UNIT 7 PROBABILISTIC REASONING

Structure **Page Nos.**

7.0 INTRODUCTION

This unit is dedicated to probability theory and its usage in decision making for various problems. Contrary to the classical decision making of True and False propositions, the probability of the truth value with a certain probability is used for making decisions. The inclusion of such a probabilistic approach is quite relevant since uncertainties are quite obvious in the real world.

As we know, the probability of an event (uncertain event I) is basically the measure of the degree of likelihood of the occurrence of event I. Let the set of all such possible events is represented as sample space S The measure of probability is a function *P ()* mapping the event outcome E_i from sample space S to some real number and satisfying few conditions such as:

(i) $0 \leq P(I) \leq 1$ for any event $I \subseteq S$

(ii) $P(S) = 1$, represents a certain outcome, and

(iii) For $E_i \cap E_j = \phi$, for all $i \neq j$ (the E_i are mutually exclusive), i.e. $P(E_1 \cup E_2 \ldots) = P(E_1) + P(E_2) + \ldots$

Using the above mentioned three conditions, we can derive the basic laws of probability. It is also to be noted that only these three conditions are not enough to compute the probability of an outcome. This additionally requires the collection of experimental data for estimating the underlying distribution.

7.1 OBJECTIVES

After going through this unit, you should be able to:

- Understand the role of probabilistic reasoning in AI
- Understand the Concept of Bayesian theory and Bayesian networks
- Perform probabilistic inference through Bayesian Networks
- Understand the other Paradigm of Uncertain Reasoning & Dempster Scheffer Theory

7.2 REASONING WITH UNCERTAIN INFORMATION

Reasoning is an important step for various decision making. The amount of information and its correctness plays a crucial role in reasoning. Decision making is easier when we have certain information i.e. the correctness of information can be ascertained. In the other situation when the certainty of information can not be ascertained, the decision-making process is likely to be erroneous or may not be correct. In this situation how decisions are made with some uncertainty (uncertain information) is the core objective of this unit. If we talk about the sources of uncertainty in the information, this could be due to various reasons including experimental error, instrument fault, unreliable source and any other reason. Once the information is received and we have to make decisions based on received uncertain information, we can not rely on models which use certain information. One of the potential solutions appears to be probabilistic reasoning for such scenarios. We can make use of probabilistic models for reasoning with uncertain information with some probability. Let's first see the basic probability concepts before discussing probabilistic reasoning.

7.3 REVIEW OF PROBABILITY THEORY

Now, you are familiar with the reasoning and how it can be useful with probability theory. Before we dive deeper into the Bayes' theorem and its applications, let us review some of the basic concepts of probability theory. These concepts will be helping us to understand other topics of this unit.

Trials, Sample Space, Events : You must have often observed that a random experiment may comprise a series of smaller sub-experiments. These are called trials. Consider for instance the following situations.

Example 1: Suppose the experiment consists of observing the results of three successive tosses of a coin. Each toss is a trial and the experiment consist of three trials so that it is completed only after the third toss (trial) is over.

Example 2: Suppose from a lot of manufactured items, ten items are chosen successively following a certain mechanism for checking. The underlying experiment is completed only after the selection of the tenth item is completed; the experiment obviously comprises 10 trials.

Example 3: If you consider Example 1 once again you would notice that each toss (trial) results into either a head (H) or a tail (T). In all there are 8 possible outcomes of the experiment viz., s_1 = (H,H,H), s_2 = (H,H,T), s_3 = (H,T,H), s_4 = (T,H,H), s_5 =(1,T,H), s_6 = (T,H,T), s_7 =(H,T,T) and s_8 = (T, T, T).

Let ζ be a fixed sample space. We have already defined an event as a collection of sample points from ζ. Imagine that the (conceptual) experiment underlying ζ is being performed. The phrase "the event E occurs" would mean that the experiment results in an outcome that is included in the event E. Similarly, non-occurrence of the event E would mean that the experiment results into an outcome that is not an element of the event E. Thus, the collection of all sample points that are not included in the event E is also an event which is complementary to E and is denoted as E^c. The event E^c is therefore the event which contains all those sample points of ζ which are not in E. As such, it is easy to see that the event E occurs if and only if the event E^c does not take place. The events E and E^c are complementary events and taken together they comprise the entire sample space, i.e., $E \;\; E^c = \zeta$.

You may recall that ζ is an event which consists of all the sample points. Hence, its complement is an empty set in the sense that it does not contain any sample point and is called the null event, usually denoted as ø so that ζ^c = ø.

Let us once again consider Example 3. Consider the event E that the three tosses produce at least one head. Thus, E = $\{s_1, s_2, s_3, s_4, s_5, s_6, s_7\}$ so that the complementary event $E^c=\{s_8\}$, which is the event of not scoring a head at all. Again in Example 3 in the case of selection without replacement, event that the white marble is picked up at least once is defined as E = $\{(r_1,w), (r_2,w), (w, r_2) (w, r_1)\}$. Hence, $E^c = \{(r_1, r_2), (r_2 , r_1)\}$ i.e. the event of not picking the white marble at all.

