Voting Aggregation Leads to (Interval) Median

Olga Kosheleva[1] and Vladik Kreinovich[2]
[1)] *Department of Teacher Education,* [2)] *Department of Computer Science*
University of Texas at El Paso, El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu

Abstract. When we have several results of measuring or estimating the same quantities, it is desirable to aggregate them into a single estimate for the desired quantities. A natural requirement is that if the majority of estimates has some property, then the aggregate estimate should have the same property. It turns out that it is not possible to require this for *all* possible properties – but we can require it for *bounds*, i.e., for properties that the value of the quantity is in between given bounds a and b. In this paper, we prove that if we restrict the above "voting" approach to such properties, then the resulting aggregate is an (interval) median. This result provides an additional justification for the use of median – in addition to the usual justification that median is the most robust aggregate operation.

Keywords: aggregation, voting aggregation, median, interval median.

1. Formulation of the Problem

Need for aggregation. For many real-real problems, there are several different decision making tools. Each of these tools has its advantages and its limitations (otherwise, if a tool does not have any advantages, it would not be used). To combine the advantages of different tools, it therefore desirable to aggregate their results.

Voting as a natural approach to aggregation. One of the most widely used methods of aggregating several results is voting: if the majority of results satisfy a certain property, then we conclude that the actual value has this property; see, e.g., (Easley and Kleinberg, 2010; Regenwetter, 2009; Tang, 2015) and references therein.

For example, in a medical classification problem, if most classifiers classify the person's data as corresponding to pneumonia, we conclude that this person has pneumonia.

What we do in this paper. In this paper, we analyze how voting can be used to aggregate several numerical estimates.

This is not easy. To understand why this task is not easy, let us recall that a similar idea has been actively used in Artificial Intelligence.

Voting is closely related to the notions of "typical" in Artificial Intelligence. Voting is closely related to the notion of a "typical" object of a class, the notion actively studied in Artificial Intelligence. Indeed, what is an intuitive meaning of a term "typical professor"? A natural meaning is that if most professors have some property, that a "typical" professor must have this property.

For example, if most professors are absent-minded, then we expect a "typical" professor to be absent-minded as well.

If we know that a certain professor is a "typical" professor – or, in other words, not an abnormal professor – then whatever property normally holds for professors should hold for this particular professor as well. This line of reasoning is known as *non-monotonic reasoning*, and it is very important in Artificial Intelligence; see, e.g., (Halpern, 2003; Jalal-Kamai et al., 2012; Kreinovich, 2004; Kreinovich, 2012; Longpré and Kosheleva, 2012).

Related problem: no one is perfectly typical. This analogy can help us illustrate the problem related to the voting approach: while some professors may be *more or less* typical, no one is *absolutely* typical. For example, even if it turns out that we have found a professor who is typical (in the voting sense) in his/her appearance, in his/her habits, this professor's specific area of research – no matter what it is – will automatically make this professor not typical.

Indeed, it could be theoretical physics – but clearly, most professors are *not* theoretical physicists. It could be computational linguistics – but most professor are *not* computational linguists, etc.

This problem is why in Artificial Intelligence, there is a vast and ongoing literature analyzing how best to describe typical (not abnormal) objects.

2. Main Definitions and the First Result Explaining Why Voting Aggregation Is Not Easy

What is given. In the simplest case, we have several estimates $x_1, \ldots, x_n$ for the value of some physical quantity. We would like to combine these estimates into a single estimate x.

In more complex situations, we have several quantities that we would like to estimate. Let us denote the number of these quantities by q. In this case, we have several tuples $x_1, \ldots, x_n$, each of which estimates all q quantities: $x_i = (x_{i1}, \ldots, x_{iq})$. Our goal is to aggregate these estimates into a single estimate $x = (e_1, \ldots, e_q)$.

Let us describe voting aggregation in precise terms. Both for the 1-D case and for the multi-D case, we would like to select an estimate x that satisfies the following condition:

if the majority of the inputs $x_1, \ldots, x_n$ satisfies a property P,
then x should satisfy this property.

In mathematics, properties are usually described by sets: namely, each property P can be described by the set S of all the objects that satisfy this property. In these terms, the above condition takes the following form:

if the majority of the inputs $x_1, \ldots, x_n$ belong to a set S,
then x should belongs to this same set S.

In principle, we can formulate this condition for *all* possible sets, but in this case, as we have mentioned earlier, there may not exist any aggregate x that satisfies this property. Thus, it make sense to consider the possibility of restricting this condition only to sets S from a certain class $\mathcal{S}$ of sets. So, we arrive at the following definition.

Definition 1. *Let $q \geq 1$, let $\mathcal{S}$ be a class of subsets of $\mathbb{R}^q$, and let $x_1, \ldots, x_n \in \mathbb{R}^q$.*

– *We say that an element $x \in \mathbb{R}^q$ is a* possible $\mathcal{S}$-aggregate *of the elements $x_1, \ldots, x_n$ if the following condition holds:*

for every $S \in \mathcal{S}$, if the majority of x_i are in this set,
then x should be in this set.

