Imprecise Probabilities from Imprecise Descriptions of Real Numbers

Jonathan Lawry
Dept of Engineering Maths,
University of Bristol,
Bristol, BS8 1TR, UK
j.lawry@bris.ac.uk

Inés González-Rodríguez
Dept. of Maths, Stats and Compt.
University of Cantabria
Santander, Spain
ines.gonzalez@unican.es

Yongchuan Tang
College of Computer Science,
Zhejiang University,
Hangzhou 310027, PR China
tyongchuan@gmail.com

Abstract

A prototype theory interpretation of the label semantics framework is proposed as a possible model of imprecise descriptions of real numbers. It is shown that within this framework conditioning given imprecise descriptions of a real variable naturally results in imprecise probabilities. An inference method is proposed from data in the form of a set of imprecise descriptions, which naturally suggests an algorithm for estimating lower and upper probabilities given imprecise data values.

Keywords. Label Semantics, Prototype Theory, Random Sets, Lower and Upper Distributions, Second Order Distributions

1 Introduction

The label semantics framework [3], [4] is an epistemic theory of the uncertainty associated with vague or imprecise descriptions of an object or value. In label semantics the focus is on the decision making process an intelligent agent must go through in order to identify which labels or expressions can actually be used to describe an object or value. In other words, in order to make an assertion describing an object in terms of some set of linguistic labels, an agent must first identify which of these labels are appropriate or assertible in this context. Given the way that individuals learn language through an ongoing process of interaction with the other communicating agents and with the environment, then we can expect there to be considerable uncertainty associated with any decisions of this kind. Furthermore, there is a subtle assumption central to the label semantic model, that such decisions regarding appropriateness or assertibility are meaningful. For instance, the fuzzy logic view is that vague descriptions like 'John is *tall*' are generally only partially true and hence it is not meaningful to consider which of a set of given labels can truthfully be used to described John's height. However, we contest that the efficacy of natural language as a means of conveying information between members of a population lies in shared conventions governing the appropriate use of words which are, at least loosely, adhered to by individuals within the population.

In our everyday use of language we are continually faced with decisions about the best way to describe objects and instances in order to convey the information we intend. For example, suppose you are witness to a robbery, how should you describe the robber so that police on patrol in the streets will have the best chance of spotting him? You will have certain labels that can be applied, for example *tall, short, medium, fat, thin, blonde, etc*, some of which you may view as inappropriate for the robber, others perhaps you think are definitely appropriate while for some labels you are uncertain whether they are appropriate or not. On the other hand, perhaps you have some ordered preferences between labels so that *tall* is more appropriate than *medium* which is in turn more appropriate than *short*. Your choice of words to describe the robber should surely then be based on these judgments about the appropriateness of labels. Yet where does this knowledge come from and more fundamentally what does it actually mean to say that a label is or is not appropriate? Label semantics proposes an interpretation of vague description labels based on a particular notion of appropriateness and suggests a measure of subjective uncertainty resulting from an agent's partial knowledge about what labels are appropriate to assert. Furthermore, it is suggested that the vagueness of these description labels lies fundamentally in the uncertainty about if and when they are appropriate as governed by the rules and conventions of language use.

The above argument brings us very close to the epistemic view of vagueness as expounded by Timothy Williamson [12]. Williamson assumes that for the extensions of a vague concept there is a precise but unknown dividing boundary between it and the ex-

tension of the negation of that concept. However, while there are marked similarities between the epistemic theory and the label semantics view, there are also some subtle differences. For instance, the epistemic view would seem to assume the existence of some objectively correct, but unknown, definition of a vague concept. Instead of this we argue that individuals when faced with decision problems regarding assertions find it useful as part of a decision making strategy to assume that there is a clear dividing line between those labels which are and those which are not appropriate to describe a given instance. We refer to this strategic assumption across a population of communicating agents as the *epistemic stance* [5], a concise statement of which is as follows:

> *Each individual agent in the population assumes the existence of a set of labeling conventions, valid across the whole population, governing what linguistic labels and expressions can be appropriately used to describe particular instances.*

In practice these rules and conventions underlying the appropriate use of labels would not be imposed by some outside authority. In fact, they may not exist at all in a formal sense. Rather they are represented as a distributed body of knowledge concerning the assertability of predicates in various cases, shared across a population of agents, and emerging as the result of interactions and communications between individual agents all adopting the epistemic stance. The idea is that the learning processes of individual agents, all sharing the fundamental aim of understanding how words can be appropriately used to communicate information, will eventually converge to some degree on a set of shared conventions. The very process of convergence then to some extent vindicates the epistemic stance from the perspective of individual agents. Of course, this is not to suggest complete or even extensive agreement between individuals as to these appropriateness conventions. However, the overlap between agents should be sufficient to ensure the effective transfer of useful information.

In this paper we consider the application of label semantics to model the description of real numbers using vague or imprecise labels. In particular, given a real valued variable x and a label L for real numbers we attempt to understand the nature of the information provided by assertions of the form 'x is L'. Indeed we will argue that from an epistemic perspective such assertions naturally result in imprecise probabilities. The model we propose will be based on a new interpretation of label semantics linking random set theory and Rosch's [9] prototype theory of concepts.

2 The Prototype Interpretation of Label Semantics

Label semantics proposes two fundamental and interrelated measures of the appropriateness of labels as descriptions of an object or value. Given a finite set of labels LA a set of compound expressions LE can then be generated through recursive applications of logical connectives. The labels $L_i \in LA$ are intended to represent words such as adjectives and nouns which can be used to describe elements from the underlying universe Ω. In other words, L_i correspond to description labels for which the expression 'x is L_i' is meaningful for any $x \in \Omega$. For example, if Ω is the set of all possible *rgb* values then LA could consist of the basic colour labels such as *red*, *yellow*, *green*, *orange* etc. In this case LE then contains those compound expression such as *red & yellow*, *not blue nor orange* etc. The measure of appropriateness of an expression $\theta \in LE$ as a description of instance x is denoted by $\mu_\theta(x)$ and quantifies the agent's subjective belief that θ can be used to describe x based on his/her (partial) knowledge of the current labeling conventions of the population. From an alternative perspective, when faced with an object to describe, an agent may consider each label in LA and attempt to identify the subset of labels that are appropriate to use. Let this set be denoted by $\mathcal{D}_x$. In the face of their uncertainty regarding labeling conventions the agent will also be uncertain as to the composition of $\mathcal{D}_x$, and in label semantics this is quantified by a probability mass function $m_x : 2^{LA} \to [0,1]$ on subsets of labels. The relationship between these two measures will be described below.

