Generating Minimum Dispersion Densities from an Interval-Valued Fuzzy Set

Carol L. Walker[1] Elbert A. Walker[1] Ronald R. Yager[2]

1. Department of Mathematical Sciences
Las Cruces, New Mexico, USA
2. Machine Intelligence Institute, Iona College
New Rochelle, New York, USA
Email: hardy@nmsu.edu, elbert@nmsu.edu, yager@panix.com

Abstract—*Let f be an interval-valued fuzzy subset of a finite set U of size n. This yields n closed intervals inside the unit interval. Picking a point in each interval and dividing by the sum of the points gives rise to a probability density on the set of intervals. For a given measure of dispersion, such as entropy, the problem is to pick points that maximize (or minimize) this dispersion. The particular case considered here is to pick points that minimize the sum of the squares of probabilities in the density. We provide an algorithm for picking such points.*

Keywords—*interval-valued fuzzy set; minimum dispersion density; probability density; sum of squares*

I. Introduction

A problem that arises in several situations is this. Given a family of probability densities, choose one from that family that has the largest value for some measure of dispersion. Some instances involving entropy are discussed in [1], [2], [3], [4], [5], [6], [7]. But there are other measures of dispersion, for which a general formulation is provided in [8].

The problem we consider is, given n subintervals $[a_1, b_1]$, $[a_2, b_2]$, ..., $[a_n, b_n]$ of $[0, 1]$, choose $x_i \in [a_i, b_i]$ such that for an n-tuple $\mathbf{x} = (x_1, ..., x_n)$ the value

$$H(\mathbf{x}) = \sum_j \left(\frac{x_j}{\sum_i x_i}\right)^2 \tag{1}$$

is minimum. This means that among all the possible choices $\mathbf{x}$ for the n-tuple of x_i's, this quantity is minimum.

One motivation for this problem is the following. Let U be a finite set and f a function from U into the set of closed intervals of $[0, 1]$. That is, $f : U \to \{[a, b] : 0 \le a \le b \le 1\}$. This yields a finite set of closed intervals $\{[a_i, b_i] : i = 1, 2, \ldots, n\}$. If not all the $a_i = 0$, than choosing $x_i \in [a_i, b_i]$ yields a probability density $P([a_i, b_i]) = x_i/\Sigma x_j$ on $\{[a_i, b_i] : i = 1, 2, \ldots, n\}$. Maximizing, (or minimizing) a dispersion measure gives a canonical way to pick out a point x_i in each interval $[a_i, b_i]$, a possible first step toward defuzzifying the interval-valued fuzzy set f.

Several remarks are in order.

1) It should be noted that the intervals do not have to be distinct, and that an interval can be a single point. Intervals such as $[0, 0]$ and $[0, a]$ are allowed, as are $[1, 1]$ and $[b, 1]$. And of course, an interval may be contained in another.
2) If all the intervals are $[0, 0]$, then each $x_i = 0$ and each probability is $\frac{0}{0}$, which is undefined so *we assume this is not the case*. Thus at least one of the intervals is not equal to $[0, 0]$, and there is at least one $S = \sum_{i=1}^n x_i$ that is positive. In this case, the interval $[0, 0]$ has associated probability 0, and its contribution to the sum of the squares is 0. Thus, in developing an algorithm for finding the x_i's that give minimum dispersion, *we can assume that no interval is* $[0, 0]$, that is, $b_i > 0$ for all i.
3) If the intersection of the intervals is non-empty, then choosing all the x_i to be any point $x_i = x > 0$ in that intersection yields minimum dispersion, namely

$$\sum_{j=1}^n \left(\frac{x}{\sum_{i=1}^n x}\right)^2 = \frac{1}{n} \tag{2}$$

In particular, the solution may not be unique. In fact, in this case, if the intersection is not a single point, then there are uncountably many solutions, namely any positive point in the intersection. That $\frac{1}{n}$ is the minimum possible dispersion is a special case of the Cauchy-Schwarz inequality, which says that

$$\left(\sum_{i=1}^n x_i y_i\right)^2 \le \left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n y_i^2\right) \tag{3}$$

Letting each $y_i = 1$, we get

$$\left(\sum_{i=1}^n x_i\right)^2 \le \left(\sum_{i=1}^n x_i^2\right) n \tag{4}$$

and

$$\frac{1}{n} \le \sum_{j=1}^n \left(\frac{x_j}{(\sum_{i=1}^n x_i)}\right)^2 \tag{5}$$

4) If $a_i = 0$ for all i, then since we are assuming $b_i > 0$ for all i, the intersection of the intervals contains a point $x > 0$, and we are in case 3 above. In particular, we can assume that $a_i > 0$ for some i. In this situation, $S = \sum x_i$ is always positive, and *there are x_i that minimize*

ISBN: 978-989-95079-6-8

$H(\mathbf{x}) = \sum_{j=1}^{n} \left(\frac{x_j}{\sum_{i=1}^{n} x_i} \right)^2$ since H is a continuous function on the compact space $\prod_{i=1}^{n} [a_i, b_i]$.

