Introduction: Recently, the scientific and engineering community has begun to recognize the importance of defining multiple types of uncertainty. The dual nature of uncertainty is described by the following concepts:

Aleatory Uncertainty: The type of uncertainty which results from the fact that a system can behave in random ways (ex. Noise).

Epistemic Uncertainty: The type of uncertainty which results from the lack of knowledge about a system and is a property of the analysts performing the analysis (Subjective uncertainty).

Probability theory has been traditionally used to characterize both types of uncertainty. E.g. aleatory uncertainty can be best dealt with using the frequentist approach associated with traditional probability theory. However, probability theory is not capable of completely capturing epistemic uncertainty. The traditional method was called Bayesian Probability. In this method, it is necessary to have information on the probability of all the events. When this is not available, uniform distribution function is often used, which means that all simple events for which a probability distribution is not known in a given sample space are equally likely. An additional assumption in the classical probability is the axiom of additivity where all probabilities that satisfy specific properties must add to 1.

i.e. \( P(A) + P(\overline{A}) = 1 \)

The 3 axioms of Bayesian Theory are:
1. \( P(\emptyset) = 0 \);
2. \( P(\Theta) = 1 \);
3. if \( A \cap B = 0 \) then \( P(A \cup B) = P(A) + P(B) \)

Dempster Shafer Theory offers an alternative to traditional probabilistic theory for the mathematical representation of uncertainty. The significant innovation of this framework is that it allows for the allocation of probability mass to sets or intervals. It does not require an assumption regarding the probability of the individual element of the interval or set. An important aspect of this theory is the combination of evidence obtained by multiple sources and the modeling of conflict between them. It provides good results for evaluation of risk and reliability in engineering applications when it is not possible to obtain a correct and precise measurement from experiments.

Dempster Shafer Theory ignores the 3\textsuperscript{rd} Bayesian Axiom and states that, i.e. \( P(A \cup B) \geq P(A) + P(B) - P(A \cap B) \).

Definition of Terms:
A set is represented by \( \Theta \) which contains all the possible elements of interest in each particular context and its elements are exhaustive and mutually exclusive events. This \( \Theta \) is called universal
set or the frame of discernment. E.g. in tossing a fair die, the frame of discernment is \{1, 2, 3, 4, 5, 6\}. This looks similar to the sample space in probability theory, but the difference is that in DS theory, the number of possible hypothesis is \(2^n\) while in probability theory it is \(|\theta|\) where \(|X|\) is the cardinality of the set X. For simplicity we assume the number of elements of \(\theta\) to be finite.

**Definition 1:** If \(\theta\) is a frame of discernment then a function \(m: 2^\theta \rightarrow [0, 1]\) is called a basic probability assignment if

\[ m(\emptyset) = 0 \]

and

\[ \sum_{A \in \theta} m(A) = 1 \]

The term \(m(A)\) is called A’s basic probability number and \(m(A)\) is the measure of the belief that is committed exactly to A.

1. It is not required that \(m(\emptyset)=1\).
2. It is not required that \(m(A) \leq m(B)\) when \(A \subset B\).
3. No relationship between \(m(A)\) and \(m(\overline{A})\) is required.
4. Also \(m(A) + m(\overline{A})\) does not always have to be 1.

**Definition 2:** Belief function Bel: \(2^\theta \rightarrow [0, 1]\)

\[ Bel(A) = \sum_{B \subset A} m(B) \]

For any \(A \subset \theta\). Bel(A) measures the total belief of all possible subsets of A.

Properties of Belief Function:

1. \(Bel(\emptyset)=0\).
2. \(Bel(\emptyset)=1\).
3. \(Bel(A_1 \cup \ldots \cup A_n) \geq \sum_{i} Bel(A_i) - \sum_{i<j} Bel(A_i \cap A_j) + \ldots + (-1)^{n+1} Bel(A_1 \cap \ldots \cap A_n)\).

