Linguistic probabilities: theory and application

Halliwell, Joe; Shen, Qiang

Published in:
Soft Computing

DOI:
10.1007/s00500-008-0304-1

Publication date:
2009

Citation for published version (APA):
Halliwell, J., & Shen, Q. (2009). Linguistic probabilities: theory and application. *Soft Computing*, *13*(2), 169-183.
https://doi.org/10.1007/s00500-008-0304-1

myjournal manuscript No.
(will be inserted by the editor)

Linguistic Probabilities: Theory and Application

Joe Halliwell[1], Qiang Shen[2]

[1] School of Informatics, University of Edinburgh, UK
[2] Department of Computer Science, University of Wales, Aberystwyth, UK

Received: date / Revised version: date

Abstract Over the past two decades a number of different approaches to "fuzzy probabilities" have been presented. The use of the same term masks fundamental differences. This paper surveys these different theories, contrasting and relating them to one another. Problems with these existing approaches are noted and a theory of "linguistic probabilities" is developed, which seeks to retain the underlying insights of existing work whilst remedying its technical defects. It is shown how the axiomatic theory of linguistic probabilities can be used to develop linguistic Bayesian networks which have a wide range of practical applications. To illustrate this a detailed and realistic example in the domain of forensic statistics is presented.

1 Introduction

It is widely acknowledged that devising computational solutions to real world problems typically demands mechanisms for representing uncertain information and reasoning with it. So it is that researchers have identified many different types of uncertainty and proposed corresponding treatments. The oldest and arguably most thoroughly researched of these schemes is that of probability theory, which models uncertainty about which propositions are, were or, in the archetypal case, will be true. Despite the practical and philosophical differences between its practitioners[1], classical probability theory is invariably commited to the position that its descriptions of the world have well-defined and exact truth conditions. But this simplifying assumption sits rather uneasily with our everyday concepts. As a familiar quote from Bertrand Russell (once a notorious proponent of logical positivism which is essentially the opposite view) puts it: "All traditional logic [and by extension classical probability theory] habitually assumes that precise symbols are being employed. It is therefore not applicable to this terrestrial life but only to an imagined celestial existence." (Russell, 1923)

This painful mismatch between the idealised model of classical logic and ordinary concepts has perhaps been felt most acutely in Artificial Intelligence where the programme of reproducing human behaviour in terms of the latter[2] has often collided with the limitations of the former. And it is this, of course, that explains the not-so-paradoxical success of Fuzzy Logic.

Historically, the relationship between fuzziness and probability has been somewhat troubled. Increasingly however, the apologetic (or inclusive) stance adopted by the fuzzy community seems to have succeeded in building a consensus that the theories are not in competition. Indeed it is now common to view them as orthogonal – each quite valid on its own terms. This recognition has, in turn, catalysed a broader acceptance of hybrid approaches. This paper focuses on one family of such hybrids which use fuzzy numbers to generalise the unit interval quantity space of classical probability theory.

The remainder of this paper is structured as follows. The following section examines the motivations behind the attempts to develop such a theory, collecting together work from psychometric researchers sympathetic to a fuzzy approach. Section 3 introduces requisite background material. Section 4 then scrutinizes the two quite distinct theories that share the name "fuzzy probability". They are found to . Building on earlier work (Halliwell and Shen, 2002), Section 5 attempts to address these problems developing a new theory which the authors term "linguistic probability" to distinguish it from previous efforts. A demonstration of the approach's practical utility is then presented. Section 6 shows how the theory may be used within a Bayesian network, and,

[1] In particular, the interpretation of probability theory remains highly contentious, although subjectivist accounts have predominated in recent years.

[2] This broad programme includes modelling expert knowledge and transparent or symbolic machine learning.

following the work presented in Halliwell et al. (2003), Section 7 develops a detailed and realistic example in the domain of forensic science, contrasting the new approach with a similar classical network. Finally, Section 8 concludes the paper and identifies areas for future work.

2 Motivations

The use of fuzzy sets to model every day descriptions such as "John is tall" is no doubt familiar. The idea behind the set of theories sharing the name "fuzzy probability" is that imprecise linguistic characterizations of probabilistic uncertainty can be treated in an analogous way. The goal then, put simply, is to develop a principled approach to statements such as

$$\text{It is quite likely to rain tomorrow.} \tag{1}$$

A possible objection at this stage is that (1) is hopelessly uninformative. If (probabilistic) information about the next day's weather is crucial to a system's successful operation there are surely better ways to obtain it. In short, why bother attempting to utilize such woefully low-grade information? The answer, of course, is the standard argument for "computing with words" (Zadeh, 1996): whilst gold-standard numerical information may be available about tomorrow's weather, there are probabilistic assessments which are too difficult, expensive or simply impossible to obtain with such precision.

For example, consider the questions: Will there be artificial intelligence in 10 years? 100 years? 1000 years? Consultation with an expert is unlikely to yield much beyond vague probabilistic statments like "It is extremely unlikely that we will have (true) artificial intelligence in ten years time." But if such information is to be used within the framework of classical probability theory, numerical estimates of the probabilities of interest are required.

In these cases the difficulty of obtaining point estimates of probability has been widely reported (Kahneman et al., 1985; Zimmer, 1983). Whilst an expert may be willing to assert that it is extremely likely that there will be intelligent constructs this time next millennium it would seem odd, a loss of academic integrity even, to state that the probability of that occurrence is 0.93. Indeed, a committee of the U.S. National Research Council (National Research Council Governing Board Commitee on the Assessment of Risk, 1981; Wallsten et al., 1986) has written that there is "an important responsibility not to use numbers, which convey the impression of precision, when the understanding of relationships is indeed less secure. Thus whilst quantitative risk assessment facilitates comparison, such comparison may be illusory or misleading if the use of precise numbers is unjustified." Subjective probability assessments are often the product of countless barely articulate intuitions and are often best expressed in words. It is misleading to seek to express them as precise numbers.

2.1 Psychometric studies

Responding to these difficulties, researchers attempted to obtain point values for probabilistic terms experimentally. The general form of these investigations is to present subjects with probabilistic terms requesting a numerical translation. It is hardly surprising that studies such as Budescu and Wallsten (1985) have concluded that point estimates of probability terms vary too greatly between subjects and exhibit too great an overlap to be useful for many problems.

Attempts to model probabilistic terms using fuzzy sets, however, have proven more successful. For example, a relatively sophisticated experimental method for eliciting fuzzy models of probabilistic terms has been developed by Wallsten et al. (1986) and the inter-subjective stability of generated terms has been examined with promising results. In addition, Zimmer (1986) has reported that verbal expressions of probabilistic uncertainty were "more accurate" than numerical values in estimating the frequency of multiple attributes by experimental studies. Whilst there are outstanding problems such as context sensitivity with the fuzzy approach to modelling probabilistic terms, these psychometric studies are unanimous in preferring it to numerical estimates.

For these and more obvious introspective reasons it is desirable to develop an account of vagueness in probabilities. To do so, however, will require a reasonably solid grasp of both classical probability theory and the mathematics of fuzzy numbers. These are rehearsed in the following Section.

3 Background

This section introduces the background material that will be required for the remainder of this paper. Although it is anticipated that many of the concepts will be familiar to the reader, it was felt that it would be best to determine these unambiguously and introduce the associated notation in a single place. The results mentioned, especially those relating to the algebra of fuzzy numbers are essential to the proofs in the following Sections.

3.1 Classical probability theory

The predominant formalisation of probability theory is that provided by Kolmogorov. These standard definitions may be found in any introductory text on probability theory e.g. Grimmet and Welsh (1986). Given an experiment or trial, such as rolling a die, the set of all possible outcomes or *sample space* will be denoted Ω. So, in the die example $\Omega = \{1, 2, 3, 4, 5, 6\}$. Clearly, various questions may be asked about the outcome of a trial. Some of these will be elementary, of the form "Was the outcome ω?", but others will be about groups of states. Returning to the die example, one might enquire "Was the outcome an odd number?" Moreover, it is often convenient to specify the probability of propositions modelled as such groups of atomic outcomes. The notion of an event space is used to capture the idea that the relevant propositions should be closed under logical operators.

Definition 1 (Event space)**.** *A set $\mathcal{E}$ is termed an event space on a set Ω of possible outcomes if and only if*

a) $\mathcal{E} \subseteq \mathbb{P}(\Omega)$
b) $\mathcal{E}$ *is non-empty.*
c) If $A \in \mathcal{E}$ *then* $A^c = \Omega \setminus A \in \mathcal{E}$
d) If $A_1, A_2, \ldots \in \mathcal{E}$ *then* $\bigcup_{i=1}^{\infty} A_i \in \mathcal{E}$

Events spaces are sometimes also referred to as "sigma algebras" and are said to be closed under complementation and countable union. Observe that since $\mathcal{E}$ is non-empty, there is some $A \in \mathcal{E}$, thus both $\Omega = A \cup A^c$ and $\emptyset = \Omega^c$ are elements of $\mathcal{E}$. With the notion of an event space in place it is possible to define the central concept of a probability measure.

