Probabilistic reasoning

Prof. Deepak C Vegda
email.:deepakvegda.it@ddu.ac.in

What is it?

It is a way of knowledge representation where we apply the concept of probability to indicate the uncertainty in knowledge.

Why do we Need it?

- When there are unpredictable outcomes.
- When specifications or possibilities of predicates becomes too large to handle.
- When an unknown error occurs during an experiment.

Uncertainty analysis

- uncertainty factor
- decision making with uncertain information
- estimation of uncertainty in complex models of risk.
- structural uncertainty and model specification.
- monitoring methods to reduce uncertainty.

Probability

- *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. The value of probability always remains between 0 and 1 that represent ideal uncertainties.*

$0 \le P(A) \le 1$, 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.

Conti…

- We can find the probability of an uncertain event by using the below formula.

- $P(\neg A)$ = probability of a not happening event.
- $P(\neg A) + P(A) = 1$.

Some words in probability

- **Event:** Each possible outcome of a variable is called an event.
- **Sample space:** The collection of all possible events is called sample space.
- **Random variables:** Random variables are used to represent the events and objects in the real world.
- **Prior probability:** The prior probability of an event is probability computed before observing new information.
- **Posterior Probability:** The probability that is calculated after all evidence or information has taken into account. It is a combination of prior probability and new information.

Basic of Probability

- sample space set
  X = {x1, x2, …, xn}
  - collection of all possible events
  - can be discrete or continuous
- probability number P(xi): likelihood of an event xi to occur
  - non-negative value in [0,1]
  - total probability of the sample space is 1
  - for mutually exclusive events, the probability for at least one of them is the sum of their individual probabilities
  - *experimental probability*
    - based on the frequency of events
  - *subjective probability*
    - based on expert assessment

Conti..

- describes *independent* events
  - do not affect each other in any way
- *joint* probability of two independent events A and B

  $P(A \cap B) = P(A) * P(B)$
- *union* probability of two independent events A and B

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

  $= P(A) + P(B) - P(A) * P(B)$

Conditional probability:

- Conditional probability is a probability of occurring an event when another event has already happened.

Conti..

- 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:
  - **Where $P(A \wedge B)$= Joint probability of a and 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:

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

Example

- In a class, there are 70% of the students who like English and 40% of the students who likes English and mathematics, and then what is the percent of students those who like English also like mathematics?

Solution:

Let, A is an event that a student likes Mathematics

B is an event that a student likes English.

$$P(A|B) = \frac{P(A \wedge B)}{P(B)} = \frac{0.4}{0.7} = 57\%$$

Hence, 57% are the students who like English also like Mathematics.

- **Random variables**
  - Domain
- **Atomic event**: complete specification of state
- **Prior probability**: degree of belief without any other evidence
- **Joint probability**: matrix of combined probabilities of a set of variables

- Alarm, Burglary, Earthquake
  - Boolean (like these), discrete, continuous
- Alarm=True ^ Burglary=True ^ Earthquake=False alarm ^ burglary ^ earthquake
- P(Burglary) = .1
- P(Alarm, Burglary) =

	alarm	¬alarm
burglary	.09	.01
¬burglary	.1	.8

- **Conditional probability**: probability of effect given causes
- **Computing conditional probs**:
  - P(a | b) = P(a ^ b) / P(b)
  - P(b): **normalizing** constant
- **Product rule**:
  - P(a ^ b) = P(a | b) P(b)
- **Marginalizing**:
  - $P(B) = \Sigma_a P(B, a)$
  - $P(B) = \Sigma_a P(B | a) P(a)$ (**conditioning**)

- P(burglary | alarm) = .47
  P(alarm | burglary) = .9
- P(burglary | alarm) =
  P(burglary ^ alarm) / P(alarm)
  = .09 / .19 = .47
- P(burglary ^ alarm) =
  P(burglary | alarm) P(alarm) =
  .47 * .19 = .09
- P(alarm) =
  P(alarm ^ burglary) +
  P(alarm ^ ¬burglary) =
  .09+.1 = .19

Bayes' theorem

- Bayes' theorem is also known as **Bayes' rule, Bayes' law**, or **Bayesian reasoning**, which determines the probability of an event with uncertain knowledge.
- In probability theory, it relates the conditional probability and marginal probabilities of two random events.
- Bayes' theorem was named after the British mathematician **Thomas Bayes**. The **Bayesian inference** is an application of Bayes' theorem, which is fundamental to Bayesian statistics.
- It is a way to calculate the value of P(B|A) with the knowledge of P(A|B).
- Bayes' theorem allows updating the probability prediction of an event by observing new information of the real world.

Example

If cancer corresponds to one's age then by using Bayes' theorem, we can determine the probability of cancer more accurately with the help of age.

Bayes' theorem can be derived using product rule and conditional probability of event A with known event B:

- As from product rule we can write:
  - $P(A \wedge B) = P(A|B) P(B)$ or

Similarly, the probability of event B with known event A:
  - $P(A \wedge B) = P(B|A) P(A)$

Conti..

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \quad \text{....(a)}$$

The above equation (a) is called as **Bayes' rule** or **Bayes' theorem**. This equation is basic of most modern AI systems for **probabilistic inference**.

- P(A|B) is known as **posterior**, which we need to calculate, and it will be read as Probability of hypothesis A when we have occurred an evidence B.
- P(B|A) is called the likelihood, in which we consider that hypothesis is true, then we calculate the probability of evidence.
- P(A) is called the **prior probability**, probability of hypothesis before considering the evidence
- P(B) is called **marginal probability**, pure probability of an evidence.