5) An x_i does not have to be an end-point of its interval. For example, if $n = 3$, and the three intervals are disjoint, then there will be a unique solution, namely the right endpoint of the left most interval, the left endpoint of the right most interval, and generally some interior point of the middle interval. We will see an explicit example of this below.

II. A Reformulation

Based on the discussion above, we are in the following situation. We have a finite set $[a_1, b_1], [a_2, b_2], ..., [a_n, b_n]$ of closed subintervals of the unit interval $[0, 1]$, all the b_i are positive, and at least one a_i is positive. We wish to choose $\mathbf{x} = (x_1, x_2, ..., x_n)$ with each $x_i \in [a_i, b_i]$ such that $H(\mathbf{x}) = \sum_{j=1}^{n} \left(\frac{x_j}{\sum_{i=1}^{n} x_i} \right)^2$ is minimum. In fact, we will provide an algorithm that will give an exact solution in terms of the end points of the intervals $[a_i, b_i]$. First we prove a lemma that provides a reformulation into a somewhat simpler problem. We term $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ a **solution** if $H(\mathbf{x})$ is minimum.

Lemma 1: Let $x_i \in [a_i, b_i]$ with $x_1 \le x_2 \le \cdots \le x_k < x_{k+1} \le \cdots \le x_n$, and let $S = \sum_{i=1}^{n} x_i$. Then there exists $h > 0$ such that

$$\left(\frac{x_1 + h}{S + h} \right)^2 + \sum_{j=2}^{n} \left(\frac{x_j}{S + h} \right)^2 < \sum_{j=1}^{n} \left(\frac{x_j}{S} \right)^2 \qquad (6)$$

Proof: Letting $f(h) = \left(\frac{x_1+h}{S+h} \right)^2 + \sum_{j=2}^{n} \left(\frac{x_j}{S+h} \right)^2$, we have

$$\begin{aligned} f'(h) &= 2 \left(\frac{x_1 + h}{S + h} \right) \left(\frac{S + h - (x_1 + h}{(S + h)^2} \right) \\ &\quad + \sum_{j=2}^{n} 2 \left(\frac{x_j}{S + h} \right) \left(\frac{-x_j}{(S + h)^2} \right) \end{aligned} \qquad (7)$$

Now $f'(h)$ is negative if and only if

$$\begin{aligned} & (x_1 + h)(S - x_1) - \sum_{j=2}^{n} x_j^2 \qquad (8) \\ = \; & (x_1 + h) \left(\sum_{j=2}^{n} x_j \right) - \sum_{j=2}^{n} x_j^2 \\ = \; & \left(x_1 \sum_{j=2}^{n} x_j \right) - \sum_{j=2}^{n} x_j^2 + h \sum_{j=2}^{n} x_j \end{aligned}$$

is negative. Since $x_1 < x_{k+1}$, $\left(x_1 \sum_{j=2}^{n} x_j \right) - \sum_{j=2}^{n} x_j^2$ is negative, and hence h can be chosen positive and small enough so that $f'(h)$ is negative. The lemma follows. ∎

If $x_1 < b_1$, then h can be chosen so that also $x_1 + h \in [a_1, b_1]$. Thus we have the following.

Corollary 2: If $x_i \in [a_i, b_i]$ such that $x_1 \le x_2 \le ... \le x_k < x_{k+1} \le ... \le x_n$ is a solution, then $x_1 = b_1$.

Note that the proof uses the fact that x_1 is the smallest x_i. If $x_2 = x_1$, then of course $x_2 = b_2$.

For intervals $[a_i, b_i]$, let $R = \min\{b_i\}_{i=1}^{n}$ and $L = \max\{a_i\}_{i=1}^{n}$. Note that if $L \le R$, then $[L, R]$ is the intersection of the intervals $[a_i, b_i]$, and picking for each x_i the same element x in $[L, R]$ yields a solution $\mathbf{x}$ with $H(\mathbf{x}) = \frac{1}{n}$, the minimum possible value. Also, from the lemma above, we see that for $\mathbf{x}$ with at least two distinct entries, $H(\mathbf{x}) > \frac{1}{n}$. Thus *we may assume that* $R < L$. Otherwise the problem is trivial. If $R < L$, then any $\mathbf{x}$ has at least two distinct entries, and by our corollary, all $x_i \ge b_1 = R$. Similarly, we may show that all x_i are $\le L$. This means that the intervals $[a_i, b_i]$ may be replaced by the intervals $[a_i \vee R, b_i \wedge L]$. This turns out to be a significant reduction of the original problem. For example, it already provides us with two elements of a solution $\mathbf{x}$. The smallest and the largest entry in any solution will be R and L, respectively, since two intervals are $[R, R]$ and $[L, L]$. Here is our reformulation.

