Wolkite University

College of Computing and Informatics

Department of Computer Science

Chapter Five :Uncertain Knowledge and Reasoning

Prepared by Adem (MSc.)

August 10, 2022

Outline

Logic and Uncertainty

- Uncertainty is Bad for Agents based on Logic
- Example: Catching a Flight
- Let action At = leave for airport t minutes before flight
- Question: Will At get me there on time?
- Problems:
  - Partial observability (road state, other drivers' plans, etc.)
  - Noisy sensors (traffic reports)
  - Uncertainty in action outcomes (flat tire, etc.)
  - Complexity of modeling and predicting traffic
- A purely logical approach leads to conclusions that are too weak for effective decision making:
  - A_{25} will get me there on time if there is no accident on the bridge and it doesn't rain and my tires remain intact, etc., etc.
  - A_{Inf} guarantees to get there in time, but who lives forever?

Uncertainty

- Observed variables or evidence:
  - agent knows certain things about the state of the world (e.g., sensor readings).
- Unobserved variables:
  - agent needs to reason about other aspects that are uncertain (e.g., where the ghost is).
- (Probabilistic) model:
  - agent knows or believes something about how the known variables relate to the unknown variables.

Reasoning under Uncertainty

- A rational agent is one that makes rational decisions (in order to maximize its performance measure)
- A rational decision depends on:
  - the relative importance of various goals
  - the likelihood they will be achieved
  - the degree to which they will be achieved

Sources of Uncertainty

- Probabilistic assertions summarize effects of
- Laziness
  - facts, observability, etc. Ignorance
- Ignorance
  - lack of explicit theories, relevant facts, observability, etc.
- Randomness
  - Inherently random behavior
- Utility theory is used to represent and infer preferences.
- Decision theory = probability theory + utility theory

Probability Theory

- the set of all possible worlds is called the sample space.
- The possible worlds are mutually exclusive and exhaustive
  - two possible worlds cannot both be the case, and
  - one possible world must be the case.
- The Greek letter Ω (uppercase omega) is used to refer to the sample space, and
- ω (lowercase omega) refers to elements of the space, that is, particular possible worlds.
- A fully specified probability model associates a numerical probability with each possible world.
- The basic axioms of probability theory say that every possible world has a
  - probability between 0 and 1 and
  - that the total probability of the set of possible worlds is 1:

Cont....

$$0 \leq P(\omega) \leq 1 \text{ for every } \omega \text{ and } \sum_{\omega \in \Omega} P(\omega) = 1.$$

- unconditional or prior refer to degrees of belief in propositions in the absence of any other information.
  - Probabilities such as and a P(doubles) P(total =11)
- conditional or posterior probability is a measure of the probability of an event given
  - that (by assumption, presumption, assertion or evidence) another event has already occurred.
  - This probability is written where P(doubles |Dies =5) where the | is pronounced "given".

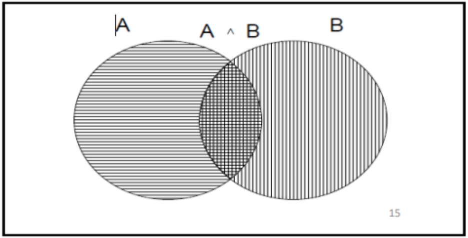
Cont...

- Frequentism (Empirical)
  - Probabilities are relative frequencies determined by observation.
  - For example, if we toss a coin many times, p(heads) is the proportion of the time the coin will come up heads
  - But what if we are dealing with events that only happen once? E.g., what is the probability that a Republican will win the presidency in 2024?
  - Reference class problem. E.g., how do we define comparable elections?
- Subjectivism (Bayesian Statistics)
  - Probabilities are degrees of belief updated by evidence.
  - How do we assign belief values to statements without evidence?
  - How do we update our degrees of belief?
  - What would make sure that agents hold consistent beliefs? E.g.,
  - The coin will land heads up and tails up at the same time.

