Joslyn, Cliff A: (1992) "Possibilistic Measurement and Set Statistics", in: Proc. 1992 Conf. of the North American Fuzzy Information Processing Society (NAFIPS 1992), v. 2, pp. 458-467, Puerto Vallerta

Possibilistic Measurement and Set Statistics*

Cliff Joslyn † ‡

Abstract

Set-based statistics are necessary to generate possibility distributions from measured data. Methods by which physical measurements can generate statistical data on real intervals are considered, including trials from multiple heterogeneous measurement devices rather than a single instrument at multiple times; classes of consistent intervals constructed from statistical data around a common point focus or interval core; and consonant intervals constructed from statistical data.

1 Introduction

My overall interest is to expand the applications of possibility theory beyond its traditional uses in the engineering of human-created technological systems (e.g. knowledge-based control systems, artificial intelligence and approximate reasoning, etc.) to include the modeling of natural, complex systems. In order to do this, it is necessary to extend the semantics of possibility beyond traditional interpretations based on the uncertainty judgments of human subjects. Instead, a semantics of possibility that has meaning with respect to natural systems is needed.

Existing empirical methods for deriving possibility distributions are frequency conversion methods, which transform some measured probabilistic data into a possibilistic form [16]. Of course such transformations must be used when *only* frequency data are available, but the resulting possibilistic representation is never ultimately *appropriate* for data initially governed by a frequency distribution. When possibilistic data are desired, it is always preferable to obtain them in a form more directly similar to their possibilistic representation.

The additivity of frequency data results from the specificity of observations of singletons, or indeed elements of any disjoint class. Therefore, the first step towards possibilistic measurement is allowing for the possibility of non-specific measurements, that is observations that are possibly non-disjoint. This is essentially the concept of set statistics, originally advanced by Wang and Liu [17], and developed more by Dubois and Prade [4, 6].

Frequency counts on subsets result in empirically derived random sets. In earlier papers, Joslyn [9, 10] and Joslyn and Klir [11] considered methods for deriving a possibility distribution from a given empirical random set. In this paper, methods for the collection of set statistics are developed, including direct collection of interval data, and also generation of intervals from point-data streams.

2 Mathematical Preliminaries

We begin with the standard evidence and possibility theory [3, 14]. Given a finite universe $\Omega = \{\omega_i\}, 1 \leq i \leq n$, the set function $m\colon 2^\Omega \mapsto [0,1]$ is an **evidence function** (otherwise known as a **basic assignment** or **basic probability assignment**) when $m(\emptyset) = 0$ and $\sum_{A \subset \Omega} m(A) = 1$. Denote a random set generated from an evidence function as $\mathcal{S} = \{\langle A_j, m_j \rangle : m_j > 0\}$, where $\langle \cdot \rangle$ is a vector, $A_j \subset \Omega, m_j = m(A_j)$, and $1 \leq j \leq N = |\mathcal{S}| \leq 2^n - 1$. Denote the **focal set** as $\mathcal{F} = \{A_j : m_j > 0\}$ with **core** $\mathbf{C}(\mathcal{F}) = \bigcap_{\mathcal{F}} A_j$. The dual belief and plausibility measures on $\forall A \subset \Omega$ are $\mathrm{Bel}(A) = \sum_{A_j \subset A} m_j$ and $\mathrm{Pl}(A) = \sum_{A_j \not\perp A} m_j$, where $A \perp B := A \cap B = \emptyset$.

*Prepared for the Conference of the North American Fuzzy Information Processing Society, December 1992, Puerto Vallarta.

†Graduate Fellow, Systems Science, SUNY–Binghamton, 327 Spring St. # 2, Portland ME, 04102, USA, (207) 774-0029, cjoslyn@bingsuns.cc.binghamton.edu, joslyn@kong.gsfc.nasa.gov.

‡Supported under NASA Grant # NGT 50757.

The **plausibility assignment** (otherwise known as the **contour function**, **falling shadow**, or **one-point coverage function**) of $\mathcal{S}$ is

$$\vec{\mathrm{Pl}} = \langle\, \mathrm{Pl}(\{\omega_i\}) \,\rangle = \langle\, \mathrm{Pl}_i \,\rangle\,, \qquad \mathrm{Pl}_i = \sum_{A_j \ni \omega_i} m_j.$$

$\vec{\mathrm{Pl}}$ is a fuzzy set that can be mapped to an equivalence class of random sets [8].

When $\forall A_j \in \mathcal{F}, |A_j| = 1$, then $\mathcal{S}$ is **specific**, and $\mathrm{Bel}(A_j) = \mathrm{Pl}(A_j) = \mathrm{Pr}(A_j)$ is an additive **probability measure** with **probability distribution** $\vec{\mathrm{Pl}} = \vec{p} = \langle\, p_i \,\rangle$ with additive normalization $\sum_i p_i = 1$. $\mathcal{S}$ is **consonant** ($\mathcal{F}$ is a **nest**) when (without loss of generality for ordering, and letting $A_0 = \emptyset$) $A_{j-1} \subset A_j$. Now $\mathrm{Pl}(A_j) = \Pi(A_j)$ is a **possibility measure**. As Pr is additive, so Π is **maximal** in the sense that $\Pi\left(\bigcup_j A_j\right) = \bigvee_j \Pi(A_j)$, where $\vee$ is the maximum operator. Denoting $A_i = \{\omega_1, \omega_2, \ldots, \omega_i\}$, and assuming that $\mathcal{F}$ is complete (i.e. $\forall \omega_i \in \Omega, \exists A_i$), then $\vec{\mathrm{Pl}} = \vec{\pi} = \langle\, \pi_i \,\rangle$ is a **possibility distribution** with maximal normalization $\bigvee_i \pi_i = 1$.