Definition 2 (Classical probability measure)**.** *A mapping* $\mathrm{P} : \mathcal{E} \to \mathbb{R}$ *is termed a probability measure on* $(\Omega, \mathcal{E})$ *if and only if for all* $A \in \mathcal{E}$

(CP1) $\mathrm{P}(A) \geq 0$
(CP2) $\mathrm{P}(\Omega) = 1$
(CP3) If $A_1, A_2, \ldots \in \mathcal{E}$ *are disjoint (i.e.* $A_i \cap A_j = \emptyset$*) then*

$$\mathrm{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathrm{P}(A_i) \tag{2}$$

Where P *is such a probability measure, the tuple* $(\Omega, \mathcal{E}, \mathrm{P})$ *is termed a probability space.*

Note that CP3 implies the existence of $\sum_{i=1}^{\infty} \mathrm{P}(A_i)$.

3.2 Fuzzy sets and numbers

The membership function of a fuzzy set, a, on some universe of discourse, D, will be denoted $\mu_a : D \to [0, 1]$.

Fuzzy numbers, then, are simply fuzzy sets of real numbers whose membership functions have the right sort of shape. The *level set* at $\alpha \in [0, 1]$ of a fuzzy set of numbers, a, will be denoted $\mathrm{L}_\alpha(a)$ i.e.

$$\mathrm{L}_\alpha(a) = \begin{cases} cl(\{x \in \mathbb{R} \ : \ \mu_(x) > 0\}) & \text{if } \alpha = 0 \\ \{x \in \mathbb{R} \ : \ \mu_a(x) \geq \alpha\} & \text{otherwise} \end{cases} \tag{3}$$

$\mathrm{L}_0(a)$ is termed the *support* of a.

Definition 3 (Fuzzy number)**.** *A fuzzy number is a fuzzy set of real numbers, a, which*

a) is normal, i.e. $\exists x \in \mathbb{R}$ *such that* $\mu_a(x) = 1$*;*
b) is convex, i.e. $\forall x, y, z \in \mathbb{R}$ *if* $x \leq y \leq z$ *then* $\mu_a(y) \geq \min(\mu_a(x), \mu_a(z))$*; and*
c) has an upper semi-continuous membership function
d) has a bounded support

Note that this definition also covers what might be termed "fuzzy intervals". It can be shown that these conditions entail that the level sets of a are closed, bounded intervals i.e.

$$\mathrm{L}_\alpha(a) = [\underline{\mathrm{L}}_\alpha(a), \overline{\mathrm{L}}_\alpha(a)] \tag{4}$$

Examples of fuzzy numbers can be found in Figure 2 which defines the linguistic probabilities that will be used in the worked example. The set of all fuzzy numbers are termed the "fuzzy reals" and denoted, $\mathbb{E}$. Embedded real numbers are denoted by a χ subscript. For example the membership function of 1_χ (the embedding of 1) is given by

$$\mu_{1_\chi}(x) = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

A similar notation will also be adopted for embedded intervals. So, for example,

$$\mu_{[0,1]_\chi}(x) = \begin{cases} 1 & \text{if } x \in [0, 1] \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

It is well known that a complete metric on the space of fuzzy numbers is given by the extended Hausdorff metric

$$d_\infty(a, b) = \sup_{\alpha \in [0,1]} \max(|\underline{\mathrm{L}}_\alpha(a) - \underline{\mathrm{L}}_\alpha(b)|, |\overline{\mathrm{L}}_\alpha(a) - \overline{\mathrm{L}}_\alpha(b)|) \tag{7}$$

Note that this metric coincides with the standard Euclidean metric for embedded real numbers. Thus proofs from real analysis immediately carry over to embedded reals.

3.2.1 The Extension Principle The Extension Principle identifies a natural way to extend maps between classical sets to maps on fuzzy sets defined over them (as a universe of discourse).

Definition 4 (Extension Principle)**.** *Given a map,*

$$f : A_1 \times A_2 \times \ldots \times A_n \to B \tag{8}$$

the natural fuzzy extension, $\tilde{f}$*, is the map determined by:*

$$\mu_{\tilde{f}(a_1,a_2,\ldots,a_n)}(y) = \sup_{f(x_1,x_2,\ldots,x_n)=y} \min(\mu_{a_1}(x_1), \mu_{a_2}(x_2), \ldots \mu_{a_n}(x_n))$$

for all fuzzy sets $a_1, a_2, \ldots, a_n$ *defined on* $A_1, A_2, \ldots, A_n$ *respectively.*

In other words, the possibility of a particular element being in the image of a fuzzy set under an extended function is the maximum of the membership values of elements mapped to it by the original function.

3.2.2 Partial orderings On the real numbers $a \leq b$ if and only if $a = \min(a, b)$. Following this observation, the Extension Principle may be used to induce a natural partial order, $\preccurlyeq$, on the fuzzy reals as follows:

$$\begin{aligned} a \preccurlyeq b &\iff a = \widetilde{\min}(a, b) \\ &\iff \mu_a(z) = \sup_{\min(x,y)=z} \min(\mu_a(x), \mu_b(y)) \; \forall z \in \mathbb{R} \end{aligned}$$

Since the set of fuzzy numbers whose membership functions are zero outside some given real interval $[a, b]$ can be characterised as

$$\{x \in \mathbb{E} \;:\; a_\chi \preccurlyeq x \wedge x \preccurlyeq b_\chi\} \tag{9}$$

it is natural to denote such an interval of fuzzy numbers by $[a_\chi, b_\chi]$.

Another partial order on the fuzzy reals is generated by the fuzzy subset relation i.e. for all $a, b \in \mathbb{E}$

$$a \subseteq b \iff \forall x \in \mathbb{R} \; \mu_a(x) \leq \mu_b(x) \tag{10}$$

Because it is somewhat confusing to talk of one fuzzy number being a superset of another, we prefer to say that the former *subsumes* the latter. One number subsumes another if it is in effect a kind of less precise version of it.

3.2.3 Arithmetical operators The Extension Principle may also be used to define fuzzy counterparts to the standard arithmetic operators of addition, multiplication, subtraction and division. If the standard arithmetic operators are considered as maps from $\mathbb{R}^2 \to \mathbb{R}$ the straightforward application of the principle yields, for example,

$$\mu_{a \oplus b}(z) = \sup_{x+y=z} \min(\mu_a(x), \mu_b(y)) \tag{11}$$

As is conventional, the extension of a real arithmetic operator will be denoted by circling its usual symbol. In the context of the fuzzy numbers it is also possible to derive these operators by examining the effects of performing interval-based calculations at each level set.

3.2.4 Algebraic properties The basic arithmetical operators described above are, of course, closed with respect to the fuzzy reals. As with their classical analogues, fuzzy addition and multiplication are commutative and associative with identities 0_χ and 1χ respectively. There are however, some important differences between classical and fuzzy arithmetic. First, fuzzy numbers are not distributive in the classical sense. Instead they are *subdistributive* i.e. for all $a, b, c \in \mathbb{E}$,

$$a \otimes (b \oplus c) \subseteq (a \otimes b) \oplus (a \otimes c) \tag{12}$$

Note however, that where the quantities involved are strictly positive (or negative) full distributivity is retained.

Second, in general, fuzzy numbers have neither additive nor multiplicative inverses, although there are (non-unique) pseudo-inverses. In particular, for all $a \in \mathbb{E}$, $0_\chi \subseteq a \oplus (0_\chi \ominus a)$ and with the usual cautions about 0, $1_\chi \subseteq a \otimes (1_\chi \oslash a)$.

3.2.5 Operators and subsumption A key property of the subsumption ordering, that has not been widely observed, is that it "carries over" any extended operator in the sense of the following Lemma.

Lemma 1. *Given an operator* $* : \mathbb{R}^n \to \mathbb{R}$ *and fuzzy numbers,* $a_1, a_2, \ldots\ a_n, b_1, b_2, \ldots\ b_n$ *such that* $a_i \subseteq b_i$ *for all* $1 \leq i \leq n$ *then*

$$\circledast(a_1, a_2 \ldots a_n) \subseteq \circledast(b_1, b_2 \ldots b_n) \tag{13}$$

Proof: *By the Extension Principle*

$$\begin{aligned} \circledast(a_1, a_2 \ldots a_n)(x) &= \sup_{x=*(x_1,x_2\ldots x_n)} \{\min_{1\leq i\leq n} a_i(x_i)\} \\ &\leq \sup_{x=*(x_1,x_2\ldots x_n)} \{\min_{1\leq i\leq n} b_i(x_i)\} \\ &= \circledast(b_1, b_2 \ldots b_n)(x) \end{aligned}$$

as required.

This result allows complex calculations (such as the Bayesian sum of products expression for joint probability distribution considered later) to be rearranged and computed from partial results just as in the classical case. Note finally that for all $a, b \in \mathbb{E}$, $a \cap b$ is subsumed by both a and b.

3.2.6 A note on the computability of fuzzy numbers It has often been observed that commonly used classes of fuzzy number are not closed under the standard arithmetic operators. So, for example, the product of two polygonal fuzzy numbers is not polygonal. This has lead some to conclude that it is not possible to have a correct and computationally tractable calculus of fuzzy numbers.

