

# Statistical Parameters Based on Fuzzy Measures

Fernando Reche , María Morales \*  and Antonio Salmerón 

Department of Mathematics and Center for the Development and Transfer of Mathematical Research to Industry (CDTIME), University of Almería, 04120 Almería, Spain; fernando.reche@ual.es (F.R.); antonio.salmeron@ual.es (A.S.)

\* Correspondence: maria.morales@ual.es

Received: 21 October 2020; Accepted: 9 November 2020; Published: 12 November 2020



**Abstract:** In this paper, we study the problem of defining statistical parameters when the uncertainty is expressed using a fuzzy measure. We extend the concept of monotone expectation in order to define a monotone variance and monotone moments. We also study parameters that allow the joint analysis of two functions defined over the same reference set. Finally, we propose some parameters over product spaces, considering the case in which a function over the product space is available and also the case in which such function is obtained by combining those in the marginal spaces.

**Keywords:** monotone statistical parameters; fuzzy measures; monotone measures; product spaces; fuzzy statistics

## 1. Introduction

Fuzzy measures [1], also known as capacities [2], non-additive measures or monotone measures [3], have shown to be a valuable tool for representing uncertainty, since they are able to cope with more general scenarios than probability measures do. Even though fuzzy measures have been successfully applied in a wide range of applications [4], no theory analogous to mathematical statistics has emerged around them in the general case, due to the difficulty of defining statistical parameters with a clear interpretation when additivity is replaced by monotonicity.

A remarkable exception is the case of the so-called imprecise probabilities [5,6], characterized by upper and lower expectations that provide rich semantics and interpretability. Dempster–Shafer belief functions [7,8], for instance, can be formulated as special cases of imprecise probabilities.

The field of fuzzy probability and statistics [9–14] has received significant attention during the last two decades. The contributions in this field can be classified into two basic groups according to the underlying approach they follow [15]. One of the groups include the methods that deal with the analysis of classical (non-fuzzy) data using methods based on fuzzy set theory, while the other group focuses on analyzing fuzzy data using statistical methods. In this context, fuzzy data refers to data in which the values correspond to fuzzy numbers [16], characterized by a membership function that returns a value between 0 and 1 indicating to which extent a given real number matches a given fuzzy number.

Examples within the first group include fuzzy clustering [17], fuzzy linear regression [18], testing fuzzy hypothesis from non-fuzzy data [19], fuzzy statistical quality control [20], time series forecasting based on fuzzy logic [21] and making statistical decisions with fuzzy utilities [22].

The second group includes methods for maximum likelihood estimation from fuzzy data [23], classification when data are labeled with Dempster–Shafer belief functions [24], distance-based statistical analysis [25], statistical hypothesis testing from fuzzy data [26], principal component analysis [27], discriminant analysis [28] and clustering [29].

In this paper, we are interested in the definition of statistical parameters when the uncertainty is represented by a general fuzzy measure. More precisely, our starting point is a measurable space

and a measurable real-valued function defined on the reference set of the space. We also assume that the measurable space is endowed with a fuzzy measure, and we will study the definition of statistical parameters over the measurable function, in a similar way as statistical parameters over a random variable can be defined from a probability measure. In this way, we attempt to handle more general scenarios than the ones covered by probability measures. To achieve this, we rely on the concept of monotone expectation [30]. We consider the case of marginal spaces as well as product spaces, and take advantage of recent advances in the construction of fuzzy measures over product spaces [31]. Our study is restricted to discrete reference sets.

The rest of the paper is organized as follows. Section 2 establishes the basic notation and definitions, and highlights the fundamental properties of product measures that are used throughout the paper. Section 3 contains the original contributions in this paper, in what concerns the definition of parameters in a marginal measurable space, while Section 4 describes our proposals for product spaces. The paper ends with conclusions in Section 5.

## 2. Preliminaries and Notation

**Definition 1.** [1] Let  $X$  be a set and  $\mathcal{A}$  be a non-empty class of subsets of  $X$  so that  $X \in \mathcal{A}$  and  $\emptyset \in \mathcal{A}$ . A function  $\mu : \mathcal{A} \rightarrow [0, 1]$  is a fuzzy measure if:

1.  $\mu(\emptyset) = 0$ .
2.  $\mu(X) = 1$ .
3.  $\forall A, B \in \mathcal{A}$  such that  $A \subseteq B$  it holds that  $\mu(A) \leq \mu(B)$ .
4. If  $\{A_n\}_{n \in \mathbb{N}} \in \mathcal{A}$  such that  $A_1 \subseteq A_2 \subseteq \dots$  and  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ , then  $\lim_n \mu(A_n) = \mu(\bigcup_{n=1}^{\infty} A_n)$ .
5. If  $\{A_n\}_{n \in \mathbb{N}} \in \mathcal{A}$  such that  $A_1 \supseteq A_2 \supseteq \dots$  and  $\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$ , then  $\lim_n \mu(A_n) = \mu(\bigcap_{n=1}^{\infty} A_n)$ .

The triplet  $(X, \mathcal{A}, \mu)$  is a measurable space, and  $X$  is called the reference set. We will only work with finite reference sets [4] in this paper. By default, we will assume that  $\mathcal{A}$  is the power set of  $X$ .

**Example 1** (Modified from [32]). Imagine there is a vehicle covering the connection between the harbor and the railway station in a city. This vehicle has four compartments: one for a car, one for a van, one for a motor-bike and another one for a bike. Assume that the gas tank of this vehicle has exactly the capacity necessary to carry the vehicle, with the four compartments busy, from the harbor to the railway station. Then we can regard this capacity to be equal to 1 unit. In this example,  $X = \{c, v, m, b\}$ , where  $c$  stands for car compartment busy,  $v$  for van compartment busy,  $m$  for motor-bike compartment busy and  $b$  for bike compartment busy. Assume also that the vehicle does not start the trip unless at least one of the compartments is busy. All the possible transportation situations are then the elements in  $\mathcal{A} = \mathcal{P}(X)$  ( $\mathcal{P}(X)$  stands for the power set of  $(X)$ ). In these conditions, for every  $A \subseteq X$ ,  $\mu(A)$  can be interpreted as the proportion of gas consumed if  $A$  happens. A possible specification of a fuzzy measure for this problem is as follows.

$$\begin{aligned} \mu(\{b\}) &= 0.1, \mu(\{v\}) = 0.4, \mu(\{c\}) = 0.3, \mu(\{m\}) = 0.2, \\ \mu(\{c, v\}) &= 0.6, \mu(\{c, b\}) = 0.35, \mu(\{c, m\}) = 0.45, \\ \mu(\{b, v\}) &= 0.42, \mu(\{b, m\}) = 0.21, \mu(\{v, m\}) = 0.68, \\ \mu(\{c, v, b\}) &= 0.7, \mu(\{c, v, m\}) = 0.75, \mu(\{c, b, m\}) = 0.5, \mu(\{v, b, m\}) = 0.69. \end{aligned}$$

Note how the fuzzy measure in Example 1 is non-additive. Therefore, the same information cannot be represented by a single probability distribution.

Every fuzzy measure over a reference set of cardinality  $n$  can be characterized by  $n!$  probability functions (not necessarily different) [33], each one of them corresponding to one possible permutation

of the reference set. Given a permutation  $\sigma$  of the set of indices  $\{1, \dots, n\}$ , we will denote by  $X^\sigma$  the ordering of the elements of  $X$  according to permutation  $\sigma$ , i.e.,  $X^\sigma = \{x_{\sigma(1)}, \dots, x_{\sigma(n)}\}$ . When it is clear from the context, we will drop  $\sigma$  from the subscripts and write  $X^\sigma = \{x_{(1)}, \dots, x_{(n)}\}$ .

**Definition 2.** [33] Let  $(X, \mathcal{A}, \mu)$  be a measurable space. The probability function associated with  $\mu$  and  $X^\sigma$  is defined as the set  $P_\sigma = \{p_\sigma(x_{(1)}), \dots, p_\sigma(x_{(n)})\}$  such that

$$p_\sigma(x_{(i)}) = \begin{cases} \mu(A_{(i)}) - \mu(A_{(i+1)}) & \text{if } i < n, \\ \mu(x_{(n)}) & \text{if } i = n, \end{cases} \tag{1}$$

where  $A_{(i)} = \{x_{(i)}, \dots, x_{(n)}\}$ .

**Definition 3.** [33] Let  $(X, \mathcal{A}, \mu)$  be a measurable space and let  $P_\sigma$  be the probability function associated with  $\mu$  and  $X^\sigma$ . The probability measure generated by  $\mu$  and  $X^\sigma$  is

$$P_\sigma(A) = \sum_{x \in A} p_\sigma(x), \quad \forall A \in \mathcal{A}. \tag{2}$$

We will use  $P_\sigma$  for both the probability function and the probability measure when it is clear from the context.

We will consider measures over marginal spaces  $(X, \mathcal{A})$  as well as product spaces  $(X_1 \times X_2, \mathcal{A}_{X_1 \times X_2})$  resulting from composing the marginal spaces  $(X_1, \mathcal{A}_{X_1})$  and  $(X_2, \mathcal{A}_{X_2})$ , with  $\mathcal{A}_{X_1 \times X_2} = \mathcal{P}(X_1 \times X_2)$ , which is not the same as  $\mathcal{P}(X_1) \times \mathcal{P}(X_2)$ .

