How to Relate Fuzzy and OWA Estimates

Tanja Magoc and Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
El Paso, Texas 79968
Emails: tmagoc@utep.edu, vladik@utep.edu

***Abstract*—In many practical situations, we have several estimates $x_1, \ldots, x_n$ of the same quantity x, i.e., estimates for which $x_1 \approx x$, $x_2 \approx x$, ..., and $x_n \approx x$. It is desirable to combine (fuse) these estimates into a single estimate for x. From the fuzzy viewpoint, a natural way to combine these estimates is: (1) to describe, for each x and for each i, the degree $\mu_\approx(x_i - x)$ to which x is close to x_i, (2) to use a t-norm ("and"-operation) to combine these degrees into a degree to which x is consistent with all n estimates, and then (3) find the estimate x for which this degree is the largest. Alternatively, we can use computationally simpler OWA (Ordered Weighted Average) to combine the estimates x_i. To get better fusion, we must appropriately select the membership function $\mu_\approx(x)$, the t-norm (in the fuzzy case) and the weights (in the OWA case).**

Since both approaches – when applied properly – lead to reasonable data fusion, it is desirable to be able to relate the corresponding selections. For example, once we have found the appropriate $\mu_\approx(x)$ and t-norm, we should be able to deduce the appropriate weights – and vice versa. In this paper, we describe such a relation. It is worth mentioning that while from the application viewpoint, both fuzzy and OWA estimates are not statistical, our mathematical justification of the relation between them uses results that have been previously applied to mathematical statistics.

I. Formulation of the Problem

Single-quantity data fusion: a problem. In many practical situations, we have several estimates $x_1, \ldots, x_n$ of the same quantity x, i.e., estimates for which $x_1 \approx x$, $x_2 \approx x$, ..., and $x_n \approx x$. It is desirable to combine (fuse) these estimates into a single estimate for x.

There are several possible approaches to solving this problem.

Fuzzy approach to data fusion. From the fuzzy viewpoint (see, e.g., [3], [4]), a natural way to combine these estimates is as follows:

- to describe, for each x and for each i, the degree
$$\mu_\approx(x_i - x) \quad (1)$$
to which x is close to x_i,
- to use a t-norm ("and"-operation) $t_\&(a, b)$ to combine these degrees into a degree
$$d(x) = t_\&(\mu_\approx(x_1 - x), \ldots, \mu_\approx(x_n - x)) \quad (2)$$
to which x is consistent with all n estimates, and then
- find the estimate x for which the degree $d(x)$ is the largest.

Computational complexity: the main limitation of the fuzzy approach. While the above approach is natural, the corresponding procedure looks computationally complex, especially for generic membership functions $\mu_\approx(x)$ and generic t-norms $t_\&(a, b)$, for which the expression $d(x)$ can be very complex, and for which the optimization of $d(x)$ may be a computationally difficult task.

Several approaches have been proposed to simplify the resulting computations; one of the most well known is the OWA approach; see, e.g., [6], [7], [8].

OWA approach to single-variable data fusion. OWA (Ordered Weighted Average) is a computationally simpler alternative to fuzzy data fusion, i.e., to combining the estimates $x_1, \ldots, x_n$ into a single estimate x.

The main idea behind OWA is as follows. We sort the values $x_1, \ldots, x_n$ into an increasing sequence
$$x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}, \quad (3)$$
select the weights $w_1, \ldots, w_n \ge 0$ for which
$$\sum_{i=1}^n w_i = 1, \quad (4)$$
and use the weighted average
$$x = \sum_{i=1}^n w_i \cdot x_{(i)} \quad (5)$$
as the desired fused estimate.

Problem. To get a better fusion:

- we must appropriately select the membership function $\mu_\approx(x)$ and the t-norm (in the fuzzy case), and
- we must appropriately select the weights (in the OWA case).

Both approaches – when applied properly – lead to reasonable data fusion.

