Systems Analysis Laboratory

Bachelor's thesis

Espoo, 20.4.2015

Katri Selonen

A study on evidence theory: a general representation of uncertainty

Thesis supervisor:

Prof. Raimo P. Hämäläinen

Thesis advisor:

Ph.D. Matteo Brunelli

Aalto University
School of Science

Aalto University
School of Science

Abstract of bachelor's thesis

Author Katri Selonen	
Title A study on evidence theory: a general representation of uncertainty	
Degree programme Engineering Physics and Mathematics	
Major Systems Sciences	**Code** T3010
Supervisor Prof. Raimo P. Hämäläinen	
Advisor Ph.D. Matteo Brunelli	
Date 20.4.2015 **Language** English	**Number of pages** 5+19

The aim of this thesis is to provide a comprehensible tutorial on evidence theory. It includes some background and most importantly the mathematical basis of the theory. Additionally, examples and comparison to other similar frameworks are discussed to enable deeper understanding. Another important aspect provided is evidence theory's current and other possible application areas.

Evidence theory is a framework for reasoning with uncertainty. It is based on measures called belief and plausibility. Those values can be assigned to groups of elements instead of single elements, which is the greatest difference compared to probability theory. Evidence theory provides tools for calculations and decision making in situations involving uncertainty.

Keywords Evidence theory, Dempster-Shafer theory, uncertainty, belief, plausibility, rule of combination

Aalto University
School of Science

kandidaatintyön tiivistelmä

Tekijä Katri Selonen	
Työn nimi Tutkimus todisteteoriasta epävarmuuden kuvauksena	
Koulutusohjelma Teknillinen fysiikka ja matematiikka	
Pääaine Systeemitieteet	**Koodi** T3010
Työn valvoja Prof. Raimo P. Hämäläinen	
Työn ohjaaja Matteo Brunelli	
Päivämäärä 20.4.2015 **Kieli** Englanti	**Sivumäärä** 5+19

Tämän kandidaatintyön tarkoituksena on tarjota helposti ymmärrettävissä oleva johdatus todisteteoriaan. Siihen sisältyy teorian taustaa sekä tärkeimpänä sen matemaattinen perusta. Lisäksi syvemmän ymmärryksen tarjoamiseksi se käsittää vertailua muihin vastaaviin teorioihin sekä havainnollistavia esimerkkejä. Työ sisältää myös todisteteorian nykyisten ja mahdollisien tulevaisuuden sovelluskohteiden käsittelyä.

Todisteteoria tarjoaa työkaluja epävarmuuden käsittelyyn. Se perustuu uskomusten ja uskottavuuden mittoihin, jotka määrätään käsillä olevien todisteiden valossa. Nämä arvot voidaan myös antaa joukoille alkioita yksittäisten sijaan. Tämä ominaisuus on selkein erottava tekijä esimerkiksi todisteteorian todennäköisyysteorian välillä. Todisteteoria tarjoaakin toisenlaisen lähestymistavan epävarmuutta sisältävien tilanteiden, kuten päätöksenteon, matemaattiselle mallintamiselle.

Avainsanat Todisteteoria, Dempsterin-Shaferin teoria, epävarmuus, uskomus, uskottavuus

Contents

Symbols and operators

Symbols

$\varnothing$	the empty set
X	the universal set, i.e. the set containing all possible elements
P_X	the power set of X, i.e. the set of all possible subsets of X, including the empty set and the set X itself

Operators

$x \in A$	the element x belongs to set A
$A \cup B$	the union of the two sets A and B
$\bigcup\limits_{i \in I} A_i$	the union of all sets A_i whose indices belong to the index set I
$A \cap B$	the intersection of the two sets A and B
$\bigcap\limits_{i \in I} A_i$	the intersection of all sets A_i whose indices belong to the index set I
$A \subseteq X$	the set A is a subset of X, i.e. X contains at least all the elements of A
$A \supseteq B$	the set A is a superset of B, i.e. B is a subset of A

1 Introduction

Many real-life situations involve uncertainty. For example, the weather of the beginning day cannot be known for sure. Meteorologists try to predict the upcoming weather with the help of information gathered from multiple sources. The resulting forecasts are based on statistics and probabilities.

Most people believe that probabilities are the only tools for modelling situations involving uncertainty. However, probability theory is quite difficult to use when the available information is subjective or there is a great amount of uncertainty involved. For instance when making decisions based on experts' opinions or sensors with poor accuracy. These kind of situations could be analysed via a framework called evidence theory.

Instead of assigning probabilities to single events, evidence theory can study the situation in bigger pieces. The available information often consists of more than a single item at once, for example a radar can give information that the target is moving, so the information regards all possible target types that can move. One of the advantages of evidence theory is that in order to do further analysis, the information does not need to be complete.

In evidence theory the values assigned to groups of events or items are called beliefs and plausibilities instead of probabilities. This thesis introduces these concepts in more detail and from a more mathematical point of view. The primary aim is to present the background and basis of evidence theory as well as some of its prospects. Its position among a few other frameworks is also included.

To enable better understanding of evidence theory as a presentation of uncertainty, some detailed examples are portrayed. The prospects of this theory are discussed via matters to improve as well as application areas, both current and future. The goal is to provide an approachable overview of evidence theory.

2 Background

Evidence theory is a mathematical theory developed for reasoning with uncertainty. It was introduced by Arthur P. Dempster in 1967 and extended, refined and recast by Glenn Shafer in 1976 [9]. Named after its creators, evidence theory is often called Dempster-Shafer theory (DST) [7].

Some history and background of evidence theory are presented in this section. An introduction to uncertainty is included. Additionally, the utility of using evidence theory for reasoning with uncertainty compared to other frameworks is discussed.

2.1 Before evidence theory

An economist called G.L.S Shackle introduced the idea of a decision theory that would account for different, less mathematical aspects in decision making [8]. At the time, economics was increasing in precision and mathematical sophistication, but predictions derived from it lacked the relation to economic reality [5]. Shackle wanted decision making to include creativity and inventiveness, so he created a decision theory for that during the 1950s and 1960s. However, the ideas by Shackle were not taken to use at the time because of their lack of suitable formalisation. Instead, probability theory gained more acceptance as an explanation in decision-making. [3]

Decades later, in 2001, Fioretti showed in his article [3] that neither subadditive probability nor infinite alleles model can handle uncertainty the way Shackle wanted. Fioretti and Klir [5] pointed out in their articles, that Shackle's idea of dealing with uncertainty is compatible with evidence and possibility theories. A couple of decades after Shackle's theory, Shafer introduced his framework for representing uncertainty in a way that was mathematically suitable for Shackle's theory. However, Shafer did not do research on the similarity of his ideas with Shackle's ideas. Shackle used different names in his ideas for uncertainty in decision making, and approached the problem from the side of what is unknown, whereas Shafer concentrated on the known parts of an event with uncertainty, but these differences were insignificant. [3]

2.2 The framework

As opposed to probability theory, in evidence theory the 'probabilities' are assigned to sets or intervals instead of singletons [7]. Singleton refers to a set with only one element. The values are considered degrees of belief and are assigned based on the body of evidence [9]. The model is designed such that the body of evidence can be of any level of precision. The values for the degree of belief can be then assigned to any subset of the event space and they can overlap with each other. With evidence theory, imprecise information of a system can be directly represented with sets or intervals [7].

