5-2019

Accuracy of Data Fusion: Interval (and Fuzzy) Case

Christian Servin
El Paso Community College, cservin@gmail.com

Olga Kosheleva
University of Texas at El Paso, olgak@utep.edu

Vladik Kreinovich
University of Texas at El Paso, vladik@utep.edu

Comments:
Technical Report: UTEP-CS-19-48

Accuracy of Data Fusion: Interval (and Fuzzy) Case

Christian Servin, Olga Kosheleva, and Vladik Kreinovich

Abstract The more information we have about a quantity, the more accurately we can estimate this quantity. In particular, if we have several estimates of the same quantity, we can fuse them into a single more accurate estimate. What is the accuracy of this estimate? The corresponding formulas are known for the case of probabilistic uncertainty. In this paper, we provide similar formulas for the cases of interval and fuzzy uncertainty.

1 Formulation of the Problem

Need for estimation. In the grand scheme of things, the main objectives of science and engineering are: to get a good understanding of the current state of the world and how this state will change, and to come up with recipes of how to make sure that this change will go in favorable directions. To describe the state of the world, we need to describe the numerical values of all the physical quantities. To describe the corresponding recommendations, we need to describe the numerical values of all the parameters of the corresponding designs and/or of the appropriate controls – and these values usually depend on the parameters describing the current state of the world.

We can estimate the corresponding physical quantities directly, by measuring them or by relying on experts who provide the related estimates. If this is not possible, we can estimate the desired quantities indirectly, by measuring and/or estimat-

Christian Servin
Computer Science and Information Technology Systems Department
El Paso Community College, 919 Hunter, El Paso, TX 79915, USA, e-mail: cservin@gmail.com

Olga Kosheleva and Vladik Kreinovich
University of Texas at El Paso, El Paso, Texas 79968, USA
e-mail: olgak@utep.edu, vladik@utep.edu

ing related quantities and using the known relation between all these quantities to compute the corresponding estimates of the desired quantities; see, e.g., [11].

Need to improve accuracy and need for data fusion. Sometimes, the available estimates already have the needed accuracy. However, in many other cases, even the measurements by state-of-the-art measuring instruments is not accurate enough. For example, in geophysics, we can use different techniques to estimate the density and other characteristics several kilometers beneath the Earth surface; however, in many cases, these estimates are not sufficient to predict the location of mineral deposits or to predict an earthquake.

In such situations, a natural way to improve accuracy is to perform more estimations of the same quantity, and then to combine ("fuse") the resulting estimates into a single more accurate one.

How accurate is the result of data fusion? A natural question is: how accurate is the result of the data fusion? The answer to this question is known for probabilistic uncertainty; we briefly describe this answer in Section 2. However, for important cases of interval and fuzzy uncertainty, no such general formulas have been known. In Section 3 of this paper, we show how to derive the corresponding formulas.

2 Accuracy of Data Fusion: Case of Probabilistic Uncertainty

Towards formulating the problem in precise terms. Let us assume that we have n estimates $x_1, \ldots, x_n$ of the same quantity x. Probabilistic uncertainty means that we know the probability distributions of the corresponding approximation errors

$$\Delta x_i \stackrel{\text{def}}{=} x_i - x.$$

We are interested in values x_i obtained by state-of-the-art measurements – otherwise, there is no big need for fusion, we could simply use more accurate measuring instruments. State-of-the-art means, in particular, that all the usual ways to improve accuracy have already been applied.

For example, if a measuring instrument has a bias, i.e., if the mean value of the approximation error is different from 0, then we can detect this bias if we *calibrate* the instrument – i.e., if we repeatedly compare its results with the results of measuring the same quantity by a stationary "standard" (maximally accurate) measuring instrument. Once the bias is known, we can simply subtract this bias from all the measurement results produced by this instrument. Thus, when we deal with state-of-the-art measuring instruments, we can safely assume that the bias is 0, i.e., that the mean value of the estimation error is 0.

