International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems
Vol. 2, No. 2 (1994) 167–196
©World Scientific Publishing Company

PROBABILITY INTERVALS: A TOOL FOR UNCERTAIN REASONING

LUIS M. DE CAMPOS, JUAN F. HUETE, and SERAFIN MORAL

Departamento de Ciencias de la Computación e I.A.
Universidad de Granada, 18071-Granada, Spain

Received 25 June 1993
Revised 10 December 1993

We study probability intervals as an interesting tool to represent uncertain information. A number of basic operations necessary to develop a calculus with probability intervals, such as combination, marginalization, conditioning and integration are studied in detail. Moreover, probability intervals are compared with other uncertainty theories, such as lower and upper probabilities, Choquet capacities of order two and belief and plausibility functions. The advantages of probability intervals with respect to these formalisms in computational efficiency are also highlighted.

Keywords: Uncertainty Management, Probability Intervals, Lower and Upper Probabilities, Combination, Marginalization, Conditioning.

1. Introduction

When dealing with numerical uncertainty in Artificial Intelligence there are many different theories available. Some of them, at least in a formal sense, are hierarchically ordered, going from the general to the specific. Hence, we have general fuzzy measures, lower and upper probabilities, Choquet capacities of order two and belief and plausibility functions, which include both necessity/possibility and probability measures. Usually, the more general the theory is, the more expressive capabilities it has, and the less efficiently the computations within this theory can be carried out.

In this paper, we study in detail a formalism for representing uncertain information: probability intervals. This formalism is easy to understand and it combines a reasonable expressive power and efficient computation. The main concepts and tools necessary for the development of a theory of uncertain information, such as precision (inclusion), combination, marginalization, conditioning and integration are studied for probability intervals. Moreover, the place of probability intervals in the hierarchy above is also analized.

The paper is divided into 7 sections. In section 2 we formally introduce probability intervals, and study their relationship with upper and lower probabilities and convex sets of probabilities. Section 3 is devoted to the combination of probability intervals, and the associated problem of inclusion. The basic concepts of marginalization and conditioning of probability intervals are analized in section 4. Section 5 studies methods of integration with respect to probability intervals. In section 6, the relationships between probability intervals and belief and plausibility functions, together with methods of approximation, are considered. Finally, section 7 contains the concluding remarks and some proposals for future work.

2. Probability Intervals

Let us consider a variable X taking its values in a finite set $D_x = \{x_1, x_2, \ldots, x_n\}$ and a family of intervals $L = \{[l_i, u_i], i = 1, \ldots, n\}$, verifying $0 \le l_i \le u_i \le 1\ \forall i$. We can interpret these intervals as a set of bounds of probability by defining the set $\mathcal{P}$ of probability distributions on D_x as

$$\mathcal{P} = \{P \in \mathcal{P}(D_x) \mid l_i \le p(x_i) \le u_i,\ \forall i\}, \tag{1}$$

where $\mathcal{P}(D_x)$ denotes the set of all the probability measures defined on a finite domain D_x. So, we will say that L is a set of *probability intervals*, and $\mathcal{P}$ is the set of *possible probabilities associated* to L.

As $\mathcal{P}$ is obviously a convex set, we can consider a set of probability intervals as a particular case of a convex set (a polytope) of probabilities with a finite set of extreme points[1,2,3,4,5].

In order to avoid the set $\mathcal{P}$ being empty, it is necessary to impose some conditions on the intervals $[l_i, u_i]$, namely that the sum of the lower bounds is less than or equal to one, and the sum of the upper bounds is greater than or equal to one:

$$\sum_{i=1}^{n} l_i \le 1 \le \sum_{i=1}^{n} u_i\,. \tag{2}$$

A set of probability intervals verifying the condition (2) will be called *proper*. We always use proper probability intervals, because non proper intervals, associated to the empty set, are useless.

In addition to a convex set $\mathcal{P}$, we can also associate with the proper intervals $[l_i, u_i]$ a pair (l, u) of lower and upper probabilities[6,7,8,9,10](also called a pair of representable measures, or a probability envelope) through $\mathcal{P}$ as follows:

$$l(A) = \inf_{P\in\mathcal{P}} P(A),\ u(A) = \sup_{P\in\mathcal{P}} P(A),\ \forall A \subseteq D_x\,. \tag{3}$$

So, probability intervals can also be considered as particular cases of lower and upper probabilities, where the set of associated probabilities is defined by restrictions affecting only the individual probabilities $p(x_i)$ (restrictions like, for example, $p(x_i)+p(x_j) \le u_{ij}$, or $p(x_i)+p(x_j)+p(x_k) \ge l_{ijk}$, are possible in general lower and upper probabilities, but they are not allowed in the case of probability intervals. Here we only allow restrictions such as $p(x_i) \ge l_i$ and $p(x_i) \le u_i$).

In order to maintain consistency between both views of probability intervals, it would be important for the restriction of $l(.)$ and $u(.)$ to the singletons (sets with only one element) to be equal to the original bounds, that is to say, that

$$l(\{x_i\}) = l_i,\ u(\{x_i\}) = u_i,\ \forall i. \tag{4}$$

These equalities are not always true: in general, we only have the inequalities

$$l(\{x_i\}) \geq l_i,\ u(\{x_i\}) \leq u_i,\ \forall i,$$

because for every probability P in $\mathcal{P}$, is $l_i \leq p(x_i) \leq u_i$, and we take the minimum and the maximum over these probabilities. But we can get the equality by modifying the bounds l_i and u_i without altering the set $\mathcal{P}$, that is to say, without changing the set of possible probabilities. First let us study which conditions the intervals $[l_i, u_i]$ should verify in order to get the equalities (4) (in Tessen[11] we can find an analogous study):

Proposition 1. Given a set of proper probability intervals $L = \{[l_i, u_i], i = 1, \ldots, n\}$, its corresponding convex set of probabilities $\mathcal{P}$ and the lower and upper probability pair (l, u) associated to L, then the equalities (4) are true if and only if the conditions

$$\sum_{j \neq i} l_j + u_i \leq 1 \text{ and } \sum_{j \neq i} u_j + l_i \geq 1,\ \forall i \tag{5}$$

hold.

Proof. As the inequalities $l(\{x_i\}) \geq l_i, u(\{x_i\}) \leq u_i\ \forall i$ are always true, then the conditions (4) are equivalent to the following ones: For each i there exist probabilities P^i and Q^i such that

$$p^i(x_i) = u_i \text{ and } l_j \leq p^i(x_j) \leq u_j,\ \forall j \neq i, \tag{6}$$

$$q^i(x_i) = l_i \text{ and } l_j \leq q^i(x_j) \leq u_j,\ \forall j \neq i. \tag{7}$$

The reason is that probabilities P^i and Q^i verifying (6) and (7) belong to $\mathcal{P}$ and they reach the maximum and minimum values u_i and l_i respectively. Now the equivalence of (6)–(7) and (5) can be easily proved after simple algebraic calculations $\square$.

A set of probability intervals verifying the conditions (5) will be called *reachable*. This name refers to the fact that the conditions (5) are equivalent to the equalities (4), which guarantee that the lower and upper bounds l_i and u_i can be reached by some probabilities in $\mathcal{P}$. Now, let us see how to modify these lower and upper bounds without changing the associated set of possible probabilities $\mathcal{P}$.

Proposition 2. Let $L = \{[l_i, u_i], i = 1, \ldots, n\}$ be a set of proper probability intervals, and $\mathcal{P}$ its associated convex set of probabilities. If we define a new set of probability intervals $L' = \{[l'_i, u'_i], i = 1, \ldots, n\}$ by means of

$$l'_i = l_i \vee \left(1 - \sum_{j \neq i} u_j\right), \; u'_i = u_i \wedge \left(1 - \sum_{j \neq i} l_j\right), \; \forall i, \tag{8}$$

then the set of probabilities associated to L' is also $\mathcal{P}$.

Proof. Let $\mathcal{P}'$ be the set of probabilities associated to L'. It is very easy to see that $l'_i \leq u'_i$, $\forall i$. So, $l_i \leq l'_i \leq u'_i \leq u_i$, $\forall i$, and then $\mathcal{P}' \subseteq \mathcal{P}$.
On the other hand, if $P \in \mathcal{P}$ then, because of the restriction $\sum_i p(x_i) = 1$, then inmediately $l'_i \leq p(x_i) \leq u'_i$ ∀i. So, $P \in \mathcal{P}'$ and thus $\mathcal{P} \subseteq \mathcal{P}'$ □.

According to proposition 2, we can replace the original set of probability intervals L by L' defined in (8) without affecting the set $\mathcal{P}$. This replacement permits us to refine the probability bounds that define $\mathcal{P}$ in such a way that these bounds can always be reached, as the next proposition shows.

Proposition 3. The probability intervals L' defined in (8) are reachable.

Proof. Let us prove that $\sum_{j \neq i} l'_j + u'_i \leq 1 \; \forall i$:
If $\forall j \neq i$ is $l_j \geq 1 - \sum_{m \neq j} u_m$, then $l'_j = l_j \; \forall j \neq i$. In these conditions, as $u'_i \leq 1 - \sum_{j \neq i} l_j$, we have $\sum_{j \neq i} l'_j + u'_i = \sum_{j \neq i} l_j + u'_i \leq 1$, and the result is true.
On the other hand, if $\exists h \neq i$ such that $l_h < 1 - \sum_{m \neq h} u_m$, then $l'_h = 1 - \sum_{m \neq h} u_m$. In these conditions, $\sum_{j \neq i} l'_j + u'_i = \sum_{j \neq i,h} l'_j + 1 - \sum_{m \neq h} u_m + u'_i = \sum_{j \neq i,h} l'_j - \sum_{j \neq i,h} u_j - u_i + u'_i + 1 = \sum_{j \neq i,h} (l'_j - u_j) + (u'_i - u_i) + 1 \leq 1$.
The proof of $\sum_{j \neq i} + u'_j + l'_i \geq 1 \; \forall i$ is similar □.

As the replacement of the original set of probability intervals L by the narrower set L' does not change the set $\mathcal{P}$ of possible probabilities, and L' constitutes a more accurate representation of these probabilities, we will perform the substitution in those cases where L does not satisfy the conditions (5), and thus we will always use reachable probability intervals.

For reachable sets of probability intervals we are guaranteed that the values $l(\{x_i\})$ and $u(\{x_i\})$ of the associated lower and upper probabilities, (l, u), coincide with the initial probability bounds l_i and u_i, as proposition 1 asserts. But what about the values of $l(.)$ and $u(.)$ for the other subsets of D_x which are not singletons? In other words, how can we calculate the values $l(A)$ and $u(A)$ for any subset A of D_x? In the next proposition we show the way in which these values can be easily calculated from the values l_i and u_i.

