Reliable Computing **10:** 249–272, 2004.

Clouds, Fuzzy Sets, and Probability Intervals

ARNOLD NEUMAIER
Institut für Mathematik, Universität Wien, Strudlhofgasse 4, A-1090 Wien, Austria,
e-mail: Arnold.Neumaier@univie.ac.at

(Received: 14 October 2002; accepted: 22 February 2003)

Abstract. Clouds are a concept for uncertainty mediating between the concept of a fuzzy set and that of a probability distribution. A cloud is to a random variable more or less what an interval is to a number. We discuss the basic theoretical and numerical properties of clouds, and relate them to histograms, cumulative distribution functions, and likelihood ratios.

We show how to compute nonlinear transformations of clouds, using global optimization and constraint satisfaction techniques. We also show how to compute rigorous enclosures for the expectation of arbitrary functions of random variables, and for probabilities of arbitrary statements involving random variables, even for problems involving more than a few variables.

Finally, we relate clouds to concepts from fuzzy set theory, in particular to the consistent possibility and necessity measures of Jamison and Lodwick.

1. Introduction

This paper proposes the concept of a cloud, combining aspects of fuzzy sets, interval analysis, and probability theory in a way suitable for calculations.

The need for a conceptual basis for combining probabilistic uncertainty, due to variability, and fuzzy uncertainty, due to missing information or missing precision in concepts, models, or measurements, has been known for a long time. The deficiencies of Monte Carlo methods in the face of partial ignorance are well documented by Ferson et al. [10], [11]. A number of alternatives have been explored. The history until about 10 years ago, and the problems involved in the various approaches known are excellently described from complementary point of views in a thesis by Williamson [37] and a book by Walley [34]. In the mean time, the basic situation has not changed much, although computational advances make approaches using copulas (Springer [32]) or histograms (Moore [24]) more tractable; see, e.g., [3], [12], [14], [22]. However, their applicability is limited to problems in very low dimension.

The scope of the present paper is to define and explain a new concept for capturing a mix of probabilistic and fuzzy uncertainty that is able to handle also larger uncertainty problems by being able to reduce calculations to global optimization and constraint satisfaction problems, for which more and more efficient algorithms become available [1], [6], [16], [30].

A cloud is a new, easily visualized concept for uncertainty with a well-defined semantics, mediating between the concept of a fuzzy set (see, e.g., [9], [18]) and that of a probability distribution (see, e.g., [36]). In general, it contains more information than a fuzzy set but less than a distribution. (This is in contrast with fuzzy intervals [18, p. 16], which, although having the same formal definition, have a completely different interpretation that contains even less information than a fuzzy set.)

A cloud is to a random variable more or less what an interval is to a number. Roughly speaking, a cloud can be described as a nested family of inner and outer confidence regions to different confidence levels. It therefore contains more information than the results of most traditional statistical calculations. Moreover, this information is in a form useful for inferences under uncertainty.

Thin clouds in dimension 1 are formally almost identical to the confidence curves discussed by Birnbaum [4], [5]. But while confidence regions (and curves) are traditionally used to assess the reliability of parameter estimates, they are here used to describe regions containing with a given probability scenarios of interest. Thus, criticism of confidence curves raised by Kiefer [17] does *not* apply in the present context.

The concept of a cloud is nearly equivalent to the *consistent possibility and necessity measures* introduced by Jamison & Lodwick [15], but our concept is phrased in a more intuitive and less technical form, which is more suited for complex calculations.

The gain in flexibility is obtained by separating the algebraic aspects of clouds (related to possibility and necessity, see Section 5) from the more basic relations to random variables (and their associated measures), and by deemphasizing the set theoretic aspects in favor of formulas featuring intervals and inequalities.

Several of the set-theoretical constructions in [15] are reformulated in a form suitable for the application of traditional numerical techniques. Complex inferences are made possible by deducing optimization problems whose solutions provide optimal lower and upper bounds for probabilities and expectations. Previously, Ferson et al. [12], Kuznetsov [20] and Weichselberger [35] (cf. also Whittle [36, Chapter 12]) used optimization within research on uncertain probability, but in different contexts.

Stetter [33] discussed in the context of a backward rounding error analysis of nonlinear problems a "cloud of potential neighborhoods." This is close in spirit but not equivalent (and less precise) to our concept of a cloud.

In the following, we assume familiarity with elementary interval analysis [23], [26], [27]. The symbol $\square S$ denotes the interval hull of a set S of real numbers, and $\|x\| = \sqrt{x^T x}$ the Euclidean norm of a vector.

Section 2 introduces the basic concepts: clouds and their relations to intervals, fuzzy sets, and probability intervals.

In Section 3, we show that histogram based approaches to uncertain probability, such as advocated, e.g., in Moore [24], Berleant et al. [2], [3], Lüthi [22], are in some sense equivalent to the special case of discrete clouds.

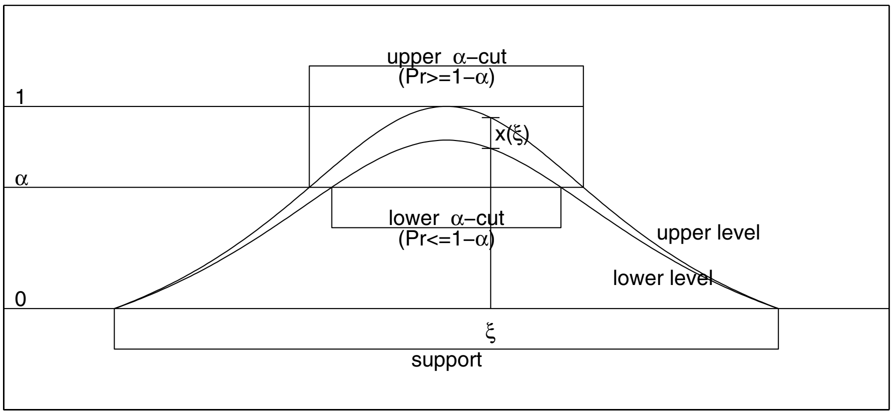

Figure 1. A cloud over $\mathbb{R}$ with an α-cut at $\alpha = 0.6$.

Similarly, we show in Section 4 that methods based on p-boxes (Ferson & Moore [13]), i.e., lower and upper cumulative distribution functions for univariate distributions are contained in the present framework of clouds. Moreover, empirical clouds for a random vector x can fairly easily be constructed when a sample of sufficiently many scenarios is available. The ease of getting good clouds from data is one of the strengths of the new concept.

Section 5 has a more theoretical character. It discusses probability intervals and properties needed to show how clouds, and some of the present development, are related to the consistent possibility and necessity measures from Jamison & Lodwick [15].

Section 6 considers the computation of functions of a random variable or vector known to lie in a cloud. The optimization approach given here is preferable to an approach based on defining operations for uncertain random variables, since it avoids the dependence problem inherent in the latter and familiar from interval arithmetic. (Interval arithmetic shows up, of course, indirectly, in methods for solving the resulting global optimization problems.)

In Section 7 we show how to compute lower and upper bounds on expectations of functions of a random vector. In Section 8, we do the same for the probability of statements defined by equations and inequalities containing a random vector belonging to a cloud. Again, in both cases, optimization problems result.

The numerical feasibility of all constructions is illustrated by applying them to the example of an exponentially decaying spherical cloud in arbitrary dimensions.

The final Section 9 lists a number of open problems for further research.

2. Clouds and Their Interpretation

Formally, cf. Figure 1, a **cloud** over a set $\mathbb{M}$ is a mapping $\mathbf{x}$ that associates with each $\xi \in \mathbb{M}$ a (nonempty, closed and bounded) interval $\mathbf{x}(\xi)$ such that

$$]0,1[\ \subseteq \bigcup_{\xi \in \mathbb{M}} \mathbf{x}(\xi) \subseteq [0,1]. \tag{2.1}$$

$\mathbf{x}(\xi) = [\underline{x}(\xi), \overline{x}(\xi)]$ is called the **level** of ξ in the cloud $\mathbf{x}$; $\underline{x}(\xi)$ and $\overline{x}(\xi)$ are the **lower** and **upper level**, respectively, and $\overline{\mathbf{x}}(\xi) - \underline{\mathbf{x}}(\xi)$ is called the **width** of ξ.

In many applications (not always, cf. Proposition 4.1, but roughly when $\overline{x}(\xi) \approx 1$ for ξ near the modes of the associated distribution), the level $\mathbf{x}(\xi)$ may be interpreted as giving lower and upper bounds on the *degree of suitability* of $\xi \in \mathbb{M}$ as a possible scenario for data modelled by the cloud $\mathbf{x}$. This degree of suitability can be given a probability interpretation by relating clouds to random variables, see (2.2) below.

A cloud is called **thin** if $\underline{x}(\xi) = \overline{x}(\xi)$ for all $\xi \in \mathbb{M}$ (the most informative case), and **fuzzy** if $\underline{x}(\xi) = 0$ for all $\xi \in \mathbb{M}$ (the least informative case). A fuzzy cloud yields a particular interpretation for a fuzzy set with membership function $\overline{x}(\xi)$. The **characteristic cloud** of a nonempty subset X of $\mathbb{M}$ is the fuzzy cloud χ_X with $\chi_X(\xi) = [0,1]$ if $\xi \in X$ and $\chi_X(\xi) = 0$ otherwise. (There is no cloud corresponding to the empty set.)