It is therefore desirable to be able to relate the corresponding selections:

- once we have found the appropriate $\mu_\approx(x)$ and t-norm, we should be able to deduce the appropriate weights, and
- once we have foud the appropriate weights, we should be able to deduce the appropriate membership function $\mu_\approx(x)$ and t-norm.

What we do in this paper. In this paper, we describe such a relation.

It is important to emphasize that:

- while from the application viewpoint, both fuzzy and OWA estimates are not statistical,
- our mathematical justification of the relation between them uses results that have been previously applied to mathematical statistics.

II. Main Idea

Reducing to the case of Archimedean t-norms. In fuzzy logic, several different types of t-norms are used. One of the mostly widely used classes of t-norms are *Archimedean* t-norms. It is known that these t-norms $t_\&(a,b)$ are isomorphic to the the simple algebraic product t-norm $a \cdot b$ – isomorphic in the following sense.

There exists a monotonic 1-1 function $f : [0,1] \to [0,1]$ for which

$$t_\&(a,b) = f^{-1}(f(a) \cdot f(b)), \tag{6}$$

where $f^{-1}(x)$ denoted the inverse function (for which $f^{-1}(x) = y$ if and only if $f(y) = x$).

Not all t-norms are Archimedean. For example, the minimum t-norm $\min(a,b)$ is not Achimedean. It is known that a general t-norm can be obtained:

- by setting Archimedean t-norms on several (maybe infinitely many) subintervals of the interval $[0,1]$, and
- by using $\min(a,b)$ as the value of $t_\&(a,b)$ for the cases when a and b do not belong to the same Archimedean subinterval.

From this general classification theorem for t-norms, we can conclude that for every t-norm and for every $\varepsilon > 0$, there exists an ε-close Archimedean t-norm. (Crudely speaking, we can replace the corresponding non-Archimedean min-parts with a close Archimedean t-norm, e.g., with $(a^{-p} + b^{-p})^{-1/p}$ for a sufficiently large p).

Thus, for an arbitrary accuracy $\varepsilon > 0$ and for an arbitrary t-norm, we can have an Archimedean t-norm which is, within this accuracy, indistinguishable from the original one. So, from the practical viewpoint, we can always safely assume that the t-norm is Archimedean, i.e., that is has the form (6).

For this t-norm, the value

$$d(x) = t_\&(\mu_\approx(x_1 - x), \ldots, \mu_\approx(x_n - x)) \tag{7}$$

can be represented as

$$d(x) = f^{-1}(f(\mu_\approx(x_1 - x)) \cdot \ldots \cdot f(\mu_\approx(x_n - x))). \tag{8}$$

Since the function $f(x)$ – and thus, its inverse function $f^{-1}(x)$ – are monotonic, the degree $d(x)$ attains its maximum if and only if the value $D(x) \stackrel{\text{def}}{=} f(d(x))$ attains its maximum. Due to (8), we have

$$D(x) = f(\mu_\approx(x_1 - x)) \cdot \ldots \cdot f(\mu_\approx(x_n - x)). \tag{9}$$

This expression can be described as

$$D(x) = \prod_{i=1}^n \rho(x_i - x), \tag{10}$$

where we denoted

$$\rho(x) \stackrel{\text{def}}{=} f(\mu_\approx(x)). \tag{11}$$

Resulting reformulation of the problem. We have two ways to fuse estimates $x_1, \ldots, x_n$ into a single estimate x:

- find x for which the value $\prod\limits_{i=1}^n \rho(x_i - x)$ is the largest possible (fuzzy approach), and
- find x as $\sum\limits_{i=1}^n w_i \cdot x_{(i)}$ (OWA approach).

The problem is:

- given the function $\rho(x)$, find the weights w_i for which the OWA estimate is close to the original fuzzy estimate; and
- given the weights w_i, find the function $\rho(x)$ for which the corresponding fuzzy estimate is close to the original OWA estimate.

