2. UNCERTAINTY QUANTIFICATION

Uncertainty is a fundamental and unavoidable feature of daily life; in order to deal with uncertainty intelligently, we need to be able to represent it and reason about it. Human reasoning usually is approximate in nature and involves various uncertainties. Uncertainty pertains to information that is not definitely fixed, not precisely determined, not dependable or that is vague or indistinct. A decision maker dealing with uncertainty has reduced confidence and assurance in decisions that are dependent upon that information. Uncertainty must be quantified in order to use it systematically in decision-making processes. To the degree that we can quantify uncertainty, it is increasingly feasible to make more reliable and assured decisions.

Uncertainty quantification (UQ) is defined as the quantitative characterization and reduction of uncertainties in applications. Many problems in the fields of natural sciences and engineering are rife with sources of uncertainty. Increasing application of computer simulation modelling to study such problems has unfolded a new construct in the form of uncertainty quantification (UQ).

2.1 REASONS OF UNCERTAINTY

The reasons responsible to introduce uncertainty in a model may be:

1. The model structure, i.e., how accurately does a mathematical model describe the true system for a real-life situation,

2. The numerical approximation, i.e., how appropriately a numerical method is used in approximating the operation of the system,

3. The initial / boundary conditions, i.e., how precise are the data / information for initial and / or boundary conditions,

4. The data for input and/or model parameters.
2.2 TYPES OF UNCERTAINTIES:
The following three types of uncertainties can be identified:
1. Uncertainty due to variability of input and/or model parameters when the characterization of the variability is available
2. Uncertainty due to variability of input and/or model parameters when the corresponding variability characterization is not available,
3. Uncertainty due to an unknown process or mechanism.

Type 1 uncertainty, which depends on chance, may be referred to as aleatory or statistical uncertainty. Type 2 and 3 are referred to as epistemic or systematic uncertainties.

It often happens in real life applications that all three types of uncertainties are present in the systems under study. Uncertainty quantification intends to work toward reducing type 2 and 3 uncertainties to type 1. The quantification for the type 1 uncertainty is relatively straightforward to perform. To evaluate type 2 and 3 uncertainties, the efforts are made to gain better knowledge of the system, process or mechanism. Methods such as fuzzy logic or evidence theory are used.

It should be clear from this description of UQ that the application of UQ is essential to the whole model development process from conception to maturity, and that UQ should not be viewed as just a discipline to study the propagation of uncertainty from the inputs to the output of interest. Thus UQ in the generic sense encompasses:

- Uncertainty analysis
- Sensitivity analysis
- Design exploration
- Design optimization
- Design validation and calibration
Previous literature on UQ has been emphasized on application-specific uncertainty analyses. For uncertainty analyses that involve many parameters with significant higher order sensitivities and computational models that do not allow intrusion, design and analysis of experiments have been found to be useful. The quantification of this uncertainty has, in recent years, grown from a collection of scientific ideas into a sub-discipline of computational science that attempts to provide a quantitative description of incomplete knowledge for use in conjunction with model-based computational resources, algorithms, and software. Uncertainty exists in all branches of science and engineering. Accordingly, in recent reports and initiatives on scientific computing, uncertainty quantification (UQ) has been recognized as a critical element necessary for continued advancement in prediction science, life-cycle design, and societal sustainability. Significant progress has been made in recent years within a subset of related problems in science and engineering.

There are three current options for accomplishing the task of Uncertainty quantification:

- Second Order probability
- Bayesian methods
- Generalized probability theories

2.3 METHODS OF UNCERTAINTY QUANTIFICATION

Propagation of Generalized probability is currently a research problem for complex technical decision problems. Following methods can be used for the quantification of uncertainty.

- Certainty Factor
- Dempster-Shafer Theory
- Fuzzy Logic
2.3.1 Certainty Factor

A certainty factor is used to express the accuracy, truthfulness or reliability of a predicate. It is neither a probability nor a truth-value. In McAllister’s scheme, a certainty factor is a number from 0.0 to 1.0; CF can be expressed as numbers in following manner:

<table>
<thead>
<tr>
<th>CF</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>Strongly suggestive</td>
</tr>
<tr>
<td>0.6</td>
<td>Suggestive</td>
</tr>
<tr>
<td>0.4</td>
<td>Weakly Suggestive</td>
</tr>
<tr>
<td>0.2</td>
<td>Slight hint</td>
</tr>
</tbody>
</table>

Table 2.1: Meaning of Certainty Factors

Shortliffe [49] introduced the notion of certainty factors for modelling belief in the Mycin medical expert system. The certainty of a hypothesis given some evidence is expressed using two factors: the measure of belief and the measure of disbelief.