Other frameworks for modelling uncertainty, especially interval-based representations, are possibility theory, probability theory and imprecise probabilities [1]. Even though evidence theory also gives a general framework for formulating these other theories and it deals with intervals, it is based only on crisp sets. Crisp set refers to the set in a classical sense: they have sharp boundaries that distinguish their members from other objects. If the boundaries are not sharp, the set is called fuzzy [1]. More about the different frameworks and the differences of using crisp and fuzzy sets are discussed in Section 4.

So far, evidence theory has been largely developed among the non-traditional theories for representing uncertainty. Evidence theory is fairly easy to understand as it relates to classical probability theory and set theory. It is also versatile in representing different evidence from multiple sources. Additionally, evidence theory provides a method amendable to mathematical analysis [9]. With increased computational power, the analyses of the surrounding world can be more complex and a method for making use of that increased capacity is needed. These together offer a greater depth of study into the scope of uncertainty. [7]

2.3 Uncertainty

There are two different kinds of uncertainty: aleatory or stochastic uncertainty and epistemic or subjective uncertainty. The aleatory uncertainty results from random behaviour of a system, whereas the epistemic uncertainty results from the lack of knowledge of a system. Epistemic uncertainty is often the by-product of an analyst or expert assessing the system or situation, as he usually does not have knowledge of every part of the system. [7]

Classical probability has been used for both types of uncertainty, but the Bayesian probability used for epistemic uncertainty is not completely suitable for that purpose [7]. For classical probability, the probabilities of all events are needed, but are often not available. In probability theory, this lack of information is compensated by Laplace's Principle of Insufficient Reason and axiom of additivity [7].

The Principle of Insufficient Reason can be interpreted to allow using uniform distribution for events with unknown probability distribution because they are assumed equally likely. The axiom of additivity introduces an assumption that all probabilities must add to one. These assumptions can lead to precise-looking information about events that are unknown. For example, if an expert gives information about one part of a three-part system, the information for the other two parts are then derived using the principle and axiom presented. This derived information regarding the two other parts can be completely false in a situation of subjective uncertainty.

Evidence theory offers an alternative, more general representation for epistemic uncertainty. It can be used in situations when there is not enough information to evaluate probabilities of events or when the information is non-specific or subjective, e.g. an expert's opinion. In evidence theory, complete information about a system is not necessary, and there is no need for further assumptions [7]. Instead of probabilities a degree of belief based on the body of evidence is assigned, and the focus is on the combination of them based on evidence instead of how the values are determined [9].

3 Mathematical formalisation

First, normalised monotone measures are introduced for better intelligibility of the following definitions. *Normalised monotone measures* are measures that exhibit a weaker property of monotonicity with respect to the set inclusion instead of the usual additivity property [1]. This substitution is necessary when working within the framework of uncertainty. A monotone measure μ is a mapping from a nonempty family C of subsets from the power set P_X of an universal set X to the range $[0, 1]$. A normalised monotone measure must also satisfy the folliwing conditions [1]:

1. Boundary conditions: $\mu(\varnothing) = 0$ and $\mu(X) = 1$
2. Monotonicity: $\forall A_i, A_j \in C : A_i \subseteq A_j \Rightarrow \mu(A_i) \leq \mu(A_j)$.

Furthermore, in the continuous case the following two conditions are also required:

3. Continuity from below: for any increasing sequence $A_1 \subseteq A_2 \subseteq ...$ of sets in C, if $\bigcup_i A_i \in C$ then $\lim_{i\to\infty} \mu(A_i) = \mu\left(\bigcup_i A_i\right)$
4. Continuity from above: for any decreasing sequence $A_1 \supseteq A_2 \supseteq ...$ of sets in C if $\bigcap_i A_i \in C$ then $\lim_{i\to\infty} \mu(A_i) = \mu\left(\bigcap_i A_i\right)$.

Evidence theory is based on belief and plausibility measures [1]. Another important function is the basic probability assignment which does not, however, refer to the classical probability [7]. A more mathematical approach as well as proofs for the following definitions and properties can be found in a book on generalised measures by Wang and Klir [13].

This section provides the mathematical formulations for the functions used in evidence theory. These formulations are necessary for the use of evidence theory as a mathematical tool in situations involving uncertainty. Additionally, some rules of combination for situations with multiple sources of information are presented.

3.1 Belief measures

A *belief measure* (Bel) is defined as a function that maps the power set of an universal set X to the range $[0, 1]$:

$$Bel : P_X \to [0, 1], \tag{1}$$

where P_X is the power set of X [13]. A belief measure has to meet also the following conditions:

$$Bel(\varnothing) = 0 \tag{2}$$

$$Bel(X) = 1 \tag{3}$$

$$Bel\left(\bigcup_{i=1}^{N} A_i\right) \geq \sum_{I \subset \{1,\ldots,N\}, I \neq \varnothing} (-1)^{|I|+1} Bel\left(\bigcap_{i \in I} A_i\right), \tag{4}$$

where $\{A_1, \ldots A_N\}$ is any finite subclass of X [13]. Conditions (2) and (3) are normalisation factors for the belief measure. The third condition (4) is similar to the additive axiom of probability but it features an inequality. This also means that a belief measure is monotone and superadditive:

$$Bel(A_1 \cup A_2) \geq Bel(A_1) + Bel(A_2) \geq \max\{Bel(A_1), Bel(A_2)\}, \tag{5}$$

where $A_1, A_2 \subset X$ and $A_1 \cap A_2 = \varnothing$ [1].