– *The set of all possible $\mathcal{S}$-aggregates is called the* $\mathcal{S}$-aggregate *of the elements $x_1, \ldots, x_n$.*

Let us first consider the case when we allow all properties (i.e., all sets). Let us first consider the case when we allow all possible sets $S \subseteq \mathbb{R}^q$, i.e., when $\mathcal{S}$ is equal to the class $U \stackrel{\text{def}}{=} 2^{\mathbb{R}^q}$ of *all* subsets of $\mathbb{R}^q$. In this case, as the following result shows, voting aggregation does not work – since the result set is often empty:

Proposition 1. *For every $q \geq 1$ and $n \geq 3$, if all n elements $x_1, \ldots, x_n$ are different from each other, then the U-aggregate of the elements $x_1, \ldots, x_n$ is empty.*

Proof. We will prove this by contradiction.

1°. Let us assume that the U-aggregate set is not empty. This means that there is an element x which is a possible U-aggregate of $x_1, \ldots, x_n$.

2°. All elements x_i belong to the set $\{x_1, \ldots, x_n\}$; thus, the majority of elements x_i belongs to this set as well. So, by definition of a possible aggregate, x should belong to this set. Thus, we must have $x = x_i$ for some i.

3°. Let us now consider the set of all the elements $x_1, \ldots, x_n$ except for the element x_i, i.e., the set $\{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n\}$. Out of n elements $x_1, \ldots, x_n$, $n-1$ belong to this set. Since $n \geq 3$, these elements constitute the majority. Thus, by definition of a possible aggregate, the element x should belong to this new set – but since $x = x_i$, it doesn't.

This contradiction proves that the U-aggregate set is indeed empty.

Let us describe all the cases when the U-aggregate set is not empty. We can actually describe all the cases when the U-aggregate set is not empty, and explicitly describe how this aggregate set looks like. We will start with the cases $n = 1$ and $n = 2$ and then consider cases when $n \geq 3$.

Proposition 2. *For every $q \geq 1$:*

– *when $n = 1$ then the U-aggregate set of x_1 is $\{x_1\}$;*

– *when $n = 2$, then the U-aggregate set of x_1, x_2 is $\{x_1, x_2\}$.*

Proof.

1°. For $n \leq 2$, the *majority* of x_i means *all* the elements x_i. Thus, the above condition means that for every set S, if $x_i \in S$ for all i, then we should have $x \in S$. This is trivially true for all elements x_i, so all elements x_i are indeed possible U-aggregates of $x_1, \ldots, x_n$.

2°. To complete the proof, let us show that no other elements are possible U-aggregates. Indeed, if x is a possible U-aggregate, then for the set $S = \{x_1, \ldots, x_n\}$, we have $x_i \in S$ for all i, and thus, we should have $x \in S$. So, x must indeed coincide with one of the elements x_i.

The proposition is proven.

Proposition 3. *For every $q \geq 1$ and for all odd $n \geq 3$:*

- *if the majority of elements $x_1, \ldots, x_n$ are equal to each other, then the U-aggregate set of $x_1, \ldots, x_n$ consists of this common element;*
- *in all other cases, the U-aggregate set is empty.*

Proof.

1°. Let us first consider the case when the majority of elements x_i are equal to each other.

1.1°. Without losing generality, we can assume that these elements are $x_1 = \ldots = x_k$ for some $k > n/2$. In this case, the majority of elements x_i belong to the set $\{x_1\}$. Thus, every possible U-aggregate x must belong to this set, and hence, x must be equal to x_1.

1.2°. Let us now prove that the element x_1 is a possible U-aggregate.

Indeed, if the majority of elements x_i belong to the set S, this means that at least some of the elements $x_1, \ldots, x_k$ must belong to this set – otherwise, we would not have a majority. Since the element x is equal to all of them, x belongs to this set S as well. Thus, x_1 is indeed a possible U-aggregate.

2°. Let us now consider the case when we do not have a majority of elements that are equal to each other. Let us show that in this case, the U-aggregate set is indeed empty.

Indeed, let x be a possible U-aggregate set. Since all elements x_i (and, thus, the majority of them) belong to the set $\{x_1, \ldots, x_n\}$, this implies that we should have $x \in \{x_1, \ldots, x_n\}$, i.e., x must be equal to one of the original inputs: $x = x_i$ for some i.

Since we do not have a majority of elements that are equal to each other, there are fewer than $n/2$ elements which are equal to x_i. Thus, the majority of elements $x_1, \ldots, x_n$ belong to the difference set $\{x_1, \ldots, x_n\} - \{x_i\}$. So, by the definition of a possible U-aggregation, the possible U-aggregate x should also belong to this set – but since $x = x_i$, it doesn't.