If however, fuzzy numbers are represented by a pair of finite series of finite polynomial shoulder functions determining the upper and lower boundaries of their alphacuts, then their arithmetic combinations also fall into this class and can be computed exactly with relative ease. As a simple example, consider the fuzzy number, a, determined by the membership function,

$$\mu_a(x) = \begin{cases} x & \text{if } x \in [0, 1] \\ 2 - x & \text{if } x \in (1, 2] \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

Then a has an equivalent representation as

$$a_{\downarrow\alpha} = [\alpha, 2 - \alpha] \tag{15}$$

Now, $a \otimes a$ has a membership function which can be obtained through solving quadratic equations and paying careful attention to boundaries

$$\mu_{a\otimes a}(x) = \begin{cases} \sqrt{x} & \text{if } x \in [0,1] \\ 2-\sqrt{x} & \text{if } x \in (1,4] \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

But this is hardly easy to represent and performing further computations with it will be increasingly complicated. On the other hand, the alphacut representation is trivially calculated as

$$(a \otimes a)_{\downharpoonright\alpha} = [\alpha^2, \alpha^2 - 4\alpha + 4] \tag{17}$$

Further computations can be performed with similar ease. Naturally, "zero-crossing" fuzzy numbers (numbers whose membership at 0 is non-zero) require some caution and introduce precision errors (as it becomes necessary to "split" the polynomials at their roots).

4 Fuzzy probabilities

Two distinct theories of "fuzzy probabilities" have been presented in the literature, one due to Lotfi Zadeh and the other due to Jain and Agogino. This Section introduces the two theories and subjects them to a critical evaluation.

4.1 Zadeh's fuzzy probabilities

The term "fuzzy probability" first appears, albeit incidentally, in the second part of Lotfi Zadeh's seminal paper on linguistic variables (Zadeh, 1975). Here, Zadeh mentions that fuzzy quantifiers, which are introduced to capture the sense of vague quantifiers such as "most" or "few" can be thought of as being like fuzzy probabilities. The idea then appears to have lain dormant for nearly ten years.

Indeed, it was not until the mid-eighties that Zadeh revisited the idea. In "Fuzzy probabilities" (Zadeh, 1984) he began to develop this loose concept more formally, combining it with some of his earliest work in the area concerning the probabilities of fuzzy events.

Zadeh defines the (fuzzy) probability of a fuzzy event in terms of it's fuzzy cardinality with respect to some universe of discourse. This fuzzy cardinality in turn is defined in terms of a fuzzy set's $\sum$ representation.

Definition 5 ($\sum$ representation of a fuzzy set)**.** *The* $\sum$ representation *of a fuzzy set, A, defined over a universe of discourse, D, presents it as a sum of memberships:*

$$A = \sum_{d\in D} \frac{\mu_A(d)}{d}$$

So, for example, one might represent the fuzzy set RED in the universe of discourse

$$\{brick, salmon, peach, rose, steak, violet\}$$

as follows:

$$RED = \frac{0.7}{brick} + \frac{0.3}{salmon} + \frac{0.3}{peach} + \frac{1}{rose} + \frac{1}{steak} + \frac{0}{violet}$$

Definition 6 (Zadeh's fuzzy cardinality)**.** *Given D, a universe of discourse and A, a fuzzy set defined over D, the* fuzzy cardinality *of A,*

$$\text{FGCount}(A) = \sum_{\alpha} \frac{\alpha}{|A_{\downharpoonright\alpha}|}$$

where it is understood that "any gap in FGCount(A) may be filled by a lower count with the same α".

Returning to the example, and filling in[3] at 0, 1 and 4, the fuzzy cardinality of RED is thus,

$$\text{FGCount}(RED) = \frac{0}{0} + \frac{0}{1} + \frac{1}{2} + \frac{0.7}{3} + \frac{0.3}{4} + \frac{0.3}{5} + \frac{0}{6}$$

It is, of course, possible to fill out the notion of "filling in", instead defining the fuzzy cardinality of A as

$$\text{FGCount}(A)_{\downharpoonright\alpha} = [|A_{\downharpoonright 1}|, |A_{\downharpoonright\alpha}|]$$

Or, equivalently,

$$\mu_{\text{FGCount}(A)}(z) = \max\left\{ \max_{x\in X}(\mu_A(x)) \ : \ X \in \mathbb{P}(D) \wedge |X| \le z \right\}$$

This formulation has the advantage that it is a fuzzy real number as defined in Section 3.

Definition 7 (Zadeh's fuzzy probability)**.** *Given D, a universe of discourse and A, a fuzzy set defined on that universe, the fuzzy probability of A,*

$$FProb(A) = \frac{FGCount(A)}{|D|_{\chi}}$$

A possible application of fuzzy probabilities in this sense is given in the following example derived from Zadeh (1984).

Scenario 1. *Albert is trying to decide whether to insure his car, c. He has a database of cars, D at his disposal containing a subset S of stolen cars and a fuzzy similarity relation, ∼, on the set of database attributes. What can he conclude about his car's chances of being stolen?*

[3] Note that Zadeh's graphs make it clear that the filling in should not reach back beyond the lowest count i.e. in this case 2.

On Zadeh's account this is reduced to the question "What is the (fuzzy) probability of a car *like* Albert's being stolen?", which is then calculated as the probability that a car is both like c and has been stolen i.e.

$$\mathrm{FProb}(S \cap \widetilde{c}) = \frac{\mathrm{FGCount}(S \cap \widetilde{c})}{|D|}$$

where $\mu_{\widetilde{c}}(d) = \mu_{\sim}(c, d)$.

Whilst Zadeh's approach may be useful for such data-centered applications[4], from the point of view of the probability theorist this approach is somewhat dubious as it seems to rest on the assumption that the set of outcomes consists of a finite number of equiprobable elements. This limitation aside, on Zadeh's theory fuzziness in a probability is secondary, merely a reflection of the primary fuzziness of the event of interest itself.

Nevertheless, it is clear that fuzzy uncertainty in a probability assessment need not derive from fuzziness in an event of interest. For example, when a teacher reassures a panicking student that it is extremely unlikely he will fail an exam, the vagueness in that description of the chance of disaster does not arise as a consequence of vagueness in the concept of what constitutes a fail. In this case the model provided by Zadeh's fuzzy probabilities is clearly inappropriate.

4.2 Bayesian fuzzy probabilities

Although survey papers have tended to conflate the two, it is exactly this point that distinguishes the different approach to "fuzzy probabilities" taken by Jain and Agogino (1990). Arguably this paper has been the most influential publication in the area, however it will be demonstrated that for technical reasons the theory it presents cannot provide a satisfactory model for qualitative probability assessments.

Jain and Agogino call their version of fuzzy probabilities "Bayesian fuzzy probabilities". The presentation here differs substantially from the original formulation of these ideas. In particular, Jain and Agogino do not explicitly use the idea of a probability measure either to place FP() in the context of an event space or to define the "mean" function m.

Definition 8 (Bayesian fuzzy probability measure)**.** *Given an event algebra, $\mathcal{E}$, defined over a set of outcomes, Ω, a function* FP() *from the set of events to the set of "convex normalized fuzzy set[s] ... of* $[0,1]$*" is a* Bayesian fuzzy probability measure *if and only if for all $A, B \in \mathcal{E}$*

(BF1) $\mathrm{FP}(A)$ *has a unique "mean" i.e. there is a function $m : \mathcal{E} \to [0,1]$ such that for all $x \in [0,1]$, $\mu_{\mathrm{FP}(A)}(x) = 1$ if and only if $x = m(A)$ (in this case* $\mathrm{FP}(()A)$ *is said to be* unimodal*)*

(BF2) $\mu_{\mathrm{FP}(A)}$ *is continuous on (0,1)*

(BF3) m (as defined in BF1) is a probability measure

(BF4) $\mathrm{FP}(\Omega) = 1_\chi$

(BF5) If A and B are disjoint then $\mathrm{FP}(A) \oplus \mathrm{FP}(B) = \mathrm{FP}(A \cup B)$

At first sight this definition seems reasonable and indeed it can and has be used as an informal theory for reasoning with fuzzy probabilities, however as a formal theory it is seriously defective as a consequence of the following Lemma.

Lemma 2. *For any event $E \in \mathcal{E}$,* $\mathrm{FP}(E)(x) = 0$ *for all $x < m(E)$. Such a membership function is termed left-crisp.*

Proof: *Consider an arbitrary event $E \in \mathcal{E}$. By definition* $\mathrm{FP}(E)$ *has a unique mode $m(E) \in [0,1]$ such that* $\mathrm{FP}(E)(m(E)) = 1 = \mathrm{FP}(E^c)(1 - m(E))$*. Suppose, for a contradiction, that* $0 < \mathrm{FP}(E)(x) \leq 1$ *for some $0 \leq x < m(E)$. Clearly, $0 < 1 - m(E) + x < 1$ and by the definition of $\oplus$*

$$\begin{aligned}
0 &= 1_\chi(1 - m(E) + x) \\
&= \mathrm{FP}(E \cup E^c)(1 - m(E) + x) \\
&= \max_{z+z'=1-m(E)+x} \min(\mathrm{FP}(E)(z), \mathrm{FP}(E^c)(z')) \\
&\geq \min(\mathrm{FP}(E)(x), \mathrm{FP}(E^c)(1 - m(E))) \\
&> 0 \quad (!)
\end{aligned}$$

So $\mathrm{FP}(E)(x) = 0$ *for all $x < m(E)$.*

Thus every BFP is necessarily left-crisp and therefore the theory cannot act as a formal model for vague probability assessments such as "quite likely" which tail-off smoothly to the left of their peak[5].

