3 DISTRIBUTIONAL REPRESENTATIONS OF RANDOM INTERVAL MEASUREMENTS

Cliff Joslyn

Abstract:
The measurement of set-valued and interval data is considered. Their random interval mathematical representations are introduced. The conditions for distributional (probabilistic and possibilistic) representations and conversions to and among them are established. Some properties of empirical random sets and possibilistic histograms related to strong probabilistic compatibility are described. Finally, the nature of probability distributions which are strongly stochastically compatible with a given possibility distribution, and the derivation of frequency distributions from empirical random sets, are discussed.

3.1 INTRODUCTION

Possibility theory (de Cooman *et al.*, 1995) is an alternative information theory to that based on probability. Although possibility theory is logically independent of probability theory, they are related: both arise in Dempster-Shafer evidence theory as fuzzy measures defined on random sets; and their distributions are both fuzzy sets. So possibility theory is a component of a broader General Information Theory (GIT), which includes all of these fields (Klir 1993).

Zadeh's concept of probabilistic-possibilistic compatibility (or consistency (Zadeh 1978)) is an example of the kind of principle which can be brought to bear on the problem of deriving a coherent, synthetic GIT. In order to accommodate the desired properties of possibilistic semantics, Joslyn (1993a) has extended this idea to a principle of strong compatibility.

Another example of a synthetic principle is the use of random sets—originally developed as a branch of stochastic geometry (Kendall 1974)—to provide a broad, unifying context within which to develop GIT (Goodman 1994). Some properties of empirical random sets and possibilistic histograms are described related to strong probabilistic compatibility. After introducing the fundamentals of possibilistic mathematics and measurement, we will discuss possibilistic histograms and the possibility of occurence, the nature of probability distributions which are strongly stochastically compatible with a given possibility distribution, and the derivation of frequency distributions from empirical random sets.

3.2 MATHEMATICAL PRELIMINARIES

Assume a finite universe of discourse $\Omega := \{\omega_i\}, 1 \le i \le n := |\Omega|$, and let $\mathcal{D} := \{[a,b) \subseteq \mathbb{R} : a, b \in \mathbb{R}, a < b\}$ be the class of half-open intervals.

3.2.1 Random Sets, Intervals, and Evidence Measures

Given a probability space $\langle X, \Sigma, \Pr \rangle$, then a function $S: X \mapsto 2^\Omega - \{\emptyset\}$, where $-$ is set subtraction, is a random subset of Ω if S is Pr-measurable, so that $\forall A \subseteq \Omega, S^{-1}(A) \in \Sigma$. Thus a general random set S associates a probability $(\Pr \circ S^{-1})(A)$ to each $A \subseteq \Omega$.

When Ω is finite, then following Dubois and Prade (1990), let $m: 2^\Omega \mapsto [0,1]$ be an evidence function if $m(\emptyset) = 0$ and $\sum_{A \subseteq \Omega} m(A) = 1$. Then $\mathcal{S} := \{\langle A_j, m_j \rangle\}$ is a random set where $1 \le j \le N := |\mathcal{S}| \le 2^n$, $m(A_j) > 0$, and $m_j := m(A_j)$.

The focal set of a random set is $\mathcal{F}(\mathcal{S}) := \{A_j\}$ with core $\mathbf{C}(\mathcal{F}(\mathcal{S})) := \bigcap_{A_j \in \mathcal{F}(\mathcal{S})} A_j$ and support $\mathbf{U}(\mathcal{F}(\mathcal{S})) := \bigcup_{A_j \in \mathcal{F}(\mathcal{S})} A_j$. $\mathcal{S}$ is called consistent when $\mathbf{C}(\mathcal{F}(\mathcal{S})) \ne \emptyset$. Let $\Gamma(\mathcal{S}) := \{G_k\} \subseteq 2^\Omega - \emptyset, 1 \le k \le Q$, be the partition of $\mathbf{U}(\mathcal{F}(\mathcal{S}))$ induced by all the mutual set intersections and subtractions of the A_j (in other words, an equivalence class of Ω representing all the atoms of the set union/intersection lattice induced by the focal sets).

A random interval $\mathcal{A}$ is a random set on $\Omega = \mathbb{R}$ with $\mathcal{F}(\mathcal{A}) \subseteq \mathcal{D}$. Denote the focal sets of $\mathcal{A}$ as $A_j = [l_j, r_j) \subseteq \mathbb{R}, l_j < r_j$. From (Joslyn 1997a) we have:

Proposition 1 Given a random interval $\mathcal{A}$, then $N + 1 \le Q \le 2N$.

Random sets are mathematically isomorphic to Dempster-Shafer bodies of evidence (Guan and Bell 1992). The plausibility and belief on $\forall A \subseteq \Omega$ are

$$\mathrm{Pl}(A) := \sum_{A_j \not\perp A} m_j, \qquad \mathrm{Bel}(A) := \sum_{A_j \subseteq A} m_j, \tag{2}$$

where $A \perp B$ denotes $A \cap B = \emptyset$. Pl and Bel are generally non-additive fuzzy measures (Wang and Klir 1992), and are dual, in that $\forall A \subseteq \Omega, \mathrm{Bel}(A) = 1 - \mathrm{Pl}(\overline{A})$. In general only plausibility will be considered below.