How we plan to solve this problem. To solve the above problem, we will take into consideration that a similar mathematical problem is already solved in *robust statistics.* Before we explain our solution, let us therefore briefly describe what is robust statistics and how a similar problem is solved there.

III. A Similar Problem Is Already Solved In Robust Statistics

Robust statistics: reminder. In this section, we will recall that a similar mathematical problem is already solved in *robust statistics* – an area of statistics in which we need to make statistical estimates under partial information about the probability distribution.

In robust statistics (see, e.g., [2]), there are several different types of techniques for estimating a shift-type parameter a based on a sample $x_1, \ldots, x_n$.

M-methods: reminder. The most widely used methods are *M-methods*, methods which are similar to the maximum likelihood approach from the traditional (non-robust) statistics. In the maximum likelihood approach, if we know that the probability density function has the form $f_0(x_i - a)$ for some unknown value a, and that the values $x_1, \ldots, x_n$ are independent, then the likelihood to get the sample $x_1, \ldots, x_n$ is equal to the product

$$\prod_{i=1}^n f_0(x_i - a). \tag{12}$$

In the Maximum Likelihood approach, we select the value $a = a_M$ for which this likelihood is the largest possible:

$$\prod_{i=1}^n f_0(x_i - a) \to \max_a. \tag{13}$$

Comment. This formula is, in effect, identical to our fuzzy fusion formula, with $f_0(x) = \rho(x)$.

M-methods: robust case. In the Maximum Likelihood approach, we know the probability density function $f_0(x)$. In the robust approach, we apply a similar method with *some* function $f_0(x)$.

Each of these robust M-methods coincides with the Maximum Likelihood method for the corresponding probability density function.

L-estimates. Another important class of robust estimates are *L-estimates*, i.e., estimates of the type

$$a_L = \frac{1}{n} \cdot \sum_{i=1}^{n} m\left(\frac{i}{n}\right) \cdot x_{(i)}, \tag{14}$$

for some function $m(x)$ for which

$$\int_0^1 m(t)\, dt = 1. \tag{15}$$

Comment. This formula is, in effect, identical to our OWA fusion formula, with

$$w_i = \frac{1}{n} \cdot m\left(\frac{i}{n}\right). \tag{16}$$

A problem which is solved in robust statistics. The question solved in robust statistics is: what is the natural correspondence between M-estimates and L-estimates?

Correspondence between M- and L-estimates: case of traditional statistics. To explain the meaning of this correspondence, let us first consider the case when we know the exact shape $f_0(x)$ of the probability density function, and we know that the actual probability density function has the form $f_0(x-a)$ for some (unknown) parameter a. In this case,

- for each function $f_0(x)$, we can use the solution of the corresponding equation (13) as an M-method estimate $a_M(U)$ for the parameter a;
- for each function $m(p)$, we can use the estimate (14) as an L-method estimate $a_L(m)$ for the parameter a.

The quality of each estimate can be estimated as the mean square of the difference between the estimate and the actual value a, i.e.,

- for M-estimates, as

$$q_M(U) = E[(a_M(U) - a)^2]; \tag{17}$$

and

- for L-estimates, as

$$q_L(m) = E[(a_L(m) - a)^2]. \tag{18}$$

For a given probability density function $f_0(x)$:

- the function $f_0(x)$ is optimal, i.e., the value

$$q_M(U) = E[(a_M(U) - a)^2] \tag{19}$$

is the smallest possible, and

- we can also find the optimal function $m(p)$, i.e., the function $m(p)$ for which the value

$$q_L(m) = E[(a_L(p) - a)^2] \tag{20}$$

is the smallest possible.

