Combining Fuzzy and Probabilistic Knowledge Using Belief Functions

Vladik Kreinovich[1], Claude Langrand[2], and Hung T. Nguyen[3]

[1]Computer Science, Univ. of Texas, El Paso, TX 79968, USA, vladik@cs.utep.edu
[2]Mathematics Dept., University of Lille, Lille, France, Claude.Langrand@univ-lille1.fr
[3]Mathematics Dept., New Mexico State U., Las Cruces,NM 88003, USA, hunguyen@nmsu.edu

Abstract

Some knowledge comes in probabilistic terms, some in fuzzy terms. These formalisms are drastically different, so it is difficult to combine the corresponding knowledge. A natural way to combine fuzzy and probabilistic knowledge is to find a formalism which enables us to express both types of knowledge, and then to use a combination rule from this general formalism. In this paper, as such a formalism, we propose to use belief functions. For the case when the universe of discourse is the set of all real numbers, we derive new explicit easy-to-compute analytical formulas for the resulting combination.

1 Introduction

Some knowledge comes in probabilistic terms, some in fuzzy terms. These formalisms are drastically different, so it is difficult to combine the corresponding knowledge. A natural way to combine fuzzy and probabilistic knowledge is to find a formalism which enables us to express both types of knowledge, and then to use a combination rule from this general formalism.

In this paper, as such a formalism, we propose to use belief functions. For the case when the universe of discourse is the set of all real numbers, we derive new explicit easy-to-compute analytical formulas for the resulting combination.

It is well known that belief functions are a natural generalization of probability distributions and hence, every probability distribution can be naturally represented as a particular case of a belief function.

For fuzzy information, there seems to be no natural immediate translation into belief functions. However, the following two-step natural representation is possible:

- First, a fuzzy information about a real number, i.e., a membership function $\mu(x)$ ("fuzzy number") with a maximum at some x_0, can be naturally represented as a "p-bound" $[F^-(x), F^+(x)]$, i.e., as a bound on the cumulative distribution function F(x):

  - $F^-(x) = 0$ for $x < x_0$, $1 - \mu(x)$ otherwise.

  - $F^+(x) = \mu(x)$ for $x < x_0$, 1 otherwise;

  In the paper, we explain why this translation is indeed natural.

- Second, since a belief function naturally leads, for each x, to lower and upper bounds $F^-(x)$ and $F^+(x)$ for the probability $F(x)$ of $(-\infty, x)$, we can invert this representation and represent each p-bound (including a p-bound coming from a fuzzy number) as a belief function.

Now that we know how to represent both types of knowledge in terms of belief functions, we must use a belief combination rule to combine these two types of knowledge. There are several different ways of combining belief functions. They are all based on the original Dempster-Shafer combination rule, the difference is how they treat inconsistent pairs of focal elements (i.e., focal elements with empty intersection). In the original rule, such pairs are ignored, and the corresponding mass is proportionally distributed between consistent pairs; in Zadeh's version, this mass is added to the entire universe of discourse (meaning complete uncertainty). It turns out that for our specific application, only the original combination rule leads to non-degenerate results.

Explicit formulas are derived for these results. It is interesting to mention that even for the case when we combine two probability distributions, we get a non-trivial combination formula. In general, for combining p-bounds, the width $F^+(x) - F^-(x)$ of the resulting combination is proportional to the product of the widths, i.e., this combination decreases the uncertainty (as it should).

## 2 p-Bounds from expert estimates: case of natural-language estimates

Words from natural language are not precise ("fuzzy"). Let us take the word "small" as an example. When the value of, say, concentration, is really small, everyone would 100% agree that this value is small indeed. When the value is really large, everyone would agree that this value is not small. For intermediate values, however, we bound to have a disagreement.

The need to translate expert knowledge from natural language to a computer-understandable language of numbers was recognized as early as the 1960s, when the first expert systems started to be designed. A special formalism called *fuzzy logic* was designed to help us capture the meaning of words. In this formalism, to represent a meaning of a word like "small", we assign, to every possible value x, a degree $\mu_{\mathrm{small}}(x)$ to which x is small. The dependence of this degree of x is called a *membership function*, or a *fuzzy set*.

Where do the values $\mu(x)$ come from? There are several dozen different techniques for eliciting these values; see, e.g., [1, 3]. Sometimes, the experts can present these numbers ("subjective probabilities") directly. If they cannot, then for every x, we can poll several (N) experts on whether they believe that this particular values x is, say, small, and if M out of N experts answer "yes", we take $\mu(x) = M/N$. What is a natural way to translate these "subjective probabilities" into p-bounds?