Worse still, there are two ways to strengthen this result to a proof that BFPs can only be embedded point probabilities (i.e. both left and right-crisp), both of which seem to reflect Jain and Agogino's intentions if not their precise formulation.

First, $\oplus$ is an operator defined on the fuzzy reals, not pairs of "convex normalized fuzzy set[s] ... of $[0,1]$". Lemma 1 rests only on the generous assumption that Jain and Agogino tacitly intended some form of $\oplus$ restricted to the unit interval. An equally reasonable assumption, albeit rather less generous, would be that the set of "convex normalized fuzzy set[s] ... of $[0,1]$" refers to $[0_\chi, 1_\chi]$. In this case suppose that for some $m(E) <$

[4] It may however be better to think of the sort of applications Zadeh considers as fuzzy information systems (Okuda et al., 1978; Tanaka et al., 1979). In this framework the database of cars would represent a series of fuzzy observations which can be used to estimate an underlying classical probabilistic model.

[5] Note that this criticism applies also to Zadeh's fuzzy probabilities.

$x \leq 1$, $\mathrm{FP}(E)(x) > 0$. Then,

$$\begin{aligned} 0 &= 1_\chi(1 - m(E) + x) \\ &= \mathrm{FP}(E \cup E^c)(1 - m(E) + x) \\ &\geq \min(\mathrm{FP}(E)(x), \mathrm{FP}(E^c)(1 - m(E))) \\ &> 0 \quad (!) \end{aligned}$$

And thus, $\mathrm{FP}(E) = m(E)_\chi$.

Second, it seems clear from their examples, that Jain and Agogino intend that for all $E \in \mathcal{E}$, $\mathrm{FP}(E) = 1_\chi \ominus \mathrm{FP}(E^c)$. Indeed this principle has considerable intuitive appeal, since it is roughly equivalent to the assertion that if you know something (however imprecise) about the probability of an event, then you know "just as much" about the probability of that event's complement[6]. But in this case, since every event is the complement of some left-crisp event, all events are also right-crisp. Again, the theory reduces to an embedding of classical point probabilities[7].

5 Linguistic probability theory

Linguistic probability theory has been developed to address the difficulties with Bayesian fuzzy probabilities. It seeks to do so rigorously, but in a way that is mindful of eventual applications. This section begins by introducing the basic elements of the theory. The concepts are then developed showing how the basic theory may be used to define discrete linguistic random variables, an analogue for the classical discrete random variable construct.

5.1 Linguistic probability measure

Linguistic probability theory differs from other theories of fuzzy probabilities by closely following the classical theory's hierarchy of concepts. This means beginning with the concept of a linguistic probability measure, modelled after the classic Kolmogorov formulation rehearsed in Section 3.

Definition 9 (Linguistic probability measure)**.** *Given an event algebra $\mathcal{E}$ defined over a set of outcomes Ω, a function $\mathrm{LP} : \mathcal{E} \to \mathbb{E}$ is termed a* linguistic probability measure *if and only if for all $A \in \mathcal{E}$*

(LP1) $0_\chi \preccurlyeq \mathrm{LP}(A) \preccurlyeq 1_\chi$
(LP2) $\mathrm{LP}(\Omega) = 1_\chi$ *and* $\mathrm{LP}(\emptyset) = 0_\chi$
(LP3) If $A_1, A_2, \ldots \in \mathcal{E}$ are disjoint, then

$$\mathrm{LP}(\bigcup_{i=1}^{\infty} A_i) \subseteq \sum_{i=1}^{\infty} \mathrm{LP}(A_i)$$

(LP4) $\mathrm{LP}(A) = 1_\chi \ominus \mathrm{LP}(A^c)$

where LP *is a linguistic probability measure on $(\Omega, \mathcal{E})$, the tuple $(\Omega, \mathcal{E}, \mathrm{LP})$ is termed a linguistic probability space.*

Like the first two axioms of classical probability theory LP1 and LP2 simply establish the scale for probabilities, but note that LP1 entails that linguistic probabilities have zero membership outside the chosen quantity space of the unit interval. The most significant parts therefore are LP3 and LP4. The underlying intuition is that vagueness in a probability acts as a soft constraint on all probabilities that are logically linked to it.

Thus, LP3 is intended to capture the intuition one might know the probability of the union of (say) two disjoint events more precisely than the probabilities of either individually. Consider, for example, tossing a coin which one knows to be biased. Here, knowledge about the probability of the result being heads (or tails) is uncertain, but the probability that result will be either heads or tails is certain (and equal to 1). Equally, the probability that the result will *not* be both heads and tails is certain (and equal to 0). In a similar vein, LP4 expresses that knowing something about the probability of an event translates into equally precise knowledge about the probability of its complement.

Like classical probability measures, linguistic probability measures are both continuous and monotonic. Furthermore, it is easy to see that a linguistic probability measure generalises its classical counterpart in the sense of the following Lemmas.

Lemma 3. *Given a classical probability measure,* P*, the map $\mathrm{LP} : \mathcal{E} \to \mathbb{E}$ determined by $\mathrm{LP}(A) = (\mathrm{P}(A))_\chi$ is a linguistic probability measure.* **Proof:** *Clearly,* $\mathrm{LP}(\Omega) = 1_\chi$, $\mathrm{LP}(\emptyset) = 0_\chi$ *and* $0_\chi \preccurlyeq \mathrm{LP}(A) \preccurlyeq 1_\chi$ *for all $A \in \mathcal{E}$ as required. Now for pairwise disjoint $A_1, A_2 \ldots \in \mathcal{E}$,*

$$\begin{aligned} \mathrm{LP}(\textstyle\bigcup_{i=1}^{\infty} A_i) &= (\mathrm{P}(\textstyle\bigcup_{i=1}^{\infty} A_i))_\chi = (\textstyle\sum_{i=1}^{\infty} P(A_i))_\chi \\ &= \textstyle\sum_{i=1}^{\infty} \mathrm{P}(A_i)_\chi = \textstyle\sum_{i=1}^{\infty} \mathrm{LP}(A_i) \end{aligned} \tag{18}$$

Finally, for all $A \in \mathcal{E}$. $\mathrm{LP}(A) = (\mathrm{P}(A))_\chi = (1 - \mathrm{P}(A^c))_\chi = 1_\chi \ominus \mathrm{LP}(A^c)$.

Thus any classical probability measure has an (embedding as an) equivalent linguistic probability measure. Similarly, any linguistic probability measure assigning only point probabilites determines a classical probability measure.

Lemma 4. *Given a linguistic probability measure,* LP*, such that for all $A \in \mathcal{E}$, $\mathrm{LP}(A) = (p_A)_\chi$ for some $p_A \in \mathbb{R}$ the map, $\mathrm{P} : \mathcal{E} \to [0, 1]$, determined by $\mathrm{P}(A) = p_A$ is a probability measure.* **Proof:** *Clearly,* $\mathrm{P}(\emptyset) = 0$ *and* $\mathrm{P}(A) \geq 0$ *for all $A \in \mathcal{E}$ as required. Now, given disjoint $A_1, A_2 \ldots \in \mathcal{E}$ and letting $A = \bigcup_{i=1}^{\infty} A_i$,*

$$(p_A)_\chi = \mathrm{LP}(A) \subseteq \sum_{i=1}^{\infty} \mathrm{LP}(A_i) = \sum_{i=1}^{\infty} (p_{A_i})_\chi = (\sum_{i=1}^{\infty} p_{A_i})_\chi \tag{19}$$

Hence, $p_A = \sum_{i=1}^{\infty} p_{A_i}$ *as required.*

[6] This will be elaborated further in the following Section

[7] Similar criticisms of the existing theories, albeit couched in very different language and developed independently can be found in Gert de Cooman's most recent work (de Cooman, 2003) on possibilistic previsions.

5.2 Discrete linguistic random variables

As in classical probaility theory discrete random variables are often more useful than events, although as we shall see, there is a strong connection between the two. The theory of discrete linguistic random variables will be developed in the context of a countable domain.