The contradiction proves that in this case, the U-aggregate set is indeed empty.

Proposition 4. *For every $q \geq 1$ and for all even $n \geq 4$:*

- *if the majority of elements $x_1, \ldots, x_n$ are equal to each other, then the U-aggregate set of $x_1, \ldots, x_n$ consists of this common element;*
- *if half of the elements $x_1, \ldots, x_n$ are equal to an element a, and the other half is equal to another element b, then the U-aggregate set is equal to $\{a, b\}$;*
- *if exactly half of the elements $x_1, \ldots, x_n$ is equal to an element a, and not all other elements are equal to each other, then the U-aggregate set is equal to $\{a\}$;*

– *in all other cases, the U-aggregate set is empty.*

Proof. The first and the last cases are proven in exactly the same way as in the proof of Proposition 3. Let us therefore consider the remaining two cases.

1°. Let us first consider the case when half of the elements $x_1, \ldots, x_n$ are equal to an element a, and the other half is equal to another element b.

1.1°. In this case, all elements x_i belongs to the set $\{a, b\}$. Thus, a possible U-aggregate x should also belong to this set. Thus, it should be equal either to a or to b.

1.2°. To complete the proof for this case, we need to prove that both a and b are indeed possible U-aggregates.

Indeed, if the majority of the elements x_i belong to a set S, then it cannot be only elements which are equal to a, since they do not form the majority. Thus, the set S must contain at least one element equal to b – i.e., we must have $b \in S$.

Similarly, if the majority of the elements x_i belong to a set S, then it cannot be only elements which are equal to b, since they do not form the majority. Thus, the set S must contain at least one element equal to a – i.e., we must have $a \in S$. So, indeed, both a and b are possible U-aggregates.

So, for this case, the proposition is proven.

2°. Let us now consider the case when exactly half of the elements $x_1, \ldots, x_n$ are equal to an element a, and not all other elements are equal to each other.

Without losing generality, we can describe this case as $x_1 = \ldots = x_{n/2} = a$, and $x_i \neq a$ for $i > n/2$.

Since not all elements x_i, $i > n/2$, are equal to each other, there is an element x_j which is different from $x_{n/2+1}$ (and both are different from a).

2.1°. Let us first prove that every U-aggregate element x must be equal to a.

Indeed, in this case, at least $n/2 + 1$ (majority) of elements x_i belong to the set $\{a, x_{n/2+1}\}$, so, for every U-aggregate x, we must have $x \in \{a, x_{n/2+1}\}$. Thus, x should be either equal to a or to $x_{n/2+1}$.

Similarly, at least $n/2 + 1$ (majority) of elements x_i belong to the set $\{a, x_j\}$, so, for every U-aggregate x, we must have $x \in \{a, x_j\}$. Thus, x should be either equal to a or to x_j.

Since we have selected x_j for which $x_j \neq x_{n/2+1}$, the condition that $x = a$ or $x = x_j$ is not satisfied when $x = x_{n/2+1}$. Thus, the option $x = x_{n/2+1}$ is not possible.

So, we conclude that, indeed, every U-aggregate element x must be equal to a.

2.2°. To complete the proof, let us show that the element a is indeed a possible U-aggregate.

Indeed, if the majority of elements $x_1, \ldots, x_n$ belong to a set S, this cannot be only elements different from a, since they do not form a majority. Thus, at least one of the elements equal to a must also belong to the set S, i.e., we should have $a \in S$. Thus, a is indeed a possible U-aggregate.

Discussion. Since in general, considering all the sets does not lead to a meaningful aggregation, we have to only allow sets from a certain family.

Structure of the paper. Let us start with considering all possible intervals. In Section 3, we analyze the 1-D case. In this case, as we will show, voting aggregation results in a median. In Section 4, we show that in the 1-D case, we cannot expand beyond intervals.

In Section 5, we extend the interval result to a multi-D case – and we also show that we cannot extend beyond multi-D intervals.

3. 1-D Interval-Based Voting Aggregation Leads to (Interval) Median

Let us first clarify what we mean by intervals.

Definition 2. *By an* interval*, we mean a finite closed interval* $[a, b] = \{x : a \le x \le b\}$ *corresponding to real numbers* $a \le b$*. The class of all interval will be denoted by* I*.*

Now, we need to clarify what we mean by a median.

