Joslyn, Cliff A: (1997) "Towards General Information Theoretical Representations of Database Problems", in: Proc. 1997 Conf. of the IEEE Society for Systems, Man and Cybernetics: Computational Cybernetics and Simulation, v. 2, pp. 1662-1667, Orlando, https://doi.org/10.1109/ICSMC.1997.638247

Towards General Information Theoretical Representations of Database Problems

Dr. Cliff Joslyn

Computer and Information Research Group
Los Alamos National Laboratory
MS B265, LANL
Los Alamos, NM 87545
joslyn@lanl.gov, http://www.c3.lanl.gov/~joslyn

June, 1997

Abstract

General database systems are described from the General Systems Theoretical (GST) framework. In this context traditional information theoretical (statistical) and general information theoretical (fuzzy measure and set theoretical, possibilistic, and random set theoretical) representations are derived. A preliminary formal framework is introduced.

1 Introduction

The Computer Research and Applications group (CIC-3) of the Los Alamos National Laboratory is deeply interested in a number of database problems related to "data mining" in general [3], and fraud and anomaly detection in particular. Such problems are typically plagued by a number of distinct challenges, including high complexity and dimensionality of the data, and multiple forms and sources of uncertainty.

Motivated by these problems, this paper lays some initial groundwork for the integration of three distinct formal fields: **database theory** [2] concerns the structures and processes of (usually relational) database information systems; **general systems theory** (GST) [5] is a general formal theory for representing and modeling systems of all kinds; and **general information theory** (GIT) [7] is the general theory for representing uncertainty and information in systems.

After some mathematical preliminaries, we first lay out some formal definitions relating database theory to GST, including a number of GIT structures which fall out naturally. We then introduce bipartate graphical representations of two dimensional projections, and finally introduce random set representations in the context of both complete and incomplete information.

2 Mathematical Preliminaries

Denote $\mathbb{N}_M := \{1, 2, \ldots, M\}$ and $\mathbb{N} := \mathbb{N}_\infty$.

Denote a vector $\vec{x} := \langle x_1, x_2, \ldots, x_M \rangle = \langle x_i \rangle, i \in \mathbb{N}_M$. For a fixed $i \in \mathbb{N}_M$, denote vector-element inclusion as $x \in \vec{x} := \exists i \in \mathbb{N}_M, x = x_i$; vector-element subtraction as $\vec{x} - x_i := \langle x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_M \rangle$ if $x_i \in \vec{x}$, or as just $\vec{x}$ otherwise; and given another vector $\vec{x}' := \langle x'_{i'} \rangle, i' \in \mathbb{N}_{M'}$ define vector-vector subtraction as $\vec{x} - \vec{x}' := (((\vec{x} - x'_1) - x'_2) - \cdots - x'_{M'})$.

We now introduce some basic concepts of GIT [4, 6, 7, 10]. Given a nonempty, finite set $\Omega = \{\omega_i\}, 1 \leq i \leq n$, then a function $\nu: 2^\Omega \mapsto [0,1]$ is a finite fuzzy measure if $\nu(\emptyset) = 0$ and $\forall A, B \subseteq \Omega, A \subseteq B \rightarrow \nu(A) \leq \nu(B)$. The trace of a fuzzy measure ν is a function $q^\nu: \Omega \mapsto [0,1]$ where $\forall \omega_i \in \Omega, q^\nu(\omega_i) := \nu(\{\omega_i\})$. If in addition there exists an operator function $\oplus: [0,1]^2 \mapsto [0,1]$, with $\oplus$ associative and commutative, such that $\forall A \subseteq \Omega, \nu(A) = \bigoplus_{\omega_i \in A} q^\nu(\omega_i)$, then $\oplus$ is said to be a distribution operator of ν and q^ν its distribution.