2.1 Consistency and Consonance

$\mathcal{S}$ is **consistent** when $\mathbf{C}(\mathcal{F}) \neq \emptyset$. Each consonant random set is consistent with core $\mathbf{C}(\mathcal{F}) = A_1$, and $\mathcal{F}$ being consistent is both necessary and sufficient for $\bigvee \mathrm{Pl}_i = 1$. Thus a consistent but non-consonant random set has a maximal possibility distribution $\vec{\mathrm{Pl}} = \vec{\pi}$, but its plausibility measure Pl is *not* a possibility measure Π. While an additive probability distribution uniquely determines a measure and random set, a maximal possibility distribution does not. However, a possibility measure Π^* that is optimally approximate can be constructed according to the formula $\forall A \subset \Omega, \Pi^*(A) = \bigvee_{\omega_i \in A} \pi_i$ [5]. When $\mathcal{S}$ is already consonant, then of course $\Pi^* = \mathrm{Pl} = \Pi$.

Dubois and Prade [3] suggest that the plausibility assignment of a consistent but non-consonant random set $\vec{\mathrm{Pl}} = \vec{\pi}$ should not be taken as a possibility distribution, but rather should be used to derive a nest from which a possibility distribution can be generated. That nest is the focal set of the constructed possibility measure Π^*, denoted $\mathcal{F}^* = \{B_k^*\}$. The evidence for each focal element, denoted $m_k^* = m(B_k^*)$, is given by the formula

$$m_k^* = \sum_{A_j \subset B_k} m_j - m_{k-1}^*$$

where $m_0^* = 0$. This method results in a greater constraint on the evidence provided by m, and thus the loss of some information available in a consistent $\mathcal{S}$ (see example in Section 4).

2.2 Consistent Transformations

When $\mathcal{F}$ is not consistent, then $\bigvee \mathrm{Pl}_i < 1$. Here a set of **focused consistent transformations** $\hat{\mathcal{S}}_i$ can be constructed from $\mathcal{S}$ [10, 11]. $\forall \omega_i \in \Omega$, $\hat{\mathcal{S}}_i$ is a consistent approximation of $\mathcal{S}$ with evidence function [10]

$$\hat{m}^i(A) = \begin{cases} m(A) + m(A - \{\omega_i\}), & \omega_i \in A \\ 0, & \omega_i \notin A \end{cases}.$$

The effect is to create a core $\mathbf{C}(\hat{\mathcal{F}}_i) = \{\omega_i\}$ with **focus** $\omega_i = \omega^*$. Under the transformation $\hat{\mathcal{S}}_i$, the sub-maximal plausibility assignment $\vec{\mathrm{Pl}} = \langle\, \mathrm{Pl}_1, \mathrm{Pl}_2, \ldots, \mathrm{Pl}_i, \ldots, \mathrm{Pl}_n \,\rangle$ is transformed into a maximal possibility distribution $\vec{\pi} = \langle\, \mathrm{Pl}_1, \mathrm{Pl}_2, \ldots, 1, \ldots, \mathrm{Pl}_n \,\rangle$. $\hat{\mathcal{S}}_i$ in turn generates a consonant random set $\hat{\mathcal{S}}_i^\pi$, determined from the constructed possibility measure Π^* of $\vec{\pi}$.

In using the transformation the task is to choose the "correct" ω^* as a focus, and to elevate the plausibility of that element to 1 as a possibilistic normalization. While there are many methods to choose ω^*, to date only the Principle of Minimal Information Distortion [10] (or Information Loss [11]) has been studied. Given a random set $\mathcal{S}$, then that focused consistent transformation $\hat{\mathcal{S}}_i$ is selected so that the total information content of $\hat{\mathcal{S}}_i^\pi$ is as close as possible to that of the original $\mathcal{S}$. Details of the measure of total information can be found elsewhere [7, 10, 15].

3 Empirical Random Sets

Assume that some phenomenal system can be described as a set $\Omega = \{\omega_i\}, 1 \le i \le n$. A traditional conception of a measurement on Ω results in the observation of an element $\omega_i \in \Omega$. For example, a thermometer calibrated in integral degrees on the interval $[0, 100]$ could yield a result of 72 degrees, $72 \in \{0, 1, \ldots, 100\}$.

Assume a counting function $c \colon \Omega \mapsto \mathcal{I}, c_i = c(\omega_i)$, where c_i is the number of observations of ω_i. Then for a total number of counts as N, the frequency distribution on Ω is $f \colon \Omega \mapsto [0,1], f(\omega_i) = f_i = c_i/N$. Since $\sum_i f_i = 1$, therefore f is a natural probability distribution on Ω with an additive measure $F \colon 2^\Omega \mapsto [0,1], F(A) = \sum_{\omega_i \in A} f_i$.