Definition 3. *For every sequence of real numbers* $x_1, \ldots, x_n$*, let* $x_{(1)} \le \ldots \le x_{(n)}$ *denote the result of sorting the numbers* x_i *in increasing order.*

- *when* n *is odd, i.e., when* $n = 2k + 1$ *for some integer* k*, then by a* median*, we mean the value* $x_{(k+1)}$*;*
- *when* n *is even, i.e., when* $n = 2k$ *for some integer* k*, then by a* median*, we mean the interval* $[x_{(k)}, x_{(k+1)}]$*.*

The median will also be called an interval median*.*

Proposition 5. *For every sequence of numbers* $x_1, \ldots, x_n$*, the* I*-aggregate set is equal to the median.*

Comment. Median is indeed often used in data processing, since it is the most robust aggregation – i.e., the aggregation which is the least vulnerable to possible outliers; see, e.g., (Huber, 2004; Huber and Ronchetti, 2009; Rousseeuw and Leroy, 1987). Not surprisingly, median is used in econometrics, as a more proper measure of "average" ("typical") income than the mean (OECD, 2016) – since a single billionaire living in a small town increases its mean income without affecting the living standards of its inhabitants (see also (Kreinovich, Nguyen, and Ouncharoen, 2015)).

Proof. In terms of intervals, the condition for a number x to be a possible I-aggregate takes the following form:

for every interval $[a, b]$, if the majority of the elements $x_{(i)}$ belong to this interval,
then x should also belong to this interval.

1°. Let us first prove that every possible U-aggregate x should belong to the median set.

1.1°. Indeed, if $n = 2k + 1$, then the majority of elements $x_{(i)}$ belong to the interval $[x_{(1)}, x_{(k+1)}]$: namely, $k + 1$ elements $x_{(1)} \le \ldots \le x_{(k+1)}$. Thus, every possible U-aggregate x must belong to the same interval, and thus, we must have $x \le x_{(k+1)}$.

Similarly, the majority of elements $x_{(i)}$ belong to the interval $[x_{(k+1)}, x_{(n)}]$: namely, $k+1$ elements $x_{(k+1)} \le \ldots \le x_{(n)}$. Thus, every possible U aggregate x must belong to the same interval, and thus, we must have $x \ge x_{(k+1)}$.

From $x \le x_{(k+1)}$ and $x \ge x_{(k+1)}$, we conclude that $x = x_{(k+1)}$, i.e., x coincides with the median.

1.2°. If $n = 2k$, then the majority of elements $x_{(i)}$ belong to the interval $[x_{(1)}, x_{(k+1)}]$: namely, $k+1$ elements $x_{(1)} \le \ldots \le x_{(k+1)}$. Thus, every possible U aggregate x must belong to the same interval, and thus, we must have $x \le x_{(k+1)}$.

Similarly, the majority of elements $x_{(i)}$ belong to the interval $[x_{(k)}, x_{(n)}]$: namely, $k+1$ elements $x_{(k)} \le \ldots \le x_{(n)}$. Thus, every possible U aggregate x must belong to the same interval, and thus, we must have $x \ge x_{(k)}$.

So, we conclude that $x_{(k)} \le x \le x_{(k+1)}$, i.e., that x is indeed an element of the median interval $[x_{(k)}, x_{(k+1)}]$.

2°. To complete the proof, let us prove that every element of the interval median is indeed a possible I-aggregate. For this, we need to show that if an interval $[a, b]$ contains the majority of elements $x_{(i)}$, then it contains the interval median.

Let us prove it by considering two possible situations: when n is odd and when n is even.

2.1°. Let us show that in the odd case, when $n = 2k+1$, if the interval $[a, b]$ contains the majority of the elements $x_{(i)}$, then $x_{(k+1)} \in [a, b]$, i.e., $a \le x_{(k+1)}$ and $x_{(k+1)} \le b$.

We will prove both inequalities by contradiction. If $a > x_{(k+1)}$, then the interval $[a, b]$ cannot contain any of the $k+1$ elements $x_{(1)} \le \ldots \le x_{(k+1)}$, and thus, must contain no more than k remaining elements $x_{(k+2)}, \ldots, x_{(n)}$ – which do not form a majority.

Similarly, if $b < x_{(k+1)}$, then the interval $[a, b]$ cannot contain any of the $k+1$ elements $x_{(k+1)} \le \ldots \le x_{(n)}$, and thus, must contain no more than k remaining elements $x_{(1)}, \ldots, x_{(k)}$, which also do not form a majority.

Thus, if the interval $[a, b]$ contains the majority of elements $x_{(i)}$, then it must contain the median $x_{(k+1)}$ – and so, the median is a possible I-aggregate of the values $x_1, \ldots, x_n$.

2.2°. Let us show that in the even case, when $n = 2k$, if the interval $[a, b]$ contains the majority of the elements $x_{(i)}$, then $[x_{(k)}, x_{(k+1)}] \subseteq [a, b]$, i.e., $a \le x_{(k)}$ and $x_{(k+1)} \le b$.

We will prove both inequalities by contradiction. If $a > x_{(k)}$, then the interval $[a, b]$ cannot contain any of the k elements $x_{(1)} \le \ldots \le x_{(k)}$, and thus, must contain no more than k remaining elements $x_{(k+1)}, \ldots, x_{(n)}$ – which do not form a majority.