We call the cloud $\mathbf{x}$ over $\mathbb{M}$ a **subcloud** of the cloud $\mathbf{y}$ over $\mathbb{M}'$, and write $\mathbf{x} \subseteq \mathbf{y}$ ("$\mathbf{x}$ is **sharper** than $\mathbf{y}$") or $\mathbf{y} \supseteq \mathbf{x}$ ("$\mathbf{y}$ is a **relaxation** of $\mathbf{x}$") iff $\mathbb{M} \subseteq \mathbb{M}'$ and $\mathbf{x}(\xi) \subseteq \mathbf{y}(\xi)$ for all $\xi \in \mathbb{M}$. Clearly,

$$x \in \mathbf{x},\ \mathbf{x} \subseteq \mathbf{y} \quad \Rightarrow \quad x \in \mathbf{y}.$$

On the other hand, the intersection is not a natural operation on clouds; cf. Remark 3.1(*iv*) below.

A (real or complex) **cloudy number** is a cloud over the set $\mathbb{R}$ of real numbers or over the set $\mathbb{C}$ of complex numbers. $\chi_{[a,b]}$ is the cloud equivalent of an interval $[a,b]$, providing support information without additional probabilistic contents. Thus cloudy numbers are generalizations of fuzzy numbers and intervals; they convey some probabilistic information, as detailed below.

Dependence or correlation between uncertain numbers (or the lack of it) can be modelled by considering them jointly as components of a clouded vector. A (real or complex) **cloudy vector** is a cloud over $\mathbb{R}^n$ or $\mathbb{C}^n$.

We say that a random variable x with values in $\mathbb{M}$ **belongs to** a cloud $\mathbf{x}$ over $\mathbb{M}$, and write $x \in \mathbf{x}$, if

$$\Pr\big(\underline{x}(x) \geq \alpha\big) \leq 1 - \alpha \leq \Pr\big(\overline{x}(x) > \alpha\big) \qquad \text{for all } \alpha \in [0,1]. \tag{2.2}$$

Here Pr denotes the probability of the statement given as argument, and it is required that the sets consisting of all $\xi \in \mathbb{M}$ with $\underline{x}(\xi) \geq \alpha$ (resp. $\overline{x}(\xi) > \alpha$) are measurable in the σ-algebra on $\mathbb{M}$ consisting of all sets $A \subseteq \mathbb{M}$ for which $\Pr(x \in A)$ is defined.

This gives clouds a canonical interpretation as the class of random variables x with $x \in \mathbf{x}$. For fuzzy clouds, this interpretation is equivalent to the interpretation of fuzzy set membership degree as an upper bound for probabilities, first advocated by Dubois et al. [8].

There is at least one random variable belonging to any given cloud. This result is surprisingly difficult to prove; the proof will be given elsewhere (Neumaier [28]).

We may restate (2.2) by saying that the probability that a random variable $x \in \mathbf{x}$ belongs to the **upper α-cut**

$$\overline{C}_\alpha := \{\xi \in \mathbb{M} \mid \overline{x}(\xi) > \alpha\}$$

is at least $1 - \alpha$, and the probability that x belongs to the **lower α-cut**

$$\underline{C}_\alpha := \{\xi \in \mathbb{M} \mid \underline{x}(\xi) \geq \alpha\}$$

is at most $1 - \alpha$. This explains the annotation in Figure 1. Condition (2.1) ensures that the lower α-cuts with $\alpha > 0$ differ from $\mathbb{M}$ and that the upper α-cuts with $\alpha < 1$ are nonempty, which is necessary for these probability statements to make sense.

Thus the upper and lower α-cuts provide outer and inner confidence regions for $\xi \in \mathbb{M}$ belonging to $\mathbf{x}$ at given confidence levels. Note that α-cuts form a family of nested sets.

In general, the shape of a cloud may be multimodal, and their α-cuts may be disconnected. For example, it is easily checked that a random variable that takes the value ξ_0 with certainty is contained in a cloud $\mathbf{x}$ iff $\mathbf{x}(\xi_0) = [0, 1]$ (and other levels are completely arbitrary). Other, more useful clouds are constructed in Sections 3 and 4.

If $\mathbb{M}$ is a metric space, the **support** $\operatorname{Supp}\mathbf{x}$ of a cloud $\mathbf{x}$ over $\mathbb{M}$ is the closure of the set of $\xi \in \mathbb{M}$ with $\mathbf{x}(\xi) \neq 0$. A cloud $\mathbf{x}$ over $\mathbb{M}$ is **compact** if its support is compact, and **bounded** if all α-cuts with $\alpha > 0$ are bounded. Any compact cloud is bounded, but not conversely.

Although, superficially, Figure 1 looks like giving bounds on a probability density (as in the approach of Yeh [38] to approximate probability), the latter is only indirectly related to clouds. However, as we shall see in Sections 3 and 4, histograms, likelihood ratios, and continuous cumulative distribution functions of real univariate random variables can be interpreted in terms of clouds.

It is important to realize that $x \in \mathbf{x}$ implies stronger probability statements than requested in (2.2):

PROPOSITION 2.1.

(*i*) *For an arbitrary statement* $\mathrm{St}(x)$ *involving a random variable* $x \in \mathbf{x}$*, and its negation* $\neg\,\mathrm{St}(x)$*,*

$$\Pr\big(\mathrm{St}(x)\big) \leq \min\big(\sup\{\overline{x}(\xi) \mid \mathrm{St}(\xi)\},\ 1 - \inf\{\underline{x}(\xi) \mid \mathrm{St}(\xi)\}\big), \tag{2.3}$$

$$\Pr\big(\mathrm{St}(x)\big) \geq \max\big(\inf\{\underline{x}(\xi) \mid \neg\,\mathrm{St}(\xi)\},\ 1 - \sup\{\overline{x}(\xi) \mid \neg\,\mathrm{St}(\xi)\}\big). \tag{2.4}$$

(*ii*) *For an arbitrary random variable* $x \in \mathbf{x}$ *and* $\alpha \in [0, 1]$,

$$\begin{aligned} &\Pr\big(\underline{x}(x) \geq \alpha\big) \leq 1 - \alpha \leq \Pr\big(\overline{x}(x) > \alpha\big), \\ &\Pr\big(\overline{x}(x) \leq \alpha\big) \leq \quad \alpha \quad \leq \Pr\big(\underline{x}(x) < \alpha\big). \end{aligned} \tag{2.5}$$

(*iii*) *If* $\mathbf{x}$ *is thin (hence a real-valued function) then*

$$\begin{aligned} &\Pr\big(\mathbf{x}(x) < \alpha\big) = \Pr\big(\mathbf{x}(x) \leq \alpha\big) = \alpha, \\ &\Pr\big(\mathbf{x}(x) > \alpha\big) = \Pr\big(\mathbf{x}(x) \geq \alpha\big) = 1 - \alpha. \end{aligned}$$

Proof.
(*i*) Let $\alpha = \inf\{\underline{x}(\xi) | \operatorname{St}(\xi)\}$. Then $\{\xi | \operatorname{St}(\xi)\} \subseteq \underline{C}_\alpha$, so that by (2.2), $\Pr(\operatorname{St}(x)) \leq \Pr(x \in \underline{C}_\alpha) \leq 1 - \alpha$, giving $\alpha \leq \Pr(\operatorname{St}(x))$. Similarly, for $\beta = \sup\{\overline{x}(\xi) | \operatorname{St}(\xi)\}$, we have $\overline{C}_\beta \cap \{\xi | \operatorname{St}(\xi)\} = \emptyset$, and by (2.2), $\Pr(\operatorname{St}(x)) \leq 1 - \Pr(x \in \overline{C}_\beta) \leq \beta$. This proves (2.3), and (2.4) then follows by applying $1 - \Pr(\operatorname{St}(x)) = \Pr(\neg \operatorname{St}(x))$ to (2.3) with $\neg \operatorname{St}(x)$ in place of $\operatorname{St}(x)$.

(*ii*) The first formula is just the definition, and (2.5) follows since probabilities of complementary statements add up to 1.

(*iii*) The thin case is an immediate consequence of (2.5) and (2.2). □

COROLLARY 2.1. *A random variable belongs to a cloud* $\mathbf{x}$ *iff it belongs to its* **mirror cloud** $\mathbf{x}' = 1 - \mathbf{x}$ *with*

$$\mathbf{x}'(\xi) = 1 - \mathbf{x}(\xi) = [1 - \overline{x}(\xi),\ 1 - \underline{x}(\xi)] \qquad \textit{for all}\ \ \xi \in \mathbb{M}.$$

Proof. This follows from Proposition 2.1 since after substituting $1 - \alpha$ for α, (2.5) becomes (2.2) for the mirror cloud. □

But note that the mirror cloud of a compact or bounded cloud over an unbounded set is neither compact nor bounded.

3. Discrete Clouds and Histograms

We call a cloud $\mathbf{x}$ **discrete** if the level $\mathbf{x}(\xi)$ only takes finitely many values $\mathbf{a}_1, \ldots, \mathbf{a}_m$. In this case, the cloud is completely specified by these intervals together with a description of the sets

$$X_l = \{\xi \in \mathbb{M} \mid \mathbf{x}(\xi) = \mathbf{a}_l\}. \tag{3.1}$$

If these sets are sufficiently simple (for example, boxes) then one has a versatile explicit description of the cloud, which may be useful in applications.