3.2 Plausibility measures

A *plausibility measure* is also defined as a function that maps the power set of an universal set X to the range $[0, 1]$: $Pl : P_X \to [0, 1]$. A plausibility measure has to meet conditions similar to the belief measure [13]:

$$Pl(\varnothing) = 0 \tag{6}$$

$$Pl(X) = 1 \tag{7}$$

$$Pl\left(\bigcap_{i=1}^{N} A_i\right) \leq \sum_{I \subset \{1,\ldots,N\}, I \neq \varnothing} (-1)^{|I|+1} Pl\left(\bigcup_{i \in I} A_i\right), \tag{8}$$

where $\{A_1, \ldots A_N\}$ is any finite subclass of X. The first two conditions are normalising factors and the third condition (8) is again similar to the additive axiom of probability. This also means that a plausibility measure is monotone and subadditive:

$$Pl(A_1 \cup A_2) \leq Pl(A_1) + Pl(A_2), \tag{9}$$

where $A_1, A_2 \subset X$ [1].

Belief and plausibility measures form a duality:

$$Pl(A) = 1 - Bel(\bar{A}) \tag{10}$$

$$Pl(\bar{A}) = 1 - Bel(A) \tag{11}$$

$$Bel(A) = 1 - Pl(\bar{A}) \tag{12}$$

$$Bel(\bar{A}) = 1 - Pl(A), \tag{13}$$

for any $A \subset X$ [1]. With the help of the dual equations (10)–(13), belief can be calculated if plausibility is known and vice versa. Another important property is that the value of plausibility is always greater than or equal to the value of belief:

$$Pl(A) \geq Bel(A) \quad \forall A \subset X. \tag{14}$$

3.3 Möbius representation

The conditions and properties for belief and plausibility measures are more straightforward to prove using a Möbius representation [13]. A Möbius representation called a basic assignment can be used to characterise the body of evidence represented by a family of sets $\{A_1, \ldots A_N\}$ and the assignment is given to the sets not the elements [1].

The *basic assignment* is denoted by m and can be characterised similarly to the belief and plausibility measures: $m : P_X \to [0, 1]$. The basic assignment also has to satisfy the following conditions that act as normalising factors [1]:

$$m(\varnothing) = 0 \tag{15}$$

$$\sum_{A \in P_X} m(A) = 1. \tag{16}$$

The basic assignment can be used to compute the belief and plausibility measures for any set $A_i \in P_X$ [1]:

$$Bel(A_i) = \sum_{A_j \subseteq A_i} m(A_j) \tag{17}$$

$$Pl(A_i) = \sum_{A_j \cap A_i \neq \varnothing} m(A_j). \tag{18}$$

An example for computing belief and plausibility and a visualisation of their differences as outlined by equations (17) and (18) can be seen in Figures 1 and 2. The grey boxes represent the sets that are included in each calculation. Examples with more detail are demonstrated in Section 5.

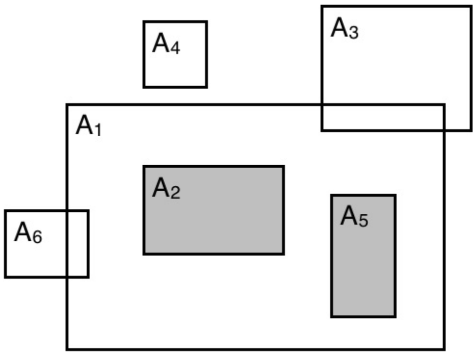

Figure 1: Computing belief $Bel(A_1)$.

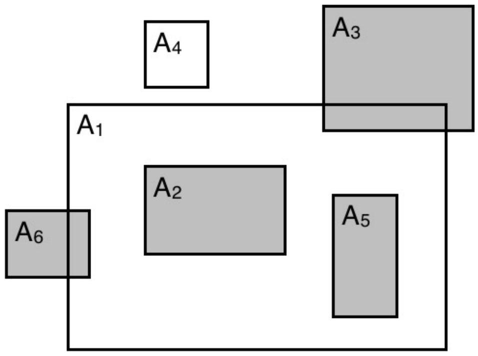

Figure 2: Computing $Pl(A_1)$.

Each of the functions Bel, Pl and m has a similar definition and they can be viewed as different representations of the same evidence. A belief measure is the strongest expression of the likelihood that an element x belongs to each A_i, whereas a plausibility measure is the weakest, as demostrated in Figures 1 and 2. The basic assignment expresses the likelihood that x belongs to each A_i with collected evidence. Despite the different representations only one of the three functions needs to be defined as the other two can be computed from any one of them using equations (10)–(13), (17) and (18). For instance, the basic assignment m can be computed using the belief function Bel:

$$m(A_i) = \sum_{A_j \subseteq A_i} (-1)^{|A_i \setminus A_j|} Bel(A_j), \tag{19}$$

where $|A_i \setminus A_j|$ is the cardinality of the difference between the two sets [13].

3.4 Combining evidence

In many situations, there are multiple sources of information available, for example multiple experts' opinions on the same event. These multiple sources of information need to be combined for further calculations and conclusions. However, there is no single universally accepted method for combining different experts' definitions for either belief, plausibility or basic assignment based on collected evidence [1].

Combination rules are a type of aggregation method, meaning they are used to summarise and simplify data coming from multiple sources. Some well-known aggregation methods include different averages and minimum and maximum values. In evidence theory, the multiple sources of information are assumed independent of each other meaning that observations made by one source do not affect the observations made by the other source. Additionally, an important feature for combination rules is associativity. Associativity means the ability to update an already combined structure with new information. [7]

Three of the simplest rules for combination of evidence are presented in this section. Examples about rules of combination are demonstrated in Section 5. The following assumes that there are two sources of information with belief assignment functions m_1 and m_2 and we define rules to combine them into an aggregate $m_{1,2}$. The combination of evidence from two different sources can be easily extended to the case with multiple sources when the rule of combination used is associative [7].

3.4.1 Dempster's rule of combination

Dempster's rule of combination is a generalisation of Bayes' rule and is formulated as follows:

$$m_{1,2}(A_i) = \frac{\sum_{A_j \cap A_k = A_i} m_1(A_j)m_2(A_k)}{1 - \sum_{A_j \cap A_k = \varnothing} m_1(A_j)m_2(A_k)}, \tag{20}$$

when $A_i \neq \varnothing$ and $m_{1,2}(A_i) = 0$, when $A_i = \varnothing$ [1]. The denominator is a normalisation factor and can be interpreted to represent the conflict among the evidence [7].

However, Dempster's rule does not take into account the reliability of the source nor other possibly relevant information [1]. For instance, different experts' estimates might not be equally reliable. Even though a reliability coefficient would be easy to include in the combination rule, there is no standard way of calculating its value [3].

