31-01-'19

UNIT-III : QUANTIFYING UNCERTAINITY, PROBABILISTIC REASONING

→ Quantifying Uncertainities

- Partial Absorbability.
- Uncertainity can be well described by Probability Theory.
- We consider utility theory to check if the case that we are considering is right or not.
- Decision Theory is the combination of Probability Theory and Utility Theory
- Structure of Agent (which uses decision theory)

  function DT-AGENT(percept) returns an action
  Persistent : belief state, probabilistic beliefs about current state of the world
  action, the agent's action
  Update belief-state based on action and percept,
  calculate outcome possibilities for actions given action description and current belief-state.
  Select action with highest expected utility given probabilities of outcomes and utility information.
  Return action
- Qualification Problem: Problem with various constraints.
- Probability can be defined as:

  Ω = Sample Space = Set of all possible states

  ω = Possible World Elements = Possible/Considered state

$0 \leq P(\omega) \leq 1$ ↳ for every ω and $\sum_{\omega \in \Omega} P(\omega) = 1$	For ϕ, $P(\phi) = \sum_{\omega \in \phi} P(\omega)$

Eg. $P(\text{Total}=11) = (P(5,6) + P(6,5)) = \frac{1}{36} + \frac{1}{36} = \frac{1}{18}$

↳ Unconditional / Prior Probability

Scanned with CamScanner

Eg. $P(cavity) = 0.2$

$P(cavity \mid toothache) = 0.6$

$P(cavity \mid toothache \land \neg cavity) = 0$

↳ Conditional / Posterior Probability

$$\boxed{\begin{array}{l} P(a/b) = \dfrac{P(a \land b)}{P(b)} \\ P(b) > 0 \\ \text{Product Rule: } P(a \land b) = P(a|b)\,P(b) \end{array}}$$

- We can't represent many attributes in propositional logic, so we make use of probability in AI.
- Unconditional Probability (one value is known / half of the probability is known in advance).
- Conditional Probability (no value is known in advance).

01-02-19

→ Language Prepositions in Probability Assertions

- The possible states that we can have are called as prepositions.
- We have random variables, domains, boolean random variables.
- Random Variables: 1st letter is upper case
  Eg. Die1
- Boolean Random Variables: Lower case letters
  true / false
- Domains: For variable weather, Domain = {Sunny, Cloudy, Snowy}

Scanned with CamScanner

- Variables may have infinite values. They may have integers or real values.
- Eg. The probability that a patient has a cavity given that she is a teenager with no toothache is 0.1. This can be represented as:

→ $P(\text{cavity} / \sim \text{toothache} \wedge \text{teen}) = 0.1$

- We use AND and OR to join prepositions.
- Eg. $P(\text{weather} = \text{sunny}) = 0.6$

  $P(\text{weather} = \text{rainy}) = 0.1$

  $P(\text{weather} = \text{cloudy}) = 0.29$

  $P(\text{weather} = \text{snow}) = 0.01$

→ $P(\text{weather}) = \langle 0.6, 0.1, 0.29, 0.01 \rangle$

- For Random Variables.

  $P(X|Y) = P(X = x_i | Y = y_i)$

- For Continuous Variables:
  - Their values are specified by a range.
  - We don't have a single exact value.
  - $P(\text{Noon Temp} = x) = \text{Uniform}_{[18C, 24C]}(x)$
  - $$\boxed{\begin{array}{l} P(X = x) \Rightarrow P(x) \\ P(x) = \lim_{dx \to 0} P(x \le X \le x + dx) / dx \end{array}}$$
  - Eg. $P(\text{Noon Temp} = x) = \text{Uniform}_{[18C, 24C]}(x)$

$$= \begin{cases} \frac{1}{8C} & \text{if } 18C \le x \le 26C \\ \text{otherwise} \end{cases}$$