Frequently, partial information about a random variable (a random vector) is described by histograms. For our purposes, a **histogram** is a finite, ordered collection of pairs (X_l, p_l), $l = 1, \ldots, m$, consisting of pairwise disjoint sets X_l partitioning

$\mathbb{M}$ and nonnegative relative frequencies p_l summing to one. A random variable x **belongs** to such a histogram if

$$\Pr(x \in X_l) = p_l \quad \text{for} \ \ l = 1, \ldots, m. \tag{3.2}$$

Our next result shows that histograms are in one-to-one correspondence with special discrete clouds, in such a way that the property (3.2) precisely characterizes the random variables belonging to these clouds.

In particular, this enables one to construct clouds from samples of discrete probability distributions, or (after discretization) of continuous distributions in low dimensions.

THEOREM 3.1. *Let* (X_l, p_l), $l = 1, \ldots, m$ *be a histogram, and define the cloud* $\mathbf{x}$ *by*

$$\mathbf{x}(\xi) := [\alpha_{l-1}, \alpha_l] \qquad \textit{for all} \ \ \xi \in X_l, \tag{3.3}$$

where

$$\alpha_0 = 0, \quad \alpha_l = \sum_{k \leq l} p_k \qquad \textit{for} \ \ L > 0. \tag{3.4}$$

Then

$$x \in \mathbf{x} \quad \Leftrightarrow \quad \Pr(x \in X_l) = p_l \qquad \textit{for} \ \ l = 1, \ldots, m. \tag{3.5}$$

Proof. The intervals (3.3) are well-defined since $\alpha_l - \alpha_{l-1} = p_l \geq 0$.

(i) We first prove the relation

$$\Pr(x \in X_l) \geq p_l. \tag{3.6}$$

This is trivial if $p_l = 0$, hence we may assume that $p_l > 0$. Let

$$A_l := X_1 \cup \cdots \cup X_l, \qquad A_0 = \emptyset.$$

The lower bound of (2.5) gives for $\alpha = \alpha_{k-1}$ the bound

$$\Pr(x \in A_{k-1}) = \Pr\Bigg(x \in \bigcup_{\alpha_l \leq \alpha_{l-1}} X_l\Bigg) = \Pr\big(\overline{x}(x) \leq \alpha_{k-1}\big) \leq \alpha_{k-1}.$$

The upper bound of (2.5) gives for $\alpha_{k-1} < \alpha < \alpha_k$ the bound

$$\Pr(x \in A_k) = \Pr\Bigg(x \in \bigcup_{\alpha_{l-1} < \alpha} X_l\Bigg) = \Pr\big(\underline{x}(x) < \alpha\big) \geq \alpha.$$

Taking the limit $\alpha \to \alpha_k$, we find $\Pr(x \in A_k) \geq \alpha_k$. Therefore

$$\Pr(x \in X_k) = \Pr(x \in A_k) - \Pr(x \in A_{k-1}) \geq \alpha_k - \alpha_{k-1} = p_k.$$

Thus (3.6) holds in general. If strict inequality holds for some l then summation over all l gives the contradiction

$$\begin{aligned} 1 &= \sum_{l=1}^{m} p_l < \sum_{l=1}^{m} \Pr(x \in X_l) = \Pr\Big(x \in \bigcup_{l=1}^{m} X_l\Big) \\ &= \Pr(x \in \mathbb{M}) = 1. \end{aligned}$$

Therefore, this is impossible, and (3.6) holds with equality. This proves the forward implication of (3.5).

(*ii*) Conversely, assume the right hand side of (3.5), and let $\alpha \in [0, 1]$. With k such that $\alpha_k \leq \alpha < \alpha_{k+1}$, we have

$$\begin{aligned} \Pr(\overline{x}(x) > \alpha) &= \Pr\Big(x \in \bigcup_{\alpha_l > \alpha} X_l\Big) = \sum_{\alpha_l > \alpha} p_l \\ &= \sum_{l>k} p_l = 1 - \alpha_k \geq 1 - \alpha. \end{aligned}$$

Similarly, with k such that $\alpha_k < \alpha \leq \alpha_{k+1}$, we have

$$\begin{aligned} \Pr(\underline{x}(x) \geq \alpha) &= \Pr\Big(x \in \bigcup_{\alpha_{l-1} \geq \alpha} X_l\Big) = \sum_{\alpha_{l-1} \geq \alpha} p_l \\ &= \sum_{l-1>k} p_l = 1 - \alpha_{k+1} \leq 1 - \alpha. \end{aligned}$$

Thus (2.2) holds, which gives $x \in \mathbf{x}$. This proves the reverse implication of (3.5). □

Remark 3.1.

(*i*) For different permutations of a histogram, this construction produces different clouds, all with the property (3.5).

(*ii*) For univariate histograms with interval supports, there is a distinguished natural permutation that arranges the supports by increasing midpoint, leading to a sigmoid-shaped cloud.

(*iii*) Another natural arrangement, by increasing probability, is possible for histograms in all dimensions. It is consistent with the interpretation of the level as degree of suitability, and may be preferable in applications even in the univariate case, where it leads to a bell-shaped cloud.

(*iv*) The theorem shows again (as already Corollary 2.1) that the same set of random variables may be characterized by different clouds. This illustrates that it is impossible to define the intersection as a natural operation on clouds.

COROLLARY 3.1. *For every random variable x taking finitely many values only, there is a discrete cloud containing x but no random variable with a different distribution.*

Proof. Label the m values that x can take arbitrarily as $\xi_1, \ldots, \xi_m$, and put $\mathbb{M} = \{\xi_1, \ldots, \xi_m\}$. Then $X_l = \{\xi_l\}$, $p_l = \Pr(x = \xi_l)$ defines a histogram. The theorem gives an associated cloud with

$$x \in \mathbf{x} \quad \Leftrightarrow \quad \Pr(x = \xi_l) = p_l \qquad \text{for } \ l = 1, \ldots m,$$

characterizing the distribution of x. □

When the value intervals of a discrete cloud are strictly overlapping, there is no longer a direct histogram interpretation. But we can view such a cloud as a kind of uncertain histogram, standing for the family of all histograms corresponding to subclouds of the form (3.3).

4. Continuous Clouds

Theorem 3.1 shows that, in principle, clouds can approximate arbitrary distributions arbitrarily well. However, for many purposes, and in particular for higher-dimensional applications, histograms are very cumbersome or expensive to use, and continuous representations are preferable.

We shall see that, using cumulative distribution functions (CDFs), certain continuous clouds (those that are thin and monotone) can describe univariate continuous distributions exactly. In the multivariate case, a continuous cloud is found to code only for information contained in a single statistic of a distribution. This suggests the construction of clouds from simple, user-defined potential functions.

PROPOSITION 4.1. *If x is a real random variable with continuous cumulative distribution function $F(\xi) = \Pr(x \le \xi)$ then $\mathbf{x}(\xi) := F(\xi)$ defines a thin cloud $\mathbf{x}$ with the property that a random variable belongs to $\mathbf{x}$ iff it has the same distribution as x.*

Proof. Since $\mathbf{x}$ is thin, (2.2) shows that $\tilde{x} \in \mathbf{x}$ iff, for all α,

$$\Pr\big(F(\tilde{x}) \ge \alpha\big) = \Pr\big(F(\tilde{x}) > \alpha\big) = 1 - \alpha,$$

i.e., iff $\Pr(F(\tilde{x}) < \alpha) = \Pr(F(\tilde{x}) \le \alpha) = \alpha$. Since F is continuous and monotone, this holds iff $\tilde{x}$ has F as a CDF. □

Note that clouds corresponding to Proposition 4.1 have $\overline{x}(\xi) \ll 1$ near the modes, and hence cannot be interpreted in terms of a degree of suitability.

The CDF $F(\xi)$ of a real random variable x has the property that $F(x)$ is a uniformly distributed random variable. While Proposition 4.1 cannot be generalized to higher dimensions, the uniformity property extends as follows, showing that thin clouds form a natural generalization of continuous CDFs. (But even in the univariate case, thin clouds are more general than continuous CDFs.)

PROPOSITION 4.2.

(*i*) *If $\varphi(x)$ is uniformly distributed in* $[0, 1]$ *and $\varphi(\xi) \in \mathbf{x}(\xi)$ for all $\xi \in \mathbb{M}$ then $x \in \mathbf{x}$.*

(*ii*) *If* $\mathbf{x}$ *is thin and* $x \in \mathbf{x}$ *then* $\mathbf{x}(x)$ *is a random variable uniformly distributed in* $[0, 1]$.

Proof.

(*i*) In this case, $\Pr(\varphi(x) \geq \alpha) = 1 - \alpha = \Pr(\varphi(x) > \alpha)$, which implies (2.2) and hence $x \in \mathbf{x}$.

(*ii*) This is simply a restatement of Proposition 2.1(*iii*). □

(*ii*) says that from a thin cloud $\mathbf{x}$ one can create a statistic which has a precise probability distribution. However, unless the level mapping $\mathbf{x}$ is bijective (which, under continuity, forces the univariate case, where Proposition 4.1 applies), this does not say anything about the marginal distribution of x on level sets $\{\xi \mid \mathbf{x}(\xi) = \alpha\}$ of the cloud $\mathbf{x}$. Indeed, an arbitrary marginal distribution on each level set is compatible with the information contained in the cloud. In particular, on dimensional grounds, multivariate distributions generally contain *much* more information than a continuous cloud can convey.