Definitions of these measures are [94]:

\[ MB[h,e] = x \] means “the measure of increased belief in the hypothesis h, based on the evidence e, is x”

\[ MD[h,e] = y \] means “the measure of increased disbelief in the hypothesis h, based on the evidence e, is y”

These values are governed by the following restrictions

\[ 0 = MB[h,e] = 1 \]
\[ 0 = MD[h,e] = 1 \]

The certainty factor of a hypothesis h given some evidence e is then defined

\[ CF[h,e] = MB[h,e] - MD[h,e], \]
Where \(-1 = CF[h,e] = 1\)

The certainty factor indicates the net belief in a hypothesis. \(CF > 0\) indicates that there is more reason to believe a hypothesis than to disbelieve it, while \(CF < 0\) indicates that there is more reason to disbelieve a hypothesis than to believe it. \(CF = 0\) indicates either a lack of evidence concerning a hypothesis or conflicting evidence.

Measures of belief and disbelief are combined and propagated according to the following functions [96]

In Certainty factor Calculus, **MYCIN rule** is well known. It is described as follows:

**Rule[1]: IF P1 AND P2 AND P3 THEN D**

This rule can be quantified with CF (Certainty Factor) in following manner

\[ CF(D) = \min(\ CF(P1), \ CF(P2), \ CF(P3)) \ast CF(Rule) \]

**Rule[2]: IF P1 OR P2 OR P3 THEN D**

\[ CF(D) = \max(\ CF(P1), \ CF(P2), \ CF(P3)) \ast CF(Rule) \]

**Rule[3]: CF(not P) = 1 - CF(P)**

Example: IF sky_is_overcast AND humid THEN rain

\(CF(Rule) = 0.8\)

\(CF(sky\_is\_overcast) = 0.9\)

\(CF(humid) = 0.7\)

\(CF(Decision) = ?\)

According to MYCIN Rule

\[ CF(Decision) = \min(0.9, 0.7) \ast 0.8 \]

\[ = 0.56 \]

**2.3.2 Dempster Shafer Theory**

Dempster-Shafer Theory (DST) is a mathematical theory of evidence [20]. It is a generalization of Bayesian inference. This theory was originally
developed by Dempster and further expanded by Shafer [91], introduced concept of reasoning in discrete domain. By allowing the representation of ignorance, DST avoids the problem of requiring the complete set of conditional probabilities under Bayesian approach. It provides a method for explicitly representing the ignorance inherent in knowledge and of reasoning in the absence of certain information. According to Haddawy [40] the belief in a proposition A is represented by a Shafer interval \([s(A) .. p(A)]\) which is a subinterval of the unit interval \([0 .. 1]\), where \(s(A)\) is called the support and \(p(A)\) is called the plausibility of the proposition A. These values can be thought of as an upper and lower bounds on the probability of A. The uncertainty of a proposition is then defined as the width of the interval: \(u(A) = p(A) - s(A)\). The precision of the probability estimate for the proposition A is then defined as: \(\text{PRE}(A) = 1 - u(A)\). The PRE of the decision is the measure, which will be used to characterize the certainty of an inference.

The variable precision logic (VLP) represents knowledge in terms of facts and rules. To apply Dempster Shafer theory to VLP, we must provide an interpretation of rules and facts in terms of Shafer intervals. Dempster Shafer theory can be defined as:

\[
S(A) + s(\sim A) + I = 1, \quad \text{Where I stands for Ignorance.}
\]

Dempster Shafer theory can be used in production rules in following manner:

IF \(C_i[s(C_1), p(C_1)]\) and \(C_2[s(C_2), p(C_2)]\) THEN \(D[s(C \rightarrow D), p(C \rightarrow D)]\)

Rule[1] prob \((C_1\text{ and } C_2) \in [s(C_1) \ast s(C_2), \ p(C_1) \ast p(C_2)]\)

Rule[2] prob \((C_1 \text{ or } C_2) \in [s(C_1) + s(C_2) - (s(C_1) \ast s(C_2)), \ p(C_1) + p(C_2) - (p(C_1) \ast p(C_2))}\)

