# Why Cauchy Membership Functions: Reliability

#### Javier Viaña

University of Cincinnati, Cincinnati, OH 45219, USA

#### **Stephan Ralescu**

University of Cincinnati, Cincinnati, OH 45219, USA

### Kelly Cohen

University of Cincinnati, Cincinnati, OH 45219, USA

### **Vladik Kreinovich**

University of Texas at El Paso, El Paso, TX 79968, USA

#### Anca Ralescu

University of Cincinnati, Cincinnati, OH 45219, USA

#### Corresponding Author: Anca Ralescu

**Copyright** © 2022 Javier Viaña, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

An important step in designing a fuzzy system is the elicitation of the membership functions for the fuzzy sets used. Often the membership functions are obtained from data in a traininglike manner. They are expected to match or be at least compatible with those obtained from experts knowledgeable of the domain and the problem being addressed. In cases when neither are possible, e.g., insufficient data or unavailability of experts, we are faced with the question of hypothesizing the membership function. We have previously argued in favor of Cauchy membership functions (thus named because their expression is similar to that of the Cauchy distributions) and supported this choice from the point of view of *efficiency* of training. This paper looks at the same family of membership functions from the point of view of *reliability*.

keywords: Fuzzy sets, Membership functions, Cauchy membership function

vianajr@mail.uc.edu

ralescs@mail.uc.edu

cohenky@ucmail.uc.edu

vladik@utep.edu

ralescal@ucmail.uc.ed

385

## **1. INTRODUCTION**

By introducing the new concept of fuzzy sets Zadeh [1], opened up new possibilities in modeling of complex systems where uncertainty and imprecision are pervasive. Building on fuzzy sets, and the associated concept of *linguistic variable* [2], *computing with words* has seamlessly integrated computation and reasoning. In many practical applications of fuzzy techniques (see, e.g., [3-7,1]), the membership functions can be obtained from the experts. In other applications, the fuzzy sets are elicited directly from the data without the intervention of a human expert, imposing some condition on the underlying summarization procedure [8,9]. When these approaches are not possible, experiments (see, e.g., [10,11]) point to membership functions shown in equation (1) to work best:

$$\mu_{(k,a)}(x) = \frac{1}{1 + \frac{(x-a)^2}{k^2}}.$$
(1)

Since the function in (1) is similar to the known expression for the probability density function f(x) of a Cauchy distribution (see, e.g., [12], shown in (2), we refer to these membership function as *Cauchy* membership functions.

$$f(x) = \text{const} \cdot \frac{1}{1 + \frac{(x-a)^2}{k^2}}.$$
 (2)

In [13], we have shown that Cauchy membership function is desirable from the point of view of *efficiency*. In this paper we explore the desirability of the Cauchy membership function from the point of view of reliability. More precisely, we answer the following question "how can we explain the empirical fact – that Cauchy membership functions work better than other functions?".

The first step in answering such a question is to define in a precise manner what it is really meant by "work better" and we suggest that along with *efficiency*, a desirable property is that of *reliability*.

# 2. WHICH MEMBERSHIP FUNCTIONS LEAD TO THE MOST RELIABLE RESULTS

In considering the issue of reliability we start with the following general idea.

### General idea

We want to select membership functions for which we will be most confident in the results of the corresponding data processing. What often makes us more confident is when two different (unrelated) techniques lead to the same result – just like:

• when we have two experts making the same statement, it makes us more confident that this statement is true, or

• when two different measurements of the same quantity agree, this makes us more confident that both measurement results are correct.

### Specific idea

As often mentioned by Zadeh, although the nature of the uncertainty captured by fuzzy sets and probability is different. It can be said that while probability addresses 'lack of data', fuzzy sets address 'lack of definition'. In other words, if in the former we can expect to reduce uncertainty as more data becomes available, in the latter, the uncertainty (in fact imprecision) is not reduced by adding more data. At most, we expect this to allow us a more precise measure of the uncertainty.

Yet, from mathematical point of view, we can establish a relation between the probabilistic and fuzzy measures of uncertainty. It has been observed that given a membership function  $\mu(x)$  continuous or discrete, usually normalized such that  $\max_{x} \mu(x) = 1$ , one can define a probability distribution based on it.