Proposition 4. Given a set of reachable probability intervals $L = \{[l_i, u_i], i = 1, \ldots, n\}$, the values of the lower and upper probability pair (l, u) associated to L

can be calculated by means of the following expressions:

$$l(A) = \sum_{x_i \in A} l_i \vee \left(1 - \sum_{x_i \notin A} u_i\right), \; u(A) = \sum_{x_i \in A} u_i \wedge \left(1 - \sum_{x_i \notin A} l_i\right), \; \forall A \subseteq D_x . \quad (9)$$

Proof. Let us first prove that $l(A) = (\sum_{x_i \in A} l_i) \vee (1 - \sum_{x_i \notin A} u_i)$. Taking into account that $l(A) = \min_{P \in \mathcal{P}} P(A) = \min_{P \in \mathcal{P}} \sum_{x_i \in A} p(x_i)$, it is very simple to check that $l(A) \geq (\sum_{x_i \in A} l_i) \vee (1 - \sum_{x_i \notin A} u_i)$.
Now, we are going to prove that the equality holds. We distinguish two cases:

1. Suppose that $\sum_{x_i \in A} l_i \geq 1 - \sum_{x_i \notin A} u_i$.

Let us define $\lambda = 1 - \sum_{x_i \in A} l_i$. We have $\sum_{x_i \notin A} l_i \leq \lambda \leq \sum_{x_i \notin A} u_i$. Then we can find numbers c_i such that $\sum_{x_i \notin A} c_i = \lambda$ and $l_i \leq c_i \leq u_i \; \forall x_i \notin$A. Therefore, if we define $p(x_i) = l_i \; \forall x_i \in A$, $p(x_i) = c_i \; \forall x_i \notin A$, we have a probability that belongs to $\mathcal{P}$ and such that $P(A) = \sum_{x_i \in A} p(x_i) = \sum_{x_i \in A} l_i$. In this case the equality holds.

2. Suppose now that $\sum_{x_i \in A} l_i \leq 1 - \sum_{x_i \notin A} u_i$.

Let us define $\lambda = 1 - \sum_{x_i \notin A} u_i$. In this case we have $\sum_{x_i \in A} l_i \leq \lambda \leq \sum_{x_i \in A} u_i$. So once again we can get numbers c_i such that $\sum_{x_i \in A} c_i = \lambda$ and $l_i \leq c_i \leq u_i \; \forall x_i \in A$. Therefore, by defining $p(x_i) = u_i \; \forall x_i \notin A$, $p(x_i) = c_i \; \forall x_i \in A$, we have a probability belonging to $\mathcal{P}$ and such that $P(A) = 1 - P(\overline{A}) = 1 - \sum_{x_i \notin A} u_i$. So, the equality is verified in this case too.
Finally, the expression for the upper measure $u(A)$ can be easily deduced by duality □.

For general lower and upper probability measures (and also for general fuzzy measures[12]), we need to give all the values $l(A)$ or $u(A)$ in order to have a complete specification of these measures, that is, we need $2^{|D_x|}$ values ($|D_x|$ stands for the cardinal of the set D_x). For several distinguished kinds of measures, such as probabilities or possibilities[13], it suffices to have the $|D_x|$ values of these measures for singletons, and the rest of the values may be calculated as

$$P(A) = \sum_{x_i \in A} p(x_i), \;\; \Pi(A) = \bigvee_{x_i \in A} \pi(x_i) , \quad (10)$$

for probabilities P and possibilities Π, respectively. The values $p(x_i)$ and $\pi(x_i)$, $i = 1, \ldots, n$, are called probability and possibility distributions respectively. For probability intervals, we need to specify only $2|D_x|$ values instead of $2^{|D_x|}$. Therefore we can consider $\{[l_i, u_i], \; i = 1, \ldots, n\}$ as the values of an 'interval probability distribution'. This fact makes probability intervals much easier to manage than lower and upper probability measures or even belief and plausibility functions.

As we have already mentioned, probability intervals can be considered as particular cases of lower and upper probability measures, where the restrictions that define the set of associated probabilities $\mathcal{P}$ only affect single values of probability. The next proposition shows that probability intervals always belong to a well-known subclass of lower and upper measures, namely, Choquet capacities of order two[14]. Remember that a pair of fuzzy measures (l, u) are Choquet capacities of order two (l is a 2-monotone capacity and u is 2-alternating capacity) if and only if

$$l(A \cup B) + l(A \cap B) \geq l(A) + l(B) \;\; \forall A \subseteq D_x\,,$$

$$u(A \cup B) + u(A \cap B) \leq u(A) + u(B) \;\; \forall A \subseteq D_x\,.$$

Moreover, it is known that pairs of Choquet capacities of order two are always lower and upper probability measures (see Campos[15] and Huber[8]).

Proposition 5. The lower and upper probability measures associated to a reachable set of probability intervals are always Choquet capacities of order two.

Proof. We will prove that $\forall A, C \subseteq D_x$ such that $A \cap C = \emptyset$, $\exists P \in \mathcal{P}$ such that

$$P(A) = l(A) \text{ and } P(A \cup C) = l(A \cup C)\,. \tag{11}$$

If this condition is true, then $\forall A, B \subseteq D_x$, is $A \cap B \subseteq A \cup B$ and therefore $\exists P \in \mathcal{P}$ such that $P(A \cap B) = l(A \cap B)$, $P(A \cup B) = l(A \cup B)$. So, we have $l(A \cup B) + l(A \cap B) = P(A \cup B) + P(A \cap B) = P(A) + P(B) \geq l(A) + l(B)$, and $l(.)$ is a 2-monotone capacity. Moreover, using the duality relation between l and u, we conclude that $u(.)$ is a 2-alternating capacity. So, if condition (11) were true, (l, u) would be Choquet capacities of order two.

Consider two sets A and C such that $A \cap C = \emptyset$. By proposition 4 we know that

$$l(A) = \sum_{x_i \in A} l_i \vee (1 - \sum_{x_i \notin A} u_i),\; l(A \cup C) = \sum_{x_i \in A \cup C} l_i \vee (1 - \sum_{x_i \notin A \cup C} u_i)\,.$$

In order to prove (11), we distinguish four cases, depending on the possible values for $l(A)$ and $l(A \cup C)$ (for the sake of simplicity, we will write $i \in A$ and $i \notin A$ instead of $x_i \in A$ and $x_i \notin A$, and similarly with $A \cup C$):

1. $l(A) = \sum_{i \in A} l_i \geq 1 - \sum_{i \notin A} u_i$ and $l(A \cup C) = 1 - \sum_{i \notin A \cup C} u_i \geq \sum_{i \in A \cup C} l_i$. In these conditions, let us define $\lambda = 1 - \sum_{i \in A} l_i - \sum_{i \notin A \cup C} u_i$. It is easy to check that $\sum_{i \in C} l_i \leq \lambda \leq \sum_{i \in C} u_i$. Then we can find numbers c_i $i \in C$ such that $\sum_{i \in C} c_i = \lambda$ and $l_i \leq c_i \leq u_i$ $\forall i \in C$. Thus, by defining $p(x_i) = l_i$ $i \in A$, $p(x_i) = u_i$ $i \notin A \cup C$, $p(x_i) = c_i$ $i \in C$, we have a probability that belongs to $\mathcal{P}$ and such that $P(A) = \sum_{i \in A} l_i = l(A)$, and $P(A \cup C) = P(A) + P(C) = \sum_{i \in A} l_i + \sum_{i \in C} c_i = \sum_{i \in A} l_i + \lambda = 1 - \sum_{i \notin A \cup C} u_i = l(A \cup C)$.

2. $l(A) = \sum_{i \in A} l_i \geq 1 - \sum_{i \notin A} u_i$ and $l(A \cup C) = \sum_{i \in A \cup C} l_i \geq 1 - \sum_{i \notin A \cup C} u_i$. In these conditions we have $\sum_{i \notin A \cup C} l_i \leq 1 - \sum_{i \in A \cup C} l_i \leq \sum_{i \notin A \cup C} u_i$. Therefore once again we find numbers c_i $i \notin A \cup C$ such that $\sum_{i \notin A \cup C} c_i = 1 - \sum_{i \in A \cup C} l_i$ and $l_i \leq c_i \leq u_i$ $\forall i \notin A \cup C$. Thus, by defining $p(x_i) = c_i$ $i \notin A \cup C$, $p(x_i) = l_i$ $i \in A \cup C$, once again we obtain a probability that belongs to $\mathcal{P}$ such that $P(A) = \sum_{i \in A} l_i = l(A)$ and $P(A \cup C) = \sum_{i \in A \cup C} l_i = l(A \cup C)$.

3. $l(A) = 1 - \sum_{i \notin A} u_i \geq \sum_{i \in A} l_i$ and $l(A \cup C) = \sum_{i \in A \cup C} l_i \geq 1 - \sum_{i \notin A \cup C} u_i$. In these conditions it can be seen that $p(x_i) = l_i \; i \in A$, $p(x_i) = u_i \; i \notin A \cup C$ and $p(x_i) = l_i = u_i \; i \in C$, defines a probability belonging to $\mathcal{P}$ such that $P(A) = \sum_{i \in A} l_i = 1 - \sum_{i \in C} l_i - \sum_{i \notin A \cup C} u_i = 1 - \sum_{i \in C} u_i + \sum_{i \notin A \cup C} u_i = 1 - \sum_{i \notin A} u_i = l(A)$, and $P(A \cup C) = \sum_{i \in A \cup C} l_i = l(A \cup C)$.

4. $l(A) = 1 - \sum_{i \notin A} u_i \geq \sum_{i \in A} l_i$ and $l(A \cup C) = 1 - \sum_{i \notin A \cup C} u_i \geq \sum_{i \in A \cup C} l_i$. In that case the inequalities $\sum_{i \in A} l_i \leq 1 - \sum_{i \notin A} u_i \leq \sum_{i \in A} u_i$ are true. Once again $\exists c_i \; i \in A$ such that $\sum_{i \in A} c_i = 1 - \sum_{i \notin A} u_i$ and $l_i \leq c_i \leq u_i \; \forall i \in A$. If we define $p(x_i) = c_i \; i \in A$, $p(x_i) = u_i \; i \notin A$, we get a probability belonging to $\mathcal{P}$ such that $P(A) = \sum_{i \in A} c_i = 1 - \sum_{i \notin A} u_i = l(A)$ and $P(A \cup C) = \sum_{i \in A} c_i + \sum_{i \in C} u_i = 1 - \sum_{i \notin A} u_i + \sum_{i \in C} u_i = 1 - \sum_{i \notin A \cup C} u_i = l(A \cup C)$.

Therefore, for the four cases we have proved (11). So, the proof is complete □.

To end this section, let us study how to obtain the extreme probabilities of the convex set $\mathcal{P}$ associated to a set of probability intervals L. This is interesting because these extreme probabilities provide an alternative representation for $\mathcal{P}$ (instead of the linear restrictions, $l_i \leq p(x_i) \leq u_i \; \forall i$, $\sum_i p(x_i) = 1$, that define $\mathcal{P}$). However, in general the representation of $\mathcal{P}$ by means of linear restrictions is more efficient than one based on extreme probabilities. The reason is that the number of extreme probabilities of the convex set $\mathcal{P}$ associated to a set of probability intervals can be very large: As indicated in Tessen[11], the maximum number $e(n)$ of extreme probabilities is

- $e(n) = \binom{n+1}{(n+1)/2} \frac{n+1}{4}$, if n is odd
- $e(n) = \binom{n+1}{n/2} \frac{n}{2}$, if n is even

For example, $e(10) = 1260$ and $e(11) = 2722$.