Scanned with CamScanner

07-02-2

- Multiple Variables
  - Join probability distribution.
  - Weather and Cavity

$P(Weather, Cavity) = P(Weather | Cavity) P(Cavity)$

↳ Product Rule

- $P(W = Sunny \wedge C = True) = P(W = Sunny | C = True) P(C = True)$

⋮

$P(W = Sunny \wedge C = False) = P(W = Sunny | C = False) P(C = False)$

↳ Product Rule for Individual Variables (Full Join Distribution)

- Probability Axioms and Relationships -

- $P(\sim a) = \sum_{\omega \in \sim a} P(\omega)$

$$= \sum_{\omega \in \sim a} P(\omega) + \sum_{\omega \in a} P(\omega) - \sum_{\omega \in a} P(\omega)$$

$$= \sum_{\omega \in \Omega} P(\omega) - \sum_{\omega \in a} P(\omega)$$

$$= 1 - P(a)$$

- It gives the relationship between the positives and the negatives.
- $\boxed{P(\sim a) = 1 - P(a)}$

- Inclusion Exclusion Principle

$$\boxed{P(a \vee b) = P(a) + P(b) - P(a \wedge b)}$$

$$\boxed{\begin{array}{l} 0 \leq P(\omega) \leq 1 \\ \sum_{\omega \in \Omega} P(\omega) = 1 \end{array}} \rightarrow \text{Kolmogorov's Axioms}$$

Scanned with CamScanner

classmate

Date

Page

07-02-'19

Joint Probability

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

→ Axioms:

- $P(A) \geq 0$
- $P(S) = 1$
- $P(A') = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B)$ (mutually exclusive)
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (not mutually exclusive)

↳ Addition Rule Of Probability

Eg. of Joint Probability:

Head in 1st coin $= P(A) = 0.5$

Head in 2nd coin $= P(A) = 0.5$

Both Coins Head $= P(A \cap B) = P(A) * P(B) = 0.5 * 0.5 = 0.25$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.5 - 0.25 = 0.75$

↳ Getting head on any one coin, on tossing two coins

Conditional Probability

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

$$= P(H) * PE) / P(E)$$

$P(H/E) = P(H)$

$P(E/H) = P(E)$

Here: E is the evidence and H is the hypothesis

Scanned with CamScanner

Eg. $P(cavity / toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$

$= 0.6$

$P(\sim cavity / toothache) = \frac{P(\sim cavity \wedge toothache)}{P(toothache)}$

$= 0.4$

→ **General Inference Procedure**

$P(X / e) = \alpha P(X, e)$

$= \alpha \sum_y P(X, e, y)$

→ **Independence**

$P(a/b) = P(a)$

$P(b/a) = P(b)$

$P(a \wedge b) = P(a) * P(b)$

→ **Bayes' Rule**

$P(b/a) = \frac{P(a/b) P(b)}{P(a)}$

→

Scanned with CamScanner

Eg. A doctor advices to have a stiff neck, say 70% of time. The doctor also knows unconditional facts. Prior probability that a patient has 1/50000 meningitis. Prior probability that any patient has a stiff neck is 1%. Calculate P(M|S).

→ $P(M) = \frac{1}{50000}$ $P(S) = \frac{1}{100}$ $P(S|M) = \frac{70}{100}$

$$P(M|S) = \frac{P(S|M)\,P(M)}{P(S)} = \frac{\frac{70}{\cancel{100}} \times \frac{1}{50000}}{\frac{1}{\cancel{100}}} = \frac{7\cancel{0}}{5000\cancel{0}}$$

$$= \frac{7}{5000} = \frac{7}{5} \times 10^{-3} = 1.4 \times 10^{-3} = 0.0014$$

$\therefore P(M|S) = 0.0014$

Eg. Suppose we are given the probability of Mike having a cold as 0.25, The probability of Mike was observed sneezing when he had cold in the past as 0.9 and the probability of Mike was observed sneezing when he did not have cold as 0.20. Find the probability of Mike having a cold given that he sneezes. P(C|S).