Similarly, we can safely assume that all known major sources of possible measurement errors have been taken care of. For example, many accurate measurements are affected by the periodic 50 or 60 Hz electromagnetic signals emitted by the usual electric outlets – so we can safely assume that these signals have been screened

away. In general, we can safely assume that all reasonably major sources of measurement errors have been eliminated and thus, that the remaining estimate comes from the joint effect of a large number of small error components. It is known that in such situations, the resulting error distribution is close to Gaussian – this follows from the Central Limit Theorem (see, e.g., [12]), according to which, when the number N of such small components tends to infinity, the distribution of the sum of these components tends to Gaussian. So, we can safely assume that each estimation error Δx_i is normally distributed.

As we have mentioned, probabilistic uncertainty means that we know the probability distribution of each estimation error. A normal distribution is uniquely determined by its mean and its standard deviation. We know that the mean of Δx_i is 0, so knowing the distribution simply means that we know the standard deviation σ_i of the corresponding estimate.

It is also reasonable to assume that measurement errors corresponding to different measurements are independent.

In these terms, the problem has the following form.

Formulation of the problem.

- We have n estimates $x_1, \ldots, x_n$ of the same quantity y;
- we know that for each i, the measurement error $x_i - x$ is normally distributed with 0 means and known standard deviation σ_i;
- we also know that the measurement errors corresponding to different distributions are independent.

Based on this information, we would like to find a combined estimate $\widetilde{x}$ and estimate how accurate is this estimate.

Fusing measurement results: derivation of the formula. Since each measurement error $x_i - x$ is normally distributed, the corresponding probability density function (pdf) $\rho_i(x)$ has the form

$$\rho_i(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma_i} \cdot \exp\left(-\frac{(x_i - x)^2}{2\sigma_i^2}\right). \tag{1}$$

Since the measurement errors corresponding to different measurements are independent, the overall pdf is equal to the product of the corresponding probability densities:

$$\rho(x) = \prod_{i=1}^{n} \rho_i(x) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi} \cdot \sigma_i} \cdot \exp\left(-\frac{(x_i - x)^2}{2\sigma_i^2}\right)\right). \tag{2}$$

As the desired fused estimate $\widetilde{x}$, it is reasonable to select the most probable value x, i.e., the value for which the expression (2) attains its largest possible value; this idea is known as *Maximum Likelihood Method*.

This maximization problem can be simplified if we take into account that the function $f(z) = -\ln(z)$ is strictly decreasing. Thus, maximizing the expression (2) is equivalent to minimizing its negative logarithm

$$-\ln(\rho(x)) = \text{const} + \sum_{i=1}^{n} \frac{(x_i - x)^2}{2\sigma_i^2}, \tag{3}$$

where const denotes terms that do not depend on x.

Differentiating the expression (3) with respect to the unknown x and equating the derivative to 0, we conclude that

$$\sum_{i=1}^{n} \frac{x - x_i}{\sigma_i^2} = 0, \tag{4}$$

i.e., equivalently, that

$$x \cdot \sum_{i=1}^{n} \sigma^{-2} = \sum_{i=1}^{n} x_i \cdot \sigma_i^{-2}. \tag{5}$$

Thus, we arrive at the following formula:

Fusion in the probabilistic case: final formula. When we fuse measurement results $x_1, \ldots, x_n$ measured with accuracies $\sigma_1, \ldots, \sigma_n$, we get the estimate

$$\widetilde{x} = \frac{\sum\limits_{i=1}^{n} x_i \cdot \sigma_i^{-2}}{\sum\limits_{i=1}^{n} \sigma_i^{-2}}. \tag{6}$$

How accurate is the fused estimate? The probability distribution for different values x is given by the formula (2), i.e., equivalently, by the formula

$$\rho(x) = \left(\frac{1}{\sqrt{2\pi} \cdot \prod\limits_{i=1}^{n} \sigma_i} \cdot \exp\left(-\sum_{i=1}^{n} \frac{(x_i - x)^2}{2\sigma_i^2} \right) \right). \tag{7}$$

One can easily see that the expression under the exponent is a quadratic function of x, and thus, the distribution for x is also Gaussian, i.e., have the form

$$\rho(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(\widetilde{x} - x)^2}{2\sigma^2} \right). \tag{8}$$