Nevertheless, in some cases it may be necessary to calculate the extreme probabilities. For example, in Cano[1], a method to propagate convex sets of probabilities in causal networks[16] is described. If we want to use it to propagate probability intervals, we must obtain the extreme probabilities.

As probability intervals are Choquet capacities of order two, then the method proposed in Campos[15] is able to give all the extreme probabilities. However this is a very inefficient method. A better alternative is the method suggested by Tessem[11]. We propose a recursive algorithm which is more efficient on average than the algorithm given by Tessem:

We maintain a global list *Prob* of the extreme probabilities found so far, and the current 'partial' probability P (this means a set of values p_i, $i = 1, \ldots, n$ verifying the restrictions $l_i \leq p_i \leq u_i \; \forall i$ but not necessarily the restriction $\sum_i p_i = 1$). We also use two local variables: a list *Expl* of explored indices and a real value λ. The initialization steps are:

- $Prob \leftarrow \emptyset$;
- $Expl \leftarrow \emptyset$;
- $\lambda \leftarrow 1 - \sum_i l_i$;
- For $i = 1$ to n do $p_i \leftarrow l_i$;

Then we call a recursive procedure *Getprob*(P, λ,*Expl*) that calculates and appends the extreme probabilities to *Prob*. *Getprob* is defined as follows:

```
Getprob(P, λ, Expl)
   For i = 1 to n do
      If not belong(i,Expl)
      then if λ ≤ u_i − l_i
            then
               v ← p_i;
               p_i ← p_i + λ;
               if not belong(P,Prob)
               then append(P,Prob);
               p_i ← v;
            else
               v ← p_i;
               p_i ← u_i;
               Getprob(P, λ − u_i + l_i, Expl ∪ {i});
               p_i ← v;
```

This algorithm uses an implicit tree search where each node is a partial probability and a child node represents a refinement of its parent node by increasing one component p_i. The leaf nodes of this tree are the extreme probabilities.

For example, for the set of probability intervals L defined on the set $D_x = \{x_1, x_2, x_3, x_4\}$, given by

$$L = \{[0, 0.3], [0.4, 0.5], [0.1, 0.5], [0.1, 0.4]\}$$

the extreme probabilities are

$$(0.3, 0.5, 0.1, 0.1),\ (0.3, 0.4, 0.2, 0.1),\ (0.3, 0.4, 0.1, 0.2),\ (0.0, 0.5, 0.4, 0.1),$$

$$(0.0, 0.5, 0.1, 0.4),\ (0.0, 0.4, 0.5, 0.1),\ (0.1, 0.4, 0.1, 0.4),\ (0.0, 0.4, 0.2, 0.4).$$

3. Inclusion and Combination of Probability Intervals

When dealing with uncertain information two important issues are precision of a piece of information and aggregation of several pieces of information. With respect to the first issue, we are going to study the concept of inclusion of probability intervals, which tries to clarify when a set of probability intervals is more precise or contains more information than another set. In relation to the issue of aggregation, we will study methods to combine two (or more) sets of probability intervals in conjunctive and disjunctive ways. To do that, we will take advantage of the interpretation of probability intervals as particular cases of lower and upper probability measures, because the concepts of inclusion and combination are defined within this theory[6,17,18].

3.1. *Inclusion of probability intervals*

Given two pairs of lower and upper probability measures (l_1, u_1) and (l_2, u_2), defined on the same domain D_x, (l_1, u_1) is said to be included in (l_2, u_2), and it is denoted by $(l_1, u_1) \subseteq (l_2, u_2)$, if and only if (see Campos[6,17] and Dubois[18])

$$[l_1(A), u_1(A)] \subseteq [l_2(A), u_2(A)],\ \forall A \subseteq D_x\,. \tag{12}$$

Because of the duality relation between l and u, (12) is equivalent to any of the following inequalities

$$l_1(A) \geq l_2(A)\ \ \forall A \subseteq D_x\,, \tag{13}$$

$$u_1(A) \leq u_2(A)\ \ \forall A \subseteq D_x\,. \tag{14}$$

Moreover, (12) is also equivalent to the inclusion of the set $\mathcal{P}_1$ of probabilities associated to (l_1, u_1) in the corresponding set $\mathcal{P}_2$ associated to (l_2, u_2), $\mathcal{P}_1 \subseteq \mathcal{P}_2$. Inclusion of (l_1, u_1) in (l_2, u_2) means that (l_1, u_1) is a more precise assessment of the information about the values of one variable than (l_2, u_2).

We will say that a set L of probability intervals is included in another set L' if the pair of lower and upper measures (l, u) associated to L is included in the corresponding pair (l', u') associated to L'. Let us see how the inclusion for probability intervals can be characterized:

Proposition 6. Let $L = \{[l_i, u_i], i = 1, \ldots, n\}, L' = \{[l'_i, u'_i], i = 1, \ldots, n\}$ be two sets of reachable probability intervals on the same domain D_x. Then L is included in L' if and only if

$$[l_i, u_i] \subseteq [l'_i, u'_i]\ \ \forall i = 1, \ldots, n\,, \tag{15}$$

or equivalently

$$l'_i \leq l_i \leq u_i \leq u'_i\ \ \forall i = 1, \ldots, n\,. \tag{16}$$

Proof. The result follows inmediately from proposition 4 and the monotonicity of the maximum operator □.

Therefore, as we could have expected, when checking the inclusion between two sets of probability intervals, only the single values l_i, l'_i, u_i and u'_i need to be considered.

3.2. *Combination of probability intervals*

With respect to the combination of lower and upper probability measures, the conjunctive and disjunctive combinations of these measures, corresponding to the logical operators 'and' and 'or' respectively, were defined by Campos[6,17]. The idea is simple: the relation of inclusion defines a partial order relation on the family of lower and upper probability pairs. The conjunction of two pairs (l, u) and (l', u'), denoted by $(l \otimes l', u \otimes u')$, is defined as the infimum of (l, u) and (l', u'), if common lower bounds exist, that is, it is the greatest pair included in both (l, u) and (l', u'). Analogously, the disjunction of (l, u) and (l', u'), denoted by $(l \oplus l', u \oplus u')$, is the supremum of (l, u) and (l', u'), the least pair including both (l, u) and (l', u'). The conjunction is the pair of lower and upper measures associated to the intersection $\mathcal{P} \cap \mathcal{P}'$ of the sets of probabilities $\mathcal{P}$ and $\mathcal{P}'$ associated to the initial lower and upper measures. Similarly, the disjunction is the pair associated to the set of probabilities $\mathcal{P} \cup \mathcal{P}'$.

The semantic of the conjunction and the disjunction is clear: the conjunction represents the conclusion we can obtain if we suppose that the two initial pieces of information are true; the disjunction is the one inferred if at least one piece of information is considered to be true.

The calculus of the disjunction $(l \oplus l', u \oplus u')$ is very easy: it can be shown (see Campos[6,17]) that

$$(l \oplus l')(A) = \min(l(A), l'(A)),\ (u \oplus u')(A) = \max(u(A), u'(A)),\ \forall A \subseteq D_x\,. \quad (17)$$

However, the calculus of the conjunction $(l \otimes l', u \otimes u')$ is not so easy. In general, we need to solve a linear programming problem for each value $(l \otimes l')(A)$ (and the values $(u \otimes u')(A)$ could be obtained by duality, see Campos[6] for details). Moreover the conjunction does not always exist. In these cases we say that the two pairs are not compatible: the information they represent cannot be simultaneously true. Clearly, compatibility holds if and only if the set $\mathcal{P} \cap \mathcal{P}'$ is not empty.

Now, we are in a position to define the combination of two sets of probability intervals as the combination of the associated lower and upper probability pairs. Next, we want to characterize compatibility and give specific formulas for the combination of probability intervals.

Proposition 7. Let $L = \{[l_i, u_i], i = 1, \ldots, n\}, L' = \{[l'_i, u'_i], i = 1, \ldots, n\}$ be two sets of reachable probability intervals on the same domain D_x. Then L and L' are compatible if and only if

$$l_i \leq u'_i \text{ and } l'_i \leq u_i\ \forall i = 1, \ldots, n, \text{ and } \sum_{i=1}^{n}(l_i \vee l'_i) \leq 1 \leq \sum_{i=1}^{n}(u_i \wedge u'_i)\,. \quad (18)$$

The proof of the result above is very easy. The following proposition shows that the conjunction of two sets of probability intervals is another set of probability intervals:

Proposition 8. Let $L = \{[l_i, u_i], i = 1, \ldots, n\}$, $L' = \{[l'_i, u'_i], i = 1, \ldots, n\}$ be two sets of reachable and compatible probability intervals on the same domain

D_x. Then their conjunction is the set of reachable probability intervals $L \otimes L' = \{[(l \otimes l')_i, (u \otimes u')_i],\ i = 1, \ldots, n\}$, where

$$(l \otimes l')_i = \max\{l_i, l'_i, 1 - \sum_{j \neq i} \min(u_j, u'_j)\}\,, \tag{19}$$

$$(u \otimes u')_i = \min\{u_i, u'_i, 1 - \sum_{j \neq i} \max(l_j, l'_j)\}\,. \tag{20}$$

Proof. The conjunction is the pair of lower and upper probabilities associated to the set of probabilities $\mathcal{P} \cap \mathcal{P}'$. Obviously, this set is

$$\mathcal{P} \cap \mathcal{P}' = \{P \in \mathcal{P}(D_x) \mid l_i \vee l'_i \leq p(x_i) \leq u_i \wedge u'_i,\ \forall i\}\,.$$

As $\mathcal{P} \cap \mathcal{P}'$ is defined by restrictions affecting only the individual probabilities $p(x_i)$, it is clear that $\{[l_i \vee l'_i, u_i \wedge u'_i], i = 1, \ldots, n\}$ is a set of probability intervals whose associated set of probabilities is just $\mathcal{P} \cap \mathcal{P}'$. Then by using propositions 2 and 3, the expressions for the equivalent but reachable probability intervals coincide with (19) and (20) □.

With respect to disjunction, although it is very easy to calculate, the problem is that this operation is not closed for probability intervals: the disjunction $L \oplus L'$ of two sets of probability intervals L and L' is always a pair of lower and upper probability measures, but it is not necessarily a set of probability intervals. Let us show this fact by means of the following example:

Example 1. Consider the following two sets of probability intervals (in fact, two single probabilities), defined on the domain $D_x = \{x_1, x_2, x_3, x_4\}$:

$$L = \{[l_1, u_1] = [0.3, 0.3], [l_2, u_2] = [0.4, 0.4], [l_3, u_3] = [0.2, 0.2], [l_4, u_4] = [0.1, 0.1]\}$$

$$L' = \{[l'_1, u'_1] = [0.0, 0.0], [l'_2, u'_2] = [0.1, 0.1], [l'_3, u'_3] = [0.5, 0.5], [l'_4, u'_4] = [0.4, 0.4]\}$$

According to (17), some of the values of $(l \oplus l')$ and $(u \oplus u')$ are:

- $(l \oplus l')_1 = 0.3 \wedge 0 = 0,\ (l \oplus l')_3 = 0.2 \wedge 0.5 = 0.2,$
- $(u \oplus u')_2 = 0.4 \vee 0.1 = 0.4,\ (u \oplus u')_4 = 0.1 \vee 0.4 = 0.4,$
- $(l \oplus l')(\{x_1, x_3\}) = 0.5 \wedge 0.5 = 0.5$

If $L \oplus L'$ were a set of probability intervals, then taking into account the result of proposition 4, we would have

$$(l \oplus l')(\{x_1, x_3\}) = ((l \oplus l')_1 + (l \oplus l')_3) \vee (1 - (u \oplus u')_2 - (u \oplus u')_4) = 0.2 \neq 0.5$$

So, in this case, $L \oplus L'$ cannot be a set of probability intervals □.