**Definition 3:** Plausibility function Pl: \(2^\theta \rightarrow [0, 1]\)

\[ Pl(A) = \sum_{B \cap A \neq \emptyset} m(B) \]

Also, \(Pl(A)=1-Bel(A)\).

and, \(Bel(A) \leq Pl(A)\).

Properties of Plausibility Function:

1. \(Pl(\emptyset)=0\).
2. \(Pl(\emptyset)=1\).
3. \(Pl(A_1 \cap \ldots \cap A_n) \leq \sum_{i} Pl(A_i) - \sum_{i<j} Pl(A_i \cup A_j) + \ldots + (-1)^{n+1} Pl(A_1 \cup \ldots \cup A_n)\).

**Definition 4:** Doubt is defined as:

\(Doubt(A) = Bel(\overline{A})\).
Pl(A) measures the total belief mass that can move into A, whereas Bel(A) measures the total belief mass that is confined to A.

Belief is also called lower probability and Plausibility is called upper probability. We can show a graphical representation the definition as follows:

If A, AB, ABC, CD and BC are events and their probabilities are known, then we can represent probability, belief and plausibility of AB as:

![Graphical representation](image)

**Rule of Combination:**

Suppose there are two distinct bodies of evidence and they are expressed by two basic assignments m1 and m2 on some focal points from power sets of the frame of discernment. The combined two belief functions can be represented by means of a joint basic assignment m1, 2.

![Combining two basic assignments](image)

From the figure, m1 is a basic assignment over θ and its focal points are B1...Bn. The other basic assignment m2 has focal points on C1...Cn. Then the joint basic assignment for A is given by:
\[ m_{1,2}(A) = \sum_{B_i \cap C_j = A} m_1(B_i) m_2(C_j) . \]

But, there is a problem in this formula when \( B_i \cap C_j = 0 \) for particular \( i \) and \( j \) such as \((i, j) = (1,2), (2,n), \ldots, (m,2) \) as shown in the figure. In this case, the total area of rectangles which are now the focal elements for \( m_{1,2} \) is less than one, which violates second condition of basic probability assignment. To make the sum of \( m_i \)'s to be equal to 1 we multiply them by the factor:

\[ (1 - \sum_{A_i \cap B_j = \emptyset} m_1(B_i) m_2(C_j))^{-1} . \]

Thus the rule of combination is expressed by:

\[ m_{1,2}(A) = \frac{\sum_{B_i \cap C_j = A} m_1(B_i) m_2(C_j)}{1 - \kappa} \]

where \( \kappa = \sum_{B_i \cap C_j = \emptyset} m_1(B_i) m_2(C_j) . \)

\( \kappa \) is also called conflict. It indicates the degree of conflict between two distinct bodies of evidence. Larger \( \kappa \) indicates that there is more conflict in evidence.

This rule satisfies both associative and commutative rules; i.e. if \( m_3(A) \) comes in, then use \( m_{12} \) and \( m_3 \). (\( m_{13}, m_2, m_{23}, m_1 \): Order does not matter.).

**Example 1:**

Suppose there are 100 cars in a parking lot consisting of type A and B. Two policemen check the type of cars in the lot. First policeman says that there are 30 A cars and 20 B cars among the 50 cars he counted. Second policeman says that there are 20 A cars and 20 cars that could A or B. Thus the basic probability assignment could be:

\[
\begin{array}{c|c|c|c}
 & m_1(A) & m_1(B) & m_1(\emptyset) \\
\hline
m_2(A) & 0.3 & 0.2 & 0.5 \\
m_2(AB) & 0.2 & 0.06 & 0.04 (\text{0 intersection}) & 0.1 \\
m_2(\emptyset) & 0.06 & 0.18 & 0.12 & 0.3 \\
\end{array}
\]

\( \kappa = 0.04 . \)

\( 1-\kappa = 0.96 . \)