→ $P(C) = 0.25 = \frac{25}{100}$ $P(S|C) = 0.9 = \frac{9}{100}$ $P(S|NC) = 0.20 = \frac{20}{100}$

P(Mike was observed sneezing / Mike has cold) = P(S/C) = 0.9

P(Mike was observed sneezing / Mike doesn't have cold) = P(S/NC) = 0.20

$$P(C/S) = \frac{P(S/C) * P(C)}{P(S)}$$

$$P(C/S) = \frac{P(S/C) \cdot P(C)}{P(S)} \quad -(1)$$

$$P(C'/S) = \frac{P(S/C') \cdot P(C')}{P(S)} \quad -(2)$$

(1) + (2) = (1)

$$\Rightarrow \frac{P(S/C) \cdot P(C)}{P(S)} + \frac{P(S/C') \cdot P(C')}{P(S)} = 1$$

$$\Rightarrow P(S/C) \cdot P(C) + P(S/C')\, P(C') = 1 \times P(S)$$

$$\Rightarrow (0.9) \cdot (0.25) + (0.2)(0.75) = P(S)$$

$$\Rightarrow P(S) = 0.225 + 0.15 = 0.375$$

$$P(C/S) = \frac{P(S/C) \cdot P(C)}{P(S)} = \frac{(0.9) \times (0.25)}{0.375}$$

$$= 0.6$$

Here, P(S) = Sneezing with cold + Sneezing without cold

$\therefore P(C/S) = 0.6$

→ **Extension Of Baye's Theorem**

→ One Hypothesis Two Evidences

$$P(H/E_1 \text{ and } E_2) = \frac{P(E_1/H) * P(E_2/H) * P(H)}{P(E_1 \text{ and } E_2)}$$

$$P(H/E_1 \text{ and } E_2) = \frac{P(H) * P(E_1/H) * P(E_2/H \text{ and } E_1)}{P(E_1) * P(E_2/E_1)}$$

$$P(H/E_1 \text{ and } E_2) = \frac{P(E_2/H \text{ and } E_1) * P(H/E_1)}{P(E_2/E_1)}$$

Scanned with CamScanner

$$\boxed{P(H \mid E_1 \text{ and } E_2) = \frac{P(E_2 \mid H \text{ and } E_1) * P(H \mid E_1)}{P(E_2 \mid E_1)}}$$

$$\boxed{P(E_1, \cdots, E_n) = \frac{P(E_1 \mid H) * \cdots * P(E_n \mid H) * P(H)}{P(E_1, \cdots, E_n)}}$$

↳ One Hypothesis, Multiple Evidences

08-02-19

→ Chain Evidence

$$P(H/E_1) = \frac{P(E_1/H) * P(E_2/E_1) * P(H)}{P(E_1/E_2) * P(E_2)}$$

→ Multiple Hypothesis and Single Evidence

$$P(H_i/E) = \frac{P(E/H_i) * P(H_i)}{\sum_{j=1}^{k} P(E/H_j) * P(H_j)}$$

→ Multiple Hypothesis Multiple Evidences

$$P(H_i/E_i \cdots E_n) = \frac{P(E_1 \cdots E_n/H_i) * P(H_i)}{\sum_{j=1}^{k} P(E_1 \cdots E_n/H_j) * P(H_j)}$$

P.T.O. →

classmate

Date
Page

Bayesian Belief Network

- Represented as: $X \rightarrow Y$
  X is the parent of Y, Y is the child of X
- In Join Probability, we consider all the ~~be~~ entries whether it is needed or not needed, but in Bayesian Bellief, we consider only the beliefs that are needed.
- It is represented as:

$$P(X_1 ------ X_n) = P(X_n \mid X_1 ----- X_{n-1}) * P(X_{n-1} \mid X_1 ----- X_{n-2}) * \cdots\cdots * P(X_2 \mid X_1) * P(X_1)$$

Eg. $P(A,B,C,D) = P(D \mid A,B,C) * P(C \mid A,B) * P(B \mid A) * P(A)$

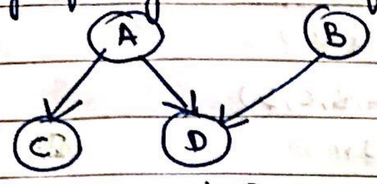
↳ Join Probability for Bayesian Belief Network using four variables.