Reformulation: Let $0 < R < L \le 1$, and let $[a_1, b_1]$, $[a_2, b_2]$, …, $[a_n, b_n]$ be subintervals of $[R, L]$ with $a_1 = b_1 = R$ and $a_n = b_n = L$. Find $x_i \in [a_i, b_i]$ which minimizes $\sum_{i=1}^{n} \left(\frac{x_i}{\sum_{j=1}^{n} x_j} \right)^2$.

The reformulated problem has the same solution (or perhaps solutions) as the original problem. In the subintervals of $[R, L]$, one is $[R, R]$, and one is $[L, L]$. Call these $[a_1, b_1]$ and $[a_n, b_n]$ respectively. So $x_1 = a_1 = b_1$, and $x_n = a_n = b_n$, and we have two of the required x_i immediately. We assume from now on that we are in this reformulated situation. There are some additional lemmas we need.

Lemma 3: If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a solution and if $a_k < b_k = L$, then $x_k < b_k$. Similarly, if $R = a_k < b_k$, then $a_k < x_k$.

Proof: Of course, $k < n$. Let $x_i \in [a_i, b_i]$, $i = 1, 2, \ldots, n$, and suppose that $a_k < b_k = b_n = x_k = L$. Let $S = \sum_{j=1}^{n} x_j$ and consider

$$f(h) = \sum_{i=1, i \ne k}^{n} \left(\frac{x_i}{S - h} \right)^2 + \left(\frac{x_k - h}{S - h} \right)^2 \qquad (9)$$

Then a calculation shows that

$$f'(h) = \frac{2}{(S - h)^3} \left(\sum_{\substack{i=1 \\ i \ne k}}^{n} x_i^2 - x_k \sum_{\substack{i=1 \\ i \ne k}}^{n} x_i + h \sum_{\substack{i=1 \\ i \ne k}}^{n} x_i \right) \qquad (10)$$

If $h = 0$, then $f'(h) < 0$ since $x_i \le x_k$ for all i, and $x_1 = R < x_k$. For h sufficiently small and positive, $f'(h) < 0$. A similar proof works for $R = a_k < b_k$. ∎

Lemma 4: If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a solution, then for no two distinct entries $x_i < x_j$ in $\mathbf{x}$ can it be that $a_i < b_i$, $a_j < b_j$, and $x_i \in [a_i, b_i)$ and $x_j \in (a_j, b_j]$.

Proof: Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, suppose that $x_1 < x_2$, and suppose that $x_1 \in [a_1, b_1)$ and $x_2 \in (a_2, b_2]$. Consider the density given by $(x_1 + h, x_2 - h, x_3, \ldots, x_n)$. Since $x_1 + h +$

ISBN: 978-989-95079-6-8

$x_2 - h + \sum_{i=3}^{n} x_i = \sum_{i=1}^{n} x_i = S$, we have

$$f(h) = \left(\frac{x_1+h}{S}\right)^2 + \left(\frac{x_2-h}{S}\right)^2 + \sum_{j=3}^{n}\left(\frac{x_j}{S}\right)^2 \quad (11)$$

and its derivative with respect to h is

$$f'(h) = \frac{2}{S}\left(\frac{x_1+h}{S}\right) - \frac{2}{S}\left(\frac{x_2+h}{S}\right) = \frac{2}{S^2}(x_1 - x_2) \quad (12)$$

This latter quantity is negative whenever $x_2 - h > x_1 + h$. Since $x_1 \in [a_i, b_i)$ and $x_2 \in (a_2, b_2]$ we can so choose h, keeping x_1 in $[a_1, b_1]$ and x_2 in $[a_2, b_2]$, and decreasing H. Thus $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ cannot be a solution. ■

This lemma just says that if $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a solution, and $x_i < x_j$, then x_i cannot be moved to the right and x_j to the left, keeping them in their original intervals, and keeping $x_i < x_j$. That is, if $x_i < x_j$ is part of a solution, then either $x_i = b_i$ or $x_j = a_j$.

Corollary 5: If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a solution, then no two distinct entries in $\mathbf{x}$ can be in the interior of their intervals.

