Statistical Reasoning

Course Code: **CSC4226** Course Title: **Artificial Intelligence and Expert System**

Dept. of Computer Science
Faculty of Science and Technology

Lecture No:	**Ten (10)**	**Week No:**	**Eleven(11)**	**Semester:**	
Lecturer:	*Dr. Hossain Md Shakhawat*				*Shakhawat@aiub.edu*

Lecture Outline

1. Quantifying Uncertainty
2. Basic Probability Notation
3. Bayes’ Rules and Its Use
4. Clinical Trials
5. The Semantics of Bayesian Network
6. Prediction in Bayesian Network

Quantifying Uncertainty

In the logic based approaches described, we have assumed that everything is either believed false or believed true.

However, it is often useful to represent the fact that we believe such that something is probably true, or true with probability (say) 0.65.

This is useful for dealing with problems where there is randomness and unpredictability (such as in games of chance) and also for dealing with problems where we could, if we had sufficient information, work out exactly what is true.

To do all this in a principled way requires techniques for probabilistic reasoning. In this section, the Bayesian Probability Theory is first described and then discussed how uncertainties are treated.

Basic Probability Notations

- **Probabilities :**

  Usually, are descriptions of the likelihood of some event occurring (ranging from **0** to **1**).

- **Event :**

  One or more outcomes of a probability experiment .

- **Probability Experiment :**

  Process which leads to well-defined results call outcomes.

- **Sample Space :**

  Set of all possible outcomes of a probability experiment.

Basic Probability Notations

- **Independent Events :**

  Two events, E_1 and E_2, are independent if the fact that E_1 occurs does not affect the probability of E_2 occurring.

- **Mutually Exclusive Events :**

  Events E_1, E_2, ..., E_n are said to be mutually exclusive if the occurrence of any one of them automatically implies the non-occurrence of the remaining $n - 1$ events.

- **Disjoint Events :**

  Another name for mutually exclusive events.

Basic Probability Notations

- **Classical Probability :**

  Also called a priori theory of probability.

  The probability of event **A** = no of possible outcomes **f** divided by the total no of possible outcomes **n** ; ie., **P(A) = f / n.**

  Assumption: All possible outcomes are equal likely.

- **Empirical Probability :**

  Empirical probability, also known as experimental **probability**, refers to **probability** that is based on historical data. In other words, **empirical probability** illustrates the likelihood of an event occurring based on historical data.

- **Conditional Probability :**

  The probability of some event **A**, given the occurrence of some other event **B**. Conditional probability is written **P(A|B)**, and read as "the probability of **A**, given **B** ".

Examples

Examples

- **Example 1**

  Sample Space - Rolling two dice

  The sums can be **{ 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 }**.

  Note that each of these are not equally likely. The only way to get a sum **2** is to roll a **1** on both dice, but can get a sum **4** by rolling out comes as **(1,3), (2,2), or (3,1)**.

  Table below illustrates a sample space for the sum obtain.

First dice	Second Dice					
	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Classical & Empirical Probability

Classical Probability

Table below illustrates frequency and distribution for the above sums.

Sum	2	3	4	5	6	7	8	9	10	11	12
Frequency	1	2	3	4	5	6	5	4	3	2	1
Relative frequency	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	1/36	1/36

The classical probability is the relative frequency of each event.

Classical probability **P(E) = n(E) / n(S); P(6) = 5 / 36, P(8) = 5 / 36**

Empirical Probability

The empirical probability of an event is the relative frequency of a frequency distribution based upon observation **P(E) = f / n**

Disjoint Events

Mutually Exclusive Events (disjoint) : means nothing in common

Two events are mutually exclusive if they cannot occur at the same time.