Another problem with Dempster's rule is the requirement of $m_{1,2}(\varnothing) = 0$. This requirement implies that the combined opinion is also included in the accepted universal set, but the universal set might be incomplete, which means that $m_{1,2}(\varnothing) \neq 0$ should be allowed [1]. The normalisation provided by this requirement and the denominator's effect of ignoring conflict can lead to counter-intuitive results, as Zadeh pointed out in his articles [17] and [18].

3.4.2 Yager's rule of combination

Yager's rule of combination is a modification of Dempster's rule and it handles the contradiction caused by the denominator of equation (20). Yager's rule of combination introduces a ground probability mass assignment $q_{1,2}$, which differs from the basic assignment m by the normalisation factor and the mass assigned to the universal set. Instead of the equality in the normalisation factor (15), the ground probability mass assignment for the empty set is $q_{1,2}(\emptyset) \geq 0$. Yager's rule is defined by [14]:

$$q_{1,2}(A_i) = \sum_{A_j \cap A_k = A_i} m_1(A_j) m_2(A_k) \tag{21}$$

$$m_{1,2}(A_i) = q_{1,2}(A_i), \text{ when } A_i \neq \emptyset \text{ and } A_i \neq X \tag{22}$$

$$m_{1,2}(X) = q_{1,2}(X) + q_{1,2}(\emptyset). \tag{23}$$

Even though the ground probability mass assignment can be used to any number of pieces of evidence, it does not make Yager's rule associative.

3.4.3 Inagaki's rule of combination

Inagaki's rule of combination combines Yager's rule and Dempster's rule based on a combination parameter k [4]:

$$m_{1,2}(A_i) = [1 + kq_{1,2}(\emptyset)]q_{1,2}(A_i) \text{ when } A_i \neq \emptyset \text{ and } A_i \neq X \tag{24}$$

$$m_{1,2}(X) = [1 + kq_{1,2}(\emptyset)]q_{1,2}(X) + [1 + kq_{1,2}(\emptyset) - k]q_{1,2}(\emptyset) \tag{25}$$

$$m_{1,2}(\emptyset) = 0. \tag{26}$$

The combination parameter k is defined on the following range [1]:

$$0 \leq k \leq \frac{1}{1 - q_{1,2}(X) - q_{1,2}(\emptyset)}. \tag{27}$$

Inagaki's rule becomes Yager's rule, when $k = 0$, and Dempster's rule when $k = 1/(1 - q(\emptyset))$. This means that Inagaki's rule is associative only when it corresponds to Dempster's rule.

3.4.4 Other rules of combination

There are plenty of rules for combining evidence because none of them are both associative and intuitive. The rules are either not completely associative or can produce counter-intuitive or otherwise incomplete results. More about the different rules can be found in Sentz's report [7]. Additionally, Zadeh has pointed out important issues of some of the earlier rules in his articles [17] and [18].

4 Other frameworks

Evidence theory is only one possible framework for reasoning with uncertainty, but it can be used as a framework for formulating other theories as well. This section provides more information on some of the available frameworks for dealing with uncertainty and their relation to evidence theory.

4.1 Possibility theory

Classical possibility theory is based on monotone measures called possibility and necessity. A *possibility measure* is subadditive and describes alternatives in a set according to given evidence. These singletons in a subset E of the universal set X are possible and those outside E are not possible:

$$Pos_E(\{x\}) = \begin{cases} 1 & \text{when } x \in E \\ 0 & \text{when } x \in \bar{E} \end{cases} \tag{28}$$

for all $x \in X$ [1]. The theory of graded possibilities extends the possibility measure to fuzzy sets.

A possibility measure has very similar characteristics as the measures in evidence theory. Wang and Klir [13] defined possibility theory in the following way:

$$Pos : P_X \to [0, 1] \tag{29}$$

and that satisfies the following conditions:

$$Pos(\varnothing) = 0 \tag{30}$$

$$Pos(X) = 1 \tag{31}$$

$$Pos(A) = \sup_{x \in A} r(x) \text{ for any nonempty set } A \in P_X, \tag{32}$$

where $r(x)$ is a possibility distribution. The possibility distribution maps the universal set to the range $[0, 1]$ and is defined as the possibility of the singleton x: $r(x) = Pos(\{x\})$ [5]. Equations (29), (30) and (31) are the same as for belief and plausibility measures. Equation (32) describes the maxitive property of a possibility measure, where $\sup_{x \in A}$ is supremum: the least upper bound, i.e. the least element of X that is greater than or equal to all the elements in A [13]. Equations (31) and (32) also imply that possibility is normalised in the sense that $\exists\, x \in X$ such that $r(x) = 1$.

A *necessity measure* Nec is the dual measure of a possibility measure [1]:

$$Nec(A_i) = 1 - Pos(\bar{A}_i). \tag{33}$$

A necessity measure is defined similar to a possibility measure (equations (29)–(31)), but the third condition is [5]: for any family $\{A_i | A_i \in P_X, i \in I\}$, where I in an arbitrary index set,

$$Nec\left(\bigcap_{i \in I} A_i\right) = \inf_{i \in I} Nec(A_i). \tag{34}$$

Here $\inf_{i \in I}$ is infimum: the greatest lower bound and a dual of supremum [13].

Additionally, there is a link between the measures used in evidence theory and the measures used in possibility theory. When the universal set X is finite, the plausibility measure induced by a consonant basic assignment is a possibility measure and the belief measure is a necessity measure [13]. The basic assignment being consonant means that it focuses on a nest. Possibility theory can then also be interpreted as a special case of evidence theory where the subsets of X are required to be nested $A_1 \subset A_2 \subset \cdots \subset X$. In addition to classical possibility theory, there is a view of possibility theory which extends it to fuzzy sets [16].

4.2 Probability theory

The relationship between probability theory and evidence theory was described briefly in Section 2.3. This section introduces the basics and highlights the use of probability theory as a framework for dealing with uncertainty. Probability theory's relation to evidence theory is also discussed.

The first appearance of probability theory was in the 15th century but the correspondence between Pascal and Fermat in the 17th century is better known by mathematicians [10]. Probability theory has since been developed by many famous mathematicians. This means that probability theory is significantly older than evidence theory.