For example, given the discrete fuzzy set (listed without loss of generality in nonincreasing order of its membership values),

$$1=\mu_{(1)}\geq\mu_{(2)}\geq\ldots\geq\mu_{(n)},$$

where the subscript (i) denotes the *i*th largest membership value, a discrete probability distribution with the probability mass function f can be defined as

$$f_{(i)} = \frac{\mu_{(i)}}{\sum_{j=1}^{n} \mu_{(j)}}.$$
(3)

For the continuous case, where  $\mu$  denotes a continuous membership function, a probability density function can be defined by

$$f(x) = \frac{\mu(x)}{\int \mu(y) \, dy}.$$
(4)

Conversely, the probability mass function and probability density distribution can be transformed into a membership function if we normalize it by dividing by its largest value, as shown in (5) and (6) respectively.

$$\mu_{(i)} = \frac{f_{(i)}}{f_{(1)}}.$$
(5)

$$\mu(x) = \frac{f(x)}{\max_{y} f(y)}.$$
(6)

Often use in image processing, for converting image histograms into fuzzy sets, Equations (5) and (6) correspond to the mechanism of the max-normalization of a histogram. The conversion from probability distributions to fuzzy sets is very important, as in practice, data is typically provided or summarized in terms of distributions. The fuzzy set representation can be thought of as being built on these distributions. An alternative approach, using the concepts of mass assignment theory [14],

and the subsequent work of [9], it can be shown that given a fuzzy set, there are several probability distributions that can be associated to it, and conversely, given a probability distribution, there are several fuzzy sets that can be associated to it. This correspondence is mediated by the focal elements of the mass assignment corresponding to the probability distribution and the *subjective probability distributions* within these focal elements. We require that the focal elements of the mass assignment coincide to the level sets of the fuzzy set, and therefore be nested.

In any case, it is reasonable to select a membership function  $\mu(x)$  for which fuzzy data processing will lead to the same result as using the corresponding subjective selection probabilities from the focal elements, and hence corresponding to the probability distribution f.

## 2.1 Data Processing: Reminder and the Resulting Explanation

What is data processing

- Whether we are using the known current values  $\tilde{x}_1, \ldots, \tilde{x}_n$  of different quantities  $x_1, \ldots, x_n$  to predict the future value of some physical quantity y,
- whether we are reconstructing the current value of some difficult-to-measure quantity y from the results  $\tilde{x}_1, \ldots, \tilde{x}_n$  of measuring related easier-to-measure quantities  $x_1, \ldots, x_n$ ,
- whether we are finding the best control y based on the known values  $x_1, \ldots, x_n$  of the related quantities,

in all these cases we have an algorithm f that transforms the known values  $\tilde{x}_1, \ldots, \tilde{x}_n$  into the desired estimate  $\tilde{y} = f(\tilde{x}_1, \ldots, \tilde{x}_n)$ .

Need to take uncertainty into account

The values  $\tilde{x}_i$  come from measurements or from expert estimates. Both measurement and expert estimates are never absolutely accurate: in general, each measurement result  $\tilde{x}_i$  is different from the actual (unknown) value  $x_i$ , i.e., there is a non-zero approximation error  $\Delta x_i \stackrel{\text{def}}{=} \tilde{x}_i - x_i$ . Because of this, the estimate  $\tilde{y}$  is, in general, different from the value  $y = f(x_1, \ldots, x_n)$  that we would have obtained if we used the actual values  $x_i = \tilde{x}_i - \Delta x_i$ . From the practical viewpoint, an important question is: how big is this difference  $\Delta y \stackrel{\text{def}}{=} \tilde{y} - y$ ?

In this section, we consider the case when our information about possible values of  $\Delta x_i$  is characterized in fuzzy terms, by a membership function.