Similarly, if $b < x_{(k+1)}$, then the interval $[a, b]$ cannot contain any of the k elements $x_{(k+1)} \le \ldots \le x_{(n)}$, and thus, must contain no more than k remaining elements $x_{(1)}, \ldots, x_{(k)}$, which also do not form a majority.

Thus, if the interval $[a, b]$ contains the majority of elements $x_{(i)}$, then it must contain the median $[x_{(k)}, x_{(k+1)}]$ – and so, every element from the interval median is a possible I-aggregate of the values $x_1, \ldots, x_n$.

The proposition is proven.

What if we require strong majority? What if instead of requiring that a typical element x satisfy all the properties that are satisfied by a *simple majority* of inputs, we instead require that the property $P(x)$ is triggered only when we have a *strong majority*: e.g., when the proportion of values x_i satisfying this property is larger than a certain threshold $t > 0.5$?

In this case, we have a similar result.

Definition 4. *Let $q \geq 1$, let $\mathcal{S}$ be a class of subsets of $\mathbb{R}^q$, let $x_1, \ldots, x_n \in \mathbb{R}^q$, and let t be a number between 0.5 and 1.*

- *We say that an element $x \in \mathbb{R}^q$ is a* possible t-$\mathcal{S}$-aggregate *of the elements $x_1, \ldots, x_n$ if the following condition holds:*

  for every $S \in \mathcal{S}$, if more than $t \cdot n$ of x_i are in this set,
  then x should be in this set.

- *The set of all possible t-$\mathcal{S}$-aggregates is called the* t-$\mathcal{S}$-aggregate set *of the elements $x_1, \ldots, x_n$.*

Proposition 6. *For every sequence of numbers $x_1, \ldots, x_n$, and for every t, the t-I-aggregate set is equal to the interval $[x_{(n-k+1)}, x_{(k)}]$, where k is the smallest integer greater than $t \cdot n$.*

Proof is similar to the proof of Proposition 5.

Another possible derivation of a median. Interval median can be also derived from other natural conditions:

- that it is a continuous function of $x_1, \ldots, x_n$,
- that it is invariant with respect to arbitrary strictly increasing or strictly decreasing re-scalings; such re-scalings make physical sense: e.g., we can measure sound energy in Watts or in decibels – which are logarithmic units; and
- that this is the narrowest such operation – else we could, e.g., take an operation returning the whole range $\left[\min\limits_i x_i, \max\limits_i x_i\right]$.

Definition 5. *Let $n \geq 1$ be fixed. By an* aggregation operation*, we mean a mapping that maps each tuple of real numbers $x_1, \ldots, x_n$ into an interval $A(x_1, \ldots, x_n) = [\underline{a}(x_1, \ldots, x_n), \overline{a}(x_1, \ldots, x_n)]$, with the following properties:*

1. *this operation is* continuous*, i.e., both functions $\underline{a}(x_1, \ldots, x_n)$ and $\overline{a}(x_1, \ldots, x_n)$ are continuous;*
2. *this operation is* scale-invariant*, meaning that:*
   - *for each strictly increasing continuous function $f(x)$, we have $\underline{a}(f(x_1), \ldots, f(x_n)) = f(\underline{a}(x_1, \ldots, x_n))$ and $\overline{a}(f(x_1), \ldots, f(x_n)) = f(\overline{a}(x_1, \ldots, x_n))$, and*
   - *for each strictly decreasing continuous function $f(x)$, we have $\underline{a}(f(x_1), \ldots, f(x_n)) = f(\overline{a}(x_1, \ldots, x_n))$ and $\overline{a}(f(x_1), \ldots, f(x_n)) = f(\underline{a}(x_1, \ldots, x_n))$;*

3. *this operation is* the narrowest *meaning that if for some operation $B(x_1,\ldots,x_n)$ that satisfies the properties* 1 *and* 2*, we have $B(x_1,\ldots,x_n) \subseteq A(x_1,\ldots,x_n)$ for all tuples, then $B(x_1,\ldots,x_n) = A(x_1,\ldots,x_n)$.*

Proposition 7. *Interval median is the only aggregation operation in the sense of Definition 5.*

Proof.

1°. One can easily check that the interval median operation $M(x_1,\ldots,x_n)$ satisfies the properties 1 and 2 from Definition 5.

2°. To complete our proof, it is thus sufficient to prove that for every operation $A(x_1,\ldots,x_n)$ that satisfies the properties 1 and 2, we have $M(x_1,\ldots,x_n) \subseteq A(x_1,\ldots,x_n)$.

Since the operation $A(x_1,\ldots,x_n)$ is continuous, and every input $(x_1,\ldots,x_n)$ with equal elements $x_i = x_j$ can be represented as a limit of inputs in which all elements are different, it is sufficient to consider the case when all the elements in the input are different, i.e., when $x_{(1)} < x_{(2)} \ldots < x_{(n)}$.

