

# Uncertain Information Fusion and Knowledge Integration: How to Take Reliability into Account

Hung T. Nguyen  
 Department of Mathematical Sciences  
 New Mexico State University  
 Las Cruces, NM 88003, USA, and  
 Faculty of Economics, Chiang Mai University  
 Chiang Mai, Thailand  
 hunguyen@nmsu.edu

Kittawit Autcharyapanitkul  
 Faculty of Economics  
 Maejo University  
 Chiang Mai, Thailand  
 kittar3@hotmail.com

Olga Kosheleva and  
 Vladik Kreinovich  
 University of Texas at El Paso  
 500 W. University  
 El Paso, Texas 79968, USA  
 olgak@utep.edu, vladik@utep.edu

**Abstract**—In many practical situations, we need to fuse and integrate information and knowledge from different sources – and do it under uncertainty. The existing methods for information fusion and knowledge integration take into account uncertainty, but, in addition to uncertainty, we also face the problem of reliability: sensors may malfunction, experts can be wrong, etc. In this paper, we show how to take reliability into account in uncertain information fusion and knowledge integration. We show this on the examples of probabilistic and fuzzy uncertainty.

## I. TRADITIONAL APPROACH TO UNCERTAIN INFORMATION FUSION AND KNOWLEDGE INTEGRATION: A BRIEF REMINDER AND THE NEED TO TAKE RELIABILITY INTO ACCOUNT

**Need for uncertain information fusion.** In many practical situations, we are interested in the values of the quantities  $x_1, \dots, x_i, \dots, x_n$ , and we know several results  $\tilde{y}_1, \dots, \tilde{y}_j, \dots, \tilde{y}_N$  of measuring or estimating either these same quantities or, more generally, several different quantities relating to  $x_i$  in a known way. In precise terms, we know that

$$\tilde{y}_j \approx y_j \stackrel{\text{def}}{=} f(x_1, \dots, x_n, a_{j1}, \dots, a_{jk}, \dots, a_{js}, c_1, \dots, c_l, \dots, c_m),$$

where:

- $f$  are a known function,
- $a_{jk}$  are the known values describing the setting of the  $j$ -th measurement or estimate, and
- $c_1, \dots, c_m$  are the unknown parameters of the corresponding dependence.

To describe this formula, in the following text, we will also use a simplifies way of describing it:  $y_j = f(x, a_j, c)$ , where  $x \stackrel{\text{def}}{=} (x_1, \dots, x_n)$ ,  $a_j \stackrel{\text{def}}{=} (a_{j1}, \dots, a_{js})$ , and  $c \stackrel{\text{def}}{=} (c_1, \dots, c_m)$ .

(The difference between the estimates  $\tilde{y}_j$  and the values  $y_j$  predicted by the model is caused by the fact that measurements are never absolutely accurate (see, e.g., [7]), and expert estimates are even less accurate than measurements.)

In such situations, it is desirable to “fuse” (combine) these estimates into estimates for the desired quantities  $x_i$ .

**Examples.** To illustrate this problem, let us give two examples. The first example illustrates the simplest case, when we are interested in a single quantity  $x_1$ , and all the estimates  $\tilde{y}_1, \dots, \tilde{y}_N$  directly estimate this quantity. In this case,  $\tilde{y}_j = x_1$  for all  $i$ , i.e.,  $s = m = 0$  and  $f(x_1) = x_1$ .

A more complex example is when we are interested in the amplitude  $x_1$  of a sinusoidal periodic process

$$x(t) = x_1 \cdot \sin(\omega \cdot t + \theta_0)$$

with the unknown frequency  $\omega$  and phase  $\theta_0$ . To find this amplitude, we measure the value of the signal at different moments of time. In this case,

$$\tilde{y}_j \approx y_i = f(x_1, a_{j1}, c_1, c_2),$$

where:

- $a_{j1}$  is the moment of time at which we make the  $j$ -th measurement,
- $c_1 = \omega$  and  $c_2 = \theta_0$  are the parameters of the sinusoidal model, and
- the function  $f$  has the form

$$f(x_1, a_{j1}, c_1, c_2) = x_1 \cdot \sin(c_1 \cdot a_{j1} + c_2).$$

**Two types of uncertainty.** As we have mentioned earlier, the estimates  $\tilde{y}_j$  are, in general, different from the values

$$y_j = f(x_1, \dots, x_n, a_{j1}, \dots, a_{js}, c_1, \dots, c_m)$$

predicted by the corresponding model. How can we describe the corresponding inaccuracy  $\Delta y_j \stackrel{\text{def}}{=} \tilde{y}_j - y_j$ ?

In some cases, we know the frequency of different values of estimation inaccuracy, i.e., in precise terms, we know the probability distribution of this inaccuracy. In other cases, all we know is the expert estimations for the size of this inaccuracy, expert estimations expressed by using imprecise (“fuzzy”) words from natural language. In such cases, a reasonable idea is to use *fuzzy logic*, techniques specifically designed for handling this uncertainty [2], [4], [10].