(a) If two events are mutually exclusive,

then probability of both occurring at same time is **P(A and B) = 0**

(b) If two events are mutually exclusive ,

then the probability of either occurring is **P(A or B) = P(A) + P(B)**

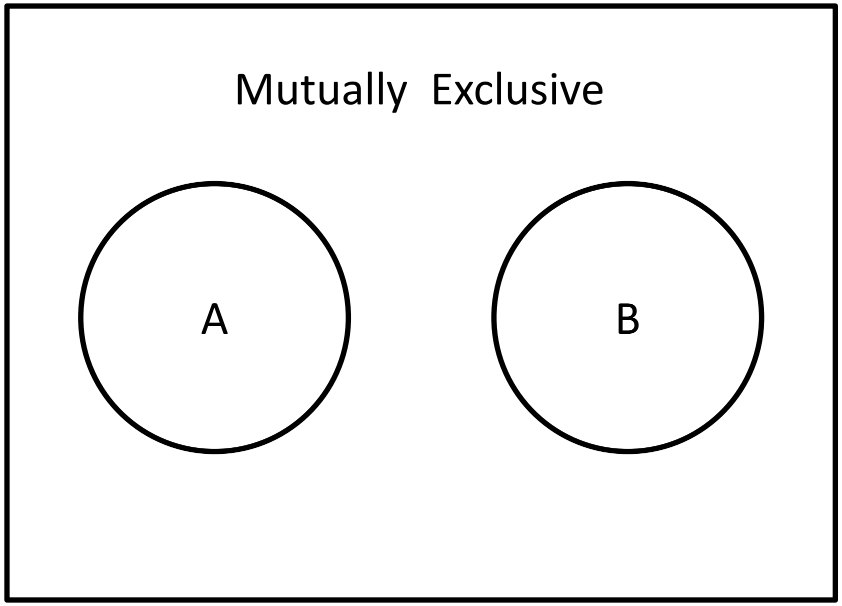
Non-Mutually Exclusive Events

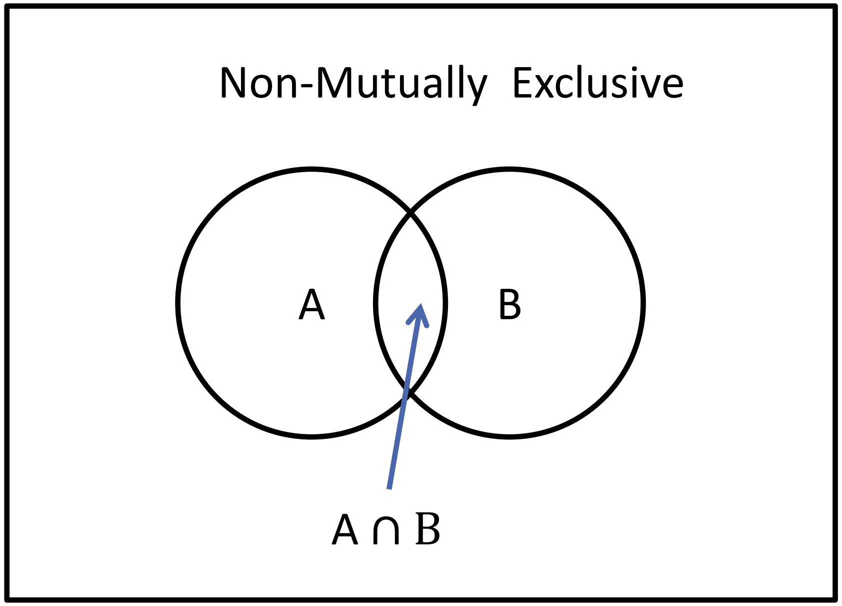
The non-mutually exclusive events have some overlap.

When **P(A)** and **P(B)** are added, the probability of the intersection (ie. "and") is added twice, so subtract once.

P(A or B) = P(A) + P(B) - P(A and B)

Mutually Exclusive
vs
Non-Mutually Exclusive Events





Probability Rules

- All probabilities are between **0** and **1** inclusive **0 <= P(E) <= 1**.
- The sum of all the probabilities in the sample space is **1**.
- The probability of an event which must occur is **1**.
- The probability of the sample space is **1**.
- The probability of any event which is not in the sample space is zero.
- The probability of an event not occurring is **P(E') = 1 - P(E)**

Practice 1

Example 1 : A single 6-sided die is rolled.

What is the probability of each outcome?

What is the probability of rolling an even number?

What is the probability of rolling an odd number?

The possible outcomes of this experiment are 1, 2, 3, 4, 5, 6.

The Probabilities are :

P(1)	=	No of ways to roll 1 / total no of sides	=	1/6
P(2)	=	No of ways to roll 2 / total no of sides	=	1/6
P(3)	=	No of ways to roll 3 / total no of sides	=	1/6
P(4)	=	No of ways to roll 4 / total no of sides	=	1/6
P(5)	=	No of ways to roll 5 / total no of sides	=	1/6
P(6)	=	No of ways to roll 6 / total no of sides	=	1/6
P(even)	=	ways to roll even no / total no of sides	=	3/6 = 1/2
P(odd)	=	ways to roll odd no / total no of sides	=	3/6 = 1/2

Practice 2

Example 2 : Roll two dices

Each dice shows one of 6 possible numbers;

Total unique rolls is 6 x 6 = 36;

List of the joint possibilities for the two dices are:

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

Roll two dices;

The rolls that add up to 4 are ((1,3), (2,2), (3,1)).