On the other hand, cloud information is much easier to collect and use. If histogram information is available, we can use Theorem 3.1. If a sufficiently large sample of scenarios is available, we can use instead a potential-based approach, which represents the level $\mathbf{x}(\xi)$ as an interval-valued function of a fixed, user-defined, real-valued potential function $V(\xi)$. Frequently, a quadratic potential

$$V(\xi) = \|R(\xi - \mu)\|^2 \tag{4.1}$$

is appropriate, where μ is a sample mean, and R is a suitable matrix; see, e.g., (4.7) below.

THEOREM 4.1. *Let* x *be a random variable with values in* $\mathbb{M}$*, and let* $V : \mathbb{M} \to \mathbb{R}$ *be bounded below. Then*

$$\mathbf{x}(\xi) := [\Pr\big(V(x) > V(\xi)\big), \Pr\big(V(x) \geq V(\xi)\big)] \tag{4.2}$$

defines a cloud $\mathbf{x}$ *with* $x \in \mathbf{x}$*, whose* α*-cuts are level sets of* V*. The cloud is thin if* $V(x)$ *has a continuous distribution.*

Proof. We may write (4.2) as

$$\mathbf{x}(\xi) := [\underline{\alpha}\big(V(\xi)\big), \overline{\alpha}\big(V(\xi)\big)], \tag{4.3}$$

where

$$\underline{\alpha}(u) = \Pr\big(V(x) > u\big), \qquad \overline{\alpha}(u) = \Pr\big(V(x) \geq u\big). \tag{4.4}$$

If $\alpha = \underline{\alpha}(u)$ for some $u \in \mathbb{R}$ then

$$\begin{aligned}\Pr\big(\underline{x}(x) \geq \alpha\big) &= \Pr\Big(\underline{\alpha}\big(V(\xi)\big) \geq \alpha = \underline{\alpha}(u)\Big) = \Pr\big(V(x) \leq u\big)\\ &= 1 - \Pr\big(V(x) > u\big) = 1 - \underline{\alpha}(u) = 1 - \alpha,\end{aligned}$$

and if no such u exists, a limit argument gives $\Pr\big(\underline{x}(x) \geq \alpha\big) \leq 1-\alpha$. Similarly, if $\alpha = \overline{\alpha}(u)$ for some $u \in \mathbb{R}$ then

$$\begin{aligned}\Pr\big(\overline{x}(x) > \alpha\big) &= \Pr\Big(\overline{\alpha}\big(V(\xi)\big) > \alpha = \overline{\alpha}(u)\Big) = \Pr\big(V(x) < u\big)\\ &= 1 - \Pr\big(V(x) \geq u\big) = 1 - \overline{\alpha}(u) = 1 - \alpha,\end{aligned}$$

and if no such u exists, a limit argument gives $\Pr(\overline{x}(x) > \alpha) \geq 1-\alpha$. Thus (2.2) holds, so that $x \in \mathbf{x}$. □

In particular, since the potential function can be chosen freely, there are many clouds containing a random variable or a random vector with a given distribution.

In most cases, the probabilities are not precisely known (since they are estimated from a sample), and it is appropriate to work in place of (4.4) with an assumed lower bound $\underline{\alpha}(u)$ for $\Pr(V(x) > u)$ and an assumed upper bound $\overline{\alpha}(u)$ for $\Pr(V(x) \geq u)$. (These can be computed from the Kolmogorov-Smirnov distribution [19], using a suitable—fixed—confidence level.) In this case, the cloud (4.3) is still a cloud $\mathbf{x}$ with $x \in \mathbf{x}$, whose α-cuts are level sets of V. We call a cloud of the form (4.3) a **potential cloud** and the functions $\underline{\alpha}, \overline{\alpha} : \mathbb{R} \to [0,1]$ its **potential level maps**.

Potential clouds are determined by the potential (which is generally taken to have a simple analytic form) and the potential level maps (which are univariate functions). This very economical representation makes them highly suitable for modelling uncertain probabilities in higher dimensions, where histograms are no longer useful.

Remark 4.1. The potential construction can be related to a Bayesian perspective of probability theory. Let x be a random variable with values in $\mathbb{M}$, and assume that expectations of functions of x can be written as

$$\langle f(\xi)\rangle = \int f(\xi)q(\xi)\,\mathrm{d}\mu(\xi), \tag{4.5}$$

where μ is a measure on $\mathbb{M}$. We can interpret $q(\xi)$ as the **relative density** of x with respect to the **prior distribution** μ. (Thus the measure characterizing the distribution of x is absolutely continuous, and $q(\xi)$ is its Radon-Nikodym derivative with respect to the prior.) The **canonical cloud** of the random variable x with respect to the prior μ is the cloud $\mathbf{x}$ defined by

$$\underline{x}(\xi) = \Pr\big(q(x) < q(\xi)\big), \qquad \overline{x}(\xi) = \Pr\big(q(x) \leq q(\xi)\big). \tag{4.6}$$

Note that the canonical cloud depends on the prior, and hence is not determined by the distribution of x alone!

For $\mathbb{M} = \mathbb{R}^n$ and continuous q, the canonical cloud is thin. In most cases where the relative density can be given explicitly, it has the form

$$q(\xi) = C\exp\big(-\beta V(\xi)\big)$$

with a continuous potential V; in this case,

$$\begin{aligned} x(\xi) &= \Pr\big(V(x) > V(\xi)\big) = \Pr\big(V(x) \geq V(\xi)\big) \\ &= C \int_{V(\xi) \geq V(\xi)} f(\xi) \exp\big(-\beta V(\xi)\big)\, \mathrm{d}\mu(\xi). \end{aligned}$$

This shows that we have a particular case of the above potential construction. If, with respect to Lebesgue measure, the distribution of x has density $\rho(\xi)$ and the prior has density $\rho_0(\xi)$ then $d\mu(\xi) = \rho_0(\xi)d\xi$,

$$q(\xi) = \rho(\xi) \,/\, \rho_0(\xi)$$

is a **likelihood ratio**, and $V(\xi) = \text{const} - \beta^{-1} \log(\rho(\xi) \,/\, \rho_0(\xi))$. In particular, the canonical cloud of a multivariate Gaussian distribution with respect to a noninformative, uniform prior corresponds to a quadratic potential (4.1). In this case, it is easy to see that R is related to the covariance matrix C of the distribution by the relation

$$R^T R = C^{-1}, \tag{4.7}$$

and can be chosen, e.g., as the upper triangular Cholesky factor of C^{-1}.

There are also relations to the surprise approach to fuzzy modelling, presented in Neumaier [25]. A surprise function defines the potential but *not* the level map. The level can be interpreted as bounds on membership levels of scenarios. The latter can be defined by means of fuzzy criteria, translated into surprise using the techniques of [25], to get a potential $V(x)$. Now a given sample of scenarios x_l determines a list of potential values $V(x_l)$, from which an empirical CDF for $V(x)$ can be constructed. Using error bounds computed from the Kolmogorov-Smirnov distribution at a suitable (fixed) confidence level, we get lower and upper levels defining the cloud. Details and examples will be discussed in another paper (Lodwick et al. [21]).

5. Probability, Possibility, and Necessity

This section relates clouds to more traditional concepts in the theory of uncertainty. We first cast Proposition 2.1 in a set theoretic form.

We define the **probability interval** $\mathbf{p}(\mathbf{x}|A)$ of $\mathbf{x}$ with respect to a subset A of $\mathbb{M}$ by

$$\mathbf{p}(\mathbf{x}|A) = [\underline{p}(\mathbf{x}|A), \overline{p}(\mathbf{x}|A)], \tag{5.1}$$

with the **lower probability**

$$\underline{p}(\mathbf{x}|A) = \inf_{\xi \notin A} \underline{x}(\xi) \tag{5.2}$$

and the **upper probability**

$$\overline{p}(\mathbf{x}|A) = \sup_{\xi\in A} \overline{x}(\xi); \tag{5.3}$$

here $\inf_{\xi\in\emptyset} = 0$, $\sup_{\xi\in\emptyset} = 1$. (Lower and upper probability are frequently discussed in the literature, starting with Dempster [7], but apart from the relation to possibility theory exhibited below, our concept seems to have little connection with other work on this topic.)

PROPOSITION 5.1. *All probability intervals* $\mathbf{p}(\mathbf{x}|A)$ *are nonempty. Moreover, for an arbitrary random variable* $x \in \mathbf{x}$ *and any measurable subset* A *of* $\mathbb{M}$*,*

$$\Pr(x \in A) \in \mathbf{P}(\mathbf{x}|A) := \mathbf{p}(\mathbf{x}|A) \cap \mathbf{p}(1 - \mathbf{x}|A) \quad \textit{for all } A, \tag{5.4}$$

and any random variable x *with this property is in* $\mathbf{x}$*.*

Proof. The definition of the lower and upper probability implies that $\overline{x}(\xi) \le \overline{p}(\mathbf{x}|A)$ for $\xi \in A$ and $\underline{x}(\xi) \ge \underline{p}(\mathbf{x}|A)$ for $\xi \notin A$; thus the union of all $\mathbf{x}(\xi)$ is a subset of $[0, \overline{p}(\mathbf{x}|A)] \cup [\underline{p}(\mathbf{x}|A), 1]$. Therefore (2.1) implies that $\underline{p}(\mathbf{x}|A) \le \overline{p}(\mathbf{x}|A)$. Thus $\mathbf{p}(\mathbf{x}|A)$ is nonempty.