We can now form a strictly increasing transformation $f(x)$ for which $f(x_{(i)}) = x_{(i)}$, and $f(x) \neq x$ for all other numbers x. Indeed:

– for $x_{(i)} \le x \le x_{(i+1)}$, we can take

$$f(x) = x_{(i)} + \left(\frac{x - x_{(i)}}{x_{(i+1)} - x_{(i)}}\right)^2 \cdot (x_{(i+1)} - x_{(i)}),$$

– for $x \le x_{(1)}$, we can take $f(x) = x_{(1)} - 2 \cdot (x_{(1)} - x)$, and

– for $x \ge x_{(n)}$, we take $f(x) = x_{(n)} + 2 \cdot (x - x_{(n)})$.

For this function, since $f(x_i) = x_i$ for all i, the condition $\underline{a}(f(x_1),\ldots,f(x_n)) = f(\underline{a}(x_1,\ldots,x_n))$ implies that $\underline{a}(x_1,\ldots,x_n) = f(\underline{a}(x_1,\ldots,x_n))$, i.e., that $f(z) = z$ for $z = \underline{a}(x_1,\ldots,x_n)$. By our selection of the function $f(x)$, this means that $z = \underline{a}(x_1,\ldots,x_n)$ must coincide with one of the values $x_{(i)}$. Similarly, the value $\overline{a}(x_1,\ldots,x_n)$ must coincide with one of the values $x_{(j)}$. Thus, we have $A(x_1,\ldots,x_n) = [\underline{a}(x_1,\ldots,x_n), \overline{a}(x_1,\ldots,x_n)] = [x_{(i)}, x_{(j)}]$ for some i and j.

From continuity, we can conclude that the corresponding indices i and j must be the same for all the tuples $x_1,\ldots,x_n$ in which all elements are different – otherwise, we would have a discontinuity.

In particular, this means that for the tuple $x_i = i$, we have $A(1,2,\ldots,n) = [i,j]$. The strictly decreasing function $f(x) = n+1-x$ keeps the tuple intact. Thus, scale-invariance implies that $A(1,2,\ldots,n) = [n+1-j, n+1-i]$. This means that $j = n+1-i$, i.e., that we have $A(x_1,\ldots,x_n) = [x_{(i)}, x_{(n+1-i)}]$. Here, we must have $i \le n+1-i$, i.e., $2i \le n+1$ and $i \le \dfrac{n+1}{2}$.

For odd $n = 2k+1$, this means that $i \le k+1$, thus $j = n+1-i \ge k+1$, so indeed, $M(x_1,\ldots,x_n) = x_{(k+1)} \in [x_{(i)}, x_{(n+1-i)}] = A(x_1,\ldots,x_n)$.

For even $n = 2k$, this means that $i \le k$, thus $j = n+1-i \ge k+1$, so indeed, $M(x_1,\ldots,x_n) = [x_{(k)}, x_{(k+1)}] \subseteq [x_{(i)}, x_{(n+1-i)}] = A(x_1,\ldots,x_n)$.

The proposition is proven.

4. 1-D Case: Can We Expand Beyond Intervals?

One can easily check that in our analysis, instead of closed intervals, we can consider general *convex subsets* of the real line – i.e., subsets S that, for every two real numbers a and x', contain all the numbers between x and x'. In addition to closed intervals, convex sets include open intervals, semi-open intervals, and intervals with infinite endpoints.

Can we go beyond intervals? It turns out that we cannot, as the following result shows.

Proposition 8. *Let a class $\mathcal{S}$ contain, in addition to all the intervals, a non-convex set S_0. Then, for $n = 3$ and for every $n \geq 5$, there exists values $x_1, \ldots, x_n$ for which the $\mathcal{S}$-aggregate set is empty.*

Proof.

1°. The fact that the set S_0 is not convex means that there exist points $a < b < c$ for which $a, c \in S_0$ but $b \notin S_0$. To construct the desired counterexample, we will then form sequences x_i in which some elements are equal to a, some to b, and some to c.

We will consider three possible cases: when $n = 3k$, when $n = 3k + 1$, and when $n = 3k + 2$.

2°. Let us first consider the case when $n = 3k$. In this case, we take k values equal to a, k values equal to b, and k values equal to c.

$2k$ of these values form a majority. Thus, the majority of elements x_i belong to the interval $[a, b]$, so any possible $\mathcal{S}$-aggregate x must also belong to this interval. Similarly, the majority of elements x_i belongs to the interval $[b, c]$, so x must also belong to the interval $[b, c]$. From $x \in [a, b]$ and $x \in [b, c]$, we conclude that $x \in [a, b] \cap [b, c] = \{b\}$, i.e., that $x = b$. However, also, the majority of elements are equal to a or to c and thus, belong to the set S_0. So, we should conclude that $x \in S_0$ – but the element $x = b$ does not belong to S_0.

This contradiction shows that in this case, $\mathcal{S}$-aggregate set is indeed empty.