Let us now consider two events E and F. We write E F, read as E "union" F, to denote the collection of sample points, which are responsible for occurrence of either E or F or both. Thus, E F is a new event and it occurs if and only if either E or F or both occur i.e. if and only if at least one of the events E or F occurs. Generalizing this idea, we can define a new event Ej, read as "union" of the k events $E_1, E_2,..., E_k$, as the event which consists of all sample points that are in at least one of the events $E_1, E_2,\ldots E_k$and it occurs if and only if at least one of the events $E_1, E_2,...,E_k$ occurs.

Again, let E and F be two given events. We write E ∩ F, read as E "Intersection" F, to denote the collection of sample points any of whose occurrence implies the occurrence of both E and F. Thus, E ∩ F is a new event and it occurs if and only if both the events E and F occur. Generalizing this idea, we can define a new event Ej read as "intersection" of the k events $E_1, E_2,...,E_k$, as the event which consists of sample points that are common to each of the events $E_1, E_2,..., E_k$, and it occurs only if all the k events $E_1, E_2,...,E_k$ occur simultaneously. Further, two events E and F are said to be mutually exclusive or disjoint if they do not have a common sample point i.e. E ∩ F = ø.

Two mutually exclusive events then cannot occur simultaneously. In the coin-tossing experiment for instance, the two events, heads and tails, are mutually exclusive: if one occurs, the other cannot occur. To have a better understanding of these events let us once again look at Example 3. Let E be the event of scoring an odd number of heads and F be the event that tail appears in the first two tosses, so that E = $\{s_1, s_5, s_6, s_7\}$ and F = $\{s_5, s_8\}$. Now E ∩ F = $\{s_5\}$, the event that only the third toss yields a head. Thus events E and F are not mutually exclusive.

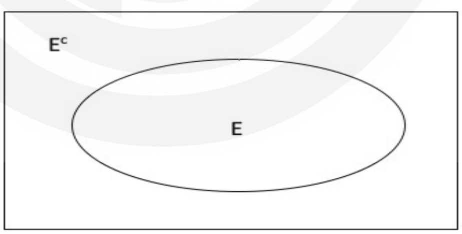


Fig. 1(a)

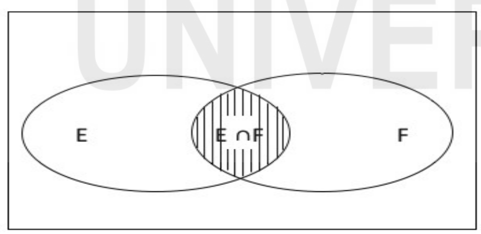


Fig.1(b)

The above relations between events can be best viewed through a Venn diagram. A rectangle is drawn to represent the sample space ζ. All the sample points are represented within the rectangle by means of points. An event is represented by the region enclosed by a closed curve containing all the sample points leading to that event. The space inside the rectangle but outside the closed curve representing E represents the complementary event E^c (See Fig.1(a) above) Similarly, in Fig.1(b), the space inside the curve represented by the broken line represent the event E U F and the shaded portion represents E ∩ F.

As is clear by now, the outcome of a random experiment being uncertain, none of the various events associated with a sample space can be predicted with certainty before the underlying experiment is performed and the outcome of it is noted. However, some events may intuitively seem to be more likely than the rest. For example, talking about human beings, the event that a person will live 20 years seems to be more likely compared to the event that the person will live 200 years. Such thoughts motivate us to explore if one can construct a scale of measurement to distinguish between likelihoods of various events. Towards this, a small but extremely significant fact comes to our help. Before we elaborate on this, we need a couple of definitions.

Consider an event E associated with a random experiment; suppose the experiment is repeated n times under identical conditions and suppose the event E (which is not likely to occur with every performance of the experiment) occurs $f_n(E)$ times in these n repetitions. Then, $f_n(E)$ is called the frequency of the event E in n repetitions of the experiment and $r_n.(E) = f_n,(E)/n$ is called the relative frequency of the event E in n repetitions of the experiment. Let us consider the following example.

Example 4: Consider the experiment of throwing a coin. Suppose we repeat the process of throwing a coin 5 times and suppose the frequencies of occurrence of head is tabulated below in Table-1:

No. of repetitions (n)	Frequency of head ($f_n(H)$	Relative frequency of head $r_n(H)$
1	0	0
2	1	1/2
3	2	2/3
4	3	3/4
5	3	3/5

Notice that the third column in Table-1 gives the relative frequencies r_n (H) of heads. We can keep on increasing the number of repetitions n and continue calculating the values of r_n (H) in Table 1. Merely to fix ideas regarding the concept of probability of an event, we present below a very naive approach which in no way is rigorous, but it helps to see things better at this stage.