The probability of rolling dices such that total of 4 is 3/36 = 1/12

and the chance of it being true is (1/12) x 100 = 8.3%.

Conditional Probability

Conditional probability P(A|B)

A conditional probability is the probability of an event given that another event has occurred.

Example : Roll two dices.

What is the probability that the total of two dice will be greater than 8 given that the first die is a 6 ?

First List of the **joint possibilities** for the two dices are:

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	**(6, 3)**	**(6, 4)**	**(6, 5)**	**(6, 6)**

There are 6 outcomes for which the first die is a 6, and of these, there are 4 outcomes that total more than 8 are (6,3; 6,4; 6,5; 6,6).

The probability of a total > 8 given that first die is 6 is therefore 4/6 = 2/3 .

This probability is written as: $\underbrace{P(\text{total}>8}_{\textbf{event}} \mid \underbrace{\text{1st die} = 6)}_{\textbf{condition}} = 2/3$

Read as "The probability that the total is > 8 given that die one is 6 is 2/3."

Written as **P(A|B)** , is the probability of event A given that the event B has occurred.

Joint Probability

- **Probability of A and B is P(A and B)**

  The probability that events **A** and **B** both occur.

  Note : Two events are independent if the occurrence of one is unrelated to the probability of the occurrence of the other.

  - **If A and B are independent**

    then probability that events **A** and **B** both occur is:

    P(A and B) = P(A) x P(B)

    ie product of probability of **A** and probability of **B**.

  - **If A and B are not independent**

    then probability that events **A** and **B** both occur is:

    P(A and B) = P(A) x P(B|A) where

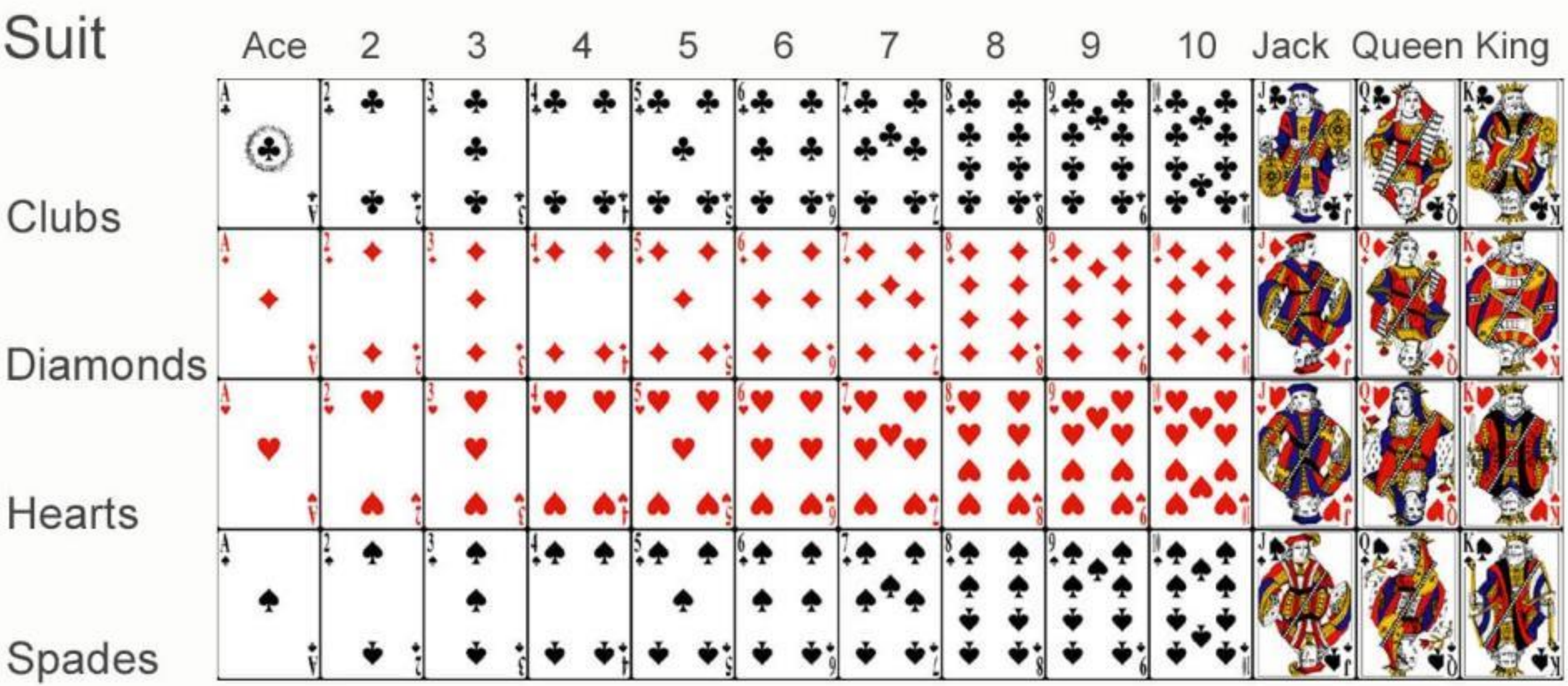
    P(B|A) is **conditional probability** of **B** given **A**