To prove (5.4) we note that

$$\overline{p}(1 - \mathbf{x}|A) = \sup_{\xi\in A} \big(1 - \underline{x}(\xi)\big) = 1 - \inf_{\xi\in A} \underline{x}(\xi) = 1 - \underline{p}(\mathbf{x}|\neg A),$$

and similarly $\underline{p}(1 - \mathbf{x}|A) = 1 - \overline{p}(\mathbf{x}|\neg A)$. Hence $\mathbf{p}(1 - \mathbf{x}|A) = 1 - \mathbf{p}(\mathbf{x}|\neg A)$. This, together with $\Pr(x \in A) \in \mathbf{p}(\mathbf{x}|A)$, which follows directly from (2.3) and (2.4), and

$$\Pr(x \in A) = 1 - \Pr(x \notin A) \in 1 - \mathbf{p}(\mathbf{x}|\neg A)$$

implies (5.4). Conversely, let x be a random variable satisfying (5.4) for all measurable subsets A of $\mathbb{M}$. Then

$$\Pr(x \in \overline{C}_\alpha) \ \ge\ \underline{p}(1 - \mathbf{x}|\overline{C}_\alpha) = \inf_{\xi\notin\overline{C}_\alpha} \big(1 - \overline{x}(\xi)\big) = 1 - \alpha,$$
$$\Pr(x \in \underline{C}_\alpha) \ \le\ \overline{p}(1 - \mathbf{x}|\underline{C}_\alpha) = \sup_{\xi\in\underline{C}_\alpha} \big(1 - \underline{x}(\xi)\big) = 1 - \alpha.$$

Thus (2.2) holds, showing $x \in \mathbf{x}$. □

Note that

$$\mathbf{P}(\mathbf{x}|\neg A) = 1 - \mathbf{P}(\mathbf{x}|A), \tag{5.5}$$

while the analogous formula for the defining upper and lower probabilities generally does not hold.

To formulate the next result, we define between intervals $\mathbf{p}, \mathbf{q}$ the order relation

$$\mathbf{p} \le_w \mathbf{q} \quad \Leftrightarrow \quad \underline{p} \le \underline{q},\ \ \overline{p} \le \overline{q},$$

and the operations

$$\min(\mathbf{p},\mathbf{q}) = [\min(\underline{p},\underline{q}), \min(\overline{p},\overline{q})], \qquad \max(\mathbf{p},\mathbf{q}) = [\max(\underline{p},\underline{q}), \max(\overline{p},\overline{q})].$$

PROPOSITION 5.2 (Properties of probability intervals).

$$\mathbf{p}(\mathbf{x}|\emptyset) = 0, \qquad \mathbf{p}(\mathbf{x}|\mathbb{M}) = 1, \tag{5.6}$$

$$A \subseteq B \quad \Rightarrow \quad \mathbf{p}(\mathbf{x}|A) \leq_w \mathbf{p}(\mathbf{x}|B), \tag{5.7}$$

$$\mathbf{p}(\mathbf{x}|A \cap B) \leq_w \min\big(\mathbf{p}(\mathbf{x}|A),\ \mathbf{p}(\mathbf{x}|B)\big), \tag{5.8}$$

$$\max\big(\mathbf{p}(\mathbf{x}|A),\ \mathbf{p}(\mathbf{x}|B)\big) \leq_w \mathbf{p}(\mathbf{x}|A \cup B), \tag{5.9}$$

$$\underline{p}(\mathbf{x}|A \cap B) = \min\big(\underline{p}(\mathbf{x}|A),\ \underline{p}(\mathbf{x}|B)\big), \tag{5.10}$$

$$\overline{p}(\mathbf{x}|A \cup B) = \max\big(\overline{p}(\mathbf{x}|A),\ \overline{p}(\mathbf{x}|B)\big). \tag{5.11}$$

Proof. (5.6) follows directly from the definition. Let $A \subseteq B$. Then $\xi \notin B \Rightarrow \xi \notin A$, hence

$$\underline{p}(\mathbf{x}|B) = \inf_{\xi \notin B} \underline{x}(\xi) \geq \inf_{\xi \notin A} \underline{x}(\xi) = \underline{p}(\mathbf{x}|A),$$

and $\xi \in A \Rightarrow \xi \in B$, hence

$$\overline{p}(\mathbf{x}|A) = \sup_{\xi \in A} \overline{x}(\xi) \leq \sup_{\xi \in B} \overline{x}(\xi) = \overline{p}(\mathbf{x}|B).$$

This implies (5.7). (5.8) follows from (5.7) since $A \cap B \subseteq A$ and $\subseteq B$, and (5.9) since $A \cup B \supseteq A$ and $\supseteq B$. (5.10) follows from (5.2) since $\xi \notin A \cap B$ iff $\xi \notin A$ or $\xi \notin B$, and (5.10) from (5.3) since $\xi \in A \cup B$ iff $\xi \in A$ or $\xi \in B$. □

In particular, $\overline{p}$ satisfies the traditional axioms [9] for a possibility measure, and $\underline{p}$ those for a necessity measure.

From this result, it is easily seen that clouds are nearly the same as the *consistent possibility and necessity measures* (CPNMs) introduced by Jamison & Lodwick [15]. To make the connection, we use their terminology but without giving details. [15] considers a triple $(\mathrm{nec}, \mu, \mathrm{pos})$ consisting of a necessity measure nec, a nonnegative measure μ and a possibility measure pos such that $\mu(A) \subseteq \mathbf{p}(A) := [\mathrm{nec}(A), \mathrm{pos}(A)]$ for all measurable $A \in \mathbb{M}$. Upon identifying $\mathbf{p}(A)$ with $\mathbf{p}(\mathbf{x}|A)$ and $\mu(A)$ with $\Pr(x \in A)$, this becomes the statement $\Pr(x \in A) \subseteq \mathbf{p}(\mathbf{x}|A)$ contained in (5.4).

Several of the set-theoretical constructions in [15] are reformulations of techniques discussed above, although in a computationally less accessible form. Our construction in Theorem 4.1 is essentially the construction in their Theorem 2 of a CPNM from a possibility nest E_λ which gives rise to a potential via $V(x) = \mathrm{const} - \inf\{r | x \in E_r\}$. (Their construction requires an additional continuity assumption—their Definition 4(3)—which is not needed in our setting.) Our composition construction in Theorem 6.1 is essentially equivalent to the construction in their Theorem 5. They also show how to compute bounds on expectations, assuming a monotonicity assumption which allows one to avoid solving a global optimization problem (as required in our Section 6). However, the set theoretic nature of their setting makes working with a CPNM clumsy to use, and they lack simple recipes for constructing probability nests from sampled scenarios.

Finally, we mention a relation to the Dempster-Shafer theory of plausibility and belief (or credibility) functions (see, e.g. Shafer [31]). If we define the **credibility** $\mathrm{Cr}(\mathbf{x}|A)$ and the **plausibility** $\mathrm{Pl}(\mathbf{x}|A)$ by

$$\mathbf{P}(\mathbf{x}|A) = [\mathrm{Cr}(\mathbf{x}|A), \mathrm{Pl}(\mathbf{x}|A)], \tag{5.12}$$

then (5.5) implies

$$\mathrm{Cr}(\mathbf{x}|A) = 1 - \mathrm{Pl}(\mathbf{x}|\neg A), \qquad \mathrm{Pl}(\mathbf{x}|A) = 1 - \mathrm{Cr}(\mathbf{x}|\neg A).$$

Moreover, the relations (5.6)–(5.9) (but not (5.10)–(5.11)) remain valid with $\mathbf{P}$ in place of $\mathbf{p}$, with corresponding implications for $\mathrm{Cr}(\mathbf{x}|A)$ and $\mathrm{Pl}(\mathbf{x}|A)$.

If $\mathbf{x}$ is a fuzzy cloud then $\mathrm{Pl}(\mathbf{x}|A) = \overline{p}(\mathbf{x}|A)$ satisfies the axioms for a possibility measure and $\mathrm{Cr}(\mathbf{x}|A) = 1 - \overline{p}(\mathbf{x}|\neg A)$ satisfies the axioms for a necessity measure; but for general clouds, this is no longer the case. Neither do Pl and Cr satisfy in general the Dempster-Shafer axioms for plausibility and belief functions, although they do for fuzzy clouds.

6. Transformation of Clouds

A frequently occurring situation is that one wants to obtain information on a random variable or random vector z defined in terms of another random variable or random vector x. We assume that x takes values in $\mathbb{M}$, and z is defined in terms of a mapping $F : \mathbb{M} \to \mathbb{M}'$ by $z := F(x)$. If the information about x is uncertain, one can only hope to get uncertain information about z. In the following, we show how the information that x belongs to a cloud $\mathbf{x}$ can be used to obtain a cloud $\mathbf{z}$ containing $F(x)$. The symbol $\circ$ denotes the composition of mappings.

THEOREM 6.1.

(*i*) *If* $\mathbf{x}, \mathbf{z}$ *are clouds satisfying*

$$\mathbf{x} \subseteq \mathbf{z} \circ F \tag{6.1}$$

then

$$x \in \mathbf{x} \quad \Rightarrow \quad F(x) \in \mathbf{z}. \tag{6.2}$$

(*ii*) *Any cloud* $\mathbf{z}$ *satisfying* (6.1) *contains the cloud* $F(\mathbf{x})$ *defined by*

$$F(\mathbf{x})(\zeta) := \square\{\mathbf{x}(\xi) \mid F(\xi) = \zeta\}. \tag{6.3}$$

Proof.