Classical probability is often defined as the number of occurrences of a certain event divided by the number of repetitions of the experiment, also called relative frequency:

$$P(X = x) = \frac{n}{N}, \tag{35}$$

where x represents the event of interest, X is a random variable describing for example the observations of the event, n is the number of observations resulting in the event of interest and N is the total number of observations [1]. The accuracy of the estimate of the probability of an event described by relative frequency increases with the number of repetitions. However, often it is not possible to make multiple repetitions, for example when studying the failure probability of a dam. This is why another definition of probability, the axiomatic definition, is often used. The probability of an event A of the universal set X is noted $P(A)$ and should satisfy the following conditions [1]:

$$P(A) \geq 0 \quad \forall A \in X \tag{36}$$

$$P(X) = 1 \tag{37}$$

$$P(A \cup B) = P(A) + P(B) \text{ for mutually exclusive events } A \text{ and } B. \tag{38}$$

These conditions also imply that the probability has also the following features:

$$P(\varnothing) = 0 \tag{39}$$

$$0 \leq P(A) \leq 1 \quad \forall A \in X. \tag{40}$$

In the case of estimating probabilities from repeating experiments, there is usually variation in the experiments themselves. Therefore, some uncertainty is involved. Uncertainty achieves an important role when making decisions based on these probabilities. Usually further assumptions are made in order to make the decision, but these assumptions are often made without any actual knowledge on the remaining system, but instead based on the properties of probability theory. Hence this method is suitable only for situations involving aleatory uncertainty.

Bayesian probabilities are tools for extending the use of probability theory to situations involving epistemic uncertainty. Commonly, engineering problems involve both objective and subjective information. Epistemic uncertainty is related to subjective information and is called prior knowledge. The combination of both types of information is called posterior knowledge. These two types of knowledge can be used for formulating the estimate for an event happening when there is certain prior knowledge of events:

$$P(E) = P(A_1)P(E|A_1) + P(A_2)P(E|A_2) + \cdots + P(A_n)P(E|A_n), \tag{41}$$

where $A_1, A_2, \ldots, A_n$ represent the subjective information and E represents the objective information [1]. This formulation is based on Bayes' theorem and $P(E|A_i)$ denotes the probability of event E occurring given A_i. This formulation can be used for computing the posterior probability as follows [1]:

$$P(A_i|E) = \frac{P(A_i)P(E|A_i)}{P(A_1)P(E|A_1) + P(A_2)P(E|A_2) + \cdots + P(A_n)P(E|A_n)}. \tag{42}$$

Even though this theorem can be used in situations involving epistemic uncertainty, a lot of prior knowledge is needed and that is often not available.

The biggest difference between probability theory and evidence theory is the use of sets instead of singletons in evidence theory [1]. This enables simpler reasoning with incomplete information. The equality in the additivity property is also removed in order to eliminate the need for further assumptions.

An important feature to note between evidence theory and probability theory is that the basic assignment in evidence theory maps the power set of X to the range $[0, 1]$, whereas probability theory's probability assignment is a mapping from X to the range $[0, 1]$. Moreover, the features presented by equations (37)–(39) are similar to the features of belief and plausibility measures. If only singletons are used, belief and plausibility measures collapse into a single measure which fulfils the conditions of a probability measure.

4.3 Imprecise probabilities and fuzzy sets

There are several other frameworks or implementations for reasoning with uncertainty and especially epistemic uncertainty. For example, theory of imprecise probabilities or allowing the use of fuzzy sets or measures can be used when modelling situations involving uncertainty [1].

The theory of imprecise probabilities is based on lower and upper probability assignments and was developed by Walley [11]. It is a generalisation of probability theory. Imprecise probabilities have many features similar to the measures used in evidence theory. For instance, the lower and upper probabilities form a duality and a Möbius representation can be used for imprecise probabilities [13]. More about imprecise probabilities can be read from Walley's work, for example article [12].

Fuzzy sets were developed by Zadeh and are useful in situations involving uncertainty as the membership in a fuzzy set is a matter of degree instead of a binary value [1]. Regarding crisp sets, each item either belongs or does not belong to a set as described by membership function values $\{0, 1\}$, whereas the membership function of a fuzzy set can take values on the interval $[0, 1]$. This means also that crisp sets are a special case of fuzzy sets. Fuzzy measures are defined with some of the properties of fuzzy sets [1]. When it is difficult to define whether an element belongs to a set or not, the fuzzy sets and measures become useful, as is often the case in real world problems. Fuzzy sets are described in more detail in Zadeh's article [15] and in several books by Klir, for example "Fuzzy sets and fuzzy logic" [6].

5 Examples and applications

As belief and plausibility measures characterise the body of evidence at hand, they can, in some applications, be interpreted as lower and upper limits on the strength of evidence at hand [1]. Belief and plausibility can be assigned to sets, but when there is enough information to assign them to singletons, their formulation collapses to the classical probabilistic formulation [7]. This way, belief can be used to develop a lower limit on probability of an event and plausibility to develop an upper limit [1]. In other words, belief and plausibility define a range of acceptable probabilities. This range can also be presented using only belief or plausibility due to their duality [13].

This section provides examples of assigning the values for belief and plausibility measures as well as examples that relate to real world problems. In addition to examples to help understand the measures, applications for evidence theory are presented. Both current applications and possible applications in the future are discussed.

5.1 Examples

5.1.1 Computing belief and plausibility

In this example, the location of the epicentre of a possible earthquake is estimated using belief and plausibility measures. There is data available on different experts' estimates about the location of the epicentre and it acts as the body of evidence. These estimates are used to estimate the likelihood that the epicentre is inside a certain area. A certain area could be for example, a densely populated area and in this case there are two areas of interest: A and B. The experts' estimates are both non-specific and conflicting with each other, as seen in Figure 3. Each E_i stands for one expert's estimate and there are estimates from a total of 15 different experts. [1]

In this case each expert is assumed to be equally credible and reliable [1]. This means that the weight of evidence assigned to each expert's estimate $E_i = 1/15$. With this information, it is possible to count values for belief and plausibility that the epicentre is inside an area of interest. For the value of belief, only estimates completely inside the area of interest are calculated, as seen in equation (17). For belief we get $Bel(A) = 2/15 = 0.13$ and $Bel(B) = 1/15 = 0.07$. As seen in equation (18), all estimates intersecting with the area of interest are used for calculating for plausibility: $Pl(A) = 5/15 = 0.33$ and $Pl(B) = 3/15 = 0.2$.