Comparing the coefficients for x^2 under the exponential function is expressions (7) and (8), we conclude that

$$\sum_{i=1}^{n} \frac{1}{\sigma_i^2} = \frac{1}{\sigma^2}, \tag{9}$$

i.e., that

$$\sigma^2 = \frac{1}{\sum\limits_{i=1}^{n} \frac{1}{\sigma_i^2}}. \tag{10}$$

In particular, for $n = 2$, we get

$$\sigma^2 = \frac{1}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}} = \frac{\sigma_1^2 \cdot \sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \tag{11}$$

Another important case is when all the measurements have the same accuracy, i.e., when $\sigma_1 = \ldots = \sigma_n$. In this case, according to (10), we have $\frac{1}{\sigma^2} = \frac{n}{\sigma_1^2}$, so $\sigma^2 = \frac{\sigma_1^2}{n}$ and $\sigma = \frac{\sigma_1}{\sqrt{n}}$.

3 Case of Interval Uncertainty

Need for interval uncertainty. The traditional probabilistic approach to measurement processing is based on the assumption that we know the probability distribution of measurement errors $\Delta x_i \stackrel{\text{def}}{=} x_i - x$.

Usually, this distribution is obtained by *calibrating* the measuring instrument, i.e., by comparing the values x_i measured by this instrument with the values measured by a much more accurate ("standard") instrument – whose measurement results are so much more accurate that we can safely ignore the difference between these results and the actual values x of the corresponding quantity.

However, there are important cases when calibration is not done. The first such case is the case of state-of-the-art measurements, when we use the most accurate measuring instruments. In this case, there are simply no more-accurate instruments which can be used for calibration, so calibration is not possible. At best, we can provide an upper bound Δ_i on the corresponding measurement error $\Delta x_i = x_i - x$: $|\Delta x_i| \le \Delta_i$. In this case, once we know the measurement result x_i, the only information that can conclude about the actual value x is that this value belongs to the interval $[x_i - \Delta_i, x_i + \Delta_i]$.

Another case is measurements on the shop floor, during the manufacturing process. In this case, theoretically, we could calibrate every single sensor, every single measuring instrument. However, calibration is a rather expensive procedure – since it involves the use of a complex standard measuring instrument. As a result, in manufacturing, such calibration is not done, and often, the only information that we have about a measuring instrument is the upper bound on its measurement error. (And if we do not even know any such upper bound, then this is not a measuring instrument at all – since the actual value can be anything, no matter what the instrument shows.)

How to fuse measurement results in case of interval uncertainty. Based on each measurement in which we get the value x_i with accuracy Δ_i, we conclude that the actual value x belongs to the interval $[x_i - \Delta_i, x_i + \Delta_i]$; see, e.g., [2, 5, 7, 11]. Thus, if we have n such measurement results, we can conclude that x belongs to the intersection

of all these n intervals, i.e., to the interval

$$[\underline{x},\overline{x}] = \bigcap_{i=1}^{n} [x_i - \Delta_i, x_i + \Delta_i] = \left[\max_i (x_i - \Delta_i), \min_i (x_i + \Delta_i)\right]. \tag{12}$$

We need to estimate *average* accuracy. Each value x_i is the result of measuring the desired quantity x with accuracy Δ_i. Thus, each value x_i can take any value from the interval $[x - \Delta_i, x + \Delta_i]$.

For the same measurement errors of two measurements, we can get different accuracies of the fusion result. For example, if we fuse the results of two measurements performed with the same accuracy Δ, then we can have two extreme situations:

- We can have the exact same measurement result in both cases $x_1 = x_2$. In this case, the corresponding intervals are the same, and their intersection is the exact same interval $[x_1 - \Delta, x_1 + \Delta]$. Thus, in this case, fusion does not improve the accuracy at all.
- On the other hand, we may have $x_1 = x + \Delta$ and $x_2 = x - \Delta$. In this case, based on the first measurement result, we conclude that the actual (unknown) value x belongs to the interval $[x_1 - \Delta, x_1 + \Delta] = [x, x + 2\Delta]$, and based on the second measurement result, we conclude that the actual (unknown) value x belongs to the interval $[x_2 - \Delta, x_2 + \Delta] = [x - 2\Delta, x]$. The intersection of these two intervals is a single point x – i.e., in this case, by fusing two measurement result, we get the exact value of the measured quantity.