Of particular interest are the elements of a product class that can be obtained from sets in the marginal space. They are called *rectangles* and are formally defined as follows:

**Definition 4.** Let  $(X_1, \mathcal{A}_{X_1})$  and  $(X_2, \mathcal{A}_{X_2})$  be two spaces where  $\mathcal{A}_{X_1}$  and  $\mathcal{A}_{X_2}$  are classes defined on  $X_1$  and  $X_2$ , respectively. The class of rectangles of  $\mathcal{A}_{X_1 \times X_2}$  is

$$\mathcal{R} = \{H \in \mathcal{A}_{X_1 \times X_2} \mid H = A \times B, \text{ where } A \in \mathcal{A}_{X_1}, B \in \mathcal{A}_{X_2}\}. \tag{3}$$

Our proposals in this paper will be based on the product measures described in [31], which make use of the concept of triangular norm and conorm.

**Definition 5.** [34] An operator  $T : [0, 1]^2 \rightarrow [0, 1]$  is a triangular norm or t-norm for short, if it satisfies the following conditions:

1.  $T(0, a) = 0, T(a, 1) = a$  for all  $a \in [0, 1]$ . (Boundary conditions)
2.  $T(a, b) = T(b, a)$ . (Commutativity)
3. If  $a \leq c$  and  $b \leq d$ , then  $T(a, b) \leq T(c, d)$ . (Monotonicity)
4.  $T(T(a, b), c) = T(a, T(b, c))$ . (Associativity)

**Definition 6.** [34] An operator  $T : [0, 1]^2 \rightarrow [0, 1]$  is a triangular conorm or t-conorm for short, if it satisfies the following properties:

1.  $S(1, a) = 1, S(a, 0) = a$  for all  $a \in [0, 1]$ . (Boundary conditions)
2.  $S(a, b) = S(b, a)$ . (Commutativity)
3. If  $a \leq c$  and  $b \leq d$ , then  $S(a, b) \leq S(c, d)$ . (Monotonicity)
4.  $S(S(a, b), c) = S(a, S(b, c))$ . (Associativity)

The usual way of integrating real functions with respect to a fuzzy measure is by means of the so-called Choquet integral, which is a generalization of Lebesgue integral to monotone measures.

**Definition 7.** [2] Let  $(X, \mathcal{A}, \mu)$  be a measurable space, and let  $h$  be a measurable real function of  $X$ . The Choquet integral of  $h$  with respect to  $\mu$  is

$$\oint_A h \circ \mu = \int_{-\infty}^0 (\mu(H_\alpha \cap A) - 1) d\alpha + \int_0^\infty \mu(H_\alpha \cap A) d\alpha \tag{4}$$

where  $A \in \mathcal{A}$  and  $H_\alpha$  are the  $\alpha$ -cuts of  $h$ , defined as

$$H_\alpha = \{x \in X / h(x) \geq \alpha\}. \tag{5}$$

If the reference set is finite, the integral can be expressed as

$$\oint h \circ \mu = h(x_{(1)})\mu(A_{(1)}) + \sum_{i=2}^n \mu(A_{(i)})[h(x_{(i)}) - h(x_{(i-1)})], \tag{6}$$

where  $X^\sigma$  is an ordering such that  $h(x_{(1)}) \leq h(x_{(2)}) \leq \dots \leq h(x_{(n)})$  and the sets  $A_{(i)}$  are of the form  $\{x_{(i)}, x_{(i+1)}, \dots, x_{(n)}\}$ . Furthermore, if  $h$  is non-negative, it can be computed as

$$\oint h \circ \mu = \sum_{i=1}^n h(x_{(i)})p_\sigma(x_{(i)}), \quad p_\sigma \in P_h, \tag{7}$$

where  $P_h$  is the probability function associated with the ordering  $X^\sigma$  induced by  $h$  (see Definition 2).

Given two measurable spaces  $(X_1, \mathcal{A}_{X_1}, \mu_1)$  and  $(X_2, \mathcal{A}_{X_2}, \mu_2)$ , the concept of product fuzzy measure is defined as follows.

**Definition 8.** [31] A product fuzzy measure of  $\mu_1$  and  $\mu_2$  is a function  $\mu_{12} : \mathcal{A}_{X_1 \times X_2} \rightarrow [0, 1]$  satisfying:

1.  $\mu_{12}(\emptyset) = 0, \mu_{12}(X_1 \times X_2) = 1$ .
2. For all  $A, B \in \mathcal{A}_{X_1 \times X_2}$  such that  $A \subseteq B$  it holds that  $\mu_{12}(A) \leq \mu_{12}(B)$ .
3. For all  $A \in \mathcal{A}_{X_1}$ , it holds that  $\mu_{12}(A \times X_2) = \mu_1(A)$ .
4. For all  $B \in \mathcal{A}_{X_2}$ , it holds that  $\mu_{12}(X_1 \times B) = \mu_2(B)$ .

The next definitions particularize the concept of a fuzzy measure product, so that it is guaranteed to be compatible with the intuitive idea of independence, in the sense that if two fuzzy measures are independent, their fuzzy measure product should be possible to be obtained using exclusively the two original fuzzy measures.

**Definition 9.** [31] Let  $(X_1, \mathcal{A}_{X_1}, \mu_1)$  and  $(X_2, \mathcal{A}_{X_2}, \mu_2)$  be measurable spaces.  $\mu_1$  and  $\mu_2$  are  $\odot$ -independent fuzzy measures if there exists a product fuzzy measure  $\mu_{12}^\odot$  such that for any  $H \in \mathcal{R}$ ,

$$\mu_{12}^\odot(H) = \mu_1(A) \odot \mu_2(B), \tag{8}$$

where  $H = A \times B$  and  $\odot$  is a t-norm.  $\mu_{12}^\odot$  is called the  $\odot$ -independent product of  $\mu_1$  and  $\mu_2$ .

**Definition 10.** [31] Let  $(X_1, \mathcal{A}_{X_1}, \mu_1)$  and  $(X_2, \mathcal{A}_{X_2}, \mu_2)$  be measurable spaces. The  $\odot$ -exterior product measure for any  $H \in \mathcal{A}_{X_1 \times X_2}$  is defined as

$$\bar{\mu}_{12}^\odot(H) = \min_{A \times B \supseteq H} \mu_1(A) \odot \mu_2(B), \tag{9}$$

where  $\odot$  is a t-norm.

**Definition 11.** [31] Let  $(X_1, \mathcal{A}_{X_1}, \mu_1)$  and  $(X_2, \mathcal{A}_{X_2}, \mu_2)$  be measurable spaces. The  $\odot$ -interior product measure for any  $H \in \mathcal{A}_{X_1 \times X_2}$  is defined as

$$\mu_{12}^{\odot}(H) = \max_{A \times B \subseteq H} \mu_1(A) \odot \mu_2(B), \tag{10}$$

where  $\odot$  is a t-norm.

Both measures conform to lower and upper bounds for any  $\odot$ -independent product fuzzy measure.

**Proposition 1.** [31] Let  $(X_1, \mathcal{A}_{X_1}, \mu_1)$  and  $(X_2, \mathcal{A}_{X_2}, \mu_2)$  be measurable spaces. Given any  $\odot$ -independent product of  $\mu_1$  and  $\mu_2$ , it holds that for all  $C \in \mathcal{A}_{X_1 \times X_2}$ ,

$$\underline{\mu}_{12}^{\odot}(C) \leq \mu_{12}^{\odot}(C) \leq \bar{\mu}_{12}^{\odot}(C). \tag{11}$$

Note that, for the particular case of the class  $\mathcal{R}$ , both measures coincide [31], i.e., for all  $H \in \mathcal{R}$ ,

$$\underline{\mu}_{12}^{\odot}(H) = \mu_{12}^{\odot}(H) = \bar{\mu}_{12}^{\odot}(H). \tag{12}$$

Product fuzzy measures can also be defined in terms of the associated probability measures [31].

**Definition 12.** [31] Let  $(X_1, \mathcal{A}_{X_1}, \mu_1)$  and  $(X_2, \mathcal{A}_{X_2}, \mu_2)$  be measurable spaces and  $P_{\sigma_1}^{\mu_1}$  and  $P_{\sigma_2}^{\mu_2}$  be the probability functions associated with  $X_1^{\sigma_1}$  and  $X_2^{\sigma_2}$ , respectively. The lower product p-measure is defined as

$$\underline{m}_{12}(C) = \min_{\sigma_1, \sigma_2} \left[ P_{\sigma_1}^{\mu_1} \otimes P_{\sigma_2}^{\mu_2}(C) \right], \tag{13}$$

for all  $C \in \mathcal{A}_{X_1 \times X_2}$ , where  $\otimes$  is the standard probabilistic product, i.e.,  $P_{\sigma_1}^{\mu_1} \otimes P_{\sigma_2}^{\mu_2}(C) = P_{\sigma_1}^{\mu_1}(C) P_{\sigma_2}^{\mu_2}(C)$ .

**Definition 13.** [31] Given the conditions in Definition 12, the upper product p-measure is defined as

$$\bar{m}_{12}(C) = \max_{\sigma_1, \sigma_2} \left[ P_{\sigma_1}^{\mu_1} \otimes P_{\sigma_2}^{\mu_2}(C) \right], \tag{14}$$

where  $\otimes$  is the standard probabilistic product.

### 3. Parameters over One Measurable Space

In this section we propose statistical parameters aimed to characterize the behavior of functions defined on a measurable space endowed with a fuzzy measure. We will separately address the case of analyzing a single function and the case of simultaneously analyzing two functions.