Joint Probability : Independent Events

Example 1: P(A and B) if events A and B are independent

- Draw a card from a deck , then replace it, draw another card.
- Find probability that 1st card is Ace of clubs (event A) and 2nd card is any Club (event B).
- Since there is only one Ace of Clubs, therefore probability P(A) = 1/52.
- Since there are 13 Clubs, the probability P(B) = 13/52 = 1/4.
- Therefore, **P(A and B) =** p(A) x p(B) = 1/52 x 1/4 = 1/208.

Joint Probability : Independent Events



Joint Probability : Dependent Events

Example 2: P(A and B) if events A and B are not independent

- Draw a card from a deck, not replacing it, draw another card.
- Find probability that both cards are Aces ie the 1st card is Ace (event A) and the 2nd card is also Ace (event B).
- Since 4 of 52 cards are Aces, therefore probability P(A) = 4/52 = 1/13.
- Of the 51 remaining cards, 3 are aces. so, probability of 2nd card is also Ace (event B) is P(B|A) = 3/51 = 1/17.
- Therefore, **P(A and B) =** p(A) x p(B|A) = 1/13 x 1/17 = 1/221

OR Operation

- **Probability of A or B is P(A or B)**

  The probability of either event **A** or event **B** occur.

  Two events are mutually exclusive if they cannot occur at same time.

  ‡ **If A and B are mutually exclusive**

  then probability that events **A** or **B** occur is:

  P(A or B) = P(A) + P(B)

  ie sum of probability of **A** and probability of **B**

  ‡ **If A and B are not mutually exclusive**

  then probability that events **A** and **B** both occur is:

  P(A or B) = P(A) + P(B) – P(A and B) where

  P(A and B) is probability that events **A** and **B** both occur while events **A** and **B** are independent

OR Operation: Mutually Exclusive

Example 1: P(A or B) if events A or B are mutually exclusive

- Rolling a die.
- Find probability of getting either, event A as 1 or event B as 6?
- Since it is impossible to get both, the event A as 1 and event B as 6 in same roll, these two events are mutually exclusive.
- The probability $P(A) = P(1) = 1/6$ and $P(B) = P(6) = 1/6$
- Hence probability of either event A or event B is :
- **P(A or B)** $= p(A) + p(B) = 1/6 + 1/6 = 1/3$

OR Operation: Not Mutually Exclusive

Example 2: P(A or B) if events A or B are not mutually exclusive

- Find probability that a card from a deck will be either an Ace or a Spade?
- probability P(A) is P(Ace) = 4/52 and P(B) is P(spade) = 13/52.
- Only way in a single draw to be Ace and Spade is Ace of Spade; which is only one, so probability P(A and B) is P(Ace and Spade) = 1/52.
- Therefore, the probability of event A or B is :

P(A or B) = P(A) + P(B) – P(A and B)

= P(ace) + P(spade) - P(Ace and Spade)

= 4/52 + 13/52 - 1/52 = 16/52 = 4/13

Bayes' Rule

Product rule can be written in two forms

$$P(a \wedge b) = P(a \mid b)P(b) \quad \text{and} \quad P(a \wedge b) = P(b \mid a)P(a) .$$

Equating the two right-hand sides and dividing by $P(a)$, we get

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

This equation is known as **Bayes' rule** (also Bayes' law or Bayes' theorem).