Corollary 6: If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a solution, and x_i is in the interior of its interval, then for every $x_j < x_i$, $x_j = b_j$ and for every $x_j > x_i$, $x_j = a_j$.

As mentioned earlier, the function

$$H(\mathbf{x}) = \sum_{j=1}^{n}\left(\frac{x_j}{\sum_{i=1}^{n} x_i}\right)^2 \quad (13)$$

assumes a minimum since it is continuous on a compact space. Whatever that minimum, it consists of some endpoints of some of the intervals $[a_i, b_i]$ and interior points of the rest. But this corollary says that all those interior points must be equal. *So any solution consists of endpoints of some of the intervals, and a common interior point of the remainder of the intervals.*

We need one more lemma.

Lemma 7: Let $0 < k < n$. Suppose that $[a, b]$ is the intersection of the intervals $[a_1, b_1]$, $[a_2, b_2]$, ..., $[a_k, b_k]$, $a < b$, and $x_i \in [a_i, b_i]$, $i = k+1, \ldots, n$. Letting $S = kx + \sum_{i=k+1}^{n} x_i$, an $x \in [a, b]$ that minimizes

$$f(x) = \left(k\left(\frac{x}{S}\right)^2 + \sum_{i=k+1}^{n}\left(\frac{x_j}{S}\right)^2\right) \quad (14)$$

is either an endpoint a or b, or an interior point x such that $f'(x) = 0$. The solution to $f'(x) = 0$ is the point

$$x_0 = \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i} \quad (15)$$

which may or may not be in the interval $[a, b]$. If $x_0 < a$, then f assumes its minimum on $[a, b]$ at a. If $x_0 > b$, then f assumes its minimum on $[a, b]$ at b. Otherwise, f assumes its minimum on $[a, b]$ at x_0.

Proof: The function f has a minimum at a, b, or at a point in the interior of the interval $[a, b]$ where the derivative of f is zero. Since the derivative of S is k, the derivative of $f(x)$ with respect to x is

$$\begin{aligned} f'(x) &= 2k\left(\frac{x}{S}\right)\left(\frac{S-kx}{S^2}\right) - 2k\sum_{i=k+1}^{n}\left(\frac{x_i}{S}\right)\left(\frac{x_i}{S^2}\right) \quad (16) \\ &= 2k\left(\frac{x}{S}\right)\left(\frac{S-kx}{S^2}\right) - 2k\sum_{i=k+1}^{n}\left(\frac{x_i}{S}\right)\left(\frac{x_i}{S^2}\right) \\ &= 2k\left(\frac{x}{S}\right)\left(\frac{\sum_{i=k+1}^{n} x_i}{S^2}\right) - 2k\sum_{i=k+1}^{n}\left(\frac{x_i}{S}\right)\left(\frac{x_i}{S^2}\right) \\ &= \left(\frac{2k}{S^3}\right)\left(x\sum_{i=k+1}^{n} x_i - \sum_{i=k+1}^{n} x_i^2\right) \end{aligned}$$

This implies that $f'(x) = 0$ only when

$$x = \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i} \quad (17)$$

The sign of the derivative $f'(x)$ is the same as the sign of $x\sum_{i=k+1}^{n} x_i - \sum_{i=k+1}^{n} x_i^2$. If $x < \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}$, then

$$\begin{aligned} x\sum_{i=k+1}^{n} x_i - \sum_{i=k+1}^{n} x_i^2 &< \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}\sum_{i=k+1}^{n} x_i - \sum_{i=k+1}^{n} x_i^2 \\ &= \sum_{i=k+1}^{n} x_i^2 - \sum_{i=k+1}^{n} x_i^2 = 0 \quad (18) \end{aligned}$$

Similarly, if $x > \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}$, then $x\sum_{i=k+1}^{n} x_i - \sum_{i=k+1}^{n} x_i^2 > 0$. Thus, $f(x)$ is decreasing when $x < \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}$, increasing when $x > \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}$,. and reaches its minimum at $x = \frac{\sum_{i=k+1}^{n} x_i^2}{\sum_{i=k+1}^{n} x_i}$. The lemma follows. ■

Note that this lemma only minimizes H *given the* x_i, $i = k+1, k+2, \ldots, n$, and note that we have assumed that $0 < k < n$.

III. An Algorithm

Corollary 5 says that any solution is an endpoint of some set of intervals and a common interior point of the intersection of the rest of the intervals. So we search out all such situations and choose the one with minimum H. The goal is to search out all such situations efficiently. Here is an algorithm that yields a solution $\mathbf{x}$ with $H(\mathbf{x})$ minimum. In the next section, we give some examples illustrating this algorithm.

