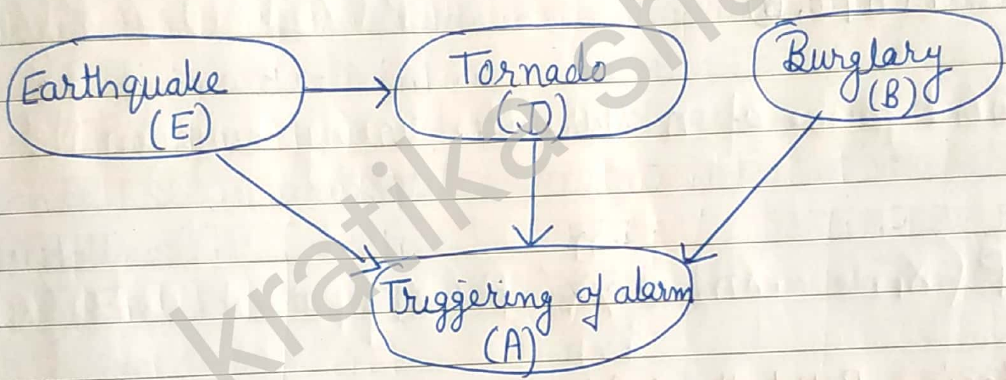
② $P(A|B)$

③ $P(A|C)$

④ $P(A|B,C)$

Simple Baysian Belief N/w : Example

Consider an alarm system installed in a house, that can be triggered by 3 events
earthquake
burglary
tornado. This can be modelled in Baysian n/w



Here E and B are independent variable and A, D are dependent variables.
probability values for complete baysian n/w:

P(E)	P(B)
0.4	0.7

E	P(D)
T	0.8
F	0.5

E	B	D	P(A)
T	T	T	1.0
T	T	F	0.9
T	F	T	0.95
T	F	F	0.85
F	T	T	0.89
F	T	F	0.7
F	F	T	0.87
F	F	F	0.3

The joint prob may be computed as:

$$P(E, B, D, A) = P(A \mid E, B, D) * P(D \mid E) * P(E) * P(B)$$
$$= 1.0 * 0.8 \times 0.4 \times 0.7$$
$$= 0.214$$

DATE

Inference using B B N/w

Using this model

what is the probability that it is an earthquake, given alarm is ringing?

$P(E|A)$

what is the prob. of alarm ringing if both earth equke & burglary occur?

$P(A|E, B)$

Basian Belief N/w : Advantages & Disadvantages:

→ It can easily handle situations where some data entries are missing.

→ easier for human to understand direct dependencies and local distributions than complete joint distribution.

classmate

DATE

~~Bayesian Belief Network (BBN)~~

~~⇒ BBN, which is~~

Baysian Network (Baysian Belief Network)

⇒ It is a data structure, use to represent dependencies among variables.

⇒ Baysian N/W can represent essentially any full joint probability distribution.

⇒ A Baysian N/W is a directed graph in which each node is annotated with quantitative probability infor. as follows:

1) Each node corresponds to random variable (discrete or continuous)
2) A set of directed links connects pairs of nodes.

$$X \rightarrow Y$$

(X is parent of Y) (Y is called child of X)

It is directed Acyclic graph (DAG)

3) Each node X_i has a conditional pro. distribution $P(X_i \mid \text{parents}(X_i))$, that quantifies the effect of the parents on the node.

⇒ It is a probabilistic graphical model that encodes probabilistic relationship among the set of variables with their probabilistic dependencies.

⇒ For example, a bayesian n/w could be used to represent the probabilistic relationships between diseases & symptoms. That is given certain symptoms, the n/w can compute the probabilities of the presence of various disease.