**What we do in this paper.** Let us consider how the existing methods take care of this uncertainty. Then, we will explain

the limitation of the current approaches: that we need to take reliability into account. After that, we will discuss how to actually take reliability into account.

**Taking uncertainty into account: from specific cases of probabilistic uncertainty to a general case.** In some cases, we know the probability distribution for the estimation error  $\Delta y_j = \tilde{y}_j - y_j$ . Each of these probability distributions can be described, e.g., by a probability density function (pdf)  $\rho_j(\Delta y_j)$ .

The estimate  $\tilde{y}_j$  is, in reality, never the exact number: it is usually plus minus the corresponding discretization level. For example, if a measuring instrument returns the result 0.376, this means any value from 0.3755 to 0.3765. Similarly, if an expert estimates the value is 1.1, this means any value from 1.05 to 1.15. In general, the estimate  $\tilde{y}_j$  means an interval  $[\tilde{y}_j - \delta_j, \tilde{y}_j + \delta_j]$ , for some small value  $\delta_j$ . The corresponding interval for the difference  $\Delta y_j = \tilde{y}_j - y_j$  has the form

$$[(\tilde{y}_j - y_j) - \delta_j, (\tilde{y}_j - y_j) + \delta_j]$$

and thus, has the same width. Thus, we can estimate the probability  $P_j$  of this estimate by multiplying the probability density  $\rho_j(\Delta y_j)$  by the width  $2\delta_j$  of the corresponding interval:  $P_j = \rho_j(\Delta y_j) \cdot (2\delta_j)$ .

Usually, all these distributions belong to the same family, they only differ by the values of the corresponding parameters. In precise terms, we have

$$\rho_j(\Delta y_j) = \rho(\Delta y_j, \theta_{j1}, \dots, \theta_{jq})$$

for an appropriate function  $\rho$  and for known values of the parameters  $\theta_{j1}, \dots, \theta_{jq}$ . For example, we may know that all the distributions are normal with 0 mean, and we know the standard deviations  $\theta_{j1} = \sigma_j$  corresponding to different estimates. In this case, we have

$$\rho(\Delta y, \theta_{j1}) = \frac{1}{\sqrt{2\pi} \cdot \theta_{j1}} \cdot \exp\left(-\frac{(\Delta y)^2}{2\theta_{j1}^2}\right).$$

In more general situations, some of the parameters  $\beta_i, \dots$  of the corresponding probability distributions are unknown. For example, we may know that the measurements come from several measuring instruments, we know that for each of these instruments, the distribution is Gaussian with 0 mean, but we do not know the standard deviations of these measuring instruments. Alternatively, we may know that the estimates from several experts, we know that for each expert, the estimation error is normally distributed with 0 mean and unknown standard deviation, but we do not know the standard deviations corresponding to different experts.

**Taking uncertainty into account: general case of probabilistic uncertainty.** In general, the set  $\{1, \dots, N\}$  of all estimations is divided into several disjoint subsets  $S_\alpha$ . The probability distribution of estimation errors  $\Delta y_j$  corresponding to each subset  $S_\alpha$  are characterized, in general, by its own expression

$$\rho_\alpha(\Delta y_j, \theta_{j1}, \dots, \theta_{jq_\alpha}, \beta_{\alpha 1}, \dots, \beta_{\alpha t_\alpha}),$$

where the values  $\theta_{\alpha 1}, \dots$  are known while the values  $\beta_{\alpha 1}, \dots$  are not known.

**Example.** If different sets  $S_\alpha$  correspond to different measuring instruments, with 0 mean and unknown standard deviations  $\beta_{\alpha 1} = \sigma_\alpha$ , then

$$\rho_\alpha(\Delta y) = \frac{1}{\sqrt{2\pi} \cdot \beta_{\alpha 1}} \cdot \exp\left(-\frac{(\Delta y)^2}{2\beta_{\alpha 1}^2}\right).$$

**How the uncertainty is taken into account now: case of probabilistic uncertainty.** As we have mentioned earlier, for each estimate  $j$ , the probability  $P_j$  of serving this estimate is proportional to the corresponding probability density.

Approximation errors corresponding to different measurement result are usually independent from each other. Thus, the overall probability of having all  $N$  estimates  $\tilde{y}_1, \dots, \tilde{y}_N$  is equal to the product of  $N$  probabilities  $P_1 \dots P_N$  and is, thus, proportional to the product  $L$  of the corresponding probability densities. This product is known as *likelihood*.