We will answer this question on the example of membership functions of three most frequent types. The first type is a function which describes words like "large", for which $\mu(x)$ is increasing from 0 at $x = 0$ to 1 for $x \to \infty$. Let us give a simple example of such function:

$$\mu_{\mathrm{large}}(x) = \begin{cases} 0 & \text{if } x \leq 1, \\ x - 1 & \text{if } 1 \leq x \leq 2, \\ 1 & \text{if } x > 2 \end{cases}$$
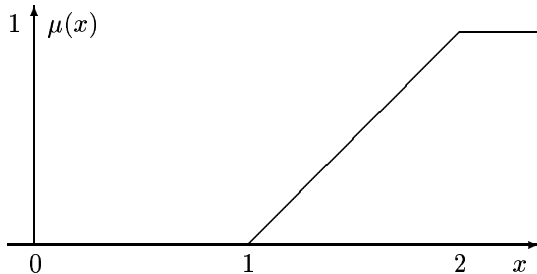
Suppose that the expert tells us that the actual value of some quantity X is large. What does it say about the possible values of the probability $F(x)$ (that $X \leq x$) for different x?

Let us start with a value $x \leq 1$. For this value, $\mu_{\mathrm{large}}(x) = 0$. This means that the values below x cannot be large, so it is reasonable to take $F(x) = 0$.

Let us now take a value $x \geq 2$. For this value, $\mu_{\mathrm{large}}(x) = 1$, which means that the value x is definitely large. Based on the expert opinion, we only know that the actual value X is large. It may be below x with probability 1 – in which case $F(x) = 1$; it may be above X with probability 1 – in which case $F(x) = 0$. So, here, the corresponding value of the p-bound – i.e., the interval of possible values of $F(x)$ – is $\mathbf{F}(x) = [0, 1]$.

What if x is in between 1 and 2, e.g., $x = 1.6$? In this case, the probability $\mu(x)$ that x is large is equal to 0.6. Since the function $\mu(x)$ is increasing, the probability $\mu(X)$ that X is large even smaller for $X < x$. Thus, out of all large values, values ≤ 0.6 should have a frequency ≤ 0.6. So, since we know that actual value X is large, we conclude that the probability $F(x)$ cannot exceed 0.6.

In general, the value $F(x)$ cannot exceed the probability $\mu(x)$, i.e., $\mu(x)$ serves as the upper part $F^+(x)$ of the p-bound. The lower part $F^-(x)$ should be 0, because we may have X so large than it is much larger than 2.

Combining these three cases, we conclude that for increasing membership functions $\mu(x)$ like "large", a natural translation of the membership function is a p-bound $[0, \mu(x)]$.

The second type of membership functions that we will consider is a function which describes words like "small", for which $\mu(x)$ is decreasing from 1 at $x = 0$ to 0 for $x \to \infty$. Let us give a simple example of such function:

$$\mu_{\mathrm{small}}(x) = \begin{cases} 1 - x & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise} \end{cases}$$
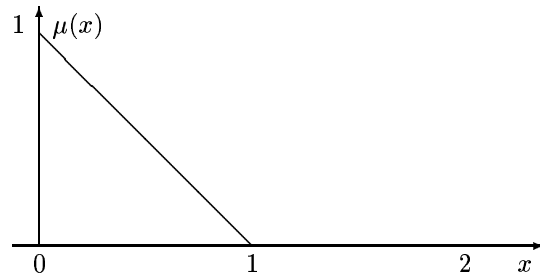

Suppose that the expert tells us that the actual value of some quantity X is small. What does it say about the possible values of the probability $F(x)$ (that $X \leq x$) for different x?

Let us start with a value $x \geq 1$. For this value, $\mu_{\mathrm{small}}(x) = 0$, which means that the value x is definitely not small. Based on the expert opinion, we only know that the actual value X is small. All values X

which can be small (i.e., for which $\mu(X) > 0$) are below 1, so they are all below x. Thus, all values of X are below x with probability 1, and $F(x) = 1$.

What if x is in between 0 and 1, e.g., $x = 0.2$? In this case, the probability $\mu(x)$ that x is small is equal to 0.8. Hence, the probability that any larger value $X > x$ is "small" also does not exceed 0.8. This means that if $F(x)$ is smaller than $1 - 0.8 = 0.2$ – e.g., equal to 0.1 – then there will be more than > 0.8 of values which are $\geq x$ – and thus, some values $X > x$ cannot be reasonably called small, in contradiction to the expert's opinion. So, if the actual value X is small, the probability $F(x)$ cannot exceed 0.2. In general, the value $F(x)$ cannot be smaller than $1 - \mu(x)$, i.e., $1 - \mu(x)$ serves as the upper part $F^-(x)$ of the p-bound. The upper part $F^+(x)$ should be 1, because we may have $X = 0$ with probability 1.

Combining these two cases, we conclude that for increasing membership functions $\mu(x)$ like "small", a natural translation of the membership function is a p-bound $[1 - \mu(x), 1]$:
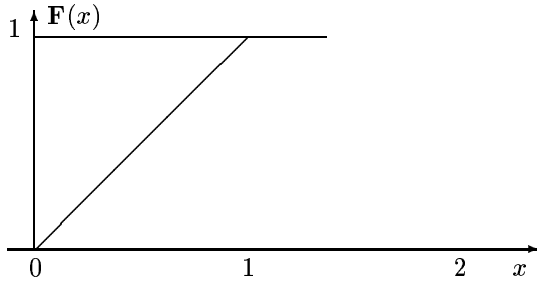
Finally, we can consider membership functions describing terms like "around x_0", which increase from 0 to 1 until they reach a certain value x_0, and then decrease from 1 to 0. For such membership functions, possible values (i.e., values for which the degree $\mu(x)$ is large enough) are concentrated around the number x_0, that is why such membership functions are called *fuzzy numbers*.