Definition 10 (Discrete linguistic random variable)**.** *Given a linguistic probability space* $(\Omega, \mathcal{E}, \mathrm{LP})$*, and a domain* D_X*, a function* $X : \Omega \to D_X$ *is termed a* discrete linguistic random variable *on* $(\Omega, \mathcal{E}, \mathrm{LP})$ *if and only if:*

a) Image$(X) = \{X(\omega) \ : \ \omega \in \Omega\}$ *is countable*
b) For all $x \in D_X$*,* $\{\omega \in \Omega \ : \ X(\omega) = x\} \in \mathcal{E}$

Definition 11 (Mass function)**.** *The* mass function *of a discrete linguistic random variable,* X *on* $(\Omega, \mathcal{E}, \mathrm{LP})$ *is the function,* $\mathrm{lp}_X : D_X \to \mathbb{E}$ *determined by*

$$\mathrm{lp}_X(x) = \mathrm{LP}(\{\omega \in \Omega \ : \ X(\omega) = x\}) \tag{20}$$

Representation theorem By definition, a linguistic mass function satisfies

a) $0_\chi \preccurlyeq \mathrm{lp}_X(x) \preccurlyeq 1_\chi$
b) $\mathrm{lp}_X(x) \subseteq 1_\chi \ominus (\sum_{x' \neq x} \mathrm{lp}_X(x'))$

for all $x \in D_X$. Note that b) also entails that $1_\chi \subseteq \sum_x \mathrm{lp}_X(x)$ since for any $x \in D_X$,

$$\begin{aligned} 1_\chi &\subseteq 1_\chi \ominus \mathrm{lp}_X(x) \oplus \mathrm{lp}_X(x) \\ &\subseteq 1_\chi \ominus \big(1_\chi \ominus \sum_{x' \neq x} \mathrm{lp}_X(x')\big) \oplus \mathrm{lp}_X(x) \\ &\subseteq \sum_{x'} \mathrm{lp}_X(x') \end{aligned}$$

Whilst these conditions are necessary they are also *sufficient* conditions in the sense of the following theorem.

Theorem 1 (Representation Theorem)**.** *If* $S = \{d_i \ : \ i \in I\}$ *is a non-empty countable set (indexed by* I*) and* $\{\pi_i \ : \ i \in I\}$ *is a set of fuzzy numbers such that for all* $i \in I$

a) $0_\chi \preccurlyeq \pi_i \preccurlyeq 1_\chi$
b) $\pi_i \subseteq 1_\chi \ominus \sum_{j \in I \ : \ j \neq i} \pi_j$

then there exists a linguistic probability space, $(\Omega, \mathcal{E}, \mathrm{LP})$ *and a discrete linguistic random variable,* X*, on* $(\Omega, \mathcal{E}, \mathrm{LP})$ *with the mass function*

$$\mathrm{lp}_X(d) = \begin{cases} \pi_i & \text{if } d = d_i \text{ for some } i \in I \\ 0_\chi & \text{otherwise} \end{cases} \tag{21}$$

Proof: *The proof proceeds by construction. Let* $\Omega = S$*,* $\mathcal{E} = \mathbb{P}(\Omega)$ *and define* $\mathrm{LP} : \mathcal{E} \to \mathbb{E}$ *by*

$$\mathrm{LP}(A) = (\sum_{i \ : \ s_i \in A} \pi_i) \cap (1_\chi \ominus \sum_{i \ : \ s_i \notin A} \pi_i) \cap [0,1]_\chi \tag{22}$$

By definition $\mathrm{LP}(\Omega) = 1_\chi$ *and* $\mathrm{LP}(\emptyset) = 0_\chi$ *as required. Now, for any* $A \in \mathcal{E}$ *since and* $0_\chi \preccurlyeq \pi_i$ *for all* $i \in I$ *there is an* $x \in [0,1]$ *such that*

$$\mu_{\sum_{i \ : \ s_i \in A} \pi_i}(x) = 1 \quad \text{and} \quad \mu_{1_\chi \ominus \sum_{i \ : \ s_i \notin A} \pi_i}(x) = 1 \tag{23}$$

Hence $\mathrm{LP}(A)$ *as defined is in* $[0_\chi, 1_\chi]$*. Now suppose* $A, B \in \mathcal{E}$ *are disjoint. By definition,*

$$\begin{aligned} \mathrm{LP}(A \cup B) &\subseteq \textstyle\sum_{i \ : \ s_i \in A \cup B} \pi_i \\ &= (\textstyle\sum_{i \ : \ s_i \in A} \pi_i) \oplus (\textstyle\sum_{i \ : \ s_i \in B} \pi_i) \\ &\subseteq \mathrm{LP}(A) \oplus \mathrm{LP}(B) \end{aligned} \tag{24}$$

Similarly, by definition,

$$\begin{aligned} &1_\chi \ominus \mathrm{LP}(A^c) \\ &= 1_\chi \ominus \{(\textstyle\sum_{i \ : \ s_i \in A^c} \pi_i) \cap (1_\chi \ominus \sum_{i \ : \ s_i \notin A^c} \pi_i) \cap [0,1]_\chi\} \\ &= (1_\chi \ominus \textstyle\sum_{i \ : \ s_i \in A^c} \pi_i) \cap (1_\chi \ominus 1_\chi \oplus \sum_{i \ : \ s_i \notin A^c} \pi_i) \cap [0,1]_\chi \\ &= \mathrm{LP}(A) \end{aligned}$$

Thus $(\Omega, \mathcal{E}, \mathrm{LP})$ *is a linguistic probability space.*

Finally, define $X : \Omega \to \mathbb{R}$ *by* $X(\omega) = \omega$*. Now, given* $s \in \mathbb{R}$*, if* $s \neq s_i$ *for all* $i \in I$*,* $X^{-1}(s) = \emptyset$ *and hence* $\mathrm{lp}_X(s) = 0_\chi$*. Otherwise,* $X_{-1}(s) = \{s_i\}$ *for some* $i \in I$ *and since* $\pi_i \subseteq 1_\chi \ominus \sum_{j \in I \ : \ j \neq i} \pi_j$

$$\mathrm{lp}_X(s) = \mathrm{LP}(\{s_i\}) = \pi_i \tag{25}$$

as required.

This theorem is important for practical applications of theory as it allows probabilistic modelling to dispense with measure theory almost all the time and concentrate on random variables which are typically the entities of interest. It is also an essential component in the proof that linguistic analogues for Bayesian networks can be constructed.

Note that the full strength of condition b) is only required to prove that the constructed linguistic random variable exactly coincides with the relevant π_i. If condition b) were replaced by the weaker condition

$$1_\chi \subseteq \sum_{i \in I} \pi_i \tag{26}$$

then $(\omega, \mathcal{E}, \mathrm{LP})$ as constructed above would still be a linguistic probability space. Thus the definition of LP can be viewed as a kind of recipe for correcting an improperly specified random variable.

Naturally if one considers certain types of domain i.e. the real numbers, it is possible to develop linguistic analogues for the classical probabilistic concepts of expectation and other moments, but this is beyond the scope of the present work.

Multivariate case As in classical probability theory it is possible to consider families of random variables and the realtionships between them.

Definition 12 (Linguistic joint mass function). *Given a set of linguistic discrete random variables* $X_1, X_2, \dots X_n$ *each defined over the linguistic probability space* $(\Omega, \mathcal{E}, \mathrm{LP})$ *and with domains* $D_{X_1}, D_{X_2}, \dots, D_{X_n}$*, the joint mass function is the function* $\mathrm{lp}_{X_1,X_2,\dots X_n} : \prod_{i=1}^{n} D_{X_i} \to [0_\chi, 1_\chi]$ *defined by*

$$\mathrm{lp}_{X_1,X_2,\dots X_n}(x_1, x_2, \dots x_n) \\ = \mathrm{LP}(\{\omega \in \Omega \ : \ X_i(\omega) = x_i \quad \forall i \in 1, 2, \dots, n\}) \tag{27}$$

This corresponds exactly to the classical definition. Note that the representation theorem proven above applies here too since $(X_1, X_2, \dots X_n)$ can be considered a single random variable ranging over the domain $\prod_{i=1}^{n} D_{X_i}$.

5.3 Conditional linguistic probability

Inference in classical probability theory depends on the notion of conditional probability.

Definition 13 (Conditional probability). *Given a probability space,* $(\Omega, \mathcal{E}, \mathrm{P})$ *and* $A, B \in \mathcal{E}$ *the* conditional probability *of* A *with respect to* B*, written* $P(A|B)$ *is given by*

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} \tag{28}$$

The classical formula for conditional probability suggests an analogous definition for linguistic conditional probabilites, namely

$$\mathrm{LP}(A|B) = \mathrm{LP}(A \cap B) \oslash \mathrm{LP}(B) \tag{29}$$

But unless $\mathrm{P}(B)$ is a point probability this linguistic conditional probability would not itself be a linguistic probability measure and would only satisfy the weaker relation

$$\mathrm{LP}(A \cap B) \subseteq \mathrm{LP}(A|B) \otimes \mathrm{LP}(B) \tag{30}$$

Note however, that the standard notation for the probability of A conditional on B, $\mathrm{P}(A|B)$ is somewhat misleading since P is a function on a set of events and $A|B$ has no set-theoretic interpretation. A more suggestive notation, and one that is in increasingly common use, writes $\mathrm{P}_B(A)$ for $\mathrm{P}(A|B)$. This makes it much clearer that a quite separate probability measure is in operation.

From this perspective, it seems reasonable to set aside concerns about defining linguistic conditional probabilities and turn instead to an investigation of what conditions a linguistic conditional and prior must obey in order to determine a viable probability measure. This question lies at the heart of the following Section.

6 Linguistic Bayesian networks

One of the most celebrated and widely used techniques to emerge from AI in the past few decades has been that of Bayesian networks, Pearl (1988) which allow information about a probabilistic system to be modelled, stored in a compact form and manipulated efficiently. Bayesian networks implicitly encode information about conditional independence relationships in their graphical structure. The theory of linguistic probabilities as developed above may be used to develop an analogous graphical model.