Under certain reasonable conditions, the optimal L-estimate can be found as follows (see, e.g., [1], [2]):

- first, we compute the cumulative distribution function $F_0(x)$ as

$$F_0(x) = \int_{-\infty}^{x} f_0(t)\, dt; \tag{21}$$

- then, we find the auxiliary function $M(p)$ as

$$M(F_0(x)) = -(\ln(f_0(x))'', \tag{22}$$

i.e., as

$$M(p) = z(F_0^{-1}(p)), \tag{23}$$

where $z(x) \stackrel{\text{def}}{=} -(\ln(f_0(x))''$, and $F_0^{-1}(p)$ denotes an inverse function;

- after that, we normalize the auxiliary function $M(p)$ to get

$$m(p) = \frac{M(p)}{\int_0^1 M(q)\, dq}. \tag{24}$$

Similarly, if we know the function $m(p)$, then, to find the corresponding function $f_0(x)$, we find a probability density function $f_0(x)$ for which $m(p)$ leads to the optimal L-estimate.

Comment. It turns out that, under reasonable conditions, for the resulting functions $f_0(x)$ and $m(p)$, the quality values

$$q_M(U) = E[(a_M(U) - a)^2] \tag{25}$$

and

$$q_L(m) = E[(a_L(p) - a)^2] \tag{26}$$

are asymptotically equal when the sample size n tends to infinity:

$$\frac{q_M(U)}{q_L(m)} = \frac{E[(a_M(U) - a)^2]}{E[(a_L(p) - a)^2]} \to 1 \text{ as } n \to +\infty. \tag{27}$$

Correspondence between M- and L-estimates: robust case. In the robust case, when we do not know the exact shape of a probability density function, we only know the *class* F_0 of possible shapes, and we know that the actual probability density function has the form $f_0(x-a)$, where $f_0(x)$ is one of the shapes from the class F_0, and a is an (unknown) parameter. In this case too, we can consider

- M-estimates $a_M(U)$ (described by the formula (13)), and
- L-estimates $a_L(m)$ (described by the formula (14)).

In the robust case, since the distribution is not known exactly, for different distributions $f_0(x)$ from the class F_0, we get different accuracies

$$E_{f_0}[(a_M(U) - a)^2] \tag{28}$$

and

$$E_{f_0}[(a_L(m) - a)^2]. \quad (29)$$

As a natural measure of quality of a given estimate, we can take the *worst-case* accuracy

$$q_M(U) = \sup_{f_0 \in F} E_{f_0}[(a_M(U) - a)^2]; \quad (30)$$

$$q_L(m) = \sup_{f_0 \in F} E_{f_0}[(a_L(m) - a)^2]. \quad (31)$$

As shown in [2], for many reasonable classes F_0 of distributions,

- we can find the optimal (*minimax*) function $f_0(x)$, i.e., the function $f_0(x)$ for which the value $q_M(U)$ is the smallest possible, and
- we can find the optimal (*minimax*) function $m(p)$, i.e., the function $m(p)$ for which the value $q_L(m)$ is the smallest possible.

These optimal M-estimates and L-estimates can be obtained as follows [1], [2]:

- first, in the class F_0, we find the probability distribution $f_0(x)$ for which the Fisher information

$$I(f_0) = \int \left(\frac{f_0'(x)}{f_0(x)}\right)^2 \cdot f_0(x)\,dx$$

  is the smallest possible;
- then, we find M-estimate and L-estimate which are optimal for this distribution $f_0(x)$.

The correspondence between the functions $f_0(x)$ and $m(p)$ can thus be described as follows.

Let us first assume that we know the function $f_0(x)$, then, to find the corresponding function $m(p)$, we do the following:

- first, we find a class F_0 of probability density functions for which $f_0(x)$ leads to the optimal M-estimate;
- then, we use this class F_0 to find the function $m(p)$ which leads to the optimal L-estimate for this class F_0.

Similarly, if we know the function $m(p)$, then, to find the corresponding function $f_0(x)$, we do the following:

- first, we find a class F_0 of probability density functions for which $m(p)$ leads to the optimal L-estimate;
- then, we use this class F_0 to find the function $f_0(x)$ which leads to the optimal M-estimate for this class F_0.

