Random Set Analysis of System Response Given Uncertain Parameters

Jim Hall
Department of Civil Engineering,
University of Bristol, UK
jim.hall@bristol.ac.uk

Jonathan Lawry
Department of Engineering Mathematics,
University of Bristol, UK
j.lawry@bristol.ac.uk

Abstract

Random set theory provides a convenient mechanism for representing uncertain knowledge, including probabilistic and set-based information, and extending it through functional relationships. Where the available information is in terms of lower and upper bounds on a set of probability measures these bounds will not necessarily correspond respectively to belief and plausibility functions. In this case when a random set cannot be obtained from the Möbius inversion we propose an Iterative Rescaling Method for constructing a random set with corresponding belief and plausibility functions that are a close outer approximation to the lower and upper probability distributions. In situations where information about an uncertain input parameter comes from more than one source, approaches to information fusion based on averaging, Dempster's rule of combination and a set union version of Dempster's rule are discussed.

Keywords. Random set theory, belief and plausibility functions, Möbius inversion, Iterative Rescaling Method

1 Random sets

Random set theory provides a general mechanism for handling interval-based measurements, fuzzy sets and discrete probability distributions. Following Dubois and Prade (1990, 1991), a finite support random set on a universal set X is a pair $(\Im, m)$ and a *mass assignment* is a mapping

$$m : \Im \rightarrow [0,1] \tag{1}$$

such that $m(\varnothing) = 0$ and

$$\sum_{A \in \Im} m(A) = 1 \tag{2}$$

Each set $A \in \Im$ contains the possible values of a variable $x \in X$, and $m(A)$ can be viewed as the probability that $x \in A$ but does not belong to any special subset of A. Given a random set $(\Im, m)$, a belief function Bel (Shafer, 1976) can be defined as the following set function

$$\forall A \in X,\ \ Bel(A) = \sum_{B \subseteq A} m(B) \tag{3}$$

and its dual plausibility function $Pl(A)$ is defined by

$$\forall A \in X, Pl(A) = 1 - Bel(\overline{A}) \tag{4}$$

$Bel(A)$ can be viewed as the lower bound on a family of probability measures and $Pl(A)$ as the upper bound, although the converse is not true, *i.e.* upper and lower probability functions are more general than belief and plausibility functions. When $\Im$ contains only singletons $Bel = Pl$ is a probability measure (with finite support).

When $\Im$ is a nested family $A_1 \subset A_2 \subset \ldots \subset A_n$ then Bel is a necessity measure ? and Pl is a possibility measure p (Zadeh 1978), and the random set is said to be *consonant*. A fuzzy set F can be defined from any random set $(\Im, m)$ as follows:

$$\forall x,\ \ m_F(x) = \sum_{x \subseteq A} m(A) = Pl(\{x\}) = p(\{x\}) \tag{5}$$

where m_F is the fuzzy membership in F.

2 Extending random sets through functional relations

Let g be a mapping $X_1 \times \ldots \times X_n \rightarrow Y$. Let $x_1,\ldots x_n$ be variables whose values are incompletely known. The question dealt with in the *Challenge Problems* of Oberkampf *et al.* (2001) is to find the range of the variable $y = g(\mathbf{x}) : \mathbf{x} = (x_1,\ldots x_n)$, and, where sufficient information exists to do so, a probability distribution over the range of y, from the available information restricting the values of $x_1,\ldots x_n$. In the *Challenge Problems* each of the variables is specified as being independent, but we first address the general case, where the dependency between $(x_1,\ldots x_n)$ can be expressed as a random relation R, which is a random set $(\hat{A},\ r)$ on the Cartesian product $X_1 \times \ldots \times X_n$, in which case the range of y is the random set $(\Im, m)$ such that:

$$\Im = \{y(R_i) \mid R_i \in \hat{A}\}, \quad y(R_i) = \{y(\mathbf{x}) \mid \mathbf{x} \in R_i\} \qquad (6a)$$

$$m(A) = \sum_{A = y(R_i)} r(R_i) \qquad (6b)$$

Special cases of Equations (6) for (i) set-valued variables (ii) consonant random Cartesian products (iii) stochastically decomposable Cartesian products and (iv) joint probability distributions, were addressed by Dubois and Prade (1991). In the case of consonant random sets the extension principle for fuzzy sets applies (Zadeh, 1975) and the image of the random relation R can be constructed from the images of the level cuts of R.

In general Equations (6) involve calculating the image of each focal element $R_i \in \hat{A}$, by applying twice the techniques of global optimisation. If the focal elements of R_i are compact sets and g is a continuous function

$$g(R_i) = [l_i, u_i] \qquad (7)$$

where

$$l_i = \min_{x \in R_i} g(x) \qquad (8)$$

$$u_i = \max_{x \in R_i} g(x) \qquad (9)$$

When each parameter x_i is specified by a marginal random set, whose focal elements are each an interval $[l_i, u_i]$, then methods of interval analysis (More 1966) are applicable.