6.1 Network representation

In order to prove that it is possible to specify a linguistic joint probability distribution in the form of a Bayesian network, it is sufficient to show that the multiplying the conditional probability table at a node by its priors yields a joint distribution. The following Lemma and Theorem present this result for discrete linguistic random variables.

Lemma 5. *Given functions,* $f_1, f_2, \dots, f_n$ *and* g *with domains* $D_1, D_2, \dots, D_n$ *and* $D^* = D \times \prod_{i=1}^{n} D_i$ *respectively, and the common range* $[0_\chi, 1_\chi]$*, such that for all* $i \in \{1, 2, \dots, n\}, \overline{x} = (x, x_1, x_2, \dots, x_n) \in D^*$,

a) $f_i(x_i) \subseteq 1_\chi \ominus \sum_{x_i' \neq x_i} f_i(x_i')$
b) $g(\overline{x}) \subseteq 1_\chi \ominus \sum_{\overline{x}' \neq \overline{x}} g(\overline{x}')$

then for all $\overline{x} \in D^*$

$$g(\overline{x}) \prod_{i=1}^{n} f_i(x_i) \subseteq 1_\chi \ominus \big(\sum_{\overline{x}' \neq \overline{x}} g(\overline{x}') \prod_{i=1}^{n} f_i(x_i') \big)$$

Proof: *The proof proceeds by induction on* n.

(Base case) Suppose $n = 1$*. For all* $x \in D, x_1 \in D_1$,

$$\begin{aligned} & g(x, x_1) f_1(x_1) \\ & = f_1(x_1) g((x, x_1)) \\ & \subseteq f_1(x_1) \big(1_\chi \ominus \textstyle\sum_{x' \neq x, x_1'} g(x', x_1')\big) \\ & \subseteq f_1(x_1) \ominus f_1(x_1) \textstyle\sum_{x' \neq x, x_1'} g(x', x_1') \\ & \subseteq 1_\chi \ominus \textstyle\sum_{x_1'' \neq x_1} f_1(x_1'') \ominus f_1(x_1) \textstyle\sum_{x' \neq x, x_1} g(x', x_1') \\ & \subseteq 1_\chi \ominus \textstyle\sum_{x_1'' \neq x_1} f_1(x_1'') \textstyle\sum_{x''} g(x'', x_1'') \ominus f_1(x_1) \textstyle\sum_{x' \neq x} g(x', x_1') \\ & = 1_\chi \ominus \textstyle\sum_{x' \neq x, x_1' \neq x_1} g(x', x_1') f_1(x_1') \end{aligned}$$

(Inductive case) Suppose the theorem holds for $n = k$*. Consider the case that* $n = k + 1$*. For all* $\overline{x} \in D \times \prod_{i=1}^{k+1} D_i$*, by the inductive hypothesis,*

$$\begin{aligned} & g(\overline{x}) \textstyle\prod_{i=1}^{k+1} f_i(x_i) \\ & \quad \subseteq f_{k+1}(x_{k+1}) \big(1_\chi \ominus \textstyle\sum_{\overline{x}' \neq \overline{x}, x_{k+1}' = x_{k+1}} g(\overline{x}') \textstyle\prod_{i=1}^{k} f_i(x_i')\big) \end{aligned}$$

Now, the sequence of inferences used to establish the base case proves the desired result.

This Lemma shows that the mass functions of a set of conditional and prior discrete linguistic variables (under certain reasonable conditions) can be combined to form a joint distribution. The details of this are spelt out in the following Theorem.

Theorem 2 (Representation Theorem for linguistic Bayesian networks)**.** *Given functions, $f_1, f_2, \ldots, f_n$ and g with domains $D_1, D_2, \ldots, D_n$ and $D^* = D \times \prod_{i=1}^{n} D_i$ respectively, and the common range $[0_\chi, 1_\chi]$, such that for all $i \in \{1, 2, \ldots, n\}, \overline{x} = (x, x_1, x_2, \ldots, x_n) \in D^*$,*

a) $f_i(x_i) \subseteq 1_\chi \ominus \sum_{x'_i \neq x_i} f_i(x'_i)$
b) $g(\overline{x}) \subseteq 1_\chi \ominus \sum_{\overline{x}' \neq \overline{x}} g(\overline{x}')$

then there exist random variables $X, X_1, X_2, \ldots, X_n$ with respective domains $D, D_1, D_2, \ldots, D_n$ such that for all $i \in 1, 2, \ldots, n$

$$\mathrm{lp}_{X_i}(x) = f_i(x) \text{ for all } x \in D_i$$

and for all $\overline{x} \in D^*$

$$\mathrm{lp}_{X, X_1, X_2, \ldots, X_n}(\overline{x}) = g(\overline{x}) \prod_{i=1}^{n} f_i(x_i)$$

Proof: *The result follows directly from the preceding Lemma and the Representation Theorem.*

6.1.1 Computational issues As has been noted, the basic fuzzy arithmetic operators need not be significantly more computationally costly than their real-line counterparts. The problem of efficient inference in linguistic bayesian networks is essentially the same as the classical one: how to push the sigmas in the sum of products expression as far in as possible. By minimising the number of operations that have to be performed to evaluate a query, fast Bayesian network propagation algorithms (such as Pearl (1988)) also minimise the amount of imprecision. When dealing with fuzzy numbers, roughly speaking, the more operations an algorithm requires the worse its precision.

7 Application to forensic statistics

This Section places the preceding developments into context by developing a small, but realistic example in the domain of forensic statistics.

7.1 The use of Bayesian Networks in Forensic Statistics

Forensic statistics is a discipline that is mainly concerned with the experimental design of forensic examinations and the analysis of the obtained results. The issues it studies include hypothesis formulation, deciding on minimal sample sizes when studying populations of similar units of evidence and determining the statistical significance of the outcome of tests. Recently, the discipline has been branching out to the study of the statistical implications of forensic examinations on defence and prosecution positions during crime investigation and criminal court proceedings.

Table 1 Interpretation of the likelihood ratio.

LR	Support of evidence to prosecution claim over defence claim
1 to 10	limited
10 to 100	moderate
100 to 1,000	moderately strong
1,000 to 10,000	strong
$>$ 10,000	very strong

In Cook et al. (1998b), a method is proposed to assess the impact of a certain piece of forensic evidence on a given case. This method is the result of a significant research effort by the Forensic Science Service (FSS), the largest provider of forensic science services in England and Wales. It involves three steps. First, formalising the respective claims of the prosecution and the defence (Cook et al., 1998a; Evett et al., 2000a). Second, computing the probability that the evidence is found given that the claim of the prosecution is true and the probability that the evidence is found given that the claim of the defence is true. Third, dividing the former probability by the latter to determine the likelihood ratio,

$$LR = \frac{P(E \mid C_p)}{P(E \mid C_d)}$$

where E, C_p, C_d respectively represent the evidence, the prosecution claim and the defence claim, and $P(E \mid C)$ is the probability that evidence E is found if claim C is true (Balding and Donnelly, 1995).

This approach has two key advantages. First, the potential benefit associated with performing forensic procedures (which are often expensive and resource intensive) may be assessed in advance by examining the effect of their possible outcomes on the likelihood ratio. Increasingly, police forces must purchase forensic services. Likelihood ratio based calculations can support this difficult decision making process. Second, the likelihood ratio can be used to justify the testimonies of forensic experts during the court proceedings. To this end, a verbal scale to help forensic experts interpret the LR is suggested by the FSS (Evett et al., 2000b). This is reproduced in Table 1 for reference.

7.2 A classical example

The likelihood ratio method is, of course, crucially dependent upon a means to compute the probabilities $P(E \mid C_p)$ and $P(E \mid C_d)$. Bayesian Networks have emerged as a helpful technique in this context (Aitken et al., 2003; Cook et al., 1999; Dawid et al., 2002). An example may best illustrate this application of Bayesian networks. Consider the following scenario:

Table 2 Variables in the one-way transfer case.

	Event	Domain
q_t	quantity of transferred fragments	{none,few,many}
q_p	quantity of persisted fragments	{none,few,many}
q_l	quantity of lifted fragments	{none,few,many}
t_c	type of contact	{none,some}
p_s	proportion of fragments shed	{none,small,large}
p_l	proportion of fragments lifted	{some,most,all}
p_l	proportion of fragments lifted	{some,most,all}

Scenario 2. *A burglar smashes the window of a shop, steals some money from the cash registry and flees the scene of the crime. A bystander witnessed this event and reports a description of the perpetrator to the police who arrest a man, matching the description of the witness half an hour after the event. The suspect, Mr. Blue, denies having been near the shop. However, q_l glass fragments, matching the type of glass of the shop's window, are retrieved from Mr. Blue's clothes.*

Figure 1 shows a Bayesian network that models the probabilistic relationship between the retrieval of q_l glass fragments from the garment of Mr. Blue in the forensic laboratory and the type of contact, t_c, between Mr. Blue and the shop's window. The number of glass fragments, q_l, that are retrieved from Mr. Blue's clothes depends on the number of glass fragments that have persisted in the clothes, q_p, and on the effectiveness of the retrieval technique, p_l, where p_l represents the proportion of glass fragments lifted from the garments under examination. The number of glass fragments, q_p, that have persisted in the clothes until the time of the examination, in turn, is dependent upon the number of glass fragments, q_t, that were transferred in the first place and the proportion of fragments, p_s, shed between the time of transfer and the time of the examination. Finally, the number of transferred fragments, q_t, depends on the type of contact t_c. The domains of these variables are reproduced in Table 2. Suppose that the prosecution case is that the defendant has had *some* contact with the window in question and that a given forensic procedure has yielded *many* matching fragments. The probabilities required to evaluate the likelihood ratio are provided in Tables 3, 4, 5 and 6. The relevant calculation is,

$$LR = \frac{\mathrm{P}(q_l = \mathit{many} \mid t_c = \mathit{some})}{\mathrm{P}(q_l = \mathit{many} \mid t_c = \mathit{none})} = \frac{0.428586}{0.038813} = 11.042\ldots$$

Thus, according to Table 1, this item of forensic evidence provides *moderate* support to the prosecution case.