$f: \Omega \mapsto [0,1]$ is a probability distribution if $\sum_{i=1}^n f(\omega_i) = 1$; and $\pi: \Omega \mapsto [0,1]$ is a possibility distribution if $\bigvee_{i=1}^n \pi(\omega_i) = 1$, where $\vee$ is the maximum operator. f induces a probability measure $F: 2^\Omega \mapsto [0,1]$, where $\forall A \subseteq \Omega, F(A) := \sum_{\omega_i \in A} f(\omega_i)$. The traditional properties of probability measures hold, in particular $\forall A, B \subseteq \Omega, F(A \cup B) = F(A) + F(B) - F(A \cap B)$.

Similarly, π induces a possibility measure $\Pi: 2^\Omega \mapsto [0,1]$, where now $\forall A \subseteq \Omega, \Pi(A) = \bigvee_{\omega_i \in A} \pi(\omega_i)$, and now the "maxitive" possibilistic property $\forall A, B \subseteq \Omega, \Pi(A \cup B) = \Pi(A) \vee \Pi(B)$ hold. F and Π are both fuzzy measures with distribution operators $+$ and $\vee$ and distributions f and π respectively.

Whereas the canonical measure of the information content of a probability distribution is the classical stochastic entropy $\mathbf{H}(f) := -\sum_{i=1}^n f(\omega_i) \log_2(f(\omega_i))$,

0-7803-4053-1/97/$10.00 © 1997 IEEE

in possibility theory the nonspecificity measure

$$\mathbf{N}(\pi) := \sum_{i=2}^{n} \pi(\omega_i) \log_2 \left(\frac{i}{i-1} \right) =$$

$$\sum_{i=1}^{n} (\pi(\omega_i) - \pi(\omega_{i+1})) \log_2(i),$$

where $\pi(\omega_{n+1}) = 0$ by convention, is most well justified. Various multidimensional (joint, marginal, and conditional) information measures are also available.

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 random set S associates a probability $(\Pr \circ S^{-1})(A)$ to each $A \subseteq \Omega$. Since here Ω is finite, then more specifically we can define an evidence function $m: 2^{\Omega} \mapsto [0,1]$, where $\sum_{A \subseteq \Omega} m(A) = 1$, and then define a random set as $\mathcal{S} := \{\langle A_j, m_j \rangle : m_j > 0\}$, where $A_j \subseteq \Omega, m_j := m(A_j)$, and $1 \leq j \leq N := |\mathcal{S}| \leq 2^n - 1$. Denote the focal set as $\mathcal{F}(\mathcal{S}) := \{A_j : m_j > 0\}$, where each A_j is a focal element. Define the core and support of $\mathcal{S}$ respectively as

$$\mathbf{C}(\mathcal{S}) := \bigcap_{A_j \in \mathcal{F}(\mathcal{S})} A_j, \qquad \mathbf{U}(\mathcal{S}) := \bigcup_{A_j \in \mathcal{F}(\mathcal{S})} A_j.$$

Given a random set $\mathcal{S}$, then denote two evidence measures called plausibility and belief

$$\mathrm{Pl}(A) := \sum_{A_j \cap A \neq \emptyset} m_j, \quad \mathrm{Bel}(A) := \sum_{A_j \subseteq A} m_j,$$

respectively for $A \subseteq \Omega$. Let $\rho := q^{\mathrm{Pl}}$ be the trace of $\mathcal{S}$.

When $\mathcal{F}(\mathcal{S})$ is specific, so that $\forall A_j \in \mathcal{F}, |A_j| = 1$, then Pl = Bel is a probability measure with distribution ρ. When $\mathcal{F}(\mathcal{S})$ is consonant, so that $\forall A_j, A_{j'} \in \mathcal{F}$, either $A_j \subseteq A_{j'}$ or $A_j \supseteq A_{j'}$, then Pl is a possibility measure with distribution ρ. Finally, when $\mathcal{F}$ is just consistent, so that $\mathbf{C}(\mathcal{S}) \neq \emptyset$, then while Pl is not a possibility measure, still ρ satisfies the properties of a possibility distribution, and a unique, best approximating possibility measure Π^* is available. Note that consonance implies consistency.