The values computed for belief and plausibility can then be used to construct estimates of respective probabilities. The intervals achieved for the probability of

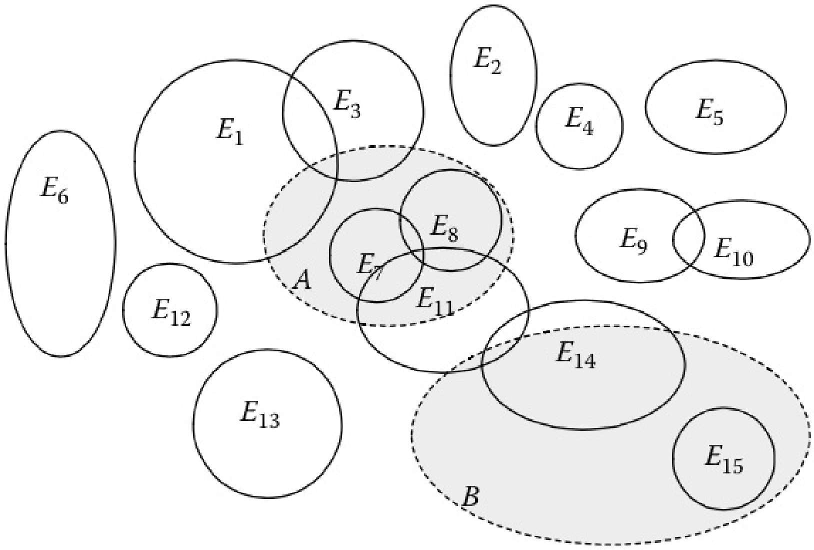

Figure 3: Expert's estimates of the locations of the epicentre [1].

the epicentre being in area A is $0.13 \leq P(A) \leq 0.33$ and for area B the interval is $0.07 \leq P(A) \leq 0.2$.

5.1.2 Using rules of combination

For the first example of using rules of combination, assume a case of target identification based on data from multiple sensors. In this case there are 100 possible target types denoted as the universal set $X = \{x_1, x_2, \ldots, x_{100}\}$. Sensor 1 can identify only 40% of these 100 target types and the source indicates that a target type belonging to this 40% has entered the observed area. Assume these 40% of target types are $A = \{x_1, x_2, \ldots, x_{40}\}$. Another sensor (sensor 2) can identify an additional 10 other target types, defining $B = \{x_1, x_2, \ldots, x_{50}\}$. Based on the evidence from sensors 1 and 2, values presented in Table 1 can be assigned. [1]

Using a rule of combination and the values presented in Table 1, a combined body of evidence can be computed. For simpler calculation, assume only two target types are involved, denoted x_1 and x_2. Suppose that sensor 1 provides a support of 0.6 that a sensed target is of type x_1 and sensor 2 provides a support of 0.95 that it is of type x_2. The combined body of evidence for the two different targets using Dempster's rule of combination (20) is shown in Table 2.

Calculating values for the combined body of evidence $m_{1,2}$ with Yager's rule of combination (21)–(23) very different values are achieved: $m_{1,2}(x_1) = 0.03$, $m_{1,2}(x_2) = 0.38$ and $m_{1,2}(X) = 0.59$ [13]. The difference is due to the large conflict between the support provided by the two sensors.

Table 1: Evidence for target identification [1].

Evidence	**Assignment** (m_i)	**Belief** (Bel_i)	**Plausibility** (Pl_i)	**Probability** (P_i)
(1) Sensor 1				
Event A	0.4	0.4	1	$[0.4, 1]$
$\bar{A}$		0	0.6	$[0, 0.6]$
Universal set X	0.6	1	1	$[1, 1]$
(2) Sensor 2				
Event B	0.5	0.5	1	$[0.5, 1]$
$\bar{B}$		0	0.5	$[0, 0.5]$
Universal set X	0.5	1	1	$[1, 1]$

Table 2: Combined evidence for target identification [1].

Evidence	Sensor 1 (m_1)	Sensor 2 (m_2)	($m_{1,2}$)	($Bel_{1,2}$)	($Pl_{1,2}$)	($P_{1,2}$)
Event $\{x_1\}$	0.6	0	0.07	0.07	0.12	$[0.07, 0.12]$
Event $\{x_2\}$	0	0.95	0.88	0.88	0.93	$[0.88, 0.93]$
Universal set X	0.6	0.05	0.05	1	1	$[1, 1]$

Another example of a rule of combination is about the counter-intuitive results Dempster's rule can give [7]. Suppose a situation of medical diagnosis for a patient. The patient is seen by two physicians who have different opinions on which medical condition the patient has based on the symptoms. The first doctor believes that is is very likely that the patient has meningitis and that there is a small chance it is a brain tumour instead, with probabilities of 0.99 and 0.01 respectively. The other doctor believes that the symptoms are a result of a concussion with a probability of 0.99, but there is a 0.01 probability of a brain tumour. Using Dempster's rule of combination, $m_{1,2}(\text{brain tumour}) = 1$, so based on equations (17) and (18) the combined probability of a brain tumour is 1, even though both doctors considered it very unlikely.

5.2 Current applications

Evidence theory has been used mainly as a tool in decision making when epistemic uncertainty is involved. It was introduced as an alternative approach to multi-criteria decision making and especially the analytical hierarchy process in an article by Beynon et al. [2]. Some of the current applications of evidence theory are with artificial intelligence and expert systems, especially as a technique for modelling reasoning under uncertainty [3].

Evidence theory has also been used in face recognition, statistical classification, target identification and other areas related to classification problems with at least some success [2]. There are applications in fields where cognitive aspects are related to the uncertainty involved. Some of these fields include biology and meteorology [3]. In addition to biology, also the medical field has had some success with applying evidence theory, for example in medical diagnosis [2].

5.3 Possible future applications

The applications of evidence theory related to computer systems might become more common in the future as the computational power continues to increase. The increasing computational power enables also applications in other fields, as more and more complex systems can be solved using mathematical tools.

A natural area for applying evidence theory would be among law and crimes, as the evidence involved there is not only on singletons. The theory could also be used in the field of biology for example for identification or classification of new species. According to Fioretti [3], social sciences is an obvious candidate for future applications, but the lack of mathematical expertise within that field is a limiting factor. Additionally, the medical field with its vast complexity probably offers multiple application subjects in the future.

The tools this theory provides could also be used for locating sunken ships or fallen aeroplanes with the help of satellite pictures. Basically, there are many suitable fields for applying evidence theory, especially in the fields involving decision making and evaluation made by humans.

6 Summary

The objective of this thesis was to provide a concise and comprehensible tutorial on evidence theory as a framework for reasoning with uncertainty. Despite the fairly big amount of literature related to the theory, there was a lack of a simple and compact introduction to it.

For historical context the creation of evidence theory and the time before it were introduced. Also, to help understand the motive for creating and using this theory, an explanation of its framework for modelling uncertainty was included. Additionally, there was a need for briefly explaining different types of uncertainty and the mathematical modelling involved to get the hang of the problem this theory was meant to solve.