Define the plausibilistic trace (one-point coverage function) of $\mathcal{S}$ as $\rho: \Omega \mapsto [0,1]$ where $\rho(\omega_i) := \mathrm{Pl}(\{\omega_i\})$, and in vector form $\vec{\rho} = \langle \rho_i \rangle$, with $\rho_i := \rho(\omega_i) = \sum_{A_j \ni \omega_i} m_j$. For random intervals, denote $\rho_{\mathcal{A}}: \mathbb{R} \mapsto [0,1]$ so that $\forall x \in \mathbb{R}, \rho_{\mathcal{A}}(x) := \mathrm{Pl}(\{x\}) = \sum_{A_j \ni x} m_j$.

3.2.2 Probability and Possibility

When $\mathcal{F}(\mathcal{S})$ is specific, so that $\forall |A_j| = 1$, then $\mathrm{Pr} := \mathrm{Pl}$ is an additive probability measure with

$$\forall A, B \subseteq \Omega, \qquad \mathrm{Pr}(A \cup B) = \mathrm{Pr}(A) + \mathrm{Pr}(B) - \mathrm{Pr}(A \cap B), \tag{3}$$

and $\vec{p} = \langle p_i \rangle := \vec{\rho}$ is a probability distribution with additive normalization $\sum_i p_i = 1$ and operator $\mathrm{Pr}(A) = \sum_{\omega_i \in A} p_i$. Statistical entropy $\mathbf{H}(\vec{p}) := -\sum_{i=1}^{n} p_i \log_2(p_i)$ is then the canonical measure of information in probability theory (Klir and Yuan 1995), and the probability distribution with maximal entropy is the maximally uninformative probability distribution denoted $\vec{p}^*$ which results when $\forall i, p_i^* = 1/n$ (Joslyn 1994). Given a random set $\mathcal{S}$, then the Maximum Entropy Principle (MEP) (Klir 1993) has been applied (Dubois and Prade 1982) to derive a canonical probability distribution $p^{\mathcal{S}}$ approximating $\mathcal{S}$, replacing each subset evidence value $m(A_j)$ with the MEP uniform probability distribution over its members, so that

$$\forall \omega \in \Omega, \qquad p^{\mathcal{S}}(\omega) := \sum_{A_j \ni \omega} \frac{m_j}{|A_j|}. \tag{4}$$

When $\mathcal{S}$ is consonant ($\mathcal{F}(\mathcal{S})$ is a nest), so that (without loss of generality for ordering, and letting $A_0 := \emptyset$) $A_{j-1} \subseteq A_j$, then $\Pi := \mathrm{Pl}$ is a possibility measure and $\eta := \mathrm{Bel}$ a necessity measure,[1] with $\forall A, B \subseteq \Omega, \Pi(A \cup B) = \Pi(A) \vee \Pi(B)$, where $\vee$ is the maximum operator. Now $\vec{\pi} = \langle \pi_i \rangle := \vec{\mathrm{Pl}}$ is a possibility distribution with maximal normalization and operator, respectively

$$\bigvee_i \pi_i = 1, \tag{5}$$

$$\Pi(A) = \bigvee_{\omega_i \in A} \pi_i. \tag{6}$$

In possibility theory entropy is replaced by nonspecificity (Klir and Yuan 1995)

$$\mathbf{N}(\vec{\pi}) := \sum_{i=1}^{n} (\pi_i - \pi_{i+1}) \log_2(i) = \sum_{i=2}^{n} \pi_i \log_2 \left(\frac{i}{i-1} \right), \tag{7}$$

where $\pi_{n+1} := 0$ by convention. The possibility distribution with maximal nonspecificity is the maximally uninformative possibility distribution (Klir 1993) denoted $\vec{\pi}^*$, resulting when $\forall i, \pi_i^* = 1$ (Joslyn 1994).

The condition (5) for $\vec{\mathrm{Pl}}$ to be a possibility distribution is actually achieved whenever $\mathcal{S}$ is even consistent (which is implied when $\mathcal{S}$ is consonant). When $\mathcal{S}$ is consistent but not consonant, then even though $\vec{\mathrm{Pl}}$ is a possibility distribution by (5), Pl is not a possibility measure Π. Then there is a unique possibilistic approximation Π^* to Pl achieved by invoking (6) on π (Joslyn 1994).

If $\mathcal{S}$ is not even consistent, then there are well-justified possibilistic normalization methods (Joslyn 1993c) such as consistent transformations (Joslyn 1993c, 1997b), which select certain elements or regions of π to be "elevated" to be in a core.

The following result will be useful below:

Proposition 8 If $\mathcal{S}$ is consistent, then $m(A) > 0 \rightarrow \mathrm{Pl}(A) = 1$.