Check Your Progress- 1

Problem -1. In each of the following exercises, an experiment is described. Specify the relevant sample spaces:

a) A machine manufactures a certain item. An item produced by the machine is tested to determine whether or not it is defective.

b) An urn contains six balls, which are colored differently. A ball is drawn from the urn and its color is noted.

c) An urn contains ten cards numbered 1 through 10. A card is drawn, its number noted and the card is replaced. Another card is drawn and its number is noted.

Problem 2. Suppose a six-faced die is thrown twice. Describe each of the following events:

i) The maximum score is 6.
ii) The total score is 9.
iii) Each throw results in an even score.
iv) Each throw results in an even score larger than 2.
v) The scores on the two throws differ by at least 2.

7.3.1 Conditional probability and independent events

Let ζ be the sample space corresponding to an experiment and E and F are two events of ζ Suppose the experiment is performed and the outcome is known only partially to the effect that the event F has taken place. Thus there still remains a scope for speculation about the occurrence of the other event E. Keeping this additional piece of information confirming the occurrence of F in view, it would be appropriate to modify the probability of occurrence of E suitably. That such modifications would be necessary can be readily appreciated through two simple instances as follows:

Example 5: Suppose, E and F are such that F E so that occurrence of F would automatically imply the occurrence of E. Thus with the information that the event F has taken place in view, it is plausible to assign probability 1 to the occurrence of E irrespective of its original probability.

Example 6: Suppose. E and F are two mutually exclusive events and thus they cannot occur together. Thus whenever we come to know that the event F has taken place, we can rule out the occurrence of E. Therefore, in such a situation, it will be appropriate to assign probability 0 to the occurrence of E.

Example 7: Suppose a pair of balanced dice A and B are rolled simultaneously so that each of the 36 possible outcomes is equally likely to occur and hence has probability   Let E be the event that the sum of the two scores is 10 or more and F be the event that exactly one of the two scores is 5.

Then E = {(4.6), (5.5), (5,6), (6,4), (6,5), (6,6)} so that P(E) = 6/36 = 1/6.

Also, F= {(1.5), (2,5), (3,5), (4,5), (6,5), (5,1), (5,2), (5,3), (5,4), (5,6)}.

Now suppose we are told that the event F has taken place (note that this is only partial information relating to the outcome of the experiment). Since each of the outcome originally had the same probability of occurring, they should still have equal probabilities. Thus given that exactly one of the two scores is 5 each of the 10 outcomes of event F has probability while the probability of remaining 26 points in the sample space is 0.

In the light of the information that the event F has taken place the sample points (4,6), (6,4), (5,5) and (6,6) in the event E must not have materialized. One of the two sample points (5,6) or (6,5) must have materialized. Therefore the probability of E would no longer be 1/6. Since all the 10 sample points in F are equally likely, the revised probability of E given the occurrence of F, which occurs through the materialization of one of the two sample points (6,5) or (5,6) should be 2/10 = 1/5.

The probability just obtained is called the conditional probability that E occurs given that F has occurred and is denoted by P(E|F). We shall now derive a general formula for calculating P(E|F).

Consider the following probability table:

Table 2

Events	E	E^c
F	P	Q
F^c	r	s

In Table 2, $P(E \cap F) = p$, $P(E^c \cap F) = q$, $P(E \cap F^c) = r$ and $P(E^c \cap F^c) = s$ and hence, $P(E)=P(E \cap F) \cup (E \cap F^c)) = P(E \cap F) + P(E \cap F^c) =p+r$ and similarly, $P(F) = q +s$.

Now suppose that the underlying random experiment is being repeated a large number of times, say N times. Thus, taking a cue from the long term relative frequency interpretation of probability, the approximate number of times the event F is expected to take place will be NP(F) = N(q+s). Under the condition that the event F has taken place, the number of times the event E is expected to take place would be NP(E ∩ F) as both E and F must occur simultaneously. Thus, the long term relative frequency of E under the condition of occurrence of F, i.e. the probability of occurrence of E under the condition of occurrence of F, should be NP(E ∩ F)/NP(F) = P(E ∩ F)/P(F). This is the proportion of times E occurs out of the repetitions where F takes place. With the above background, we are now ready to define formally the conditional probability of an event given another.

Definition: Let E and F be two events from a sample space ζ. The conditional probability of the event E given the event F, denoted by P(E|F), is defined as P(E|F) = P(E ∩ F)/P(F), whenever P(F) > 0.

When P(F) = 0, we say that P(E|F) is undefined. We can also write from Eqn. P(E ∩ F) = P(E|F)P(F).

Referring back to Example 3, we see that P(E) = 6/36,P(F) = 10/36; since, E ∩ F= {(5,6), (6,5)}, P(E ∩ F) = 2/36, P(E|F) = (2/36)/(10/36) = 2/10 = 1/5, which is the same as that obtained in Example 3. Another result can be generalized to k events E_1 E_2, ..., E_k, where k $\geq$2. And now an exercise for you.