In order to get a set of probability intervals as the result of the disjunction of two sets of probability intervals, we can try to find the set of probability intervals that is the best approximation of $L \oplus L'$. So, we look for the set of probability intervals, say $(L \oplus L')^a$, such that, first $L \oplus L'$ is included in $(L \oplus L')^a$ (in order not to add information), and second every other set of probability intervals including $L \oplus L'$ must also include $(L \oplus L')^a$ (we should lose as little information as possible). The following proposition shows that we can always find a set of probability intervals verifying these conditions, and it gives its concrete expression too:

Proposition 9. Let $L = \{[l_i, u_i], i = 1, \ldots, n\}, L' = \{[l'_i, u'_i], i = 1, \ldots, n\}$ be two sets of reachable probability intervals on the same domain D_x, and let $L \oplus L'$ be their disjunction. Define the reachable set $(L \oplus L')^a$ of probability intervals as

$$(L \oplus L')^a = \{[l_i \wedge l'_i, u_i \vee u'_i],\ i = 1, \ldots, n\}. \tag{21}$$

Then $L \oplus L' \subseteq (L \oplus L')^a$ and for every other set of probability intervals L'' such that $L \oplus L' \subseteq L''$, we have $(L \oplus L')^a \subseteq L''$.

Proof. First, it is very easy to see that $(L \oplus L')^a$ verifies the conditions (5) that characterize reachability, provided that L and L' are reachable sets of probability intervals. Now, let us prove that $L \oplus L' \subseteq (L \oplus L')^a$: From (17), we know that $(l \oplus l')(A) = \min(l(A), l'(A))\, \forall A \subseteq D_x$; from (21) and (9) we deduce that $(l \oplus l')^a(A) = \sum_{i \in A}(l_i \wedge l'_i) \vee (1 - \sum_{i \notin A}(u_i \vee u'_i))$. Then we obtain $l(A) \geq (l \oplus l')^a(A)$ and $l'(A) \geq (l \oplus l')^a(A)$. Therefore $(l \oplus l')(A) \geq (l \oplus l')^a(A)\ \forall A \subseteq D_x$, and from (13) we conclude that $L \oplus L' \subseteq (L \oplus L')^a$.
Finally, let us prove that if L'' is a set of probability intervals such that $L \oplus L' \subseteq L''$, then $(L \oplus L')^a \subseteq L''$: the condition $L \oplus L' \subseteq L''$ means that $l''(A) \leq (l \oplus l')(A) \leq (u \oplus u')(A) \leq u''(A)$, $\forall A \subseteq D_x$. In particular, we have $l''_i \leq (l \oplus l')_i = l_i \wedge l'_i \leq u_i \vee u'_i = (u \oplus u')_i \leq u''_i\ \forall i$. But $(l \oplus l')^a_i = l_i \wedge l'_i$ and $(u \oplus u')^a_i = u_i \vee u'_i$, and therefore we have $l''_i \leq (l \oplus l')^a_i \leq (u \oplus u')^a_i \leq u''_i\ \forall i$. From proposition 6, this is equivalent to the inclusion of $(L \oplus L')^a$ in L'' $\Box$.

From proposition 9, if we want to have a disjunctive combination closed for probability intervals, the best choice is to define it as $(L \oplus L')^a$ in (21).

4. Marginalization and Conditioning of Probability Intervals

In most of the problems, our interest is not usually restricted to only one variable since we deal with several variables defined on different domains exhibiting some relationship among each other. In these cases we have one joint piece of information on the set of variables (or a number of pieces of information relative to several subsets of variables). In such situations, we need a tool to obtain information on one variable or a subset of variables from the joint information. Such a tool is the marginalization operator. Moreover, it is also necessary to have a mechanism available to update our information about one or several variables once we know for sure the values taken by other variables. This is a conditioning operator. In this section we define and study the concepts of marginalization and conditioning for probability intervals. We will study the simple case in which we have only two variables, but the generalization to deal with more variables is straightforward.

So, let us consider two variables X and Y taking values in the sets $D_x = \{x_1, x_2, \ldots, x_n\}$ and $D_y = \{y_1, y_2, \ldots, y_m\}$ respectively, and a set of reachable bidimensional probability intervals $L = \{[l_{ij}, u_{ij}], i = 1, \ldots, n, j = 1, \ldots, m\}$, defined on the cartesian product $D_x \text{x} D_y$, representing the joint available information on these two variables.

4.1. *Marginalization of probability intervals*

First, we want to define the marginals of these bidimensional probability intervals. To do that, we can use the interpretation of a set of probability intervals as a pair of lower and upper probabilities (l, u). Given (l, u), the marginal measures (l_x, u_x) on D_x (for the marginals on D_y is analogous) are defined[19,20] as:

$$l_x(A) = l(A \text{x} D_y), \;\; u_x(A) = u(A \text{x} D_y), \; \forall A \subseteq D_x \,. \tag{22}$$

This definition, which obviously reproduces the usual definition of marginalization for probability measures, also preserves duality between l_x and u_x. Moreover, it can be proved that marginalization, as defined above, is a closed operation for most of the subclasses of lower and upper probability measures (necessities/possibilities, belief/plausibility functions, Choquet capacities of order two,...), that is, the marginal measures belong to the same subclass as the bidimensional measures[19,20].

Alternatively, we could use the interpretation of probability intervals as convex sets of probabilities, and define the marginal of L on D_x as the set $\mathcal{P}_x$ of marginal probabilities of the probabilities in the convex set $\mathcal{P}$, being $\mathcal{P}$ the set of probabilities associated to L, that is to say,

$$\mathcal{P}_x = \{P \in \mathcal{P}(D_x) \,|\, \exists Q \in \mathcal{P} \text{ such that } p(x_i) = \sum_{j=1}^{m} q(x_i, y_j) \; \forall i\} \,. \tag{23}$$

Fortunately, both definitions are equivalent, in the sense that $\mathcal{P}_x$ is precisely the set of probabilities associated to (l_x, u_x), as the following proposition asserts.

Proposition 10. Given a set $L = \{[l_{ij}, u_{ij}], i = 1, \ldots, n, j = 1, \ldots, m\}$ of reachable bidimensional probability intervals, the corresponding convex set of probabilities $\mathcal{P}$ and the lower and upper probability pair (l, u) associated to L, then the marginal measures (l_x, u_x) defined in (22) and the set of probabilities $\mathcal{P}_x$ defined in (23) verify the following relation:

$$l_x(A) = \min_{P \in \mathcal{P}_x} P(A), \;\; u_x(A) = \max_{P \in \mathcal{P}_x} P(A), \; \forall A \subseteq D_x \,. \tag{24}$$

The proof is very simple and hence we have not included it. Proposition 10 shows that we can define the marginals of a set of probability intervals consistently with the two interpretations of probability intervals. Moreover, it can be proved that these marginals are, in fact, probability intervals:

Proposition 11. Let $L = \{[l_{ij}, u_{ij}], i = 1, \ldots, n, j = 1, \ldots, m\}$ be a set of reachable bidimensional probability intervals. Then the marginal lower and upper measures (l_x, u_x) defined in (22) are associated to the set of reachable probability intervals $L_x = \{[l_i, u_i], i = 1, \ldots, n\}$, defined as follows:

$$l_i = \sum_{j=1}^{m} l_{ij} \vee (1 - \sum_{k \neq i} \sum_{j=1}^{m} u_{kj}),\ i = 1, \ldots, n\,, \tag{25}$$

$$u_i = \sum_{j=1}^{m} u_{ij} \wedge (1 - \sum_{k \neq i} \sum_{j=1}^{m} l_{kj}),\ i = 1, \ldots, n\,. \tag{26}$$

Proof. First, it is very easy to see that the set $\mathcal{P}_x$ of probabilities associated to the marginal measures (l_x, u_x) of L is

$$\mathcal{P}_x = \{P \in \mathcal{P}(D_x) \mid \sum_{j=1}^{m} l_{ij} \leq p(x_i) \leq \sum_{j=1}^{m} u_{ij},\ \forall i\}\,.$$

So, $\mathcal{P}_x$ is defined by restrictions affecting only the single values of probability $p(x_i)$. Therefore $\mathcal{P}_x$ is associated to the set of probability intervals $\{[\sum_j l_{ij}, \sum_j u_{ij}], i = 1, \ldots, n\}$. Now, using (8), the equivalent but reachable intervals are precisely those defined in (25) and (26) $\square$.

Observe that the calculus of the marginal probability intervals on one variable is very easy: we simply sum the values l_{ij} and u_{ij} on the other variable; then the equivalent but reachable intervals are obtained using the formulas (25) and (26). If we want to calculate the values of the marginal measures l_x and u_x for subsets which are different from singletons, they can be obtained using the result of proposition 4.

4.2. *Conditioning of probability intervals*

In order to define the conditioning of probability intervals we will once again use their interpretation as lower and upper probabilities, because several definitions of conditioning in this framework are available (see Moral[21] for a review). We will use the following definition of conditioning[22,23,24]: Given a pair of lower and upper probabilities (l, u) defined on a domain D, and a subset $B \subseteq D$, the conditional lower and upper measures given that we know B, $(l(.|B), u(.|B))$ are defined as

$$l(A|B) = \frac{l(A \cap B)}{l(A \cap B) + u(\overline{A} \cap B)},\ \ u(A|B) = \frac{u(A \cap B)}{u(A \cap B) + l(\overline{A} \cap B)},\ \forall A \subseteq D\,. \tag{27}$$

In our case, we have a set of bidimensional probability intervals $L = \{[l_{ij}, u_{ij}], i = 1, \ldots, n, j = 1, \ldots, m\}$, and we want to calculate the conditional probability intervals for one variable, say X, given that we know the value of the other variable,

$Y = y_j$. Then the previous expressions (27) become

$$l_{i|j} = l(x_i|y_j) = l(\{x_i\}\mathrm{x}D_y|D_x\mathrm{x}\{y_j\}) = \frac{l(\{(x_i, y_j)\})}{l(\{(x_i, y_j)\}) + u((D_x - \{x_i\})\mathrm{x}\{y_j\})},$$

$$u_{i|j} = u(x_i|y_j) = u(\{x_i\}\mathrm{x}D_y|D_x\mathrm{x}\{y_j\}) = \frac{u(\{(x_i, y_j)\})}{u(\{(x_i, y_j)\}) + l((D_x - \{x_i\})\mathrm{x}\{y_j\})}.$$