#### 3.1. The Case of Only One Function

Our proposals rely on the extension of the concept of mathematical expectation associated with probability measures, to the more general case of fuzzy measures. Consider a measurable space  $(X, \mathcal{A}, \mu)$  where  $\mu$  is a fuzzy measure, and the class  $\mathfrak{P}$  of all the additive measures over  $X$ . One way to extend the concept of mathematical expectation [5,35] is based on defining the set

$$\mathcal{M}_P(\mu) = \{P \in \mathfrak{P} | P(A) \geq \mu(A), \forall A \in \mathcal{A}\} \tag{15}$$

of all the probability measures that dominate the fuzzy measure  $\mu$ .

Since all the elements in  $\mathcal{M}_P(\mu)$  are additive measures, the expectation of a function  $h$  with respect to a fuzzy measure  $\mu$  can be defined as

$$E_{\mu}(h) = \min_{P \in \mathcal{M}_P(\mu)} E_P(h), \tag{16}$$

where  $E_P(h)$  is the mathematical expectation of  $h$  with respect to the probability measure  $P$ .

The problem of this definition is that it is not always well defined, since there can exist a fuzzy measure  $\mu$  for which  $\mathcal{M}_P(\mu) = \emptyset$ . It happens, for instance, when the sum of the fuzzy measure  $\mu$  over the unitary subsets of  $X$  is greater than 1, as it is not possible to find a probability measure bounding  $\mu$  from above.

A class of fuzzy measures that are compatible with the definition of expectation in Equation (16) are those that conform a lower envelope of a set of probability measures [6], i.e.,  $\mu(A) = \min\{P(A) | P \in \mathcal{M} \subseteq \mathfrak{P}\}$ , because in that case  $\mathcal{M}_P(\mu) \neq \emptyset$ .

A more general definition of expectation, based on Choquet integral [2], was given in [30] with the aim of extending the probabilistic concept of expectation to non-additive settings.

**Definition 14.** [30] Let  $(X, \mathcal{A}, \mu)$  be a measurable space and let  $h$  be a non-negative, real valued measurable function of  $X$ . The monotone expectation of  $h$  with respect to the fuzzy measure  $\mu$  is defined as

$$E_\mu(h) = \int h \circ \mu. \tag{17}$$

Since a fuzzy measure can always be characterized by a set of probability measures, it is clear from Definition 14 and Equation (7) that the monotone expectation is equal to the mathematical expectation obtained with the probability function associated with the fuzzy measure  $\mu$  and the ordering induced by the function  $h$  (see Definition 2), i.e.,

$$E_\mu(h) = E_{P_{\mu,h}}(h), \tag{18}$$

where  $P_{\mu,h}$  denotes the probability function associated with  $\mu$  and the ordering induced by  $h$ . In the particular case of considering a finite reference set, the monotone expectation can be expressed as

$$E_\mu(h) = \sum_{i=1}^n h(x_{(i)}) p_\sigma(x_{(i)}), \quad p_\sigma \in P_{\mu,h}. \tag{19}$$

The relation between the monotone expectation and the mathematical expectation is also illustrated in Proposition 2.

**Proposition 2.** [30] Let  $(X, \mathcal{A}, \mu)$  be a measurable space and let  $\{P_\sigma, \sigma \in S_n\}$  be the set of all the probability functions associated with the fuzzy measure  $\mu$ . Then, for any non-negative real valued, measurable function  $h$  of  $X$  it holds that

$$\min_{\sigma} E_{P_\sigma}(h) \leq E_\mu(h) \leq \max_{\sigma} E_{P_\sigma}(h). \tag{20}$$

### 3.1.1. Monotone Variance

In the same way as the monotone expectation extends in a natural way the concept of mathematical expectation to non-additive measures, we will pursue the extension of other statistical parameters in a similar way.

We will start off considering the extension of the concept of variance to a non-monotone context. A direct approach is to define an extension of the variance using Choquet integral, as in the case of the monotone expectation, which yields

$$\text{Var}_\mu(h) = E_\mu[(h - E_\mu(h))^2]. \tag{21}$$

However, the definition of variance in Equation (21) is problematic, since the distribution associated with  $\mu$  and the ordering induced by  $h$  is not, in general, the same as the one induced by  $(h - E_\mu(h))^2$ . The reason is that functions  $h$  and  $(h - E_\mu(h))^2$  are not comonotone, and therefore they may induce different orderings of the reference set. Hence, the monotone variance defined in this

way could not be considered as a measure of dispersion with respect to the monotone expectation, as the underlying probability distribution can be different (see Definition 2).

Taking this into account, we propose a definition of monotone variance that preserves the underlying probability measure associated with  $\mu$  and the ordering induced by  $h$ .

**Definition 15.** Let  $(X, \mathcal{A}, \mu)$  be a measurable space and let  $h$  be a non-negative real valued measurable function of  $X$ . We define the monotone variance of  $h$  with respect to the fuzzy measure  $\mu$  as

$$\text{Var}_\mu(h) = \text{Var}_{P_{\mu,h}}(h), \tag{22}$$

where  $P_{\mu,h}$  is the probability function associated with  $\mu$  and the ordering induced by  $h$ .

It is clear from the definition that  $\text{Var}_\mu(h) \geq 0$  and that it is equal to the traditional variance when  $\mu$  is a probability measure.

**Example 2.** Consider the fuzzy measure over the reference set  $X = \{x_1, x_2, x_3\}$  and its associated probability distributions in Table 1, and the function  $h$  defined as  $h(x_1) = 0.4, h(x_2) = 0.1$  and  $h(x_3) = 0.7$ . The ordering of  $X$  induced by  $h$  is thus  $(x_2, x_1, x_3)$ , i.e., the ordering induced by permutation  $\sigma = (2, 1, 3)$ , which corresponds to the probability distribution  $P_{(2,1,3)}$ . Therefore, according to Equation (22), the monotone variance of  $h$  is just the variance of  $h$  computed using probability distribution  $P_{(2,1,3)}$ , resulting in

$$\text{Var}_\mu(h) = 0.0621.$$

**Table 1.** A fuzzy measure and the associated probability distributions corresponding to all the possible permutations of the indices  $(1, 2, 3)$ .

$\mathcal{A}$	$\mu$	$P_{(1,2,3)}$	$P_{(1,3,2)}$	$P_{(2,1,3)}$	$P_{(2,3,1)}$	$P_{(3,1,2)}$	$P_{(3,2,1)}$
$x_1$	0.2	0.6	0.6	0.3	0.2	0.4	0.2
$x_2$	0.1	0.1	0.1	0.4	0.4	0.1	0.3
$x_3$	0.3	0.3	0.3	0.3	0.4	0.5	0.5
$x_1, x_2$	0.5	0.7	0.7	0.7	0.6	0.5	0.5
$x_1, x_3$	0.6	0.9	0.9	0.6	0.6	0.9	0.7
$x_2, x_3$	0.4	0.4	0.4	0.7	0.8	0.6	0.8

Our definition of monotone variance preserves some properties of the traditional variance, likewise the monotone expectation preserves some properties of the mathematical expectation. In particular, the result in Theorem 1 is of practical value as it simplifies the calculation, and it is also of interest because it links the concepts of monotone variance and monotone expectation.

**Theorem 1.** Let  $(X, \mathcal{A}, \mu)$  be a measurable space and let  $h$  be a non-negative real valued measurable function of  $X$ , then it holds that

$$\text{Var}_\mu(h) = E_\mu(h^2) - E_\mu^2(h). \tag{23}$$

**Proof.** According to Equation (22),

$$\text{Var}_\mu(h) = \text{Var}_{P_{\mu,h}}(h),$$

i.e., the variance of  $h$  computed according to probability distribution  $P_{\mu,h}$ , which can be calculated as

$$\text{Var}_{P_{\mu,h}}(h) = E_{P_{\mu,h}}(h^2) - \left[ E_{P_{\mu,h}}(h) \right]^2,$$

and thus

$$\text{Var}_\mu(h) = E_{P_{\mu,h}}(h^2) - [E_{P_{\mu,h}}(h)]^2. \tag{24}$$

The functions  $h$  and  $h^2$  are comonotone, and therefore they induce the same ordering of the reference set and hence yield the same associated probability distribution (see Definition 2). Thus, it holds that  $P_{\mu,h} = P_{\mu,h^2}$  and therefore,

$$E_{P_{\mu,h}}(h^2) = E_{P_{\mu,h^2}}(h^2).$$

In addition, according to Equation (18),  $E_\mu(h^2) = E_{P_{\mu,h^2}}(h^2) = E_{P_{\mu,h}}(h^2)$  and  $E_\mu(h) = E_{P_{\mu,h}}(h)$ . Now, replacing  $E_{P_{\mu,h}}(h^2)$  by  $E_\mu(h^2)$  and  $E_{P_{\mu,h}}(h)$  by  $E_\mu(h)$  in Equation (24) we obtain Equation (23).  $\square$

**Example 3.** As a continuation of Example 2, we will compute  $\text{Var}_\mu(h)$  using Equation (23).

$$\begin{aligned} E_\mu(h^2) &= E_{P_{(2,1,3)}}(h^2) \\ &= 0.3 \cdot 0.4^2 + 0.4 \cdot 0.1^2 + 0.3 \cdot 0.7^2 = 0.199. \end{aligned}$$

$$\begin{aligned} E_\mu(h) &= E_{P_{(2,1,3)}}(h) \\ &= 0.3 \cdot 0.4 + 0.4 \cdot 0.1 + 0.3 \cdot 0.7 = 0.37. \end{aligned}$$

Hence,

$$\text{Var}_\mu(h) = 0.199 - 0.37^2 = 0.0621.$$

The next result shows that the monotone variance behaves in a similar way as traditional variance in relation to affine transformations.