Check Your Progress 2

Problem-1: In a class, three students tossed one coins (one each) for 3 times. Write down all the possible outcomes which can be obtained in this experiment.

Problem-2: In problem 1, what is the probability of getting 2 more than 2 heads at a time. Also write the probability of getting three tails at a time.

Problem-3: In problem 1 calculate the Relative frequency of tail $r_n(T)$.

7.4 INTRODUCTION TO BAYESIAN THEORY

Bayes' theorem is widely used to calculate the conditional probabilities of events without a joint probability. It is also used to calculate the conditional probability where intuition fails. In simple terms, probability of a given hypothesis H conditional on E can be defined as P_E (H) = P(H \& E)/ P(E), where P(E) > 0, and the term P(H \& E) also exists. Here P_E is referred to as a probability function. To simply understand the Bayes' theorem, have a look at the following definitions.

Joint Probability: This refers to the probability of two or more events simultaneously occurring, e.g. P(A and B) or P(A, B).

Marginal Probability: It is the probability of an event occurring irrespective of outcome of the other random variables e.g. P(A).

Conditional probability: A conditional probability is defined as the probability of occurrence of an event provided that another event has occurred. e.g. P(A | B).

The conditional probability can also be written in terms of joint probability as P(A|B) = P(A, B)/ P(B). In other way, if one conditional probability is given, other can be calculated as P(A|B) = P(B|A)*P(A) /P(B).

Let ‘S’ be a sample space in consideration. Let events ‘A1’ , ‘A2’ , ... ‘An’ is the set of mutually exclusive events in sample space ‘S’. Let ‘B’ be an event from sample space ‘S’ provided P(B) > 0, then according to Bayes’ theorem.

$P(A_k \mid B) = P(A_k \cap B) / P(A_1 \cap B)+P(A_2 \cap B) \ldots\ldots\ldots\ldots P(A_n \cap B)$, this can also be written in terms of Bayes’ theorem.

$P(A_k \mid B) = P(A_k)P(B \mid A_k) / P(A_1).P(B \mid A_1)+P(A_2).P(B \mid A_2) \ldots\ldots\ldots\ldots P(A_n).P(B \mid A_n)$

7.5 BAYE’S NETWORKS

The probabilistic models are being used in defining the relationships among variables and are used to calculate probabilities.The Bayes’ network is a simpler form of applying Bayes’ theorem to complex real world problems. This uses a probabilistic graphical model which captures the conditional dependence explicitly and is represented using directed edges in a graph. Here if we take fully conditional models, we may need a big amount of data to address all possible events/ cases and in such scenario probabilities may not be calculated practically. On the other hand, simple assumptions like conditional independence of random variables may turn out to be effective, giving a way for Bayes’ Network.

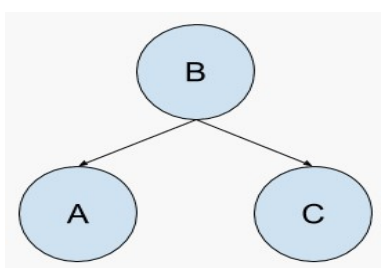
While representing a Bayes’ Network graphically, nodes represent the distribution of probabilities for random variables. The edges in the graph represent the relationship among random variables. The key benefits of a Bayes’ Network are model visualization, relationships among random variables and computations of complex probabilities.

Example 8: Let us now create a Bayesian Network for an example problem. Let us consider three random variables A, B and C. It is given that A is dependent on B, and C is dependent on B. The conditional dependence can be stated as P(A|B|) and P(C|B) for both the given statements respectively. Similarly the conditional independence can be stated here as P(A|B, C) and P(C|B, A) for both statements respectively.

Here, we can also write P(A|C, B) = P(A|B) as A is unaffected by the C. Also the joint probability of A and C given B can be written as product of conditional probabilities as P(A, C|B) = P(A|B) * P(C|B).

Now using Bayes' theorem, the joint probability of P(A, B, C) can be written as P(A,B,C)= P(A|B)*P(C|B)*P(B).

The corresponding graph is shown below in figure 1. Here each random variable is represented as a node and edges between nodes are conditional probabilities.