Under certain special conditions the Vertex method (Dong and Shah 1987) applies and can be used to greatly reduce computational expense. Suppose each focal element R_i of the random relation $(\hat{A},\ r)$ is a p-dimensional box, whose 2^n vertices are indicated as v_j, $j = 1, \ldots, 2^n$. If $y = g(\mathbf{x})$ is continuous in R_i and also no extreme points exist in this region (including its boundaries), then

$$g(R_i) = \begin{matrix} [\min_j \{g(v_j) : j = 1, \ldots 2^n\}, \\ \max_j \{g(v_j) : j = 1, \ldots 2^n\}] \end{matrix} \qquad (10)$$

Thus function f has to evaluated 2^n times for each focal element R_i. This computational burden can be further reduced if g is continuous, its partial derivatives are continuous and if g is a strictly monotonic function with respect to each parameter x_i, in which case

$$\exists! v_j : g(v_j) = \min_j \{g(v_j) : j = 1, \ldots 2^n\} \qquad (11)$$

$$\exists! v_k : g(v_k) = \max_k \{g(v_k) : k = 1, \ldots 2^n\} \qquad (12)$$

and v_j and v_k can be identified merely by consideration of the direction of increase of g. Thus g has to be calculated only twice for each focal element R_i (Tonon *et al*. 2000).

3 Random set approximations to lower and upper probability distributions

Consider the distribution of a continuous random variable $\boldsymbol{x}$. By definition

$$\Pr(\boldsymbol{x} \in A) = \int_A f(x, \mathbf{a}) dx \qquad (13)$$

where f is a probability density function and $\mathbf{a}$ is a vector of parameters of f, $\mathbf{a} = [a_1, a_2, \ldots, a_n]$. The cumulative distribution function $F(x)$ is

$$F(x) = \Pr(\mathbf{x} = x) = \int_{-\infty}^{x} f(t, \mathbf{a}) dt \qquad (14)$$

Definition 1: If each parameter a_i in $\mathbf{a}$ is specified by a closed interval $[l_i, u_i]$ then $\mathbf{a}$ is constrained by an n-dimensional box Q. We define the lower probability $P_*(A)$ as the capacity functional

$$P_*(A) = \inf_{\mathbf{a} \in Q} \int_A f(x, \mathbf{a}) dx, \quad \forall A \in \mathcal{P}(X) \qquad (15)$$

where $\mathcal{P}(X)$ is the power set of X. The upper probability $P^*(A)$ is

$$P^*(A) = \sup_{\mathbf{a} \in Q} \int_A f(x, \mathbf{a}) dx, \quad \forall A \in \mathcal{P}(X) \qquad (16)$$

$P_*(\overline{A})$ and $P^*(A)$ will be located at the same value of $\mathbf{a}$, so $P_*(\overline{A}) = 1 - P^*(A)$ and it suffices to consider lower probabilities. The capacity $P_*(A)$ as defined by Equation (15) is not necessarily a belief function.

Lemma 1: (see Shafer 1976 for proof) Suppose X is a finite set and r and s are functions on the power set $\mathcal{P}(X)$, then

$$r(A) = \sum_{B \subseteq A} s(B) \qquad (17)$$

for all $A \subset X$ if and only if

$$s(A) = \sum_{B \subseteq A} (-1)^{|A-B|} r(B). \qquad (18)$$

where $|A\text{-}B|$ denotes the cardinality of the set $A\text{-}B$ *i.e.* $|A \cap B|$. Equation (18) is referred to as the *Möbius inversion*. Observing that Equation (17) is equivalent to the definition of a belief function *Bel* from a mass assignment m (Equation (3)), Shafer (1976) proposed the use of the Möbius inversion as a general mechanism for reconstructing m from *Bel i.e*.

$$m(A) = \sum_{B \subseteq A} (-1)^{|A \cap \overline{B}|} Bel(B) \quad (19)$$

Lemma 2: Suppose that r is an unknown function on $\mathcal{P}(X)$ *i.e.* not necessarily a belief function, if the function s generated from Equation (18) conforms to the axioms of a mass assignment *i.e.*

$$\sum_{A \subseteq X} s(A) = 1 \quad (20)$$

and

$$s(A) \geq 0,\ \forall A \subseteq X \quad (21)$$

(and $s(\varnothing)$ is automatically equal to zero), then s is a mass assignment m and r is a belief function Bel.

Lemma 2 provides a mechanism for establishing whether the capacity functional P_* defined by Equation (15) is a belief function. If application of Equation (15) does not yield a mass assignment then P_* is not a belief function.