If we group together estimates corresponding to each group  $S_\alpha$ , we get the following expression for the likelihood:

$$L = \prod_{\alpha} \prod_{j \in S_\alpha} \rho_\alpha(\Delta y_j, \theta_{j1}, \dots, \theta_{jq_\alpha}, \beta_{\alpha 1}, \dots, \beta_{\alpha t_\alpha}),$$

where

$$\Delta y_j = \tilde{y}_j - f(x_1, \dots, x_n, a_{j1}, \dots, a_{js}, c_1, \dots, c_m).$$

We need to find the desired values  $x_1, \dots, x_n$ , as well as all the remaining unknowns  $c_1, \dots, c_m, \beta_{\alpha 1}, \dots$ . A reasonable idea is to find the values for which the above probability is the largest, i.e., equivalent, the likelihood  $L$  takes the largest possible value. This idea is known as the *Maximum Likelihood Method*.

**Specific case of Gaussian (normal) distributions.** There are usually many different reasons for an estimation error. For example, for measurements, there is noise in each part of the measuring instrument – and all these noises contribute to the overall estimation error.

In situations when the overall estimation error is a sum of many different independent components, it is usually possible to invoke the Central Limit Theorem, according to which for large  $N$ , the distribution of the sum of  $N$  small independent random variables is close to Gaussian; see, e.g., [8]. And indeed, in many practical cases, the probability distribution of the measurement error is close to Gaussian [5], [6].

For the measurement error, it is usually safe to assume that the mean error (*bias*) is 0, since this bias can be detected if we several times compare the results of measuring instrument with a more accurate (“standard”) one, and thus, can be eliminated by simply re-scaling the measuring instrument.

It is therefore same to assume that each estimation error is normally distributed, with 0 mean and some standard deviation  $\sigma_j$ . The corresponding probability density function has the form

$$\rho_j(\Delta y_j) = \frac{1}{\sqrt{2\pi} \cdot \sigma_j} \cdot \exp\left(-\frac{(\Delta y_j)^2}{2\sigma_j^2}\right).$$

Thus, the likelihood takes the form

$$L = \prod_{j=1}^N \rho_j(\Delta y_j) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi} \cdot \sigma_j} \cdot \exp\left(-\frac{(\Delta y_j)^2}{2\sigma_j^2}\right) = \frac{1}{(\sqrt{2\pi})^N \cdot \prod_{j=1}^N \sigma_j} \cdot \exp\left(-\frac{1}{2} \cdot \sum_{j=1}^N \frac{(\Delta y_j)^2}{\sigma_j^2}\right).$$

When all the standard deviations  $\sigma_j$  are known, maximizing the above expression for the likelihood is equivalent to minimizing the sum in the exp part of this expression:

$$\sum_{j=1}^N \frac{(\Delta y_j)^2}{\sigma_j^2} = \sum_{j=1}^N \frac{(\tilde{y}_j - f(x, a_j, c))^2}{\sigma_j^2} \rightarrow \min_{x,c}.$$

This is the usual Least Squares approach.

In particular, when all the estimates have the same accuracy – e.g., come from using similar measuring instruments or the same expert – then  $\sigma_j = \sigma$  for all  $j$ , and the above optimization problem can be further simplified, into:

$$\sum_{j=1}^N (\Delta y_j)^2 = \sum_{j=1}^N (\tilde{y}_j - f(x, a_j, c))^2 \rightarrow \min_{x,c}.$$

For example, when, in addition, we only have one quantity of interest  $x_1$  and all estimates  $\tilde{y}_j$  directly estimate this quantity, then the formula takes the form

$$\sum_{j=1}^n (\tilde{y}_j - x_1)^2 \rightarrow \min_{x_1}.$$

Differentiating with respect to  $x_1$  and equating the derivative to 0, we can conclude that the fused estimate becomes the arithmetic mean

$$x_1 = \frac{1}{N} \cdot \sum_{j=1}^N \tilde{y}_j.$$

In cases when we do not know the approximation errors  $\sigma_\alpha$ , maximizing the likelihood  $L$  (or, equivalently, minimizing log-likelihood  $-\ln(L)$ ) over  $\sigma_\alpha$  leads to

$$\sigma_\alpha^2 = \frac{1}{N_\alpha} \cdot \sum_{j \in S_\alpha} (\tilde{y}_j - y_j)^2,$$

where  $N_\alpha$  is the overall number of estimates  $j$  from the  $\alpha$ -th group  $S_\alpha$ .

Substituting these values into the expression for log-likelihood, we conclude that minimizing log-likelihood is equivalent to minimizing the sum of the logarithms of these standard deviations, i.e., minimizing the sum

$$\sum_{\alpha} \ln \left( \sum_{j \in N_\alpha} (\tilde{y}_j - f(x, a_j, c))^2 \right) \rightarrow \min_{x,c}.$$

#### Taking uncertainty into account: case of fuzzy uncertainty.

In the fuzzy cases, instead of probabilities, for each estimate  $j$  and for each possible value of the estimation error  $\Delta y_j$ , we

know the degree  $\mu_j(\Delta y_j)$  to which this value of the estimation error is possible. The corresponding function is known as the *membership function*.