7.6 PROBABILISTIC INFERENCE

The probabilistic inference is very much dependent on the conditional probability of the specified events provided the information of occurrence of other events is available. For example, two events E and F such that P(F)>0, the conditional probability of event E when F has occurred can be written as :

$$P(E/F) = \frac{P(E \cap F)}{P(F)}$$

When an experiment is repeated a large number of times (say n), the above expression can be given a frequency interpretation. Let the number of occurrences of an event F is represented as No. (F) and the probability of a joint event of E and F as No. (E ∩ F). The relative frequencies of both these events can be computed as f_r:

$f_r(E \cap F) = \frac{No.(E \cap F)}{n}$ and similarly,

$$f_r(F) = \frac{No.(F)}{n}$$

Here, if n is large, the ratio of above two expressions represent the proportion of times the event E occurs relative to the occurrence of F. This can also be understood as the approximate conditional occurrence of event F with E.

$$f_r(E \cap F) / f_r(F) \simeq P(E \cap F) / P(F)$$

We can also write the conditional probability of event F while it is given that event E has already occurred, as

$$P(E / F) = P(E \cap F) / P(F)$$

Using above two equations we can also write

$$P(F / E) = P(E / F)\, P(F) / P(E)$$

The above expression is also one form of Bayes' Rule. Here the notion is simple: the probability of an event F occurring when we know the probability of an event E which has already occurred is the same as the probability of occurring of event E when the probability of occurrence of event F is known.

7.7 BASIC IDEA OF INFERENCING WITH BAYE'S NETWORKS

We are now aware of the Bayes theorem, probability and Bayes networks. Let's now talk about how inferences can be made using Bayes networks.A network here represents the degree of belief of proposition and their causal interdependence. The inference in a network can be done by propagating the given probabilities of related information through the network giving the output to one of the conclusion nodes. The network representation also reduces the time and space requirements for huge computations involving the probabilities of uncertain knowledge of propositional variables. Further, one can not make the inference from such a large data in real time. The solution to such a problem can be found using the network representation. Here the network of nodes represents variables connected by edges which represents causal influences (dependencies) among nodes. Here the edge weights can be used to represent the strength of influences or in other terms the conditional probabilities.

To use this type of probabilistic inference model, one first needs to assign probabilities to all basic facts in the underlying knowledge base. This requires the definition of an appropriate sample space and the assignment of a priori and conditional probabilities. In addition to this some methods must be selected to compute the combined probabilities when pooling evidence in a sequence of inference steps. In the end, when the outcome of an inference chain results in one or more proposed conclusions, the alternatives must be compared and one should be chosen on the basis of likelihood.

7.8 OTHER PARADIGM OF UNCERTAIN REASONING

The other ways of dealing with uncertainty are the ones with no theoretical proof. These are mostly based on intuition. These are selected over formal methods as a pragmatic solution to a particular problem, when the formal methods impose difficult or impossible conditions. One such ad hoc procedure is used to diagnose meningitis and infectious blood disease, the system is called MYCIN. The MYCIN uses If and then rules to assess various forms of patient evidence. It also measures both belief and disbelief to represent degree of confirmation and disconfirmation respectively in a given hypothesis. The ad hoc methods have been used in a larger number of knowledge-based systems than formal methods. This is due to the difficulties encountered in acquiring a large number of reliable probabilities related to the given domain and to the complexities to the ensuing calculations.

One other paradigm is to use Heuristic reasoning methods. These are based on the use of procedures, rules and other forms of encoded knowledge to achieve specified goals under certainty. Using both domain specific and general heuristics, one of several alternative conclusions may be chosen through the strength of positive vs negative evidence presented in the form of justification or endorsement.

The in depth and detailed discussion on this is not in the scope of this unit/ course.

7.9 DEMPSTER SCHEFFER THEORY

Let us now discuss a mathematical theory based only on the evidence, known as Dempster-Schafer (D-S) theory given by Dempster and extended by Shafer in "Mathematical Theory of Evidences". This uses a belief function to combine separate and independent evidence pieces to quantify the belief in a statement. The D-S theory is a generalization of Bayesian probability theory where multiple possible events are assigned probabilities opposed to mutually exclusive singletons. The D-S theory assumes the existence of ignorance in knowledge creating uncertainty which in turn induces belief. Here, uncertainty of the hypothesis is represented by the belief function. The main characteristic of the theory is:

1. Multiple possible events are permitted to assign probabilities.
2. These events should be exhaustive and exclusive.

Here, the multiple sources of information are assigned some degree of belief and then aggregated using the D-S combination rule. This also limits the theory for intensive computation because of the lack of independent assumptions from such a large number of information sources.

Let us now define a few terms used in D-S theory which will be useful for us.

7.9.1 Evidence

These are events related to one hypothesis or set of hypotheses. Hare, a relation is not permitted between various pieces of evidence or set of hypotheses. Also, the relation between the set of hypotheses and the piece of evidence is only quantified by a source of data. In context of D-S theory, we have four types of evidences as following:

a) Consonant Evidence: These are basically appearing in a nested structure where each subset is included into the next bigger subset and so on. Here with each increasing subset size, the information refines the evidentiary set over the time.
b) Consistent Evidence: This assures the presence of at least one common element to all the subsets.
c) Arbitrary Evidence: A situation where there is not a common element occurring in the subsets though some of the subsets may have a few common element(s).
d) Disjoint Evidence: There is no subset having common elements.