Example 1: Consider the following example in which $X = \{a, b, c\}$ and

$P_*(\{a\}) = 0$	$P^*(\{a\}) = 0.6$
$P_*(\{b\}) = 0$	$P^*(\{b\}) = 0.6$
$P_*(\{c\}) = 0$	$P^*(\{c\}) = 0.6$
$P_*(\{a,b\}) = 0.4$	$P^*(\{a,b\}) = 1$
$P_*(\{a,c\}) = 0.4$	$P^*(\{a,c\}) = 1$
$P_*(\{b,c\}) = 0.4$	$P^*(\{b,c\}) = 1$

Notice that the sets of constraints are consistent since they are, for example, satisfied by the probability distribution $P(\{a\}) = 0.4$, $P(\{b\}) = 0.4$, $P(\{c\}) = 0.2$. Applying the Möbius inversion to the singletons gives $m(\{a\}) = m(\{b\}) = m(\{c\}) = 0$. Next considering sets with cardinality 2 we obtain $m(\{a,b\}) = 0.4$, $m(\{a,c\}) = 0.4$ and $m(\{b,c\}) = 0.4$, giving a sum greater than 1, so the condition in Equation (20) has been violated. If the Möbius inversion is applied at cardinality 3 a negative mass $m(\{a,b,c\}) = -0.2$ is obtained.

In order to use Equations (6) for projecting uncertainty through some function g we require a mechanism for constructing a random set that is consistent with the probability bounds in Equations (15) and (16). In other words, we require a belief function that approximates P_* from below such that $Bel(A) = P_*(A)\ \forall A \in X$. It would also be desirable if the constraints generated by this belief measure and its associated plausibility measure where as close as possible to the original upper and lower probabilities. In the sequel we adopt the following measure of accuracy E for a belief function Bel approximating P_*:

$$E(Bel) = \frac{1}{2^{|X|}} \sum_{A \subseteq X} \left(P_*(A) - Bel(A) \right) \quad (22)$$

In the case that we have coherent upper and lower probabilities then this measure has the pleasant property that

$$E(Bel) = \frac{1}{2^{|X|}} \sum_{A \subseteq X} \left(P_*(A) - Bel(A) \right)$$
$$= \frac{1}{2^{|X|}} \sum_{A \subseteq X} \left(1 - P^*(\overline{A}) - 1 + Pl(\overline{A}) \right) = \frac{1}{2^{|X|}} \sum_{A \subseteq X} \left(Pl(\overline{A}) - P^*(\overline{A}) \right)$$
$$= \frac{1}{2^{|X|}} \sum_{A \subseteq X} \left(Pl(A) - P^*(A) \right) \quad (23)$$

Notice that for Example 1 any belief function consistent with P_* must have a mass assignment of the following form:

$$m(\{a\}) = m(\{b\}) = m(\{c\}) = 0$$
$$m(\{a,b\}) = x, m(\{a,c\}) = y, m(\{b,c\}) = z$$
$$m(\{a,b,c\}) = 1 - x - y - z$$

subject to $x, y, z \leq 0.4$ and $x + y + z \leq 1$. In this case the error associated with any consistent belief function is given by:

$$E(Bel) = \frac{1}{8}\left((0.4 - x) + (0.4 - y) + (0.4 - z) \right)$$
$$= \frac{1}{8}\left(1.2 - (x + y + z) \right) \quad (24)$$

and this is clearly minimal for any mass assignment such that $x + y + z = 1$ giving a value of 0.025.

We now propose an algorithm based on a simple heuristic search for finding a good approximating belief function (on the basis of E) for P_*.

Algorithm: Iterative Rescaling Method (IRM)
Order the subsets of X according to increasing cardinality and arbitrarily for subsets with the same cardinality. Specifically, suppose

$\mathcal{F} = \mathcal{P}(X) - \{\varnothing\} = \{A_i | i = 1, \ldots, 2^{|X|}-1\}$

are ordered such that if $i = j$ then $|A_i| = |A_j|$.

1. For A_i evaluate $m(A_i) = P_*(A_i) - \sum_{B \in \mathcal{F}: B \subset A_i} m(B)$
2. If $m(A_i) \geq 0$ then:
   Let $i = i+1$ and goto 3
   Else:
   For all $B \subset A_i$ determine the largest value of $k < |A_i|$ for which
   $$\sum_{B \subset A_j : |B| \leq k} m(B) \leq P_*(A_j).$$
   For all $B \subset A_i$ with $|B| = k$ leave $m(B)$ unchanged. For all $B \subset A_i$ with $|B| > k$ rescale $m(B)$ according to

$$m(B) = m(B)\left(\frac{P_*(A_i) - \sum_{B\subset A_j:|B|\le k} m(B)}{\sum_{B\subset A_j:|B|>k} m(B)}\right). \quad (25)$$

Set $m(A_i) = 0$, let $i = i+1$ and goto 3

3. If $i > 2^{|X|} - 1$ then:
   Terminate
   Else:
   Goto 1

It is natural to think of the IRM algorithm as being implemented in stages where for each stage masses are calculated according to the Möbius inversion until a negative mass occurs. At this point the relevant subsets are rescaled and the algorithm continues in the next stage until the next negative value is encountered. The following table illustrates the IRM algorithm for the constraints given in Example 1:

Table 1 :IRM applied to Example 1

Möbius inversion	Random set from IRM
$m(\{a\}) = 0$	$m(\{a\}) = 0$
$m(\{b\}) = 0$	$m(\{b\}) = 0$
$m(\{c\}) = 0$	$m(\{c\}) = 0$
$m(\{a,b\}) = 0.4$	$m(\{a,b\}) = 1/3$
$m(\{a,c\}) = 0.4$	$m(\{a,c\}) = 1/3$
$m(\{b,c\}) = 0.4$	$m(\{b,c\}) = 1/3$
$m(\{a,b,c\}) = -0.2$ rescale factor = 1/1.2	$m(\{a,b,c\}) = 0$

Given the argument presented earlier we see that in this case IRM provides a belief function minimising E.