### Linearization

In many practical situations, the approximation errors are relatively small. So, we can expand the expression for  $\Delta y$ :

$$\Delta y = \tilde{y} - y = f(\tilde{x}_1, \dots, \tilde{x}_n) - f(\tilde{x}_1 - \Delta x_1, \dots, \tilde{x}_n - \Delta x_n)$$

Javier Viaña, et al.

in Taylor series in terms of  $\Delta x_j$ , and ignore terms which are quadratic (or of higher order) in terms of  $\Delta x_j$ . In this approximation:

$$f(\widetilde{x}_1 - \Delta x_1, \dots, \widetilde{x}_n - \Delta x_n) = f(\widetilde{x}_1, \dots, \widetilde{x}_n) - \sum_{j=1}^n c_j \cdot \Delta x_j,$$

where we denoted  $c_j \stackrel{\text{def}}{=} \frac{\partial f}{\partial x_j} (\tilde{x}_1, \dots, \tilde{x}_n)$ . In this case, we get

$$\Delta y = \sum_{j=1}^{n} c_j \cdot \Delta x_j.$$
<sup>(7)</sup>

This is the case we consider in this section.

How to describe and process fuzzy uncertainty

We assume that for each estimate  $\tilde{x}_j$ , we have a numerical estimate  $\Delta_j$  of the corresponding approximation error. This assumption is in good accordance with the usual practice, according to which we say something like

" $x_j$  is approximately 1.0, with an error of about 0.1";

in this example, the estimate  $\tilde{x}_i$  is equal to 1.0, and  $\Delta_i = 0.1$ .

If we select a membership function  $\mu(x)$  corresponding to the case  $\Delta_j = 1$ , then for each *i* for which  $\Delta_j \neq 1$ , as a membership function for  $\Delta x_j$ , it s reasonable to take

$$\mu_j(\Delta x_j) = \mu\left(\frac{\Delta x_j}{\Delta_j}\right) \tag{8}$$

To process this fuzzy uncertainty, we can use Zadeh's extension principle, according to which the resulting membership function  $\mu_y(\Delta y)$  has the form

$$\mu_{y}(\Delta y) = \max\left\{\min(\mu_{1}(\Delta x_{1}),\ldots,\mu_{n}(\Delta x_{n})): \sum_{j=1}^{n} c_{j} \cdot \Delta x_{j} = \Delta y\right\}.$$

Since we have no information about the membership function  $\mu(x)$ , we have no reason to conclude that positive or negative values of x are more possible. Thus, it makes sense to assume that such values are equally possible, i.e., that  $\mu(x) = \mu(-x)$  for all x. It is known for such even functions  $\mu(x)$ , when all the membership function have the same shape – i.e., have the form (8) – then the resulting membership function also has the same form  $\mu_y(\Delta y) = \mu\left(\frac{\Delta y}{\Delta}\right)$ , where we denoted

$$\Delta = \sum_{j=1}^{n} |c_j| \cdot \Delta_j.$$
(9)

How to process the corresponding subjective probabilities

Based on each membership function (8), we form the corresponding probability density functions

$$f_i(\Delta x_i) = \operatorname{const} \cdot \mu_j(\Delta x_j) = \operatorname{const} \cdot \mu\left(\frac{\Delta x_j}{\Delta_j}\right).$$

One can easily check that if by  $\xi$  we denote a random variable corresponding to  $\Delta_j = 1$ , with probability density f(x), then the distribution of the random variable  $\xi_i$  corresponding to  $\Delta_j \neq 1$  is equivalent to the distribution of  $\Delta_i \cdot \xi$ . We therefore write that  $\xi_j = \Delta_j \cdot \xi^{(j)}$ , where  $\xi^{(j)}$  is distributed according to the distribution f(x) (corresponding to  $\Delta_j = 1$ ).

Since we have no reason to expect positive or negative correlation between these random variables, it makes sense to assume that they are independent. Thus, due to the formula (7), the random variable  $\xi_{y}$  corresponding to  $\Delta y$  has the form

$$\xi_y = \sum_{j=1}^n c_j \cdot \Delta_j \cdot \xi^{(j)}.$$
(10)

So, the condition that the resulting probability density will lead, after renormalization, to the membership function  $\mu_y(\Delta y) = \mu\left(\frac{\Delta y}{\Delta}\right)$ , with the value  $\Delta$  described by the formula (9), is equivalent to requiring that:

- for *n* independent identically distributed random variables  $\xi^{(i)}$ , with common probability density f(x),
- the distribution of their linear combination (10) is equivalent to the distribution of  $\Delta \cdot \xi$ , where  $\Delta$  is determined by the formula (9).