As an example, we will consider the following function corresponding to "around 1":

$$\mu_{\approx 1}(x) = \begin{cases} x & \text{if } 0 \leq x \leq 1, \\ 2 - x & \text{if } 1 \leq x \leq 2, \\ 0 & \text{otherwise} \end{cases}$$
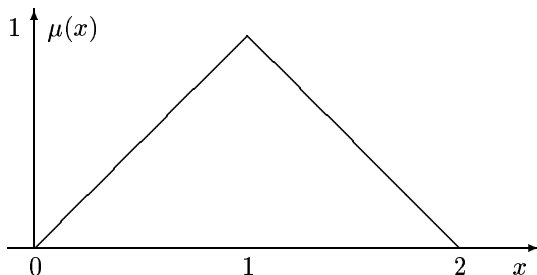
For a membership function of this type, with a maximum at some value x_0, similar arguments lead to the following p-bound $[F^-(x), F^+(x)]$:
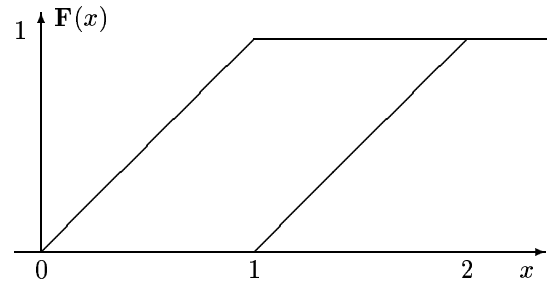
$$F^-(x) = \begin{cases} 0 & \text{if } x \leq x_0, \\ 1 - \mu(x) & \text{if } x > x_0 \end{cases}$$

$$F^+(x) = \begin{cases} \mu(x) & \text{if } x \leq x_0, \\ 1 & \text{if } x > x_0 \end{cases}$$

In particular, for the above membership function "around 1", the corresponding p-bound has the following form:

$$F^-(x) = \begin{cases} 0 & \text{if } x \leq 1, \\ x - 1 & \text{if } 1 \leq x \leq 2, \\ 1 & \text{if } x > 2 \end{cases}$$

$$F^+(x) = \begin{cases} x & \text{if } x \leq 1, \\ 1 & \text{if } x > 1 \end{cases}$$

These three cases can be described in a way which is similar to our transformation of measurements into p-bounds. Indeed, how can we describe a fuzzy set that corresponds to a certain property like "around 1"? A natural way to characterize a fuzzy set is to describe, for every level α, the set $X_\alpha = \{x \mid \mu(x) \geq \alpha\}$ of all the values which have this property with degree at least α. Such sets are called α-*cuts* – because on the graph, they really correspond to horizontal cuts. For example, for the above membership function "around 1", the α-cuts are $X_\alpha = [\alpha, 2 - \alpha]$:
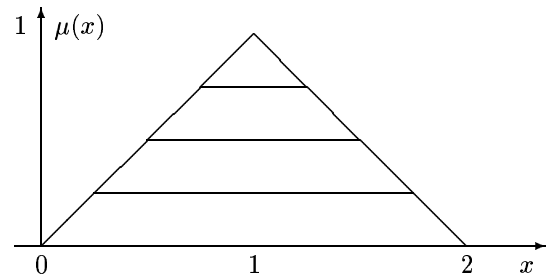
If we, e.g., have α-cuts $X_{0.1}$, $X_{0.2}$, etc., corresponding to $\alpha = 0.1$, $\alpha = 0.2$, etc., this means, crudely speaking, that all experts agree that $x \in X_0$, that 90% of them agree that $x \in X_{0.1}$, that 80% of experts agree that

$x \in X_{0.2}$, etc., until we reach we level $X_{0.9}$ in which only 10% of the experts agree; see, e.g., [2]. So, we have a natural subdivision of experts into 10 groups: 10% believe that x is somewhere on the interval $X_{0.1}$ – and no narrower bounds are possible; 10% believe that x is somewhere on the interval $X_{0.2}$ – and no narrower bounds are possible, etc. We thus have a typical Dempster-Shafer knowledge base. One can easily see that if we use the above algorithm to transform this knowledge base into a p-bound, we get exactly the p-bound that we cam up with.

3 How can Dempster-Shafer aggregation method be applied to combining CDF's

Idea. An arbitrary Dempster-Shafer knowledge base can be naturally represented as a p-bound. So, if we want to apply the Dempster-Shafer (DS) combination rule to CDF's, we must do the following:

- first, we transform CDF into a DS knowledge base;

- second, we apply DS combination rule to get a new DS knowledge base;

- finally, we translate the resulting DS knowledge base into a CDF.

We know how to perform the second and the third steps. The first step is somewhat ambiguous because there are many ways to reconstruct a DS knowledge base from a p-bound.