Theorem 1: The IRM algorithm results in a belief function Bel such that $\forall A \subseteq X$ $Bel(A) = P_*(A)$

Proof: (Note: In the following proof we shall abuse notation slightly and use $Bel(A_i)$ to denote $\sum_{B\subseteq A_j} m(A_j)$ even thought for intermediate stages of the IRM algorithm the latter may not formally correspond to a belief value.)

First note that for any $m(A_i) < 0$ the rescaling factor

$$\left(P_*(A_i) - \sum_{B\subset A_j:|B|\le k} m(B) \Big/ \sum_{B\subset A_j:|B|>k} m(B)\right) \in [0,1]$$

since by definition

$$P_*(A_i) \ge \sum_{B\subset A_j:|B|\le k} m(B) \text{ and } P_*(A_i) \le \sum_{B\subset A_j} m(B).$$

Suppose the IRM algorithm requires v rescalings at $A_{i_1},\cdots,A_{i_k}$ where $1 \le i_1 \le i_2 \le \cdots \le i_v \le 2^{|X|} - 1$. Then we prove the result by induction on rescalings.

Initial Case For rescaling 1 we have that $\forall j < i_1$ $m(A_j)$ currently has value $P_*(A_j) - \sum_{B\subset A_j} m(B)$ so that $Bel(A_j) = P_*(A_j)$. Since the rescaling factor is always in [0,1] then rescaling 1 has the effect that $\forall j < i_1$ $m(A_j)$ either decreases or remains unchanged and hence $\forall j < i_1$ $Bel(A_j) = P_*(A_j)$ after rescaling. Furthermore, $m(A_{i_1})$ is set to zero and the masses of the subsets of A_{i_1} are rescaled such that $Bel(A_{i_1}) = P_*(A_{i_1})$. Hence, after rescaling 1 it holds that $\forall j < i_1$ $Bel(A_j) = P_*(A_j)$.

Inductive Step Suppose that after rescaling t it holds that $\forall j < i_t$ $Bel(A_j) = P_*(A_j)$ then after rescaling $t+1$ $\forall j < i_t$ $m(A_j)$ either decreases or remains unchanged and hence by the inductive hypothesis $\forall j < i_t$ $Bel(A_j) = P_*(A_j)$.

$\forall j \in (i_t, i_{t+1})$ $m(A_j)$ is unaffected by rescaling t. Before rescaling $t+1$ $m(A_j)$ is set to $P_*(A_j) - \sum_{B\subset A_j} m(B)$ so that $Bel(A_j) = P_*(A_j)$. Therefore, after rescaling $t+1$ $\forall j \in (i_t, i_{t+1})$ $Bel(A_j) = P_*(A_j)$ since the scaling factor is in [0,1]. Furthermore, after rescaling $t+1$ $Bel(A_{i_{t+1}}) = P_*(A_{i_{t+1}})$ and therefore $\forall j = i_{t+1}$ $Bel(A_j) = P_*(A_j)$ as required.

In order to apply the IRM algorithm to constructing a random set we discretise X by sampling it over a grid in order to obtain a discrete random set. Suppose that X is a finite interval $[x_l, x_r]$. A s-algebra B can be defined on X by partitioning $[x_l, x_r]$ into s disjoint sub-intervals: $[x_l, x_2]$, $(x_2, x_3]$,…, $(x_{s-1}, x_s]$, $(x_s, x_r]$, according to the desired (or feasible) accuracy. B is therefore a family of 2^s sets. If P_* and P^* are continuous distributions, then the random set on B will be a discrete approximation, and the accuracy of this approximation will increase with increasing granularity in the definition of the partition. However, in practice the error in the IRM is found to increase slowly with the number of sets in B, so an optimal partition will balance the effect of these two approximations.

In order to illustrate the proposed approach a rather coarse partition has been applied to the interval [0.1,20.0], as shown in Table 2. The upper and lower probability distribution on the space corresponds to the family of lognormal distributions: $\ln x \sim N(m,s)$ with the value of the mean m and standard deviation s given by closed intervals. The set bounds have been chosen so that even though x is on $[0,\infty]$, more than 0.9999 of the total probability mass is contained within these bounds and the results are accurate to the quoted precision.

Table 2 Partition of the space of X in Example 2

Set	Bounds
$\{a\}$	[0.15, 1.3]
$\{b\}$	(1.3,2.0]
$\{c\}$	(2.0,3.0]
$\{d\}$	(3.0,18.0]

Example 2: Table 3 presents the results where $m \in [0.6,0.8]$ and $s \in [0.4,0.5]$. In this case $m(A_i)$ was first less than zero when $|A_i| = 3$ and rescaling was then applied four times to subsets of cardinality 2. For this example the error is $E(Bel) = 0.0136$. For problems of the size shown here the error function can be minimised using an optimisation method, and the result of such an optimisation is given in the column labelled m_{opt}, which gives a belief function with an error of 0.0131. Naturally, global optimisation is only practical for small scale problems, but for this example the error from IRM is reassuringly close to the minimum error.