Thus, using the rule of combination,
\[ m_{12}(A) = \frac{(0.06+0.06+0.18+0.1)}{(1-k)} = 0.42. \]
\[ m_{12}(B) = \frac{(0.04+0.12)}{(1-k)} = 0.17. \]
\[ m_{12}(AB) = \frac{0.1}{(1-k)} = 0.1. \]
\[ m_{12}(\theta) = \frac{0.3}{(1-k)} = 0.31. \]

\[ \text{Bel}(A) = m_{12}(A) = 0.42. \text{ (42 A cars)} \]
\[ \text{Bel}(B) = m_{12}(B) = 0.17. \text{ (17 B cars)} \]
\[ \text{Bel}(AB) = m_{12}(A) + m_{12}(B) + m_{12}(AB) = 0.69. \text{ (69 cars)} \]

\[ \text{Pl}(A) = m_{12}(A) + m_{12}(AB) + m_{12}(\theta) = 0.83. \text{ (83 A cars)} \]
\[ \text{Pl}(B) = m_{12}(B) + m_{12}(AB) + m_{12}(\theta) = 0.58. \text{ (58 B cars)} \]
\[ \text{Pl}(\theta) = 1.00. \text{ (100 cars)} \]

Thus by the rule of combination, we get a range of values such as 42-83 A cars and 17-58 B cars.

Example 2:

Comparison of Dempster Shafer Theory and Bayesian Theory.

Suppose if \( m_1(A) = 0.6 \) and \( m_1(\overline{A}) = 0.2 \) and \( m_2(A) = 0.8 \) and \( m_2(\overline{A}) = 0.1 \).

For the rest of the values, let \( m_1(\theta) = 0.2 \) and \( m_2(\theta) = 0.1 \).

<table>
<thead>
<tr>
<th></th>
<th>( m_1(A) )</th>
<th>( m_1(\overline{A}) )</th>
<th>( m_1(\theta) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_2(A) )</td>
<td>0.8</td>
<td>0.48</td>
<td>0.16</td>
</tr>
<tr>
<td>( m_2(\overline{A}) )</td>
<td>0.1</td>
<td>0.06</td>
<td>0.02</td>
</tr>
<tr>
<td>( m_2(\theta) )</td>
<td>0.1</td>
<td>0.06</td>
<td>0.02</td>
</tr>
</tbody>
</table>

\( K = 0.16 + 0.6 = 0.22. \)

\( 1-k = 0.78. \)

Using the formulas from example (1), we can find the following:

\[ m_{12}(A) = \frac{(0.48+0.06+0.16)}{(1-k)} = 0.9. \]
\[ m_{12}(\overline{A}) = \frac{(0.02+0.02+0.02)}{(1-k)} = 0.08. \]
\[ m_{12}(\theta) = \frac{0.02}{(1-k)} = 0.03. \]

Using the formula for belief and plausibility functions:

\[ \text{Bel}(A) = 0.9. \]
\[ \text{Bel}(\overline{A}) = 0.08. \]
\[ \text{Bel}(\theta) = 1. \]
\[ \text{Pl}(A) = 0.92. \]
\[ \text{Pl}(\overline{A}) = 0.1. \]
\[ \text{Pl}(\theta) = 1. \]

If we represent it graphically:
If we solve this problem using Bayesian Theory:

Covariance of A by m1 and by m2 is:

\[ P = (P_1^{-1} + P_2^{-1})^{-1} = \frac{1}{0.1} + \frac{1}{0.05} = \frac{10 + 20}{10 + 20} = 0.033. \]

Here, we are actually calculating the confidence in the value of A using two different sources.

Now, mean,

\[ m = P(P_1^{-1}m1 + P_2^{-1}m2) \]

= 0.033(10*0.7 + 20*0.85) = 0.8.

Comparing the range of uncertainties of Dempster Shafer (0.9-0.92) and Bayesian (0.8+or-covariance), we can conclude that Dempster Shafer provides the estimate with a smaller degree of uncertainty. Dempster Shafer combines more efficiently the two pieces of evidence than Bayesian.

**Example 3:**

If Doctor1 says that a patient has 99% chances of having cough and 1% chance of having flu and Doctor2 says that the same patient has 99% chances of having backache and 1% chance of having flu, then we can surely say that the patient has the lowest probability of having flu. Let us solve this using Dempster Shafer Theory.