**Proposition 3.** Assume the conditions in Theorem 1 and let  $t$  be a function defined as  $t = ah + b$  with  $a \in \mathbb{R}_0^+$  and  $b \in \mathbb{R}$ . It holds that

$$\text{Var}_\mu(t) = a^2 \text{Var}_\mu(h). \tag{25}$$

**Proof.** First, we have to show that  $t$  and  $h$  are comonotone, i.e., that for all  $x, y \in X$ ,  $(h(x) - h(y))$  and  $(t(x) - t(y))$  have the same sign:

$$(h(x) - h(y))(t(x) - t(y)) = (h(x) - h(y))(ah(x) + b - ah(y) - b) = a(h(x) - h(y))^2 \geq 0,$$

since  $a \in \mathbb{R}_0^+$ . Therefore, the probability distribution associated with the measure  $\mu$  is the same for both functions, i.e.,  $P_{\mu,t} = P_{\mu,h}$  and thus

$$\text{Var}_\mu(t) = \text{Var}_{P_{\mu,t}}(t) = \text{Var}_{P_{\mu,h}}(t) = a^2 \text{Var}_{P_{\mu,h}}(h) = a^2 \text{Var}_\mu(h).$$

$\square$

The next results analyze when the monotone variance is equal to 0.

**Theorem 2.** Let  $(X, \mathcal{A}, \mu)$  be a measurable space and let  $h$  be a non-negative real valued measurable function of  $X$ . Let  $P_{\mu,h}$  be the probability function associated with  $\mu$  and  $h$ . Then, the following three conditions are equivalent:

1.  $\text{Var}_\mu(h) = 0$ .
2.  $\exists! i (1 \leq i \leq n)$  such that  $p_\sigma(x_i) = 1$  and  $p_\sigma(x_j) = 0, \forall j \neq i$ , with  $p_\sigma \in P_{\mu,h}$ .
3.  $\exists i (1 \leq i \leq n)$  such that

$$\mu(H_{\alpha_j}) = 1, \quad \forall j \leq i \quad \text{and} \quad \mu(H_{\alpha_j}) = 0, \quad \forall j > i$$



where  $H_{\alpha_i} = \{x \in X | h(x) \geq h(x_i)\}, i = 1, \dots, n$ .

**Proof.** Let us assume without loss of generality that

$$h(x_1) \leq h(x_2) \leq \dots \leq h(x_n). \tag{26}$$

(1)  $\implies$  (2)

Since  $p_\sigma(x_i) \geq 0, i = 1, \dots, n$ , and  $\sum_{i=1}^n p_\sigma(x_i) = 1$ , there must be at least one  $i \in \{1, \dots, n\}$  such that  $p_\sigma(x_i) \neq 0$ .

Suppose that  $\text{Var}_\mu(h) = 0$  and there exist two different  $j, k \in \{1, \dots, n\}, j < k$ , such that  $p_\sigma(x_j) \neq 0$  and  $p_\sigma(x_k) \neq 0$ . Then it holds that

$$\text{Var}_\mu(h) = \text{Var}_{p_\sigma}(h) = p_\sigma(x_j)(h(x_j) - E_\mu(h))^2 + p_\sigma(x_k)(h(x_k) - E_\mu(h))^2 = 0,$$

which means that  $h(x_j) = E_\mu(h)$  and  $h(x_k) = E_\mu(h)$ . However, according to the assumption in Equation (26), it holds that  $h(x_j) \leq h(x_{j+1}) \leq \dots \leq h(x_{k-1}) \leq h(x_k)$ . Hence,

$$E_\mu(h) = h(x_j) \leq h(x_{j+1}) \leq \dots \leq h(x_{k-1}) \leq h(x_k) = E_\mu(h),$$

which means that

$$\begin{aligned} h(x_j) = h(x_{j+1}) = \dots = h(x_k) = E_\mu(h) &\implies H_{\alpha_j} = H_{\alpha_{j+1}} = \dots = H_{\alpha_k} \\ &\implies p_\sigma(x_j) = p_\sigma(x_{j+1}) = \dots = p_\sigma(x_{k-1}) = 0, \end{aligned}$$

which is a contradiction with the assumption that  $p_\sigma(x_j) \neq 0$ . Thus, there is only one  $p_\sigma(x_i) \neq 0$  and furthermore,  $p_\sigma(x_i) = 1$ .

(2)  $\implies$  (3)

Assume  $\exists! i$  such that  $p_\sigma(x_i) \neq 0$ . Then,

$$p_\sigma(x_1) = p_\sigma(x_2) = \dots = p_\sigma(x_{i-1}) = p_\sigma(x_{i+1}) = \dots = p_\sigma(x_n) = 0$$

and therefore

$$\mu(H_{\alpha_1}) = \mu(H_{\alpha_2}) = \dots = \mu(H_{\alpha_i})$$

and

$$\mu(H_{\alpha_{i+1}}) = \mu(H_{\alpha_{i+2}}) = \dots = \mu(H_{\alpha_n}).$$

On the other hand, since  $\mu(H_{\alpha_i}) - \mu(H_{\alpha_{i+1}}) = 1$ , it follows that  $\mu(H_{\alpha_j}) = 1$  if  $j \leq i$  and  $\mu(H_{\alpha_j}) = 0$  if  $j > i$ . (3)  $\implies$  (1)

It is straightforward from the definition of monotone variance.  $\square$

**Corollary 1.** If  $h$  is constant, then  $\text{Var}_\mu(h) = 0$ , for any fuzzy measure  $\mu$ .

**Example 4.** Assume a function  $h$  defined on  $X = \{x_1, x_2, x_3\}$  as  $h(x_1) = 0.4, h(x_2) = 0.1$  and  $h(x_3) = 0.7$ , and an associated probability distribution  $p_\sigma$  such that  $p_\sigma(x_1) = 1, p_\sigma(x_2) = 0$  and  $p_\sigma(x_3) = 0$ . We will see how the monotone variance is equal to 0. However, first we need to calculate the monotone expectation.

$$E_\mu(h) = 1 \cdot 0.4 + 0 \cdot 0.1 + 0 \cdot 0.7 = 0.4.$$

Thus,

$$\text{Var}_\mu(h) = 1 \cdot (0.4 - 0.4)^2 + 0 \cdot (0.1 - 0.4)^2 + 0 \cdot (0.7 - 0.4)^2 = 0.$$

Now we will calculate the value of the measure  $\mu$  over the sets  $H_{\alpha_i} = \{x \in X | h(x) \geq h(x_i)\}, i = 1, 2, 3$ , i.e.,  $H_{\alpha_1} = \{x_1, x_3\}$ ,  $H_{\alpha_2} = \{x_1, x_2, x_3\}$  and  $H_{\alpha_3} = \{x_3\}$ .

We can obtain the values of  $\mu$  from  $p_\sigma$  using Definition 2. The result is

$$\begin{aligned} \mu(\{x_3\}) &= p_\sigma(x_3) = 0, \\ \mu(\{x_1, x_3\}) &= p_\sigma(x_1) + \mu(\{x_3\}) = 1 + 0 = 1, \\ \mu(\{x_1, x_2, x_3\}) &= p_\sigma(x_2) + \mu(\{x_1, x_3\}) = 0 + 1 = 1. \end{aligned}$$

Therefore,  $\mu(H_{\alpha_1}) = 1$ ,  $\mu(H_{\alpha_2}) = 1$  and  $\mu(H_{\alpha_3}) = 0$ .

### 3.1.2. Monotone Moments

Following the same idea underlying the definition of monotone variance, we can extend the concepts of central and non-central moments from a probabilistic setting to a monotone one.

**Definition 16.** Let  $(X, \mathcal{A}, \mu)$  be a measurable space and let  $h$  be a non-negative real valued measurable function of  $X$ . We define the  $k$ -th non-central monotone moment of  $h$  with respect to  $\mu$  as

$$g_\mu^k(h) = E_\mu(h^k). \tag{27}$$

Note that Equation (27) is well defined, since  $h$  and  $h^k$  are comonotone, and therefore the corresponding probability function is the same for both of them, regardless of the value of  $k$ .

The definition of central monotone moments is, however, more problematic. If we follow the same idea as in Definition 16, and define the central monotone moment as  $E_\mu(h - E_\mu(h))^k$ , we find the problem that functions  $h$  and  $(h - E_\mu(h))^k$  are not comonotone, and that would mean that different underlying probability distributions would be used to compute  $E_\mu(h)$  and  $E_\mu(h - E_\mu(h))^k$ . We will therefore generalize the definition of monotone variance to values of  $k \neq 2$ , utilizing the probability function associated with  $\mu$  and  $h$ .

**Definition 17.** Let  $(X, \mathcal{A}, \mu)$  be a measurable space and let  $h$  be a non-negative real valued measurable function of  $X$ . We define the  $k$ -th central monotone moment of  $h$  with respect to  $\mu$  as

$$\gamma_\mu^k(h) = E_{P_{\mu,h}} \left[ (h - E_\mu(h))^k \right], \tag{28}$$

where  $P_{\mu,h}$  is the probability function associated with  $\mu$  and  $h$ .