This condition can be described in terms of the characteristic functions  $\chi_{\alpha}(\omega) \stackrel{\text{def}}{=} E[\exp(i \cdot \omega \cdot \alpha)]$ , were  $E[\cdot]$  denotes the mean value and  $i \stackrel{\text{def}}{=} \sqrt{-1}$ . Indeed, from (10), we conclude that for

$$E[\exp(\mathbf{i}\cdot\boldsymbol{\omega}\cdot\boldsymbol{\xi}_{\gamma})] = E[\exp(\mathbf{i}\cdot\boldsymbol{\omega}\cdot\boldsymbol{\Delta}\cdot\boldsymbol{\xi})] = \chi_0(\boldsymbol{\Delta}\cdot\boldsymbol{\omega}), \tag{11}$$

where  $\chi_0$  denotes the characteristic function of the random variable  $\xi$ , we have

$$\begin{split} E[\exp(\mathbf{i}\cdot\boldsymbol{\omega}\cdot\boldsymbol{\xi}_{y})] &= E\left[\exp\left(\mathbf{i}\cdot\boldsymbol{\omega}\cdot\sum_{j=1}^{n}c_{j}\cdot\Delta_{j}\cdot\boldsymbol{\xi}^{(j)}\right)\right] \\ &= E\left[\prod_{j=1}^{n}\exp\left(\mathbf{i}\cdot\boldsymbol{\omega}\cdot c_{j}\cdot\Delta_{j}\cdot\boldsymbol{\xi}^{(j)}\right)\right]. \end{split}$$

Javier Viaña, et al.

Since the variables  $\xi^{(j)}$  are independent, the expected value of the product is equal to the product of expected values, i.e.,

$$E[\exp(\mathbf{i}\cdot\boldsymbol{\omega}\cdot\boldsymbol{\xi}_{y})] = \prod_{j=1}^{n} E\left[\exp\left(\mathbf{i}\cdot\boldsymbol{\omega}\cdot\boldsymbol{c}_{j}\cdot\boldsymbol{\Delta}_{j}\cdot\boldsymbol{\xi}^{(j)}\right)\right] = \prod_{j=1}^{n} \chi_{0}(\boldsymbol{c}_{j}\cdot\boldsymbol{\Delta}_{j}\cdot\boldsymbol{\omega}).$$
(12)

Comparing the expression (11) and (12), we conclude that

$$\chi_0\left(\left(\sum_{j=1}^n |c_j| \cdot \Delta_j\right) \cdot \omega\right) = \prod_{j=1}^n \chi_0(c_j \cdot \Delta_j \cdot \omega).$$
(13)

For any a > 0, for  $\omega = 1$ ,  $\Delta_1 = a$ , and  $c_1 = -1$ , we get  $\chi_0(a) = \chi_0(-a)$ , so the function  $\chi_0(a)$  is even. For any a > 0 and b > 0, for n = 2,  $\omega = 1$ ,  $\Delta_1 = a$ , and  $\Delta_2 = b$ , we conclude that

$$\chi_0(a+b) = \chi_0(a) \cdot \chi_0(b).$$
(14)

Taking logarithms of both sides, we get Cauchy's functional equation

$$\ell(a+b) = \ell(a) + \ell(b),$$

where  $\ell(a) \stackrel{\text{def}}{=} \ln(\chi_0(a))$  is measurable. It is known that the only measurable solutions of Cauchy's functional equation are linear functions, so  $\ell(a) = k \cdot a$  for some constant k, or  $\ln(\chi_0(a)) = k \cdot a$  and hence,  $\chi_0(a) = \exp(k \cdot a)$ . Since the function  $\chi_0(a)$  is even, we have  $\chi_0(a) = \exp(k \cdot |a|)$ .

The characteristic function is a Fourier transform of the probability density function. So, by applying the inverse Fourier transform to the characteristic function, we can reconstruct the probability density function. For the above expression, we obtain

$$f(x) = \frac{\text{const}}{1 + \frac{x^2}{k^2}},$$

from which after normalizing it back to the membership function, we get

$$\mu(x) = \frac{1}{1 + \frac{x^2}{k^2}},\tag{15}$$

which is exactly what we called Cauchy membership function.