An Algorithm:

1) If $\cap_{i=1}^{n}[a_i, b_i] = [a, b] \neq \varnothing$, then let each x_i be any element in $[a, b]$.
2) If $\cap_{i=1}^{n}[a_i, b_i] = \varnothing$, replace $[a_i, b_i]$ by $[a_i \vee R, b_i \wedge L]$, and call these new intervals again $[a_i, b_i]$. (This is the reformulation mentioned earlier, which says we may assume $0 < R < L \leq 1$ with $[a_1, b_1]$, $[a_2, b_2]$, ...,

ISBN: 978-989-95079-6-8

$[a_n, b_n]$ subintervals of $[R, L]$, with $a_1 = b_1 = R$, and $a_n = b_n = L$.)

3) Find the intersection of all distinct pairs of intervals from $[a_i, b_i]$, $i = 2, 3, ..., n-1$, such that these intersections are proper closed intervals; that is, are of the form $[a, b]$ with $a < b$. Let $\mathcal{I}$ be the set of these intersections together with all the proper intervals $[a_i, b_i]$. (The intersection of any nonempty family $\mathcal{S}$ of the intervals $[a_i, b_i]$ is either a member of $\mathcal{S}$ or can be obtained as the intersection of two members of $\mathcal{S}$.)
4) For each interval $I \in \mathcal{I}$, form the family $\mathcal{S}_I$ of all intervals $[a_i, b_i]$ containing I, so for each intersection $[a, b]$, $\mathcal{S}_{[a,b]}$ is the family of all the intervals in $\{[a_i, b_i] : i = 2, 3, ..., n-1\}$ containing $[a, b]$. Note that $\mathcal{S}_{[a,b]}$ has a multiplicity, namely the number of (not necessarily distinct) intervals containing $[a, b]$.
5) For those intervals $[a_i, b_i]$ not in $\mathcal{S}_I$ and such that $a_i < b_i$, choose an endpoint in the interior of $[R, L]$. If $a_i < a_j < b_i < b_j$, do not choose a_j and b_i, as such endpoints cannot be part of a solution minimizing H by Lemma 4. If no possible choices of endpoints are possible, proceed to another $\mathcal{S}_I$. Choose all endpoints that are of the form $[a, a]$, which include $[a_1, b_1]$ and $[a_n, b_n]$. This gets for each $[a_i, b_i] \notin \mathcal{S}_I$, an endpoint x_i, except for those I for which no proper selection of endpoints is possible.
6) For each $I \in \mathcal{I}$, with each possible set of endpoints gotten from step 5, use Lemma 7 to calculate the resulting H. There may be several sets of possible endpoints for a given $I \in \mathcal{I}$.
7) Repeat for all $I \in \mathcal{I}$. Now choose from the resulting candidates the one that gives minimum H.

Note that the resulting solution given by this algorithm is an exact expression in the endpoints of the given intervals. For all we know, there may be more than one solution, even when $R < L$, but we suspect not. And, of course, there may be a more efficient algorithm.

IV. Examples

We give three examples illustrating the algorithm. The first illustrates the simplest nontrivial case.

Example 8: Consider the three intervals $I_1 = [0.25, 0.33]$, $I_2 = [0.5, 0.67]$, $I_3 = [0.75, 1]$.

1) The intersection of the intervals is empty.
2) $R = 0.33$ and $L = 0.75$. Replacing the intervals by $[a_i \vee R, b_i \wedge L]$ yields the intervals

$$\begin{aligned} &[0.33, 0.33] \\ &[0.5, 0.67] \\ &[0.75, 0.75] \end{aligned} \tag{19}$$

3) $\mathcal{I}$ has only one element, namely $[0.5, 0.67]$.
4) The only interval containing $[0.5, 0.67]$ is $[0.5, 0.67]$ itself.
5) The only endpoints to be chosen are 0.33 and 0.75.
6) Lemma 7 says that the minimum H is the minimum value of

$$\left(\frac{0.33}{1.08 + x}\right)^2 + \left(\frac{x}{1.08 + x}\right)^2 + \left(\frac{0.75}{1.08 + x}\right)^2 \tag{20}$$

for $x \in [0.5, 0.67]$. Setting the derivative equal to 0 gives

$$x = \frac{0.33^2 + 0.75^2}{0.33 + 0.75} = 0.621\,67 \tag{21}$$

which is in $[0.5, 0.67]$. So by Lemma 7 the solution is $\mathbf{x} = (0.33, 0.621\,67, 0.75)$. Calculating, we get $H(0.33, 0.621\,67, 0.75) = 0.365\,33$. Notice that $0.36533 > \frac{1}{n} = \frac{1}{3}$.
7) There was only one I to deal with.