3.1 General Measuring Devices

However, most real measuring devices are not like this, due to necessary measurement uncertainty. Most measurements produce an observation of some *subset* $A \subset \Omega$, perhaps an interval $A \subset \Re$. The observation of the interval A leaves uncertainty as to the "actual" value $\omega \in A$.

It may be that not all subsets are observable. Thus a **general measuring device** is defined as a class $\mathcal{C} = \{A_{j'}\} \subset 2^\Omega, 1 \le j' \le N'$. The nature of the measuring device will depend on the elements and structure of $\mathcal{C}$.

Assume a collection of set observations $A^k \in \mathcal{C}, 1 \le k \le M$. In general, for some k_1, k_2, it may be that $A^{k_1} = A^{k_2}$. Therefore the A^k form a multi-set, denoted as a vector $\vec{A} = \langle A^1, A^2, \ldots, A^M \rangle$. The **empirically derived focal set** $\mathcal{F}^E \subset \mathcal{C}$ is the set of subsets that are actually observed in $\vec{A}$. $\mathcal{F}^E$ is derived by eliminating the duplicates in $\vec{A}$. Let $\mathcal{F}^E = \{A_j\}$, where $\mathcal{F}^E \subset \mathcal{C}, 1 \le j \le N \le N', N \le M$ and $\forall A_j \in \mathcal{F}^E, A_j \in \vec{A}$, and inclusion of an element in a vector is defined as would be expected.

Now establish a set-counting function $C \colon \mathcal{F}^E \mapsto \mathcal{I}, C_j = C(A_j)$, where $\forall A_j \in \mathcal{F}^E, C_j$ is the number of occurrences of A_j in $\vec{A}$. Finally the **set-frequency function** is arrived at

$$m^E \colon \mathcal{F}^E \mapsto [0,1], \qquad m^E(A_j) = m^E_j = \frac{C_j}{\sum_{A_j \in \mathcal{F}^E} C_j} = C_j/M.$$

The intention is obvious: since $\sum_j m^E_j = 1$ and $\emptyset \notin \mathcal{F}^E$, therefore m^E is a natural evidence function on Ω generating an **empirically derived random set** denoted $\mathcal{S}^E$.

3.2 Disjoint Measuring Devices

Generally, scientists strive to construct measuring devices for which $\mathcal{C}$ is disjoint; that is, $\forall A_1, A_2 \in \mathcal{C}, A_1 \perp A_2$. In such classical measuring devices, $\mathcal{C}$ is an equivalence class on Ω, yielding an observation of an $A^k \in \mathcal{C}$ unambiguous.

Virtually all traditional measuring devices are of this type. A typical example could be a thermometer, where $\Omega \subset \Re$ is some distance along a glass tube marked at certain points, say d_j, with a certain number of degrees. The A could then be the disjoint, equal length, half-open intervals $A_j = [d_j, d_{j+1})$. Observation of a specific position of the mercury (an $\omega \in A_j$) yields a specific A_j reading for the temperature. The size of the A_j relative to the size of the tube indicates the precision of the thermometer. While any particular interval A_j is usually identified with one degree reading (either d_j or d_{j+1}), it must always be kept in mind that it in fact indicates the *entire* interval $[d_j, d_{j+1})$.

Because the A_j are disjoint, observation of any one particular interval admits to no uncertainty *at the level of description of* $\mathcal{C}$. Thus in this case $\mathcal{C}$ itself can be considered as a new universe of discourse $\Omega' = \mathcal{C} = \{A_j\}$. Because the A_j are disjoint, so will the A^k.

Now m^E is the frequency of the disjoint A_j, and is thus a true probability distribution, and not an evidence function proper. Measurements from a classical measuring device are usually parameterized in time k, yielding the observations A^k as time-series point data. An additive distribution and measure are derived as for frequencies above

$$c' \colon \Omega' \mapsto \mathcal{I}, \qquad f' \colon \Omega' \mapsto [0,1], \qquad F' \colon 2^{\Omega'} \mapsto [0,1]$$

$$c'(A_j) = c'_j = C(A_j), \qquad f'(A_j) = f'_j = m^E_j, \qquad \sum_j f'_j = 1, \qquad F'(B \subset \Omega') = \sum_{A_j \in B} f'_j.$$

4 Instrument Ensembles

One way to generate measurements of intersecting subsets is to use an ensemble of classical instruments. That ensemble can be considered as either multiple, heterogeneous instruments taking separate measurements at the same time, or as a single instrument which is changing its structure over time.