Definition 1. *Label Expressions*
Given a finite set of labels LA the corresponding set of label expressions LE is defined recursively as follows:

- *If $L \in LA$ then $L \in LE$*
- *If $\theta, \varphi \in LE$ then $\neg\theta, \theta \wedge \varphi, \theta \vee \varphi \in LE$*

The mass function m_x on sets of labels then quantifies the agent's belief that any particular subset of labels contains all and only the labels with which it is appropriate to describe x.

Definition 2. *Mass Function on Labels*
$\forall x \in \Omega$ a mass function on labels is a function $m_x : 2^{LA} \to [0,1]$ such that $\sum_{F \subseteq LA} m_x(F) = 1$

The appropriateness measure, $\mu_\theta(x)$, and the mass function m_x are then related to each other on the basis that asserting 'x is θ' provides direct constraints on $\mathcal{D}_x$. For example, asserting 'x is $L_1 \wedge L_2$', for labels $L_1, L_2 \in LA$ is taken as conveying the infor-

mation that both L_1 and L_2 are appropriate to describe x so that $\{L_1, L_2\} \subseteq \mathcal{D}_x$. Similarly, '$x$ is $\neg L$' implies that L is not appropriate to describe x so $L \notin \mathcal{D}_x$. In general we can recursively define a mapping $\lambda : LE \to 2^{2^{LA}}$ from expressions to sets of subsets of labels, such that the assertion 'x is θ' directly implies the constraint $\mathcal{D}_x \in \lambda(\theta)$ and where $\lambda(\theta)$ is dependent on the logical structure of θ.

Definition 3. *λ-mapping*
$\lambda : LE \to 2^{2^{LA}}$ is defined recursively as follows: $\forall L_i \in LA$, $\forall \theta,\ \varphi \in LE$

- $\lambda(L_i) = \{F \subseteq LA : L_i \in F\}$
- $\lambda(\theta \wedge \varphi) = \lambda(\theta) \cap \lambda(\varphi)$
- $\lambda(\theta \vee \varphi) = \lambda(\theta) \cup \lambda(\varphi)$
- $\lambda(\neg\theta) = \lambda(\theta)^c$

Based on the λ mapping we then define $\mu_\theta(x)$ as the sum of m_x over those sets of labels in $\lambda(\theta)$.

Definition 4. *Appropriateness Measure*
The appropriateness measure defined by mass function m_x is a function $\mu : LA \times \Omega \to [0,1]$ satisfying

$$\forall \theta \in LE,\ \forall x \in \Omega\ \mu_\theta(x) = \sum_{F \in \lambda(\theta)} m_x(F)$$

where $\mu_\theta(x)$ is used as shorthand notation for $\mu(\theta, x)$.

Prototype theory and imprecise probabilities have already been linked by Walley and de Cooman [11] who identified labels based on prototypes as a special case of monotonic predicates which they argue naturally induce possibility distributions. A prototype theory interpretation of Label Semantics has recently been proposed [6], [7], [10] in which the basic labels LA correspond to natural categories each with an associated set of prototypes. A label L_i is then deemed to be an appropriate description of an element $x \in \Omega$ provided x is *sufficiently similar* to the prototypes of L_i. The requirement of being 'sufficiently similar' is clearly imprecise and is modelled here by introducing an uncertain threshold on distance from prototypes.

A distance function d is defined on Ω such that $d : \Omega^2 \to [0, \infty)$ and satisfies $d(x,x) = 0$ and $d(x,y) = d(y,x)$ for all elements $x, y \in \Omega$. This function is then extended to sets of elements such that for $S, T \subseteq \Omega$, $d(S,T) = \inf\{d(x,y) : x \in S \text{ and } y \in T\}$. For each label $L_i \in LA$ let there be a set $P_i \subseteq \Omega$ corresponding to prototypical elements for which L_i is certainly an appropriate description. Within this framework L_i is deemed to be appropriate to describe an element $x \in \Omega$ provided x is sufficiently close or similar to a prototypical element in P_i. This is formalized by the requirement that x is within a maximal distance threshold ϵ of P_i. i.e. L_i is appropriate to describe x if $d(x, P_i) \leq \epsilon$ where $\epsilon \geq 0$. From this perspective an agent's uncertainty regarding the appropriateness of a label to describe a value x is characterised by his or her uncertainty regarding the distance threshold ϵ. Here we assume that ϵ is a random variable and that the uncertainty is represented by a probability density function δ for ϵ defined on $[0, \infty)$. Within this interpretation a natural definition of the complete description of an element $\mathcal{D}_x$ and the associated mass function m_x can be given as follows:

Definition 5. *Prototype Interpretations of $\mathcal{D}_x$ and m_x*
For $\epsilon \in [0, \infty)$ $\mathcal{D}_x^\epsilon = \{L_i \in LA : d(x, P_i) \leq \epsilon\}$ and $\forall F \subseteq LA$ $m_x(F) = \delta(\{\epsilon : \mathcal{D}_x^\epsilon = F\})$[1]

Appropriateness measures can then be evaluated according to definition 4. Alternatively, we can define a random set neighbourhood for each expression $\theta \in LE$ corresponding to those elements of Ω which can be appropriately described as θ, and then define $\mu_\theta(x)$ as the single point coverage function of this random set as follows:

Definition 6. *Random Set Neighbourhood of an Expression*
For $\theta \in LE$ and $\epsilon \in [0, \infty)$, $\mathcal{N}_\theta^\epsilon \subseteq \Omega$ is defined recursively as follows: $\forall L_i \in LA$, $\forall \theta, \varphi \in LE$

- $\mathcal{N}_{L_i}^\epsilon = \{x \in \Omega : d(x, P_i) \leq \epsilon\}$
- $\mathcal{N}_{\theta \wedge \varphi}^\epsilon = \mathcal{N}_\theta^\epsilon \cap \mathcal{N}_\varphi^\epsilon$
- $\mathcal{N}_{\theta \vee \varphi}^\epsilon = \mathcal{N}_\theta^\epsilon \cup \mathcal{N}_\varphi^\epsilon$
- $\mathcal{N}_{\neg\theta}^\epsilon = (\mathcal{N}_\theta^\epsilon)^c$

Theorem 1. *Random Neighbourhood Representation Theorem [7]*

$$\forall \theta \in LE,\ \forall x \in \Omega\ \mu_\theta(x) = \delta(\{\epsilon : x \in \mathcal{N}_\theta^\epsilon\})$$

Proof. Initially we show by induction that $\forall \theta \in LE$, $\forall \epsilon \geq 0$ $\mathcal{N}_\theta^\epsilon = \{x : \mathcal{D}_x^\epsilon \in \lambda(\theta)\}$. Let $LE^{(1)} = LA$ and for $k > 1$ $LE^{(k)} = LE^{(k-1)} \cup \{\theta \wedge \varphi, \theta \vee \varphi, \neg\theta : \theta, \varphi \in LE^{(k-1)}\}$. We now proceed by induction on k.
Limit Case: $k = 1$ For $L_i \in LA$ we have by definition 6 that $\mathcal{N}_{L_i}^\epsilon = \{x : d(x, P_i) \leq \epsilon\} = \{x : L_i \in \mathcal{D}_x^\epsilon\} = \{x : \mathcal{D}_x^\epsilon \in \lambda(L_i)\}$ by definition 3.
Inductive Step: Assume true for k For $\Psi \in$

[1] For Lesbegue measurable set $I \subseteq [0, \infty)$, we denote $\delta(I) = \int_I \delta(\epsilon) d\epsilon$ i.e. we also use δ to denote the probability measure induced by density function δ.