When the CDF's come from measurement, then CDF is actually the result of combining several (N) measurement results, with equal probabilities $1/N$, into a single CDF. In this case, a natural way is to represent the CDF as a combination of different measurement results with probability $1/N$. In other words, we subdivide the interval $[0, 1]$ into small subintervals $[0, \Delta p]$, $[\Delta p, 2\Delta p]$, etc. To each of these narrow intervals, we assign the corresponding x-interval $[F^{-1}(p), F^{-1}(p + \Delta p)]$. These narrow intervals are our focal points, and each has the probability Δp.

Formula. In this case, DS-aggregation results in a new probability distribution whose density $\rho_{\text{new}}(x)$ is proportional to the *maximum* of the densities $\rho_1(x)$ and $\rho_2(x)$ of the two aggregated probability distributions:

$$\rho_{\text{new}}(x) = \frac{1}{N} \cdot \max(\rho_1(x), \rho_2(x)),$$

where

$$N = \int_{-\infty}^{\infty} \max(\rho_1(x), \rho_2(x)) \, dx.$$

Idea of the proof. For each distribution $F_i(x)$, a focal element containing x has the length $\Delta p / \rho_i(x)$. Thus, if, say, $\rho_1(x) = 2\rho_2(x)$, then a focal element corresponding to $F_2(x)$ contains two focal elements corresponding to $F_1(x)$. We therefore get two non-empty intersections close to x, i.e., exactly as many as corresponds to the largest of the two probability densities. As a result, the focal element containing x has the length $\Delta p / \max(\rho_1(x), \rho_2(x))$. This formula is true irrespective of which of the two probability density functions $\rho_i(x)$ is larger at x and how much larger.

What is the density $\rho(x)$ corresponding to this new intersection DS-knowledge base? In general, the length of each focal element is proportional to $\Delta p' / \rho(x)$, hence $\rho(x)$ is indeed proportional to $\max(\rho_1(x), \rho_2(x))$.

Examples. Let us first check this formula on the example when $\rho_1(x) = \rho_2(x)$. In this case, we get $\max(\rho_1(x), \rho_2(x)) = \rho(x)$, hence $N = 1$ and $\rho_{\text{new}}(x) = \rho_1(x)$. This example shows that this aggregation operation is idempotent.
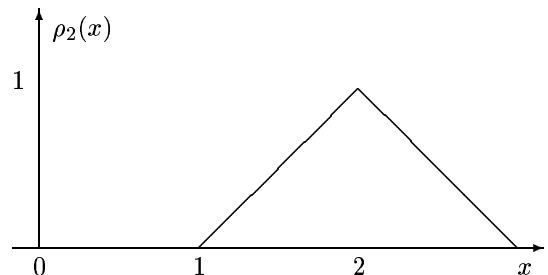
Let us now give one non-trivial example of this operation. Let $\rho_1(x)$ be a unimodal distribution with a triangular density function:

$$\rho_1(x) = \begin{cases} x & \text{if } 0 \leq x \leq 1, \\ 2 - x & \text{if } 1 \leq x \leq 2, \\ 0 & \text{otherwise} \end{cases}$$
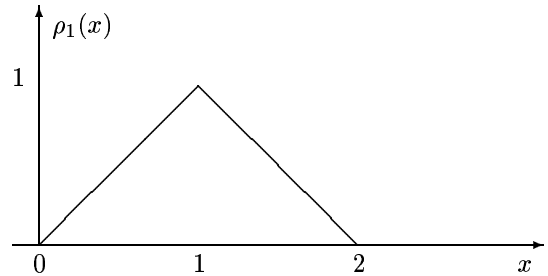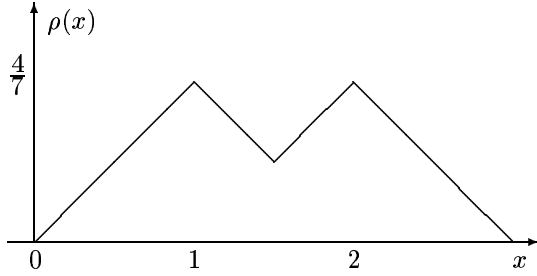

As $\rho_2(x)$, let us take a similar distribution, but shifted by 1:

$$\rho_2(x) = \begin{cases} x - 1 & \text{if } 1 \leq x \leq 2, \\ 3 - x & \text{if } 2 \leq x \leq 3, \\ 0 & \text{otherwise} \end{cases}$$

Here, $N = 7/4$, and the aggregation result $\rho(x)$ is the following bi-modal distribution:

$$\rho(x) = \begin{cases} (4/7) \cdot x & \text{if } 0 \leq x \leq 1, \\ (4/7) \cdot (2 - x) & \text{if } 1 \leq x \leq 1.5, \\ (4/7) \cdot (x - 1) & \text{if } 1.5 \leq x \leq 2, \\ (4/7) \cdot (3 - x) & \text{if } 2 \leq x \leq 3, \\ 0 & \text{otherwise} \end{cases}$$


How can we describe this operation in terms of CDF? Since p-bounds are described in terms of the cumulative distribution function (CDF), not probability density function $\rho(x)$, it is desirable to describe this combination operation in terms of CDF. For that, we can use the fact that the probability density is a slope of the CDF. So, if we start with the two CDF's, then, in essence, on each subinterval of the real line, we pick the shape corresponding to the steepest of the two CDF's.