Since the accuracy of the fused result depends on the actual measurement results, the only think that we can estimate is the *average* value of the corresponding estimation error.

What probability distribution should we use? For each measurement, all we know about the measurement error is that it is located somewhere between $-\Delta_i$ and Δ_i. There is no reason to believe that some values from this interval $[-\Delta_i, \Delta_i]$ are more probable than others. Thus, it is reasonable to conclude that all these values should have the exact same probability – i.e., that the measurement error should be uniformly distributed on the corresponding interval. This natural conclusion is known as *Laplace Indeterminacy Principle*; see, e.g., [3].

Similarly, if we have n measurements, then all we know about n corresponding measurement errors $\Delta x_1, \ldots, \Delta x_n$ is that the corresponding vector $(\Delta x_1, \ldots, \Delta x_n)$ is located somewhere in the box

$$[-\Delta_1, \Delta_1] \times \ldots \times [-\Delta_n, \Delta_n].$$

Thus, it is reasonable to conclude that we have a uniform distribution on this box – i.e., in other words, that all measurement errors are independent random variables.

What is known and what we will do in this paper. It is known (see, e.g., [13]) that if we fuse several interval estimates with the same accuracy $\Delta_1 = \ldots = \Delta_n$, then the average accuracy of the fused result is, for large n, asymptotically equal to $\dfrac{\Delta_1}{n}$.

We can see that, in this case, the average measurement error decreases more than in the probabilistic case, when the average measurement error decreases as $1/\sqrt{n}$ (much slower).

In this paper, we consider the general case of possibly different accuracies Δ_i.

Case of two fused measurements $n = 2$: analysis of the problem. Let us start with the simplest case when we fuse two measurement results x_1 and x_2, measured with accuracies, correspondingly, Δ_1 and Δ_2. Without losing generality, let us assume that the measurements are numbered from the most accurate one to the least accurate one, i.e., in this case, that $\Delta_1 \le \Delta_2$.

According to the formula (12), based on the results of these two measurements, we can compute the following upper bound u on the actual value x of the measured quantity

$$u = \min(x_1 + \Delta_1, x_2 + \Delta_2). \tag{13}$$

The accuracy of the measurement can be described by the difference $u - x$ between this upper bound and the actual (unknown) value x. We are interested in the average (expected) value Δ of the difference $u - x$.

Due to symmetry with respect to the change of $x \to -x$ that swaps lower and upper bounds, we will get the exact same average value for the difference between x and the lower bound $\ell = \max(x_1 - \Delta_1, x_2 - \Delta_2)$. Thus, it is sufficient to compute the expected value Δ of the upper-bound difference $u - x$.

To compute this average value, let us find the probability distribution of the difference $\Delta x \stackrel{\text{def}}{=} u - x$. Here, for both i, we have $x_i = x + \Delta x_i$, thus,

$$u = \min(x + \Delta x_1 + \Delta_1, x + \Delta x_2 + \Delta_2) = x + \min(\Delta x_1 + \Delta_1, \Delta x_2 + \Delta_2). \tag{14}$$

Thus,

$$\Delta x = u - x = \min(\Delta x_1 + \Delta_1, \Delta x_2 + \Delta_2). \tag{15}$$

Let us compute, for each real number z, the probability

$$\text{Prob}(z \le \Delta x) = \text{Prob}(z \le \min(\Delta x_1 + \Delta_1, \Delta x_2 + \Delta_2)). \tag{16}$$

It is easy to see that z is smaller than the minimum of the two numbers if and only it is smaller than both of them, so

$$\text{Prob}(z \le \Delta x) = \text{Prob}(z \le \Delta x_1 + \Delta_1 \,\&\, z \le \Delta x_2 + \Delta_2). \tag{17}$$

Let us rewrite each of the resulting inequalities so that it will have Δx_i on one side on the corresponding inequality and all other terms on the other side. Then, we get

$$\text{Prob}(z \le \Delta x) = \text{Prob}(\Delta x_1 \ge z - \Delta_1 \,\&\, \Delta x_2 \ge z - \Delta_2). \tag{18}$$

Since we assumed that the measurement errors Δx_1 and Δx_2 are independent, the probability that both inequalities (related to both measurement errors) hold is simply equal to the product of the probabilities corresponding to individual measurement errors:

$$\text{Prob}(z \le \Delta x) = \text{Prob}(\Delta x_1 \ge z - \Delta_1) \cdot \text{Prob}(\Delta x_2 \ge z - \Delta_2). \tag{19}$$

Each of the measurement errors Δx_i is uniformly distributed on the corresponding interval $[-\Delta_i, \Delta_i]$ of width $2\Delta_i$. Uniform distribution means that the probability to be on each subinterval is proportional to the width of this subinterval – to be precise, it is equal to the ratio of the width of the subinterval to the width of the original interval.