Rule[3] prob \((\sim C) \in [1-p(C), 1-s(C)]\)

Rule[4] prob \((D) \in [s(C) \ast s(C \rightarrow D), (1 - s(C)) + s(C) \ast p(C \rightarrow D)]\)

Example: Consider a rule IF C THEN D, value of shafer interval \([s(C), p(C)] = [0.7, 0.9]\) and \([s(C \rightarrow D), p(C \rightarrow D)] = [0.8, 1]\) then compute \([s(D), p(D)]\).
According to Dempster Shafer Theory
\[ \text{Prob}(D) \in [0.7 \cdot 0.8, (1-0.7) + (0.7 \cdot 1)] \]
\[ \text{Prob}(D) \in [0.56, 1] \]

2.3.2.1 Dempster's Rule

Dempster's rule is a method of integrating information from independent sources. In Dempster's calculus, propositions are represented as subsets of the exhaustive set of possibilities \( \Theta \), the frame of discernment. A basic probability assignment \( m \) maps elements of \( 2^\Theta \) to a value in the interval \([0..1]\). Given a proposition which represents some subset of the possibilities contained in the frame of discernment, \( m(A) \) represents the probability mass constrained to stay in \( A \) but otherwise free to move, called its basic probability mass. This represents our ignorance because we cannot further subdivide our belief and restrict movement of the probability mass. The residual uncertainty of the domain, \( m(\Theta) \), represents the residual uncertainty of the domain, \( m \) is defined as follows:

\[
\begin{align*}
  m(\emptyset) &= 0 \\
  m(A) &\in (0..1) \\
  \Sigma m(A) &= 1
\end{align*}
\]

Where \( \emptyset \) denotes the empty set. The support for a proposition \( A \) is then defined as the total mass attributed to \( A \) and all its subsets. This represents the total belief in \( A \). The plausibility is one minus the support for the complement of \( A \).

\[
\begin{align*}
  \text{Support}(A) &= s(A) = \Sigma m(A) \\
  \text{Plausibility} &= p(A) = 1 - s(\sim A)
\end{align*}
\]

Example:
We consider 40 fair, 20 wheatish, 30 dark, and 10 women with covered face, making a total of 100 women. Based on this, the basic probability assignment $m$ is defined as

\begin{align*}
    m(\text{fair}) &= 0.4 \\
    m(\text{wheatish}) &= 0.2 \\
    m(\text{dark}) &= 0.3 \\
    m(\text{covered face}) &= m(\text{fair V wheatish V dark}) = m(\emptyset) = 0.1
\end{align*}

The support for the proposition $A = \text{fair V wheatish}$ is then

$$s(A) = m(\text{fair}) + m(\text{wheatish}) = 0.6$$

and the support for $\neg A = \text{dark}$ is

$$s(\neg A) = m(\text{dark}) = 0.3.$$ 

The plausibility of $A$ is $p(A) = 1 - s(\neg A) = 0.7$. So the $\text{prob}(A) \in [0.6 \ldots 0.7]$. The uncertainty in the domain is just $m(\emptyset) = 0.1$. Dempster-Shafer theory allows us to represent ignorance by assigning a probability mass to a statement, which contains several alternatives, in this case $m(\emptyset)$. In this way we can account for a lack of observations and uncertain observations, both of which represent forms of incomplete information.

Two basic probability assignments $m_1$ and $m_2$, which provide evidence concerning a common frame of discernment, are combined using Dempster's orthogonal sum rule to yield a new basic probability assignment. The combined effects of $m_1$ and $m_2$ are represented by a unit square. Suppose we have a frame of discernment $\Theta = [a,b,c]$ and the propositions $A \equiv aVb$ and $\neg A \equiv c$. Figure 2.1 shows the unit square representing the combination of two basic probability mass functions $m_1$ and $m_2$. The belief committed exactly to the combination of $A \cap \Theta$ is then the area of the rectangles corresponding to the intersection, i.e., $m_1(A).m_2(\Theta).m_2(A)$. The total mass allocated to a given subset $C$ of $\Theta$ is
Referring to the above example, the total mass allocated to \( A \) is

\[
m(A) = m_1(A) \cdot m_2(A) + m_1(A) \cdot m_2(\emptyset) + m_1(\emptyset) \cdot m_2(A)
\]