Usually, all these membership functions belong to the same family, they only differ by the values of the corresponding parameters. In precise terms, we have

$$\mu_j(\Delta y_j) = \mu(\Delta y_j, \theta_{j1}, \dots, \theta_{jq})$$

for an appropriate function  $\mu$  and for known values of the parameters  $\theta_{j1}, \dots, \theta_{jq}$ . For example, we may know that all the membership functions are triangular.

In more general situations, some of the parameters  $\beta_i, \dots$  of the corresponding membership functions are unknown. For example, we may know that the measurements come from several experts, we know that the membership functions for each of the experts is triangular with 0 maximum, but we do not know the spread of these membership functions.

In general, the set  $\{1, \dots, N\}$  of all estimations is divided into several disjoint subsets  $S_\alpha$ . The membership functions characterizing the estimation errors  $\Delta y_j$  from each subset  $S_\alpha$  are described, in general, by their own expression

$$d_j = \mu_\alpha(\Delta y_j, \theta_{j1}, \dots, \theta_{jq_\alpha}, \beta_{\alpha 1}, \dots, \beta_{\alpha t_\alpha}),$$

where the values  $\theta_{\alpha 1}, \dots$  are known while the values  $\beta_{\alpha 1}, \dots$  are not known.

**How the uncertainty is taken into account now: case of fuzzy uncertainty.** We are interested in the degree to which  $\Delta y_1$  is a possible value of the first estimation error and  $\Delta y_2$  is a possible value of the second estimation error, etc. In line with the general fuzzy techniques, to find this degree  $D$ , we apply an appropriate “and”-operation  $f_\&(a, b)$  to the degrees corresponding to different values  $j$ :

$$D = f_\&(D_1, D_2, \dots, D_\alpha, \dots),$$

where

$$D_\alpha = f_\&\{d_j : j \in S_\alpha\}.$$

We need to find the desired values  $x_1, \dots, x_n$ , as well as all the remaining unknowns  $c_1, \dots, c_m, \beta_{\alpha 1}, \dots$ . A reasonable idea is to find the values for which the above possibility degree  $D$  is the largest.

**From the computational viewpoint, the general fuzzy optimization problem can be reduced to an appropriate Maximum Likelihood problem.** When the “and”-operation is the algebraic product  $f_\&(a, b) = a \cdot b$ , then the above optimization takes the same form as for the probabilistic uncertainty, the only difference is that we have membership functions instead of the probability density functions.

In principle, however, we can have many different “and”-operations. From this viewpoint, the optimization problem corresponding to fuzzy information fusion is much more general – and thus, more complex than the Maximum Likelihood problem corresponding to probabilistic uncertainty. However, it is possible to reduce the general fuzzy case to the Maximum-Likelihood-type case of the product.

Indeed, it is known that every “and”-operation can be approximated, with any given accuracy, by an *Archimedean* “and”-operation, i.e., by an “and”-operation of the type  $f_{\&}(a, b) = g^{-1}(g(a) \cdot g(b))$  for some increasing functions  $g(x)$ . Thus, from the practical viewpoint, we can safely assume that the actual “and”-operation is Archimedean.

For an Archimedean “and”-operation, we have

$$D = g^{-1} \left( \prod_{\alpha} g(D_{\alpha}) \right).$$

Similarly, for every  $\alpha$ , we have

$$D_{\alpha} = g^{-1} \left( \prod_{j \in S_{\alpha}} g(d_j) \right)$$

and thus,

$$g(D_{\alpha}) = \prod_{j \in S_{\alpha}} g(d_j).$$

Substituting the formula for  $g(D_{\alpha})$  into the expression for  $D$ , we conclude that

$$D = g^{-1} \left( \prod_{\alpha} \prod_{j \in S_{\alpha}} g(d_j) \right).$$

Since the function  $g(x)$  is increasing, maximizing the degree  $D$  is equivalent to maximizing the expression  $g(D)$ , which has a somewhat simpler form:

$$g(D) = \prod_{\alpha} \prod_{j \in S_{\alpha}} g(d_j) =$$

$$\prod_{\alpha} \prod_{j \in S_{\alpha}} g(\mu_{\alpha}(\Delta y_j, \theta_{j_1}, \dots, \theta_{j_{q_{\alpha}}}, \beta_{\alpha_1}, \dots, \beta_{\alpha_{t_{\alpha}}}).$$

One can see that we arrive at the exact same expression as for the Maximim Likelihood, but with an auxiliary function  $g(\mu_{\alpha}(\dots))$  instead of the pdf  $\rho_{\alpha}(\dots)$ .

**Need to take reliability into account: general formulation of the problem.** In the above text, we implicitly assumed that all the measurement results and all expert estimates are absolutely reliable: they may contain some estimation error, but they do reflect the desired quantities  $x_1, \dots, x_n$ .

In reality, measurements and expert estimates are not always reliable: sometimes, the corresponding are related not to the desired quantities  $x_i$ , but to some other quantities. For example, in underwater sonar measurements, when we measure the distance to an object by the time that it takes for a signal to bounce back to us, the sensors sometimes record the signal reflected by some other object; see, e.g., [9] and references therein.