Taking into account the expressions for the lower and upper measures associated to a set of probability intervals given in the proposition 4, the set of conditional probability intervals on X given $Y = y_j$ is

$$L(X|Y = y_j) = \{[l_{i|j}, u_{i|j}], i = 1, \ldots, n\},$$

where

$$l_{i|j} = \frac{l_{ij}}{l_{ij} + (\sum_{k\neq i} u_{kj} \wedge (1 - \sum_k \sum_{h\neq j} l_{kh} - l_{ij}))}, \tag{28}$$

$$u_{i|j} = \frac{u_{ij}}{u_{ij} + (\sum_{k\neq i} l_{kj} \vee (1 - \sum_k \sum_{h\neq j} u_{kh} - u_{ij}))}. \tag{29}$$

If we define $L_{\bullet\bullet}$, $U_{\bullet\bullet}$, $L_{k\bullet}$, $L_{\bullet h}$, $U_{k\bullet}$, $U_{\bullet h}$ by means of the following expressions:

$$L_{\bullet\bullet} = \sum_{k=1}^{n}\sum_{h=1}^{m} l_{kh},\ L_{k\bullet} = \sum_{h=1}^{m} l_{kh},\ k = 1, \ldots, n,\ L_{\bullet h} = \sum_{k=1}^{n} l_{kh},\ h = 1, \ldots, m,$$

$$U_{\bullet\bullet} = \sum_{k=1}^{n}\sum_{h=1}^{m} u_{kh},\ U_{k\bullet} = \sum_{h=1}^{m} u_{kh},\ k = 1, \ldots, n,\ U_{\bullet h} = \sum_{k=1}^{n} u_{kh},\ h = 1, \ldots, m,$$

then the conditional probability intervals $[l_{i|j}, u_{i|j}]$ can also be expressed as

$$l_{i|j} = \frac{l_{ij}}{(U_{\bullet j} - (u_{ij} - l_{ij})) \wedge (1 + L_{\bullet j} - L_{\bullet\bullet})}, \tag{30}$$

$$u_{i|j} = \frac{u_{ij}}{(L_{\bullet j} + (u_{ij} - l_{ij})) \vee (1 + U_{\bullet j} - U_{\bullet\bullet})}. \tag{31}$$

Note that the calculus of the conditional probability intervals is very simple. Moreover, as the next proposition proves, these intervals are always reachable, and therefore it is not necessary to transform them in reachable intervals by using propositions 2 and 3.

Proposition 12. Given a set $L = \{[l_{ij}, u_{ij}], i = 1, \ldots, n, j = 1, \ldots, m\}$ of reachable bidimensional probability intervals, then for each $j = 1, \ldots, m$, the set of conditional probability intervals $L(X|Y = y_j)$ is always reachable.

Proof. Let us denote by $\mathcal{P}(X|j)$ the set of probabilities associated to the conditional probability intervals $L(X|Y = y_j)$ given in (28) and (29), that is

$$\mathcal{P}(X|j) = \{P \in \mathcal{P}(D_x) \,|\, l_{i|j} \leq p(x_i) \leq u_{i|j}\ \forall i\}.$$

Then, as we did in proposition 1, to prove the reachability it suffices to prove that for each i there exist probabilities P^i and Q^i that belong to $\mathcal{P}(X|j)$ whose values for the singleton $\{x_i\}$ coincide with $l_{i|j}$ and $u_{i|j}$ respectively, that is

$$p^i(x_i) = l_{i|j} \text{ and } l_{k|j} \leq p^i(x_k) \leq u_{k|j} \; \forall k \neq i\,,$$

$$q^i(x_i) = u_{i|j} \text{ and } l_{k|j} \leq q^i(x_k) \leq u_{k|j} \; \forall k \neq i\,.$$

We will only prove the first condition; the proof for the second one is analogous. The proof is based on the following result about Choquet capacities of order two (see Campos[22]): If (l, u) is a pair of Choquet capacities of order two, $\mathcal{P}$ being their set of associated probabilities, then the conditional measures defined in (27) can be written as

$$l(A|B) = \min_{P \in \mathcal{P}} P(A|B),\; u(A|B) = \max_{P \in \mathcal{P}} P(A|B),\; \forall A,\; \forall B\,.$$

As the measures associated to a set of probability intervals, according to proposition 5, are always Choquet capacities of order two, then the result above can be applied. So,

$$l_{k|j} = \min_{P \in \mathcal{P}} P(x_k|y_j) \leq P(x_k|y_j) \leq \max_{P \in \mathcal{P}} P(x_k|y_j) = u_{k|j},\; \forall P \in \mathcal{P},\; \forall k, j\,.$$

Then given i, there exists a probability P that belongs to $\mathcal{P}$ such that $l_{i|j} = P(x_i|y_j)$. The conditional probability $P(.|y_j)$ is precisely the probability P^i that we are looking for $\Box$.

To end this section, let us consider the following simple example, which illustrates some of the concepts studied:

Example 2. We are performing a study in a car factory. Our objective is to know the production rates of vehicles classified in two classes, say Motor (Hp 90, 115) and Model (Md Alpha, Beta). In order to know the exact production rates, we decided to ask the Production Manager, but unfortunately he was on holiday. So, we asked a member of his team, Mr.XX. He did not have exact information about the production rates, and answered our question in the following terms: 'The production rate for a 90 Hp Alpha model is between 30% and 40%, and no more than 20% for the 90 Hp Beta model. For the 115 Hp Alpha model the production rate is exactly 20%, and between 30% and 50% for the 115 Hp Beta model'. This information can be represented as the following set of reachable bidimensional probability intervals:

	Md Alpha	Md Beta
Hp 90	[0.3,0.4]	[0.0,0.2]
Hp 115	[0.2,0.2]	[0.3,0.5]

If we want to obtain information about either Motor or Model, we must marginalize. By using (25) and (26), the marginals are

Hp 90	Hp 115
[0.3,0.5]	[0.5,0.7]

Md Alpha	Md Beta
[0.5,0.6]	[0.4,0.5]

We decided to extend our study by asking the staff in the assembly chain, aiming to improve our information. We selected a person in the Engine section, Mr.YY, who told us that 'the production rate for the 115 Hp cars is at least 60%'. Mr.ZZ, from the Equipment section, told us that 'the rates for both the Alpha and Beta models are between 40% and 60%'.

Mr.YY and Mr.ZZ's answers are represented as two sets of probability intervals, as follows:

Hp 90	Hp 115
[0.0,0.4]	[0.6,1]

Md Alpha	Md Beta
[0.4,0.6]	[0.4,0.6]

In order to refine our knowledge about Motor, we may combine Mr.XX's marginal information about Motor with Mr.YY's information using the conjunctive operator. The result, using (19) and (20), is

Hp 90	Hp 115
[0.3,0.4]	[0.6,0.7]

that is, between 30% and 40% of the cars are equipped with the 90 Hp engine and the cars with the 115 Hp engine represent between 60% and 70% of the production. Another way of expressing this information is the following: 30% of the cars are equipped with the 90 Hp engine, 60% are equipped with the 115 Hp engine, and we are unsure for the remaining 10%: they could be equipped either with 90 Hp or 115 Hp engines.

With regards to the Model, we can also combine Mr.XX's marginal information about Model with Mr.ZZ's information. However, in this case Mr.ZZ's answer does not provide new information (it includes the other information) and therefore the combination does not change the information provided by Mr.XX.

Finally, if we want to obtain information about the proportion of cars of a given model that equip the two possible engines, we calculate the conditional probability intervals about Motor given Model. They are

Model =	Alpha
Hp 90	Hp 115
[0.6,0.67]	[0.33,0.4]

Model =	Beta
Hp 90	Hp 115
[0,0.4]	[0.6,1]

that is, for Alpha models, 60% use the 90 Hp engine, 33% use the 115 Hp engine and 7% could use the two engines; 60% of the Beta models use the 115 Hp engine and we do not know what happens with the remaining 40% □.

5. Integration with Respect to Probability Intervals

In probability theory, the concept of mathematical expectation or integral with respect to a probability measure plays an important role from both a theoretical and a practical point of view. Indeed, integration is useful, for example, to derive the probability of an event A, $P(A)$, from the conditional probabilities $P(A|B_i)$ of that event given a set of mutually exhaustive and exclusive events $B_1, \ldots, B_m$, and the probabilities of these events $P(B_i)$. Concepts such as the entropy of a probability distribution, or the quantity of information about one variable that another variable contains are defined with the help of an integral. Basically, an integral, with respect to a probability measure, is a tool able to summarize all the pieces of information provided by a function in a single value; this value is a kind of average of the function in terms of the probability measure. Integration is also essential in decision-making problems with uncertainty. The following two simple examples illustrate this point:

Example 3. Suppose we can choose to participate in one of two different lotteries. The two lotteries have three possible outcomes x_1, x_2 or x_3. If we choose x_i and this is the outcome of the lottery, we obtain a prize. The prizes are the same for each lottery, and they are 10$ for x_1, 5$ for x_2 and 20$ for x_3. The probabilities of each outcome for each lottery are:

Lottery 1			Lottery 2		
$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_1)$	$p(x_2)$	$p(x_3)$
0.75	0.15	0.1	0.4	0.4	0.2

Which lottery is chosen? Using several assumptions about what is 'rational behavior', we should choose the lottery that produces a better prize on average. So, if we calculate the expected prize EP for each lottery (as the integral of the prize function with respect to each probability), we get

- $EP(\text{lottery 1}) = 10.25$
- $EP(\text{lottery 2}) = 10$

Therefore lottery 1 provides a better expected prize and we should choose it $\square$.

Example 4. After doing a number of tests, it is determined that the illness a given patient is suffering is one among three possible illnesses, say x_1, x_2 or x_3, with probabilities 0.6, 0.3 and 0.1 respectively. Once the doctor chooses a diagnostic, he applies the appropriate treatment. However, to wrongly diagnose an illness has a cost that depends on both the true illness that the patient is suffering and the diagnosed one. These costs are reflected in the following table:

		true illness		
	c_{ij}	x_1	x_2	x_3
diagnosed illness	x_1	0	60	100
	x_2	30	0	90
	x_3	40	50	0

Which diagnostic and subsequent treatment should the doctor choose? Having probabilistic information about the true illness, the 'rational' choice is the one that minimizes the average cost. So, if we select illness x_i, the average cost of this choice is $C(x_i) = p_1 * c_{i1} + p_2 * c_{i2} + p_3 * c_{i3}$, that is, it is the mathematical expectation of the cost function corresponding to the choice x_i with respect to the probability of illness. In our case $C(x_1) = 28$, $C(x_2) = 27$ and $C(x_3) = 39$, and therefore the best diagnostic is illness x_2 $\square$.

In this section we are going to study the concept of integration when the underlying uncertainty measure is not a probability measure but a set of probability intervals. So, in the previous examples, even if we do not have purely probabilistic information, we are able to perform comparisons and make decisions, on the basis of the 'average behavior'.

As is usual in this paper, we will use the interpretation of probability intervals as particular cases of pairs of lower and upper probability measures, which in turn are particular cases of fuzzy measures, since we have several methods of integration for fuzzy measures available (fuzzy integrals). The two main fuzzy integrals are the Sugeno integral[12] and the Choquet integral[14]. We will use the Choquet integral, because it is closer in spirit to the mathematical expectation than the Sugeno integral, and therefore it seems to us to be more appropriate for probability intervals. Moreover, the Choquet integral can be defined for any real-valued function whereas the Sugeno integral is only defined for functions taking values in the interval [0,1] (see Campos[25,26] for an in-depth study of Choquet and Sugeno integrals).