For finer partitions it becomes impractical to write down the mass assignment and corresponding belief and plausibility functions. However, they can be partially visualised by plotting the cumulative belief and plausibility functions as defined below. The remainder of figures in this paper adopt this format.

Definition 2: Suppose that X is partitioned into s disjoint sub-intervals $[x_1, x_2]$, $(x_2,x_3]$,…, $(x_{s-1},x_s]$, $(x_s, x_{s+1}]$, and members of the corresponding power set are labelled $X_{i,j}$, $i = 1,…,s$, $j = 2,…, s+1$, $i>j$, according to the left and right hand bounds x_i and x_j. Note that this notation does not result in a unique label for every set in $\mathcal{P}(X)$ but is sufficient for the following definitions. We define the cumulative belief $CBel(x)$ at some point x in $[x_1, x_s]$ as

$$CBel(x) = \sum_{x \geq x_j} m(X_{i,j}) \tag{26}$$

and the cumulative plausibility $CPl(x)$ is defined as

$$CPl(x) = \sum_{x \geq x_i} m(X_{i,j}) \tag{27}$$

The lower and upper cumulative probability distribution functions, $F_*(x)$ and $F^*(x)$ respectively, have the conventional definitions *i.e.*

$$F_*(x) = \inf_{\mathbf{a} \in Q} \int_{-\infty}^{x} f(t,\mathbf{a})dt \tag{28}$$

and

$$F^*(x) = \sup_{\mathbf{a} \in Q} \int_{-\infty}^{x} f(t,\mathbf{a})dt \tag{29}$$

Cumulative lower and upper probability distributions and the corresponding bounding cumulative belief and plausibility functions for a partition of 10 intervals are illustrated in Figure 1. In this case IRM gives and error $E(Bel) = 0.0260$. The computational expense of problems on this scale can be reduced by applying the fast Möbius inversion of Thoma (1991).

Table 3 IRM random set approximation to upper and lower probability distribution (Example 2)

Set	P_*	P^*	Stage 1	Stage 2	Stage 3	Stage 4	m	Bel	Pl	m_{opt}
$\{a\}$	0.0895	0.2498	0.0895	0.0895	0.0895	0.0895	0.0895	0.0895	0.2498	0.0895
$\{b\}$	0.2743	0.3928	0.2743	0.2743	0.2743	0.2743	0.2743	0.2743	0.3989	0.2743
$\{c\}$	0.2668	0.3776	0.2668	0.2668	0.2668	0.2668	0.2668	0.2668	0.3967	0.2668
$\{d\}$	0.1063	0.2752	0.1063	0.1063	0.1063	0.1063	0.1063	0.1063	0.2893	0.1063
$\{a, b\}$	0.3947	0.5921	0.0310	0.0174	0.0120	0.0120	0.0120	0.3757	0.6009	0.0189
$\{a, c\}$	0.4506	0.5165	0.0943	0.0530	0.0530	0.0488	0.0488	0.4051	0.5619	0.0240
$\{a, d\}$	0.2959	0.4163	0.1001	0.1001	0.0692	0.0638	0.0638	0.2595	0.4396	0.0513
$\{b, c\}$	0.5837	0.7041	0.0427	0.0240	0.0240	0.0240	0.0194	0.5604	0.7405	0.0106
$\{b, d\}$	0.4835	0.5494	0.1029	0.1029	0.0711	0.0711	0.0575	0.4381	0.5949	0.0608
$\{c, d\}$	0.4079	0.6053	0.0349	0.0349	0.0349	0.0321	0.0260	0.3991	0.6243	0.0313
$\{a, b, c\}$	0.7248	0.8937	-0.0736	0.0000	0.0000	0.0000	0.0000	0.7107	0.8937	0.0000
$\{a, b, d\}$	0.6224	0.7332	0.0000	-0.0681	0.0000	0.0000	0.0000	0.6033	0.7332	0.0212
$\{a, c, d\}$	0.6072	0.7257	0.0000	0.0000	-0.0123	0.0000	0.0000	0.6011	0.7257	0.0380
$\{b, c, d\}$	0.7502	0.9105	0.0000	0.0000	0.0000	-0.0243	0.0000	0.7502	0.9105	0.0000
$\{a, b, c, d\}$	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0357	1.0000	1.0000	0.0069

Table 4 IRM random set approximation to upper and lower probability distribution (Example 3)