3 Projections and Extensions of Relations

We now recapitulate some of the essential points of database theory [2] from the perspective of GST.

Assume a set of M finite dimensions $X_i := \{x_{j_i}^i\}$ for $1 \leq i \leq M, 1 \leq j_i \leq n_i = |X_i|$. Without ambiguity, denote $x^i \in X_i$. The overall space is therefore $X := \bigtimes_{i=1}^{n} X_i$ with $n := |X| = \prod_{i=1}^{M} n_i$, with individual states $\vec{x} := \langle x_{j_i}^i \rangle \in X$.

Assume a subset of indices $K \subseteq \mathbb{N}_M$, and let $\neg K := \mathbb{N}_M - K$. Where possible without ambiguity, denote K as a simple list of its elements, for example $K = 124 := K = \{1,2,4\}$, or $K = 10,11,12 := K = \{10,11,12\}$.

Define the **projection** of the space X through K as $X \downarrow K = X^K := \bigtimes_{k \in K} X_k$, which is the K variables of X. Denoting $k' \in \neg K$, then $X \downarrow K$ has corresponding projected vectors

$$\vec{x} \downarrow K = \vec{x}^K := \langle x_{j_k}^k \rangle = \vec{x} - \langle x_{k'} \rangle \in X \downarrow K,$$

which are the K variables of the $\vec{x}$.

Note that $|\vec{x}^K| = |K| \leq M$. Also, if $\exists! i' \in \mathbb{N}_M, K = \{i'\}$, then $X \downarrow K$ is just the collection of "singleton" vectors $\left\{ \left\langle x_{j_{i'}}^{i'} \right\rangle \right\}$ for all $x_{j_{i'}}^{i'} \in X_{i'}$. Then just denote $X \downarrow K := X_{i'}$.

More generally, assume two sets of indices $K, K' \subseteq \mathbb{N}_M$. Then the projection of $X \downarrow K$ again through K' is just $X \downarrow (K \cap K')$. Thus we identify $X = X \downarrow \mathbb{N}_M$, so that $X \downarrow K = X \downarrow (\mathbb{N}_M \cap K)$. Of course if $K \cap K' = \emptyset$ then the projection is empty.

Now consider $X \downarrow (K \cup K')$, and define this as the **extension** to K' of $X \downarrow K$, denoted $(X \downarrow K) \uparrow K'$. Essentially this is just adding the K' variables back in to $X \downarrow K$, so that

$$(X \downarrow K) \uparrow K' := X \downarrow (K \cup K') = (X \downarrow K') \uparrow K.$$

The vector

$$(\vec{x} \downarrow K) \uparrow K' := \vec{x} - \langle x_l \rangle, \qquad l \in \neg(K \cup K')$$

is just the sub-vector of $\vec{x}$ containing both the K and the K' variables. Here, if $K' \subseteq K$ then the extension is meaningless.

4 Databases

Define a collection $\mathbf{X} := \langle \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N \rangle = \langle \vec{x}_j \rangle, j \in \mathbb{N}_N$, as a vector of the records $\vec{x}_j \in X$. $\mathbf{X}$ is usually represented tabularly as a matrix with N rows (data items) and M columns (fields, dimensions). Denote the projection of $\mathbf{X}$ through K as $\mathbf{X} \downarrow K := \langle \vec{x}_j \downarrow K \rangle$.

Assume a subset of indices on the observations $Y \subseteq \mathbb{N}_N$. Then denote $\mathbf{Y} := \langle \vec{x}_{l_j} \rangle, \forall l_j \in Y$. Call this "subsetting". Usually Y must be specified according to some condition or values of the $\vec{x}$, for example $Y := \{\vec{x} \in \mathbf{X} : x_1 = x_0^1\}$, for some $x_0^1 \in X_1$. We will shorten this sometimes, for example here by denoting $Y := \{x_1 = x_0^1\}$.