3.2.3 Possibilistic-Probabilistic Compatibility

Measures of compatibility[2] between a probability and a possibility distribution are available (Delgado and Moral, 1987). The best known of these is Zadeh's (1978) measure $\gamma(\vec{p}, \vec{\pi}) := \vec{p} \cdot \vec{\pi} = \sum_{i=1}^{n} p_i \pi_i$, where $\gamma(\vec{p}, \vec{\pi}) = 1$ indicates maximal compatibility and $\gamma(\vec{p}, \vec{\pi}) = \bigwedge_i \pi_i$ minimal compatibility, and $\wedge$ is the minimum operator.

Many methods are available to convert a given probability distribution to a possibility distribution, and vice versa (Klir and Yuan 1995, Sudkamp 1992). One of the most prominent is the maximum normalization or ratio scale method (Klir 1990). Given a probability distribution p, then let $\pi^m \colon \Omega \mapsto [0,1]$ be a possibility distribution where $\pi^m(\omega_i) = \pi_i^m := p_i / \bigvee_{i=1}^{n} p_i$. We have the following results (Joslyn 1994):

Proposition 9 $f_i = \pi_i^m / \sum_{i=1}^{n} \pi_i^m$.

Proposition 10 If $\gamma(\vec{p}, \vec{\pi}^m) = 1$ then $\vec{p} = \vec{p}^*$ and $\vec{\pi}^m = \vec{\pi}^*$.

Joslyn (1994) has considered the semantics of possibility theory from a number of different perspectives, including graduated, physical, and modal conceptual frameworks. In particular, he has considered what an appropriate relation between probabilistic and possibilistic representations of the same problem domain would be (1993a), and has asserted the following strong principle.

Principle 11 (Strong Probability-Possibility Compatibility (PPC)**)** A given probability Pr and possibility measure Π are strongly compatbile iff $\forall A \subseteq \Omega, \Pr(A) > 0 \leftrightarrow \Pi(A)$.

It follows that $\Pr(A) = 0 \leftrightarrow \Pi(A) < 1$, $\eta(A) > 0 \rightarrow \Pr(A) > 0$, and $\eta(A) = 0 \rightarrow \Pr(A) = 0$. At the distribution level it follows that $\forall \omega \in \Omega, p(\omega) > 0 \leftrightarrow \pi(\omega) = 1$ and $p(\omega) = 0 \leftrightarrow \pi(\omega) < 1$. Finally, if the distributions $\vec{p}$ and $\vec{\pi}$ are strongly compatibile, then $\gamma(\vec{p}, \vec{\pi}) = 1$.

Note that the PPC (11) is not a definition or a theorem, but is rather a *principle* asserted as a *semantic* criterion, and is thus necessarily extra-theoretical. Detailed arguments justifying this position are offered elsewhere (Joslyn 1993a, 1994). Suffice it here to say that the PPC states that something having non-zero probability is *likely*, and therefore given sufficient time *eventual*, and therefore equivalent to its being *completely possible*. Conversely, a properly possible event $(0 < \Pi(A) < 1)$ must be of probability measure zero, and probability zero may or may not indicate proper possibility.

3.3 POSSIBILISTIC MEASUREMENT FROM CONSISTENT RANDOM INTERVALS

Measurement methods for possibility distributions based on the observation of random intervals have been developed (Joslyn 1992, 1993b, 1994, 1997a, 1997b). These stand in sharp constrast to traditional measurement of point values or disjoint subset observations.

3.3.1 Probabilistic Measurement

Assume a counting function $c: \Omega \mapsto \mathcal{W}$ such that $c_i := c(\omega_i)$ is the count of the occurrences of ω_i in a statistical record. Then a frequency distribution is a function $f: \Omega \mapsto [0,1]$ where $f(\omega_i) = f_i := c_i / \sum_i c_i$. Denote the vector $\vec{f} := \langle f_i \rangle$. The function $P: 2^\Omega \mapsto [0,1]$ is a frequency measure where $\forall A \subseteq \Omega, P(A) := \sum_{\omega_i \in A} f_i$. Now $\vec{f}$ is a natural probability distribution with normalization $\sum_i f_i = 1$, and P is a natural probability measure as in (3).

3.3.2 Random Set Measurement and Possibilistic Histograms

Consider now instead observing subsets $B_s \subseteq \Omega, 1 \leq s \leq M$ denoted as a vector $\vec{B} := \langle B_s \rangle$. The set of observed subsets produced by eliminating any duplicates in $\vec{B}$ is an empirical focal set $\mathcal{F}^E := \{A_j\}$, where $N \leq M$ and

$\forall A_j \in \mathcal{F}^E, \exists B_s \in \vec{B}, B_s = A_j$, and inclusion of an element in a vector is defined as appropriate.