Set	P_*	P^*	Stage 1	Stage 2	Stage 3	m	Bel	Pl
$\{a\}$	0.0000	0.9478	0.0000	0.0000	0.0000	0.0000	0.0000	0.9478
$\{b\}$	0.0011	0.9688	0.0011	0.0011	0.0011	0.0011	0.0011	0.9688
$\{c\}$	0.0000	0.9574	0.0000	0.0000	0.0000	0.0000	0.0000	0.9760
$\{d\}$	0.0000	0.4218	0.0000	0.0000	0.0000	0.0000	0.0000	0.4448
$\{a,b\}$	0.0011	1.0000	0.0000	0.0000	0.0000	0.0000	0.0011	1.0000
$\{a,c\}$	0.0312	0.9574	0.0312	0.0312	0.0312	0.0312	0.0312	0.9760
$\{a,d\}$	0.0000	0.9478	0.0000	0.0000	0.0000	0.0000	0.0000	0.9707
$\{b,c\}$	0.0522	1.0000	0.0512	0.0512	0.0512	0.0282	0.0293	1.0000
$\{b,d\}$	0.0426	0.9688	0.0416	0.0415	0.0415	0.0229	0.0240	0.9688
$\{c,d\}$	0.0000	0.9989	0.0000	0.0000	0.0000	0.0000	0.0000	0.9989
$\{a,b,c\}$	0.5782	1.0000	0.4947	0.4947	0.4947	0.4947	0.5552	1.0000
$\{a,b,d\}$	0.0426	1.0000	-0.0000	0.0000	0.0000	0.0000	0.0240	1.0000
$\{a,c,d\}$	0.0312	0.9989	0.0000	-0.0000	0.0000	0.0000	0.0312	0.9989
$\{b,c,d\}$	0.0522	1.0000	0.0000	0.0000	-0.0415	0.0000	0.0522	1.0000
$\{a,b,c,d\}$	0.9999	1.0000	0.0000	0.0000	0.0000	0.4217	0.9999	1.0000

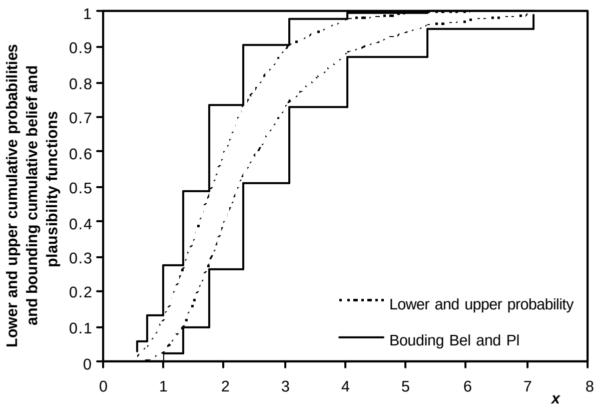

Figure 1 P_* and P^* for lognormal distributions with $m \in [0.6,0.8]$ and $s \in [0.4,0.5]$ and bounding belief and plausibility functions obtained from IRM

Example 3: Table 4 presents the results where $m \in [0.1, 1.0]$ and $s \in [0.1, 0.5]$. Here the error given by the IRM belief function is $E(Bel)$= 0.0052.

4 Combining information from different sources

If more than one source of information is available relating to some uncertain parameter x_i a mechanism is required to combine the various sources. The two most common approaches for constructing a random set from various sources are (i) averaging and (ii) Dempster's rule of combination. The set union version of Dempster's rule of combination provides a third alternative.

4.1 Averaging

Suppose there are n alternative random sets describing some variable x, each one corresponding to an independent source of information. One interpretation of the situation is to suppose that only one of the sources of information is correct, so, in the absence of any information about which source is true, an unbiased combination of the n random sets should be adopted. For each focal element $A \in \mathcal{P}(X)$

$$m(A) = \frac{1}{n}\sum_{i=1}^{n} m_i(A) \tag{30}$$

Baldwin *et al*. (1995) use an analogous approach, which they refer to as a 'voting model', for the construction of fuzzy sets by assimilating multiple sources of information to which a consonance condition is applied.

In the case when each of n sources of information is a single but in each case different set $A_1, A_2,\ldots, A_n$, then the averaging treats each of these sets as a focal element and assigns a mass of $1/n$ to each A_i of these focal elements. If the sets are nested then the resulting random set will be consonant.

In the case when the unknown parameters are specified as n lower and upper probability distributions then to combine the various items of evidence involves first finding a random set $(\Im_i, m_i) : i = 1,\ldots,n$ with corresponding belief and plausibility distributions that bound the lower and upper probability distributions, using the IRM, and then obtaining a merged random set $(\Im, m)$ such that

$$\Im = \bigcup_i \Im_i \qquad (31)$$

and $m(A)\ \forall\ A \in \Im$ is obtained from Equation 30.

Example 4: Suppose that a parameter b corresponds to lognormal distributions $\ln\ b \sim N(m, s)$ with the mean m and standard deviation s specified, respectively, by closed intervals $M = [m_1, m_2]$ and $S = [s_1, s_2]$ where the three information sources are as follows (*Challenge Problem* 5a)

$M_1 = [0.6, 0.8]$, $M_2 = [0.2, 0.9]$, $M_3 = [0.0, 1.0]$
$S_1 = [0.3, 0.4]$, $S_2 = [0.2, 0.45]$, $S_3 = [0.1, 0.5]$

In Figure 2 the three lower and upper cumulative probability distributions are plotted, together with the cumulative belief and plausibility distributions corresponding to the combined random set obtained by averaging. In this example the IRM generated errors $E(Bel_1) = 0.0222$, $E(Bel_2) = 0.0398$, $E(Bel_3) = 0.0083$.