In our case, we have a set L of probability intervals, and the associated pair of lower and upper measures (l, u). So, we can define the Choquet integral with respect to the two fuzzy measures $l(.)$ or $u(.)$. We will denote them as the lower $E_l(h)$ and the upper $E_u(h)$ Choquet integrals, and they form an interval $[E_l(h), E_u(h)]$. This interpretation as an interval is justified by the following equalities (which are true for Choquet capacities of order two[14,27], that relate the values $E_l(h)$ and $E_u(h)$ with the integrals $E_P(h)$ with respect to probabilities P that belong to the set $\mathcal{P}$ associated to L:

$$E_l(h) = \min_{P \in \mathcal{P}} E_P(h), \; E_u(h) = \max_{P \in \mathcal{P}} E_P(h) . \tag{32}$$

It is easy to see that the specific expressions for $E_l(h)$ and $E_u(h)$ for the particular case of reachable probability intervals are the following:

$$E_l(h) = \sum_{i=1}^{n} p_i h(x_i) , \tag{33}$$

$$E_u(h) = \sum_{i=1}^{n} q_i h(x_i) , \tag{34}$$

where:
$h : D_x \rightarrow \Re^+$ is a real function such that $h(x_1) \leq h(x_2) \leq \ldots \leq h(x_n)$,

$(p_1, p_2, \ldots, p_n) = (u_1, u_2, \ldots, u_{k-1}, 1 - L_{k+1} - U_{k-1}, l_{k+1}, \ldots, l_n)$ and k is the index such that $l_k \leq 1 - L_{k+1} - U_{k-1} \leq u_k$, and $L_i = \sum_{j=i}^{n} l_j$, $U_i = \sum_{j=1}^{i} u_j$ $\forall i$,
$(q_1, q_2, \ldots, q_n) = (l_1, l_2, \ldots, l_{h-1}, 1 - L^{h-1} - U^{h+1}, u_{h+1}, \ldots, u_n)$, and h is the index such that $l_h \leq 1 - L^{h-1} - U^{h+1} \leq u_h$, and $L^i = \sum_{j=1}^{i} l_j$, $U^i = \sum_{j=i}^{n} u_j$ $\forall i$.

An easy algorithm to calculate the weights p_i in (33) is the following:

$S \leftarrow 0$;
For $i = 1$ to $n - 1$ do $S \leftarrow S + u_i$;
$S \leftarrow S + l_n$;
$k \leftarrow n$;
While $S \geq 1$ do $S \leftarrow S - u_{k-1} + l_{k-1}$; $p_k \leftarrow l_k$; $k \leftarrow k - 1$;
For $i = 1$ to $k - 1$ do $p_i \leftarrow u_i$;
$p_k \leftarrow 1 - S + l_k$; $\square$.

The analogous algorithm to obtain the weights q_i in (34) is:

$S \leftarrow 0$;
For $i = 1$ to $n - 1$ do $S \leftarrow S + l_i$;
$S \leftarrow S + u_n$;
$k \leftarrow n$;
While $S \leq 1$ do $S \leftarrow S + u_{k-1} - l_{k-1}$; $p_k \leftarrow u_k$; $k \leftarrow k - 1$;
For $i = 1$ to $k - 1$ do $p_i \leftarrow l_i$;
$p_k \leftarrow 1 - S + u_k$; $\square$.

To end this section, let us consider modified versions of the examples 3 and 4 (see Bolaños[28] for a study of decision-making problems within the theory of evidence, and Loui[29], Wakker[30] for other approaches):

Example 5. Consider the same situation of example 3, but now the information about the chances of lottery 2 is not completely precise: All we know about lottery 2 is the following set of probability intervals:

$[l_1, u_1]$	$[l_2, u_2]$	$[l_3, u_3]$
[0.2,0.4]	[0.4,0.6]	[0.1,0.2]

Then, by calculating the interval of expected prize for lottery 2, using (33) and (34), we get the interval [8,10]. Therefore, we still prefer lottery 1, which gives an expected prize of 10.25 $\square$.

Example 6. Let us suppose that in example 4 the information about the three possible illnesses is not a probability, but the following set of probability intervals:

Illness	x_1	x_2	x_3
$[l_i, u_i]$	$[0.5, 0.7]$	$[0.2, 0.4]$	$[0.1, 0.2]$

Then, if we calculate the intervals of expected cost for each choice, we obtain:

$$C(x_1) = [22, 38],\ C(x_2) = [24, 36],\ C(x_3) = [34, 40].$$

Although we can not extract a definitive conclusion, it is clear that the worst diagnostic is x_3. Between x_1 and x_2, perhaps the choice depends on the decision-maker's attitude to risk: An optimistic person would possibly prefer x_1 because this choice guarantees a lower expected cost lesser than x_2. However, a pessimistic person would prefer x_2 since x_2 gives an upper expected cost lesser than x_1. Obviously, other intermediate criteria are also possible $\square$.

6. Probability Intervals and Belief/Plausibility Functions

Belief and plausibility functions constitute an interesting formalism for representing uncertainty. Despite their different interpretations[31,32,33], here we consider them in a formal sense as Choquet capacities of infinite order. So, they are also Choquet capacities of order two. We already know that probability intervals are also Choquet capacities of order two. However, in general, probability intervals are not belief/plausibility functions. As belief and plausibility functions, although easier to manage than general lower and upper probabilities or Choquet capacities of order two, they require more complex processing than probability intervals, the first problem that we consider in this section is that of approximating belief and plausibility functions by probability intervals.

So, given a pair (bel, Pl) of belief and plausibility functions, we look for the set of probability intervals L^e, such that (bel, Pl) is included in L^e and every other set of probability intervals L including (bel, Pl) must also include L^e, that is:

$$\text{Find } L^e \text{ such that} \tag{35}$$

1. $(bel, Pl) \subseteq L^e$, and
2. $\forall L$ such that $(bel, Pl) \subseteq L$ then $L^e \subseteq L$.

The solution to this problem is very easy, as the next proposition shows:

Proposition 13. The solution to the problem (35), which is the best probability interval approximation L^e of a belief/plausibility pair (bel, Pl) is $L^e = \{[l_i^e, u_i^e], i = 1, \ldots, n\}$, where

$$l_i^e = bel(x_i),\ u_i^e = Pl(x_i),\ \forall i = 1, \ldots, n. \tag{36}$$

Proof. Let m be the basic probability assigment (b.p.a) associated to (bel, Pl), that is to say, $bel(A) = \sum_{B\subseteq A} m(B)$ and $Pl(A) = \sum_{B\cap A\neq\emptyset} m(B)$.
According to (13), to prove the inclusion of (bel, Pl) in L^e we must prove that $l^e(A) \leq bel(A)\ \forall A$. From proposition 4 we know that $l^e(A) = \sum_{x_i\in A} l_i^e \vee (1 - \sum_{x_i\notin A} u_i^e)$.
As $\sum_{x_i\in A} l_i^e = \sum_{x_i\in A} bel(x_i) = \sum_{x_i\in A} m(x_i) \leq \sum_{B\subseteq A} m(B) = bel(A)$ and
$1 - bel(A) = Pl(\overline{A}) = \sum_{B\cap\overline{A}\neq\emptyset} m(B) \leq \sum_{x_i\notin A}\sum_{B\supseteq\{x_i\}} m(B) = \sum_{x_i\notin A} Pl(x_i) = \sum_{x_i\notin A} u^{e_i}$.

then $l^e(A) \leq bel(A)$, and (bel, Pl) is included in L^e.
Now, let us suppose that L is a set of probability intervals that includes (bel, Pl). Then $l(A) \leq bel(A) \leq Pl(A) \leq u(A)\ \forall A$. In particular we have $l_i \leq bel(x_i) = l_i^e \leq u_i^e = Pl(x_i) \leq u_i\ \forall i$, and from proposition 6 this means that L^e is included in L □.

Remark: If we consider a pair (l, u) of lower and upper probabilities instead of a pair (bel, Pl), its approximation by probability intervals is the same: $l_i^e = l(x_i), u_i^e = u(x_i)\ \forall i$ □.

Now, let us consider a different problem, whilst still relating belief and plausibility functions and probability intervals: If we have a set L of probability intervals, can we find a belief/plausibility pair whose values for the singleton subsets coincide with the values of L? In other words, we must look for the conditions that a set of probability intervals must verify in order to be considered as a partial specification of a belief/plausibility pair. This problem was solved by Lemmer and Kyburg[34], who found a necessary and sufficient condition. Their result, adapted to our notation, is the following:

Proposition 14 [*Lemmer and Kyburg 1991*]. Given a set $L = \{[l_i, u_i], i = 1, \ldots, n\}$ of probability intervals, we can find a pair (bel_L, Pl_L) of belief and plausibility functions such that

$$bel_L(x_i) = l_i \text{ and } Pl_L(x_i) = u_i,\ \forall i = 1, \ldots, n\,, \tag{37}$$

if and only if the following three conditions are verified:

$$\sum_{i=1}^{n} l_i \leq 1\,, \tag{38}$$

$$\sum_{j \neq i} l_j + u_i \leq 1\ \ \forall i\,, \tag{39}$$

$$\sum_{i=1}^{n} l_i + \sum_{i=1}^{n} u_i \geq 2\,. \tag{40}$$

Moreover, Lemmer and Kyburg proposed an algorithm that constructs the b.p.a. corresponding to bel_L and Pl_L whenever the three conditions are met (however, in general there exist several pairs (bel, Pl) verifying (37), and the pair (bel_L, Pl_L) obtained by using the algorithm in Lemmer and Kyburg[34] is not necessarily the least specific). In our case, the first two conditions are always verified, because we are considering proper and reachable sets of probability intervals. The only condition that we need to check is the third one.

The problem that remains to be considered is the following: if condition (40) is not verified for a set of probability intervals L, then we can not consider L as a partial specification of any belief/plausibility pair. In that case, it makes sense to look for another set of probability intervals L^m satisfying (40) which is an approximation of L. In one sense, this is the reverse of the problem of approximating belief and plausibility functions by probability intervals, because once L^m is obtained, we can use the algorithm given by Lemmer and Kyburg[34] to obtain belief and plausibility functions that constitute an approximation of the original set L.

So, given a set $L = \{[l_i, u_i], i = 1, \ldots, n\}$ of probability intervals that does not satisfy (40), we look for another set of probability intervals that includes L, verifying (40), and included in every other set of probability intervals including L and verifying (40). This set would be the minimum (in the sense of the inclusion relation) of all the intervals including L that satisfy (40).