All these four evidence types can be understood by looking at the below given figure 2.(a-d).

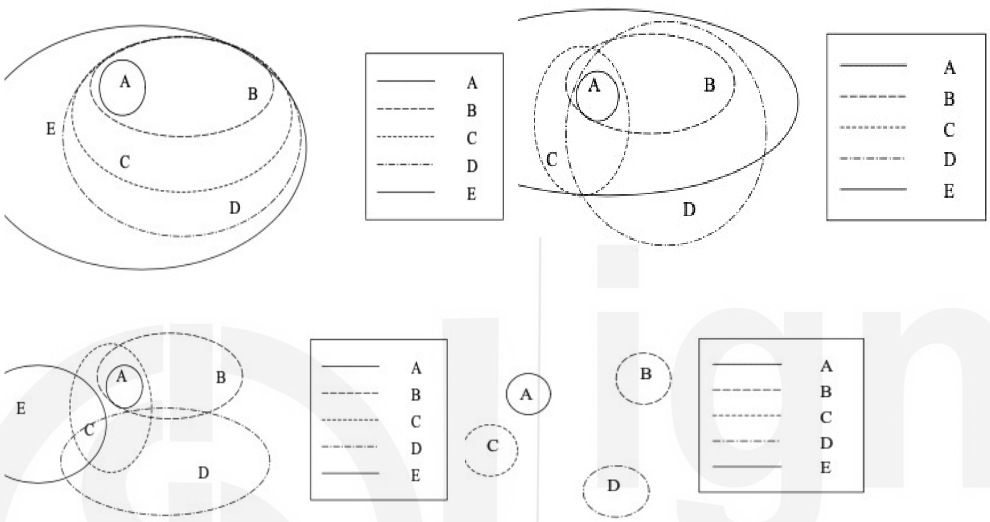


Figure 2. (a-d)

The source of information can be an entity or person giving some relevant state information. Here the information source is a non biased source of information. The information received from such sources is combined to provide more reliable information for further use. The D-S theory models are able to handle the varying precision regarding the information and hence no additional assumptions are needed to represent the information.

7.9.2 Frame of Discernment

Let us consider a random variable 'θ' whose true value is not known. Let 'θ' = { θ_1, θ_2 ….. θ_n } represent mutually exclusive and discretized values of the possible outcome of 'θ'. Conventionally, the uncertainty about 'θ' is given by the assigning probability p_i to the elements θ_i, i = 1: n, satisfying sum p_i = 1. In the case of D-S theory, the probabilities are assigned to the subsets of 'θ' and the individual element 'θ_i' along with it.

7.9.3 The Power Set P(θ= 2^{θ})

This is defined as the set of all subsets of 'θ' including singletons, defining the frame of 'θ'. The subset of this powerset may contain a single or conjunctions of hypotheses. Here, with respect to the power set, the complete probability assignment is called basic probability assignment.

The core functions in D-S theory are :

1. Basic Probability Assignment function

This is represented by m and maps the power set to the interval 0 and 1. Here, the basic probability assignment (bpa) to the null set is 0 and for all subsets of the power set sum = 1. For a given set A, m(A) represents the measure of belief assigned by the available evidences in support of A, where $A \in 2^\{\theta\}$. Mathematically, the bpa can be represented as follows.

1. m : $2^\{\theta\}$--> [0, 1] (interval)
2. $m(\phi) = 0$ (null)
3. $m(A) \geq 0, \forall A \in 2^\{\theta\}$
4. sum{m(A) $\forall A \in 2^\{\theta\}$} = 1.

This is to note here that, the element of power set with m(A) > 0 is termed as focal element(s).

Example9: Let θ = {a, b, c}; then the power set is P(θ) = {ϕ, a,b,c, (a,b), (a,c), (b,c),

(a,b,c)}. The information source assigned the m-values as m(a) = 0.2, m(c) = 0.1 and m(a,b) = 0.4. Here the mentioned three subsets are focal elements.

2. The Belief Function

The assignment of the basic probability we can define the lower and upper bounds of the intervals representing the precise probability of a set. This is also bounded by continuous measures of nonadditive nature known as Belief and Plausibility.

The lower bound (belief) for set A is defined as the sum of all basic probability assignments of proper subset B of set A. The measurement of the amount of support by the information source given to support a specific element as a correct one is done by the belief function, mathematically Bel(A) = $\text{sum}_{\{B \subset A\}}\ m(B)\ \ \forall\ A \subset \theta$.

3. The Plausibility Function

The upper bound (plausibility) for set A is defined as the sum of all basic probability assignments of B intersecting set A, mathematically Pl(A) = $\text{sum}_{\{B \cap A \neq \phi\}}\ m(B)$. Here, the plausibility function measures the level of information by a source contradicting an element as a correct answer specifically.

Apart from the above-mentioned functions a few terms also require some attention while referring

to the D-S theory. The Uncertainty Interval, shows the range where the true probability may be

found. This is calculated as the difference of belief and plausibility level i.e. Pl(A) - Bel(A).