It turns out that for the resulting functions $f_0(x)$ and $m(p)$, the quality values $q_M(U)$ and $q_L(m)$ are also asymptotically equal when the sample size n tends to infinity:

$$\frac{q_M(U)}{q_L(m)} \to 1 \text{ as } n \to +\infty. \quad (32)$$

Correspondence between M- and L-estimates: explicit description. We have mentioned that the robust M- and L-estimates coincide with M- and L-estimates for an appropriate probability density function $f_0(x)$. Thus, the robust-case correspondence between M- and L-estimates can be described by exactly the same formulas as for the traditional statistical case.

Examples. Several examples are given in [1] and [2]. For example, the Gaussian function

$$f_0(x) = \exp\left(-\frac{1}{2} \cdot x^2\right) \quad (33)$$

is proportional to the probability density of the normal distribution. Hence,

$$F_0(x) = \int_{-\infty}^{x} f_0(t)\,dt \quad (34)$$

is proportional to the cumulative distribution function of a normal distribution. Here,

$$\ln(f_0(x)) = -\frac{1}{2} \cdot x^2, \quad (35)$$

hence

$$\ln(f_0(x))'' = 1, \quad (36)$$

and

$$z(x) = -\ln(f_0(x))'' = 1. \quad (37)$$

So,

$$M(p) = z(F_0^{-1}(p)) = 1. \quad (38)$$

The integral of $M(p) = 1$ over the interval $[0, 1]$ is 1, so

$$m(p) = M(p) = 1. \quad (39)$$

IV. Relation Between Fuzzy and OWA Estimates: Our Main Idea

Let us apply the solution from robust statistics to the case of fuzzy and OWA estimates. We have seen that, mathematically,

- M-estimates correspond to fuzzy estimates, and
- L-estimates correspond to OWA estimates.

We can therefore use the solution provided by robust statistics to find the desired correspondence between the utility function and the spectral risk measures.

Resulting solution. Specifically, once we know the membership function $\mu_{\approx}(x)$ and the function $f(x)$ describing the t-norm, i.e., for which

$$t_{\&}(a, b) = f^{-1}(f(a) \cdot f(b)), \quad (40)$$

then we can find the corresponding OWA weights as follows:

- first, we compute an auxiliary function

$$f_0(x) = f(\mu_{\approx}(x)); \quad (41)$$

- then, we compute the second auxiliary function

$$F_0(x) = \int_{-\infty}^{x} f_0(t)\,dt; \quad (42)$$

- after that, we find the third auxiliary function $M(p)$ from the formula

$$M(F_0(x)) = -(\ln(f_0(x))'', \quad (43)$$

  i.e., as

$$M(p) = z(F_0^{-1}(p)), \quad (44)$$

where $z(x) = -(\ln(f_0(x))''$, and $F_0^{-1}(p)$ denotes an inverse function;

- finally, we compute

$$I \stackrel{\text{def}}{=} \int_0^1 M(q)\,dq, \tag{45}$$

then $m(p) = \dfrac{M(p)}{I}$, and select the desired weights

$$w_i = \frac{1}{n} \cdot m\left(\frac{i}{n}\right). \tag{46}$$

Comment. The above procedure describes how, knowing the membership function and the t-norm, we can find the corresponding weights w_i. What if we know the weights w_i and we want to find the membership function and the t-norm?

First, by extrapolation, we find a function $m(p)$ for which

$$m\left(\frac{i}{n}\right) = n \cdot w_i. \tag{47}$$

To find $f(x)$, we can use the above formula $M(F_0(x)) = -(\ln(f_0(x)))''$, where $f_0(x) = F_0''(x)$, and $M(p) = I \cdot m(p)$ for $I = \int_0^1 M(q)\,dq$.

Thus, given w_i, we can find the function $f_0(x)$ as follows:

- first, we find the auxiliary function $F_0(x)$ and the auxiliary value I by solving the equation

$$I \cdot m(F_0(x)) = -(\ln(F_0'(x)))''; \tag{48}$$

- then, we find $f_0(x) = F_0'(x)$.