Denoting the number of times that a given set $B_s = A_j$ occurs in $\vec{B}$ as $C(A_j)$, then the set-frequency function is

$$m^E : \mathcal{F}^E \mapsto [0,1], \qquad m^E(A_j) = \frac{C_j}{\sum_{A_j \in \mathcal{F}^E} C_j} = \frac{C_j}{M}, \tag{12}$$

where $C_j := C(A_j)$. m^E is clearly an evidence function, which in turn generates an empirical random set denoted $\mathcal{S}^E$. If $\mathcal{F}^E$ is a disjoint class, then $\mathcal{S}^E$ generates a probability distribution on $\Gamma(\mathcal{S})$. But if $\mathcal{S}^E$ is consistent, then the empirical possibility distribution is

$$\pi(\omega) = \sum_{A_j \ni \omega} m_j^E = \frac{\sum_{A_j \ni \omega} C_j}{M}. \tag{13}$$

When $\mathcal{F}^E \subseteq \mathcal{D}$, then this method produces an empirical random interval $\mathcal{A}^E$. When $\mathcal{A}^E$ is consistent, then $\pi_{\mathcal{A}} := \rho_{\mathcal{A}}$ derived according to (13) can be properly described as a possibilistic histogram. These are similar to ordinary (stochastic) histograms, but generated from possibly overlapping interval observations, and thus governed by the mathematics of random sets. In particular, each possibilistic histogram is a fuzzy interval (Joslyn 1997a). In the sequel it will be assumed that $\mathcal{A}^E$ is consistent, and thus $\pi_{\mathcal{A}}$ from (13) is a possibility distribution.

An example is shown in Fig. 3.1. On the left top, four observed intervals are shown. The middle two occur with frequency 1/2, while each of the top and bottom have frequency 1/4. Together they determine $\mathcal{A}^E$. The step function on the right is the possibilistic histogram π derived from (13). It can be briefly stated in vector form as $\vec{\pi}_{\mathcal{A}} = \langle 1/4, 1, 1/2, 1/4 \rangle$, where the values are taken on each of the piecewise constant segments

$$\Gamma(\mathcal{S}) = \langle [1, 1.5), [1.5, 2], (2, 3.5], (3.5, 4] \rangle, \tag{14}$$

of the step function, as shown in the figure.

3.3.3 Realization

We will next consider a simple thought experiment in depth in order to motivate random interval observations.

Consider Fig. 3.2. There is a length of screen Ω against which an experimenter throws Z balls, each striking somewhere between points a and b. A wall rests in front of the screen, and each of two observers can see a different portion

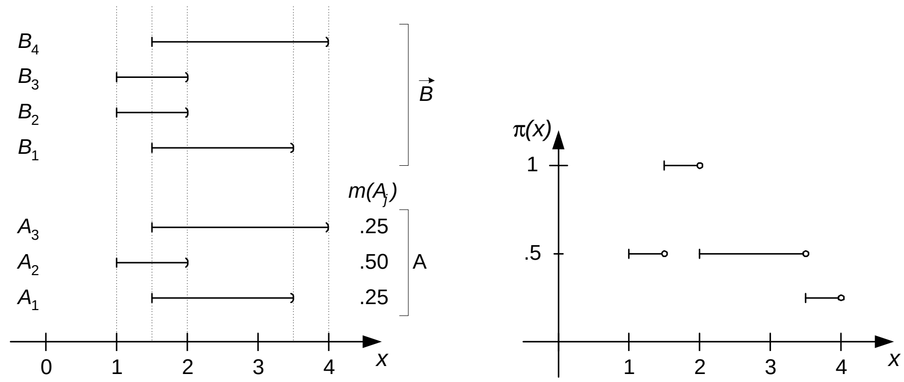

Figure 3.1 (Left) Observed focal elements of a random interval. (Right) Possibilistic histogram π^E.

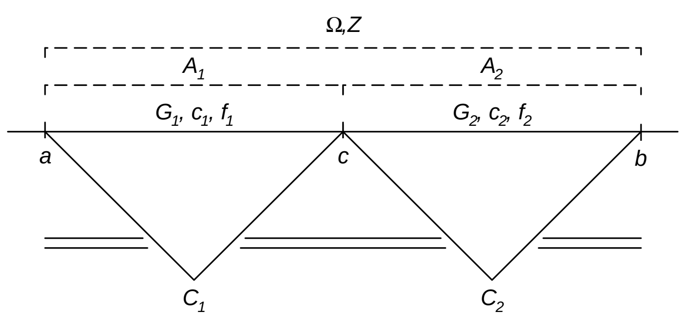

Figure 3.2 Observations resulting in frequency distributions.

of the screen through her own hole in the wall. Observer O_1 sees $A_1 := [a, c)$ and observer O_2 sees $A_2 := [c, b]$. Each observer then reports the count of the number of balls C_1, C_2 she sees through her hole. This is then predicated to be the same as the actual number of balls c_1, c_2 striking each section of the screen.

Here we recognize c_i as the counting function from Sec. 3.3.1 over the universe of discourse $\Gamma := \{A_1, A_2\}$, where $C_i := c_i$ and $Z = \sum_i c_i = \sum_i C_i$, so that frequencies $f_i = c_i/Z$ can be derived.