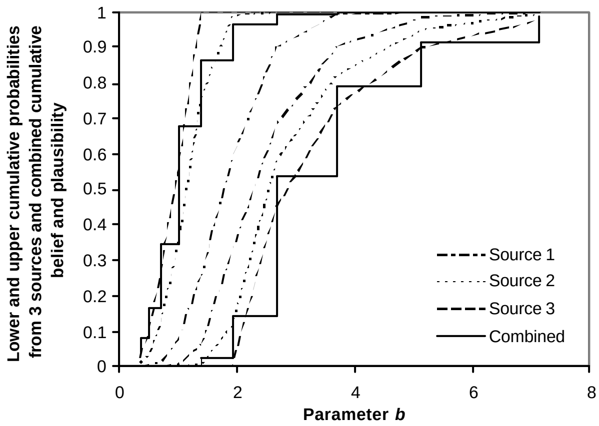

Figure 2 Belief and plausibility distributions corresponding to combined lower and upper probability distributions from averaging

4.2 Dempster's rule of combination

Dempster's rule (Shafer 1976) is a well-known mechanism for fusing evidence from different independent sources. Dempster's rule is thought of as being applicable to the situation where each information source provides some imprecise yet correct and consistent information about an unknown quantity or proposition. Suppose there are two items of evidence expressed as two mass assignments m_1 and m_2 on $P(X)$

$$m_{1,2}(A) = \frac{\sum_{B\cap C=A} m_1(B).m_2(C)}{1-K} \qquad (32)$$

for $A \neq 0$, where

$$K = \sum_{B\cap C=\varnothing} m_1(B).m_2(C) \qquad (33)$$

and $m_{1,2}(\varnothing)=0$.

Example 5: Dempster's rule has been applied to the same problem as Example 4 (*Challenge Problem* 5a). The cumulative belief and plausibility distributions corresponding to the combined random set are illustrated in Figure 3.

In the case when each of n sources of information is a single, but in each case different, set $A_1, A_2,\ldots, A_n$, then Dempster's rule of combination will result in a mass of unity being applied to the intersection $A_1\cap A_2\cap\ldots\cap A_n$, provided $A_1\cap A_2\cap\ldots\cap A_n \neq 0$. When $A_1\cap A_2\cap\ldots\cap A_n \neq 0$ then no combination is defined.

Dempster's rule has been criticised from generating counter-intuitive results in situations where the information to be combined is not consistent (Zadeh 1986, Walley 1991). These situations do not correspond with the semantic description given above of the situations to which the rule is applicable, so the counter-intuitive results are unsurprising. In situations of significant inconsistency or conflict, averaging or the set union version of Dempster's rule of combination is more applicable.

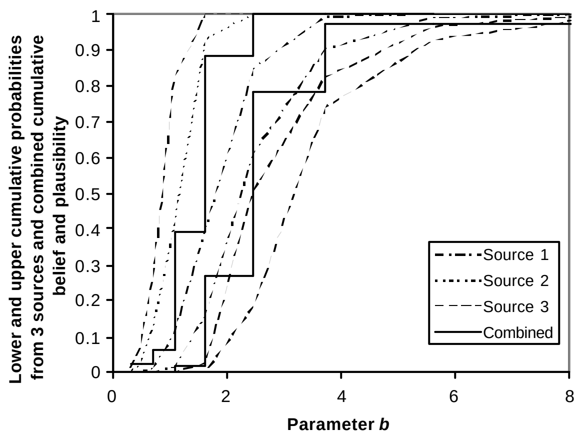

Figure 3 Bounds on three cumulative probability distributions and the combined distribution from the Dempster's rule of combination (Example 5)

4.3 Set union version of Dempster's rule of combination

Suppose there are two items of evidence expressed as two mass assignments m_1 and m_2 on $P(X)$, then the set union version of Dempster's rule of combination is

$$m_{1,2}(A) = \sum_{B\cup C=A} m_1(B).m_2(C) \qquad (34)$$

In this case the combination of evidence is consistent with the belief that one or other (or both) of the experts are correct. This is in contrast to Dempster's Rule where, up to inconsistency, it is assumed that both experts are correct.

In the case when each of n sources of information is a single, but in each case different, set A_1, A_2,…, A_n, then the set union version of Dempster's rule of combination will result in a mass of unity being applied to the union $A_1 \cup A_2 \cup \ldots \cup A_n$, which is clearly also applicable in situations of conflicting evidence.

Example 6: The set union version of Dempster's rule has been applied to the same problem as Example 4 (*Challenge Problem* 5a). The cumulative belief and plausibility distributions corresponding to the combined random set are illustrated in Figure 4.