It can be represented as:

$$\boxed{P(X_1 ----- X_n) = \prod_{i=1}^{n} P(X_i \mid Parent\_node(X_i))}$$

- Node that does not have a parent is called as local probability distribution and it is unconditional (since they don't depend on any other node).
- Nodes with no children are called as Hypothesis nodes.
- Nodes with no parents are independent nodes.
- Nodes with children are dependent nodes.
- A node that has been observed is called evidence node.

– Eg. of Bayesian Belief Network



Here, A and B are unconditional since they have no parent and C and D are conditional since they have a parent.

P(A)	P(B)
0.3	0.6

A	P(C)
T	0.4
F	0.2

A	B	P(D)
T	T	0.7
T	F	0.4
F	T	0.2
F	F	0.01

can be represented as

$P(A)=0.3$, $P(B)=0.6$, $P(C|A)=0.4$, $P(C|\sim A)=0.2$

$P(D|A,B)=0.7$, $P(D|A,\sim B)=0.4$, $P(D|\sim A,B)=0.2$, $P(D|\sim A,\sim B)=0.01$

$P(A,B,C,D)=?$

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

Here: $P(D|A,B,C) \Rightarrow P(D|A,B)$

$P(C|A,B) \Rightarrow P(C|A)$

$P(B|A) \Rightarrow P(B)$

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

$= 0.7 * 0.4 * 0.6 * 0.3$

$= 0.0504$

→ <u>Inference Using Bayesian Network</u>

* What is likelyhood or probability that hypothesis is 'A' given 'C'?

→ $P(A/C)$ must ~~to~~ be calculated

Scanned with CamScanner

Here: $P(A|C) = \dfrac{P(A,C)}{P(C)}$

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

[i.e. $P(A,C)$, take $\sim$ for B, D]

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

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

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

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

$= 0.4 * 0.4 * 0.4 * 0.\cancel{8}3 = \cancel{0.0192}\ \cancel{0.0256}$ 0.0192

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

$= (1-0.7) * 0.4 * 0.6 * 0.3 = \cancel{0.0252}\ 0.0216$

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

$= (1-0.4) * 0.4 * 0.4 * 0.3 = 0.0288$

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

$= (1-0.7) * 0.4 * 0.6 * 0.3 = 0.0216$

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

$= 0.4 * 0.4 * 0.4 * 0.3 = 0.0192$

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

$= 0.2 * 0.2 * 0.6 * 0.7 = 0.0168$

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

$= 0.01 * 0.2 * 0.4 * 0.7 = 0.00056$

classmate
Date
Page

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

$= (1-0.2) * 0.2 * 0.6 * 0.7 = 0.0672$

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

$= (1-0.01) * 0.2 * 0.4 * 0.7 = 0.0554$

$P(A, C) = 0.0504 + 0.0192 + 0.0216 + 0.0288 = 0.1200$

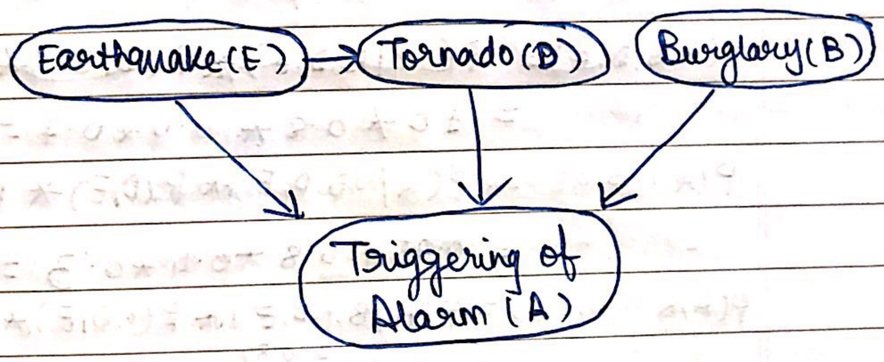
$P(C) = 0.0504 + 0.0216 + 0.0192 + 0.0288 + 0.0168 +$

$0.00056 + 0.0672 + 0.0554 = 0.2600$

$\therefore P(A/C) = \dfrac{P(A,C)}{P(C)} = \dfrac{0.12}{0.26} = \dfrac{12}{26} = \dfrac{6}{13} = 0.4615$

Ans: $P(A|C) = 0.4615$

- Eg.