Unfortunately, it is not generally possible to find such a minimum set, but only several minimal sets, that is to say, sets of probability intervals $L^m = \{[l_i^m, u_i^m], i = 1, \ldots, n\}$ verifying:

$$L \subseteq L^m\,,$$

$$\sum_{i=1}^n l_i^m + \sum_{i=1}^n u_i^m \geq 2\,, \tag{41}$$

$$\text{There is no } L' \neq L^m \text{ that satisfies (40) and } L \subseteq L' \subseteq L^m\,.$$

The following proposition characterizes these minimal sets of probability intervals:

Proposition 15. Let $L = \{[l_i, u_i], i = 1, .., n\}$ be a set of reachable probability intervals such that

$$\sum_{i=1}^n l_i + \sum_{i=1}^n u_i < 2\,.$$

Then every set of probability intervals $L^m = \{[l_i^m, u_i^m], i = 1, .., n\}$ verifying

$$l_i^m = l_i,\ \forall i\,,$$

$$u_i^m \geq u_i,\ \forall i\,, \tag{42}$$

$$\sum_{i=1}^n l_i^m + \sum_{i=1}^n u_i^m = 2\,,$$

is minimal, that is, it verifies (41). The converse is also true.

Proof. We are going to prove the equivalence between (42) and (41):
From $l_i^m = l_i$ and $u_i^m \geq u_i\ \forall i$, it is obvious that $L \subseteq L^m$. The condition $\sum_{i=1}^n l_i^m + \sum_{i=1}^n u_i^m \geq 2$ is also clear. Finally, if $L' \neq L^m$ is such that $L \subseteq L' \subseteq L^m$ then $l_i = l_i^m = l_i'$ and $u_i \leq u_i' \leq u_i^m\ \forall i$, but $u_k' < u_k^m$ for some k. In these conditions $\sum_{i=1}^n l_i' + \sum_{i=1}^n u_i' = \sum_{i=1}^n l_i + \sum_{i=1}^n u_i' < \sum_{i=1}^n l_i + \sum_{i=1}^n u_i^m = 2$, and L' does not satisfy (40). So we have proved that (42) implies (41).

On the other hand, from $L \subseteq L^m$ we have $l_i^m \leq l_i$ and $u_i^m \geq u_i\ \forall i$. If $l_j^m < l_j$ for some j, then define L' by means of $l_i' = l_i^m\ \forall i \neq j$, $l_j' = l_j$, $u_i' = u_i^m\ \forall i$. In these conditions we have $L' \neq L^m$, $L \subseteq L' \subseteq L^m$ but $\sum_{i=1}^n l_i' + \sum_{i=1}^n u_i' = \sum_{i=1}^n l_i' + \sum_{i=1}^n u_i^m > \sum_{i=1}^n l_i^m + \sum_{i=1}^n u_i^m \geq 2$. So, L' satisfies (40), in contradiction with the hypothesis. Therefore $l_i^m = l_i\ \forall i$.
Finally if $\sum_{i=1}^n l_i^m + \sum_{i=1}^n u_i^m > 2$ then $\sum_{i=1}^n l_i + \sum_{i=1}^n u_i^m > 2 > \sum_{i=1}^n l_i + \sum_{i=1}^n u_i$. Thus, $\sum_{i=1}^n u_i^m > 2 - \sum_{i=1}^n l_i > \sum_{i=1}^n u_i$. We can get numbers c_i such that $u_i \leq c_i \leq u_i^m\ \forall i$ and $\sum_{i=1}^n c_i = 2 - \sum_{i=1}^n l_i$. Then L', defined as $l_i' = l_i$ and $u_i' = c_i\ \forall i$, is such that $L \subseteq L' \subseteq L^m$ and it satisfies (40), which is again a contradiction with the hypothesis. Therefore $\sum_{i=1}^n l_i^m + \sum_{i=1}^n u_i^m = 2$ and then (41) implies (42). The proof is complete $\square$.

From (42) we can deduce that any set of probability intervals of the form $[l_i, u_i + \lambda_i]$, where $\lambda_i \geq 0\ \forall i$ and $\sum_{i=1}^n \lambda_i = 2 - \sum_{i=1}^n (l_i + u_i)$, is a minimal approximation of L. Moreover, it can be seen that all of these minimal approximations are always reachable, provided that the original probability intervals are.

Another interesting result about the minimal approximations is that they are associated to pairs of belief/plausibility functions whose focal elements always have a cardinality less than or equal to 2:

Proposition 16. If (bel, Pl) is a pair of belief/plausibility functions such that $bel(x_i) = l_i^m$, $Pl(x_i) = u_i^m\ \forall i$, and $\sum_{i=1}^n l_i^m + \sum_{i=1}^n u_i^m = 2$, then for every focal element B of (bel, Pl), is $|B| \leq 2$.

Proof. As $\sum_{i=1}^n l_i^m + \sum_{i=1}^n u_i^m = 2$ then $\sum_{i=1}^n (u_i^m - l_i^m) = 2(1 - \sum_{i=1}^n l_i^m)$.
So, on the one hand:
$1 - \sum_{i=1}^n l_i^m = 1 - \sum_{i=1}^n bel(x_i) = 1 - \sum_{i=1}^n m(x_i) = \sum_{\{B\,|\,|B|\geq 2\}} m(B)$.
On the other hand:
$\sum_{i=1}^n (u_i^m - l_i^m) = \sum_{i=1}^n (Pl(x_i) - bel(x_i)) = \sum_{i=1}^n \sum_{\{B\,|\,|B|\geq 2,\, x_i \in B\}} m(B) = \sum_{\{B\,|\,|B|\geq 2\}} |B| m(B)$
Then $\sum_{\{B\,|\,|B|\geq 2\}} |B| m(B) = 2 \sum_{\{B\,|\,|B|\geq 2\}} m(B)$ and therefore
$\sum_{\{B\,|\,|B|\geq 2\}} (|B| - 2) m(B) = 0$. As $|B| - 2 \geq 0$, all the terms in the sum are non-negative. The conclusion is that if $|B| > 2$ then $m(B) = 0$. So, the focal elements must have a cardinality less than or equal to 2 $\square$.

If we want to select only one approximation among all the minimal approximations of L, we must use an additional criterion. We propose to use the so called symmetry principle[35]. Intuitively this principle says that if we have several possible solutions, then we should look for an intermediate solution among the extreme ones. In our case, the n extreme minimal approximations L^{m_i}, $i = 1, \ldots, n$ of $L = \{[l_i, u_i], i = 1, .., n\}$ are:

$$L^{m_i} = \{[l_j^{m_i}, u_j^{m_i}] \,|\, l_j^{m_i} = l_j, u_j^{m_i} = u_j\ \forall j \neq i,\ l_i^{m_i} = l_i, u_i^{m_i} = u_i + \lambda\} \qquad (43)$$

where $\lambda = 2 - \sum_{i=1}^{n}(l_i + u_i)$. So, the arithmetic mean of these extreme minimal approximations, given by

$$L^{\mu} = \{[l_i^{\mu}, u_i^{\mu}] \mid l_i^{\mu} = l_i, u_i^{\mu} = u_i + \frac{\lambda}{n},\ i = 1, \ldots, n\}, \tag{44}$$

looks appropriate as the single approximation of L. Let us see an example:

Example 7. Consider the following set of reachable probability intervals defined on the domain $D_x = \{x_1, x_2, x_3, x_4\}$:

$$L = \{[0, 0.3], [0.1, 0.2], [0.3, 0.4], [0.1, 0.4]\}.$$

As $\sum_{i=1}^{4} l_i + \sum_{i=1}^{4} u_i = 1.8 < 2$, then L can not be a partial specification of any pair of belief and plausibility functions. The extreme minimal approximations (43) are

- $L^{m_1} = \{[0, 0.5], [0.1, 0.2], [0.3, 0.4], [0.1, 0.4]\}$
- $L^{m_2} = \{[0, 0.3], [0.1, 0.4], [0.3, 0.4], [0.1, 0.4]\}$
- $L^{m_3} = \{[0, 0.3], [0.1, 0.2], [0.3, 0.6], [0.1, 0.4]\}$
- $L^{m_4} = \{[0, 0.3], [0.1, 0.2], [0.3, 0.4], [0.1, 0.6]\}$

The single approximation L^{μ} defined in (44) is

$$L^{\mu} = \{[0, 0.35], [0.1, 0.25], [0.3, 0.45], [0.1, 0.45]\}.$$

If we apply the algorithm of Lemmer and Kyburg[34] to L^{μ}, we obtain a pair (bel, Pl) whose associated b.p.a. m is given by

- $m(x_2) = 0.1,\ m(x_3) = 0.3,\ m(x_4) = 0.1,$
- $m(\{x_1, x_2\}) = m(\{x_1, x_3\}) = 0.05,$
- $m(\{x_2, x_3\}) = m(\{x_2, x_4\}) = m(\{x_3, x_4\}) = 0.05,$
- $m(\{x_1, x_4\}) = 0.25$ $\square$.

Finally, let us study several interesting particular cases of sets of probability intervals which verify (40):

Example 8. $L^c = \{[l_i, u_i], i = 1, .., n\}$, where $l_i = 0\ \forall i$, and $u_i = 1\ \forall x_i \in B$, $u_i = 0\ \forall x_i \notin B$, where $B \neq \emptyset$ is any subset of the domain D_x which is not a singleton. Obviously L^c is reachable and it verifies (40). The only pair (bel, Pl) compatible with this partial specification (that is to say, verifying (37)) is associated to the b.p.a. m given by

$$m(B) = 1,\ m(A) = 0,\ \forall A \neq B.$$

This kind of belief and plausibility functions (in fact they are also necessity and possibility measures) is known as crisp measures focused on a subset, and they represent the following piece of information about the unknown value of the variable X: 'the value of X is in B' □.

Example 9. Consider a probability P defined on D_x, with probability distribution $p(x_i)$, $i = 1, \dots, n$. Let us define a set L of probability intervals by means of

$$l_i = (1-\epsilon)p(x_i), \; u_i = (1-\epsilon)p(x_i) + \epsilon, \; i = 1, \dots, n,$$

where $0 \leq \epsilon \leq 1$.
It is very easy to see that L is reachable, and it verifies (40). Moreover, there is only one pair (bel, Pl) which is compatible with L, and its associated b.p.a. is

$$m(x_i) = (1-\epsilon)p(x_i), \; i = 1, \dots, n, \; m(D_x) = \epsilon$$

This pair (bel, Pl) corresponds to the discounting operation defined by Shafer[32] for belief and plausibility functions, applied to the probability P (which is a particular case of belief measure that is equal to the plausibility measure). The semantic of this set of probability intervals corresponds to an ill-known probability, where the percentage of error is quantified by the value 100ϵ (we have a confidence level of $100(1-\epsilon)\%$ that the probability P is the correct one) □.