It is therefore desirable to take this reliability into account when we fuse information and integrate knowledge.

**What we do know about reliability?** Usually, situations when the estimate is not related to the desired quantities are rare. From past experience, we can estimate how rare they can be. Thus, we can assume that for every  $j$ , we know:

- in the probabilistic case, the probability  $p_j$  that the  $j$ -th estimate is indeed related to the desired quantities, and
- in the fuzzy case, the degree of confidence  $q_j$  to which the  $j$ -th estimate is related to the desired quantity.

How can use this information in uncertain information fusion and knowledge integration?

**What we do in this paper.** In this paper, we describe how to take reliability into account.

## II. ANALYSIS OF THE PROBLEM: PROBABILISTIC CASE

**How did we solve the original problem?** In the above text, we had the following unknowns:

- the desired quantities  $x_1, \dots, x_n$ ,
- the unknown parameters  $c_1, \dots, c_m$  in the formula describing the dependence of the measurement results  $\tilde{y}_j$  on the desired quantities, and
- the parameters  $\beta_{\alpha_1}, \dots$  that describe the probability distributions of different values  $\Delta y_j$  of the estimation error.

To find all these parameters from observations, we used the Maximum Likelihood method.

**Natural idea: use Maximum Likelihood method in case of reliability as well.** If we take reliability into account, then there are other things that we do not know: e.g., we do not know which estimates are related to the desired values  $x$  and which are not. In other words, we now have more unknowns than before.

A natural idea is to again use the Maximum Likelihood approach – this time, to find *all* the unknowns: both the previous unknowns and the new unknowns.

**So what are the new unknowns?** If we take reliability into account, then we have following additional unknowns:

- First, for every  $j$ , we do not know whether the  $j$ -th estimate  $\tilde{y}_j$  is related to the desired quantity or not. This can be described by introducing, for each estimate  $j$ , a new binary variable  $z_j$  which is:
  - equal to 1 if this estimate is related to the desired quantities, and
  - equal to 0 if the estimate  $\tilde{y}_j$  is not related. to the desired quantities  $x_1, \dots, x_n$ .

The quantities  $z_j$  are new unknowns.

- Second, for those  $j$  for which the estimate is not related to the desired quantities, we do not know what quantity  $y_j$  is being estimated. Such values  $y_j$  should also be added to the list of unknown.

Thus, we should use the Maximum Likelihood approach to estimate not only the values of the previous unknowns  $x$ ,  $c$ , and  $\beta$ , but also the values of the new unknowns:

- the values  $z_j \in \{0, 1\}$  corresponding to all estimates  $j = 1, \dots, N$ , and
- the values  $y_j$  corresponding to estimates for which

$$z_j = 0.$$

**Let us describe the corresponding probabilities.** In situations in which we take reliability into account, it is still reasonable to assume that situations corresponding to different estimates  $j$  are independent. Thus, the overall probability – that we will maximize – is still equal to the product  $P_1 \cdot \dots \cdot P_N$  of the probabilities  $P_j$  corresponding to different estimates.

The difference from the previous case is that the expressions for the probabilities  $P_j$  are now different. In the previous case, when we fixed the values of all the unknowns  $x$ ,  $c$ , and  $\beta$ , then we concluded that the probability  $P_j$  is proportional to the value of the pdf:

$$P_j \sim \rho_\alpha(\Delta y_j, \theta_j, \beta_\alpha),$$

where  $\Delta y_j = \tilde{y}_j - f(x, a_j, c)$ .

In the new (general) case, once we know the values of all the unknowns, i.e., once we know the values  $x$ ,  $c$ ,  $\beta$ ,  $z_j$ , and  $y_j$  for those  $j$  for which  $z_j = 0$ , what is the probability  $P_j$  to have the corresponding values  $\tilde{y}_j$  and  $z_j$ ?

A natural assumption is that the values  $z_j$  and  $\tilde{y}_j$  are independent. (Indeed, if they were dependent, we would be able, based on the estimates  $\tilde{y}_j$ , to tell whether this estimate depends on the desired quantities or not – so we would not face the situation in which we do not know it.) Thus, the probability of having the values  $z_j$  and  $\tilde{y}_j$  is equal to the product of the probability to have  $z_j$  and the probability to have  $\tilde{y}_j$ .

The probability  $p(z_j)$  to have  $z_j$  is easy to describe:

- the probability to have  $z_j = 1$  is equal to  $p_j$ , and
- the probability to have  $z_j = 0$  is equal to the remaining probability  $1 - p_j$ .

The probability to have a given estimate  $\tilde{y}_j$  is still proportional to  $\rho_\alpha(\Delta y_j, \theta_j, \beta_\alpha)$ , the only difference is that now, the expression for  $\Delta y_j$  is more complicated:

- when  $z_j = 1$ , then we still have  $\Delta y_j = \tilde{y}_j - f(x, a_j, c)$ ;
- when  $z_j = 0$ , then we have  $\Delta y_j = \tilde{y}_j - y_j$  for the given value  $y_j$ .