Now consider a modification shown in Fig. 3.3. The holes in the wall have been moved so that O_1 sees $[a, d] = G_1 \cup G_2 = A_1$, O_2 sees $[c, b] = G_2 \cup G_3 = A_2$, and both see $[c, d] = G_2 = A_1 \cap A_2$. As before, each observer reports the count C_1, C_2 she observes, and we again wish to derive the actual number of hits c_k in each of the disjoint regions G_k so that we can then determine frequencies f_k.

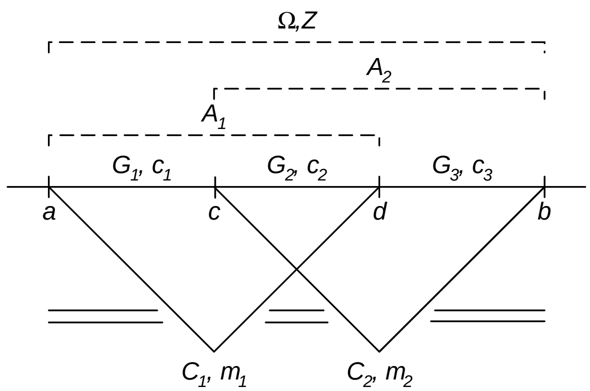

Figure 3.3 Observations resulting in empirical random sets.

We must make a number of assumptions now. First, the records of the hits reported by each observer cannot be *correlated*, rather only a *statistical* description of the collection of observations, in this case the total number of hits seen by each observer, can be reported. If the experimenter knew for any *particular* ball toss which of the observers (or both) reported, then he would be able to disambiguate where the ball struck within a particular $G_k \in \Gamma$.

Second, our knowledge of the outcome comes only from the reports of the observers. In particular, the experimenter has no independent knowledge as to the position where the balls hit. Thus while there may be "real" positions of the ball hits against the wall, all that is known is whether a hit was seen in A_1 or A_2. Furthermore, the total number Z of balls thrown by the experimenter is not known. This we also wish to infer from our knowledge of the C_j.

Under these conditions clearly the c_k are underdetermined. This is a linear problem where

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \vec{c}^T = \begin{pmatrix} Z \\ C_1 \\ C_2 \end{pmatrix}, \tag{15}$$

additionally constrained by $\forall c_i \geq 0$, with solution

$$\begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} Z \\ C_1 \\ C_2 \end{pmatrix} = \vec{c}^T. \tag{16}$$

Knowledge of any three of $c_1, c_2, c_3, Z, C_1, C_2$ determines the other three, but all we have are C_1, C_2.

In particular, we wish to derive knowledge of Z in order to determine the c_k. While C_1, C_2 do not determine Z, they do bound it, and bounds are also available on all the c_k and f_k. These bounds are determined as follows. First,

$$c_1 = Z - C_2 \geq 0 \quad \rightarrow \quad Z \geq C_2, \qquad c_3 = Z - C_1 \geq 0 \quad \rightarrow \quad Z \geq C_1 \tag{17}$$

$$c_2 = C_1 + C_2 - Z \geq 0 \quad \rightarrow \quad Z \leq C_1 + C_2, \tag{18}$$

so $Z \in [C_1 \vee C_2, C_1 + C_2]$. Next, $c_2 = C_1 + C_2 - Z$, so

$$c_2 \in [(C_1 + C_2) - C_1 \vee C_2, (C_1 + C_2) - (C_1 + C_2)] = [C_1 \wedge C_2, 0]. \tag{19}$$

Then $c_1 = Z - C_2$ and $c_2 = Z - C_1$, so

$$c_1 \in [(C_1 \vee C_2) - C_2, (C_1 + C_2) - C_2], \qquad c_3 \in [(C_1 \vee C_2) - C_1, (C_1 + C_2) - C_1]. \tag{20}$$

Assume without loss of generality that $C_1 \leq C_2$. Then it follows that

$$c_1 \in [0, C_1 \wedge C_2], \qquad c_3 \in [C_2 - C_1, C_1 \vee C_2], \tag{21}$$

which generalizes to

$$c_1 \wedge c_3 \in [0, C_1 \wedge C_2], \qquad c_1 \vee c_3 \in [|C_2 - C_1|, C_1 \vee C_2]. \tag{22}$$

For frequencies we have

$$f_2 \in \left[\frac{C_1 \wedge C_2}{C_1 \vee C_2}, 0\right], \quad f_1 \wedge f_3 \in \left[0, \frac{C_1 \wedge C_2}{C_1 + C_2}\right], \quad f_1 \vee f_3 \in \left[\frac{|C_1 - C_2|}{C_1 \vee C_2}, \frac{C_1 \vee C_2}{C_1 + C_2}\right], \tag{23}$$

and finally it is interesting to note that

$$f_1+f_2 \in \left[\frac{C_1 \wedge C_2}{C_1 \vee C_2}, \frac{C_1 \wedge C_2}{C_1 + C_2}\right], \; f_1+f_3 \in \left[\frac{|C_1 - C_2|}{C_1 \vee C_2}, 1\right], \; f_2+f_3 \in \left[1, \frac{C_1 \vee C_2}{C_1 + C_2}\right]. \tag{24}$$