One of the main parts was the mathematical formulation of the theory and its components. The concepts of belief and plausibility measures used for describing the body of evidence in a mathematical notation were introduced, as well as a Möbius representation for simpler calculations. The main aspect to be still improved – combining evidence from multiple sources – was also discussed. Additionally, examples for using the mathematical formulation for solving real-life problems and the issues with rules of combination were displayed.

Another important aspect of this thesis was to compare evidence theory to other frameworks created for modelling uncertainty. This was important because probability theory has been used for these problems for a long time and evidence theory has a straight connection to it. In addition to probability theory, both the connection to imprecise probabilities and fuzzy sets and evidence theory as a framework for formulating them was introduced. Also, a different yet very similar framework, possibility theory, was presented.

In addition to the development of rules of combination, the prospects of evidence theory were discussed via applications. Some of the current fields of application as well as possible application areas in the future were viewed. Increasing computational power was presented as the main reason for being able to apply evidence theory for more and more situations in the future.

Evidence theory is also subject matter of a number of conference series, for instance:

- UAI: Conference on Uncertainty in Artificial Intelligence
- IPMU: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems
- BELIEF: International Conference on Belief Functions.

References

[1] Bilal M Ayyub and George J Klir. *Uncertainty Modeling and Analysis in Engineering and the Sciences.* CRC Press, 2006.

[2] Malcolm Beynon, Bruce Curry, and Peter Morgan. The Dempster–Shafer theory of evidence: an alternative approach to multicriteria decision modelling. *Omega*, 28(1):37–50, 2000.

[3] Guido Fioretti. A mathematical theory of evidence for G.L.S. Shackle. *Mind & Society*, 2(1):77–98, 2001.

[4] Toshiyuki Inagaki. Interdependence between safety-control policy and multiple-sensor schemes via Dempster-Shafer theory. *IEEE Transactions on Reliability*, 40(2):182–188, 1991.

[5] George J Klir. Uncertainty in economics: the heritage of G.L.S. Shackle. *Fuzzy Economic Review*, 7(2):3–21, 2002.

[6] George J Klir and Bo Yuan. *Fuzzy Sets and Fuzzy Logic*, volume 4. Prentice Hall New Jersey, 1995.

[7] Kari Sentz and Scott Ferson. *Combination of Evidence in Dempster-Shafer Theory*, volume 4015. Sandia Reports, 2002.

[8] George L S Shackle. *Uncertainty in Economics and Other Reflections.* Cambridge University Press, 1955.

[9] Glenn Shafer. *A Mathematical Theory of Evidence*, volume 1. Princeton University Press, Princeton, 1976.

[10] Isaac Todhunter. *A History of the Mathematical Theory of Probability.* Cambridge University Press, 2014.

[11] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*, volume 42. Chapman and Hall/CRC, 1990.

[12] Peter Walley. Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning*, 24(2):125–148, 2000.

[13] Zhenyuan Wang and George J Klir. *Generalized Measure Theory*, volume 25. Springer Science & Business Media, 2010.

[14] Ronald R Yager. On the Dempster-Shafer framework and new combination rules. *Information Sciences*, 41(2):93–137, 1987.

[15] Lotfi A Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[16] Lotfi A Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1(1):3–28, 1978.

[17] Lotfi A Zadeh. Review of a mathematical theory of evidence. *AI Magazine*, 5(3):81, 1984.

[18] Lotfi A Zadeh. A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination. *AI Magazine*, 7(2):85, 1986.

A Yhteenveto (A summary in Finnish)

Monet tosielämän tilanteet sisältävät epävarmuutta. Esimerkiksi alkavan päivän säätä ei voida tietää varmasti. Meteorologit muodostavat ennusteensa statistiikkaan ja todennäköisyyksiin nojautumalla. Ne puolestaan nostavat tietonsa monista eri lähteistä. Monet uskovatkin todennäköisyysteorian olevan paras tai jopa ainoa keino tälläisten epävarmuutta sisältävien tilanteiden mallintamiseen. Todisteteoria (engl. evidence theory) tarjoaa toisen menetelmän epävarmuuden käsittelyyn.

Todisteteoriasta on melko paljon kirjallisuutta, mutta selkeä ja tiivis johdatus aiheeseen puuttuu. Monet teokset käsittelevät vain jotakin todisteteorian osa-aluetta, joten teoriasta kokonaisuudessan kertovia teoksia on vähän. Teoksissa vaihtelevat lisäksi matemaattisten ilmausten ja merkintöjen määrä ja laatu.

Tämä opinnäytetyö tarjoaa helposti ymmärrettävän ja ytimekkään katsauksen todisteteoriaan. Katsaukseen sisältyy teorian taustan lisäksi esittely sen matemaattisesta perustasta. Lisäksi ymmärtämisen helpottamiseksi se käsittää vertailua muihin vastaaviin menetelmiin sekä havainnollistavia esimerkkejä. Työ sisältää myös todisteteorian nykyisten ja mahdollisien tulevaisuuden sovelluskohteiden käsittelyä.

Todisteteoria tarjoaa apuvälineitä epävarmuuden käsittelylle. Se on kehitetty 1960- ja 1970-luvuilla, ja teoriaa kutsutaankin joskus kehittäjiinsä mukaan Dempsterin–Shaferin teoriaksi. Jo ennen todisteteoriaa muun muassa taloustieteilijä G.L.S Shackle oli kehittänyt vastaavanlaisen teorian, mutta formaalin matemaattisen muotoilun puutteen vuoksi matemaatikot eivät kiinnostuneet siitä.

Epävarmuutta on kahdenlaista: stokastista ja subjektiivista. Stokastinen epävarmuus on systeemin satunnaisuutta, kun taas subjektiivinen epävarmuus johtuu tiedon puutteesta. Tiedon puute liittyy usein tilanteisiin, joissa saatavilla oleva tieto perustuu esimerkiksi asiantuntijan lausuntoon. Todisteteoria kehitettiin avuksi erityisesti tilanteisiin, joissa esiintyy paljon subjektiivista epävarmuutta. Todennäköisyysteoriaa on käytetty molempien epävarmuustyyppien mallintamiseen, mutta edes subjektiivisen epävarmuuden huomioon ottava bayesiläinen todennäköisyysteoria ei sovellu siihen kovin hyvin.