Since the combination of \( m_1 \) and \( m_2 \) must again be a basic probability assignment, the following must hold.

\[
\sum_{A \cap B = \emptyset} m_1(A_i) m_2(B_j) + \sum_{A \cap B = \emptyset} m_1(A_i) m_2(B_j) = 1
\]

Thus Dempster's orthogonal sum rule for the new probability function for all subset \( C \) of \( \Theta \) is

\[
\sum_{A \cap B = C} m_1(A_i) m_2(B_j).
\]

The denominator is a normalizing factor, which removes the probability assigned to the empty set.
Now consider the case in which $\Theta$ contains only two values $A$ and $\neg A$. The support and plausibility values are then:

$s(A) = m(A)$

$p(A) = 1 - m(\neg A) = m(A) + m(\Theta)$.

Substituting these into Dempster's rule and simplifying, we obtain:

$$S(A) = 1 - \frac{(1 - s_1(A)(1 - s_2(A))}{1 - [s_1(A)(1 - p_2(A) + (1 - p_1(A))s_2(A)]}$$

Dempster's rule can be expressed in terms of Shafer intervals as:

$$[a..b] \oplus [c..d] = \left[\begin{array}{cc}
\tilde{a} & \tilde{c} \\
1 - (ad + \tilde{bc}) & 1 - (\tilde{a}d + bc)
\end{array}\right]$$

Where $\oplus$ denotes the orthogonal sum of the two intervals $[a .. b]$ and $[c .. d]$. The formula is both associative and communicative so it can be applied pairwise in any order.

2.3.2.2 Dempster-Shafer Interpretation of VPL

The VPL system represents knowledge in terms of rules and facts. To apply DST to VPL, we must provide an interpretation of rules and facts in terms of Shafer intervals. The belief in a fact is represented by associating a Shafer interval with each fact. Shafer intervals are viewed as representing upper and lower bounds on the probability of a fact; thus $A_{[s,p]} = \text{prob}(A) \in [s .. p]$, where $A$ is proposition. A censored production rule should then imply information concerning the Shafer interval of its decision given the Shafer intervals of its premises and censors. Implication is interpreted as expressing conditional probability.
Four belief values are associated with each censored production rule:

\[ P \rightarrow D \mid C: \alpha, \beta, \gamma, \delta \]

The \( \alpha \) value is lower bounded on the probability of \( D \) given \( P \) and \( \neg C \); the \( \beta \) value is the lower bound on the probability of \( \neg D \) given \( P \) and \( C \); the \( \gamma \) value is the lower bound on the probability of \( D \) given \( P \); and the \( \delta \) value is the lower bound on the probability of \( \neg D \) given \( P \). This is summarized below:

\[
\begin{align*}
\text{Prob}(D|P & \& \neg C) \in [\alpha ..1] \\
\text{Prob}(\neg D|P & \& C) \in [\beta ..1] \\
\text{Prob}(D|P) \in (\gamma ..1) \\
\text{Prob}(\neg D|P) \in \delta ..1]
\end{align*}
\]

The \( \gamma \) and \( \delta \) values are constrained by the restriction that \( \gamma + \delta \leq 1 \), thus

\[ \text{Prob}(D|P) \in [\gamma ..(1-\delta)] \]

For example, the following rule might be used to express the fact that I read the paper before going to work unless I oversleep, which occurs once or twice a week:

\[ \text{Weekday - morning} \rightarrow \text{Read-paper} \mid \text{Oversleep: 0.9, 1, 0.6, 0.2} \]

Where the belief factors are interpreted as follows:

- The "0.9" states that on weekday morning when I do not oversleep I read the paper at least 0.9 of the times because there are other factors which would keep me from reading the paper, such as the paper boy throwing it on the roof which are not being considered;
- the "1" states that on weekday mornings when I oversleep I certainly do not read the paper;
- the "0.6" states that I read the paper at least three out of five weekday mornings (because I oversleep at most twice a week);
- the "0.2" states that I do not read the paper at least one out of five weekday mornings (because I oversleep at least once a week)
This formalism also allows representation of incomplete and inconsistent censors. Figure 2.2 shows the set theoretic interpretation of a censored production rule. Most of the premise intersects with the decision, indicating that the inference $P \rightarrow D$ is quite strong. Most of that part of the premise, which is outside of the decision, is covered by the censor. This censor is both incomplete and inconsistent. The censor part of a rule is incomplete if it does not cover all possible exceptions. It is inconsistent if it has a nonempty intersection with the decision. The degree of incompleteness is expressed by the amount by which $\alpha < 1$ and the degree of inconsistency by the amount by which $\beta < 1$.