As a numerical example, let $C_1 = 10, C_2 = 15$. Then (abusing interval notation for vectors slightly)

$$Z \in [15, 25], \qquad \vec{c} \in [\langle 0, 10, 5\rangle, \langle 10, 0, 15\rangle], \qquad \vec{f} \in [\langle 0, 2/3, 1/3\rangle, \langle 2/5, 0, 3/5\rangle]. \tag{25}$$

In each of these cases, the range of Z (or the others), represents the amount of uncertainty present in the overall measurement, and in particular the extent to which we are ignorant of the number of balls which struck in the intersection G_2. Consider the limits conditions on Z. If $Z = C_1 \vee C_2$ then from (18) $c_2 = C_1 \wedge C_2$, so that no balls struck in the G_k not included in the A_i for which $C_i = C_1 \vee C_2$. If $Z = C_1 + C_2$ then $c_2 = 0$, so that no balls struck in the intersection. Finally, because $C_1, C_2 \geq 0$, if $C_1 + C_2 = C_1 \vee C_2$ then $C_1 = 0$ or $C_2 = 0$, so that all the balls struck in the G_k not included in the A_i for which $C_i = 0$.

Of course, we recognize this as a special case of the measurement of a random interval, with C_j the set-counting function for the region A_j, where $M = \sum_{j=1}^N C_j$. Since Dempster's early work (1967) it has been known that no additive probability distribution p is generally available, but rather a class of measures Pr are identified which are bounded by [Bel, Pl].

The random set interpretation resolves the uncertainty present in the intervals by representing the measurement as a single entity in a more abstract space. Thus no matter the value of Z, the constraints $C_1 = c_1 + c_2, C_2 = c_2 + c_3$ are always present. Or in other words, consider the interval expression of the quantity $q := C_1 + C_2 = (c_1 + c_2) + (c_2 + c_3)$, generalized as $q = c_1 \wedge c_3 + 2c_2 + c_1 \vee c_3$. Recalling that $a, b \geq 0 \to a \wedge b + |a - b| = a + b$, we have

$$\begin{aligned} q &\in [0, C_1 \wedge C_2] + 2[C_1 \wedge C_2, 0] + [|C_1 - C_2|, C_1 \vee C_2] && (26) \\ &= [2(C_1 \vee C_2) + |C_1 - C_2|, C_1 \wedge C_2 + C_1 \vee C_2] && (27) \\ &= [C_1 + C_2, C_1 + C_2] = C_1 + C_2 && (28) \end{aligned}$$

uniquely. Furthermore, in the general case $\vec{c}$ is substantially more underdetermined, in that the matrix in (15) is $N \times Q$, and thus from (1) is rarely even square, let alone invertible.

So in general, if a measured random set is consistent or specific, a direct distributional representation is available in possibilistic or probabilistic forms respectively. But otherwise, if a distributional representation is required, then the direct approach will be infeasible. Instead we must look for other methods which transform the given data into a representational form.

3.4 STRONGLY COMPATIBLE PROBABILITY DISTRIBUTIONS

From the natural language perspective, the primary semantic criterion for possibility is that the occurrence of an event requires maximal (unitary) possibility. In a possibilistic histogram the occurring events are exactly those $B_s \in \vec{B}$ which have been observed. So this condition is easily met by possibilistic histograms.

Corollary 29 If $\mathcal{F}^E$ is consistent, then $\forall B_s \in \vec{B}, \Pi(B_s) = 1$.

Proof: Fix B_s. Then $C(B_s) \geq 1$, so $m(B_s) \geq 1/M > 0$. The result follows from the corrolary (8) and the consistency of $\mathcal{F}^E$. ■

Probability distributions which conform to the PPC with a possibilistic histogram should also be considered. Under the PPC (11), it is necessary that $p(\omega) = 0$ wherever $\pi(\omega) < 1$, that is $\forall \omega \notin \mathbf{C}(\mathcal{F}^E)$. In the example in Fig. 3.1, that would yield $p > 0$ only on the interval $\mathbf{C}(\pi) = [1.5, 2)$. No further information would be provided by π, and so the MEP would yield the uniform probability density

$$p^*(\omega) = \begin{cases} 2, & \omega \in [1.5, 2) \\ 0, & \text{elsewhere} \end{cases} . \tag{30}$$

This result makes complete sense in the context of the nature of subset measurements. Given a consistent set of observed intervals, if they are all to be believed then all that can be said is that the event actually happened *somewhere* in the core. There the possibility is unitary, and by the PPC the probability is positive. But there is no further information about the *likelihood* of the event being anywhere *particular* inside the core, thus requiring the maximally uninformative probability distribution p^*.

The fact that $\forall \omega \in \mathbf{U}(\pi), \omega \notin \mathbf{C}(\pi), 0 < \pi < 1$ indicates that it is *somewhat* possible for *another* observation, perhaps at another time, to be found somewhere between the core and the edge of the support, but not *completely* possible, since nothing can be said to have been actually observed there yet. Thus the subset measurements give *no* likelihood information about the occurrence of an ω in this region, and by the PPC $p = 0$ there.