P(E)	P(B)
0.4	0.7

E	P(D)	
T	0.8	D/E
F	0.5	D/~E

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

Scanned with CamScanner

What is the probability that it is an earthquake, given the alarm is ringing? $P(E|A) = ?$

$\rightarrow$ $P(E|A) = \dfrac{P(E, A)}{P(A)}$

$$P(E,A) = \sum_{\{D,B\} \in \{T,F\}} P(A, B, D, E)$$

$$= P(A,B,D,E) + P(A,\sim B, D, E) + P(A, B, \sim D, E) + P(A, \sim B, \sim D, E)$$

$$P(A) = \sum_{B,D,E \in \{T,F\}} P(A, B, D, E)$$

$$= P(A,B,D,E) + P(A,\sim B,D,E) + P(A,B,\sim D,E) + P(A,B,D,\sim E) + P(A,\sim B,\sim D,E) + P(A,\sim B,D,\sim E) + P(A,B,\sim D,\sim E) + P(A,\sim B,\sim D,\sim E)$$

$$P(A,B,D,E) = P(A|B,D,E) * P(D|E) * P(E) * P(B)$$
$$= 1.0 * 0.8 * 0.4 * 0.7 = 0.2240.$$

$$P(A,\sim B,D,E) = P(A|\sim B,D,E) * P(D|E) * P(E) * P(\sim B)$$
$$= 0.95 * 0.8 * 0.4 * 0.3 = 0.0192.$$

$$P(A,B,\sim D,E) = P(A|B,\sim D,E) * P(\sim D|E) * P(E) * P(B)$$
$$= 0.9 * \overset{(1-0.8)}{\cancel{0.5}} * 0.4 * 0.7 = 0.0504.$$

$$P(A,B,D,\sim E) = P(A|B,D,\sim E) * P(D|\sim E) * P(\sim E) * P(B)$$
$$= 0.89 * 0.5 * 0.6 * 0.7 = 0.1869.$$

$$P(A,\sim B,\sim D,E) = P(A|\sim B,\sim D,E) * P(\sim D|E) * P(E) * P(\sim B)$$
$$= 0.85 * (1-0.8) * 0.4 * 0.3 = 0.0240.$$

$$P(A,\sim B,D,\sim E) = P(A|\sim B,D,\sim E) * P(D|\sim E) * P(\sim E) * P(\sim B)$$
$$= 0.87 * 0.5 * 0.6 * 0.3 = 0.0783.$$

$$P(A,B,\sim D,\sim E) = P(A|B,\sim D,\sim E) * P(\sim D|\sim E) * P(\sim E) * P(B)$$
$$= 0.7 * (1-0.5) * 0.6 * 0.7 = 0.1470.$$

$$P(A,\sim B,\sim D,\sim E) = P(A|\sim B,\sim D,\sim E) * P(\sim D|\sim E) * P(\sim E) * P(\sim B)$$
$$= 0.3 * (1-0.5) * 0.6 * 0.3 = 0.0270.$$

Scanned with CamScanner

classmate
Date
Page

$P(E,A) = 0.2240 + 0.0192 + 0.0504 + 0.0240 = 0.3176$

$P(A) = 0.2240 + 0.0192 + 0.0504 + 0.1869 + 0.0240 + 0.0783$
$+ 0.1470 + 0.0270 = 0.7568$

$P(E|A) = \frac{P(E,A)}{P(A)} = \frac{0.3176}{0.7568} = 0.4197$

Ans:- Probability that it is an earthquake given the alarm is ringing $= P(E|A) = 0.4197$