From the membership function for the approximation error to the membership function for the actual quantity

According to equation (15), the membership function for each approximation error  $\Delta x$  should have the form

$$\mu_{\Delta x}(\Delta x) = \frac{1}{1 + \frac{(\Delta x)^2}{k^2}}.$$
(16)

Substituting the expression  $\Delta x = \tilde{x} - x$  into the formula (16), we get the membership function corresponding to each quantity x:

$$\mu_x(x) = \frac{1}{1 + \frac{(x-a)^2}{k^2}},\tag{17}$$

for a constant  $a \stackrel{\text{def}}{=} \widetilde{x}$ .

## **3. CONCLUSIONS**

For each membership function, we can process the corresponding uncertainty in two different ways. First, we can apply Zadeh's extension principle. Alternatively, we can:

- transform the corresponding membership functions into probability density functions,
- · process the corresponding random variable, and then
- transform the probability density function for the result back into a membership function.

The only case when these two results coincide, and thus when we have additional confidence in this joint result, is when we use the Cauchy membership functions shown in equation (17).

## 4. ACKNOWLEDGMENTS

This work was supported in part by a grant from the "la Caixa" Banking Foundation (ID 100010434), whose code is LCF/BQ/AA19/11720045, by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and HRD-1834620 and HRD-2034030 (CAHSI Includes), and by the AT&T Fellowship in Information Technology, and by the NSF CBET grant 1936908. It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478.

## References

- [1] Zadeh LA. "Fuzzy sets", Information and Control. 1965; 8:338–353.
- [2] Zadeh LA. The concept of a linguistic variable and its application to approximate reasoning—i. Information sciences. 1975;8:199–249.

- [3] Belohlavek R, Dauben JW, Klir GJ. Fuzzy Logic and Mathematics: A Historical Perspective, Oxford University Press, New York. 2017.
- [4] Klir G, Yuan B. Fuzzy Sets and Fuzzy Logic, Prentice Hall, Upper Saddle River, New Jersey. 1995.
- [5] Mendel JM. Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions, 2nd Edition. Springer, Cham, Switzerland. 2017.
- [6] Nguyen HT, Walker CL, Walker EA. A First Course in Fuzzy Logic, Chapman and Hall/CRC, Boca Raton, Florida. 2019.
- [7] Novák V, Perfilieva I, Močkoř J. Mathematical Principles of Fuzzy Logic. Kluwer Academic Publishers. Boston, Dordrecht. 1999.
- [8] Visa S, Ralescu A. Data-driven fuzzy sets for classification. International Journal of Advanced Intelligence Paradigms. 2008;1:3–30.
- [9] Ralescu A, Visa S. Obtaining fuzzy sets using mass assignment theory consistency with interpolation. In NAFIPS 2007-2007 Annual Meeting of the North American Fuzzy Information Processing Society. IEEE. 2007; 436–440.
- [10] Viaña J, Cohen K. "Fuzzy-based, noise-resilient, explainable algorithm for regression", Proceedings of the Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2021, West Lafayette, Indiana. 2021;7–9.
- [11] Viaña J, Ralescu S, Cohen K, Ralescu A, Kreinovich V. "Extension to multi- dimensional problems of a fuzzy-based explainable and noise-resilient algorithm", Proceedings of the 14th International Workshop on Constraint Programming and Decision Making CoProd'2021, Szeged, Hungary, September 12, 2021.
- [12] Sheskin DJ. Handbook of Parametric and Non-Parametric Statistical Procedures, Chapman Hall/CRC, London, UK. 2011.
- [13] Viaña J, Ralescu S, Cohen K, Kreinovich V, Ralescu A. Why Cauchy Membership Functions: Efficiency. Adv. Artif. Intell. Mach. Learn. 2021;1:86-93.
- [14] Baldwin JF. A theory of mass assignments for artificial intelligence. In Dimiter Driankov, Peter W. Eklund, and Anca L. Ralescu, editors, Fuzzy Logic and Fuzzy Control, Berlin, Heidelberg. Springer Berlin Heidelberg. 1994; 22–34.