If $\mathcal{S}^E$ is inconsistent, and thus a consistent approximation must be made, then for a focus $\omega_0 \in \Omega, \mathbf{C}(\mathcal{S}^E) = \{\omega_0\}$, and so p will be a Dirac-delta function at ω_0.

3.5 FREQUENCY DISTRIBUTIONS FROM EMPIRICAL RANDOM SETS

It is interesting to consider how a purely "probabilistic" treatment would approach set-statistics. In order to simplify the problem, return to the case of two overlapping observations on a discrete universe, letting $\Omega = \{a, b, c\}$ and

$\mathcal{F}(\mathcal{S}) = \{\{a,b\},\{b,c\}\}$ with set-frequencies m_1, m_2 so that $m_1 + m_2 = 1$. Under the assumption that Pr should have an additive probability distribution $p: \Omega \mapsto [0,1]$, then

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} p(a) \\ p(b) \\ p(c) \end{pmatrix} = \begin{pmatrix} 1 \\ m_1 \\ m_2 \end{pmatrix}. \qquad (31)$$

which has the solution $p(a) = m_1, p(c) = m_2, p(b) = 0$. This is entirely unsatisfactory, and maximally incompatible with the possibilistic results above: it *eliminates* probability exactly on b, the point where there is the *most* evidence, and where in the possibilistic histogram $\pi(b) = 1$.

As above with the observation of the failure of the direct approach for distributional representations, in general the matrix in (31) is $N + 1 \times n$, and thus rarely are there *any* feasible solutions for non-negative probabilities. Other methods which transform the set-based data must be used, for example translating the counts on subsets into counts on elements, thus establishing a mapping $C \mapsto c$. There are a number of ways in which that could be done.

3.5.1 *Duplicated Counts*

We could say that a nonspecific observation is really an observation of *every* element of the subset. Then each observation of a subset B_s would contribute one element count for every $\omega \in B_s$. Then the overall element count is

$$\forall \omega \in \Omega, \quad c(\omega) = \sum_{A_j \ni \omega} C_j. \qquad (32)$$

Corollary 33

$$f(\omega) = \frac{c(\omega)}{\sum_{A_j \in \mathcal{F}^E} C_j |A_j|}. \qquad (34)$$

Proof:

$$f(\omega) = \frac{c(\omega)}{\sum_{\omega \in \Omega} c(\omega)} = \frac{c(\omega)}{\sum_{\omega \in \Omega} \sum_{A_j \ni \omega} C_j} = \frac{c(\omega)}{\sum_{A_j \in \mathcal{F}^E} C_j |A_j|}. \qquad (35)$$

■

By this method, the example in Fig. 3.1 yields the frequency distribution

$$\vec{f} = \langle 2/9, 4/9, 2/9, 1/9 \rangle. \qquad (36)$$

similarly valued over the piecewise constant segments in (14). Note that this is identical to $\vec{\pi}$ for elements having the same numerator, but the denominator changed from 4 (which is $\sum C_j$) to 9 (which is $\sum c(\omega) = \sum C_j |A_j|$). In fact, the effect of this count duplication method is to establish a maximum normalized ratio scale between π and f.

Theorem 37 Given a consistent $\mathcal{F}^E$ with a frequency distribution f determined by (32), then $\forall \omega \in \Omega$,

$$f(\omega) = \frac{\pi(\omega)}{\sum \pi(\omega)}, \qquad \pi(\omega) = \frac{f(\omega)}{\bigvee f(\omega)} \tag{38}$$

Proof: From the possibilistic histogram formula (13) and (32), $\forall \omega \in \Omega$, $M\pi(\omega) = \sum_{A_j \ni \omega} C_j = c(\omega)$. Therefore from the corollary (33),

$$f(\omega) = \frac{c(\omega)}{\sum_{\omega \in \Omega} c(\omega)} = \frac{M\pi(\omega)}{\sum_{\omega \in \Omega} M\pi(\omega)} = \frac{\pi(\omega)}{\sum_{\omega \in \Omega} \pi(\omega)}. \tag{39}$$

The second result follows from the ratio scale frequency conversion (9). ■

Thus the disadvantages of duplicating counts like this are clear. First, frequency additivity is violated because

$$\sum_{\omega \in A_j} c(\omega) = \sum_{\omega \in A_j} \sum_{A_k \ni w} C_k \geq C_j. \tag{40}$$

Also, the PPC is generally violated in virtue of the ratio scale frequency conversion, as shown in Prop. (10).

3.5.2 Distributed Counts

Instead of a subset count contributing multiple element counts, the single subset count can be additively distributed amongst the $\omega \in A$. Since there is no further information about how to distribute the count, then by the MEP a uniform distribution should be used. Then the element count for each $\omega \in \Omega$ is

$$\forall \omega \in \Omega, \quad c(\omega) = \sum_{A_j \ni \omega} \frac{C_j}{|A_j|}. \tag{41}$$

Corollary 42 $f(\omega) = c(\omega)/M$.