For each threshold z_i, the inequality $\Delta x_i \ge z_i$ is satisfied on the subinterval $[z_i, \Delta_i]$ of width $\Delta_i - z_i$. Thus, the probability that this inequality is satisfied is equal to the ratio $\dfrac{\Delta_i - z_i}{2\Delta_i}$. In particular, for $z_i = z - \Delta_i$, we get

$$\text{Prob}(z \ge z - \Delta_i) = \frac{\Delta_i - (z - \Delta_i)}{2\Delta_i} = \frac{2\Delta_i - z}{2\Delta_i}. \tag{20}$$

Thus, based on the formula (19), we have

$$\text{Prob}(z \le \Delta x) = \frac{2\Delta_1 - z}{2\Delta_1} \cdot \frac{2\Delta_2 - z}{2\Delta_2} = \frac{(2\Delta_1 - z) \cdot (2\Delta_2 - z)}{4\Delta_1 \cdot \Delta_2}. \tag{21}$$

So, the cumulative distribution function (cdf) $F(z) = \text{Prob}(\Delta x \le z)$ is equal to

$$F(z) = 1 - \text{Prob}(z \le \Delta x) = 1 - \frac{(2\Delta_1 - z) \cdot (2\Delta_2 - z)}{4\Delta_1 \cdot \Delta_2}. \tag{22}$$

The corresponding probability density function $\rho(z)$ can be obtained if we differentiate the cdf:

$$\rho(z) = \frac{dF(z)}{dz} = \frac{(2\Delta_1 - z) + (2\Delta_1 - z)}{4\Delta_1 \cdot \Delta_2} = \frac{2\Delta_1 + 2\Delta_2 - 2z}{4\Delta_1 \cdot \Delta_2} = \frac{\Delta_1 + \Delta_2 - z}{2\Delta_1 \cdot \Delta_2}. \tag{23}$$

The difference $\Delta x = u - x$ is always greater than or equal to 0 – since u is the upper bound for x. This difference cannot exceed the value $2\Delta_1$ corresponding to the worst-possible case of the first measurement when $x_1 = x + \Delta_1$. Thus, the difference Δx takes all possible value from the interval $[0, 2\Delta_1]$. The average (expected) value Δ of this difference can thus be computed as

$$\Delta = \int_0^{2\Delta_1} z \cdot \rho(z)\, dz = \frac{1}{2\Delta_1 \cdot \Delta_2} \cdot \int_0^{2\Delta_1} [z \cdot (\Delta_1 + \Delta_2) - z^2]\, dz =$$

$$\frac{1}{2\Delta_1 \cdot \Delta_2} \cdot \left[\frac{z^2}{2} \cdot (\Delta_1 + \Delta_2) - \frac{z^3}{3}\right]_0^{2\Delta_1} = \frac{1}{2\Delta_1 \cdot \Delta_2} \cdot \left[\frac{4\Delta_1^2}{2} \cdot (\Delta_1 + \Delta_2) - \frac{8\Delta_1^3}{3}\right] =$$

$$\frac{1}{2\Delta_1 \cdot \Delta_2} \cdot \left(2\Delta_1^3 + 2\Delta_1^2 \cdot \Delta_2 - \frac{8}{3} \cdot \Delta_1^3\right) = \frac{2\Delta_1^2 \cdot \Delta_2 - \dfrac{2}{3} \cdot \Delta_1^3}{2\Delta_1 \cdot \Delta_2} = \Delta_1 - \frac{1}{3} \cdot \frac{\Delta_1^2}{\Delta_2}. \tag{24}$$

Thus, we arrive at the following conclusion.