Once we know the auxiliary function $f_0(x)$, we can take different t-norms, and for each of these t-norms, we can find an appropriate membership function $\mu_\approx(x)$.

For an algebraic product t-norm, we have $f(x) = x$, thus $f_0(x) = f(\mu_\approx(x)) = \mu_\approx(x)$, so the desired membership function is $\mu_\approx(x) = f_0(x)$.

For a general Archimedean t-norm $t_\&(a,b)$, we first find the function $f(x)$ for which $t_\&(a,b) = f^{-1}(f(a) \cdot f(b))$. Then, from the equality $f_0(x) = f(\mu_\approx(x))$, we conclude that

$$\mu_\approx(x) = f^{-1}(f_0(x)). \tag{49}$$

Example. for the Gaussian membership function and algebraic t-norm, the condition

$$\prod_{i=1}^n \rho(x_i - x) \to \max_x \tag{50}$$

leads to

$$\prod_{i=1}^n \exp\left(-\frac{1}{2} \cdot (x_i - x)^2\right) \to \max_x \tag{51}$$

i.e., equivalently, to

$$\exp\left(-\sum_{i=1}^n \frac{1}{2} \cdot (x_i - x)^2\right) \to \max_x, \tag{52}$$

and to

$$\sum_{i=1}^n \frac{1}{2} \cdot (x_i - x)^2 \to \min_x. \tag{53}$$

Differentiating the minimized expression by x and equating the derivative to 0, we conclude that

$$x = \frac{1}{n} \cdot \sum_{i=1}^n x_i, \tag{54}$$

i.e., that the corresponding fused estimate is simply the arithmetic average of the original estimates $x_1, \ldots, x_n$.

As we have mentioned, for the Gaussian $f_0(x)$, the corresponding function $m(p)$ is equal to 1. Thus, the corresponding weights are

$$w_i = \frac{1}{n} \cdot m\left(\frac{i}{n}\right) = \frac{1}{n}. \tag{55}$$

So, the corresponding OWA estimate is equal to

$$x = \frac{1}{n} \cdot \sum_{i=1}^n x_{(i)}. \tag{56}$$

The sum does not change if we simply re-order the values x_i. Thus, in this case, the fuzzy and OWA estimates are not only asymptotically equivalent – the corresponding estimates (54) and (56) are actually identical.

Comment. It is worth mentioning that the same mathematical result from robust statistics also has economic applications; see, e.g., [5].

ACKNOWLEDGMENT

This work was also supported in part by the National Science Foundation grants HRD-0734825 and DUE-0926721, by Grant 1 T36 GM078000-01 from the National Institutes of Health, and by Grant 5015 from the Science and Technology Centre in Ukraine (STCU), funded by European Union.

The authors are thankful to Hung T. Nguyen for valuable discussions.

REFERENCES

[1] J. Collins and D. Wiens, "Mimimax properties of M-, R-, and L-estimators of location in Lévy neighborhoods", *The Annals of Statistics*, 1989, Vol. 17, No. 1, pp. 327–336.

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

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

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

[5] S. Sriboonchitta, H. T. Nguyen, and V. Kreinovich, "How to Relate Spectral Risk Measures and Utilities", *Proceedings of the Third Conference of Thailand Econometric Society*, Chiang Mai, Thailand, January 7–8, 2010.

[6] V. Torra and Y. Narukawa, *Modeling Decisions: Information Fusion and Aggregation Operators*, Springer, Berlin, 2007.

[7] R. R. Yager, "On ordered weighted averaging aggregation operators in multi-criteria decision making," *IEEE Transactions on Systems, Man and Cybernetics*, 1988, Vol. 18, pp. 183–190.

[8] R. R. Yager and J. Kacprzyk (eds), *The Ordered Weighted Averaging Operators: Theory and Applications*, Kluwer, Norwell, Massachusetts, 1997,