DATE

⇒ Joint probability of n variables : (Dependent or independent with out bayesian n/w)

$$P(X_1, \ldots, X_n) = P(X_n | X_1, \ldots X_{n-1}) * P(X_1, \ldots X_{n-1})$$

or

$$P(X_1, \ldots X_n) = P(X_n | X_1 \ldots X_{n-1}) * P(X_{n-1} | X_1 \ldots X_{n-2}) * \ldots * P(X_2 | X_1) * P(X_1)$$

⇒ Joint prob. of n variables using Baysian Network

$$\boxed{P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \text{parent-nodes}(X_i))}$$

⇒ It is the product of the local distributions of each node and its parents.

⇒ This expression is reduction of joint probability formula of n variables as some of the terms corresponding to independent variables will not be required.

⇒

DATE

Probabilistic Reasoning Over time

Time and Uncertainty:

→ We saw techniques for probabilistic reasoning in the content of static worlds, in which each random variable has a single fixed value.

→

States and Observations:

Transition and sensor models.

After deciding set of states and evidence variables for a given problem, the next step is to specify how the world evolves (the transition model)

Time & Probabilistic reasoning over time

Time & Uncertainity

→ Static world

→ repairing a car

→ diabetic patient.

states & observations

X_t – set of state variable at time t unobservable

E_t – set of obs. evidence variable.

→ time slice
↓
set of variables
↓
some are obs some are not

Security guard

E_t – single evidence variable
Umbrella U_t

X_t – single state variable
$Rain_t$

Diabetic Eg:

E_t – MBS_t, PR_t

X_t – BS_t, $StoCont$.

time Slice
$t_0 = 0$

Umbrella world

State R_0 -- R_2

evidence U_1 -- U_3 -

a:b $X_{a:b}$ X_a to X_b set of variable

Tran. & Sensor Model

→ how the world evolves (the transition model)

→ how the evidence variables get their values (sensor model)

→ The transition model specifies the Pro. dis. over the latest variables

$$P(X_t \mid X_{0:t-1})$$

Unbounded in size as t increases

↓ soln

Markov Assumption

The Current state depends on only a finite fixed no. of previous state.

Mar Process or chains

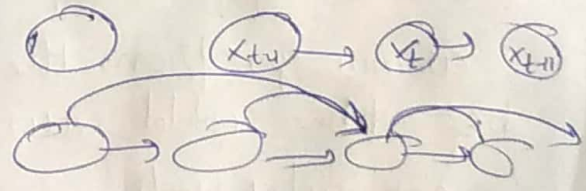
F-O-M-P

$$P(X_t \mid X_{0:t-1}) = P(X_t \mid X_{t-1})$$

Conditional distribution

SOMP

$$P(X_t \mid X_{t-2}, X_{t-1})$$



Baysian n/w structure for FOMP & SOMP

→ still there is a problem;
infinitely many possible values of t.
diff. distribution for each time step?

" Changes in the world state are caused by a stationary process -

~~stationary~~ ↓ static

(In static process, the state itself does not change)

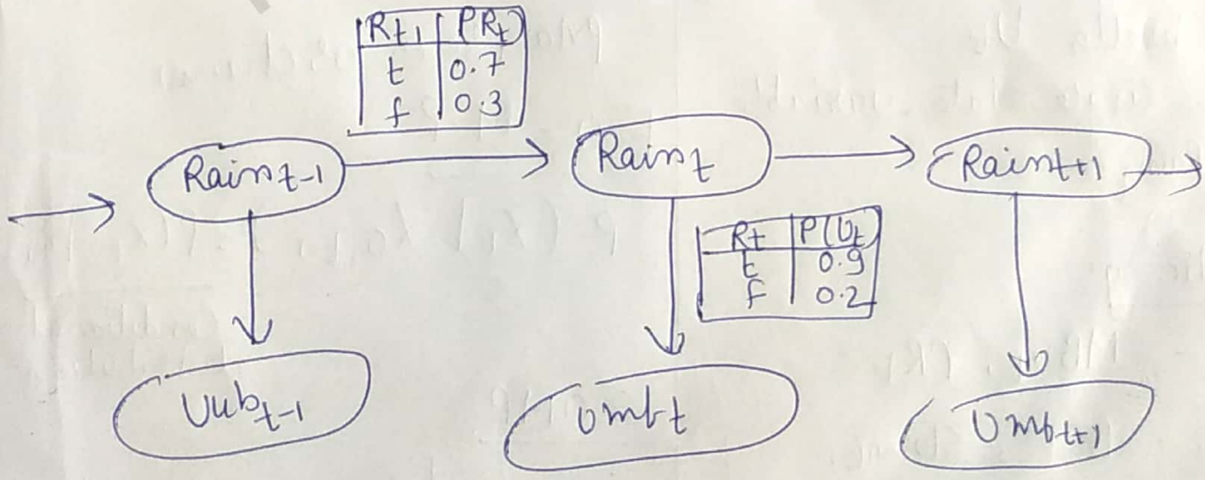
→ Condi pro Umb. rain $P(R_t | R_{t-1})$
same for all t.