The following result establishes the relation between central and non-central monotone moments.

**Proposition 4.** Let  $(X, \mathcal{A}, \mu)$  be a measurable space and let  $h$  be a non-negative real valued measurable function of  $X$ . It holds that

$$\gamma_\mu^k(h) = \sum_{j=0}^k (-1)^j \binom{k}{j} [g_\mu(h)]^j g_\mu^{k-j}(h). \tag{29}$$

**Proof.** Assume  $X = \{x_1, \dots, x_n\}$ .

$$\begin{aligned} \gamma_\mu^k(h) &= E_{P_{\mu,h}} \left[ (h - E_\mu(h))^k \right] = \sum_{i=1}^n (h(x_i) - E_\mu(h))^k P_{\mu,h}(x_i) \\ &= \sum_{i=1}^n \sum_{j=0}^k (-1)^j \binom{k}{j} E_\mu^j(h) h^{k-j}(x_i) P_{\mu,h}(x_i) \\ &= \sum_{j=0}^k (-1)^j \binom{k}{j} E_\mu^j(h) \sum_{i=1}^n h^{k-j}(x_i) P_{\mu,h}(x_i) \\ &= \sum_{j=0}^k (-1)^j \binom{k}{j} E_\mu^j(h) E_\mu(h^{k-j}) = \sum_{j=0}^k (-1)^j \binom{k}{j} [g_\mu(h)]^j g_\mu^{k-j}(h). \end{aligned}$$

□

### 3.2. The Case of Two Functions

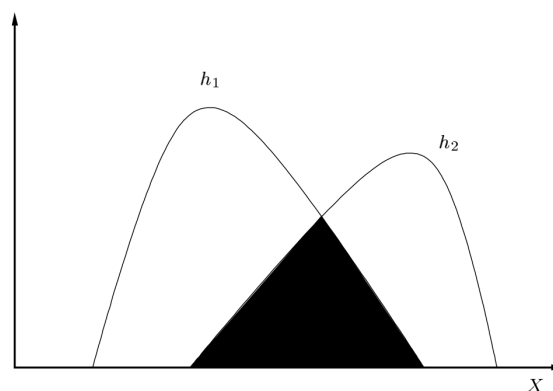
In this section we approach the simultaneous analysis of two functions  $h_1$  and  $h_2$  over the same reference set,  $X$ . Our goal is to model the information that both functions have in common, or the way in which they interact with one another.

Generalizing the concept of covariance, for instance, by using  $E_\mu[(h_1 - E_\mu(h_1))(h_2 - E_\mu(h_2))]$ , raises the problem that the underlying probability distribution used to compute the monotone expectation is not the one induced by  $h_1$  nor by  $h_2$  for the same fuzzy measure  $\mu$ , and therefore it is not clear that this monotone covariance in fact measures the relationship between both functions at all. We will therefore explore a different approach, in which we will model the degree of similarity between  $h_1$  and  $h_2$ , by measuring the common region determined by both functions.

**Definition 18.** Let  $(X, \mathcal{A}, \mu)$  be a measurable space and let  $h_1$  and  $h_2$  be non-negative real valued measurable functions of  $X$ . We define the common expectation of  $h_1$  and  $h_2$  with respect to  $\mu$  as

$$\psi_\mu(h_1, h_2) = E_\mu[\min\{h_1, h_2\}]. \tag{30}$$

The concept of common expectation is illustrated in Figure 1. More precisely, the value of the common expectation of  $h_1$  and  $h_2$  is the measure, according to  $\mu$ , of the function under which the shaded area is.



**Figure 1.** An illustration of the concept of common expectation of  $h_1$  and  $h_2$ .

**Example 5.** We want to obtain the global grade for two students out of the individual grades they obtained in four different courses  $\{x_1, x_2, x_3, x_4\}$ . In the final grade we want to reflect if a student shows a good performance in the two scientific courses,  $\{x_1, x_2\}$ , the humanistic ones,  $\{x_3, x_4\}$ , or in the combination

$\{x_2, x_3\}$ , corresponding to a social sciences profile. These criteria are encoded in the fuzzy measure in Table 2, while the grades obtained by both students (between 0 and 1) in each of the courses are shown in Table 3.

**Table 2.** A fuzzy measure matching the criteria in Example 5.

Reference Subsets	Measure
$\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}$	0.2
$\{x_1, x_2\}$	0.6
$\{x_1, x_3\}, \{x_1, x_4\}$	0.3
$\{x_2, x_3\}$	0.5
$\{x_2, x_4\}$	0.4
$\{x_3, x_4\}$	0.7
$\{x_1, x_2, x_3\}$	0.9
$\{x_1, x_2, x_4\}$	0.6
$\{x_1, x_3, x_4\}$	0.7
$\{x_2, x_3, x_4\}$	0.8

**Table 3.** Grades obtained by the students in Example 5 in the individual courses.

Student	$x_1$	$x_2$	$x_3$	$x_4$
$h_1$	0.9	0.8	0.3	0.2
$h_2$	0.2	0.3	0.8	0.9

The calculation of the respective monotone expectations and variances result in

$$E_\mu(h_1) = 0.61, \quad E_\mu(h_2) = 0.65,$$

$$Var_\mu(h_1) = 0.0769, \quad Var_\mu(h_2) = 0.0765,$$

which are quite similar, while the common expectation is  $\psi_\mu(h_1, h_2) = 0.25$ .

The next proposition states the basic properties of the common expectation.

**Proposition 5.** Let  $(X, \mathcal{A}, \mu)$  be a measurable space and let  $h_1$  and  $h_2$  be non-negative real valued measurable functions of  $X$ . Then,  $\psi_\mu$  satisfies the following properties:

- $\psi_\mu(h_1, h_2) = \psi_\mu(h_2, h_1)$ .
- $\psi_\mu(h_1, h_2) \leq \min\{E_\mu(h_1), E_\mu(h_2)\}$ .
- If  $\forall x \in X, h_1(x) \leq h_2(x)$ , then for any non-negative real valued measurable function  $h$  of  $X$ ,  $\psi_\mu(h_1, h) \leq \psi_\mu(h_2, h)$ .
- If  $\forall x \in X, h_1(x) \leq h_2(x)$ , then  $\psi_\mu(h_1, h_2) = E_\mu(h_1)$ .
- $\psi_\mu(h_1, h_2) = 0 \iff \{x \in X | h_1(x) > 0\} \cap \{x \in X | h_2(x) > 0\} = \emptyset$ .

**Proof.**

- It is straightforward from Equation (30).
- It follows from the facts that  $E_\mu$  is a monotone functional and that  $\min\{h_1, h_2\}$  is bounded from above by both  $h_1$  and  $h_2$ .
- It is a direct consequence of the monotonicity of operator  $\min$  and functional  $E_\mu$ .
- If  $h_1 \leq h_2$ , then  $\min\{h_1, h_2\} = h_1$ , and therefore both expectations are the same.
- If  $\{x \in X | h_1(x) > 0\} \cap \{x \in X | h_2(x) > 0\} = \emptyset$  then the minimum of both functions is the identically null function, which is known to be the only one that has null monotone expectation [30].

□

The common expectation is not normalized, and therefore its value alone is not enough to determine if it can be regarded as high or low. For instance, in Example 5 we obtained  $\psi_\mu(h_1, h_2) = 0.25$ , but that value does not tell us if it is high or low. However, it is clear that the common expectation can be bounded from above, since it is known that for any positive real numbers  $a$  and  $b$ , it holds that  $\min\{a, b\} \leq \sqrt{a \cdot b} \leq \max\{a, b\}$  and the equality is reached only when  $a = b$ . Hence, we can normalize the common expectation using these bounds, which yields three possible definitions of coefficients of concordance between  $h_1$  and  $h_2$ .

**Definition 19.** Let  $(X, \mathcal{A}, \mu)$  be a measurable space and let  $h_1$  and  $h_2$  be non-negative real valued measurable functions of  $X$ . We define the coefficients of concordance  $\rho_1, \rho_2$  and  $\rho_3$  between  $h_1$  and  $h_2$  with respect to  $\mu$  as

$$\rho_1^\mu(h_1, h_2) = \frac{\psi_\mu(h_1, h_2)}{\sqrt{E_\mu(h_1)E_\mu(h_2)}}, \tag{31}$$

$$\rho_2^\mu(h_1, h_2) = \frac{\psi_\mu(h_1, h_2)}{E_\mu(\max\{h_1, h_2\})}, \tag{32}$$

$$\rho_3^\mu(h_1, h_2) = \frac{\psi_\mu(h_1, h_2)}{\min\{E_\mu(h_1), E_\mu(h_2)\}}. \tag{33}$$

The next proposition shows the basic properties of the three concordance coefficients (when it is clear from the context, we will drop the measure and the functions, thus denoting  $\rho_i^\mu(h_1, h_2)$  by  $\rho_i$ ).