Projecting reduces the number of columns of the database, and subsetting the number of rows of the

database. These can be combined, considering the subsetted database $\mathbf{Y}$ being projected through the K dimensions. Thus denote a **database** proper as a system $\mathcal{D}(Y,K) := \mathbf{Y} \downarrow K$, or just $\mathcal{D}$ without ambiguity.

Clearly $\mathcal{D}$ is a sufficient representation of a real database. Since the X_i are assumed finite, scalar variables will be represented ordinally by some appropriate binning. In the limit, this will reflect the finite precision of the computer representation. Nor are we currently considering relational databases, but rather a simple flat table, perhaps produced by a large join over many component tables.

Note that $\mathcal{D} \downarrow K' = (\mathbf{Y} \downarrow K) \downarrow K' = \mathcal{D}(Y, K \cap K')$. So, as above, if $\exists! i' \in \mathbb{N}_N, K' = \{i'\}$, then we denote $\mathcal{D} \downarrow K' \subseteq X_{i'}$. In other words, projecting $\mathcal{D}$ through a single dimension i' just identifies that portion of $X_{i'}$ which $\mathcal{D}$ touches.

Furthermore, consider that

$$\exists i' \in \mathbb{N}_N, \quad \exists! j_{i'} \in \mathbb{N}_{n_{i'}}, \quad \mathcal{D} \downarrow \{i'\} = \left\{ x^{i'}_{j_{i'}} \right\},$$

or in other words, for a given dimension i' the values of all the records is identical, namely $x^{i'}_{j_{i'}}$. Then $\mathcal{D}$ and $\mathcal{D} \downarrow \neg\{i'\} = \mathcal{D} \downarrow K - \{i'\}$ are equivalent databases, since the value of the $X_{i'}$ dimension is constant across all its records. So in particular, if $Y := \{x_{i'} = x^{i'}_0\}$ for some $x^{i'}_o \in X_{i'}$, then this effectively reduces the dimensionality of $\mathcal{D}$ by one, since then $\mathcal{D}(Y,K) = \mathcal{D}(Y, K - \{i'\})$.

4.1 Information Functions in Databases

Any database $\mathcal{D}(Y,K)$ induces the following:

- A counting function $c_{Y,K}: X \downarrow K \mapsto \mathbb{N}_{|Y|}$, or just c without ambiguity, where $c(\vec{x})$ is the number of times that each record type $\vec{x} \in X \downarrow K$ has been observed in $\mathbf{Y}$. Note that $\sum_{\vec{x} \in X \downarrow K} c(\vec{x}) = |Y|$.
- A joint frequency function $f_{Y,K}: X \downarrow K \mapsto [0,1]$, or just f, where

$$f(\vec{x}) := \frac{c(\vec{x})}{\sum_{\vec{x} \in X \downarrow K} c(\vec{x})} = \frac{c(\vec{x})}{|Y|}.$$

- A joint possibility distribution function $\pi_{Y,K}: X \downarrow K \mapsto [0,1]$, or just π, where

$$\pi(\vec{x}) := \frac{c(\vec{x})}{\bigvee_{\vec{x} \in X \downarrow K} c(\vec{x})}.$$

- A relation $R_{Y,K} \subseteq X \downarrow K$, where $\vec{x} \in R \leftrightarrow c_{Y,K}(\vec{x}) > 0$. The $\vec{x} \in R_{Y,K}$ are just the projected record types which have been observed at least once in the subset $\mathbf{Y}$.
- An incidence function $I: X \mapsto \{0,1\}$, with

$$I(\vec{x}) := \lceil f(\vec{x}) \rceil = \lceil \pi(\vec{x}) \rceil = \begin{cases} 1, & \vec{x} \in R, \\ 0, & \vec{x} \notin R \end{cases}.$$

For $g \in \{c, f, \pi, I\}$, denote $g_{N,K} := g_{\mathbb{N}_N, K}$ and $g_{Y,M} := g_{Y,\mathbb{N}_M}$. Note that $f_{N,M}$ and $\pi_{N,M}$ are just the global, joint frequency and possibility distributions of the $\vec{x} \in \mathbf{X}$.

c and f are generally additive functions, while π and I are maxitive.