Let $\mathcal{C}^k = \left\{ A^k_{j'_k} \right\}, 1 \le j'_k \le N'_k = |\mathcal{C}_k|$ be disjoint classes on Ω, and $\mathbf{F} = \{\mathcal{C}^k\}$ be the family of such classes, $1 \le k \le M$. The natural partial order on $\mathbf{F}$ is

$$\mathcal{C}^1 \prec \mathcal{C}^2 \quad := \quad \forall A^2_{j_2} \in \mathcal{C}^2, \quad \exists \{A^1_{j_1}\} \subset \mathcal{C}^1, \quad A^2_{j_2} = \bigcup A^1_{j_1}.$$

When $\mathcal{C}^1 \prec \mathcal{C}^2$ then $\mathcal{C}^1$ refines $\mathcal{C}^2$, and $\mathcal{C}^2$ coarsens $\mathcal{C}^1$. For example, $\mathcal{C}^1$ could be a thermometer reading in tenths of degrees, while $\mathcal{C}^2$ could belong to a mutually calibrated thermometer reading in whole degrees. $\mathbf{F}$ is **consonant** whenever the $\mathcal{C}^k$ are all comparable under $\prec$ (they are all mutual refinements or coarsenings).

Leting A^k be the subset observed in device $\mathcal{C}^k$, then the vector of observations over $\mathbf{F}$ is $\vec{A} = \langle A^k \rangle, |\vec{A}| = M$, and $\vec{A}$ generates the empirical random set $\mathcal{S}^E$ as described in Section 3.1. If any of the $\mathcal{C}^k$ share common members (in particular, if any of them are equal), then some of the A^k may be equal, yielding multiple observations in $\vec{A}$ of certain subsets. Otherwise, all subsets will be observed a single time, and will not necessarily be disjoint.

Assume observations from two devices, say $A^1 \in \mathcal{C}^1$ and $A^2 \in \mathcal{C}^2$. It is expected that $A^1 \not\perp A^2$. In the event that $A^1 \perp A^2$, then at least one of the devices $\mathcal{C}^1$ or $\mathcal{C}^2$ would be regarded as being in error, or perhaps even the assumption of the "reality" of the quantity being measured would be questioned. Thus, while there is nothing in the mathematics that would preclude such a result, pragmatic conditions require that $\mathcal{F}^E$ be consistent, so that $\mathcal{S}^E$ has a natural possibility distribution π and at worst a constructed possibility measure Π^*. In the event that $\mathcal{F}^E$ is nevertheless not consistent, and there are pragmatic reasons for accepting the results of the measurement, then the focused consistent transformation method outlined in Section 2.2 is available to construct consistent random sets $\hat{\mathcal{S}}_i$.

When $\mathbf{F}$ is consonant, then without loss of generality for ordering, $\mathcal{C}^1 \prec \mathcal{C}^2 \prec \ldots \prec \mathcal{C}^M$. Here if $\mathcal{F}^E$ is consistent, then it must also be consonant, with $A_1 \subset A_2 \subset \ldots \subset A_N$. Of course, in this case a possibilistic analysis is less useful than it would be otherwise, since there is an absolute gain in accuracy in the movement towards the finest measurement A^1. Nevertheless, the mathematical analysis is available.

Example 1: Let $\Omega = [0,5] \subset \Re$ and define a family $\mathbf{F}$ of four measuring devices

$$\begin{array}{ll} \mathcal{C}^1 = \{[0,1),[1,2),[2,3),[3,4),[4,5]\}, & \mathcal{C}^2 = \{[0,1),[1,2),[2,3.5),[3.5,5]\}, \\ \mathcal{C}^3 = \{[0,1.5),[1.5,3.5),[3.5,4),[4,5]\}, & \mathcal{C}^4 = \{[0,1.5),[1.5,4),[4,5]\}, \end{array}$$

so that $M = 4$. $\mathbf{F}$ is not consonant, but $\mathcal{C}^3 \prec \mathcal{C}^4$. Measurements are made on each instrument yielding a vector of four measurements (Figure 1)

$$\vec{A} = \langle [1,2),[1,2),[1.5,3.5),[1.5,4) \rangle.$$

After eliminating duplicates, the set of observed intervals $\mathcal{F}^E$ is derived with $N = 3 < M$ and random

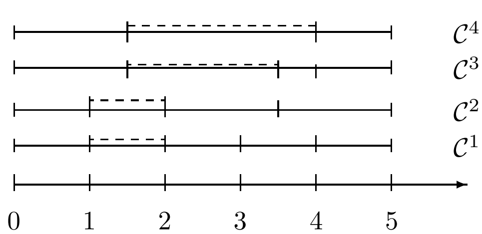

Figure 1: Measurements on four instruments.

set $\mathcal{S}^E$

$$\mathcal{F}^E = \{[1,2),[1.5,3.5),[1.5,4)\}, \qquad \mathcal{S}^E = \{\langle [1,2),.5 \rangle, \langle [1.5,3.5),.25 \rangle, \langle [1.5,4),.25 \rangle\}.$$

$\mathcal{F}^E$ is consistent with core $\mathbf{C}(\mathcal{F}^E) = [1.5, 2)$, the region on which $\pi = 1$. $\pi(\omega)$ is determined by $\pi(\omega) = \sum_{A_j \ni \omega} m_j^E$, so that

$$\pi(\omega) = \begin{cases} .5, & \omega \in [1, 1.5) \\ 1, & \omega \in [1.5, 2) \\ .5, & \omega \in [2, 3.5) \\ .25, & \omega \in [3.5, 4) \\ 0, & \text{elsewhere} \end{cases}$$

as shown in Figure 2.