$LE^{(k+1)}$ either $\Psi \in LE^{(k)}$, in which case the result holds trivially by the inductive hypothesis, or one of the following holds for $\theta, \varphi \in LE^{(k)}$:

- $\Psi = \theta \wedge \varphi$ so that $\mathcal{N}^{\epsilon}_{\Psi} = \mathcal{N}^{\epsilon}_{\theta\wedge\varphi} = \mathcal{N}^{\epsilon}_{\theta} \cap \mathcal{N}^{\epsilon}_{\varphi}$ (by definition 6) $= \{x : \mathcal{D}^{\epsilon}_{x} \in \lambda(\theta)\} \cap \{x : \mathcal{D}^{\epsilon}_{x} \in \lambda(\varphi)\}$ (by the inductive hypothesis) $= \{x : \mathcal{D}^{\epsilon}_{x} \in \lambda(\theta) \cap \lambda(\varphi)\} = \{x : \mathcal{D}^{\epsilon}_{x} \in \lambda(\theta \wedge \varphi)\}$ (by definition 3).
- $\Psi = \theta \vee \varphi$ so that $\mathcal{N}^{\epsilon}_{\Psi} = \mathcal{N}^{\epsilon}_{\theta\vee\varphi} = \mathcal{N}^{\epsilon}_{\theta} \cup \mathcal{N}^{\epsilon}_{\varphi}$ (by definition 6) $= \{x : \mathcal{D}^{\epsilon}_{x} \in \lambda(\theta)\} \cup \{x : \mathcal{D}^{\epsilon}_{x} \in \lambda(\varphi)\}$ (by the inductive hypothesis) $= \{x : \mathcal{D}^{\epsilon}_{x} \in \lambda(\theta) \cup \lambda(\varphi)\} = \{x : \mathcal{D}^{\epsilon}_{x} \in \lambda(\theta \vee \varphi)\}$ (by definition 3).
- $\Psi = \neg\theta$ so that $\mathcal{N}^{\epsilon}_{\Psi} = \mathcal{N}^{\epsilon}_{\neg\theta} = (\mathcal{N}^{\epsilon}_{\theta})^c$ (by definition 6) $= \{x : \mathcal{D}^{\epsilon}_{x} \in \lambda(\theta)\}^c$ (by the inductive hypothesis) $= \{x : \mathcal{D}^{\epsilon}_{x} \notin \lambda(\theta)\} = \{x : \mathcal{D}^{\epsilon}_{x} \in \lambda(\theta)^c\} = \{x : \mathcal{D}^{\epsilon}_{x} \in \lambda(\neg\theta)\}$ (by definition 3).

Now by definition 4 we have that $\forall\theta \in LE$ $\mu_\theta(x) = \sum_{F\in\lambda(\theta)} m_x(F) = \sum_{F\in\lambda(\theta)} \delta(\{\epsilon : \mathcal{D}^{\epsilon}_{x} = F\})$ (by definition 5) $= \delta(\{\epsilon : \mathcal{D}^{\epsilon}_{x} \in \lambda(\theta)\}) = \delta(\{\epsilon : x \in \mathcal{N}^{\epsilon}_{\theta}\})$ (by above). □

For example, for $L_i \in LA$ $\mathcal{N}^{\epsilon}_{L_i} = \{x : d(x, P_i) \leq \epsilon\}$. Hence, $\mu_{L_i}(x) = \Delta(d(x, P_i))$ where $\Delta(\epsilon) = \delta([\epsilon, \infty))$.

Theorem 1 shows a clear link between appropriateness measures and Goodman and Nguyen's characterisation of fuzzy set membership functions as single point coverage functions of random sets [1], [2], [8].

Theorem 2. *Restricted Consonance [7]*
Let $LE^{\wedge,\vee}$ be those expressions in LE which can be generated from LA using only the connectives $\wedge$ and $\vee$. Then $\forall\theta \in LE^{\wedge,\vee}$, $\forall 0 \leq \epsilon \leq \epsilon'$ $\mathcal{N}^{\epsilon}_{\theta} \subseteq \mathcal{N}^{\epsilon'}_{\theta}$

Proof. Let $LE^{\wedge,\vee,(1)} = LA$ and for $k > 1$ let $LE^{\wedge,\vee,(k)} = LE^{\wedge,\vee,(k-1)} \cup \{\theta \wedge \varphi, \theta \vee \varphi : \theta, \varphi \in LE^{\wedge,\vee,(k-1)}\}$. We now proceed by induction on k.
Limit Case: $k = 1$ For $L_i \in LA$, since $\epsilon' \geq \epsilon$ then trivially $\mathcal{N}^{\epsilon}_{L_i} = \{x : d(x, P_i) \leq \epsilon\} \subseteq \{x : d(x, P_i) \leq \epsilon'\} = \mathcal{N}^{\epsilon'}_{L_i}$
Inductive Step: Assume true for k For $\Psi \in LE^{(k+1)}$ either $\Psi \in LE^{(k)}$, in which case the result holds trivially by the inductive hypothesis, or one of the following holds for $\theta, \varphi \in LE^{(k)}$:

- $\Psi = \theta \wedge \varphi$: In this case $\mathcal{N}^{\epsilon}_{\Psi} = \mathcal{N}^{\epsilon}_{\theta\wedge\varphi} = \mathcal{N}^{\epsilon}_{\theta} \cap \mathcal{N}^{\epsilon}_{\varphi}$ (by definition 6) $\subseteq \mathcal{N}^{\epsilon'}_{\theta} \cap \mathcal{N}^{\epsilon'}_{\varphi}$ (by the inductive hypothesis) $= \mathcal{N}^{\epsilon'}_{\Psi}$ (by definition 6).
- $\Psi = \theta \vee \varphi$: In this case $\mathcal{N}^{\epsilon}_{\Psi} = \mathcal{N}^{\epsilon}_{\theta\vee\varphi} = \mathcal{N}^{\epsilon}_{\theta} \cup \mathcal{N}^{\epsilon}_{\varphi}$ (by definition 6) $\subseteq \mathcal{N}^{\epsilon'}_{\theta} \cup \mathcal{N}^{\epsilon'}_{\varphi}$ (by the inductive hypothesis) $= \mathcal{N}^{\epsilon'}_{\Psi}$ (by definition 6).