Proof: Because

$$\sum_{\omega \in \Omega} c(\omega) = \sum_{\omega \in \Omega} \sum_{A_j \ni \omega} \frac{C_j}{|A_j|} = \sum_{A_j \in \mathcal{F}^E} \frac{C_j |A_j|}{|A_j|} = \sum_{A_j \in \mathcal{F}^E} C_j = M, \tag{43}$$

therefore

$$f(\omega) = \frac{c(\omega)}{\sum_{\omega \in \Omega} c(\omega)} = \frac{c(\omega)}{M}. \tag{44}$$

■

By this method, the example in Fig. 3.1 yields a frequency distribution

$$\vec{f} = \langle 1/4, 11/24, 5/24, 1/12 \rangle \,. \tag{45}$$

Not surprisingly, this method is closely related to the applications of the MEP as discussed above.

Theorem 46 Assume an empirical random set $\mathcal{S}^E$ and let f be a frequency distribution determined by (41). Then f is the maximum entropy probability distribution $p^{\mathcal{S}^E}$ from (4).

Proof: From (41), (42), the set-frequency definition (12), and the maximum entropy probability distribution formula (4), then $\forall \omega \in \Omega$,

$$f(\omega) = \frac{c(\omega)}{M} = \sum_{A_j \ni \omega} \frac{C_j}{|A_j| M} = \sum_{A_j \ni \omega} \frac{m_j^E}{|A_j|} = p^{\mathcal{S}^E}(\omega). \tag{47}$$

■

Notes

1. Since results for necessity are dual to those of possibility, only possibility will be discussed in the sequel.
2. The term used in the literature is actually "consistency", so to avoid confusion with random set consistency, we will use "compatibility".

References

de Cooman, G., Ruan, D., and Kerre, E.E. eds., *Foundations and Applications of Possibility Theory*, World Scientific, Singapore, 1995

Dempster, A. P., Upper and Lower Probabilities Induced by a Multivalued Mapping, *Annals of Mathematical Statistics*, 38, 1967, 325-339.

Delgado, M., and Moral, S., On the Concept of Possibility-Probability Consistency, *Fuzzy Sets and Systems*, 21, 1987, 311-318.

Dubois, D., and Prade, H. On Several Representations of an Uncertain Body of Evidence, *Fuzzy Information and Decision Processes*, 1982, pp. 167-181.

Dubois, D., and Prade, H., Consonant Approximations of Belief Functions, *Int. J. Approximate Reasoning*, 4, 1990, 419-449.

Goodman, I. R., New Characterization of Fuzzy Logic Operators Producing Homomorphic-Like Relations with One-Point Coverages of Random Sets, *Advances in Fuzzy Theory and Technology*, ed. P.P. Wang, 133-159, Bookwrights Press, Raleigh NC, 1994.

Guan, J., and Bell, D.A., *Evidence Theory and its Applications*, North Holland, 1992.

Joslyn, C., Possibilistic Measurement and Set Statistics, *Proc. NAFIPS 1992*, v. 2, Puerto Vallerta, 458-467, 1992.

Joslyn, C., Possibilistic Semantics and Measurement Methods in Complex Systems, *Proc. 2nd Int. Symposium on Uncertainty Modeling and Analysis*, ed. B. Ayyub, IEEE Computer Society Press, College Park, MD, 208-215, 1993a.

Joslyn, C., Some New Results on Possibilistic Measurement, *Proc. NAFIPS 1993*, Allentown PA, 227-231, 1993b.

Joslyn, C., Empirical Possibility and Minimal Information Distortion, *Fuzzy Logic: State of the Art*, ed. R. Lowen and M. Roubens, Kluwer, Dordrecht, 143-152, 1993c.

Joslyn, C., *Possibilistic Processes for Complex Systems Modeling*, PhD dissertation, SUNY at Binghamton, available from UMI Disseration Services, Ann Arbor MI, 1994.

Joslyn, C., Measurement of Possibilistic Histograms from Interval Data, *Int. J. General Systems*, 26:(2-3), in press, 1997a.

Joslyn, C., Possibilistic Normalization of Inconsistent Random Intervals, *Advances in Systems Science and Applications*, in press, 1997b.

Kendall, D.G., Foundations of a Theory of Random Sets, *Stochastic Geometry*, ed. E.F. Harding and D.G. Kendall, Wiley, New York, 322-376, 1974.

Klir, G., Principle of Uncertainty and Information Invariance, *Int. J. of General Systems*, 17(2-3), 249-275, 1990.

Klir, G., Developments in Uncertainty Based Information, *Advances in Computers*, 36, ed. M. Yovitz, Academic Press, 255-332, 1993.

Klir, G., and Yuan, B., *Fuzzy Sets and Fuzzy Logic*, Prentice-Hall, New York, 1995.

Sudkamp, T., On Probability-Possibility Transformations, *Fuzzy Sets and Systems*, 51, 73-81, 1992.

Wang, Z., and Klir, G., *Fuzzy Measure Theory*, Plenum Press, New York, 1992.

Zadeh, L.A., Fuzzy Sets as the Basis for a Theory of Possibility, *Fuzzy Sets and Systems*, 1, 3-28, 1978.