Summarizing: the overall probability is proportional to the product  $E_1 \cdot \dots \cdot E_N$  of the following expressions  $E_j$  corresponding to different estimates  $j$ :

- when  $z_j = 1$ , then

$$E_j = p_j \cdot \rho_\alpha(\tilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha);$$

- when  $z_j = 0$ , then

$$E_j = (1 - p_j) \cdot \rho_\alpha(\tilde{y}_j - y_j, \theta_j, \beta_\alpha).$$

**What can we conclude from the Maximum Likelihood approach?** We need to find the values of all the parameters  $x$ ,  $c$ ,  $\beta$ ,  $z_j$ , and  $y_j$  that maximize the product of the above expressions.

Let us start with finding the unknown values  $y_j$  corresponding to  $z_j = 0$ . For each  $j$ , only the value  $E_j$  depends on  $y_j$ . Thus, the product  $E_1 \cdot \dots \cdot E_N$  is the largest if this value  $E_j$  is the largest. In its turn, this value is the largest if it corresponds

to the largest value of the probability density  $\rho_\alpha(\Delta y_j, \dots)$ . Its largest values is thus equal to

$$E_j = (1 - p_j) \cdot \max_y \rho_\alpha(y, \theta_j, \beta_\alpha).$$

The probability of the estimation error is usually the largest when this error is 0 and decreases when  $|\Delta y_j|$  decreases. In such cases, the maximum is attained when  $y = 0$  and thus,

$$E_j = (1 - p_j) \cdot \rho_\alpha(0, \theta_j, \beta_\alpha).$$

Now that we have found the optimal values of  $y_j$ , let us find the optimal values of  $z_j$ . Similarly to the above case, for each  $j$ , only the value  $E_j$  depends on  $z_j$ . Thus, the product  $E_1 \cdot \dots \cdot E_N$  is the largest if this value  $E_j$  is the largest. To find out which value  $z_j \in \{0, 1\}$  makes the expression  $E_j$  the largest let us compare the values of  $E_j$  corresponding to  $z_j = 0$  and to  $z_j = 1$ :

- when  $z_j = 0$ , we have

$$E_j = (1 - p_j) \cdot \max_y \rho_\alpha(y, \theta_j, \beta_\alpha);$$

- when  $z_j = 1$ , we have

$$E_j = p_j \cdot \rho_\alpha(\tilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha).$$

The largest of these two expressions is equal to

$$E_j = \max \left( (1 - p_j) \cdot \max_y \rho_\alpha(y, \theta_j, \beta_\alpha), p_j \cdot \rho_\alpha(\tilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha) \right).$$

To find the values of the desired parameters  $x$ ,  $c$ , and  $\beta$ , we therefore need to maximize the product of such maxima.

So, we arrive at the following conclusion.

### III. HOW TO TAKE RELIABILITY INTO ACCOUNT IN UNCERTAIN INFORMATION FUSION AND KNOWLEDGE INTEGRATION: PROBABILISTIC CASE

**General case.** In the general case, we know:

- the function  $f(x, a, c)$  describing the dependence of the estimated quantities on the desired quantities,
- the families  $\rho_\alpha(\Delta y, \theta, \beta_\alpha)$  that describe the probabilities of estimation errors  $\Delta y_j$  for estimates  $j$  from different groups  $S_\alpha$ , and
- for each  $j$ , we know the probability  $p_j$  that the  $j$ -th estimate is indeed related to the desired quantities  $x_1, \dots, x_n$ .

In this case, according to the Maximum Likelihood method, we should select values  $x$ ,  $c$ , and  $\beta$  that maximize the product

$$E_1 \cdot \dots \cdot E_N,$$

where, for each  $j \in S_\alpha$ , we have

$$E_j = \max \left( (1 - p_j) \cdot \max_y \rho_\alpha(y, \theta_j, \beta_\alpha), p_j \cdot \rho_\alpha(\tilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha) \right).$$

In particular, for probability distributions for which zero estimation error is the most probable, we have

$$E_j = \max((1 - p_j) \cdot \rho_\alpha(0, \theta_j, \beta_\alpha), p_j \cdot \rho_\alpha(\tilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha)).$$

**How to actually find the corresponding values?** Let us assume that we already know how to solve the optimization problem corresponding to the case when all the estimates are absolutely reliable. How can we transform this algorithm into an algorithm for solving the new problem?

A natural idea is to use component-wise maximization, when we first maximize over one group of variables, then over another group, etc., until the process converges; see, e.g., [1]:

- 1) first, we pick  $z_j = 1$  for all  $j$  and use the usual Maximum Likelihood techniques to optimize over  $x$ ,  $c$ , and  $\beta$ ;
- 2) once we find the corresponding values of  $x$ ,  $c$ , and  $\beta$ , we optimize over  $z_j$ : namely, we select  $z_j = 1$  if

$$p_j \cdot \rho_\alpha(\tilde{y}_j - f(x, a_j, c), \theta_j, \beta_\alpha) \geq (1 - p_j) \cdot \max_y \rho_\alpha(y, \theta_j, \beta_\alpha);$$

for all other  $j$ , we select  $z_j = 0$ ;

- 3) then, only taking into account the estimates  $j$  selected on the previous step, we again use the maximum Likelihood method to find new estimates for  $x$ ,  $c$ , and  $\beta$ , and go back to Step 2.