□

3 Imprecise Descriptions of Real Numbers

In this section we apply label semantics and prototype theory to model inference from imprecise descriptions of real numbers. Adopting random set neighbourhoods to represent extensions of concepts we will consider what imprecise probabilities result from conditioning given linguistic descriptions of a real variable. This approach is grounded in a clear interpretation of vague linguistic descriptions, in contrast to fuzzy methods in which membership functions and consequently probabilities of fuzzy events have no clear operational semantics [4].

Here we take $\Omega = \mathbb{R}$ and $d(x, y) = \|x - y\|$ and we consider descriptions based on number labels of the following form:

Definition 7. *Number Labels*
We consider a set LA of number labels L_i describing $\mathbb{R}$ with prototype sets P_i each corresponding to an interval of $\mathbb{R}$

The appropriateness measure for a number expressions $\theta \in LE$ (generated as in definition 1) is defined directly as the single point coverage function of $\mathcal{N}^{\epsilon}_{\theta}$ as in theorem 1. This allows us to relax the requirement in label semantics that LA is finite.

Here we particularly consider appropriateness measures generated by two types of density δ; normal distributions and uniform distributions.

Let $f(c, \sigma, \epsilon)$ denote the normal density function with mean c and standard deviation σ so that:

$$f(c, \sigma, \epsilon) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\epsilon-c)^2}{2\sigma^2}}$$

From this we can define a density δ as a normalised normal density of the form:

$$\delta(c, \sigma, \epsilon) = \frac{f(c, \sigma, \epsilon)}{1 - k} \text{ where } k = \int_{-\infty}^{0} f(c, \sigma, \epsilon) d\epsilon$$

From this we also have that:

$$\Delta(c, \sigma, \epsilon) = \frac{erfc(\frac{\epsilon-c}{\sigma\sqrt{2}})}{erfc(\frac{-c}{\sigma\sqrt{2}})}$$

where $erfc$ is the complementary error function

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-t^2} dt$$

Now let $\mu^{(c,\sigma)}_{\theta}(x)$ denote the appropriateness measure for θ generated by a normalised normal distribution δ with mean c and standard deviation σ. Figure 1 shows the appropriateness measure for a number label with prototypes $P_i = [5, 7]$ based on a normalised normal distribution with $c = 2$ and $\sigma = 1$.

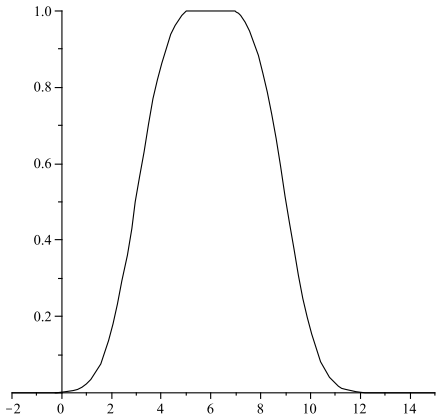

Figure 1: Appropriateness for label with prototypes $[5, 7]$ generated by a normalised normal distribution with mean $c = 2$ and standard deviation $\sigma = 1$

Theorem 3. *Let $c \leq c'$ then for $L_i \in LA$ it holds that:*

$$\forall x \in \mathbb{R}, \forall \sigma \in \mathbb{R}\ \mu_{L_i}^{(c,\sigma)}(x) \leq \mu_{L_i}^{(c',\sigma)}(x)$$

Proof. It is sufficient to show that $\Delta(c, \sigma, \epsilon)$ is an increasing function of c. Let $t = \frac{c}{\sigma\sqrt{2}}$ and $s = \frac{\epsilon}{\sigma\sqrt{2}}$ then:

$$\Delta(c, \sigma, \epsilon) = h(t, s) = \frac{erfc(s-t)}{erfc(-t)}$$

Hence it is sufficient to show that h is an increasing function of t.

$$\frac{\partial h}{\partial t} = \frac{2e^{-(s-t)^2}}{\sqrt{2}erfc(t)} + \frac{2erfc(s-t)e^{-t^2}}{erfc(t)^2\sqrt{\pi}} \geq 0$$

as required. □

Another interesting case is where δ is the uniform distribution on an interval $[k, r]$ for $r > k \geq 0$. This results in trapezoidal (or triangular) appropriateness measures. In this case we have:

$$\delta(k, r, \epsilon) = \begin{cases} 0 : \epsilon < k \\ \frac{1}{r-k} : \epsilon \in [k, r] \\ 0 : \epsilon > r \end{cases}$$

$$\text{and } \Delta(k, r, \epsilon) = \begin{cases} 1 : \epsilon < k \\ \frac{r-\epsilon}{r-k} : \epsilon \in [k, r] \\ 0 : \epsilon > r \end{cases}$$

Now let $\mu_\theta^{(k,r)}(x)$ denote the appropriateness measure for θ generated by a uniform distribution δ on $[k, r]$. Figure 2 shows the appropriateness for a number label with prototypes $[a, b]$ based on a uniform δ.

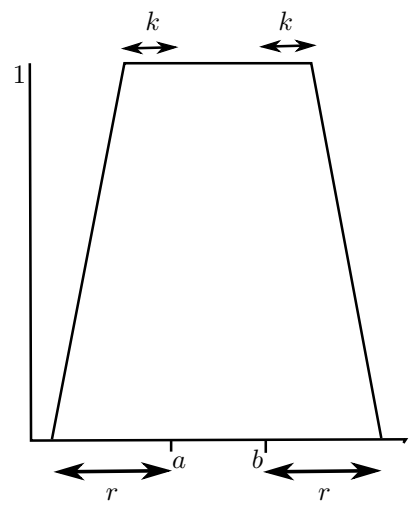

Figure 2: Appropriateness for label with prototypes $[a, b]$ generated by a uniform distribution δ on $[k, r]$

Theorem 4. *Let $0 \leq k < r$, $0 \leq k' < r'$, $k \leq k'$, and $r \leq r'$ then for $L_i \in LA$ it holds that:*

$$\forall x \in \mathbb{R}\ \mu_{L_i}^{(k,r)}(x) \leq \mu_{L_i}^{(k',r')}(x)$$

Proof. Trivially from the above it holds that $\forall \epsilon \geq 0$ $\Delta(k, r, \epsilon) \leq \Delta(k', r', \epsilon)$ and hence $\Delta(k, r, d(x, P_i)) \leq \Delta(k', r', d(x, P_i))$ as required. □