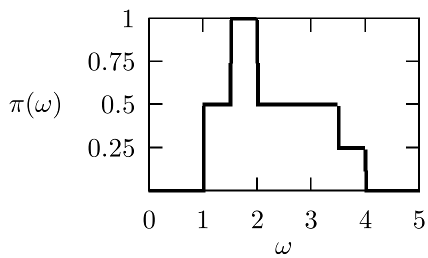

Figure 2: π determined from $\mathcal{S}^E$.

Dubois and Prade's method described in Section 2.1 results in the consonant random set

$$\{\langle [1.5, 2), 0 \rangle, \langle [1, 3.5), .75 \rangle, \langle [1, 4), .25 \rangle\}$$

and possibility distribution shown in Figure 3. Comparing Figures 2 and 3, it can be seen that reliance on the consonant class and its greater constraint results in a loss of distinctions of possibility values over portions of the possibility curve.

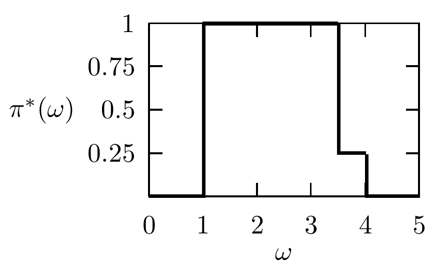

Figure 3: π^* determined from Dubois and Prade's method.

Because $\mathcal{F}^E$ is finite, π is piecewise continuous, consisting of a union of constant segments. Also, because $\mathbf{C}(\mathcal{F}^E)$ is connected, π is unimodal at $\mathbf{C}$. Therefore π in this example, and in the sections to follow, has the form of a centrally peaked staircase. As $|\mathcal{F}^E| \longrightarrow \infty$, π approaches the traditional forms for possibility distributions (e.g., fuzzy numbers [2]).

5 Consistent Intervals from Focused Point Data

Even given a single measuring device and time-series data gathered on it (as discussed in Section 3.2), which is our normal concept of measurement, interval data can still be generated. Since classical instruments generate observations of disjoint intervals that can be regarded as distinct points in a higher-level state space, therefore in the following sections a single measuring device that yields observations of points in a lower-level state space, a closed interval $\Omega \subset \Re$, will be considered.

Denote an observation as a data point $d \in \Omega$, and the collection of data as a **data stream**, a multiset denoted as the vector $\vec{D} = \langle d_i \rangle, 1 \le i \le n$. The set generated by eliminating duplicates in $\vec{D}$ is the **data set** $D = \{d'_i\}, 1 \le i \le n' \le n$.

A possibilistic analysis of $\vec{D}$ will be approached by using its **order statistics** [1]. For a given data stream $\vec{D}$, the order statistics, denoted $d_{(i)}$, are a permutation of the d_i such that $d_{(1)} \le d_{(2)} \le \ldots \le d_{(n)}$. $d_{(1)}$ and $d_{(n)}$ are called the **extremes**, and the **range** interval is $W = \left[d_{(1)}, d_{(n)}\right]$. The order statistics of the data set D' are $d'_{(i)}, 1 \le i \le n'$. The $d'_{(i)}$ naturally generate the disjoint intervals $\delta_i = \left[d'_{(i)}, d'_{(i+1)}\right), 1 \le i \le n' - 1$. For completeness, let $\delta_{n'} = \left[d'_{(n')}, d'_{(n')}\right] = \left\{d'_{(n')}\right\}$. Let the set of disjoint intervals be $\Delta = \{\delta_i\}$, so that $\bigcup_{\delta_i \in \Delta} \delta_i = \Omega$.

5.1 Focused Data Intervals

Δ thus represents a classical measuring device with the δ_i partitioning W, and so the greatest problem with deriving a possibility distribution from Δ is the lack of a focus, or any core. Thus we posit the existence of a focus $u \in W$. The purpose of u is to provide a value on which all the intervals (yet to be determined) agree; a value for which $\pi(u) = 1$. u naturally divides W into left and right sub-intervals denoted $W_l = \left[d_{(1)}, u\right)$ and $W_r = \left(u, d_{(n)}\right]$ so that $W_l \cup [u, u] \cup W_r = W$.

Given a focus $u \in W$, then $\forall d_{(i)} \ne u, d_{(i)} \in W_l$ or $d_{(i)} \in W_r$. Denote the intervals $A^i, 1 \le i \le n$ as follows:

$$A^i = \begin{cases} \left[d_{(i)}, u\right], & d_{(i)} \in W_l \\ \left[u, d_{(i)}\right], & d_{(i)} \in W_r \\ [u, u], & d_{(i)} = u \end{cases}.$$

Since

$$d_{(i_1)}, d_{(i_2)} \in W_l, \quad i_1 \le i_2 \quad \rightarrow \quad A^{i_2} \subset A^{i_1}; \qquad \text{and} \qquad d_{(i_1)}, d_{(i_2)} \in W_r, \quad i_1 \le i_2 \quad \rightarrow \quad A^{i_1} \subset A^{i_2},$$

therefore each of the sets of intervals

$$\mathcal{F}_l = \{A^i : d_{(i)} \in W_l\}, \qquad \mathcal{F}_r = \{A^i : d_{(i)} \in W_r\},$$

are nests. Since $\forall i, u \in A^i$, the total set $\{A^i\}$ is consistent, forming a focal set $\mathcal{F}^E = \mathcal{F}_l \cup \mathcal{F}_r$ with core $\mathbf{C}(\mathcal{F}^E) = [u, u] = \{u\}$. $\mathcal{S}^E$ is then constructed from the counts of the $d_{(i)} \in \vec{D}$ of the corresponding interval A^i.