This process continues until the process converges, i.e., until the values of the desired variables  $x_1, \dots, x_n$  obtained on the next iteration are sufficiently close to the values from the previous iteration.

**Case of normal distributions.** Let us consider the consequences of this approach for the typical case when all the estimation error are normally distributed with 0 mean and known standard deviations  $\sigma_j$ .

In this case, substituting the explicit formulas for the normal pdf into the above expressions, we conclude that the second term in the expression for  $E_j$  is larger when

$$1 - p_j \leq p_j \cdot \exp\left(-\frac{(\Delta y_j)^2}{2\sigma_j^2}\right),$$

i.e., equivalently, when

$$\frac{1 - p_j}{p_j} \leq \exp\left(-\frac{(\Delta y_j)^2}{2\sigma_j^2}\right)$$

and, taking negative logarithm of both sides, when

$$\frac{(\Delta y_j)^2}{2\sigma_j^2} \leq \ln\left(\frac{p_j}{1 - p_j}\right),$$

i.e., when

$$|\Delta y_j| \leq \sigma_j \cdot \sqrt{2 \ln\left(\frac{p_j}{1 - p_j}\right)}.$$

Thus, for the case of normal distributions, the above algorithm takes the following simplified form:

- 1) first, we pick  $z_j = 1$  for all  $j$  and use the usual Least Squares method to find the values  $x$  and  $c$  for which the sum

$$\sum_{j=1}^N \frac{(\tilde{y}_j - f(x, a_j, c))^2}{\sigma_j^2}$$

is the smallest possible;

- 2) once we find the corresponding values of  $x$  and  $c$ , we select  $z_j = 1$  if

$$|\tilde{y}_j - f(x, a_j, c)| \leq \sigma_j \cdot \sqrt{2 \ln\left(\frac{p_j}{1 - p_j}\right)};$$

for all other  $j$ , we select  $z_j = 0$ ;

- 3) then, only taking into account the estimates  $j$  selected on the previous step, we again use the Least Squares Method to find new estimates for  $x$  and  $c$  by minimizing the sum

$$\sum_{j:z_j=1} \frac{(\tilde{y}_j - f(x, a_j, c))^2}{\sigma_j^2},$$

and go back to Step 2.

This process continues until the process converges, i.e., until the values of the desired variables  $x_1, \dots, x_n$  obtained on the next iteration are sufficiently close to the values from the previous iteration.

#### IV. HOW TO TAKE RELIABILITY INTO ACCOUNT: FUZZY CASE

**Just like in the probabilistic case, we have a problem with additional unknowns.** The original fuzzy problem has the following unknowns:

- the desired quantities  $x_1, \dots, x_n$ ,
- the unknown parameters  $c_1, \dots, c_m$  in the formula describing the dependence of the measurement results  $\tilde{y}_j$  on the desired quantities, and
- the parameters  $\beta_{\alpha 1}, \dots$  that characterize the membership functions describing the estimation errors  $\Delta y_j$ .

Now, we have to also find the two new types of unknowns:

- the values  $z_j \in \{0, 1\}$  that describe whether the  $j$ -th estimate is indeed related to the desired quantities  $x_1, \dots, x_n$ , and
- for estimates  $\tilde{y}_j$  which are not related to the desired quantities (i.e., for which  $z_j = 0$ ), the actual values  $y_j$  of the physical quantities which are estimated by these estimates.

**How to find all these unknowns?** A natural idea is to select the values of all these unknowns for which the degree  $D$  of possibility is the largest. This degree of possibility has the form

$$D = f_{\&}(D_1 \dots, D_\alpha, \dots),$$

where

$$D_\alpha = f_{\&}\{d_j : j \in S_\alpha\},$$

and  $d_j$  is the degree to which the values  $\tilde{y}_j, z_j$  (and, if needed,  $y_j$ ) are possible.