Example 10. Another way to express partial confidence in a probability measure P could be by considering the set of probability intervals $L = \{[l_i, u_i], i = 1, \dots, n\}$ defined by means of:

$$l_i = (p(x_i) - \epsilon) \vee 0, \; u_i = (p(x_i) + \epsilon) \wedge 1, \; i = 1, \dots, n,$$

where $p(x_i), i = 1, \dots, n$ is the probability distribution of P, and $0 \leq \epsilon \leq 1$.
It can be proved, although it is a bit harder than in the previous examples, that L is reachable and verifies (40). However, in this case there is more than one pair (bel, Pl) compatible with L. For instance, if $p(x_1) = 0.7$, $p(x_2) = 0.2$, $p(x_3) = 0.1$, $p(x_4) = 0.0$, and $\epsilon = 0.15$, then the pair (bel_1, Pl_1) with b.p.a. m_1 obtained using the algorithm in[34] is:

- $m_1(x_1) = 0.55$, $m_1(x_2) = 0.05$, $m_1(\{x_1, x_2\}) = 0.0833$,
- $m_1(\{x_1, x_3\}) = m_1(\{x_2, x_3\}) = 0.0333$,
- $m_1(\{x_1, x_4\}) = m_1(\{x_2, x_4\}) = m_1(\{x_3, x_4\}) = 0.0333$,
- $m_1(\{x_1, x_2, x_3\}) = 0.1$, $m_1(\{x_1, x_2, x_3, x_4\}) = 0.05$.

But the pair (bel_2, Pl_2) with associated b.p.a. m_2 defined by

- $m_2(x_1) = 0.55$, $m_2(x_2) = 0.05$, $m_2(\{x_1, x_2\}) = 0.05$,

- $m_2(\{x_1, x_4\}) = m_2(\{x_2, x_4\}) = m_2(\{x_3, x_4\}) = 0.05$
- $m_2(\{x_1, x_2, x_3\}) = 0.2$,

is also compatible with L $\square$.

Example 11. Suppose that we only know lower bounds for an unknown probability P defined on D_x, that is $l_i \leq p(x_i)\ \forall i$, where $\sum_{i=1}^{n} l_i \leq 1$. This information can obviously be represented by means of the set of probability intervals $\{[l_i, 1], i = 1, \ldots, n\}$. However, this set is not reachable. The equivalent reachable set of probability intervals is $L = \{[l_i, u_i], i = 1, \ldots, n\}$, where $u_i = 1 - \sum_{j \neq i} l_j\ \forall i$. This set of probability intervals was used by Fertig[2] to define interval influence diagrams. The lower and upper probabilities (l, u) associated to L are in this case very simple:

$$l(A) = \sum_{x_i \in A} l_i,\ u(A) = 1 - \sum_{x_i \notin A} l_i,\ \forall A \subseteq D_x\,,$$

which are belief and plausibility functions with associated b.p.a. m given by

$$m(x_i) = l_i,\ i = 1, \ldots, n,\ m(D_x) = 1 - \sum_{i=1}^{n} l_i\,.$$

Moreover, it is very easy to see that L verifies (40), and the only pair (bel, Pl) compatible with L is precisely (l, u). Finally, is interesting to remark that the sets of probability intervals considered in this example, generated only by lower bounds, are equivalent to those considered in example 9, generated by a probability and and parameter ϵ, by defining

$$\epsilon = 1 - \sum_{i=1}^{n} l_i,\ p(x_i) = \frac{l_i}{\sum_{j=1}^{n} l_j},\ \forall i\ \square.$$

7. Concluding Remarks.

In this paper we have studied in depth probability intervals as a formalism to represent uncertain information. Basic concepts for the management of uncertain information, such as combination, marginalization, conditioning and integration have been considered for probability intervals. Moreover, the relationship of this formalism with others, such as lower and upper probabilities, Choquet capacities of order two and belief and plausibility functions has also been clarified. Our opinion is that probability intervals, because of their computational simplicity and expressive power, are interesting tools for uncertain reasoning.

In further work we aim to include:

—The study of the concept of independence[19]. Independence permits us to modularise our knowledge in such a way that we only need to consult the pieces of information which are relevant to the question we are interested in, instead of having to

explore a complete knowledge base. For example, in Bayesian networks, independence is used to decompose an intractable joint probability distribution into smaller distributions, and then local methods may be used to perform inference. Having an appropriate definition of independence for probability intervals we could define belief networks based on probability intervals[2,11,16]. For this task, the study of ways to obtain bidimensional distributions from marginal and conditional ones, and the related problems of generalizing total probability and Bayes theorems to probability intervals will also be necessary (for several approaches to perform inference with imprecise or partially known probabilities see Refs. 9, 10, 36, 37, 38, 39, 40).
—Estimation methods for probability intervals, and their use for learning in belief networks. We can get probability intervals directly from experts or from empirical information (a set of examples) using statistical techniques as confidence intervals, but more research is necessary in this direction.

Acknowledgements

This work has been supported by the DGYCIT under Project n. PS89-0152

References

1. J.E. Cano, S. Moral and J.F. Verdegay, "Propagation of convex sets of probabilities in directed acyclic networks", in *Uncertainty in Intelligent Systems*, eds. B. Bouchon-Meunier, L. Valverde and R.R. Yager (North-Holland, Amsterdam, 1993) pp. 15–26.
2. K.W. Fertig and J.S. Breese, "Interval influence diagrams", in *Uncertainty in Artificial Intelligence* **5**, eds. M. Henrion, R.D. Shachter, L.N. Kanal and J.F. Lemmer (North-Holland, Amsterdam, 1990) pp. 149–161.
3. H.E. Kyburg, "Bayesian and non-bayesian evidential updating", *Artificial Intelligence* **31** (1987) 271–293.
4. I. Levi, *The Enterprise of Knowledge* (The MIT Press, Cambridge, Massachusetts, 1980).
5. H.E. Kyburg and M. Pittarelli, "Some problems for convex bayesians", *Proc. 8th Conference on Uncertainty in Artificial Intelligence*, Stanford, 1992, pp. 149–154.
6. L.M. de Campos, M.T. Lamata and S. Moral, "Logical connectives for combining fuzzy measures", in *Methodologies for Intelligent Systems* **3**, eds. Z.W. Ras and L. Saitta (North-Holland, Amsterdam, 1988) pp. 11–18.
7. J.E. Cano, S. Moral and J.F. Verdegay, "Partial inconsistency of probability envelopes", *Fuzzy Sets and Systems* **52** (1992) 201–216.
8. P.J. Huber, *Robust Statistics* (Wiley, New York, 1981).
9. P. Walley, *Statistical reasoning with imprecise probabilities* (Chapman and Hall, London, 1991).
10. R. Quinlan, "Inferno: a cautious approach to uncertain inference" *The Computer Journal* **26** (1983) 255–269.
11. B. Tessem, "Interval representation on uncertainty in Artificial Intelligence", Ph.D. Thesis, University of Bergen, Norway, 1989.
12. M. Sugeno, "Theory of fuzzy integrals and its application", Ph.D. Thesis, Tokio Inst. of Technology, 1974.
13. L.A. Zadeh, "Fuzzy sets as a basis for a theory of possibility", *Fuzzy Sets and Systems* **1** (1978) 3–28.

14. G. Choquet, "Theory of capacities", *Ann. Inst. Fourier* **5** (1953) 131–295.
15. L.M. de Campos and M.J. Bolaños, "Representation of fuzzy measures through probabilities", *Fuzzy Sets and Systems* **31** (1989) 23–36.
16. J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference* (Morgan and Kaufmann, San Mateo, 1988).
17. L.M. de Campos, "Caracterización y estudio de medidas e integrales difusas a partir de probabilidades", Ph.D. Thesis, Universidad de Granada, 1988.
18. D. Dubois and H. Prade, "A set-theoretic view of belief functions", *International Journal of General Systems* **12** (1986) 193–226.
19. L.M. de Campos and J.F. Huete, "Independence concepts in upper and lower probabilities", in *Uncertainty in Intelligent Systems*, eds. B. Bouchon-Meunier, L. Valverde and R.R. Yager (North-Holland, Amsterdam, 1993) pp. 49–59.
20. M.T. Lamata, "Modelos de decisión con información general", Ph.D. Thesis, Universidad de Granada, 1985.
21. S. Moral and L.M. de Campos, "Updating uncertain information", in *Uncertainty in Knowledge Bases, Lecture Notes in Computer Science* **521**, eds. B. Bouchon-Meunier, R.R. Yager and L.A. Zadeh, (Springer Verlag, Berlin, 1991), pp. 58–67.
22. L.M. de Campos, M.T. Lamata and S. Moral, "The concept of conditional fuzzy measure", *International Journal of Intelligent Systems* **5** (1990) 237–246.
23. R. Fagin and J.Y. Halpern, "A new approach to updating beliefs", Research Report RJ 7222, IBM Almaden Research Center, 1990.
24. J.Y. Jaffray, "Bayesian updating and belief functions", *Proc. 3th IPMU Conference*, Paris, 1990, pp. 449–451.
25. L.M. de Campos, M.T. Lamata and S. Moral, "A unified approach to define fuzzy integrals", *Fuzzy Sets and Systems* **39** (1991) 75-90.
26. L.M. de Campos and M.J. Bolaños, "Characterization and comparison of Sugeno and Choquet integrals", *Fuzzy Sets and Systems* **52** (1992) 61–67.
27. P.J. Huber and V. Strassen, "Minimax tests and the Neyman-Pearson lemma for capacities", *Ann. Statist.* **1** (1973) 251–263.
28. M.J. Bolaños, M.T. Lamata and S. Moral, "Decision making problems in a general environment", *Fuzzy Sets and Systems* **25** (1988) 135–144.
29. R.P. Loui, "Interval-based decisions for reasoning systems", *Uncertainty in Artificial Intelligence*, eds. L.N. Kanal and J.F. Lemmer (North-Holland, Amsterdam, 1986), pp. 459–472.
30. P. Wakker, "A behavioral foundation for fuzzy measures", *Fuzzy Sets and Systems* **37** (1990) 327–350.
31. A.P. Dempster, "Upper and lower probabilities induced by a multivalued mapping", *Ann. Math. Stat.* **38** (1967) 325–339.
32. G. Shafer, *A mathematical theory of evidence* (Princeton University Press, Princeton, NJ, 1976).
33. P. Smets, "The transferable belief model and random sets", *International Journal of Intelligent Systems* **7** (1992) 37–46.
34. J.F. Lemmer and H.E. Kyburg, "Conditions for the existence of belief functions corresponding to intervals of belief", *Proc. 9th National Conference on A.I.*, 1991, pp. 488–493.
35. S. Moral and L.M. de Campos, "Partially specified belief functions", *Proc. 9th Conference on Uncertainty in Artificial Intelligence*, Whashington, 1993, pp. 492–499.
36. N.J. Nilsson, "Probabilistic logic", *Artificial Intelligence* **28** (1986) 71–87.
37. G. Paass, "Probabilistic logic", in *Non-Standard logics For Automated Reasoning*, eds. D. Dubois, Ph. Smets, A. Mamdani and H. Prade (Academic Press, London, 1988) pp. 231–251.

38. S. Amarger, D. Dubois and H. Prade, "Constraint propagation with imprecise conditional probabilities", *Proc. 7th Conference on Uncertainty in AI*, Los Angeles, 1991, pp. 26–34.
39. D. Dubois, H. Prade and J.M. Toucas, "Inference with imprecise numerical quantifiers", in *Intelligent Systems: State of the Art and Future Directions*, eds. Z. Ras and M. Zemankova (Ellis-Horwood, 1990) pp. 52–72.
40. H. Thone, U. Guntzer and W. Kieβling, "Towards precision of probabilistic bounds propagation", *Proc. 8th Conference on Uncertainty in AI*, San Francisco, 1992, pp. 315–322.