Proposition 1 $\forall K \subseteq \mathbb{N}_M$,

$$\sum_{\vec{x}^K \in X^K} c_{N,K}(\vec{x}) = |Y|, \qquad \sum_{\vec{x}^K \in X^K} f_{N,K}(\vec{x}) = 1,$$

$$\bigvee_{\vec{x}^K \in X^K} \pi_{N,K}(\vec{x}) = 1, \qquad \bigvee_{\vec{x}^K \in X^K} I_{N,K}(\vec{x}) = 1.$$

Obviously $I(\vec{x}) = 1 \leftrightarrow c(\vec{x}) > 0 \leftrightarrow f(\vec{x}) > 0 \leftrightarrow \pi(\vec{x}) > 0$. From the frequency point of view, I (and thus R) represents the set of states with positive frequency, or probability, or in other words the set of states which is *possible*, or have been *seen*. Note that c determines f and π; f and $|Y|$ or π and $\bigvee_{\vec{x} \in X \downarrow K} c(\vec{x})$ determines c; and either c, f or π determine I; but I determines neither c nor f.

Let $g \in \{c, f, \pi, I\}$. For a fixed K, the **marginal** over K is defined as the projection of any of these functions through these dimensions. Define $g^K: X \downarrow K \mapsto \text{ran}(g)$, with $g^K(\vec{x}^K) := \sum_{\vec{x} \in X^K} g(\vec{x})$ for $g \in \{f, c\}$, and $g^K(\vec{x}^K) := \bigvee_{\vec{x} \in X^K} g(\vec{x})$ for $g \in \{\pi, I\}$.

Marginalization is preserved across projections for the additive information functions, but not for the maxitive. In particular, $c_{N,K}$ and $f_{N,K}$ are just the K-marginals of c, f.

Proposition 2

$$g \in \{c, f\} \rightarrow g_{Y,K} \downarrow K' = g_{Y,K \cap K'}.$$

$$g \in \{\pi, I\} \rightarrow g_{Y,K} \downarrow K' \neq g_{Y,K \cap K'}.$$

But the properties of both distributional operators are preserved across subsetting.

Proposition 3 Fix $Y \subseteq \mathbb{N}_M$, and let $\eta := \{Y_j\}$ be a partition of Y. Then $\forall K \subseteq \mathbb{N}_N$,

$$g \in \{f, c\} \rightarrow \sum_{\vec{x} \in \mathcal{D}} g(\vec{x}) = \sum_{Y_j \in \eta} \sum_{\vec{x} \in \mathcal{D}(Y_j, K)} g_{Y_j,K}(\vec{x}),$$

$$g \in \{\pi, I\} \rightarrow \bigvee_{\vec{x} \in \mathcal{D}} g(\vec{x}) = \bigvee_{Y_j \in \eta} \bigvee_{\vec{x} \in \mathcal{D}(Y_j, K)} g_{Y_j,K}(\vec{x}).$$

Once we fully attach the entropy and nonspecificity measures, in both their simple and multi-variate forms, to the additive and maxitive functions respectively, then we are more or less equipped to apply many of the standard set of GST and statistical tools for multivariate modeling, for example: mask analysis and reconstructibility analysis [5], in either the probabilistic or possibilistic contexts [1]; K-systems analysis [9]; and hierarchical log-linear modeling [8].