Generally, each $d_{(i)}$ will generate a single count for the interval A^i. However, if $\exists i_1, i_2, A^{i_1} = A^{i_2}$ then multiple counts will be generated as discussed in Section 4. If $u = d_{(1)}$ or $u = d_{(n)}$ then $\mathcal{F}^E$ will actually be consonant.

Example 2: As above, let $\Omega = [0, 5]$, and assume that $n = 6$ point observations in Ω are taken giving the data stream $\vec{D} = \langle 2, 1, 4, 1.5, 2, 4.5 \rangle$. The order statistics are

$$d_{(1)} = 1, \quad d_{(2)} = 1.5, \quad d_{(3)} = d_{(4)} = 2, \quad d_{(5)} = 4, \quad d_{(6)} = 4.5$$

and $W = [1, 4.5]$. The corresponding data set is $D' = \{1, 1.5, 2, 4, 4.5\}$ so that $n' = 5 < n$, with order statistics and disjoint intervals

$$d'_{(1)} = 1, \quad d'_{(2)} = 1.5, \quad d'_{(3)} = 2, \quad d'_{(4)} = 4, \quad d'_{(5)} = 4.5$$

$$\Delta = \{[1, 1.5), [1.5, 2), [2, 4), [4, 4.5), [4.5, 4.5]\}$$

Assuming that $u \in [2, 4]$, then the focal and random sets (Figure 4, with $u = 3$) are

$$\mathcal{F}^E = \mathcal{F}_l \cup \mathcal{F}_r = \{[1, u], [1.5, u], [2, u]\} \cup \{[u, 4], [u, 4.5]\},$$

$$\mathcal{S}^E = \{\langle [1, u], 1/6 \rangle, \langle [1.5, u], 1/6 \rangle, \langle [2, u], 1/3 \rangle, \langle [u, 4], 1/6 \rangle, \langle [u, 4.5], 1/6 \rangle\}.$$

The possibility distribution is shown in Figure 5.

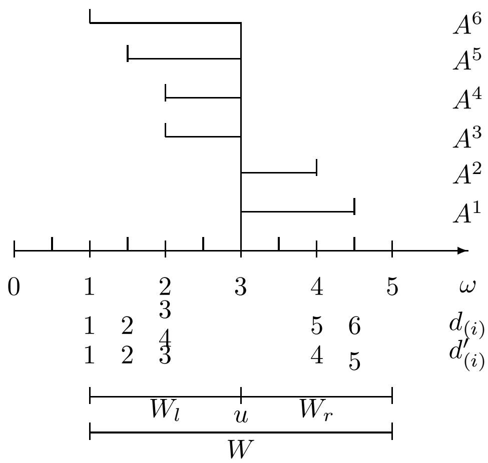

Figure 4: Consistent family from focused data set.

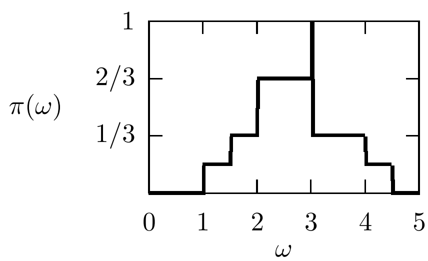

Figure 5: Derived possibility distribution.

5.2 Choice of Focus

So far the method by which the focus u can be chosen has not been discussed. While a number of methods suggest themselves, selection of methods will depend on user methodology and further empirical research. However, in Example 2 the first four methods below all yield $u \in [2, 4]$, which is the inner interval of Δ (see Section 6).

Sample Mean: Selection of

$$u = \bar{D} = \sum d_i / n$$

is a possibility, although one that is not in keeping with possibilistic concepts. In our example, this would yield $u = 2.5$.

Range Midpoint: The midpoint of W, denoted $\bar{W}$, is much more in keeping with possibilistic concepts:

$$u = \bar{W} = \frac{d_{(1)} + d_{(n)}}{2}.$$

It expresses something like the concept of a "possibilistic sample mean". This would yield $u = 2.75$ in the example.

Closest to Range Midpoint: There may be some value in having u actually be one of the data points, so that $u \in D'$. This can be done by selecting that $d'_i \in D'$ closest to $\bar{W}$ (yielding $u = 2$ in our example):

$$u = \min_{d'_i \in D'} |d'_i - \bar{W}|.$$

Data-Set Midpoint: The middle point of the data set itself can be chosen, that is

$$u = d'_{\left(\frac{n'+1}{2}\right)}$$

if n' is odd. If n' is even, then either

$$u = d'_{(n'/2)} \text{ or } u = d'_{\left(\frac{n'}{2}+1\right)}.$$

Alternatively, if n' is even then the midpoint of the central interval can be selected:

$$u = \frac{d'_{(n'/2)} + d'_{\left(\frac{n'}{2}+1\right)}}{2}.$$

Information Principles: Finally, the Information Principles introduced in Section 2.2 can be applied to the problem [11, 12]. Again, details will not be given here. Selection of u can be regarded as a problem of ampliative reasoning, of making an inductive inference beyond the given information. Then the Principle of Maximum Uncertainty can be invoked, which states that u should be chosen so as to maximize the total uncertainty of the resulting random set, or of the final possibility distribution.