Example 9: Consider the four intervals $I_1 = [0.25, 0.33]$, $I_2 = [0.5, 0.67]$, $I_3 = [0.69, 0.72]$, $I_4 = [0.75, 1]$.

1) The intersection of the intervals is empty.
2) $R = 0.33$ and $L = 0.75$. Replacing the intervals by $[a_i \vee R, b_i \wedge L]$ yields the intervals

$$\begin{aligned} &[0.33, 0.33] \\ &[0.5, 0.67] \\ &[0.69, 0.72] \\ &[0.75, 0.75] \end{aligned} \tag{22}$$

3) $\mathcal{I}$ has two elements, namely $[0.5, 0.67]$ and $[0.69, 0.72]$.
4) The only interval containing $[0.5, 0.67]$ is $[0.5, 0.67]$ itself and the only endpoints to choose with this interval are 0.33, 0.69, and 0.75.
5) Lemma 7 says that the minimum of H for $[0.5, 0.67]$ occurs for

$$x_0 = \frac{0.33^2 + 0.69^2 + 0.75^2}{0.33 + 0.69 + 0.75} = 0.648\,305 \tag{23}$$

since $0.648\,305$ is in $[0.5, 0.67]$. So by Lemma 7 the minimum solution for $[0.5, 0.67]$ occurs for $\mathbf{x} = (0.33, 0.648\,305, 0.69, 0.75)$. Calculating, we get $H(0.33, 0.648\,305, 0.69, 0.75) = 0.268\,082$.
6) The only interval containing $[0.69, 0.72]$ is $[0.69, 0.72]$ itself and the only endpoints to choose with this interval are 0.33, 0.67, and 0.75.
7) Lemma 7 says that the minimum H for this interval occurs for

$$x_0 = \frac{0.33^2 + 0.67^2 + 0.75^2}{0.33 + 0.67 + 0.75} = 0.640\,171 \tag{24}$$

if this is in the interval, but it is to the left of the interval, and thus by Lemma 7 the minimum occurs for $x = 0.69$. So the solution is $\mathbf{x} = (0.33, 0.67, 0.69, 0.75)$. Calculating, we get $H(0.33, 0.67, 0.69, 0.75) = 0.268\,14$.
8) Notice that $0.268\,14 > 0.268\,082 > \frac{1}{n} = 0.25$. There are no more intervals to check, so the minimum is $H(\mathbf{x}) = 0.268\,082$, which occurs for $\mathbf{x} = (0.33, 0.648\,305, 0.69, 0.75)$.

From the two preceding examples, it is easy to see what happens for any family of *mutually disjoint* intervals: for each proper interval $[a_i, b_i]$, choose all the right endpoints to the left of the interval and all the left endpoints to the right of the interval. Use the point in $[a_i, b_i]$ determined by Lemma 7 and compute the resulting $H(\mathbf{x})$. Then compare these and find the minimum.

In the case the intervals are not all mutually disjoint, the situation is a bit more complicated and the more subtle points in the algorithm come into play. The next example is more illustrative of the algorithm.

Example 10: Consider the seven intervals

$$\begin{aligned} J_1 &= [0.2, 0.3] \\ J_2 &= [0.35, 0.5] \\ J_3 &= [0.15, 0.375] \\ J_4 &= [0.1, 0.3] \\ J_5 &= [0.35, 0.7] \\ J_6 &= [0.25, 0.5] \\ J_7 &= [0.4, 0.8] \end{aligned} \quad (25)$$

1) The intersection of the intervals is empty.
2) $R = 0.3$ and $L = 0.4$. Replacing $[a_i, b_i]$ by $[a_i \vee R, b_i \wedge L]$ yields the intervals

$$\begin{aligned} I_1 &= [0.3, 0.3] \\ I_2 &= [0.35, 0.4] \\ I_3 &= [0.3, 0.375] \\ I_4 &= [0.3, 0.3] \\ I_5 &= [0.35, 0.4] \\ I_6 &= [0.3, 0.4] \\ I_7 &= [0.4, 0.4] \end{aligned} \quad (26)$$