Case of two fused measurements $n = 2$: conclusion. When we fuse two measurements with interval uncertainties $\Delta_1 \le \Delta_2$, then the average accuracy Δ of the fused result is

$$\Delta = \Delta_1 - \frac{1}{3} \cdot \frac{\Delta_1^2}{\Delta_2}. \tag{25}$$

In particular, when $\Delta_2 \to \infty$, we get $\Delta \to \Delta_1$, which makes perfect sense: very inaccurate measurements do not add any information, so accuracy is not improved.

When $\Delta_1 = \Delta_2$, we get $\Delta = \frac{2}{3} \cdot \Delta_1$. In other words, the average inaccuracy decreases by a factor of 1.5 – as opposed to the probabilistic case, when it only decreases by a smaller factor of $\sqrt{2} \approx 1.41$. Let us show that for $n = 2$, interval uncertainty always leads to a larger decrease of the average measurement error.

For $n = 2$, the result of interval fusion is, on average, more accurate, than the result of the probabilistic fusion: a proof. Let us start with the same accuracy values $\Delta_1 = \sigma_1$ and $\Delta_2 = \sigma_2$, and let us compare the results of applying the formulas (11) and (25). We want to prove that the result of applying the formula (25) is always smaller, i.e., that

$$\Delta_1 - \frac{1}{3} \cdot \frac{\Delta_1^2}{\Delta_2} < \sqrt{\frac{\Delta_1^2 \cdot \Delta_2^2}{\Delta_1^2 + \Delta_2^2}} = \frac{\Delta_1 \cdot \Delta_2}{\sqrt{\Delta_1^2 + \Delta_2^2}}. \tag{26}$$

If we divide both sides of this inequality by Δ_1 and express both sides in terms of the ratio $r \stackrel{\text{def}}{=} \frac{\Delta_2}{\Delta_1} \ge 1$, then the inequality (26) gets the following equivalent form

$$1 - \frac{1}{3r} < \frac{r}{\sqrt{1+r^2}}. \tag{27}$$

Multiplying both sides by r and by $\sqrt{1+r^2}$, we get yet another equivalent inequality

$$\left(r - \frac{1}{3}\right) \cdot \sqrt{1+r^2} < r^2. \tag{28}$$

Squaring both sides, we get the following equivalent inequality

$$(r^2+1) \cdot \left(k^2 - \frac{2}{3}r + \frac{1}{9}\right) < r^4. \tag{29}$$

Opening parentheses, we get

$$r^4 - \frac{2}{3}r^3 + \frac{1}{9}r^2 + r^2 - \frac{2}{3}r + \frac{1}{9} < r^4, \tag{30}$$

i.e., by moving all the terms to the right-hand side and combining terms proportional to the same power of r, that, equivalently,

$$\frac{2}{3}r^3-\frac{10}{9}r^2+\frac{2}{3}r-\frac{1}{9}>0. \tag{31}$$

Let us prove that this inequality holds for all r. Indeed, this inequality clearly holds for $r=1$: then, the left-hand side is equal to

$$\frac{2}{3}-\frac{10}{9}+\frac{2}{3}-\frac{1}{9}=\frac{6-10+6-1}{9}=\frac{1}{9}>0. \tag{32}$$

Let us prove that the left-hand side of (31) is increasing and thus, it is positive for all $r>1$ as well. Indeed, the derivative of this left-hand side is equal to

$$2r^2-\frac{20}{9}r+\frac{2}{3}. \tag{33}$$

The discriminant of this quadratic equation is equal to

$$\left(\frac{20}{9}\right)^2-4\cdot 2\cdot\frac{2}{3}=\frac{400}{81}-\frac{16}{3}=\frac{400-16\cdot 27}{81}=\frac{400-432}{81}=-\frac{32}{81}<0. \tag{34}$$

So, the quadratic expression (33) for the derivative is always non-negative. Thus, the inequality (31) holds for all $r\ge 1$ – and since it is equivalent to the desired inequality (26), the desired inequality also holds always. The statement is proven.