**Proposition 6.** Assume the conditions in Definition 19. The coefficients of concordance satisfy the following conditions:

1.  $0 \leq \rho_i \leq 1, i = 1, 2, 3$ .
2.  $h_1 = h_2 \Rightarrow \rho_1 = \rho_2 = \rho_3 = 1$ .
3.  $\rho_2 \leq \rho_1 \leq \rho_3$ .
4.  $\rho_1 = \rho_2 = \rho_3 = 0$  iff  $h_1$  and  $h_2$  have empty intersection, i.e.,  $\{x \in X/h_1(x) > 0\} \cap \{x \in X/h_2(x) > 0\} = \emptyset$ .
5. If  $h_1 \leq h_2$ , then

$$\begin{aligned} \rho_1 &= \sqrt{\frac{E_\mu(h_1)}{E_\mu(h_2)}}, \\ \rho_2 &= \frac{E_\mu(h_1)}{E_\mu(h_2)}, \\ \rho_3 &= 1. \end{aligned}$$

6. If  $h_1 = kh_2$ , with  $k > 1$ , then

$$\begin{aligned} \rho_1 &= \frac{1}{\sqrt{k}}, \\ \rho_2 &= \frac{1}{k}, \\ \rho_3 &= 1. \end{aligned}$$

**Proof.**

1. It is clear that

$$\min\{E_\mu(h_1), E_\mu(h_2)\} \leq \sqrt{E_\mu(h_1)E_\mu(h_2)} \leq \max\{E_\mu(h_1), E_\mu(h_2)\}.$$

Furthermore, since  $E_\mu$  is a monotone functional,  $\max\{E_\mu(h_1), E_\mu(h_2)\} \leq E_\mu(\max\{h_1, h_2\})$ . According to property 2 in Proposition 5,  $\psi_\mu(h_1, h_2) \leq \min\{E_\mu(h_1), E_\mu(h_2)\}$ , and thus

$$\psi_\mu(h_1, h_2) \leq \min\{E_\mu(h_1), E_\mu(h_2)\} \leq \sqrt{E_\mu(h_1)E_\mu(h_2)} \leq \max\{E_\mu(h_1), E_\mu(h_2)\}.$$

Therefore

$$\begin{aligned} \rho_1^\mu(h_1, h_2) &= \frac{\psi_\mu(h_1, h_2)}{\sqrt{E_\mu(h_1)E_\mu(h_2)}} \leq 1, \\ \rho_2^\mu(h_1, h_2) &= \frac{\psi_\mu(h_1, h_2)}{E_\mu(\max\{h_1, h_2\})} \leq 1, \\ \rho_3^\mu(h_1, h_2) &= \frac{\psi_\mu(h_1, h_2)}{\min\{E_\mu(h_1), E_\mu(h_2)\}} \leq 1. \end{aligned}$$

On the other hand, since  $h_1$  and  $h_2$  are non-negative, so it is  $\psi_\mu(h_1, h_2)$ , which means that  $\rho_1^\mu(h_1, h_2) \geq 0$ ,  $\rho_2^\mu(h_1, h_2) \geq 0$  and  $\rho_3^\mu(h_1, h_2) \geq 0$ .

2. If  $h_1 = h_2 = h$ , then  $\min\{h_1, h_2\} = \max\{h_1, h_2\} = h$  and  $\psi_\mu(h_1, h_2) = E_\mu(h)$ ; hence, the three coefficients are equal to 1.
3. From the proof of property 1, we know that

$$\begin{aligned} \min\{E_\mu(h_1), E_\mu(h_2)\} \leq \sqrt{E_\mu(h_1)E_\mu(h_2)} \leq E_\mu(\max\{h_1, h_2\}) &\Rightarrow \\ \frac{\psi_\mu(h_1, h_2)}{E_\mu(\max\{h_1, h_2\})} \leq \frac{\psi_\mu(h_1, h_2)}{\sqrt{E_\mu(h_1)E_\mu(h_2)}} \leq \frac{\psi_\mu(h_1, h_2)}{\min\{E_\mu(h_1), E_\mu(h_2)\}} &\Rightarrow \\ \rho_2 \leq \rho_1 \leq \rho_3. \end{aligned}$$

4.  $\rho_1 = \rho_2 = \rho_3 = 0$  iff  $\psi_\mu(h_1, h_2) = 0$  which, according to property 5 in Proposition 5, can only happen if  $\{x \in X/h_1(x) > 0\} \cap \{x \in X/h_2(x) > 0\} = \emptyset$ .
5. If  $h_1 \leq h_2$ , then  $\min\{h_1, h_2\} = h_1$ ,  $\max\{h_1, h_2\} = h_2$  and  $E_\mu(h_1) \leq E_\mu(h_2)$ . Therefore,

$$\rho_1 = \frac{E_\mu(h_1)}{\sqrt{E_\mu(h_1)E_\mu(h_2)}} = \sqrt{\frac{E_\mu(h_1)}{E_\mu(h_2)}}, \quad \rho_2 = \frac{E_\mu(h_1)}{E_\mu(h_2)} \quad \text{and} \quad \rho_3 = \frac{E_\mu(h_1)}{\min\{E_\mu(h_1), E_\mu(h_2)\}} = 1.$$

6. If  $h_1 = kh_2$  with  $k > 0$ , then  $\min\{h_1, h_2\} = h_2$ ,  $\max\{h_1, h_2\} = kh_2$  and  $E_\mu(h_1) = kE_\mu(h_2)$ . Therefore,

$$\rho_1 = \frac{E_\mu(h_2)}{\sqrt{kE_\mu(h_2)E_\mu(h_2)}} = \frac{1}{\sqrt{k}}, \quad \rho_2 = \frac{E_\mu(h_2)}{E_\mu(kh_2)} = \frac{1}{k} \quad \text{and} \quad \rho_3 = \frac{E_\mu(h_2)}{\min\{kE_\mu(h_2), E_\mu(h_2)\}} = 1.$$

□

**Example 6.** As a continuation of Example 5, we can use the data in Tables 2 and 3 to compute the coefficients of concordance, obtaining

$$\rho_1(h_1, h_2) = 0.397, \quad \rho_2(h_1, h_2) = 0.301, \quad \rho_3(h_1, h_2) = 0.410.$$

Note how the three coefficients have low values, which is consistent with the data in the example, as in spite of the similar values for the monotone expectation and variance corresponding to both students, they have a clearly different profile, scientific in the case of  $h_1$  and humanistic in the case of  $h_2$ .

#### 4. Parameters Defined over Product Spaces

In this section we explore scenarios where we have two measurable spaces each of them equipped with a different fuzzy measure. We will consider the definition of statistical parameters on the product space.

Likewise, in Section 3, we will separately study the case of one or two real functions. In both cases, it is necessary to obtain a fuzzy measure over the product space. We will rely on the proposals in [31] to obtain the product measures.

##### 4.1. The Case of One Function

The methods proposed in [31] for constructing fuzzy measures over product spaces, rather than single measures, usually yield a set of them, bounded by an upper and lower measure. Similarly, our proposals here will consist of intervals of parameters rather than a single one.

We will start defining the concept of joint expectation making use of the interior and exterior product measures (see Definitions 10 and 11).

**Definition 20.** Let  $(X_1, \mathcal{A}_{X_1}, \mu_1)$  and  $(X_2, \mathcal{A}_{X_2}, \mu_2)$  be measurable spaces,  $h : X_1 \times X_2 \rightarrow [0, 1]$  and  $\underline{\mu}_{12}^\odot, \overline{\mu}_{12}^\odot$  the  $\odot$ -interior and  $\odot$ -exterior product measures. We define the joint lower and upper  $\odot$ -expectations as

$$\underline{E}_{12}^\odot(h) = \oint h \circ \underline{\mu}_{12}^\odot \quad (\text{lower}), \tag{34}$$

$$\overline{E}_{12}^\odot(h) = \oint h \circ \overline{\mu}_{12}^\odot \quad (\text{upper}). \tag{35}$$

**Proposition 7.** Let  $(X_1, \mathcal{A}_{X_1}, \mu_1)$  and  $(X_2, \mathcal{A}_{X_2}, \mu_2)$  be measurable spaces,  $h : X_1 \times X_2 \rightarrow [0, 1]$  and  $\underline{\mu}_{12}^\odot, \overline{\mu}_{12}^\odot$  the  $\odot$ -interior and  $\odot$ -exterior product measures. It holds that

$$\underline{E}_{12}^\odot(h) \leq \overline{E}_{12}^\odot(h). \tag{36}$$

Furthermore, if  $\mu_{12}^\odot$  is any  $\odot$ -independent product measure of  $\mu_1$  and  $\mu_2$  (see Definition 9), it also holds that

$$\underline{E}_{12}^\odot(h) \leq E_{12}^\odot(h) \leq \overline{E}_{12}^\odot(h), \tag{37}$$

where  $E_{12}^\odot(h) = \oint h \circ \mu_{12}^\odot$ .

**Proof.** Note that  $\underline{E}_{12}^\odot, \overline{E}_{12}^\odot$  and  $E_{12}^\odot$  are monotone expectations, namely,  $E_{\underline{\mu}_{12}^\odot}, E_{\overline{\mu}_{12}^\odot}$  and  $E_{\mu_{12}^\odot}$  respectively. Therefore, Equations (36) and (37) are a direct consequence of the monotonicity of the monotone expectation and Proposition 1.  $\square$

The concept of joint  $\odot$ -expectations is analogous to the concept of monotone expectation in a marginal space, with the difference that, in the case of the product space, the underlying fuzzy measure is not known, but instead we have an interval of measures bounded by the interior and exterior  $\odot$ -product measures.

We can define joint expectations using other product measures, as the  $p$ -measures given in Definitions 12 and 13.