4 Information from Imprecise Descriptions

In this section we discuss the issue of conditioning given information in the form of imprecise descriptions of a real valued variable x. In other words, suppose we learn that 'x is *high*' or 'x is *high* $\wedge$ $\neg$*very high*' or more generally 'x is θ', what can we infer from such information about the value of x? To answer this question it is necessary to have a clear operational interpretation of statements 'x is θ'. For example, Zadeh [14] proposes that such statements define a possibility distribution on x when imprecise descriptions are represented by fuzzy sets. However, such a claim remains unconvincing while there is no clear operational meaning for fuzzy set membership functions. For the prototype model proposed in this paper a statement 'x is θ' is clearly interpreted as $x \in \mathcal{N}_\theta^\epsilon$. In other words, an imprecise description of x

restricts x to a random set neighbourhood generated by that description. Consequently, given the information 'x is θ' the remaining uncertainty concerning the value of x has two distinct sources. Firstly, for a specific value of ϵ, $\mathcal{N}_\theta^\epsilon$ is imprecise in the sense that typically it is a union of intervals of $\mathbb{R}$ rather than a precise value. Secondly, the value of the threshold ϵ is uncertain resulting in uncertainty about the definition of $\mathcal{N}_\theta^\epsilon$. Here, we shall argue that in the absence of any further information about x these two sources of uncertainty naturally result in lower and upper cumulative distributions. Furthermore, in the presence of a known prior probability distribution on x, conditioning on 'x is θ' results in a second order probability distribution on the cumulative probabilities for x.

Definition 8. *Upper and Lower Distributions*
Given a real valued random variable x for which we know only that 'x is θ' for some $\theta \in LE$ we define upper and lower cumulative distribution functions for the probability that $x \leq y$ as follows:

$$\underline{F}(y|\theta) = \delta_\theta(\{\epsilon : \mathcal{N}_\theta^\epsilon \subseteq (-\infty, y]\}) \text{ and}$$
$$\overline{F}(y|\theta) = \delta_\theta(\{\epsilon : \mathcal{N}_\theta^\epsilon \cap (-\infty, y] \neq \emptyset\}) \text{ and where}$$
$$\delta_\theta(\epsilon) = \begin{cases} \frac{\delta(\epsilon)}{\int_{\epsilon:\mathcal{N}_\theta^\epsilon \neq \emptyset} \delta(\epsilon) d\epsilon} : \mathcal{N}_\theta^\epsilon \neq \emptyset \\ 0 : \text{otherwise} \end{cases}$$

In definition 8 δ_θ is the posterior density on ϵ resulting from updating δ based on the information that $\mathcal{N}_\theta^\epsilon \neq \emptyset$. A possible justification for this normalisation of δ is that if we learn 'x is θ' this would intuitively imply that $\mathcal{N}_\theta^\epsilon \neq \emptyset$ since otherwise our information would be contradictory. In other words, accepting the assertion 'x is θ' implicitly implies accepting that the threshold ϵ must be such that $\mathcal{N}_\theta^\epsilon \neq \emptyset$. Clearly such conditioning is only possible if $\delta(\{\epsilon : \mathcal{N}_\theta^\epsilon \neq \emptyset\}) > 0$ otherwise the lower and upper probabilities given in definition 8 are undefined.

Theorem 5. *For $\theta \in LE^{\wedge,\vee}$ then $\forall y \in \mathbb{R}$*

$$\overline{F}(y|\theta) = \sup\{w\mu_\theta(x) : x \leq y\}$$
$$\underline{F}(y|\theta) = 1 - \sup\{w\mu_\theta(x) : x > y\}$$
$$\text{where } w = \frac{1}{\int_{\epsilon:\mathcal{N}_\theta^\epsilon \neq \emptyset} \delta(\epsilon) d\epsilon}$$

Proof. Straightforward from theorem 2 and definition 8 □

Theorem 5 shows that for θ not involving negation $\underline{F}(y|\theta)$ and $\overline{F}(y|\theta)$ are necessity and possibility measures respectively generated by the normalised possibility distribution $w\mu_\theta(x)$.

Corollary 1. *Let $\overline{F}^{(c,\sigma)}$ and $\underline{F}^{(c,\sigma)}$ be the upper and lower cumulative distributions as given in definition 8 and where δ is the normalised normal distribution with parameters c and σ. Then $\forall L_i \in LA$, $\forall c \leq c'$, $\forall \sigma \in \mathbb{R}$, $\forall y \in \mathbb{R}$*

$$\underline{F}^{(c',\sigma)}(y|L_i) \leq \underline{F}^{(c,\sigma)}(y|L_i) \text{ and}$$
$$\overline{F}^{(c,\sigma)}(y|L_i) \leq \overline{F}^{(c',\sigma)}(y|L_i)$$

Proof. Straightforward from theorems 5 and 3 □

Corollary 2. *Let $\overline{F}^{(k,r)}$ and $\underline{F}^{(k,r)}$ be the upper and lower cumulative distributions as given in definition 8 and where δ is the a uniform distribution on $[k, r]$. Then $\forall L_i \in LA$, $0 \leq k < r$, $0 \leq k' < r'$, $k \leq k'$, and $r \leq r'$, $\forall y \in \mathbb{R}$*

$$\underline{F}^{(k',r')}(y|L_i) \leq \underline{F}^{(k,r)}(y|L_i) \text{ and}$$
$$\overline{F}^{(k,r)}(y|L_i) \leq \overline{F}^{(k',r')}(y|L_i)$$

Proof. Straightforward from theorems 5 and 4 □

Now suppose we have prior information that x is distributed according to density function $p(x)$. In this case if we learn 'x is θ' then we should generate a posterior distribution by updating $p(x)$ given the new constraint that $x \in \mathcal{N}_\theta^\epsilon$. Let $F(y|\mathcal{N}_\theta^\epsilon)$ denote the corresponding updated cumulative distribution. However, the values of $F(y|\mathcal{N}_\theta^\epsilon)$ are uncertain given the remaining uncertainty about the value of the threshold ϵ. Hence, updating a prior distribution on x given an imprecise description of x results in a second order probability distribution as follows:

Definition 9. *Second Order Distribution*
Given a prior density $p(x)$ for x we define a second order cumulative distribution on the cumulative probability that $x \leq y$ as follows: $\forall p \in [0, 1]$

$$\tilde{F}_{y,\theta}(p) = \delta_\theta(\{\epsilon : F(y|N_\theta^\epsilon) \leq p\}) \text{ where}$$
$$F(y|N_\theta^\epsilon) = \int_{-\infty}^{y} p(x|\mathcal{N}_\theta^\epsilon) dx \text{ and where}$$
$$p(x|\mathcal{N}_\theta^\epsilon) = \begin{cases} \frac{p(x)}{\int_{\mathcal{N}_\theta^\epsilon} p(x) dx} : x \in \mathcal{N}_\theta^\epsilon \\ 0 : \text{otherwise} \end{cases}$$

If a precise posterior distribution is required conditional on θ, then one possibility is to take the expected value of posterior distributions given $\mathcal{N}_\theta^\epsilon$, as ϵ varies.