This simple equation underlies most modern AI systems for probabilistic inference.

General Case of Bayes' Rule

The more general case of Bayes' rule for multivalued variables can be written in the **P** notation as follows:

$$\mathbf{P}(Y \mid X) = \frac{\mathbf{P}(X \mid Y)\mathbf{P}(Y)}{\mathbf{P}(X)} ,$$

As before, this is to be taken as representing a set of equations, each dealing with specific values of the variables. We will also have occasion to use a more general version conditionalized on some background evidence **e**:

$$\mathbf{P}(Y \mid X, \mathbf{e}) = \frac{\mathbf{P}(X \mid Y, \mathbf{e})\mathbf{P}(Y \mid \mathbf{e})}{\mathbf{P}(X \mid \mathbf{e})} .$$

Bayes' Theorem: Revisited

Let **S** be a sample space.

Let **A_1, A_2, ... , A_n** be a set of mutually exclusive events from **S**.

Let **B** be any event from the same **S**, such that **$P(B) > 0$**.

Then Bayes' Theorem describes following two probabilities :

$$P(A_k|B) = \frac{P(A_k \cap B)}{P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B)}$$ and

by invoking the fact $P(A_k \cap B) = P(A_k).P(B|A_k)$ the probability

$$P(A_k|B) = \frac{P(A_k).P(B|A_k)}{P(A_1).P(B|A_1) + P(A_2).P(B|A_2) + \cdots + P(A_n).P(B|A_n)}$$

Applying Bayes' rule: The simple case

$$P(cause \mid effect) = \frac{P(effect \mid cause)P(cause)}{P(effect)} \, .$$

The conditional probability **P(*effect* | *cause*)** quantifies the relationship in the **causal** direction,
whereas **P(cause | *effect*)** describes the **diagnostic** direction

In a task such as medical diagnosis, we often have conditional probabilities on causal relationships (that is, the doctor knows P(symptoms | disease)) and want to derive a diagnosis, P(disease | symptoms).

Applying Bayes’ rule: A Simple Example

A doctor knows that
the disease **meningitis causes the patient to have a stiff neck, say, 70% of the time**.
The doctor also knows some unconditional facts: the prior probability
that **a patient has meningitis is 1/50,000**,
and the **prior probability that any patient has a stiff neck is 1%.**

Letting
s be the proposition that the **patient has a stiff neck** and
m be the proposition that the **patient has meningitis**, we have

$$
\begin{aligned}
P(s \mid m) &= 0.7 \\
P(m) &= 1/50000 \\
P(s) &= 0.01 \\
P(m \mid s) &= \frac{P(s \mid m)P(m)}{P(s)} = \frac{0.7 \times 1/50000}{0.01} = 0.0014\,.
\end{aligned}
$$

Another Example

Problem : Marie's marriage is tomorrow.

- in recent years, each year it has rained only 5 days.
- the weatherman has predicted rain for tomorrow.
- when it actually rains, the weatherman correctly forecasts rain 90% of the time.
- when it doesn't rain, the weatherman incorrectly forecasts rain 10% of the time.

The question : What is the probability that it will rain on the day of Marie's wedding?

Solution

Solution : The sample space is defined by two mutually exclusive events – "it rains" or "it does not rain". Additionally, a third event occurs when the "weatherman predicts rain".

The events and probabilities are stated below.

- Event A1 : rains on Marie's wedding.
- Event A2 : does not rain on Marie's wedding
- Event B : weatherman predicts rain.
- $P(A1)$ = 5/365 =0.0136985 [Rains 5 days in a year.]
- $P(A2)$ = 360/365 = 0.9863014 [Does not rain 360 days in a year.]
- $P(B|A1)$ = 0.9 [When it rains, the weatherman predicts rain 90% time.]
- $P(B|A2)$ = 0.1 [When it does not rain, weatherman predicts rain 10% time.]

Calculation

We want to know **P(A1|B)**, the probability that it will rain on the day of Marie's wedding, given a forecast for rain by the weatherman.