Axioms of Probability

- Probability Theory is governed by the following axioms:
  - All probabilities are real values between 0 and 1: for all φ, $0 \leq P(\varphi) \leq 1$
  - Valid propositions have probability 1 $P(True) = P(\alpha \vee \neg \alpha) = 1$
  - The probability of disjunction is defined as follows:

$$P(\alpha \vee \beta) = P(\alpha) + P(\beta) - P(\alpha \wedge \beta)$$



Cont...

- A random variable is a function that maps from the domain of possible worlds Ω (called sample space) to the real numbers written as X: $\Omega \rightarrow$
- Denoted by capital letters
  - R: Is it raining?
  - W: What's the weather?
  - Die: What is the outcome of rolling two dice?
  - V: What is the speed of a car (in MPH)?
- Names for values are always lowercase
  - A Boolean random variable has the range true, false .
  - the range of weather to be the set sun, cloud, rain,snow
- Domain values must be mutually exclusive and exhaustive
  - R $\in$ True, False
  - W $\in$ Sunny, Cloudy, Rainy, Snow
  - Die $\in$(1,1), (1,2), ... (6,6)

Notations

- Random variables are written in upper roman letters: X, Y etc.
- Realizations of a random variable are written in corresponding lower case letters. E.g.x1,x2,..xn could be of outcomes of the random variable X.
- The probability value of the realization X is written as P(X = X) .
- When clear from context,this will be abbreviated as P(X) .
- The probability distribution of the (discrete) random variable X is denoted as p(X).
- This corresponds e.g.to a vector of numbers, one for each of the probability valuesP(X =Xi) (and not to a single scalar value!).

Events and Propositions

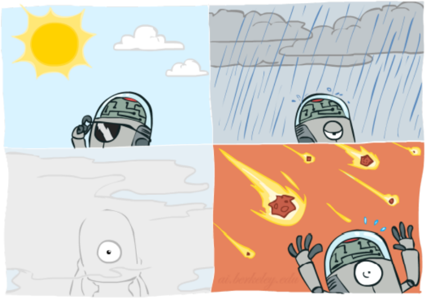
- Events
  - Probabilistic statements are defined over events, world states or sets of states
    - "It is raining"
    - "The weather is either cloudy or snowy"
    - "The sum of the two dice rolls is 11"
- Events are described using propositions:
  - R = True
  - W = "Cloudy" $\vee$ W = "Snowy'
  - D $\in$ (5,6),(6,5)

Probability distributions

- For discrete variables,
  - the probability distribution can be encoded by a discrete list of the probabilities of the outcomes, known as the probability mass function.

$\mathbf{P}(W)$

W	P
sun	0.6
rain	0.1
fog	0.3
meteor	0.0



Joint distributions

- A joint probability distribution over a set of random variables specifies the probability of each outcome

$$P(X_1 = x_1, \ldots, X_n = x_n) = \sum_{\{\omega: X_1(\omega)=x_1, \ldots, X_n(\omega)=x_n\}} P(\omega)$$

$$\mathbf{P}(T, W)$$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

Cont..

- From a joint distribution,the probability of any event can be calculated.
  - Probability that it is hot and sunny?
  - Probability that it is hot?
  - Probability that it is hot or sunny?
- Interesting events often correspond to partial assignments, e.g.P(hot).

Marginal distributions

- The marginal distribution of a subset of a collection of random variables is the joint probability distribution of the variables contained in the subset.

$\mathbf{P}(T, W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$\mathbf{P}(T)$

T	P
hot	0.5
cold	0.5

$$P(t) = \sum_w P(t, w)$$

$\mathbf{P}(W)$

W	P
sun	0.6
rain	0.4

$$P(w) = \sum_t P(t, w)$$

- Intuitively, marginal distributions are sub-tables which eliminate variables.

Cont...