Definition 10. *Expected Density*
Given prior density $p(x)$ for x we can define an expected density for x conditional on θ by taking the

expected value of $p(x|\mathcal{N}_\theta^\epsilon)$ as ϵ varies:

$$p(x|\theta) = E_{\delta_\theta}(p(x|\mathcal{N}_\theta^\epsilon))$$

Notice that the above is a clearly motivated definition of conditional probability given imprecise linguistic information, consistent with a random set and prototype theory view of vague concepts. This is a distinct advantage over earlier work on the probability of fuzzy events [13], in which definitions do not appear to be linked to any underlying interpretation of fuzziness.

The following theorem shows that the expected cumulative distribution obtained from definition 10 is consistent with the lower and upper distributions given in definition 8.

Theorem 6. *For $y \in \mathbb{R}$ and $\theta \in LE$, $\underline{F}(y|\theta) \leq F(y|\theta) \leq \overline{F}(y|\theta)$ where $F(y|\theta) = \int_{-\infty}^{y} p(x|\theta)dx = E_{\delta_\theta}(F(y|\mathcal{N}_\theta^\epsilon))$.*

Proof.

$$\begin{aligned} F(y|\theta) = \int_{-\infty}^{y} p(x|\theta)dx &= \int_{-\infty}^{y}\int_{0}^{\infty} p(x|\mathcal{N}_\theta^\epsilon)\delta_\theta(\epsilon)d\epsilon dx \\ &= \int_{0}^{\infty} F(y|\mathcal{N}_\theta^\epsilon)\delta_\theta(\epsilon)d\epsilon \\ &= \int_{\epsilon:\mathcal{N}_\theta^\epsilon\cap(-\infty,y]\neq\emptyset} F(y|\mathcal{N}_\theta^\epsilon)\delta_\theta(\epsilon)d\epsilon \\ &\leq \int_{\epsilon:\mathcal{N}_\theta^\epsilon\cap(-\infty,y]\neq\emptyset} \delta_\theta(\epsilon)d\epsilon = \overline{F}(y|\theta) \end{aligned}$$

Alternatively

$$\begin{aligned} F(y|\theta) &= \int_{\epsilon:\mathcal{N}_\theta^\epsilon\subseteq(-\infty,y]} F(y|\mathcal{N}_\theta^\epsilon)\delta_\theta(\epsilon)d\epsilon + \\ &\quad \int_{\epsilon:\mathcal{N}_\theta^\epsilon\not\subseteq(-\infty,y]} F(y|\mathcal{N}_\theta^\epsilon)\delta_\theta(\epsilon)d\epsilon \\ &= \int_{\epsilon:\mathcal{N}_\theta^\epsilon\subseteq(-\infty,y]} \delta_\theta(\epsilon)d\epsilon + \int_{\epsilon:\mathcal{N}_\theta^\epsilon\not\subseteq(-\infty,y]} F(y|\mathcal{N}_\theta^\epsilon)\delta_\theta(\epsilon)d\epsilon \\ &\geq \int_{\epsilon:\mathcal{N}_\theta^\epsilon\subseteq(-\infty,y]} \delta_\theta(\epsilon)d\epsilon = \underline{F}(y|\theta) \end{aligned}$$

□

Example 1. *Consider the number label L_i = about 2 for which $P_i = \{2\}$. Let $\delta(\epsilon) = \begin{cases} 1 : \epsilon \in [0,1] \\ 0 : otherwise \end{cases}$ then the lower and upper cumulative distributions given the information 'x is about 2' are as follows:*

$$\underline{F}(y|L_i) = \begin{cases} 0 : y \leq 2 \\ 1 - \mu_{L_i}(y) : y > 2 \end{cases} \quad \text{and}$$

$$\overline{F}(y|L_i) = \begin{cases} 0 : y \leq 1 \\ \mu_{L_i}(y) : 1 < y \leq 2 \\ 1 : y > 1 \end{cases} \quad \text{and where}$$

$$\mu_{L_i}(y) = \begin{cases} 0 : x < 1 \\ x - 1 : x \in [1,2] \\ 3 - x : x \in (2,3] \\ 0 : x > 3 \end{cases}$$

Suppose we now further learn that x is distributed according to a uniform distribution on $[0,10]$ then we can infer a second order distribution the probability that $x \leq y$ as follows: Initially note that $\mathcal{N}_{L_i}^\epsilon = [2-\epsilon, 2+\epsilon]$ so that for $\epsilon \leq 1$

$$p(x|\mathcal{N}_{L_i}^\epsilon) = \begin{cases} \frac{1}{2\epsilon} : x \in [2-\epsilon, 2+\epsilon] \\ 0 : otherwise \end{cases} \quad \text{and hence}$$

$$F(y|\mathcal{N}_{L_i}^\epsilon) = \begin{cases} 1 : y > 2+\epsilon \\ \frac{y+\epsilon-2}{2\epsilon} : y \in [2-\epsilon, 2+\epsilon] \\ 0 : y < 2-\epsilon \end{cases}$$

From this we obtain four cases of $\tilde{F}_{y,L_i}$ as follows:

For $y < 1$

$$\forall p \in [0,1] \ \tilde{F}_{y,L_i}(p) = 1$$

For $1 \leq y \leq 2$ (see figure 3)

$$\tilde{F}_{y,L_i}(p) = \begin{cases} 1 : p > \frac{y-1}{2} \\ \frac{2-y}{1-2p} : p \leq \frac{y-1}{2} \end{cases}$$

For $2 < y \leq 3$ (see figure 4)

$$\tilde{F}_{y,L_i}(p) = \begin{cases} 0 : p < \frac{y-1}{2} \\ \frac{2p-y+1}{2p-1} : \frac{y-1}{2} \leq p < 1 \\ 1 : p = 1 \end{cases}$$

For $y > 3$

$$\tilde{F}_{y,L_i}(p) = \begin{cases} 0 : p < 1 \\ 1 : p = 1 \end{cases}$$

The expected density $p(x|L_i)$ is given by (figure 5):

$$p(x|L_i) = \begin{cases} 0 : x < 1 \\ -\frac{1}{2}\ln(2-x) : 1 \leq x < 2 \\ -\frac{1}{2}\ln(x-2) : 2 < x \leq 3 \\ 0 : x > 3 \end{cases}$$

Figure 6 shows the upper and lower cumulative distributions given 'x is about 2' together with the expected cumulative distribution assuming that x is distributed according to a uniform distribution on $[0,10]$.