The answer can be determined from Bayes' theorem, shown below.

$$P(A1|B) = \frac{P(A1).P(B|A1)}{P(A1).P(B|A1)+P(A2).P(B|A2)} = \frac{(0.014)(0.9)}{[(0.014)(0.9)+(0.986)(0.1)]}$$

$$= 0.111$$

So, despite the weatherman's prediction, there is a good chance that Marie will not get rain on at her wedding.

Thus Bayes theorem is used to calculate conditional probabilities.

Example: Clinical Trial

In a clinic, the probability of the patients having HIV virus is **0.15**.

A blood test done on patients :

If patient has virus, then the test is **+ve** with probability **0.95**.

If the patient does not have the virus, then the test is **+ve** with probability **0.02**.

Assign labels to events : **H** = patient has virus; **P** = test +ve

Given : $P(H) = 0.15$; $P(P|H) = 0.95$; $P(P|\neg H) = 0.02$

Find :

If the test is **+ve** what are the probabilities that the patient

i) has the virus ie $P(H|P)$; ii) does not have virus ie $P(\neg H|P)$;

If the test is **-ve** what are the probabilities that the patient

iii) has the virus ie $P(H|\neg P)$; iv) does not have virus ie $P(\neg H|\neg P)$;

Calculations

i) For $P(H|P)$ we can write down Bayes Theorem as

$P(H|P) = [\ P(P|H)\ P(H)\ ] / P(P)$

We know $P(P|H)$ and $P(H)$ but not $P(P)$ which is probability of a +ve result.

There are two cases, that a patient could have a +ve result, stated below :

1. Patient has virus and gets a +ve result : $H \cap P$
2. Patient does not have virus and gets a +ve result: $\neg H \cap P$

Find probabilities for the above two cases and then add

ie $P(P) = P(H \cap P) + P(\neg H \cap P)$.

But from the second axiom of probability we have :

$P(H \cap P) = P(P|H)\ P(H)$ and $P(\neg H \cap P) = P(P|\neg H)\ P(\neg H)$.

Therefore putting these we get :

$$P(P) = P(P|H)\ P(H) + P(P|\neg H)\ P(\neg H) = 0.95 \times 0.15 + 0.02 \times 0.85 = 0.1595$$

Now substitute this into Bayes Theorem and obtain $P(H|P)$

$$P(H|P) = \frac{P(P|H)\ P(H)}{P(P|H)\ P(H) + P(P|\neg H)\ P(\neg H)} = 0.95 \times 0.15 / 0.1595 = 0.8934$$

Calculations

ii) Next is to work out $P(\neg H|P)$

$$P(\neg H|P) = 1 - P(H|P) = 1 - 0.8934 = 0.1066$$

iii) Next is to work out $P(H|\neg P)$; again we write down Bayes Theorem

$$P(H|\neg P) = \frac{P(\neg P|H)\ P(H)}{P(\neg P)}$$

here we need $P(\neg P)$ which is $1 - P(P)$

$$= (0.05 \times 0.15)/(1\text{-}0.1595) = 0.008923$$

iv) Finally, work out $P(\neg H|\neg P)$

It is just $1 - P(H|\neg P) = 1 - 0.008923 = 0.99107$

Bayesian Networks

A Bayesian network (or a belief network) is a probabilistic graphical model that represents a set of variables and their probabilistic independencies. For example, a Bayesian network could represent the **probabilistic relationships** between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases.

Bayesian Networks are also called : Bayes nets, Bayesian Belief Networks (BBNs) or simply Belief Networks. Causal Probabilistic Networks (CPNs).

Components of Bayesian Network

A Bayesian network consists of :

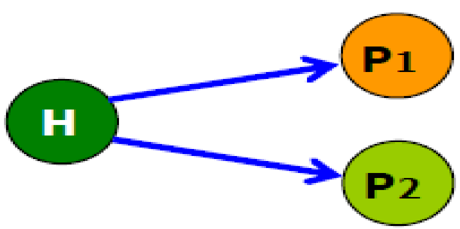
- a set of nodes and a set of directed edges between nodes.
- the edges reflect cause-effect relations within the domain.
- The effects are not completely deterministic (e.g. disease -> symptom).
- the strength of an effect is modeled as a probability.

Bayesian Network: Clinical Trial Example

Imagine, the patient is given a second test independently of the first; means the second test is done at a later date by a different person using different equipment. So, the error on the first test does not affect the probability of an error on the second test.

In other words the two tests are independent. This is depicted using the diagram below :

A simple example of a Bayesian Network.