**Definition 21.** Let  $(X_1, \mathcal{A}_{X_1}, \mu_1)$  and  $(X_2, \mathcal{A}_{X_2}, \mu_2)$  be measurable spaces,  $h : X_1 \times X_2 \rightarrow [0, 1]$  and  $\underline{m}_{12}, \overline{m}_{12}$  the lower and upper product  $p$ -measures respectively. We define the lower and upper joint probabilistic expectations as

$$E_{\underline{m}_{12}}(h) = \oint h \circ \underline{m}_{12} \quad (\text{lower}), \tag{38}$$

$$E_{\overline{m}_{12}}(h) = \oint h \circ \overline{m}_{12} \quad (\text{upper}). \tag{39}$$

Since we have a function defined over the product space and fuzzy measures defined over the marginal spaces, it is natural to define marginal expectations. We will utilize the concept of  $\oplus$ -marginal of a function [31].

**Definition 22.** [31] Let  $h$  be a function defined on  $X_1 \times X_2$  and taking values on  $[0, 1]$ . We define the  $\oplus$ -marginals of  $h$  as

$$h_{X_1}^\oplus(x_{1i}) = \bigoplus_{x_{2j} \in X_2} h(x_{1i}, x_{2j}) = h(x_{1i}, x_{21}) \oplus h(x_{1i}, x_{22}) \oplus \dots \oplus h(x_{1i}, x_{2m}), \tag{40}$$

$$h_{X_2}^\oplus(x_{2j}) = \bigoplus_{x_{1i} \in X_1} h(x_{1i}, x_{2j}) = h(x_{11}, x_{2j}) \oplus h(x_{12}, x_{2j}) \oplus \dots \oplus h(x_{1n}, x_{2j}), \tag{41}$$

where  $\oplus$  is a  $t$ -conorm (see Definition 6),  $n$  is the cardinality of  $X_1$  and  $m$  is the cardinality of  $X_2$ .

**Definition 23.** Let  $(X_1, \mathcal{A}_{X_1}, \mu_1)$  and  $(X_2, \mathcal{A}_{X_2}, \mu_2)$  be measurable spaces and let  $h$  be a function defined on  $X_1 \times X_2$  and taking values on  $[0, 1]$ . We define the marginal  $\oplus$ -expectations as

$$E_{X_i}^\oplus(h) = \oint h_{X_i}^\oplus \circ \mu_i, \quad i = 1, 2, \tag{42}$$

where  $h_{X_i}^\oplus$  are the  $\oplus$ -marginals of  $h$ .

#### 4.2. The Case of Two Functions

We will now assume that we have two different functions, one for each marginal space, and define parameters that combine the information provided by the marginal spaces.

**Definition 24.** Let  $(X_1, \mathcal{A}_{X_1}, \mu_1)$  and  $(X_2, \mathcal{A}_{X_2}, \mu_2)$  be measurable spaces, and let  $h_1, h_2$  be functions defined on  $X_1$  and  $X_2$  respectively, taking values on  $[0, 1]$ . We define the upper and lower global expectation of  $h_1$  and  $h_2$  as

$$\underline{\phi}_\odot^*(h_1, h_2) = \oint h_{12}^* \circ \underline{\mu}_{12}^\odot \quad (\text{lower}), \tag{43}$$

$$\overline{\phi}_\odot^*(h_1, h_2) = \oint h_{12}^* \circ \overline{\mu}_{12}^\odot \quad (\text{upper}). \tag{44}$$

where  $\star$  and  $\odot$  are arbitrary  $t$ -norms (see Definition 5),  $h_{12}^*(x_1, x_2) = h_1(x_1) \star h_2(x_2), \forall (x_1, x_2) \in X_1 \times X_2$  and  $\underline{\mu}_{12}^\odot$  and  $\overline{\mu}_{12}^\odot$  are the interior and exterior product measures of  $\mu_1$  and  $\mu_2$  respectively.

The next proposition shows that both expectations coincide when  $\star$  is the min  $t$ -norm.

**Proposition 8.** Assume the conditions in Definition 24. If  $\star$  is the min  $t$ -norm, it holds that

$$\underline{\phi}_\odot^*(h_1, h_2) = \overline{\phi}_\odot^*(h_1, h_2). \tag{45}$$

**Proof.** According to ([31], Proposition 8), the  $\alpha$ -cuts generated by  $h^*$  belong to  $\mathcal{R}$  when  $\star$  is the min  $t$ -norm. Furthermore, Equation (12) establishes that  $\underline{\mu}_{12}^\odot = \overline{\mu}_{12}^\odot$  for the elements of  $\mathcal{R}$ , which proves the result.  $\square$

As a consequence of Proposition 8, when using the min  $t$ -norm we will just write  $\phi_\odot^*$  for both  $\underline{\phi}_\odot^*$  and  $\overline{\phi}_\odot^*$ .



The global expectation is in fact an extension of the monotone expectation in the sense expressed by the next theorem.

**Theorem 3.** Let  $(X, \mathcal{A}, \mu)$  be a measurable space and let  $h$  be a function defined on  $X$  and taking values on  $[0, 1]$ . Consider the product space  $X \times X$  and let both  $\star$  and  $\odot$  be the min t-norm. Then,

$$\phi_{\odot}^{\star}(h, h) = \phi_{\min}^{\min}(h, h) = E_{\mu}(h). \tag{46}$$

**Proof.** Assume  $X = \{x_1, x_2, \dots, x_n\}$ . Then

$$\forall(x_1, x_2) \in X \times X \Rightarrow h^{\min}(x, y) = \min\{h(x_1), h(x_2)\}.$$

Without loss of generality, we can assume that

$$h(x_1) < h(x_2) < \dots < h(x_n),$$

in which case the  $\alpha$ -cuts generated by  $h^{\min}$  are of the form

$$H_{\alpha_i} = \{(x_k, x_l) \in X \times X \mid k, l \geq i\},$$

which are elements of the class  $\mathcal{R}$  with their two projections being identical.

Since we are using the min t-norm to construct the product measure, it turns out that the measure in each  $\alpha$ -cut of the product space is equal to the measure assigned by  $\mu$  to the  $\alpha$ -cuts in the marginal space, and thus

$$\begin{aligned} \phi_{\min}^{\min}(h, h) &= \oint h^{\min} \circ \mu_{12}^{\min} = \sum_{i=1}^{n \times n} \mu_{12}^{\min}(H_{\alpha_i})(\alpha_i - \alpha_{i-1}) \\ &= \sum_{i=1}^n [\mu(H_{\alpha_i}^{\downarrow X})] (\alpha_i - \alpha_{i-1}) = E_{\mu}(h). \end{aligned}$$

□

**Example 7.** Consider a reference set  $X = \{x_1, x_2, x_3\}$  and the function defined as  $h(x_1) = 0.1, h(x_2) = 0.4, h(x_3) = 0.7$ . The function  $h^{\min}$  is displayed in Table 4.

**Table 4.** Values of the function  $h^{\min}$ .

	$x_1$	$x_2$	$x_3$
$x_1$	0.1	0.1	0.1
$x_2$	0.1	0.4	0.4
$x_3$	0.1	0.4	0.7

It can be seen how the diagonal contains the original values of  $h$ , and its  $\alpha$ -cuts are

$$\begin{aligned} H_{0.1} &= X \times X, & \mu_{12}^{\min}(H_{0.1}) &= \min\{\mu(X), \mu(X)\}, \\ H_{0.4} &= \{x_2, x_3\} \times \{x_2, x_3\}, & \mu_{12}^{\min}(H_{0.4}) &= \min\{\mu(\{x_2, x_3\}), \mu(\{x_2, x_3\})\}, \\ H_{0.7} &= \{x_3\} \times \{x_3\}, & \mu_{12}^{\min}(H_{0.7}) &= \min\{\mu(\{x_3\}), \mu(\{x_3\})\}, \end{aligned}$$

and therefore

$$\begin{aligned} \phi_{\min}^{\min}(h, h) &= \mu_{12}^{\min}(H_{0.1})(0.1 - 0) + \mu_{12}^{\min}(H_{0.4})(0.4 - 0.1) + \mu_{12}^{\min}(H_{0.7})(0.7 - 0.4) \\ &= \mu(X)(0.1 - 0) + \mu(\{x_2, x_3\})(0.4 - 0.1) + \mu(\{x_3\})(0.7 - 0.4) \\ &= E_{\mu}(h). \end{aligned}$$

Likewise for the common expectation, the global expectation is not normalized, but it can be easily normalized in the same way as we did for the common expectation case, as stated in the next definition.

**Definition 25.** Let  $(X_1, \mathcal{A}_{X_1}, \mu_1)$  and  $(X_2, \mathcal{A}_{X_2}, \mu_2)$  be measurable spaces and let  $h_1$  and  $h_2$  be functions defined on  $X_1$  and  $X_2$  respectively and taking values on  $[0, 1]$ . We define the global coefficients of concordance of  $h_1$  and  $h_2$  as

$$\Phi_1(h_1, h_2) = \frac{\phi_{\min}^{\min}(h_1, h_2)}{\min\{E_{\mu_1}(h_1), E_{\mu_2}(h_2)\}}, \tag{47}$$

$$\Phi_2(h_1, h_2) = \frac{\phi_{\min}^{\min}(h_1, h_2)}{\sqrt{E_{\mu_1}(h_1)E_{\mu_2}(h_2)}}. \tag{48}$$

**Example 8.** (Continuation of Example 5)

Using the data in Table 3 we can obtain the function  $h_{12}^{\min}$ , the values of which are given in Table 5.