4.2 Example

As an example, let $N = 50$ and $M = 3$, with $X_1 := \{a, b, c\}, X_2 := \{x, y\}$, and $X_3 := \{\alpha, \beta\}$. Fix $Y := \mathbb{N}_N, K := \mathbb{N}_N$. Then let c, f, π, and I be defined by the table:

X_3	X_2	X_1	c	f	π	I
α	x	a	1	0.02	0.05	1
		b	0	0.00	0.00	0
		c	8	0.16	0.40	1
	y	a	4	0.08	0.20	1
		b	3	0.06	0.15	1
		c	0	0.00	0.00	0
β	x	a	8	0.16	0.40	1
		b	0	0.00	0.00	0
		c	20	0.40	1.00	1
	y	a	4	0.08	0.02	1
		b	2	0.04	0.10	1
		c	0	0.00	0.00	0

The projections are then:

X_2	X_1	c^{12}	f^{12}	π^{12}	I^{12}
x	a	9	0.18	0.32	1
	b	0	0.00	0.00	0
	c	28	0.56	1.00	1
y	a	5	0.16	0.29	1
	b	5	0.10	0.18	1
	c	0	0.00	0.00	0
X_3	X_1	c^{13}	f^{13}	π^{13}	I^{13}
α	a	5	0.10	0.25	1
	b	3	0.06	0.15	1
	c	8	0.16	0.40	1
β	a	12	0.24	0.60	1
	b	2	0.04	0.10	1
	c	20	0.40	1.00	1
X_3	X_2	c^{23}	f^{23}	π^{23}	I^{23}
α	x	9	0.18	0.32	1
	y	7	0.14	0.25	1
β	x	28	0.56	1.00	1
	y	6	0.12	0.21	1

X_1	c^1	f^1	π^1	I^1
a	17	0.34	0.61	1
b	5	0.10	0.18	1
c	28	0.56	1.00	1

X_2	c^2	f^2	π^2	I^2
x	37	0.74	1.00	1
y	13	0.48	0.35	1
X_3	c^3	f^3	π^3	I^3
α	16	0.32	0.47	1
β	34	0.68	1.00	1

5 Graphical Representations

When $|K| = 2$, then $\mathcal{D}$ has a simple graphical representation. Assume, without loss of generality, that $K = \{1, 2\}$, then X_1, X_2 can be represented as nodes of the two parts of a bipartate graph $G := \langle X_1, X_2, E \rangle$, where $E := R$ so that $\langle x^1, x^2 \rangle \in E \leftrightarrow I(\langle x^1, x^2 \rangle) = 1$. We then label the edge $\langle x^1, x^2 \rangle$ with $c^{12}(\langle x^1, x^2 \rangle)$, $f^{12}(\langle x^1, x^2 \rangle)$, or $\pi^{12}(\langle x^1, x^2 \rangle)$ as appropriate, and the label on a node $x^i \in X_i$ is the marginal $c^i(x^i)$, $f^i(x^i)$, or $\pi^i(x^i)$.

Fig. 1 shows f^{12} and π^{12} for the example above. Note that the unlabeled version of the graph is simply a representation of I or R: those nodes which have ever been seen, and are thus connected with some positive probability or possibility, are present.

6 Random Set Representations

For continued simplicity, again assume $K = \{1, 2\}$. For $i \in K$, denote $i' := 3 - i$. Define two **structure functions** $S_i : X_i \mapsto 2^{X_{i'}}$, where

$$S_i(x^i) := \left\{ x^{i'} \in X_{i'} : I^{12}\left(\left\langle x^i, x^{i'} \right\rangle\right) = 1 \right\} \subseteq X_{i'}.$$

In other words, $S_i(x^i)$ is the image of I in $X_{i'}$, or the subset of all nodes in $X_{i'}$ that are linked to x^i with positive probability or possibility.

Each marginal frequency distribution f^i induces a random subset of $X_{i'}$, where $\forall A \subseteq X_{i'}$,

$$m(A) := \begin{cases} \sum_{x^i \in S_i^{-1}(A)} f^i(x^i), & \exists x^i \in X_i, S(x^i) = A \\ 0, & \text{otherwise} \end{cases}$$

In other words, each marginal probability $f^i(x^i)$ is projected into a subset probability over the collection of nodes in the other dimension that it is positively linked to.