![Figure 2.2: Set theoretic interpretation of $P \rightarrow D \mid C$](image)

The above discussion has presented a representation of belief for propositional logic, but the VPL system uses a predicate logic representation. In this representation, terms containing only ground instances are equivalent to propositional logic and this present no additional problems. However, a semantics for expressions with free variables is needed. Rules of the form $A(x,y) \rightarrow B(x)$, with an associated belief $[s..p]$ are interpreted as $\forall x,y \ \text{prob}(B(x)|A(x,y)) \in [s..p]$. This is essentially a shorthand for listing rules.
over the entire domain of x and y. Similarly, a fact A(x) with belief \([s.. p]\) is interpreted as \(\forall x \ \text{prob}(A(x)) \in [s.. p]\).

2.3.3 Fuzzy Logic

Fuzzy logic is a well-defined reasoning system that is based on the use of fuzzy sets rather than on the binary values. A fuzzy set is a class of elements with loosely defined boundaries. Formally, a fuzzy set \(F\) in a universe of discourse \(U\) is defined by a membership function, \(\mu_F : U \rightarrow [0 \ 1]\). The membership function expresses the degree of membership of elements in the fuzzy subset. For example \(\mu = 0\) indicates no membership, \(\mu = 1\) indicates full membership and \(0 < \mu < 1\) indicates partial membership of the fuzzy set.

A fuzzy IF...THEN rule may be expressed as: \(Y\) is \(B\) if \(X\) is \(A\), in which the antecedent, consequent or both are fuzzy rather than crisp. Where \(X\) and \(Y\) are variables whose domains are \(U\) and \(V\) respectively, \(A\) and \(B\) are fuzzy predicates or relations in \(U\) and \(V\), which play the role of elastic constraints on \(X\) and \(Y\). Thus IF...THEN rules in a fuzzy system are nothing but relations between fuzzy sets.

Example: Volume is low if pressure is high. In this case \(Y = \) volume, \(X = \) pressure, \(B = \) low and \(A = \) high.

A fuzzy conditional statement “IF \(X\) is \(A\) THEN \(Y\) is \(B\)” (of \(A(x) \Rightarrow B(y)\) for short) where “IF \(X\) is \(A\)” is represented by the membership function \(\mu_A(x)\): domain(\(X\)) \(\rightarrow [0 \ 1]\) which restricts the possible values of the variable \(X\) to a fuzzy set of values having (each to a certain degree) the property \(A\). In a similar way “\(Y\) is \(B\)” is represented by the membership function \(\mu_B(y)\): domain(\(Y\)) \(\rightarrow [0 \ 1]\). Then the meaning of the rule \(A(x) \Rightarrow B(y)\) is represented by a fuzzy relation defined on domain(\(X\)) \(\times\) domain(\(Y\)), i.e. \(\mu_R(x,y)\): domain(\(X\)) \(\times\) domain(\(Y\)) \(\rightarrow [0 \ 1]\)

Here the Gödels type of implication is chosen:

\(\mu_R(x,y) = 1, \ \text{if} \ \mu_A(x) \leq \mu_B(y)\)
2.4 CONCLUSION

Our aim of the research is to discover Quantified Hierarchical Censored Production Rules. By using an exception rule as Censored Production Rule, we are free to ignore the censor (exception) condition when the resources needed to establish its presence are tight or simply no information is available as to whether it holds or does not hold. As time permits, the censor condition C is evaluated establishing the conclusion B with higher certainty if C does not hold otherwise if C holds then the conclusion is ~ B. As we have discussed in Chapter 1, different approaches have been developed to deal with different types of uncertainty. For example, the best approach to handle type of uncertainty “ignorance” is Dempster Shafer Theory. It provides a method for explicitly representing the ignorance inherent in knowledge and of reasoning in the absence of certain information. According to the requirement of our problem we have used Dempster Shafer Theory as a method of uncertainty quantification.

\[ = \max\{\min(\mu_A(x), \mu_R(x,y))\} \ldots (2) \]