(*i*) Let $F^{-1}(A)$ denote the preimage of a subset $A \in \mathbb{M}'$ under F. Then

$$\begin{aligned}
\underline{p}(\mathbf{x}|F^{-1}(A)) &= \inf_{\xi \notin F^{-1}(A)} \underline{x}(\xi) \geq \inf_{\xi \notin F^{-1}(A)} \underline{z}\big(F(\xi)\big) \\
&= \inf_{F(\xi) \notin A} \underline{z}\big(F(\xi)\big) \geq \inf_{\zeta \notin A} \underline{z}(\zeta) = \underline{p}(\mathbf{z} \mid A), \\
\overline{p}\big(\mathbf{x}|F^{-1}(A)\big) &= \sup_{\xi \in F^{-1}(A)} \overline{x}(\xi) \leq \sup_{\xi \in F^{-1}(A)} \overline{z}\big(F(\xi)\big) \\
&= \sup_{F(\xi) \in A} \overline{z}\big(F(\xi)\big) \leq \sup_{\zeta \in A} \overline{z}(\zeta) = \overline{p}(\mathbf{z} \mid A).
\end{aligned}$$

Therefore

$$\Pr\big(F(\xi) \in A\big) = \Pr\big(\xi \in F^{-1}(A)\big) \in \mathbf{p}\big(\mathbf{x}|F^{-1}(A)\big) \subseteq \mathbf{p}(\mathbf{z}|A),$$

which implies (6.2).

(*ii*) If $\mathbf{z}$ satisfies (6.2) and $F(\xi) = \zeta$ then

$$\mathbf{x}(\xi) \subseteq (\mathbf{z} \circ F)(\xi) = \mathbf{z}\big(F(\xi)\big) = \mathbf{z}(\zeta),$$

hence $\mathbf{z}(\zeta)$ contains all $\mathbf{x}(\xi)$ with $F(\xi) = \zeta$. Therefore it contains $F(\mathbf{x})(\zeta)$. Since this holds for all ζ, we find $F(\mathbf{x}) \subseteq \mathbf{z}$. □

EXAMPLE 6.1. For the thin cloud in $\mathbb{R}^n$ defined by $\mathbf{x}(\xi) = e^{-\lambda\|\xi\|}$ for some $\lambda > 0$, we determine the optimal cloud containing $z = a^T x$. By (6.3), we have

$$\begin{aligned}
F(\mathbf{x})(\zeta) &= \Box\{e^{-\lambda\|\xi\|} \mid a^T\xi = \zeta\} \\
&= \Box\{e^{-\lambda r} \mid \exists\xi : a^T\xi = \zeta,\ \|\xi\| = r\} \\
&= \Box\{e^{-\lambda r} \mid r \geq |\zeta| / \|a\|\} = [0, e^{-\lambda|\zeta|/\|a\|}].
\end{aligned}$$

Thus $\mathbf{z} = F(\mathbf{x})$ is defined by

$$\mathbf{z}(\zeta) = [0, e^{-\beta|\zeta|}], \qquad \beta = \lambda / \|a\|.$$

While (6.3) gives the optimal cloud representing the information on $z = F(x)$ that can be inferred from $x \in \mathbf{x}$, it is difficult to compute in practice. It is therefore of interest to find a suboptimal cloud with a computationally more tractable shape. We first give a recipe that checks whether a proposed cloud has the required property (6.1), and then show how this can be used to construct such a cloud.

COROLLARY 6.1. *Let* $x \in \mathbf{x}$, *and let*

$$\begin{aligned}
\underline{X} &= \{\xi \in \mathbb{M} \mid \underline{x}(\xi) < \underline{z}\big(F(\xi)\big)\}, \\
\overline{X} &= \{\xi \in \mathbb{M} \mid \overline{x}(\xi) > \overline{z}\big(F(\xi)\big)\}.
\end{aligned}$$

Then

$$\underline{X} = \overline{X} = \emptyset \quad \Rightarrow \quad F(x) \in \mathbf{z}.$$

Proof. If $\underline{X} = \overline{X} = \emptyset$ then, for all $\xi \in \mathbb{M}$,

$$\underline{z}\big(F(\xi)\big) \leq \underline{x}(\xi), \qquad \overline{x}(\xi) \leq \overline{z}\big(F(\xi)\big),$$

which is equivalent with $\mathbf{x}(\xi) \subseteq \mathbf{z} \circ F(\xi)$. □

Note that checking the emptiness of $\underline{X}$ or $\overline{X}$ is a standard constraint satisfaction problem that can be solved by interval techniques. If we are prepared instead to solve two global optimization problems, we get following explicit construction. (We write $s_+ = \max(s, 0)$).

THEOREM 6.2. *Let* $\mathbf{z}$ *be a cloud on* $\mathbb{M}'$ *and let*

$$\delta \geq \left(1 - \inf_{\xi \in \mathbb{M}} \frac{\underline{x}(\xi)}{\underline{z}(F(\xi))}\right)_+, \qquad \varepsilon \geq \left(1 - \inf_{\xi \in \mathbb{M}} \frac{1 - \overline{x}(\xi)}{1 - \overline{z}(F(\xi))}\right)_+. \tag{6.4}$$

Then the cloud $\mathbf{z}'$ *defined by*

$$\mathbf{z}'(\zeta) = [(1 - \delta)\underline{z}(\zeta),\ \varepsilon + (1 - \varepsilon)\overline{z}(\zeta)] \tag{6.5}$$

satisfies $F(x) \in \mathbf{z}'$.

Proof. (6.4) implies that $\underline{x}(\xi) \geq (1-\delta)\underline{z}(F(\xi))$ and $1-\overline{x}(\xi) \geq (1-\varepsilon)(1-\overline{z}(F(\xi)))$, hence $\overline{x}(\xi) \leq \varepsilon + (1 - \varepsilon)\overline{z}(F(\xi))$. Thus Corollary 6.1 applies with $\mathbf{z}'$ in place of $\mathbf{z}$. □

Thus we first guess a cloud $\underline{z}$ and then correct it to have the required properties. The constraint satisfaction problems or global optimization problems can be solved by standard branch and bound methods [1], [6], [16], [30]. Suboptimal bounds may be obtained cheaper using simple interval arithmetic or centered forms, or, for clouds constructed from a quadratic potential, using ellipsoid arithmetic [29].

7. Expectations from Clouds

Applications frequently require the estimation of expected values $\langle f(x)\rangle$ of some expression $f(x)$ in a random variable x, or its mean square deviations $\langle (f(x) - \mu)^2\rangle$ of $f(x)$ from some nominal value μ. If the random variable is uncertain and only known to lie in a given cloud $\mathbf{x}$, it is therefore of interest to bound the expectation of arbitrary expressions in $x \in \mathbf{x}$.

The situation is simplest for thin clouds.

THEOREM 7.1. *Let* $\mathbf{x}$ *be a thin cloud over* $\mathbb{M}$, *and let* $f : \mathbb{M} \to \mathbb{R}$ *be a measurable function. Then, for any grid*

$$0 = \alpha_0 < \alpha_1 < \cdots < \alpha_n = 1, \tag{7.1}$$

we have

$$\langle f(x)\rangle \in \sum_{l=1}^{n}(\alpha_l - \alpha_{l-1})\square\{f(\xi) \mid \alpha_{l-1} < \mathbf{x}(\xi) \le \alpha_l\}. \tag{7.2}$$

Moreover, with lower and upper Riemann integrals,

$$\underline{\int} d\alpha \inf\{f(\xi) \mid \mathbf{x}(\xi) = \alpha\} \le \langle f(x)\rangle \le \overline{\int} d\alpha \sup\{f(\xi) \mid \mathbf{x}(\xi) = \alpha\}. \tag{7.3}$$

Proof. Writing $\alpha_{-1} = -1$ and $\mathbf{f}_l := \square\{f(\xi) \mid \alpha_{l-1} < \mathbf{x}(\xi) \le \alpha_l\}$, we have

$$f(\xi) \in \sum_{l=0}^{n}\mathbf{f}_l\chi\big(\alpha_{l-1} < \mathbf{x}(\xi) \le \alpha_l\big)$$

for all $\xi \in \mathbb{M}$, hence, by linearity of the expectation,

$$\begin{aligned}\langle f(\xi)\rangle &\in \sum_{l=0}^{n}\mathbf{f}_l \Pr\big(\alpha_{l-1} < \mathbf{x}(\xi) \le \alpha_l\big)\\ &= \sum_{l=0}^{n}\mathbf{f}_l\Big(\Pr\big(\mathbf{x}(\xi) \le \alpha_l\big) - \Pr\big(\mathbf{x}(\xi) \le \alpha_{l-1}\big)\Big)\\ &= \sum_{l=1}^{n}\mathbf{f}_l(\alpha_l - \alpha_{l-1})\end{aligned}$$

by Proposition 2.1, giving (7.2). Letting the grid become arbitrarily fine then gives (7.3). $\square$

For clouds that are not thin, the above argument breaks down since only inequalities for probabilities are available. However, a more sophisticated linear programming approach produces useful bounds. Again, we consider a grid (7.1) and define

$$\underline{A}_i := \{\xi \mid \alpha_{i-1} < \underline{x}(\xi) \le \alpha_i\}, \qquad \overline{A}_i := \{\xi \mid \alpha_{k-1} < \overline{x}(\xi) \le \alpha_k\}.$$