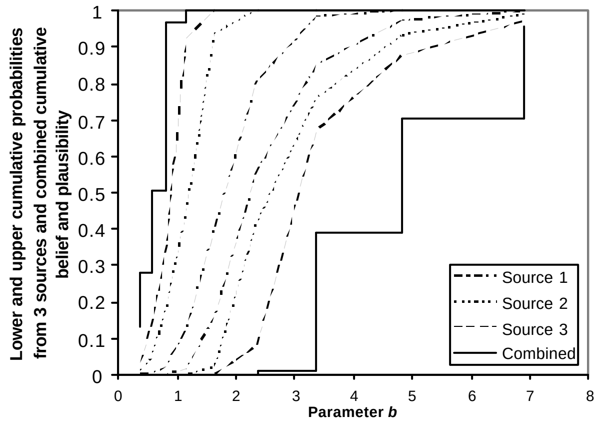

Figure 4 Bounds on three cumulative probability distributions and the combined distribution from the set union version of Dempster's rule (Example 6)

Example 7: To illustrate the final stage in the random set analysis consider the function

$$y = (a + b)^a \tag{35}$$

Suppose that evidence of parameter a is provided by three independent sources who specify that a is in the closed interval A, for which they provide the estimates A_1 = [0.5, 0.7], A_2 = [0.3, 0.8], A_3 = [0.1, 1.0]. Parameter b is given by the probability distributions addressed in Examples 4, 5 and 6. Figure 5 illustrates three estimates of the cumulative belief and plausibility distributions of y, corresponding to estimates of b generated using the IRM in Examples 4, 5 and 6.

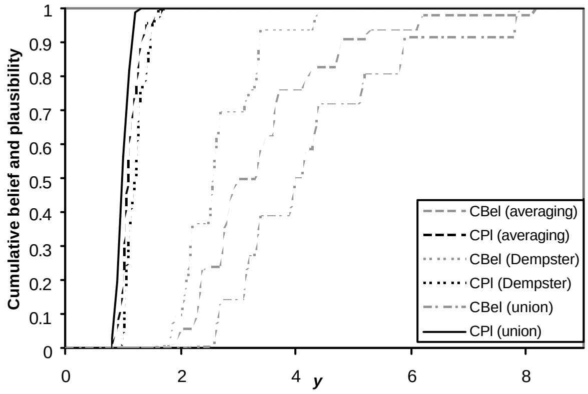

Figure 5 Cumulative belief and plausibility distributions on $y = (a + b)^a$

5 Conclusions

Random set theory provides a general framework encompassing interval bounds, fuzzy sets and discrete probability distributions. The random set extension principle introduced by Dubois and Prade (1991) enables random sets to be projected through functional relationships describing system behaviour in order to generate a random set on the system response. In general this will involve the solution to a double optimisation problem, though when the vertex method is applicable the computational burden is significantly reduced.

This paper has addressed two aspects of a random set approach to the *Challenge Problems* of Oberkampf *et al.* (2001). The Iterative Rescaling Method has been proposed for generating a random set and corresponding belief and plausibility functions that are an outer approximation to lower and upper probability distributions. The Iterative Rescaling Method generates a low average error on problems of a practical scale. Three approaches to combining information from different sources have been illustrated with examples. Of these the averaging has the attraction of general applicability, computational ease and accessible semantic interpretation.

References

J.F. Baldwin, T.P. Martin, B.W. Pilsworth. *Fril - Fuzzy and Evidential Reasoning in AI*, Research Studies Press, Taunton, 1995.

W. Dong and H.C. Shah. Vertex method for computing functions of fuzzy variables. *Fuzzy Sets & Syst.*, 24:65-78, 1987.

D. Dubois and H. Prade. Consonant approximations of belief functions. *Int. J. Approx. Reasoning*, 4:419-449, 1990.

Dubois, D. and H. Prade. Random sets and fuzzy interval analysis. *Fuzzy Sets & Syst*., 42:87-101, 1991 .

W.L. Oberkampf, J.C. Helton, C.A. Joslyn, S.F. Wojtkiewicz and S. Ferson. Challenge problems: uncertainty in system response given uncertain parameters. Draft version, November 1, 2001. www.sandia.gov/epistemic/challenge.htm

R.E. Moore. *Interval Analysis*, Prentice Hall, Englewood Cliffs, 1966.

G. Shafer. *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, 1976.

H.M. Thoma. Belief function computations. in *Conditional Logic in Expert Systems*, I.R. Goodman, M.M. Gupta, H.T. Nguyen and G.S. Rogers (eds), Elsevier, New York, 1991.

F. Tonon, A Bernardini and A. Mammino. Reliability analysis of rock mass response by means of random set theory. *Reliability Eng. and Syst. Safety*, 70(3):263-282, 2000.

P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

L.A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning. *Information Science* 8:199-249, 1975; p:43-80, 1975.

L.A. Zadeh. Fuzzy sets as the basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3-28, 1978.

L.A. Zadeh. A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination. *A.I. Magazine*, 7(2):85-90, 1986.