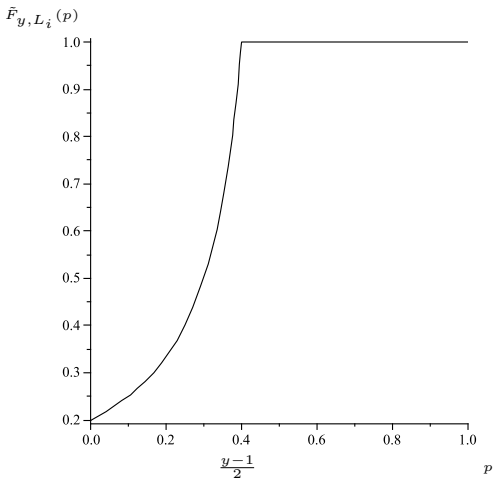

Figure 3: $\tilde{F}_{y,L_i}$ for $1 < y \leq 2$

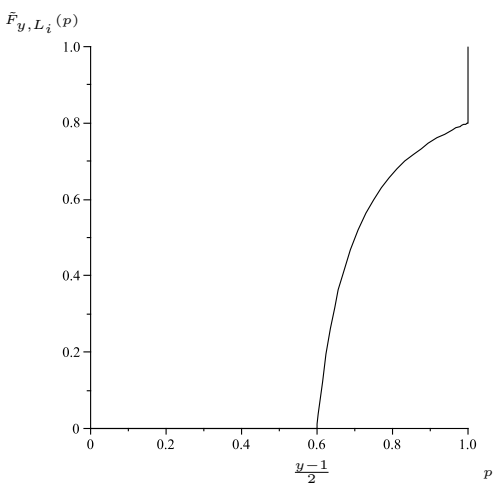

Figure 4: $\tilde{F}_{y,L_i}$ for $2 < y \leq 3$

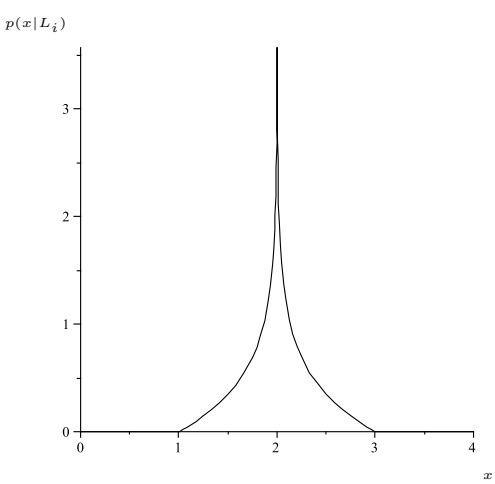

Figure 5: Expected density for x conditional on L_i

5 Information from Imprecise Data

In this section we consider inference on the basis of data taking the form of imprecise descriptions of real

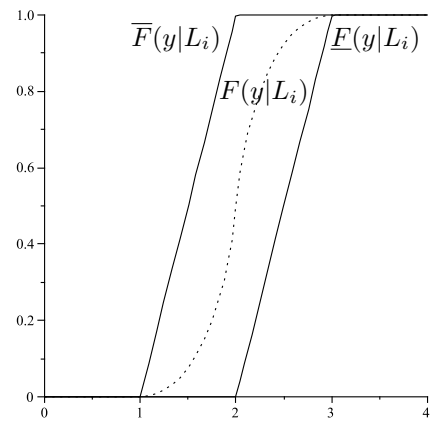

Figure 6: Lower and upper cumulative distribution $\underline{F}(y|L_i)$, $\overline{F}(y|L_i)$ together with expected cumulative distribution $F(y|L_i)$ assuming a uniform prior on x

values. Let x be a real valued random variable, and let $DB = \{\theta_1, \ldots, \theta_N\}$ where $\theta_i \in LE$ be a set of independently generated descriptions of x. Given DB we define lower and upper cumulative distributions for x as follows:

Definition 11.

$$\forall y \in \mathbb{R} \ \underline{F}(y|DB) = \frac{1}{N}\sum_{i=1}^{N} \underline{F}(y|\theta_i)$$

$$\overline{F}(y|DB) == \frac{1}{N}\sum_{i=1}^{N} \overline{F}(y|\theta_i)$$

Definition 12. *Given a density $p(x)$ we can define a expected density cumulative distribution conditional on DB according to:*

$$\forall x \in \mathbb{R} \ p(x|DB) = \frac{1}{N}\sum_{i=1}^{N} p(x|\theta_i) \ \textit{and}$$

$$\forall y \in \mathbb{R} \ F(y|DB) = \frac{1}{N}\sum_{i=1}^{N} F(y|\theta_i)$$

The underlying intuition behind these definitions is as follows: In order to estimate x one approach would be to randomly select a description θ_i from DB and then condition on the information 'x is θ_i'. Assuming that each element of DB is equally likely to be selected (i.e. has equal weighting) then the expected information we would learn about x is as given in definitions 11 and 12.

One natural example of this approach is where we have an independent sample $\{x_1, \ldots, x_N\}$ of values of x for which we are assuming there is an associated

uncertain error ϵ with density δ, so that each x_i effectively identifies a random set interval $[x_i - \epsilon, x_i + \epsilon]$. In this case we define $DB = \{L_1, \ldots, L_N\}$ where L_i is a number label with prototype $P_i = \{x_i\}$ (i.e. $L_i = about\ x_i$).

Example 2. *A sample of* 100 *values was drawn at random from the normal mixture distribution* $g = \frac{N(2,3)+N(8,0.5)}{2}$. *DB was then taken to correspond to the set of labels* L_i *with prototype* $P_i = \{x_i\}$ *for each value* x_i *in the sample.* δ *was assumed to be a uniform distribution on* $[k, r]$ *where* k *and* r *are effectively treated as parameters in the estimating of distributions from* DB.

To compare the upper and lower cumulative distributions obtained from DB *with that of the generating distribution* g *we introduce two measure as follows:*

$$IE := \frac{1}{N}\sum_{i=1}^{N} \chi_{[\underline{F}(x_i|DB),\overline{F}(x_i|DB)]}(G(x_i))$$

where $\chi_{[\underline{F}(x_i|DB),\overline{F}(x_i|DB)]}$ *is the characteristic function for the interval* $[\underline{F}(x_i|DB), \overline{F}(x_i|DB)]$ *and* G *is the cumulative distribution function for density* g. *Hence,* IE *provides a measure of the extent to which the generating cumulative density* G *is contained within the estimated upper and lower envelope across the original sample.*

We also evaluate the average range of the upper and lower distribution envelope according to:

$$Range = \frac{1}{N}\sum_{i=1}^{N}(\overline{F}(x_i|DB) - \underline{F}(x_i|DB))$$

Table 1 shows the IE *and Range values for a number of different* k, r *values. Notice that by corollary 2 it follows immediately that as* k *and* r *increase the* IE *values decrease. Figure 7 shows the upper and lower envelope together with* G *for* $k = 0$ *and* $r = 2.2$, *these corresponding to the values in table 1 for which* IE *is* 0 *and Range is minimal.*

Table 2 compares $p(x|DB)$ *with* $g(x)$ *according to* MSE *defined as follows:*

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(p(x_i|DB) - g(x_i))^2$$

Figure 8 shows $p(x|DB)$ *and* $g(x)$ *for* $k = 0.4$, $r = 0.5$ *these corresponding to the values in table 2 with lowest* MSE.