Alternatively, selection of u can be regarded as one of transformation from the frequency distribution of $\vec{D}$ to a possibility distribution. Then the Principle of Uncertainty Invariance [13] or Minimal Information Distortion [10] can be used, which states that u should be chosen so as to make the total uncertainty of $\mathcal{S}^E$ as close as possible to the entropy $\vec{D}$.

6 Interval Cores

A potential disadvantage of the methods in Section 5 is the reliance on a singleton-valued core set $\mathbf{C}(\mathcal{F}^E) = \{u\}$, while the other elements of the method are the intervals δ_i and A^i. Instead, methods that yield an interval-valued core can be considered. A disadvantage of these methods is that they may eliminate some data points, thus loosing some information from the resulting $\mathcal{S}^E$.

Let $\vec{D}, D'$ and Δ be given as above. Now identify the core as an interval in the range with endpoints $\mathbf{C}_l$ and $\mathbf{C}_r$, so that $\mathbf{C} = [\mathbf{C}_l, \mathbf{C}_r] \subset W$. Assume for the moment that $\nexists d_{(i)} \in \mathbf{C}$. Then the left and right ranges can be redefined as $W_l = \left[d_{(1)}, \mathbf{C}_l\right)$ and $W_r = \left(\mathbf{C}_r, d_{(n)}\right]$, so that $W_l \cup \mathbf{C} \cup W_r = W$. Also redefine the intervals A^i as follows:

$$A^i = \begin{cases} \left[d_{(i)}, \mathbf{C}_r\right], & d_{(i)} \in W_l \\ \left[\mathbf{C}_l, d_{(i)}\right], & d_{(i)} \in W_r \end{cases} .$$

Again $\mathcal{F}_l$ and $\mathcal{F}_r$ are nests, so that $\mathcal{F}^E = \mathcal{F}_l \cup \mathcal{F}_r$ is consistent with core $\mathbf{C}(\mathcal{F}^E) = \bigcap A^i = \mathbf{C}$.

If $\exists \{d_{(k)}\} \subset \mathbf{C}$, then a new data set $\vec{D}^- = \vec{D} - \{d_{(k)}\}$ is defined, where the operation $-$ of a set from a vector is just the elimination of $\forall d_{(k)}$ from $\vec{D}$. Corresponding new $d^-_{(i)}, D'^-$, etc. can be generated without special treatment.

6.1 Choice of C

As with the selection of point foci, there are a variety of methods by which an interval core can be selected.

Central Disjoint Interval: If n' is even, then a central disjoint interval is naturally generated from the data set D':

$$\mathbf{C} = \delta_{n'/2}.$$

Since $d'_{(n'/2)}, d'_{\left(\frac{n'}{2}+1\right)} \in \mathbf{C}$, all instances of them will be eliminated from $\vec{D}$ in forming $\vec{D}^-$.

Modified Central Interval: If n' is odd, then there are two disjoint intervals on either side of $d'_{\left(\frac{n'+1}{2}\right)}$. Thus a core would be selected

$$\mathbf{C} = \delta_{\frac{n-1}{2}} \cup \delta_{\frac{n+1}{2}},$$

that eliminates instances of the three data points $d'_{\left(\frac{n'-1}{2}\right)}, d'_{\left(\frac{n'+1}{2}\right)}$, and $d'_{\left(\frac{n'+3}{2}\right)}$ from $\vec{D}$.

Alternatively, the midpoints of the two disjoint intervals around $d'_{\left(\frac{n'+1}{2}\right)}$ can be selected as the endpoints of $\mathbf{C}$:

$$\mathbf{C} = \left[\frac{d'_{\left(\frac{n'-1}{2}\right)} + d'_{\left(\frac{n'+1}{2}\right)}}{2}, \frac{d'_{\left(\frac{n'+1}{2}\right)} + d'_{\left(\frac{n'+3}{2}\right)}}{2} \right].$$

Disjoint Interval Around Focus: Given a method from Section 5.2 to select a point focus u, then $\mathbf{C}$ can just be selected as the data-generated disjoint interval around u:

$$\mathbf{C} = \delta_i, \quad u \in \delta_i.$$

As above, instances of $d'_{(i)}$ and $d'_{(i+1)}$ will be eliminated from $\vec{D}$.

Confidence Interval Around Focus: It may be appropriate for the user to involve some traditional statistical information. Again, given some focus u, then $\mathbf{C}$ can be selected as the interval within a standard deviation of u:

$$\mathbf{C} = \left[u - \sigma(\vec{D}), u + \sigma(\vec{D})\right].$$

Information Principles: Methods of Uncertainty Maximization or Invariance can be applied, as discussed in Section 5.2.