Sensor Model

$$P(E_t \mid X_{0:t-1}, E_{0:t-1}) = P(E_t \mid X_t)$$

Sensor Model or observation Model.

E_t could depend on previous variables as well as the current state variables.



Direction of the dependence b/w state and sensors:
arrow go from the actual state to sensor values
bcz the world causes the sensors to take on particular value.

Inference in Temporal Model

① **Filtering:** Task of computing the belief state – posterior probability over the current state, given all evidence up to present.

$$\boxed{P(X_t \mid e_{1:t})}$$

② **Prediction:** posterior probability over a future state, given all evidence up to present.

$$\boxed{P(X_{t+k} \mid e_{1:t})} \quad \text{for some } k > 0$$

③ **Smoothing:** posterior probability over a past state given all evidence up to present.

$$\boxed{P(X_k \mid e_{1:t})}$$

④ **Most likely Explanation (Best Sequence)**

Best state sequence given all evidence upto present.

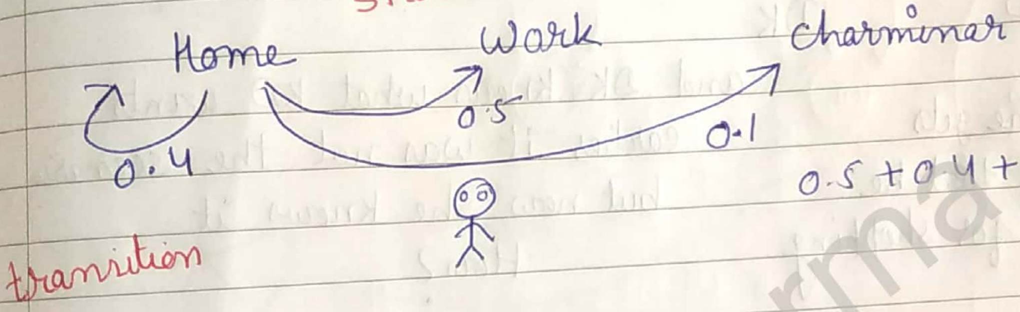
$$\boxed{\arg\max_{x_{1:t}} P(X_{1:t} \mid e_{1:t})}$$

Markov Models

→ Andrew Markov (1856-1922)
→ Markov processes - chain of events that are memoryless

Previous event → current event → Next event

states



$0.5 + 0.4 + 0.1 = 1$

$\frac{5}{10} = 0.5$

10 → total 10 times we observed the person

From

$$\begin{vmatrix} & \\ & \end{vmatrix} = \text{to} \begin{vmatrix} 0.4 & 0.6 & 1.0 \\ 0.5 & 0.3 & 0.0 \\ 0.1 & 0.1 & 0.0 \end{vmatrix} \begin{vmatrix} 1.0 \\ 0.0 \\ 0.0 \end{vmatrix}$$

Transition probability Matrix

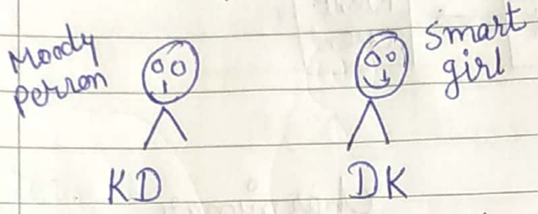
prediction for future

starting at home

→ future prediction depends only on present state
(no past, no future)

Hidden Markov Model - Actual states are hidden from observes.