Thus,

\[ m1(A) = 0.99 \text{ (cough)} \]
\[ m1(B) = 0.1 \text{ (flu)} \]
\[ m_2(C) = 0.99 \text{ (backache)} \]
\[ m_2(B) = 0.01 \text{ (flu)} \]

Then,

<table>
<thead>
<tr>
<th></th>
<th>( m_2(C) )</th>
<th>( m_2(B) )</th>
<th>( m_2(\theta) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_1(A) )</td>
<td>0.99</td>
<td>0.01</td>
<td>0</td>
</tr>
<tr>
<td>( m_1(B) )</td>
<td>0.01</td>
<td>0.0099</td>
<td>0.0001</td>
</tr>
<tr>
<td>( m_1(\theta) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ K = 0.9999. \]
\[ 1 - \kappa = 0.0001. \]

Now using the rules of combination,

\[ m_{12}(A) = 0; \]
\[ m_{12}(C) = 0; \]
\[ m_{12}(B) = 1. \]

From this result, combination of both the doctors’ prescription and using Dempster Shafer Theory, the chances of the patient having the flu is 100%. But this is wrong. Thus, Dempster Shafer Theory fails to give satisfactory results when a high degree of conflict exists among the sources of evidence.

**Applications of Dempster Shafer Theory:**

1. **Sensor Fusion:** The main function of sensor fusion techniques is to extract more reliable and accurate information from an object using two or more sensors. Since there are not only operating limits of a sensor but also noise corrupted environment, no single sensor can guarantee to deliver acceptably accurate information all the time. We can use Dempster Shafer Theory in sensor fusion. According to this scheme, sensor classifier produces two parameters for each hypothesis, supportability variable and plausibility variable. The difference is the ignorance of the hypothesis. Dempster Shafer’s rule of combination provides a means of computing composite supportability/plausibility intervals for each hypothesis. As a result, a supportability/plausibility interval vector is obtained with a normalization vector. If one hypothesis is strongly supported by a group of sensor classifiers, then the credibility interval is reduced for that hypothesis. The normalization vector indicates the consistency of each sensor.

   Potential advantage of using Dempster Shafer theory in Sensor Fusion:
   1. Different level of abstraction.
   2. Representation of Ignorance.
   3. Conflict Resolution.

2. **Expert Systems:** Uncertainty in rule based systems can be represented by associating a basic assignment with each rule.

   \[ m \]

   \[ E \rightarrow H \]

   We can have an expert system which aims at selecting a forecast technique or its combination based on the characteristics of a forecasting situation given by the user. The objective of forecasting is to predict the future. Since its decision is directly related with the future of the business, it is very important to forecasting or marketing managers in the field. This system has five modules, knowledge acquisition, knowledge base, belief base, Dempster Shafer module and Inference Engine.

3. **Fault Detection and Identification:** The FDI algorithm is developed for detecting and identifying a failure in a system. Suppose that there are more than one estimable parameters,
the Dempster Shafer Theory is useful method to detect and identify a failure from the signatures of the parameters. In the last stage of FDI algorithm, the decision is made with an index which measures the reliability of the decision, which is called the degree of certainty.

**Future of Dempster Shafer Theory:**

1. **Relationship between Fuzzy Logic, Neural Networks, Bayesian and Dempster Shafer Theory.**

2. **Modification of Dempster Shafer rule using weighted sources of evidence:** In Dempster Shafer rule, the weights of independent sources of evidence are considered to be the same. Also, if belief functions m1 and m2 are identical, then their combination may not be equal to either of them. As a result, in repetitive combination of belief with same bpa, the resultant bpa either goes to zero or unity (Flu example). If we use weighted sources of evidence, we can solve this problem. Weighted Dempster Shafer allows taking into account the different reliabilities of the sensors. A high (low) weight is assigned to a more reliable (less reliable) source of evidence. Research is devoted to dynamically update these weights on-line.