3°. When $n = 3k + 1$ and $n > 4$, we get $k \geq 2$. In this case, we select k elements equal to a, k elements equal to b, and $k + 1$ elements equal to c. If we pick only two of these three groups of elements, we get at least $2k$ elements. So, to continue with the arguments similar to what we had in Part 2 of this proof, it is sufficient to make sure that $2k$ elements form a majority, i.e., that $2k > \dfrac{3k+1}{2}$. This inequality is equivalent to $4k > 3k + 1$ and to $k > 1$ and is, thus, true. So, this case is proven as well.

4°. When $n = 3k + 2$ and $n > 4$, we get $k \geq 1$. In this case, we select k elements equal to a, $k + 1$ elements equal to b, and $k + 1$ elements equal to c. If we pick only two of these three groups of elements, we get at least $2k + 1$ elements. So, to continue with the arguments similar to what we had in Part 2 of this proof, it is sufficient to make sure that $2k + 1$ elements form a majority, i.e., that $2k + 1 > \dfrac{3k+2}{2}$. This inequality is equivalent to $4k + 2 > 3k + 2$ and to $k > 0$ and is, thus, always true. So, this case is proven as well.

The proposition is proven.

Discussion. The above proposition excluded values $n = 1$, $n = 2$, and $n = 4$. For $n = 1$ and $n = 2$, as we have mentioned earlier, all the values x_i are possible aggregates even when we consider all possible sets S.

For $n = 4$, it is also possible to have a non-convex set S_0 for which the $(I \cup \{S_0\})$-aggregate set is always non-empty.

Proposition 9. *For $n = 4$, there exists a non-convex set S_0 for which the $(I \cup \{S_0\})$-aggregate set is always non-empty.*

Proof. Indeed, let us take a 2-point set $S_0 = \{0, 1\}$. In this case, majority means at least 3 elements. So, we must consider tuples of 4 elements in which 3 are equal to 0 or 1.

For $n = 4$, the interval median is the interval $[x_{(2)}, x_{(3)}]$. All these elements are possible I-aggregates. So, prove our result, it is sufficient to show that at least one of them is also a possible $\{S_0\}$-aggregate.

If two or fewer element x_i are equal to 0 or 1, then the S_0-related condition does not require anything from a possible aggregate element. The only time when this condition needs to be taken into account is when 3 out of the 4 elements $x_{(1)}, \ldots, x_{(4)}$ are equal to 0 or 1. In this case, however, at least one of the bounds $x_{(2)}$ and $x_{(3)}$ is equal to 0 or 1, and thus, belongs to the set S_0. This bound is therefore a possible $\{S_0\}$-aggregate – thus, the $(I \cup \{S_0\})$-aggregate set is indeed non-empty.

The proposition is proven.

5. Multi-D Interval-Based Voting Aggregation

Discussion. Let us now consider the multi-D situation. In this case, a natural multi-D analog of intervals are boxes.

Definition 6. *For every $q \geq 1$, by a* box*, we mean a set $[a_1, b_1] \times \ldots \times [a_q, b_q]$, where $[a_i, b_i]$ are intervals. The class of all boxes will be denoted by B.*

Proposition 10. *For every sequence of tuples $x_1, \ldots, x_n$, the B-aggregate set is the box*

$$M_1 \times \ldots \times M_q,$$

where for every i, M_i is the interval median of the i-th components $x_{1i}, \ldots, x_{ni}$.

Proof.

$1°$. Let us first prove that every possible B-aggregate tuple belongs to the median box $M_1 \times \ldots \times M_q$.

Let us fix one of the dimensions. Without losing generality, we can assume that this dimension is the first one. Then, we consider numbers $x_{11}, \ldots, x_{n1}$.

For all other dimensions $j \neq 1$, let us consider the smallest possible intervals

$$[A_j, B_j] \stackrel{\text{def}}{=} \left[\min_k(x_{kj}), \max_k(x_{kj})\right]$$

that contain all given values x_{kj}.

For each possible B-aggregate tuple $x = (e_1, \ldots, e_q)$, the desired property holds for all the boxes of the type $[a_1, b_1] \times [A_2, B_2] \times \ldots \times [A_q, B_q]$. Since all other intervals forming this box are the largest possible, the condition that a tuple x_i is contained in this box is equivalent to the condition that $x_{i1} \in [a_1, b_1]$.

Thus, for these boxes, the definition of a possible B-aggregate of the tuples $x_1, \ldots, x_n$ implies that the first component e_1 of the tuple x is a possible I-aggregate of the components $x_{11}, \ldots, x_{n1}$. We already know that this implies that e_1 belongs to the interval median M_1 of these components.

Similarly, we can prove that e_2 belongs to M_2, etc., thus indeed $x \in M_1 \times \ldots \times M_q$.

2°. Vice versa, let us prove that every tuple $x \in M_1 \times \ldots \times M_q$ is a possible B-aggregate.