Consider our example for $K = \{1, 2\}$. Then S^1 and S^2 induce respectively the random sets

$$\mathcal{S}^1 := \{\langle \{x, y\}, .34 \rangle, \langle \{x\}, .56 \rangle, \langle \{y\}, .10 \rangle\},$$

$$\mathcal{S}^2 := \{\langle \{a, c\}, .74 \rangle, \langle \{a, b\}, .26 \rangle\},$$

with traces $\rho^1 = \langle .90, .44 \rangle, \rho^2 = \langle 1.00, .26, .74 \rangle$. Notice that $\mathcal{S}^2$ is consistent, but S^1 is not, so that ρ^2 is a possibility distribution. This example is shown graphically in Fig. 2.

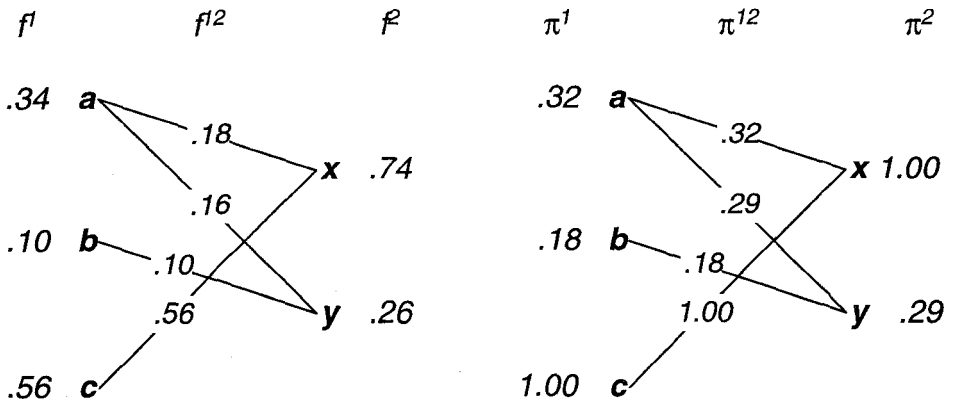

Figure 1: Graphical representations of $\mathcal{D}(N, \{1,2\})$. (Left) f^{12}. (Right) π^{12}.

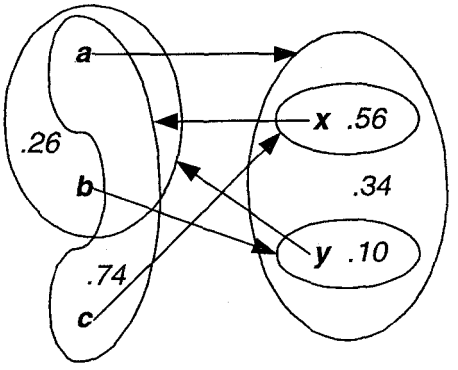

Figure 2: Graphical representations of the random sets $\mathcal{S}^1, \mathcal{S}^2$ induced by $\mathcal{D}(N, 12)$.

The $\mathcal{S}^i$ reveal a combination of structural and numerical information about the joint distribution, and in particular the distribution of joint values which are not positively related, or for which $I(\vec{x}) = 0$. In particular, if all joint values are possible, or have been seen, then the induced random sets will be vacuous.

Proposition 4 If $\forall j \in \mathbb{N}_{n_i}, \forall j' \in \mathbb{N}_{n_{i'}}$, $I(\langle x_j, x_{j'} \rangle) = 1$, then $\mathcal{S}_i = \{\langle X_{i'}, 1 \rangle\}$.

Furthermore, this structural information is completely derivable from the joint distribution f^{12}. Instead, what we are especially interested in is situations where there is incomplete information, which only the random set model can represent.