7 Consonant Intervals from Focused Point Data

It may be desirable to go as far as generating consonant, not just consistent, families from a data stream $\vec{D}$. However, as the methods progress from consistent families with point focuses, through consistent families with interval cores, to consonant classes, the constraint on $\mathcal{S}^E$ increases, thus loosing information available in the original $\vec{D}$. This is reflected in the loss of some data points in the interval core methods, and in roughly half the number of available intervals from the following consonant methods. Thus as with the case of an ensemble of measuring devices (Section 4), use of strictly consonant cases may be less useful than simply consistent cases.

Again, a number of methods present themselves.

Inner Nested Intervals from Interval Core: Assume that an interval core $\mathbf{C} = [\mathbf{C}_l, \mathbf{C}_r]$ has been determined according to some method discussed in Section 6.1. Denote $A^1 = \mathbf{C}$, and construct a set of intervals $A^k = \left[A_l^k, A_r^k\right]$ such that $A_l^k, A_r^k \in D'$ and $A^k \subset A^{k+1}$. Given an interval A^k, then A^{k+1} is the nearest interval determined by D' containing A^k

$$A_l^{k+1} = \max_{d'_{(i)} \in D'} d'_{(i)} < A_l^k, \qquad A_r^{k+1} = \min_{d'_{(i)} \in D'} d'_{(i)} > A_r^k.$$

The A^k are available up to a maximal $A^{\lfloor n'/2 \rfloor} = W$. $\mathcal{F}^E = \{A^k\}$ is then a consonant class. The count of A^k can be determined as the maximum number of occurrences of either endpoint of A^k in $\vec{D}$.

Inner Nested Intervals from Point Focus: Assume instead that a point core $u \in W$ has been determined according to some method discussed in Section 5.2. Now simply let $A^1 = [u, u]$ and apply the method above.

Outer Nested Intervals: Proceed in the opposite direction from above. Now define $A^1 = W$, and construct A^{k+1} from A^k as follows:

$$A_l^{k+1} = \min_{d'_{(i)} \in D'} d'_{(i)} > A_l^k, \qquad A_r^{k+1} = \max_{d'_{(i)} \in D'} d'_{(i)} < A_r^k.$$

Acknowledgements

Prof. George Klir and Dr. Peter Cariani provided useful discussion on these subjects. I would also like to thank an anonymous reviewer for his or her helpful comments.

References

[1] David, Herbert A: (1981) *Order Statistics*, Wiley, New York

[2] Dubois, Didier: (1987) "Fuzzy Numbers: An Overview", in: *Analysis of Fuzzy Information*, v. **1 of 3**, pp. 3-39, CRC Press, Boca Raton

[3] Dubois, Didier and Prade, Henri: (1988) *Possibility Theory*, Plenum Press, New York

[4] Dubois, Didier and Prade, Henri: (1989) "Fuzzy Sets, Probability, and Measurement", *Eur. J. of Operational Research*, v. **40**:2, pp. 135-154

[5] Dubois, Didier and Prade, Henri: (1990) "Consonant Approximations of Belief Functions", *Int. J. Approximate Reasoning*, v. **4**, pp. 419-449

[6] Dubois, Didier and Prade, Henri: (1992) "Evidence, Knowledge and Belief Functions", *Int. J. Approximate Reasoning*, v. **6**:3, pp. 295-320

[7] Geer, James and Klir, George: (1991) "Discord in Possibility Theory", *Int. J. Gen. Sys.*, v. **19**, pp. 119-132

[8] Goodman, IR and Nguyen, HT: (1986) *Uncertainty Models for Knowledge-Based Systems*, North-Holland, Amsterdam

[9] Joslyn, Cliff: (1991) "Towards an Empirical Semantics of Possibility Through Maximum Uncertainty", in: *Proc. IFSA 1991*, v. **A**, pp. 86-89

[10] Joslyn, Cliff: (1992) "Empirical Possibility and Minimal Information Distortion", in: *Fuzzy Logic: State of the Art*, ed. R. Lowen, Kluwer, in press

[11] Joslyn, Cliff and Klir, George: (1992) "Minimal Information Loss Possibilistic Approximations of Random Sets", in: *Proc. 1992 FUZZ-IEEE Conference*, ed. Jim Bezdek, pp. 1081-1088, IEEE, San Diego

[12] Klir, George: (1990) "Uncertainty Principles in Reasoning", in: *Proc. 1990 NAFIPS*, pp. 212-214

[13] Klir, George: (1990) "Principle of Uncertainty and Information Invariance", *Int. J. Gen. Sys.*, v. **17**:2-3, pp. 249-275

[14] Klir, George and Folger, Tina: (1987) *Fuzzy Sets, Uncertainty, and Information*, Prentice Hall

[15] Klir, George and Parviz, Behvad: (1992) "A Note on the Measure of Discord", in: *Proc. 8th Conf. on Uncertainty in AI*, Stanford

[16] Klir, George and Parviz, Behvad: (1992) "Possibility-Probability Conversions: An Empirical Study", in: *Progress in Cybernetics and Systems*, ed. R. Trappl, pp. 19-26, Hemisphere, New York

[17] Wang, PZ and Liu, XH: (1984) "Set-Valued Statistics", *J. Eng. Math.*, v. **1**:1, pp. 43-54