7.3 A linguistic example

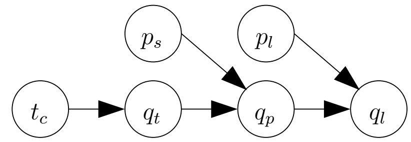

Fig. 1 Bayesian Network of a one-way transfer case.

Table 3 Classical prior probabilities $\mathrm{P}(p_s)$ and $\mathrm{P}(p_l)$.

p_s	$\mathrm{P}(p_s)$
none	0.03
small	0.3
large	0.67

p_l	$\mathrm{P}(p_l)$
none	0.06
few	0.29
many	0.65

Table 4 Classical conditional probabilities $P(q_t \mid t_c)$.

t_c	$\mathrm{P}(q_t = \mathit{none} \mid t_c)$	$\mathrm{P}(q_t = \mathit{few} \mid t_c)$	$\mathrm{P}(q_t = \mathit{many} \mid t_c)$
none	0.9	0.05	0.05
some	0.1	0.25	0.65

Table 5 Classical conditional probabilities $P(q_p \mid q_t, p_s)$.

q_t	p_s	$\mathrm{P}(q_p = \mathit{none} \mid q_t, p_s)$	$\mathrm{P}(q_p = \mathit{few} \mid q_t, p_s)$	$\mathrm{P}(q_p = \mathit{many} \mid q_t, p_s$
none	none	1	0	0
	small	1	0	0
	large	1	0	0
few	none	0	1	0
	small	0.1	0.9	0
	large	0.3	0.7	0
many	none	0	0	1
	small	0.05	0.1	0.85
	large	0.07	0.48	0.45

Table 6 Classical conditional probabilities $P(q_l \mid q_p, p_l)$.

q_p	p_l	$\mathrm{LP}(q_l = \mathit{none} \mid q_p, p_l)$	$\mathrm{LP}(q_l = \mathit{few} \mid q_p, p_l)$	$\mathrm{LP}(q_l = \mathit{many} \mid q_p, p_l)$
none	some	1	0	0
	most	1	0	0
	all	1	0	0
few	some	0.05	0.95	0
	most	0.05	0.95	0
	all	0.02	0.6	0.38
many	some	0.08	0.46	0.46
	most	0.2	0.2	0.6
	all	0	0	1

Table 7 Linguistic prior probabilities $\mathrm{LP}(p_s)$ and $\mathrm{LP}(p_l)$

p_s	$\mathrm{LP}(p_s)$
none	nearly impossible
small	quite unlikely
large	quite likely

p_l	$\mathrm{LP}(p_l)$
none	nearly impossible
few	quite unlikely
many	quite likely

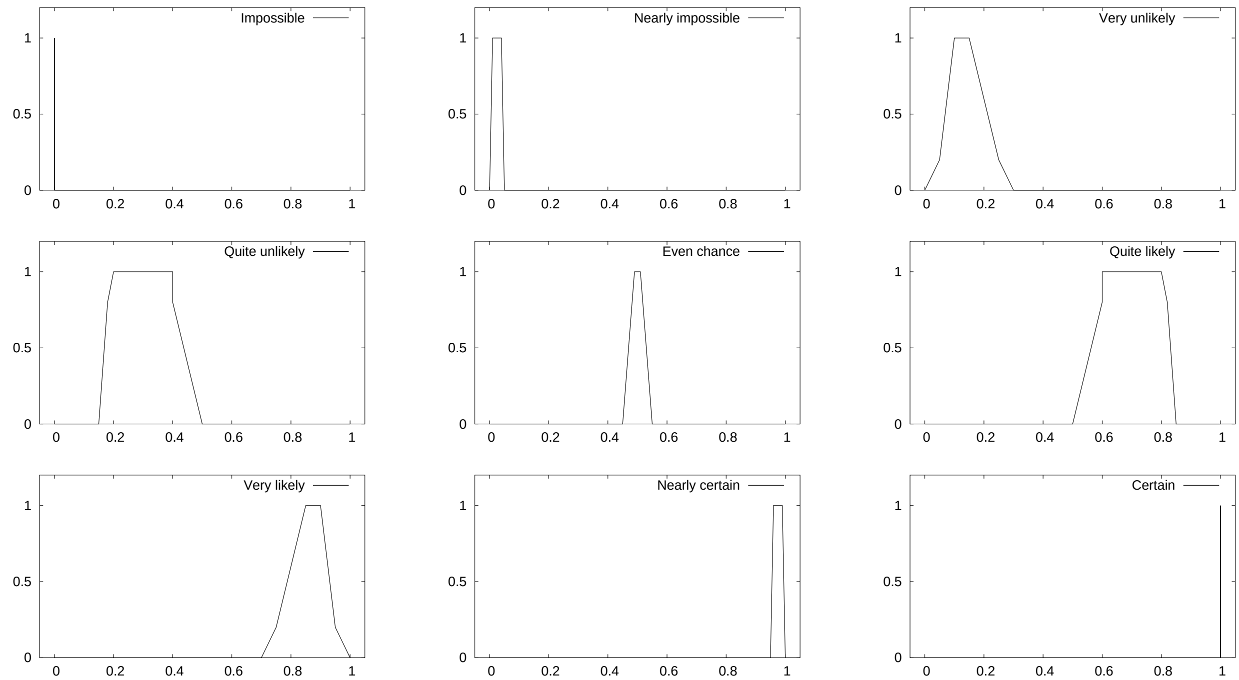

Fig. 2 The linguisitic probabilites used in the worked example.

Table 8 Linguistic conditional probabilities $\mathrm{LP}(q_t \mid t_c)$.

t_c	$\mathrm{LP}(q_t = none \mid t_c)$	$\mathrm{LP}(q_t = few \mid t_c)$	$\mathrm{LP}(q_t = many \mid t_c)$
none	nearly certain	nearly impossible	nearly impossible
some	impossible	quite unlikely	quite likely

Table 9 Linguistic conditional probabilities $P(q_p \mid q_t, p_s)$.

q_t	p_s	$\mathrm{LP}(q_p = none \mid q_t, p_s)$	$\mathrm{LP}(q_p = few \mid q_t, p_s)$	$\mathrm{LP}(q_p = many \mid q_t, p_s)$
none	none	certain	impossible	impossible
	small	certain	impossible	impossible
	large	certain	impossible	impossible
few	none	impossible	certain	impossible
	small	very unlikely	very likely	impossible
	large	quite unlikely	quite likely	impossible
many	none	impossible	impossible	certain
	small	nearly impossible	very unlikely	very likely
	large	nearly impossible	even chance	even chance

Table 10 Linguistic conditional probabilities $P(q_l \mid q_p, p_l)$.

q_p	p_l	$\mathrm{LP}(q_l = none \mid q_p, p_l)$	$\mathrm{LP}(q_l = few \mid q_p, p_l)$	$\mathrm{LP}(q_l = many \mid q_p, p_l)$
none	some	certain	impossible	impossible
	most	certain	impossible	impossible
	all	certain	impossible	impossible
few	some	nearly impossible	nearly certain	impossible
	most	nearly impossible	nearly certain	impossible
	all	nearly impossible	very likely	very unlikely
many	some	nearly impossible	even chance	even chance
	most	very unlikely	very unlikely	quite likely
	all	impossible	impossible	certain

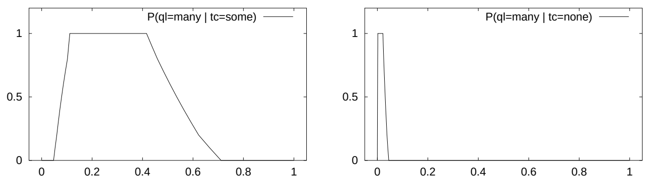

Fig. 3 Computed linguistic probabilities for LP($q_l = many \mid t_c = some$) and LP($q_l = many \mid t_c = none$).

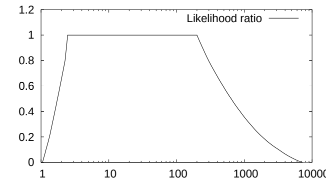

Fig. 4 The computed fuzzy likelihood ratio plotted on a logarithmic scale.