3) We get the new interval $I_8 = [0.35, 0.375]$, and $\mathcal{I} = \{I_2, I_3, I_6, I_8\}$. ($I_2 = I_5$, so we need not list I_5.)
4) For each interval $I \in \mathcal{I}$, forming the family $\mathcal{S}_I$ of all intervals containing I yields
   a) $\mathcal{S}_{I_2} = \{I_2, I_5, I_6\}$
   b) $\mathcal{S}_{I_3} = \{I_3, I_6\}$
   c) $\mathcal{S}_{I_6} = \{I_6\}$
   d) $\mathcal{S}_{I_8} = \{I_2, I_3, I_5, I_6\}$
5) For each of these four families there happens to be only one choice for endpoints for each of the intervals not in $\mathcal{S}_I$.
   a) $\mathcal{S}_{I_2} = \{I_2, I_5, I_6\}$ pairs with $\mathcal{P}_{I_2} = \{P_1 = 0.3, P_3 = 0.375, P_4 = 0.3, P_7 = 0.4\}$
   b) $\mathcal{S}_{I_3} = \{I_3, I_6\}$ pairs with $\mathcal{P}_{I_3} = \{P_1 = 0.3, P_2 = 0.35, P_4 = 0.3, P_5 = 0.35, P_7 = 0.4\}$
   c) $\mathcal{S}_{I_6} = \{I_6\}$ pairs with $\mathcal{P}_{I_6} = \{P_1 = 0.3, P_2 = 0.35, P_3 = 0.375, P_4 = 0.3, P_5 = 0.35, P_7 = 0.4\}$. But by Lemma 4, the endpoints P_2 and P_3 cannot both be part of a solution, so we discard this option.
   d) $\mathcal{S}_{I_8} = \{I_2, I_3, I_5, I_6\}$ pairs with $\mathcal{P}_{I_8} = \{P_1 = 0.3, P_4 = 0.3, P_7 = 0.4\}$
6) For each of these three remaining options $\mathcal{S}_{I_2}$, $\mathcal{S}_{I_3}$, and $\mathcal{S}_{I_8}$ we compute the values

$$f_I(x) = k_I \left(\frac{x}{S_I}\right)^2 + \sum_{x_j \in \mathcal{P}_I} \left(\frac{x_j}{S_I}\right)^2 \quad (27)$$

where k_I is the number of intervals in $\mathcal{S}_I$, $S_I = k_I x + \sum_{x_j \in \mathcal{P}_I} x_j$, and x is either an endpoint of the intersection of $\mathcal{S}_I$ that lies in the interior of $[R, L]$, or

$$x = \frac{\sum_{x_i \in \mathcal{P}_I} x_i^2}{\sum_{x_i \in \mathcal{P}_I} x_i}. \quad (28)$$

   a) For $\mathcal{S}_{I_2} = \{I_2, I_5, I_6\}$ we have $I_2 = [0.35, 0.4]$ and $\mathcal{P}_{I_2} = \{0.3, 0.375, 0.3, 0.4\}$. Then

$$\begin{aligned} x &= \frac{0.3^2 + 0.375^2 + 0.3^2 + 0.4^2}{0.3 + 0.375 + 0.3 + 0.4} \\ &= 0.349\,55 \notin [0.35, 0.4] \end{aligned} \quad (29)$$

   By Lemma 7, since $0.349\,55 < 0.35$, on $I_2 = [0.35, 0.4]$, the minimum is assumed at 0.35. We have

$$3\,(0.35) + 0.3 + 0.375 + 0.4 = 2.\,125 \quad (30)$$

   and so the value of H for this set of points is

$$\begin{aligned} f_{I_2}(0.35) &= 3\left(\frac{0.35}{2.\,125}\right)^2 + \left(\frac{0.3}{2.\,125}\right)^2 \quad (31) \\ &+ \left(\frac{0.375}{2.\,125}\right)^2 + \left(\frac{0.4}{2.\,125}\right)^2 \\ &= 0.167\,89 \end{aligned}$$

   b) For $\mathcal{S}_{I_3} = \{I_3, I_6\}$, we have $I_3 \cap I_6 = [0.3, 0.375]$ and $\mathcal{P}_{I_3} = \{0.3, 0.35, 0.3, 0.35, 0.4\}$, and

$$\begin{aligned} x &= \frac{0.3^2 + 0.35^2 + 0.3^2 + 0.35^2 + 0.4^2}{0.3 + 0.35 + 0.3 + 0.35 + 0.4} \\ &= 0.344\,12 \in [0.3, 0.375] \end{aligned} \quad (32)$$

   By Lemma 7, since $0.344\,12 \in [0.3, 0.375]$, on $[0.3, 0.375]$, the minimum is assumed at $0.344\,12$.which has multiplicity 2. We have

$$2\,(0.344\,12) + 0.3 + 0.35 + 0.3 + 0.35 + 0.4 = 2.\,388\,2 \quad (33)$$

   and so the value of H for this set of points is

$$\begin{aligned} f_{I_3}(0.344\,12) &= 2\left(\frac{0.344\,12}{2.\,388\,2}\right)^2 + \left(\frac{0.3}{2.\,388\,2}\right)^2 \\ &+ \left(\frac{0.35}{2.\,388\,2}\right)^2 + \left(\frac{0.3}{2.\,388\,2}\right)^2 \\ &+ \left(\frac{0.35}{2.\,388\,2}\right)^2 + \left(\frac{0.4}{2.\,388\,2}\right)^2 \\ &= 0.144\,09 \end{aligned} \quad (34)$$