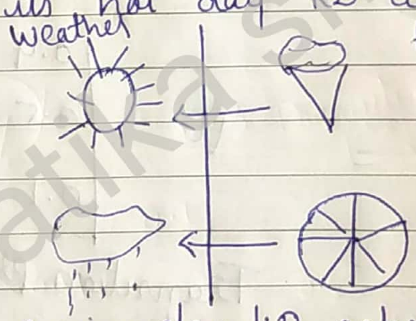
→ It is a statistical markov model in which system is being modeled assumed to be a markov process with unobserved state.



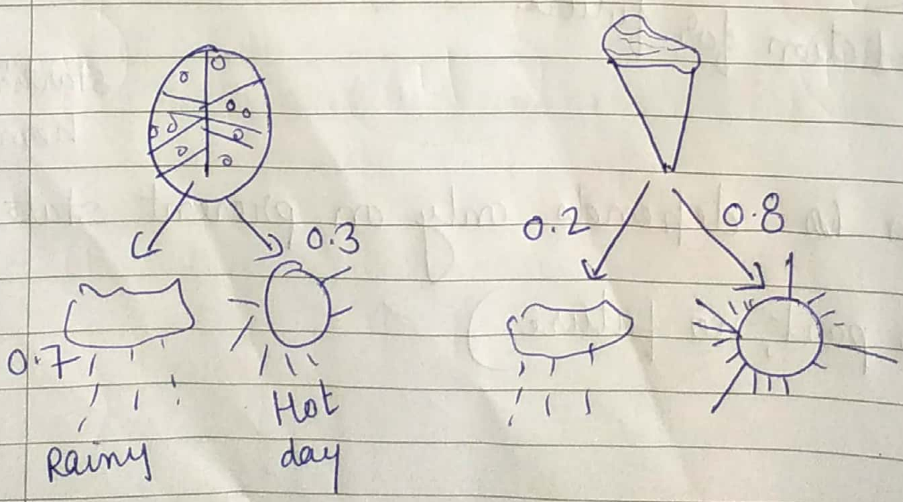
KD loves food, if he gets his choice of food he brings DK for shopping

and DK knows what KD wants earlier it was not the scenario but now she knows it
How?

DK knows KD is foodie person so she has started observing KD's food choice
when its hot day KD asks for icecreame