Todisteteoriassa 'todennäköisyyksiä' asetetaan joukoille yksittäisten alkioiden sijaan. Nämä joukot voivat olla minkä tahansa kokoisia, eikä tietoa kaikista alkioista tarvita. Näitä asetettavia arvoja kutsutaan uskomuksen asteiksi, ja ne pohjautuvat käsillä oleviin todisteisiin. Yksi todisteteorian vahvuuksista onkin toimiminen epätäydellisellä tai puutteellisella informaatiolla niin, että edes lisäoletuksia ei tarvita.

Matemaattiset mitat nimeltä uskomus (engl. belief) ja uskottavuus (engl. plausibility) muodostavat perustan todisteteorialle. Ne voivat saada arvoja nollan ja yhden väliltä ja toimivat kuvauksina todisteiden vahvuudesta. Myös erästä Möbiuksen kuvausta voidaan käyttää vastaavasti. Nämä kolme funktiota kertovat samoista todisteista hieman eri tavalla, ja muut kaksi voidaan laskea, kun yksi tunnetaan. Kaikilla kolmella funktiolla on samoja ominaisuuksia, ja ne vastaavat osittain todennäköisyysteoriassa käytettyä todennäköisyyden mittaa. Tarkasteltavan joukon uskomuksen mitta voidaan määritellä kaikkina sen sisältävinä joukkoina, kun taas uskottavuuden mittaan lasketaan myös osittain ulkopuolelle jäävät joukot.

Yksi todisteteorian tärkeistä osista on todisteiden yhdistely. Monissa tilanteissa tarjolla on useammasta eri lähteestä tietoa, joka koskee samaa asiaa. Tällöin uskomuksia tai uskottavuuksia tarvitsee yhdistellä kokonaiskuvan aikaansaamikseksi. Ei kuitenkaan ole olemassa yhtä yleismaailmallisesti hyväksyttyä sääntöä yhdistelylle. Yhdistelymenetelmälle tärkeä ominaisuus olisi sen kyky päivittää olemassa olevaa tulosta uudella informaatiolla. Dempsterin sääntö yhdistelylle on ensimmäinen kehitetty menetelmä ja sillä on tämä ominaisuus. Dempsterin sääntöön liittyy kuitenkin muita ongelmia, kuten joskus järjenvastaiset tulokset. Nämä muut ongelmat on helppo korjata pienillä muokkauksilla, mutta samalla edellä mainittu ominaisuus menetetään. Tutkimusta eri yhdistelykeinojen heikkouksista on tehty jonkin verran ja lisäksi on kehitetty paljon uusia menetelmiä. Tästä huolimatta ei ole onnistuttu luomaan erittäin hyvää sääntöä todisteiden yhdistelylle.

Muitakin malleja epävarmuuden kuvaamiselle on kehitetty. Näistä tunnetuin on todennäköisyysteoria, mutta se ei muun muassa informaatiovaatimuksiensa puolesta sovi subjektiiviselle epävarmuudelle kovin hyvin. Todisteteorian ja todennäköisyysteorian välillä on kuitenkin yhteys: mikäli uskomukset ja uskottavuudet määrätään yksittäisille alkioille joukkojen sijaan, ne supistuvat yhdeksi arvoksi, joka vastaa todennäköisyysmittaa. Erityisesti subjektiiviselle epävarmuudelle on kehitetty myös mahdollisuusteoria (engl. possibility theory), joka onkin ominaisuuksiltaan hyvin samanlainen kuin todisteteoria. Lisäksi tietyillä ehdoilla mahdollisuusteoriassa käytetyt mitat ja todisteteorian mitat vastaavat täysin toisiaan.

Todisteteorialla on monia sovelluskohteita. Sitä on käytetty pääasiassa päätöksenteon apuvälineenä tilanteissa, joissa on mukana subjetiivista epävarmuutta. Sitä on sovellettu esimerkiksi monikriteerisessä päätöksenteossa ja erityisesti analyyttisessä hierarkiaprosessissa apuvälineenä. Muita onnistuneita sovelluskohteita löytyy tietokoneohjelmista, etenkin tekoälyn kehityksestä. Todisteteoriaa on sovellettu myös osin menestyksekkäästi luokitteluun, esimerkiksi kasvon- ja hahmontunnistukseen sekä tilastolliseen luokitteluun.

Sovelluksia todisteorialle löytyy käytännössä kaikilta aloilta, joihin liittyy ihmisten tekemiä analyysejä. Tällaisia aloja ovat muun muassa biologia ja meteorologia. Lisäksi lääketiede on hyvin todisteteorian työkaluille soveltuva ala, ja teoriaa onkin jo sovellettu muun muassa taudinmäärityksessä. Lääketieteestä löytyy varmasti sopivia sovelluskohteita myös jatkossa.

Tulevaisuudessa todisteteorialle löytyy todennäköisesti paljon lisää sovelluskohteita. Tietokoneiden alati kasvava laskentateho mahdollistaa yhä monimutkaisempien systeemien mallintamisen matemaattisin keinoin. Tällöin myös todisteteoriaa on mahdollista käyttää yhä suurempien systeemien mallintamiseen. Laskentatehon kasvusta huolimatta esimerkiksi lääketieteen ongelmia on jatkossakin kovin työlästä mallintaa muilla epävarmuutta mallintavilla menetelmillä. Luokittelun ja biologian sovellukohteista voidaan johtaa yhteys esimerkiksi uusien eliölajien luokitteluun. Muita suuria tietomääriä sisältäviä sovelluskohteita voisivat olla myös uponneiden alusten paikannus satelliitikuvilla.

Laki ja rikokset ovat luonnollisia sovellusaloja todisteteorialle, sillä niissä usein todisteet liittyvät suuriin kokonaisuuksiin. Myös yhteiskuntatieteistä löytyisi sovelluskohteita ihmisläheisyyden takia, mutta matemaattisen ammattitaidon puute alalla on rajoittava tekijä. Muiltakin aloilta, erityisesti päätöksentekoon liittyviltä ja ihmisten tekemiä arvioita sisältävilä aloilta, löytyy sovelluskohteita. Todisteteoriaa voi varmasti soveltaa näiden lisäksi muillakin tieteenaloilla.

Todisteteoria soveltuu moniin epävarmuutta sisältäviin tilanteisiin, joissa on tähän asti käytetty todennäköisyysteoriaa apuvälineenä tai joihin ei ole ollut mahdollista soveltaa matemaattista menetelmää ollenkaan. Todisteteoriassa on myös vielä kehitettävää, erityisesti yhdistelysäännöissä. Mikäli tutkimuksissa tähän asti löydetyt asiat osataan ottaa huomioon, teoria soveltuu kuitenkin hyvin käyttöön jo nykyään.