ISBN: 978-989-95079-6-8

d. For $\mathcal{S}_{I_8} = \{I_2, I_3, I_5, I_6\}$ we have $I_2 \cap I_3 \cap I_4 \cap I_6 = [0.35, 0.375]$ and $\mathcal{P}_{I_8} = \{0.3, 0.3, 0.4\}$, and

$$x = \frac{0.3^2 + 0.3^2 + 0.4^2}{0.3 + 0.3 + 0.4} = 0.34 \notin [0.35, 0.375] \quad (35)$$

By Lemma 7, since $0.34 < 0.35$, the minimum is assumed at 0.35,which have multiplicity 4. We have

$$4\,(0.35) + 0.3 + 0.3 + 0.4 = 2.\,4 \quad (36)$$

and so the value of H for this set of points is

$$\begin{aligned} f_{I_8}(0.35) &= 4\left(\frac{0.35}{2.\,4}\right)^2 + \left(\frac{0.3}{2.\,4}\right)^2 \quad (37) \\ &+ \left(\frac{0.3}{2.\,4}\right)^2 + \left(\frac{0.4}{2.\,4}\right)^2 \\ &= 0.144\,10 \end{aligned}$$

Note that the values computed are all $> \frac{1}{n} = \frac{1}{7}$, as they should be. Comparing the values 1.6789, 1.4409, and 1.4410, we see that H assumes its minimum as indicated below.

$$\begin{aligned} 0.3 &\in [0.3, 0.3] \quad (38) \\ 0.35 &\in [0.35, 0.4] \\ 0.344\,12 &\in [0.3, 0.375] \\ 0.3 &\in [0.3, 0.3] \\ 0.35 &\in [0.35, 0.4] \\ 0.344\,12 &\in [0.3, 0.4] \\ 0.4 &\in [0.4, 0.4] \end{aligned}$$

In terms of the original intervals, we have

$$\begin{aligned} 0.3 &\in [0.2, 0.3] \quad (39) \\ 0.35 &\in [0.35, 0.5] \\ 0.344\,12 &\in [0.15, 0.375] \\ 0.3 &\in [0.1, 0.3] \\ 0.35 &\in [0.35, 0.7] \\ 0.344\,12 &\in [0.25, 0.5] \\ 0.4 &\in [0.4, 0.8] \end{aligned}$$

V. Comments

We have not shown for $R < L$ that there is a unique solution. This seems to be an interesting technical problem but may yield to more sophisticated analytical techniques. We also have not shown that our algorithm is the fastest possible. This was done in [7] for the case of computing entropy under interval uncertainty, and it may be possible to do something similar for minimizing the sum of squares.

The algorithm provided to find a solution is tedious, but could easily be programmed. It is emphasized that each member of a solution is an exact expression in terms of the endpoints of the given intervals.

There are many measures of dispersion, and in [8], several are discussed. Also, a general definition of dispersion is formulated.

References

[1] J. Abellan and S. Moral. Maximum of entropy for credal sets. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11:587-597, 2003.

[2] E. Jaynes. *Probability Theory: The Logic of Science*. Cambridge Univ. Press, 2003.

[3] A. Meyerowitz, F. Richman, E. Walker. Calculating maximum-entropy probability densities for belief functions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2:377-389, 1994.

[4] H. Nguyen and E. Walker. On decision-making using belief functions. In *Advances in the Dempster-Shafer Theory of Evidence*, R. Yager, M. Fedrizzi, and J. Kacprzyk, Eds., John Wiley & Sons: New York, 311-330, 1994.

[5] H. Nguyen and E. Walker. *A First Course in Fuzzy Logic*, 3rd ed. Chapman & Hall/CRC: Boca Raton, Florida, 2006.

[6] C. Walker, E. Walker, and R. Yager. Generating a Maximum Entropy Probability Density from an Interval-Valued Fuzzy Set. *2008 Proceedings of the NAFIPS International Conference*, New York, 2224-27, May 2008.

[7] G. Xiang, M. Ceberio, and V. Kreinovich. Population Variance and Entropy under Interval Uncertainty: Linear-Time Algorithms. *Reliable Computing*, 13(6):467-488, 2007.

[8] R. Yager. On the dispersion measure of OWA operators. *Technical Report #MII-2901*, Machine Intelligence Institute, Iona College.

ISBN: 978-989-95079-6-8