Let us give a simple example. Let us consider the following two CDF's: $F_1(x)$ corresponds to a uniform distribution on the interval $[0, 1]$, and $F_2(x)$ is a combination of two uniform distributions: on the interval $[0, 2/3]$ (with probability $1/3$) and on the interval $[2/3, 1]$ (with probability $2/3$). The corresponding CDF's are:

$$F_1(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ x & \text{if } 0 \leq x \leq 1, \\ 1 & \text{if } x \geq 1 \end{cases}$$
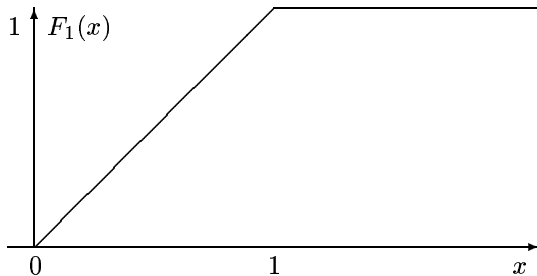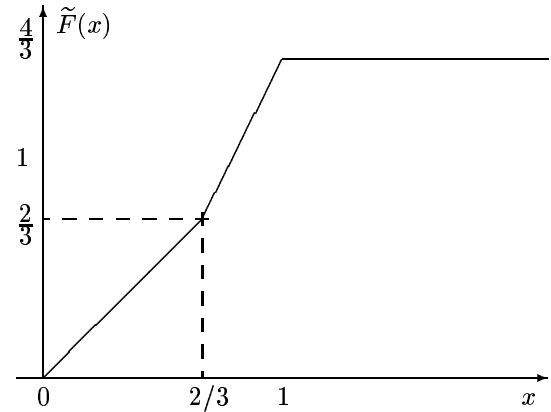

$$F_2(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ (1/2) \cdot x & \text{if } 0 \leq x \leq 2/3, \\ 2x - 1 & \text{if } 2/3 \leq x \leq 1, \\ 1 & \text{if } x \geq 1 \end{cases}$$
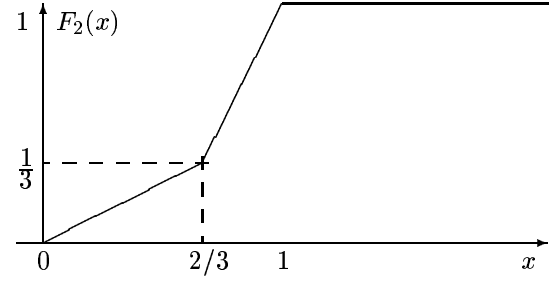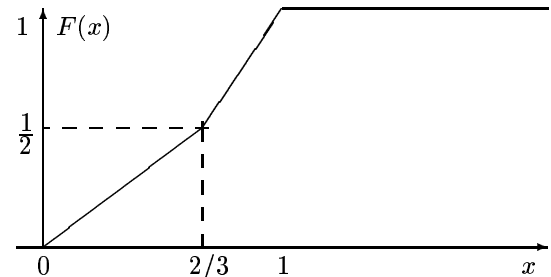

In this example, for $x \leq 2/3$, $F_1(x)$ is steeper, while for $x \geq 2/3$, $F_2(x)$ is steeper. Thus, for $x \leq 2/3$, we copy the CDF $F_1(x)$, and for $x \geq 2/3$, we copy the CDF $F_2(x)$. As a result, we get the following "CDF" $\widetilde{F}(x)$:

$$\widetilde{F}(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ x & \text{if } 0 \leq x \leq 2/3, \\ 2x - 2/3 & \text{if } 2/3 \leq x \leq 1, \\ 4/3 & \text{if } x \geq 1 \end{cases}$$

Finally, the resulting "CDF" needs to be normalized, so we get the following aggregated CDF $F(x)$:

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ (1/2) \cdot x & \text{if } 0 \leq x \leq 2/3, \\ 2x - 1 & \text{if } 2/3 \leq x \leq 1, \\ 1 & \text{if } x \geq 1 \end{cases}$$

Properties. This operation is idempotent, commutative, sensitive, rather easy to compute. However, one important property does not hold: this operation is not associative, i.e., it is not true that $(\rho * \rho') * \rho'' =$

$\rho*(\rho'*\rho'')$ for all possible distributions ρ, ρ', and ρ''. As an example of non-associativity, we take three uniform distributions: $\rho(x)$ corresponds to a uniform distribution on the interval $[-1,0]$, $\rho'(x)$ – on $[-0.5, 0.5]$, and $\rho''(x)$ – on the interval $[0,1]$. Each of these three intervals has a unit length, so each probability density function has a value 1 within this interval.