As we have already shown, maximizing the degree  $D$  is equivalent to maximizing the value

$$g(D) = \prod_{\alpha} \prod_{j \in S_{\alpha}} g(d_j).$$

When  $z_j = 1$ , we are interested in the degree to which  $\tilde{y}_j$  is related to the desired quantities  $x_i$  and the difference  $\Delta y_j = \tilde{y}_j - f(x, a_j, c)$  is possible. We know the degree  $q_j$  to which  $z_j = 1$ : this degree is equal to  $q_j$ . Thus,

$$d_j = f_{\&}(q_j, \mu_{\alpha}(\tilde{y}_j - f(x, a_j, c), \theta_j, \beta_{\alpha})),$$

hence

$$g(d_j) = g(q_j) \cdot g(\mu_{\alpha}(\tilde{y}_j - f(x, a_j, c), \theta_j, \beta_{\alpha})).$$

When  $z_j = 0$ , then we are interested in our degree of confidence that  $\tilde{y}_j$  is *not* related to the desired quantities  $x_i$  (this degree is equal to  $1 - q_j$ ) and that the difference  $\tilde{y}_j - y_j$  is possible. Thus,

$$d_j = f_{\&}(1 - q_j, \mu_{\alpha}(\tilde{y}_j - y_j, \theta_j, \beta_{\alpha})),$$

and

$$g(d_j) = g(1 - q_j) \cdot g(\mu_{\alpha}(\tilde{y}_j - y_j, \theta_j, \beta_{\alpha})).$$

Similar to the probabilistic case, the maximum is attained:

- when for  $z_j = 0$ , the membership function describing the estimation error reaches its maximum, and
- when we select  $z_j = 0$  or  $z_j = 1$  depending on which terms is larger.

Thus, we need to maximize the product  $E_1 \cdot \dots \cdot E_N$ , where, for each  $j \in S_{\alpha}$ , the expression  $E_j$  takes the form

$$E_j = \max \left( g(1 - q_j) \cdot \max_y g(\mu_{\alpha}(y, \theta_j, \beta_{\alpha})), \right. \\ \left. g(q_j) \cdot g(\mu_{\alpha}(\tilde{y}_j - f(x, a_j, c), \theta_j, \beta_{\alpha})) \right).$$

*Comment.* It should be mentioned that, in contrast to the previous case, when we did not take reliability into account, this problem is not mathematically the same as for the probabilistic case:

- there, we had the weights  $p_j$  and  $1 - p_j$  that add up to 1, while
- here, the weights  $g(q_j)$  and  $g(1 - q_j)$  do not necessarily add up to 1.

However, we can still use component-wise minimization to solve the corresponding optimization problem.

**Algorithm: fuzzy case.** Let us assume that we know how to solve the particular case of this problem when everything is perfectly reliable – e.g., we can do it by reducing this problem to the appropriate Maximum Likelihood problem. We will call the corresponding algorithm *original*.

Then, if we take reliability into account, we should do the following:

- 1) first, we pick  $z_j = 1$  for all  $j$  and use the original optimization method to optimize over  $x, c$ , and  $\beta$ ;
- 2) once we find the corresponding values of  $x, c$ , and  $\beta$ , we optimize over  $z_j$ : namely, we select  $z_j = 1$  if

$$g(q_j) \cdot g(\mu_{\alpha}(\tilde{y}_j - f(x, a_j, c), \theta_j, \beta_{\alpha})) \geq \\ d(1 - q_j) \cdot \max_y g(\mu_{\alpha}(y, \theta_j, \beta_{\alpha}));$$

for all other  $j$ , we select  $z_j = 0$ ;

- 3) then, only taking into account the estimates  $j$  selected on the previous step, we again use the original optimization method to find new estimates for  $x, c$ , and  $\beta$ , and go back to Step 2.

This process continues until the process converges, i.e., until the values of the desired variables  $x_1, \dots, x_n$  obtained on the next iteration are sufficiently close to the values from the previous iteration.

#### ACKNOWLEDGMENT

This work was supported in part by the Faculty of Economics, Chiang Mai University, 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.

#### REFERENCES

- [1] J. C. Bezdek and R. J. Hathaway, ‘‘Convergence of alternating optimization’’, *Journal Neural, Parallel, and Scientific Computations*, 2003, Vol. 11, No. 4, pp. 351–368.
- [2] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
- [3] H. T. Nguyen, V. Kreinovich, and P. Wojciechowski, ‘‘Strict Archimedean t-Norms and t-Conorms as Universal Approximators’’, *International Journal of Approximate Reasoning*, 1998, Vol. 18, Nos. 3–4, pp. 239–249.
- [4] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.
- [5] P. V. Novitskii and I. A. Zograph, *Estimating the Measurement Errors*, Energoatomizdat, Leningrad, 1991 (in Russian).
- [6] A. I. Orlov, ‘‘How often are the observations normal?’’, *Industrial Laboratory*, 1991, Vol. 57, No. 7, pp. 770–772.
- [7] S. G. Rabinovich, *Measurement Errors and Uncertainty. Theory and Practice*, Springer Verlag, Berlin, 2005.
- [8] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.
- [9] A. Welte, L. Jaulin, M. Ceberio, and V. Kreinovich, ‘‘Robust data processing in the presence of uncertainty and outliers: case of localization problems’’, *Proceedings of the IEEE Series of Symposia in Computational Intelligence SSCI’2016*, Athens, Greece, December 6–9, 2016.
- [10] L. A. Zadeh, ‘‘Fuzzy sets’’, *Information and Control*, 1965, Vol. 8, pp. 338–353.