**Table 5.** Values of the function  $h_{12}^{\min}$ .

Course	$x_1$	$x_2$	$x_3$	$x_4$
$x_1$	0.2	0.3	0.8	0.9
$x_2$	0.2	0.3	0.8	0.8
$x_3$	0.2	0.3	0.3	0.3
$x_4$	0.2	0.2	0.2	0.2

Using the fuzzy measure in Table 2, we find that  $\phi_{\min}^{\min} = 0.6$  and the global concordance coefficients are  $\Phi_1(h_1, h_2) = 0.953$  and  $\Phi_2(h_1, h_2) = 0.984$ .

The value of the global expectation (0.6) is very close to the values of the monotone expectations for each student in Example 5 (0.61 and 0.65 respectively). It can be interpreted as the fact that the grades of both students are acceptable individually and also globally, which is reflected in high values of the global coefficients of concordance. Note how the global expectation is not detecting the fact that both students have different profiles (scientific and humanistic), while the common expectation detected this fact yielding a much lower value (0.25) resulting in lower values of the coefficients of concordance as well.

### 5. Conclusions

With the introduction of the concept of monotone variance, we have complemented the already known concept of monotone expectation. It can be regarded as a measure of dispersion with respect to a central position measure. We have also introduced the concepts of central and non-central monotone moments, that can serve as a vehicle to define further statistical parameters based on fuzzy measures as, for instance, shape measures. The potential application scope is certainly wide, as it covers non-additive scenarios like the ones described in the examples in this paper, and just to mention some of them, such scenarios can be found in Engineering and Social Sciences applications.

The common expectation and concordance coefficients can be interpreted as measures of match between the functions, and in that sense can provide information about to which extent one function explains the other one. A possible application of these concepts is the development of prediction models when the measures are not additive.

Thanks to the developments in [31] we have been able to extend the concept of monotone expectation to product spaces, where, in addition, we have shown how to marginalize the information provided by a function over a product space using the marginal  $\oplus$ -expectations.

All the developments in this paper are restricted to finite reference sets. Even though it covers a wide variety of practical applications, it is worth exploring the formulation of the results obtained here to uncountable reference sets, which seems to be a promising research line. The first step in this direction would be the extension of the results in [31] to continuous domains.

**Author Contributions:** Investigation, F.R., M.M. and A.S.; writing—original draft, F.R., M.M. and A.S.; writing—review and editing, F.R., M.M. and A.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Spanish Ministry of Science and Innovation through grants TIN2016-77902-C3-3-P, PID2019-106758GB-C32 and by ERDF-FEDER funds.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Sugeno, M. Theory of Fuzzy Integrals and Its Applications. Ph.D. Thesis, Tokyo Institute of Technology, Tokyo, Japan, 1974.
2. Choquet, G. Theory of capacities. *Annales de l'Institut Fourier* **1954**, *5*, 131–295. [[CrossRef](#)]
3. Li, J. On null-continuity of monotone measures. *Mathematics* **2020**, *8*, 205. [[CrossRef](#)]
4. Beliakov, G.; James, S.; Wu, J. *Discrete Fuzzy Measures; Studies in Fuzziness and Soft Computing*; Springer: Berlin, Germany, 2020; Volume 382.
5. Walley, P. *Statistical Reasoning with Imprecise Probabilities*; Chapman and Hall: London, UK, 1991.
6. Walley, P. *BI Statistical Methods. Volume 1: Foundations*; Prescience Press: New York, NY, USA, 2015.
7. Dempster, A.P. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Stat.* **1967**, *38*, 325–339. [[CrossRef](#)]
8. Shafer, G. *A Mathematical Theory of Evidence*; Princeton University Press: Princeton, NJ, USA, 1976.
9. Buckley, J. *Fuzzy Probability and Statistics; Studies in Fuzziness and Soft Computing*; Springer: Berlin, Germany, 2006; Volume 196.
10. Coppi, R.; Gil, M.A.; Kiers, H.A.L. The fuzzy approach to statistical analysis. *Comput. Stat. Data Anal.* **2006**, *51*, 1–14. [[CrossRef](#)]
11. D'Urso, P.; Gil, M.A. Fuzzy data analysis and classification. *Adv. Data Anal. Classif.* **2017**, *11*, 645–657. [[CrossRef](#)]
12. Intan, R.; Mukaidono, M. Fuzzy conditional probability relations and their applications in fuzzy information systems. *Knowl. Inf. Syst.* **2004**, *6*, 345–365. [[CrossRef](#)]
13. Nguyen, H.; Wu, B. *Fundamentals of Statistics with Fuzzy Data; Studies in Fuzziness and Soft Computing*; Springer: Berlin, Germany, 2006; Volume 198.
14. Viertl, R. *Statistical Methods for Fuzzy Data*; Wiley: Chichester, UK, 2011.
15. Kruse, R.; Held, P.; Moewes, C. On fuzzy data analysis. *On Fuzziness—A Homage to Lotfi A. Zadeh, Volume 1; Studies in Fuzziness and Soft Computing*; Springer: Heidelberg, Germany, 2013; Volume 298, pp. 343–347.
16. Dijkman, J.G.; van Haeringen, H.; Lange, S.J. Fuzzy numbers. *J. Math. Anal. Appl.* **1983**, *92*, 301–341. [[CrossRef](#)]
17. D'Urso, P. Informational paradigm, management of uncertainty and theoretical formalisms in the clustering framework: A review. *Inf. Sci.* **2017**, *400–401*, 30–62. [[CrossRef](#)]
18. Tanaka, H.; Uejima, S.; Asai, K. Linear regression analysis with fuzzy model. *IEEE Trans. Syst. Man Cybern.* **1982**, *12*, 903–907.
19. Parchami, A.; Taheri, S.M.; Mashinchi, M. Fuzzy  $p$ -value in testing fuzzy hypotheses with crisp data. *Stat. Pap.* **2010**, *51*, 209. [[CrossRef](#)]
20. Grzegorzewski, P.; Hryniewicz, O. Soft methods in statistical quality control. *Control Cybern.* **2000**, *29*, 119–140.
21. Zhang, R.; Ashuri, B.; Deng, Y. A novel method for forecasting time series based on fuzzy logic and visibility graph. *Adv. Data Anal. Classif.* **2017**, *11*, 759–783. [[CrossRef](#)]

22. Gil, M.A.; López-Díaz, M. Fundamentals and Bayesian analyses of decision problems with fuzzy-valued utilities. *Int. J. Approx. Reason.* **1996**, *15*, 95–115. [[CrossRef](#)]
23. Denoeux, T. Maximum likelihood estimation from fuzzy data using the EM algorithm. *Fuzzy Sets Syst.* **2011**, *183*, 72–91. [[CrossRef](#)]
24. Quost, B.; Denoeux, T.; Li, S. Parametric classification with soft labels using the evidential EM algorithm: Linear discriminant analysis versus logistic regression. *Adv. Data Anal. Classif.* **2017**, *11*, 659–690. [[CrossRef](#)]
25. Blanco-Fernández, A.; Casals, M.R.; Colubi, A.; Corral, N.; García-Bárcana, M.; Gil, M.A.; González-Rodríguez, G.; López, M.T.; Lubiano, M.A.; Montenegro, M.; et al. A distance-based statistical analysis of fuzzy number-valued data. *Int. J. Approx. Reason.* **2014**, *55*, 1487–1501. [[CrossRef](#)]
26. Wu, H.C. Statistical hypotheses testing for fuzzy data. *Inf. Sci.* **2005**, *279*, 446–459. [[CrossRef](#)]
27. Calcagni, A.; Lombardi, L.; Pascali, E. A dimension reduction technique for two-mode non-convex fuzzy data. *Soft Comput.* **2016**, *20*, 749–762. [[CrossRef](#)]
28. Colubi, A.; González-Rodríguez, G.; Gil, M.A.; Trutschnig, W. Nonparametric criteria for supervised classification of fuzzy data. *Int. J. Approx. Reason.* **2011**, *52*, 1272–1282. [[CrossRef](#)]
29. Coppi, R.; D’Urso, P.; Giordani, P. Fuzzy and possibilistic clustering for fuzzy data. *Comput. Stat. Data Anal.* **2012**, *56*, 915–927. [[CrossRef](#)]
30. Bolaños, M.J.; De Campos, L.M.; González, A. Convergence properties on monotone expectation and its applications to the extension of fuzzy measures. *Fuzzy Sets Syst.* **1989**, *33*, 201–212. [[CrossRef](#)]
31. Reche, F.; Morales, M.; Salmerón, A. Construction of fuzzy measures over product spaces. *Mathematics* **2020**, *8*, 1605. [[CrossRef](#)]
32. Reche, F.; Salmerón, A. Operational approach to general fuzzy measures. *Int. J. Uncertain. Fuzziness-Knowl.-Based Syst.* **2000**, *8*, 369–382. [[CrossRef](#)]
33. De Campos, L.M.; Bolaños, M.J. Representation of fuzzy measures through probabilities. *Fuzzy Sets Syst.* **1989**, *31*, 23–36. [[CrossRef](#)]
34. Schweizer, B.; Sklar, A. *Probability Metric Spaces*; Elsevier North Holland: New York, NY, USA, 1983.
35. Huber, P.H. *Robust Statistics*; Wiley Series in Probability and Mathematical Statistics; John Wiley and Sons: Hoboken, NJ, USA, 1981.

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).