In this example, $\max(\rho(x), \rho'(x))$ is equal to 1 on the interval $[-1, 0.5]$:
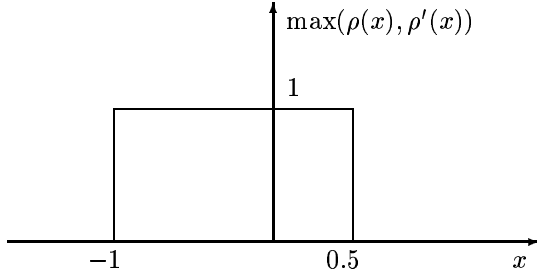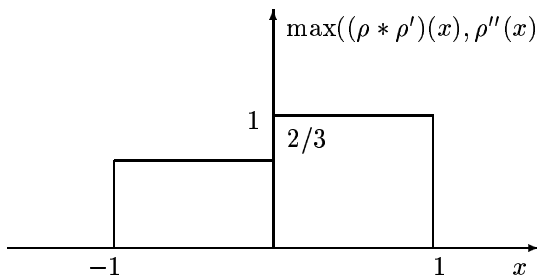

The integral N of the corresponding function $\max(\rho(x), \rho'(x))$ is $3/2$, hence the normalized function $(\rho * \rho')(x)$ has the following form:

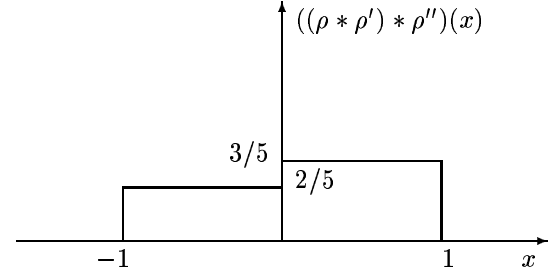$$(\rho * \rho')(x) = \begin{cases} 2/3 & \text{if } -1 \le x \le 0.5, \\ 0 & \text{otherwise} \end{cases}$$
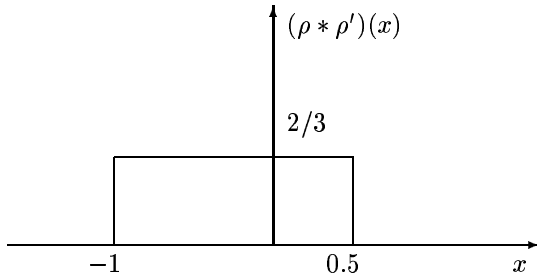
Now,

$$\max((\rho * \rho')(x), \rho''(x)) = \begin{cases} 2/3 & \text{if } -1 \le x \le 0, \\ 1 & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise} \end{cases}$$
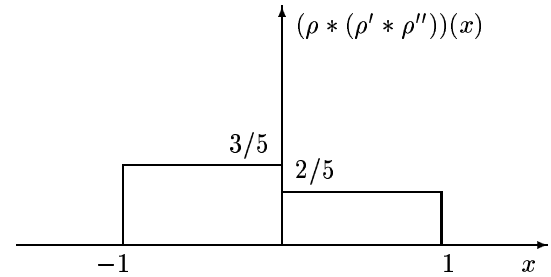

Here, $N = 5/3$, hence:

$$((\rho * \rho') * \rho'')(x) = \begin{cases} 2/5 & \text{if } -1 \le x \le 0, \\ 3/5 & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise} \end{cases}$$

Similarly, we conclude that:

$$(\rho * (\rho' * \rho''))(x) = \begin{cases} 3/5 & \text{if } -1 \le x \le 0, \\ 2/5 & \text{if } 0 \le x \le 1, \\ 0 & \text{otherwise} \end{cases}$$

Hence, $(\rho * \rho') * \rho'' = \rho * (\rho' * \rho'')$. So, the aggregation operation is indeed non-associative.

Comment. In terms of densities, the DS aggregation operation for probability distributions consists of two steps: taking the maximum of the values of the probability density functions and normalization. Since $\max(a, b)$ is clearly an associative operation, non-associativity comes from normalization.

4 Dempster-Shafer approach to aggregating p-bounds

Formulas. When we discussed aggregation of CDF's, we have mentioned that every CDF can be represented as a Dempster-Shafer knowledge base, with percentiles values $F_i^{-1}(p)$ (or, to be more precise, with small intervals around percentile values), as focal elements.

For a p-bound, we do not have a single percentile value, we have an *interval* $[(F_i^+)^{-1}(p), (F_i^-)^{-1}(p)]$ of possible percentile values. Similar to the case of CDF's, every p-bound can be represented as Dempster-Shafer knowledge base, with the corresponding percentile intervals as focal elements. If we apply Dempster-Shafer combination rule to combine these percentiles, we get the following formula for the result $[F^-(x), F^+(x)]$ of aggregating two p-bounds $[F_1^-(x), F_1^+(x)]$ and $[F_2^-(x), F_2^+(x)]$. First, we compute

pre-normalized "CDF"s as follows:

$$\widetilde{F}^-(x) = \int \rho_1^-(x) \cdot (F_2^+(x) - F_2^-(x))\, \mathrm{d}x +$$

$$\int \rho_2^-(x) \cdot (F_1^+(x) - F_1^-(x))\, \mathrm{d}x;$$

$$\widetilde{F}^+(x) = \int \rho_1^+(x) \cdot (F_2^+(x) - F_2^-(x))\, \mathrm{d}x +$$