With the unknown probabilities

$$p_{ik} := \Pr(x \in \underline{A}_i \cap \overline{A}_k) \tag{7.4}$$

and arbitrary computable enclosures

$$\{f(\xi) \mid \xi \in \underline{A}_i \cap \overline{A}_k\} \in \mathbf{f}_{ik}, \tag{7.5}$$

we then have

$$f(\xi) \in \sum_{i,k}\mathbf{f}_{ik}\chi(\xi \in \underline{A}_i \cap \overline{A}_k)$$

for all $\xi \in \mathbb{M}$, hence

$$\langle f(x)\rangle \in \sum_{i,k} \mathbf{f}_{ik} p_{ik}. \tag{7.6}$$

The sum in (7.6) can be restricted to all pairs in the set

$$E := \{(i,k) \mid \underline{A}_i \cap \overline{A}_u \neq \emptyset\} \tag{7.7}$$

since $p_{ik} = 0$ for $(i,k) \notin E$. Since $\xi \in \underline{A}_i \cap \overline{A}_k$ implies $\alpha_{i-1} < \underline{x}(\xi) \leq \overline{x}(\xi) \leq \alpha_k$, E only contains pairs with $i \leq k$. Moreover, for narrow clouds and fine grids, all pairs in E have $k - i$ a small integer. To exploit relations (7.6) we need to use restrictions on the p_{ik} implied by $x \in \mathbf{x}$. Proposition 2.1 gives the inequalities

$$\sum_{\substack{(i,k)\in E\\ k\leq l}} p_{ik} \leq \alpha_l \leq \sum_{\substack{(i,k)\in E\\ i\leq l}} p_{ik}, \tag{7.8}$$

and we also have the trivial inequalities

$$p_{ik} \geq 0, \qquad \sum_{(i,k)\in E} p_{ik} = 1. \tag{7.9}$$

(7.6) now implies that

$$\underline{f} = \min\left\{ \sum_{(i,k)\in E} \underline{f}_{ik} p_{ik} \;\middle|\; (7.8), (7.9)\right\}, \tag{7.10}$$

$$\overline{f} = \max\left\{ \sum_{(i,k)\in E} \overline{f}_{ik} p_{ik} \;\middle|\; (7.8), (7.9)\right\} \tag{7.11}$$

define an enclosure

$$\langle f(x)\rangle \in [\underline{f}, \overline{f}].$$

Clearly, the bounds become optimal in the limit where the grid becomes arbitrarily fine. (7.10) and (7.11) can be computed by linear programming techniques. Bounds (7.5) needed to determine the coefficients can be obtained by global optimization, or in simple cases by interval arithmetic or analytic estimates. It is worth remarking that upon refining a grid, part of the work spent in the old bound computation for (7.5) can be re-used if the points where the extrema are achieved are stored; only the parts not containing the extrema need to be recalculated.

For other uses of optimization techniques in related contexts, see Kuznetsov [20], Weichselberger [35], Ferson et al. [12].

EXAMPLE 7.1. For the thin cloud in $\mathbb{R}^n$ defined by $\mathbf{x}(\xi) = e^{-\lambda\|\xi\|}$ for some $\lambda > 0$, we have $\mathbf{x}(\xi) = \alpha$ iff $\|\xi\| = \lambda^{-1} \log \alpha^{-1}$, hence (7.3) gives

$$\begin{aligned}\langle a^T x\rangle &\leq \int^{-} \mathrm{d}\alpha \sup\{a^T\xi \mid \|\xi\| = \lambda^{-1}\log\alpha^{-1}\}\\ &= \int_0^1 \mathrm{d}\alpha \|a\|\lambda^{-1}\log\alpha^{-1}\\ &= \frac{\|a\|}{\lambda}\left([\alpha\log\alpha^{-1}]_0^1 - \int_0^1 \mathrm{d}\alpha(\log\alpha^{-1})'\right)\\ &= \frac{\|a\|}{\lambda}\left(0 - \int_0^1 \mathrm{d}\alpha(-1)\right) = \frac{\|a\|}{\lambda}.\end{aligned}$$

Replacing a by $-a$ gives a lower bound, combining to the enclosure

$$\langle a^T x\rangle \in [-q, q], \qquad q = \lambda^{-1}\|a\|. \tag{7.12}$$

Note that our expectation bounds are only vaguely related to the lower and upper previsions discussed, e.g., in Walley [34].

8. Probabilities from Clouds

Apart from expectations, applications frequently require bounds on probabilities. The formulas from Proposition 2.1 are not yet in a form directly useful for computations. Indeed, probability statements of interest in practice are usually about membership of a random variable in a set implicitly defined by equations and inequalities. In terms of extended intervals $\mathbf{b} = [\underline{b}, \overline{b}]$ with lower bounds $\underline{b} \in \mathbb{R} \cup \{-\infty\}$ and upper bounds $\overline{b} \in \mathbb{R} \cup \{\infty\}$, we may write an equation $f(x) = b$ as enclosure $f(x) \in [b, b]$, and inequalities $f(x) \geq b$ or $f(x) \leq b$ as enclosures $f(x) \in [b, \infty]$ or $f(x) \in [-\infty, b]$, respectively. Therefore, enclosures of the form $f(x) \in \mathbf{b}$ capture both equations and inequalities. A set defined by the simultaneous validity of several equations or inequalities can therefore always be written as the set of x satisfiying a vector enclosure

$$F(x) \in \mathbf{F}, \tag{8.1}$$

where $\mathbf{F} \in \mathbb{I}^*\mathbb{R}^m$ is an extended box and $F : \mathbb{M} \to \mathbb{R}^m$. The following result reduces the computation of bounds for the probability of the statement (8.1), given a random variable $x \in \mathbf{x}$, to bounding the objective function of two optimization problems.

THEOREM 8.1. *Let* $\mathbf{x}$ *be a cloud over* $\mathbb{M}$. *Then, for every random variable* $x \in \mathbf{x}$, *we have*

$$\Pr(F(x) \in \mathbf{F}) \in [\underline{p}, \overline{p}] \tag{8.2}$$

for arbitrary $\underline{p}, \overline{p} \in [0, 1]$ *satisfying*

$$\underline{p} \leq \inf\{\underline{x}(\xi) \mid F(\xi) \notin \mathbf{F}\}, \tag{8.3}$$

$$\overline{p} \geq \sup\{\overline{x}(\xi) \mid F(\xi) \in \mathbf{F}\}. \tag{8.4}$$

Proof. Let $A = \{\xi \in \mathbb{M} \mid F(\xi) \in \mathbf{F}\}$. Then

$$\begin{aligned}\Pr\bigl(F(x) \in \mathbf{F}\bigr) &= \Pr(x \in A) \in \mathbf{p}(\mathbf{x}|A) \geq \underline{p}(\mathbf{x}|A)\\ &= \inf_{\xi \notin A} \underline{x}(\xi) = \inf_{F(\xi)\notin \mathbf{F}} \underline{x}(\xi) \geq \underline{p},\end{aligned}$$

and similarly

$$\Pr\bigl(F(x) \in \mathbf{F}\bigr) \leq \overline{p}(\mathbf{x}|A) = \sup_{\xi \in A} \overline{x}(\xi) = \sup_{F(\xi)\in \mathbf{F}} \overline{x}(\xi) \leq \overline{p}. \qquad \square$$

To get the best possible probability bounds valid for *all* random variables in a cloud, we need to take the intersection of (8.2) for the cloud $\mathbf{x}$ and its mirror cloud $1 - \mathbf{x}$, with $\underline{p}$ and $\overline{p}$ chosen to satisfy equality in (8.3) and (8.4). Note that, although optimal, the bounds (8.2) are usually conservative when applied to a *particular* random variable belonging to a cloud, since the bounds must account for all other random variables in the same cloud, too.

The computation of the bounds (8.3) and (8.4) requires the solution of two global optimization problems, which can be done by interval techniques combined with branch and bound (Kearfott [16]). The process can be speeded up, however, since in many cases the bounds are needed only with low accuracy. Moreover, for many practically relevant conditions (8.1), the resulting optimization problems have a unique local extremum, making the global search much more tractable than for general global optimization problems. In simple cases, the optimal bounds can even be calculated analytically:

EXAMPLE 8.1. For the thin cloud in $\mathbb{R}^n$ defined by $\mathbf{x}(\xi) = e^{-\lambda\|\xi\|}$ for some $\lambda > 0$, we want to bound the probability that $x \in \mathbf{x}$ lies in the half plane defined by $a^T x \geq \alpha$. The optimal bounds are given by

$$\begin{aligned}\underline{p} &= \inf\{e^{-\lambda\|\xi\|} \mid a^T\xi < \alpha\} = 0,\\ \overline{p} &= \sup\{e^{-\lambda\|\xi\|} \mid a^T\xi \geq \alpha\} = e^{-\lambda\max(\alpha,0)/\|a\|}\end{aligned}$$

since $\max(\alpha, 0) \leq |a^T\xi| \leq \|a\|\,\|\xi\|$. Thus we get a nontrivial bound

$$\Pr(a^T x \geq \alpha) \leq e^{-\lambda\alpha/\|a\|} \qquad \text{if } \alpha > 0. \tag{8.5}$$

The bound decays exponentially with increasing α.