P(Cavity, Toothache)	
Cavity = false ∧*Toothache = false*	0.8
Cavity = false ∧ *Toothache = true*	0.1
Cavity = true ∧ *Toothache = false*	0.05
Cavity = true ∧ *Toothache = true*	0.05

Marginal Prob. Distr.

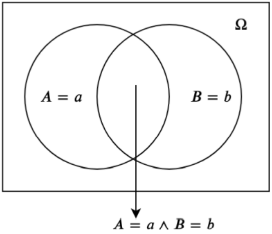
P(Cavity)	
Cavity = false	0.8+0.1 = 0.9
Cavity = true	0.05+0.05=0.1

P(Toothache)	
Toothache = false	0.8+0.0.5= 0.85
Toothache = true	0.1+0.05= 0.15

Conditional distributions

- P(A) is the unconditional (or prior) probability of fact A
- An agent can use the unconditional probability of A to reason about A in the absence of further information
- If further evidence B becomes available, the agent must use the conditional (or posterior) probability:

$$P(a|b) = \frac{P(a,b)}{P(b)}.$$



Cont...

Joint Prob. Distr.

P(Cavity, Toothache)	
$Cavity = false \land Toothache = false$	0.8
$Cavity = false \land Toothache = true$	0.1
$Cavity = true \land Toothache = false$	0.05
$Cavity = true \land Toothache = true$	0.05

$$P(a|b) = \frac{P(a \land b)}{P(b)}$$

Marginal Prob. Distr.

P(Cavity)	
$Cavity = false$	0.9
$Cavity = true$	0.1

P(Toothache)	
$Toothache = false$	0.85
$Toothache = true$	0.15

What is $P(Cavity = true \mid Toothache = false)$?

0.05 / 0.85 = 0.059

What is $P(Cavity = false \mid Toothache = true)$?

0.1 / 0.15 = 0.667

Conditional distributions

P(Cavity, Toothache)	
Cavity = false ∧ Toothache = false	0.8
Cavity = false ∧ Toothache = true	0.1
Cavity = true ∧ Toothache = false	0.05
Cavity = true ∧ Toothache = true	0.05

$$P(a|b) = \frac{P(a \wedge b)}{P(b)}$$

A conditional distribution is a distribution over the values of one variable given fixed values of other variables

P(Cavity | Toothache = true)	
Cavity = false	0.667
Cavity = true	0.333

P(Cavity | Toothache = false)	
Cavity = false	0.941
Cavity = true	0.059

P(Toothache | Cavity = true)	
Toothache= false	0.5
Toothache = true	0.5

P(Toothache | Cavity = false)	
Toothache= false	0.889
Toothache = true	0.111

Normalization trick

- To get the whole conditional distribution P(X| y) at once, select all entries in the joint distribution matching Y = y and renormalize them to sum to one

P(Cavity, Toothache)	
Cavity = false ∧ *Toothache = false*	0.8
Cavity = false ∧ *Toothache = true*	0.1
Cavity = true ∧ *Toothache = false*	0.05
Cavity = true ∧ *Toothache = true*	0.05

Select P(X,y)

Toothache, Cavity = false	
Toothache= false	0.8
Toothache = true	0.1

Sum is $P(y) = 0.9$

Renormalize sum to 1 (= divide by $P(y)$)

P(Toothache | Cavity = false)	
Toothache= false	0.889
Toothache = true	0.111

Equivalent to

$$\boldsymbol{P}(X \mid y) = \alpha\, \boldsymbol{P}(X, y)$$

with $\alpha = 1/P(y)$

Bayes' Rule

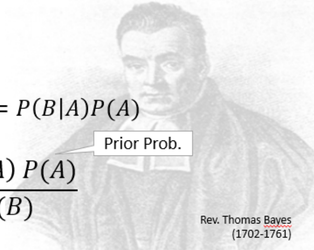
- The product rule gives us two ways to factor a joint distribution:

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

Posterior Prob.

Prior Prob.