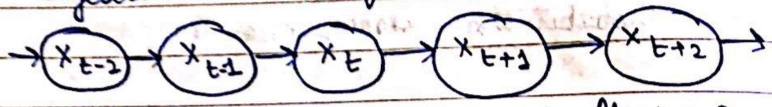
21-02-'19

Probabilistic Reasoning Over Time

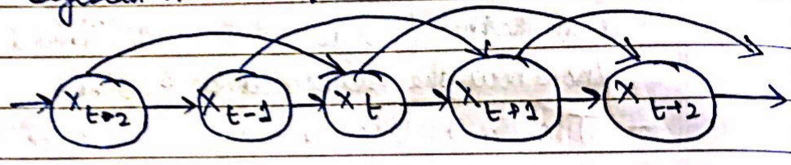
- Agent should have past, present and future beliefs.
- With the belief state and transition model, agent will know how the states will be evolved.
- Time and Uncertainity.
- States and Observations
- X_t : Unobserved State Variables
  E_t : Observable Evidence Variables
  $X_{a:b}$
  $U_{1:3} = U_1, U_2, U_3$
  ↳ We are and underground security guard and we want to know if it is raining outside, if the person who comes carries an umbrella (comes for inspection), then it rains outside.
- Transition Model and Sensor Model
- $P(X_t | X_{0:t-1})$
- A transition model is how an agent is moving from one state to another state, in the state space.

Scanned with CamScanner

- First Order Markov Process: Next state depens only on the previous state, not the earlier states before it. $P(X_t \mid X_{t-1})$
- Second Order Markov Process: Next state depends only on the previous two states in transition model not the earlier states before it. $P(X_t \mid X_{t-2}, X_{t-1})$
- Transition Model consists all the states, so we reduce the states using the Markov Process.
- Bayesian Network for first order Markov Process



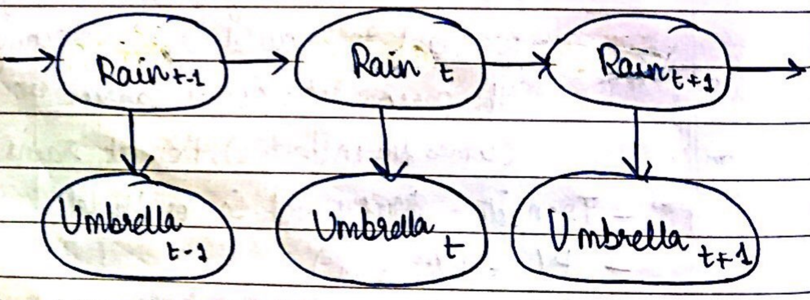
- Bayesian Network for second order Markov Process



- Transition and Sensor Model

R_{t-1}	$P(R_t)$
t	0.7
f	0.3

R_t	$P(U_t)$
t	0.9
f	0.1



Transition Model $P(Rain_t \mid Rain_{t-1})$

Sensor or Model $P(Umbrella_t \mid Rain_t)$

- Complete Join Distribution of variables

$$P(X_{0:t}, E_{1:t}) = \underbrace{P(X_0)}_{\text{initial State}} \prod_{i=1}^{t} \left[\underbrace{P(X_i \mid X_{i-1})}_{\text{transition model}} \; \underbrace{P(E_i \mid X_i)}_{\text{Sensor model}} \right]$$

- Using first order Markov Process, we ~~can~~ cannot assess the exact probability. The probability is approximate.
- So, it accuracy can be increased by increasing the order of the Markov process, ~~ie,~~ or increasing the no. of states. (like season, temperature, etc.).

- Inference of Temporal Models
  - (i) Filtering $P(X_t \mid e_{1:t})$ - From states that rational agent has, it takes a state
  - (ii) Prediction $P(X_{t+k} \mid e_{1:t})_{k>0}$ - Predicts the state that it has to have.
  - (iii) Smoothing $P(X_k \mid e_{1:t})$ $0<k<t$
  - (iv) Most likely $P(X_{1:t} \mid e_{1:t})$
    Explanation Learning