Consider the situation where we have good statistical information about one field X_1 in a database, but only weak structural or coupling information about how that variable is related to another X_2. Thus, continuing our example, we are given f^1, and I, but not the other marginal f^2 or the joint f^{12}. In our example, this would be represented by the table

f^{12}	X_2		f^1
X_1	x	y	
a	?	?	.34
b	0	?	.10
c	?	0	.56
f^2	?	?	1.00

where ? indicates any number in $(0, 1]$, or in other words just that $I^{12} = 1$ in that cell. This is shown graphically in Fig. 3.

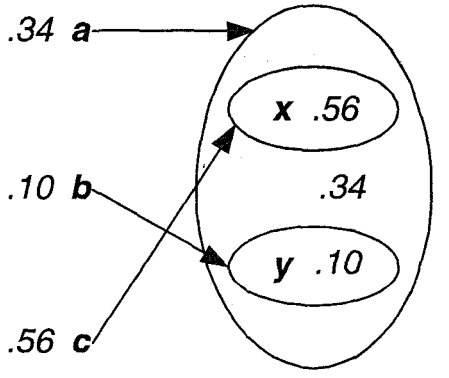

Figure 3: Random set $\mathcal{S}^1$ when only f^1 and I are known.

In this case while we can determine $f^{12}(c, x) = .56$ and $f^{12}(b, y) = .10$, still $f^{12}(a, x)$ and $f^{12}(a, y)$ have a degree of freedom between them, and each is bounded within $(0, .34)$.

In general the f^{12} are vastly underdetermined by f^1 and I, and only the random set representation S^1 is sufficient. In particular, ρ will be an upper bound on a class of functions which approximate a probability distribution which f^i would impose on $f^{i'}$.

7 Conclusions

This examination is, of course, preliminary and cursory, and a number of important questions and directions immediately present themselves:

- Definition of the final, crucial, database operation of refining and coarsening of any of the dimensions X_i (these operations change the n_i, and thus the number of overall possible states, while not affecting N or M);
- Similarly, treatment of fuzzy projections, allowing for gradations of inclusion in coarsened categories;
- The expected behavior of joint and conditional information measures under each of the operations of projection, subsetting, and refining;
- Extension of induced random sets to $|K| > 2$;
- The relationship between the properties of databases and the topological structure of the corresponding induced random sets;
- And finally, and most importantly, the relationship between the directly derived information functions f and π and the corresponding induced random set distributions ρ, especially when ρ is itself either additively f or maxitive π.

References

[1] Cavallo, Roger E and Klir, George: (1982) "Reconstruction of Possibilistic Behavior Systems", *Fuzzy Sets and Systems*, v. **8**

[2] Date, C.J.: (1981) *An Introduction to Database Systems*, Addison-Welsey, Reading, MA.

[3] Fayyad, Usama M., G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy, eds.: (1996) *Advances in Knowledge Discovery and Data Mining*, MIT Press, Menlo Park, CA.

[4] Joslyn, Cliff: (1996) "Aggregation and Completion of Random Sets with Distributional Fuzzy Measures", *Int. J. of Uncertainty, Fuzziness, and Knowledge-Based Systems*, v. **4**:4, pp. 307-329

[5] Klir, George: (1985) *Architecture of Systems Problem Solving*, Plenum, New York

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

[7] Klir, George, and David Harmanec: (1996) "Generalized Information Theory: Recent Developments", *Kybernetes*, v. **25**:7-8, pp. 50-67.

[8] Knoke, David, and P.J. Burke: (1980) "Log-Linear Models", in series: *Quantitative Applications in the Social Sciences*, Sage Publications, New York.

[9] Shaffer, Gary P.: (1997) "K-Systems Analysis: An Anti-Silver Bullet Approach to Ecosystems Modeling", *Advances in Systems Science and Applications*, pp. 32-39.

[10] Wang, Zhenyuan and Klir, George J: (1992) *Fuzzy Measure Theory*, Plenum Press, New York