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

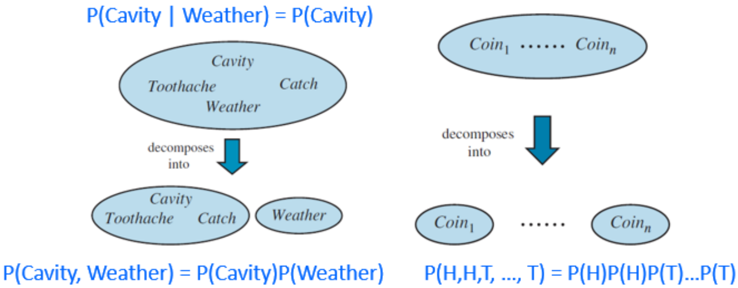


Independence

- Two events a and b are independent if and only if
  - $P(a \wedge b) = P(a)\ P(b)$
  - This is equivalent to $P(a \mid b) = P(a)$ and $P(b \mid a) = P(b)$
  - $P(a| b) = P(a)$
  - $P(b|a) = P(b)$
  - $P(a, b) = P(a)P(b)$
- Independence is denoted as $A \perp B$.

Cont...

- Independence is an important simplifying assumption for modeling,
- e.g., Cavity and Weather can be assumed to be independent



Cont...

- Conditional independence: a and b are conditionally independent given c (i.e., if we know c) iff
- $P(a \wedge b | c) = P(a | c) P(b | c)$

$$P(a|b,c) = P(a|c), \text{or}$$

$$P(b|a,c) = P(b|c), \text{or}$$

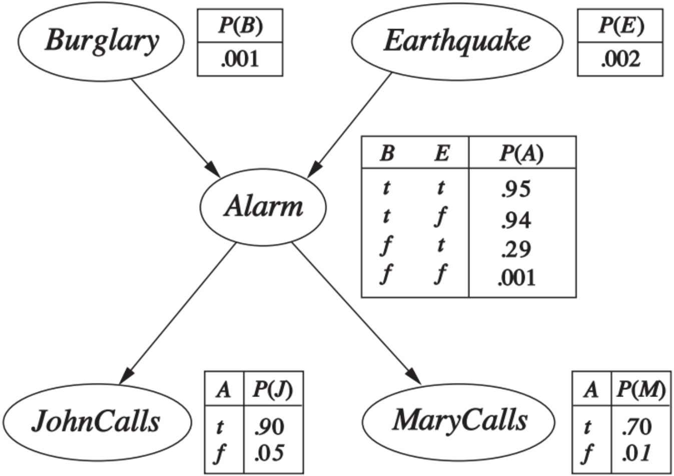
$$P(a,b|c) = P(a|c)P(b|c)$$

- Conditional independence is denoted as $A \perp B | C$.

Bayesian network

- A Bayesian network is a directed acyclic graph (DAG) in which:
  - Each node corresponds to a random variable.
    - Can be observed or unobserved.
    - Can be discrete or continuous.
  - Each edge indicates dependency relationships.
  - Ifthere is an arrowfrom node to node , is said to be a parent of .
  - Each node is annotated with a conditional probability distribution

Cont...



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

Example

- I am at work, neighbor John calls to say my alarm is ringing, but neighbor Mary does not call. Sometimes it's set off by minor earthquakes. Is there a burglar?
- Variables: Burglar Earthquake Alarm JohnCalls MaryCalls
- Network topology from "causal" knowledge:
  - A burglar can setthe alarm off
  - An earthquake can setthe alaram off
  - The alarm can causeMary to call
  - The alarm can cause John to call

Summary

- Uncertainty arises because of laziness and ignorance.
- It is inescapable in complex non-deterministic or partially observable environments.
- Probabilistic reasoning provides a framework for managing our knowledge and beliefs, with the Bayes' rule acting as the workhorse for inference.
- A Bayesian Network specify full joint distribution.
- They are often exponentially smaller than an explicitly enumerated joint distribution.