Tables 7, 8, 9 and 10 present a linguistically specified version of the network discussed in Section 2. The qualitative probability terms themselves are in turn graphed in Figure 2. Computations are performed in exactly the same sequence as in the classical case, but with fuzzy arithmetic operators and numbers. This yields fuzzy values for LP($q_l = many \mid t_c = some$), LP($q_l = many \mid t_c = none$) and the likelihood ratio. These are presented in Figures 3 and 4 respectively. Note that the membership functions, as expected, subsume their classical counterparts as calculated in Section 2.

The value calculated for LP($q_l = many \mid t_c = none$) is particularly interesting. Practical forensic applications typically use conservative (high) estimates for $\mathrm{P}(E|C_d)$ (i.e. the denominator in the likelihood calculation) thereby biasing the case in favour of the defence (Cook et al., 1999). Additionally, the probabilities typically associated with the subsets of events modelling the case where evidence originates not with the crime, but with some other source, are vanishingly small. Moreover, these probabilities are typically the most difficult to obtain experimentally. The use of linguistic probabilities to represent such probabilities allows the uncertainty that prompts this conservatism to be explicitly included in the model.

The fuzzy value calculated for the likelihood ratio has an extremely broad plateau (α-cut at 1), dramatically exhibiting the sensitivity of this statistic to small perturbations in the subjective probabilities on which it is based. That the set's membership function is greater than zero in each of the Forensic Science Service's recommended interpretation classes that are reproduced in Table 1 is, of course, partly a result of the rather "low-resolution" term set used for convenience of presentation here. Nevertheless, to re-iterate the central argument of this paper, the effects of propagating uncertainties should not be brushed aside. It is clear from the graph that the support provided by the evidence is roughly speaking moderate to strong[8], but that the newly acknowledged uncertainties in the subjective probability estimates are certainly consistent with much more limited support.

[8] Note that, given a fuzzification of the likelihood ratio quantity space, it would be possible to automatically generate this description.

8 Conclusions

This paper has sought to establish a series of propositions. First, that it is desirable to find a principled way of reasoning with approximate probability assessments represented as fuzzy sets. Second, that the two existing approaches commonly grouped together under the rubric of "fuzzy probabilities" are quite distinct: Zadeh's theory deals with the fuzzy probability induced by fuzzy events whereas Jain and Agogino seek to represent approximate knowledge about the probability of a crisp event. Third, that technical problems with the latter render it unsuited to its intended purpose. And finally, that it is possible to address these problems and develop a useful theory of vague probability assessments or linguistic probabilities, which is both rich enough to support analogues of classical probabilitic models and computationally tractable. This is the main contribution of this article.

The resulting theory of linguistic probabilities has been used to establish a linguistically-specified analogue to classical Bayesian networks. The domain of forensic statistics was chosen as an ideal setting for a discussion of the importance of this application. Although Bayesian networks have proven their worth as a knowledge acquisition tool, they rely on a large number of prior and conditional probabilities and it is often practically infeasible or too expensive to determine these through detailed experiments. In all such domains, knowledge engineers must fall back on subjective probability estimates, and it is at this point that the need for a more expressive representation is at its most pressing. The example drawn from forensic statistics presented earlier demonstrates this point well.

There are two clear lines of research stemming from the results presented in this paper. First, we hope to flesh out the loose interpretation currently attached to linguistic probabilities into a full semantics. Of course, in its broadest sense this is a significant challenge as is evidenced by the continuing disputes about the interpretation of classical probability theory. Nevertheless, it would be highly desirable to give an account of a decision theory based on linguistic probabilities that explains the fact that fuzzy expected values are not necessarily comparable. It would also be interesting to formally address the epistemic origins of vague probability assessments. In short: how does an agent come to have (only) approximate beliefs about the probability of some event?

An answer to this question would help to pave the way for applications in probabilistic modelling and machine learning.

Second, although preliminary studies have provided a proof-of-concept, work is ongoing to build more sophisticated and realistic linguistic Bayesian networks (Halliwell et al., 2003). In particular, we hope to combine this work with ongoing research into the automatic combination of network fragments for crime scenario creation and selection. A further aspect of this work will involve the translation of computed linguistic probabilities back into natural language through the use of linguistic hedges(Gómez Marin-Blazquez and Shen, 2002). Working with forensic experts in this way will also provide an opportunity to evaluate the claim that linguistic probabilities are a useful knowledge acquisition tool.

Acknowledgments

This work was partly supported by UK EPSRC grant EP/D057086. The authors are grateful to Drs Jeroen Keppens and Ruiqing Zhao for their assistance in carrying out this research, but will take full responsibility for the views expressed in this paper.

References

Aitken, C., Taroni, F., Garbolino, P., 2003. A graphical model for the evaluation of cross-transfer evidence in DNA profiles. Theoretical Population Biology 63 (3), to appear.

Balding, D. J., Donnelly, P., 1995. Inference in forensic identification. Journal of the Royal Statistical Society Series A 158, 21–53.

Budescu, D. V., Wallsten, T. S., 1985. Consistency in interpretation of probabilistic phrases. Organizational Behaviour and Human Decision Processes 36, 391–405.

Cook, R., Evett, I., Jackson, G., Jones, P., Lambert, J., 1998a. A hierarchy of propositions: deciding which level to address in casework. Science and Justice 38, 231–239.

Cook, R., Evett, I., Jackson, G., Jones, P., Lambert, J., 1998b. A model for case assessment and interpretation. Science and Justice 38, 151–156.

Cook, R., Evett, I., Jackson, G., Jones, P., Lambert, J., 1999. Case pre-assessment and review in a two-way transfer case. Science and Justice 39, 103–111.

Dawid, A. P., Mortera, J., Pascali, V. L., van Boxel, D. W., 2002. Probabilistic expert systems for forensic inference from genetic markers. Scandinavian Journal of Statistics 29, 577–595.

de Cooman, G., 2003. A behavioural model for vague probability assessments. Personal communication expected to appear for journal publication soon.

Evett, I., Jackson, G., Lambert, J., 2000a. More on the hierarchy of propositions: exploring the distinction between explanations and propositions. Science and Justice 40, 3–10.

Evett, I., Jackson, G., Lambert, J., McCrossan, S., 2000b. The impact of the principles of evidence interpretation on the structure and content of statements. Science and Justice 40, 233–239.

Gómez Marin-Blazquez, J., Shen, Q., 2002. From approximative to descriptive fuzzy classifiers. IEEE Transactions on Fuzzy Systems 10 (4), 487–497.

Grimmet, G., Welsh, D., 1986. Probability: an introduction. Oxford University Press, Oxford.

Halliwell, J., Keppens, J., Shen, Q., 2003. Linguistic bayesian networks for reasoning with subjective probabilities in forensic statistics. In: Proceedings of the 9th International Conference on AI and Law. ACM, pp. 42–50.

Halliwell, J., Shen, Q., 2002. Towards linguistic probability theory. In: Proceedings of the 11th International Conference on Fuzzy Systems. pp. 596–601.

Jain, P., Agogino, A. M., 1990. Stochastic sensitivity analysis using fuzzy influence diagrams. In: Schachter, R. D., Levitt, T. S., Kanal, L. N., Lemmer, J. F. (Eds.), Uncertainty in Artificial Intelligence 4. North-Holland, Amsterdam, pp. 79–92.

Kahneman, D., Slovic, P., Tversy, A., 1985. Judgement Under Uncertainty: Heuristics and Biases. Cambridge University Press, Cambridge, UK.

National Research Council Governing Board Commitee on the Assessment of Risk, 1981. The handling of risk assessments in NRC reports. Washington D.C.

Okuda, T., Tanaka, H., Asai, K., 1978. A formulation of fuzzy decision problems with fuzzy information, using probability measures of fuzzy events. Information Control 38, 135–147.

Pearl, J., 1988. Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann, San Mateo, CA.

Russell, B., 1923. Vagueness. Australasian Journal of Psychological Philosophy 1, 84–92.

Tanaka, H., Okuda, T., Asai, K., 1979. Fuzzy information and decision in statistical model. In: Gupta, M., Rage, P., Yager, R. (Eds.), Advances in Fuzzy Sets Theory and Applications. North Holland, Amsterdam, pp. 303–320.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., Forsyth, B., 1986. Measuring the vague meanings of probability terms. Journal of Experimental Psychology: General 115 (4), 348–365.

Zadeh, L., 1975. The concept of a linguistic variable and its application to approximate reasoning. Part 2. Information Science 8, 301–353.

Zadeh, L., 1984. Fuzzy probabilities. Information Processing Management 20, 363–372.

Zadeh, L., 1996. Fuzzy logic = computing with words. IEEE Transactions on Fuzzy Systems 2, 103–111.

Zimmer, A. C., 1983. Verbal vs. numerical processing of subjective probabilities. In: Scholz, R. W. (Ed.), Decision making under uncertainty. North-Holland, Amsterdam.

Zimmer, A. C., 1986. What uncertainty judgements can tell about the underlying subjective probabilities. In: Kanal, L. N., Lemmer, J. F. (Eds.), Uncertainty in Artificial Intelligence. North-Holland, Amsterdam.