Example

Question: what is the probability that a patient has diseases meningitis with a stiff neck?

- **Given Data:**

  A doctor is aware that disease meningitis causes a patient to have a stiff neck, and it occurs 80% of the time. He is also aware of some more facts, which are given as follows:
  - The Known probability that a patient has meningitis disease is 1/30,000.
  - The Known probability that a patient has a stiff neck is 2%.

Example

Question: From a standard deck of playing cards, a single card is drawn. The probability that the card is king is 4/52, then calculate posterior probability P(King|Face), which means the drawn face card is a king card.

$$P(king|face) = \frac{P(Face|king) * P(King)}{P(Face)} \quad ....(i)$$

P(king): probability that the card is King= 4/52= 1/13

P(face): probability that a card is a face card= 3/13

P(Face|King): probability of face card when we assume it is a king = 1

Independent

Two random variables A and B are independent if

P(A^B) = P(A) P(B) hence if P(A|B) = P(A)

 Two random variables A and B are independent given C, if

P(A^B|C) = P(A|C) P(B|C) hence if P(A|B,C) = P(A|C)

Bayesian Belief Network

- is key computer technology for dealing with probabilistic events and to solve a problem which has uncertainty. We can define a Bayesian network as:

  "A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph.“

- It is also called a **Bayes network, belief network, decision network**, or **Bayesian model**.

Conti..

- Bayesian Network can be used for building models from data and experts opinions, and it consists of two parts:
  - **Directed Acyclic Graph**
  - **Table of conditional probabilities.**
- The generalized form of Bayesian network that represents and solve decision problems under uncertain knowledge is known as an **Influence diagram**.

Joint probability distribution

- If we have variables x1, x2, x3,....., xn, then the probabilities of a different combination of x1, x2, x3.. xn, are known as Joint probability distribution.
- **$P[x_1, x_2, x_3, \ldots\ldots, x_n]$**, it can be written as the following way in terms of the joint probability distribution.

$= P[x_1| x_2, x_3, \ldots\ldots, x_n]P[x_2, x_3, \ldots\ldots, x_n]$

$= P[x_1| x_2, x_3, \ldots\ldots, x_n]P[x_2|x_3, \ldots\ldots, x_n] \ldots. P[x_{n-1}|x_n]P[x_n]$.

Example

P[D, S, A, B, E]= P[D | S, A, B, E]. P[S, A, B, E]

=P[D | S, A, B, E]. P[S | A, B, E]. P[A, B, E]

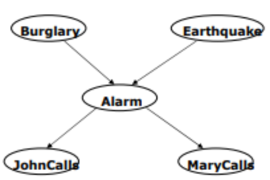
= P [D| A]. P [S| A, B, E]. P[A, B, E]

= P[D | A]. P[S | A]. P[A| B, E]. P[B, E]

= P[D | A]. P[S | A]. P[A| B, E]. P[B | E]. P[E]

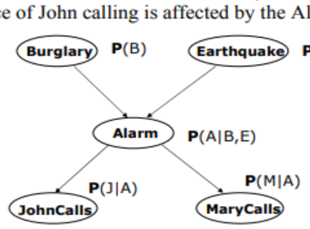
Belief Network

- Assume your house has an alarm system against burglary. You live in the seismically active area and the alarm system can get occasionally set off by an earthquake. You have two neighbors, Mary and John, who do not know each other. If they hear the alarm they call you, but this is not guaranteed.
- We want ity distribution of events Alarm, Mary calls and

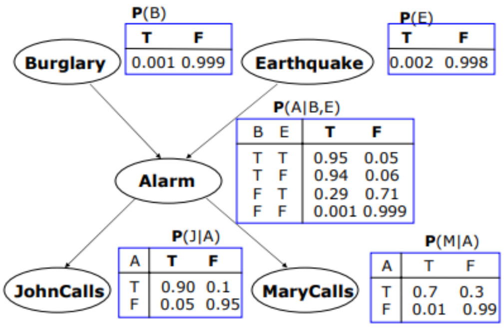


BBN

1. **Directed acyclic graph**
   - **Nodes** = random variables
     Burglary, Earthquake, Alarm, Mary calls and John calls
   - **Links** = direct (causal) dependencies between variables.
     The chance of Alarm is influenced by Earthquake, The chance of John calling is affected by the Alarm



Bayesian belief network.

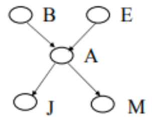


Full joint distribution in BBNs

Full joint distribution is defined in terms of local conditional distributions (obtained via the chain rule):

$$\mathbf{P}(X_1, X_2, .., X_n) = \prod_{i=1,..n} \mathbf{P}(X_i \mid pa(X_i))$$

Example:



Assume the following assignment of values to random variables

$B = T, E = T, A = T, J = T, M = F$

Then its probability is:

$P(B = T, E = T, A = T, J = T, M = F) =$

$P(B = T)P(E = T)P(A = T \mid B = T, E = T)P(J = T \mid A = T)P(M = F \mid A = T)$

P(J, M, A, ¬B, ¬E) = ?

P(J, M, A, ¬B, ¬E)
= P (J|A) *P (M|A)*P (A|¬B ^ ¬E) *P (¬B) *P (¬E).