6 Summary and Conclusions

The prototype theory interpretation of label semantics has been introduced as a possible model for imprecise descriptions of real numbers. Based on this interpretation it has been shown that conditioning given information in the form 'x is θ', for $\theta \in LE$, naturally results in imprecise probabilities. Also, within this framework, we have proposed a possible approach to inference from data in the form of imprecise descrip-

Table 1: Table showing IE and *Range* for different values of k and r

k	r	IE	*Range*
0.7	1.1	0.1	0.2926
0.8	1.1	0.09	0.3036
0.9	1.1	0.07	0.31324
1	1.1	0.03	0.3221
0.8	1.2	0.04	0.3122
0.9	1.2	0.02	0.3213
1	1.2	0.01	0.3298
1.1	1.2	0	0.3584
0.9	1.3	0.01	0.3286
1	1.3	0	0.3367
0.9	1.4	0.01	0.3286
0.7	1.4	0.02	0.3171
0.8	1.4	0	0.3245
0.6	1.5	0.02	0.31413
0.7	1.5	0	0.3245
0.6	1.6	0	0.3212
0.5	1.7	0	0.3176
0.4	1.8	0	0.3135
0.3	1.9	0	0.3087
0.2	2	0	0.3034
0.1	2.1	0	0.2979
0	2.2	0	0.2920

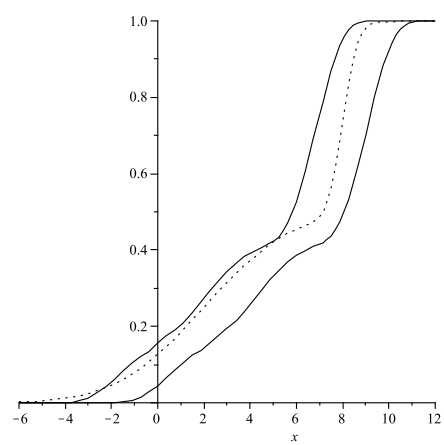

Figure 7: Upper and lower cumulative distributions based on uniform δ with $k = 0$ and $r = 2.2$, compared with cumulative distribution for the generating distribution g (dashed line)

Table 2: Table showing MSE for different values of k and r

k	r	MSE
0.1	0.2	0.0039
0.1	0.3	0.00229
0.2	0.3	0.001689
0.1	0.4	0.001596
0.2	0.4	0.001224
0.3	0.4	0.001005
0.1	0.5	0.00119
0.2	0.5	0.000929
0.3	0.5	0.00077
0.4	0.5	0.000698
0.1	0.6	0.000968
0.2	0.6	0.000817
0.3	0.6	0.000733
0.4	0.6	0.0007535
0.5	0.6	0.000925

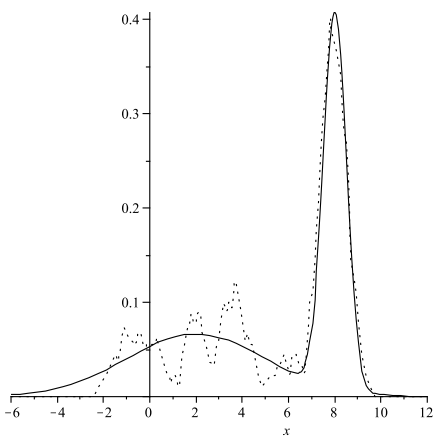

Figure 8: Density estimate based on uniform δ with $k = 0.4$ and $r = 0.5$ (dashed line), compared with generating distribution g

tions of a real variable. This naturally suggests an algorithm of estimating distributions given imprecise data values.

Acknowledgements

Inés González-Rodríguez is supported by MEC-Programa José Castillejo Grant JC2007-00152 and MEC-FEDER Grant TIN2007-67466-C02-01.
Yongchuan Tang is funded by the National Natural Science Foundation of China (NSFC) under Grant 60604034, the joint funding of NSFC and MSRA under Grant 60776798, and the Science and Technology Program of Zhejiang Province under Grant 2007C223061

References

[1] I.R. Goodman, H.T. Nguyen, (1985), *Uncertainty Model for Knowledge Based Systems*, North Holland.

[2] I.R. Goodman, (1982), 'Fuzzy Sets as Equivalence Classes of Random Sets', in *Fuzzy Set and Possibility Theory* (ed. R. Yager), pp327-342

[3] J. Lawry, (2004), 'A Framework for Linguistic Modelling', *Artificial Intelligence*, Vol. 155, pp1-39.

[4] J. Lawry, (2006), *Modelling and Reasoning with Vague Concepts*, Springer

[5] J. Lawry, (2008), 'Appropriateness Measures: An Uncertainty Model for Vague Concepts', *Synthese*, Vol. 161, No.2, pp255-269

[6] J. Lawry, Y. Tang, (2008), 'Relating Prototype Theory and Label Semantics', in *Soft Methods for Handling Variability and Imprecision*, (Eds. D. Dubois, M.A. Lubiano, H. Prade, M.A. Gil, P. Grzegorzewski, O. Hryniewicz), pp35-42

[7] J. Lawry, Y. Tang, (2009), 'Uncertainty Modelling for Vague Concepts: A Prototype Theory Approach', *submitted*

[8] H.T. Nguyen, (1984), 'On Modeling of Linguistic Information using Random Sets', *Information Science*, Vol. 34, pp265-274

[9] E.H. Rosch, (1975), 'Cognitive Representation of Semantic Categories', *Journal of Experimental Psychology: General*, Vol. 104, pp192-233

[10] Y. Tang, J. Lawry, (2009), 'Linguistic Modelling and Information Coarsening based on Prototype Theory and Label Semantics', *International Journal of Approximate Reasoning*, in Press

[11] P. Walley, G. de Cooman, (2001), 'A Behavioural Model of Linguistic Uncertainty', *Information Sciences*, Vol. 134, pp1-37

[12] T. Williamson, (1994), *Vagueness*, Routledge.

[13] L.A. Zadeh, (1968), 'Probability Measures of Fuzzy Events' *Journal of Mathematical Analysis and Applications* Vol. 23, pp421-427

[14] L.A. Zadeh, (1978), 'Fuzzy Sets as a Basis for a Theory of Possibility', *Fuzzy Sets and Systems*, Vol. 1, pp3-28