Event **H** is the cause of the two events **P1** and **P2**.

The arrows represent the fact that **H** is driving **P1** and **P2**.

The network contained 3 nodes.

If both **P1** and **P2** are **+ve**

then find the probability that patient has the virus ?

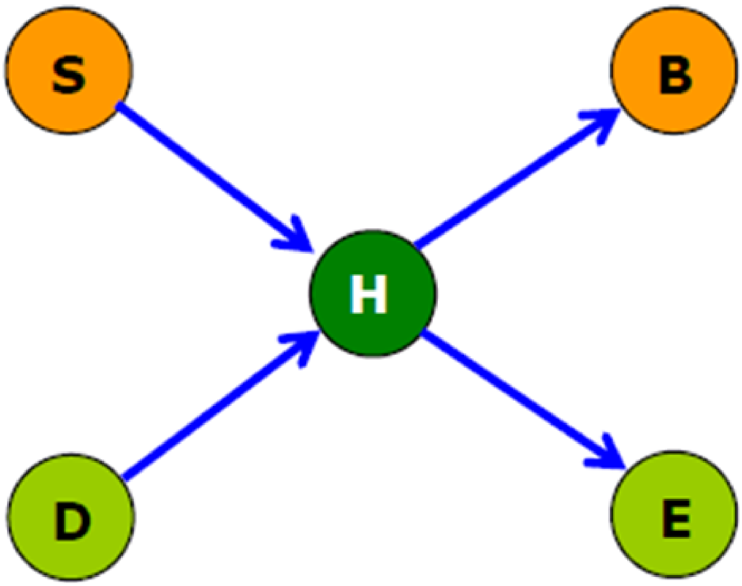
In other words asked to find $P(H|P1 \cap P2)$.

Complicated Bayesian Network

Given the following facts about heart disease.

- Either smoking or bad diet or both can make heart disease more likely.
- Heart disease can produce either or both of the following two symptoms:
  - high blood pressure
  - an abnormal electrocardiogram
- Here smoking and bad diet are regarded as causes of heart disease. The heart disease in turn is a cause of high blood pressure and an abnormal electrocardiogram.

Network of Heart Disease



The symbols define :

S = smoking,

D = bad diet,

H = heart disease,

B = high blood pressure,

E = abnormal electrocardiogram

Here **H** has two causes **S** and **D**.

Find probability of **H,** given each of the four possible combinations of **S** and **D**.

Practice Question

Find probability of **H**, given each of the four possible combinations of **S** and **D**.

A medical survey gives us the following data :

$P(S)$	$= 0.3$	$P(D)$	$= 0.4$
$P(H \mid S \cap D)$	$= 0.8$		
$P(H \mid \neg S \cap D)$	$= 0.5$		
$P(H \mid S \cap \neg D)$	$= 0.4$		
$P(H \mid \neg S \cap \neg D)$	$= 0.1$		
$P(B \mid H)$	$= 0.7$	$P(B \mid \neg H)$	$= 0.1$
$P(E \mid H)$	$= 0.8$	$P(E \mid \neg H)$	$= 0.1$

Given these information, an answer to the question concerning this network :

what is the probability of heart disease ?

References

1. Chapter 13: Quantifying Uncertainty , Pages 480-530
"Artificial Intelligence: A Modern Approach," by Stuart J. Russell and Peter Norvig

Books

1. “Artificial Intelligence: A Modern Approach,” by Stuart J. Russell and Peter Norvig.
2. "Artificial Intelligence: Structures and Strategies for Complex Problem Solving", by George F. Luger, (2002)
3. "Artificial Intelligence: Theory and Practice", by Thomas Dean.
4. "AI: A New Synthesis", by Nils J. Nilsson.
5. “Programming for machine learning,” by J. Ross Quinlan,
6. “Neural Computing Theory and Practice,” by Philip D. Wasserman, .
7. “Neural Network Design,” by Martin T. Hagan, Howard B. Demuth, Mark H. Beale, .
8. “Practical Genetic Algorithms,” by Randy L. Haupt and Sue Ellen Haupt.
9. “Genetic Algorithms in Search, optimization and Machine learning,” by David E. Goldberg.
10. "Computational Intelligence: A Logical Approach", by David Poole, Alan Mackworth, and Randy Goebel.
11. “Introduction to Turbo Prolog”, by Carl Townsend.