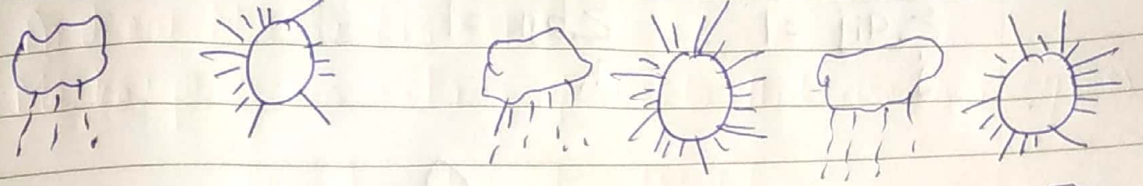


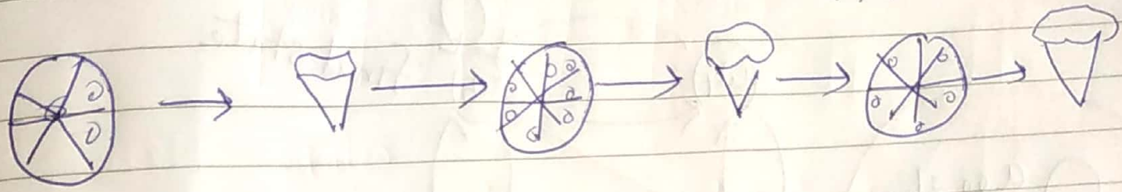
when its rainy day KD asks for pizza



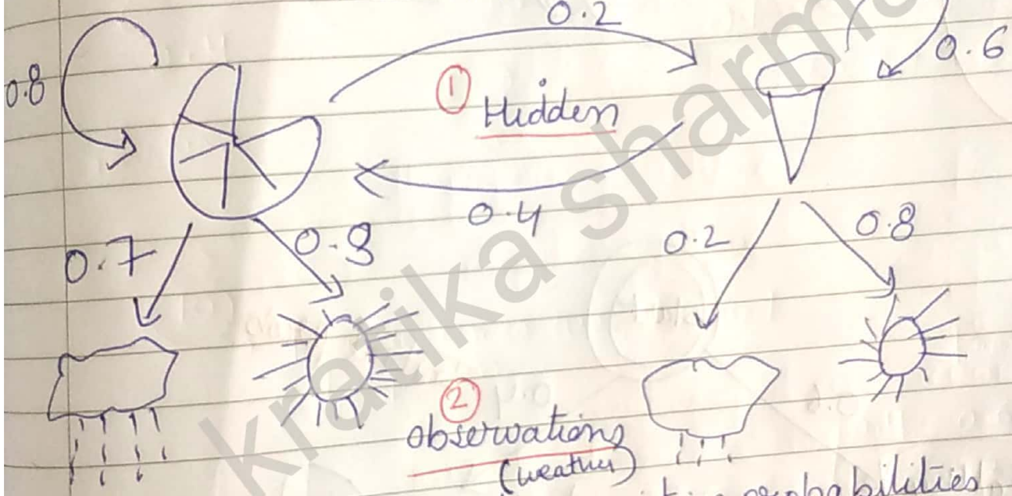
DATE

Mon	Tues	Wed	Thur	Fri	Sat.





Hidden Markov Model

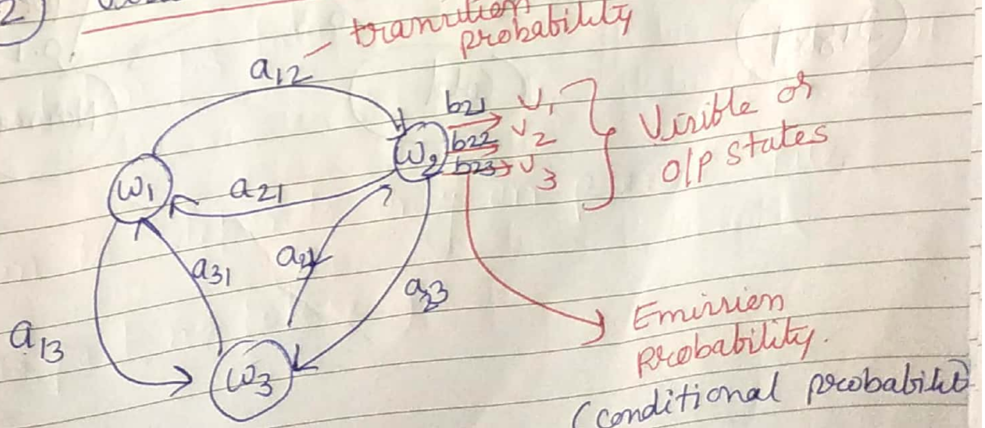


there no. are c/d transition probabilities

→ HMM behaves like FSM. consists of two states

① Hidden (W) states (food)

② Visible state / output / or observed states (V)



DATE

When to terminate ?

$\sum_j a_{ij} = 1$ $\sum_K b_{jK} = 1$