Indeed, let $x = (e_1, \ldots, e_q) \in M_1 \times \ldots \times M_q$, and let us assume that the majority of the tuples x_i belong to the box $B = [a_1, b_1] \times \ldots \times [a_q, b_q]$. This implies, for every component i, that the majority of the values $x_{1i}, \ldots, x_{ni}$ belong to the interval $[a_i, b_i]$. We already know, from the 1-D case, that this implies that $e_i \in [a_i, b_i]$ for every i. Thus, we indeed have

$$x = (e_1, \ldots, e_q) \in [a_1, b_1] \times \ldots \times [a_q, b_q] = B.$$

The proposition is proven.

Can we replace boxes with more general sets? Can we use more general sets, e.g., convex polytopes? In general, no, and here is a simple 2-D counter-example.

Definition 7. *Let P denote the class of all convex polytopes.*

Proposition 11. *For $q \geq 2$, and for $n = 3$ or $n \geq 5$, there exist an input $x_1, \ldots, x_n$ for which the P-aggregate set is empty.*

Proof. For $n = 3$, let us have $x_1 = (0, 0, 0, \ldots, 0)$, $x_2 = (0.1, 0.9, 0, \ldots, 0)$, and $x_3 = (1, 1, 0, \ldots, 0)$. Clearly here,

- the median M_1 of the first components 0, 0.1, and 1 is 0.1, and
- the median M_2 of the second components 0, 0.9, and 1 is 0.9.

For all other components, the median of the values 0, 0, and 0 is clearly 0.

Thus, the median box $M_1 \times M_2 \times M_3 \times \ldots \times M_q$ consists of a single point $x_2 = (0.1, 0.9, 0, \ldots, 0)$.

Here, the majority of the points (namely, x_1 and x_3) belong to the convex straight line segment $S = \{(x, x, 0, \ldots, 0) : 0 \leq x \leq 1\}$, but the median does not belong to this segment – and thus, if we add this segment to boxes, the resulting aggregate set will be empty.

For general n, we can have several points equal to $x_1 = (0, 0, 0, \ldots, 0)$, several points equal to $x_1 = (0.1, 0.9, 0, \ldots, 0)$, and several points equal to $x_3 = (1, 1, \ldots, 0)$ – just as we had in the proof of Proposition 8.

The proposition is proven.

Acknowledgements

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, and by an award "UTEP and Prudential Actuarial Science Academy and Pipeline Initiative" from Prudential Foundation.

The authors are thankful to the anonymous referees for valuable suggestions.

References

Easly, D. and J Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, Cambridge University Press, 2010.

Halpern, J. Y. *Reasoning about Uncertainty*, MIT Press, 2003.

Huber, P. J., and E. M. Ronchetti, *Robust Statistics*, Wiley, Hoboken, New Jersey, 2009.

Huber, P. J., *Robust Statistics*, Wiley, Hoboken, New Jersey, 2004.

Jalal-Kamali, A., O. Nebesky, M. H. Durcholz, V. Kreinovich, and L. Longpré, Towards a "generic" notion of genericity: from "typical" and "random" to meager, shy, etc., *Journal of Uncertain Systems*, 6(2):104–113, 2012.

Kreinovich, V. Toward formalizing non-monotonic reasoning in physics: the use of Kolmogorov complexity and algorithmic information theory to formalize the notions 'typically' and 'normally' ", In: L. Sheremetov and M. Alvarado (eds.), *Proceedings of the Workshops on Intelligent Computing WIC'04 associated with the Mexican International Conference on Artificial Intelligence MICAI'04*, Mexico City, Mexico, April 26–27, 2004, pp. 187–194.

Kreinovich, V. Negative results of computable analysis disappear if we restrict ourselves to random (or, more generally, typical) inputs, *Mathematical Structures and Modeling*, 25:100–113, 2012.

Kreinovich, V., H. T. Nguyen, and R. Ouncharoen, From mean and median income to the most adequate way of taking inequality into account, In: V.-N. Huynh, V. Kreinovich, S. Sriboonchitta, and K. Suriya (eds.), *Econometrics of Risk*, Springer Verlag, Berlin, Heidelberg, 2015, pp. 63–73.

Longpré, L., and O. Kosheleva, Towards unique physically meaningful definitions of random and typical objects, *Mathematical Structures and Modeling*, 26:49–56, 2012.

Organization for Economic Cooperation and Development (OECD), *OECD Statistics*, available at http://stats.oecd.org/

Regenwetter, M. Perspectives on preference aggregation. *Perspectives on Psychological Science*, 4(4):403–407 (2009).

Rousseeuw, P., and A. Leroy, *Robust Regression and Outlier Detection*, J. Wiley, New York, 1987.

Tang, K. Evolutionary ensemble learning in uncertain environments, *Proceedings of the 2015 IEEE Series of Symposia on Computational Intelligence SSCI'2015*, Cape Town, South Africa, December 7–10, 2015.