7.9.4 Rule of Combination

In the D-S theory, the measure of Plausibility and Belief are taken from the combined assignments. The D-S rule of combination takes multiple belief functions and combines them using m i.e. respective basic probability assignments. The D-S combination rule is basically a conjunctive operation i.e. AND. Here the joint $m_{\{12\}}$ (combination) is obtained using aggregating two basic probability assignments m_1 and m_2 as following:

$$m_{\{12\}}(A) = 1/\text{sum}_{\{\{B \cap A = A\}\ \ m1\ (B)\ m2\ (C)\}}\{1\text{-}K\}\ ,$$

Where, $A \neq \phi$,

$M_{\{12\}}(\phi) = 0$,

And $K = \text{sum}_{\{B \cap C = \phi\}}\ m_1(B)\ m_2(C)$.

Here, in the above expression, K is the basic probability mass which is associated with the conflict calculated as a sum of products of the basic probability assignments of all sets having null intersection. The normalization factor is represented as 1-K in the denomination. The rule is associative, commutative but not continuous or idempotent in nature.

Example 10: In a multinational company 100 applicants appeared for a job interview. The company setup two interview boards for applicants.

While assessing the grades of the class of 100 students, two of the class teachers responded the overall result as follow. First teacher assessed that 40 students will get A and 20 students will get B grade amongst the total 60 students he interviewed. Whereas second teacher stated that 30 students will get A grade and 30 students will get either A or B amongst the 60 students he took the interview. Combining both evidences to find the resultant evidence, we will do following calculations. Here frame of discernment $\theta = \{A, B\}$ and Power set $2^{\theta} = \{\emptyset, A, B, (A, B)\}$,

Evidence (1) =Ev1	Evidence (2) =Ev2
$m_1(A) = 0.4$	$m_2(A) = 0.3$
$m_1(B) = 0.2$	$m_2(A, B) = 0.3$
$m_1(\theta) = 0.4$	$m_2(\theta) = 0.4$

Plausibility function (PI):

$A \cap A = A \neq \emptyset$ hence $m_1(A) = 0.4$

$A \cap B = \emptyset$

$A \cap \theta = A \neq \emptyset$ hence $m_1(\theta) = 0.4$

$Pl_1(A) = m_1(A) + m_1(\theta) = 0.4 + 0.4 = 0.8$

$A \cap A = A \neq \emptyset$ hence $m_2(A) = 0.3$

$A \cap B = \emptyset$

$A \cap \theta = A \neq \emptyset$ hence $m_2(\theta) = 0.4$

$Pl_2(A) = m_2(A) + m_2(\theta) = 0.3 + 0.4 = 0.7$

$B \cap A = \emptyset$

$B \cap B = B \neq \emptyset$ hence $m_1(B) = 0.2$

$B \cap \theta = B \neq \emptyset$ hence $m_1(\theta) = 0.4$

$Pl_1(B) = m_1(B) + m_1(\theta) = 0.2 + 0.4 = 0.6$

$(A, B) \cap A = A \neq \emptyset$ $m_2(A) = 0.3$

$(A, B) \cap B = B \neq \emptyset$, $m_2 (B) = 0$

$(A, B) \cap (A, B) = (A, B) \neq \emptyset$ $m_2(A, B) = 0.3$

$(A, B) \cap \theta = (A, B) \neq \emptyset$ hence $m_2 (\theta) = 0.4$

$Pl_1(A, B) = m_2 (A) + m_2 (A, B) + m_2 (\theta) = 0.3 + 0.3 + +0.4 = 1.0$

$\theta \cap A = A \neq$ □ hence $m_1 (A) = 0.4$

$\theta \cap B = B \neq$ □ hence $m_1 (B) = 0.2$

$\theta \cap \theta = \theta \neq$ □ hence $m_1 (\theta) = 0.4$

$Pl_1(\theta) = m_1 (A) + m_1 (B) + m_1 (\theta) = 0.4 + 0.2 + 0.4 = 1.0$

$\theta \cap A = A \neq$ □ hence $m_2 (A) = 0.3$

$\theta \cap (A, B) = (A, B) \neq$ □, $m_2 (A, B) = 0.3$

$\theta \cap \theta = \theta \neq$ □ hence $m_2 (\theta) = 0.4$

$Pl_2(\theta) = m_2 (A) + m_2 (A, B) + m_2 (\theta) = 0.3 + 0.3 + 0.4 = 1.0$

D-S Rule of Combination: Table 3 shows combination of concordant evidences using D-S theory.