In many cases, it is only required to check whether a risk probability is small enough. In this case, the following, simpler criterion applies. Again, the criterion is best possible, given only the information in the cloud.

COROLLARY 8.1. *If the constraints*

$$F(\xi) \in \mathbf{F}, \qquad \overline{x}(\xi) \geq \delta \tag{8.6}$$

have no solution then

$$\Pr\bigl(F(x) \in \mathbf{F}\bigr) < \delta. \tag{8.7}$$

Proof. In this case, the supremum in (8.4) is $< \delta$, and the claim follows from (8.2). □

Checking the solvability of (8.5) is a standard constraint satisfaction problem, that can again be treated by branch and bound.

9. Open Problems

We end by listing a number of open problems for further research, together with a few comments.

PROBLEM 1. Provide good implementations of the procedures presented here for calculating with clouds.

PROBLEM 2. Combine a cloud $\mathbf{x}$ containing x with a cloud $\mathbf{y}$ containing y to a cloud $\mathbf{z}$ containing $z = (x, y)$. It would be interesting to have constructions of optimal clouds containing z that are compatible with the information in the clouds $\mathbf{x}$ and $\mathbf{y}$ and contain parameters which allow the user to control the amount and form of assumed dependence. This involves the use of copulas [32] for specifying the degree of dependence of x and y.

PROBLEM 3. Find ways to get optimal enclosures for expectations $\langle f(x, y)\rangle$ and probabilities $\Pr(F(x, y) \in \mathbf{F})$, given the knowledge of a cloud $\mathbf{x}$ containing x and a cloud $\mathbf{y}$ containing y, possibly together with some dependence information.

PROBLEM 4. The same as Problems 2 and 3, but for more than two arguments.

PROBLEM 5. We have seen in Examples 6.1, 7.1, and 8.1 that for the example of an exponentially decaying spherical cloud in arbitrary dimensions, the optimization problems posed in the present paper can be solved in closed form. Find other useful special cases where this is possible.

PROBLEM 6. Translate lower and upper bounds on a probability density (as in Yeh [38]) into a good description in terms of clouds.

Acknowledgements

I want to thank Weldon Lodwick and Dave Jamison for stimulating discussions, without which this paper would not exist, Scott Ferson and Hermann Schichl for useful comments on a preliminary version of this paper, and Vladik Kreinovich for bringing to my attention the thesis by Williamson [37].

References

1. Adjiman, C. S., Dallwig, S., Floudas, C. A., and Neumaier, A.: A Global Optimization Method, αBB, for General Twice-Differentiable Constrained NLPs—I. Theoretical Advances, *Comput. Chem. Engin.* **22** (1998), pp. 1137–1158.

2. Berleant, D.: Automatically Verified Arithmetic on Probability Distributions and Intervals, in: Kearfott, R. B. and Kreinovich, V. (eds), *Applications of Interval Computations*, Kluwer Academic Publishers, Dordrecht, 1996, pp. 227–244.
3. Berleant, D. and Cheng, H.: A Software Tool for Automatically Verified Operations on Intervals and Probability Distributions, *Reliable Computing* **4** (1) (1998), pp. 71–82.
4. Birnbaum, A.: A Unified Theory of Estimation I, *Ann. Math. Statist.* **32** (1961), pp. 112–135.
5. Birnbaum, A.: Confidence Curves: An Omnibus Technique for Estimation and Testing Statistical Hypotheses, *J. Amer. Statist. Assoc.* **56** (1961), pp. 246–249.
6. Bliek, C., Spellucci, P., Vicente, L. N., Neumaier, A., Granvilliers, L., Monfroy, E., Benhamou, F., Huens, E., Van Hentenryck, P., Sam-Haroud, D., and Faltings, B.: *Algorithms for Solving Nonlinear Constrained and Optimization Problems: The State of the Art*, A progress report of the COCONUT project, 2001,
`http://www.mat.univie.ac.at/~neum/glopt/coconut/StArt.html`
7. Dempster, A. P.: Upper and Lower Probabilities Induced by a Multi-Valued Mapping, *Ann. Math. Statist.* **39** (1968), pp. 957–966.
8. Dubois, D., Moral, S., and Prade, H.: A Semantics for Possibility Theory Based on Likelihoods, *J. Math. Anal. Appl.* **205** (1997), pp. 359–380.
9. Dubois, D. and Prade, H.: *Possibility Theory*, Plenum Press, New York, 1988.
10. Ferson, S.: What Monte Carlo Methods Cannot Do, *Human and Ecological Risk Assessment* **2** (1996), pp. 990–1007.
11. Ferson, S., Ginzburg, L., and Akçakaya, R.: Whereof One Cannot Speak: When Input Distributions Are Unknown, *Risk Analysis*, to appear.
12. Ferson, S., Ginzburg, L., Kreinovich, V., and Lopez, J.: Absolute Bounds on the Mean of Sum, Product, etc.: A Probabilistic Extension of Interval Arithmetic, in: *Extended Abstracts of the 2002 SIAM Workshop on Validated Computing, Toronto, Canada, May 23–25, 2002*, pp. 70–72.
13. Ferson, S. and Moore, D. R. J.: *Bounding Uncertain Analysis*, Manuscript, 2002.
14. Ferson, S., Root, W. T., and Kuhn, R.: *RAMAS Risk Calc: Arithmetic with Uncertain Numbers*, Applied Biomathematics, Setauket, New York, 1998,
`http://www.ramas.com/riskcalc.htm`
15. Jamison, K. D. and Lodwick, W. A.: *The Construction of Consistent Possibility and Necessity Measures*, UCD/CCM Report No. 171, Center for Computational Mathematics, Univ. of Colorado at Denver, 2001.
16. Kearfott, R. B.: *Rigorous Global Search*, Kluwer Academic Publishers, Dordrecht, 1996.
17. Kiefer, J.: Conditional Confidence Statements and Confidence Estimators, *J. Amer. Statist. Assoc.* **72** (1977), pp. 789–808.
18. Klir, G. and Yuan, B.: *Fuzzy Sets and Fuzzy Logic. Theory and Application*, Prentice Hall, Upper Saddle River, 1995.
19. Kolmogoroff, A.: Confidence Limits for an Unknown Distribution Function, *Ann. Math. Statist.* **12** (1941), pp. 461–463.
20. Kuznetsov, V. P.: *Interval Statistical Models*, Radio i Svyaz Publ., Moscow, 1991 (in Russian).
21. Lodwick, W. A., Jamison, K. D., and Neumaier, A.: *Multivariate Fuzzy Data Analysis Using Clouds*, in preparation.
22. Lüthi, J.: Histogram-Based Characterization of Workload Parameters and Its Consequences on Model Analysis, in: *Proc. MASCOTS'98 Workshop on Workload Characterization in High-Performance Computing Environments*, 1998, pp. 1/52–1/64,
`http://www.informatik.unibw-muenchen.de/inst4/luethi/research.html`
23. Moore, R. E.: *Methods and Applications of Interval Analysis*, SIAM, Philadelphia 1979.
24. Moore, R. E.: Risk Analysis without Monte Carlo Methods, *Freiburger Intervall-Ber.* **84** (1) (1984), pp. 1–48.
25. Neumaier, A.: Fuzzy Modeling in Terms of Surprise, *Fuzzy Sets and Systems* **135** (2003), pp. 21–38, `http://www.mat.univie.ac.at/~neum/papers.html#surprise`
26. Neumaier, A.: *Interval Methods for Systems of Equations*, Cambridge Univ. Press, Cambridge, 1990.
27. Neumaier, A.: *Introduction to Numerical Analysis*, Cambridge Univ. Press, Cambridge, 2001.

28. Neumaier, A.: *On the Structure of Clouds*, 2004, submitted, http://www.mat.univie.ac.at/~neum/papers.html#struc
29. Neumaier, A.: The Wrapping Effect, Ellipsoid Arithmetic, Stability and Confidence Regions, *Computing Supplementum* **9** (1993), pp. 175–190.
30. Sahinidis, N. V.: BARON: A General Purpose Global Optimization Software Package, *J. Global Optim.* **8** (1996), pp. 201–205.
31. Shafer, G.: *A Mathematical Theory of Evidence*, Princeton Univ. Press, Princeton, 1976.
32. Springer, M. D.: *The Algebra of Random Variables*, Wiley, New York, 1979.
33. Stetter, H. J.: Algebraic Predicates for Empirical Data, in: Ganzha, V. G., Mayr, E. W., and Vorozhtsov, E. V. (eds), *Computer Algebra in Scientific Computing—CASC 2001*, Springer, 2001.
34. Walley, P.: *Statistical Reasoning with Imprecise Probabilities*, Chapman & Hall, London, 1991.
35. Weichselberger, K.: Interval Probability on Finite Sample-Spaces, in: Rieder, H. (ed.), *Robust Statistics, Data Analysis and Computer-Intensive Methods*, Springer, New York, 1996, pp. 391–409.
36. Whittle, P.: *Probability via Expectation*, Springer, Berlin, 1992.
37. Williamson, R. C.: *Probabilistic Arithmetic*, Ph.D. Thesis, University of Queensland, 1989, http://axiom.anu.edu.au/~williams/papers/thesis300dpi.ps
38. Yeh, A.: Finding the Likely Behaviors of Static Continuous Nonlinear Systems, *Ann. Math. Artif. Intell.* **2** (1990), pp. 383–398.