$$\int \rho_2^+(x) \cdot (F_1^+(x) - F_1^-(x))\, \mathrm{d}x.$$

where $\rho_i^-(x)$ and $\rho_i^+(x)$ are probability density functions corresponding to CDF's $F_i^-(x)$ and $F_i^+(x)$. We can rewrite these formulas exclusively in terms of CDF's if we use the notion of a Stieltjes integral $\int f(x)\,\mathrm{d}F(x)$ (which is equivalent to $\int f(x) \cdot \rho(x)\,\mathrm{d}x$):

$$\widetilde{F}^-(x) = \int (F_2^+(x) - F_2^-(x))\, \mathrm{d}F_1^-(x) +$$

$$\int (F_1^+(x) - F_1^-(x))\, \mathrm{d}F_2^-(x);$$

$$\widetilde{F}^+(x) = \int (F_2^+(x) - F_2^-(x))\, \mathrm{d}F_1^+(x) +$$

$$\int (F_1^+(x) - F_1^-(x))\, \mathrm{d}F_2^+(x).$$

After we compute each "CDF"s $\widetilde{F}(x)$, we normalize it by dividing by the normalizing constant – which happens to be the value $\widetilde{F}^+(\infty)$ of the pre-normalized "CDF" $\widetilde{F}^+(x)$ when $x \to \infty$:

$$F^-(x) = \frac{\widetilde{F}^-(x)}{\widetilde{F}^+(\infty)}; \quad F^+(x) = \frac{\widetilde{F}^+(x)}{\widetilde{F}^+(\infty)}.$$

These formulas can be naturally generalized to the case when we aggregate an arbitrary number n of p-bounds. In this case, we get:

$$\widetilde{F}^-(x) = \sum_{i=1}^n \int \rho_i^-(x) \cdot \prod_{j \neq i}(F_j^+(x) - F_j^-(x))\, \mathrm{d}x;$$

$$\widetilde{F}^+(x) = \sum_{i=1}^n \int \rho_i^+(x) \cdot \prod_{j \neq i}(F_j^+(x) - F_j^-(x))\, \mathrm{d}x.$$

Alternatively:

$$\widetilde{F}^-(x) = \sum_{i=1}^n \int \prod_{j \neq i}(F_j^+(x) - F_j^-(x))\, \mathrm{d}F_i^-(x);$$

$$\widetilde{F}^+(x) = \sum_{i=1}^n \int \prod_{j \neq i}(F_j^+(x) - F_j^-(x))\, \mathrm{d}F_i^+(x).$$

This aggregation operation sounds somewhat complex, but it leads to a simple formula for the interval width $w(x) = F^+(x) - F^-(x)$ of the resulting p-bound:

$$w(x) = k \cdot w_1(x) \cdot \ldots \cdot w_n(x),$$

where k is a normalizing constant, and $w_i(x) = F_i^+(x) - F_i^-(x)$ are the widths of the aggregated p-bounds.

Idea of the proof. For every point x, how can we compute the increase, e.g., in the lower p-bound at x? By the definition of the lower p-bound, this increase comes from all the focal elements that end at x. For each of the combined p-bounds, there is only one focal element that ended at x, and this increase is proportional to $\rho_i^-(x)$. In accordance with the DS combination rule, the new focal elements are non-empty intersections of focal elements of both knowledge bases. An intersection focal element ends exactly at x if one of the intersected intervals ends at x. If it is the first intersected focal element that ends at x, then it has a non-empty intersection with all focal elements of the second p-bound that contain x. These focal elements correspond to probabilities from $F_2^-(x)$ to $F_2^+(x)$, hence there are exactly $(F_2^+(x) - F^-(x))/\Delta p$ of them. The increase in $\widetilde{F}^-(x)$ caused by such elements is therefore proportional to $\rho_1^-(x) \cdot (F_2^+(x) - F_2^-(x))$. Similarly, we get a formula for the increase caused by intersection focal points in which the second intersected interval ends at x, and thus, the desired formula for $\widetilde{F}^-(x)$.

The formula for $\widetilde{F}^+(x)$ is obtained in a similar manner, only we must count not focal points that end at x but those that start at x.

Properties and examples. Let us start with checking idempotence. When we combine a p-bound $\mathbf{F}_1(x)$ with itself, we get a new p-bound with the width $w(x) = k \cdot w_1(x)^2$. The only way for the p-bound to stay the same is when $w(x) = k \cdot w_1^2(x) = w_1(x)$, i.e., when $w_1(x) \equiv \text{const}$. For CDF's, a similar operation was idempotent, because a CDF can be viewed as a CDF with a contact (0) width. However, as soon as the width stops being constant, we lose the idempotence property.