Evidences	m1(A)=0.4		m1(B)=0.2		m1(θ)=0.4
m2(A)=0.3	m1-2 (A)	0.12	m1-2 (□)	0.06	m1-2 (A) 0.12
m2(A,B)=0.3	m1-2 (A)	0.12	m1-2 (B)	0.06	m1-2 (A,B) 0.12

m2(θ)=0.4	m1-2 (A)	0.16	m1-2 (B)	0.08	m1-2 (θ)	0.16

k = 0.06 and 1 − k = 0.94 Combined masses are worked out

Bel1-2(A) = m1-2(A) = 0.553

Bel1-2(B) = m1-2(B) = 0.149

$Bel_{1-2}(A, B) = m_{1-2}(A) + m_{1-2}(B) + m_{1-2}(A, B) = 0.553 + 0.149 + 0.128 = 0.83$ $Bel_{1-2}(\theta) = m_{1-2}(A) + m_{1-2}(B) + m_{1-2}(A, B) + m_{1-2}(\theta) = 0.553 + 0.149 + 0.128 +$

$0.170 = 1$

$Pl_{1-2}(A) = m_{1-2}(A) + m_{1-2}(A, B) + m_{1-2}(\theta) = 0.553 + 0.128 + 0.170 = 0.851,$

(85 students in A Grade)

$Pl_{1-2}(B) = m_{1-2}(B) + m_{1-2}(A, B) + m_{1-2}(\theta) = 0.149 + 0.128 + 0.170 = 0.447,$

(45 students in B Grade)

$Pl_{1-2}(A, B) = m_{1-2}(A) + m_{1-2}(B) + m_{1-2}(AB) + m_{1-2}(\theta) = 0.553 + 0.149 + 0.128 + 0.170 = 1.0$

$Pl_{1-2}(\theta) = m_{1-2}(A) + m_{1-2}(B) + m_{1-2}(A, B) += 0.553 + 0.149 + 0.128 + 0.170 =$

1.00. (100 students in total)

According to rule of combination, concluded ranges are then 55 to 85 students will get

``A'' grade and 15 to 45 students will get ``B'' grade.

Key advantages of D-S theory:

- The level of uncertainty reduces with addition of information.
- Addition of more evidences reduces ignorance
- We can represent diagnose hierarchies using D-S theory.

Check Your Progress 3

Problem-1. Differentiate between Join, Marginal and conditional probability with an example of each.

Problem-2. Explain Dempster Shafer theory with a suitable example.

Problem-3. What are different type of evidences? Give suitable example of each.

7.10 SUMMARY

This unit relates to the discussion over Reasoning with uncertain information, whih involves Review of Probability Theory, and Introduction to Bayesian Theory. Unit also covers the concept of Baye's Networks, which is later used for the purpose of inferencing. Finally, the unit discussed about the Other Paradigm of Uncertain Reasoning, including the Dempster Scheffer Theory

7.11 SOLUTIONS/ANSWERS

Check Your Progress- 1

Problem -1. In each of the following exercises, an experiment is described. Specify the relevant sample spaces:

a) A machine manufactures a certain item. An item produced by the machine is tested to determine whether or not it is defective.
b) An urn contains six balls, which are colored differently. A ball is drawn from the urn and its color is noted.
c) An urn contains ten cards numbered 1 through 10. A card is drawn, its number noted and the card is replaced. Another card is drawn and its number is noted.

Solution - *Please refer to section 7.3 to answer these problems.

Problem 2. Suppose a six-faced die is thrown twice. Describe each of the following events:

i) The maximum score is 6.
ii) The total score is 9.
iii) Each throw results in an even score.
iv) Each throw results in an even score larger than 2.
v) The scores on the two throws differ by at least 2.

Solution - *Please refer to section 7.3 to answer these problems.

Check Your Progress 2

Problem-1: In a class,three students tossed one coins (one each) for 3 times. Write down all the possible outcomes which can be obtained in this experiment.

Solution - *Please refer to example 4 and section 7.3 to solve these problems

Problem-2: In problem 1, what is the probability of getting 2 more than 2 heads at a time. Also write the probability of getting three tails at a time.

Solution - *Please refer to example 4 and section 7.3 to solve these problems

Problem-3: In problem 1 calculate the Relative frequency of tail $r_n(T)$.

Solution - *Please refer to example 4 and section 7.3 to solve these problems

Check Your Progress 3

Problem-1. Differentiate between Join, Marginal and conditional probability with an example of each.

Solution - *Please refer to section 7.9 and example 10 to answer these problems.

Problem-2. Explain Dempster Shafer theory with a suitable example.

Solution - *Please refer to section 7.9 and example 10 to answer these problems.

Problem-3. What are different type of evidences? Give suitable example of each.

Solution - *Please refer to section 7.9 and example 10 to answer these problems.

7.12 FURTHER READINGS

1. David Barber ,"Bayesian Reasoning And Machine Learning", Cambridge University Press
2. John J. Craig, "Introduction to Robotics", Addison Wesley publication
3. Ela Kumar, " Artificial Intelligence", IK International Publications
4. Ela Kumar, " Knowledge Engineering ", IK International Publications