Case when all the measurements are equally accurate $\Delta_1=\ldots=\Delta_n$. In this case, similar to the case $n=2$, we conclude that

$$F(z)=1-\frac{2\Delta_1-z}{2\Delta_1}\cdot\ldots\cdot\frac{2\Delta_n-z}{2\Delta_n}=1-\left(\frac{2\Delta_1-z}{2\Delta_1}\right)^n. \tag{35}$$

Here, z can take any value from 0 to $2\Delta_1$, so the ratio $y\stackrel{\text{def}}{=}\frac{z}{2\Delta_1}$ takes values from the interval $[0,1]$. In terms of y, we have

$$F(y)=1-(1-y)^n, \tag{36}$$

so

$$\rho(y)=\frac{dF(y)}{dy}=n\cdot(1-y)^{n-1}. \tag{37}$$

Thus, the average value $E[y]$ of y is equal to

$$E[y]=\int_0^1 n\cdot(1-y)^{n-1}\cdot y\,dy. \tag{38}$$

Introducing an auxiliary variable $u=1-y$ for which $y=1-u$ and for which u also changes between 0 and 1, we get

$$E[y]=n\cdot\int_0^1(1-u)\cdot u^{n-1}\,du=n\cdot\int_0^1(u^{n-1}-u^n)\,du=n\cdot\left[\frac{u^n}{n}-\frac{u^{n+1}}{n+1}\right]_0^1=$$

$$n \cdot \left(\frac{1}{n} - \frac{1}{n+1}\right) = n \cdot \frac{(n+1)-n}{n \cdot (n+1)} = n \cdot \frac{1}{n \cdot (n+1)} = \frac{1}{n+1}. \tag{39}$$

Thus, for the accuracy $\Delta = r \cdot (2\Delta_1)$ of the fusion result, we get

$$\Delta = \frac{2}{n+1} \cdot \Delta_1. \tag{40}$$

For the case of $n = 2$, this is exactly what we got based on our general formula.

General case. In the case, one can actually also get an explicit formula for each n: indeed, here, we have, for

$$\Delta_1 \le \ldots \le \Delta_n, \tag{41}$$

that for the values $z \in [0, 2\Delta_1]$, we get the following cumulative distribution function:

$$F(z) = 1 - \frac{2\Delta_1 - z}{2\Delta_1} \cdot \ldots \cdot \frac{2\Delta_n - z}{2\Delta_n}. \tag{41}$$

The right-hand side of this formula is a polynomial. Thus, by differentiation, we can get an explicit polynomial formula for the derivative $\rho(z) = \dfrac{dF(z)}{dz}$ and thus, an explicit polynomial formula for the resulting value

$$\Delta = \int_0^{2\Delta_1} \rho(z) \cdot z\, dz. \tag{42}$$

Case of fuzzy estimates. It is known that in fuzzy logic (see, e.g. [1, 4, 6, 8, 9, 10, 14]), the usual way of processing fuzzy estimates – by using Zadeh's extension principle – is equivalent to processing α-cut intervals for all $\alpha \in [0,1]$; see, e.g., [1, 4, 6, 8, 9, 10]. Thus, the above formulas can be applied to the fuzzy case as well.

Acknowledgments

This work was supported in part by the National Science Foundation via grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science) and HRD-1242122 (Cyber-ShARE Center of Excellence).

References

1. R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.

2. L. Jaulin, M. Kiefer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control, and Robotics*, Springer, London, 2001.
3. E. T. Jaynes and G. L. Bretthorst, *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge, UK, 2003.
4. G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
5. G. Mayer, *Interval Analysis and Automatic Result Verification*, de Gruyter, Berlin, 2017.
6. J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.
7. R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009.
8. H. T. Nguyen and V. Kreinovich, "Nested intervals and sets: concepts, relations to fuzzy sets, and applications", In: R. B. Kearfott and V. Kreinovich (eds.), *Applications of Interval Computations*, Kluwer, Dordrecht, 1996, pp. 245–290.
9. H. T. Nguyen, C. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2019.
10. V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.
11. S. G. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, Springer, New York, 2005.
12. D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.
13. G. W. Walster and V. Kreinovich, "For unknown–but–bounded errors, interval estimates are often better than averaging", *ACM SIGNUM Newsletter*, 1996, Vol. 31, No. 2, pp. 6–19.
14. L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.