Let us give a simple example of why DS-combination rule is not idempotent. Let us take, as the aggregated p-bound, a DS knowledge base with three focal elements $\mathbf{x}_1 = [0, 2]$, $\mathbf{x}_2 = [1, 3]$, and $\mathbf{x}_3 = [3, 4]$, to each of which we assign the same mass $p_1 = p_2 = p_3 = 1/3$. This DS knowledge base corresponds to the following CDF:

$$F_1^-(x) = \begin{cases} 0 & \text{if } x < 2, \\ 1/3 & \text{if } 2 \leq x < 3, \\ 2/3 & \text{if } 3 \leq x < 4, \\ 1 & \text{if } x \geq 4 \end{cases}$$

$$F_1^+(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1/3 & \text{if } 0 \leq x < 1, \\ 2/3 & \text{if } 1 \leq x < 3, \\ 1 & \text{if } x \geq 3 \end{cases}$$
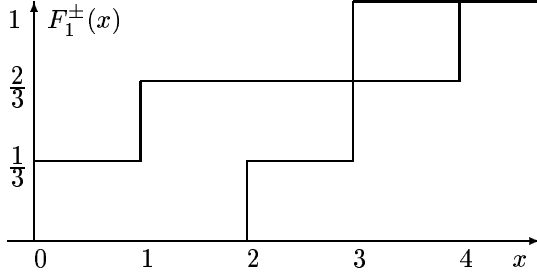
In accordance with the Dempster-Shafer combination rule, we take all intervals $\mathbf{x}_i$ from the first knowledge case, all intervals $\mathbf{y}_j$ from the second knowledge base (which in this case is the same, i.e., $\mathbf{y}_j = \mathbf{x}_j$), and assign the mass $p_i \cdot q_j$ (in this case, $1/9$) to all non-empty intersections. Since these masses do not add up to 1, we normalize them so that they will.

In this case, the following intersections are non-empty: $\mathbf{x}_1 \cap \mathbf{y}_1 = [0,2]$; $\mathbf{x}_1 \cap \mathbf{y}_2 = [1,2]$; $\mathbf{x}_2 \cap \mathbf{y}_1 = [1,2]$; $\mathbf{x}_2 \cap \mathbf{y}_2 = [1,3]$; $\mathbf{x}_3 \cap \mathbf{y}_3 = [3,4]$. These five intervals get the same mass, so after normalization, they are each assigned the same mass $1/5$. The resulting p-bounds are as follows:

$$F^-(x) = \begin{cases} 0 & \text{if } x < 2, \\ 3/5 & \text{if } 2 \le x < 3, \\ 4/5 & \text{if } 3 \le x < 4, \\ 1 & \text{if } x \ge 4 \end{cases}$$

$$F^+(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1/5 & \text{if } 0 \le x < 1, \\ 4/5 & \text{if } 1 \le x < 3, \\ 1 & \text{if } x \ge 3 \end{cases}$$
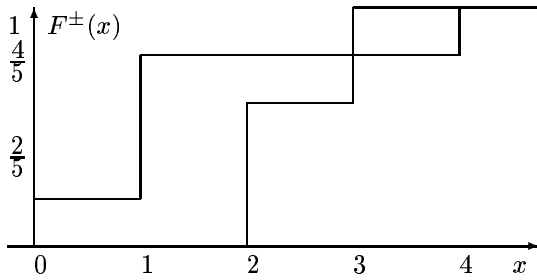
On this example, we see that the width of the aggregated p-bound is indeed proportional to the square of the original one:

- In the original p-bound, the width was $2/3$ on $[1,2]$ and twice smaller ($1/3$) elsewhere on $[0,4]$.

- In the aggregated p-bound, the weight is $4/5$ on $[1,2]$, and four times smaller elsewhere on $[0,4]$.

Comparing the original p-bound with the aggregated one, we can see that not only the aggregated p-bound

is different: it is neither enclosed in the original one, not enclosing the original one. Indeed, for $x \in (0,1)$, we have:

$$[F^-(x), F^+(x)] = \left[0, \frac{1}{5}\right] \subset [F_1^-(x), F_1^+(x)] = \left[0, \frac{1}{3}\right].$$

On the other hand, for $x \in (1,2)$, we have:

$$[F^-(x), F^+(x)] = \left[0, \frac{4}{5}\right] \supset [F_1^-(x), F_1^+(x)] = \left[0, \frac{2}{3}\right].$$
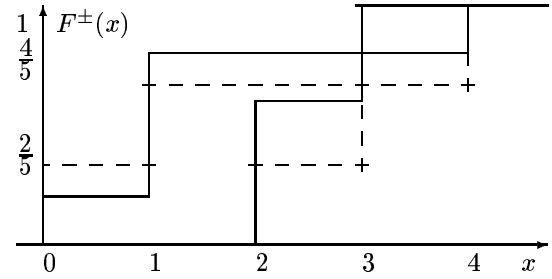
Summarizing: this operation is not idempotent. It is commutative, sensitive, rather easy to compute, but – similar to the case of CDF's – not associative.

Acknowledgments

References

[1] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, NJ, 1995.

[2] H. T. Nguyen and V. Kreinovich, "Nested Intervals and Sets: Concepts, Relations to Fuzzy Sets, and Applications", In: R. B. Kearfott et al (eds.), *Applications of Interval Computations*, Kluwer, Dordrecht, 1996, pp. 245–290.

[3] H. T. Nguyen and E. A. Walker, *First Course in Fuzzy Logic*, CRC Press, Boca Raton, FL, 1999.